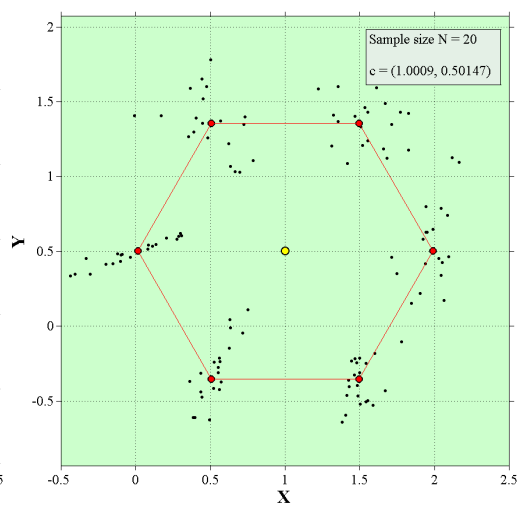
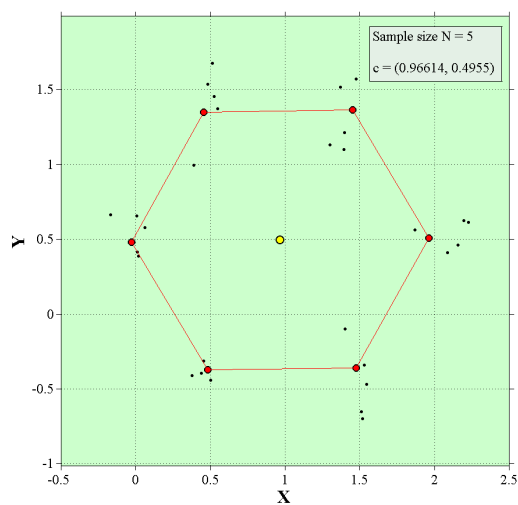
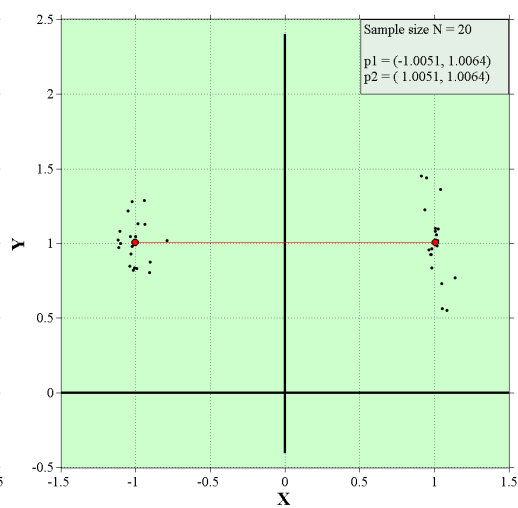
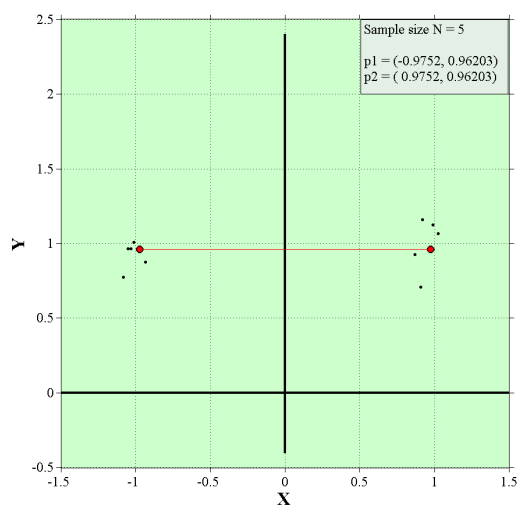




Машинное обучение и анализ данных

Journal of Machine Learning and Data Analysis

Журнал «Машинное обучение и анализ данных» публикует новые теоретические и обзорные статьи с результатами научных исследований в области теоретических основ информатики и её приложений. Цель журнала — развитие теории машинного обучения, интеллектуального анализа данных и методов проведения вычислительных экспериментов. Принимаются статьи на английском и русском языках.

Журнал включен в российский индекс научного цитирования РИНЦ. Информация о цитировании статей находится на сайте Российского индекса научного цитирования. ISSN 2223-3792, номер свидетельства о регистрации ЭЛ № ФС 77-55486.

- Архив журнала <http://www.ccas.ru/jmlda/>
- Новостной сайт <http://jmlda.org/>
- Электронная система подачи статей <http://jmlda.org/papers/>

Тематика журнала:

- классификация, кластеризация, регрессионный анализ,
- алгебраический подход к проблеме синтеза корректных алгоритмов,
- многомерный статистический анализ,
- выбор моделей и сложность,
- предсказательное моделирование,
- статистическая теория обучения,
- методы прогнозирования временных рядов,
- методы обработки и распознавания сигналов,
- методы оптимизации в задачах машинного обучения и анализа данных,
- методы визуализации данных,
- обработка и распознавание речи и изображений,
- анализ и понимание текста,
- информационный поиск,
- прикладные задачи анализа данных.

Редакционный совет:

Ю.Г. Евтушенко, акад.,
Ю.И. Журавлёв, акад.,
В.Л. Матросов, акад.,
К.В. Рудаков, чл. корр.

Редколлегия:

К. В. Воронцов, д.ф.-м.н.,
А. Г. Дьяконов, д.ф.-м.н.,
Л. М. Местецкий, д.т.н.,
В. В. Моттль, д.т.н.,
М. Ю. Хачай, д.ф.-м.н.

Координаторы:

М. П. Кузнецов,
А. П. Мотренко.

Редактор: В. В. Стрижов, к.ф.-м.н. (strijov@ccas.ru)

Вычислительный центр Российской академии наук
Московский физико-технический институт
Факультет управления и прикладной математики
Кафедра «Интеллектуальные системы»

Москва, 2014

Содержание

Е. А. Рябенко

Мультипликативный метод неотрицательного матричного разложения с АБ-дивергенцией и его сходимость 800

О. А. Кушнир, О. С. Середин, А. В. Степанов

Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений 817

И. А. Матвеев, А. Н. Трекин

Метод обнаружения автомобилей на аэрокосмических снимках 828

А. О. Ларин, О. С. Середин, В. В. Кондрашов

Комбинирование одноклассовых классификаторов в задачах сегментации изображений, содержащих объекты топологии плат микросборок 835

Е. В. Дюкова, А. Г. Никифоров, П. А. Прокофьев

Статистически эффективная схема распараллеливания алгоритмов дуализации . 843

А. С. Пушняков

О комбинаторных оценках максимальных ϵ -разбиений метрических конфигураций 854

А. И. Панов

Алгебраические свойства операторов распознавания в моделях зрительного восприятия 863

Н. В. Дербенев, Д. А. Козлюк, В. В. Никитин, В. О. Толчеев

Экспериментальное исследование методов выявления нечетких дубликатов научных публикаций 875

С. Д. Двоенко, Д. О. Пшеничный

Оптимальная коррекция метрических нарушений в матрицах парных сравнений . 885

Н. Г. Загоруйко, О. А. Кутненко, А. О. Зырянов, Д. А. Леванов

Обучение распознаванию без переобучения 891

Ф. А. Корнилов

Поиск структурных различий изображений: алгоритмы и методы исследования . 902

А. Н. Каркищенко, В. Б. Мнухин

Симметризация точек изображения, заданных статистическими выборками 920

Мультипликативный метод неотрицательного матричного разложения с АБ-дивергенцией и его сходимость*

Е. А. Рябенко

riabenko.e@gmail.com

Москва, Вычислительный центр им. А. А. Дородницына РАН

Мультипликативный метод неотрицательного матричного разложения для случая, когда точность приближения модели измеряется с помощью АБ-дивергенции, вблизи границы неотрицательной области может сходиться к нестационарной точке. Предлагается модифицированный мультипликативный метод, в котором за счет отделения элементов матриц от нуля константой ϵ удается показать не только монотонность невозрастания функции потерь, но и тот факт, что любая предельная точка этого метода является стационарной точкой отделенной от нуля задачи. Разреживание получаемых таким методом матриц дает решение, являющееся стационарной точкой исходной задачи с точностью до $\mathcal{O}(\epsilon)$. Для частного случая, соответствующего норме Фробениуса, показано, что метод всегда сходится.

Ключевые слова: *неотрицательное матричное разложение, мультипликативные обновления, АБ-дивергенция, сходимость.*

Multiplicative Method for Nonnegative Matrix Factorization with AB-Divergence and its Convergence*

E. A. Riabenko

Moscow, Dorodnicyn Computing Centre of RAS

Multiplicative method for nonnegative matrix factorization with AB-divergence could converge to nonstationary point when the elements of matrices approach zero. A modified multiplicative method that bounds the matrices from zero by constant ϵ is proposed, and there proved not only monotonic descent, but the fact that its every limiting point is a stationary point of the modified bounded problem. Setting particular elements of the resulting matrix to zero yields the solution that is $\mathcal{O}(\epsilon)$ -close to the stationary point of the original problem. For a special case of Frobenius norm, it is proved that the method always converges.

Keywords: *nonnegative matrix factorization, multiplicative updates, AB-divergence, convergence.*

Введение

Матричные разложения используются для решения задач сжатия данных, восстановления сигналов, заполнения пропусков, для выявления структурных особенностей коллекций данных. Приложения, связанные с получением и анализом матричных разложений, различаются ограничениями, накладываемыми на факторы. В задаче неотрицательного матричного разложения, рассматриваемой в данной работе, ключевую роль играют ограничения на знак элементов факторных матриц. Впервые она была рассмотрена в ра-

*Работа выполнена при финансовой поддержке РФФИ (проект № 11-07-00480) и программы ОМН РАН «Алгебраические и комбинаторные методы математической кибернетики и информационные системы нового поколения».

боте [1] в приложении к задаче византийских генералов из теории отказоустойчивости, однако основной интерес к этой теме возник после работ [2, 3], авторы которых обобщили постановку задачи и предложили простой алгоритм ее решения. Неотрицательные матричные разложения используются при анализе изображений, текстов, в вычислительной биологии, медицине и других прикладных областях. Обзор применений можно найти в [4].

Задача формулируется следующим образом. Дана матрица $P \in \mathbb{R}_+^{m \times n}$ с неотрицательными элементами (подразумевается, что в ней отсутствуют нулевые строки и столбцы), а также некоторое натуральное число $r < \min(m, n)$. Требуется найти матрицы $A^* \in \mathbb{R}_+^{m \times r}$ и $X^* \in \mathbb{R}_+^{r \times n}$ с неотрицательными элементами, такие, что их произведение $Q^* = A^* X^*$ максимально близко к исходной матрице P в смысле некоторой функции потерь $D(P, Q)$:

$$(A^*, X^*) = \arg \min_{A \geq 0, X \geq 0} D(P, AX). \quad (1)$$

Выбор функции потерь оказывает существенное влияние как на саму задачу, так и на получаемое решение [5].

Наибольший интерес представляют аддитивные функции потерь:

$$D(P, Q) = \sum_{i,j} d(p_{ij}, q_{ij}),$$

где $d(p, q) \geq 0$, причем $d(p, q) = 0$ тогда и только тогда, когда $p = q$.

Чаще всего в качестве функции потерь используется норма Фробениуса:

$$D_F(P, Q) = \sum_{i,j} (p_{ij} - q_{ij})^2.$$

Одна из причин ее популярности является оптимальность получаемых в результате ее минимизации оценок для моделей с аддитивным гауссовским шумом — такие оценки совпадают с оценками максимального правдоподобия. Однако для других видов шума, а также в присутствии выбросов, оценки, доставляющие минимум норме Фробениуса, могут оказываться несостоятельными. В случае аддитивного пуассоновского шума задача максимизации правдоподобия эквивалентна задаче минимизации дивергенции Кульбака–Лейблера между данными и моделью [6]:

$$D_{KL}(P, Q) = \sum_{i,j} \left(p_{ij} \ln \frac{p_{ij}}{q_{ij}} - p_{ij} + q_{ij} \right).$$

Кроме того, если столбцы матриц P , A и X нормированы и рассматриваются как вероятностные распределения, минимизация дивергенции Кульбака–Лейблера также соответствует максимизации правдоподобия модели. Первые алгоритмы неотрицательного матричного разложения, предложенные в работах [2, 3], были построены именно для этих мер качества.

В данной работе мы рассмотрим более широкий класс функций потерь — АБ-дивергенции, параметрическое семейство, содержащее как приведенные функции, так и многие другие [7].

Для решения оптимизационной задачи (1) наиболее популярными являются итерационные методы с мультипликативным шагом, в которых обновления переменных A и X имеют вид умножения на положительное число. Преимущество мультипликативных обновлений заключается в том, что неотрицательность решения сохраняется естественным образом без дополнительных вычислительных затрат.

В данной работе предлагается мультипликативный метод получения неотрицательного матричного разложения с АБ-дивергенцией в качестве функции потерь и доказывается несколько фактов о его сходимости, в частности, монотонность невозрастания ошибки и близость получаемого решения к стационарной точке задачи.

АБ-дивергенция

Для измерения качества моделей в процессе настройки ее параметров существуют разнообразные функции потерь, как возникшие из контекста прикладных задач, так и мотивированные соображениями теории информации, выпуклого анализа или информационной геометрии. Представление различных функций потерь в виде параметрических семейств позволяет унифицировать работу с ними, увидеть между ними связь и лучше понять, как выбор меры качества влияет на получаемое решение. Среди таких семейств широко используются альфа- [8] и бета-дивергенции [9], задающие непрерывные по параметрам множества сепарабельных функций потерь, включающих в том числе норму Фробениуса, дивергенции Кульбака-Лейблера, Итакура-Саито и другие. Класс АБ-дивергенций, предложенный в работе [7], обобщает эти семейства и задается в виде двухпараметрического семейства функций следующего вида:

$$D_{AB}^{(\alpha, \beta)}(P, Q) = \sum_{i,j} d_{AB}^{(\alpha, \beta)}(p_{ij}, q_{ij}),$$

$$d_{AB}^{(\alpha, \beta)}(p, q) = \begin{cases} \frac{1}{\alpha\beta} \left(\frac{\alpha}{\alpha+\beta} p^{\alpha+\beta} + \frac{\beta}{\alpha+\beta} q^{\alpha+\beta} - p^{\alpha} q^{\beta} \right), & \text{если } \alpha, \beta, \alpha + \beta \neq 0; \\ \frac{1}{\alpha^2} \left(p^{\alpha} \ln \frac{p^{\alpha}}{q^{\alpha}} - p^{\alpha} + q^{\alpha} \right), & \text{если } \alpha \neq 0, \beta = 0; \\ \frac{1}{\alpha^2} \left(\ln \frac{q^{\alpha}}{p^{\alpha}} + \left(\frac{q^{\alpha}}{p^{\alpha}} \right)^{-1} - 1 \right), & \text{если } \alpha = -\beta \neq 0; \\ \frac{1}{\beta^2} \left(q^{\beta} \ln \frac{q^{\beta}}{p^{\beta}} - q^{\beta} + p^{\beta} \right), & \text{если } \alpha = 0, \beta \neq 0; \\ \frac{1}{2} (\ln p - \ln q)^2, & \text{если } \alpha = \beta = 0. \end{cases}$$

Частные случаи (2)–(5) здесь являются предельными для случая (1) и могут быть получены по правилу Лопиталя.

АБ-дивергенции — одно из наиболее обширных известных параметрических семейств функций потерь. Оно включает многие широко применяемые меры близости (табл. 1). Оценки, получаемые при минимизации функций этого класса, являются оценками максимального правдоподобия при различных распределениях шума как аддитивного и мультипликативного, так и смешанного, состоящего из обеих этих компонент. Выбирая параметры α и β , мы тем самым определяем такие свойства получаемой модели, как разреженность и устойчивость, а также неявно задаем вид шума, подходящий для исследуемой задачи. Пример такого использования АБ-дивергенций можно найти в работе [10].

Будем решать задачу неотрицательного матричного разложения, используя АБ-дивергенцию в качестве функции потерь:

$$(A^*, X^*) = \arg \min_{A \geq 0, X \geq 0} D_{AB}^{(\alpha, \beta)}(P, AX). \quad (2)$$

Мультипликативные алгоритмы минимизации функции потерь

Поскольку используемые на практике функции потерь не выпуклы по совокупности аргументов A, X , задача (1) решается поочередной минимизацией $D(P, AX)$ по A и X с помощью алгоритмов следующего вида:

Таблица 1: Некоторые функции из семейства АБ-дивергенций

(α, β)	Функция потерь $d_{AB}^{(\alpha, \beta)}(p, q)$	Название
$(1, 0)$	$d_{KL}(p, q) = p \ln \frac{p}{q} - p + q$	Дивергенция Кульбака-Лейблера
$(1, -1)$	$d_{IS}(p, q) = \ln \frac{q}{p} + \frac{p}{q} - 1$	Дивергенция Итакура-Саито
$(1, 1)$	$\frac{1}{2}d_E(p, q) = \frac{1}{2}(p - q)^2$	Евклидово расстояние
$(0, 5, 0, 5)$	$2d_H(p, q) = 2(\sqrt{p} - \sqrt{q})^2$	Расстояние Хелингера
$(2, -1)$	$\frac{1}{2}d_P(p, q) = \frac{1}{2} \frac{(p - q)^2}{q}$	χ^2 Пирсона
$(-1, 2)$	$\frac{1}{2}d_N(p, q) = \frac{1}{2}d_H(q, p) = \frac{1}{2} \frac{(p - q)^2}{p}$	χ^2 Неймана
$(0, 0)$	$\frac{1}{2}d_E(\ln p, \ln q) = \frac{1}{2}(\ln(p) - \ln(q))^2$	Лог-евклидово расстояние

Вход: $A^0 \geq 0, X^0 \geq 0$

цикл // $t = 0, 1, 2, \dots$

зафиксировав A^t , найдем такое X^{t+1} , что $D(P, A^t X^{t+1}) \leq D(P, A^t X^t)$;

зафиксировав X^{t+1} , найдем такое A^{t+1} , что $D(P, A^{t+1} X^{t+1}) \leq D(P, A^t X^{t+1})$;

Существуют различные подходы к построению таких обновлений. Можно на каждом шаге находить значение матрицы-множителя, минимизирующее $D(P, AX)$, когда вторая матрица-множитель фиксирована. Алгоритм с такими обновлениями будет реализовывать метод блочно-покоординатного спуска. В случае двух блоков любая предельная точка такого алгоритма является стационарной [11]. Для некоторых функций потерь точные минимумы по каждой из матриц найти легко: например, когда точность приближения измеряется при помощи нормы Фробениуса, покомпонентные минимумы находятся аналитически (это свойство лежит в основе метода HALS [12]). При использовании других функций потерь, в том числе рассматриваемых здесь АБ-дивергенций, покомпонентные минимумы можно найти только численно и приближенно. Это может потребовать значительных вычислительных затрат, притом что предельная точка такого алгоритма может и не являться стационарной. Вместо поиска точного покомпонентного минимума функции $D(P, AX)$ достаточно на каждой итерации делать только один шаг в направлении ее уменьшения. Для этого может использоваться обычный градиентный метод:

Вход: $A^0 \geq 0, X^0 \geq 0$

цикл // $t = 0, 1, 2, \dots$

$$x_{kj}^{t+1} = x_{kj}^t - \nu \frac{\partial D(P, AX)}{\partial x_{kj}}, \quad k = 1, \dots, r, \quad j = 1, \dots, n;$$

$$a_{ik}^{t+1} = a_{ik}^t - \eta \frac{\partial D(P, AX)}{\partial a_{ik}}, \quad i = 1, \dots, m, \quad k = 1, \dots, r;$$

где ν, η — положительные константы, задающие длины шага в направлении градиентного спуска. В общем случае длина шага в направлении спуска может выбираться индивидуаль-

но для каждого элемента матриц, а также меняться в зависимости от номера итерации t :

$$\begin{aligned}x_{kj}^{t+1} &= x_{kj}^t - \nu_{kj}^t \frac{\partial D(P, AX)}{\partial x_{kj}}; \\a_{ik}^{t+1} &= a_{ik}^t - \eta_{ik}^t \frac{\partial D(P, AX)}{\partial a_{ik}}.\end{aligned}$$

Кроме того, градиентный спуск может вестись в трансформированном пространстве параметров:

$$\begin{aligned}x_{kj}^{t+1} &= \varphi^{-1} \left(\varphi(x_{kj}^t) - \nu_{kj}^t \frac{\partial D(P, AX)}{\partial \varphi(x_{kj})} \right); \\a_{ik}^{t+1} &= \varphi^{-1} \left(\varphi(a_{ik}^t) - \eta_{ik}^t \frac{\partial D(P, AX)}{\partial \varphi(a_{ik})} \right),\end{aligned}$$

где $\varphi(z)$ — некоторое биективное отображение. Переход в трансформированное пространство позволяет увеличить скорость сходимости, когда покомпонентные гессианы минимизируемого функционала в нем имеют меньшее число обусловленности (известно, что скорость сходимости градиентных методов ограничивается разностью минимального и максимального собственных значений гессиана [13]).

Поскольку градиентный спуск не гарантирует сохранения неотрицательности компонент, на каждом шаге необходимо проецировать обновленные значения A и X на неотрицательную область. В таком случае для того, чтобы обеспечить монотонное убывание функции потерь, необходимо дополнительно оптимизировать параметры ν_{kj}^t, η_{ik}^t вдоль направления спуска, что может потребовать значительных вычислительных затрат. Вместо этого можно выбрать ν_{kj}^t, η_{ik}^t так, чтобы обновления были мультипликативными, т. е. чтобы x_{kj}, a_{ik} на каждом шаге умножались на некоторое положительное число. В этом случае неотрицательность решения будет сохраняться без дополнительных операций. Наиболее распространенные алгоритмы получения неотрицательного матричного разложения основаны именно на этом подходе.

Простейший способ выбора ν_{kj}^t, η_{ik}^t , приводящий к мультипликативным обновлениям в исходном пространстве параметров, заключается в следующем. Пусть ∇_A — градиент $D(P, AX)$ по A , $\nabla_A^+ = \max(\nabla_A, 0)$, $\nabla_A^- = \max(-\nabla_A, 0)$, где максимумы берутся покомпонентно, т. е. $\nabla_A = \partial D / \partial A = \nabla_A^+ - \nabla_A^-$. Выбирая $\eta_{ik}^t = a_{ik}^t / [\nabla_A^+]_{ik}$, получим мультипликативные обновления вида

$$a_{ik}^{t+1} = a_{ik}^t \frac{[\nabla_A^-]_{ik}}{[\nabla_A^+]_{ik}}.$$

Аналогично для X , взяв $\nu_{kj}^t = x_{kj}^t / [\nabla_X^+]_{kj}$, получаем:

$$x_{kj}^{t+1} = x_{kj}^t \frac{[\nabla_X^-]_{kj}}{[\nabla_X^+]_{kj}}.$$

Алгоритм, решающий задачу (2) путем поочередных мультипликативных шагов по A и X , был предложен в работе [7]. Градиентный спуск в нем ведется в пространстве, трансформированном функцией деформированного логарифма:

$$\varphi(z) = \ln_{1-\alpha}(z) = \begin{cases} \frac{z^\alpha - 1}{\alpha}, & \text{если } \alpha \neq 0; \\ \ln z, & \text{если } \alpha = 0, \end{cases}$$

обратной функцией к которой является деформированная экспонента:

$$\varphi^{-1}(z) = \exp_{1-\alpha}(z) = \begin{cases} \exp(z), & \text{если } \alpha = 0; \\ (1 + \alpha z)^{1/\alpha}, & \text{если } \alpha \neq 0, 1 + \alpha z \geq 0; \\ 0, & \text{если } \alpha \neq 0, 1 + \alpha z < 0. \end{cases}$$

Алгоритм выглядит следующим образом:

$$\begin{aligned} x_{kj} &\leftarrow x_{kj} \left(\exp_{1-\alpha} \left(\frac{\sum_{i=1}^m a_{ik} q_{ij}^{\alpha+\beta-1} \ln_{1-\alpha}(p_{it}/q_{it})}{\sum_{i=1}^m a_{ik} q_{ij}^{\alpha+\beta-1}} \right) \right)^{\omega(\alpha, \beta)} ; \\ a_{ik} &\leftarrow a_{ik} \left(\exp_{1-\alpha} \left(\frac{\sum_{j=1}^n x_{kj} q_{ij}^{\alpha+\beta-1} \ln_{1-\alpha}(p_{it}/q_{it})}{\sum_{j=1}^n x_{kj} q_{ij}^{\alpha+\beta-1}} \right) \right)^{\omega(\alpha, \beta)}, \end{aligned} \quad (3)$$

где

$$\omega(\alpha, \beta) = \begin{cases} 1, & \text{если } \alpha = 0, \beta = 1; \\ 0, & \text{если } \alpha = 0, \beta \neq 1; \\ \frac{\alpha}{1-\beta}, & \text{если } \alpha \neq 0, \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ 1, & \text{если } \alpha \neq 0, \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{\alpha}{\alpha+\beta-1}, & \text{если } \alpha \neq 0, \frac{\beta}{\alpha} > \frac{1}{\alpha}. \end{cases}$$

При $\alpha \neq 0$ вид обновлений упрощается:

$$\begin{aligned} x_{kj} &\leftarrow x_{kj} \left(\frac{\sum_{i=1}^m a_{ik} p_{ij}^{\alpha} q_{ij}^{\beta-1}}{\sum_{i=1}^m a_{ik} q_{ij}^{\alpha+\beta-1}} \right)^{\omega'(\alpha, \beta)} ; \\ a_{ik} &\leftarrow a_{ik} \left(\frac{\sum_{j=1}^n x_{kj} p_{ij}^{\alpha} q_{ij}^{\beta-1}}{\sum_{j=1}^n x_{kj} q_{ij}^{\alpha+\beta-1}} \right)^{\omega'(\alpha, \beta)} ; \end{aligned}$$

$$\omega'(\alpha, \beta) = \begin{cases} \frac{1}{1-\beta}, & \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \frac{1}{\alpha}, & \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{1}{\alpha + \beta - 1}, & \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha}. \end{cases}$$

В матричном виде их можно записать следующим образом:

$$\begin{aligned} X &\leftarrow X \otimes ((A^T Z) \oslash (A^T Q^{[\alpha+\beta-1]}))^{\omega'(\alpha, \beta)} ; \\ A &\leftarrow A \otimes ((Z X^T) \oslash (Q^{[\alpha+\beta-1]} X^T))^{\omega'(\alpha, \beta)}. \end{aligned} \quad (4)$$

Здесь $Z = P^{[\alpha]} \otimes Q^{[\beta-1]}$, символом $\otimes$ обозначается операция поэлементного (Адамарова) произведения матриц, символом $\oslash$ — поэлементного деления, $[\cdot]$ — поэлементного возведения в степень.

Видно, что при $\alpha = 0$, $\beta \neq 1$ алгоритм (3) не модифицирует A и X ; авторы [7] предлагают ограничить $\omega(\alpha, \beta)$ снизу небольшой положительной константой, чтобы алгоритм можно было использовать при α , близких к нулю, и β , не равных единице. В данной работе мы не будем подробнее останавливаться на этом частном случае.

Отметим, что в случаях $\alpha = 1$, $\beta = 1$, когда АБ-дивергенция превращается в норму Фробениуса, и $\alpha = 1$, $\beta = 0$, когда получается дивергенция Кульбака-Лейблера, алгоритм (3) совпадает с мультипликативными алгоритмами для соответствующих функций потерь из работ [2, 3].

Сходимость

Рассмотрим определения и понятия сходимости, используемые для неотрицательного матричного разложения [4].

Лагранжиан оптимизационной задачи (1) имеет вид $L(A, X, \mu, \nu) = D(P, AX) - \mu \otimes A - \nu \otimes X$, где μ и ν — матрицы множителей Лагранжа размеров $m \times r$ и $r \times n$ соответственно. Согласно теореме Каруша–Куна–Таккера [13], если (A, X) — локальный минимум, то найдутся такие μ и ν с неотрицательными элементами, что выполняются условия

$$\begin{aligned} A &\geq 0, & X &\geq 0; \\ \nabla_A L &= 0, & \nabla_X L &= 0; \\ \mu \otimes A &= 0, & \nu \otimes X &= 0. \end{aligned}$$

Упрощая вторую пару условий, получаем

$$\mu = -\nabla_A D, \quad \nu = -\nabla_X D.$$

С учетом этого, а также требований $\mu_{ik} \geq 0$, $\nu_{kj} \geq 0$, условия Каруша–Куна–Таккера можно записать в следующем виде:

$$A \geq 0, \quad X \geq 0; \tag{5-а}$$

$$\nabla_A D \geq 0, \quad \nabla_X D \geq 0; \tag{5-б}$$

$$A \otimes \nabla_A D = 0, \quad X \otimes \nabla_X D = 0. \tag{5-в}$$

Определение 1. Будем называть (A, X) стационарной точкой задачи неотрицательного матричного разложения с функцией потерь $D(P, AX)$ тогда и только тогда, когда для A и X выполняются условия Каруша–Куна–Таккера (5-а), (5-б), (5-в).

Поскольку рассматриваемые функции потерь $D(P, AX)$ не являются выпуклыми по совокупности аргументов (A, X) , выполнение (5-а)–(5-в) является только необходимым условием локального минимума. Однако, как правило, в задачах невыпуклой оптимизации, к которым относится и рассматриваемая задача, для большинства алгоритмов удается показать только сходимость к стационарной точке, а более сильных утверждений о сходимости доказать не удастся [14]. На практике при минимизации невыпуклых функций именно такой вид сходимости используется как доказательство применимости алгоритма.

Мультипликативные алгоритмы широко используются в неотрицательном матричном разложении несмотря на то, что их сходимость к стационарной точке показать достаточно

сложно — для многих алгоритмов она не доказана вообще. Вместо этого как некоторый аналог сходимости зачастую используется тот факт, что на каждой итерации алгоритма значение функции потерь не возрастает. Поскольку $D(P, A^t, X^t)$ ограничена снизу (нулем), ее невозрастание позволяет надеяться, что алгоритм сойдется. При этом невозрастание функции потерь не является ни необходимым, ни достаточным условием сходимости к стационарной точке: во-первых, не гарантируется, что предельная точка алгоритма с невозрастающей функцией потерь является стационарной, во-вторых, нет гарантии, что алгоритм сойдется к своей предельной точке [15].

Для доказательства невозрастания функции потерь чаще всего используются дополнительные функции. Пусть при фиксированной матрице A минимизируется $D(X) = D(P, AX)$; функция $G(X, Y)$ называется дополнительной для $D(X)$, если

$$G(X, Y) \geq D(X), \quad G(X, X) = D(X) \quad \forall X, Y.$$

Определим

$$X^{t+1} = \arg \min_X G(X, X^t). \quad (6)$$

Тогда по построению

$$D(X^t) = G(X^t, X^t) \geq G(X^{t+1}, X^t) \geq G(X^{t+1}, X^{t+1}) = D(X^{t+1}).$$

В силу симметричности задачи факторизации функция $G(X, Y)$, дополнительная для $D(X)$, будет дополнительной и для $D(A^T) = D(P^T, X^T A^T)$:

$$G(A^T, B^T) \geq D(A^T), \quad G(A^T, A^T) = D(A^T) \quad \forall A, B.$$

Таким образом, для доказательства невозрастания функции потерь алгоритма достаточно привести такую функцию $G(X, Y)$, что обновления X можно представить как решение задачи (6), а обновления A — как решение аналогичной задачи:

$$A^{t+1} = \arg \min_A G(A, A^t). \quad (7)$$

Мультипликативный алгоритм минимизации нормы Фробениуса

Рассмотрим сначала наиболее исследованный частный случай $\alpha = 1$, $\beta = 1$, при котором АБ-дивергенция превращается в норму Фробениуса:

$$(A^*, X^*) = \arg \min_{A \geq 0, X \geq 0} D_F(P, AX) = \arg \min_{A \geq 0, X \geq 0} \|P - AX\|_F^2. \quad (8)$$

Именно для этого случая в одной из первых работ по неотрицательному матричному разложению [2] был впервые предложен мультипликативный алгоритм следующего вида:

$$\begin{aligned} X &\leftarrow X \otimes ((A^T P) \oslash (A^T A X)); \\ A &\leftarrow A \otimes ((P X^T) \oslash (A X X^T)), \end{aligned} \quad (9)$$

а в [3] показано монотонное невозрастание функции потерь при его использовании.

У этого алгоритма существует несколько проблем, возникающих на границе области неотрицательности элементов A и X . Чтобы выполнялись условия Каруша-Куна-Таккера (5-б), (5-в), необходимо, чтобы в предельной точке алгоритма не было таких k, j , что

$\nabla_{x_{jk}} D < 0$, $x_{jk} = 0$. Если элементы матрицы начального приближения X^0 строго положительны, то это не может произойти, так как, если на шаге с номером t выполняется

$$\nabla_{x_{kj}} D = \left[(A^t)^T A^t X^t \right]_{kj} - \left[(A^t)^T P \right]_{kj} < 0,$$

то на следующем шаге

$$x_{kj}^{t+1} = x_{kj}^t \frac{\left[(A^t)^T P \right]_{kj}}{\left[(A^t)^T A^t X^t \right]_{kj}} > x_{kj}^t > 0.$$

Аналогичные рассуждения справедливы и для A . Следовательно, чтобы гарантировать выполнение условий стационарности, достаточно, чтобы все элементы искомых матриц были инициализированы строго положительными, а в ходе мультипликативных обновлений они автоматически будут оставаться такими.

Однако это избавляет нас не от всех проблем: сходимость алгоритма оказывается под вопросом не только тогда, когда какие-то из элементов матриц в точности равны нулю, но и тогда, когда они к нулю стремятся. В этом случае алгоритм также может останавливаться в нестационарных точках. Чтобы показать это, рассмотрим обновление для столбца $\mathbf{x}_j$ матрицы X , обозначая как $\mathbf{p}_j$ соответствующий столбец матрицы P , и перепишем мультипликативное обновление в аддитивном виде:

$$\begin{aligned} \mathbf{x}_j^{t+1} &= \mathbf{x}_j^t \otimes ((A^T \mathbf{p}_j) \otimes (A^T A \mathbf{x}_j^t)) = \\ &= \mathbf{x}_j^t \otimes ((A^T \mathbf{p}_j + A^T A \mathbf{x}_j^t - A^T A \mathbf{x}_j^t) \otimes (A^T A \mathbf{x}_j^t)) = \\ &= \mathbf{x}_j^t \otimes (1 - (A^T A \mathbf{x}_j^t - A^T \mathbf{p}_j) \otimes (A^T A \mathbf{x}_j^t)) = \\ &= \mathbf{x}_j^t - \mathbf{x}_j^t \otimes (A^T A \mathbf{x}_j^t) \otimes (A^T A \mathbf{x}_j^t - A^T \mathbf{p}_j) = \\ &= \mathbf{x}_j^t - d_{\mathbf{x}_j}^T \nabla_{\mathbf{x}_j} D(\mathbf{p}_j, A \mathbf{x}_j). \end{aligned}$$

Здесь $d_{\mathbf{x}_j} \in \mathbb{R}_+^r$ — вектор-столбец с элементами $d_k = \frac{x_{kj}^t}{[A^T A \mathbf{x}_j^t]_k}$. Чтобы алгоритм обеспечивал достаточный спуск¹, необходимо, чтобы все d_k были отделены от нуля, что для данного вектора в общем случае неверно; поэтому обновления X могут прекращаться в точке, которая не является стационарной.

Чтобы избежать описанных проблем со сходимостью, достаточно явным образом отделить решения от нуля положительной константой. Это делает следующая модификация алгоритма (9):

$$\begin{aligned} X &\leftarrow \max(\varepsilon, X \otimes ((A^T P) \otimes (A^T A X))) ; \\ A &\leftarrow \max(\varepsilon, A \otimes ((P X^T) \otimes (A X X^T))) , \end{aligned} \tag{10}$$

где $\varepsilon > 0$ — произвольная константа, а максимум берется покомпонентно.

В работе [16] доказаны следующие свойства такого алгоритма.

Утверждение 1. Для любых $\varepsilon > 0$, $P \geq 0$, $(A^0, X^0) \geq \varepsilon$ функция потерь $\|P - AX\|_F^2$ монотонно не возрастает в процессе обновлений (10).

¹В задаче минимизации функции $f(x)$, $x \in \mathbb{R}^n$, направление $d \in \mathbb{R}^n$ является направлением спуска, если $d^T \nabla f(x) < 0$, и направлением достаточного спуска, если существует такая константа C , что $d^T \nabla f(x) \leq -C \|\nabla f(x)\|^2$.

Утверждение 2. Любая предельная точка алгоритма (10) является стационарной точкой оптимизационной задачи

$$(A_\varepsilon^*, X_\varepsilon^*) = \arg \min_{A \geq \varepsilon, X \geq \varepsilon} \|P - AX\|_F^2.$$

Утверждение 3. Пусть $(A_\varepsilon, X_\varepsilon)$ — предельная точка алгоритма (10). Проведем разреживание, обнуляя элементы полученных матриц, в точности равные ε . Определим $A = A_\varepsilon \otimes [A_\varepsilon > \varepsilon]$, $X = X_\varepsilon \otimes [X_\varepsilon > \varepsilon]$. Для всех i, j, k выполняется следующее:

$$\left\{ \begin{array}{l} \left[\begin{array}{l} a_{ik} = 0, \quad \nabla_{a_{ik}} D \geq -\mathcal{O}(\varepsilon); \\ a_{ik} > 0, \quad |\nabla_{a_{ik}} D| \leq \mathcal{O}(\varepsilon); \end{array} \right. \\ \left[\begin{array}{l} x_{kj} = 0, \quad \nabla_{x_{kj}} D \geq -\mathcal{O}(\varepsilon); \\ x_{kj} > 0, \quad |\nabla_{x_{kj}} D| \leq \mathcal{O}(\varepsilon). \end{array} \right. \end{array} \right.$$

То есть, хотя полученное таким образом решение (A, X) и не является в точности стационарной точкой исходной задачи (8), условия Каруша–Куна–Таккера (5-а)–(5-в) выполняются с точностью до $\mathcal{O}(\varepsilon)$.

Таким образом, для алгоритма (10) открытым остается только вопрос существования в генерируемой им последовательности (A^t, X^t) предельной точки. Используем для доказательства ее существования следующее свойство, установленное в работе [17].

Утверждение 4. Любая последовательность (A^t, X^t) , генерируемая алгоритмом (10), принадлежит компакту C при условии, что $(A^0, X^0) \in C$.

Отсюда, с учетом теоремы Больцано–Вейерштрасса, получаем следствие.

Следствие 1. Последовательность (A^t, X^t) , порождаемая алгоритмом (10), имеет хотя бы одну предельную точку.

Таким образом, для модифицированного мультипликативного алгоритма (10) неотрицательного матричного разложения с нормой Фробениуса показано, что он всегда сходится, а его предельная точка лежит сколь угодно близко к стационарной точке исходной задачи (8). То есть небольшими изменениями в алгоритме (9) удастся обеспечить его вычислительную устойчивость (отсутствие деления на ноль) и гарантировать сходимость в окрестность стационарной точки, причем параметром, определяющим размер этой окрестности, можно управлять, обеспечивая баланс между вычислительной устойчивостью алгоритма и точностью получаемого решения.

Мультипликативный алгоритм минимизации АБ-дивергенции

Вернемся к рассмотрению задачи (2) неотрицательного матричного разложения с АБ-дивергенцией. Относительно алгоритма (3) в работе [7] доказано следующее утверждение.

Утверждение 5. Функция $D_{AB}^{(\alpha, \beta)}(P, AX)$ монотонно не возрастает при обновлениях (4) для любого начального приближения $(A^0, X^0) \geq 0$.

Для доказательства используется следующая дополнительная функция:

$$G(Q^{t+1}, Q^t, P) = \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^r \frac{a_{ik}^t x_{kj}^t}{q_{ij}^t} \bar{d}_{AB}^{(\alpha, \beta)}(\hat{q}_{ij}, q_{ij}^t, p_{ij});$$

$$\bar{d}_{AB}^{(\alpha, \beta)}(\hat{q}_{ij}, q_{ij}^t, p_{ij}) = \begin{cases} \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{(q_{ij}^t)^{\alpha+\beta}}{\alpha(\alpha+\beta)} - \frac{p_{ij}^\alpha \hat{q}_{ij}^\beta}{\alpha\beta} + \frac{(q_{ij}^t)^{\alpha+\beta-1}}{\alpha} (\hat{q}_{ij} - q_{ij}^t), \\ \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{\hat{q}_{ij}^{\alpha+\beta}}{\alpha(\alpha+\beta)} - \frac{p_{ij}^\alpha \hat{q}_{ij}^\beta}{\alpha\beta}, \\ \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{\hat{q}_{ij}^{\alpha+\beta}}{\alpha(\alpha+\beta)} - \frac{p_{ij}^\alpha (q_{ij}^t)^\beta}{\alpha\beta} - \frac{p_{ij}^\alpha (q_{ij}^t)^{\beta-1}}{\alpha} (\hat{q}_{ij} - q_{ij}^t); \\ \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha}, \end{cases} \quad (11)$$

$$\hat{q}_{ij}^t = q_{ij}^t \frac{a_{ik}^{t+1} x_{kj}^{t+1}}{a_{ik}^t x_{kj}^t}.$$

При $\beta = 0$ и $\alpha = -\beta$ функция доопределяется при помощи правила Лопиталя.

Заметим, что знаменатели в (4) могут принимать нулевые значения. Чтобы избежать деления на ноль, в работе [7] было предложено прибавлять небольшую положительную константу ε к знаменателям выражений в (4), однако вопрос влияния такой модификации на сходимость алгоритма и свойства получаемого решения не рассматривался. Вместо добавления ε к знаменателям множителей мультипликативного алгоритма модифицируем (4) аналогично (10), отделив принимаемые элементами матриц значения от нуля:

$$\begin{aligned} X &\leftarrow \max \left(\varepsilon, X \otimes ((A^T Z) \oslash (A^T Q^{[\alpha+\beta-1]}))^{[\omega'(\alpha, \beta)]} \right); \\ A &\leftarrow \max \left(\varepsilon, A \otimes ((Z X^T) \oslash (Q^{[\alpha+\beta-1]} X^T))^{[\omega'(\alpha, \beta)]} \right). \end{aligned} \quad (12)$$

Рассмотрим задачу минимизации функции потерь в отделенной от нуля области:

$$(A_\varepsilon^*, X_\varepsilon^*) = \arg \min_{A \geq \varepsilon, X \geq \varepsilon} D_{AB}^{(\alpha, \beta)}(P, AX). \quad (13)$$

Теорема 1. Для любого $\varepsilon > 0$ функция $D_{AB}^{(\alpha, \beta)}(P, AX)$ монотонно не возрастает при обновлениях (12) для любого начального приближения $(A^0, X^0) \geq \varepsilon$.

Доказательство. Пусть $A^t \geq \varepsilon$. Обозначим за $X^{t+1} = \arg \min_{X \geq \varepsilon} G(A^t X, A^t X^t, P)$, где G — дополнительная функция (11). По определению дополнительной функции имеем:

$$\begin{aligned} D_{AB}^{(\alpha, \beta)}(P, A^t X^t) &= G(A^t X^t, A^t X^t, P) \geq \min_{X \geq \varepsilon} G(A^t X, A^t X^t, P) = \\ &= G(A^t X^{t+1}, A^t X^t, P) \geq G(A^t X^{t+1}, A^t X^{t+1}, P) = D_{AB}^{(\alpha, \beta)}(P, A^t X^{t+1}). \end{aligned}$$

Покажем, что обновления (12) совпадают с решениями задач (6), (7) минимизации дополнительной функции. Рассмотрим функцию

$$G(A^t X, A^t X^t, P) = \sum_{i,j,k} \frac{a_{ik}^t x_{kj}^t}{q_{ij}^t} \bar{d}_{AB}^{(\alpha,\beta)} \left(q_{ij}^t \frac{x_{kj}^t}{x_{kj}^t}, q_{ij}^t, p_{ij} \right).$$

Обозначим

$$g_{kj}(x_{kj}, x_{kj}^t, A^t, P) = x_{kj}^t \sum_{i=1}^m \frac{a_{ik}^t}{q_{ij}^t} \bar{d}_{AB}^{(\alpha,\beta)} \left(q_{ij}^t \frac{x_{kj}^t}{x_{kj}^t}, q_{ij}^t, p_{ij} \right).$$

Поскольку функция $G(A^t X, A^t X^t, P)$ состоит из независимых слагаемых по x_{kj} ,

$$\begin{aligned} X^{t+1} &= \arg \min_{X \geq \varepsilon} G(A^t X, A^t X^t, P) \iff \\ x_{kj}^{t+1} &= \arg \min_{x_{kj} \geq \varepsilon} g_{kj}(x_{kj}, x_{kj}^t, A^t, P) \quad \forall k, j. \end{aligned}$$

Далее, поскольку функция

$$\bar{d}_{AB}^{(\alpha,\beta)} \left(q_{ij}^t \frac{x_{kj}^t}{x_{kj}^t}, q_{ij}^t, p_{ij} \right) = \begin{cases} \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{(q_{ij}^t)^{\alpha+\beta}}{\alpha(\alpha+\beta)} - \frac{p_{ij}^\alpha (q_{ij}^t)^\beta x_{kj}^\beta}{\alpha\beta (x_{kj}^t)^\beta} + \frac{(q_{ij}^t)^{\alpha+\beta}}{\alpha} \left(\frac{x_{kj}^t}{x_{kj}^t} - 1 \right), & \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{(q_{ij}^t)^{\alpha+\beta} x_{kj}^{\alpha+\beta}}{\alpha(\alpha+\beta) (x_{kj}^t)^{\alpha+\beta}} - \frac{p_{ij}^\alpha \hat{q}_{ij}^\beta}{\alpha\beta}, & \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{p_{ij}^{\alpha+\beta}}{\beta(\alpha+\beta)} + \frac{(q_{ij}^t)^{\alpha+\beta} x_{kj}^{\alpha+\beta}}{\alpha(\alpha+\beta) (x_{kj}^t)^{\alpha+\beta}} - \frac{p_{ij}^\alpha (q_{ij}^t)^\beta}{\alpha\beta} - \frac{p_{ij}^\alpha (q_{ij}^t)^\beta}{\alpha} \left(\frac{x_{kj}^t}{x_{kj}^t} - 1 \right), & \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha} \end{cases}$$

является выпуклой по x_{kj} по построению [7], то выпукла по x_{kj} и $g_{kj}(x_{kj}, x_{kj}^t, A^t, P)$; следовательно, ее безусловный минимум можно получить из условия

$$\frac{\partial g_{kj}}{\partial x_{kj}} = x_{kj}^t \sum_{i=1}^m \frac{a_{ik}^t}{q_{ij}^t} \frac{\partial \bar{d}_{AB}}{\partial x_{kj}} = 0.$$

Распишем его:

$$\frac{\partial \bar{d}_{AB}}{\partial x_{kj}} = \begin{cases} \frac{(q_{ij}^t)^\beta}{\alpha} \frac{1}{x_{kj}^t} \left((q_{ij}^t)^\alpha - p_{ij}^\alpha \left(\frac{x_{kj}^t}{x_{kj}^t} \right)^{\beta-1} \right), & \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \frac{(q_{ij}^t)^\beta}{\alpha} \frac{x_{kj}^{\beta-1}}{(x_{kj}^t)^\beta} \left((q_{ij}^t)^\alpha \left(\frac{x_{kj}^t}{x_{kj}^t} \right)^\alpha - p_{ij}^\alpha \right), & \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{(q_{ij}^t)^\beta}{\alpha} \frac{1}{x_{kj}^t} \left((q_{ij}^t)^\alpha \left(\frac{x_{kj}^t}{x_{kj}^t} \right)^{\alpha+\beta-1} - p_{ij}^\alpha \right), & \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha}. \end{cases}$$

$$\frac{\partial g_{kj}}{\partial x_{kj}} = \begin{cases} \frac{1}{\alpha} \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} \left((q_{ij}^t)^\alpha - p_{ij}^\alpha \left(\frac{x_{kj}}{x_{kj}^t} \right)^{\beta-1} \right), & \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \frac{1}{\alpha} \left(\frac{x_{kj}}{x_{kj}^t} \right)^\beta \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} \left((q_{ij}^t)^\alpha \left(\frac{x_{kj}}{x_{kj}^t} \right)^\alpha - p_{ij}^\alpha \right), & \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \frac{1}{\alpha} \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} \left((q_{ij}^t)^\alpha \left(\frac{x_{kj}}{x_{kj}^t} \right)^{\alpha+\beta-1} - p_{ij}^\alpha \right), & \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha}; \end{cases}$$

$$\frac{\partial g_{kj}}{\partial x_{kj}} = 0 \Leftrightarrow \begin{cases} \left(\frac{x_{kj}}{x_{kj}^t} \right)^{1-\beta} \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\alpha+\beta-1} = \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} p_{ij}^\alpha, & \text{если } \frac{\beta}{\alpha} < \frac{1}{\alpha} - 1; \\ \left(\frac{x_{kj}}{x_{kj}^t} \right)^\alpha \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\alpha+\beta-1} = \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} p_{ij}^\alpha, & \text{если } \frac{\beta}{\alpha} \in \left[\frac{1}{\alpha} - 1, \frac{1}{\alpha} \right]; \\ \left(\frac{x_{kj}}{x_{kj}^t} \right)^{\alpha+\beta-1} \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\alpha+\beta-1} = \sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} p_{ij}^\alpha, & \text{если } \frac{\beta}{\alpha} > \frac{1}{\alpha}. \end{cases}$$

Отсюда получаем:

$$x_{kj}^{t+1} = \arg \min_{x_{kj} \geq \varepsilon} g_{kj}(x_{kj}, x_{kj}^t, A^t, P) = \max \left(\varepsilon, x_{kj}^t \left(\frac{\sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\beta-1} p_{ij}^\alpha}{\sum_{i=1}^m a_{ik}^t (q_{ij}^t)^{\alpha+\beta-1}} \right)^{\omega'(\alpha, \beta)} \right),$$

что в точности совпадает с обновлениями алгоритма (12); следовательно, обновления по X из (12) приводят к монотонному невозрастанию функции потерь.

Для обновлений по A доказательство строится аналогично. $\blacksquare$

Рассмотрим теперь вопрос качества решения, получаемого модифицированным алгоритмом (12). Для задачи (13) условия Каруша–Куна–Таккера записываются следующим образом:

$$\begin{aligned} A &\geq \varepsilon, & X &\geq \varepsilon, \\ \nabla_A D_{AB}^{(\alpha, \beta)}(P, AX) &\geq 0, & \nabla_X D_{AB}^{(\alpha, \beta)}(P, AX) &\geq 0, \\ (A - \varepsilon) \otimes \nabla_A D_{AB}^{(\alpha, \beta)}(P, AX) &= 0, & (X - \varepsilon) \otimes \nabla_X D_{AB}^{(\alpha, \beta)}(P, AX) &= 0. \end{aligned}$$

Следующая теорема утверждает, что в предельной точке последовательности, получаемой при помощи алгоритма (12), эти условия выполняются.

Теорема 2. Любая предельная точка последовательности, порождаемой алгоритмом (12) для любого начального приближения $(A^0, X^0) \geq \varepsilon$, является стационарной точкой отделенной от нуля задачи (13).

Доказательство. Пусть $(\bar{A}, \bar{X})$ — предельная точка последовательности (A^t, X^t) , порождаемой алгоритмом (12), а $\bar{Q} = \bar{A}\bar{X}$. Поскольку функция потерь ограничена снизу (нулем), из ее монотонного невозрастания следует сходимость $D_{AB}^{(\alpha, \beta)}(P, A^t X^t)$ к $D_{AB}^{(\alpha, \beta)}(P, \bar{A}\bar{X})$. Кроме того,

$$\bar{x}_{kj} = \max \left(\varepsilon, c_{kj}^{\omega'(\alpha, \beta)} \bar{x}_{kj} \right), \quad (15)$$

где

$$c_{kj} \equiv \frac{\sum_{i=1}^m \bar{a}_{ik} \bar{q}_{ij}^{\beta-1} p_{ij}^{\alpha}}{\sum_{i=1}^m \bar{a}_{ik} \bar{q}_{ij}^{\alpha+\beta-1}},$$

причем это выражение всегда определено, так как $(A^0, X^0) \geq \varepsilon$ и выражение в знаменателе не может быть равным нулю. Пользуясь выражением для градиента АБ-дивергенции:

$$\left[\nabla_X D_{AB}^{(\alpha, \beta)}(P, AX) \right]_{kj} = \frac{1}{\alpha} \sum_{i=1}^m q_{ij}^{\beta-1} a_{ik} (q_{ij}^{\alpha} - p_{ij}^{\alpha}) = \frac{1}{\alpha} \left(1 - \frac{\sum_{i=1}^m a_{ik} q_{ij}^{\beta-1} p_{ij}^{\alpha}}{\sum_{i=1}^m a_{ik} q_{ij}^{\alpha+\beta-1}} \right),$$

необходимые и достаточные условия стационарности для компоненты $\bar{x}_{kj}$ решения задачи (13) можно записать следующим образом:

$$\begin{aligned} \bar{x}_{kj} &\geq \varepsilon; \\ c_{kj} &\leq 1; \\ (\bar{x}_{kj} - \varepsilon)(c_{kj} - 1) &= 0. \end{aligned}$$

В то же время, из выражения (15) следует, что, поскольку $\bar{x}_{kj}$ — предельная точка, либо $c_{kj}^{\omega'(\alpha, \beta)} \leq 1$ (а раз $\omega'(\alpha, \beta) \leq 1$, то и $c_{kj} \leq 1$) и $\bar{x}_{kj} = \varepsilon$, либо $\bar{x}_{kj} > \varepsilon$ и $c_{kj} = 1$; следовательно, условия стационарности для $\bar{x}_{kj}$ выполняются. Это справедливо для всех k, j , т. е. для всей матрицы $\bar{X}$.

Для $\bar{A}$ доказательство строится аналогично. ■

Пусть $(A_{\varepsilon}, X_{\varepsilon})$ — предельная точка алгоритма (12). Определим следующие матрицы:

$$\begin{aligned} A_0 &= A_{\varepsilon} \otimes [A_{\varepsilon} > \varepsilon]; \\ X_0 &= X_{\varepsilon} \otimes [X_{\varepsilon} > \varepsilon], \end{aligned}$$

т. е. обнулим в $A_{\varepsilon}, X_{\varepsilon}$ все элементы, равные ε . Покажем, что (A_0, X_0) аппроксимирует стационарную точку исходной задачи (2) с точностью порядка ε .

Теорема 3. Для матриц (A_0, X_0) , полученных из $(A_{\varepsilon}, X_{\varepsilon})$ обнулением элементов, равных ε , верно следующее:

$$\left\{ \begin{aligned} &\left[\begin{aligned} a_{0ik} &= 0, & \left(\nabla_A D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right)_{ik} &\geq -\mathcal{O}(\varepsilon); \\ a_{0ik} &> 0, & \left| \nabla_A D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right|_{ik} &\leq \mathcal{O}(\varepsilon); \end{aligned} \right. \\ &\left[\begin{aligned} x_{0kj} &= 0, & \left(\nabla_X D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right)_{kj} &\geq -\mathcal{O}(\varepsilon); \\ x_{0kj} &> 0, & \left| \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right|_{kj} &\leq \mathcal{O}(\varepsilon), \end{aligned} \right. \end{aligned} \right.$$

т. е. в точке (A_0, X_0) условия Каруша-Куна-Таккера для исходной задачи (2) практически выполняются.

Доказательство. Для стационарной точки $(A_\varepsilon, X_\varepsilon)$ задачи (13) выполняются условия Каруша–Куна–Таккера:

$$\left\{ \begin{array}{l} \left[\begin{array}{l} a_{\varepsilon ik} = \varepsilon, \quad \left(\nabla_A D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) \right)_{ik} \geq 0; \\ a_{\varepsilon ik} > \varepsilon, \quad \left| \nabla_A D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) \right|_{ik} = 0; \end{array} \right. \\ \left[\begin{array}{l} x_{\varepsilon kj} = \varepsilon, \quad \left(\nabla_X D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) \right)_{kj} \geq 0; \\ x_{\varepsilon kj} > \varepsilon, \quad \left| \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) \right|_{kj} = 0. \end{array} \right. \end{array} \right. \quad (16)$$

Построим верхнюю оценку для расстояния между значениями градиента $\nabla_X D_{AB}^{(\alpha, \beta)}$ в точках (A_0, X_0) и $(A_\varepsilon, X_\varepsilon)$, обозначая $Q_0 = A_0 X_0$ и $Q_\varepsilon = A_\varepsilon X_\varepsilon$:

$$\begin{aligned} \max_{k,j} \left| \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) - \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right|_{k,j} &= \\ &= \frac{1}{\alpha} \max_{k,j} \left| A_\varepsilon^T (Q_\varepsilon^{[\alpha+\beta-1]} - P^{[\alpha]} \otimes Q_\varepsilon^{[\beta-1]}) - A_0^T (Q_0^{[\alpha+\beta-1]} - P^{[\alpha]} \otimes Q_0^{[\beta-1]}) \right|_{k,j} \leq \\ &\leq \frac{1}{\alpha} \max_{k,j} \left| A_\varepsilon^T Q_\varepsilon^{[\alpha+\beta-1]} - A_0^T Q_0^{[\alpha+\beta-1]} \right|_{k,j} + \\ &\quad + \frac{1}{\alpha} \max_{k,j} \left| A_\varepsilon^T (Q_\varepsilon^{[\beta-1]} \otimes P^{[\alpha]}) - A_0^T (Q_0^{[\beta-1]} \otimes P^{[\alpha]}) \right|_{k,j} \leq \\ &\leq \frac{1}{\alpha} \max_{k,j} \left| \sum_{i=1}^m \left((a_{0ik} + \varepsilon) q_{\varepsilon ij}^{\alpha+\beta-1} - a_{0ik} q_{0ij}^{\alpha+\beta-1} \right) \right| + \\ &\quad + \frac{1}{\alpha} \max_{k,j} \left| \sum_{i=1}^m p_{ij}^\alpha \left((a_{0ik} + \varepsilon) q_{\varepsilon ij}^{\beta-1} - a_{0ik} q_{0ij}^{\beta-1} \right) \right|. \end{aligned} \quad (17)$$

Запишем выражения для элементов матриц Q_0 и Q_ε :

$$\begin{aligned} q_{0ij} &= \sum_{l=1}^r a_{0il} x_{0lj} = \sum_{l: \substack{a_{0il} \neq 0, \\ x_{0lj} \neq 0}} a_{0il} x_{0lj}; \\ q_{\varepsilon ij} &= \sum_{l=1}^r a_{\varepsilon il} x_{\varepsilon lj} = \sum_{l: \substack{a_{\varepsilon il} > \varepsilon, \\ x_{\varepsilon lj} > \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} + \sum_{l: \substack{a_{\varepsilon il} > \varepsilon, \\ x_{\varepsilon lj} = \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} + \sum_{l: \substack{a_{\varepsilon il} = \varepsilon, \\ x_{\varepsilon lj} > \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} + \sum_{l: \substack{a_{\varepsilon il} = \varepsilon, \\ x_{\varepsilon lj} = \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj}. \end{aligned}$$

По построению матриц A_0, X_0 верно следующее:

$$\begin{aligned} \sum_{l: \substack{a_{\varepsilon il} > \varepsilon, \\ x_{\varepsilon lj} > \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} &= \sum_{l: \substack{a_{0il} \neq 0, \\ x_{0lj} \neq 0}} a_{0il} x_{0lj} = q_{0ij}; \\ \sum_{l: \substack{a_{\varepsilon il} > \varepsilon, \\ x_{\varepsilon lj} = \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} &= \varepsilon \sum_{l: \substack{a_{0il} \neq 0, \\ x_{0lj} = 0}} a_{0il}; \\ \sum_{l: \substack{a_{\varepsilon il} = \varepsilon, \\ x_{\varepsilon lj} > \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} &= \varepsilon \sum_{l: \substack{a_{0il} = 0, \\ x_{0lj} \neq 0}} x_{0lj}; \\ \sum_{l: \substack{a_{\varepsilon il} = \varepsilon, \\ x_{\varepsilon lj} = \varepsilon}} a_{\varepsilon il} x_{\varepsilon lj} &= \varepsilon^2 \cdot \# \{l: a_{\varepsilon il} = \varepsilon, x_{\varepsilon lj} = \varepsilon\} = \varepsilon^2 \cdot \# \{l: a_{0il} = 0, x_{0lj} = 0\}, \end{aligned}$$

откуда получаем:

$$\begin{aligned} q_{\varepsilon ij} &= \sum_{l: \substack{a_{0il} \neq 0, \\ x_{0lj} \neq 0}} a_{0il} x_{0lj} + \varepsilon \left(\sum_{l: \substack{a_{0il} \neq 0, \\ x_{0lj} = 0}} a_{0il} + \sum_{l: \substack{a_{0il} = 0, \\ x_{0lj} \neq 0}} x_{0lj} \right) + \varepsilon^2 \cdot \# \{l: a_{0il} = 0, x_{0lj} = 0\} \equiv \\ &\equiv q_{0ij} + \varepsilon d_{ij} + \varepsilon^2 f_{ij}. \end{aligned}$$

Используя это представление, запишем первые несколько членов ряда Тейлора функции $q_{\varepsilon ij}^y$ при разложении по ε в нуле:

$$q_{\varepsilon ij}^y = (q_{0ij} + \varepsilon d_{ij} + \varepsilon^2 f_{ij})^y = q_{0ij}^y + y q_{0ij}^{y-1} d_{ij} \varepsilon + \mathcal{O}(\varepsilon^2).$$

Подставим это разложение при $y = \alpha + \beta - 1$ в выражение под знаком суммы в первом слагаемом правой части (17):

$$\begin{aligned} (a_{0ik} + \varepsilon) q_{\varepsilon ij}^{\alpha+\beta-1} - a_{0ik} q_{0ij}^{\alpha+\beta-1} &= (a_{0ik} + \varepsilon) \left(q_{0ij}^{\alpha+\beta-1} + (\alpha + \beta - 1) q_{0ij}^{\alpha+\beta-2} d_{ij} \varepsilon + \mathcal{O}(\varepsilon^2) \right) - \\ - a_{0ik} q_{0ij}^{\alpha+\beta-1} &= \varepsilon (\alpha + \beta - 1) q_{0ij}^{\alpha+\beta-2} d_{ij} (a_{0ik} + \varepsilon) + \mathcal{O}(\varepsilon^2) = \mathcal{O}(\varepsilon). \end{aligned}$$

Аналогично для второго слагаемого при $y = \beta - 1$:

$$\begin{aligned} p_{ij}^\alpha \left((a_{0ik} + \varepsilon) q_{\varepsilon ij}^{\beta-1} - a_{0ik} q_{0ij}^{\beta-1} \right) &= p_{ij} \left((a_{0ik} + \varepsilon) \left(q_{0ij}^{\beta-1} + (\beta - 1) q_{0ij}^{\beta-2} d_{ij} \varepsilon + \mathcal{O}(\varepsilon^2) \right) - \right. \\ &\left. - a_{0ik} q_{0ij}^{\beta-1} \right) = p_{ij} \left(\varepsilon (\beta - 1) q_{0ij}^{\beta-2} d_{ij} (a_{0ik} + \varepsilon) + \mathcal{O}(\varepsilon^2) \right) = \mathcal{O}(\varepsilon). \end{aligned}$$

Таким образом, для максимального расстояния между матрицами градиентов $\nabla_X D_{AB}^{(\alpha, \beta)}$ справедливо

$$\max_{k,j} \left| \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) - \nabla_X D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right|_{k,j} \leq \mathcal{O}(\varepsilon).$$

Аналогичным образом легко показать, что

$$\max_{i,k} \left| \nabla_A D_{AB}^{(\alpha, \beta)}(P, A_\varepsilon X_\varepsilon) - \nabla_A D_{AB}^{(\alpha, \beta)}(P, A_0 X_0) \right|_{i,k} \leq \mathcal{O}(\varepsilon).$$

Подставляя полученное в условия Каруша–Куна–Таккера для точки $(A_\varepsilon, X_\varepsilon)$ (16), получаем утверждение теоремы. ■

Заключение

Функции потерь, которые могут быть применены в задаче неотрицательного матричного разложения, соответствуют различным вероятностным предположениям о характере шума в модели. Несмотря на то что выбор функции потерь существенным образом влияет на решение задачи, на практике исследователи зачастую выбирают норму Фробениуса или дивергенцию Кульбака–Лейблера, поскольку методы для них лучше всего известны и проработаны. Разработка методов, использующих такие параметрические семейства функций потерь, как АБ-дивергенции, позволяет унифицировать работу с различными функциями потерь, делая выбор меры качества более обоснованным. Имея в распоряжении такие методы, вопрос определения модели шума можно свести к выбору оптимальных значений параметров дивергенции.

Нулевые элементы матриц являются проблемой для многих методов неотрицательного матричного разложения. Во многих задачах естественно предположение о разреженности матриц-множителей, но добиться ее мультипликативными методами нельзя, а более сложные методы разработаны только для отдельных функций потерь. Предложенный в данной работе метод позволяет, с одной стороны, получить разреженные матрицы, а с другой — гарантировать, что получаемое решение близко к стационарной точке задачи.

Литература

- [1] *Paatero P., Tapper U.* Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values // *Environmetrics*, 1994. Vol. 5. Pp. 111–126.
- [2] *Lee D., Seung S.* Learning the parts of objects by non-negative matrix factorization // *Nature*, 1999. Vol. 401, no. 6755. Pp. 788–791.
- [3] *Lee D., Seung S.* Algorithms for non-negative matrix factorization // *Advances in neural information processing*, 2001. Vol. 13. Pp. 556–562.
- [4] *Wang Y., Zhang Y.* Non-negative matrix factorization: A comprehensive review // *IEEE Transactions on Knowledge and Data Engineering*, 2012. Vol. 25, no. 6. Pp. 1336–1353.
- [5] *Zhang Z.-Y.* Divergence functions of non negative matrix factorization: A comparison study // *Communications in Statistics — Simulation and Computation*, 2011. Vol. 40, no. 10. Pp. 1594–1612.
- [6] *Févotte C., Cemgil A.* Nonnegative matrix factorizations as probabilistic inference in composite models // *17th European Signal Processing Conference*, 2009. Glasgow, Scotland. Pp. 1913–1917.
- [7] *Cichocki A., Cruces S., Amari S.* Generalized alpha-beta divergences and their application to robust nonnegative matrix factorization // *Entropy*, 2011. Vol. 13, no. 1. Pp. 134–170.
- [8] *Cichocki A., Lee H., Kim Y., Choi S.* Nonnegative matrix factorization with α -divergence // *Pattern Recognition Letters*, 2008. Vol. 29, no. 9. Pp. 1433–1440.
- [9] *Kompass R.* A generalized divergence measure for nonnegative matrix factorization // *Neural Computation*, 2007. Vol. 19, no. 3. Pp. 780–791.
- [10] *Рябенко Е. А.* Настройка нелинейной модели данных экспериментов с экспрессионными ДНК-микрочипами // *Математическая биология и биоинформатика*, 2012. Т. 7. № 2. С. 554–566.
- [11] *Grippo L., Sciandrone M.* On the convergence of the block nonlinear Gauss–Seidel method under convex constraints // *Operations Research Letters*, 2000. Vol. 26, no. 3. Pp. 127–136.
- [12] *Cichocki A., Anh-Huy P.* Fast local algorithms for large scale nonnegative matrix and tensor factorizations // *IEICE Transactions on Fundamentals of Electronics Communications and Computer Sciences*, 2009. Vol. E92-A, no. 3. Pp. 708–721.
- [13] *Bertsekas D. P.* *Nonlinear programming*. Massachusetts: Athena Scientific Belmont, 1999. 780 p.
- [14] *Horst R., Pardalos P. M., Nguyen V. T.* *Introduction to global optimization*. 2nd ed. New York: Springer, 2001. 358 p.
- [15] *Badeau R., Bertin N., Vincent E.* Stability analysis of multiplicative update algorithms and application to nonnegative matrix factorization // *IEEE Transactions on Neural Networks*, 2010. Vol. 21, no. 12. Pp. 1869–1881.
- [16] *Gillis N.* Nonnegative matrix factorization: Complexity, algorithms and applications. PhD Thesis. Departement d'Ingenierie Mathematique, Universite catholique de Louvain, 2011.
- [17] *Hibi R., Takahashi N.* A modified multiplicative update algorithm for euclidean distance-based nonnegative matrix factorization and its global convergence // *Neural Information Processing*, 2011. Vol. 7063. Pp. 655–662.

Экспериментальное исследование параметров регуляризации и аппроксимации скелетных графов бинарных изображений

О. А. Кушнир¹, О. С. Середин², А. В. Степанов³

¹kushnir-olesya@rambler.ru, ²oseredin@yandex.ru, ³tokbl@mail.ru

Тула, Тульский государственный университет

В работе систематизируются и уточняются сведения о параметрах стабилизации непрерывного скелетного графа — коэффициентах регуляризации и аппроксимации. Также приводится алгоритм вычисления диаметра минимальной описанной вокруг скелета окружности, являющегося масштабным множителем для применения коэффициента аппроксимации. Проводятся экспериментальные исследования влияния коэффициентов регуляризации и аппроксимации на топологию скелета и выдвигаются предположения по определению их адекватных значений для выделения в исходном скелете его базового подграфа — устойчивого дескриптора формы фигуры при наличии шумовых изменений ее границы.

Ключевые слова: *бинарное растровое изображение, скелетный граф, базовый подграф, регуляризация и аппроксимация скелета, минимальная описанная окружность.*

Experimental study of regularization and approximation parameters for binary images skeletons

Kushnir O. A.¹, Seredin O. S.², Stepanov A. V.³

Tula State University

The article is devoted to the problem of avoiding the complexity and instability of binary images continuous skeletons by means of the regularization and approximation coefficients. The algorithm for computing the minimal diameter of circle circumscribed about the skeleton is suggested. Minimal diameter is a scaled multiplier which is used for computation of the approximation coefficient. Experimental study of how the regularization and approximation parameters affect the skeleton topology has been done, and assumptions about the appropriate values for finding the base subgraph of skeleton for the stable shape description are proposed.

Keywords: *binary image, skeleton base, pruning, approximation, minimal circumscribed circle.*

Введение

Одним из способов описания бинарного изображения является скелетный граф [1]. Однако недостатком этого подхода является чувствительность к шумам (рис. 1 [2]). Так, скелетное описание чрезвычайно восприимчиво к локальным свойствам края объекта: с каждой точкой локального максимума кривизны границы связана отдельная ветвь скелета. Получается, что две области, имеющие сходную форму, но разную изрезанность края, принципиально различны в смысле топологической структуры их скелетов. Следовательно, скелет является неустойчивым дескриптором, который нельзя без дополнительных преобразований использовать для описания и сравнения форм [4, 3, 5].

Существует несколько процедур для управления сложностью топологии скелета: регуляризация, аппроксимация, склейка. Данная статья посвящена анализу коэффициентов регуляризации и аппроксимации. Нам неизвестны систематические исследования для подбора оптимальных (разумных) значений этих коэффициентов. Мы считаем, что именно

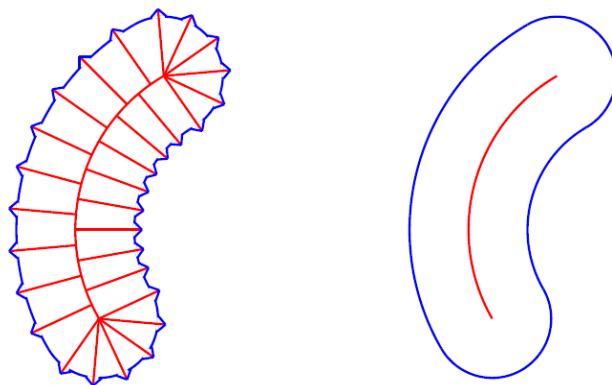


Рис. 1: Иллюстрация неустойчивости скелетного описания

неудачный выбор данных параметров (или отсутствие выбора: например, берутся параметры по умолчанию для решения любой задачи) зачастую отвращает специалистов в области анализа изображений от использования скелетной морфологии. Суть параметров кратко описана ниже.

Регуляризация скелета. Выделение в скелете части, которая называется базовым скелетом, и удаление ребер, появление которых обусловлено шумовыми эффектами, называется регуляризацией скелета. Иногда используется термин «стрижка», англ. *pruning* [2, 4, 6, 7]. Для непрерывных скелетов используется метод регуляризации скелета с контролируемой точностью [5, 8]. Суть его состоит в том, что евклидово расстояние между границей исходной фигуры и границей, которая будет получена после удаления очередного терминального ребра скелета, сравнивается с заданным параметром регуляризации. Ребро может быть удалено и, следовательно, не будет составлять базовый скелет, если это расстояние не превосходит коэффициент регуляризации. Силуэт фигуры, соответствующей усеченному скелету, восстанавливается путем объединения максимальных вписанных в фигуру окружностей с центрами на базовом скелете.

Кроме того, следует отметить содержательно близкий стрижке метод, основанный на понятии разметки скелета [9].

Для дискретных скелетов нельзя точно вычислить радиусы вписанных в фигуру окружностей. Поэтому для них используется метод, основанный на сравнении разности площадей исходной фигуры и фигуры, соответствующей усеченному скелету, с коэффициентом регуляризации [10]. Понятно, что для одной и той же фигуры в случае применения дискретного и непрерывного скелета значение параметра регуляризации будет разным.

В работе [11] было определено понятие интервалов стабилизации числа вершин скелета в зависимости от коэффициента регуляризации. На одном из таких интервалов остается только то множество ребер, которое образует базовый скелет изображения. Значение коэффициента регуляризации в начале этого интервала автор называет оптимальным. Отметим, что в работе не было проведено исследование по определению релевантного интервала стабилизации для каждого класса тестовой выборки, и коэффициент стрижки подбирался вручную.

Аппроксимация. Для решения определенного рода задач (например, задач сравнения изображений [12]) бывает недостаточно выполнить только регуляризацию скелета. Дополнительным механизмом стабилизации скелетного графа является процедура аппроксимации скелета. Идея процедуры аппроксимации для дискретных скелетов описана в [10]. Поясним, что в дискретном скелете вершины имеют либо степень 1, либо степень больше

двух, а ребра, соединяющие вершины, могут иметь произвольную форму, т.е. быть прямыми и кривыми линиями. Аппроксимация представляет собой замену ребра произвольной формы прямыми отрезками либо одним прямым отрезком, основываясь на определении степени кривизны изначального ребра.

В непрерывных скелетах вершины могут иметь любую степень, начиная с 1. Все ребра непрерывного скелета упрощенно будем считать прямыми. Последовательно соединенные между собой ребра (цепочка ребер), все вершины которых, за исключением двух крайних, имеют степень 2, называют ветвью скелета. Аппроксимация представляет собой замену скелетной ветви прямым ребром на основе оценки кривизны ветви. Для того, что избежать неоднозначности понимания процесса аппроксимации непрерывного скелета, приведем подробное описание.

Процесс аппроксимации непрерывного скелета изображен на рис. 2. Дана скелетная ветвь A, B, C, D, E . Вычислен диаметр минимальной описанной около скелета окружности $D_{\min}$ и задана пороговая величина, представляющая собой некоторую часть от $D_{\min}$ и изменяющаяся в диапазоне $[0 \dots 1]$. Назовем эту величину коэффициентом аппроксимации и обозначим K_a . Находим вершину ветви, наиболее удаленную от прямой, проведенной через концы цепочки A и E (вершина C). Сравниваем расстояние CO от вершины C до прямой AE с величиной $K_a * D_{\min}$ (она равна расстоянию от прямой AE до штрихпунктирной линии на рис. 2). Если расстояние $CO > K_a * D_{\min}$, то вершина C остается в списке вершин скелета и делит таким образом ветвь A, B, C, D, E на две ветви. Далее процедура аппроксимации повторяется для каждой из получившихся ветвей. Если же расстояние $CO \leq K_a * D_{\min}$, то вся ветвь заменяется одним прямым ребром с концами в точках A и E , и все вершины ветви за исключением концевых удаляются из скелета. Таким образом проводится аппроксимация всех ветвей скелетного графа.

Отметим, что в [10] коэффициент аппроксимации выбирался приблизительно, априорно.

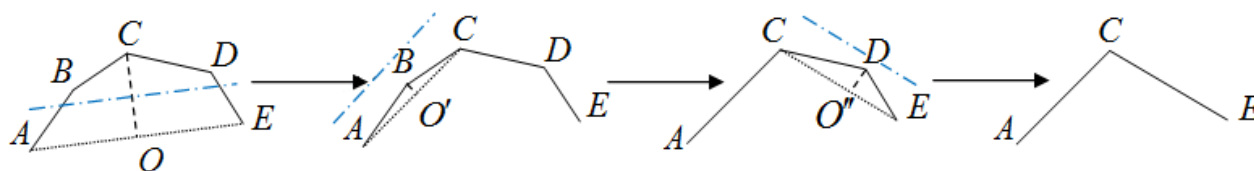


Рис. 2: Определение степени кривизны ветви при аппроксимации скелета

Склейка. Идею склейки внутренних вершин скелета можно кратко описать следующим образом: каждое ребро скелета с нетерминальными вершинами, такое, что расстояние между вершинами достаточно мало, считается связующим и заменяется одной внутренней вершиной — узлом. Положение нового узла рассчитывают так, что восстановленный по полученному скелету силуэт отличается от исходной фигуры не более чем на заданную величину — параметр стабилизации скелета [13], который в данной работе рассматриваться не будет.

На рис. 3 показан процесс получения устойчивого скелета для бинарного изображения.

Таким образом, задача данной работы — экспериментальное исследование влияния коэффициентов регуляризации и аппроксимации на топологию скелета и выработка рекомендаций по определению адекватных значений.

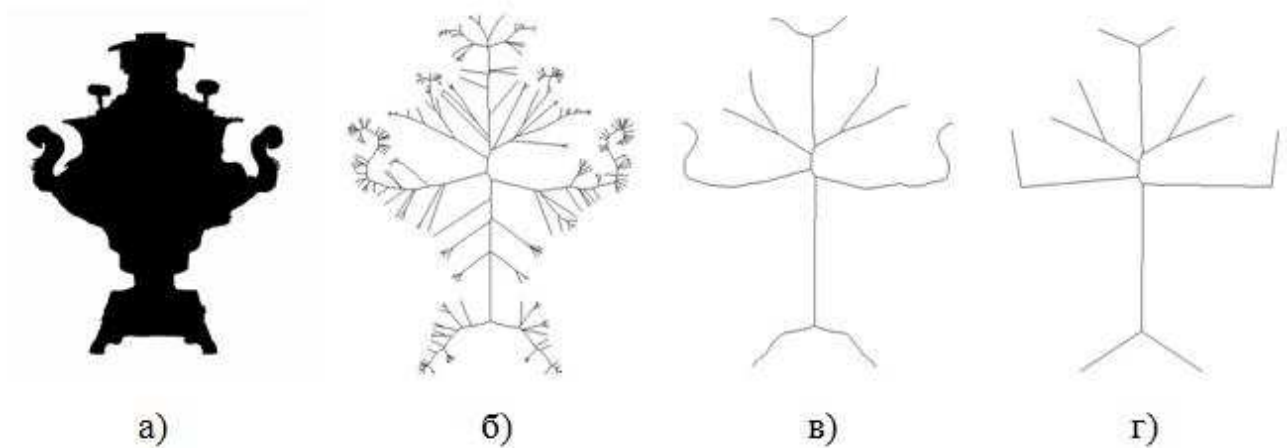


Рис. 3: Исходное изображение (а), его скелет (б), скелет после регуляризации (в), регуляризованный скелет после аппроксимации (г)

Диаметр минимальной описанной вокруг скелета окружности как масштабный множитель в задаче аппроксимации

Как указывалось выше, коэффициент аппроксимации задает, какую величину относительно диаметра минимальной описанной вокруг скелета окружности $D_{\min}$ взять в качестве порогового значения при определении степени изгиба ветви. Минимальная описанная вокруг скелета окружность (также известная как минимальная покрывающая окружность) представляет собой окружность минимального диаметра, содержащую все вершины заданного исходного множества в евклидовом пространстве [14]. Вычисление диаметра минимальной описанной вокруг скелета окружности является нетривиальной задачей, некоторые из известных решений которой представлены далее.

1. Метод ограничивающей сферы Риттера. Его идея состоит в следующем: случайным образом выбирается вершина x из исходного множества P и наиболее удаленная от нее вершина y . Затем находится наиболее удаленная от вершины y вершина z , и отрезок, соединяющий эти вершины, объявляется начальным искомым диаметром сферы с центром в середине данного отрезка. Далее, если нет вершины p , лежащей вне ограничивающей сферы, сфера на текущем шаге принимается за искомую. Иначе она расширяется так, чтобы охватить сферу на предыдущем шаге и вершину p . Процесс продолжается, пока не останется ни одной вершины снаружи сферы. Сложность данного метода составляет $O(nd)$ (где n — число исходных вершин, d — число измерений), что довольно эффективно. Но в то же время, метод дает на 5%~20% большее значение относительно оптимума [15].
2. Метод скачущего пузыря. При обнаружении очередной вершины p вне ограничивающей сферы (пузыря) центр сферы сдвигается в направлении p и одновременно увеличивается радиус r . Причем рост r осуществляется таким образом, чтобы исключить возможность превышения оптимального значения. Сложность данного метода составляет $O(nd) + O(d/\varepsilon^2)$ (где n — число исходных вершин, d — число измерений, ε — точность вычисления результата), при этом обеспечивается высокоточный результат. Наиболее простой вариант данного алгоритма выполняется за $O(nd)$ и дает на 1%~2% большее значение относительно оптимума [16].

Несмотря на привлекательность второго из описанных подходов, в нашем распоряжении имеется мало информации касательно конкретной реализации алгоритма. Поэтому для решения задачи вычисления диаметра минимальной описанной вокруг скелета окружности предлагаются следующие методы:

1. Замена значения диаметра минимальной описанной вокруг скелета окружности $D_{\min}$ значением расстояния между двумя наиболее удаленными вершинами графа $d_{\max}$. Наиболее удаленные вершины находятся с помощью перебора по всем вершинам с исключением просмотренных из дальнейшего рассмотрения. Сложность метода — $O(n^2)$. Очевидно, что диаметр окружности $D_{\min}$ всегда будет больше либо равен $d_{\max}$, поскольку любое расстояние внутри окружности меньше или равно ее диаметру. Проиллюстрируем это рис. 4, где показан скелетный граф с вершинами A, B, C, D, E . Максимальное расстояние $d_{\max}$ между вершинами графа A и D , отмеченное красным штрихпунктиром, меньше диаметра $D_{\min}$, обозначенного синим пунктиром. Если бы в графе существовала вершина A' , то максимальное расстояние $d_{\max}$ было бы равно $D_{\min}$.

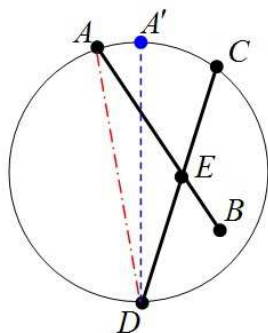


Рис. 4: Иллюстрация различия величин минимальной описанной вокруг скелета окружности $D_{\min}$ и расстояния между двумя наиболее удаленными вершинами графа $d_{\max}$

Таким образом, данный метод дает некоторую погрешность при определении масштаба фигуры, а также менее эффективен, но прост в реализации и может использоваться как временная замена другим методам.

2. Метод последовательных приближений для вычисления минимальной описанной вокруг скелета окружности $D_{\min}$. Идея вычисления диаметра по данному методу: находится выпуклый многоугольник, содержащий все точки скелета, после чего итерационно вычисляются координаты центра и радиус описываемой окружности, пока не будет достигнута некоторая заданная точность ε .

Алгоритм вычисления диаметра методом последовательных приближений

- (a) По методу Джарвиса [17] определяется последовательность вершин V_{hull} скелета, образующих выпуклую оболочку многоугольника, содержащего все вершины V этого скелета, а именно:
 - i. Проходом по всем вершинам определяется номер левой нижней вершины v_1 множества V .
 - ii. v_1 берется в качестве начальной вершины, добавляется во множество V_{hull} и исключается из множества V .
 - iii. Текущей вершиной v_c устанавливается v_1 .
 - iv. В качестве следующей вершины v_n выбирается вершина, не входящая в оболочку.

- v. Далее осуществляется проход по оставшимся вершинам v_i множества V : вычисляется псевдоскалярное произведение m_{ps} векторов, построенных на вершинах v и v_n , а также на v и v_i . В отличие от скалярного произведения знак псевдоскалярного произведения позволяет отследить, положительный или отрицательный угол составляют друг с другом вектора на плоскости. В случае если заданы 2 вектора $a = (a_1, a_2)$ и $b = (b_1, b_2)$, их псевдоскалярное произведение равно определителю матрицы $\begin{pmatrix} a_1 & a_2 \\ b_1 & b_2 \end{pmatrix}$. Когда m_{ps} становится меньше или равно 0, вершиной v_n устанавливается v_i .
- vi. Вершина v_n добавляется во множество V_{hull} и исключается из множества V .
- vii. Текущей вершиной v_c устанавливается v_n .
- viii. Если v_c — это v_1 , то выпуклая оболочка V_{hull} с n вершинами найдена, иначе переход на шаг iv.
- (б) Если оболочка состоит из 1 вершины, то диаметр равен 0.
- (в) Иначе выбирается любая вершина v_{sa} оболочки V_{hull} и перебором с вычислительной сложностью $O(n)$ определяется максимально удаленная от нее вершина v_{sb} .
- (г) Аналогичным перебором определяется максимально удаленная от v_{sb} вершина v_{sc} и находится значение расстояния между v_{sb} и v_{sc} — d_{max} .
- (д) Центром p_0 искомой окружности устанавливается середина отрезка, соединяющего v_{sb} и v_{sc} .
- (е) Начальный шаг $step$ устанавливается $0, 2 * d_{\text{max}}$.
- (ж) Перебором среди вершин оболочки V_{hull} находится вершина $p_{l_{\text{max}0}}$, наиболее удаленная от p_0 , а расстояние до нее становится текущим радиусом искомой окружности r_0 .
- (з) Далее координаты центра искомой окружности определяются по формуле: $p_{i+1} = p_i + v * (step / ||v||)$, где p_i — текущий найденный центр искомой окружности, v — вектор с началом в вершине p_i и концом в наиболее удаленной от нее вершине $p_{l_{\text{max}i}}$. Таким образом, p_i сдвигается в направлении вершины $p_{l_{\text{max}i}}$ на расстояние $step$.
- (и) Перебором среди вершин оболочки V_{hull} находится вершина $p_{l_{\text{max}(i+1)}}$, наиболее удаленная от p_{i+1} , и расстояние до нее r_{i+1} .
- (к) Если $r_{i+1} \geq r_i$, то шаг $step$ уменьшается в 2 раза.
- (л) Если $step < \varepsilon$, то $r_{i+1} * 2$ — искомый минимальный диаметр, а p_{i+1} — искомый центр описанной окружности, иначе переход к п. (з).

Вычислительная сложность данного метода $O(nd/\varepsilon)$ при рассмотрении всех вершин скелета и $O(hn) + O(hd/\varepsilon)$ — при использовании алгоритма Джарвиса для построения выпуклой оболочки V_{hull} (где n — число исходных вершин, h — число вершин в оболочке и $h \ll n$, d — число измерений, ε — точность вычисления результата).

3. Метод описания данных опорными векторами (Support Vector Data Description — SVDD) с линейным ядром [18] позволяет построить гиперсферу (на плоскости — окружность) вокруг объектов (вершин скелета). Вычислительная сложность $O(n^3)$. Второй и, особенно, третий методы более трудоемки, чем первый, но дают более точные значения диаметра минимальной описанной вокруг скелета окружности, при этом метод последовательных приближений представляет компромисс между точностью результата и скоростью выполнения.

Экспериментальные исследования параметров

Параметр регуляризации. С целью определения зависимости числа ребер скелетного графа от значения коэффициента регуляризации строились графики для скелетов

изображений разных классов. Варьируемые значения коэффициента регуляризации менялись в интервале от 1 до 100 с шагом 1. Выбор верхнего значения коэффициента объясняется тем предположением, что для изображений с разрешением 600×800 евклидово расстояние между самым значительным шумовым выбросом на границе фигуры и незашумленной границей не будет превышать 100.

Изображения двух классов фигур, взятых для примера, представлены на рис. 5.



Рис. 5: Изображения двух классов фигур для проведения экспериментов с параметрами регуляризации и аппроксимации

На рис. 6 приводятся графики зависимости числа ребер скелетного графа от значения коэффициента регуляризации (он отвечает за удаление шумовых вершин из скелета). Числа в «легенде» диаграммы показывают значение исследуемой величины (количество ребер) для соответствующих графиков при коэффициенте регуляризации, равном 100. Рисунок демонстрирует, что вид графиков для изображений одного класса практически одинаков и графики изображений одного класса располагаются ближе друг к другу, чем графики другого. Предполагаем, что данное наблюдение можно использовать в задачах классификации изображений с небольшим количеством хорошо отделимых друг от друга классов. Также на рисунке наблюдаются интервалы стабилизации числа ребер, характерные для каждого из классов.

Про поведение графиков в общем можно сказать, что они показывают стремительное уменьшение числа ребер при малых значениях параметра регуляризации. При дальнейшем его увеличении изменение количества уже не столь существенно.

Отдельный интерес представляет собой зависимость параметра регуляризации от масштаба изображения (величины диаметра описанной около скелета окружности). Наблюдения показывают, что для одинаковых изображений разной величины значение адекватного коэффициента регуляризации пропорционально масштабу изображения, что представляется логичным, исходя из смысла коэффициента. Поэтому при решении задач классификации изображений будет разумно предварительно масштабировать все скелеты изображений, которые надо сравнивать между собой, приводя диаметры минимальных описанных около них окружностей к некоторой взятой за единицу величине или к значению диаметра минимальной описанной около скелета окружности одного из изображений.

Параметр аппроксимации. На рис. 7 представлены графики зависимости числа ребер скелетного графа от значения коэффициента аппроксимации. В «легенде» диаграммы приведены значения исследуемой величины (количество ребер) для соответствующих графиков при коэффициенте аппроксимации, равном 0,03. Видно, что количество ребер в скелете при изменении коэффициента аппроксимации достаточно быстро становится постоянным. Также стоит отметить, что и здесь графики для изображений одного класса изменяются примерно одинаково.

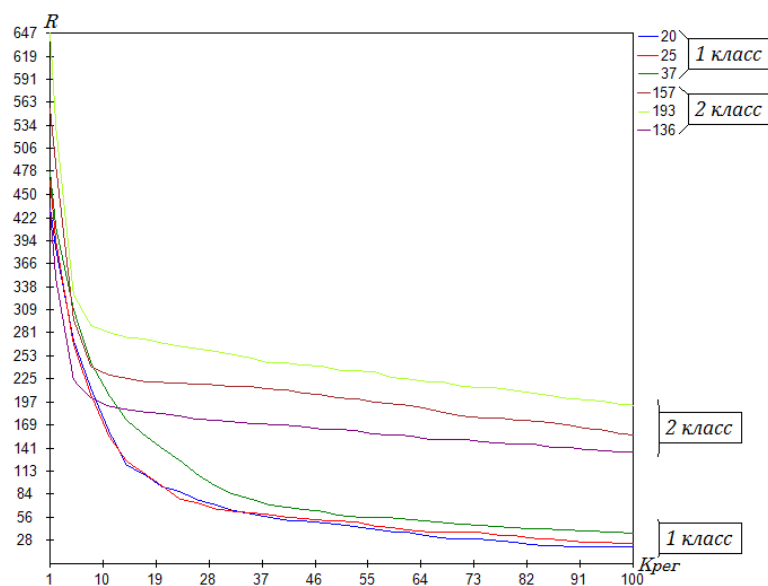


Рис. 6: Графики зависимости числа ребер скелетного графа от значения параметра регуляризации

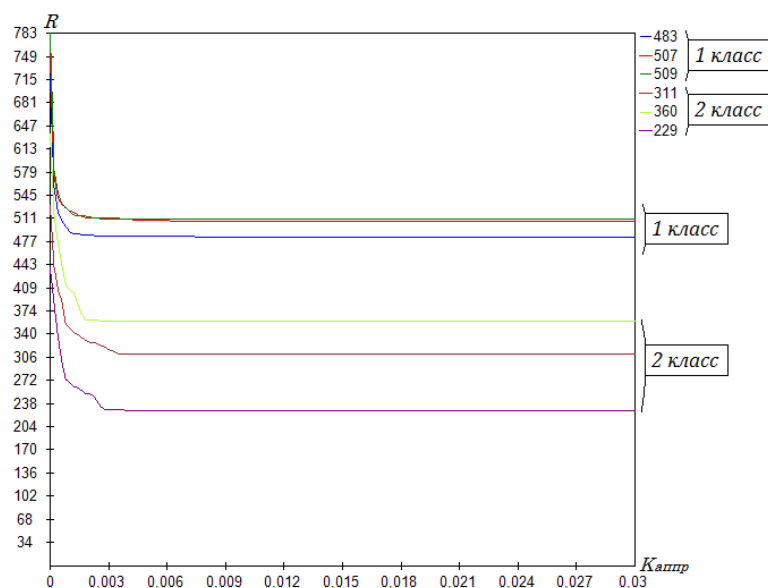


Рис. 7: Графики зависимости числа ребер скелетного графа от значения параметра аппроксимации

Совместный анализ параметров. Для исследования зависимости количества ребер от двух коэффициентов при их одновременном изменении был разработан специальный программный комплекс, позволяющий в интерактивном режиме наблюдать результат стабилизации скелета при заданных параметрах аппроксимации и регуляризации (рис. 8).

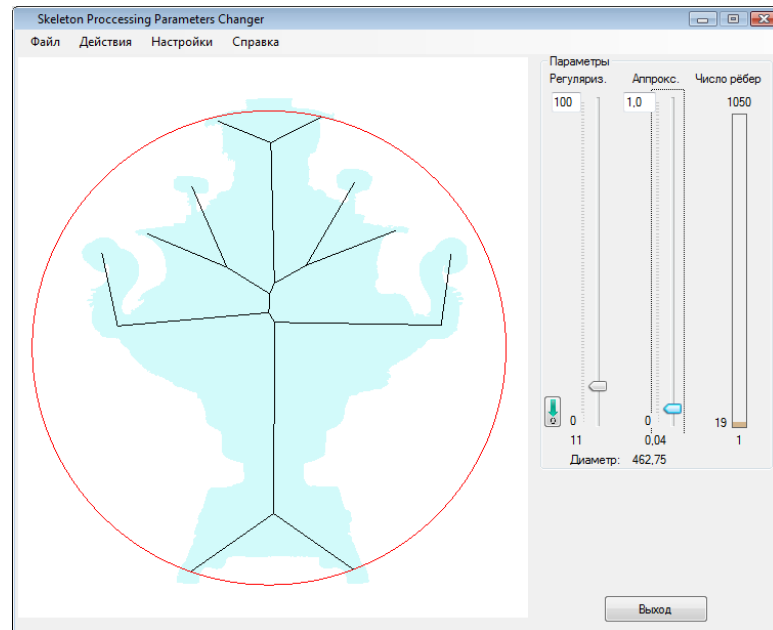


Рис. 8: Интерфейс модуля программного комплекса для исследования параметров стабилизации скелета

Для тестового изображения был построен график зависимости числа ребер скелетного графа от значений коэффициентов аппроксимации и регуляризации (рис. 9). Наши наблюдения показали, что для задач распознавания изображений по их скелетным графам предпочтительно выбирать значения параметров, обеспечивающих минимальное число ребер, достаточное для стабильного описания скелета. Это соответствует внутреннему «плато» на профиле изменения количества ребер. Из графика хорошо видно, что одно и то же число ребер обеспечивается достаточно широким диапазоном варьируемых параметров, которое можно назвать, по терминологии, принятой в [11], интервалом стабилизации. Из всего интервала мы рекомендуем выбирать наименьшие по величине значения параметров.

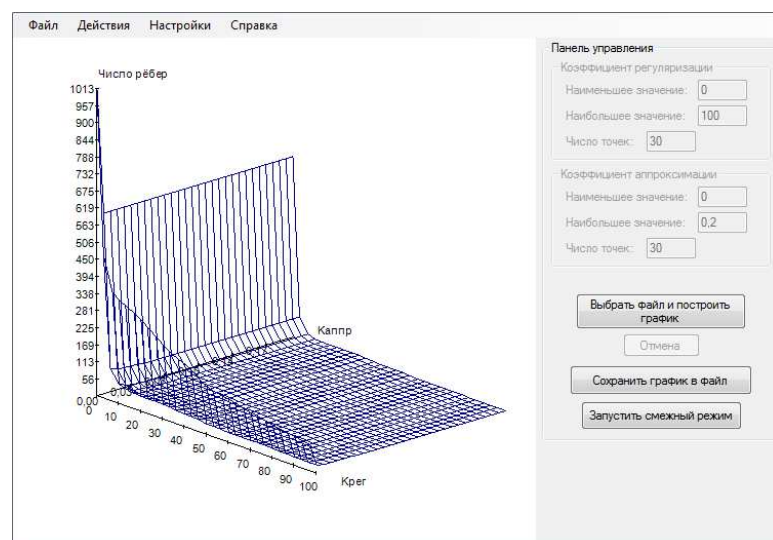


Рис. 9: График зависимости числа ребер скелетного графа от значений коэффициентов аппроксимации и регуляризации

Выводы

В работе сделаны первые шаги к теоретическому исследованию и выработке алгоритма по определению адекватных значений параметров регуляризации и аппроксимации скелетных графов бинарных изображений. В дальнейшем планируется построить полностью автоматическую версию алгоритма и применить ее для задач классификации на основе парного выравнивания цепочек скелетных примитивов [12]. Кроме того, в простейших задачах с небольшим количеством классов, хорошо отличающихся по сложности скелета, сами подобранные значения коэффициентов могут играть роль дескрипторов.

Литература

- [1] *Blum H.* A transformation for extracting new descriptors of shape // *Models for the Perception of Speech and Visual Form*. MIT Press, 1967. Pp. 362–380.
- [2] *Choi S.W., Lee S.W.* Stability analysis of medial axis transform under relative Hausdorff distance // *15th International Conference on Pattern Recognition Proceedings*. IEEE, 2000. Vol. 3. Pp. 135–138.
- [3] *Домахина Л. Г.* Регуляризация скелета для задачи сравнения формы // *Математические методы распознавания образов: 14-я Всероссийская конференция. Сб. докладов*. Владимирская обл., Суздаль, 21–26 сентября 2009 г. М.: МАКС Пресс, 2009. С. 342–345.
- [4] *Shaked D., Bruckstein A. M.* Pruning medial axes // *Computer Vision and Image Understanding*, 1998. Vol. 69, no. 2. Pp. 156–169.
- [5] *Местецкий Л. М.* Непрерывная морфология бинарных изображений: фигуры, скелеты, циркуляры. М.: ФИЗМАЛИТ, 2009. 288 с.
- [6] *Ogniewicz, R.* Automatic medial axis pruning by mapping characteristics of boundaries evolving under the euclidean geometric heat flow onto voronoi skeletons. Harvard Robotics Laboratory Technical Report, 1995. Pp. 95–114.
- [7] *Attali D., di Baja G. S., and Thiel E.* Skeleton simplification through non significant branch removal // *Image Processing and Communications 3.3–4*, 1997. Pp. 63–72.
- [8] *Местецкий Л. М., Рейер И. А.* Непрерывная гранично-скелетная модель дискретного изображения с контролируемой точностью аппроксимации // *Доклады XI Всероссийской конференции «Математические методы распознавания образов» (ММРО-11)*. Москва, 2003. С. 367–371.
- [9] *Жукова К. В., Рейер И. А.* Параметрическое семейство базовых скелетов многоугольной фигуры // *J. Machine Learning and Data Analysis*, 2012. Т. 1. С. 4.
- [10] *Рогов А. А., Быстров М. Ю.* Структурное распознавание бинарных изображений с использованием скелетов // *Математические методы распознавания образов: 15-я Всероссийская конференция. Сб. докладов*. Петрозаводск, 11–17 сентября 2011 г. М.: МАКС Пресс, 2011. С. 420–423.
- [11] *Макарова Е. Ю.* Классификация лекарственных растений по форме листа на основе скелетного представления // *Математические методы распознавания образов: 15-я Всероссийская конференция. Сб. докладов*. Петрозаводск, 11–17 сентября 2011 г. М.: МАКС Пресс, 2011. С. 412–415.
- [12] *Кушнир О. А., Середин О. С.* Функция парного сравнения скелетных графов, заданных цепочками примитивов. // *Известия ТулГУ. Сер. Технические науки*. Тула: Издательство ТулГУ, 2013. Вып. 2. С. 197–207.
- [13] *Домахина Л. Г., Охлопков А. Д.* Изоморфные скелеты растровых изображений // *Тр. 18-й межд. конференции ГРАФИКОН*, 2008.
- [14] *Elzinga J., Hearn D. W.* The minimum covering sphere problem // *Management Science*, 1972. Vol. 19, Issue 1. Pp. 96–104.

- [15] *Ritter J.* An efficient bounding sphere // Graphicsgems / Ed. Andrew S. Glassner. Boston, MA: Academic Press, 1990. Pp. 301–303.
- [16] *Tian B.* Bouncing Bubble: A fast algorithm for Minimal Enclosing Ball problem // *Scholarly Essay*, 2012. 21 p.
- [17] *Кормен Т., Лейзерсон Ч., Ривест Р., Штайн К.* Алгоритмы. Построение и анализ. М.: Вильямс, 2005. 1296 с.
- [18] *Tax D.* One-class classification; Concept-learning in the absence of counterexamples. Ph.D thesis. Delft University of Technology, ASCI Dissertation Series, 2001. 146 p.

Метод обнаружения автомобилей на аэрокосмических снимках*

И. А. Матвеев¹, А. Н. Трекин²

matveev@ccas.ru

¹ВЦ РАН; ²МФТИ

Представлен алгоритм обнаружения автомобилей на цветных изображениях, полученных аэрофотосъемкой. Подходом к решению является то, что большинство автомобилей окрашено в один цвет и имеет сходные размеры. Это позволяет выделять автомобили на фоне как области с цветовыми характеристиками, отличающимися от фоновых и имеющие определенные размер и форму. Области строятся путем иерархического объединения более мелких на основании сходства цветовых и пространственных характеристик. Проведено тестирование алгоритма на наборе изображений, содержащих в общей сложности 2226 автомобилей.

Ключевые слова: сегментация изображений; иерархическая кластеризация

Vehicle detection in aerial color images*

I. A. Matveev¹, A. N. Trekin²

¹CC RAS; ²MIPT

The paper presents an algorithm for detecting cars in color images obtained by aerial photography. The key to the solution is that majority of cars are painted in single color and have similar size and shape. This allows to discriminate the areas occupied by vehicles from the background as regions with certain color characteristics and geometrical properties. Regions in image are constructed by hierarchical clustering induced by color likelihood and spatial neighborhood. The algorithm was tested with the set of urban and rural images, which contains totally 2226 vehicles.

Keywords: image segmentation; hierarchical clustering

Введение

Быстрое развитие средств аэрофотосъемки (особенно беспилотных летательных аппаратов), позволяющее оперативно получать значительное количество изображений, равно как и развитие вычислительных средств, дающих возможность обрабатывать эту информацию в реальном или близком к реальному масштабе времени, ставит новые интересные задачи. Одна из таких задач — распознавание объектов автомобильного транспорта на изображениях аэрофотосъемки [1]. Входными данными служат изображения в масштабе 5–10 см/пиксель (так что автомобиль занимает по длине область размером 30–100 пикселей), цветные (3 канала — красный, зеленый, синий), глубина цвета 8 бит. Требуется найти автомобили (их положение, ориентация, размеры) с минимальным количеством ошибок первого и второго рода (пропуск/ложное обнаружение).

Существующие подходы к решению задачи можно разделить на три основных класса: выделение границ областей на изображении и обнаружение характерных контуров машин (см., например, [2, 3]), использование корреляторов в схеме АдаБуст (например, [4]) и вы-

Работа выполнена при финансовой поддержке РФФИ, проект № 12-07-00778.

деление областей с определенными характеристиками (см., например, [5, 6]). Первые два подхода используются на черно-белых изображениях, третий – как правило, на цветных. В представленной работе используется именно третий подход — выделение областей на изображении на основании цветовых характеристик составляющих их точек. В статьях, опирающихся на этот метод, изученных авторами, точки областей интереса (автомобилей) дискриминируются от фона на основании разнообразных эвристик (кластеризация в цветовом пространстве, байесовский подход, работа с последовательностью изображений и др.), после чего из этих точек составляются области. В данной работе используется иерархическое объединение точек в укрупняющиеся области на основании сравнения цветовых и пространственных характеристик этих областей.

Метод решения

Метод основан на выделении областей с близкими цветовыми характеристиками. Такой подход оправдан тем, что большинство автомобилей имеет однотонную окраску и при этом тон этой окраски, как правило, отличается от фонового (дорога, растительность). Основные стадии работы алгоритма таковы:

- выделение тоновых характеристик (исключение влияния затенения на цвет);
- формирование областей одинакового тона;
- выделение однотонных областей — гипотетических автомобилей;
- вычисление признаков областей и отбраковка.

Рассмотрим каждый шаг подробнее.

Выделение тоновых характеристик. Основное модельное ограничение, эксплуатируемое в данном подходе — однотонность окраски автомобиля. Автомобили, раскрашенные в несколько цветов, составляют весьма малую часть (не более 1%) от общего количества, и погрешность (доля пропущенных автомобилей), создаваемая этим ограничением, представляется приемлемой. Видимый цвет окрашенных участков автомобиля задается цветом окраски и затенением (углом падения света на окрашенную поверхность). Бликами пренебрегаем.

Обозначим три (красную, зеленую и синюю) цветовые компоненты исходных точек как R , G , B , составленный из них вектор цвета $\mathbf{C} = (R, G, B)$. Нормализованные компоненты также составляют трехмерный вектор $\mathbf{c} = (r, g, b)$. Для компенсации затенения, т. е. уравнивания цвета одинаково окрашенных, но по-разному освещенных участков, можно использовать несколько подходов. Из всех вариантов выбрана следующая нормировка:

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}, \quad b = \frac{B}{R + G + B}. \quad (1)$$

Запись в векторном виде: $\mathbf{c} = \mathbf{C} / \|\mathbf{C}\|_{L_1}$. Учитывая то, что значения цветовых компонент задаются в диапазоне $[0; 255]$, необходимо также домножить на 255, получая в итоге:

$$\mathbf{c} = \frac{255\mathbf{C}}{\|\mathbf{C}\|_{L_1}}. \quad (2)$$

Можно заметить, что нормированные компоненты, в отличие от исходных, линейно зависимы: $r + g + b \equiv 1$. Таким образом проявляется исключение яркости. На рис. 1 приведен пример исходного изображения (часть исходного изображения аэрофотосъемки).

Дальнейшая работа происходит в нормированных цветах.

Формирование областей одинакового тона. Задача состоит в выделении областей одинакового цветового тона, т. е. в объединении соседних точек, имеющих одинаковый или



Рис. 1: Пример изображения

близкий цветовой тон, в области. Подход к решению — последовательное слияние близких по цветовым характеристикам областей. Сначала опишем алгоритм в целом.

Область W может включать в себя единственную точку изображения $W = \{(x, y)\}$ или группу точек $W = \{(x, y)_1, \dots, (x, y)_k\}$. Для каждой области заданы ее *соседи* (такие же области) $N(W) = \{W_1, \dots, W_s\}$, определяемые по принципу *4-связности*. Область может иметь от одного до неопределенно большого (ограниченного сверху количеством точек изображения) числа соседей. Вырожденный случай — отсутствие соседей — достигается если все изображение объединено в одну область. В каждой точке (x, y) изображения три компоненты яркости $\mathbf{c}(x, y) = (r, g, b)$ назовем *цветовой характеристикой* этой точки. Аналогично цветовые характеристики области $\mathbf{c}(W)$ — такие же, как у точки три цветовых компонента, являющиеся усреднением по всем точкам, включенным в область $\mathbf{c}(W) = \langle \mathbf{c}(x, y) \rangle_{(x, y) \in W}$. Для каждой области в процессе построения определяется общая яркость $\mathbf{B}(W) = \sum_{(x, y) \in W} \mathbf{C}(x, y)$ по трем компонентам включенных в нее то-

чек и их количество $M(W) = |W|$, что позволяет быстро вычислять среднюю яркость: $\mathbf{c}(W) = \mathbf{B}(W)/M(W)$. Области могут быть сравнены по их цветовым характеристикам (а также по размерам). Если характеристики близки, т. е. различие между характеристиками множеств меньше некоторого порога $\rho(W_1, W_2) < T$, и области являются соседями $N(W_1) \ni W_2$ (т. е. хотя бы одна точка одной 4-связна с хотя бы одной точкой другой), то области объединяются $W_3 = W_1 \cup W_2 = \{(x, y) : (x, y) \in W_1 \text{ or } (x, y) \in W_2\}$. Каждое *объединение* производится для двух и только двух областей W_1 и W_2 , в результате появляется новая (*родительская*, хотя данный термин не совсем удачен) область W_3 , которая имеет тех же соседей, что и исходные области $N(W_3) = (N(W_1) \cup N(W_2)) \setminus W_1 \setminus W_2$, а ее яркости и число точек являются суммами соответствующих характеристик исходных областей $\mathbf{B}(W_3) = \mathbf{B}(W_1) + \mathbf{B}(W_2)$, $M(W_3) = M(W_1) + M(W_2)$. Такая новая область делается

активной (т. е. может участвовать в дальнейших объединениях), а две объединенные в нее области становятся пассивными и в дальнейших объединениях не участвуют. На каждом шаге t работы алгоритма существует упорядоченный *список активных областей* $L_t = \{W_{t,1}, \dots, W_{t,s}\}$, т. е. таких, которые еще не слились ни с какими другими и являются отдельными.

В начальный момент при инициализации список активных объектов заполняется точками изображения $L_0 = \{W_{0,1}, \dots, W_{0,XY}\}$, $W_{0,s} = \{(x, y)_s\}$, $M_{0,s} \equiv 1$, $\mathbf{B}(W_{0,s}) = \mathbf{c}((x, y)_s)$, $N(W_{0,s}) = \{W_{0,s-X}, W_{0,s-1}, W_{0,s+1}, W_{0,s+X}\}$. Последняя запись означает, что соседями каждого элементарного множества становятся четыре других элементарных множества, полученных из соответствующих точек раstra. На каждом шаге алгоритма из списка активных областей выбирается очередная и производится анализ возможности объединения с каждым из ее соседей. Если такое объединение возможно, оно происходит. В каждом *проходе* алгоритма просматривается список всех активных областей. Один проход алгоритма объединения (т. е. анализ всех активных областей, имеющихся в момент начала прохода, но не новых, появившихся во время прохода) называется также *поколением*. Алгоритм заканчивает работу, когда при очередном проходе не произведено ни одного объединения, т. е. никакие две активные области не могут более слиться при заданных требованиях на слияние.

Требования на слияние при разных проходах могут различаться. Список активных областей организован, как очередь FIFO (первый вошел, первый вышел), новые активные области добавляются в конец. Таким образом, после n -го прохода в результате произведенных объединений список активных областей удлиняется и $n + 1$ -й проход производится просто как продолжение сканирования списка активных областей, без возврата в начало списка. В конце концов случается так, что при очередном проходе не происходит ни одного слияния и алгоритм должен был бы завершить работу. Однако при этом остаются активные области, которые были пропущены при первом сканировании (если активная область была пропущена, то она более не рассматривается, поскольку возврата в начало списка не производится). Эти активные области могут быть объединены (с учетом того, что правила объединения могут изменяться в разных проходах). Поэтому имеет смысл произвести сканирование списка активных областей с самого начала. Такое сканирование (последовательность проходов или поколений) начиная с первой активной области до тех пор, пока очередной проход не будет безрезультатен, называется *эрой*. Окончательно алгоритм прекращает работу, когда в результате прохождения эры не было зафиксировано ни одного слияния (в этом случае эра вырождена в один проход, также безрезультатный).

Представим работу алгоритма в виде блок-схемы (рис. 2).

Оценим алгоритмическую сложность. Рассмотрим идеальный случай изображения, заполненного единственным цветом (которое будет объединено в единственную область). Пусть изображение имеет размер N пикселей. В первом проходе они объединятся в $N/2$ пар (краевыми эффектами пренебрегаем), во втором — в $N/4$ четверок и так далее. Всего может быть осуществлено $N - 1$ операций объединения. Очевидно, что число объединений равно разности числа точек изображения и числа полученных в конце работы областей. Значит, для изображений более сложной структуры объединений будет меньше. Таким образом, по объединениям сложность алгоритма линейная. Сложность по просмотрам. В самом худшем случае при каждом просмотре будет происходить лишь одно объединение, и каждый следующий просмотр будет анализировать лишь на одну пару меньше, и число сравнений областей будет пропорциональным квадрату числа точек изображения. Таким образом, сложность по просмотрам — квадратичная. Однако для реальных изобра-

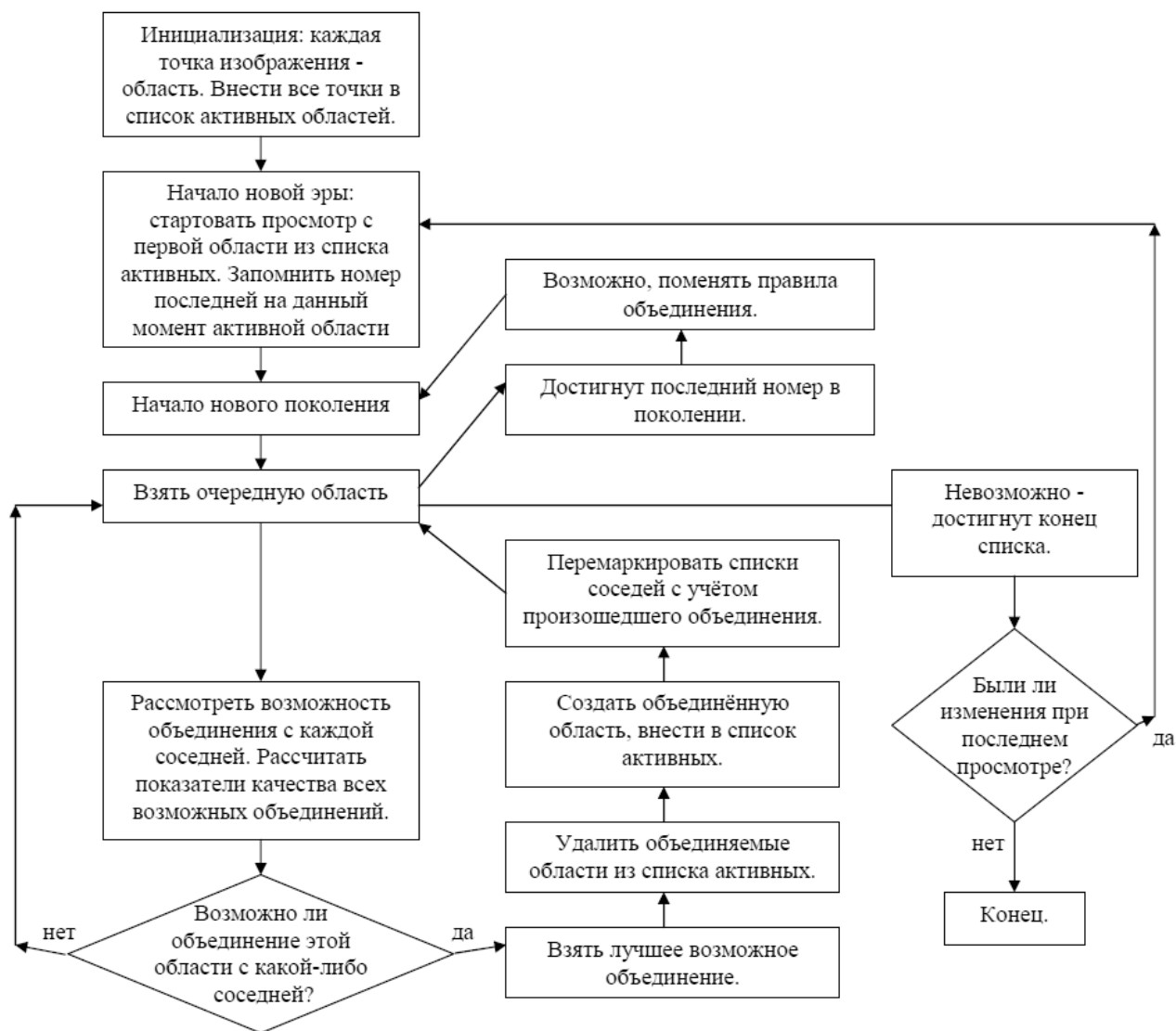


Рис. 2: Блок-схема алгоритма объединения областей

жений с большими областями приблизительно равных оттенков, которые объединятся в первых же поколениях, квадратичное возрастание сложности произойдет лишь на окончательных стадиях работы метода, когда число активных областей будет очень невелико по сравнению с их начальным количеством. Общее число операций возрастет незначительно.

На рис. 3 дан пример работы алгоритма на фрагменте изображения рис.1. Видно, что из присутствующих на фрагменте два автомобиля выделены в виде областей, но одна из этих областей включает в себя шумовой выброс.

Выделение однотонных областей — гипотетических автомобилей

Автомобили выделяются, прежде всего, как области, имеющие определенную площадь, т.е. содержащие определенное число точек. Учитывая известный масштаб изображения можно выбрать пороги количества точек в объектах, соответствующие размеру автомобиля на изображении. В представленных расчетах были использованы пороги $500 < M < 5000$, где M — число точек объекта. На рис. 4а изображены области изображения рис. 1, попадающие в это ограничение. Видно, что из присутствующих на исходном изображении восьми автомобилей восемь остались в виде областей примерно совпадаю-

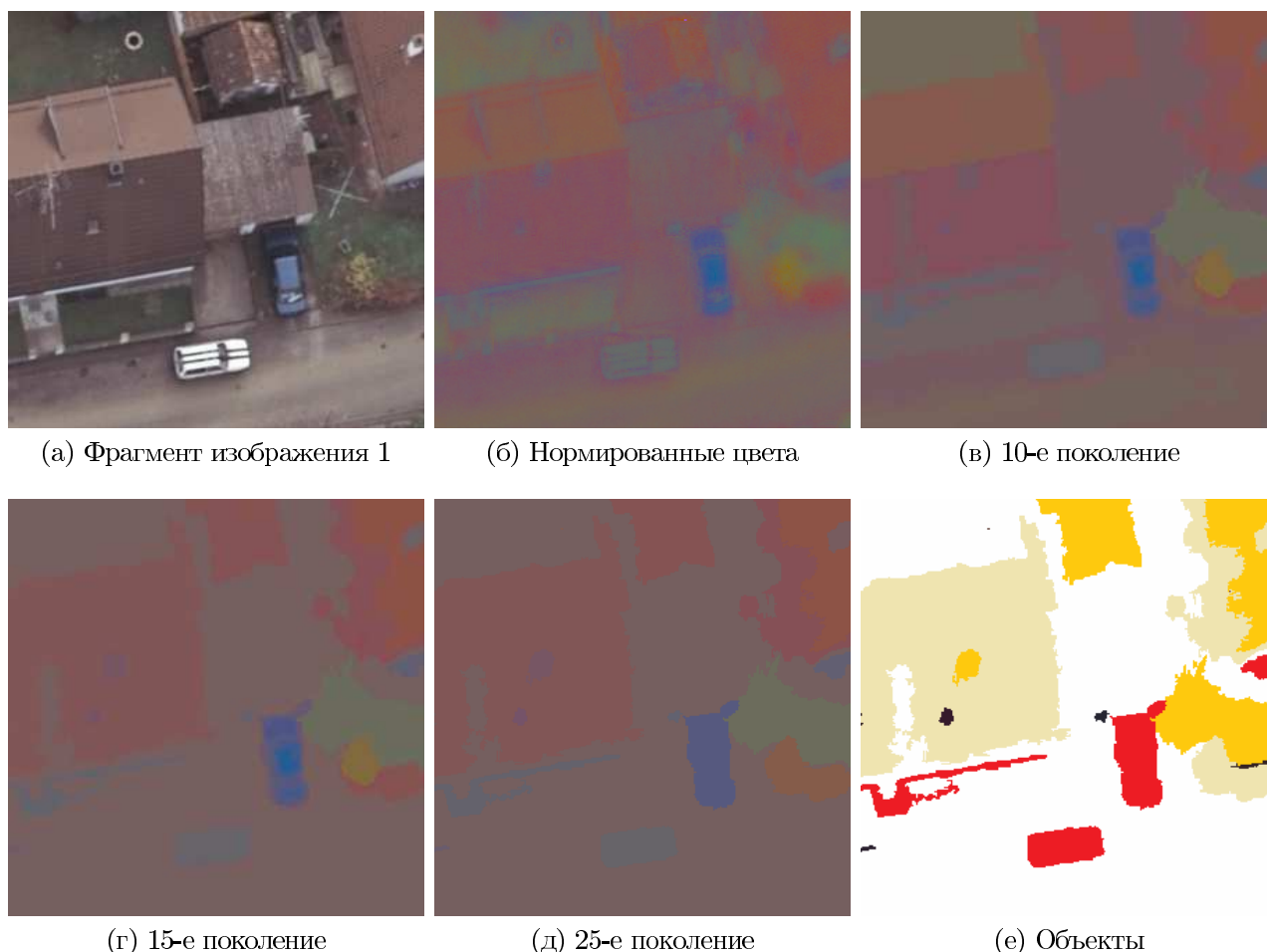


Рис. 3: Работа алгоритма объединения

щих размеров, а один распался на две области. Видно, что наряду с автомобилями имеется достаточно много других областей. Основная их доля отсекается на следующем шаге при анализе формы.

Вычисление признаков областей и отбраковка

Прежде всего, изображение автомобиля — компактная область точек приблизительно прямоугольной формы с отношением большой и малой сторон, отличающимся от единицы, но не слишком большим. Исходя из этого, можно воспользоваться моделью эквивалентного эллипса. Требования к эллипсу таковы:

- отношение большой и малой полуосей: $1,3 < a/b < 3$ (машина не должна быть квадратной и не должна быть очень длинной);
- $\pi ab < 2M$ (площадь эквивалентного эллипса инерции не должна намного превосходить площадь области, т. е. область должна быть «сплошной», не содержать дыр и длинных выступов).

На рис. 4 справа представлены области, прошедшие отбраковку по эллиптичности. Видно, что семь из восьми машин, имеющихся на рис. 1 выделены, восьмая потеряна. Кроме того, имеется шесть лишних объектов (не машин).

Экспериментальная проверка

Были проведены тесты на наборе 46 изображений городской и сельской местности, содержащих в общей сложности 2226 автомобилей. На этих изображениях было обнаружено



(а) области изображения, представленного на рис.1, (б) области, удовлетворяющие ограничению по форме, удовлетворяющие ограничениям по площади (числу ме; темные — соответствуют автомобилям, светлые — ошибки ложного обнаружения)

Рис. 4: Области, выделенные кластеризацией

1613 автомобилей (72,5%, т.е. ошибка первого рода составляет 17,5%) и было допущено 2452 ложных срабатывания. Эти характеристики не являются удовлетворительными, алгоритм нуждается в дальнейшей доработке.

Литература

- [1] Бондур В. Г., Старченко С. А. Методы и программы обработки и классификации аэрокосмических изображений // *Изв. ВУЗов. Геодезия и аэрофотосъемка*. 2001. № 3. С.118–143.
- [2] Zhao T., Nevatia J. Car Detection in Low Resolution Aerial Image // *Proc. IEEE ICCV*, 2001. Vol. 1. Pp.710–717.
- [3] Sotelo M. A., Nuevo J., Bergasa L. M. et al. Road Vehicle Recognition in Monocular Images // *Proc. IEEE ISIE*, 2005. Vol.4. Pp.1471–1476.
- [4] Negri P., Clady X., Hanif S. M. et al. A Cascade of Boosted Generative and Discriminative Classifiers for Vehicle Detection // *EURASIP J. Advances in Signal Processing*, 2008. Article ID 782432.
- [5] Tsai L.-W., Hsieh J.-W., Fan K.-C. Vehicle Detection Using Normalized Color and Edge Map // *IEEE Trans. Image Processing*, 2007. Vol. 16, no. 3. Pp.850–864.
- [6] Mejia-Inigo R., Barilla-Perez M. E., Montes-Venegas H. A. Color-based Texture Image Segmentation for Vehicle Detection // www.intechopen.com/download/pdf/15388.

Комбинирование одноклассовых классификаторов в задачах сегментации изображений, содержащих объекты топологии плат микросборок*

А. О. Ларин¹, О. С. Середин², В. В. Кондрашов²

ekzebox@gmail.com

¹Московский физико-технический институт; ²Тульский государственный университет

Решение задачи автоматизации установок лазерной подгонки резистивных элементов требует предварительной сегментации элементов топологической структуры платы микросборки на основе изображения, полученного с видеокамеры. Ввиду специфики работы системы, применение стандартных методов сегментации изображения может оказаться затруднительным. В данной работе для решения этой задачи предлагается использовать результаты распознавания одноклассового классификатора Тэкса на основе признаков LBP.

Ключевые слова: сегментация изображений, топология плат микросборок, распознавание образов, одноклассовая классификация, метод описания данных опорными векторами, локальные бинарные шаблоны.

Combining of one-class classifiers for segmentation images with microassembly layout objects*

A. O. Larin¹, O. S. Seredin², V. V. Kondrashov²

¹Moscow Institute of Physics and Technology; ²Tula State University

Automation solution for laser trimming of resistive element facilities requires presegmentation of microassembly topology elements based on image obtained from a video camera. Because the specifics of the system, the application of standard methods of image segmentation can be difficult. In this paper, to solve this problem, the authors proposed an approach based on using Support Vector Data Description method and LBP (local binary patterns) features.

Keywords: image segmentation, microassembly topology, pattern recognition, one-class classification, support vector data description, local binary patterns.

Введение

Автоматизация работы установок лазерной подгонки резистивных элементов [1] является необходимым средством повышения точности доводки резисторов до требуемого номинала, снижения уровня брака, сокращения временных затрат операторов и их утомляемости и, как следствие, снижения себестоимости изделий. Такая автоматизация требует внедрения механизмов анализа и принятия решений с минимальным участием оператора либо полностью в автоматическом режиме. Для построения системы адаптивного управления процессом лазерной подгонки с учетом получаемой в режиме реального времени информации о ходе технологического процесса необходимо решение довольно большого спектра подзадач из разных областей знаний.

Одной из основных этапов такого технологического процесса является автоматическое детектирование объектов топологии плат микросборок в режиме реального времени,

Работа выполнена при финансовой поддержке РФФИ, проект №12-07-92000-ННС_а, 14-07-00527.

закключающееся в выделении границ основных элементов, частности проводников и резисторов, на изображениях, получаемых по видеоканалу (рис. 1).

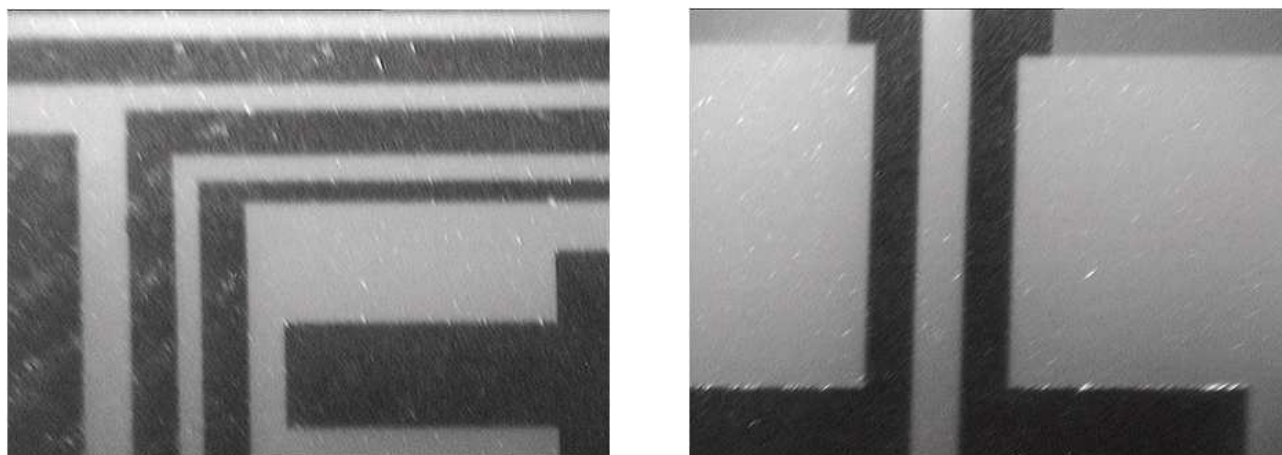


Рис. 1: Исходные изображения участков плат микросборок

Применение стандартных алгоритмов машинного зрения может быть затруднено спецификой изображений различных плат с топологическим рисунком на установках с современными оптическими системами:

- видеоизображение постоянно меняется за счет перемещения области обзора камеры;
- значительная зависимость от условий освещения (неоднородность подсветки всей платы, несоответствие длин волн подсветки и полосы пропускания светофильтров, зеркал гальваносканеров объективов и т. д.);
- случайные помехи;
- видимая неоднородность структуры подложки и элементов микросборки;
- искажение изображения (абберации, изменение геометрии и т. д.).

Предлагаемая методика сегментации изображения

В данной работе мы предлагаем альтернативный подход сегментации изображений, основанный на комбинировании одноклассовых классификаторов. Идея анализа заключается в следующем. Элементы на платах микросборок (далее — платах) в силу своей геометрической структуры представлены границами 4-х типов: левой, верхней, правой и нижней. В микросборках диагональные элементы используются значительно реже, чем в обычных печатных платах. Это связано с тем, что платы микроэлектроники в основном используются для мощных низкочастотных устройств либо для СВЧ-схем, где часто проявляется эффект «гашения СВЧ-волны», чему сопутствует значительное выделение тепла. Использование острых углов в соединениях проводников в таких случаях настоятельно не рекомендуется отраслевыми ГОСТами и РД [2]. Таким образом, используя комбинацию 4-х одноклассовых классификаторов, каждый из которых будет обучен для детектирования определенного типа границы, представляется возможным выделить на изображении платы большую часть областей, принадлежащих контурам элементов.

Применив к полученной в результате информации, например, метод связанных компонент, возможно получить полную интерпретацию фрагмента чертежа платы. Следует также отметить, что результатом исходной классификации является не только факт обнару-

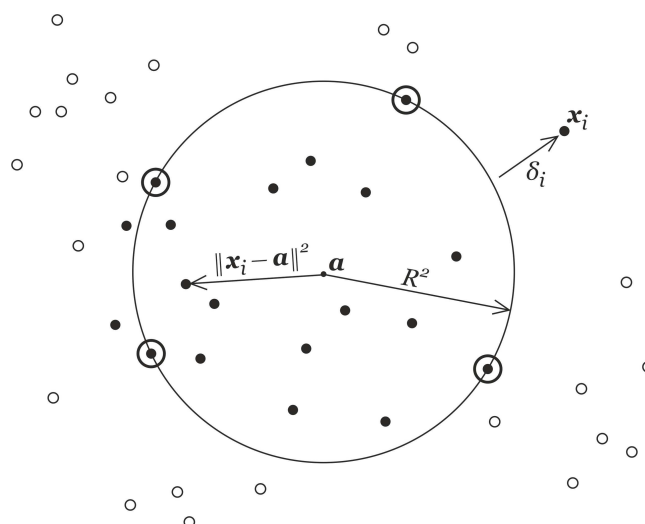


Рис. 2: Сферическая модель описания данных метода SVDD

жения границы на изображении, но и информация о его типе, что в перспективе предоставляет дополнительные возможности для улучшения метода, обобщающего результат, полученный после классификации.

Метод описания данных опорными векторами

В качестве одноклассового классификатора предлагается использовать метод описания данных опорными векторами (Support Vector Data Description, SVDD) [3], построенный на аналогии с SVM-методом В. Вапника.

Моделью описания исходного набора данных $\mathbf{x}_i \in \mathbb{R}^n$, $i = 1, \dots, N$ в этом методе является гиперсфера, представляющая ближайшую внешнюю границу вокруг данных, заданная следующими параметрами: центром $\mathbf{a} \in \mathbb{R}^n$ и радиусом $R \in \mathbb{R}$. Задающие параметры подбираются так, чтобы большая часть объектов обучающей совокупности не выходила за пределы гиперсферы, а объекты, оказавшиеся за ее границей, были «оштрафованы» (см. рис. 2). При этом радиус гиперсферы с учетом полученного «штрафа» должен быть минимальным.

Таким образом, необходимо минимизировать структурную ошибку модели:

$$\begin{cases} R^2 + C \sum_{i=1}^N \delta_i \rightarrow \min_{R, \mathbf{a}, \delta}, \\ \|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2 + \delta_i, \delta_i \geq 0, i = 1, \dots, N. \end{cases} \quad (1)$$

Двойственная задача по отношению к (1) имеет вид:

$$\begin{cases} \sum_{i=1}^N \lambda_i (\mathbf{x}_i \cdot \mathbf{x}_i) - \sum_{i=1}^N \sum_{j=1}^N \lambda_i \lambda_j (\mathbf{x}_i \cdot \mathbf{x}_j) \rightarrow \max_{\lambda}, \\ \sum_{i=1}^N \lambda_i = 1, 0 \leq \lambda_i \leq C, i = 1, \dots, N, \end{cases} \quad (2)$$

где λ – множители Лагранжа.

Объекты, лежащие на границе гиперсферы (множители Лагранжа которых принадлежат интервалу $(0, C)$), так называют опорными, и именно эти объекты используются для описания всей обучающей совокупности.

Новый объект считается принадлежащим классу интереса, если его расстояние до центра гиперсферы меньше ее радиуса. Из этого следует, что функция одноклассового решающего правила распознавания нового объекта $\mathbf{z}$ будет иметь вид индикаторной функции:

$$d(\mathbf{z}; \boldsymbol{\lambda}, R) = I(\|\mathbf{z} - \mathbf{a}\|^2 \leq R^2), \quad (3)$$

где

$$\|\mathbf{z} - \mathbf{a}\|^2 = (\mathbf{z} \cdot \mathbf{z}) - 2 \sum_{i=1}^{N_{SV}} \lambda_i (\mathbf{z} \cdot \mathbf{x}_i) + \sum_{i=1}^{N_{SV}} \sum_{j=1}^{N_{SV}} \lambda_i \lambda_j (\mathbf{x}_i \cdot \mathbf{x}_j), \quad (4)$$

$$R^2 = (\mathbf{x}_k \cdot \mathbf{x}_k) - 2 \sum_{i=1}^{N_{SV}} \lambda_i (\mathbf{x}_k \cdot \mathbf{x}_i) + \sum_{i=1}^{N_{SV}} \sum_{j=1}^{N_{SV}} \lambda_i \lambda_j (\mathbf{x}_i \cdot \mathbf{x}_j), \quad (5)$$

где N_{SV} – количество опорных объектов, а $\mathbf{x}_k, \lambda_k \in (0, C)$ – любой из опорных объектов.

Для возможности описания данных более «гибкой формой», нежели сфера, необходимо использовать идею метода потенциальных функций [4, 5] для перехода в спрямляющее пространство признаков большей размерности. Наиболее часто используемыми потенциальными функциями являются полиномиальная

$$K(\mathbf{x}_i, \mathbf{x}_j) = (1 + \mathbf{x}_i \cdot \mathbf{x}_j)^p$$

и радиальная базисная функция Гаусса

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|^2}{s^2}\right). \quad (6)$$

Таким образом, чтобы получить улучшенную модель описания данных по методу Тэкса, необходимо заменить в (2), (4) и (5) операцию вычисления скалярного произведения двух векторов вычислением значения потенциальной функции двух аргументов.

Для решения задачи сегментации топологии плат предпочтительной является потенциальная функция Гаусса (6), показавшая наилучшие результаты в предыдущих сравнительных экспериментах.

Подбор оптимального параметра s для потенциальной функции Гаусса

В своей работе Тэкс предложил алгоритм, позволяющий подбирать оптимальное значение параметра s функции Гаусса. Суть алгоритма заключается в поиске параметра s , соответствующего заданной ошибке распознавания, в пределах от минимального значения параметра до максимального. Эти значения могут быть определены по следующим формулам:

$$s_{min} = \min_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|, \quad i \neq j,$$

$$s_{max} = \max_{i,j} \|\mathbf{x}_i - \mathbf{x}_j\|.$$

В данной работе поиск параметра осуществлялся по итерационной схеме с заданным шагом до тех пор, пока ошибка распознавания, полученная при использовании процедуры скользящего контроля, не стала соответствовать желаемой. Значение параметра s в данной

работе совпало с тем, что было получено в задаче цветовой параметризации объектов в пространстве RGB [6].

Пространство признаков LBP

Предварительные эксперименты с использованием параметризации на основе цветового представления в пространстве RGB, описанной в предыдущей работе [6], показали необходимость выбора другого признакового пространства для процедуры классификации. В качестве такого пространства были выбраны широко применяемые для решения задач компьютерного зрения локальные бинарные шаблоны (Local Binary Patterns, LBP) [7], основанные на применении к изображению μ -LBP оператора.

LBP оператор построен на операции сравнения значений яркости пикселя с соседними пикселями и определяется двумя основными параметрами: радиусом соседней области r и количеством соседних пикселей, участвующих в сравнении p , поэтому его общая запись — $LBP_{p,r}$. Результат вычисления LBP оператора, как показано на рис. 3, — десятичное число [8].

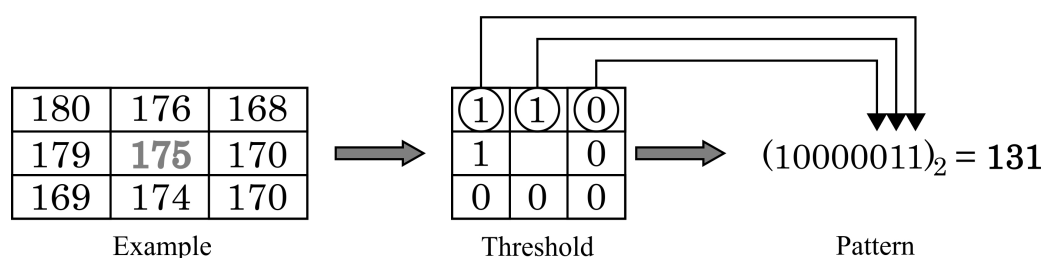


Рис. 3: Пример вычисления оператора $LBP_{8,1}$

Особенности LBP оператора определяют факт того, что размер набора значений, вычисляемых оператором, равен 2^p , что достаточно удобно при использовании малого числа соседних пикселей, но вызывает трудности, когда число p возрастает. Для решения этой проблемы в работе [7] вводится понятие «фундаментальных» LBP шаблонов (некоторые из них представлены на рис. 4) и μ -LBP оператора. Под «фундаментальными» шаблонами понимаются такие паттерны, которые в двоичном представлении содержат не более двух переходов 0–1 и 1–0, а суть μ -LBP заключается как раз в том, что вычисляемое им значение заменяется на некоторое значение по умолчанию, если сам паттерн не является «фундаментальным». Результатом данного нововведения является существенное сокращение возможных чисел, возвращаемых оператором.

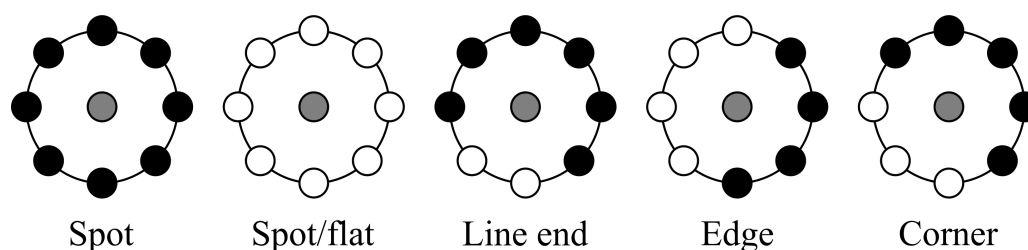


Рис. 4: Фундаментальные паттерны LBP

Признаковое пространство, порожденное применением к изображению LBP-оператора, обладает следующими преимуществами:

— простота вычислений, низкие требования к мощности вычислительной техники;

- устойчивость к изменениям освещенности и общей яркости изображения;
- устойчивость к малым поворотам изображения ($\approx 5^\circ$ – 8°);
- положительные результаты использования признаков LBP в решении различных задач компьютерного зрения.

Существуют две основные стратегии применения LBP-оператора к изображению с целью получения нового признакового пространства. Первая, наиболее простая, — применение LBP оператора с определенными параметрами последовательно к каждому пикселю изображения. Таким образом, исходный массив признаков, представленный значениями яркости в пикселях, заменяется массивом вычисленных значений LBP-оператора. Этот массив и будет новым образом объекта (см. рис. 5). Другой распространенной стратегией является использование частотных LBP-гистограмм, когда новый образ объекта является частотной гистограммой значений LBP-оператора для всех пикселей на изображении. Второй способ показал большую эффективность в задаче распознавания лиц на изображении, поэтому в данной работе мы будем использовать именно его.

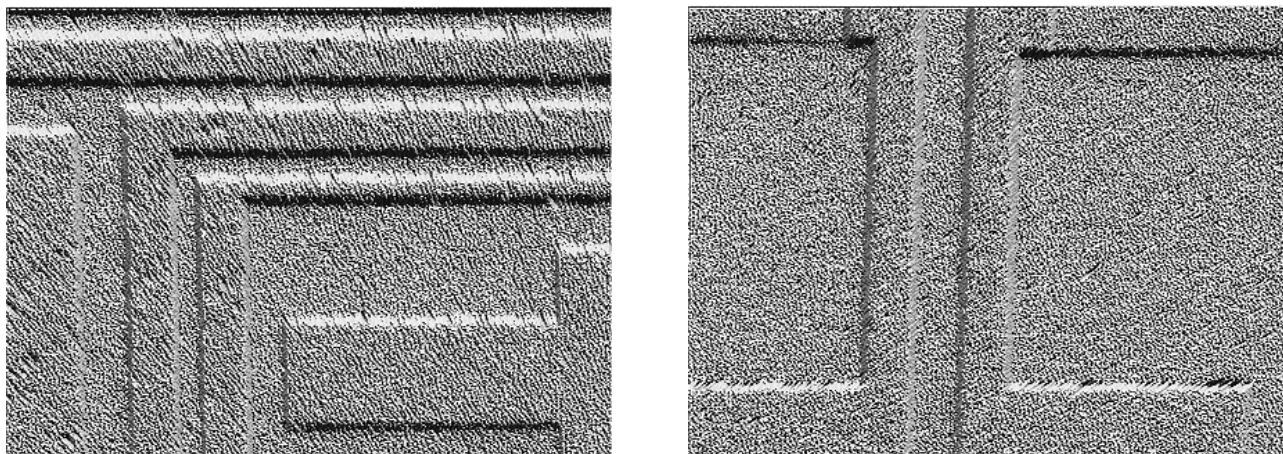


Рис. 5: Результат попиксельного применения оператора $LBP_{8,1}$ к изображениям плат

Экспериментальное исследование

На исходных изображениях были выбраны характерные фрагменты, представляющие все 4 типа границ элементов на плате, примерно по 10 на каждую границу определенного типа (см. рис. 6, 7). Затем на выбранных фрагментах было обучено 4 одноклассовых классификатора Тэкса с заранее выбранными параметрами.

Размер обучающего фрагмента следует выбирать таким образом, чтобы на нем были четко различимы оба типа граничащих материалов и граница между ними, которая, как правило, достаточно сильно размыта. Таким образом, размер обучающего фрагмента зависит от масштаба объектов и разрешения камеры. В нашем случае минимальный допустимый размер обучающего фрагмента равен 10×10 пикселей.

Методика классификации:

- каждому пикселю на всем анализируемом изображении ставились в соответствие 4 счетчика, изначально равные нулю;
- последовательно по всем позициям изображения перемещалось «окно» размера обучающего фрагмента (10×10 пикселей) с некоторым шагом (2 пикселя);

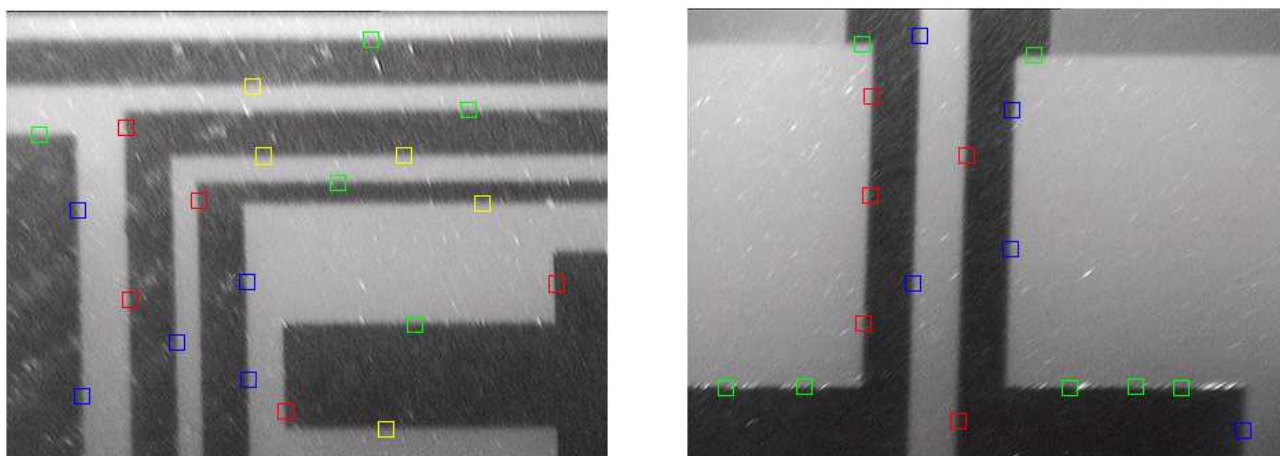


Рис. 6: Примеры выделения фрагментов границ плат для обучения

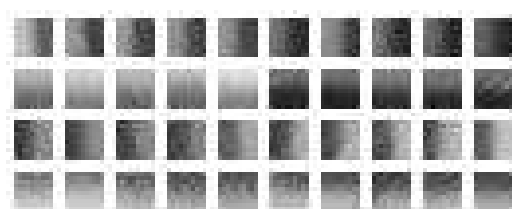


Рис. 7: Примеры обучающих фрагментов границ плат

- для пикселей внутри окна вычислялась частотная LBP-гистограмма;
- получившаяся гистограмма передавалась последовательно на вход 4-м классификаторам;
- в случае, если классификатор опознавал гистограмму, соответствующий счетчик для каждого пикселя внутри исходного фрагмента увеличивался на 1.

Пиксель исходного изображения был отнесен к некоторому типу границы, если величина соответствующего счетчика превышала половину от максимально возможного значения, а остальные счетчики были равны нулю.

Результаты распознавания границ проводников на плате представлены на рис. 8.

Заключение

Предложен альтернативный метод сегментации изображений, основанный на комбинировании одноклассовых классификаторов, а также приведены предварительные результаты экспериментов. Опираясь на полученные результаты (рис. 8), мы планируем продолжить исследование предложенного подхода. Дальнейшие исследования будут направлены на сравнение существующих методов сегментации с методом, основанным на комбинировании одноклассовых классификаторов, которые будут проводиться на предварительно размеченной экспертом базе тестовых изображений плат.

Также необходимы дополнительные исследования, направленные на регуляризацию (стабилизацию) получающихся результатов. В настоящий момент классификация пикселей происходит без учета их совместного расположения.

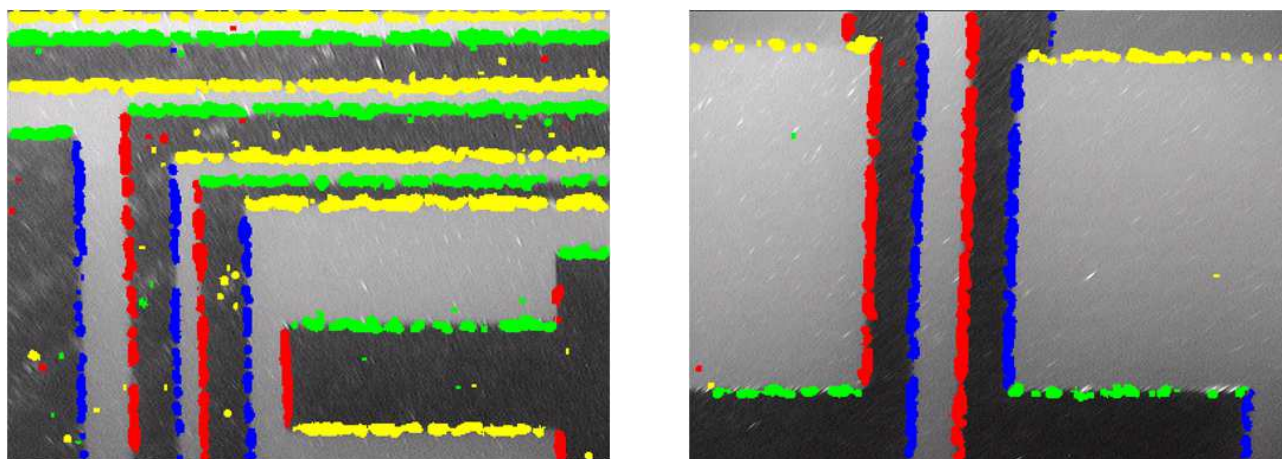


Рис. 8: Результаты классификации пикселей, принадлежащих границам 4-х типов на плате

Литература

- [1] Кондрашов В. В., Копылов А. В., Середин О. С. Автоматизация процесса лазерной подгонки пленочных резистивных элементов // *Известия ТулГУ. Технические науки*. Тула: Издательство ТулГУ, 2013. № 2. С. 146–159.
- [2] Коледов Л. А. Технология и конструкции микросхем, микропроцессоров и микросборок. М.: Радио и связь, 1989. 400 с.
- [3] Tax D. One-class classification; Concept-learning in the absence of counterexamples // Ph.D Thesis. Delft University of Technology, ASCI Dissertation Series. 2001. 146 p.
- [4] Vapnik V. Statistical Learning Theory. N.Y.: J. Wiley, 1998. 768 p.
- [5] Айзерман М., Браверман Э., Розоноэр Л. Метод потенциальных функций в теории обучения машин. М.: Наука, 1970. 386 с.
- [6] Ларин А. О., Середин О. С. Модели представления объектов в пространстве цветовых компонент при анализе изображений // *Известия ТулГУ. Технические науки*. Тула: Издательство ТулГУ, 2013. № 2. С. 207–216.
- [7] Ahonen T., Hadid A., Pietikainen M. Face recognition with local binary patterns // *Proc. of 9th Euro15 We*, 2004. Pp. 469–481.
- [8] Hadid A., Pietikainen M., Ahonen T. A Discriminative feature space for detecting and recognizing faces // *Machine Vision Group Infotech Oulu and Dept. of Electrical and Information Engineering*. 2005. 8 p.

Статистически эффективная схема распараллеливания алгоритмов дуализации*

Е. В. Дюкова¹, А. Г. Никифоров², П. А. Прокофьев¹

edjukova@mail.ru, ankifor@gmail.com, p_prok@mail.ru

¹ВЦ РАН, ²ВМиК МГУ

Одной из центральных задач дискретной математики является дуализация, которая может быть сформулирована как построение всех неприводимых покрытий булевой матрицы. Это задача является вычислительно сложной. В данной работе предлагается эффективная в типичном случае схема распараллеливания асимптотически оптимальных алгоритмов дуализации. Предлагаемая схема основана на статистическом анализе множества неприводимых покрытий булевой матрицы.

Ключевые слова: перечислительная задача; дуализация; неприводимое покрытие булевой матрицы; трансверсаль гиперграфа; эффективность в среднем; асимптотически оптимальный алгоритм; параллельные вычисления; балансировка нагрузки; сильная масштабируемость.

Statistically Efficient Parallel Scheme for Dualization Algorithms*

E. Djukova¹, A. Nikiforov², P. Prokofyev¹

Moscow, ¹Dorodnitsyn Computing Centre of Russian Academy of Sciences, ²Moscow State University

Background: Dualization is a fundamental problem in discrete mathematics. It is equivalent to irreducible coverings enumeration of a given boolean matrix. This problem is of high computational complexity. Therefore, using of parallel computing is practical. However, existing approaches to dualization algorithms parallelizing are trivial and not much efficient.

Methods: An efficient on average parallel scheme ($\mathfrak{B}$ -scheme) for asymptotically optimal dualization algorithms is suggested. The proposed scheme is based on statistical analysis of irreducible coverings set.

Results: The experiments show that $\mathfrak{B}$ -scheme provides a balanced load of processors on the average and that gained speedup of parallel dualization algorithm is almost the highest possible.

Concluding Remarks: $\mathfrak{B}$ -scheme outperforms other considered approaches to dualization algorithms parallelization. However, the number of processors used by $\mathfrak{B}$ -scheme should be rather small; otherwise, no performance improvement is gained. Getting rid of processors number limitation and expanding the range of approach applicability will be a topic of future works.

Keywords: enumeration; dualization; irreducible covering of a boolean matrix; hypergraph transversal; on the average efficiency; asymptotically optimal algorithm; parallel computations; load balancing; strong scaling.

Введение

Пусть L — булева матрица размера m на n . Набор столбцов H матрицы L называется покрытием, если в подматрице L^H матрицы L , образованной столбцами из H , нет строки,

Работа частично поддержана грантами РФФИ №13-01-00787-а и №14-07-00819-а и грантом президента РФ НШ-4908.2014.1.

состоящей из одних нулевых элементов. Покрытие называется неприводимым, если любое его собственное подмножество не является покрытием. Обозначим через $P(L)$ множество всех неприводимых покрытий матрицы L .

Задача построения $P(L)$ называется дуализацией и является одной из центральных перечислительных задач дискретной математики. Дуализация возникает, в частности, в задачах логического анализа данных, например при разработке логических распознающих алгоритмов (тестовых алгоритмов, алгоритмов голосования по представительным наборам, по антипредставительным наборам, по покрытиям класса).

Существуют другие формулировки рассматриваемой задачи, в частности использующие понятия теории графов и теории булевых функций. Приведем эти формулировки.

1. Пусть дан гиперграф G . Требуется построить множество всех минимальных трансверселей этого гиперграфа.
2. Пусть монотонная булева функция f задана конъюнктивной нормальной формой, не содержащей отрицания переменных. Требуется построить сокращенную дизъюнктивную нормальную форму функции f .

Эффективность алгоритмов перечисления принято оценивать сложностью шага, т. е. сложностью построения одного решения [1]. Говорят, что алгоритм работает с полиномиальной задержкой, если каждый его шаг выполняется за полиномиальное время. Алгоритм с полиномиальной задержкой для задачи дуализации до сих пор не построен. В [2] предложен подход к построению асимптотически оптимальных алгоритмов дуализации, который получил развитие в работах [3, 4]. Асимптотически оптимальные алгоритмы делают полиномиальные шаги, но некоторые из этих шагов являются «лишними». Шаг считается «лишним», если построен набор столбцов, не являющийся неприводимым покрытием, или построен набор столбцов, найденный ранее. При этом доля «лишних» шагов стремится к нулю для почти всех матриц данного размера при увеличении размера матрицы.

Время работы алгоритмов дуализации очень быстро растет с ростом размера входных данных, поэтому актуальным является использование параллельных вычислений. Существуют простые и очевидные схемы распараллеливания алгоритмов, решающих указанную задачу. В данной работе показано, что главный недостаток таких схем — неравномерная загрузка процессоров, что является причиной недостаточного ускорения времени работы параллельного алгоритма по сравнению с его однопоточной версией. Поэтому важной является задача создания более эффективных схем распараллеливания.

Ниже предлагается эффективная схема распараллеливания асимптотически оптимальных алгоритмов ($\mathfrak{B}$ -схема), которая основана на статистическом анализе множества неприводимых покрытий булевой матрицы.

Положим $J_n = \{1, \dots, n\}$. Через $P_j(L)$, $j \in J_n$, обозначим множество неприводимых покрытий булевой матрицы L , для которых j является наименьшим номером столбца. Очевидно, что $\{P_j(L) \mid j \in J_n\}$ является разбиением множества $P(L)$ на непересекающиеся подмножества. Предлагаемый подход к построению параллельного алгоритма дуализации базируется на использовании статистических свойств множеств $P_j(L)$, $j \in J_n$, для случайных матриц.

В $\mathfrak{B}$ -схеме каждое из множеств $P_1(L), P_2(L), \dots, P_n(L)$ строится одним из p процессоров, при этом каждый процессор может построить несколько множеств. Установлено, что схема является эффективной при сравнительно небольшом числе процессоров относительно числа столбцов матрицы L .

Работа этой схемы проиллюстрирована на примере асимптотически оптимального алгоритма АО2 [3], который на каждом шаге строит неприводимое покрытие. Однако каждое покрытие из $P(L)$ может быть построено неоднократно. Проверка, было ли построено неприводимое покрытие на предыдущих шагах, осуществляется за полиномиальное время. В [3, 4, 5] показана практическая применимость асимптотически оптимальных алгоритмов: при определенных условиях эти алгоритмы работают быстрее других существующих алгоритмов дуализации.

Проводится сравнение $\mathfrak{B}$ -схемы с ранее известной $\mathfrak{U}$ -схемой, в которой каждый из p процессоров обрабатывает одинаковое число столбцов, равное примерно n/p . Эксперименты показывают, что $\mathfrak{B}$ -схема осуществляет более равномерную загрузку процессоров по сравнению с $\mathfrak{U}$ -схемой и обеспечивает лучшее ускорение. Это ускорение близко к максимально возможному.

Алгоритм дуализации АО2

Обозначим через M_{mn} множество булевых матриц размера m на n . Пусть $L = (a_{ij}) \in M_{mn}$. Очевидно, что покрытие H длины r является неприводимым тогда и только тогда, когда в L^H найдется перестановочная подматрица порядка r (матрица $r \times r$, в каждой строке и в каждом столбце которой в точности один единичный элемент). Для перестановочной подматрицы Q с единичными элементами $a_{i_1 j_1}, \dots, a_{i_r j_r}$ введем следующее обозначение: $Q = \{(i_k, j_k) \mid k \in J_r\}$. Перестановочную матрицу Q назовем максимальной, если она не содержится ни в какой другой перестановочной подматрице матрицы L . Максимальная перестановочная подматрица $Q = \{(i_1, j_1), \dots, (i_r, j_r)\}$ называется правильной, если набор столбцов $H(Q)$ с номерами $\{j_1, \dots, j_r\}$ принадлежит $P(L)$.

Основной особенностью алгоритма АО2 является то, что на каждом шаге этого алгоритма строится правильная подматрица Q и осуществляется проверка, не было ли покрытие $H(Q)$ построено на предыдущих шагах.

Будем говорить, что $j \in J_n$ покрывает $i \in J_m$ ($i \in J_m$ покрывает $j \in J_n$), если $a_{ij} = 1$. Обозначим $R \subseteq J_m$ и $C \subseteq J_n$. Паре (R, C) поставим в соответствие подматрицу $L(R, C)$ матрицы L , образованную строками с номерами из R и столбцами с номерами из C . Будем говорить, что $i_2 \in J_m$ охватывает $i_1 \in J_m$ в C , если $a_{i_1 j} \leq a_{i_2 j}, \forall j \in C$ (обозначение $i_1 \preceq_C i_2$).

Алгоритм осуществляет обход дерева решений в глубину. Узлами этого дерева являются тройки (Q, R, C) , где Q — перестановочная подматрица (для висячих вершин подматрица Q является правильной). Корню дерева соответствуют $Q = \emptyset, R = J_m, C = J_n$. На каждом шаге алгоритма строится ветвь дерева решений. Зафиксируем некоторую ветвь длины r . Пусть узлу t -го уровня, $t \in J_r$, соответствуют (Q_t, R_t, C_t) . Назовем правильной подматрицу $Q_r = \{(i_1, j_1), \dots, (i_r, j_r)\}$ верхней, если $\exists t \in J_r$ и $\exists i \in R_t, i > i_r$ такие, что подматрица $Q = Q_r \setminus \{(i_t, j_t)\} \cup \{(i, j_t)\}$ является правильной. Определим предикат $\text{upreg}(Q)$ равный единице тогда и только тогда, когда подматрица Q является верхней. Очевидно, если подматрица Q является верхней, то неприводимое покрытие $H(Q)$ не было построено на предыдущих шагах алгоритма.

Ниже приведено описание алгоритма АО2, которое для удобства разбито на четыре процедуры.

Процедура 1 запускает алгоритм АО2 путем вызова процедуры 2 с параметрами, соответствующими корню дерева решений.

Процедура 2 является рекурсивной процедурой. Она строит поддерево решений с корнем в $(Q \cup \{(i_0, j_0)\}, R', C'), R' \subseteq R, C' \subseteq C$, где R' и C' формируются после вызова проце-

Процедура 1 AO2

Вход: L ;**Выход:** $\emptyset$;1: BuildSubtree($\emptyset, J_m, J_n, L, 0, 0$)

Процедура 2 BuildSubtree

Вход: Q, R, C, L, i_0, j_0 ;**Выход:** $\emptyset$;1: $(R, C) \leftarrow \text{ReduceSubmatrix}(R, C, L, i_0, j_0)$ 2: $C \leftarrow \text{ProcessUnityColumns}(Q, R, C, L)$ 3: **для всех** $j \in C$ 4: **для всех** $i \in \{u \in R \mid a_{ij} = 1\}$ 5: BuildSubtree($Q \cup \{(i, j)\}, R, C, L, i, j$)

Процедура 3 ReduceSubmatrix

Вход: R, C, L, i_0, j_0 ;**Выход:** R, C ;1: **если** $i_0 > 0$ и $j_0 > 0$ **то**2: $C \leftarrow \{j \in C \mid j > j_0\}$ 3: $C \leftarrow \{j \in C \mid a_{i_0 j} = 0\}$ 4: $R \leftarrow \{i \in R \mid a_{i j_0} = 0\}$ 5: $R \leftarrow \{i \in R \mid \nexists u \in R, u > i : a_{ij} = a_{uj}, \forall j \in C\}$ 6: $R \leftarrow \{i \in R \mid \nexists u \in R, u \neq i : u \preceq_C i\}$ 7: $C \leftarrow \{j \in C \mid \exists i \in R : a_{ij} = 1\}$

Процедура 4 ProcessUnityColumns

Вход: Q, R, C, L ;**Выход:** C ;1: **для всех** $j \in \{v \in C \mid a_{ij} = 1, \forall i \in R\}$ 2: **если** $\text{upreg}(Q \cup \{(i, j)\}) = 1$ **то**3: **печать** $H(Q \cup \{(i, j)\})$

дур 3 и 4. На первой итерации алгоритма $i_0 = 0$ и $j_0 = 0$, что соответствует корню дерева решений.

Процедура 3 вызывается из процедуры 2 и формирует упомянутые выше R' и C' путем

- 1) удаления из C покрытых i_0 элементов;
- 2) удаления из R покрытых j_0 элементов;
- 3) удаления из $L(R, C)$ повторов строк;
- 4) удаления из R охватывающих в C элементов;
- 5) удаления из $L(R, C)$ столбцов, состоящих из одних нулей.

Процедура 4 вызывается из процедуры 2 и строит висячие вершины, которые являются дочерними для узла (Q, R, C) . В частности, процедура находит в подматрице $L(R, C)$ столбцы, состоящие из одних единиц, и сопоставляет каждому из них правильную под-

матрицу. Если эта подматрица является верхней, то покрытие, ей соответствующее, распечатывается.

Статистические свойства множеств $P_j(L)$

Зафиксируем размер $m \times n$ булевых матриц. В качестве пространства элементарных событий Ω рассмотрим множество всевозможных пар (L, H) , где $L \in M_{mn}$, $H \in P(L)$.

Введем на Ω случайную величину $\eta(L, H)$, равную наименьшему номеру столбца в H . Обозначим $f(j) = P\{\eta(L, H) = j\}$. Требуется найти функцию $f(j)$.

Случайная величина ξ имеет бета-биномиальное распределение $\mathfrak{B}(\alpha, \beta, r)$, если она принимает значения q из множества $\{0, 1, \dots, r\}$ с вероятностью

$$\varphi(q; \alpha, \beta, r) = \binom{r}{q} \frac{B(q + \alpha, r - q + \beta)}{B(\alpha, \beta)},$$

где B — бета-функция, а $\alpha > 0, \beta > 0$ и $r \in \mathbb{N}$ — параметры распределения. Обозначим

$$\psi_{\alpha\beta}(j) = \varphi(j - 1; \alpha, \beta, n - 1), \quad j \in J_n.$$

Проведем статистический эксперимент. Пусть $\mathbf{x} = (x_1, \dots, x_N)$ — выборка из распределения $f(j)$. Проверим статистическую гипотезу $H_0 = \{f(j) = \psi_{\alpha\beta}(j)\}$ о виде распределения. В случае, когда гипотеза верна, случайная величина $\eta - 1$ имеет распределение $\mathfrak{B}(\alpha, \beta, n - 1)$.

Для проверки гипотезы H_0 используется видоизмененный критерий Хи-квадрат. Статистика $Z(\mathbf{x})$ этого критерия может быть определена как $(1/N) \sum_{j=1}^n (f^*(j) - \psi_{\alpha\beta}(j))^2 / \psi_{\alpha\beta}(j)$, где $f^*(j)$ — доля элементов выборки $\mathbf{x} = (x_1, \dots, x_N)$, равных j . При вычислении $\psi_{\alpha\beta}(j)$ значения параметров α и β принимаются равными оценкам максимального правдоподобия.

Пусть задан уровень значимости γ и пусть $\chi_{\delta, r}^2$ — квантиль уровня δ распределения Хи-квадрат с r степенями свободы. Гипотеза H_0 отвергается тогда и только тогда, когда $Z(\mathbf{x}) > \chi_{1-\gamma, n-3}^2$ (напомним, что по выборке $\mathbf{x}$ оцениваются два параметра функции $\psi_{\alpha\beta}(j)$). Следуя этому критерию, можно ошибочно отклонить гипотезу H_0 , когда она верна, с вероятностью, приближенно равной γ .

Достигнутым уровнем значимости критерия Хи-квадрат называется величина $\gamma^*(\mathbf{x}) = 1 - \chi_{n-3}^2(Z(\mathbf{x}))$, где $\chi_{n-3}^2(x)$ — функция распределения Хи-квадрат с $(n - 3)$ степенями свободы. Близость значения $\gamma^*(\mathbf{x})$ к 0 говорит о том, что гипотезу H_0 вероятнее всего следует отклонить.

Возникает вопрос, как построить выборку $\mathbf{x} = (x_1, \dots, x_N)$ из распределения $f(j)$. Предлагается следующий способ. Построим t случайных матриц из M_{mn} : $L_1, \dots, L_t$. Из каждого множества $P(L_k), k \in J_t$, выберем по u неприводимых покрытий. Возьмем $N = tu$ значений случайной величины η на парах матрица–покрытие и сформируем выборку.

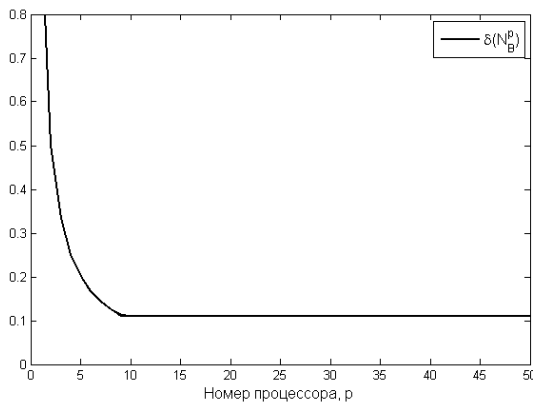
Проведены эксперименты с выборками, построенными по матрицам размера $m \times n$ при $m = 20, 35, 50, 65$ и 80 и $n = 20, 35, 50, 65$ и 80 . Для получения выборки формировалось по 200 случайных матриц одинакового размера и из множества $P(L)$ каждой матрицы L случайно выбиралось по 10 неприводимых покрытий. В результате получалась выборка длиной 2000. В табл. 1 приведены полученные в ходе экспериментов значения достигнутого уровня значимости $\gamma^*(\mathbf{x})$ и оценок максимального правдоподобия α и β .

Таблица 1. Достигнутый уровень значимости критерия и значения оценок максимального правдоподобия параметров, построенных по выборкам для матриц различного размера $m \times n$

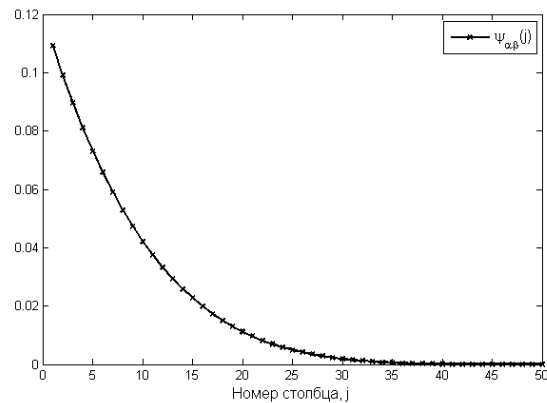
(a) Достигнутый уровень значимости $\gamma^*(x)$					
$m \setminus n$	20	35	50	65	80
20	0	0,106	0,326	0,502	0,751
35	0,001	0,517	0,992	0,842	0,974
50	0	0,919	0,515	0,647	0,983
65	0	0,638	0,719	0,327	0,556
80	0	0,303	0,969	0,79	0,252

(b) Оценка параметра α					
$m \setminus n$	20	35	50	65	80
20	1,193	1,048	0,988	0,994	0,959
35	1,086	1,006	1,005	0,957	0,993
50	1,051	1,049	0,975	1,01	0,993
65	0,994	1,015	0,966	1,001	0,985
80	0,978	0,959	0,996	0,985	1,011

(c) Оценка параметра β					
$m \setminus n$	20	35	50	65	80
20	6,062	5,118	4,762	4,941	4,75
35	6,694	5,901	5,649	5,427	5,51
50	7,283	6,51	6,12	6,26	6,14
65	7,013	6,887	6,432	6,633	6,491
80	7,727	6,887	6,938	6,962	6,69



(a)



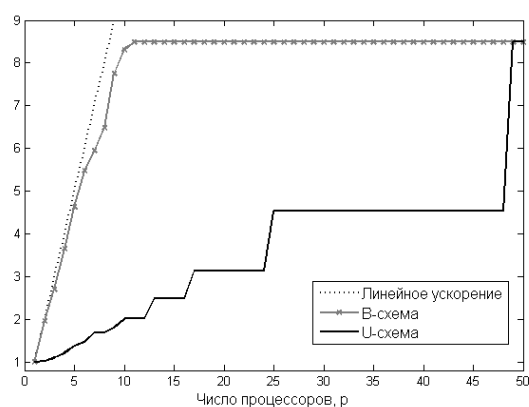
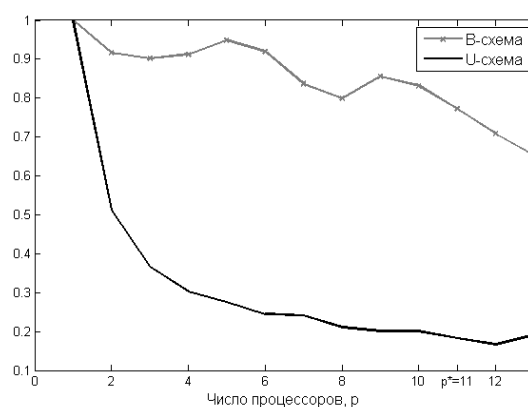
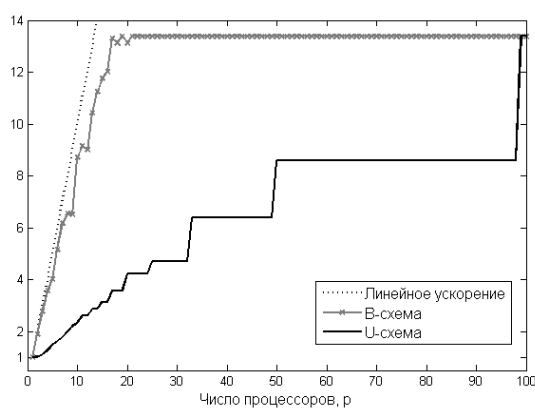
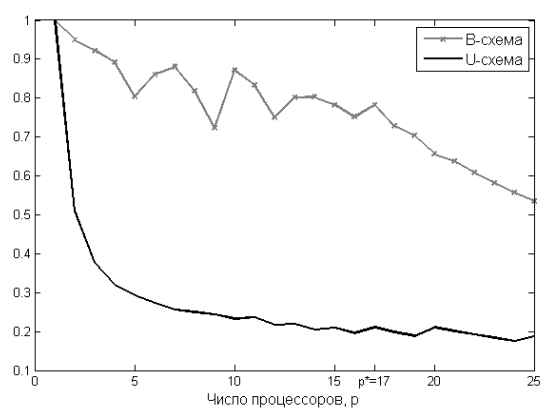
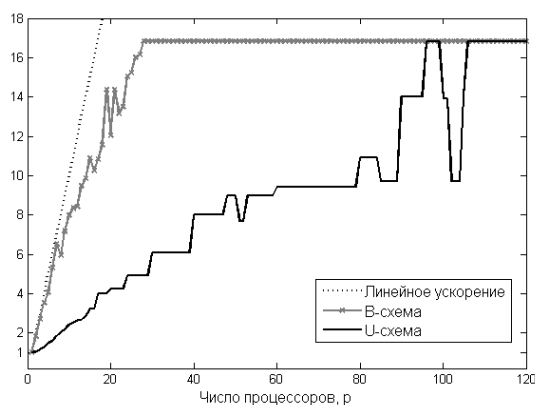
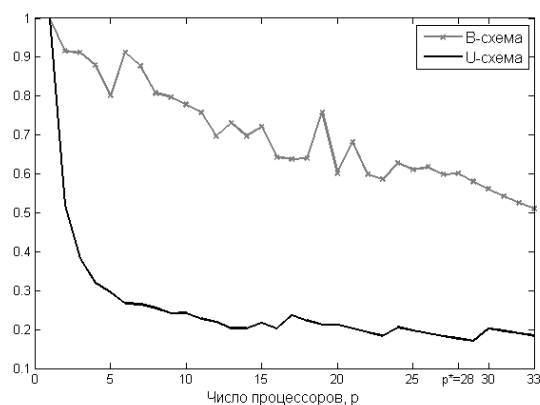
(б)

Рис. 1. Пример зависимостей $\delta(N_B^p)$ от p и $\psi_{\alpha\beta}$ от j

Проведенные эксперименты позволяют сделать следующие выводы. При $n = 20$ гипотезу H_0 следует отклонить (см. табл. 1, а). Возможной причиной этого является нерепрезентативность выборки, связанная с небольшим число неприводимых покрытий у матриц при $n = 20$. Матрицы размера 80×80 , как правило, имеют большое число неприводимых покрытий (порядка 10^7), что также мешает получению репрезентативной выборки ограниченного объема используемым в работе способом. Достигнутый уровень значимости в этом случае достаточно мал. В остальных случаях гипотезу H_0 можно принять с большой уверенностью.

В-схема распараллеливания

В этом разделе показано, как функцию вероятности $f(j)$ случайной величины $\eta(L, H)$ можно применять для эффективного распределения вычислительных ресурсов в начале

(а) $S(p)$ при $m = 50, n = 50$ (б) $E(p)$ при $m = 50, n = 50$ (в) $S(p)$ при $m = 50, n = 100$ (г) $E(p)$ при $m = 50, n = 100$ (д) $S(p)$ при $m = 30, n = 120$ (е) $E(p)$ при $m = 30, n = 120$ **Рис. 2.** Сильная масштабируемость

работы алгоритма дуализации. Пусть $L \in M_{mn}$ и пусть дано $p \leq n$ процессоров. Каждому из множеств $P_j(L), j \in J_n$, сопоставим процессор с номером N_j^p , который должен это множество построить. Вектор $\mathbf{N}^p = (N_1^p, \dots, N_n^p)$ назовем расписанием. Уровнем загрузки

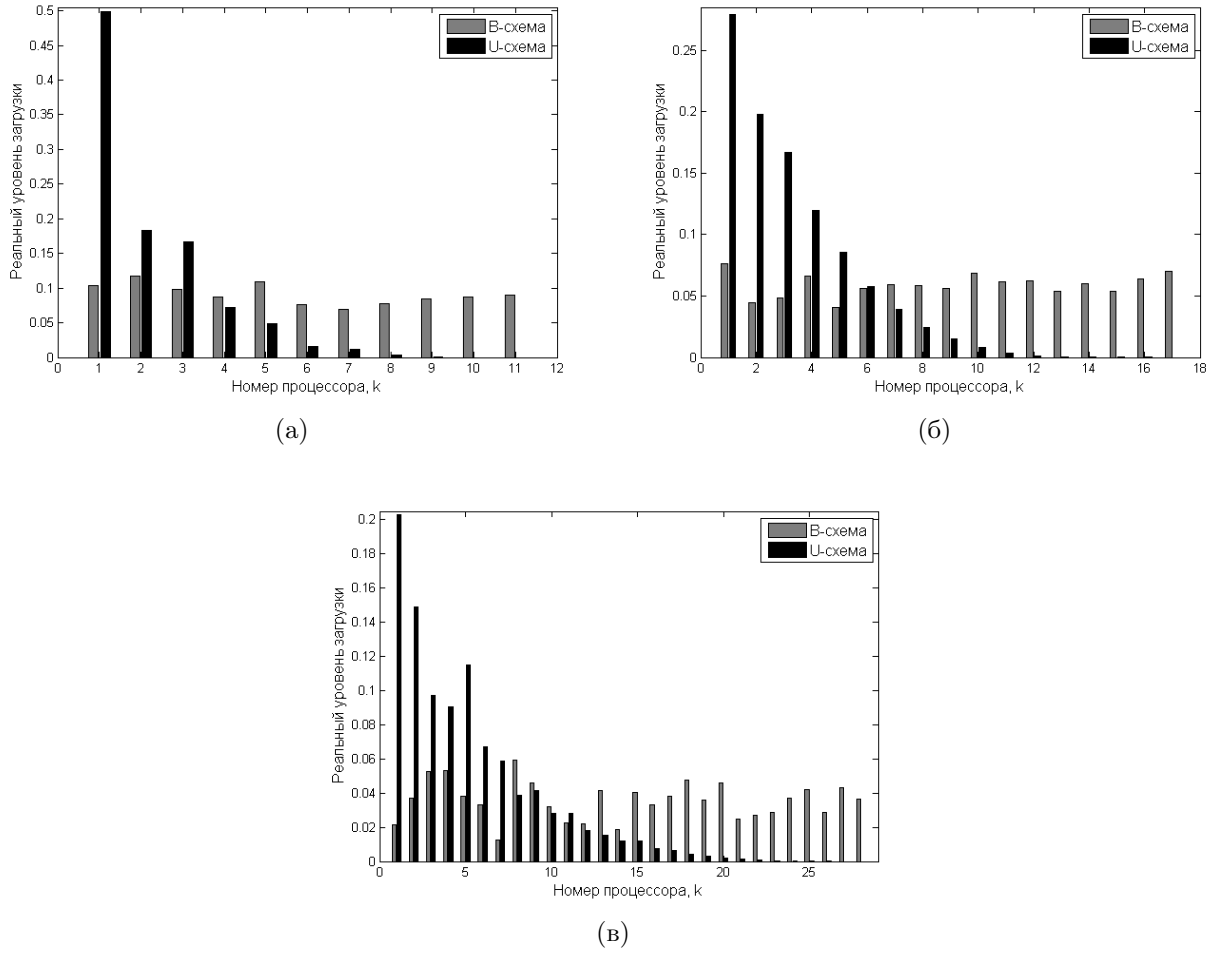


Рис. 3. Графики $\sigma_p(k)$ для $p = p^*$ для различных конфигураций матриц: (а) $m = 50, n = 50, p = 11$; (б) $m = 50, n = 100, p = 17$; (в) $m = 30, n = 120, p = 28$

k -го процессора назовем величину

$$\delta_k(L, \mathbf{N}^p) = \sum_{j \in J_n: N_j^p = k} |P_j(L)| / |P(L)|. \quad (1)$$

Введем функцию вероятности в пространстве Ω как $P\{L, H\} = 2^{-mn} |P(L)|^{-1}$, $L \in M_{mn}$, $H \in P(L)$. Нетрудно показать, что математическое ожидание $E(|P_j(L)| / |P(L)|)$ равно $P\{\eta(L, H) = j\}$, так как имеет место равенство:

$$|P_j(L)| = \sum_{H \in P(L)} \text{Ind}(\eta(L, H) = j), \quad (2)$$

где

$$\text{Ind}(x) = \begin{cases} 1, & \text{если } x \text{ истинно;} \\ 0 & \text{иначе.} \end{cases}$$

Из формул (1) и (2) следует, что средний уровень загрузки k -го процессора $E\delta_k(L, \mathbf{N}^p)$ равен

$$\sum_{j \in J_n: N_j^p = k} P\{\eta(L, H) = j\}. \quad (3)$$

Для построения эффективной в типичном случае схемы распараллеливания требуется решить задачу минимизации

$$\delta(\mathbf{N}^p) = \max_{k \in J_p} \mathbb{E} \delta_k(L, \mathbf{N}^p) \rightarrow \min_{\mathbf{N}^p}. \quad (4)$$

Ниже приведено описание процедуры 5, которая ищет приближенное решение задачи 4 при помощи жадного алгоритма. На вход процедура принимает число процессоров p , число столбцов n обрабатываемой матрицы и функцию вероятности $f(j) = \mathbb{P}\{\eta(L, H) = j\}$.

Процедура 5 AllocateResources

Вход: p, n, f ;

Выход: $\mathbf{N}^p$;

- 1: для $k = 1, \dots, p$
 - 2: $\delta_k \leftarrow 0$
 - 3: для $j = 1, \dots, n$
 - 4: $k_0 \leftarrow \arg \min_{k \in J_p} \delta_k$
 - 5: $N_j^p \leftarrow k_0$
 - 6: $\delta_k \leftarrow \delta_k + f(j)$
-

В предыдущем разделе показано, что $f(j) \approx \psi_{\alpha\beta}(j)$, где параметры α и β вычисляются с использованием оценок максимального правдоподобия. Пусть

$$\mathbf{N}_B^p \leftarrow \text{AllocateResources}(p, n, \psi_{\alpha\beta}).$$

Тогда схему распараллеливания, использующую расписание $\mathbf{N}_B^p$, назовем $\mathfrak{B}$ -схемой. Эта схема не будет эффективна при большом числе процессоров. На рис. 1, а изображен пример зависимости $\delta(\mathbf{N}_B^p)$ от p . Ясно, что при p не меньшем некоторого p^* функция $\delta(\mathbf{N}_B^p)$ является постоянной. В этом случае

$$\delta(\mathbf{N}_B^p) = \mathbb{E} \delta_1(L, \mathbf{N}_B^p) = \psi_{\alpha\beta}(1),$$

что следует из поведения функции $\psi_{\alpha\beta}$, график которой приведен на рисунке 1, б для случая $\alpha = 1, \beta = 6$ и $n = 50$. Другими словами, при $p \geq p^*$ недостаточно задач для равномерной загрузки процессора.

Эксперименты и сравнение схем распараллеливания

Все численные эксперименты выполнены на системе с Intel Core i3-2370M и 8192 MB RAM. В предложенных схемах распараллеливания межпроцессорный обмен осуществляется на начальном этапе, когда управляющий процессор раздает задания.

Опишем тривиальную схему распараллеливания асимптотически оптимальных алгоритмов дуализации ($\mathfrak{U}$ -схему), с которой будет сравниваться $\mathfrak{B}$ -схема. Пусть дано p процессоров и $L \in M_{mn}$. По аналогии с $\mathfrak{B}$ -схемой работа $\mathfrak{U}$ -схемы основана на использовании расписания $\mathbf{N}^p = (N_1^p, \dots, N_n^p)$, где $N_j^p = \lceil pj/n \rceil, j \in J_n$, и $\lceil x \rceil$ равно ближайшему целому, не меньшему x . Другими словами, первые примерно n/p столбцов должны быть обработаны первым процессором, следующие n/p столбцов — вторым и т. д.

Пусть $L \in M_{mn}$. Фиксируем схему распараллеливания. Обозначим через $T_p(k)$ время работы k -го из p процессоров. Тогда время работы $T(p)$ параллельного алгоритма будет

равно максимальному из $T_p(k)$. Фактическим уровнем загрузки k -го процессора назовем величину $\sigma_p(k) = T_p(k) / \sum_{k \in J_p} T_p(k)$.

Для исследования сильной масштабируемости предложенных схем распараллеливания (зависимости основных показателей работы алгоритма от количества процессоров при фиксированном размере входа задачи) рассмотрим такие показатели, как ускорение алгоритма $S(p) = T(1)/T(p)$ и средняя загрузка $E(p) = S(p)/p$. Значения $E(p)$, близкие к единице, свидетельствуют о равномерности загрузки процессоров.

На рис. 2 приведены результаты работы параллельных версий алгоритма АО2 (с $\mathfrak{B}$ -схемой и $\mathfrak{U}$ -схемой) для матриц размера 50×50 , 50×100 и 30×120 . Согласно графикам функции $S(p)$, $\mathfrak{B}$ -схема эффективна лишь при $p \leq p^*$. Например, для матриц 50×50 величина p^* равна 11. Видно, что при $p \leq p^*$ ускорение $S(p)$ для $\mathfrak{B}$ -схемы близко к линейному, что является несомненным преимуществом перед $\mathfrak{U}$ -схемой. На рис. 3 в виде столбчатых диаграмм приведены функции $\sigma_{p^*}(k)$, $k \in J_{p^*}$, демонстрирующие уровень загруженности процессоров. Во всех рассмотренных случаях $\mathfrak{B}$ -схема существенно превосходит $\mathfrak{U}$ -схему по равномерности загрузки процессоров.

Заключение

В данной работе предложена эффективная схема распараллеливания ($\mathfrak{B}$ -схема) асимптотически оптимальных алгоритмов дуализации.

Пусть на вход алгоритму подана матрица L с n столбцами. При наличии $p \leq n$ процессоров схема строит расписание $N^p = (N_1^p, \dots, N_n^p)$, где N_j^p равно номеру процессора, который должен построить подмножество неприводимых покрытий $P_j(L)$, $j \in J_n$, которые начинаются с j -го столбца.

На пространстве элементарных событий $\Omega = \{(L, H) \mid L \in M_{mn}, H \in P(L)\}$ введена случайная величина $\eta(L, H)$, равная наименьшему номеру столбца в H . Выдвинута статистическая гипотеза о том, что величина $\eta(L, H) - 1$ имеет бета-биномиальное распределение $\mathfrak{B}(\alpha, \beta, n - 1)$, и эта гипотеза подтверждена экспериментально. $\mathfrak{B}$ -схема использует бета-биномиальную аппроксимацию распределения случайной величины $\eta(L, H)$ для построения расписания N^p .

Рассмотрена более простая схема распараллеливания — $\mathfrak{U}$ -схема, в которой все процессоры строят примерно одинаковое число множеств $P_j(L)$.

Эксперименты показывают, что одним из главных преимуществ $\mathfrak{B}$ -схемы перед $\mathfrak{U}$ -схемой является значительно более равномерная загрузка процессоров и почти линейное ускорение $S(p)$. Однако указанная схема является эффективной при числе процессоров, существенно меньшем n . Поэтому в рамках дальнейших исследований планируется избавиться от указанного ограничения, разработав схемы для произвольного числа процессоров.

Литература

- [1] Johnson D. S., Yannakis M., Papadimitriou C. H. On generating all maximal independent sets // *Inform. Process. Lett.*, 1988. Vol. 27. P. 119–123.
- [2] Дюкова Е. В. Об асимптотически оптимальном алгоритме построения тупиковых тестов // *ДАН СССР*, 1977. Т. 223. № 4. С. 527–530.
- [3] Дюкова Е. В. О сложности реализации дискретных (логических) процедур распознавания // *Журнал вычислительной математики и математической физики*, 2004. Т. 44. № 3. С. 551–561.

- [4] Дюкова Е. В., Инякин А. С. Асимптотически оптимальное построение тупиковых покрытий целочисленной матрицы // *Математические вопросы кибернетики*, 2008. Т. 17. С. 235–246.
- [5] Murakami K., Uno T. Efficient algorithms for dualizing large-scale hypergraphs. Tokyo: Institute of Informatics, 2011.

References

- [1] Johnson D. S., Yannakis M., Papadimitriou C. H. 1988. On generating all maximal independent sets. *Inform. Process. Lett.* 27:119–123.
- [2] Djukova E. V. On an asymptotically optimal algorithm for constructing irredundant tests. 1977. *Dokl. Akad. Nauk SSSR* 233(4):423–426.
- [3] Dyukova E. V. On the implementation complexity of the realization of discrete (logical) recognition procedures. 2004. *Comp. Math. Math. Phys.* 44(3):532–541.
- [4] Djukova E. V., Inyakin A. S. 2008. Asimptoticheski optimal'noe postroenie tupikovykh pokrytiy tselochislennoy matritsy [Asymptotically optimal construction of irredundant coverings of an integer matrix]. *Matematicheskie voprosy kibernetiki* [Mathematical Problems of Cybernetics] 17:235–246.
- [5] Murakami K., Uno T. 2011. Efficient algorithms for dualizing large-scale hypergraphs. Tokyo: Institute of Informatics.

О комбинаторных оценках максимальных ε -разбиений метрических конфигураций

А. С. Пушняков

aleksey.pushnyakov@phystech.edu

Московский физико-технический институт, г. Москва

Рассматривается метрическое пространство с конечным числом точек. Вводится понятие максимального ε -разбиения. Рассматриваются нижние оценки на мощность максимального множества диаметра не более ε' при ограничении сверху на число расстояний, превосходящих ε . Показано, что в случае $\varepsilon' < 2\varepsilon$ нельзя гарантировать линейную по мощности пространства оценку. В случае $\varepsilon' \geq 2\varepsilon$ получена наилучшая оценка.

Ключевые слова: метрическая конфигурация; максимальное ε -разбиение; лемма Холла

On combinatorial bounds for maximal ε -partitions of a finite metric space

A. S. Pushnyakov

Moscow Institute of Physics and Technology, Moscow, Russia

A finite metric space (X, ρ) is studied. By ε -cluster, a subset of X with diameter at most ε is meant. Let there be an upper bound for the number of distances which are greater than ε . For maximal cardinality of ε' -cluster, lower bounds are considered. An important question is to find dependence between ε and ε' . It is shown that in case where $\varepsilon' < 2\varepsilon$, one cannot guarantee any linear bound. In case where $\varepsilon' \geq 2\varepsilon$, the best possible bound is obtained. A maximal ε' -partition is a partition into ε' -clusters constructed according to greedy procedure described below. Using Hall's marriage theorem, the existence of special matching between every two elements of maximal ε' -partition is proved. Considering maximal matching between ε' -cluster with maximal cardinality and its complement, one can calculate number of pairs (x, y) such that $\rho(x, y) > \varepsilon$ and obtain lower bound for maximal cardinality of ε' -cluster. In some particular cases, the value of ε' can be decreased. For instance, in case of Euclidean metric, one can assume $\varepsilon' = \sqrt{2}\varepsilon$ and obtain linear bound. However, it is unknown whether this bound could be improved.

Keywords: finite metric space; maximal ε -partition; Hall's marriage theorem

Введение

Использование метрической информации является основой решения многих задач интеллектуального анализа данных. В задачах классификации и кластеризации метрика на множестве объектов обычно согласуется с *принципом компактности*: близкие объекты скорее должны лежать в одном классе, нежели в разных (см., например, [1, 2]). Таким образом, можно полагать, что в случае *хорошей* метрики, множество объектов распадется на подмножества *небольшого* диаметра.

В данной работе рассматривается произвольное метрическое пространство конечной мощности — метрическая конфигурация. Исследуется существование подмножества *малого* (по сравнению с диаметром всего пространства) диаметра ε' и *достаточно большой* мощности при ограничении на число расстояний превосходящих, значение другого *малого* параметра ε . Иными словами, ставится вопрос, всегда ли при известных и достаточно малых ε' и ε можно найти достаточно большой кластер, покрывающий *значительную* долю точек, например, близкую к единице. Отметим, что поиск такого рода кластеров или метрических сгущений важен для интеллектуального анализа данных, визуализации [3] и анализа самих задач классификации [4]. Соответствующим вычислительным методам посвящено значительное число работ (см., например, [5]). Особый интерес представляет связь ε' и ε . Далее будет показано, что в случае $\varepsilon' < 2\varepsilon$ нельзя гарантировать линейную по мощности пространства оценку. В случае $\varepsilon' = 2\varepsilon$ будет получена оценка, стремящаяся к единице при стремящейся к нулю доле расстояний, превосходящих ε . В случае $\varepsilon' > 2\varepsilon$ оценка не изменится, более того, будет показано, что она является наилучшей.

Однако введение дополнительных ограничений на метрику позволяет уменьшить ε' . В последнем разделе будут рассмотрены два частных случая: метрика является ультраметрикой и метрика вложима некоторое евклидово пространство [6]. В случае ультраметрики получаемая наилучшая оценка для $\varepsilon' = \varepsilon$, а для евклидовой метрики получена оценка для $\varepsilon' = \sqrt{2}\varepsilon$.

Для получения свойств искомого кластера рассматриваются разбиения метрической конфигурации на подмножества диаметра не более ε' , среди которых *жадным* образом выделяется *максимальное*: среди

всех подмножеств диаметра не более ε' вначале выбирается максимальное по мощности, затем из оставшихся точек снова набирается максимальное по мощности и т. д. Нижняя оценка на мощность первого подмножества в определяемом ниже *максимальном* ε' -разбиении и есть искомая оценка.

Основные определения

Рассмотрим метрическое пространство (X, ρ) с конечным числом точек. Напомним, что *диаметром* множества $Y \subset X$ называется величина $D(Y) = \sup_{x, y \in Y} \rho(x, y)$, причем, в нашем случае $|X| < \infty$, и супремум можно заменить на максимум.

Определение 1. Набор подмножеств $\{X_0, \dots, X_s\}$ будем называть ε -разбиением метрической конфигурации (X, ρ) , если выполнены следующие условия:

- 1) $\bigcup_{i=0}^s X_i = X$;
- 2) $X_i \cap X_j = \emptyset$ при всех $0 \leq i < j \leq s$;
- 3) $D(X_i) \leq \varepsilon$.

Зафиксируем произвольное число $\varepsilon > 0$. Из всех подмножеств X выберем такие, что их диаметр не превосходит ε :

$$W_\varepsilon = \{V \subseteq X : D(V) \leq \varepsilon\}. \quad (1)$$

Так как полученное семейство подмножеств конечно (в силу конечности множества X), то существует максимальное по мощности множество $X_0 \in W_\varepsilon$ (если таких несколько, то выберем любое из них). Далее выберем X_1 следующим образом: среди всех $V \in W_\varepsilon$ таких, что $V \cap X_0 = \emptyset$, берем любое из максимальных по мощности. Пусть уже построены множества $X_0, \dots, X_k$, тогда X_{k+1} строится аналогично: среди всех $V \in W_\varepsilon$ таких, что $V \cap X_i = \emptyset$ при всех $i = 0, \dots, k$, берем любое из максимальных по мощности. В итоге получаем разбиение X на подмножества X_i , $i = 0, \dots, s$, которое будем называть *максимальным ε -разбиением* метрической конфигурации X .

Определение 2. Набор подмножеств $\{X_0, \dots, X_s\}$ будем называть *максимальным ε -разбиением* метрической конфигурации (X, ρ) , если выполнены следующие условия:

- 1) набор является ε -разбиением (X, ρ) ;
- 2) при всех $0 \leq i \leq s$ для любого $Y \in W_\varepsilon$ такого, что $Y \cap \bigcup_{j=0}^{i-1} X_j = \emptyset$, выполнено $|Y| \leq |X_i|$.

Отметим, что максимальное разбиение может быть неединственным. Это иллюстрирует следующий простой пример. Пусть $X = \{x, y, z\}$, $\rho(x, y) = \rho(x, z) = \varepsilon$, $\rho(y, z) = 2\varepsilon$. Тогда ε -разбиения $\{x, y\}, \{z\}$ и $\{x, z\}, \{y\}$ являются максимальными.

Свойства максимальных ε -разбиений

Пусть дано некоторое максимальное ε -разбиение $\{X_0, \dots, X_s\}$. Вначале докажем почти очевидное утверждение.

Утверждение 1. Для всех целых неотрицательных i и любой точки $x \in \bigcup_{k=i+1}^s X_k$ существует $y \in X_i$ такое, что $\rho(x, y) > \varepsilon$.

Доказательство. Предположим противное: пусть нашлась такая точка $x \in \bigcup_{k=i+1}^s X_k$, что $\rho(x, y) \leq \varepsilon$, для

любого $y \in X_i$. Так как $x \notin \bigcup_{k=0}^i X_k$, то при построении множества X_i к нему можно добавить точку x , и диаметр $X_i \cup \{x\}$ не будет превосходить ε , что противоречит максимальнойности X_i согласно построению. ■

Теперь докажем более сложное утверждение, использующее свойство максимальнойности X_i и лемму Холла о паросочетаниях [7]. Напомним, что *паросочетанием* в графе $G = (E, V)$ называется подмножество ребер $M \subset E$ такое, что никакие два ребра из M не инциденты одной и той же вершине.

Лемма 1 (Холл). В двудольном графе $G = (E, V)$ с долями V_0 и V_1 существует паросочетание, покрывающее все вершины V_1 , тогда и только тогда, когда для любого подмножества $U \subset V_1$ мощность U не превосходит мощности множества вершин, смежных с вершинами U .

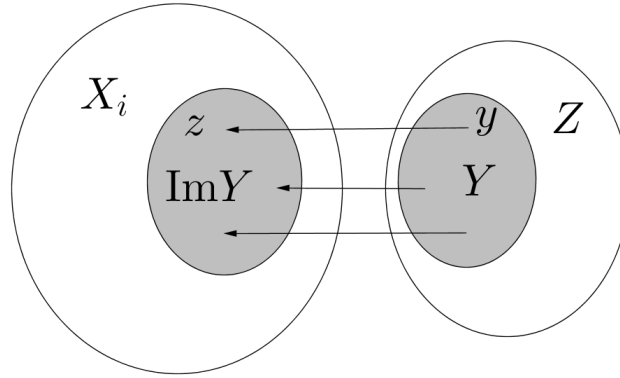


Рис. 1.

Утверждение 2. Для всех целых неотрицательных i и любого подмножества $Z \subseteq \bigcup_{k=i+1}^s X_k$ такого, что $D(Z) \leq \varepsilon$, существует инъективное отображение $\varphi : Z \rightarrow X_i$, удовлетворяющее условию $\rho(x, \varphi(x)) > \varepsilon$ для любого $x \in Z$.

Доказательство. Представим множества X_i и Z как доли двудольного графа G (рис. 1), в котором вершины $u \in X_i$ и $v \in Z$ соединены ребром тогда и только тогда, когда $\rho(u, v) > \varepsilon$. Рассмотрим произвольное подмножество $Y \subseteq Z$, $|Y| = k$ и соответствующее ему подмножество $\text{Im } Y \subseteq Z$, определяемое следующим образом:

$$\text{Im } Y = \{z \in X_i : \exists y \in Y, \rho(z, y) > \varepsilon\}. \quad (2)$$

Иными словами, $\text{Im } Y$ — множество вершин графа, смежных с вершинами Y . Покажем, что $|Y| \leq |\text{Im } Y|$. Будем следовать идее доказательства утверждения 1. Предположим противное. Тогда рассмотрим множество $X'_i = (X_i \setminus \text{Im } Y) \cup Y$; в силу предположения имеем $|X'_i| > |X_i|$. Так как для любых $y \in Y$, $x \in X'_i$ верно $\rho(x, y) \leq \varepsilon$, то получаем, что $D(X'_i) \leq \varepsilon$. А это в свою очередь противоречит выбору X_i как максимального по мощности.

Теперь осталось заметить, что построение инъективного отображения $\varphi : Z \rightarrow X_i$, удовлетворяющего условию, эквивалентно нахождению паросочетания в графе G , покрывающего Z . Наличие последнего следует из леммы Холла, условие применения которой не что иное, как только что доказанное неравенство $|Y| \leq |\text{Im } Y|$. ■

До сих пор мы пользовались только свойством симметричности метрики. Следующее утверждение уже использует неравенство треугольника. Пусть $D(V) \leq \varepsilon$, а $Y \cap V = \emptyset$. Аналогично (2) в более общем случае определим:

$$\text{Im}_\varepsilon(Y, V) = \{v \in V : \exists y \in Y, \rho(y, v) > \varepsilon\}. \quad (3)$$

Утверждение 3. Пусть выполнено $D(V) \leq \varepsilon$, $D(Y) \leq 2\varepsilon$, $Y \cap V = \emptyset$ и $\text{Im}_\varepsilon(Y, V) \neq V$, тогда верно $D(Y \cup V) \leq 2\varepsilon$.

Доказательство. Так множество $V \setminus \text{Im}_\varepsilon(Y, V)$ непусто, то возьмем любой его элемент z_0 . Тогда для любого $v \in V$ верно $\rho(v, z_0) \leq \varepsilon$, и для любого $y \in Y$ также выполнено $\rho(y, z_0) \leq \varepsilon$. По неравенству треугольника получаем, что $\rho(v, y) \leq \rho(v, z_0) + \rho(z_0, y) \leq 2\varepsilon$ для любых $v \in V$, $y \in Y$. А так как $D(V) \leq \varepsilon$, $D(Y) \leq 2\varepsilon$, то $D(Y \cup V) \leq 2\varepsilon$. ■

Постановка задачи о максимальном кластере

Теперь применим вышеописанную конструкцию для решения следующей задачи. Пусть выбран некоторый порог ε , и известно, что из всех $\binom{|X|}{2}$ попарных расстояний число таких, которые больше ε , мало в следующем смысле:

$$|\Lambda_\varepsilon| = |\{(x, y) : \rho(x, y) > \varepsilon\}| \leq \frac{\delta |X|^2}{2}, \quad (4)$$

где $\delta > 0$ — некоторое число. В задаче требуется определить некоторый новый порог ε' такой, чтобы значительная часть точек пространства попала в некоторое множество диаметра не более ε' . Иначе говоря, требуется найти гарантированную оценку снизу на мощность максимального множества диаметра не более ε' , которое и будем называть *максимальным кластером*.

В начале приведем отрицательный результат, который определит дальнейший выбор ε' .

Утверждение 4. Пусть $\varepsilon \leq \varepsilon' < 2\varepsilon$, тогда для любых $\alpha > 0$, $\delta > 0$ существует метрическое пространство (X, ρ) такое, что $|\Lambda_\varepsilon| \leq \delta |X|^2$, и $\max\{|V| : D(V) \leq \varepsilon'\} < \alpha |X|$.

Доказательство. Положим $X = \bigcup_{i=1}^s Y_i$, где $Y_i = \{x_{i1}, \dots, x_{im}\}$, $|X| = ms$. Определим метрику следующим образом:

$$\rho(x_{ij}, x_{kl}) = \begin{cases} 2\varepsilon, & i \neq k, j = l; \\ 0, & i = k, j = l; \\ \varepsilon, & \text{иначе.} \end{cases} \quad (5)$$

Очевидно, что данная функция является метрикой. Тогда имеем $D(Y_i) \leq \varepsilon \leq \varepsilon'$, то есть $\max\{|V| : D(V) \leq \varepsilon'\} \geq m$, но для любого $(m+1)$ -элементного подмножества $V \subseteq X$, в нем найдутся x_{ij} и x_{kl} такие, что $i \neq k, j = l$, и тогда $D(V) > \varepsilon'$. Значит, $\max\{|V| : D(V) \leq \varepsilon'\} = m$. Тогда $|\Lambda_\varepsilon| = m \binom{s}{2}$, и

$$\begin{aligned} \frac{|\Lambda_\varepsilon|}{|X|^2} &= \frac{ms(s-1)}{2s^2m^2} \leq \frac{1}{2m}, \\ \frac{\max\{|V| : D(V) \leq \varepsilon'\}}{|X|} &= \frac{m}{ms} = \frac{1}{s}. \end{aligned}$$

Осталось только положить $m = \lfloor 2/\delta \rfloor + 1$ и $s = \lfloor 1/\alpha \rfloor + 1$. ■

Таким образом, чтобы получить достаточно большое значение $\max\{|V| : D(V) \leq \varepsilon'\}$, необходимо брать $\varepsilon' \geq 2\varepsilon$. Мы остановимся на случае $\varepsilon' = 2\varepsilon$. Очевидно, что в случае $\varepsilon' > 2\varepsilon$ оценка как функция от δ не ухудшится.

Оценка мощности максимального кластера

Построим, как было описано выше, максимальное 2ε -разбиение множества X на множества X_i , $i = 0, \dots, s$, диаметра не более 2ε .

Сформулируем сначала почти очевидное неравенство.

Утверждение 5. При выполнении неравенства (4) верно неравенство $|X_0| \geq |X|(1 - \delta)$.

Доказательство. Общее количество расстояний не превосходящих ε не меньше, чем $|X|(|X| - 1)/2 - |X|^2\delta/2$. Представим X как граф, в котором вершины u и v соединены ребром тогда и только тогда, когда $\rho(u, v) \leq \varepsilon$. Тогда суммарная степень вершин не менее $|X|(|X| - 1 - \delta|X|)$, а значит, есть вершина z_0 степени не менее $|X| - 1 - \delta|X|$. Тогда в замкнутом шаре $B_\varepsilon(z_0) = \{x : \rho(z_0, x) \leq \varepsilon\}$ есть хотя бы $|X| - \delta|X|$ точек. Осталось заметить, что $D(B_\varepsilon(z_0)) \leq 2\varepsilon$. ■

Теперь перейдем к получению более точной оценки. Мы будем оценивать снизу число расстояний превосходящих ε , то есть мощность множества Λ_ε . Как и раньше, удобно рассматривать (X, ρ) как взвешенный полный граф. Будем называть ребро (u, v) *коротким*, если $\rho(u, v) \leq \varepsilon$, все остальные ребра — *длинными*.

В этом графе рассмотрим двудольный подграф $G_1 = (X, E)$ с долями X_0 и $X \setminus X_0$, причем $(u, v) \in E$ тогда и только тогда, когда u и v лежат в разных долях, и $\rho(u, v) > 2\varepsilon$. Согласно утверждению 2 для любого $Y \subseteq X \setminus X_0$ такого, что $D(Y) \leq 2\varepsilon$, существует инъекция по ребрам G_1 в X_0 . Далее мы покажем, что существует подмножество $X'_0 \subseteq X_0$ мощности не более $|X \setminus X_0|$ такое, что для любого $Y \subseteq X \setminus X_0$ такого, что $D(Y) \leq 2\varepsilon$, существует инъекция по ребрам G_1 в X'_0 .

Рассмотрим следующую конструкцию. Пусть $M \subseteq E$ — максимальное паросочетание в $G_1 = (X, E)$, определенном выше. Путь в G_1 назовем *чередующимся*, если в нем ребра из M и из $E \setminus M$ чередуются. Пусть U_1 — множество вершин в X_0 , не покрытое паросочетанием M , аналогично определяется U_2 . В силу максимальной M , между U_1 и U_2 ребер нет. Далее, обозначим за Y_1 вершины из X_0 , покрытые M , и до которых существует *чередующийся* путь из U_1 ; остальные вершины X_0 образуют Z_1 . Аналогично определим Y_2 как покрытые M вершины из $X \setminus X_0$, до которых существует *чередующийся* путь из U_1 ; остальные вершины $X \setminus X_0$ образуют Z_2 .

На рис. 2 схематично изображена структура графа G_1 . Сплошными линиями изображены ребра паросочетания M . Пунктирными линиями обозначено отсутствие ребер между соответствующими компонентами. Следующее утверждение объясняет, почему граф устроен именно так.

Утверждение 6. Для вышеописанной конструкции выполнено:

- 1) между парами компонент (U_1, U_2) , (U_1, Z_2) , (Y_1, U_2) и (Y_1, Z_2) нет ребер;

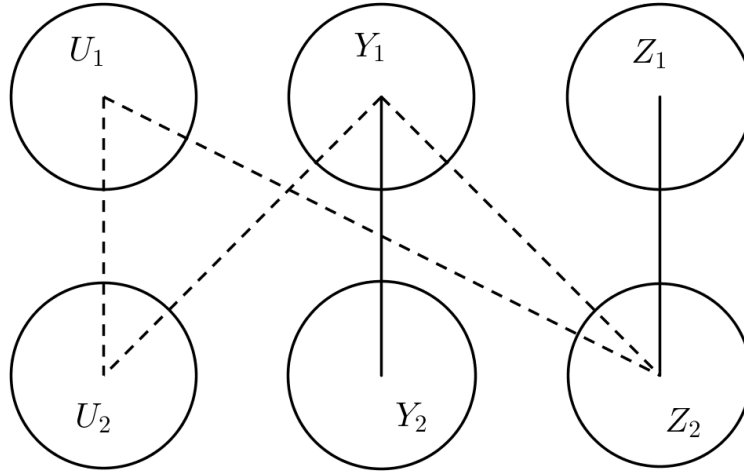


Рис. 2.

2) ребра паросочетания соединяют вершины Y_1 с вершинами Y_2 и вершины Z_1 с вершинами Z_2 .

Доказательство. Так как любом *чередующемся* пути из U_1 в Y_1 первое ребро не лежит в M , то тогда все ребра этого пути, идущие в X_0 , будут лежать в M . Тогда последнее ребро будет принадлежать M . Следовательно, предпоследняя вершина на этом пути лежит в Y_2 (так как до Z_2 нет чередующихся путей из U_1). Более того, среди ребер M нет таких, которые соединяют Y_2 и Z_1 , иначе бы нашелся *чередующийся* путь из U_1 в Z_1 . Следовательно, паросочетание M содержит только ребра между Y_1 и Y_2 и между Z_1 и Z_2 . Значит, $|Y_1| = |Y_2|$ и $|Z_1| = |Z_2|$. Далее, если бы между Z_2 и Y_1 были ребра, то они лежали бы в $E \setminus M$, и мы бы нашли *чередующийся* путь из U_1 в Z_2 . Значит между Z_2 и Y_1 нет ребер.

Теперь покажем, что ребер из U_2 в Y_1 нет. Предположим противное. Пусть $a_0 \in U_2$, $a_1 \in Y_1$, $(a_0, a_1) \in E \setminus M$. Пусть $a_1 a_2 \dots a_{2k+1}$ — *чередующийся* путь из Y_1 в U_1 , тогда $a_0 a_1 \dots a_{2k+1}$ — *чередующийся* путь из U_1 в U_1 . Причем ребра (a_{2i-1}, a_{2i}) лежат в паросочетании, а остальные — нет. Тогда заменим в паросочетании M все ребра (a_{2i-1}, a_{2i}) на (a_{2i}, a_{2i+1}) , и мощность M увеличится на единицу, что противоречит максимальной M . Противоречие. ■

Теперь используем вышеописанную конструкцию для доказательства следующего утверждения.

Утверждение 7. Для любого $V \subseteq X \setminus X_0$ такого, что $D(V) \leq 2\varepsilon$, существует инъекция φ по ребрам G_1 в $X'_0 \stackrel{\text{def}}{=} Y_1 \cup Z_1$.

Доказательство. Обозначим $V_u = V \cap U_2$, $V_y = V \cap Y_2$ и $V_z = V \cap Z_2$, причем $V = V_u \sqcup V_y \sqcup V_z$. Для всех $u \in V_y$ в качестве $\varphi(u)$ возьмем вершину, соединенную с u по ребру из M , то есть $(u, \varphi(u)) \in M$ и $\varphi(u) \in Y_1$. Отметим, что из $U_2 \cup Z_2$ нет ребер в $U_1 \cup Y_1$. Тогда, применяя утверждение 2 для $V_u \cup V_z$, получаем, что для всех $u \in V_u \cup V_z$ верно включение $\varphi(u) \in Z_1$. ■

Отметим также, что $|X_1| \leq |X'_0| \leq |X \setminus X_0|$. Теперь перейдем непосредственно к оценке числа *длинных* ребер.

Утверждение 8. Для любого $y \in X \setminus X'_0$ существует не менее $|X \setminus X_0|$ *длинных* ребер во множество вершин $X'_0 \cup (X \setminus X_0)$.

Доказательство. Рассмотрим множество $C_y = \{x \in X \setminus X_0 : \rho(x, y) \leq \varepsilon\}$. По неравенству треугольника для любых $x_1, x_2 \in C_y$ имеем $\rho(x_1, x_2) \leq \rho(x_1, y) + \rho(x_2, y)$, а значит, $D(C_y) \leq 2\varepsilon$ и $|C_y| \leq |X_1|$. Тогда по утверждению 7 существует инъекция $\varphi : C_y \rightarrow X'_0$. И для каждого $x \in C_y$ имеем $\rho(y, \varphi(x)) \geq \rho(x, \varphi(x)) - \rho(x, y) > 2\varepsilon - \varepsilon = \varepsilon$. Мы получили $|C_y|$ *длинных* ребер, идущих в X'_0 , и $|X \setminus X_0| - |C_y|$, идущих в $X \setminus X_0$. ■

Следствие 1. Число *длинных* ребер, исходящих из $X_0 \setminus X'_0$, не менее $|X \setminus X_0| \cdot |X_0 \setminus X'_0|$.

Утверждение 9. Число *длинных* ребер между вершинами множества $X'_0 \cup (X \setminus X_0)$ не менее $|X \setminus X_0| \cdot |X'_0|$.

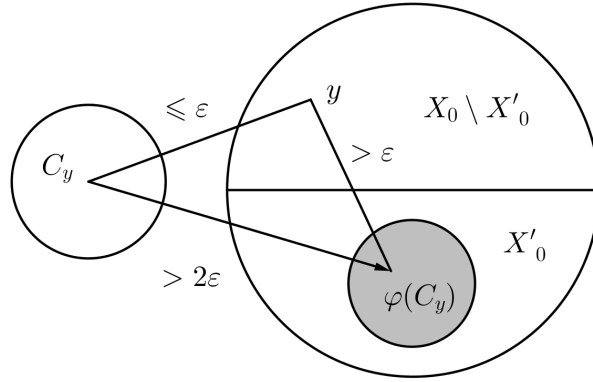


Рис. 3.

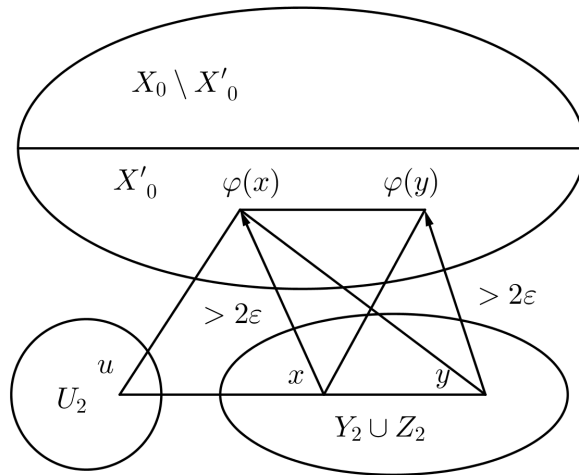


Рис. 4.

Доказательство. Для каждого $x \in Y_2 \cup Z_2$ обозначим за $\varphi(x)$ такую вершину, что ребро $(x, \varphi(x))$ лежит в максимальном паросочетании M ; по построению $\rho(x, \varphi(x)) > 2\epsilon$. Также имеем $\bigcup_{x \in Y_2 \cup Z_2} \varphi(x) = Y_1 \cup Z_1 = X'_0$.

Рассмотрим вначале любые две вершины $x, y \in Y_2 \cup Z_2$. Покажем, что среди ребер (x, y) , $(\varphi(x), y)$, $(x, \varphi(y))$, $(\varphi(x), \varphi(y))$ есть хотя бы два *длинные*. Действительно, если $\rho(x, \varphi(y))$ и $\rho(x, \varphi(x))$ *длинные*, то утверждение верно. Если же, например, $\rho(x, \varphi(y)) \leq \epsilon$, то по неравенству треугольника получаем, что $\rho(x, y) \geq \rho(y, \varphi(y)) - \rho(x, \varphi(y)) > 2\epsilon - \epsilon = \epsilon$ и $\rho(\varphi(x), \varphi(y)) \geq \rho(x, \varphi(x)) - \rho(x, \varphi(y)) > 2\epsilon - \epsilon = \epsilon$, то есть ребра (x, y) и $(\varphi(x), \varphi(y))$ *длинные*. Также имеем, что все ребра $(x, \varphi(x))$ *длинные*. Тогда получаем следующую оценку на λ_1 — число *длинных* ребер между вершинами множества $X'_0 \cup Y_2 \cup Z_2$:

$$\lambda_1 \geq |X'_0| + 2 \frac{|X'_0|(|X'_0| - 1)}{2} = |X'_0|^2.$$

Теперь рассмотрим произвольную вершину $u \in U_2$ и пару $(x, \varphi(x)) \in (Y_2 \cup Z_2) \times X'_0$. Опять же используя неравенство треугольника для вершин x , $\varphi(x)$ и u получаем, что одно из ребер (u, x) и $(u, \varphi(x))$ является *длинным*. Тогда мы получаем следующую оценку на число ребер из U_2 в $X'_0 \cup Y_2 \cup Z_2$:

$$\lambda_2 \geq |U_2| \cdot |X'_0|.$$

Складывая полученные неравенства и вспоминая, что $|U_2| + |X'_0| = |X \setminus X_0|$, получаем требуемое неравенство. ■

Из следствия 1 и утверждения 9 сразу получаем центральное утверждение.

Следствие 2. Число *длинных* ребер в X не меньше, чем $|X_0| \cdot |X \setminus X_0|$.

Теперь воспользуемся неравенством (4) и получим:

$$|X| \cdot |X_0| - |X_0|^2 = |X_0| \cdot |X \setminus X_0| \leq |\Lambda_\varepsilon| \leq \frac{\delta |X|^2}{2}. \quad (6)$$

Будем считать, что $0 < \delta \leq 1/2$. Согласно утверждению 5, имеем $X_0 \geq |X|/2$. Тогда, решая неравенство (6) и выбирая решения, подходящие под условие $X_0 \geq \frac{|X|}{2}$, получаем:

$$|X_0| \geq \frac{|X|(1 + \sqrt{1 - 2\delta})}{2}. \quad (7)$$

Следующий простой пример показывает, что полученная оценка на число *длинных* ребер не улучшаема в общем случае. Пусть $X = X_0 \sqcup X_1$ и для любых $x \in X_0, y \in X_1$ $\rho(x, y) = 3\varepsilon$, а все остальные расстояния равны ε . Тогда число *длинных* ребер в точности равно $|X_0| \cdot |X_1|$. Этот пример также подходит для случая $\varepsilon' > 2\varepsilon$ при замене 3ε на $\varepsilon' + 1$.

Наконец, проверим, что мы получили оценку более сильную, чем в утверждении 5. В силу $0 < \delta \leq 1/2$ имеем $\sqrt{1 - 2\delta} \geq 1 - 2\delta$ и тогда получаем:

$$\frac{|X|(1 + \sqrt{1 - 2\delta})}{2} \geq |X|(1 - \delta),$$

причем равенство достигается только при $\delta = \frac{1}{2}$. Также исследуем асимптотическое поведение оценки (7) при $\delta \rightarrow +0$:

$$\frac{|X|(1 + \sqrt{1 - 2\delta})}{2} = \frac{|X|(1 + 1 - \delta + o(\delta))}{2} \approx |X| \left(1 - \frac{\delta}{2}\right).$$

Некоторые частные случаи

Выше, в утверждении 4, было показано, что в случае $\varepsilon' < 2\varepsilon$ линейную по $|X|$ оценку на $|X_0|$ гарантировать нельзя. Однако при некоторых дополнительных условиях на метрику значение ε' может быть уменьшено. Далее будут рассмотрены два частных случая: метрика является ультраметрикой и метрика вложима в евклидово пространство.

Напомним, что метрика ρ на X называется *ультраметрикой*, если для любых $x, y, z \in X$ выполнено $\rho(x, y) \leq \max\{\rho(y, z), \rho(x, z)\}$. Для ультраметрик известно следующее свойство.

Утверждение 10. Если ρ — ультраметрика на X , то любой треугольник является равнобедренным. Если треугольник не равносторонний, то основание меньше боковых сторон.

Покажем, что в случае ультраметрики для получения не улучшаемой оценки (7) достаточно взять $\varepsilon' = \varepsilon$. Построим максимальное ε -разбиение множества X на множества $X_i, i = 0, \dots, s$.

Утверждение 11. Если ρ — ультраметрика на X , и выполнено неравенство (4) при $\delta \leq \frac{1}{2}$, тогда верно (7).

Доказательство. Рассмотрим любую точку $x \in X \setminus X_0$. Согласно утверждению 1 найдем $\varphi(x) \in X_0$ такую, что $\rho(x, \varphi(x)) > \varepsilon$. Рассмотрим произвольную точку $y \in X_0$ такую, что $y \neq \varphi(x)$. Тогда, применяя утверждение 10 к треугольнику $xy\varphi(x)$, получаем $\rho(x, y) > \varepsilon$. Значит, число *длинных* ребер из точки x во множество X_0 равно $|X_0|$. Таким образом, получаем, что число *длинных* ребер не меньше $|X_0| \cdot |X \setminus X_0|$. Отсюда с учетом $\delta \leq 1/2$ сразу получается (7). ■

Перейдем к случаю евклидовых метрик. Напомним, что метрическое пространство (X, ρ) называется *вложимым* в евклидово пространство, если существует изомерия из (X, ρ) в некоторое евклидово пространство E . Далее будем для удобства считать, что $X \subset E$.

Рассмотрим произвольную точку $x_0 \in X$ и векторы $v_i = x_i - x_0, i = 1, \dots, |X| - 1$. Выразим их скалярные произведения через попарные расстояния

$$\Gamma_{ij} = (v_i, v_j) = \frac{1}{2}(\rho^2(x_i, x_0) + \rho^2(x_j, x_0) - \rho^2(x_i, x_j)), \quad i, j = 1, 2, \dots, |X| - 1. \quad (8)$$

Справедливо следующее утверждение.

Утверждение 12. Метрическое пространство $(X, \rho), |X| < \infty$ вложимо в некоторое евклидово пространство тогда и только тогда, когда матрица $\Gamma = \|\Gamma_{ij}\|_{i,j=1}^{|X|-1}$ является неотрицательно определенной.

Доказательство. Пусть метрическое пространство (X, ρ) вложимо в некоторое евклидово пространство. Матрица Γ есть матрица Грама векторов v_i , откуда сразу следует, что Γ является неотрицательно определенной.

Пусть теперь матрица Γ является неотрицательно определенной. Тогда существует такие ортогональная матрица S и диагональная матрица $D = \text{diag}(\lambda_1, \dots, \lambda_{|X|-1})$, $\lambda_i \geq 0$, что $\Gamma = S^T D S$. Рассмотрим матрицу $U = \sqrt{D} S = \|u_1, \dots, u_{|X|-1}\|$, тогда $\Gamma = U^T U$. Искомое вложение в $\mathbb{R}^{|X|-1}$ строим следующим образом: точка x_0 переходит в начало координат, а любая другая x_i переходит в точку с координатами u_i .

Проверим условие изометрии. Евклидово расстояние в $\mathbb{R}^{|X|-1}$ обозначим за ρ_e . Для любых $i \geq 1$ имеем:

$$\rho_e^2(x_i, x_0) = (u_i, u_i) = \Gamma_{ii} = \rho^2(x_i, x_0).$$

Для любых $i, j \geq 1$, применяя предыдущее равенство, имеем:

$$\rho_e^2(x_i, x_j) = (u_i, u_i) + (u_j, u_j) - 2(u_i, u_j) = \rho_e^2(x_i, x_0) + \rho_e^2(x_j, x_0) - 2\Gamma_{ij} = \rho^2(x_i, x_j).$$

■

Получим удобное следствие из предыдущего утверждения.

Утверждение 13. Пусть $a, b, c, d \in X$, тогда $\rho^2(a, b) + \rho^2(b, c) + \rho^2(c, d) + \rho^2(d, a) \geq \rho^2(a, c) + \rho^2(b, d)$.

Доказательство. Пусть $x = b - a$, $y = c - a$, $z = d - a$. Рассмотрим матрицу Грама $\Gamma(x, y, z)$ векторов x, y, z ; как следствие утверждения 12, она является неотрицательно определенной, и в частности, при $\xi = (1, -1, 1)^T$ выполнено $\xi^T \Gamma(x, y, z) \xi \geq 0$. Подставляя (8), получаем:

$$\begin{aligned} & \rho^2(a, b) + \rho^2(a, c) + \rho^2(d, a) + (\rho^2(a, b) + \rho^2(d, a) - \rho^2(b, d)) - \\ & - (\rho^2(a, c) + \rho^2(d, a) - \rho^2(c, d)) - (\rho^2(a, c) + \rho^2(a, b) - \rho^2(c, b)) \geq 0; \\ & \rho^2(a, b) + \rho^2(a, c) + \rho^2(d, a) + \rho^2(c, b) + \rho^2(c, d) \geq \rho^2(a, c) + \rho^2(b, d). \end{aligned}$$

■

Следствие 3. Пусть $a, b, c, d \in X$ и $\rho(a, c) > \sqrt{2}\epsilon$ и $\rho(b, d) > \sqrt{2}\epsilon$, тогда

$$\max\{\rho(a, b), \rho(a, d), \rho(b, c), \rho(c, d)\} > \epsilon.$$

Теперь положим $\epsilon' = \sqrt{2}$ и рассмотрим максимальное ϵ' -разбиение множества X на множества X_i , $i = 0, \dots, s$. Покажем, что в этом случае можно гарантировать линейную по $|X|$ оценку на $|X_0|$.

Утверждение 14. Пусть ρ вложима в евклидово пространство, и выполнено неравенство (4) при $\delta \leq 1/4$, тогда верна следующая оценка:

$$|X_0| \geq |X|(1 - 2\sqrt{\delta}).$$

Доказательство. Для каждого нечетного по номеру подмножества X_{2k+1} в максимальном $\sqrt{2}\epsilon$ -разбиении построим согласно утверждению 2 инъекцию φ по ребрам длины больше $\sqrt{2}\epsilon$ во множество X_{2k} и получим некоторое паросочетания M , мощность которого:

$$|M| = \sum_{1 \leq 2k+1 \leq s} |X_{2k+1}| \geq \frac{1}{2} \sum_{k=1}^s |X_k| = \frac{1}{2}(|X| - |X_0|).$$

Для каждой пары ребер $x\varphi(x)$ и $y\varphi(y)$ из M рассмотрим четырехугольник $xu\varphi(x)\varphi(y)$, и по следствию 3 среди ребер xu , $x\varphi(y)$, $y\varphi(x)$, $\varphi(x)\varphi(y)$ найдется хотя бы одно *длинное* (больше, чем ϵ), причем, каждое из этих ребер однозначно определяется парой ребер паросочетания M . Тогда получим, что общее число *длинных* ребер не меньше, чем

$$|M| + \binom{|M|}{2} \geq \frac{1}{2}|M|^2 \geq \frac{1}{8}(|X| - |X_0|)^2.$$

Тогда в силу (4) получаем:

$$\frac{1}{8}(|X| - |X_0|)^2 \leq \frac{\delta|X|^2}{2}, \quad |X_0| \geq |X|(1 - 2\sqrt{\delta}).$$



Заключение

При ограничениях на число расстояний, превосходящих ε , получена неупрощаемая нижняя оценка на мощность максимального по включению множества диаметра не более 2ε . Показано, что при $\varepsilon' < 2\varepsilon$ в общем случае нельзя гарантировать существования линейной по мощности метрической конфигурации нижней оценки на мощность множества диаметра не более ε' . Рассмотрены частные случаи ультраметрик и евклидовых метрик, для которых значение ε' может быть уменьшено.

Литература

- [1] Аркадьев А. Г., Браверман Э. М. Обучение машины классификации объектов. — Наука, 1971.
- [2] Загоруйко Н. Г. Гипотезы компактности и λ -компактности в методах анализа данных // *Сибирский журнал индустриальной математики*, 1998. Т. 1. № 1. С. 114–126.
- [3] Выявление и визуализация метрических структур на множествах пользователей и ресурсов интернет / К. В. Воронцов, К. В. Рудаков, В. А. Лексин, А. Н. Ефимов // *Искусственный Интеллект*, 2006. Т. 2. С. 285–288.
- [4] Рудаков К. В., Черепнин А. А., Чехович Ю. В. О метрических свойствах пространств задач классификации // *Доклады Академии наук*, 2007. Т. 416. С. 457–460.
- [5] Стрижов В. В., Кузнецов М. П., Рудаков К. В. Метрическая кластеризация последовательностей аминокислотных остатков в ранговых шкалах // *Математическая биология и биоинформатика*, 2012. Т. 7. С. 345–359.
- [6] Деца М. М., Лоран М. Геометрия разрезов и метрик. — М.: МЦНМО, 2001. С. 736.
- [7] Hall P. On representatives of subsets // *J. Lond. Math. Soc.*, 1935. — Vol. 10. — P. 26–30.

References

- [1] Arkadev A. G., Braverman E. M. 1971. *Learning in pattern classification machines*. Nauka.
- [2] Zagoruiko N. G. 1998. Compactness and λ -compactness hypotheses in data analysis methods. *Sibirskii Zhurnal Industrial'noi Matematiki* 1(1):114–126.
- [3] 2006. Vorontsov, K. V., K. V. Rudakov, V. A. Leksin, A. N. Efimov, eds. Web usage mining based on web users and web sites similarity measures. *Artificial Intelligence*. Donetsk. 2:285–288.
- [4] Rudakov K. V., Cherepnin A. A., Chekhovich Y. V. 2007. On metric properties of spaces in classification problems. *Doklady Mathematics* 76:790–793.
- [5] Strijov V. V., Kuznetsov M. P., Rudakov K. V. 2012. Rank-scaled metric clustering of amino-acid sequences. *Matematicheskaya Biologiya i Bioinformatika [Mathematical Biology and Bioinformatics]* 7(1):345–359.
- [6] Deza M., Laurent M. 1997. *Geometry of cuts and metrics*. Springer. Vol. 15.
- [7] Hall P. 1935. On representatives of subsets. *J. Lond. Math. Soc.* 10:26–30.

Алгебраические свойства операторов распознавания в моделях зрительного восприятия*

А. И. Панов

pan@isa.ru

Институт системного анализа Российской академии наук (ИСА РАН), Москва, Россия

В статье рассматривается задача моделирования зрительного восприятия, приводятся основные принципы построения нисходящих моделей. На основе этих принципов приводится формальное определение распознающего блока, как базового математического объекта большинства моделей восприятия. Описывается алгоритм его работы, на основе которого строятся операторы распознавания. Проводится постановка классической статической и динамической задач распознавания и исследуются свойства корректности построенных операторов с использованием алгебраического подхода Ю. И. Журавлева.

Ключевые слова: компьютерное когнитивное моделирование; распознавание образов; иерархическое динамическое распознавание; корректность операторов распознавания

Algebraic properties for recognition operators in modeling visual perception*

A. I. Panov

Institute for Systems Analysis RAS, Moscow, Russia

The article discusses the problem of visual perception modeling and presents the main principles of top-down models developed as a result of neurophysiological research. These principles include (a) hierarchical information representation; (b) the ability to predict the type of the incoming signal; (c) the ability to recognize both static and dynamic scenes; and (d) the controllability of the perception process.

The basis of most perception models is a so-called recognizing unit. In this paper, it is formally defined according to the described principles. Recognition operators are built upon the recognizing unit algorithm which is proposed in the article.

Classical static and dynamic recognition problems are formulated. Correctness of the operators is examined using Yu. I. Zhuravlev's algebraic theory. The article presents theorems about correctness of several operator classes. These theorems indicate the possibility to build a hierarchy of basic elements that can recognize the incoming signals. This, in its turn, indicates the existence of the corresponding learning algorithm that uses the recognizing blocks.

The obtained results prove that (a) the recognizing blocks algorithm properly models human perception subsystem and does not contradict the modern pattern recognition theory; and (b) it is possible to use Yu. I. Zhuravlev's algebraic theory to verify if the operators are correct.

Keywords: computational cognitive modeling; pattern recognition; dynamic hierarchical recognition; correctness of recognition operators

Введение

В отличие от классической задачи распознавания (классификации) образов, задача моделирования зрительного восприятия человека и высших животных имеет ряд существенных особенностей. В моделях восприятия, так же как и при распознавании, каждому входному сигналу ставится в соответствие класс (категория, понятие), которому принадлежит данный сигнал. Однако основным критерием качества работы системы, реализующей модель восприятия, является биологическое правдоподобие, а не точность и полнота классификации. Под биологическим правдоподобием понимается два критерия: совпадение определенных параметров поведения системы с параметрами поведения человека в условиях конкретной задачи и построение структуры системы и ее основных компонент в соответствии с физиологическими и нейрофизиологическими особенностями строения функциональных систем [1] организма человека, отвечающих за тот или иной моделируемый аспект процесса восприятия.

Наибольший интерес представляют в настоящее время так называемые биологически оправданные (biological-inspired) модели зрительного восприятия, которые за основу берут нейрофизиологические данные о строении зрительных участков коры головного мозга и строении глаза человека. В данном направлении большое внимание уделяется такому аспекту восприятия, как внимание, системы моделирования которого получили в последнее время широкое практическое применение не только в компьютерном зрении, но и в таких областях робототехники, как навигация, локализация, человеко-машинное взаимодействие [2]. В биологически оправданных моделях, несмотря на широкое распространение нейросетевой парадигмы, в качестве математического аппарата используются не только нейронные и байесовские сети, но также обобщенные ресурсные и фильтрующие функции, которые строятся по данным когнитивных экспериментов и, как предполагается, моделируют вычисления, производимые отдельным нейроном или их ансамблем в процессе восприятия.

Наибольший интерес с точки зрения моделирования восприятия и внимания у человека представляют так называемые нисходящие (top-down) подходы, в которых делается попытка связывания высокоуровневой модели мира и низкоуровневых процессов распознавания. Иными словами, в таких подходах предпринимается попытка создания системы, которая обладает как моделью представления знаний, подтверждаемой психологическими и лингвистическими данными (например, такой как в работе [3]), так и моделью процессов восприятия и внимания, подтверждаемой в свою очередь нейрофизиологическими данными. При такой постановке задачи, возникает большое количество различных подзадач, одной из которых является исследование математических свойств формальных объектов, лежащих в основе моделей, разрабатываемых в рамках нисходящих подходов.

В данной статье приводятся основные принципы работы биологически оправданных моделей восприятия, используемые для построения формальных математических объектов. Проводится исследование этих объектов с использованием алгебраического подхода, разработанного Ю. И. Журавлевым для построения общей теории операторов распознавания [4], в частности проводится анализ корректности получаемых множеств биологически оправданных операторов распознавания.

Основные принципы моделей зрительного восприятия

В данной статье будут рассматриваться математические объекты, которые возникают при описании моделей зрительного восприятия, построенных на следующих основных принципах:

- 1) иерархичность;
- 2) способность выдвигать гипотезы;
- 3) способность распознавать как динамические так и статические явления;
- 4) управляемость.

Первый принцип был выдвинут в работах когнитивных психологов Трисман и Джедд [5] и заключается в том, что на уровне работы сетчатки имеется набор базовых признаков или протобъектов (на уровне вторичных зрительных отделов коры головного мозга) [6], из которых в процессе научения образуются более сложные признаки. Из полученных сложных признаков строятся еще более сложные и т. д. При этом процесс восприятия представляет собой последовательную активацию части получающейся иерархии, начиная с базовых признаков и заканчивая сложным объектом, предъявляемым зрительной системе. Основным критерием принадлежности разных признаков одному объекту (сложному признаку) является пространственная и временная когерентность. Иерархичность процесса восприятия также проявляется и в функциональной иерархичности коры головного мозга, что подтверждается большим количеством нейрофизиологических данных [7, 8].

Основной задачей системы, моделирующей восприятие, на каждом уровне иерархии, таким образом, становится выявление повторяющихся временных и пространственных шаблонов в поступающем наборе сигналов и низкоуровневых признаков.

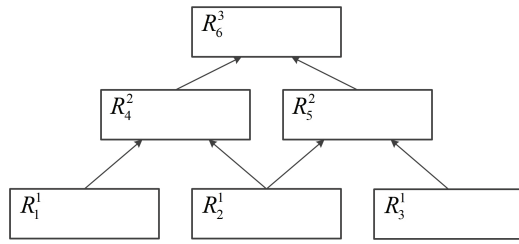
По данным анализа движения глаз испытуемых доказано, что любой процесс восприятия, как динамического так и статического явления, представляет собой развернутый во времени процесс, каждый этап которого с той или иной степенью точности предсказывается на основе предыдущих этапов [7, 9]. Именно в этом заключается второй принцип: модель должна включать в себя процессы выдвижения гипотез о том, какая часть иерархии признаков будет активирована в следующий момент времени.

Третий принцип определяет важность параметра времени: модель должна с самого начала уметь работать с меняющимися во времени признаками, не выделяя явно случай статического изображения. Наконец, четвертый принцип основан на теории активного зрения и том факте, что каждый этап распознавания признака на каком-либо уровне иерархии в процессе восприятия чередуется с активным этапом моторной реакции. Особенно ярко этот факт проявляется в случае зрительного восприятия при наблюдении саккадических движений глаза.

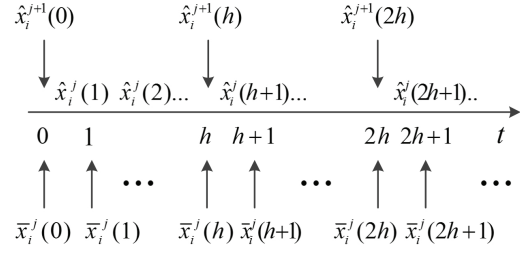
Учитывая перечисленные принципы, на которых строятся большинство существующих моделей восприятия (не только зрительного), в следующем параграфе вводится определение распознающего блока, являющегося основным структурным элементом данных моделей. Далее приводится классическая постановка задачи распознавания, множество получаемых операторов распознавания исследуется на корректность относительно поставленной задачи.

Распознающий блок

Будем рассматривать сложный математический объект R_i^j , который будем называть распознающим блоком уровня j с индексом i или просто распознающим блоком. Каждый распознающий блок, исходя из своего названия, распознает, или, как мы будем говорить,



(а) Пример иерархии распознающих блоков

(б) Циклы вычисления распознающего блока R_i^j **Рис. 1.** Функционирование распознающих блоков

измеряет, некоторые признаки. Измерение заключается в сопоставлении признака — весовому значению, характеризующему тот факт, удастся ли собрать (измерить) признак из составляющих его низкоуровневых входных признаков. Такой вес будем называть весом присутствия признака во входном векторе. Входной вектор, в свою очередь, представляет собой весовой вектор присутствия низкоуровневых признаков, по которым измеряются выходные признаки. Распознающий блок обладает состоянием, которое представляет собой также весовой вектор присутствия входных признаков, но в следующий момент времени. Такой вектор будем называть вектором ожиданий. Запишем все вышесказанное строго.

Пусть заданы множества $\{R_i^j\}$ и $\{f_k\}$. Множество $\{R_i^j\}$ будем называть совокупностью распознающих блоков, а множество $\{f_k\}$ — совокупностью допустимых признаков. Введем бинарное отношение $\dashv$, определенное на паре множеств $\{f_k\}$ и $\{R_i^j\}$, и будем читать $f_k \dashv R_i^j$ как «признак f_k измеряется распознающим блоком R_i^j ». Множество всех измеряемых распознающим блоком R_i^j признаков будем обозначать F_i^{*j} , т.е. $\forall f^* \in F_i^{*j} f^* \dashv R_i^j, F_i^{*j} \subseteq \{f_k\}$.

Рассмотрим связный ориентированный (ярусный) граф $G_R = (V, E)$, где V — множество вершин, E — множество ребер. Каждую вершину v , принадлежащую j -ому ярусу графа G_R , будем связывать с соответствующим распознающим блоком R_i^j уровня j , а ребро $e = (v, u) \in E$ будем интерпретировать как иерархическую связь между соответствующим вершине v дочерним блоком $R_{i_1}^{j_1}$ и соответствующим вершине u блоком-родителем $R_{i_2}^{j_2}$ (рис. 1(а)).

Рассмотрим распознающий блок R_i^j . Определим множество $F_i^j \subseteq \{f_k\}$ таких признаков, что для любого $f \in F_i^j$ существует распознающий блок R_k^{j-1} уровня $j-1$, дочерний по отношению к блоку R_i^j , такой, что $f \dashv R_k^{j-1}$. Такое множество F_i^j будем называть совокупностью входных признаков распознающего блока R_i^j . Для каждого признака $f^* \in F_i^{*j}$ введем функцию измерения $\hat{f}(x_1, \dots, x_q) = x^*$, где $x^* \in [0, 1]$ — весовое значение присутствия измеряемого признака f^* , а $x_1, \dots, x_q \in [0, 1]$ — весовое значение присутствия признаков из множества входных признаков F_i^j . Множество таких функций для распознающего блока R_i^j обозначим как $\hat{F}_i^j$.

Опишем распознающий блок R_i^j с точки зрения классической теории динамических систем [10]. Пусть мощность множества измеряемых признаков F_i^{*j} и множества функций измерения $\hat{F}_i^j$ равна l , а мощность множества входных признаков F_i^j равна q . Введем упорядоченное множество локальных моментов времени T_i^j для распознающего блока R_i^j . Для каждого распознающего блока определим характерный масштаб времени h_i^j , за который происходит один цикл вычисления в распознающем блоке R_i^j .

В начале s -го цикла вычисления (момент времени $\tau_s \in T_i^j$) распознающий блок R_i^j получает на вход вектор длины l ожиданий $\hat{x}_i^{j+1}(\tau_s)$, вычисляемый по формуле среднего от векторов ожиданий, получаемых от родительских относительно блока R_i^j распознающих

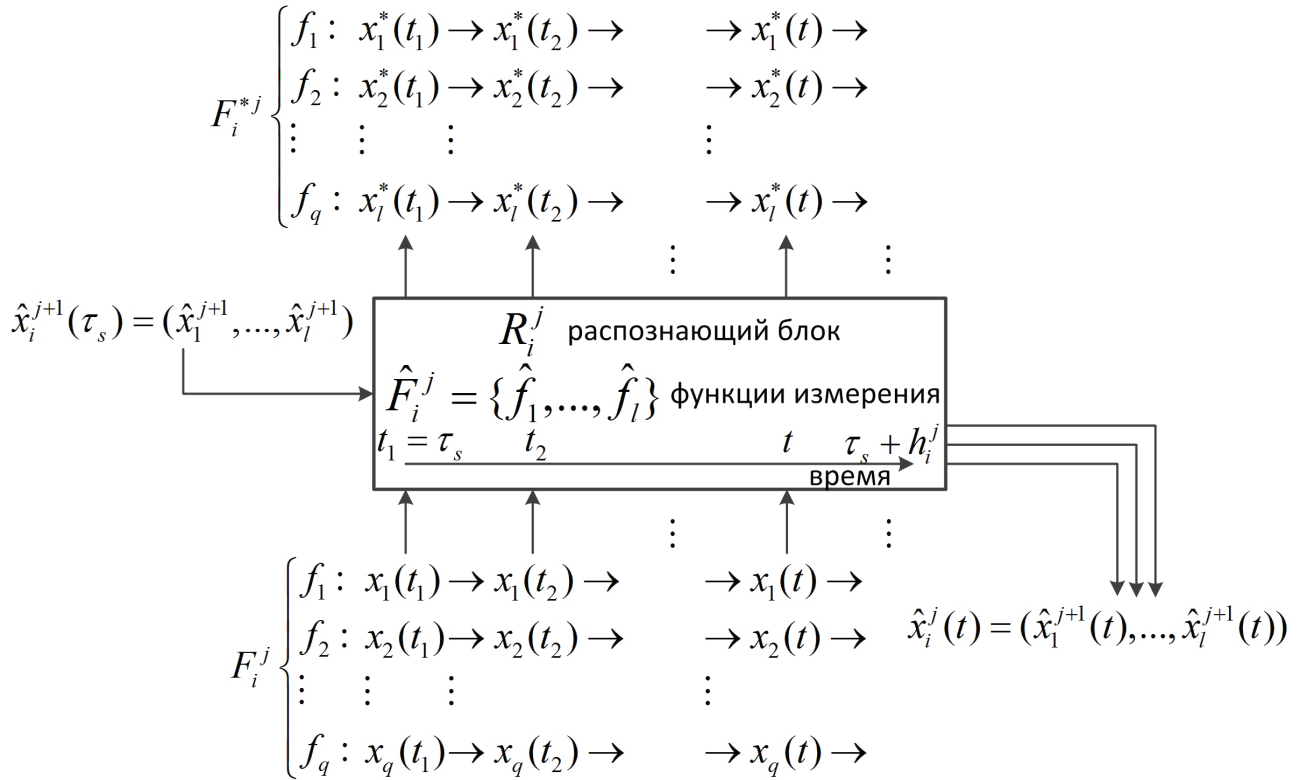


Рис. 2. Схема входных и выходных отображений распознающего блока

блоков R_k^{j+1} :

$$\hat{x}_i^{j+1}(\tau_s) = \frac{1}{N} \sum_{k \in K^{j+1}} \hat{x}_k^{j+1}(\tau_s),$$

где N — количество родительских блоков, K^{j+1} — множество индексов родительских относительно R_i^j распознающих блоков. Далее в каждый момент времени $t \in T_i^j$, $\tau_s \leq t \leq \tau_s + h_i^j$, распознающий блок R_i^j получает на вход весовой вектор $\bar{x}_i^j(t)$ длины l присутствия входных признаков из множества F_i^j , вычисляет выходной весовой вектор $\bar{x}_i^{*j}(t)$ длины l присутствия измеряемых признаков из множества F_i^{*j} , вычисляет вектор длины q ожиданий $\hat{x}_i^j(t)$ присутствия входных признаков в следующий момент времени (рис. 1(б)).

Обозначим множество возможных мгновенных значений выходных векторов распознающего блока R_i^j как X_i^{*j} . Очевидно, что X_i^{*j} является векторным пространством. Обозначим множество возможных мгновенных значений весовых векторов присутствия входных признаков как X_i^j . Очевидно, что X_i^j также является векторным пространством. Определим входное воздействие $\omega : T \rightarrow X_i^j$ и выходную величину $\gamma : T \rightarrow X_i^{*j}$ в смысле теории динамических систем. Будем считать, что множество всех возможных мгновенных значений векторов ожиданий образует множество состояний $\hat{X}_i^j$ распознающего блока R_i^j . Определим функцию переходов $\varphi(t; \tau_s, \hat{x}_i^{j+1}, \omega) = \hat{x}_i^j$ в смысле теории динамических систем. Множество $\hat{X}_i^j$ в таком случае интерпретируется как множество состояний распознающего блока R_i^j . Также определим выходное отображение $\eta : T \times \hat{X}_i^j \rightarrow X_i^{*j}$ в смысле теории динамических систем, определяющее выходные векторы $\bar{x}_i^{*j}(t) = \eta(t, \hat{x}_i^j(t))$ (рис. 2).

Будем рассматривать распознающий блок R_i^j как динамическую систему с дискретным временем, т. е. считать множество моментов времени T множеством целых чисел. Каждой функции измерения $\hat{f}_k$ из множества $\hat{F}_i^j$ будем ставить в соответствие набор матриц пред-

сказания $Z_k = \{Z_1^k, \dots, Z_m^k\}$ размерности $q \times h_i^j$, где h_i^j - характерное время распознающего блока R_i^j . Столбец $\bar{z}_u^r = (z_{u1}^r, \dots, z_{uq}^r)$ матрицы Z_r^k интерпретируется как вектор предсказания присутствия входных признаков из множества F_i^j в момент времени $\tau_s + u$, при этом $z_{uv}^k \in \{0, 1\}$, т.е. вектор $\bar{z}_u^r$ является булевым вектором. Сама матрица Z_r^k задает, таким образом, последовательность событий, наличие которой свидетельствует о присутствии измеряемого функцией f_k признака. Множество всех матриц предсказания распознающего блока R_i^j будем обозначать как $\mathcal{Z}_i^j$.

В листинге 1 приведен пороговый алгоритм $\mathfrak{A}_{th}$ вычислительного цикла распознающего блока, в котором рассчитываются значения функции переходов $\varphi(\tau_s + t; \tau_s, \hat{x}_i^{j+1}, \omega)$, $0 \leq t \leq h_i^j$, и выходного отображения $\eta(\tau_s + t, \hat{x}_i^j(\tau_s + t))$.

Статическая задача классификации

В начале рассмотрим статический случай, т.е. зафиксируем момент времени t , равный началу некоторого s -го вычислительного цикла τ_s , и будем считать, что характерное время h_i^j распознающего блока R_i^j равно 1, т.е. все матрицы из множества $\mathcal{Z}_i^j$ состоят из одного столбца. В этом случае, распознающий блок R_i^j можно рассматривать как статический оператор распознавания $R_i^j(\hat{x}_i^{j+1}, \mathcal{Z}_i^j, \bar{x}_i^j) = \bar{x}_i^{*j}$. Напомним, что $\bar{x}_i^{*j}$ — это весовой вектор присутствия измеряемых признаков $f_1^*, \dots, f_l^*$ из множества F_i^{*j} . Далее кратко будем записывать $R(\hat{x}, \mathcal{Z}, \bar{x}) = \bar{x}^*$, опуская индексы j и i .

Введем совокупность задач $\{Q\}$ аналогично работе [4]. Задача $Q(\hat{x}, \bar{x}, \alpha_1, \dots, \alpha_l) \in \{Q\}$ состоит в построении оператора, вычисляющего по поступившему вектору ожиданий $\hat{x}$ и входному вектору $\bar{x}$ значения $\alpha_1, \dots, \alpha_l \in \{0, 1\}$ присутствия признаков $f_1^*, \dots, f_l^*$. Другими словами, искомый алгоритм $\mathcal{A}^*$ переводит набор $(\hat{x}, \bar{x})$ в вектор $\bar{\alpha} = (\alpha_1, \dots, \alpha_l)$, который будем называть информационным вектором входного вектора $\bar{x}$. Пусть множество $\{\mathcal{A}\}$ состоит из алгоритмов, переводящих пары $(\hat{x}, \bar{x})$ в векторы $\bar{\beta}$, составленные из элементов 0, 1, Δ : $\mathcal{A}(\hat{x}, \bar{x}) = \bar{\beta}$. Если $\beta_i \in \{0, 1\}$, то β_i — значение величины α_i , вычисленное алгоритмом $\mathcal{A}$. Если $\beta_i = \Delta$, то алгоритм $\mathcal{A}$ не вычислил значение α_i .

Определение 1. Алгоритм $\mathcal{A}$ называется корректным для задачи Q , если выполнено равенство

$$\mathcal{A}(\hat{x}, \bar{x}) = \bar{\alpha}.$$

Алгоритм $\mathcal{A}$, не являющийся корректным для Q , называется некорректным.

Далее будем считать, что множество $\{\mathcal{A}\}$ является совокупностью, вообще говоря, некорректных алгоритмов.

Утверждение 1 (аналог теоремы 1 из [4]). Каждый алгоритм $\mathcal{A} \in \{\mathcal{A}\}$ представим как последовательность выполнения алгоритмов R и C , где $R(\hat{x}, \bar{x}) = \bar{x}^*$, $\bar{x}^*$ — вектор действительных чисел, $C(\bar{x}^*) = \bar{\beta}$, $\beta_i \in \{0, 1, \Delta\}$.

Доказательство. Пусть D — алгоритм перехода вектора $\bar{\beta}$ к числовому вектору $\bar{y}$. В качестве D можно рассмотреть, например, $y_i = \beta_i$, если $\beta_i \in \{0, 1\}$, и $y_i = 1/2$, если $\beta_i = \Delta$. Очевидно, что существует обратный алгоритм D^{-1} перехода от $\bar{y}$ к $\bar{\beta}$. Положим $R = \mathcal{A} \cdot D$, $C = D^{-1}$. Тогда очевидно, что $\mathcal{A} = R \cdot C = (\mathcal{A} \cdot D) \cdot D^{-1} = \mathcal{A}$. ■

Из утверждения 1 следует, что множество алгоритмов $\{\mathcal{A}\}$ порождает множества $\{R\}$ и $\{C\}$, которые будем называть множеством операторов распознавания и множеством решающих правил, соответственно. В качестве операторов из множества $\{R\}$ будем рассматривать операторы $R(\hat{x}, \mathcal{Z}, \bar{x})$.

Алгоритм 1 Пороговый алгоритм $\mathcal{A}_{th}$ **Вход:** $\tau_s, \hat{x}_i^{j+1}(\tau_s), \omega$ **Выход:** φ, η

- 1: $\hat{F}^* = \emptyset$ // множество активных функций измерения
- 2: $Z^* = \emptyset$ // множество подходящих матриц предсказания
- 3: $t = 0$
- 4: $c_1 \in (0, 1), c_2 \in (0, 1)$ // константы — параметры алгоритма
- 5: **для всех** компонент $\hat{x}_{ik}^{j+1}$ вектора $\hat{x}_i^{j+1}(\tau_s) = (\hat{x}_{i1}^{j+1}, \hat{x}_{i2}^{j+1}, \dots, \hat{x}_{il}^{j+1})$
- 6: **если** $\hat{x}_{ik}^{j+1} \geq c_1$ **то**
- 7: $\hat{F}^* := \hat{F}^* \cup \{\hat{f}_k\}$ // помещаем соответствующую функцию измерения в множество активных функций измерения
- 8: **для всех** функций измерения $\hat{f}_k \in \hat{F}^*$
- 9: **для всех** матриц предсказания Z_r^k из соответствующего функции измерения $\hat{f}_k$ множества Z_k
- 10: **если** $(\|\bar{z}_1^r - \bar{x}_i^j\|)/(\|\bar{z}_1^r\| + \|\bar{x}_i^j\|) < c_2$ **то**
- 11: $Z^* := Z^* \cup \{Z_r^k\}$ // пополняем множество подходящих матриц предсказания
- 12: $\hat{x}_i^j(\tau_s) = \hat{x}_i^{j+1}(\tau_s)$
- 13: **пока** $t \leq h_i^j - 1$
- 14: $\bar{x}_i^j = \omega(\tau_s + t)$
- 15: $\bar{x}_{is}^{*j} = (0, \dots, 0)$
- 16: **для всех** компонент x_{is}^{*j} вектора $\bar{x}_i^{*j} = (x_{i1}^{*j}, \dots, x_{il}^{*j})$
- 17: $x_{is}^{*j} = |Z^*|_s / \max_{\hat{f}_v} |Z^*|_v$ // $|Z^*|_s$ — количество матриц предсказания, соответствующих функции измерения $\hat{f}_s$
- 18: $\eta(\tau_s + t, \hat{x}_i^j(\tau_s + t)) = \bar{x}_i^{*j}$
- 19: **для всех** матриц предсказания Z_r^k из множества Z^*
- 20: **если** $(\|\bar{z}_{t+2}^r - \omega(\tau_s + t + 1)\|)/(\|\bar{z}_{t+2}^r\| + \|\omega(\tau_s + t + 1)\|) \geq c_2$ **то**
- 21: $Z^* := Z^* \setminus \{Z_r^k\}$ // исключаем не подходящие на следующем шаге матрицы предсказания
- 22: $t = t + 1$
- 23: **если** $t \leq h_i^j - 1$ **то**
- 24: $\hat{x}_i^j = (0, \dots, 0)$ // следующее состояние
- 25: **для всех** компонент $\hat{x}_{is}^j$ вектора ожиданий $\hat{x}_i^j = (\hat{x}_{i1}^j, \hat{x}_{i2}^j, \dots, \hat{x}_{iq}^j)$
- 26: $\hat{x}_{is}^j = \frac{1}{|Z^*|} \sum_{\hat{f}_v \in \hat{F}^*} \sum_{Z_r^v \in Z^*} \hat{x}_v^{j+1} \cdot z_{(t+1)s}^r$ // $|Z^*|$ — мощность множества подходящих матриц предсказания Z^* , $\bar{z}_{t+1}^r = (t+1)$ -й столбец входящей в множество Z^* матрицы предсказаний Z_r^v признака f_v
- 27: $\varphi(\tau_s + t; \tau_s, \hat{x}_i^{j+1}, \omega) = \hat{x}_i^j(\tau_s + t) = \hat{x}_i^j$

Определение 2. Решающее правило C^* называется корректным на множестве входных векторов X , если для всякого вектора $\bar{x}$ из X существует хотя бы один числовой вектор $\bar{x}^*$ такой, что $C^*(\bar{x}^*) = \bar{\alpha}$, где $\bar{\alpha}$ — информационный вектор входного вектора $\bar{x}$.

В множестве операторов $\{R\}$ введем операции умножения на скаляр, сложения и умножения. Пусть r' — скаляр, $R', R'' \in \{R\}$. Определим операторы $r' \cdot R'$, $R' + R''$ и $R \cdot R''$ следующим образом:

$$r' \cdot R' = (r' \cdot x_1^{*'}, \dots, r' \cdot x_l^{*'}); \quad (1)$$

$$R' + R'' = (x_1^{*'} + x_1^{*''}, \dots, x_l^{*'} + x_l^{*''}); \quad (2)$$

$$R' \cdot R'' = (x_1^{*'} \cdot x_1^{*''}, \dots, x_l^{*'} \cdot x_l^{*''}). \quad (3)$$

Утверждение 2. Замыкание $L\{R\}$ множества $\{R\}$ относительно операций (1) и (2) является векторным пространством.

Утверждение 3. Замыкание $\mathfrak{U}\{R\}$ множества $\{R\}$ относительно операций (1), (2) и (3) является ассоциативной линейной алгеброй с коммутативным умножением.

Определение 3. Множества $L\{A\}$ и $\mathfrak{U}\{A\}$ алгоритмов $\mathcal{A} = R \cdot C^*$ соответственно таких, что $R \in L\{R\}$ и $R \in \mathfrak{U}\{R\}$, соответственно называются линейными и алгебраическими замыканиями множества $\{A\}$.

Зафиксируем пару $(\hat{x}, \bar{x})$ вектора ожидания и входного вектора. Аналогично [4] будем рассматривать задачи $Q(\hat{x}, \bar{x})$, обладающие следующим свойством относительно множества операторов распознавания $\mathcal{R}$.

Определение 4. Если множество векторов $\{R(\hat{x}, \bar{x})\}$, где R пробегает некоторое множество операторов распознавания $\mathcal{R}$, содержит базис в пространстве числовых векторов длины l , то задача $Q(\hat{x}, \bar{x}, \bar{\alpha})$ называется полной относительно $\mathcal{R}$.

Утверждение 4 (аналог теоремы 2 из [4]). Если множество задач $\{Q\}$ состоит лишь из задач, полных относительно $\mathfrak{R}$, то линейное замыкание $L\{R \cdot C^*\}$ (C^* — произвольное фиксированное корректное решающее правило, R пробегает множество $\mathcal{R}$) является корректным относительно $\{Q\}$.

Доказательство. При фиксированном l пространство числовых векторов длины l состоит из l векторов. В множестве $\{R(\hat{x}, \bar{x})\}$ выделим совокупность базисных векторов $e_1, \dots, e_l$. Тогда существуют числа $c_1, \dots, c_l$ такие, что

$$\bar{x}^* = \sum_{i=1}^l c_i \cdot e_i,$$

где $\bar{x}^*$ — такой вектор, что $C^*(\bar{x}^*) = \bar{\alpha}$ — информационный вектор входного вектора $\bar{x}$. Из корректности C^* следует, что такой вектор существует. Разложение вектора $\bar{x}^*$ возможно в силу того, что векторы $e_1, \dots, e_l$ образуют базис в пространстве числовых векторов длины l . Если векторы $e_1, \dots, e_l$ построены из пары $(\hat{x}, \bar{x})$ при помощи операторов $R_1, \dots, R_l$, то алгоритм $\mathcal{A} = (\sum_{i=1}^l c_i \cdot e_i) \cdot C^*$ переводит $(\hat{x}, \bar{x})$ в информационный вектор $\bar{\alpha}$. ■

Следствие 1. Пусть $\{A\}$ — совокупность некорректных алгоритмов, $\{R\}$ — соответствующее множество операторов распознавания, C^* — фиксированное корректное решающее правило. Тогда $L\{A\} = L\{R \cdot C^*\}$ является корректным относительно множества задач $\{Q\}$, если $\{Q\}$ состоит из задач, полных относительно $\{R\}$.

Будем рассматривать только такие задачи $Q(\hat{x}, \bar{x}, \bar{\alpha})$, для которых удовлетворяется следующее условие: $\exists k$ такое, что x_k является k -ым элементом вектора $\bar{x}$ и $x_k > c_1$. Такое условие является естественным, иначе вектор $\bar{x}$, в котором отсутствуют весовые значения больше c_1 , не может рассматриваться как достоверный с точки зрения порогового алгоритма 1.

Теорема 1. *Линейное замыкание $L\{\mathcal{A}\}$ семейства алгоритмов $\{\mathcal{A}\} = \{R \cdot C^*\}$ с произвольным корректным решающим правилом C^* и операторами распознавания R , определенными пороговым алгоритмом 1, является корректным на $\{Q\}$.*

Доказательство. В силу утверждения 4 достаточно доказать, что произвольная задача $Q \in \{Q\}$ является полной относительно $\{R\}$. Доказательство полноты Q состоит в прямом построении операторов $R_i^\varepsilon, i = 1, 2, \dots, l$ из $L\{R\}$, переводящих пару $(\hat{x}, \bar{x}), \hat{x} = (\hat{x}_1, \dots, \hat{x}_l), \bar{x} = (x_1, \dots, x_q)$ в числовой вектор $\bar{x}_i^* = (x_{i1}^*, \dots, x_{il}^*)$, в котором $x_{ii}^* = 1$, а $|x_{ij}^*| < \varepsilon$ при $j \neq i$. Построение проводится для любого сколь угодно малого ε .

Пусть мощность множества $\mathcal{Z}_i$ признака f_i равна N , норма $\|x\|$ равна $M \leq q$, максимальная компонента вектора $\bar{x}$ равна x_{max} . Зафиксируем величину i и коэффициенты $c_1 = \min_x \hat{x}_k, c_2 = \frac{M}{1+M}$. Рассмотрим матрицы предсказания из множеств $\mathcal{Z}_1, \dots, \mathcal{Z}_l$ признаков $f_1, \dots, f_l$, удовлетворяющие следующим условиям:

- 1) в каждой матрице предсказаний $Z_r^i \in \mathcal{Z}_i$ в столбце $\bar{z}_{t-\tau}^r = (z_{t-\tau,1}^r, \dots, z_{t-\tau,q}^r)$ компонента $z_{t-\tau,k}^r = 1$, если $x_k = x_{max}$, и $z_{t-\tau,k}^r = 0$, если $x_k < x_{max}$;
- 2) в каждой матрице предсказаний $Z_r^j \in \mathcal{Z}_j, j \neq i$ в столбце $\bar{z}_{t-\tau}^r = (z_{t-\tau,1}^r, \dots, z_{t-\tau,q}^r)$ компонента $z_{t-\tau,k}^r = 0$ при любых k .

Вычислим величину x_{ii}^* . Т.к. $c_1 = \min_{\hat{x}} \hat{x}_k$, то условие $\hat{x}_i \geq c_1$ на строчке 5 алгоритма 1 автоматически выполняется и функция измерения $\hat{f}_i$ попадает в множество $\hat{F}^*$. Из условия 1) следует, что каждая матрица $Z_r^i \in \mathcal{Z}_i$ попадает в множество Z^* на строчке 10 алгоритма 1:

$$\frac{\|\bar{z}_{t-\tau}^r - \bar{x}\|}{\|\bar{z}_{t-\tau}^r\| + \|\bar{x}\|} < \frac{\sum_k |z_{t-\tau,k}^r - x_k|}{1 + M} < \frac{M}{1 + M} = c_2,$$

так как минимум один компонент в $\bar{z}_{t-\tau}^r$ равен 1 и существует элемент $x_k > 1/2$. В этом случае $x_{ii}^* = \gamma \cdot N$, где γ — весовой коэффициент.

Оценим величины x_{ij}^* . Т.к. $c_1 = \min_{\hat{x}} \hat{x}_k$, то условие $\hat{x}_j \geq c_1$ на строчке 5 алгоритма 1 автоматически выполняется и все функции измерения $\hat{f}_j$ попадают в множество $\hat{F}^*$. Из условия 2) следует, что каждая матрица $Z_r^j \in \mathcal{Z}_j$ не попадает в множество Z^* на строчке 10 алгоритма 1:

$$\frac{\|\bar{z}_{t-\tau}^r - \bar{x}\|}{\|\bar{z}_{t-\tau}^r\| + \|\bar{x}\|} = \frac{M}{M} = 1 > \frac{M}{1 + M} > c_2.$$

В этом случае $x_{ij}^* = 0$.

Рассмотрим оператор распознавания $\frac{1}{\gamma \cdot N} R(\hat{x}, \mathcal{Z}, \bar{x})$, который переводит задачу Q в вектор $\bar{x}_i^*$, причем $\bar{x}_{ii}^* = 1$, а $\bar{x}_{ij}^* = 0 < \varepsilon, j \neq i$. Таким образом, условия на числовой вектор $\bar{x}_i^*$ выполняются. Полнота задачи Q доказана. ■

Фиксация момента времени не в начале вычислительного цикла, а на любом другом значении $\tau_s < t \leq \tau_s + h_i^j$, приводит к операторам вида $R_i^j(\hat{x}_i^j(t), \mathcal{Z}_i^j, \bar{x}_i^j(t))$, который кратко будем записывать R^t . Для этих операторов постановка задачи распознавания выглядит таким же образом как и для операторов R , формулировки определений полноты и корректности идентичны. Теорема о корректности линейного замыкания $L\{R^t \cdot C^*\}$ доказывается аналогично.

Динамическая задача классификации

Теперь рассмотрим динамическую постановку задачи. Будем фиксировать не конкретный момент времени t , а промежуток времени $\Delta t = [\tau_s, \tau_s + h_i^j]$. В этом случае распознающий блок R_i^j можно рассматривать как динамический оператор распознавания

$\hat{R}_i^j(\hat{x}_i^{j+1}(\tau_s), \mathcal{Z}_i^j, \omega_{\Delta t}) = \gamma_{\Delta t}$, принимающий функцию входного воздействия ω , ограниченную на промежутке времени Δt , и выдающую функцию выходной величины γ на том же временном промежутке. Так как мы предполагаем время дискретным, т.е. множество моментов времени T является множеством целых чисел, то действие динамического оператора $\hat{R}_i^j$ можно заменить последовательным действием статических операторов $R_i^j(\hat{x}_i^{j+1}(\tau_s), \mathcal{Z}_i^j, \bar{x}_i^j(\tau_s)), R_i^j(\hat{x}_i^j(\tau_s + 1), \mathcal{Z}_i^j, \bar{x}_i^j(\tau_s + 1)), \dots, R_i^j(\hat{x}_i^j(\tau_s + h_i^j - 1), \mathcal{Z}_i^j, \bar{x}_i^j(\tau_s + h_i^j - 1))$, в результате выдающих последовательность $\{\bar{x}_i^{*j}(t)\} = \{\bar{x}_i^{*j}(\tau_s), \bar{x}_i^{*j}(\tau_s + 1), \dots, \bar{x}_i^{*j}(\tau_s + h_i^j - 1)\}$. Так как параметр h_i^j фиксирован, то конечные последовательности векторов $\omega_{\Delta t}$ и $\gamma_{\Delta t}$ можно считать матрицами размерности $l \times h_i^j$. Далее будем опускать индексы i и j .

Формулировка задачи в динамическом случае будет выглядеть следующим образом: задача $\hat{Q}(\hat{x}, \omega_{\Delta t}, \bar{\alpha})$ состоит в построении алгоритма $\hat{A}$, вычисляющего по поступившему начальному вектору ожиданий $\hat{x}$ матрице входных воздействий $\omega_{\Delta t}$ последовательность векторов $\beta_{\Delta t}$, монотонно сходящуюся к информационному вектору $\bar{\alpha}$, т.е. искомый оператор распознавания $\hat{R}$ должен выдавать матрицу весовых значений присутствия измеряемых признаков $\gamma_{\Delta t}$, столбцы которой должны сходиться (с учетом корректного решающего правила) к информационному вектору: $\lim_{t \rightarrow \tau_s + h} \bar{x}^*(t) = \bar{\alpha}$. Введем соответствующие определения.

Определение 5. Алгоритм $\hat{A}(\hat{x}, \bar{x}) = \beta_{\Delta t} = (\bar{\beta}_1, \dots, \bar{\beta}_h)$ называется корректным для задачи $\hat{Q}$, если выполнено условие

$$\|\bar{\beta}_1 - \bar{\alpha}\| \geq \|\bar{\beta}_2 - \bar{\alpha}\| \geq \dots \geq \|\bar{\beta}_h - \bar{\alpha}\| = 0.$$

$\|\bar{\beta}_i - \bar{\alpha}\| = \sum_j (\beta_{ij} - \alpha_j)$, где $\beta_{ij} - \alpha_j = 0$, если $\beta_{ij} = \alpha_j$, $\beta_{ij} - \alpha_j = \frac{1}{2}$, если $\beta_{ij} = \Delta$, и $\beta_{ij} - \alpha_j = 0$ иначе. Алгоритм $\hat{A}$, не являющийся корректным для $\hat{Q}$, называется некорректным.

Утверждение 5. Каждый алгоритм $\hat{A} \in \{\hat{A}\}$ представим как последовательность выполнения алгоритмов $\hat{R}$ и $\hat{C}$, где $\hat{R}(\hat{x}, \mathcal{Z}, \omega_{\Delta t}) = \gamma_{\Delta t}$, $\gamma_{\Delta t}$ — матрица действительных чисел, $\hat{C}(\gamma_{\Delta t}) = \beta_{\Delta t}$, $\beta_{\Delta t}$ — матрица значений $\beta_{ij} \in \{0, 1, \Delta\}$.

Доказательство. Проводится аналогично утверждению 1 с тем отличием, что рассматриваются не отдельные векторы, а матрицы. ■

Корректное решающее правило $\hat{C}^*$ для матрицы $\gamma_{\Delta t}$ определяется через набор корректных правил для векторов $(\hat{C}_1^*, \dots, \hat{C}_h^*)$ таких, что $\|C_1^*(\bar{x}^*(\tau_s)) - \bar{\alpha}\| \geq \|C_2^*(\bar{x}^*(\tau_s + 1)) - \bar{\alpha}\| \geq \dots \geq \|C_h^*(\bar{x}^*(\tau_s + h - 1)) - \bar{\alpha}\| = 0$. В простейшем случае $\forall i C_i^*(\bar{x}^*(\tau_s + i)) = \bar{\alpha}$. Аналогично статическому случаю вводятся определения линейного $L\{\hat{R}\}$ и алгебраического $\mathcal{U}\{\hat{R}\}$ замыкания над множеством $\{\hat{R}\}$.

Зафиксируем начальный вектор ожиданий $\hat{x}$ и последовательность входных векторов $\omega_{\Delta t}$. Если, как и в статическом случае, будем рассматривать только такие задачи $\hat{Q}(\hat{x}, \omega_{\Delta t}, \bar{\alpha})$, для которых в матрице $\omega_{\Delta t}$ в каждом столбце с номером $s \exists k$ такое, что x_{sk} является k -м элементом вектора $\bar{x}(\tau_s + s)$ и $x_{sk} > c_1$, то можно сформулировать следующую теорему.

Теорема 2. Линейное замыкание $L\{\hat{A}\}$ семейства алгоритмов $\{\hat{A}\} = \{\hat{R} \cdot \hat{C}^*\}$ с произвольным корректным решающим правилом $\hat{C}^*$ и операторами распознавания $\hat{R}$, определенными пороговым алгоритмом 1, является корректным на $\{\hat{Q}\}$.

Доказательство. В силу того, что динамический оператор $\hat{R}$ представим в виде последовательного применения статических операторов R_i к столбцам матрицы $\omega_{\Delta t}$, то для

доказательства теоремы необходимо подобрать такие операторы R_i , которые выдают последовательность $\gamma_{\Delta t}$, сходящуюся (с учетом применения простейшего корректного решающего правила $\hat{C}^* = (C_1^*, \dots, C_i^*)$) к информационному вектору $\bar{\alpha}$.

Рассмотрим алгебраическое замыкание $L\{R_i\}$ операторов вида $R_i(\hat{x}(\tau_s + i), \mathcal{Z}_i, \omega_{\Delta t}(\tau_s + i))$ с фиксированным вектором $\hat{x}(\tau_s + i)$ и $\omega(\tau_s + i)$. Из задачи $\hat{Q}(\hat{x}, \omega_{\Delta t}, \bar{\alpha})$ выделим подзадачу $Q_i(\hat{x}(\tau_s + i), \omega_{\Delta t}(\tau_s + i), \bar{\alpha})$. В силу теоремы 1 можно построить такой оператор $R_i^* \in L\{R_i\}$, что $C_i^*(R_i^*(\omega(\tau_s + i))) = \bar{\alpha}$. Формируя таким образом линейные замыкания и формулируя подзадачи для каждого момента времени, получим необходимую последовательность $\gamma_{\Delta t} = (C_1^*(R_1^*(\omega(\tau_s))), \dots, C_h^*(R_h^*(\omega(\tau_s + h - 1)))) = (\bar{\alpha}, \dots, \bar{\alpha})$, которая очевидным образом сходится к $\bar{\alpha}$. Корректность, таким образом, доказана. ■

Заключение

На основании исследуемых в статье математических свойств распознающих блоков, реализующих основные принципы моделей восприятия, можно сделать следующие выводы:

- 1) динамические характеристики моделей восприятия описываются в терминах классической теории управления;
- 2) базовые структурные элементы данных моделей представляют собой операторы распознавания, которые можно изучать в рамках классических алгебраических теорий;
- 3) базовые структурные элементы моделей восприятия обладают свойством корректности относительно входных данных и требуемых результатов классификации, что означает существования такого процесса обучения, в рамках которого будут сформирована иерархия базовых элементов, корректно распознающая (классифицирующая) поступающие сигналы.

Литература

- [1] Анохин П. К. 1975. Очерки по физиологии функциональных систем. — М.: Медицина.
- [2] Borji A., Itti L. State-of-the-Art in Visual Attention Modeling // *IEEE Trans. Pattern Anal. Machine Intelligence*, 2013. Vol. 35. No. 1. Pp. 185–207.
- [3] Осипов Г. С., Панов А. И., Чудова Н. В. Управление поведением как функция сознания. I. Картина мира и целеполагание // *Известия РАН. Теория и системы управления*, 2014. № 4. С. 83–96.
- [4] Журавлев Ю. И. Корректные алгебры над множеством некорректных (эвристических) алгоритмов. Часть I // *Кибернетика*, 1977. № 4. С. 5–17.
- [5] Triesman A. M., Gelade G. A Feature integration theory of attention // *Cognitive Psychology*, 1980. Vol. 12. Pp. 97–136.
- [6] Rensink R. The dynamic representation of scenes // *Vis. Cogn.*, 2000. No. 7. Pp. 17–42.
- [7] George D., Hawkins J. Towards a mathematical theory of cortical micro-circuits // *PLoS Comput. Biol.*, 2009. Vol. 5. No. 10. Pp. 1–26.
- [8] Болотова Ю. А., Спицын В. Г., Фомин А. Э. Применение иерархической временной памяти в распознавании изображений // *Известия Томского политехнического университета*, 2011. Т. 318. № 5. С. 60–63.
- [9] Величковский Б. М. Когнитивная наука: основы психологии познания. Т. 1. — М.: Смысл, 2006. 448 с.

- [10] Калман Р., Фалб П., Арбуб М. Очерки по математической теории систем. — М.: Мир, 1971. 400 с.

References

- [1] Anokhin P. K. 1978. *Philosophical aspects of the theory of functional systems*.
- [2] Borji A., Itti L. 2013. State-of-the-art in visual attention modeling. *IEEE Trans. Pattern Anal. Machine Intelligence* 35(1):185–207.
- [3] Osipov G. S., Panov A. I., Chudova N. V. 2014. Behavior control as a function of consciousness. I. World model and goal setting. *Int. J. Comput. Syst. Sci.* 53(4).
- [4] Zhuravlev Yu. I. Correct algebras over sets of incorrect (heuristic) algorithms. I. 1977. *Cybernetics* 13(4):489–497.
- [5] Triesman A. M., Gelade G. 1980. A feature integration theory of attention. *Cognitive Psychology* 12:97–136.
- [6] Rensink R. 2000. The dynamic representation of scenes. *Visual Cognition* 7:17–42.
- [7] George D., Hawkins J. 2009. Towards a mathematical theory of cortical micro-circuits. *PLoS Comput. Biol.* 5(10):1–26.
- [8] Bolotova Ju. A., Spicin V. G., Fomin A. E. 2011. Application of hierarchical temporal memory in image recognition. *Izvestia Tomskogo Politechnicheskogo Universiteta* 318(5):60–63. [In Russian.]
- [9] Velichkovsky B. M. 2006. *Cognitive science: Foundations of epistemic psychology*. Vol. 1.
- [10] Kalman R. E. and Falb P. L. and Arbib M. A. 1969. *Topics in mathematical system theory*.

Экспериментальное исследование методов выявления нечетких дубликатов научных публикаций

Н. В. Дербенев, Д. А. Козлюк, В. В. Никитин, В. О. Толчеев

nicvic@mail.ru, {dmitry.kozliuk, vadimirtoej}@gmail.com, tolcheevvo@mail.ru

НИИ «МЭИ», Россия, Москва, ул. Красноказарменная, 14

Рассмотрены известные методы, в том числе авторский — обобщенный коэффициент ассоциативности (ОКА), для выявления нечетких дубликатов в научных публикациях и критерии эффективности их работы. Сформулирован целевой критерий работы методов, сочетающий требования к полноте и к точности. Составлена выборка пар документов, представленных библиографическими описаниями (заголовками и аннотациями), получены экспертные оценки схожести документов в парах. Проведены эксперименты по установлению наибольшей точности результатов различных методов при ограничении на полноту. Для коэффициента ассоциативности Джаккарда и ОКА, продемонстрировавших наилучшие результаты, предложены и апробированы способы повышения точности до 74 % при полноте 90 %. Результаты проверены путем анализа полнотекстовых описаний части документов исследуемой выборки, доступных публично.

Ключевые слова: анализ текстовой информации; нечеткие дубликаты; наукометрия

Experimental Research of Near-Duplicate Detection Methods for Scientific Papers

N. V. Derbenev, D. A. Kozliuk, V. V. Nikitin, V. O. Tolcheev

Moscow Power Engineering Institute, Krasnokazarmennaya 14, Moscow, Russia

Near-duplicate detection problem focuses on determining pairs of semantically equivalent documents which differ syntactically. For scientific papers, the case in question corresponds either to plagiarism among different authors or to a single author publishing “cloned” papers to achieve higher citation rank. Given a set of document pairs with *a priori* expert opinions, efficiency of any near-duplicate detection method can be measured by the number of correct near-duplicate detections (recall) and false-positive detection count (precision). These metrics cannot be maximized simultaneously — complex criterion are used.

Improvements can be made by choosing a method with the highest precision for a given recall and then preprocessing documents in favor of specifics of the selected algorithm. An efficiency criteria limiting both recall and precision is proposed. Then, a sample set of title and abstract pairs (publically available bibliographic descriptions) is formed and expert assessments are acquired for them. After that, the sample set is used to evaluate performance of various known near-duplicate detection methods subjected to the criteria proposed.

Precision of 74% at 90% recall for Jaccard and generalized similarity coefficients appeared to be reachable by removing frequent words of authors’ vocabulary from documents’ abstracts. Generalized Similarity Coefficient (GSC) method was introduced in the authors’ previous work. Along with Winnowing, GSC scored best in method comparison without preprocessing. The results were checked by examining a subset of documents with full texts available (about 20 % of sample set). Verification confirmed high sustained precision by revealing documents with near-duplicate titles and abstracts to have identical content.

Keywords: information retrieval; near-duplicates detection; scientometrics

Введение

В настоящее время программно-алгоритмические средства, ориентированные на обнаружение неоригинальных публикаций, способны решать важные практические проблемы: устранять одинаковые тексты из выдачи поисковых систем, идентифицировать совпадающие web-страницы, исключать повторяющуюся информацию из баз данных, предотвращать публикацию идентичных статей в научных журналах, обнаруживать заимствование чужих результатов и идей.

Применение таких средств представляется эффективным для анализа массивов научной информации (статей, докладов на конференциях, диссертаций). В основе современного порядка финансирования научных работ лежат наукометрические показатели. Стремление достичь их закономерно приводит к росту публикационной активности ученых, сопровождающемуся снижением качества научных статей, распространением неоригинальных публикаций и заимствованием фрагментов. Данная работа посвящена проблеме обнаружения дубликатов и нечетких дубликатов (очень похожих по терминологическому составу документов) в массивах научных публикаций.

Отметим, что полные дубликаты могут быть достаточно просто обнаружены с помощью специализированного программного обеспечения. Основные сложности возникают при идентификации нечетких дубликатов. Эта задача лишь частично решается применением автоматизированных средств. Для выработки окончательного суждения о статусе документа (уникальный или неуникальный) необходимо задействовать экспертов.

В системах обработки и анализа информации используется два типа текстов: полнотекстовые документы и библиографические описания, которые содержат сведения об авторах, названия, аннотации, ключевые слова и другие вспомогательные данные. Чаще всего в открытом доступе имеются только библиографические описания. Применение полнотекстовых версий для обнаружения нечетких дубликатов зачастую затрудняется их отсутствием в открытых источниках. Некоторые издания практикуют платное предоставление полных текстов статей. Поэтому, несмотря на сильную усеченность и меньшую информативность библиографических описаний по сравнению с полнотекстовыми версиями, именно по первым целесообразно проводить выявление нечетких дубликатов. Вместе с тем большинство методов, которые применяются для анализа библиографических описаний, могут применяться также для выявления нечетких дубликатов среди полнотекстовых публикаций.

Формирование целевого критерия

В зависимости от специфики используемых методов в работе для описания документа применяется два вида моделей: 1) символьная последовательность, задающая порядок следования слов (символов) и определяющая их позицию в тексте; 2) «мешок слов», в котором не учитываются связи между терминами (словами) и их местоположение:

$$\mathbf{X}_j = [x_j^{(1)}, \dots, x_j^{(M)}]^T, \quad (1)$$

где i — номер термина ($i = 1, \dots, M$; M — количество терминов после удаления служебных слов), j — номер документа в выборке ($j = 1, \dots, N$; N — количество документов в выборке). Значение $x_j^{(i)}$ будет существенно зависеть от используемого метода взвешивания термина. В некоторых алгоритмах $x_j^{(i)}$ обозначает не только вес слова, но и позицию i слова в документе.

Для формализации понятия схожести двух статей введем меру близости $\rho(\mathbf{X}_j, \mathbf{X}_l)$, значения которой изменяются в интервале $[0; 1]$. Мера близости должна равняться единице в случае, если документы $\mathbf{X}_j$ и $\mathbf{X}_l$ — дубликаты, и стремиться к нулю, если $\mathbf{X}_j$ и $\mathbf{X}_l$ —

уникальные публикации. Два документа $\mathbf{X}_j$ и $\mathbf{X}_l$ являются *полными дубликатами*, если мера их близости равна единице, и два документа $\mathbf{X}_j$ и $\mathbf{X}_l$ являются *нечеткими дубликатами (дублями)*, если мера их близости превосходит экспериментально (или экспертно) установленный порог θ ($\rho(\mathbf{X}_j, \mathbf{X}_l) \geq \theta$).

Выбор конкретного значения порога осуществляется, прежде всего, в зависимости от критерия качества (целевого критерия) обнаружения нечетких дубликатов. В задачах выявления нечетких дубликатов целевой критерий обычно основывается на расчете показателя «полнота-точность».

Коэффициент полноты (recall) характеризует долю найденных (с помощью метода автоматического выявления дубликатов) неуникальных публикаций среди их общего количества в выборке (т. е. по сравнению с числом нечетких дубликатов, определенных экспертами):

$$R = \frac{a}{a + c}. \quad (2)$$

Здесь a — количество выявленных (методом) документов выборки, являющихся нечеткими дубликатами; $a + c$ — общее число нечетких дубликатов в выборке (c — количество нечетких дубликатов в выборке, которые не найдены методом).

Коэффициент точности (precision) характеризует долю нечетких дубликатов среди документов, которые определены методом в качестве неуникальных:

$$R = \frac{a}{a + b}. \quad (3)$$

Здесь $a + b$ — общее количество документов, которые идентифицированы методом в качестве неуникальных (b — количество документов, не являющихся, по мнению экспертов, нечеткими дубликатами, но отнесенных к ним методом).

Показателем, объединяющим полноту и точность, является F_1 -мера: $F_1 = \frac{2PR}{P+R}$.

Наиболее действенным способом влияния на R и P является изменение порога θ , после которого публикации считаются дубликатами. Чем ближе его значение к нулю, тем больше полнота и меньше точность. С увеличением порога наблюдается обратная зависимость: полнота уменьшается, но при этом увеличивается точность.

Рассмотрим основные варианты целевых критериев, используемых при поиске нечетких дубликатов.

- 1) Максимальное значение одного из показателей (*полноты* или *точности*) при заданных ограничениях на другой показатель:

$$\exists \theta_* \in [0; 1] : P(\theta_*)_* = \max_{\theta \in [0; 1]} P(\theta), R(\theta_*) \geq C$$

или

$$\exists \theta_* \in [0; 1] : R(\theta_*)_* = \max_{\theta \in [0; 1]} R(\theta), P(\theta_*) \geq C.$$

- 2) Максимальное суммарное значение *полноты* и *точности*:

$$\exists \theta_* \in [0; 1] : \forall \theta \in [0; 1] (R(\theta) + P(\theta)) \leq (R(\theta_*) + P(\theta_*)).$$

- 3) Баланс между *полнотой* и *точностью* (близость значений этих параметров):

$$\exists \theta_* \in [0; 1] : \forall \theta \in [0; 1] R(\theta_*) \approx P(\theta_*).$$

В публикациях по проблеме обнаружения нечетких дубликатов исследователи основное внимание уделяют достижению высокой полноты поиска, которая обычно должна быть не менее 90 % [3, 5]. Это позволяет идентифицировать в анализируемом массиве практически все имеющиеся дубликаты. Вторым по важности показателем является точность поиска, которая должна быть максимально возможной при введенном ограничении на полноту. Таким образом, целевой критерий в данной работе имеет вид: достичь наибольшей точности обнаружения нечетких дубликатов при условии, что полнота составляет не менее 90 %.

Учитывая взаимообратный характер изменения полноты и точности, достижение выше сформулированного критерия возможно, прежде всего, за счет использования дополнительной лексической информации об используемых терминах, частоте их появления, местоположении и т. п.

Краткий обзор известных методов выявления нечетких дубликатов

Несмотря на интенсивную разработку методов обнаружения нечетких дубликатов, на настоящий момент не удалось создать универсальный подход, обеспечивающий наилучшие показатели полноты и точности на различных выборках. Сложности разработки обусловлены спецификой обработки текстовой информации: отсутствием механизмов логико-математического описания смысла излагаемого материала, большой размерностью и трудоемкостью задачи, малым размером выборок и т. п.

К настоящему времени опубликовано большое число теоретических и экспериментальных исследований, направленных на изучение преимуществ и ограничений различных процедур выявления нечетких дубликатов [1, 2, 3, 4]. Имеются достаточно полные обзоры, в которых дается описание алгоритмов и приводятся результаты их сравнительного анализа [5, 6].

В большинстве публикаций принято разделять известные процедуры на два класса: синтаксические методы (анализ последовательностей, состоящих из символов, слов или предложений) и лексические методы (анализ информативных терминов). Вместе с тем, представляется целесообразным использовать более детализированные систематизации. Одна из таких систематизаций объединяет методы по принципу общего подхода к принятию решений при выявлении нечетких дубликатов: специализированные расстояния (*расстояние Хэмминга, Левенштейна, Дамерау-Левенштейна, Джаро, Джаро-Винклера*), шинглы и их модификации (*супершинглы, мегашинглы, Winnowing, SpotSigs*), процедуры на основе расчета весов терминов (*метод опорных слов, I-match*), меры близости (*коэффициенты ассоциативности, косинусодальная мера, двухуровневая мера совпадения Монге-Элкана (Monge-Elkan matching scheme), двухуровневая функция сравнения на основе «мягкого» tfidf-взвешивания*) [3, 5, 7, 8, 9, 11].

Как справедливо указывается в литературе, выбор конкретного метода в существенной степени зависит от цели разработки, особенностей предметной области, исходной информации и имеющихся ограничений. Применительно к проблеме, рассматриваемой в данной работе, при выборе методов для экспериментальных исследований использовались следующие критерии: эффективность обработки коротких документов, чувствительность к операциям редактирования (вставка, замена, удаление терминов), трудоемкость. С этих позиций были отобраны следующие хорошо известные в литературе процедуры: метод шинглов, метод Winnowing, коэффициент ассоциативности Джаккарда, расстояние Джаро-Винклера, нормированное расстояние Левенштейна, а также обобщенный коэф-

фициент ассоциативности (ОКА), предложенный и исследованный в нашей предыдущей публикации [10].

Учитывая, что большинство выбранных методов широко известны и описаны в специализированной литературе [1, 3, 4, 11], остановимся более подробно на процедуре ОКА.

Обобщенный коэффициент ассоциативности документов рассчитывается как линейная комбинация коэффициентов ассоциативности между названиями и между аннотациями. Формулы для расчета выбираются отдельно из нескольких вариантов в зависимости от схожести терминологического состава названий или аннотаций соответственно. Набор формул расчета для выбора обоснован экспериментами, в которых проводился сравнительный анализ известных коэффициентов ассоциативности и исследовались их возможные комбинации.

Введем следующие обозначения: A_{title} и A_{abstract} — число совпавших терминов соответственно в названиях и аннотациях двух анализируемых документов; B_{title} и B_{abstract} — число терминов в названиях и аннотациях, имеющих в первом документе и отсутствующих во втором; C_{title} и C_{abstract} — число терминов в названиях и аннотациях, имеющих во втором документе и отсутствующих в первом. Для расчета ОКА предложена следующая формула [10]:

$$\begin{aligned} \text{ОКА} &= \frac{1}{2}(K_1 + K_2) = \\ &= \frac{1}{2} \left\{ \frac{A_{\text{title}}}{\max(A_{\text{title}}; B_{\text{title}}; C_{\text{title}})} + \min \left(\frac{A_{\text{abstract}}}{A_{\text{abstract}} + B_{\text{abstract}}}; \frac{A_{\text{abstract}}}{A_{\text{abstract}} + C_{\text{abstract}}} \right) \right\} \quad (4) \end{aligned}$$

Коэффициент K_1 вычисляется только по терминам, встречающимся в названиях статей ($K_1 \in [0; 1]$); коэффициент K_2 рассчитывается только по терминам из аннотаций ($K_2 \in [0; 1]$). Усреднение суммы коэффициентов и приводит значения ОКА в диапазон $[0; 1]$.

В данной работе процедура выявления нечетких дубликатов осуществлялась в несколько этапов:

- 1) Составление выборок, состоящих из пар документов одного автора.
- 2) Экспертная оценка пар на наличие нечетких дубликатов.
- 3) Выявление из выборки нечетких дубликатов с помощью вышеуказанных методов.
- 4) Оценка полноты и точности методов. Определение соответствия требованию целевого критерия.
- 5) Улучшение показателя «полнота-точность» путем учета дополнительных характеристик текстовых документов.

Экспериментальные исследования методов выявления нечетких дубликатов

Для создания специализированного документального массива, в котором осуществляется поиск нечетких дубликатов, использовалась российская научная электронная библиотека eLibrary.ru. С помощью авторского указателя было найдено 1070 авторов, которые имеют не менее пяти публикаций (включая соавторство) в области автоматизации и вычислительной техники с аннотациями на русском языке. В итоге был создан документальный массив, состоящий из 7257 библиографических описаний. Для анализа выбирались наиболее тематически близкие авторские работы (для оценки близости использовался коэффициент ассоциативности Джаккарда больше 0,3 [12]). Полные дубликаты, содержащиеся в eLibrary.ru (например, из-за опечаток при указании страниц,

Таблица 1. Сравнительные результаты работы методов выявления нечетких дубликатов

Метод	Пороговое значение	Полнота	Точность
Коэффициент ассоциативности Джаккарда	0,54	0,908	0,712
Шинглы (длина 2 слова)	0,38	0,908	0,718
Жаро	0,74	0,908	0,701
Жаро-Винклер	0,74	0,908	0,701
Расстояние Левенштейна	0,51	0,939	0,697
Winnowing ($k = 2, t = 4$)	0,35	0,917	0,720
ОКА	0,57	0,908	0,724

фамилий авторов и т. п.), из рассмотрения удалялись. Далее случайным образом было отобрано 150 пар библиографических документов (общий размер выборки — 300 документов). Библиографические описания, вошедшие в выборку, были проанализированы тремя экспертами. Статьи относились к нечетким дубликатам, если все эксперты принимали согласованное решение (в противном случае тексты считались оригинальными). Всего нечеткими дубликатами эксперты сочли 98 пар (65 %). Выборка размещена по адресу: <http://uii.mpei.ru?term=49>.

Задача экспериментальных исследований заключалась в определении с помощью автоматизированных средств тех нечетких дубликатов, которые ранее были выявлены экспертами (с учетом требований целевого критерия).

Для проведения автоматизированного анализа и выявления нечетких дубликатов был разработан специализированный программный комплекс, в котором предусмотрена обработка документов, заданных как библиографическими описаниями, так и полными текстами. В комплексе реализованы все вышеуказанные методы обнаружения дублей и основные процедуры предварительной обработки текстовых данных.

Экспериментальные исследования были организованы следующим образом: на первом этапе рассчитывались показатели полнота-точность анализируемых методов (метод шинглов, метод Winnowing, коэффициент ассоциативности Джаккарда, расстояние Джаро-Винклера, расстояние Левенштейна, ОКА) и определялось их соответствие целевому критерию, на втором этапе изучалось влияние различных характеристик текстовых документов на значения полноты и точности. Отметим, что в наших исследованиях использовались стандартные методы Левенштейна и Джаро-Винклера, не предусматривающие учета местоположения термина (в названии или в аннотации). Структура текста использовалась только при расчете ОКА (именно эта идея была положена в основу разработки данной процедуры). Вместе с тем, в перспективе, планируется провести разработку модификаций методов Левенштейна и Джаро-Винклера, учитывающих структуру текста и исследовать влияние этой информации на целевой критерий.

В таблице представлены значения полноты и точности различных методов на исследуемой выборке для случая предварительного удаления стоп-слов и выполнения стемминга. Для метода шинглов и Winnowing результаты приведены для значений настраиваемых параметров, которым соответствуют наилучшие полнота и точность (размер шингла равен 2 слова; размер k -граммы $k = 2$, значение шумового порога $t = 4$). На основе проведенного исследования можно сделать вывод, что все анализируемые методы удовлетворяют требованию целевого критерия. При этом наилучших результатов достигают два метода: Winnowing и ОКА.

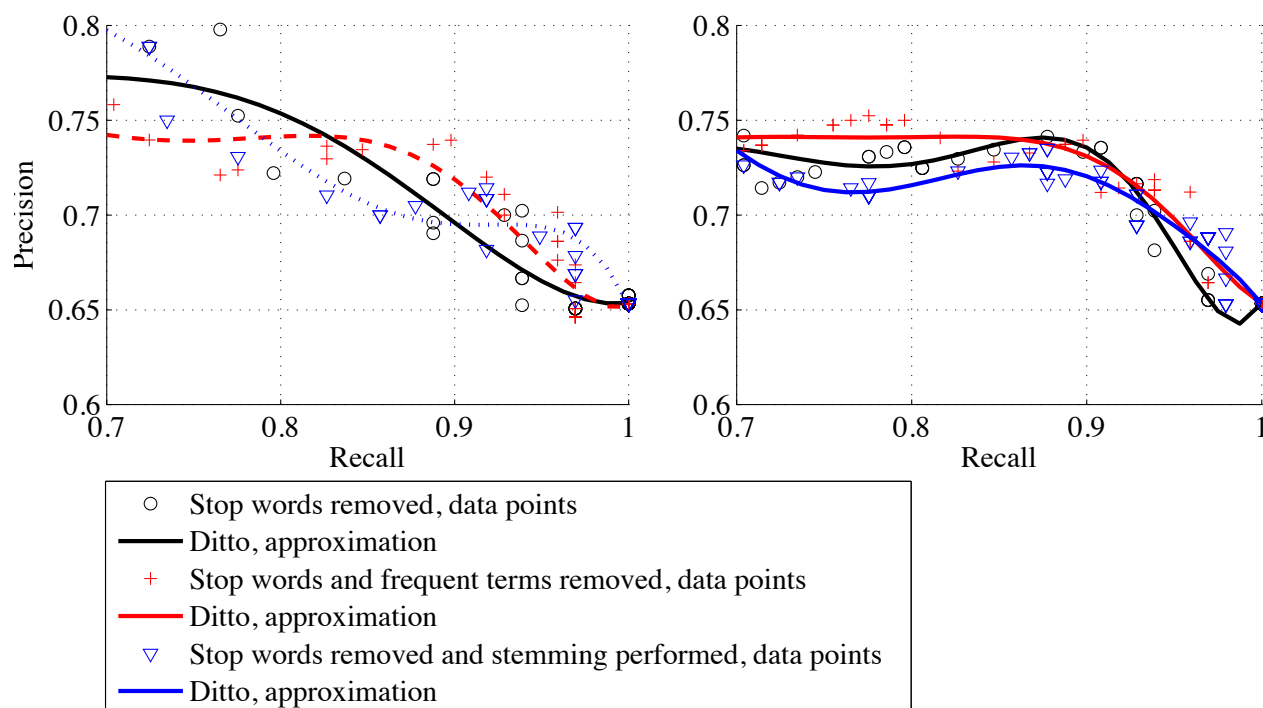


Рис. 1. Графики «полнота–точность» для коэффициента ассоциативности Джаккарда (слева) и для обобщенного коэффициента ассоциативности (справа)

Графики «полнота–точность» методов выявления нечетких дубликатов при различных видах предварительной обработки представлены на рис. 1. Конкретные значения полноты и точности зависят от выборки, поэтому важен качественный характер аппроксимаций, главным образом, радиус их кривизны на интересующем участке. Рационально выбирать такой метод и сочетание параметров, чтобы точность не только была достаточной при заданной полноте, но и при изменении полноты не менялась бы резко (то есть, достигнутая точность должна быть устойчива к возмущениям, вносимым особенностями выборки).

На втором этапе проводилась оценка целесообразности использования для более качественного определения нечетких дубликатов следующих характеристик текста:

- стоп-слов, встречающихся в двух сравниваемых документах с одинаковой частотой (наличие одних и тех же служебных слов: союзов, предлогов и т. п.);
- служебных символов, включая знаки препинания и сокращения.

Проведенные исследования подтвердили известные в литературе результаты, которые свидетельствуют о том, что учет вышеуказанных характеристик не позволяет значительно улучшить показатель «полнота–точность» выявления нечетких дубликатов [4].

Нами также изучалась возможность повышения полноты и точности за счет создания специального терминологического словаря автора. Такой словарь составляется из всех информативных слов, которые содержатся в пяти авторских публикациях, включенных в сформированный документальный массив. Аналогичные работы проводились ранее группой авторов при решении задач выявления идентичных строк в базах данных при сравнении наименований и адресов юридических лиц [8, 9], а также в [4] для анализа web-документов.

Анализ частоты встречаемости терминов из словаря позволяет определить степень «общности–специфичности» различных слов в статьях автора. Наше предположение за-

ключается в том, что исключение высокочастотных слов из расчета коэффициента ассоциативности Джаккарда и ОКА, способно увеличить точность выявления нечетких дубликатов без ухудшения значения полноты. Эксперименты показали, что без учета высокочастотных общих слов точность коэффициента ассоциативности Джаккарда и ОКА возросли и составили соответственно 0,74 и 0,745. У остальных методов, кроме расстояния Левенштейна, было отмечено снижение точности, значения которой вышли из пределов, допустимых по целевому критерию. Отметим, что при расчете ОКА высокочастотные слова исключались только из текста аннотации, т. е., при вычислении коэффициента K_2 в формуле (4). В работах [8, 9] при использовании более сложных двухуровневых функций сравнения на основе «мягкого» tfidf-взвешивания также указывается на улучшение показателей полнота-точность.

Таким образом, использование специализированного документального массива, содержащего не только две статьи, которые рассматриваются в качестве кандидатов в нечеткие дубликаты, но и другие работы автора, позволяет улучшить точность выявления неоригинальных публикаций с помощью методов, основанных на коэффициентах ассоциативности. Это достигается за счет предварительной оценки частоты встречаемости терминов в авторском словаре.

Обсуждение результатов экспериментальных исследований

Главный вопрос — насколько можно доверять полученным результатам, полезны ли они при выявлении не гипотетических нечетких дубликатов (по библиографическим описаниям), а конкретных полнотекстовых публикаций, зеркально повторяющих друг друга? Другими словами, можно ли утверждать, что в выборке, по которой проводились исследования, действительно встречаются случаи «недобросовестных» действий автора (авторов) по клонированию в различных журналах совершенно одинакового материала.

Для ответа на этот вопрос в интернете был проведен поиск полнотекстовых версий 300 статей из выборки (при этом необходимо было найти полнотекстовые варианты сразу двух статей автора). Из всех 150-и пар документов, в открытом доступе нам удалось обнаружить только 31 пару полнотекстовых статей (около 20 % от всего объема выборки). С помощью разработанного программного комплекса [13] было проведено сравнение документов методом шинглов. В результате было обнаружено 3 пары полных дубликатов, в которых авторы лишь незначительно изменили название и аннотацию. При этом основные тексты публикаций оказались абсолютно идентичными (эти выводы подтверждены также экспертно). Следует отметить, что библиографические описания для этих пар публикаций были признаны нечеткими дубликатами всеми вышеперечисленными методами.

Процент самоплагиата, который мы получили, достаточно низкий, что может говорить о хорошем качестве статей в области автоматизации и вычислительной техники. Следует отметить, что мы рассматривали только статьи с близкими библиографическими описаниями.

Авторы, конечно же, понимают, что делать обобщения и заключения на весьма небольшом статистическом материале крайне не разумно и планируют в ближайшее время подтвердить или опровергнуть полученные оценки на новых выборках.

Отметим также, что разработанные программно-алгоритмические средства, на наш взгляд, уже в настоящее время могут быть эффективно использованы редколлегиями журналов (и, возможно, диссертационными советами) при анализе статей авторов и соискателей. Такой анализ разумно проводить на предварительном этапе по библиографическим описаниям. В случае выявления с помощью автоматизированных методов потенциальных

нечетких дубликатов необходимо реализовывать полнотекстовую проверку и экспертное изучение публикаций.

Литература

- [1] Рубцов Д. Н., Барахнин В. Б. О возможности борьбы с дубликатами при запросах к разнородным библиографическим источникам // *Тр. 11-й Всеросс. научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»*. — Петрозаводск: изд-во ПетрГУ, 2007. С. 293–298.
- [2] Зеленков Ю. Г., Сегалович И. В. Сравнительный анализ методов определения нечетких дубликатов для Web-документов // *Тр. 9-й Всеросс. научной конф. «Электронные библиотеки: перспективные методы и технологии, электронные коллекции»*. — Переславль-Залесский: Изд-во ИПС РАН, 2007. С. 166–174.
- [3] Chowdhury A., Frieder O., Grossman D., McCabe C. Collection statistics for fast duplicate document detection // *ACM Trans. Inform. Syst.*, 2002. Vol. 20, No. 2. P. 171–191.
- [4] Heintze N. Scalable document fingerprinting // *2nd USENIX Electronic Commerce Workshop Proceedings*, 1996. P. 191–200.
- [5] Косинов Д. И. Использование статистической информации при выявлении схожих документов // *Сборник «Интернет-математика»*. — Екатеринбург: Изд-во Уральского университета, 2007. С. 84–90.
- [6] Kumar J., Govindarajulu P. Duplicate and near duplicate documents: A review // *European J. Scientific Research*, 2009. Vol. 32, No. 4. P. 514–527.
- [7] Дербенев Н. В., Толчеев В. О. Выявление нечетких дубликатов в наукометрическом анализе // *Информационные технологии*, 2011. № 12. С. 24–29.
- [8] Cohen W. W., Ravikumar P., and Fienberg S. E. A comparison of string distance metrics for name matching tasks, 2003. P. 73–78.
- [9] Bilenko M., Mooney R., Cohen W., Ravikumar P., and Fienberg S. Adaptive name matching in information integration. *IEEE Intelligent Systems*, 2003. Vol. 18. P. 16–23.
- [10] Broder A., Glassman S., Manasse M., Zweig G. Syntactic clustering of the Web. *6th World Wide Web Conference (International)*, 1997. P. 393–404.
- [11] Дербенев Н. В., Толчеев В. О. Разработка метода выявления нечетких дубликатов по библиографическим описаниям // *Тр. Междунар. конф. «Интеллектуализация обработки информации»*. — Черногория, Будва. М.: Изд-во «ТОРУС ПРЕСС», 2012. С. 613–616.
- [12] Дербенев Н. В., Толчеев В. О. Сравнительный анализ коэффициентов ассоциативности для выявления нечетких дубликатов текстовых документов // *Тр. 18-й Междунар. научно-технич. конф. «Информационные средства и технологии»*. — М.: Изд-во МЭИ, 2010. С. 266–270.
- [13] Дербенев Н. В., Козлюк Д. А., Никитин В. В., Толчеев В. О. Разработка программно-алгоритмических средств выявления плагиата в учебных и научных кафедральных работах // *Мат-лы 6-й Всеросс. мультikonf. по проблемам управления — МКПУ-2014*. — Дивноморское, 2014. Т. 1. С. 59–63.

References

- [1] Chowdhury A., Frieder O., Grossman D., McCabe C. 2002. Collection statistics for fast duplicate document detection. *ACM Trans. Inform. Syst.* 20(2):171–191.

- [2] Heintze N. 1996. Scalable document fingerprinting. *2nd USENIX Electronic Commerce Workshop Proceedings* . 191–200.
- [3] Roubtsov D. N., Barakhnin V. B. 2007. On the possibility of duplicates struggle when performing queries to heterogeneous bibliographic sources. *XI All-Russian Scientific Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections” Proceedings*. Petrozavodsk. 293–298.
- [4] Kosinov D. 2007. Use of statistical parameters in similar documents detection. *“Internet-Mathematics” Digest*. Yekaterinburg. 84–90.
- [5] Zelenkov Yuri. G., Segalovich Ilya V. 2007. Comparative analysis of near-duplicate detection methods of Web documents . *IX All-Russian Scientific Conference “Digital Libraries: Advanced Methods and Technologies, Digital Collections” Proceedings*. Pereslavl-Zalessky. 166–174.
- [6] Kumar J., Govindarajulu P. 2009. Duplicate and near duplicate documents: A review. *European J. Scientific Research* 32(4):514–527.
- [7] Derbenev N. V., Tolcheev V. O. 2011. Using a method of detecting near duplicates in scientometric analysis. *Information Technologies* 12:24–29.
- [8] Cohen W. W., Ravikumar P., and Fienberg S. E. 2003. A comparison of string distance metrics for name matching tasks. 73–78.
- [9] Bilenko M., Mooney R., Cohen W., Ravikumar P., and Fienberg S. 2003. Adaptive name matching in information integration. *IEEE Intelligent Systems* 18:16–23.
- [10] Derbenev N. V., Tolcheev V. O. 2012. Development of a method of near-duplicates detection on the base of bibliographic descriptions. *Conference (International) “Intellectualization of Information Processing” Proceedings*. Montenegro, Budva. 613–616.
- [11] Broder A., Glassman S., Manasse M., Zweig G. 1997. Syntactic clustering of the Web. *6th World Wide Web Conference (International)*. 393–404.
- [12] Derbenev N. V., Tolcheev V. O. 2010. Comparative analysis of associativity coefficients for near-duplicate detection within textual documents. *XVIII Conference (International) “Information Tools and Techniques” Proceedings*. Moscow. 266–270.
- [13] Derbenev N. V., Kozliuk D. A., Nikitin V. V., Tolcheev V. O. 2014. Development of software for plagiarism detection in academic and scientific papers. *6th All-Russian Multiconference on Control Problems — MCCP-2014 Proceedings*. Divnomorskoye. 1:59–63.

Оптимальная коррекция метрических нарушений в матрицах парных сравнений*

С. Д. Двоенко, Д. О. Пшеничный

sergedv@yandex.ru; denispshenichny@yandex.ru

Тульский государственный университет, г. Тула

В задачах интеллектуального анализа экспериментальные данные часто представлены результатами парных сравнений объектов между собой. В отсутствие исходного признакового пространства условием корректного погружения множества объектов в метрическое пространство является неотрицательная определенность матрицы парных сравнений элементов множества друг с другом. В этом случае близости интерпретируются как скалярные произведения, а соответствующие различия — как расстояния. В работе предлагается способ коррекции нормированных матриц парных близостей с тем, чтобы скорректированная матрица была положительно определенной и минимально отличающейся от исходной.

Ключевые слова: парные сравнения; собственные значения; метрика; детерминант; матрица

Optimal correction of metrical violations in matrices of pairwise comparisons*

S. D. Dvoenko, D. O. Pshenichny

Tula State University, Tula

Background: In modern data mining, experimental data are usually represented as objects' mutual pairwise comparisons. The condition of correct immersion of set objects in metrical space in the absence of the initial feature space is nonnegative definiteness of matrix of pairwise comparisons between these objects. In this case, similarities are interpreted as scalar products and dissimilarities as distances.

Methods: This paper suggests a method of correction for normalized matrices of similarities; so, the corrected matrix is positively definite and minimally deviated from the initial one.

Results: The proposed method detects objects, which contribute violations in the metrics. Pairwise comparisons of these objects with subset of other objects are corrected. This approach also allows to choose this subset of elements. It is proved that such a correction always exists and can be optimal.

Concluding Remarks: In contrast to traditional approach based on Karhunen–Loeve discrete decomposition, the proposed method is able to correct only few of matrix elements.

Keywords: pairwise comparisons; eigenvalues; metrics; determinant; matrix

Введение

Условием корректного погружения множества в метрическое пространство является неотрицательная определенность матрицы парных близостей его элементов [1]. В этом случае, применяя соответствующие «беспризнаковые» версии алгоритмов машинного обучения и кластер-анализа [2, 3] к матрице сходства или к соответствующей ей матрице различий, мы получим математически корректный результат обработки.

Работа выполнена при финансовой поддержке РФФИ, проект № 13-07-00010.

Машинное обучение и анализ данных, 2014. Т. 1, № 7.

Machine Learning and Data Analysis, 2014. Vol. 1 (7).

В данной работе рассматривается подход к регулируемой коррекции матриц парных сравнений с целью устранения метрических нарушений. Предполагается, что элементы множества, представленные матрицей парных сравнений, требуется погрузить в некоторое метрическое пространство, например, евклидово.

Предлагаемый в работе подход — это развитие идей, описанных в [4, 5]. В работе [4] нами был предложен метод коррекции нормированных матриц парных близостей. Он основан на последовательном погружении элементов множества в специальным образом построенное координатное пространство, где корректное положение очередного добавляемого элемента определяется относительно всех ранее добавленных элементов радиусом соответствующей гиперсферы всех возможных его положений.

В такой процедуре главный минор нормированной матрицы парных близостей уменьшается, начиная с единицы, оставаясь положительным, при добавлении очередного элемента множества. Если на множестве элементов возникают метрические нарушения, то главный минор текущего размера оказывается знакопеременным, постепенно уменьшаясь по модулю. Отрицательность очередного главного минора означает, что очередной добавленный элемент множества внес метрическое нарушение. Его следует исправить, корректируя значения парных сравнений нового элемента с предыдущими элементами до получения неотрицательного главного минора текущего размера.

В [4] показан метод коррекции всех парных сравнений вносящего метрические нарушения элемента. В работе [5] описан метод, позволяющий корректировать лишь одно парное сравнение такого элемента. В данной работе предложен метод, позволяющий корректировать произвольное подмножество парных сравнений нарушающего элемента с остальными минимальным образом.

Практика показывает, что просмотр главных миноров в их исходном порядке приводит к тому, что приходится корректировать большое число строк и столбцов матрицы. Чтобы уменьшить это количество, было предложено найти такую ранжировку объектов, что первый отрицательный главный минор соответствующей ранжировке матрицы близостей находился бы как можно ближе к концу последовательности миноров. Способ получения такой ранжировки описан в [5].

Задача оптимальной коррекции метрических нарушений

Пусть имеется нормированная матрица парных близостей $S(n, n)$, для элементов которой выполняются следующие условия: все диагональные элементы равны единице ($s_{ii} = 1$, $i = 1, \dots, n$), все элементы симметричны относительно главной диагонали ($s_{ij} = s_{ji}$, $i, j = 1, \dots, n$), все элементы матрицы не превышают по модулю единицы ($|s_{ij}| \leq 1$). Требуется разработать алгоритм, позволяющий для этой матрицы найти такую матрицу $S'(n, n)$, что она также будет иметь единичную главную диагональ, симметричные относительно нее и не превосходящие по модулю единицы элементы, при этом она будет положительно определенной, а отклонение ее от S будет в некотором смысле минимальным. Очевидно, что если матрица S является положительно определенной, то получаемая в результате работы алгоритма матрица S' должна в точности совпадать с S . Далее мы будем рассматривать матрицу S и ее миноры, поэтому там, где это не вызывает затруднений, будем обозначать значение минора как и сам минор.

Оптимальная коррекция заданных парных сравнений

Как и ранее, будем просматривать главные миноры $S_k = S(k, k)$, $k = 1, \dots, n$ матрицы $S(n, n)$ по порядку. Если очередной минор S_k оказывается отрицательным, то скорректи-

руем элементы его последней строки и столбца, индексы которых принадлежат множеству $A \subseteq \{1, \dots, k-1\}$, и продолжим просмотр. Рассмотрим случай, когда главный минор $S_k < 0$, а все предыдущие главные миноры $S_i > 0, i = 1, \dots, k-1$.

Вычислим значение этого минора, используя разложение определителя по последней строке и последнему столбцу, и, выполняя преобразования, получим выражение:

$$S_k = S_{k-1} \left(1 - \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} s_{ki} s_{jk} r_{ij} \right), \quad (1)$$

где r_{ij} — элемент (i, j) матрицы $R = S(k-1, k-1)^{-1}$, являющейся обратной матрице $S(k-1, k-1)$, получающейся при взятии первых $k-1$ строк и $k-1$ столбцов матрицы S .

По условию значение минора S_k отрицательно $S_k < 0$. Пусть мы хотим скорректировать минор S_k так, чтобы его значение стало равно некоторому заданному числу $\tilde{C}$, где $0 < \tilde{C} < S_{k-1}$.

Обозначим эти искомые элементы s_{ki}, s_{ik} матрицы S_k как $x_i, i = 1, \dots, k-1$, которые будут соответствовать скорректированным значениям элементов минора S_k . Согласно (1), получим условие, налагаемое на элементы последней строки и столбца минора S_k :

$$\sum_{i=1}^{k-1} \sum_{j=1}^{k-1} x_i x_j r_{ij} = C, \quad (2)$$

где $C = 1 - \tilde{C}/S_{k-1}$.

Так как мы хотим иметь минимальное отклонение скорректированных элементов от исходных, то (2) — это ограничение равенства в следующей задаче оптимизации:

$$\left. \begin{aligned} f(\mathbf{x}) &= \sum_{i=1}^{k-1} (s_{ik} - x_i)^2 \rightarrow \min \\ g(\mathbf{x}) &= \sum_{i=1}^{k-1} \sum_{j=1}^{k-1} x_i x_j r_{ij} = C. \end{aligned} \right\} \quad (3)$$

Решая (3) методом множителей Лагранжа, получаем, что требуется решить следующую систему уравнений:

$$\left. \begin{aligned} \lambda \sum_{i=1, i \in A}^{k-1} x_i r_{ip} + \sum_{i=1, i \notin A}^{k-1} s_{ki} r_{ip} &= s_{kp} - x_p, p \in A \\ \sum_{i=1, i \in A}^{k-1} \sum_{j=1, j \in A}^{k-1} x_i x_j r_{ij} &+ + \sum_{i=1, i \in A}^{k-1} \sum_{j=1, j \notin A}^{k-1} x_i s_{jk} r_{ij} + \\ &+ \sum_{i=1, i \notin A}^{k-1} \sum_{j=1, j \in A}^{k-1} s_{ki} x_j r_{ij} + \sum_{i=1, i \notin A}^{k-1} \sum_{j=1, j \notin A}^{k-1} s_{ki} s_{jk} r_{ij} = C. \end{aligned} \right\} \quad (4)$$

Количество уравнений в этой системе зависит от индексов, образующих множество A , т.е. от того, сколько элементов последней строки и столбца минора S_k требуется скорректировать. От того, какие именно элементы последней строки и столбца минора S_k корректируются, зависит состав первых $|A|$ уравнений системы (4).

Решая эту систему численным методом решения систем алгебраических уравнений, получим оптимально скорректированную строку и столбец минора.

Эксперименты

Проведем эксперименты над двумя нормированными матрицами парных близостей. Пусть первая из них — модельная матрица

$$S(3, 3) = \begin{pmatrix} 1 & 0,5 & 0,5 \\ 0,5 & 1 & -0,9 \\ 0,5 & -0,9 & 1 \end{pmatrix}$$

с главными минорами $S_1 = 1, S_2 = 0.75, S_3 = -0.76$. Поправим предложенным методом последнюю строку и столбец целиком так, чтобы третий минор стал равен $\tilde{C} = 0.1$.

Вычислим сначала матрицу $R = S(2, 2)^{-1} = \begin{pmatrix} 4/3 & -2/3 \\ -2/3 & 4/3 \end{pmatrix}$. Далее вычислим $C = 13/15$. Получим систему уравнений:

$$\begin{cases} \lambda \left(\frac{4}{3}x_1 - \frac{2}{3}x_2 \right) = 0,5 - x_1; \\ \lambda \left(-\frac{2}{3}x_1 + \frac{4}{3}x_2 \right) = -0,9 - x_2; \\ \frac{4}{3}x_1^2 - \frac{4}{3}x_1x_2 + \frac{4}{3}x_2^2 = \frac{13}{15}. \end{cases}$$

Решая ее, получаем $x_1 = 0.285487, x_2 = -0.624637$. Подставив эти значения в исходную матрицу, убедимся, что ее определитель равен 0.1.

Теперь скорректируем предложенным методом только первый элемент последней строки и столбца матрицы $S(3, 3)$. Мы также хотим, чтобы значение определителя скорректированной матрицы было равно 0.1. Получим систему уравнений:

$$\begin{cases} \lambda \left(\frac{4}{3}x_1 - \frac{2}{3} \cdot (-0,9) \right) = 0,5 - x_1 \\ \frac{4}{3}x_1^2 - \frac{4}{3}x_1 \cdot (-0,9) + \frac{4}{3} \cdot (-0,9)^2 = \frac{13}{15}. \end{cases}$$

Решая эту систему, получаем $x_1 = -0,243845$. Подставив эти значения в исходную матрицу, убедимся, что ее определитель также равен 0.1.

Наконец скорректируем описанным методом только второй элемент последней строки и столбца матрицы $S(3, 3)$ так, чтобы ее определитель снова стал равен 0,1.

Теперь требуется решить такую систему уравнений:

$$\begin{cases} \lambda \left(-\frac{2}{3} \cdot 0,5 + \frac{4}{3}x_2 \right) = -0,9 - x_2; \\ \frac{4}{3} \cdot 0,5^2 - \frac{4}{3} \cdot 0,5x_2 + \frac{4}{3}x_2^2 = \frac{13}{15}. \end{cases}$$

Также решая эту систему, получаем $x_2 = -0.430074$. Подставив эти значения в исходную матрицу, убедимся, что ее определитель снова равен 0.1.

Второй эксперимент поставлен на матрице данных $S(11, 11)$, которая является корреляционной матрицей статистических взаимосвязей между энергетическими свойствами биоритмов головного мозга для 11 частот (тета-, альфа-, бета-ритмы энцефалограммы головного мозга), полученных В. Д. Небылицыным в ходе исследований по эффекту навязывания ритма [7]:

$$\begin{pmatrix} 1 & 0,562 & 0,568 & 0,152 & 0,347 & 0,250 & 0,264 & -0,020 & -0,212 & -0,086 & -0,076 \\ 0,562 & 1 & 0,784 & 0,057 & 0,196 & 0,218 & 0,009 & -0,017 & -0,002 & 0,163 & 0,284 \\ 0,568 & 0,784 & 1 & 0,288 & 0,475 & 0,264 & 0,066 & 0,144 & 0,114 & 0,228 & 0,151 \\ 0,152 & 0,057 & 0,288 & 1 & 0,686 & 0,293 & 0,034 & 0,048 & -0,069 & -0,064 & 0,175 \\ 0,347 & 0,196 & 0,475 & 0,686 & 1 & 0,429 & 0,070 & 0,152 & 0,036 & 0,028 & 0,216 \\ 0,250 & 0,218 & 0,264 & 0,293 & 0,429 & 1 & 0,788 & 0,197 & 0,154 & 0,109 & 0,035 \\ 0,264 & 0,009 & 0,066 & 0,034 & 0,070 & 0,788 & 1 & 0,109 & 0,054 & -0,002 & -0,018 \\ -0,020 & -0,017 & 0,144 & 0,048 & 0,152 & 0,197 & 0,109 & 1 & 0,807 & 0,830 & 0,699 \\ -0,212 & -0,002 & 0,114 & -0,069 & 0,036 & 0,154 & 0,054 & 0,807 & 1 & 0,904 & 0,728 \\ -0,086 & 0,163 & 0,228 & -0,064 & 0,028 & 0,109 & -0,002 & 0,830 & 0,904 & 1 & 0,768 \\ -0,076 & 0,284 & 0,151 & 0,175 & 0,216 & 0,035 & -0,018 & 0,699 & 0,728 & 0,768 & 1 \end{pmatrix}.$$

Данная матрица содержит 10 положительных и одно отрицательное собственное число.

Сначала был выполнен поиск оптимальной ранжировки, в результате которого получилась следующая перестановка элементов множества: 7, 4, 8, 1, 3, 11, 5, 9, 6, 10, 2. Главные миноры матрицы, соответствующей оптимальной ранжировке, имеют следующие значения: 1,000000, 0,998844, 0,985015, 0,893156, 0,539420, 0,255770, 0,108704, 0,024602, 0,004752, 0,000481, $-0,000057$.

Таким образом, оказалось, что последний элемент в данной ранжировке вносит метрическое нарушение. В исходной перестановке элементов множества это второй элемент. Скорректируем все элементы последней строки и столбца матрицы, соответствующей полученной перестановке так, чтобы полученное значение минора составляло примерно 0,1 от значения предыдущего минора S_{10} . Пусть $S'_{11} = 0,00005$. В результате коррекции и обратной перестановки строк и столбцов матрицы была получена следующая скорректированная матрица (жирным в ней выделены измененные элементы):

$$\begin{pmatrix} 1 & \mathbf{0,552} & 0,568 & 0,152 & 0,347 & 0,250 & 0,264 & -0,020 & -0,212 & -0,086 & -0,076 \\ \mathbf{0,552} & 1 & \mathbf{0,759} & \mathbf{0,065} & \mathbf{0,213} & \mathbf{0,191} & \mathbf{0,030} & \mathbf{-0,006} & \mathbf{0,009} & \mathbf{0,169} & \mathbf{0,254} \\ 0,568 & \mathbf{0,759} & 1 & 0,288 & 0,475 & 0,264 & 0,066 & 0,144 & 0,114 & 0,228 & 0,151 \\ 0,152 & \mathbf{0,065} & 0,288 & 1 & 0,686 & 0,293 & 0,034 & 0,048 & -0,069 & -0,064 & 0,175 \\ 0,347 & \mathbf{0,213} & 0,475 & 0,686 & 1 & 0,429 & 0,070 & 0,152 & 0,036 & 0,028 & 0,216 \\ 0,250 & \mathbf{0,191} & 0,264 & 0,293 & 0,429 & 1 & 0,788 & 0,197 & 0,154 & 0,109 & 0,035 \\ 0,264 & \mathbf{0,030} & 0,066 & 0,034 & 0,070 & 0,788 & 1 & 0,109 & 0,054 & -0,002 & -0,018 \\ -0,020 & \mathbf{-0,006} & 0,144 & 0,048 & 0,152 & 0,197 & 0,109 & 1 & 0,807 & 0,830 & 0,699 \\ -0,212 & \mathbf{0,009} & 0,114 & -0,069 & 0,036 & 0,154 & 0,054 & 0,807 & 1 & 0,904 & 0,728 \\ -0,086 & \mathbf{0,169} & 0,228 & -0,064 & 0,028 & 0,109 & -0,002 & 0,830 & 0,904 & 1 & 0,768 \\ -0,076 & \mathbf{0,254} & 0,151 & 0,175 & 0,216 & 0,035 & -0,018 & 0,699 & 0,728 & 0,768 & 1 \end{pmatrix},$$

где евклидово расстояние между полученной матрицей и исходной составило 0,058.

Теперь скорректируем отрицательный минор методом коррекции всего вектора парных сравнений, который ранее был нами предложен [4]. Значение евклидова расстояния между исходной матрицей и скорректированной данным способом составило 0,058 (при этом значение определителя скорректированной матрицы составило 0,0000005). Наконец, если в отрицательном миноре скорректировать только одиночный элемент $s_{11,1}$, то отклонение составит 0,088.

Таким образом, предложенный метод на данной матрице действительно обеспечивает оптимальный результат коррекции.

Заключение

В данной работе предложен алгоритм коррекции неположительно определенных матриц парных близостей. Этот алгоритм позволяет определить, какие объекты вносят наибольшее искажение в метрику пространства и скорректировать их парные сравнения.

Предложенный метод корректировки парных сравнений объекта позволяет найти наиболее близкие скорректированные парные сравнения к значениям исходных парных сравнений, при этом вклад в нарушение этим объектом метрического пространства будет уменьшен, причем величина этого вклада может регулироваться.

Предложенный метод позволяет корректировать не все парные сравнения объекта, а лишь некоторые, оставляя неизменными значения остальных парных сравнений. Данный метод является обобщением ранее описанных метода коррекции всего вектора парных сравнений и метода коррекции одиночных элементов.

Литература

- [1] Young G., Householder A. S. Discussion of a set of points in terms of their mutual distances // *Psychometrika*, 1938. Vol. 3. No. 1. P. 19–22.
- [2] Двоенко С. Д. Распознавание элементов множества, представленных взаимными расстояниями и близостями // ММРО-14. — М.: МАКС Пресс, 2009. С. 112–115.
- [3] Двоенко С. Д. Кластеризация множества, описанного парными расстояниями и близостями между его элементами // Сибирский журнал индустриальной математики, 2009. Т. 12. № 1(37). С. 61–73.
- [4] Двоенко С. Д., Пшеничный Д. О. Об устранении отрицательных собственных значений матриц парных сравнений // ИОИ-2012. — М.: Торус Пресс, 2012. С. 13–16.
- [5] Двоенко С. Д., Пшеничный Д. О. О метрической коррекции матриц парных сравнений // Машинное обучение и анализ данных, 2013. Т. 1. № 5. С. 606–620.
- [6] Гантмахер Ф. Р. Теория матриц. — М.: Наука, 1967. 576 с.
- [7] Небылицын В. Д. Главные характеристики нервной системы. — М.: Просвещение, 1966.

References

- [1] Young G., Householder A. S. 1938. Discussion of a set of points in terms of their mutual distances. *Psychometrika* 3(1):19–22.
- [2] Dvoenko S. D. 2009. Recognition of the set elements presented by mutual distances and similarities. 14th Conference “Mathematical Methods of Pattern Recognition” Proceedings. Moscow: MAKSS Press. 112–115.
- [3] Dvoenko S. D. 2009. Clusterization of the set presented by distances and similarities between its elements. *Siberian J. Industrial Mathematics* 12(1(37)):61–73. [In Russian.]
- [4] Dvoenko S. D., Pshenichny D. O. 2012. On negative eigenvalues removing from matrixes of pairwise comparisons. 9th Conference (International) “Intellectualization of Information Processing” Proceedings. Moscow: TORUS PRESS. 13–16. [In Russian.]
- [5] Dvoenko S. D., Pshenichny D. O. 2013. On metrical correction of the matrices of pairwise comparisons. *JMLDA* 1(5):606–620.
- [6] Gantmacher F. R. 2000. The theory of matrices. AMS Chelsea Publishing. 660 p.
- [7] Nebylitsyn V. D. 1966. Main characteristics of the nervous system. Moscow: Prosveshchenie. [In Russian.]

Обучение распознаванию без переобучения*

Н. Г. Загоруйко^{1,2}, О. А. Кутненко^{1,2}, А. О. Зырянов², Д. А. Леванов¹

zag@math.nsc

¹Институт математики им. С. Л. Соболева СО РАН, Новосибирск; ²Новосибирский государственный университет, Новосибирск

Описывается процесс обучения, состоящий в выборе подмножества наиболее информативных объектов («столпов») и признаков и в автоматическом обнаружении момента начала переобучения. Сходство между объектами оценивается функцией конкурентного сходства (FRiS-функцией). На каждом шаге увеличения числа столпов в фиксированном признаковом пространстве вычисляется качество описания выборки (или оценка разделимости классов). Сформулирована и подтверждена гипотеза о том, что точка перегиба кривой, описывающей разделимость классов, может служить сигналом о начале переобучения. На этом основании разработан алгоритм FRiS-C-GRAD обучения без переобучения. Описываются результаты тестирования алгоритма на модельных задачах.

Ключевые слова: анализ данных; функция конкурентного сходства; разделимость образов; переобучение; цензурирование объектов и признаков

Learning to recognition without overfitting*

N. G. Zagoruiko^{1,2}, O. A. Kutnenko^{1,2}, A. O. Zyryanov², D. A. Levanov¹

¹Sobolev Institute of Mathematics, SB RAS, Novosibirsk; ²Novosibirsk State University, Novosibirsk

Background: The problem of dealing with overfitting of recognition algorithms is one of the central problems in data mining. The algorithm that selects the most informative objects and features and generates a signal at the point in which overfitting starts is described. At this signal, the learning process stops.

Methods: To calculate the similarity between objects, the ternary relative measure called the function of rival similarity (FRiS) is used. Selecting informative features is made by the algorithm FRiS-Grad. Learning includes the consistent increase of number of reference objects (“stolps”) and the formation of clusters of objects closed to stolps in a fixed feature space. On each step of augmentation of number of stolps, the estimation of quality of the description of sample (or divisibility of classes) is calculated.

Results: The hypothesis that the inflection point of the curve describing the separability of classes may signal about the beginning of overfitting is formulated and confirmed. On this basis, the algorithm FRiS-C-GRAD of learning without overfitting is developed. A method of decision-making that takes into account the weight of clusters is proposed. The results of testing of the algorithm on simulated problems are described.

Concluding Remarks: Using FRiS-function has been useful in constructing decision rules, automatic classification (taxonomy), and selecting informative features and in obtaining quantitative estimation of compactness. In this paper, the utility of using this measure of similarity for addressing the protection of overfitting is illustrated.

Keywords: data mining; function of rival similarity; pattern separation; overfitting; objects and features censoring

Введение

Проблема борьбы с переобучением алгоритмов распознавания (overfitting, переподгонкой) является одной из центральных проблем теории распознавания образов. Переобучение наступает тогда, когда алгоритм обучения уточняет решающие правила, учитывая все более мелкие особенности распределения классов, вплоть до отдельных объектов («выбросов») с резко выделяющимися характеристиками. Аналитические методы цензурирования таких объектов, использующие знания о законах распределения обучающей выборки [1], в большинстве современных задач распознавания неприменимы, так как эти законы, как правило, неизвестны, количество признаков N часто на порядки превышает количество объектов M и т. д.

В регрессионном анализе борьбе с выбросами посвящено большое количество методов регуляризации [2], основанных на выборе оптимальных стратегий сглаживания кривых, аппроксимирующих дискретные значения результатов измерений. Качество сглаживания оценивается «прямым» методом Cross Validation с использованием тестовой выборки, не участвовавшей в обучении. Однако отмечено [3], что модели, выбранные этим методом, часто показывают плохие результаты при их практическом применении.

Из работ по переобучению и регуляризации следует, что вероятность переобучения тем выше, чем больше размерность признакового пространства. Так что в число мер по предотвращению переобучения следует включить наряду с выбором типичных представителей классов и сокращение количества признаков. В работе [4] формулировалась задача выбора «нуклеуса», т. е. такого подмножества объектов и признаков, которые отражают закономерности обучающей выборки лучше, чем все M исходных объектов в пространстве всех N исходных признаков.

В данной работе делается попытка построить алгоритм, который автоматически выбирает нуклеус, т. е. подмножество наиболее информативных объектов и признаков, и обнаруживает момент начала переобучения.

Начнем с рассмотрения процесса описания выборки в фиксированном признаковом пространстве. Используются решающие правила прецедентного типа. В качестве прецедентов алгоритмом FRiS-Stolp [5] выбирается множество объектов (столпов), описывающее обучающую выборку. Данный алгоритм основан на использовании меры сходства между объектами в виде функции конкурентного сходства (FRiS-функции) [6]. При изменении количества столпов l меняется качество описания H обучающей выборки и надежность распознавания P тестовой выборки. Выдвигается и проверяется гипотеза о том, что между функциями $H = f(l)$ и $P = f(l)$ имеется закономерная связь, используя которую можно найти такое количество столпов l^* , что дальнейшее увеличение числа столпов ведет к переобучению. Выбор информативных признаковых пространств делается алгоритмом FRiS-Grad [7].

Вначале опишем упомянутые выше понятия и алгоритмы, которые используются при решении данной задачи.

Функция конкурентного сходства (FRiS-функция)

Распространенные меры сходства объекта z с объектом a , использующие только расстояние $r(z, a)$ между ними, не адекватны природе задач классификационного типа. Похожа ли сосна на ель? По сравнению с кустом сирени, пожалуй, да. А по сравнению с кедром — скорее нет. Для ответа на вопрос «Похож или не похож?» или «Близко или

далеко?» нужно знать ответ на вопрос «По сравнению с чем?» В связи с этим при решении вопроса о том, чему равно сходство объекта z с объектом a по сравнению с объектом b , предлагается использовать тернарную относительную меру, которую назовем функцией конкурентного сходства или FRiS-функцией [6]:

$$F(z, a|b) = \frac{r(z, b) - r(z, a)}{r(z, b) + r(z, a)}. \quad (1)$$

Эта функция (FRiS, от Function of Rival Similarity) меняется от -1 (при совпадении z с b) до 1 (при совпадении z с a). Если $r(z, a) = r(z, b)$, то $F = 0$. Сходство z с a в конкуренции с b и сходство z с b в конкуренции с a (т. е. отличие z от a) связаны соотношением $F(z, a|b) = -F(z, b|a)$.

При распознавании принадлежности объекта z к одному из двух образов A или B в эвристических алгоритмах распознавания, основанных на сходстве объектов, важно знать не только расстояние $r(z, A)$ от z до образа A , но и расстояние $r(z, B)$ от z до конкурирующего образа B . То есть сходство в распознавании образов является категорией не абсолютной, а относительной. Отметим, что в зависимости от особенностей решаемой задачи расстояние $r(z, A)$ от объекта z до образа A может вычисляться по-разному. В качестве него может использоваться и расстояние $r(z, a)$ до ближайшего объекта a образа A , и среднее расстояние до всех объектов образа, и среднее расстояние до k ближайших объектов образа, и расстояние до центра тяжести образа и т. д. Сходство в шкале порядка, используемое в методе kNN , отвечает на вопрос: «На объекты какого образа объект z похож больше всего?». Конкурентное сходство, измеряемое с помощью FRiS-функции, отвечает на этот вопрос и, кроме того, на такой вопрос: «Какова абсолютная величина сходства z с объектами образа A в конкуренции с объектами образа B ?» Оказалось, что дополнительная информация, которую дает абсолютная шкала по сравнению со шкалой порядка, позволяет существенно улучшить методы анализа данных.

Конкурентное сходство объектов с образами будем определять по тому же принципу, что и конкурентное сходство объектов с объектами:

$$F(z, A|B) = \frac{r(z, B) - r(z, A)}{r(z, B) + r(z, A)}. \quad (2)$$

В случае нормально распределенных образов с одинаковыми матрицами ковариации конкурентное сходство объекта с этими образами можно вычислять через сходство с их средними. Если же образы имеют очень сложную структуру, то ориентироваться при вычислении FRiS-функции можно лишь на локальную окрестность (ближайших соседей) того объекта, для которого она вычисляется. Одним из способов проявить особенности данных в задаче распознавания является переход к их сжатому описанию с помощью множества эталонных представителей каждого образа, сохраняющему основные закономерности, необходимые для хорошего распознавания как объектов исходной выборки, так и новых объектов. Такие эталонные объекты далее будем называть столпами. И вычислять конкурентное сходство объекта z с образом A в конкуренции с образом B можно как $F(z, S_A|S_B)$, где S_A (S_B) — множество столпов образа A (B). При распознавании принадлежности объекта z к одному из двух образов A или B в (2) будем использовать расстояния от объекта z до ближайших столпов данных образов.

Выбор эталонных объектов (алгоритм FRiS-Stolp)

Для распознавания образов в заданном фиксированном пространстве необходимо выбрать объекты-эталон (столпы), с которыми будут сравниваться контрольные объекты. Рассматривается задача выбора столпов для описания обучающей выборки, состоящей из M объектов, разделенных на K образов (классов). Если $K > 2$, то при построении столпов для образа A_k , $k \in \{1, \dots, K\}$, объекты всех остальных образов объединяются в один виртуальный образ $B_k = \bigcup_{i=1, \dots, K, i \neq k} A_i$. Так что будем рассматривать алгоритм FRIS-Stolp для выбора столпов двух классов — A и B . Этот алгоритм работает при произвольном виде распределения образов и при любом соотношении количества объектов к количеству признаков. Набор столпов считается достаточным для описания выборки, если сходство F всех объектов обучающей выборки с ближайшими своими столпами в конкуренции с ближайшими объектами других образов не меньше некоторого порогового значения F^* , например, $F^* = 0$.

В качестве столпов выбираются объекты, обладающие высоким значением обороноспособности по отношению к объектам своего образа, т. е. минимизирующие ошибки I типа («пропуск цели»). С помощью рис. 1 проиллюстрируем методику вычисления оценки обороноспособности (или веса) объекта на примере задачи распознавания двух образов $A = \{a_1, \dots, a_{M_A}\}$ и $B = \{b_1, \dots, b_{M_B}\}$, представленных наборами из M_A и M_B объектов обучающей выборки, соответственно.

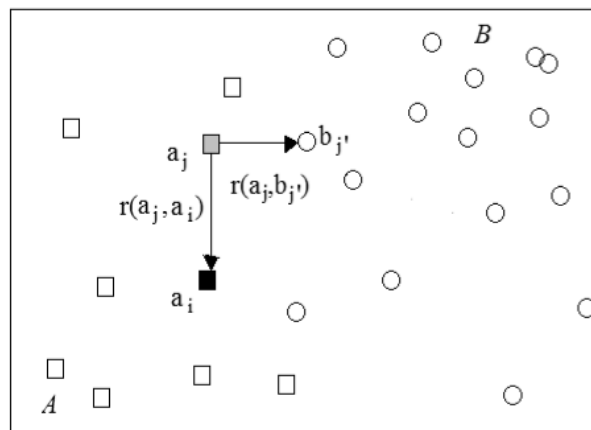


Рис. 1. Оценка веса объекта $a_i \in A$.

Отметим, что при определении по (1) $F(a_j, a_i | b_{j'})$, $j = i$, возникает неопределенность типа $0/0$ в случае совпадения объектов $a_j \in A$ и $b_{j'} \in B$. В реальных задачах совпадающие объекты разных образов, как правило, исключаются из обучающей выборки.

Начнем с выбора первого столпа для образа A в конкуренции с образом B . Оценим качество исполнения роли столпа всеми объектами a_i , $i = 1, \dots, M_A$, по очереди. Для этого найдем расстояния $r(a_j, a_i)$ от всех объектов a_j , $j = 1, \dots, M_A$, класса A до объекта a_i . В качестве конкурента каждому объекту a_j класса A выберем ближайший объект $b_{j'} \in B$, т. е. $j' = \arg \min_{m=1, \dots, M_B} r(a_j, b_m)$, и по (1) получим значение $F(a_j, a_i | b_{j'})$ функции сходства объекта a_j с $a_i \in A$ в конкуренции с $b_{j'} \in B$ (см. рис. 1). Получим вес объекта a_i :

$$V(a_i) = \sum_{j=1}^{M_A} F(a_j, a_i | b_{j'}). \quad (3)$$

Выделим объекты $a_j \in A$, $j = 1, \dots, M_A$, сходство которых с a_i не меньше заданного порога F^* , т. е. $F_j^+ = F(a_j, a_i | b_{j'}) - F^* \geq 0$. Объект a_i вместе с объектами, которые он надежно защищает, образует кластер. Через

$$C(a_i) = \sum_{j=1}^{M_A} F_j^+ |_{F_j^+ \geq 0} \quad (4)$$

обозначим вес данного кластера. Отметим, что вес единичного кластера, т. е. кластера, содержащего объект, защищающий только самого себя, равен 1.

Величина обороноспособности (или вес) $V(a_i)$ является оценкой качества исполнения объектом a_i роли столпа класса A . После получения таких оценок всеми M_A объектами на роль первого столпа образа A выбирается объект, имеющий максимальный вес. Аналогично выбирается первый столп образа B .

Если в кластеры вошли не все M объектов, то среди оставшихся («незащищенных») выбирается объект на роль третьего столпа. Третьим столпом назначается объект любого класса, сходство с которым незащищенных объектов этого класса в конкуренции с ближайшими объектами другого класса максимально. Для выбора $l + 1$ -го столпа ($l > 2$) пересчитываются только веса незащищенных объектов того класса, объект которого стал l -ым столпом. Поскольку кластеры возникают поочередно и формирование состава каждого следующего кластера осуществляется в условиях «отсутствия» многих исходных объектов, то после выбора очередного столпа уточняется состав кластеров того образа, объект которого был выбран на роль столпа: защищенные объекты включаются в кластеры, образованные ближайшими к ним столпами своего образа. Теперь каждый из столпов стоит в центре своего кластера, т. е. подмножества объектов, которые на него похожи больше, чем на любой другой столп. Процесс уточнения описания выборки путем увеличения количества столпов продолжается до включения в кластеры всех объектов.

Вне зависимости от вида распределения обучающей выборки столпами выбираются объекты, расположенные в центрах локальных сгустков и защищающие максимально возможное количество объектов с заданной надежностью. При нормальных распределениях столпами в первую очередь будут выбраны объекты, ближайšie к точкам математического ожидания. Следовательно, при приближении закона распределения к нормальному решение задачи построения решающих функций стремится к статистически оптимальному. Если распределения полимодальны и образы линейно неразделимы, столпы будут стоять в центрах мод.

Качество описания выборки (оценка разделимости классов)

В основе всех методов распознавания образов лежит понятие компактности [8]. В [9] разработан комплекс критериев для оценивания качества обучающих выборок, включающий в том числе и такие характеристики выборки, как компактность и отделимость образов.

В [10] предложен критерий получения количественной оценки компактности с помощью FRiS-функции. Компактность играет большую роль в алгоритмах автоматической классификации (таксономии) [11]. В алгоритмах выбора признаков компактность оказалась критерием информативности более мощным, чем распространенный метод минимума ошибок при Cross Validation. Важна она и для решения данной задачи. О том, насколько хорошо некоторая система столпов описывает классы в заданном признаковом пространстве, будем судить по оценке качества описания обучающей выборки, опирающейся на понятие компактности.

Качество описания обучающей выборки (или оценка разделимости классов) зависит от набора выбранных эталонов. В случае, когда каждый объект выборки является столпом, распределение восстанавливается безошибочно, но для распознавания такое описание не пригодно.

Пусть $s_l \in \bigcup_{k=1}^K A_k$ — l -ый столп в описании выборки, полученном приведенным выше алгоритмом. Через

$$H(L) = \frac{K}{L \times M} \sum_{l=1}^L V(s_l) \quad (5)$$

обозначим качество описания обучающей выборки L столпами. Из (1), (3), (5) следует, что $H \in (-1, 1]$. Например, максимальная разделимость классов, равная 1, достигается, если все объекты каждого образа отображаются в свою отдельную точку.

На каждом шаге добавления в описание нового эталона увеличивается доля защищенных этими эталонами объектов. Процесс описания выборки столпами останавливается, когда все объекты выборки будут защищены своими столпами.

Распознавание контрольного объекта

Сжатое описание образов через множество столпов можно использовать для распознавания новых объектов. В данной работе предлагается метод принятия решений с учетом взвешенных расстояний. Поскольку разные столпы защищают разное количество объектов, то участие в принятии решения столпов с одинаковым весом плохо согласуется с принятым в статистике правилом учета априорной вероятности классов.

Каждый столп защищает свою часть выборки — кластер. Веса кластеров используются в процессе распознавания. Вес кластера $C(s_l)$ определяется формулой (4) как сумма сходств объектов, входящих в данный кластер, со своим столпом s_l в конкуренции с ближайшим объектом другого класса.

Распознавание контрольного объекта z с использованием L столпов состоит в следующем: определяются взвешенные расстояния $r_l = r(z, s_l)/C(s_l)$ от z до всех столпов s_l , $l = 1, \dots, L$, описывающих обучающую выборку. Выбираются два минимальных значения r_{l_1} и r_{l_2} таких, что столпы s_{l_1} и s_{l_2} принадлежат разным классам. Объект z считается принадлежащим классу, взвешенное расстояние до столпа которого оказалось меньше. По величине сходства $F(z, s_{l_1}|s_{l_2})$ можно судить о достоверности принятого решения: чем больше конкурентное сходство, тем выше вероятность того, что решение правильно.

Обнаружение начала переобучения. Экспериментальная проверка гипотезы

Вполне возможно, что в конце процесса обучения появляются столпы, вокруг которых формируются малочисленные кластеры, вплоть до того, что некоторые сильно выделяющиеся объекты становятся столпами, которые защищают только самих себя. Наличие мелких и единичных кластеров ведет к ухудшению качества распознавания контрольной последовательности. Такое явление называется переобучением.

Обратим внимание на зависимость качества описания H от количества столпов. Функция $H = f(l)$ с увеличением l сначала быстро растет, затем рост замедляется и при некотором значении l^* начинает уменьшаться. Не исключено, что наблюдая поведение функции $H(l)$, можно прогнозировать надежность распознавания контрольной выборки, что позволит выбирать оптимальное значение l^* . Для проверки этой гипотезы нужно при

каждом новом значении l вычислять качество описания H и надежность P распознавания контрольной выборки.

Обнаружено, что обычно в точке l^* , в которой достигается первый локальный максимум функции $H(l)$, заканчивается обучение и начинается переобучение. Объекты, которые к этому моменту оказались не включенными в кластеры, из дальнейшего рассмотрения исключаются (т. е. цензурируются).

Тестирование проводилось на модельной задаче распознавания двух образов, каждый из которых представлял собой суперпозицию нескольких (от 2-х до 4-х) нормально распределенных кластеров в двумерном пространстве признаков. Рассматривалось 10 распределений, которые отличались друг от друга количеством образующих нормальных компонентов, их дисперсиями, координатами математических ожиданий и количеством объектов в компонентах. Каждый образ был представлен 250 объектами. При каждом распределении выборка 100 раз случайным способом делилась на две части: обучающую (по 50 объектов первого и второго образов), и контрольную (по 200 объектов каждого образа). Таким образом, общее количество экспериментов при различных численных реализациях исходных данных было равно 1000.

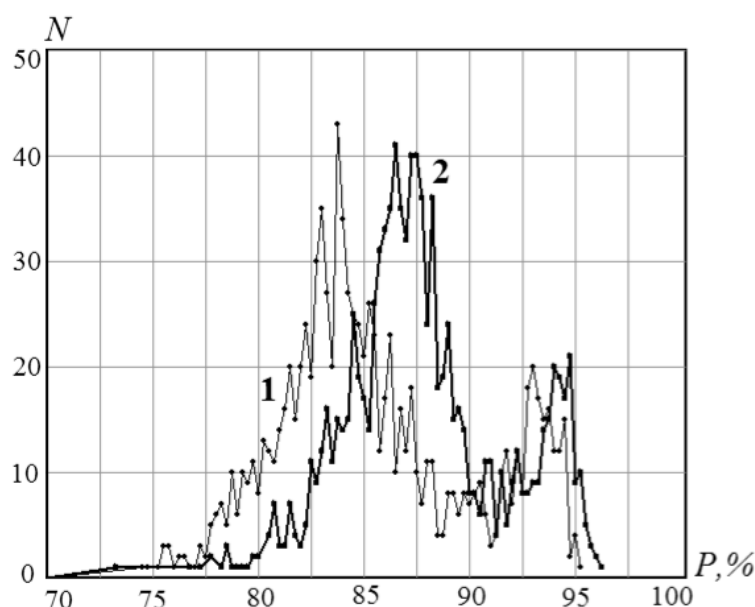


Рис. 2. Распределение надежности $P(\%)$ распознавания контрольной выборки без цензурирования (1) и с цензурированием (2)

Эксперименты показали, что описание обучающей выборки с цензурированием объектов приводит к повышению качества распознавания. Более простые решающие правила повышают надежность распознавания контрольной выборки. На рис. 2 показаны гистограммы распределения надежности $P(\%)$ распознавания контрольной выборки. По оси ординат отложено абсолютное число экспериментов N (из 1000), в которых была достигнута данная надежность $P(\%)$. Кривая 1 соответствует надежности без цензурирования, среднее значение надежности равно 85,60%. Кривая 2 соответствует надежности распознавания с использованием цензурирования. Здесь среднее значение равно 87,86%.

Среднее значение количества столпов l^* , при котором достигался первый локальный максимум критерия качества описания выборки H , было равно 10,81. А среднее значение

числа столпов, достаточных для защиты всех объектов выборки, равнялось 28, 91. Таким образом, цензурирование повышает надежность распознавания и уменьшает трудоемкость обучения.

Сокращение пространства признаков (алгоритм FRiS-GRAD)

В настоящее время часто встречаются задачи, в которых количество признаков N превышает количество объектов M . При этом информация, полезная для решения конкретной классификационной задачи, обычно представлена в нескольких признаках $n \ll N$. Выбор этих n признаков позволяет в дальнейшем не только существенно сократить затраты машинных ресурсов, но и повышает компактность образов и надежность их распознавания. Признаки могут зависеть друг от друга, что не позволяет по оценкам индивидуальной информативности каждого признака выбрать подмножество в виде списка из n наиболее информативных признаков. Если n задано, точное решение можно получить, проверив все сочетания из N признаков по n , что в реальных задачах практически невозможно. По этой причине используются эвристические алгоритмы направленного перебора.

В разработанном нами алгоритме GRAD [12] используется прием, предложенный в [13]. Сначала методом полного перебора формируются информативные системы признаков (гранулы) малой размерности, а затем эти гранулы используются в качестве входных элементов — обобщенных признаков для алгоритма AdDel [14], который представляет собой комбинацию двух известных жадных алгоритмов Addition [15] и Deletion [16]. Эти алгоритмы дают оптимальное решение на каждом шаге, но не обеспечивают глобального оптимума.

В алгоритме AdDel в процессе прямого хода (алгоритм Addition) набирается некоторое количество информативных признаков (гранул признаков) и затем часть из них исключается в процессе обратного хода (алгоритм Deletion). Такое чередование алгоритмов Addition и Deletion продолжается до достижения заданного количества признаков n . В алгоритме FRiS-GRAD [7], использующего вместо отдельных признаков гранулы, в полученной системе некоторые признаки могут встретиться более одного раза.

Как показали эксперименты, по мере увеличения числа признаков качество распознавания вначале растет, потом рост прекращается и начинается его снижение за счет добавления неинформативных, шумящих признаков. Перегиб кривой качества позволяет автоматически определить количество признаков в системе.

Информативность признака или системы признаков может оцениваться разными способами. В описываемом алгоритме выбора признаков FRiS-C-GRAD информативность каждой гранулы и системы признаков проверяется по критерию качества H описания выборки при разных количествах столпов l . В итоге автоматически формируется нуклеус обучающей выборки, содержащий минимальное и достаточное количество признаков n и оптимальное количество столпов l^* , которые обеспечивают построение решающих правил, избегающих переобучения.

Это сокращенное описание выборки в виде множества столпов в фиксированном пространстве информативных признаков далее может использоваться для решения задачи распознавания. Объект z относится к образу, сходство со столпом которого в пространстве выбранных информативных признаков оказалось наибольшим, а величина сходства рассматривается как вероятность правильности принятого решения.

Заключение

Использование FRiS-функции было полезным при построении решающих правил, автоматической классификации (таксономии) и выборе информативных признаков [17], при получении количественной оценки компактности [18]. В данной работе показана полезность применения этой меры сходства и для решения задачи защиты от переобучения. Описан алгоритм, который выбирает подмножество наиболее информативных объектов и признаков и останавливает процесс обучения в точке, в которой начинается переобучение.

Дальнейшие исследования предполагается ориентировать как на разработку других принципов выбора эталонных объектов, так и новых критериев качества описания обучающей выборки, основанных на функции конкурентного сходства.

Литература

- [1] Лемешко Б. Ю., Лемешко С. Б., Постовалов С. Н., Чимитова Е. В. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход. Новосибирск: НГТУ, 2011. 888 с.
- [2] Тихонов А. Н., Гончарский А. В., Степанов В. В., Ягола А. Г. Регуляризирующие алгоритмы и априорная информация. — М.: Наука, 1983.
- [3] Колмогоров А. Н. К вопросу о пригодности найденных статистическим путем формул прогноза // *Заводская лаборатория*, 1933. № 1. С. 164–167.
- [4] Иващенко А. Г., Степашко В. С. Помехоустойчивость моделирования. — Киев: Наукова думка, 1985.
- [5] Загоруйко Н. Г., Кутненко О. А. Количественная мера компактности образов и метод ее повышения // *9-я Междунар. конф. «Интеллектуализация обработки информации»*. — М.: ТОРУС ПРЕСС, 2012. С. 29–32.
- [6] Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. Methods of recognition based on the function of rival similarity // *Pattern Recognition Image Analysis*, 2008. Vol. 18. № 1. P. 1–6.
- [7] Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. Attribute selection through decision rule construction (algorithm FRiS-GRAD) // *9th Conference (International) "Pattern Recognition and Image Analysis: New Information Technologies" Proceedings*. Nizhny Novgorod, 2008. Vol. 2. P. 335–338.
- [8] Аркадьев А. Г., Браверман Э. М. Обучение машины распознаванию образов. — М.: Наука, 1964.
- [9] Субботин С. А. Комплекс характеристик и критериев сравнения обучающих выборок для решения задач диагностики и распознавания образов // *Математичні машини і системи*, 2010. № 1. С. 25–39.
- [10] Борисова И. А., Дюбанов В. В., Загоруйко Н. Г., Кутненко О. А. Сходство и компактность // *Труды 14-й Всеросс. конф. «Математические методы распознавания образов»*, 2009. С. 89–92.
- [11] Борисова И. А. Алгоритм таксономии FRiS-Tax // *Научный вестник НГТУ*, 2007. № 3. С. 3–12.
- [12] Загоруйко Н. Г., Кутненко О. А. Алгоритм GRAD для выбора признаков // *Труды VIII Междунар. конф. «Применение многомерного статистического анализа в экономике и оценке качества»*. — Москва: МЭСИ, 2006. С. 81–89.
- [13] Pudil P., Novicova J., Kittler J. Floating search methods in feature selection. *Pattern Recognit. Lett.*, 1994, Vol. 15, № 11. P. 1119–1125.

- [14] Загоруйко Н. Г. Прикладные методы анализа данных и знаний. — Новосибирск: ИМ СО РАН, 1999. 268 с.
- [15] Merill T., Green O. M. On the effectiveness of receptions in recognition systems // *IEEE Trans. Inform. Theory*, 1963. Vol. IT-9. P. 11–17.
- [16] Барабаш Ю. Л., Варский Б. В., Зиновьев В. Т., Кириченко В. С., Спезгин В. Ф., Трахтман В. Ю. Автоматическое распознавание образов. — Киев: КВАИУ, 1963. 168 с.
- [17] Borisova I. A., Dyubanov V. V., Kutnenko O. A., Zagoruiko N. G. Use FRiS-Function for taxonomy, attribute selection and decision rule construction // *Knowledge Processing Data Analysis*. Berlin–Heidelberg: Springer-Verlag, 2011. P. 256–270.
- [18] Загоруйко Н. Г., Борисова И. А., Дюбанов В. В., Кутненко О. А. Количественная мера компактности и сходства в конкурентном пространстве // *Сибирский журнал индустриальной математики*, 2010. Т. XIII. № 1(41). С. 59–71.

References

- [1] Lemesko B. Yu., Lemesko S. B., Postovalov S. N., Chimitova E. V. 2011. *Statistical analysis of the data, modeling and research the laws of probability. Computer approach*. Novosibirsk: NGTU. 888 p.
- [2] Tixonov A. N., Goncharski A. V., Stepanov V. V., Yagola A. G. 1983. *Regularization algorithms and a priori information*. Moscow: Nauka.
- [3] Kolmogorov A. N. 1933. On the question of the suitability of the formulas of forecast found by statistical. *Zavodskaja Laboratorija* 1:164–167.
- [4] Ivashenko A. G., Stepashko V. S. 1985. *Noiseproof modeling*. Kiev: Naukova dumka.
- [5] Zagoruiko N. G., Kutnenko O. A. 2012. A quantitative measure of the compactness of the images and the method of its increase. *9th Conference (International) "Intellectualization of Information Processing" Proceedings*. Moscow: TORUS PRESS. 29–32.
- [6] Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. 2008. Methods of recognition based on the function of rival similarity. *Pattern Recognition Image Analysis* 18(1):1–6.
- [7] Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A. 2008. Attribute selection through decision rule construction (algorithm FRiS-GRAD). *9th Conference (International) "Pattern Recognition and Image Analysis: New Information Technologies" Proceedings*. Nizhniy Novgorod. 2:335–338.
- [8] Arkad'ev A. G., Braverman E. M. 1964. *Machine learning to pattern recognition*. Moscow: Nauka.
- [9] Subbotin S. A. 2010. Complex characterization and comparison criteria of training samples for diagnostics and pattern recognition. *Matematichni Mashini Sistemi* 1:25–39.
- [10] Borisova I. A., Dyubanov V. V., Zagoruiko N. G., Kutnenko O. A. 2009. The similarities and compactness. *14th National Conference "Mathematical Methods of Pattern Recognition" Proceedings*. 89–92.
- [11] Borisova I. A. 2007. Clustering algorithm FRiS-Tax. *Scientific Bulletin of NGTU* 3:3–12.
- [12] Zagoruiko N. G., Kutnenko O. A. 2006. Algorithm GRAD for feature selection. *8th Conference (International) "Application of Multivariate Statistical Analysis in Economics and Quality Assessment" Proceedings*. Moscow. 81–89.
- [13] Pudil P., Novicova J., Kittler J. 1994. Floating search methods in feature selection. *Pattern Recognition Lett.* 15(1):1119–1125.
- [14] Zagoruiko N. G. 1999. *Applied methods of data and knowledge analysis*. Novosibirsk: Institute of Mathematics SD RAS. 268 p.

- [15] *Merill T., Green O. M.* 1963. On the effectiveness of receptions in recognition systems. *IEEE Trans. Inform. Theory*. IT-9:11–17.
- [16] *Barabash Yu. L., Varsky B. V., Zinov'ev V. T., Kirichenko V. S., Sapegin V. F., Trahtman V. Yu.* 1963. *Automatic pattern recognition*. Kiev: KVAIU. 168 p.
- [17] *Borisova I. A., Dyubanov V. V., Kutnenko O. A., Zagoruiko N. G.* 2011. Use FRiS-Function for taxonomy, attribute selection and decision rule construction. *Knowledge Processing and Data Analysis*. Berlin–Heidelberg: Springer-Verlag. 256–270.
- [18] *Zagoruiko N. G., Borisova I. A., Dyubanov V. V., Kutnenko O. A.* 2010. A quantitative measure of compactness and similarities in the competitive space. *Siber. J. Industrial Math.* XIII(1(41)):59–71.

Поиск структурных различий изображений: алгоритмы и методы исследования*

Ф. А. Корнилов

fakornilov@mail.ru

Институт математики и механики им. Н. Н. Красовского Уральского отделения Российской академии наук (ИММ УрО РАН), 620990, Российская Федерация, г. Екатеринбург, ул. Софьи Ковалевской, 16

В работе рассматривается задача поиска структурных различий на разновременных космических снимках земной поверхности, под которыми понимаются появившиеся или исчезнувшие объекты местности. В качестве отправной точки служит морфологический анализ изображений Ю. П. Пытьева. Вводится общая формулировка понятия структуры изображения и структурных различий. Приводится ряд алгоритмов, решающих поставленную задачу для полутоновых изображений, а также предлагается вариант морфологического проектора для работы с цветными изображениями. Для алгоритма, основанного на морфологическом проекторе, приводятся методика построения распределения яркости выходного изображения в случае возмущения входных изображений аддитивным шумом, и определяются формула оптимального порога и оценки ошибок первого и второго рода результатов работы алгоритма для нее. Формулируется вычислительная методика сравнения алгоритмов поиска структурных различий изображений, для чего вводится специальная математическая модель, позволяющая максимально приблизить эксперимент к реальным задачам. С помощью этой методики определяется наилучший процент верного обнаружения (ложной тревоги), который можно получить при зафиксированном максимально допустимом уровне ложной тревоги (минимально допустимом уровне верного обнаружения), а также приводятся результаты сравнения качества работы предложенных алгоритмов между собой. Эксперименты на реальных данных показывают, что предложенные алгоритмы подходят для прикладного применения.

Ключевые слова: *обработка изображений; структурные различия; математическая морфология; регуляризация; марковские случайные поля*

Detection of structural differences in images: algorithms and methods of research*

F. A. Kornilov

Institute of Mathematics and Mechanics named after N. N. Krasovkiy, Ural Branch of the Russian Academy of Sciences, 16 S. Kovalevskaja Str., Ekaterinburg 620990, Russia

Background: The article is devoted to the problem of structural differences detection at satellite images taken at different times. Here, structural changes mean appeared or disappeared landscape objects. The starting point of the investigation is the morphological analysis of images theory by Yu. P. Pytiev.

Methods: In the work, the general notion of the image's structure and structural difference are introduced. Several algorithms for solving the problem for grayscale images are proposed and also, the morphological projector's modification for dealing with color images is given. For the algorithm based on the morphological projector in the case of image's distortion with additive noise, the method for computing the output image's values distribution is given. Also the formula for optimal threshold and estimation of the I and II type errors are given.

Results: The computational technique for comparison of the algorithms of structure changes detections in images is formulated. For such a purpose, the special mathematical model is introduced. This model provides an opportunity to maximally closer move to applied real problems. The proposed technique gives the way for computing the best value of the true positives rate (false alarm rate), with given fixed maximal value of the false alarm rate (minimal value of the true positives rate). Also, the comparison results of the performance of the proposed algorithms are given.

Concluding Remarks: Experiments with real data demonstrate that the proposed algorithms are good enough for applications.

Keywords: *image processing; structural differences; mathematical morphology; regularization; markov random fields*

Введение

Автоматический анализ данных дистанционного зондирования, в частности космических снимков в оптическом диапазоне, является важной проблемой при решении многих практических задач [1]. Необходимость своевременного обновления топографических карт с наименьшими затратами делает крайне актуальной задачу поиска структурных различий на разновременных снимках одного и того участка земной поверхности. В неформальном, содержательном смысле структурные различия — это существенные изменения местности, при которых объекты интереса появляются, исчезают или изменяют свою форму. При этом изменения освещенности и цвета объектов структурными различиями не считаются. Поскольку процедура создания картографических материалов является достаточно трудоемкой и затратной, то решение об обновлении, в частности, топографических карт принимается в случае, если на местности произошли существенные изменения объектового состава. Таким образом, одной из актуальных задач анализа цифровых космических снимков становится поиск структурных различий на двух разновременных снимках с целью выявления изменений в составе объектов сцены на них. Данная статья является продолжением серии работ, посвященных задаче поиска структурных различий изображений [2, 3].

Задача поиска различий на двух изображениях является классической для анализа изображений и возникает в различных областях компьютерного зрения, таких как сжатие видеоданных, системы видеонаблюдения и других. Основными подходами к ее решению являются методы попиксельного анализа, не учитывающие специфику объектов; методы, основанные на анализе контуров объектов [4] и методы, основанные на текстурной классификации [5]. Фундаментальное отличие этих подходов сводится к различному пониманию структуры изображения, среди многообразия подходов к которому следует выделить морфологический анализ изображений Ю.П. Пытьева [6], который в данной работе является основой для построения алгоритмов поиска структурных различий изображений.

Основные определения

Под *изображением* f будем понимать функцию $f : X \rightarrow \mathbb{R}$, а семейством допустимых изображений будем называть конечномерное евклидово пространство Φ таких функций f ,

наделенное естественным скалярным произведением $(f, g) = \sum_X f(x)g(x)$ и нормой $\|f\| = (\sum_X f^2(x))^{1/2}$. Значение $f(x)$ будем называть *яркостью* изображения f в точке $x \in X$, а множество пикселей X (поле зрения) — подмножество плоскости $\mathbb{R}^2$, $S = |X|$ (количество пикселей) — *размером* (или *площадью*) изображения.

Структура изображения — это некоторый способ описания информации о представленной на изображении сцене. Разумеется, для конкретного снимка таких описаний может быть бесчисленное множество и выбор подходящего описания напрямую зависит от требований задачи. Структура изображения должна учитывать существенные изменения сцены и быть инвариантной к несущественным в контексте данной задачи изменениям. Помимо этого, должен существовать способ сравнения структур изображений между собой с целью определения наличия *структурных* различий между ними. Существует множество различных определений структуры изображений, среди которых важное место занимает теория морфологического анализа Ю.П.Пытьева [6], в котором развиты методы, инвариантные относительно преобразований, характеризующих влияние условий регистрации изображений. В его основе лежит понятие *формы* изображения f , под которой понимается оператор проектирования (морфологический проектор) P_f в пространстве изображений Φ на множество изображений, полученных из исходного путем применения всех возможных преобразований яркости:

$$P_f g(x) = \sum_i \frac{\sum_{x' \in X} g(x') \cdot \chi_i^f(x')}{\sum_{x' \in X} \chi_i^f(x')} \cdot \chi_i^f(x), \quad (1)$$

где

$$\chi_i^f(x) = \begin{cases} 1, & \text{если } f(x) = i; \\ 0, & \text{иначе.} \end{cases}$$

Выбор конкретного определения структуры должен быть связан с практическими требованиями задачи. В данной работе поиск структурных различий осуществляется с целью оценки площади произошедших изменений, при этом изображения содержат дискретный набор яркостей, поэтому в качестве носителей структуры будут выбраны множества уровней яркости изображения:

Определение 1. Структурой Ω изображения f называется семейство $\mathcal{L}_f = \{L_f(i)\}_i$ множеств уровня $L_f(i) = \{x \in X \mid f(x) = i\}$ функции f .

В данной работе основное внимание уделяется относительной радиометрической (яркостной) коррекции исходных снимков. Функции, пригодные для яркостной коррекции изображений связаны с идеей сохранения структуры одного из изображений с приведением яркостей его пикселей максимально близко к яркостям второго изображения. Применительно к задаче поиска структурных различий это означает построение изображений f' и g' , которые были бы по структуре близки к f и g , а по яркости f' близко к g , а g' близко к f ; и дальнейшего вычисления различий уже для построенных изображений с g и f соответственно:

$$\begin{aligned} f' &= \arg \min(\|h - g\|^2 + \chi(h; f) \mid h \in \Phi); \\ g' &= \arg \min(\|h - f\|^2 + \chi(h; g) \mid h \in \Phi), \end{aligned} \quad (2)$$

где под χ понимается характеристическая функция структурной эквивалентности, которая отражает тот факт, что изображение h получено путем изменения значений уровней яркости изображения f с сохранением его структуры. Необходимость построения второго изображения связана с тем, что операция яркостной коррекции несимметрична, т. е. ее применение позволяет найти лишь «появившиеся» изменения (объекты, отсутствующие на первом изображении, но присутствующие на втором). Нахождение «исчезнувших» объектов требует проведения симметризации яркостной коррекции.

В работе будут рассматриваться алгоритмы, в основе которых лежат следующие характеристические функции структурной эквивалентности:

1. характеристическая функция χ_π : $\chi_\pi(h; f) = 0$, если форма h не сложнее f в смысле морфологии Ю. П. Пытьева, и иначе равна бесконечности, т. е. равенство функции χ_π нулю означает, что имеется некоторая функция преобразования яркости $F: \mathbb{R} \rightarrow \mathbb{R}$ такая, что $h(x) = F(f(x)) \forall x \in X$;

2. характеристическая функция χ_1 : $\chi_1(h; f) = 0$, если $h(x) = kf(x) + b \forall x \in X$ и для некоторых вещественных k, b , и иначе равна бесконечности;

3. характеристическая функция χ_2 : $\chi_2(h; f) = 0$, если $h(x) = kf(x)^2 + b \cdot f(x) + c \forall x \in X$ и для некоторых вещественных k, b, c , и иначе равна бесконечности;

4. характеристическая функция χ_m :

$$\chi_m(r; f) = \lambda \sum_{x \in X} \sum_{y \in N_x} G_f(h(x) - h(y), x, y);$$

$$G_f(i, x_1, x_2) = \begin{cases} \min(i^2, \alpha^2), & \text{если } |f(x_1) - f(x_2)| < \alpha; \\ 0, & \text{иначе,} \end{cases}$$

где N_x — окрестность точки x , λ и α — положительные параметры. Параметр λ отвечает за близость приведения изображения f к g : чем он больше, тем больше алгоритм будет стремиться сохранить структуру изображения f , из-за чего яркостное приведение f к g будет менее точным. Если же $\lambda \rightarrow 0$, то структура изображения f будет разрушаться для наиболее полного совпадения яркостей f и g (очевидно, что если $\lambda = 0$, то $f' = g$). Параметр α характеризует структуру изображения f , задавая допустимый перепад яркости между различными линиями уровня: если яркости двух точек отличаются меньше, чем на α , то они принадлежат одному уровню.

Теперь можно сформулировать определение структурного различия:

Определение 2. Структурным различием называется множество пикселей, удовлетворяющее следующему условию:

$$SD(f, g) = \{x \in X \mid |f'(x) - g(x)| > T \text{ или } |f(x) - g'(x)| > T\},$$

где f' и g' определены формулами (2), T — некоторое пороговое значение.

Можно показать, что применение формулы (2) эквивалентно построению морфологической проекции Ю. П. Пытьева при определенном выборе характеристической функции χ_π . Доказательство этого факта и аналогичное утверждение для класса линейных и квадратичных функций приводится в [2].

Введем на пространстве изображений Φ отношение частичного порядка. Для произвольных $f, g \in \Phi$ будем говорить, что структура изображения f не сложнее структуры изображения g , и писать $\mathcal{L}_f \preceq \mathcal{L}_g$, если для каждого значения i функции f найдется разбиение $L_f(i) = \bigcup_{j=1}^n L_g(i_j)$, $n \geq 1$, причем $i_k \neq i_l \forall k, l \in [1, n]$. Если говорить неформально, то

на изображении g сохранены все границы изображения f , но к ним, возможно, добавлены новые. Очевидно, что $|\mathcal{L}_f| \leq |\mathcal{L}_g|$ (здесь $|\cdot|$ — мощность множества). Определение структуры изображения (1) наиболее близко по сути к форме изображения в морфологическом анализе Ю.П. Пытьева, и можно доказать следующее утверждение:

Утверждение 1. $\mathcal{L}_f \preceq \mathcal{L}_g$ тогда и только тогда, когда $f = P_g f$.

Благодаря этому утверждению можно разбить все возможные варианты соотношения структур исходных изображений на следующие три случая.

1. Изображения f и g независимы: $\mathcal{L}_f \not\preceq \mathcal{L}_g$ и $\mathcal{L}_g \not\preceq \mathcal{L}_f$. Тогда $f \neq P_g f$ и $g \neq P_f g$. В этом случае будем говорить, что структура этих изображений различна, т. е. обнаружено структурное различие изображений в целом.

2. Изображения f и g изоморфны: $\mathcal{L}_f \preceq \mathcal{L}_g$ и $\mathcal{L}_g \preceq \mathcal{L}_f$. Тогда $f = P_g f$ и $g = P_f g$. В этом случае будем говорить, что изображения имеют одинаковую структуру, т. е. структурных различий нет.

3. Для изображений f и g выполнено либо $\mathcal{L}_f \preceq \mathcal{L}_g$ и $\mathcal{L}_g \not\preceq \mathcal{L}_f$, либо $\mathcal{L}_f \not\preceq \mathcal{L}_g$ и $\mathcal{L}_g \preceq \mathcal{L}_f$. Поскольку эти случаи симметричны, то, не нарушая общности, далее везде будем считать, что выполняются условия $\mathcal{L}_f \preceq \mathcal{L}_g$ и $\mathcal{L}_g \not\preceq \mathcal{L}_f$ ($f = P_g f$ и $g \neq P_f g$). В отличие от случаев независимых и изоморфных изображений здесь можно говорить не только об обнаружении структурного различия, но и о его локализации, т. е. об определении наличия или отсутствия структурного различия в произвольной зафиксированной точке x_c .

Математическая модель структурных различий

Пусть дана пара изображений $f^0, g^0 \in \Phi$. Первое изображение f^0 (будем называть его фоном) содержит N_f уровней яркости. Значение яркости i -го уровня обозначим f_i^0 , площадь уровня — $S_i = |L_{f^0}(f_i^0)|$ ($i \in [1, N_f]$). Изображение g^0 строится путем добавления к i -ому уровню фона объекта, состоящего из N_g^i уровней яркости. Значение яркостей уровней разбиения $j \in [0, N_g^i]$ обозначим $g_{i,j}^0$, $S_{i,j} = |L_{g^0}(g_{i,j}^0)|$, $S_i = \sum_{j=0}^{N_g^i} S_{i,j}$ (индексы j от 1 до

N_g^i соответствуют объекту, 0 — фону). Таким образом, $L_{f^0}(f_i^0) = \bigcup_{j=0}^{N_g^i} L_{g^0}(g_{i,j}^0)$, $i \in [1, N_f]$.

Чтобы избежать краевых эффектов, добавим условие, что площадь добавленного объекта меньше площади фона. Таким образом, изображение g состоит из $N_f + \sum_{i=1}^{N_f} N_g^i$ уровней яркости, где первое слагаемое соответствует количеству значений функции f^0 , второе слагаемое соответствует добавленному объекту. Будем считать, что в данной точке есть структурное различие, если эта точка принадлежит объекту. Для моделирования реальной ситуации в задаче поиска структурных различий к каждому изображению добавляем дискретный, стационарный и независимый в каждой точке аддитивный шум с известными параметрами: для первого изображения — n_η , для второго — n_ξ (здесь n — означает случайную величину, η и ξ — плотности распределения). Обозначим $f(\cdot) = f^0(\cdot) + n_\eta$ и $g(\cdot) = g^0(\cdot) + n_\xi$. В результате имеем два семейства случайных величин (два случайных поля):

$$\begin{aligned} f &= \{f(x_i)\}_{i=1}^S = \{f^0(x_i) + n_\eta\}_{i=1}^S, \\ g &= \{g(x_i)\}_{i=1}^S = \{g^0(x_i) + n_\xi\}_{i=1}^S, \end{aligned} \quad (3)$$

которые можно интерпретировать как модели случайных изображений, к которым будем применять алгоритм поиска структурных различий.

Поскольку изображения f и g — случайные поля, то значение разностного изображения $R = |f' - g|$ (более подробно описания алгоритма будет дано в следующей главе) в точке $x_c \in X$ также является величиной случайной, и основной задачей будет построение способа ее вычисления. Однако значения случайной величины $R(x_c)$ существенно зависят от того, есть структурное различие в точке x_c или его нет, поэтому необходимо рассматривать соответствующие два условных распределения.

У с л о в и е 1: точка x_c принадлежит объекту, т. е. в этой точке есть структурное различие.

У с л о в и е 2: точка x_c принадлежит фону, т. е. структурного различия в этой точке нет.

Распределение вероятности $r(R(x_c) \mid x_c \text{ принадлежит объекту})$ обозначим $r^+(R)$. Второе распределение $r(R(x_c) \mid x_c \text{ принадлежит фону})$ обозначим $r^-(R)$.

З а д а ч а 1. Получить условные распределения вероятностей $r^+(R)$ и $r^-(R)$ величины $R(x_c)$ для модельной задачи.

Алгоритмы поиска структурных различий изображений

Общая схема алгоритма поиска структурных различий изображений

В данном пункте приводится описание разработанных алгоритмов поиска структурных различий изображений. Их общая схема такова:

1. Исходные изображения сканируются с шагом 1 локальным окном заданного размера $d \times d$. Центральную точку этого окна будем обозначать x_c . В случае необходимости производится геометрическое выравнивание фрагментов.

2. Для каждого положения окна по двум фрагментам сравниваемых изображений f и g строятся две функции преобразования яркости F_{fg} и F_{gf} . Использование схемы с двумя функциями позволяет добиться симметризации результата, т. е. становится неважным, сравнивается первое изображение со вторым или наоборот.

3. С помощью функций F_{fg} и F_{gf} строятся преобразованные изображения $f' = F_{fg}(f)$ и $g' = F_{gf}(g)$, при этом яркость изображения f' «выровнена» по яркости изображения g с сохранением структуры изображения f , и аналогично — для изображения g' .

4. Строятся разностные изображения $R_{fg}(x) = |f'(x) - g(x)|$ и $R_{gf}(x) = |g'(x) - f(x)|$. Для этих изображений яркость точки характеризует величину структурного несоответствия исходных изображений, т. е. чем ярче точка, тем вероятнее, что в ней присутствует структурное различие.

5. Для завершения симметризации строится результирующее разностное изображение R , яркость каждой точки которого есть максимум яркостей точек разностных изображений с соответствующими координатами: $R(x) = \max(R_{fg}(x), R_{gf}(x))$.

6. Производится пороговая обработка изображения R , параметр порога будем обозначать T . Если стоит задача локализации структурного различия, то оценивается яркость центральной точки: при выполнении $R(x_c) \geq T$ в данной точке присутствует структурное различие, и в соответствующую точку выходного результирующего изображения $\mathbf{R}(x_c)$ (изображения-результата работы работы алгоритма) записывается значение 255, если нет, то 0. Для задачи обнаружения структурного различия условие $R(x) \geq T$ проверяется для каждой точки сканирующего окна, и если число таких точек значительно, то считается, что в данном окне обнаружено структурное различие, и в $\mathbf{R}(x_c)$ записывается значение

255, иначе 0. Таким образом, обработка каждого сканирующего окна дает на результирующем изображении одну точку, яркость которой 255 или 0. Далее из этих точек формируются связные области структурных различий — конечный результат. В целях борьбы со случайными выбросами можно фильтровать найденные области, отбрасывая те, площадь которых незначительна.

Функция преобразования яркости — это попиксельно действующий оператор относительной радиометрической коррекции исходных изображений, который строится для каждого положения сканирующего окна на основе наборов яркостей изображений в нем, и применяется для «выравнивания» яркостей фрагментов с сохранением их структуры. Такой подход позволяет игнорировать не считающиеся структурными различиями объекты, изменившие свой цвет, но сохранившие форму. В качестве таких функций в данной работе используются:

1. морфологический проектор (1);
2. регуляризованный вариант морфологического проектора, предназначенный для устранения влияния шума с сохранением точности границ найденных различий, и связанный со сглаживанием характеристических функций уровней яркости:

$$P_f g(x) = \frac{\sum_i \tilde{f}_i(x)}{\sum_j \tilde{\chi}_i^f(x)} = \sum_i \left(\frac{\sum_{x' \in X} g(x') \cdot \tilde{\chi}_i^f(x')}{\sum_{x' \in X} \tilde{\chi}_i^f(x')} \cdot \frac{\tilde{\chi}_i^f(x)}{\sum_j \tilde{\chi}_i^f(x)} \right), \quad (4)$$

где $\tilde{\chi}_i^f$ вычисляется по формуле

$$\tilde{\chi}_i^f(x) = \exp \left(-\frac{(f(x) - i)^2}{\sigma_c^2} \right) \cdot \exp \left(-\frac{(x - x_c)^2}{\sigma_d^2} \right);$$

3. линейная функция преобразования яркости:

$$f'(x) = k \cdot f(x) + b, \quad (5)$$

где коэффициенты находятся с помощью МНК по двум фрагментам изображений, ограниченных сканирующим окном, из соотношения

$$k \cdot f(\cdot) + b = g(\cdot);$$

4. квадратичная функция преобразования яркости:

$$f'(x) = a \cdot f^2(x) + b \cdot f(x) + c, \quad (6)$$

где коэффициенты находятся с помощью МНК по двум фрагментам изображений, ограниченных сканирующим окном, из соотношения

$$a \cdot f^2(\cdot) + b \cdot f(\cdot) + c = g(\cdot).$$

В основе алгоритма поиска структурных различий с непрерывной структурной характеристической функцией χ_m лежит, как и в предыдущих случаях, яркостная коррекция входных снимков, заключающаяся в приведении набора яркостей первого изображения к яркости второго с сохранением структуры первого изображения. Для улучшения результата также производится приведение каждого изображения к самому себе, что позволяет

избавиться от присутствующего шума. Реализация такого алгоритма сводится к задаче глобальной минимизации функционалов (2), что эквивалентно минимизации следующей энергетической функции:

$$\begin{aligned} E(f'; f, g) &= \|f'(x) - g(x)\| + \chi_m(f'; f) = \\ &= \|f'(x) - g(x)\| + \lambda \sum_{x \in X} \sum_{y \in N_x} G_f(f'(x) - f'(y), x, y); \end{aligned} \quad (7)$$

$$G_f(i, x_1, x_2) = \begin{cases} \min(i^2, \alpha^2), & \text{если } |f(x_1) - f(x_2)| < \alpha; \\ 0, & \text{иначе,} \end{cases}$$

где N_x — окрестность точки x , λ и α — положительные параметры (описаны выше). Для нахождения минимума функции $E(f'; f, g)$ будет использоваться аппарат марковских случайных полей (МСП) [7]. Одним из наиболее простых в реализации методов приближенного решения подобной задачи является алгоритм Метрополиса [8]. В качестве энергетической функции для данного метода будем использовать введенную функцию $E(f'; f, g)$.

Теоретическое исследование алгоритма

Описав схему алгоритма и используемые в нем функции преобразования яркости, можно перейти к решению задачи локализации структурных различий для модельной задачи 1 для алгоритма поиска структурных различий, основанного на морфологическом проекторе. Поскольку алгоритм поиска структурных различий оценивает только одну зафиксированную точку $x_c \in X$, то при использовании морфологического проектора важно только множество $L_f(f(x_c))$, которому эта точка принадлежит. Пусть $k = |L_f(f(x_c))|$. Очевидно, что $k \in [1, S]$. Введем обозначения: $L_f^{i,j}(f(x_c)) = \{x \in L_f(f(x_c)) \mid f^0(x) = f_i^0 \text{ и } g^0(x) = g_{i,j}^0\}$; под $B(n, m, p)$ будем понимать биномиальное распределение вероятностей: вероятность появления события ровно n раз в серии из m испытаний при условии, что в единичном испытании вероятность его появления равна p ; конфигурация $a = (a_{1,0}, a_{1,1}, \dots, a_{1,N_g^1}, \dots, a_{N_f,0}, a_{N_f,1}, \dots, a_{N_f,N_g^{N_f}})$, где $a_{q,j}$ соответствует мощности пересечения множества $L_f(f(x_c))$ и уровня яркости объекта с номером j , полученного из разбиения уровня фона с номером q (индекс 0 соответствует фону на втором изображении). Из-за разницы яркостей незашумленного изображения f^0 в точках разных уровней после добавления шума распределения вероятностей яркостей в этих уровнях также будут различными, поэтому далее будем обозначать $p_q(f = i) = p(f(x) = i \mid f^0(x) = f_q^0)$ — вероятность появления яркости i в точке, которая на изображении f^0 принадлежит уровню яркости с номером q . Теперь можно доказать следующую теорему:

Теорема 1. Пусть дано случайное поле f , заданное по формуле (3).

1. В задаче 1 при выполнении условий 1 и $x_c \in L_f^{n,m}(f(x_c))$ ($n \in [1, N_f]$ и $m \in [1, N_g^n]$) вероятность появления реализации поля f такой, что $|L_f(f(x_c))| = k$, $|L_f^{n,m}(f(x_c))| = a_{n,m} + 1$ и $|L_f^{q,j}(f(x_c))| = a_{q,j}$ для $(q, j): q \in [1, N_f]$, $j \in [0, N_g^q]$, $(q, j) \neq (n, m)$, равна

$$\begin{aligned} p_{k,a}^{n,m} &= \sum_i \left(p_n(f = i) \prod_{q=1}^{n-1} \prod_{j=0}^{N_g^q} B(a_{q,j}, S_{q,j}, p_q(f = i)) \times \right. \\ &\times \prod_{j=0}^{m-1} B(a_{n,j}, S_{n,j}, p_n(f = i)) B(a_{n,m}, S_{n,m} - 1, p_n(f = i)) \times \end{aligned}$$

$$\times \prod_{j=m+1}^{N_g^n} B(a_{n,j}, S_{n,j}, p_n(f=i)) \prod_{q=n+1}^{N_f} \prod_{j=0}^{N_g^q} B(a_{q,j}, S_{q,j}, p_q(f=i)) \Bigg),$$

где

$$\begin{aligned} \sum_{q=1}^{N_f} \sum_{j=1}^{N_g^q} a_{q,j} &= k-1; \\ 0 \leq a_{q,j} &\leq S_{q,j} \quad \forall (q,j): q \in [1, N_f], j \in [0, N_g^q], (q,j) \neq (n,m); \\ 0 &\leq a_{n,m} \leq S_{n,m} - 1. \end{aligned}$$

2. В задаче 1 при выполнении условий 2 и $x_c \in L_f^{n,0}(f(x_c))$ ($n \in [1, N_f]$) вероятность появления реализации поля f такой, что $|L_f(f(x_c))| = k$, $|L_f^{n,0}(f(x_c))| = a_{n,0} + 1$ и $|L_f^{q,j}(f(x_c))| = a_{q,j}$ для $(q,j): q \in [1, N_f], j \in [0, N_g^q], (q,j) \neq (n,0)$, равна

$$\begin{aligned} p_{k,a}^{n,0} &= \sum_i \left(p_n(f=i) \prod_{q=1}^{n-1} \prod_{j=0}^{N_g^q} B(a_{q,j}, S_{q,j}, p_q(f=i)) \times \right. \\ &\times B(a_{n,0}, S_{n,0} - 1, p_n(f=i)) \prod_{j=1}^{N_g^n} B(a_{n,j}, S_{n,j}, p_n(f=i)) \times \\ &\times \left. \prod_{q=n+1}^{N_f} \prod_{j=0}^{N_g^q} B(a_{q,j}, S_{q,j}, p_q(f=i)) \right), \end{aligned}$$

где

$$\begin{aligned} \sum_{q=1}^{N_f} \sum_{j=1}^{N_g^q} a_{q,j} &= k-1; \\ 0 &\leq a_{n,0} \leq S_{n,0} - 1; \\ 0 \leq a_{q,j} &\leq S_{q,j}, \quad \forall (q,j): q \in [1, N_f], j \in [0, N_g^q] \quad (q,j) \neq (n,0). \end{aligned}$$

Таким образом, пространство всех изображений разбивается на классы в зависимости от параметров (k, a) путем перебора всех возможных вариантов реализации поля (изображения) f , а теорема 1 позволяет определить вероятности для классов из этого разбиения. Можно доказать основную теорему:

Теорема 2. 1. В задаче 1 при условии 1 распределение вероятностей случайной величины $R(x_c)$ имеет вид:

$$r^+(i) = \frac{\sum_{n=1}^{N_f} \sum_{m=1}^{N_g^n} (S_{n,m} r_{n,m}(i))}{\sum_{u=1}^{N_f} \sum_{v=1}^{N_g^u} S_{u,v}},$$

где

$$r_{n,m}(i) = \sum_{k,a} (p_{k,a}^{n,m} p(|G_{k,a}^{n,m} + E_k| = i));$$

величина $p_{k,a}^{n,m}$ определяется теоремой 1:

$$G_{k,a}^{n,m} = \frac{1}{k} \left(\sum_{q=1}^{N_f} \sum_{j=0}^{N_g^q} (a_{q,j} g_{q,j}^0) + (1-k) g_{n,m}^0 \right); \quad E_k = \frac{1}{k} \sum_{j=1}^{k-1} \varepsilon_j + \frac{1-k}{k} \varepsilon_c;$$

ε_j — распределение вероятностей случайной величины $g(x_j)$, соответствующей точке $x_j \in L_f(f(x_c))$, ε_c — распределение вероятностей случайной величины $g(x_c)$.

2. В задаче 1 при условии 2 распределение вероятностей случайной величины $R(x_c)$ имеет вид:

$$r^-(i) = \frac{\sum_{n=1}^{N_f} (S_{n,0} r_{n,0}(i))}{\sum_{u=1}^{N_f} S_{u,0}},$$

где

$$r_{n,0}(i) = \sum_{k,a} (p_{k,a}^{n,0} p(|G_{k,a}^{n,0} + E_k| = i)),$$

величина $p_{k,a}^{n,0}$ определяется теоремой 1:

$$G_{k,a}^{n,0} = \frac{1}{k} \left(\sum_{q=1}^{N_f} \sum_{j=0}^{N_g^q} (a_{q,j} g_{q,j}^0) + (1-k) g_{n,0}^0 \right);$$

определения E_k , ε_j и ε_c такие же, как в п. 1.

Опираясь на теорему 2, можно определить способ вычисления оптимального порога алгоритма. В качестве критерия оптимальности будет использоваться *percentage correct classification*, т. е. минимум суммы ошибок первого и второго рода. Имеет место следующая теорема.

Теорема 3. В условиях задачи 1 порог алгоритма поиска структурных различий на основе морфологического проектора, вычисленный по следующей формуле:

$$T_{opt} = \arg \min_T \left(\sum_{i < T} r^+(i) + \sum_{i > T} r^-(i) \right),$$

является оптимальным в смысле критерия *percentage correct classification*.

Доказательство теорем 1–3 приведено в [3]. Результаты численного эксперимента по применению теоремы 3 изображены на рис. 1 для случая объекта, состоящего из двух уровней яркости, графики распределений вероятностей $r^+(R)$ и $r^-(R)$, полученных с помощью теоремы 2. Параметры изображений $S = 441$, $S_1 = 60$, $S_2 = 60$, $g_1 = 120$, $g_2^0 = 80$, $g_0^0 = 0$. Оба изображения содержат дискретный шум, распределение которого аппроксимируется нормальным распределением с параметрами $N(0, 10)$. График распределения $r^+(R)$ состоит из двух пиков, соответствующих двум уровням объекта.

Морфологический проектор для цветных изображений

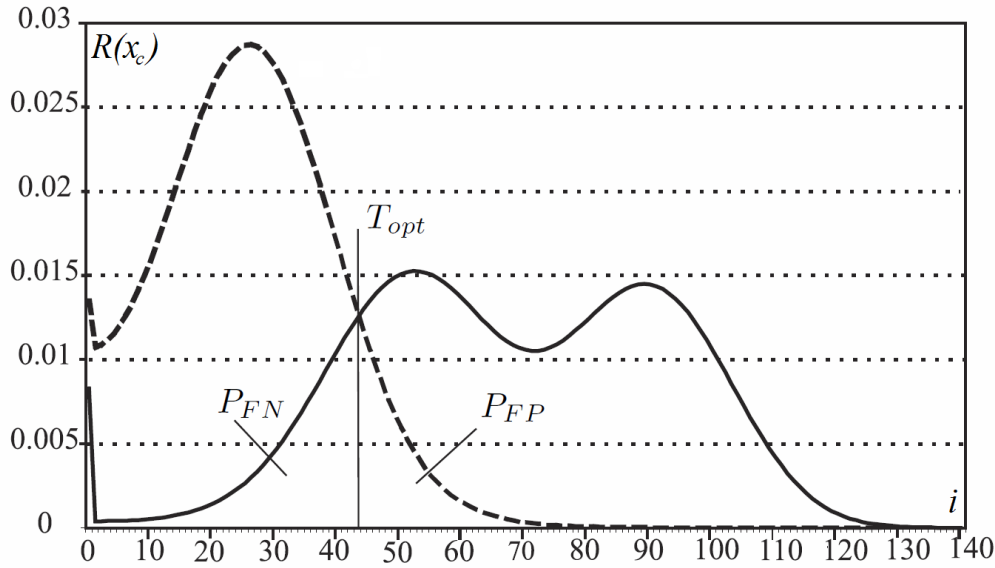


Рис. 1. Графики распределений $r^+(R)$ (сплошная линия) и $r^-(R)$ (пунктирная линия) для изображений с параметрами $S = 441$, $S_1 = 60$, $S_2 = 60$, $g_1^0 = 120$, $g_2^0 = 80$, $g_0^0 = 0$ (добавление двухуровневого объекта). Оптимальный порог $T_{opt} = 43$. Оценки ошибок: $P_{FN} = 0.15$, $P_{FP} = 0.1$

Представленные в данной работе алгоритмы предназначены для поиска структурных различий на полутоновых снимках земной поверхности. Тем не менее определенный интерес представляет решение данной задачи для цветных изображений. Простейшим вариантом является поканальная обработка цветных снимков предложенными методами. Однако, такой подход не дает ничего нового по сравнению со схемой обесцвечивания и применения стандартных методов, в то время как цветное изображение содержит гораздо больше полезной информации. Идея предлагаемого метода идентична вышеописанным подходам: первоначально требуется произвести радиометрическую коррекцию входных изображений. Ю. П. Пытьевым в работе [6] тема устранения влияния условий регистрации для цветных изображений раскрыта достаточно широко. Здесь будет предложен один из возможных вариантов проектора.

Поскольку теперь речь будет идти о цветных изображениях, то придется внести изменения во введенное определение изображения. А именно, в данном пункте под *изображением* $\mathbf{f}$ будем понимать функцию $\mathbf{f}: X \rightarrow \mathbb{R}^3$, а значение $\mathbf{f}(x) = (f_R(x), f_G(x), f_B(x))$ будем называть *яркостью* изображения $\mathbf{f}$ в точке $x \in X$, где f_R , f_G и f_B — яркости по красному, зеленому и синему каналу, соответственно.

Суть морфологического проектора заключается в усреднении яркостей второго изображения по уровням яркости первого (1). Однако представляется довольно затруднительным задать уровень яркости как некоторое множество в пространстве $\mathbb{R}^3$: любое разбиение является и подходящим, и неудачным, что делает задачу такого разбиения неоднозначной. Однако известно, что предлагаемый в данной работе алгоритм оценивает центральную точку окна, поэтому можно задавать уровень яркости как точки, близкие по яркости к центральной x_c :

$$\hat{L}_{\mathbf{f}}(x_c) = \{x \in X \mid \max(|f_R(x) - f_R(x_c)|, |f_G(x) - f_G(x_c)|, |f_B(x) - f_B(x_c)|) \leq t_c\},$$

где t_c — параметр яркостной близости. В таком случае каждой точке первого изображения соответствует свой набор точек и свой уровень яркости, но эти наборы точек могут иметь непустое пересечение. Тогда формула морфологического проектора для цветных изображений выглядит следующим образом:

$$P_f g(x) = \sum_i \frac{\sum_{x' \in X} g(x') \hat{\chi}^f(x', x)}{\sum_{x' \in X} \hat{\chi}^f(x', x)} \cdot \chi_i^f(x), \quad (8)$$

где

$$i = (i_R, i_G, i_B); \quad \chi_i^f(x) = \begin{cases} 1, & \text{если } f(x) = i; \\ 0, & \text{иначе;} \end{cases} \quad \hat{\chi}^f(x', x) = \begin{cases} 1, & \text{если } x' \in \hat{L}_f(x); \\ 0, & \text{иначе.} \end{cases}$$

Алгоритм поиска структурных различий с функцией преобразования яркости (8) полностью совпадает с описанным выше. На рис. 2 представлены результаты его работы для двух цветных изображений городской застройки:

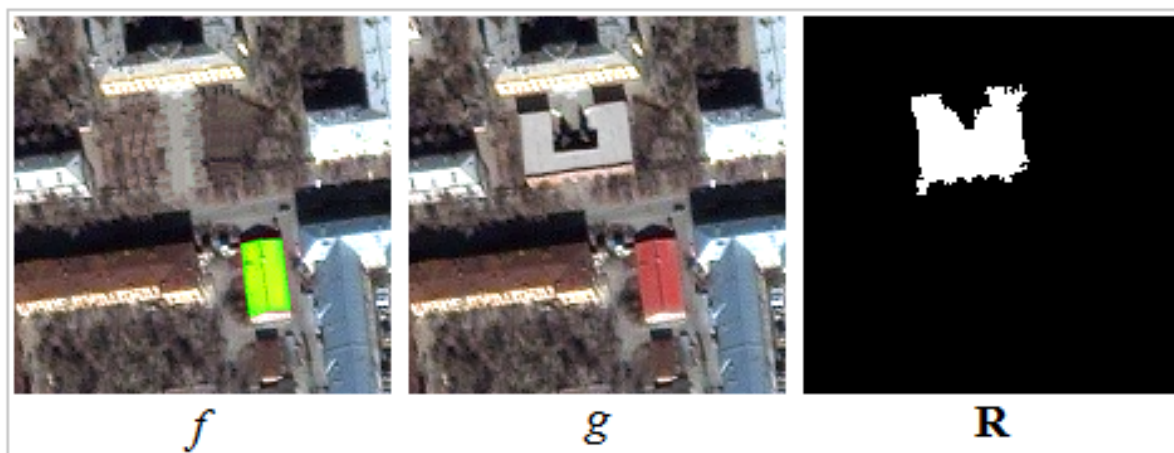


Рис. 2. Результаты работы алгоритма поиска структурных различий для цветных изображений: исходные изображения f и g и найденные области структурных различий R

Параметры алгоритма были выбраны вручную: яркостная близость для линии уровня $t_c = 16$, порог $T = 55$. Можно видеть, что алгоритм успешно нашел появившийся дом, проигнорировав строение, крыша которого изменила свой цвет. Таким образом, можно говорить о применимости данного метода для поиска структурных различий на разновременных цветных снимках земной поверхности.

Эмпирическое исследование алгоритмов и сравнение результатов их работы

Работа каждого алгоритма напрямую зависит от выбранного набора его параметров. Ключевое место среди них занимает оптимальный порог, однако оценка требуется и для всех остальных: например, для размера сканирующего окна. Среди всех возможных значений параметров требуется выбрать те, при которых алгоритм показывает наилучшее качество работы в зависимости от уровня шума, присутствующего на изображениях. Подходящим критерием качества работы является анализ ROC-кривых [9] (ROC — *Receiver*

Таблица 1. Максимальные значения TP (минимальные FP) при зафиксированном уровне FP = 0,1 (TP = 0,9) для алгоритмов поиска структурных различий, и оптимальные параметры, при которых они получены

Алгоритм	TP _{max}	FP _{min}	Параметры
Рег. морфологический проектор	0.892091	0.113738	$d = 29, \sigma_c = 2$
Квадратичная функция	0.890949	0.113725	$d = 23$
Морфологический проектор	0.885229	0.120545	$d = 27$
Линейная функция	0.884299	0.122274	$d = 23$
Глобальная оптимизация	0.80634	0.332012	$d = 19, \lambda = 1816, \alpha = 5$

Operating Characteristic, операционная характеристика приемника), появившийся в теории обработки сигналов в середине прошлого века. Количественную интерпретацию ROC-кривой дает показатель AUC (англ. *Area Under the Curve*, площадь под ROC-кривой) — площадь, ограниченная ROC-кривой и осью доли ложных положительных классификаций. Чем выше показатель AUC, тем качественнее классификатор, при этом значение 0,5 соответствует случайному решению.

В первой части данной статьи была предложена математическая модель структурных различий, подходящая для теоретического исследования алгоритмов. Однако, несмотря на ее близость к реальным изображениям, спутниковые снимки все же намного сложнее. Тем не менее для проведения численного эксперимента можно расширить эту модель путем усложнения «фона» исходных изображений и добавляемых «объектов», в качестве которых теперь будут рассматриваться не уровни яркости, а фрагменты спутниковых снимков, что позволит учесть следующие соображения:

1. Сравнение алгоритмов необходимо проводить на реальных, а не на синтезированных изображениях. Работа каждого алгоритма будет оцениваться путем сравнения (с помощью анализа ROC-кривых) полученных результатов с ручной разметкой исходных изображений. Однако получение ручной разметки реальных снимков представляет собой сложную и трудоемкую задачу. Причиной является сложность дешифрирования таких снимков, даже в ручном режиме. Иными словами, на снимках могут (и будут) присутствовать области, которые нельзя однозначно отнести к областям структурных различий или областям их отсутствия.

2. Возможность варьировать и строго контролировать тип и уровень шума, добавляемого к изображениям. То есть необходимо, что бы исходные изображения имели «нулевой» уровень шума.

3. Возможность выбора размера и типа объектов, присутствующих на изображениях.

Построения пары изображений, которые станут входными данными для алгоритма поиска структурных различий, и изображения с ручной разметкой осуществляется следующим образом. Выбирается космический снимок подходящего участка земной поверхности, который разделяется на две части: первый фрагмент и его копия становятся исходными изображениями, второй фрагмент будет донором для генерации структурных различий. Поскольку исходные изображения изначально идентичны, то одинаковы и присутствующие на них уровни яркости, вызванные влиянием шума. Таким образом, эти уровни яркости становятся в некотором смысле «содержательными», т. е. определяют «фон» на обоих изображениях. Применение алгоритмов поиска структурных различий к таким изображениям, очевидно, дает нулевое результирующее разностное изображение R . Структур-

ными различиями становятся объекты, вырезанные из второго фрагмента снимка и добавленные к исходным изображениям. При этом добавление объекта ко второму изображению эквивалентно его появлению; добавление к первому изображению — исчезновению объекта; добавление двух разных объектов к одной и той же области обоих снимков — изменению формы. Типы переносимых объектов также определяются требованиями задачи.

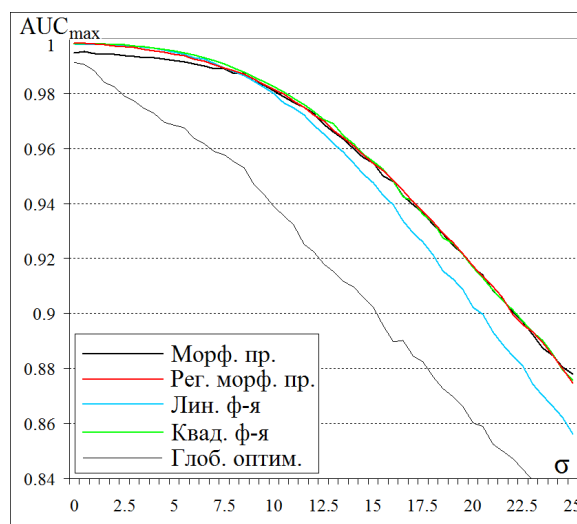


Рис. 3. Графики зависимости максимальной по всем параметрам величины площади под ROC-кривой AUC от величины стандартного отклонения шума σ для алгоритмов поиска структурных различий изображений

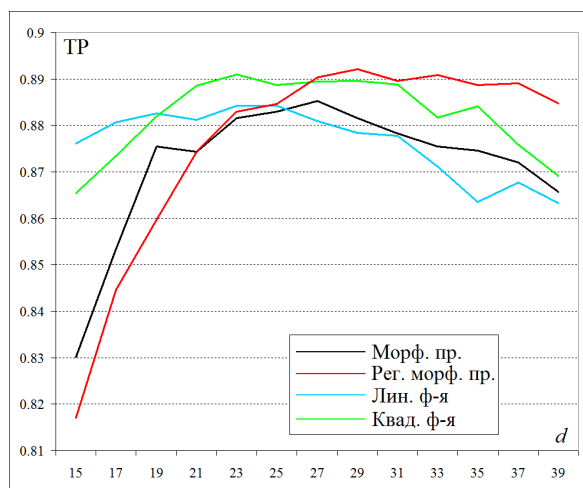


Рис. 4. Графики наибольшего уровня верного обнаружения TP алгоритмов для заданного максимально допустимого уровня ложной тревоги $FP = 0,1$ при различных значениях размера сканирующего окна d

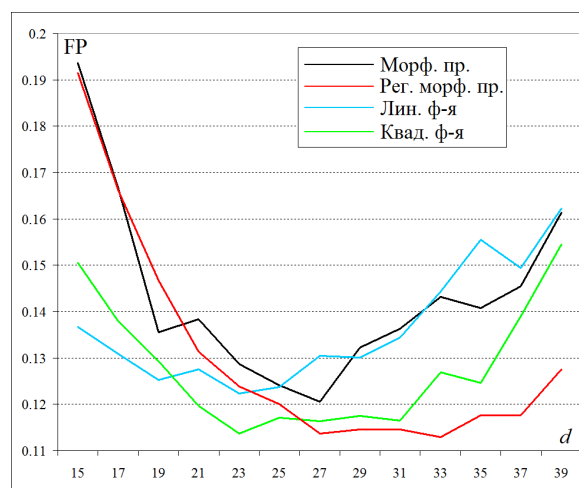


Рис. 5. Графики наименьшего уровня ложной тревоги FP алгоритмов для заданного минимально допустимого уровня верного обнаружения $TP = 0,9$ при различных значениях размера сканирующего окна d

Ввиду своей крайне высокой вычислительной сложности, численный эксперимент проводился с использованием технологии CUDA [10] и суперкомпьютера «Уран» ИММ УрО

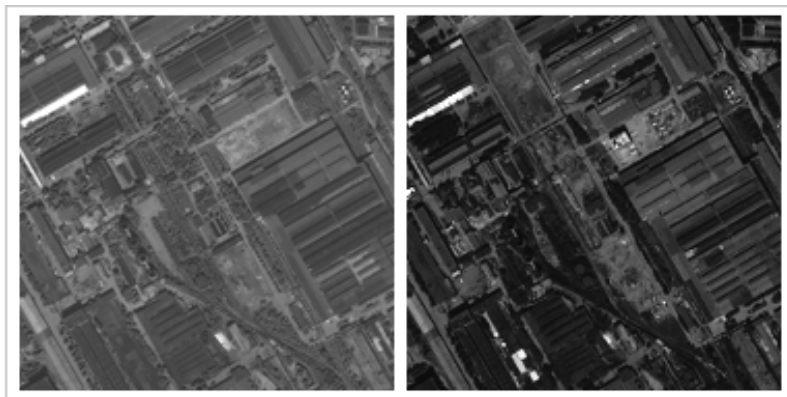


Рис. 6. Исходные полутонные снимки городской застройки, сделанные с разницей в 4 года

РАН [11]. Схемы серий испытаний для различных алгоритмов отличаются только набором исследуемых параметров, и состоят из следующих шагов:

1. Перебираются значения параметров алгоритма. Интервал варьирования параметров установлен на основе опыта работы с алгоритмом так, чтобы гарантированно содержать оптимальные значения.

2. Генерируется пара изображений (в соответствии с моделью) с добавлением шума, стандартное отклонение которого выбирается из интервала $\sigma \in [0, 25]$ с шагом 0.5.

3. К полученной паре изображений применяется алгоритм с выбранными значениями параметров. Результат работы алгоритма сравнивается с ручной разметкой для нахождения величины площади под ROC-кривой AUC. Шаги 2–3 повторяются 10 раз и полученные результаты усредняются. Максимальное значение AUC находится для каждого значения σ по всем параметрам алгоритмов, т. е. это наилучшее качество работы алгоритма при заданном уровне шума. Полученные графики изображены на рис. 3.

В реальной задаче может потребоваться знание следующих показателей: какой наилучший процент верного обнаружения (ложной тревоги) можно получить при зафиксированном максимально допустимом уровне ложной тревоги (минимально допустимом уровне верного обнаружения)? Для этого зададим допустимый уровень ложной тревоги $FP = 0.1$ и для каждого алгоритма определим наибольшее значение верного обнаружения TP . Полученные графики в зависимости от размера сканирующего окна d представлены на рис. 4. Аналогично, на рис. 5 представлены графики наименьшего уровня ложной тревоги FP алгоритмов для заданного уровня верного обнаружения $TP = 0.9$. В обоих случаях можно видеть, что до значения $d = 19$ лучшие результаты показывает алгоритм, основанный на линейной функции преобразования яркости, от $d = 21$ до $d = 25$ — алгоритм, основанный на квадратичной функции преобразования яркости, а далее — алгоритм, основанный на регуляризованном морфологическом проекторе, что подтверждает представленные на рис. 3 результаты. В табл. 1 представлены максимальные значения верного обнаружения TP (ложной тревоги FP) для всех алгоритмов и оптимальные параметры, при которых они получены.

Наконец, на рис. 6 изображены два фрагмента реальной пары космоснимков городской застройки, а на рис. 7 — фрагменты созданного вручную эталона (ручной разметки присутствующих на снимках структурных различий) и результатов работы алгоритмов. Черным цветом отмечены верно классифицированные положительные примеры, белым — верно классифицированные отрицательные случаи, красным — пропуск (ошибка I рода),

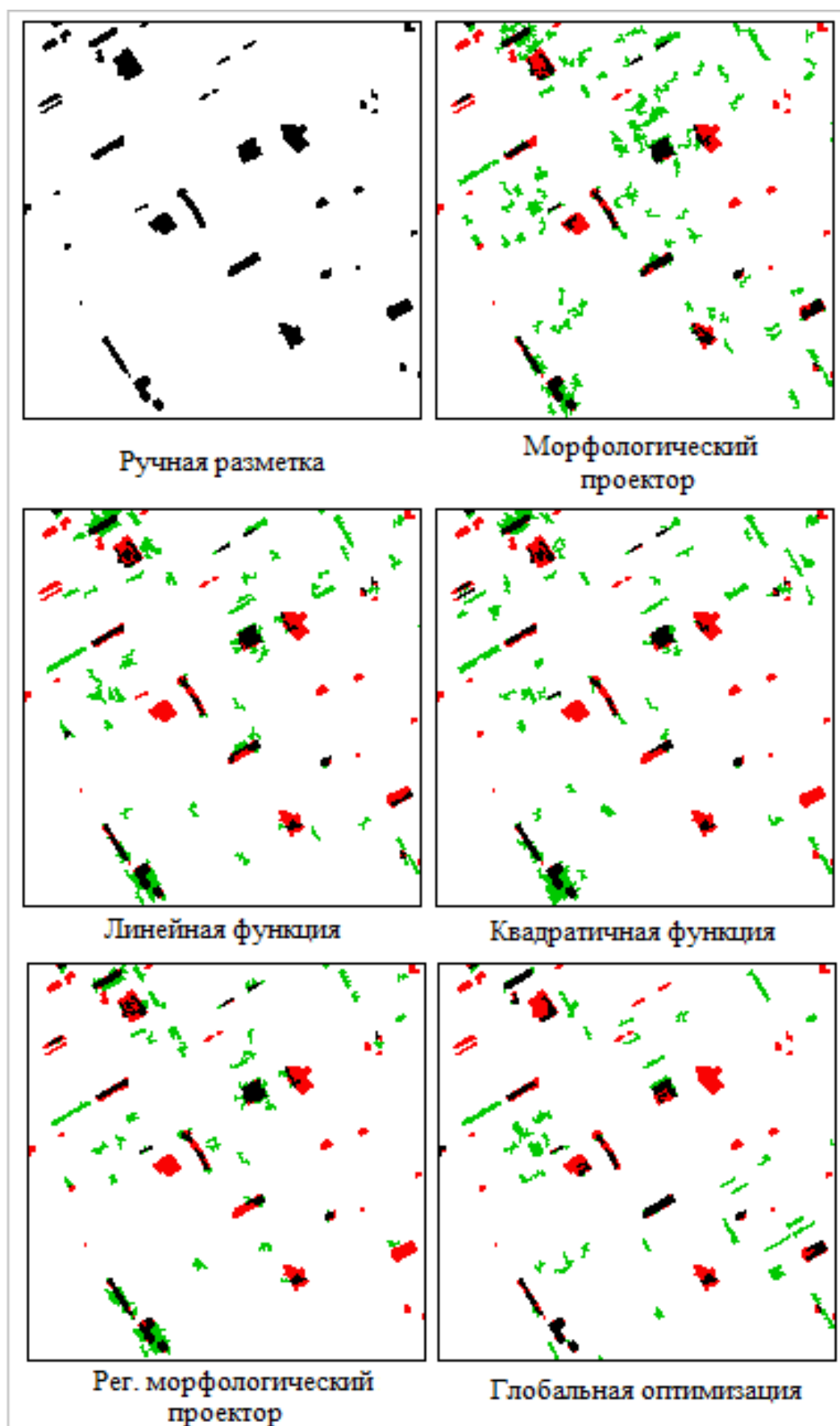


Рис. 7. Ручная разметка и результаты работы алгоритмов поиска структурных различий с оптимальными параметрами: черным цветом отмечено верное обнаружение, белым — верно классифицированные отрицательные случаи, красным — пропуск, зеленым — ложная тревога

зеленым — ложная тревога (ошибка II рода). Результаты таковы: морфологический проектор достаточно неплохо нашел имеющиеся структурные различия, но и выдал больше всех ложных срабатываний; линейная и квадратичная функции оказались близки — среднее количество найденных различий и умеренное ложных тревог; регуляризованный морфологический проектор нашел практически те же различия, что и предыдущие алгоритмы, но зато выдал наименьшее количество ложных срабатываний; глобальная оптимизация показала наилучший результат по обнаружению и локализации различий, но уровень ложных тревог оказался чуть выше среднего.

Заключение

В данной работе был подведен итог серии работ, посвященных задаче поиска структурных различий изображений, а также предложены вариант морфологического проектора для цветных изображений и методика исследования алгоритмов с целью определения оптимальных параметров работы. В последней была сформулирована особая математическая модель, подходящая для проведения численного эксперимента, максимально приближенного к реальным задачам. На основе всех полученных данных был разработан программный комплекс, решающий задачу поиска структурных различий изображений. Данный комплекс был интегрирован в среду визуализации и обработки данных дистанционного зондирования Земли ENVI, и внедрен в работу ОАО «НИИП центр Природа».

Резюмируя, можно отметить, что в ходе исследовательской работы удалось получить удовлетворительное решение поставленной задачи с помощью различных алгоритмов, но в тоже время на основе единого теоретического подхода. Можно сформулировать следующие направления дальнейшей работы.

1. Получить теоретические формулы для оптимального порога алгоритмов поиска структурных различий. В настоящий момент такая формула есть лишь для алгоритма, основанного на морфологическом проекторе. Определив эти формулы для остальных алгоритмов, можно будет сформулировать утверждение о теоретическом преимуществе одного из них. Эмпирическое преимущество осталось за алгоритмом, основанном на регуляризованном морфологическом проекторе.

2. Использование глобальной оптимизации дает очень неплохой и перспективный результат, поэтому важным направлением является усовершенствование данного метода и его дальнейшее применение в прикладных задачах.

3. Разработать версии алгоритмов, работающие с многоканальными изображениями, а также использующими информацию о тестурах и границах поверхностей. Также стоит использовать также данные различных геофизических полей, например, поле высот строений и сооружений, которое можно получить, используя стереоснимки.

Литература

- [1] Kostousov V. B., Kandoba I. N., Skripnuk V. V., Shabanov G. A. The system for automated deciphering of cosmic Earth surface photographs // *Int. Archives Photogrammetry Remote Sensing*. Amsterdam, 2000. Vol. XXXIII. Part B4. P. 425–432.
- [2] Корнилов Ф. А., Перевалов Д. С. Задача обнаружения структурных различий изображений // *Алгоритмы и программ. средства парал. вычислений: Сб. науч. тр.* — Екатеринбург: УрО РАН, 2011. Вып. 11. С. 37–56.
- [3] Корнилов Ф. А. Исследование алгоритма поиска структурных различий изображений // *Тр. Ин-та математики и механики УрО РАН*. — Екатеринбург: ИММ УрО РАН, 2013. Т. 19. № 1. С. 144–159.

- [4] Li W., Li X., Wu Y., Hu Z. A novel framework for urban change detection using VHR satellite images // IEEE The 18th Conference (International) on Pattern Recognition (ICPR'06). — Hong Kong, 2006. P. 312–315.
- [5] Li L., Leung M. K. H. Integrating intensity and texture differences for robust change detection // IEEE Trans. Image Processing, February 2002. Vol. 11. No. 2. P. 105–112.
- [6] Пытьев Ю. П., Чуличков А. И. Методы морфологического анализа изображений. — М.: ФизМатЛит, 2010. 336 с.
- [7] Li S. Z. Markov random field modeling in image analysis. — 3rd ed. — Berlin–Heidelberg: Springer, 2009. 362 p.
- [8] Винклер Г. Анализ изображений, случайные поля и динамические методы Монте-Карло. Математические основы. — Новосибирск: Изд-во СО РАН, филиал «Гео», 2002. 343 с.
- [9] Green D. M., Swets J. M. Signal detection theory and psychophysics. — New York: John Wiley and Sons Ltd., 1966. 455 p.
- [10] Боресков А. В. и др. Параллельные вычисления на GPU. Архитектурная и программная модель CUDA. — М.: Изд-во Моск. ун-та. Сер. «Суперкомпьютерное образование», 2012. 336 с.
- [11] Суперкомпьютерный центр «Уран» ИММ УрО РАН. <http://parallel.uran.ru/>.

References

- [1] Kostousov V. B., Kandoba I. N., Skripnuk V. V., Shabanov G. A. 2000. The system for automated deciphering of cosmic Earth surface photographs. *Int. Archives Photogrammetry Remote Sensing*. Amsterdam. XXXIII(part B4):425–432.
- [2] Kornilov F. A., Perevalov D. S. 2011. The problem of detecting structural differences on the images. *Algorithms Software Parallel Computing*. Ekaterinburg: IMM UB RAS. 11:37–56.
- [3] Kornilov F. A. 2013. Study algorithm detection structural differences on the images. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*. Ekaterinburg: IMM UB RAS. 19(1):144–159.
- [4] Li W., Li X., Wu Y., Hu Z. 2006. A novel framework for urban change detection using VHR satellite images. *IEEE The 18th Conference (International) on Pattern Recognition (ICPR'06)*. Hong Kong. 312–315.
- [5] Li L., Leung M. K. H. 2002. Integrating intensity and texture differences for robust change detection. *IEEE Trans. Image Processing* 11(2):105–112.
- [6] Pitev Y. P., Chulichkov A. I. 2010. *The methods of morphological analysis of images*. Moscow: FizMatLit. 336 p.
- [7] Li S. Z. 2009. *Markov random field modeling in image analysis*. 3rd ed. Berlin–Heidelberg: Springer. 362 p.
- [8] Wincler G. 2002. *Image analysis, random fields and markov chain Monte Carlo methods: A mathematical introduction*. Novosibirsk: SB RAS press, “Geo.” 343 p.
- [9] Green D. M., Swets J. M. 1966. *Signal detection theory and psychophysics*. New York: John Wiley and Sons Ltd. 455 p.
- [10] Boreskov A. V. , et al. 2012. *Parallel computing with GPU. Architecture and programming model CUDA*. Moscow: Moscow University Press. Ser. Supercomputer education. 336 p.
- [11] Supercomputer center “Uran” IMM UB RAS. Available at: <http://parallel.uran.ru/>.

Симметризация точек изображения, заданных статистическими выборками*

А. Н. Каркищенко¹, В. Б. Мнухин²

¹karkishalex@gmail.com; ²mnukhin.valeriy@mail.ru

Южный Федеральный Университет, Россия, г. Ростов-на-Дону 344006,
ул. Большая Садовая, 105/42

Работа посвящена построению отражательно и вращательно симметричных конфигураций точек изображения по заданным, но не обладающим этим свойством точкам, про которые известно, что они должны быть симметричны. Данная задача называется задачей симметризации и находит многочисленные применения. В предыдущих работах авторов рассматривались методы ее решения при условии, что каждая точка задана некоторой оценкой своего положения, которое может не совпадать с истинным. Были предложены решения для случаев отражательной и вращательной симметрии с разной степенью априорной неопределенности. Оптимальность получаемых решений выражалась в том, что симметризация достигалась минимальным отклонением «симметризованных» точек от первоначальных положений по евклидовой метрике. В отличие от предшествующих публикаций в данной работе мы рассматриваем более общий случай, когда точки задаются статистическими выборками координат. При этом требуется, соответственно, «статистическое» понимание оптимальности решений. Предлагаются методы построения по статистическим выборкам гарантированно симметричных конфигураций, оптимальных в том смысле, что они обладают наибольшей плотностью вероятности появления. Показано, что эти задачи сводятся к стандартной постановке задачи минимизации в пространстве с метрикой Махаланобиса. Предлагается также нормированная мера симметричности исходных статистических данных. Для иллюстрации методов симметризации приводятся результаты моделирования.

Ключевые слова: симметризация; отражательная симметрия; вращательная симметрия; характерные точки; мера симметричности

Symmetrization of the image points defined by statistical sampling*

A. N. Karkishchenko, V. B. Mnukhin

Southern Federal University, 105/42 Bolshaya Sadovaya Str., Rostov-on-Don 344006, Russia

The paper is devoted to the methods of constructing reflectionally and rotationally symmetrical configurations of characteristic points in an image calculated from a certain set of given points that do not possess this property. The problem has many applications and is called the problem of symmetrization. In their previous studies, the authors examined the methods for its solution provided that each point analyzed is given with some estimate of its position, which may not coincide with the true one. The solutions have been proposed for the cases of reflectional and rotational symmetry with varying degrees of *a priori* uncertainty. Optimality of the solutions is in the fact that symmetrization is achieved in a minimum Euclidean metric deviation of “symmetrized” points from the original locations.

Работа выполнена при поддержке РФФИ, проекты № 13-07-00327-а и № 13-07-13112-офи_м_РЖД.

In contrast to previous publications, in this paper, the more general case is considered when the points are specified by statistical samples of coordinates of their possible position. It requires, respectively, the “statistical” understanding of the optimality of solutions. Therefore, the methods proposed in this paper, being based on statistical samples, are available for constructing symmetrical configurations of points which are optimal in the sense that they have the highest probability density of appearance. The paper also shows that this problem can be reduced to the standard setting of the minimization problem in a space with Mahalanobis metric. Also, it is proposed normalized (from zero to one) measure to assess the degree of symmetry of the source statistics, which naturally follows from the proposed methods. To illustrate the methods, the results of modeling are given in the article.

Keywords: symmetrization; reflectional symmetry; rotational symmetry; feature points; symmetry measure

Введение

Симметрия играет значительную роль в правильном восприятии как естественных, так и в большей степени искусственно созданных объектов. Практически все объекты в той или иной степени симметричны, либо имеют симметричные части. В последнее время многочисленные усилия при анализе формы объектов, заданных в оцифрованном виде, были сосредоточены на выявлении симметрии 2D и 3D объектов [9, 6, 7]. Информация о симметрии эффективно используется в многочисленных приложениях — в компактном описании моделей [8], обработке сканированных изображений [11], сегментации изображений [10], установлении соответствия форм [9] и др. Как правило, на «атомарном» уровне методы сводятся к исследованию симметричности так называемых характерных точек, смысл которых зависит от решаемой задачи.

Симметризация может иметь многочисленные приложения. Например, при конструировании формы симметричных объектов исходные данные при проектировании могут не отвечать требованиям симметрии, заложенным в конструкторской документации или присутствующим в некоторой идеальной модели. Ручная корректировка данных с целью приведения их к «симметричному» виду очень затратна по времени, громоздка по исполнению и практически нереализуема для крупномасштабных моделей. Эта проблема оказывается еще более сложной для моделей, в которых симметрия проявляется лишь в определенных положениях или позициях.

Многие задачи могут получить более простое решение в случае явно симметричных изображений. Так, качество автоматического распознавания изображений может быть улучшено благодаря предварительной симметризации повторяющихся элементов рассматриваемого объекта. Симметризация может также привести к снижению уровня шума и повысить производительность алгоритмов сжатия изображений. В общем случае можно констатировать, что хорошо симметризованный объект легче воспринимается, классифицируется и понимается как человеком, так и машинными алгоритмами.

Суть проблемы симметризации состоит в том, что локализация точек на изображении всегда осуществляется с некоторой погрешностью, величина которой зависит от многих факторов — используемого алгоритма локализации, качества изображения в целом, зашумления в области, содержащей локализуемую точку и др. Как следствие, координаты обнаруженных точек, про которые априори известно, что они симметричны, в реальности этому условию не удовлетворяют. Поэтому целесообразно воспользоваться известной априорной информацией о симметричности точек для уточнения их положения. При этом

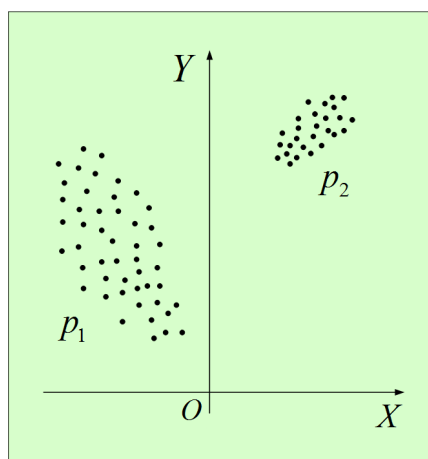


Рис. 1. Статистические выборки для парных точек

само уточнение необходимо осуществлять в некотором смысле наилучшим образом, скажем, симметричность должна достигаться минимальным изменением их положения.

В работах [5, 1] рассмотрено построение симметричных точек в случаях отражательной симметрии, когда каждая точка задается единственным образом. В работе [2] рассмотрено использование симметризации характерных точек для более точной локализации точек, описывающих лицо человека. Задача связана с тем, что точность детекции и распознавания лиц существенно зависит от точности определения центров зрачков на лице, положение которых отвечает естественной симметрии. Это же относится и к другим точкам, на основе которых осуществляется сегментация изображения лица.

Более сложная задача, связанная с искусственной симметризацией характерных точек, возникает в том случае, если сами точки задаются не одной локализованной точкой, а некоторой выборкой. Такая выборка может быть получена при многократном применении одного или разных алгоритмов локализации при различных начальных условиях — различных по размерам и форме областей, содержащих характерную точку, разных сочетаниях параметров алгоритма, различающихся в мелких деталях изображений (например, последовательность кадров видеоряда) и др. Можно считать, что оценка математического ожидания такой выборки будет существенно точнее определять истинное положение характерной точки. Симметризация точек в этом случае также может быть осуществлена точнее, но при этом целесообразно учитывать не только оценки математических ожиданий, но также дисперсии и корреляционные моменты координат характерных точек. Ниже предлагается метод, который позволяет решать указанную задачу.

Симметризация отражательно симметричных точек, заданных статистическими выборками

Пусть заданы отражательно симметричные точки с осью симметрии, совпадающей для простоты с осью Oy . Будем считать, что положение каждой точки оценивается статистической выборкой ее возможных положений (рис. 1), так что правдоподобный вывод об истинном положении этой точки может быть сделан на основе анализа такой выборочной совокупности. В простейшем случае за положение точки можно взять оценку математического ожидания, вычисленную по выборке. Однако при этом, очевидно, соответствующие точки не будут располагаться симметрично. Применение к математическим ожиданиям

способа симметризации, описанного, например, в [1], также нельзя рассматривать в данном случае как приемлемое, поскольку не учитываются дисперсии случайных координат и их корреляции. Рассмотрим более точное решение этой задачи, не имеющее указанных недостатков. Для простоты будем считать вначале, что имеются две точки, которые априори должны быть симметричны относительно оси Oy и не лежат на ней. Обозначим их $p_1(x_1, y_1)$ и $p_2(x_2, y_2)$. Эти точки заданы выборками: $\{p_1^1, p_1^2, p_1^3, \dots\}$, $\{p_2^1, p_2^2, p_2^3, \dots\}$. Задача состоит в том, чтобы найти оценки положения характерных точек, так чтобы эти положения были наиболее вероятны (с точки зрения выборок), и при этом они удовлетворяли условию симметричности. Дадим формальную постановку задачи.

Пусть $f_1(x_1, y_1)$ — плотность вероятности появления точки $p_1(x_1, y_1)$, а $f_2(x_2, y_2)$ — плотность вероятности для $p_2(x_2, y_2)$, представленные статистическими выборками. Обозначим через $\mathbf{Z} = (x_1 \ y_1 \ x_2 \ y_2)^T$ случайный вектор, состоящий из координат точек p_1 и p_2 . Естественно считать, что случайные векторы $(x_1 \ y_1)$ и $(x_2 \ y_2)$ независимы, поэтому $\mathbf{Z}$ подчиняется вероятностному распределению $f(\mathbf{Z}) = f_1(x_1, y_1)f_2(x_2, y_2)$. Искомый вектор, описывающий наиболее вероятное положение пары симметричных точек, должен, очевидно, удовлетворять условиям: $-x_1 = x_2$, $y_1 = y_2$. Нетрудно заметить, что множество векторов, координаты которых обладают этим свойством, образует «симметричное» подпространство в линейном пространстве R^4 . Обозначим его R_{Sym}^4 . Таким образом, R_{Sym}^4 состоит из векторов вида $(x \ y \ -x \ y)^T$, $x, y \in R$, соответствующих характерным точкам, расположенным симметрично относительно оси ординат. Тогда задача отыскания наиболее вероятного положения симметричных точек формально запишется так:

$$\begin{cases} f(\mathbf{Z}) \rightarrow \max, \\ \mathbf{Z} \in R_{Sym}^4. \end{cases} \quad (1)$$

Как показывают эксперименты [3], выборочные совокупности во многих случаях хорошо аппроксимируются нормальными распределениями:

$$f_1(\mathbf{z}_1) = \frac{1}{2\pi\sqrt{|C_1|}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_1 - \mathbf{m}_1)^T C_1^{-1}(\mathbf{z}_1 - \mathbf{m}_1) \right\},$$

$$f_2(\mathbf{z}_2) = \frac{1}{2\pi\sqrt{|C_2|}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_2 - \mathbf{m}_2)^T C_2^{-1}(\mathbf{z}_2 - \mathbf{m}_2) \right\},$$

где $\mathbf{z}_1 = (x_1 \ y_1)^T$, $\mathbf{z}_2 = (x_2 \ y_2)^T$, а $\mathbf{m}_1 = (m_{x_1} \ m_{y_1})^T$, $\mathbf{m}_2 = (m_{x_2} \ m_{y_2})^T$, C_1 и C_2 — математические ожидания и ковариационные матрицы, которые статистически оцениваются по соответствующим выборкам. Введем для удобства матрицу $S = \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}$. Тогда условие принадлежности вектора $\mathbf{Z}$ подпространству R_{Sym}^4 можно записать следующим образом: $\mathbf{z}_1 = \mathbf{z}$, $\mathbf{z}_2 = S\mathbf{z}$ или $\mathbf{Z} = \begin{pmatrix} \mathbf{z} \\ S\mathbf{z} \end{pmatrix}$, где $\mathbf{z} = (x \ y)^T \in R^2$. Задача (1) в этом случае принимает вид:

$$f(\mathbf{z}) = f_1(\mathbf{z})f_2(S\mathbf{z}) =$$

$$= \frac{1}{4\pi^2\sqrt{|C_1| \cdot |C_2|}} \exp \left\{ -\frac{1}{2} \left[(\mathbf{z} - \mathbf{m}_1)^T C_1^{-1}(\mathbf{z} - \mathbf{m}_1) + (S\mathbf{z} - \mathbf{m}_2)^T C_2^{-1}(S\mathbf{z} - \mathbf{m}_2) \right] \right\} \rightarrow \max$$

или, что эквивалентно,

$$(\mathbf{z} - \mathbf{m}_1)^T C_1^{-1}(\mathbf{z} - \mathbf{m}_1) + (S\mathbf{z} - \mathbf{m}_2)^T C_2^{-1}(S\mathbf{z} - \mathbf{m}_2) \rightarrow \min.$$

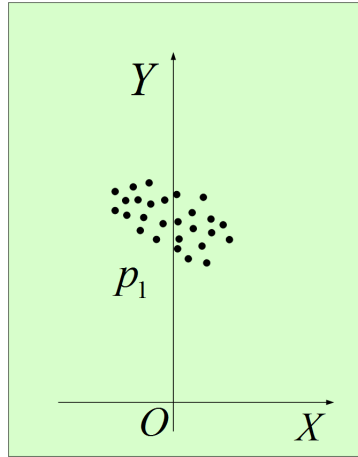


Рис. 2. Статистическая выборка для непарной точки

Решение данной задачи приводит к следующей системе уравнений:

$$(C_1^{-1} + SC_2^{-1}S) \mathbf{z} = C_1^{-1}\mathbf{m}_1 + SC_2^{-1}\mathbf{m}_2. \quad (2)$$

Если $\mathbf{z}^* = (x^* \ y^*)^T$ — решение данной системы, то, следовательно, координаты симметризованных точек будут иметь вид: $p_1^*(x^*, y^*)$ и $p_2^*(-x^*, y^*)$.

Для удобства вычислений отметим, что ковариационные матрицы и матрицы, обратные к ним, имеют вид:

$$C_k = \begin{pmatrix} \sigma_{x_k}^2 & \rho_k \sigma_{x_k} \sigma_{y_k} \\ \rho_k \sigma_{x_k} \sigma_{y_k} & \sigma_{y_k}^2 \end{pmatrix}, \quad C_k^{-1} = \begin{pmatrix} \frac{1}{(1-\rho_k^2)\sigma_{x_k}^2} & -\frac{\rho_k}{(1-\rho_k^2)\sigma_{x_k}\sigma_{y_k}} \\ -\frac{\rho_k}{(1-\rho_k^2)\sigma_{x_k}\sigma_{y_k}} & \frac{1}{(1-\rho_k^2)\sigma_{y_k}^2} \end{pmatrix}, \quad k = 1, 2,$$

где $\sigma_{x_k}, \sigma_{y_k}$ — среднеквадратические отклонения, ρ_k — коэффициенты корреляции, а система (2) в явной форме выглядит следующим образом:

$$\left\{ \begin{aligned} & \left(\frac{1}{(1-\rho_1^2)\sigma_{x_1}^2} + \frac{1}{(1-\rho_2^2)\sigma_{x_2}^2} \right) x + \left(-\frac{\rho_1}{(1-\rho_1^2)\sigma_{x_1}\sigma_{y_1}} + \frac{\rho_2}{(1-\rho_2^2)\sigma_{x_2}\sigma_{y_2}} \right) y = \\ & \quad = \frac{m_{x_1}}{(1-\rho_1^2)\sigma_{x_1}^2} - \frac{\rho_1 m_{y_1}}{(1-\rho_1^2)\sigma_{x_1}\sigma_{y_1}} - \frac{m_{x_2}}{(1-\rho_2^2)\sigma_{x_2}^2} + \frac{\rho_2 m_{y_2}}{(1-\rho_2^2)\sigma_{x_2}\sigma_{y_2}}, \\ & \left(-\frac{\rho_1}{(1-\rho_1^2)\sigma_{x_1}\sigma_{y_1}} + \frac{\rho_2}{(1-\rho_2^2)\sigma_{x_2}\sigma_{y_2}} \right) x + \left(\frac{1}{(1-\rho_1^2)\sigma_{y_1}^2} + \frac{1}{(1-\rho_2^2)\sigma_{y_2}^2} \right) y = \\ & \quad = -\frac{\rho_1 m_{x_1}}{(1-\rho_1^2)\sigma_{x_1}\sigma_{y_1}} + \frac{m_{y_1}}{(1-\rho_1^2)\sigma_{y_1}^2} - \frac{\rho_2 m_{x_2}}{(1-\rho_2^2)\sigma_{x_2}\sigma_{y_2}} + \frac{m_{y_2}}{(1-\rho_2^2)\sigma_{y_2}^2}. \end{aligned} \right.$$

Заметим, что поскольку C_1^{-1} и $SC_2^{-1}S$ — положительно определенные матрицы, то положительно определенной является и главная матрица системы уравнений (2), а поэтому она невырождена и, следовательно, система (2) имеет единственное решение. Однако при некоторых выборках эта матрица может оказаться плохо обусловленной, и значит, решение может содержать значительные погрешности. Общий характер зависимости координат симметричных точек от величин математических ожиданий и дисперсий становится более

понятным, если рассмотреть частный случай некоррелированных случайных величин (эллипсоиды рассеяния точек ориентированы по координатным осям). Полагая $\rho_1 = \rho_2 = 0$, из системы уравнений, представленной в явном виде, получаем:

$$\begin{cases} x = -\frac{\sigma_{x_2}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} m_{x_1} + \frac{\sigma_{x_1}^2}{\sigma_{x_1}^2 + \sigma_{x_2}^2} m_{x_2}, \\ y = \frac{\sigma_{y_2}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} m_{y_1} + \frac{\sigma_{y_1}^2}{\sigma_{y_1}^2 + \sigma_{y_2}^2} m_{y_2}. \end{cases}$$

Из данных выражений следует, что $x \in [-m_{x_1}, m_{x_2}]$ и $y \in [m_{y_1}, m_{y_2}]$, причем каждая координата расположена ближе к математическому ожиданию того распределения, у которого соответствующая дисперсия меньше. Данная зависимость очевидна, поскольку чем меньше дисперсия, тем более точно характеризуется область, содержащая выборочные значения. Поэтому оценки математических ожиданий такого распределения вызывают большее «доверие». В частности, если соответствующие дисперсии обоих распределений равны, то $x = (-m_{x_1} + m_{x_2})/2$ и $y = (m_{y_1} + m_{y_2})/2$.

Если на изображении заданы n пар симметричных точек, то описанная процедура симметризации может быть независимо применена последовательно к каждой паре.

Рассмотрим теперь симметризацию «непарных» характерных точек, лежащих на оси Oy подобно тому, как это показано на рис. 2. Допустим, что статистическая выборка, задающая такую точку, описывается плотностью

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\frac{(x-m_x)^2}{\sigma_x^2} - \frac{2\rho(x-m_x)(y-m_y)}{\sigma_x\sigma_y} + \frac{(y-m_y)^2}{\sigma_y^2} \right] \right\}.$$

Для решения достаточно найти моду условной плотности при $x = 0$. Полагая $x = 0$, найдем точку y , в которой плотность достигает максимального значения. Дифференцируя по y и приравнявая производную нулю, после преобразований находим: $y = m_y - \frac{\sigma_y}{\sigma_x} \rho m_x$. Таким образом, для «непарных» точек соответствующая симметризованная точка имеет координаты $(0, m_y - \frac{\sigma_y}{\sigma_x} \rho m_x)$.

На рис. 3 представлены результаты построения симметризованных точек на модельных примерах. На левом графике симметризация производилась по выборкам, содержащим по 5 экспериментально полученных точек, в то время как на правом - по 20 точек. В том и другом случае истинные точки имеют координаты $p_1(-1, 1)$ и $p_1(1, 1)$. Точки, найденные по выборке объема 5, — $p_1^*(-0.9752, 0.96203)$ и $p_2^*(0.9752, 0.96203)$, а по выборке объема 20, — $p_1^*(-1.0051, 1.0064)$ и $p_2^*(1.0051, 1.0064)$.

Связь с расстоянием Махаланобиса и мера симметричности

Покажем, что задача «симметризации» статистически заданных нормально распределенных характерных точек эквивалентна задаче отыскания проекции вектора на подпространство в линейном пространстве с метрикой Махаланобиса.

Обозначим, как и раньше, $\mathbf{z}_k = (x_k \ y_k)^T$, $\mathbf{m}_k = (m_{x_k} \ m_{y_k})^T$, $k = 1, 2$. В силу сделанных выше предположений положение точек на плоскости описывается функциями плотности вероятности

$$f_k(\mathbf{z}_k) = \frac{1}{2\pi\sqrt{|C_k|}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_k - \mathbf{m}_k)^T C_k^{-1}(\mathbf{z}_k - \mathbf{m}_k) \right\}, \quad k = 1, 2,$$

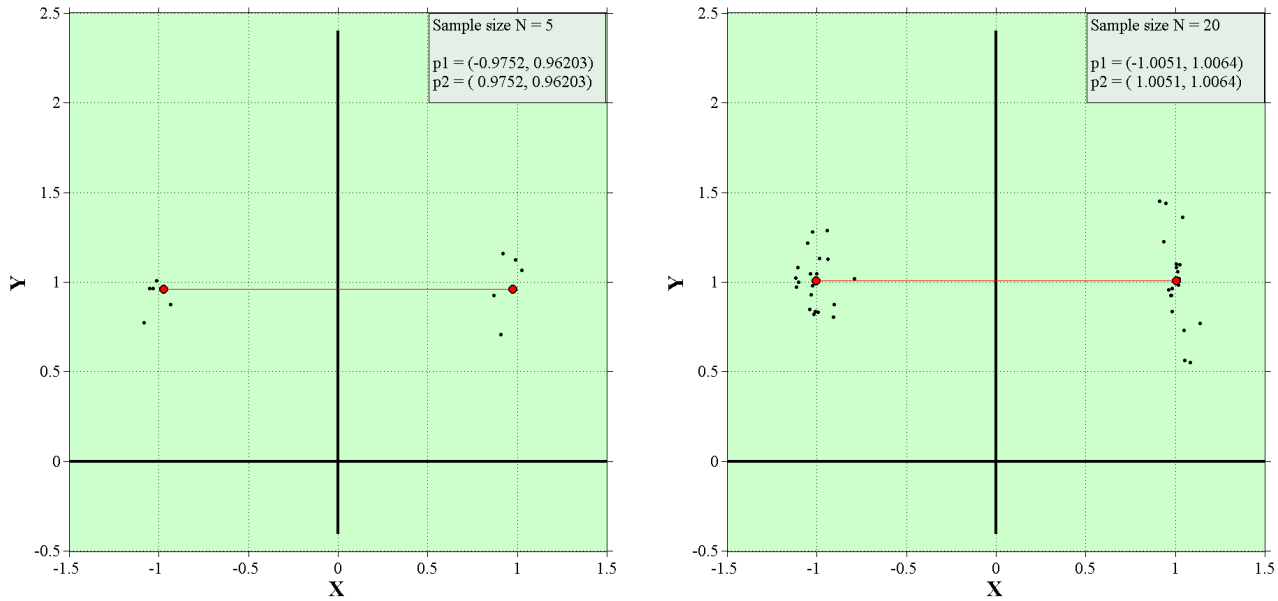


Рис. 3. Отражательная симметризация, основанная на статистических выборках точек

с ковариационными матрицами C_1 и C_2 . Если обозначить

$$\mathbf{Z} = (\mathbf{z}_1^T \ \mathbf{z}_2^T)^T = (x_1 \ y_1 \ x_2 \ y_2)^T, \quad \mathbf{m} = (m_{x_1} \ m_{y_1} \ m_{x_2} \ m_{y_2})^T, \quad C = \begin{pmatrix} C_1 & \mathbf{0} \\ \mathbf{0} & C_2 \end{pmatrix},$$

то совместное распределение пары точек p_1 и p_2 можно записать в компактном виде:

$$f(\mathbf{Z}) = f_1(\mathbf{z}_1) f_2(\mathbf{z}_2) = \frac{1}{4\pi^2 \sqrt{|C|}} \exp \left\{ -\frac{1}{2} (\mathbf{Z} - \mathbf{m})^T C^{-1} (\mathbf{Z} - \mathbf{m}) \right\}.$$

Тогда задача (1) эквивалентна следующей экстремальной задаче:

$$\begin{cases} (\mathbf{Z} - \mathbf{m})^T C^{-1} (\mathbf{Z} - \mathbf{m}) \rightarrow \min, \\ \mathbf{Z} \in R_{Sym}^4, \end{cases}$$

где R_{Sym}^4 - введенное выше подпространство векторов, состоящее из точек, расположенных симметрично относительно оси Oy .

Минимизируемая функция в последнем выражении представляет собой квадрат расстояния Махаланобиса от точки $\mathbf{Z}$ до точки $\mathbf{m}$, поэтому поиск минимума сводится к задаче вычисления в пространстве с метрикой Махаланобиса проекции вектора $\mathbf{m}$ на подпространство R_{Sym}^4 , т.е.

$$\mathbf{X}_s = \arg \min_{\mathbf{Z} \in R_{Sym}^4} \|\mathbf{Z} - \mathbf{m}\|_{C^{-1}}. \quad (3)$$

При $C^{-1} = I$ данная формула описывает задачу симметризации в обычном (не статистическом) смысле, т.е. когда обе точки имеют единственные оценки своего положения, задаваемые вектором $\mathbf{m}$.

Аналогичную форму можно придать и задаче симметризации нормально распределенных характерных точек, лежащих на оси Oy ; в этом случае $\mathbf{Z} = (x \ y)^T$, а $\mathbf{m} = (m_x \ m_y)^T$, а в качестве «симметричного» подпространства выступает подпространство R_{Sym}^2 векторов вида $(0 \ y)^T$, $y \in R$.

Опираясь на данную выше интерпретацию симметризации точек как решения экстремальной задачи, можно достаточно естественным образом ввести меру первоначальной «симметричности» выборок, задающих симметризуемые точки. Рассмотрим пространство R^4 со скалярным произведением $(\mathbf{x}, \mathbf{x})_G = (\mathbf{x}, G\mathbf{x})$, где G — положительно определенный самосопряженный оператор. Если P — оператор проектирования на некоторое подпространство $S \subseteq R^4$, то норму проекции некоторого вектора $\mathbf{x}$ на S , т.е. $\|P\mathbf{x}\|_G$, можно рассматривать как меру близости $\mathbf{x}$ к S . Чем больше норма, тем ближе $\mathbf{x}$ к S . При этом для количественного измерения близости более удобно использовать нормированную величину

$$\mu(\mathbf{x}, S) = \frac{\|P\mathbf{x}\|_G}{\|\mathbf{x}\|_G} = \|P\mathbf{x}^0\|_G,$$

где $\mathbf{x}^0$ — орт вектора $\mathbf{x}$. Очевидно, $0 \leq \mu(\mathbf{x}, S) \leq 1$. Понятно, что равенство $\mu(\mathbf{x}, S) = 1$ соответствует случаю, когда $\mathbf{x} \in S$.

В нашем случае, учитывая, что $\dim R_{Sym}^4 = 2$, условие $\mathbf{Z} \in R_{Sym}^4$ можно записать в виде $\mathbf{Z} = A\mathbf{u}$, где $A: R^2 \rightarrow R_{Sym}^4$ — оператор, матрица которого имеет вид

$$A = \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}^T.$$

Следовательно, задача (3) может быть записана в следующей эквивалентной форме:

$$\|\mathbf{Z} - \mathbf{m}\|_{C^{-1}}^2 = \|A\mathbf{u} - \mathbf{m}\|_{C^{-1}}^2 = (A\mathbf{u} - \mathbf{m}, C^{-1}(A\mathbf{u} - \mathbf{m})) \xrightarrow{\mathbf{u} \in R^2} \min.$$

Ее решение имеет вид: $\mathbf{Z} = A(A^T C^{-1} A)^{-1} A^T C^{-1} \mathbf{m}$. Заметим, что $(A^T C^{-1} A)^{-1}$ существует, поскольку все входящие в произведение операторы имеют полный ранг. Оператор $A(A^T C^{-1} A)^{-1} A^T C^{-1}$ является оператором ортогонального проектирования на подпространство R_{Sym}^4 со скалярным произведением $(\mathbf{x}, \mathbf{x})_{C^{-1}} = (\mathbf{x}, C^{-1}\mathbf{x})$. Следовательно, величину

$$\mu(\mathbf{m}, R_{Sym}^4) = \frac{\|\mathbf{Z}\|_{C^{-1}}}{\|\mathbf{m}\|_{C^{-1}}}$$

можно рассматривать как нормированную меру симметричности исходных точек, задаваемых как среднее по представляющим их выборкам.

Анализ характера изменения меры μ в зависимости от размера выборок позволяет делать некоторые качественные выводы. Простейшие из них состоят в следующем. Если процедура отыскания точек не имеет систематических погрешностей, то можно ожидать, что с увеличением объема выборок математические ожидания будут стремиться к истинным координатам симметричных точек, а следовательно, μ будет сходиться по вероятности к единице. Если же этого не происходит, то либо алгоритм, определяющий положение точек, т.е. формирующий выборки, дает систематические погрешности, либо рассматриваемые точки вопреки предположению не являются симметричными. Систематические ошибки могут возникать, например, в том случае, если формирование исходных условий для определения положения точек зависит от определения положения ряда предыдущих точек. В этом случае помимо инструментальных ошибок, связанных с локализацией точек на зашумленном изображении, происходит последовательное накопление ошибок, вызванных неточностями в локализации предшествующих точек.

Графики на рис. 4 на модельных данных показывают изменение меры симметричности в зависимости от объема выборки при симметризации двух точек. При этом верхний

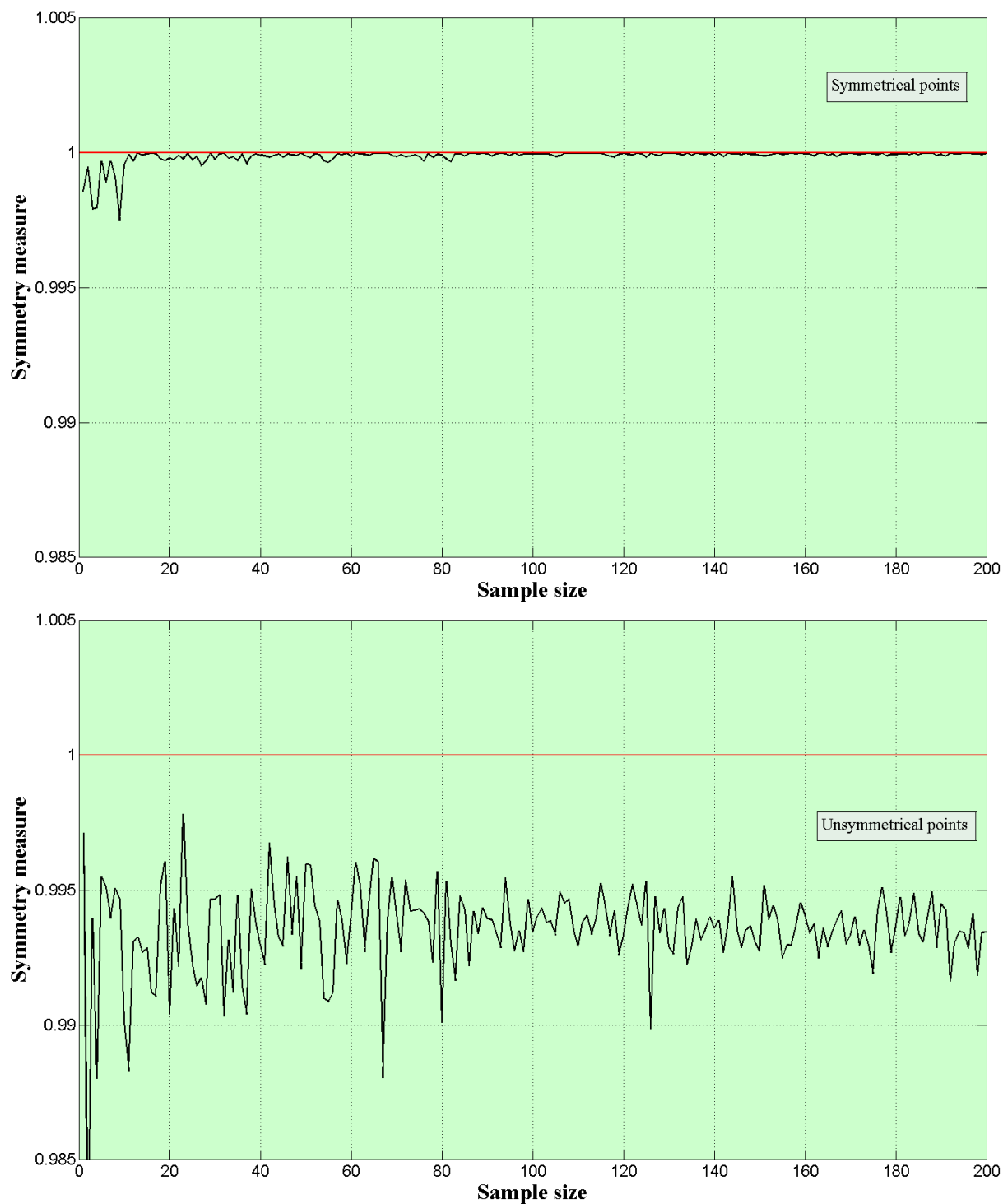


Рис. 4. Сравнение зависимости меры симметричности от объема выборки для симметричных и для несимметричных точек

график соответствует априорно симметричным точкам — $p_1(-1, 1)$, $p_2(1, 1)$, а нижний — несимметричным, $p_1(-1, 1)$, $p_2(1, 1.2)$. В обоих случаях для удобства сравнения масштабы на графиках выбраны одинаковыми. Видно, что в первом случае с ростом числа наблюдений мера μ быстро стремится к единице. В то же время во втором случае статистическая сходимость к единице, как и следовало ожидать, отсутствует.

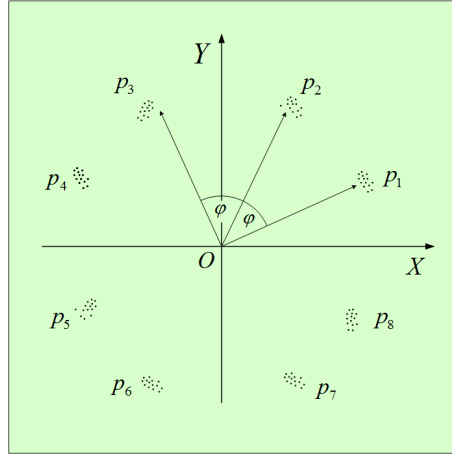


Рис. 5. Статистические выборки вращательно симметричных точек

Вращательная симметризация точек, заданных статистическими выборками

Пусть оценки положения характерных точек заданы статистическими выборками (рис. 6), и при этом априорно известно, что точки должны удовлетворять вращательной симметрии порядка n с центром вращения в начале координат. Найдем оценки положения всех характерных точек так, чтобы эти положения были наиболее вероятны, и при этом они удовлетворяли условию вращательной симметричности.

Пусть $p_k(x_k, y_k)$, $k = 1, \dots, n$, — неизвестные пока точки, которые должны быть вращательно симметричны. Сопоставим каждой точке вектор ее координат: $p_k \leftrightarrow z_k = (x_k \ y_k)^T$. Тогда вращательная симметричность точек означает, что $z_{k+1} = R_\varphi z_k$, $k = 1, \dots, n-1$, где $R_\varphi = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix}$ — оператор поворота на угол $\varphi = 2\pi/n$ против часовой стрелки. Отсюда следует, в частности, $z_k = R_\varphi^{k-1} z_1$ для всех k .

Введем более удобное задание точек:

$$\mathbf{p} = (p_1 \ p_2 \ \dots \ p_n)^T \leftrightarrow (x_1 \ y_1 \ x_2 \ y_2 \ \dots \ x_n \ y_n)^T = \mathbf{Z}.$$

Множество таких векторов образует линейное пространство R^{2n} . С учетом этого условие вращательной симметричности можно записать в общем виде для всех k :

$$\mathbf{Z} = \begin{pmatrix} z_1 \\ z_2 \\ \dots \\ z_n \end{pmatrix} = \underbrace{\begin{pmatrix} I \\ R_\varphi \\ \dots \\ R_{(n-1)\varphi} \end{pmatrix}}_A z_1 = A z_1.$$

Множество векторов $\mathbf{Z}$, получающихся при всех $z_1 \in R^2$, образует подпространство R_{Sym} в R^{2n} , причем размерность этого подпространства равна 2, а столбцы матрицы A образуют базис этого подпространства. Обозначим через $f_k(x_k, y_k)$ плотность вероятности появления точки $p_k(x_k, y_k)$, построенную по статистической выборке. Будем считать, что появления точек статистически независимы, т.е. случайные векторы $z_k = (x_k \ y_k)^T$, $k = 1, \dots, n$, независимы. Тогда случайный вектор $\mathbf{Z} = (x_1 \ y_1 \ \dots \ x_n \ y_n)^T$ описывается плотностью вероятности $f(\mathbf{Z}) = f_1(x_1, y_1) \dots f_n(x_n, y_n)$. При этом искомый вектор должен

принадлежать подпространству R_{Sym} . Это позволяет записать задачу отыскания наиболее вероятного положения вращательно симметричных точек следующим образом:

$$\begin{cases} f(\mathbf{Z}) \rightarrow \max, \\ \mathbf{Z} \in R_{Sym}. \end{cases} \quad (4)$$

Будем далее считать, что выборочные совокупности положений характерных точек подчиняются двумерным нормальным распределениям:

$$f_k(x_k, y_k) = \frac{1}{2\pi\sqrt{|C_k|}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_k - \mathbf{m}_k)^T C_k^{-1} (\mathbf{z}_k - \mathbf{m}_k) \right\}, \quad k = 1, \dots, n,$$

где $\mathbf{m}_k = (m_{x_k} \ y_{y_k})^T$ и C_k — соответственно математические ожидания и ковариационные матрицы, которые получены в результате статистического оценивания по выборке. Тогда задачу (4) можно записать в эквивалентном виде:

$$\begin{cases} \sum_{k=1}^n (\mathbf{z}_k - \mathbf{m}_k)^T C_k^{-1} (\mathbf{z}_k - \mathbf{m}_k) \rightarrow \min, \\ \mathbf{Z} \in R_{Sym}. \end{cases}$$

Учитывая, что условие $\mathbf{Z} \in R_{Sym}$ эквивалентно выполнению системы равенств $\mathbf{z}_k = R_\varphi^{k-1} \mathbf{z}_1$, $k = 1, \dots, n$, последнюю задачу можно переписать так:

$$\sum_{k=1}^n (R_\varphi^{k-1} \mathbf{z}_1 - \mathbf{m}_k)^T C_k^{-1} (R_\varphi^{k-1} \mathbf{z}_1 - \mathbf{m}_k) \rightarrow \min,$$

или, подробнее, как

$$\begin{aligned} & \sum_{k=1}^n \left[(R_\varphi^{k-1} \mathbf{z}_1)^T C_k^{-1} R_\varphi^{k-1} \mathbf{z}_1 - (R_\varphi^{k-1} \mathbf{z}_1)^T C_k^{-1} \mathbf{m}_k - \mathbf{m}_k^T C_k^{-1} R_\varphi^{k-1} \mathbf{z}_1 + \mathbf{m}_k^T C_k^{-1} \mathbf{m}_k \right] = \\ & = \sum_{k=1}^n \left[\mathbf{z}_1^T (R_\varphi^{k-1})^T C_k^{-1} R_\varphi^{k-1} \mathbf{z}_1 - 2\mathbf{m}_k^T C_k^{-1} R_\varphi^{k-1} \mathbf{z}_1 + \mathbf{m}_k^T C_k^{-1} \mathbf{m}_k \right] = \\ & = \mathbf{z}_1^T \underbrace{\left[\sum_{k=1}^n (R_\varphi^{k-1})^T C_k^{-1} R_\varphi^{k-1} \right]}_{(*)} \mathbf{z}_1 - 2 \left[\sum_{k=1}^n \mathbf{m}_k^T C_k^{-1} R_\varphi^{k-1} \right] \mathbf{z}_1 + \sum_{k=1}^n \mathbf{m}_k^T C_k^{-1} \mathbf{m}_k \rightarrow \min. \end{aligned}$$

Для отыскания минимума продифференцируем последнее выражение по $\mathbf{z}_1$ и приравняем производную к нулю. Учитывая, что матрица (*) симметрична, после очевидных преобразований получаем систему уравнений для отыскания уточненных координат точки $\mathbf{z}_1$:

$$\left[\sum_{k=1}^n (R_\varphi^{k-1})^T C_k^{-1} R_\varphi^{k-1} \right] \mathbf{z}_1 = \sum_{k=1}^n (R_\varphi^{k-1})^T C_k^{-1} \mathbf{m}_k.$$

Заметим, что если вероятностные распределения положения всех характерных точек невырождены, то данная система имеет единственное решение. Действительно, в силу положительной определенности ковариационных матриц определитель системы отличен от нуля. Однако при плохой обусловленности главной матрицы решения могут содержать значительные ошибки. В этой связи полезно дать нижнюю оценку для модуля определителя системы. Для этого воспользуемся следующим детерминантным неравенством [4]:

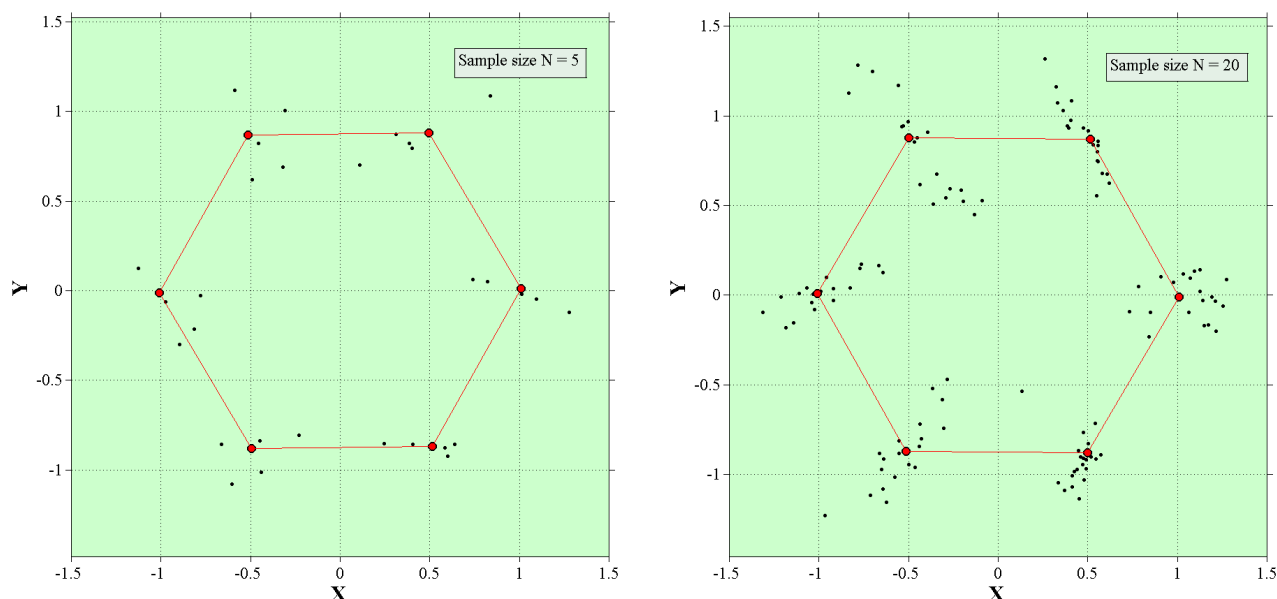


Рис. 6. Вращательная симметризация, основанная на статистических выборках точек

пусть матрицы $A_1, \dots, A_n$ — положительно определенные, и $\alpha_k > 0$, $k = 1, \dots, n$, $\alpha_1 + \dots + \alpha_n = 1$. Тогда $\left| \sum_{k=1}^n \alpha_k A_k \right| \geq \prod_{k=1}^n |A_k|^{\alpha_k}$.

Полагая $\alpha_k = \frac{1}{n}$ для всех k и учитывая, что в случае невырожденности вероятностных распределений $|C_k| > 0$, $k = 1, \dots, n$, получим для определителя системы:

$$\left| \sum_{k=1}^n (R_\varphi^{k-1})^T C_k^{-1} R_\varphi^{k-1} \right| \geq n^2 \left(\prod_{k=1}^n |C_k| \right)^{-\frac{1}{n}} > 0.$$

На рис. 6 показаны примеры вращательной симметризации при объеме нормально распределенных выборок равном 5 (слева) и 20 (справа). Для удобства восприятия симметризованные точки изображены в виде вершин правильного многоугольника.

Вращательная симметризация точек, заданных статистическими выборками, с неизвестным центром

Рассмотрим случай, когда вращательно симметричные характерные точки заданы статистическими выборками, но при этом неизвестны координаты центра $\mathbf{c}$ вращательной симметрии (рис. 8). Как и выше, будем считать, что порядок симметрии известен и равен n , а следовательно, имеется n статистических выборок, дающих некоторую информацию об истинном положении точек.

Будем, как и раньше считать, что вектор $\mathbf{Z} = (x_1 \ y_1 \ \dots \ x_n \ y_n)^T$, задающий координаты неизвестных точек $p_k(x_k \ y_k)$, $k = 1, \dots, n$, подчиняется распределению $f(\mathbf{Z}) = f_1(x_1, y_1) \dots f_n(x_n, y_n)$. Пусть, как и выше, $\mathbf{z} = (x_k \ y_k)^T$ — вектор координат точки. Если R_φ — оператор поворота на угол $\varphi = 2\pi/n$, то $\mathbf{z}_{k+1} - \mathbf{c} = R_\varphi(\mathbf{z}_k - \mathbf{c})$, $k = 1, \dots, n-1$. Отсюда получаем $\mathbf{z}_{k+1} - \mathbf{c} = R_{k\varphi}(\mathbf{z}_1 - \mathbf{c})$ или $\mathbf{z}_{k+1} = R_{k\varphi}\mathbf{z}_1 + (I - R_{k\varphi})\mathbf{c}$, $k = 0, \dots, n-1$.

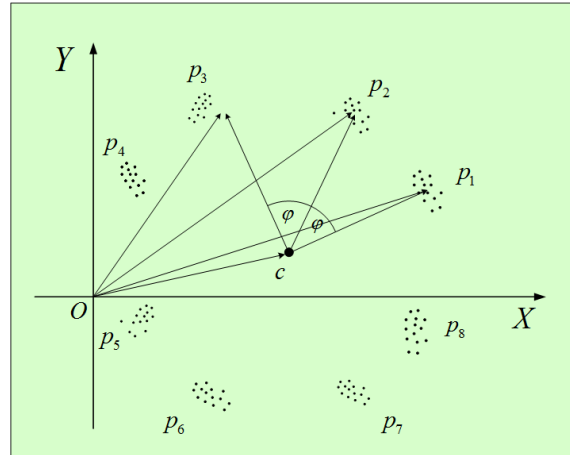


Рис. 7. Статистические выборки вращательно симметричных точек с неизвестным центром

Эти соотношения удобно записать сразу для всех k :

$$\mathbf{Z} = \begin{pmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \dots \\ \mathbf{z}_n \end{pmatrix} = \begin{pmatrix} I \\ R_\varphi \\ \dots \\ R_{(n-1)\varphi} \end{pmatrix} \mathbf{z}_1 + \begin{pmatrix} 0 \\ I - R_\varphi \\ \dots \\ I - R_{(n-1)\varphi} \end{pmatrix} \mathbf{c}.$$

Если ввести матрицу $B = (I \ I \ \dots \ I)^T$, то последнее равенство можно переписать в матричном виде: $\mathbf{Z} = A\mathbf{z}_1 + (B - A)\mathbf{c}$. Множество векторов $\mathbf{Z}$, которые получаются при фиксированном $\mathbf{c}$ и всех возможных $\mathbf{z} \in R^2$, обозначим $R_{Sym}(\mathbf{c})$. Таким образом, $R_{Sym}(\mathbf{c})$ является линейным многообразием в R^{2n} , параметрически зависящим от неизвестных координат центра вращательной симметрии.

Тогда задача отыскания наиболее вероятного положения точек вокруг наиболее вероятного положения центра вращательной симметрии имеет вид:

$$\begin{cases} f(\mathbf{Z}) \rightarrow \max, \\ \mathbf{Z} \in R_{Sym}(\mathbf{c}). \end{cases}$$

Пусть

$$f_k(x_1, y_k) = f_k(\mathbf{z}_k) = \frac{1}{2\pi\sqrt{|C_k|}} \exp \left\{ -\frac{1}{2}(\mathbf{z}_k - \mathbf{m}_k)^T C_k^{-1} (\mathbf{z}_k - \mathbf{m}_k) \right\},$$

где $\mathbf{m}_k = (m_{x_k} \ m_{y_k})^T$ — математическое ожидание, а C_k — ковариационная матрица координат точки $p(x_k, y_k)$, $k = 1, \dots, n$. Тогда

$$\begin{aligned} f(\mathbf{Z}) &= f_1(x_1, y_1) \dots f_n(x_n, y_n) = \\ &= \frac{1}{(2\pi)^n \sqrt{|C_1| \cdot \dots \cdot |C_n|}} \exp \left\{ -\frac{1}{2} \cdot \sum_{k=1}^n (\mathbf{z}_k - \mathbf{m}_k)^T C_k^{-1} (\mathbf{z}_k - \mathbf{m}_k) \right\}. \end{aligned}$$

Обозначим $\mathbf{m} = (\mathbf{m}_1^T \ \dots \ \mathbf{m}_n^T)^T$ и введем в рассмотрение блочно-диагональную матрицу $C = \text{diag}(C_1, \dots, C_n)$; тогда с учетом условия $\mathbf{Z} = A\mathbf{z}_1 + (B - A)\mathbf{c}$ получаем

$$\begin{aligned} f(\mathbf{Z}) &= \frac{1}{(2\pi)^n \sqrt{|C_1| \cdot \dots \cdot |C_n|}} \exp \left\{ -\frac{1}{2} \cdot (\mathbf{Z} - \mathbf{m})^T C^{-1} (\mathbf{Z} - \mathbf{m}) \right\} = \\ &= \frac{1}{(2\pi)^n \sqrt{|C_1| \cdot \dots \cdot |C_n|}} \exp \left\{ -\frac{1}{2} \cdot (A\mathbf{z}_1 + (B - A)\mathbf{c} - \mathbf{m})^T C^{-1} (A\mathbf{z}_1 + (B - A)\mathbf{c} - \mathbf{m}) \right\}, \end{aligned}$$

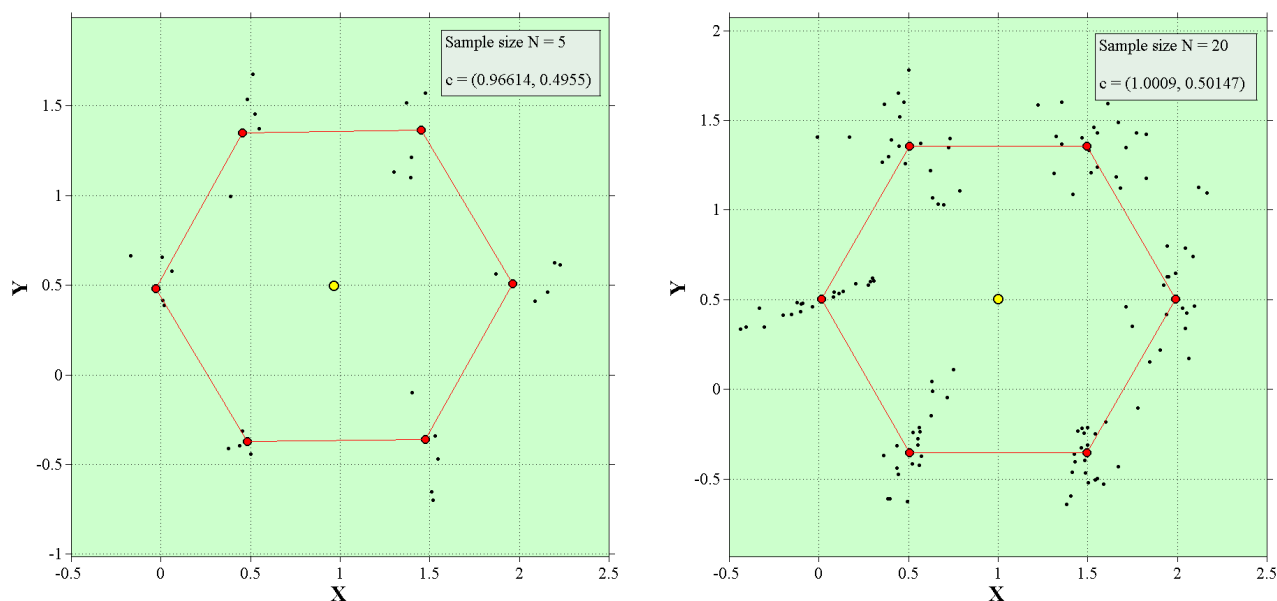


Рис. 8. Вращательная симметризация точек с неизвестным центром вращения

и задача сводится, таким образом, к безусловной минимизации квадратичной формы:

$$\Phi(z_1, c) = (Az_1 + (B - A)c - m)^T C^{-1} (Az_1 + (B - A)c - m) \xrightarrow{z_1, c \in R^2} \min.$$

Дифференцируя $\Phi(z_1, c)$ последовательно по z_1 и по c и приравнявая полученные выражения к нулю, после упрощений получаем:

$$\begin{cases} A^T C^{-1} A(z_1 - c) + A^T C^{-1} Bc = A^T C^{-1} m, \\ B^T C^{-1} A(z_1 - c) + B^T C^{-1} Bc = B^T C^{-1} m. \end{cases}$$

Таким образом, если z_1^* и c^* — решение последней системы, то координаты симметризованных точек вычисляются по формуле $z_{k+1}^* = R_{k\varphi} z_1^* + (I - R_{k\varphi}) c^*$, $k = 0, \dots, n - 1$.

На рис. 8 представлены результаты моделирования симметризации вращательно симметричных точек с неизвестным центром симметрии при объемах выборок 5 и 20. При этом для случая выборок объемом 5 (слева) координаты найденного центра $c^*(0.96614, 0.4955)$, а для выборок объемом 20 (справа), соответственно, — $c^*(1.0009, 0.50147)$. Истинным положением центра при моделировании была точка $c^*(1, 0.5)$. Как и выше, симметризованные точки представлены вершинами правильного многоугольника.

Заключение

Симметризация характерных точек изображения относительно отражательной и вращательной симметрии имеет многочисленные приложения, поскольку эти типы симметрии являются наиболее простыми и распространенными. Ранее авторами рассматривались вопросы симметризации точек по отношению к отражательной и вращательной симметрии при различных условиях. При этом всегда считалось, что каждая точка представлена одной единственной оценкой своего истинного положения. Все методы обеспечивали построение гарантированно симметричного положения новых точек при минимальном уклонении от заданных.

В данной работе рассмотрена более общая задача, когда каждая точка задается некоторой выборкой оценок своего положения. Поэтому в такой постановке задача носит в большей степени статистический характер. Предложены методы построения симметричных точек, которые максимизируют вероятность появления точек при гарантированном их симметричном расположении как в случае отражательной, так и вращательной симметрии. Показано, что такая постановка сводится к оптимизационной задаче в «нестатистическом» случае с тем отличием, что минимальное отклонение ищется не в пространстве с евклидовой метрикой, а в пространстве с метрикой Махаланобиса.

Предложен простой способ оценки априорной симметричности исходных данных на основе введенной меры симметричности. Показано, что с ростом объема выборок, задающих симметризуемые точки, мера симметричности стремится к единице в статистическом смысле.

Вместе с тем в данной работе не рассмотрены некоторые важные случаи, а именно: отражательная симметризация относительно произвольной фиксированной оси, а также наиболее полезный с практической точки зрения случай неизвестной оси симметрии. Для случая, когда характерные точки задаются координатами одного, возможно, неточного своего положения, данные задачи были подробно исследованы авторами ранее [1]. Аналогичные рассуждения при задании характерных точек статистическими выборками являются более громоздкой задачей, поскольку в отличие от случая фиксированной оси, когда симметризация пар характерных точек производится независимо, для построения неизвестной оси симметрии и последующей симметризации требуется учет статистической информации обо всех точках одновременно. Авторы предполагают рассмотреть эту и другие связанные с ней задачи в следующей работе.

Отдельным вопросом, заслуживающим особое внимание, является агрегирование в статистических выборках информации о положении точек, полученной из разных источников, о чем упоминалось во введении к статье. Координаты точек, порождаемые различными источниками информации, образуют неоднородные (гетерогенные) выборки, что накладывает дополнительные сложности, связанные с отличающимися статистическими характеристиками элементов неоднородной выборки. В некоторых случаях (например, когда выборку значений координат точек можно разделить на классы с одинаковыми статистическими характеристиками) можно рассчитывать на получение оптимального решения. Поэтому представляется целесообразным корректно распространить методы симметризации также и на случай неоднородных статистических выборок.

Авторы выражают благодарность неизвестным рецензентам за полезные замечания и пожелания, способствовавшие улучшению статьи, а также стимулирующие дальнейшие исследования в данном направлении.

Литература

- [1] Podolak J., Shilane P., Giesen J., Gross M., Guibas L. Example-based 3d scan completion // Symposium on Geometry Processing, 2005. P. 23–32.
- [2] Martinet A., Soler C., Holzschuch N., Sillion F. Accurate detection of symmetries in 3d shapes // ACM Trans. Graph., 2006. Vol. 25(2). P. 439–464.
- [3] Mitra N. J., Guibas L. J., Pauly M. Partial and approximate symmetry detection for 3d geometry // ACM Trans. Graph., 2006. Vol. 25(3). P. 560–568.
- [4] Mitra N. J., Guibas L. J., Pauly M. Symmetrization // ACM Trans. Graph., 2007. Vol. 26(3).

- [5] *Thrun S., Wegbreit B.* Shape from symmetry // 10th IEEE Conference (International) on Computer Vision Proceedings, 2005. P. 1824–1831.
- [6] *Simari P., Kalogerakis E., Singh K.* Folding meshes: Hierarchical mesh segmentation based on planar symmetry // Eurographics Symposium on Geometry Processing Proceedings, 2006. P. 111–119.
- [7] *Karkishchenko A. N., Mnukhin V. B.* Reflective Symmetrization of feature points in images // 11th Conference (International) on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013) Proceedings. — Samara, 2013. P. 209–212.
- [8] *Каркищенко А. Н., Мнухин В. Б.* Восстановление симметричности точек на изображениях объектов с отражательной симметрией // Машинное обучение и анализ данных, 2013. Т. 1. № 5. С. 619–629.
- [9] *Каркищенко А. Н., Гречухин И. А.* Локализация характерных точек на основе естественной симметрии изображений // Тр. Первой научно-технич. конф. «Интеллектуальные системы управления на железнодорожном транспорте (ИСУЖТ-2012)» — М.: ОАО НИИАС, 2012. С. 261–266.
- [10] *Каркищенко А. Н., Гречухин И. А.* Статистическое распознавание лиц по геометрии характерных точек для систем транспортной безопасности // *Управление большими системами: Сб. тр.* — М.: ИПУ РАН, 2012. Вып. 38. С. 65–77.
- [11] *Прасолов В. В.* Задачи и теоремы линейной алгебры. М.: Наука, 2008. 536 с.

References

- [1] *Karkishchenko A. N., Mnukhin V. B.* 2013. Recovery of points symmetry in images of objects with reflectional symmetry. *Machine Learning Data Analysis*. 1(5):619–629.
- [2] *Karkishchenko A. N., Grechukhin I. A.* 2012. Localizing feature points based on natural symmetry of images. *1st Scientific Conference “Intelligent Control Systems for Rail Transport. ISUZHT 2012 Proceedings”* Moscow: JSC NIIAS. 261–266.
- [3] *Karkishchenko A. N., Grechukhin I. A.* 2012. Statistical face recognition on the geometry of the characteristic points for transport safety systems. *Managing Large Systems Proceedings*. Moscow: Institute of Control Sciences. 38:65–77.
- [4] *Prasolov V. V.* 2008. *Problems and theorems in linear algebra*. Moscow: Nauka. 536 p.
- [5] *Karkishchenko A. N., Mnukhin V. B.* 2013. Reflective symmetrization of feature points in images. *11th Conference (International) on Pattern Recognition and Image Analysis: New Information Technologies (PRIA-11-2013) Proceedings*. Samara. 209–212.
- [6] *Martinet A., Soler C., Holzschuch N., Sillion F.* 2006. Accurate detection of symmetries in 3d shapes. *ACM Trans. Graph.* 25(2):439–464.
- [7] *Mitra N. J., Guibas L. J., Pauly M.* 2006. Partial and approximate symmetry detection for 3d geometry. *ACM Trans. Graph.* 25(3):560–568.
- [8] *Mitra N. J., Guibas L. J., Pauly M.* 2007. Symmetrization. *ACM Trans. Graph.* 26(3).
- [9] *Podolak J., Shilane P., Giesen J., Gross M., Guibas L.* 2005. Example-based 3d scan completion. *Symposium on Geometry Processing Proceedings*. 23–32.
- [10] *Simari P., Kalogerakis E., Singh K.* 2006. Folding meshes: Hierarchical mesh segmentation based on planar symmetry. *Eurographics Symposium on Geometry Processing Proceedings*. 111–119.
- [11] *Thrun S., Wegbreit B.* 2005 Shape from symmetry. *10th IEEE Conference (International) on Computer Vision Proceedings*. 1824–1831.