

Обучение распознаванию без переобучения*

Н. Г. Загоруйко^{1,2}, *О. А. Кутненко*^{1,2}, *А. О. Зырянов*², *Д. А. Леванов*¹
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. Zyryanov*², *D. A. Levanov*¹

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-эталонные (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.

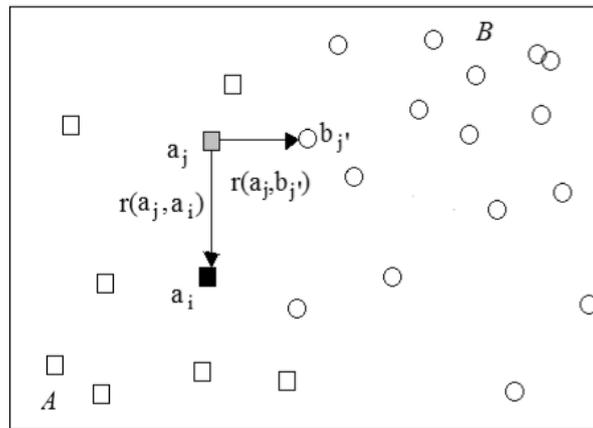


Рис. 1. Оценка веса объекта $a_i \in A$.

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. цензируются).

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.

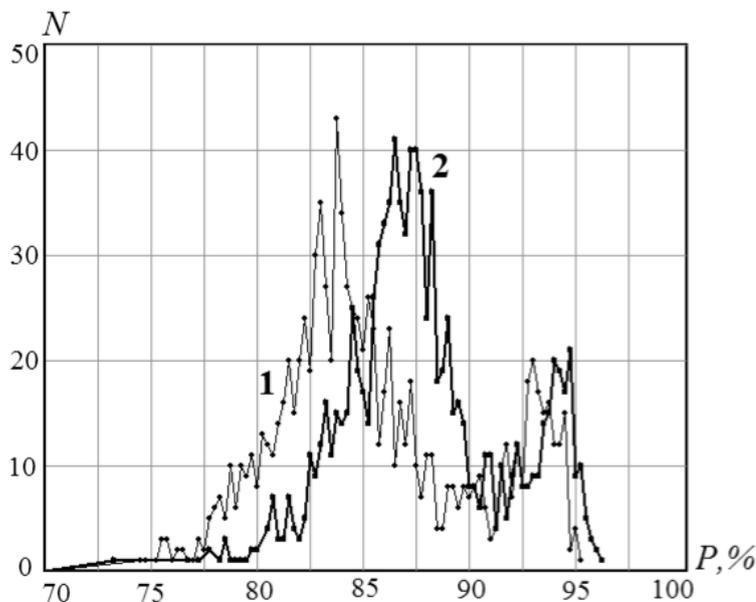


Рис. 2. Распределение надежности $P(\%)$ распознавания контрольной выборки без цензурирования (1) и с цензурированием (2)

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., Степанко 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] Zagoruiiko 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] Zagoruiiko 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., Сапегин 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] *Lemeshko B. Yu., Lemeshko 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] *Ivaxnenko 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.