

Выбор оптимальной модели классификации физической активности по измерениям акселерометра*

М. С. Попова¹, В. В. Стрижов²

В данной работе решается проблема построения оптимальных устойчивых моделей в задаче классификации физической активности человека. Каждый тип физической активности конкретного человека описывается набором признаков, сгенерированных по временным рядам с акселерометра. В условиях мультиколлинеарности признаков выбор устойчивых моделей классификации затруднен из-за необходимости оценки большого числа параметров этих моделей. Оценка оптимального значения параметров также затруднена в связи с тем, что функция ошибок имеет большое количество локальных минимумов в пространстве параметров. В работе исследуются модели, принадлежащие классу двуслойных нейронных сетей. Ставится задача нахождения Парето оптимального фронта на множестве допустимых моделей. Предлагаются критерии оптимального, последовательного и устойчивого прореживания нейронной сети, критерий наращивания сети, а также строится стратегия пошаговой модификации модели с использованием предложенных критериев. В вычислительном эксперименте модели, порождаемые предложенной стратегией, сравниваются по трем критериям качества — сложность, точность и устойчивость.

Ключевые слова: *классификация, нейронные сети, сложность, устойчивость, оптимальность по Парето, критерии прореживания и наращивания.*

Selection of optimal physical activity classification model using measurements of accelerometer*

Maria Popova¹, Vadim Strijov²

In this paper we solve the problem of selecting optimal stable models for classification of physical activity. Each type of physical activity of a particular person is described by a set of features generated from the accelerometer time series. In conditions of feature's multicollinearity selection of stable models is hampered by the need to evaluate a large number of parameters of these models. Evaluation of optimal parameter values is also difficult due to the fact that the error function has a large number of local minima in the parameter space. In the paper we choose the optimal models from the class of two-layer artificial neural networks. We solve the problem of finding the Pareto optimal front of the set of models. The paper presents a stepwise strategy of building optimal stable models. The strategy includes steps of deleting and adding parameters, criteria of pruning and growing the model and criteria of breaking the process of building. The computational experiment compares models generated by the proposed strategy on three quality criteria — complexity, accuracy and stability.

Keywords: *classification, artificial neural networks, complexity, accuracy, stability, Pareto efficiency, growing and pruning criteria.*

1. Введение

Решается задача последовательной модификации нейронной сети. Требуется получить нейронную сеть с небольшим числом связей между нейронами, которая бы достаточно

1 — Московский физико-технический институт (Moscow Institute of Physics and Technology), maria_popova@phystech.edu

2 — ВЦ РАН им. Дородницына (Dorodnicyn Computing Centre of RAS), strijov@ccas.com

Работа поддержана РФФИ, грант 14-07-31045.

точно решала задачу классификации физической активности человека по показаниям акселерометра и обладала бы свойством устойчивости к возмущениям данных. Ввиду этого возникает задача минимизации сложности модели без потери точности классификации [1].

Существуют два базовых подхода к решению задачи выбора сетей оптимальной структуры: *наращивание структуры сети* (network growing) [2] и *прореживание структуры сети* (network pruning) [3, 4, 5].

Согласно первому подходу в качестве начальной модели выбирается сеть с небольшим числом нейронов, решающая поставленную задачу с большим значением функции ошибки, после чего в сеть добавляются новые нейроны и связи между ними. В [2] описаны некоторые методы наращивания, приведен сравнительный анализ генетических алгоритмов с алгоритмом байесовской оптимизации. В алгоритмах метода прореживания модифицируется многослойная сеть с избыточным числом нейронов и связей между ними. Классическими алгоритмами прореживания нейронных сетей являются «optimal brain damage» [4] и «optimal brain surgery» [5], основанные на вычислении вторых производных функции ошибки. Также получили развитие *гибридные алгоритмы*, в которых объединяются оба упомянутых выше подхода [6, 7, 8].

В данной работе предлагается стратегия пошаговой модификации нейронной сети, комбинирующая этапы добавления и удаления параметров [9, 10, 11]. Стратегия включает в себя критерии прореживания и наращивания структуры сети, критерии останова этапов добавления и удаления параметров, а также критерий останова процедуры модификации. Согласно предложенной стратегии процедура модификации начинается с нейронной сети избыточной сложности и чередует шаги удаления и добавления параметров до тех пор, пока этот процесс не стабилизируется согласно критерию останова процедуры модификации. Критерии прореживания и наращивания позволяют на каждом шаге процедуры модификации выбирать параметр, добавление или удаление которого улучшит качество нейронной сети, которое оценивается по трем критериям качества — сложность, точность и устойчивость [12, 13, 14]. Также предлагается рассматривать процедуру пошаговой модификации нейронной сети как путь в многомерном кубе.

В вычислительном эксперименте вычисляются значения критериев качества для нейронных сетей, порожденных предложенной стратегией. В качестве тестового примера рассматривается задача классификации физической активности человека по измерениям акселерометра [15].

2. Постановка задачи

Дана выборка $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{t}_i)\}, i \in \mathcal{I} = \{1 \dots m\}$, состоящая из m объектов, каждый из которых описывается n признаками $\mathbf{x}_i \in \mathbb{R}^n$ и принадлежит одному из z классов $\mathbf{t}_i \in \{0, 1\}^z$. Также задано разбиение множество индексов выборки $\mathcal{I} = \mathcal{L} \sqcup \mathcal{T}$ на обучающую $(\mathbf{x}_i, \mathbf{t}_i), i \in \mathcal{L}$ и контрольную $(\mathbf{x}_i, \mathbf{t}_i), i \in \mathcal{T}$. Требуется выбрать устойчивую модель классификации оптимальной сложности.

Определение 1. Моделью назовем отображение:

$$\mathbf{f} : (\mathbf{w}, \mathbf{X}) \mapsto \mathbf{y},$$

где $\mathbf{w} = [w_1, \dots, w_j, \dots, w_k]^\top, j \in \mathcal{J} = \{1, \dots, k\}$ — вектор параметров модели, $\mathbf{X} \in \mathbb{R}^{n \times m}$ — матрица плана, $\mathbf{y} \in \{0, 1\}^z$ — зависимая переменная.

Предполагается, что переменная \mathbf{y} — мультиномиально распределенная случайная величина, а переменная \mathbf{w} имеет нормальное распределение с нулевым математическим ожи-

данием:

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1}), \quad (1)$$

где \mathbf{A}^{-1} — ковариационная матрица параметров общего вида, положительно-определенная: $\mathbf{w}^\top \mathbf{A} \mathbf{w} > \mathbf{0}$ для любого $\mathbf{w} \in \mathbb{R}^k$.

В данной работе рассматриваются модели, принадлежащие классу двуслойных нейронных сетей с функциями активации \tanh и softmax :

$$\begin{aligned} \mathbf{a}(\mathbf{x}) &= \mathbf{W}_2^\top \tanh(\mathbf{W}_1^\top \mathbf{x}), \\ \mathbf{p}(\mathbf{x}) &= \frac{\exp(\mathbf{a}(\mathbf{x}))}{\sum_j \exp(a_j(\mathbf{x}))}. \end{aligned}$$

Вектор \mathbf{p} интерпретируется как вектор вероятностей: p_ξ есть вероятность того, что вектор \mathbf{x} принадлежит классу с номером ξ :

$$\mathbf{p}(\mathbf{x}) = \{p_\xi\}, \quad 0 \leq p_\xi \leq 1, \quad \sum p_\xi = 1, \quad \xi = 1 \dots z.$$

Под вектором параметров двуслойной нейронной сети будем понимать $\mathbf{w} = \text{vec}(\mathbf{W}_1^\top | \mathbf{W}_2^\top)$, где $\mathbf{W}_1, \mathbf{W}_2$ — матрицы весов первого и второго слоя нейронной сети. Вектор $\mathbf{y} = [y_1, \dots, y_\xi, \dots, y_z]^\top$ определим следующим образом:

$$y_\xi = \begin{cases} 1, & \text{если } \xi = \underset{\xi \in \{1, \dots, z\}}{\text{argmax}}(p_\xi), \\ 0, & \text{иначе.} \end{cases}$$

Под структурными параметрами двухслойной нейронной сети будем понимать количество нейронов в скрытом слое нейронной сети — N_h . Матрица весов первого слоя имеет размерность $n \times N_h$, матрица весов второго слоя имеет размерность $N_h \times z$. Далее будем считать, что структурные параметры фиксированы и одинаковы для всех рассматриваемых моделей.

Определение 2. Параметр w_j модели \mathbf{f} назовем активным, если $w_j \neq 0$.

Определение 3. Структурой \mathcal{A} модели \mathbf{f} назовем множество индексов активных параметров этой модели $\mathcal{A} = \{j : w_j \neq 0\} \subseteq \mathcal{J}$.

Каждая структура $\mathcal{A} \subseteq \mathcal{J}$ однозначно задает некоторую модель:

$$\mathbf{f}_{\mathcal{A}} : \hat{\mathbf{w}}_{\mathcal{A}} \in \mathbb{R}^k,$$

где $\mathbf{f}_{\mathcal{A}}$ — модель со структурой \mathcal{A} , а $\hat{\mathbf{w}}_{\mathcal{A}} \in \mathbb{R}^k$ — оптимальный вектор параметров модели $\mathbf{f}_{\mathcal{A}}$, определение которому будет дано ниже. Объединение всех $\mathbf{f}_{\mathcal{A}}$ назовем множеством допустимых моделей:

$$\mathfrak{F} = \bigcup_{\mathcal{A} \subseteq \mathcal{J}} \{\mathbf{f}_{\mathcal{A}}\}. \quad (2)$$

Оптимальную модель $\hat{\mathbf{f}}_{\mathcal{A}}$ будем выбирать из множества допустимых моделей \mathfrak{F} .

Согласно гипотезе (1) о распределении многомерных случайных величин \mathbf{y} и \mathbf{w} в качестве функции ошибки выберем функцию:

$$S(\mathbf{w} | \mathcal{K}) = - \sum_{i \in \mathcal{K}} \sum_{\xi=1}^z t_{i\xi} \ln(p_\xi(\mathbf{x}_i, \mathbf{w})), \quad (3)$$

максимизирующую логарифм правдоподобия случайной величины \mathbf{y} и заданную на разбиении выборки \mathcal{D} , определенном некоторым множеством индексов $\mathcal{K} \subseteq \mathcal{I}$, $\mathbf{t}_i = [t_{i1}, \dots, t_{i\xi}, \dots, t_{iz}]^\top$.

Определение 4. *Оптимальным вектором параметров модели \mathbf{f}_A назовем такой вектор $\hat{\mathbf{w}}_A$, который является решением следующей задачи оптимизации:*

$$\hat{\mathbf{w}}_A = \underset{\mathbf{w}_A \in \mathbb{R}^k}{\operatorname{argmin}} S(\mathbf{w}_A | \mathcal{L}). \quad (4)$$

Для оценки качества моделей и сравнения их друг с другом введем три критерия качества — сложность, устойчивость и точность.

Определение 5. *Сложностью $C = C(\hat{\mathbf{w}})$ модели \mathbf{f} с вектором параметров $\hat{\mathbf{w}} = [w_1, \dots, w_k]$ назовем мощность множества активных параметров этой модели:*

$$C(\mathbf{w}) = \sum_{i=1}^k [w_i \neq 0].$$

Чем больше мощность множества активных параметров, тем сложнее модель. Максимально возможная сложность модели равна размерности пространства параметров k .

Определение 6. *Устойчивостью $\eta = \eta(\hat{\mathbf{w}})$ модели \mathbf{f} с вектором параметров $\hat{\mathbf{w}}$ назовем число η , равное числу обусловленности матрицы \mathbf{A} 1, т.е. $\eta(\hat{\mathbf{w}}) = \frac{\lambda_{\max}}{\lambda_{\min}}$, где λ_{\max} — максимальное, а λ_{\min} — минимальное собственные числа матрицы \mathbf{A} .*

Чем лучше обусловлена матрица \mathbf{A} , тем более устойчива модель. У идеально устойчивой модели $\lambda_{\min} = \lambda_{\max}$, $\eta = 1$.

Определение 7. *Под точностью S модели \mathbf{f} с вектором параметров $\hat{\mathbf{w}}$ будем понимать величину функции ошибки (2) на контрольной выборке.*

Чем больше значение функции ошибки, тем меньше точность модели.

Введем на множестве допустимых моделей \mathcal{F} отношение доминирования. Будем говорить, что модель \mathbf{f}_a доминирует модель \mathbf{f}_b и обозначать $\mathbf{f}_a \succ \mathbf{f}_b$, если

$$C_a \leq C_b, \quad \eta_a \leq \eta_b, \quad S_a \leq S_b,$$

где C_i , η_i , S_i — сложность, устойчивость и точность модели \mathbf{f}_i , $i \in \{a, b\}$.

Определение 8. *Модель $\mathbf{f} \in \mathcal{F}$ назовем оптимальной по Парето, если не существует $\mathbf{f}' \in \mathcal{F}$ такой, что $\mathbf{f}' \succ \mathbf{f}$.*

Определение 9. *Множество оптимальных по Парето моделей назовем Парето оптимальным фронтом $\operatorname{POF}_{\mathfrak{F}}$ множества допустимых моделей \mathfrak{F} .*

Задача выбора оптимальной модели состоит в том, чтобы найти Парето оптимальный фронт $\operatorname{POF}_{\mathfrak{F}}$ множества допустимых моделей \mathfrak{F} .

3. Стратегия пошаговой модификации модели

Определение 10. *Стратегией пошаговой модификации модели называется процедура последовательного изменения модели, в которой на каждом шаге решается оптимизационная задача вида:*

$$\hat{j} = \underset{j \in \mathcal{A}}{\operatorname{argopt}} Q(\hat{\mathbf{w}}_A),$$

где Q — один из вышеприведенных критериев качества или их набор.

Стратегия задается следующими математическими объектами:

- набором критериев оптимизации — сложность, точность, устойчивость $\{C, S, \eta\}$,
- набором ограничений на структуру и параметры модели $\mathcal{A} \subseteq \mathcal{J}$, $\mathbf{w} = \hat{\mathbf{w}}_{\mathcal{A}}$ из (4),
- критериями останова шагов удаления (10) и добавления (11),
- критерием останова процедуры выбора модели (12).

Действуя согласно стратегии, мы будем изменять структуру модели, удаляя из неё элементы и добавляя их согласно (12).

Для определения индекса параметра \hat{j} , который должен быть удален из модели или добавлен в неё, ниже предлагаются несколько критериев оптимизации модели.

3.1 Критерий оптимального прореживания

Этот критерий позволяет выяснить индекс параметра, удаление которого приведет к минимизации приращения функции ошибки (2). Для функции ошибки используется локальная аппроксимация вблизи локального минимума вектора параметров \mathbf{w}_0 :

$$E(\mathbf{w}_0 + \Delta\mathbf{w}) = E(\mathbf{w}_0) + \mathbf{g}^\top(\mathbf{w}_0)\Delta\mathbf{w} + \frac{1}{2}\Delta\mathbf{w}^\top \mathbf{H}\Delta\mathbf{w} + O(\|\Delta\mathbf{w}\|^3),$$

где $\Delta\mathbf{w}$ — возмущение вектора параметров в данной точке \mathbf{w}_0 ; $\mathbf{g}(\mathbf{w}_0)$ — вектор градиента, вычисленный в точке \mathbf{w}_0 , $\mathbf{H} = \mathbf{H}(\mathbf{w}_0)$ — матрица вторых производных функции ошибки. Предполагается, что матрица вторых производных $\mathbf{H} = \mathbf{H}(\mathbf{w})$ диагональная, а функция ошибки в окрестности глобального или локального минимума является квадратичной. Исходя из этих гипотез, аппроксимация функции ошибки записывается в следующем виде:

$$\Delta E = E(\mathbf{w}_0 + \Delta\mathbf{w}) - E(\mathbf{w}_0) = \frac{1}{2}\Delta\mathbf{w}^\top \mathbf{H}\Delta\mathbf{w}.$$

Пусть w_j — некоторый параметр. Удаление этого параметра (присваивание ему нулевого значения) эквивалентно выполнению условия:

$$\mathbf{e}_j^\top \Delta\mathbf{w} + w_j = 0,$$

где \mathbf{e}_j^\top — вектор, все элементы которого равны нулю, за исключением j -го, который равен единице. Таким образом, получаем задачу условной минимизации:

$$\Delta E = \frac{1}{2}\Delta\mathbf{w}^\top \mathbf{H}\Delta\mathbf{w} \rightarrow \min, \quad \mathbf{e}_j^\top \Delta\mathbf{w} + w_j = 0.$$

Для решения этой задачи строим лагранжиан:

$$L = \frac{1}{2}\Delta\mathbf{w}^\top \mathbf{H}\Delta\mathbf{w} - \lambda_j(\mathbf{e}_j^\top \Delta\mathbf{w} + w_j).$$

Продифференцировав L по $\Delta\mathbf{w}$, получаем значение Лагранжиана L_j для элемента w_j :

$$L_j = \frac{w_j^2}{2[\mathbf{H}^{-1}]_{j,j}},$$

где \mathbf{H}^{-1} — матрица обратная гессиану \mathbf{H} ; $[\mathbf{H}^{-1}]_{j,j}$ — j -ый диагональный элемент этой матрицы. Значение лагранжиана L_j называется выпуклостью w_j . Выпуклость L_j описывает рост среднеквадратичной ошибки, вызываемый удалением параметра w_j .

Критерию оптимального прореживания отвечает параметр $w_{\hat{j}}$, соответствующий минимальному значению выпуклости:

$$\hat{j} = \operatorname{argmin}_{j \in \mathcal{A}} L_j. \quad (5)$$

3.2 Критерий последовательного прореживания

В качестве второго критерия предлагается простой критерий последовательного удаления параметров w_j — компонент вектора \mathbf{w} . Основной идеей этого критерия является принцип локально-оптимального выбора — критерию отвечает параметр w_j , без которого функция ошибки (2) оказывается минимальной.

Для нахождения параметра, отвечающего этому критерию, решается задача

$$\hat{j} = \operatorname{argmin}_{j \in \mathcal{A}} E(\mathbf{w}_{\mathcal{A}} \setminus w_j | \mathcal{T}). \quad (6)$$

3.3 Критерий устойчивого прореживания

Помимо вышеописанных критериев предлагается критерий устойчивого прореживания, основанный на модификации метода Белсли.

Пусть \mathbf{W} — матрица реализаций оптимального вектора параметров $\hat{\mathbf{w}}$, определенно-го в (4) и рассматриваемого согласно (2) как многомерная случайная величина. Пусть эта матрица имеет размерность $r \times k$. Выполним её сингулярное разложение:

$$\mathbf{W} = \mathbf{U}\mathbf{S}\mathbf{V}^T, \quad (7)$$

где \mathbf{U} и \mathbf{V} — ортогональные матрицы размера $r \times r$ и $k \times k$ соответственно, а $\mathbf{\Lambda}$ — матрица, на диагонали которой стоят сингулярные числа матрицы \mathbf{W} .

По определению ковариационная матрица вектора параметров \mathbf{w} вычисляется как:

$$\mathbf{A}^{-1} = \operatorname{cov}(\mathbf{W}) = E(\mathbf{W}^T \mathbf{W}) - E(\mathbf{W})E(\mathbf{W}^T) = E(\mathbf{W}^T \mathbf{W}).$$

Последнее равенство выполняется в силу предположения о том, что математическое ожидание вектора параметров равно нулю: $E(\mathbf{w}) = \mathbf{0}$. По матрице реализаций \mathbf{W} многомерной случайной величины \mathbf{w} ковариационная матрица может быть оценена следующим образом:

$$\mathbf{A}^{-1} = \frac{1}{r} \mathbf{W}\mathbf{W}^T.$$

У ковариационной матрицы есть нулевые строки с индексами из множества $\mathcal{J} \setminus \mathcal{A}$, где \mathcal{J} — множество индексов всех параметров модели, а \mathcal{A} — множество индексов активных параметров. Таким образом, ковариационная матрица является неполноранговой.

Теперь, используя сингулярное разложение (7) матрицы \mathbf{W} получим выражение для матрицы \mathbf{A}^{-1} :

$$\mathbf{A}^{-1} = (\mathbf{W}\mathbf{W}^T) = (\mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \mathbf{V}\mathbf{\Lambda}^T \mathbf{U}^T) = (\mathbf{U}\mathbf{\Lambda}\mathbf{\Lambda}^T \mathbf{U}^T) = \mathbf{U}\mathbf{\Lambda}^2 \mathbf{U}^T.$$

Индексом обусловленности η_{ζ} назовём отношение максимального элемента λ_{\max} матрицы $\mathbf{\Lambda}$ к ζ -ому по величине элементу λ_{ζ} этой матрицы:

$$\eta_{\zeta} = \frac{\lambda_{\max}}{\lambda_{\zeta}}.$$

Так как ковариационная матрица \mathbf{A}^{-1} неполноранговая, то некоторые значения индексов обусловленности неопределены. Однако мы рассматриваем только те значения η_ζ , для которых $\zeta \in \mathcal{A}$.

Оценками дисперсии параметров будут диагональные элементы \mathbf{A}^{-1} :

$$\sigma(w_\zeta) = \mathbf{A}_{\zeta\zeta}^{-1}.$$

Долевой коэффициент $q_{\zeta j}$ определим как вклад j -го признака в дисперсию ζ -го элемента вектора параметров \mathbf{w} :

$$q_{\zeta j} = \frac{u_{\zeta j}^2 \lambda_{jj}^2}{\sigma(w_\zeta)}.$$

Находим индексы обусловленности и долевы коэффициенты для набора активных параметров \mathcal{A} . Большие значения индексов обусловленности указывают на зависимость между признаками. Поэтому для нахождения параметра, отвечающего этому критерию прорезивания, находим максимальный индекс обусловленности:

$$\hat{\zeta} = \operatorname{argmax}_{\zeta \in \mathcal{A}} \eta_\zeta.$$

Затем находим максимальный долевой коэффициент, соответствующий найденному максимальному индексу обусловленности $\eta_{\hat{\zeta}}$:

$$\hat{j} = \operatorname{argmax}_{j \in \mathcal{A}} q_{\hat{\zeta} j}, \quad (8)$$

Параметр $w_{\hat{j}}$ и есть параметр, отвечающий критерию устойчивого прорезивания.

3.4 Критерий последовательного наращивания

Критерий последовательного добавления параметров как и критерий 6 основан на принципе локально-оптимального выбора — критерию отвечает параметр, при добавлении которого в сеть, функция ошибки (2) минимальна.

Для нахождения параметра, отвечающего этому критерию, решается задача:

$$\hat{j} = \operatorname{argmin}_{j \in \mathcal{T} \setminus \mathcal{A}} E(\mathbf{w}_{\mathcal{A}} \cup w_j | \mathcal{T}). \quad (9)$$

3.5 Описание базовой стратегии

Стратегия пошаговой модификации модели состоит из двух этапов — Del и Add. Перед началом процедуры модификации все параметры модели активны.

Этап Del. Ищем параметр с индексом \hat{j} , отвечающий одному из критериев прорезивания (5), (6) или (7) и удаляем его из множества активных параметров:

$$\mathcal{A} = \mathcal{A} \setminus \hat{j}.$$

Этап Del повторяем до тех пор, пока ошибка $E(\mathbf{w}_{\mathcal{A}} | \mathcal{T})$ превосходит свое минимальное значение на данном этапе не более, чем на некоторое заданное значение δE_1 . Следующее условие является критерием останова шага Del:

$$E(\mathbf{w}_{\mathcal{A}} | \mathcal{T}) \geq E_{\min} + \delta E_1, \quad (10)$$

где E_{\min} — некоторое заданное значение.

Этап Add. В модели ищем параметр \hat{j} , отвечающий критерию наращивания (8) и добавляем найденный параметр во множество активных параметров:

$$\mathcal{A} = \mathcal{A} \cup \hat{j}.$$

Критерием останова шага Add является выполнение условия:

$$E(\hat{\mathbf{w}}_{\mathcal{A}}|\mathcal{T}) \geq E_{\min} + \delta E_2, \quad (11)$$

где E_{\min} — некоторое заданное значение. На Рис. 1 приведен график, демонстрирующий изменение функции ошибки при удалении параметров из модели. Аналогичным образом ведет себя функция ошибки при добавлении параметров в модель. Из графика видно, что эта зависимость имеет минимум, а значит модели с большим числом параметров не являются наиболее точными. На Рис. 2 показано как согласно критериям останова (10) и (11) сменяются шаги удаления и добавления. Процедура модификации продолжается до тех

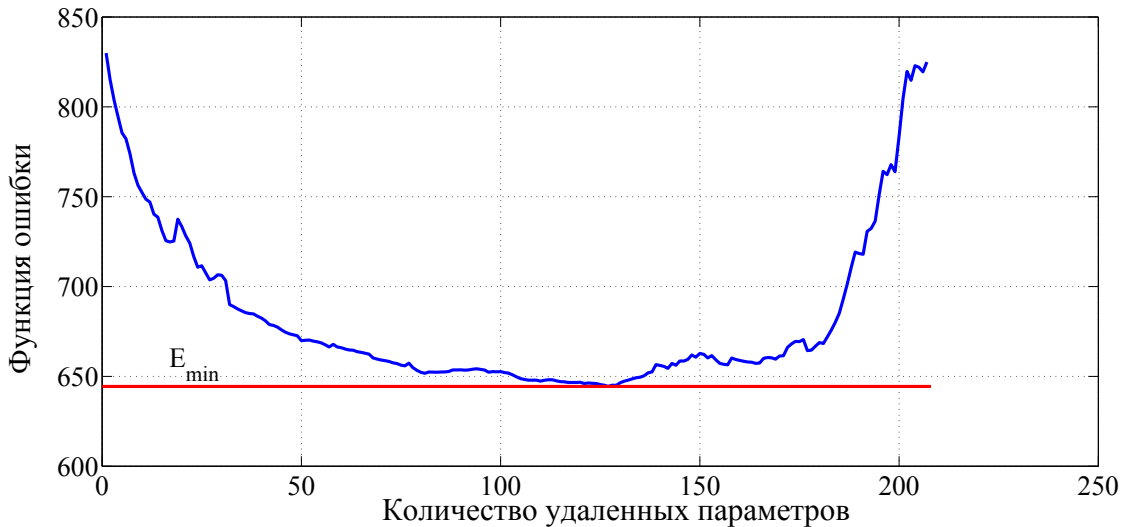


Рис. 1. Изменение функции ошибки при удалении параметров из модели

пор, пока процесс не стабилизируется. В качестве критерия стабилизации предлагается использовать энтропию

$$H(\mathcal{A}, \mathcal{A}') = - \sum_{j=1}^k \rho(a_j, a'_j) \ln(\rho(a_j, a'_j)), \quad (12)$$

множества попарных нормированных расстояний Хэмминга между элементами наборов $\mathcal{A} = \{a_1, \dots, a_k\}$ и $\mathcal{A}' = \{a'_1, \dots, a'_k\}$, полученных на двух последовательных итерациях алгоритма следующим образом:

$$a_j = \begin{cases} 1, & \text{если } w_j \neq 0, \\ 0, & \text{если } w_j = 0. \end{cases}$$

Процесс считается стабильным, если энтропия $H(\mathcal{A}, \mathcal{A}')$ не превосходит заданного порога.

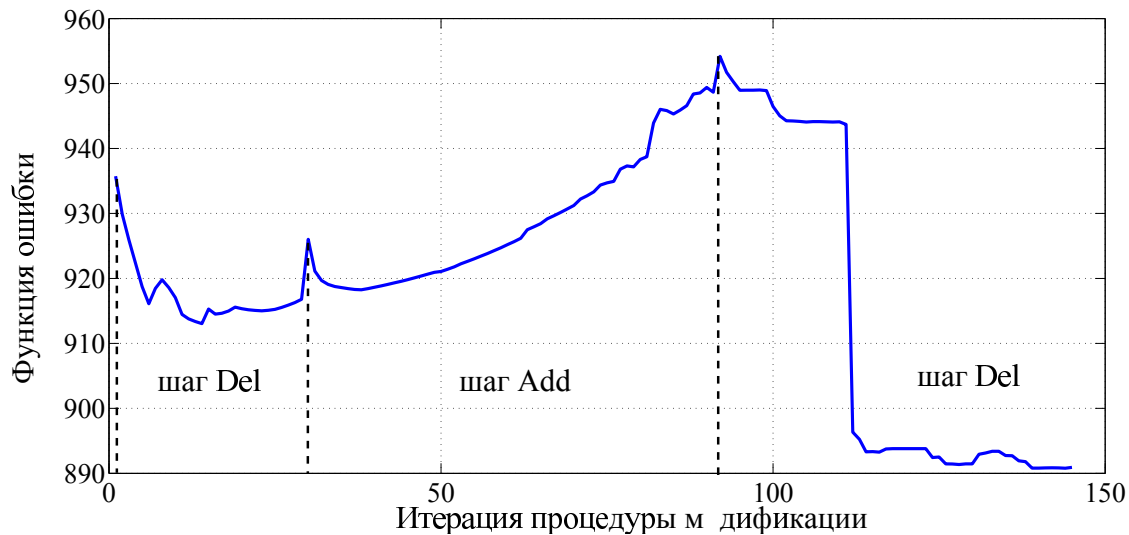


Рис. 2. Смена шагов Del и Add

4. Путь в k -мерном кубе

В данной задаче мы имеем дело с вектором параметров размерности k , что означает, что существует 2^k вариантов структур модели. Из этих 2^k возможных вариантов структур мы и выбираем оптимальные. Все варианты можно представить в виде вершин \mathbf{v} k -мерного куба \mathfrak{V} . И тогда стратегия задает путь \mathbf{V} по его вершинам. Этот путь заканчивается в некоторой вершине $\hat{\mathbf{v}}$, к которой сходится процедура модификации. Будем искать оптимальные модели в некоторой окрестности вершины $\hat{\mathbf{v}}$. Так как рассматривать все возможные варианты слишком трудоёмко, то в качестве окрестности $\hat{\mathbf{v}}$ мы будем рассматривать путь по вершинам куба, ведущий к $\hat{\mathbf{v}}$, полученный по описанной выше стратегии.

Пример 1. В этом примере использовалась выборка $\{\mathbf{x}_i, y_i\}$, $i \in \{1, \dots, 177\}$. Каждый объект выборки описывается 6 признаками χ_1, \dots, χ_6 и принадлежит одному из трех классов. Схематично взаимное расположение векторов χ_1, \dots, χ_6 изображено на рис 3. Для класси-

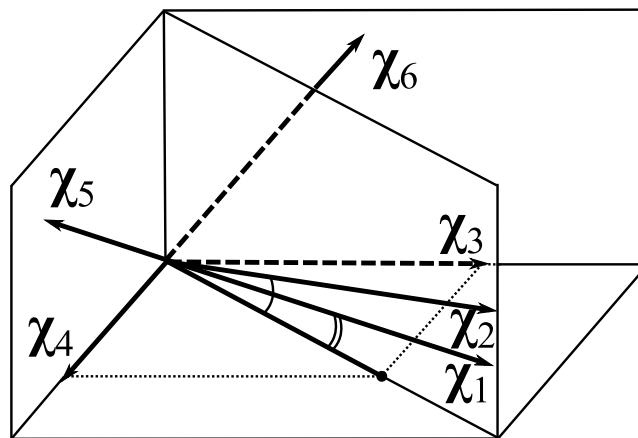


Рис. 3. Данные

фикации такой выборки модифицировалась двуслойная нейронная сеть с одним нейроном

в скрытом слое. Совокупное число параметров такой сети равно девяти. Нейронная сеть модифицировалась за 11 итераций. На рис. 4 изображен путь по вершинам девятимерного куба. По вертикали отложен номер параметра, по горизонтали — номер итерации. Черная клетка означает, что параметр с индексом j активный, белая клетка — параметр неактивный. Например, на пятой итерации из сети был удален параметр 9, а на одиннадцатой итерации этот параметр был снова добавлен в сеть.

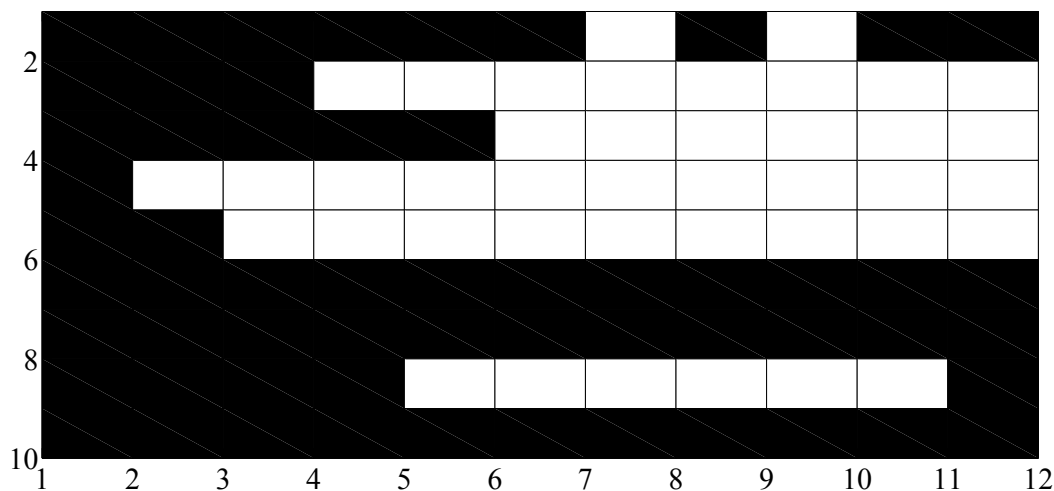


Рис. 4. Путь в кубе

5. Вычислительный эксперимент

С целью получить значение критериев качества описанной стратегии был проведен вычислительный эксперимент. Использовались данные акселерометра мобильного телефона. Данные состоят из 5418 векторов признаков, которые были получены в результате обработки соответствующих временных рядов. Было выделено 43 признака и 6 классов физической активности: ходьба, бег, сидение, стояние, подъем и спуск. Временные ряды записывались акселерометром мобильного телефона, который находился в кармане у человека, выполняющего один из типов физической активности. Для выделения признаков временные ряды разделялись на десятисекундные сегменты. Из этих сегментов извлекались признаки, такие как: проекции среднего ускорения на координатные оси, среднеквадратические отклонения от проекций среднего ускорения на каждую из трех координатных осей, время между пиками синусоидального сигнала в миллисекундах и др. С более подробным описанием признаков и процессом их генерации можно ознакомиться в [15].

В вычислительном эксперименте оптимизировалась двуслойная нейронная сеть с пятью нейронами в скрытом слое. Размерность вектора параметров такой модели $k = 245$. Нейронная сеть оптимизировалась по стратегии, описанной в главе 3. Был получен набор из 771 модели. В процедуре модификации использовались каждый из трех критериев прорезживания — оптимального, последовательного и устойчивого. Для всех моделей были вычислены численные значения критериев качества. Был построен Парето оптимальный фронт трех критериев. На Рис. 5 в координатах "устойчивость-сложность" изображены все полученные модели. Синим цветом обозначены модели, которые были получены по стратегии с применением критерия устойчивого прорезживания, зеленым цветом — критерия

последовательного прореживания, красным цветом — оптимального прореживания. Парето оптимальные модели обозначены черным крестом. Аналогично на Рис. 6 и Рис. 7 изображены все полученные модели в координатах "точность-сложность" и "точность-устойчивость" соответственно. Из Рис. 5 видно, что самые устойчивые модели получаются при использовании критерия устойчивого прореживания. В Таблице 1 приведены значения критериев качества моделей, которые являются точками останова процедуры модификации для каждого из трех критериев прореживания. На Рис. 8 приведена интерпретация

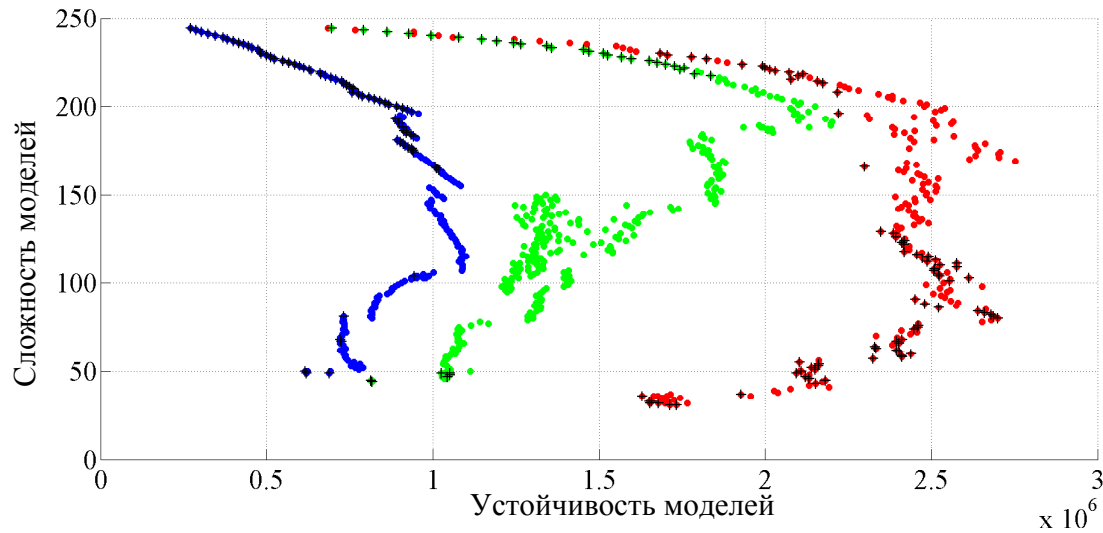


Рис. 5. Множество моделей в координатах устойчивость-сложность

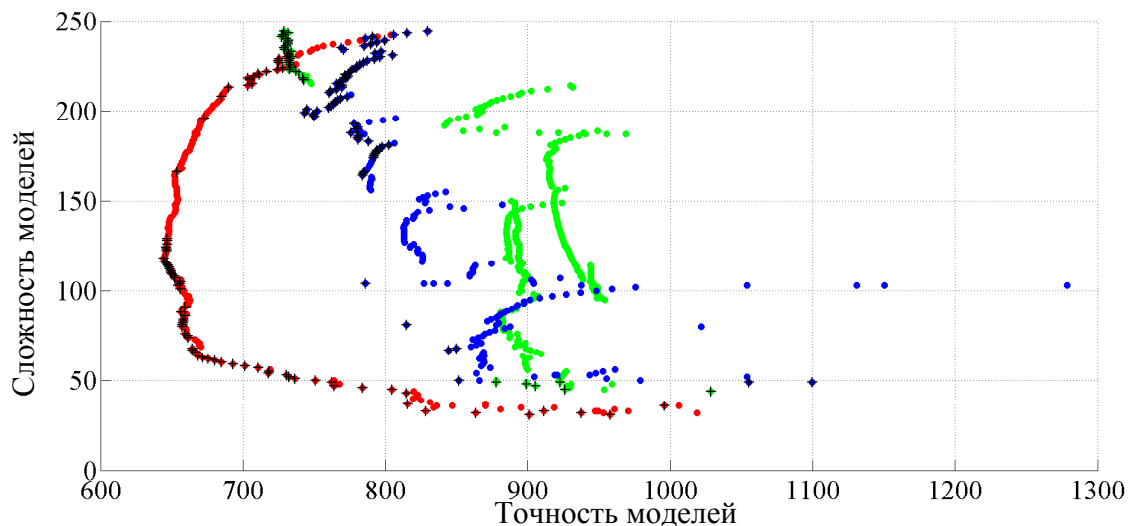


Рис. 6. Множество моделей в координатах точность-сложность

полученных результатов. В красной области графика Парето оптимальные модели неинтересны для рассмотрения, так как в этой области имеет место недообучение — модели излишне сложны. Парето оптимальные модели с незначительной сложностью находятся в зеленой области графика. Также была визуализирована процедура пошаговой модифи-

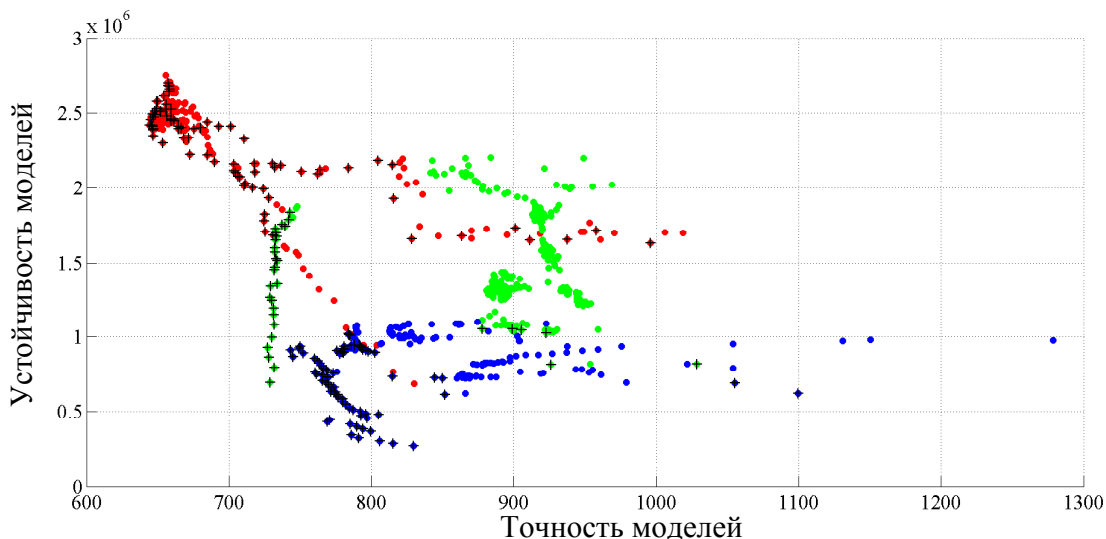


Рис. 7. Множество моделей в координатах точность-устойчивость

Таблица 1. Сложность, точность, устойчивость моделей

Стратегия	Сложность	Точность	Устойчивость
Оптимальное прореживание	50	877	$1.2 \cdot 10^6$
Последовательное прореживание	36	870	$2 \cdot 10^6$
Устойчивое прореживание	50	866	$6 \cdot 10^5$

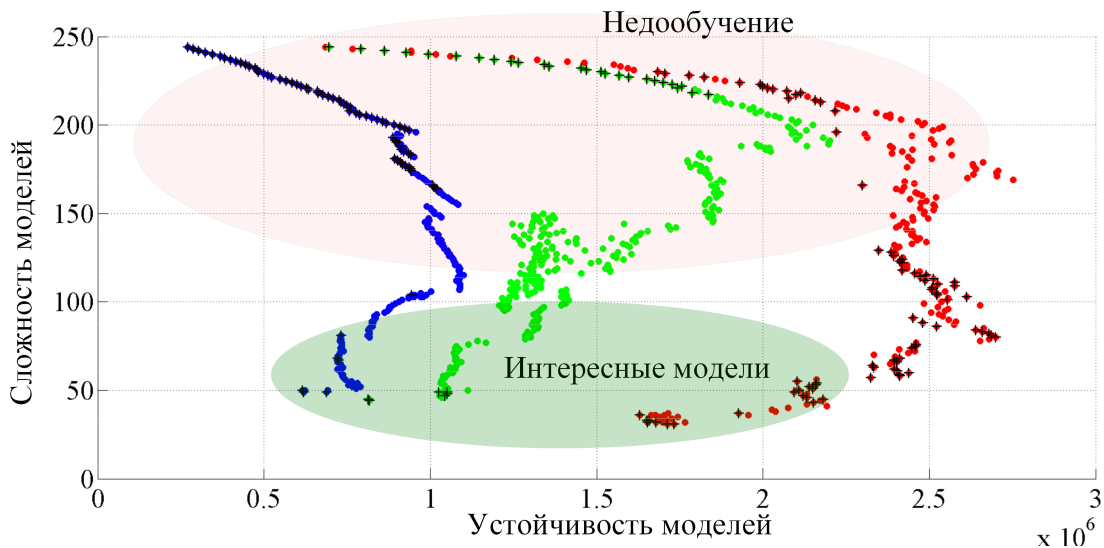


Рис. 8. Интерпретация результатов

каци модели как путь в k -мерном кубе. На Рис. 9, 10, 11 также как и в примере 1, по вертикали отложен номер параметра, по горизонтали — номер итерации. Черная клетка означает, что параметр активный, белая клетка — параметр неактивный. По Рис. 9, 10, 11 можно восстановить последовательность, в которой параметры удалялись из модели и добавлялись в неё. Из Рис. 10, 11 видно, что стратегия с критериями оптимального и

последовательного прореживания (которые выбирают для удаления параметр, минимизирующий функцию ошибки) оставляет в моделях параметры с номерами с 216 по 245. Это связано с тем, что параметры с такими номерами относятся ко второму слою нейронной сети, а удаление большого числа параметров второго слоя приводит к росту функции ошибки.

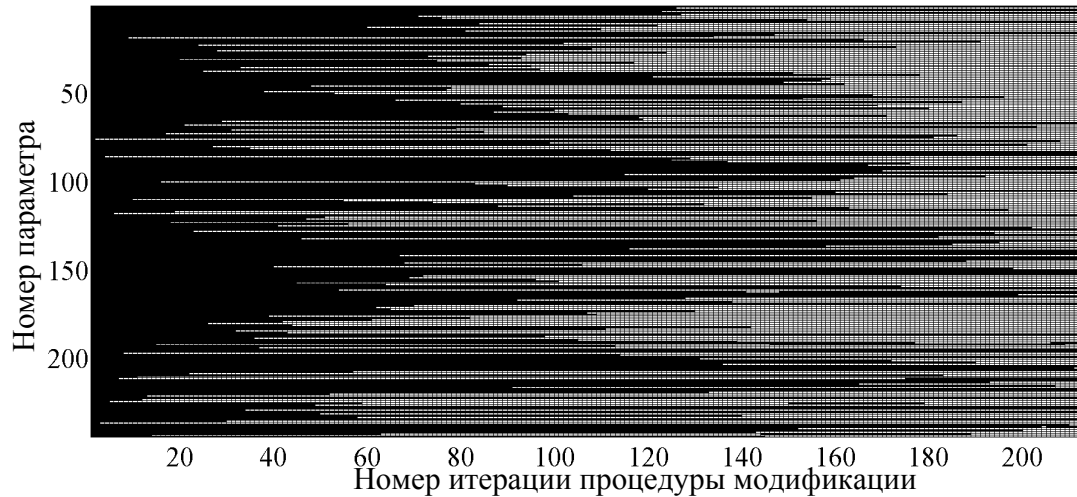


Рис. 9. Путь в кубе, устойчивое прореживание

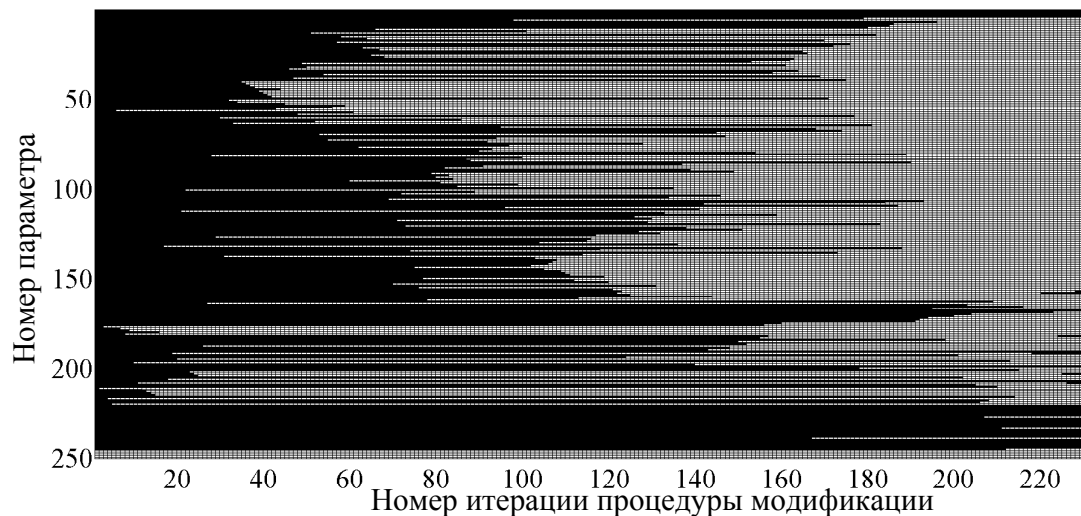


Рис. 10. Путь в кубе, последовательное прореживание

Заключение

В работе была предложена стратегия пошаговой модификации моделей классификации согласно трем критериям качества — сложность, точность и устойчивость. В рамках стратегии были предложены критерии добавления и удаления параметров в модель, критерии останова шагов добавления и удаления, а также критерий останова процедуры модификации. Процедура пошаговой модификации модели была рассмотрена и визуализирована

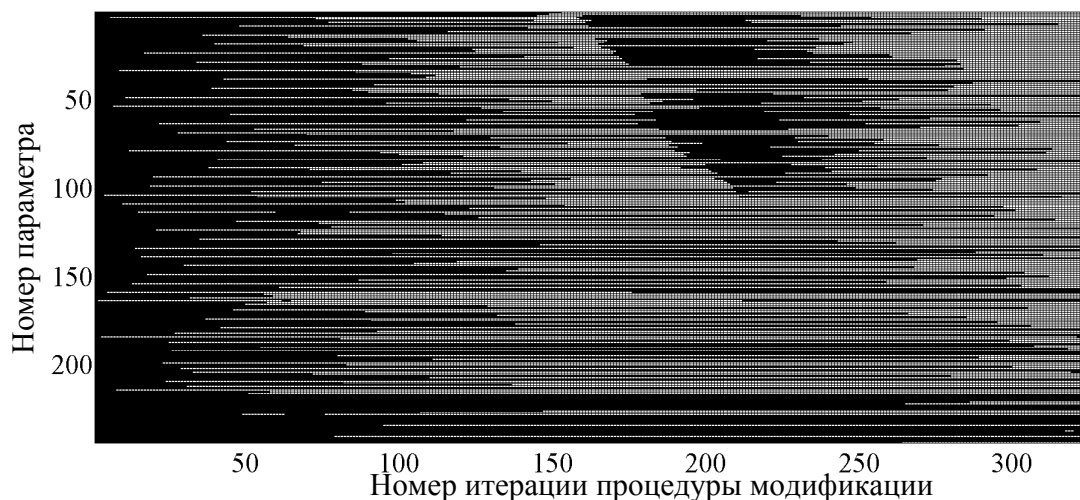


Рис. 11. Путь в кубе, оптимальное прореживание

как путь в многомерном кубе. Был проведен вычислительный эксперимент, в ходе которого был получен набор моделей и найден Парето оптимальный фронт критериев качества этого набора. Вычислительный эксперимент показал, что наилучшие по рассматриваемым критериям качества модели получаются при использовании критерия устойчивого прореживания. Это связано с тем, что критерий устойчивого прореживания позволяет получать более устойчивые модели, удаляя коррелирующие параметры и тем самым повышая устойчивость и обобщающую способность модели классификации. Программная реализация стратегии пошаговой модификации нейронной сети в среде разработки MatLab находится в свободном доступе по адресу [16].

Литература

- [1] *Myung I. J.* The Importance of Complexity in Model Selection // *Journal of Mathematical Psychology*, 2000. Vol. 44, No. 1. Pp. 190–204.
- [2] *MacLeod C., Maxwell M.* Incremental evolution in ANNs: Neural nets which grow // *Artificial Intelligence Review*, 2001. Vol. 16, No. 3. Pp. 201–224.
- [3] *Karnin E. D.* A simple procedure for pruning back-propagation trained neural networks // *IEEE Transactions on Neural Networks*, 1990. Vol. 1, No. 2. Pp. 239–242.
- [4] *LeCun Y., Denker L. S., Solla S. A.* Optimal Brain Damage // *Advances in neural information processing systems*, 1990. Vol. 2, No. 2. Pp. 598–605.
- [5] *Hassibi B., Stork D. G., Wolff G. J.* Optimal brain surgeon and general network pruning // *Proceedings of 1993 IEEE International Conference on Neural Networks*, 1993. Pp. 293–299.
- [6] *Yang S., Chen Y.* An evolutionary constructive and pruning algorithm for artificial neural networks and its prediction applications // *Neurocomputing*, 2012. Vol. 86, Pp. 140–149.
- [7] *Pu X., Pengfei Sun P.* A New Hybrid Pruning Neural Network Algorithm Based on Sensitivity Analysis for Stock Market Forecast // *Journal of Information and Computational Science*, 2013. Vol. 3, Pp. 883–892.
- [8] *Hong-Gui H., Qi-li C., Jun-Fei Q.* An efficient self-organizing RBF neural network for water quality prediction // *Neural Networks*, 2011. Vol. 24, No. 7. Pp. 717–725.

- [9] *Knerr S., Personnaz L., Dreyfus G.* Single-layer learning revisited: a stepwise procedure for building and training a neural network // *Neurocomputing: Algorithms, Architectures and Applications*, 1990. Vol. 68, No. 1. Pp. 41–50.
- [10] *Strijov V., Krymova E., Weber S. V.* Evidence optimization for consequently generated models // *Mathematical and Computer Modelling*, 2010. Vol. 57, No. 1-2. Pp. 50–56.
- [11] *Леонтьева Л. Н.* Последовательный выбор признаков при восстановлении регрессии // *Машинное обучение и анализ данных*, 2012. Т. 1, № 3. С. 335–346.
- [12] *Сандуляну Л. Н., Стрижов В. В.* Выбор признаков в авторегрессионных задачах прогнозирования // *Информационные технологии*, 2012. Т. 7, С. 11–15.
- [13] *Зайцев А. А., Токмакова А. А.* Оценка гиперпараметров линейных регрессионных моделей методом максимального правдоподобия при отборе шумовых и коррелирующих признаков // *Машинное обучение и анализ данных*, 2012. Т. 1, № 3. С. 347–353.
- [14] *Токмакова А. А., Стрижов В. В.* Оценивание гиперпараметров линейных и регрессионных моделей при отборе шумовых и коррелирующих признаков // *Информатика и её применения*, 2012. Т. 6, № 4. С. 66–75.
- [15] *Kwapisz J. R., Weiss G. M., Moore S.* Activity recognition using cell phone accelerometers // *SIGKDD Explorations*, 2010. Vol. 12, No. 2. Pp. 74–82.
- [16] *Попова М. С.* 2014. Реализация стратегии пошаговой модификации нейронной сети. <http://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group174/Popova2014OptimalModelSelection/code/main.m>

References

- [1] Myung, I. J. 2000. The Importance of Complexity in Model Selection. *Journal of Mathematical Psychology*, 44(1):190–204.
- [2] MacLeod, C., and Maxwell, M. 2001. Incremental evolution in ANNs: Neural nets which grow. *Artificial Intelligence Review*, 16(3):201–224.
- [3] Karnin, E. D. 1990. A simple procedure for pruning back-propagation trained neural networks. *IEEE Transactions on Neural Networks*, 1(2):239–242.
- [4] LeCun, Y., Denker, L. S., and Solla, S. A. 1990. Optimal Brain Damage. *Advances in neural information processing systems*, 2(2):598–605.
- [5] Hassibi, B., Stork, D. G., and Woff G J. 1993. Optimal brain surgeon and general network pruning. *Proceedings of 1993 IEEE International Conference on Neural Networks*. 293–299.
- [6] Yang, S., and Chen, Y. 2012. An evolutionary constructive and pruning algorithm for artificial neural networks and its prediction applications. *Neurocomputing*, 86(1):140–149.
- [7] Pu, X., and Pengfei-Sun, P. 2013. A New Hybrid Pruning Neural Network Algorithm Based on Sensitivity Analysis for Stock Market Forecast. *Journal of Information and Computational Science*, 3(1):883–892.
- [8] Hong-Gui, H., Qi-li, C., and Jun-Fei, Q. 2011. An efficient self-organizing RBF neural network for water quality prediction. *Neural Networks*, 24(7):717–725.
- [9] Knerr, S., Personnaz, L., and Dreyfus, G. 1990. Single-layer learning revisited: a stepwise procedure for building and training a neural network. *Neurocomputing: Algorithms, Architectures and Applications*, 68(1):41–50.
- [10] Leont’eva, L. N. 2012. Posledovatel’nyj vybor priznakov pri vosstanovlenii regressii [Sequential selection of features when restoring regression]. *Mashinnoe obuchenie i analiz dannyh* [J. Machine Learning and Data Analysis], 1(3):335–346.

- [11] Sanduljanu, L. N., Strizhov, V. V. 2012. Vybory priznakov v avtoregressionnykh zadachah prognozirovaniya [Feature selection in autoregressive prediction problems]. *Informacionnye tehnologii [Information Technology]*, 7:11–15.
- [12] Strijov, V., Krymova, E., and Weber, S. V. 2013. Evidence optimization for consequently generated models. *Mathematical and Computer Modelling*, 57(1-2):50–56.
- [13] Zajcev, A. A., and Tokmakova, A. A. 2012. Ocenka giperparametrov linejnykh regressionnykh modelej metodom maksimal'nogo pravdopodobija pri otbore shumovyh i korrelirovannykh priznakov [Estimation of the hyperparameters of linear and regression models for the selection of the noise and correlated features]. *Mashinnoe obuchenie i analiz dannyh [J. Machine Learning and Data Analysis]*, 1(3):347–353.
- [14] Tokmakova, A. A., Strizhov, V. V. 2012. Ocenivanie giperparametrov linejnykh i regressionnykh modelej pri otbore shumovyh i korrelirovannykh priznakov [Estimation of the hyperparameters of linear and regression models for the selection of the noise and correlated features]. *Informatika i ejo primenenija [Informatics and Applications]*, 6(4):66–75.
- [15] Kwapisz, J. R., Weiss, G. M., and Moore, S. 2010. Activity recognition using cell phone accelerometers. *SIGKDD Explorations*, 12(2):74–82.
- [16] Popova, M. S. 2014. Realizacija strategii poshagovoj modifikacii nejronnoj seti [Realization of a stepwise strategy for neural network modification]. <http://sourceforge.net/p/mlalgorithms/code/HEAD/tree/Group174/Popova2014OptimalModelSelection/code/main.m>