Оптимизационные методы линейной и монотонной коррекции в алгебраическом подходе к проблеме распознавания*

К. В. Воронцов

Вычислительный центр РАН, Москва

Описан оптимизационный метод построения корректных алгоритмов на основе технологии алгебраического подхода к проблеме распознавания. Рассмотрено применение данного метода в случаях линейных и монотонных корректирующих операций при решении задач классификации и восстановления регрессии. Предложены методы решения оптимизационных задач, поставленных в работе [1].

Алгебраический подход к проблеме распознавания, развиваемый школой академика РАН Ю. И. Журавлёва, основан на идее построения корректных алгоритмов с помощью корректирующих операций над несколькими эвристическими алгоритмами. При практической реализации этой идеи возникает необходимость уменьшения сложности получаемой алгоритмической суперпозиции. В данной работе это достигается путём использования методов оптимизации для настройки (выбора параметров) алгоритмических операторов и корректирующей операции. Данная техника изучалась ранее в работе [1] для семейства монотонных корректирующих операций. Являясь фактически продолжением [1], настоящая работа преследует две основные цели.

Во-первых, предлагается общая методология настройки алгоритмических операторов, при которой учитывается не только исходная обучающая выборка, но и структура суперпозиции. При этом техника настройки основывается на сведении данной задачи к уже известным, что позволяет избежать разработки новых специальных методов оптимизации для каждой из используемых моделей алгоритмических операторов. Рассматриваются также комбинированные стратегии настройки, совмещающие настройку на обучающую выборку и компенсацию совокупной ошибки остальных алгоритмических операторов. Использование данной техники иллюстрируется на примере линейных и монотонных корректирующих операций.

---

*Работа выполнена при финансовой поддержке Российского фонда фундаментальных исследований (код проекта 96-01-00552)

Во-вторых, рассмотрение семейства монотонных корректирующих операций доводится здесь до конца и завершается построением эффективных численных методов, которые остались за рамками работы [1]. Описываются методы построения монотонных корректирующих операций, удовлетворяющих дополнительным требованиям непрерывности и гладкости.

1 Оптимизационный метод синтеза корректных алгоритмов

1. Рассмотрим следующую постановку задачи распознавания [2, 3]. Имеется множество начальных информаций $\mathfrak{I}_i$ и множество финальных информаций $\mathfrak{I}_f$. Требуется построить алгоритм, реализующий отображение из $\mathfrak{I}_i$ в $\mathfrak{I}_f$, удовлетворяющее локальным ограничениям вида $A(x_k) = y_k$, $k = 1, \ldots, q$, и дополнительным ограничениям вида $A \in \mathfrak{M}^u$, где $\{x_k\}_{k=1}^q$ — последовательность элементов множества $\mathfrak{I}_i$, $\{y_k\}_{k=1}^q$ — последовательность элементов множества $\mathfrak{I}_f$ и $\mathfrak{M}^u$ — заданное множество отображений из $\mathfrak{I}_i$ в $\mathfrak{I}_f$. Алгоритм, удовлетворяющий локальным и дополнительным ограничениям, называется *корректным.* Последовательность пар $\{x_k, y_k\}_{k=1}^q$ называется *обучающей выборкой.*

Итак, рассматриваемые задачи распознавания полностью определяются пятёркой $\langle \mathfrak{I}_i, \mathfrak{I}_f, \mathfrak{M}^u, \{x_k\}_{k=1}^q, \{y_k\}_{k=1}^q \rangle$. Мы также будем рассматривать два распространённых частных случая: задачи классификации с двумя непересекающимися классами ($\mathfrak{I}_f = \{0, 1\}$) и задачи восстановления регрессии ($\mathfrak{I}_f = \mathbb{R}$).

2. Алгебраический подход к проблеме распознавания основан на следующей схеме синтеза корректных алгоритмов [2, 3]. Наряду с множествами $\mathfrak{I}_i$ и $\mathfrak{I}_f$ вводится пространство оценок $\mathfrak{I}_e$. Затем выбирается модель алгоритмических операторов $\mathfrak{M}^0 \subseteq \mathfrak{M}^0_* = \{B : \mathfrak{I}_i \to \mathfrak{I}_e\}$, семейство решающих правил $\mathfrak{M}^1 \subseteq \bigcup\limits_{p=0}^{\infty} \{C : \mathfrak{I}_e^p \to \mathfrak{I}_f\}$ и семейство корректирующих операций $\mathfrak{F} \subseteq \bigcup\limits_{p=0}^{\infty} \{F : \mathfrak{I}_e^p \to \mathfrak{I}_e\}$. Все три семейства отображений строятся таким образом, чтобы $\mathfrak{M}^1 \circ \mathfrak{F}(\mathfrak{M}^0) \subseteq \mathfrak{M}^u$. Тем самым гарантируется, что все алгоритмы удовлетворяют дополнительным ограничениям «по построению». Общие подходы к построению $\mathfrak{F}$-расширений моделей алгоритмов развиваются в теории универсальных и локальных ограничений [4, 5, 6].

Заметим, что суперпозиция $B(x) \equiv F(B_1(x), \ldots, B_p(x))$ осуществляет отображение из $\mathfrak{I}_i$ в $\mathfrak{I}_e$, и поэтому также является алгоритмическим оператором, который будет обозначаться $F(B_1, \ldots, B_p)$. Далее рассматривается случай, когда алгоритм A строится в виде суперпозиции алгоритмических операторов $B_1, \ldots, B_p$ из $\mathfrak{M}^0$, корректирующей операции F из $\mathfrak{F}$ и решающего правила C из $\mathfrak{M}^1$, то есть $A = C(F(B_1, \ldots, B_p))$.

3. Пусть задан функционал качества алгоритмических операторов $Q\colon \mathfrak{M}^0_* \to \mathbb{R}$, принимающий нулевое значение $Q(B) = 0$ тогда и только тогда, когда существует решающее правило C из $\mathfrak{M}^1$, при котором алгоритм $A = C \circ B$ корректен. В этом случае задача синтеза корректного алгоритма сводится к тому, чтобы найти алгоритмические операторы $B_1, \dots, B_p$ из $\mathfrak{M}^0$ и корректирующую операцию F из $\mathfrak{F}$, при которых достигается минимум функционала $Q(F(B_1, \dots, B_p))$. Для решения поставленной задачи предлагается общий итерационный метод, основанный на поочерёдной настройке алгоритмических операторов и корректирующей операции, который состоит в следующем.

В качестве нулевого приближения выбирается оператор B_1 из $\mathfrak{M}^0$, найденный путём минимизации функционала $Q(B)$. Следующие операторы B_2, $B_3, \dots$ строятся по очереди, причём после добавления очередного оператора производится повторная оптимизация ранее построенных операторов и корректирующей операции. При этом на каждом шаге итерационного процесса решается одна из двух задач: функционал качества минимизируется либо по оператору B_r при фиксированных F и $B_1, \dots, B_{r-1}$, $B_{r+1}, \dots, B_p$:

$$B_r^* = \underset{B_r \in \mathfrak{M}^0}{\arg\min}\, Q(F(B_1, \dots, B_r, \dots, B_p)), \quad r = 1, \dots, p, \tag{1.1}$$

либо по корректирующей операции F при фиксированных $B_1, \dots, B_p$:

$$F^* = \underset{F \in \mathfrak{F}}{\arg\min}\, Q(F(B_1, \dots, B_p)). \tag{1.2}$$

Описанный процесс представляет собой вариант покоординатного спуска с тем отличием, что определение каждой «координаты» $B_1, \dots, B_p$ и F требует решения отдельной, как правило многопараметрической, оптимизационной задачи.

Предлагаемый подход к минимизации функционала (1.1) состоит в том, чтобы приспособить для этого методы, изначально предназначенные для решения более простой задачи

$$B_r^* = \underset{B_r \in \mathfrak{M}^0}{\arg\min}\, Q(B_r). \tag{1.3}$$

Возможность такого сведения вытекает из следующих общих соображений. При построении корректирующей операции обычно предполагается, что каждый из операторов B_r предварительно настроен на обучающую выборку, скажем, в результате решения задачи вида (1.3) или аналогичной ей. В то же время при настройке оператора B_r должна быть сделана поправка, учитывающая его место в суперпозиции $F(B_1, \dots, B_p)$, то есть влияние остальных операторов и корректирующей операции. Способ технической реализации такой поправки в общем случае зависит от семейства $\mathfrak{F}$ и множеств $\mathfrak{I}_e$ и $\mathfrak{I}_f$. Однако в конечном итоге её введение сводится либо к настройке на компенсацию совокупной ошибки остальных операторов, либо к более точной настройке на выделенном подмножестве объектов при возможном увеличении погрешностей на других объектах обучения. Типичными способами реализации подобных дополнительных требований являются: введение весовых коэффициентов объектов обучения, исключение некоторого подмножества объектов из обучающей выборки и модификация целевого вектора $[y_k]_{k=1}^q$. Большинство из используемых моделей алгоритмических операторов и методов их оптимизации допускают введение таких поправок.

Указанный факт позволяет не рассматривать конкретные методы оптимизации (1.1) для отдельных моделей алгоритмических операторов, а ограничиться получением постановок задач, аналогичных тем, которые возникают при решении задачи (1.3). При таком сведении корректировка численных методов либо вообще не требуется, либо является несложным техническим упражнением.

Наличие пары оптимизационных задач (1.1) и (1.3), для решения которых подходит один и тот же численный метод, естественно приводит к комбинированной постановке единой задачи оптимизации. Степень близости полученной задачи к первой или второй можно параметризовать числовым параметром $\lambda \in [0,1]$. Один из способов параметризации состоит в построении функционала качества

$$Q_\lambda(B_r) = (1-\lambda)\,Q(B_r) + \lambda\,Q(F(B_1,\ldots,B_r,\ldots,B_p)). \tag{1.4}$$

Другой способ основан на более детальном рассмотрении постановок задач, к которым сводятся (1.1) и (1.3). Допустим, что они принадлежат одному классу задач и характеризуются векторами параметров α_1 и α_2 соответственно. Если множество допустимых значений параметров является выпуклым множеством в линейном векторном пространстве, то задача с вектором параметров $(1-\lambda)\alpha_2 + \lambda\alpha_1$ также принадлежит данному классу, и её решение может рассматриваться в качестве компромиссной стратегии настройки.

От способа параметризации в общем случае требуется только, чтобы при граничных значениях параметра λ реализовались «чистые» стратегии настройки, а именно:

— при $\lambda = 0$ должна решаться задача (1.3), состоящая в настройке алгоритмического оператора B_r на исходную обучающую выборку без учёта его дальнейшего использования в качестве аргумента корректирующей операции;

— при $\lambda = 1$ должна решаться задача (1.1), в которой оператор B_r настраивается исключительно на компенсацию неточностей, допущенных остальными операторами.

Каждая из этих двух «чистых» стратегий настройки имеет свои недостатки.

В первом случае, многократно решая одну и ту же задачу, мы получим одинаковые операторы $B_1,\ldots,B_p$. Даже при варьировании метода настройки нет никакой гарантии, что мы не получим набор из почти одинаковых операторов. Тогда никакая корректирующая операция не позволит построить алгоритм, качественно отличающийся от уже имеющихся. Это приведёт либо к неоправданному наращиванию числа операторов, либо вообще к отказу от идеи коррекции.

Во втором случае модель операторов $\mathfrak{M}^0$ используется для компенсации суммарной ошибки остальных операторов, но не для аппроксимации зависимости между начальными и финальными информациями. Подобное использование представляется нецелесообразным для проблемно-ориентированных моделей, структура которых «по построению» учитывает специфические особенности этой зависимости.

При промежуточных значениях параметра λ образуются комбинированные стратегии настройки, приводящие к компромиссу между настройкой на обучающую выборку и компенсацией совокупной ошибки остальных операторов. Появление дополнительного параметра повышает гибкость и настраиваемость суперпозиции в целом. На практике этот параметр назначается из априорных соображений, учитывающих особенности выбранной модели $\mathfrak{M}^0$ и семейства $\mathfrak{F}$, их универсальность и адекватность решаемой задаче. Приемлемое значение этого параметра предлагается подбирать в ходе вычислительных экспериментов.

2 Линейные корректирующие операции

Возьмём в качестве пространства оценок $\mathfrak{I}_e$ множество действительных чисел $\mathbb{R}$ и определим семейство линейных корректирующих операций:

$$\mathfrak{F}_L = \left\{ F(B_1, \ldots, B_p) = \sum_{i=1}^{p} \alpha_i B_i \;\middle|\; \alpha_i \in \mathbb{R},\ i = 1, \ldots, p \right\}.$$

Каждая линейная корректирующая операция однозначно задаётся набором из p параметров $\alpha_1, \ldots, \alpha_p$. Иногда на параметры накладывают дополнительные ограничения, например требуют их неотрицательности, и/или чтобы сумма их модулей равнялась единице. Линейные корректирующие операции над множествами некорректных алгоритмов были впервые введены Ю. И. Журавлёвым [2, 3] при рассмотрении задачи классификации.

1. Рассмотрим сначала использование семейства линейных корректирующих операций для решения задачи классификации с двумя непересекающимися классами, $\mathfrak{I}_f = \{0, 1\}$. Возьмём в качестве $\mathfrak{M}^1$ однопараметрическое семейство пороговых решающих правил $\{C(B) = \theta(B - c) \mid c \in \mathbb{R}\}$, где θ — функция Хэвисайда.

Положим функционал качества Q равным числу ошибочных классификаций объектов обучения при условии оптимального выбора решающего правила:

$$Q(B) = \min_{c \in \mathbb{R}} \sum_{k=1}^{q} |\theta(B(x_k) - c) - y_k|.$$

Перенумеруем элементы обучающей выборки $\{x_k, y_k\}_{k=1}^{q}$ таким образом, чтобы первые q_0 элементов принадлежали первому классу, а остальные $(q - q_0)$ элементов — второму классу. Таким образом, $y_1 = \ldots = y_{q_0} = 0$ и $y_{q_0+1} = \ldots = y_q = 1$. Легко видеть, что значение функционала равно наименьшему числу невыполненных неравенств в системе

$$\begin{cases} B(x_k) \leqslant c, & k = 1, \ldots, q_0; \\ B(x_k) > c, & k = q_0 + 1, \ldots, q; \end{cases}$$

Допустим без ограничения общности, что в наборе алгоритмических операторов $B_1, \ldots, B_p$ все операторы кроме последнего фиксированы и решается задача оптимизации оператора B_p. Введём матричные обозначения:

$$\begin{array}{ll} Z_0 = [B_i(x_k)]_{k=1,q_0}^{i=1,p-1}, & Z_1 = [B_i(x_k)]_{k=q_0+1,q}^{i=1,p-1}, \\ z_0 = [B_p(x_k)]_{k=1,q_0}, & z_1 = [B_p(x_k)]_{k=q_0+1,q}, \\ a = [\alpha_i]_{i=1,p-1}. & \end{array}$$

Тогда значение функционала $Q(B_p)$ равно наименьшему числу невыполненных неравенств в системе

$$\begin{cases} z_0 \leqslant c; \\ z_1 > c; \end{cases} \tag{2.1}$$

а значение функционала $Q(F(B_1, \ldots, B_p))$ — наименьшему числу невыполненных неравенств в системе

$$\begin{cases} Z_0 a + z_0 \alpha_p \leqslant c; \\ Z_1 a + z_1 \alpha_p > c. \end{cases} \tag{2.2}$$

Задача (1.2) минимизации функционала $Q(F(B_1, \ldots, B_p))$ по корректирующей операции F при фиксированных алгоритмических операторах сводится к поиску максимальной совместной подсистемы системы линейных неравенств (2.2) с неизвестными a, α_p и c. Для её решения можно использовать стандартные методы [7, 8, 9].

Задача минимизации этого же функционала по оператору B_p при фиксированных $B_1, \ldots, B_{p-1}$ и F сводится к поиску максимальной совместной подсистемы системы неравенств

$$\begin{cases} z_0 \leqslant (c - Z_0 a)/\alpha_p; \\ z_1 > (c - Z_1 a)/\alpha_p; \end{cases} \quad \text{либо} \quad \begin{cases} z_0 \geqslant (c - Z_0 a)/\alpha_p; \\ z_1 < (c - Z_1 a)/\alpha_p; \end{cases}$$

при $\alpha_p > 0$ и $\alpha_p < 0$ соответственно. Данная задача формально эквивалентна задаче (2.1).

Построим комбинированный критерий настройки при произвольном $\lambda \in [0, 1]$. Выберем $[\lambda q]$ номеров $k_1, \ldots, k_{[\lambda q]}$ из множества $\{1, \ldots, q\}$ и потребуем, чтобы для объектов обучающей выборки $x_{k_1}, \ldots, x_{k_{[\lambda q]}}$ выполнялись соответствующие им неравенства системы (2.2), а для остальных элементов — неравенства системы (2.1).

При $\lambda = 0$ получаем систему (2.1), приводящую к настройке оператора B_p на исходную обучающую выборку. При $\lambda = 1$ получаем систему (2.2), приводящую к настройке на компенсацию неточности остальных операторов. При промежуточных значениях λ имеем всевозможные комбинированные стратегии настройки. Использование такой стратегии сводится к выбору подмножества точек обучающей выборки, на которых желательно достичь лучшего качества распознавания. Методом решения, так же как для (2.1) и (2.2), по-прежнему остаётся поиск максимальной совместной подсистемы.

2. Рассмотрим теперь использование семейства линейных корректирующих операций для решения задачи восстановления регрессии, $\mathfrak{I}_f = \mathbb{R}$.

Возьмём в качестве семейства решающих правил $\mathfrak{M}^1$ множество, состоящее из единственного тождественного отображения $C(B) \equiv B$. Фактически это означает, что решающие правила не используются.

Положим функционал качества Q равным среднеквадратичной невязке:

$$Q(B) = \sum_{k=1}^{q} (B(x_k) - y_k)^2.$$

Допустим без ограничения общности, что в наборе алгоритмических операторов $B_1, \dots, B_p$ все операторы, кроме последнего, фиксированы, и решается задача оптимизации оператора B_p. Введём матричные обозначения: $Z = [B_i(x_k)]_{k=1,q}^{i=1,p-1}$, $z = [B_p(x_k)]_{k=1,q}$, $a = [\alpha_i]_{i=1,p-1}$, $y = [y_k]_{k=1,q}$. Тогда

$$Q(B_p) = \|z - y\|^2; \tag{2.3}$$

$$Q(F(B_1, \dots, B_p)) = \|Za + z\alpha_p - y\|^2. \tag{2.4}$$

Задача (1.2) минимизации функционала $Q(F(B_1, \dots, B_p))$ по корректирующей операции F при фиксированных алгоритмических операторах сводится к нахождению вектора (a, α_p) и может быть решена методом наименьших квадратов [10].

Рассмотрим задачу минимизации этого функционала по оператору B_p при фиксированных $B_1, \dots, B_{p-1}$ и F. Путём элементарных преобразований он приводится к виду

$$Q(F(B_1, \dots, B_p)) = \alpha_p^2 \, \|z - \widetilde{y}\|^2, \quad \text{где} \quad \widetilde{y} = (y - Za)/\alpha_p.$$

Для произвольного $\lambda \in [0, 1]$ определим функционал Q_λ как взвешенную сумму функционалов (2.3) и (2.4):

$$Q_\lambda(B_p) = (1 - \lambda) \, \|z - y\|^2 + \lambda \alpha_p^2 \, \|z - \widetilde{y}\|^2.$$

Приводя подобные, получаем, что задача его минимизации эквивалентна задаче минимизации функционала

$$Q'_\lambda(B_p) = \|z - y'(\lambda)\|^2, \quad \text{где} \quad y'(\lambda) = \frac{(1 - \lambda) y + \lambda \alpha_p^2 \widetilde{y}}{1 - \lambda + \lambda \alpha_p^2}. \tag{2.5}$$

При $\lambda = 0$ оператор B_p настраивается на исходную обучающую выборку, при $\lambda = 1$ — на компенсацию неточности остальных операторов. При промежуточных значениях λ получаются комбинированные стратегии настройки. Задачи минимизации трёх функционалов (2.3), (2.4) и (2.5) формально эквивалентны и отличаются только целевым вектором: y, $\widetilde{y}$, $y'(\lambda)$.

3 Монотонные корректирующие операции

Пусть $\mathfrak{I}_f$ и $\mathfrak{I}_e$ — частично упорядоченные множества. В качестве $\mathfrak{F}$ возьмём семейство всех монотонных отображений из $\mathfrak{I}_e^p$ в $\mathfrak{I}_f$ при произвольном натуральном p, что позволит нам избежать отдельного рассмотрения решающих правил. Отношение порядка на $\mathfrak{I}_e^p$ и монотонность отображений понимаются в общепринятом смысле.

1. Обозначим множество индексов $\{1,\dots,q\}$ через $\mathbb{Q}$ и введём последовательность $\{a_k\}_{k=1}^q$, состоящую из q векторов $a_k = [B_i(x_k)]_{i=1,p}$, $k \in \mathbb{Q}$.

Определение 1. Пара индексов $(j,k) \in \mathbb{Q}^2$ называется *дефектной парой* алгоритмического оператора B, если $y_j < y_k$ и $B(x_j) \geqslant B(x_k)$. Множество всех дефектных пар оператора B обозначим через $\mathbb{D}(B)$.

Число дефектных пар можно использовать как функционал качества, поскольку $\mathbb{D}(B) = \varnothing$ тогда и только тогда, когда существует монотонное отображение F такое, что $F(B(x_k)) = y_k$ для всех k из $\mathbb{Q}$. Положим $Q(B) = |\mathbb{D}(B)|$.

Определение 2. Множество $\mathbb{D}(B_1,\dots,B_p) = \mathbb{D}(B_1) \cap \ldots \cap \mathbb{D}(B_p)$ называется *дефектом* набора операторов $B_1,\dots,B_p$.

Определение 3. Тройка индексов $(j,s,k) \in \mathbb{Q}^3$ называется *дефектной тройкой* для набора алгоритмических операторов $B_1,\dots,B_p$, если:

(а) пара (j,k) принадлежит дефекту $\mathbb{D}(B_1,\dots,B_p)$;

(б) вектор a_s несравним с a_j и a_k;

(в) выполнена цепочка неравенств $y_j \leqslant y_s \leqslant y_k$.

Дефектная тройка (j,s,k) называется *строго дефектной*, если $y_j < y_s < y_k$. Пара (j,k) называется *основанием* дефектной тройки (j,s,k),

В работе [1] показано, что задача (1.3) настройки оператора B_r сводится к поиску максимальной совместной подсистемы в системе неравенств

$$B_p(x_j) < B_p(x_k) \quad \text{для всех } (j,k)\colon y_j < y_k, \tag{3.1}$$

а задача (1.1) — к поиску совместной подсистемы максимального веса в системе неравенств

$$B_p(x_j) < B_p(x_k) \quad \text{с весом } w_{jk} \quad \text{для всех } (j,k) \in \mathbb{D}(B_1,\dots,B_{p-1}), \tag{3.2}$$

где вес подсистемы определяется как сумма весов входящих в неё неравенств. Веса отдельных неравенств вычисляются по формуле $w_{jk} = t_{jk}^0 + \frac{1}{2}t_{jk} + 1$, где t_{jk} и

t^0_{jk} — соответственно число дефектных и строго дефектных троек с основанием (j,k) для набора операторов $(B_1,\dots,B_{p-1})$. Нашей ближайшей целью будет рассмотрение методов решения задач (3.1) и (3.2), которые остались за рамками работы [1].

Основной недостаток полученных постановок в том, что они не являются стандартными для задач распознавания. Обычно каждое ограничение на оператор B_p относится только к одному объекту обучения, а число ограничений имеет порядок q. В случае монотонных корректирующих операций каждое ограничение относится к паре объектов. Число ограничений получается порядка q^2, причём ограничения очевидным образом зависимы: выполнение условия (3.1) для пар (j,k) и (k,s) по транзитивности влечёт их выполнение для пары (j,s). При построении эффективного численного метода этот факт либо должен учитываться явно, либо система (3.1) должна быть сведена к системе ограничений, относящихся к отдельным объектам, а не парам объектов. Именно второго подхода будем придерживаться в дальнейшем.

2. Рассмотрим сначала задачу классификации, для которой такое сведéние оказывается наиболее естественным. Положим $\mathfrak{I}_f=\{0,1\}$, $\mathfrak{I}_e=\mathbb{R}$.

Каждое из неравенств (3.1) должно выполняться при условии $y_j<y_k$, откуда следует $y_j=0$ и $y_k=1$. В результате каждое неравенство системы (3.1) распадается на два: $B_p(x_j)\leqslant c$ и $B_p(x_k)>c$, где c — заданная пороговая константа.

Введём следующие обозначения:

$$\begin{aligned}
I_0 &= \{k\in\mathbb{Q}\mid y_k=0\},\\
I_1 &= \{k\in\mathbb{Q}\mid y_k=1\},\\
D_0^p &= \{k\in\mathbb{Q}\mid \exists j\ (k,j)\in\mathbb{D}(B_1,\dots,B_p)\},\\
D_1^p &= \{k\in\mathbb{Q}\mid \exists j\ (j,k)\in\mathbb{D}(B_1,\dots,B_p)\}.
\end{aligned}$$

Множества индексов D_0^p и D_1^p перечисляют все объекты, образующие дефект и принадлежащие первому и второму классам соответственно. Очевидно, что $D_0^{p-1}\subseteq I_0$ и $D_1^{p-1}\subseteq I_1$.

Задача минимизации $Q(B_p)$ сводится к поиску максимальной совместной подсистемы в системе неравенств

$$\begin{cases} B_p(x_k)\leqslant c, & k\in I_0;\\ B_p(x_k)>c, & k\in I_1. \end{cases} \tag{3.3}$$

Задача минимизации функционала $Q(F(B_1,\dots,B_p))$ по оператору $B_p\in\mathfrak{M}^0$ при фиксированных $B_1,\dots,B_{p-1}$ сводится к поиску совместной подсистемы с максимальным весом в системе неравенств

$$\begin{cases} B_p(x_k)\leqslant c, & \text{с весом } w_k, & k\in D_0^{p-1};\\ B_p(x_k)>c, & \text{с весом } w_k, & k\in D_1^{p-1}; \end{cases} \tag{3.4}$$

где веса w_k являются оценкой числа неравенств системы (3.2), выполняемых при выполнении k-го неравенства данной системы. Например, можно положить

$$w_k = \frac{1}{2}\sum_{j=1}^{q}(w_{jk} + w_{kj}), \tag{3.5}$$

считая, что $w_{jk} = 0$ для всех (j, k), не принадлежащих $\mathbb{D}(B_1, \ldots, B_{p-1})$.

Комбинированная стратегия настройки при заданном $\lambda \in [0, 1]$ сводится к поиску совместной подсистемы с максимальным весом в системе неравенств

$$\begin{cases} B_p(x_k) \leqslant c, & \text{с весом } 1 + \lambda w_k, \quad k \in D_0^{p-1}; \\ B_p(x_k) \leqslant c, & \text{с весом } 1 - \lambda, \quad k \in I_0 \setminus D_0^{p-1}; \\ B_p(x_k) > c, & \text{с весом } 1 + \lambda w_k, \quad k \in D_1^{p-1}; \\ B_p(x_k) > c, & \text{с весом } 1 - \lambda, \quad k \in I_1 \setminus D_1^{p-1}; \end{cases} \tag{3.6}$$

При $\lambda = 0$ система (3.6) преобразуется в (3.3) и оператор B_p настраивается на исходную обучающую выборку. При $\lambda = 1$ она преобразуется в (3.4) и настройка идёт на компенсацию неточности остальных операторов. При промежуточных значениях λ получаются комбинированные стратегии настройки. Все три задачи (3.3), (3.4) и (3.6) решаются одним и тем же численным методом поиска совместной подсистемы максимального веса и отличаются только способом приписывания весов неравенствам.

3. Рассмотрим теперь задачу восстановления регрессии, $\mathfrak{I}_f = \mathfrak{I}_e = \mathbb{R}$. Обычно в этом случае используют среднеквадратичный функционал качества:

$$Q_0(B) = \sum_{k=1}^{q}(B(x_k) - y_k)^2 \to \min_{B \in \mathfrak{M}^0}. \tag{3.7}$$

В то же время минимизация числа дефектных пар оператора $F(B_1, \ldots, B_p)$ при фиксированных $B_1, \ldots, B_{p-1}$ сводится к поиску совместной подсистемы максимального веса в системе неравенств (3.2). Различие в постановках задач (3.7) и (3.2) делает невозможным как их решение одним и тем же методом, так и построение комбинированной стратегии настройки оператора B_p. Поэтому имеет смысл свести задачу (3.2) к минимизации функционала, аналогичного (3.7). Разумеется, точное сведение невозможно. Однако выбор функционала качества изначально основан на эвристиках, поэтому любое удобное его изменение, в общем сохраняющее цель настройки, представляется допустимым.

Воспользуемся тем, что близость значений $B_p(x_k)$ к значениям y_k является достаточным условием для выполнения соотношений (3.2). Потребуем, чтобы расстояния $|B_p(x_k) - y_k|$ были не слишком велики для всех k, образующих дефектные пары из множества $\mathbb{D}(B_1, \ldots, B_p)$.

Введём множества индексов

$$\begin{aligned} D^+_{p-1,k} &= \{j \in \mathbb{Q} \mid (k, j) \in \mathbb{D}(B_1, \ldots, B_{p-1})\}, \quad k = 1, \ldots, q; \\ D^-_{p-1,k} &= \{j \in \mathbb{Q} \mid (j, k) \in \mathbb{D}(B_1, \ldots, B_{p-1})\}, \quad k = 1, \ldots, q; \\ D_{p-1} &= \{k \in \mathbb{Q} \mid D^+_{p-1,k} \cup D^-_{p-1,k} \neq \varnothing\}; \end{aligned}$$

и рассмотрим систему неравенств

$$y_k - \delta_k < B_p(x_k) < y_k + \delta_k, \qquad k \in D_{p-1}, \tag{3.8}$$

где $\delta_k = \frac{1}{2}\min\left(\Delta y_k^+, \Delta y_k^-\right)$, $\Delta y_k^+ = \min\limits_{j \in D_{p-1,k}^+}(y_j - y_k)$, $\Delta y_k^- = \min\limits_{j \in D_{p-1,k}^-}(y_k - y_j)$, и предполагается, что если множество $D_{p-1,k}^+$ (или $D_{p-1,k}^-$) пусто, то значение Δy_k^+ (или Δy_k^-) не определено и при вычислении δ_k не учитывается.

Система (3.8) в общем случае несовместна. Как и для задачи классификации, припишем каждому неравенству весовой коэффициент (3.5). Тем самым задача свелась к поиску совместной подсистемы максимального веса в системе неравенств (3.8). Для её решения применим приближённый метод, не гарантирующий максимальности найденной подсистемы. Потребуем, чтобы значения $|B_p(x_k) - y_k|/\delta_k$ были одновременно малы для всех $k \in D_{p-1}$. С учётом весовых коэффициентов w_k запишем это требование в виде

$$Q_1(B_p) = \sum_{k=1}^{q} \gamma_k (B_p(x_k) - y_k)^2 \to \min_{B_p \in \mathfrak{M}^0}, \tag{3.9}$$

где $\gamma_k = w_k/\delta_k^2$ при $k \in D_{p-1}$ и $\gamma_k = 0$ в противном случае.

В итоге путём двух последовательных модификаций мы свели задачу поиска совместной подсистемы максимального веса (3.2) к задаче минимизации функционала среднеквадратичной ошибки (3.9). Первая модификация состояла в усилении ограничений на оператор B_p, что позволило уменьшить число ограничений. Вторая модификация, напротив, привела к ослаблению требований к оператору B_p, но позволила видоизменить функционал качества.

Полученный функционал (3.9) имеет совершенно прозрачную интерпретацию: настройка производится на объектах, образующих дефект операторов $B_1, \ldots, B_{p-1}$, причём приоритет отдаётся (посредством весовых коэффициентов γ_k) тем из них, которые лежат в основании наибольшего числа дефектных троек.

Сходство функционалов (3.7) и (3.9) позволяет построить комбинированную стратегию настройки при заданном $\lambda \in [0, 1]$:

$$\begin{aligned} Q_\lambda(B_p) &= (1-\lambda)\, Q_0(B_p) + \lambda\, Q_1(B_p) = \\ &= \sum_{k=1}^{q} (1 - \lambda + \gamma_k \lambda)(B_p(x_k) - y_k)^2. \end{aligned} \tag{3.10}$$

При $\lambda = 0$ этот функционал переходит в $Q_0(B_p)$ и оператор B_p настраивается на исходную обучающую выборку. При $\lambda = 1$ он преобразуется в $Q_1(B_p)$, и настройка идёт на компенсацию неточности остальных операторов. При промежуточных значениях λ реализуются комбинированные стратегии настройки. Для минимизации всех трёх функционалов (3.7), (3.9) и (3.10) можно применять одни и те же методы, например метод наименьших квадратов.

4. Рассмотрим наконец задачу построения монотонной корректирующей операции F, доставляющей минимум функционалу качества

$$Q(F) = |\mathbb{D}(F(B_1, \ldots, B_p))|$$

при фиксированных алгоритмических операторах $B_1, \ldots, B_p$.

Если дефект набора операторов $B_1, \ldots, B_p$ непуст, то построить монотонную функцию F, удовлетворяющую системе равенств $F(a_k) = y_k$, $k \in \mathbb{Q}$, невозможно. Однако можно воспользоваться методом, описанным в [1], чтобы, минимальным образом изменив значения y_k, обеспечить выполнение условия

$$a_j < a_k \to y_j \leqslant y_k \quad \text{для всех } (j,k) \in \mathbb{Q}^2. \tag{3.11}$$

Итак, будем считать, что существует монотонная функция F, проходящая через заданные точки $\{a_k, y_k\}_{k=1}^q$. Опишем один из методов построения такой функции для случая $\mathfrak{I}_e = \mathbb{R}$.

С каждым из векторов a_k свяжем два множества, которые назовём соответственно верхней и нижней *областью монотонности* вектора a_k:

$$\begin{aligned} M_k^1 &= \{a \in \mathbb{R}^p \mid a_k \leqslant a\}; \\ M_k^0 &= \{a \in \mathbb{R}^p \mid a \leqslant a_k\}. \end{aligned}$$

Возьмём произвольную p-арную непрерывную неубывающую на $\mathbb{R}_+^p$ функцию $\mu(\rho_1, \ldots, \rho_p)$, принимающую нулевое значение тогда и только тогда, когда $\rho_1 = \ldots = \rho_p = 0$. На практике в качестве функции μ можно брать максимум, сумму, корень из суммы квадратов, и другие.

Для произвольных векторов a из $\mathbb{R}^p$ и a_k из $\{a_k\}_{k=1}^q$ определим расстояние от a до верхней и нижней областей монотонности вектора a_k соответственно:

$$\begin{aligned} r^1(a, a_k) &= \mu\left((a_k^1 - a^1)_+, \ldots, (a_k^p - a^p)_+\right), \\ r^0(a, a_k) &= \mu\left((a^1 - a_k^1)_+, \ldots, (a^p - a_k^p)_+\right), \end{aligned}$$

где индекс «+» обозначает операцию срезки: $z_+ = z$ при $z \geqslant 0$, и $z_+ = 0$ при $z \leqslant 0$.

Непосредственно из определений вытекают следующие свойства введённых расстояний:

(а) функция $r^1(a, a_k)$ невозрастающая по первому аргументу;

(б) функция $r^0(a, a_k)$ неубывающая по первому аргументу;

(в) вектор $a \in \mathfrak{I}_e^p$ принадлежит области монотонности тогда и только тогда, когда расстояние до неё равно нулю: $a \in M_k^\beta \Leftrightarrow r^\beta(a, a_k) = 0$, где $\beta = 0, 1$;

(г) если функция μ непрерывна, то функции r^1 и r^0 также непрерывны.

Перенумеруем объекты обучающей выборки так, чтобы выполнялась цепочка неравенств $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_q$. Для произвольного $Y \in [y_1, y_q)$ определим функции

$$\begin{aligned}
h_Y^1(a) &= \min_{\{k: y_k > Y\}} r^1(a, a_k);\\
h_Y^0(a) &= \min_{\{k: y_k \leqslant Y\}} r^0(a, a_k);\\
f_Y(a) &= \frac{h_Y^0(a)}{h_Y^0(a) + h_Y^1(a)};
\end{aligned}$$

Определение функций h_Y^1 и h_Y^0 корректны, так как множества индексов, по которым берутся минимумы, непустые в силу условия $y_1 \leqslant Y < y_q$.

Функция $f_Y(a)$ представляет собой непрерывную монотонную «ступеньку», равную единице на объединении верхних областей монотонности множества векторов $\{a_k \mid y_k > Y\}$ и нулю на объединении нижних областей монотонности множества векторов $\{a_k \mid y_k \leqslant Y\}$. В пространстве между этими областями она принимает всевозможные промежуточные значения.

Теорема 1. *Если $y_1 \leqslant Y < y_q$, то функция $f_Y(a)$ непрерывная, неубывающая, принимает значения из отрезка $[0, 1]$, причём на векторах a_k — только 0 или 1:*

$$f_Y(a_k) = \theta(y_k - Y), \quad k \in \mathbb{Q}.$$

Доказательство. Покажем, что функция $h_Y^1(a)$ невозрастающая. Возьмём произвольные векторы a и b из $\mathbb{R}^p$, $a \leqslant b$. Из условия $y_1 \leqslant Y < y_q$ и определения функции h_Y^1 следует, что найдутся такие k и j из $\mathbb{Q}$, для которых $h_Y^1(a) = r^1(a, a_k)$ и $h_Y^1(b) = r^1(b, a_j)$. В силу минимальности расстояния $r^1(b, a_j)$ выполняется неравенство $r^1(b, a_j) \leqslant r^1(b, a_k)$. Поскольку функция r^1 невозрастающая по первому аргументу, справедливо также неравенство $r^1(b, a_k) \leqslant r^1(a, a_k)$. Объединяя оба неравенства, получаем $r^1(b, a_j) \leqslant r^1(a, a_k)$, или, что то же самое, $h_Y^1(b) \leqslant h_Y^1(a)$. Но это в силу произвольности векторов a и b означает, что функция $h_Y^1(a)$ невозрастающая.

Аналогично доказывается, что функция $h_Y^0(a)$ является неубывающей. Следовательно функция $f_Y(a)$ неубывающая как суперпозицией неубывающих функций.

Докажем, что функции h_Y^1 и h_Y^0 не могут одновременно обратиться в нуль. Допустим, что это не так. Тогда из условия $y_1 \leqslant Y < y_q$ следует, что найдутся k и j из $\mathbb{Q}$, для которых имеет место равенство $r^1(a, a_k) = r^0(a, a_j) = 0$. Отсюда следует, что $y_j \leqslant Y < y_k$. С другой стороны, поскольку расстояния до областей монотонности равны нулю, $a_k \leqslant a \leqslant a_j$. Полученные соотношения противоречат условию (3.11). Таким образом, знаменатель в определении функции $f_Y(a)$ никогда не обращается в нуль.

Все функции $r^1(a, a_k)$, $r^0(a, a_k)$, $k = 1, \ldots, q$, непрерывны по первому аргументу. Функции $h_Y^1(a)$ и $h_Y^0(a)$ непрерывны, так как минимум непрерывных функций есть непрерывная функция. Следовательно функция $f_Y(a)$ непрерывна как суперпозиция непрерывных функций.

Для произвольного вектора $a \in \mathbb{R}^p$ в силу положительности $h_Y^1(a)$ и $h_Y^0(a)$ справедлива цепочка неравенств $0 \leqslant h_Y^0(a) \leqslant h_Y^0(a) + h_Y^1(a)$, откуда следует $0 \leqslant f_Y(a) \leqslant 1$.

Рассмотрим произвольный вектор a_k из последовательности $\{a_k\}_{k=1}^q$. Если $y_k > Y$, то $h_Y^1(a_k) = r^1(a_k, a_k) = 0$, следовательно $f_Y(a_k) = 1$. Если же $y_k \leqslant Y$, то $h_Y^0(a_k) = r^0(a_k, a_k) = 0$, значит $f_Y(a_k) = 0$. Объединяя оба случая, заключаем, что на векторах a_k функция $f_Y(a)$ вычисляется по формуле $f_Y(a_k) = \theta(y_k - Y)$.

Теорема доказана.

Следствие. Для задач классификации искомая корректирующая операция имеет вид $F(a) = \theta\left(f_{\frac{1}{2}}(a) - \frac{1}{2}\right)$, где θ — функция Хэвисайда.

Для задач восстановления регрессии корректирующая операция строится в виде суммы q элементарных ступенчатых функций вида $f_Y(a)$:

$$F(a) = y_1 + \sum_{k=1}^{q-1} (y_{k+1} - y_k)\, f_{y_k}(a).$$

Теорема 2. *Функция $F(a)$ определена на всём множестве $\mathbb{R}^p$, непрерывна, монотонно не убывает и удовлетворяет условию $F(a_k) = y_k$ для всех $k \in \mathbb{Q}$.*

Доказательство. Функция $F(a)$ непрерывная и неубывающая как сумма q непрерывных неубывающих функций. Согласно предыдущей теореме справедливо тождество $f_{y_k}(a_j) = \theta(y_j - y_k)$ для всех j и k из $\mathbb{Q}$. Поэтому для любого $j \in \mathbb{Q}$

$$F(a_j) = y_1 + \sum_{k=1}^{q-1} (y_{k+1} - y_k)\, \theta(y_j - y_k) = y_1 + \sum_{k=1}^{j-1} (y_{k+1} - y_k) = y_j.$$

Теорема доказана.

Построенная функция $F(a)$ обладает важным для корректирующих операций свойством гладкости. Причём под гладкостью в данном случае удобнее понимать не дифференцируемость, а отсутствие у функции резких скачков в промежутках между точками $\{a_k\}_{k=1}^q$. Для оценок гладкости в таком, менее строгом, понимании можно использовать конечно-разностные методы. Вычислительные эксперименты показали, что по значению конечно-разностного функционала гладкости построенная функция оказывается достаточно близкой к классическим многомерным сплайнам.

Автор выражает глубокую признательность чл.-корр. РАН К. В. Рудакову за постановку задачи и ключевую идею использования методов оптимизации для построения проблемно-ориентированных базисов.

Список литературы

[1] Воронцов К.В. О проблемно-ориентированной оптимизации базисов задач распознавания // Ж. вычисл. матем. и матем. физ. 1998. Т. 38. № 5. С. 870-880.

[2] Журавлёв Ю.И. Корректные алгебры над множеством некорректных (эвристических) алгоритмов. I–III. // Кибернетика. 1977. № 4. С. 14–21; 1977. № 6. С. 21–27; 1978. № 2. С. 35–43.

[3] Журавлёв Ю.И. Об алгебраическом подходе к решению задач распознавания или классификации. // Пробл. кибернетики. 1979. Вып. 33. С. 5–68.

[4] Рудаков К.В. Универсальные и локальные ограничения в проблеме коррекции эвристических алгоритмов // Кибернетика. 1987. № 2. С. 30–35.

[5] Рудаков К.В. Полнота и универсальные ограничения в проблеме коррекции эвристических алгоритмов классификации // Кибернетика. 1987. № 3. С. 106–109.

[6] Рудаков К.В. О симметрических и функциональных ограничениях для алгоритмов классификации // Докл. АН СССР. 1987.Т. 297. № 1. С. 43-46.

[7] Казанцев В.С. Задачи классификации и их программное обеспечение. М.: Наука. 1990.

[8] Катериночкина Н.Н. Поиск максимального верхнего нуля монотонной функции алгебры логики // Докл. АН СССР. 1975. Т. 224. № 3. С. 557-560.

[9] Мазуров В.Д. Метод комитетов в задачах оптимизации и классификации. М.: Наука. 1990.

[10] Лоусон Ч., Хенсон Р. Численное решение задач метода наименьших квадратов. М.: Наука. 1986.