О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания

чл.-корр. РАН К. В. Рудаков, К. В. Воронцов

В сообщении описаны специальные оптимизационные методы построения корректных алгоритмов на основе технологии алгебраического подхода к проблеме распознавания. Предложен и исследован общий метод синтеза проблемно-ориентированных базисов для задач распознавания и прогнозирования, сводящий проблему построения таких базисов к последовательности классических оптимизационных задач. Для семейства монотонных корректирующих операций доказана теорема о сходимости метода. Получен алгоритм эффективного синтеза монотонных непрерывных корректирующих операций, достаточных для построения корректных алгоритмов.

1. Рассматривается следующая постановка 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 $A(x_k) = y_k$, $k = 1, \dots, q$, и универсальным 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 $\mathfrak{I}_f$, для которого выполнены условия из [1]. Алгоритм, удовлетворяющий локальным и универсальным ограничениям, называют *корректным.*

Итак, рассматриваемые задачи распознавания Z полностью определяются пятёркой $\langle \mathfrak{I}_i, \mathfrak{I}_f, \mathfrak{M}^u, \{x_k\}_{k=1}^q, \{y_k\}_{k=1}^q \rangle$.

2. Традиционный подход к решению задач распознавания состоит в использовании некоторой фиксированной модели алгоритмов $\mathfrak{M} \subseteq \mathfrak{M}^u$ и поиску в ней алгоритма A наилучшего качества. Если качество построенного алгоритма оказывается неприемлемым, используется алгебраический подход к проблеме синтеза корректных алгоритмов [2, 3].

В алгебраическом подходе 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 универсальным ограничениям «по построению». Общие подходы к построению $\mathfrak{F}$-расширений моделей алгоритмов развиваются в теории универсальных и локальных ограничений [1, 4, 5]. Ниже будет рассматриваться 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 $A = C \circ F(B_1, \ldots, B_p)$.

Определение 1. Конечное множество алгоритмических операторов $B_1, \ldots, B_p$ называется *базисом* для задачи Z при заданных $\mathfrak{F}$ и $\mathfrak{M}^1$, если найдутся корректирующая операция $F \in \mathfrak{F}$ и решающее правило $C \in \mathfrak{M}^1$ такие, что алгоритм A является корректным для задачи Z.

Отметим, что базис определяется для фиксированной финальной информации. Тем самым мы ограничиваемся рассмотрением проблемно-ориентированных базисов, то есть таких наборов алгоритмических операторов, которые подходят для решения только данной конкретной задачи, а не всего класса регулярных задач [2]. В классических работах по алгебраическому подходу при конструктивном доказательстве теорем существования корректные алгоритмы строились без использования методов оптимизации. В данном сообщении рассматриваются методы построения базисов, основанные на решении специальной последовательности оптимизационных задач.

3. Пусть задан функционал качества $Q : \mathfrak{M}^0_* \to \mathbb{R}$. В общем случае задача построения базиса для задачи Z заключается в том, чтобы найти алгоритмические операторы $B_1, \ldots, B_p$ из $\mathfrak{M}^0$ и корректирующую операцию F из $\mathfrak{F}$, при которых достигается минимум функционала $Q(F(B_1, \ldots, B_p))$.

Для решения поставленной задачи предлагается подход, основанный на поочерёдной настройке алгоритмических операторов и корректирующей операции. В качестве нулевого приближения в модели $\mathfrak{M}^0$ выбирается оператор B_1, на котором достигается минимум функционала $Q(B)$. Следующие базисные операторы B_2, $B_3, \ldots$ 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 $B_1, \ldots, B_{r-1}$, $B_{r+1}, \ldots, B_p$:

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

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

$$F^* = \arg\min_{F \in \mathfrak{F}} Q(F(B_1, \ldots, B_p)). \tag{2}$$

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

Наряду с (1) и (2) для каждого $r = 1, \ldots, p$ рассмотрим исходную задачу оптимизации

$$(3) \qquad B_r^* = \arg\min_{B_r \in \mathfrak{M}^0} Q(B_r).$$

Оказывается, что для семейств корректирующих операций, обычно применяемых на практике (линейные, полиномиальные, монотонные), задачи (1) и (3) фактически эквивалентны. Это позволяет без существенных изменений перенести весь арсенал методов оптимизации, разработанных для различных эвристических моделей $\mathfrak{M}^0$, на задачи построения проблемно-ориентированных базисов.

Наличие пары оптимизационных задач (1) и (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 $\lambda \in [0, 1]$, введя функционал качества

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

При $\lambda = 0$ решается задача (3), и настройка очередного алгоритмического оператора B_r производится на исходные прецеденты без учёта его дальнейшего использования в качестве аргумента корректирующей операции. При $\lambda = 1$ решается задача (1), и оператор B_r настраивается исключительно на компенсацию неточностей, допущенных при использовании ранее построенных операторов и корректирующей операции. Каждая из этих двух «чистых» стратегий настройки имеет свои недостатки. При промежуточных значениях параметра λ образуются смешаные компромиссные критерии настройки.

4. Перейдём к рассмотрению описанного метода построения базиса задачи Z в случае монотонных корректирующих операций.

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. Множество индексов $\{1, \ldots, q\}$ обозначим через $\mathbb{Q}$.

Определение 2. 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 3. Множество $\mathbb{D}(B_1, \ldots, B_p) = \mathbb{D}(B_1) \cap \ldots \cap \mathbb{D}(B_p)$ назовём *дефектом* набора операторов $B_1, \ldots, B_p$.

Для любой p-арной монотонной корректирующей операции F справедливо включение $\mathbb{D}(B_1, \dots, B_p) \subseteq \mathbb{D}(F(B_1, \dots, B_p))$. Если дефект пуст, то найдётся монотонная корректирующая операция F, для которой множество $\mathbb{D}(F(B_1, \dots, B_p))$ также пусто. Таким образом, чтобы получить корректный алгоритм, достаточно построить набор алгоритмических операторов с пустым дефектом. Для этого очередной оператор B_r, $r = 2, \dots, p$ необходимо выбирать так, чтобы он в максимальной степени удовлетворял, вообще говоря несовместной, системе неравенств

$$B_r(x_j) < B_r(x_k) \quad \text{для всех } (j,k) \in \mathbb{D}(B_{[r]}), \tag{4}$$

где $(B_{[r]}) = (B_1, \dots, B_{r-1}, B_{r+1}, \dots, B_p)$. Задача оптимизации очередного базисного оператора B_r сводится к поиску максимальной совместной подсистемы системы (4).

При определённых ограничениях на модель $\mathfrak{M}^0$ процесс построения корректного алгоритма сходится за конечное число шагов:

Теорема 1. *Пусть модель $\mathfrak{M}^0$ такова, что для любого множества пар*

$$\{(j_i, k_i) \in \mathbb{Q}^2 \mid j_i \neq k_i,\ i = 1, \dots, m\}, \quad m \geqslant 1,$$

найдётся алгоритмический оператор $B \in \mathfrak{M}^0$, для которого ни одна из этих пар не является дефектной. Тогда при $p_* = \left]\frac{1}{m} C_q^2\right[$ существуют набор алгоритмических операторов $(B_1, \dots, B_{p_*}) \in (\mathfrak{M}^0)^{p_*}$ и корректирующая операция $F \in \mathfrak{F}$ такие что $Q(F(B_1, \dots, B_{p_*})) = 0$.

5. Из общих соображений, касающихся надёжности алгоритмов распознавания [6], вытекает целесообразность оптимизации функционала $Q(F(B_1, \dots, B_p))$ при фиксированном числе операторов p. В этом случае дефект $\mathbb{D}(B_1, \dots, B_p)$ может оказаться непустым. Специально для этого случая проведём более детальное исследование множества $\mathbb{D}(F(B_1, \dots, B_p))$ и его соотношения с множеством $\mathbb{D}(B_1, \dots, B_p)$, что позволит ввести весовые коэффициенты для неравенств системы (4) и свести задачу оптимизации оператора B_r к поиску совместной подсистемы максимального веса.

Введём последовательность $\{a_k\}_{k=1}^q$, состоящую из q векторов $a_k = [B_i(x_k)]_{i=\overline{1,p}}$.

Определение 4. Тройка индексов $(j, s, k) \in \mathbb{Q}^3$ называется *дефектной тройкой* набора алгоритмических операторов $B_1, \dots, B_p$, если: (а) пара (j,k) дефектная для всех B_i, $i = 1, \dots, p$; (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 $y_j < y_s < y_k$. Пару (j,k) будем называть *основанием* дефектной тройки (j,s,k), пары (j,s) и (s,k) — её *рёбрами*, а индекс s — её *вершиной*. Очевидно, основание любой дефектной тройки принадлежит множеству $\mathbb{D}(B_1, \dots, B_p)$.

Теорема 2. *Если в $\mathbb{Q}^3$ имеется хотя бы одна строго дефектная тройка, то для любой монотонной корректирующей операции F справедливо строгое включение $\mathbb{D}(B_1, \dots, B_p) \subset \mathbb{D}(F(B_1, \dots, B_p))$. Если в $\mathbb{Q}^3$ нет дефектных троек, то су-*

ществует такая монотонная корректирующая операция F, что $\mathbb{D}(B_1, \ldots, B_p) = \mathbb{D}(F(B_1, \ldots, B_p))$.

Таким образом, множество $\mathbb{D}(F(B_1, \ldots, B_p))$ состоит из дефектных пар трёх типов: (1) элементов дефекта, (2) рёбер дефектных троек, (3) всех остальных пар. Отсутствие пар второго типа приводит к отсутствию пар третьего типа, а отсутствие пар первого типа — к отсутствию любых дефектных пар. На этом основании предлагается следующий принцип минимизации функционала качества $Q(F(B_1, \ldots, B_p))$: при настройке базисного оператора B_r устранять дефектные пары первого типа, стремясь вместе с ними устранить наибольшее число пар второго типа, и вообще не принимать во внимание пары третьего типа.

Обозначим через t_{jk} число дефектных троек набора операторов $(B_{[r]})$, имеющих основание (j, k). Некоторые из них могут оказаться строго дефектными; обозначим их число через t^0_{jk}.

Теорема 3. *Пусть множество $\mathbb{D}(B_{[r]})$ состоит только из дефектных пар первых двух типов. Тогда для числа d_{jk} дефектных пар, устраняемых при выполнении условия $B_r(x_j) < B_r(x_k)$, справедлива оценка $t^0_{jk} + 1 \leqslant d_{jk} \leqslant t^0_{jk} + t_{jk} + 1$.*

Каждой паре (j, k) из дефекта набора операторов $(B_{[r]})$ сопоставим оценку w_{jk} числа дефектных пар, исключаемых вместе с (j, k). Например, можно положить $w_{jk} = t^0_{jk} + \frac{1}{2} t_{jk} + 1$. Весовой коэффициет w_{jk} показывает, насколько предпочтительнее устранить дефект именно на данной паре. Таким образом, задача построения очередного базисного оператора B_r при неполном исчерпывании дефекта сводится к поиску совместной подсистемы максимального веса в системе неравенств (4) с весами w_{jk}.

Для построения смешаного критерия настройки зададим число λ из отрезка $[0, 1]$. Определим веса неравенств как функцию параметра λ:

$$w_{jk}(\lambda) = \begin{cases} 1 + \lambda(t^0_{jk} + \frac{1}{2} t_{jk}), & (j,k) \in \mathbb{D}(B_{[r]}), \\ 1 - \lambda, & (j,k) \notin \mathbb{D}(B_{[r]}) \text{ и } y_j < y_k. \end{cases}$$

Оптимизация оператора B_r по смешаному критерию сводится к поиску совместной подсистемы максимального веса для системы взвешенных неравенств

$$B_r(x_j) < B_r(x_k) \quad \text{с весом } w_{jk}(\lambda) \quad \text{для всех } (j,k)\text{: } y_j < y_k.$$

6. Теперь рассмотрим задачу построения собственно монотонного отображения F, доставляющего минимум функционалу качества $Q(F(B_1, \ldots, B_p))$ при фиксированных алгоритмических операторах.

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 [7], чтобы, минимальным образом изменив значения y_k, обеспечить выполнение условия $a_j \leqslant a_k \to y_j \leqslant y_k$ для всех $(j, k) \in \mathbb{Q}^2$.

Итак, будем считать, что существует монотонная функция F, проходящая через заданные q точек (a_k, y_k). Опишем один из методов построения такой функции.

Положим для определённости $\mathfrak{I}_e = \mathbb{R}$. В качестве множества финальных информаций будем рассматривать либо $\mathfrak{I}_f = \{0, 1\}$ для задач классификации в два непересекающихся класса, либо $\mathfrak{I}_f = \mathbb{R}$ для задач восстановления регрессии.

Возьмём произвольную p-арную неубывающую на $\mathbb{R}^p_+$ непрерывную функцию $\mu(\rho_1, \ldots, \rho_p)$ такую, что $\mu(\rho_1, \ldots, \rho_p) = 0$ тогда и только тогда, когда $\rho_1 = \ldots = \rho_p = 0$. Для произвольного вектора $a = (a^1, \ldots, a^p)$ из $\mathbb{R}^p$ и произвольного вектора $a_k = (a^1_k, \ldots, a^p_k)$ из последовательности $\{a_k\}^q_{k=1}$ определим две функции расстояния:

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

где индекс «+» обозначает операцию срезки: $z_+ = z$ при $z \geqslant 0$, и $z_+ = 0$ при $z \leqslant 0$. Для произвольного действительного Y определим функции

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

Теорема 4. *Для любого Y функция $f_Y(a)$ непрерывная, неубывающая, принимает значения из отрезка $[0, 1]$, причём на векторах a_k — только 0 или 1:*

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

Данная теорема позволяет построить искомую корректирующую операцию в случае задачи классификации. Для этого достаточно положить $F(a) = \theta\left(f_{\frac{1}{2}}(a) - \frac{1}{2}\right)$, где θ — функция Хэвисайда.

Для задач восстановления регрессии корректирующая операция строится в виде суммы q элементарных ступенчатых функций вида $f_Y(a)$. Допустим без ограничения общности, что исходные векторы расположены в порядке возрастания финальных информаций: $y_1 \leqslant y_2 \leqslant \ldots \leqslant y_q$. Определим корректирующую операцию F:

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

Теорема 5. *Функция $F(a)$ непрерывная, кусочно-дифференцируемая, монотонно неубывающая и удовлетворяет условиям $F(a_k) = y_k$ для всех $k \in \mathbb{Q}$.*

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

Список литературы

[1] Рудаков К.В. // Кибернетика. 1987. № 2. С. 30–35.

[2] Журавлёв Ю. И. // Проблемы кибернетики. 1979. Вып. 33. С. 5–68.

[3] Журавлёв Ю. И. I–II. // Кибернетика. 1977. № 4. С. 14–21. 1977. № 6. С. 21–27.

[4] Рудаков К.В. // Кибернетика. 1987. № 3. С. 106–109.

[5] Рудаков К.В. // ДАН СССР. 1987.Т. 297, № 1. С. 43-46.

[6] Матросов В.Л. // ДАН СССР. 1980. Т. 253, № 1. С. 25-30.

[7] Воронцов К.В. // ЖВМ и МФ. 1998. Т. 38, № 5. С. 870-880.

О методах оптимизации и монотонной коррекции в алгебраическом подходе к проблеме распознавания

(реферат)

чл.-корр. РАН К. В. Рудаков, К. В. Воронцов

В сообщении описаны специальные оптимизационные методы построения корректных алгоритмов на основе технологии алгебраического подхода к проблеме распознавания. Предложен и исследован общий метод синтеза проблемно-ориентированных базисов для задач распознавания и прогнозирования, сводящий проблему построения таких базисов к последовательности классических оптимизационных задач. Для семейства монотонных корректирующих операций доказана теорема о сходимости метода. Получен алгоритм эффективного синтеза монотонных непрерывных корректирующих операций, достаточных для построения корректных алгоритмов.