РЕШЕНИЕ ЗАДАЧИ ОБУЧЕНИЯ ПО ПРЕЦЕДЕНТАМ МЕТОДОМ МИНИМИЗАЦИИ СЛОЖНОСТИ МОДЕЛИ

А.С. Анафиев

Таврический национальный университет им. В.И. Вернадского,
факультет математики и информатики
пр-т Вернадского, 4, г. Симферополь, Крым, Украина, 95007
e-mail: *aydera@mail.ru*

Abstract

We consider an approach to the learning by precedents problem based on the minimization of the required function model complexity. Such notions as function model, function class template, function model complexity, function complexity are introduced. The described method is used to solve the problem of pseudo-Boolean function reconstruction by precedents.

Основные определения. Постановка задачи

Пусть имеется множество объектов X, множество ответов Y и множество $\mathfrak{A}$ всех функций, действующих из X в Y.

Рассмотрим класс функций $A \subseteq \mathfrak{A}$. Каждой функции $f \in A$ будет ставить в соответствие некоторые вещественное число ω_f, которое будем называть *весом* функции f. Если для некоторой функции $f \in A$ явно не указан ее вес, то он считается равным единице.

Определение 1. *Замыканием* класса функций A будем называть подмножество $[A] \subseteq \mathfrak{A}$ всех функций представимых в виде суперпозиции конечного числа функций из A. Другими словами, для любой функции $f \in [A]$ справедливо следующее соотношение

$$f = f_1 \circ \ldots \circ f_m, f_i \in A, \qquad i = \overline{1, m}. \tag{1}$$

Определение 2. Класс функций P называется *замкнутым*, если $[P] = P$.

Определение 3. Класс функций $P \in \mathfrak{A}$ называется *функционально полным* в $\mathfrak{A}$, если $[P] = \mathfrak{A}$.

Определение 4. Формулу вида (1) будем называть *моделью (описанием)* функции f над классом функций A, а число функций в такой модели - *рангом* модели.

Модель функции f над классом A будем обозначать через $M(f, A)$, ранг модели через $r(M(f, A))$, а множество всех моделей класса A через M_A.

Определение 5. Отображение $T : \underbrace{A \times \ldots \times A}_{s} \to M_A$ называется *шаблоном* класса A, а число s - местностью данного шаблона.

Другими словами, шаблон - это функция, переменными которой являются функции класса A, а результатом - модель некоторой функции из $[A]$.

Определение 6. Рассмотрим модель $M(\cdot, A)$ и шаблон T. Если найдется такой набор функций $h_1, \ldots, h_s \in A$, что $T(h_1, \ldots, h_s) = M(\cdot, A)$, то будем говорить, что модель $M(\cdot, A)$ *соответствует* шаблону T или *порождается* шаблоном T.

Определение 7. Множество всех моделей соответствующих шаблону T будем называть *образом*шаблона T и обозначатьΔ_T.

Очевидно, что $\Delta_T \subseteq M_A$.

Определение 8. Шаблон $T : \Delta_T = M_A$ называется *полным шаблоном* для класса A.

Определение 9. Число

$$\mathfrak{C}(M(f, A)) = \sum_{i=1}^{m} \omega_{f_i},$$

где $\omega_{f_i} \in \mathbb{R}$ - вес функции $f_i \in M(f, A)$, $i = \overline{1, m}$ и $m = r(M(f, A))$, будем называть *сложностью модели* $M(f, A)$ функции f.

Определение 10. Число

$$\mathfrak{C}(M(f, A)|T) = \begin{cases} \mathfrak{C}(M(f, A)), & \text{если} \quad \exists M(f, A) : M(f, A) \in \Delta_T, \\ \infty, \text{иначе}, \end{cases}$$

будем называть *сложностью* модели $M(f, A)$ функции f *относительно шаблона* T.

Определение 11. Сложностью $K(f|A)$ функции $f \in \mathfrak{A}$ по классу A будем называть величину

$$K(f|A) = \begin{cases} \min\limits_{M(f,A)} \mathfrak{C}(M(f, A)), & f \in [A], \\ \infty, & f \notin [A]. \end{cases}$$

Другими словами, сложность функции $f \in \mathfrak{A}$ по классу A есть сложность модели, которая имеет наименьшую сложность среди всех моделей функции f над классом A, если функция f принадлежит замыканию $[A]$, и равна бесконечности в противном случае.

Определение 12. Сложностью $K(f|A, T)$ функции $f \in \mathfrak{A}$ по классу A относительно шаблона T будем называть величину

$$K(f|A, T) = \begin{cases} \min\limits_{M(f,A) \in \Delta_T} \mathfrak{C}(M(f, A)|T), & f \in [A], \\ \infty, & f \notin [A]. \end{cases}$$

Утверждение 1. $\forall f, T$ справедливо следующее неравенство

$$\min_{M(f,A)} \mathfrak{C}(M(f, A)) \leqslant \mathfrak{C}(M(f, A)|T).$$

Утверждение 2. $\forall f, T$ справедливо следующее неравенство

$$K(f|A) \leqslant K(f|A, T).$$

Данные утверждения позволяют сделать следующий вывод. Сложность модели не превосходит сложности модели относительно произвольного шаблона (которую в некоторых случаях гораздо проще вычислить, чем сложность модели в целом) является верхней оценкой для сложности модели. Аналогичные рассуждения справедливы и для сложности функции.

Теорема 1.

$$\arg\min_{h_1,\dots,h_s\in A}\mathfrak{C}(M(f,A)|T(h_1,\dots,h_s))=\arg\min_{h_1,\dots,h_s\in A}\sum_{i=1}^{s}\omega_{h_i},$$

где ω_{h_i} *— вес функции* h_i *в классе* A.

Доказательство. Т.к. значением шаблона для произвольного набора $(h_1,\dots,h_s)$ является модель, то

$$\mathfrak{C}(M(f,A)|T(h_1,\dots,h_s))=\sum_{g\notin H}\omega_g+\sum_{i=1}^{s}\omega_{h_i},$$

где $H=h_1,\dots,h_s$ и ω_g — означает вес функции g.

Учитывая, что слагаемое $\sum\limits_{g\notin H}\omega_g$ не зависит от набора $(h_1,\dots,h_s)$, то отсюда следует утверждение теоремы.

Теорема доказана. $\square$

Замечание. Далее, если явно не указан вес для функции класса A, то он считается равным единице.

Пример 1. Рассмотрим класс функций

$$\mathfrak{B}=\{a,a\in\mathbb{R},f_1(x,y)=x+y,f_2(x,y)=x\cdot y\},$$

$x,y\in X^n\subseteq\mathbb{R}$. Класс $\mathfrak{B}$, является классом полиномов над множеством X_n с вещественными коэффициентами.

Определим вес для каждой функции из £ следующим образом: $\omega_a=0$, $\omega_{x+y}=1$ и $\omega_{x\cdot y}=2$.

Рассмотрим полином $p=5x_1^3x_3+x_2^2+2$. Моделью данного полинома будет формула

$$M(p,\mathfrak{B})=f_1(f_2(2)),f_1(f_2(x_2,x_2),f_2(x_2,f_2(5,f_2(x_1,f_2(x_1,x_1))))),$$

$r(M(p,\mathfrak{B}))=10$ и $\mathfrak{C}(M(p,\mathfrak{B}))=14$.

$T(h_1,h_2,h_3)=h_1x+h_2x^2+h_3x^3$ является шаблоном для класса £. Вычислим значения этого шаблона на некоторых наборах функций:

$$T(3,2,1)=3x+2x^2+x^3,$$

$$T(5,6,x+y)=5x+6x^2+x^3(x+y).$$

Постановка задачи. Рассмотрим задачу $z=\langle X,Y,\mathfrak{A},X^l\rangle$ обучения по прецедентам, где X - множество объектов, Y - множество ответов, $\mathfrak{A}$ - множество отображений из X в Y и $X^l = \{x_1, \dots, x_l\}$ конечная обучающая выборка. Функция $f : X \to Y$, не обязательно принадлежащая $\mathfrak{A}$, задана не полностью, известны лишь значения данной функции $y_i = f(x_i)$ на объектах X^l. Необходимо построить функцию $h^* \in \mathfrak{A}$ такую, что:

- $h^*(x_i) = y_i$, $i = \overline{1, l}$. В зависимости от поставленной задачи равенство можно понимать и как приближенное. Другими словами на алгоритм h^* накладываются так называемые локальные ограничения [1], - ограничения, связанные с конечным числом обучающих объектов и допускающих эффективную проверку за конечное число шагов;
- на функцию h^* могут накладываться дополнительные ограничения общего характера, такие как линейность, монотонность, непрерывность, гладкость и т.д., а также их сочетания. Другими словами, предполагается, что $h^* \in A \subseteq \mathfrak{A}$, A - заданное подмножество функций, определяемое спецификой задачи;
- $h^* = \arg\min\limits_{h \in A} \mathfrak{C}(M(h, A))$.

Поставленная задача является задачей выбора решающего правила h^*, которое строится по принципу *Minimum Description Length* (MDL) [2], поскольку минимизирует функционал $\mathfrak{C}(M(h, A))$.

Задачу $z = \langle X, Y, \mathfrak{A}, X^l \rangle$ в общем случае является трудно разрешимой. Поэтому будем рассматривать различные типы подобных задач с дополнительными предположениями относительно компонент задачи $X, Y, \mathfrak{A}$ и функционала качества$\mathfrak{C}$. В частности, будем рассматривать задачи $z = \langle X, Y, \mathfrak{A}, X^l, T \rangle$, где функционалом качества служит сложность $\mathfrak{C}(M(h, A)|T)$ модели $M(h, A)$ относительно шаблона T класса A. В этом случае задача z будем иметь следующий вид:

$$\begin{cases} \mathfrak{C}(M(h, A)|T) \to \min, \\ h(x_i) = y_i, \quad i = \overline{1, l}, \end{cases}$$

где T – некоторый фиксированный шаблон класса A.

Таким образом, задача $z = \langle X, Y, \mathfrak{A}, X^l, T \rangle$ сводится к некоторой, в зависимости от вида множеств $X, Y, \mathfrak{A}, A$ и шаблона T, оптимизационной задаче. Для решения последней можно использовать хорошо известные алгоритмы из теории исследования операций.

1. Задачи типа $z = \langle \mathbb{B}^n, \mathbb{R} \rangle$

Задача типа $z = \langle \mathbb{B}^n, \mathbb{R} \rangle$ – это задача вида $z = \langle X, Y, \mathfrak{A}, X^l \rangle$, в которой $X = \mathbb{B}^n$ и $Y = \mathbb{R}$. Таким образом, множество $\mathfrak{A}$ является множеством всех псевдобулевых функций от n переменных.

Определение 13. Функция $f : \mathbb{B}^n \to \mathbb{R}$, где $\mathbb{B}^n = \{0, 1\}_n$ и $\mathbb{R}$ – множество вещественных чисел, называется *псевдобулевой функцией от n переменных.*

Множество всех псевдобулевых функций от n переменных будем обозначать через $PS_2(n)$.

Известно [3], что любая псевдобулевая функция f представима идинственным образом полиномом вида

$$P = \sum_{\{i_1,\ldots,i_k\}\subseteq\{1,2,\ldots,n\}} C_{i_1\ldots i_k} x_{i_1} \cdot \ldots \cdot x_{i_k}$$

где $C_{i_1\ldots i_k} \in \mathbb{R}$, $\{i_1, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$.

Отсюда следует, что любую псевдобулевую функцию $f(x_1, \ldots, x_n)$ можно единственным образом представить в виде полинома вида

$$f(x_1, \ldots, x_n) = a_1 K_1 + \ldots + a_m K_m + a_0, \quad a_i \in \mathbb{R},\ i = \overline{0, m}, \tag{2}$$

где K_j, $j = \overline{1, m}$, – попарно различные монотонные (не содержащие отрицаний переменных) элементарные конъюнкции над множеством переменных $\{x_1, \ldots, x_n\}$. Максимальный из рангов конъюнкций K_j, $j = \overline{1, m}$, называется *степенью* полинома и обозначается $deg(f)$, а сумма всех рангов называется *размером* и обозначается $size(f)$.

Определение 14. Псевдобулевая функция f называется *линейной*, если $deg(f) \leqslant 1$.

Множество всех линейных псевдобулевых функций от n переменных будем обозначать через $LPS_2(n)$, а множество всех линейных псевдобулевых функций, которые представимы полиномом вида (2), где $a_i \geqslant 0$ ($a_i \leqslant 0$), $i = \overline{0, m}$, через $LPS_2^+(n)$ ($LPS_2^-(n)$).

Рассмотрим класс псевдобулевых функций $\mathfrak{B} = \{f_c \equiv c,\ c \in \mathbb{R};\ x+y;\ xy\}$. Данный класс является полным в $PS_2(n)$, т.к. любую функцию можно записать как полином вида (2), который является суперпозицией функции xy, $x + y$ и констант c, $c \in \mathbb{R}$.

Рассмотрим следующий класс шаблонов класса $PS_2(n)$.

$\mathfrak{T}_1 = \{T_s | T_s = T_s(h_0, h_1, \ldots, h_s) = h_0 + h_1 x_1 + \ldots + h_s x_s\}$, где $h_i = f_c$, $c \in \mathbb{R}$, $i = \overline{0, s}$, $s = \overline{0, n}$. При $s = n$ образ шаблона T_n будет совпадать с классом функций $LPS_2(n)$, а при $s = 0$ образом шаблона T_0 будет классом псевдобулевых функций-констант.

Так, например, значение шаблона $T_3(1, 3, 0, 4)$ будет модель $M = 1 + 3x_1 + 4x_3$, а значением шаблона $T_2(1, 1)$ – модель $M = 1 + x_1$.

Рассмотрим следующую задачу

$$z = \langle \mathbb{B}^n, \mathbb{R}, LPS_2(n), *, T_n^+ \rangle, \tag{3}$$

где $T_n^+(h_0, h_1, \ldots, h_n) = T_n(h_0, h_1, \ldots, h_n)$, в котором функции $h_i = f_c$, $c \in \mathbb{R}$, $c \geqslant 0$.
Замечание. Символ $*$ означает произвольность компонента задачи, в данном случае, произвольность обучающей выборки.

Для решения задачи (3), мы должны построить функцию $h_* \in LPS_2^+(n)$ наименьшей сложности, которая удовлетворяет обучающей выборки, предполагая, что восстанавливаемая функция является линейной. Другими словами, построить модель функции наименьшей сложности, удовлетворяющую локальным ограничениям задачи.

Задачу 3 можно записать в следующей эквивалентной форме:

$$\begin{cases} \mathfrak{C}(h|LPS_2^+, T_n^+) \to \min, \\ h(x_i) = y_i, \quad i = \overline{1,l}. \end{cases} \tag{4}$$

Зададим веса функций класса $\mathfrak{B}$ следующим образом: $\omega_{x+y} = \omega_{xy} = 1$ и $\omega_{f_c} = c$. Тогда, на основании теоремы 1, задачу (4) можно записать, учитывая, что $x_i = (x_{1i}, \ldots, x_{ni})$, следующим образом:

$$\begin{cases} c_1 + c_2 + \ldots + c_n \to \min, \\ c_1x_{11} + c_2x_{21} + \ldots + c_nx_{n1} = y_1, \\ c_1x_{12} + c_2x_{22} + \ldots + c_nx_{n2} = y_2, \\ \qquad\qquad \ldots \\ c_1x_{1l} + c_2x_{2l} + \ldots + c_nx_{nl} = y_l, \\ c_i \geqslant 0, \quad i = \overline{1,l}. \end{cases} \tag{5}$$

Задача (5) является задачей линейного программирования. Одним из методов решения которой является симплекс-метод. Если данная задача имеет решение и $c_{min} = (c_1^*, \ldots, c_n^*)$, то искомым решением задачи (3) будет функция $h^* = T_n^+(c_1^*, \ldots, c_n^*) = c_1x_1 + \ldots + c_nx_n$.

Пример 2. Пусть $f = 6x_1 + x_2 + 2x_3$. Будем считать, что функция f нам задана не полностью, а известны лишь значения этой функции на следующих наборах $f(0,1,1) = 3$, $f(1,0,1) = 8$ и $f(0,0,0) = 0$. Найдем h^*. Для этого необходимо решить следующую задачу:

$$\begin{cases} c_1 + c_2 + c_3 \to \min, \\ c_2 + c_3 = 3, \\ c_1 + c_3 = 8, \quad c_i \geqslant 0, \ i = \overline{1,3}, \end{cases}$$

решая которую, получим $c_{min} = (5, 0, 3)$ и $h^*(x) = 5x_1 + 3x_3$.

Учитывая специфику задач линейного программирования, где число ограничений l меньше (намного меньше) числа переменных n, в решении c_{min} задачи 5 будет как минимум $n - l$ нулевых координат (нулевых слагаемых в полиномиальном представлении функции h^*). Таким образом, решая задачу (3) методом минимизации сложности модели, мы минимизируем и размер (в данном случае размер равен числу слагаемых в полимиальном представлении функции h^*) функции h^*, который играет важное значение в методах поиска оптимального решения в задачах оптимизации псевдобулевых функций.

Заключение

В данной работе рассмотрен один из подходов решения задачи $z = \langle X, Y, \mathfrak{A}, X^l, T \rangle$ обучения по прецедентам, где в качестве функционала качества выбирается сложность модели функции $\mathfrak{C}(M(\cdot, A)|T)$ относительно шаблона T. На основании этого, в зависимости от вида множеств $X, Y, \mathfrak{A}, A$ и шаблона T, данная задача сводится к некоторой оптимизационной задаче, для решения которой можно использовать хорошо известные алгоритмы из теории исследования операций.

Рассмотрен случай задачи z, когда $X = \mathbb{B}^n$, $Y = \mathbb{R}$, $\mathfrak{A} = LPS_2^+(n)$ и $T = T_n^+(h_0, h_1, \ldots, h_n)$. Показано, что в этом случае задача z обучения по прецедентам сводится к задачи линейного программирования и решается с помощью симплекс-метода.

Список литературы

1. Журавлев Ю.И., Рудаков К.В. Об алгебраической коррекции процедур обработки (преобразования) информации. В сб. «Проблемы прикладной математики и информатики» под ред. О.М. Белоцерковского и др. 1987. С. 187-198.
2. Kearns M.J., Mansour Y., Ng A.Y., Ron D. An experimental and theoretical comparison of model selection methods Computational Learning Theory.- 1995.- P.21-30.
3. Boros E., Hammer P.L. Pseudo-boolean Optimization. Discrete Applied Mathematics 123(1-3), 2002, pp. 155-225.