

## НЕКОТОРЫЕ ПОЛОЖЕНИЯ И ЗАДАЧИ ТЕОРИИ ШАБЛОНОВ

© Анафиев А.С.

ТАВРИЧЕСКИЙ НАЦИОНАЛЬНЫЙ УНИВЕРСИТЕТ ИМ. В.И. ВЕРНАДСКОГО  
ФАКУЛЬТЕТ МАТЕМАТИКИ И ИНФОРМАТИКИ  
ПР-Т ВЕРНАДСКОГО, 4, Г. СИМФЕРОПОЛЬ, 95007, УКРАИНА  
E-MAIL: aydera@mail.ru

**Abstract.** The main definitions, positions and problems of template theory are considered. The new types of templates are introduced. Some properties of polynomial templates are considered. New tasks which play an important role of the construction of high-quality decision rules are emphasized.

### ВВЕДЕНИЕ

Рассмотрим стандартную постановку задачи обучения по прецедентам. Дано множество объектов  $X$ , множество ответов  $Y$  и некоторая неполная информация о целевой зависимости  $f^* : X \rightarrow Y$  представленная в виде конечного набора прецедентов  $\{(x_i \in X, y_i \in Y)\}_{i=1}^{\ell}$ , где  $y_i = f^*(x_i)$ ,  $i = 1..{\ell}$ . При этом набор  $\{x_1, \dots, x_{\ell}\} \subset X$  будем называть обучающей выборкой и обозначать  $X^{\ell}$ .

Задача обучения по прецедентам заключается в том, чтобы восстановить неизвестную целевую зависимость  $f^*$  между объектами и ответами, т.е. построить алгоритм  $a^* : X \rightarrow Y$ , который будет удовлетворять следующим требованиям.

- Алгоритм  $a^*(x)$  должен воспроизводить на объектах обучающей выборки заданные ответы:  $a(x_i) = y_i$ ,  $i = 1..{\ell}$ . Причем, равенство здесь можно понимать и как приближенное, в зависимости от специфики задачи.
- На алгоритм  $a^*$  могут накладываться разного рода априорные ограничения, такие как, линейность, монотонность, гладкость и т.д., а также и их сочетания. Более того, может быть задана некоторая модель данного алгоритма [1].
- Алгоритм  $a^*$  должен достаточно хорошо восстанавливать неизвестную целевую зависимость  $f^*$  не только на объектах обучающей выборки, но и на всем множестве объектов  $X$ . Другими словами, алгоритм  $a^*$  должен обладать обобщающей способностью.

При этом фраза «достаточно хорошо восстанавливает» означает, что искомый алгоритм  $a^*$  должен быть лучшим в некотором семействе алгоритмов относительно некоторого заранее известного функционала качества.

Кроме этого, при решении задачи обучения по прецедентам заранее фиксируется класс решающих правил  $A$ , из которого и будет выбираться оптимальный алгоритм, удовлетворяющий вышеописанным требованиям.

При этом, в большинстве случаев, класс решающих правил  $A$  выбирается достаточно широким, что приводит к неприменимости многих оценок, например, оценок Вапника-Червоненкиса [3], для оценки качества полученного алгоритма и получения «рецепта» – как выбирать оптимальное (близкое к оптимальному) решающее правило в заданном семействе  $A$ .

Для сужения класса решающих правил  $A$  с целью построения более качественных решающих правил была предложена теория шаблонов [2]. Рассмотрим основные определения и положения данной теории.

### 1. НЕКОТОРЫЕ ОПРЕДЕЛЕНИЯ И ПОЛОЖЕНИЯ ТЕОРИИ ШАБЛОНОВ

**Определение 1.** Оператор  $t : \Theta \rightarrow A$  будем называть *шаблоном* множества алгоритмов  $A$  относительно параметрического множества  $\Theta$ . Если при этом  $\Theta = P^k$ , то шаблон  $t$  будем называть *k-параметрическим* и обозначать  $t^k$ .

Приведем здесь также определение модели алгоритмов [1].

**Определение 2.** *Моделью алгоритмов* называется параметрическое семейство отображений  $A = \{\varphi(x, \theta) \mid \theta \in \Theta\}$ , где  $\varphi : X \times \Theta \rightarrow Y$  – некоторая фиксированная функция,  $\Theta$  – множество допустимых значений параметра  $\theta$ , называемое пространством параметров или пространством поиска.

Данное определение можно переформулировать, используя понятие шаблона.

**Определение 3.** *Моделью алгоритмов* называется параметрическое семейство отображений  $A = \{f(x) \mid f = t(\theta), \theta \in \Theta\}$ , где  $t(\theta)$  – некоторый фиксированный шаблон,  $\Theta$  – множество допустимых значений параметра  $\theta$ .

Задавая модель алгоритма с помощью шаблона, мы акцентируем свое внимание на конструктивных особенностях решающих правил и на возможность сужения области неопределенности, задаваемой моделью алгоритмов. При этом параметрическое множество определяет количество параметров и их допустимые значения, а шаблон показывает каким образом связаны между собой параметры и переменные. Тогда возникают следующие две интересные задачи: первая (прямая) – как можно использовать структуру и вид связей между параметрами и переменными для построения качественных решающих правил адекватных данной выборке; и вторая (обратная) – задача нахождения структуры выборки описываемой в виде шаблона.

Также можно рассматривать модель отдельной функции.

**Определение 4.** Тройку  $M_f = \langle \theta, t, \Theta \rangle$  такую, что  $t(\theta) = f$ , где  $t$  – шаблон множества  $A$  относительно параметрического множества  $\Theta$  и  $\theta \in \Theta$ , будем называть *моделью функции*  $f \in A$ .

**Определение 5.** Множество

$$\Delta_t(\Theta) = \{f \in A \mid M_f = \langle \theta, t, \Theta \rangle\}$$

будем называть *образом шаблона*  $t$  на множестве  $A$  относительно параметрического множества  $\Theta$ .

**Определение 6.** Шаблон  $t$  называется *универсальным шаблоном выборки*  $X^\ell = (x_i, y_i)_{i=1}^\ell$ , если  $\forall f \in \Delta_t$  и  $\forall i = \overline{1, \ell}$  выполняется  $f(x_i) = y_i$ .

Будем говорить, что шаблон  $t$  *удовлетворяет выборке*  $X^\ell = (x_i, y_i)_{i=1}^\ell$ , если  $\exists f \in \Delta_t$  такая, что  $\forall i = \overline{1, \ell}$  выполняется  $f(x_i) = y_i$ .

**Замечание.** В некоторых задачах равенство  $f(x_i) = y_i$  можно понимать как приближенное.

**Определение 7.** Шаблон  $t_{min}$  называется *минимальным шаблоном для выборки*  $X^\ell$ , если он имеет наименьшее число параметров среди всех удовлетворяющих выборке  $X^\ell$  шаблонов.

Тогда становится актуальной задача построения минимального или близкого к минимальному шаблонов удовлетворяющих обучающей выборке! И какова сложность отыскания таких шаблонов для различного рода задач обучения по прецедентам.

Рассмотрим несколько различных видов шаблонов.

Обозначим  $\tilde{\theta}^{(k)} = (\theta_1, \dots, \theta_k)$ .

**Определение 8.** Шаблон  $t^k$  называется  $\alpha$ -шаблоном, если его можно представить в виде

$$t^k(\tilde{\theta}^{(k)}) = h \left( h_1 \left( \psi_1 \left( \tilde{\theta}^{(k)} \right), \varphi_1(x_1, \dots, x_n) \right), \dots, h_m \left( \psi_m \left( \tilde{\theta}^{(k)} \right), \varphi_m(x_1, \dots, x_n) \right) \right), \quad (1)$$

где

- 1)  $x_1, \dots, x_n$  – символы переменных функций из  $A$ ;
- 2)  $h$  –  $m$ -местный функциональный символ;
- 3)  $h_j$  – 2-местный функциональный символ;
- 4)  $\varphi_j$  –  $s$ -местные функциональные символы,  $j = \overline{1, m}$ ,  $s \in \{0, \dots, n\}$ , причем  $\varphi_i \neq \varphi_j$ ,  $i \neq j$ , функции  $\varphi_j(x_1, \dots, x_n)$  будем называть *свойствами множества  $X$  выделяемые шаблоном  $t^k$* ;
- 5)  $\psi_j$  –  $p$ -местные функциональные символы,  $j = \overline{1, m}$ ,  $p \in \{0, \dots, k\}$ .

При этом, если  $m = k$  и  $\psi_j \left( \tilde{\theta}^{(k)} \right) = \theta_j$ ,  $j = \overline{1, k}$ , то шаблон  $t^k$  называется *простым*.

Пусть  $\mathfrak{F}$  алгебраическое поле с операциями сложения  $+$  и умножения  $\cdot$ . Простой шаблон с  $h(z_1, \dots, z_k) = z_1 + \dots + z_k$ ,  $h_j(z_1, z_2) = z_1 \cdot z_2$  и функциями-свойствами  $\varphi_j(x_1, \dots, x_n) = x_{i_1} \cdot \dots \cdot x_{i_r}$ ,  $j = \overline{1, k}$ , называется *полиномиальным шаблоном* множества функций  $A = \{f : \mathfrak{F}^n \rightarrow \mathfrak{F}\}$  над полем  $\mathfrak{F}$  относительно параметрического множества  $\Theta = \mathfrak{F}^k$ .

Согласно определению, каждый полиномиальный шаблон  $t^k$  можно представить в виде

$$t(\theta_1, \dots, \theta_k) = \theta_1 \cdot x_{i_1^1} \cdot \dots \cdot x_{i_{r_1}^1} + \dots + \theta_k \cdot x_{i_1^k} \cdot \dots \cdot x_{i_{r_k}^k} = \sum_{j=1}^k \theta_j \cdot x_{i_1^j} \cdot \dots \cdot x_{i_{r_j}^j}. \quad (2)$$

Причем произведения  $s_j = x_{i_1^j} \cdot \dots \cdot x_{i_{r_j}^j}$ ,  $j = \overline{1, k}$ , будут свойствами множества  $X$ , выделяемые шаблоном  $t$ . Число  $r_j$  будем называть *рангом* свойства  $s_j$ , а относительно шаблона  $t$  будем говорить, что он порожден множеством свойств  $S = \{s_1, s_2, \dots, s_k\}$  и обозначать  $t \Leftarrow S$ .

Полиномиальный шаблон над полем  $\mathfrak{F}$  будем обозначать  $t_{\mathfrak{F}}$ , а через  $t^{k,r}$  –  $k$ -параметрический полиномиальный шаблон, который способен выделять из множества  $X$  только свойства ранга не больше, чем  $r$ .

Получены следующие оценки ёмкости класса решающих правил порожденных полиномиальными шаблонами [2].

**Определение 9.** Поля, элементы которых можно представить в виде  $\log M$ -разрядного двоичного кода,  $M < \infty$ , будем называть  $M$ -полями.  $M$ -поле, содержащееся во множестве рациональных (целых) чисел, будем называть  $M$ -полем рациональных (целых) чисел.

**Теорема 1.** Пусть  $\mathfrak{F}$  некоторое  $M$ -поле, тогда ёмкость класса  $h(\Delta_{\mathfrak{F}}^{k,r})$  образа множества всех  $k$ -параметрических полиномиальных шаблонов, порожденных свойствами ранга не выше  $r$ , удовлетворяет неравенству

$$h(\Delta_{\mathfrak{F}}^{k,r}) \leq \min \left( \begin{array}{l} k(r \log(n+1) + \log M), \\ k(n \log(r+1) + \log M), \\ P(r) \log M \end{array} \right),$$

где  $P(r)$  – число свойств множества  $X$ , выделяемые всеми шаблонами из  $\mathfrak{T}_{\mathfrak{F}}^{k,r}$ , ранга не больше, чем  $r$ .

**Следствие 1.** Ёмкость  $h(P_{k,r}(n))$  класса  $P_{k,r}(n)$  полиномов над  $M$ -полем рациональных чисел  $\mathbb{Q}$  от  $n$  переменных степени не выше  $r$  с  $k$  отличными от нуля коэффициентами удовлетворяет неравенству

$$h(P_{k,r}(n)) \leq \min \left( \begin{array}{l} k(r \log(n+1) + \log M), \\ k(n \log(r+1) + \log M), \\ \frac{(n+r)!}{n!r!} \log M \end{array} \right).$$

**Следствие 2.** Ёмкость  $h(P_{k,r}^2(n))$  класса  $P_{k,r}^2(n)$  полиномов Жегалкина от  $n$  переменных степени не выше  $r$  с  $k$  отличными от нуля коэффициентами удовлетворяет неравенству

$$h(P_{k,r}^2(n)) \leq \min \left( k(r \log(n+1) + 1), k(n \log(r+1) + 1), \sum_{i=0}^r C_n^i \right).$$

Как показано в [2], учет числа параметров при оценивании ёмкости класса решающих правил позволяет намного улучшить известные оценки. Кроме этого, данная теорема подчеркивает целесообразность построения минимальных (с минимальным числом параметров) шаблонов.

Приведем численную оценку зависимости вероятности неслучайной настройки на выборку от числа параметров  $k$ -параметрического полиномиального шаблона.

**Теорема 2.** Для того, чтобы вероятность  $P(\Delta_{\mathfrak{F}}, \ell, \delta\ell)$  настройки на какие-нибудь  $\ell - \delta\ell$  элементов выборки  $X^\ell$  длины  $\ell$  алгоритмами семейства  $\Delta_{\mathfrak{F}}$ , – образа  $k$ -параметрического полиномиального шаблона, – была меньше некоторого заданного  $\eta > 0$

необходимо, чтобы число параметров  $k$  удовлетворяло условию

$$k < \frac{1}{\log \ell} (\ell - \delta \ell + \log \eta - \log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell))),$$

где  $\gamma(\ell - \delta \ell) = \frac{1}{1 - 2^{-(\ell - \delta \ell) + 1}} \rightarrow 1$  при  $\ell - \delta \ell \rightarrow \infty$ ,  $\delta \ell$  — число ошибок, допускаемых на обучающей выборке алгоритмом из семейства  $\Delta_{tk}$  построенным в результате обучения.

*Доказательство.* По теореме 2.7 из [2]

$$P(\Delta_{tk}, \ell, \delta \ell) < \gamma(\ell - \delta \ell) \cdot C_\ell^{\delta \ell} \cdot 2^{k(\log \ell - 1)} \cdot 2^{-R'(X^\ell, \Delta_{tk})}$$

или, согласно определению степени частичной закономерности  $R'$  [2],

$$P(\Delta_{tk}, \ell, \delta \ell) < \gamma(\ell - \delta \ell) \cdot C_\ell^{\delta \ell} \cdot 2^{k(\log \ell - 1)} \cdot 2^{-(\ell - k - \delta \ell)}.$$

Тогда, если

$$\gamma(\ell - \delta \ell) \cdot C_\ell^{\delta \ell} \cdot 2^{k(\log \ell - 1)} \cdot 2^{-(\ell - k - \delta \ell)} < \eta, \quad (*)$$

то и  $P(\Delta_{tk}, \ell, \delta \ell) < \eta$ . Из (\*) следует, что

$$\log (\gamma(\ell - \delta \ell) \cdot C_\ell^{\delta \ell} \cdot 2^{k(\log \ell - 1)} \cdot 2^{-(\ell - k - \delta \ell)}) < \log \eta,$$

$$\log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell)) + \log (2^{k(\log \ell - 1)}) + \log (2^{-(\ell - k - \delta \ell)}) < \log \eta,$$

$$\log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell)) + k(\log \ell - 1) - (\ell - k - \delta \ell) < \log \eta,$$

$$k(\log \ell - 1) - \ell + k + \delta \ell < \log \eta - \log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell)),$$

$$k \log \ell < \ell - \delta \ell + \log \eta - \log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell)).$$

Откуда следует утверждение теоремы

$$k < \frac{\ell - \delta \ell + \log \eta - \log (C_\ell^{\delta \ell} \cdot \gamma(\ell - \delta \ell))}{\log \ell}.$$

□

**Следствие 3.** Для того, чтобы вероятность  $P(\Delta_{tk}, \ell, 0)$  безошибочной настройки на выборку  $X^\ell$  длины  $\ell \geq 10$  алгоритмами семейства  $\Delta_{tk}$  была меньше некоторого заданного  $\eta > 0$  необходимо, чтобы число параметров  $k$  удовлетворяло условию

$$k < \frac{\ell + \log \eta}{\log \ell}.$$

*Доказательство.* Следует из теоремы 2, учитывая, что  $\delta \ell = 0$  и  $\gamma(\ell) \approx 1$  при  $\ell \geq 10$  [1]. □

Кроме этого в [2] была получена численная оценка зависимости вероятности неслучайной настройки на выборку от числа параметров  $k$ -параметрического полиномиального шаблона для булевого случая (рассматривается булева выборка  $X_{\{0,1\}}^\ell$ ).

**Следствие 4.** Для того, чтобы вероятность  $P(\Delta_k, \ell, \delta\ell)$  безошибочной настройки на выборку  $X_{\{0,1\}}^\ell$  длины  $\ell \geq 10$  алгоритмами семейства  $\Delta_k$ , – образа  $k$ -параметрического шаблона над полем  $\{0,1\}$ , – была меньше некоторого заданного  $\eta > 0$ , необходимо, чтобы число параметров  $k$  удовлетворяло условию

$$k < \ell - \log \ell + \log \eta.$$

Приведем таблицу, которая показывает каким должно быть число параметров  $k$ , чтобы вероятность неслучайной безошибочной настройки на выборку  $X_{\{0,1\}}^\ell$  длины  $\ell$   $k$ -параметрическим полиномиальным шаблоном была выше 90%.

$\ell$	$k \leq$
10	3
20	12
30	22
40	31
50	41
60	51

$\ell$	$k \leq$
70	61
80	70
90	80
100	90
1000	987
10000	9983

## 2. ТЕОРЕМА СУЩЕСТВОВАНИЯ ПОЛИНОМИАЛЬНОГО ШАБЛОНА

**Теорема 3** (существования полиномиального шаблона). Для произвольной корректной выборки  $X^\ell$  длины  $\ell$  существует  $k$ -параметрический полиномиальный шаблон  $t_k^k$ ,  $k \leq \ell$ , удовлетворяющий выборке  $X^\ell$ .

Здесь возникают следующие интересные задачи:

- задача поиска условий существования и вида обучающей выборки  $X^\ell$ , для которой не существует  $k$ -параметрического полиномиального шаблона с  $k < \ell$  параметрами удовлетворяющего выборке  $X^\ell$ ;
- задача нахождения минимального  $k$ , при котором для выборки  $X^\ell$  существует удовлетворяющий ей  $k$ -параметрический полиномиальный шаблон.

## ЗАКЛЮЧЕНИЕ

В работе приведены основные определения, положения и проблемы теории шаблонов. Введены новые типы шаблонов, такие как, универсальные и  $\alpha$ -шаблоны. Получено условие на количество параметров  $k$ -параметрического полиномиального шаблона  $t^k$ , при котором вероятность настройки на выборку алгоритмами семейства  $\Delta_k$  будет меньше некоторого  $\eta > 0$ .

Выделены новые задачи теории шаблонов, которые играют важную роль при построении качественных решающих правил и сужении области неопределенности при решении задач обучения по прецедентам.

**СПИСОК ЛИТЕРАТУРЫ**

1. *Donskoy V. I.* The Estimations Based on the Kolmogorov Complexity and Machine Learning from Examples // Proceedings of the Fifth International Conference «Neural Networks and Artificial Intelligence» (ICNNAI'2008), Minsk, 2008, - p.p. 292-297.
2. *Анафиев А.С.* Теория шаблонов в задачах обучения по прецедентам и выбора моделей. Диссертация на соискание ученой степени кандидата физико-математических наук. – 2007. – 135 с.
3. *Вапник В.Н.* Восстановление зависимостей по эмпирическим данным. М.: Наука, 1979. – 448 с.
4. *Воронцов К.В.* Вычислительные методы обучения по прецедентам. Курс лекций по машинному обучению. – 2009 г. – 42 с.  
<http://www.machinelearning.ru/wiki/images/8/8d/Voron-ML-Intro.pdf>

*Статья поступила в редакцию 21.12.2008*