УДК 519.7

ОЦЕНКА VCD ОБРАЗА k-ПАРАМЕТРИЧЕСКОГО ШАБЛОНА $t[+,\cdot]$

Анафиев А.С.

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.

In this article the notion of the template of set functions and various kinds of such templates are introduced. Are considered primary k-parametrical templates of a kind $t = t[+,\cdot]$.The *Vapnick-Chevronenkis Dimension* (VCD) for image of such templates is estimated. As well as VCD for the class of polynomials above a field of real (rational) numbers and the class of Zhegalkin polynomials are estimated.

Основные определения

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

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

Другими словами, шаблон — это оператор, аргументами которого являются элементы параметрического множества — *параметры*, а значением — функция множества $\mathfrak{A}$.

Определение 2. Множество всех k-параметрических шаблонов множества $\mathfrak{A}$ относительно параметрического множества Θ будем обозначать $T^k(\Theta)$ или просто T^k, т.е.

$$T^k(\Theta) = \{t \mid t : \Theta \to \mathfrak{A}, \Theta = P^k\}$$

Определение 3. Тройку $M = (\theta \mid t, \Theta)$ такую, что $t(\theta) = f$, где t — шаблон множества $\mathfrak{A}$ относительно параметрического множества Θ и $\theta \in \Theta$, будем называть *параметрической моделью* функции $f \in \mathfrak{A}$.

Определение 4. Множество

$$\Delta_t = \bigcup_{\theta \in \Theta} (\theta \mid t, \Theta)$$

будем называть *образом шаблона* t на множестве $\mathfrak{A}$.

Определение 5. Множество

$$\Delta_T = \bigcup_{t \in T} \Delta_t$$

будем называть *образом множества шаблонов* T на множестве $\mathfrak{A}$.

Определение 6. Множество шаблонов T называется *полным* на множестве $\mathfrak{A}$, если $\Delta_T = \mathfrak{A}$.

Определение 7. Шаблон t относительно параметрического множества Θ называется *корректным*, если для любого $\theta \in \Theta$ существует единственная функция $f \in \mathfrak{A}$ такая, что $f = t(\theta)$; и наоборот, для любой функции $f \in \Delta_t \subseteq \mathfrak{A}$ существует единственный параметр $\theta \in \Theta$ такой, что $t(\theta) = f$.

Определение 8. Шаблон t относительно параметрического множества Θ называется *простым*, если для любого $i = \overline{1,k}$ не существует такой функции $\hat{g} : P^r \to P$,что $\theta_i = \hat{g}(\theta_{i_1}, \ldots, \theta_{i_r})$, где $\Theta = P^k, i_j \neq i, j = \overline{1,r}$.

Другими словами, простой шаблон — это такой шаблон, все параметры которого не зависят друг от друга.

Рассмотрим k-местный простой шаблон t над параметрическим множеством $\Theta = P^k$. Тогда шаблон можно представить в виде

$$t(\theta_1, \ldots, \theta_k) = g(h_1(\theta_1; x_1, \ldots, x_n), \ldots, h_k(\theta_k; x_1, \ldots, x_n)),$$

где $x_1, \ldots, x_n$ — параметры данного шаблона, $g : P^k \to \mathfrak{A}$ и $h_j : P \to P, j = \overline{1,k}$.

Тогда для задания шаблона необходимо определить $k+1$ функцию $g, h_1, \ldots, h_k$. В этом случае шаблон t будем обозначать как $t[g; h_1, \ldots, h_k]$. Если при этом все $h_i = h, i = \overline{1,k}$, то шаблон t будем обозначать $t[g, h]$.

Пусть у вас имеется некоторая n-местная операция $\circ$ во множестве X. Будем говорить, что функция $f_\circ$ *реализует операцию* $\circ$, если $f(x_1, \ldots, x_n) = x_1 \circ x_2 \circ \cdots \circ x_n, x_i \in X, i = \overline{1,n}$. В обозначении шаблона $t[g, h_1, \ldots, h_k]$, если g или h_j, $j = \overline{1,k}$, реализуют некоторую операцию, то вместо функционального символа будем использовать символ этой операции.

1. Сложность класса решающих правил по В.Н. Вапнику

Пусть, как и прежде, X — множество объектов, Y — множество ответов. Класс решающих правил A, используемых для восстановления зависимостей на произвольной выборке $\widetilde{X}^l = \{(x^i \in X; y^i \in Y)\}_{i=1}^l$, может быть охарактеризован емкостью или, как еще говорят, VCD-размерностью (*Vapnik-Chervonenkis Dimension*), которая определяется следующим образом [1]. Каждое решающее правило $a \in A$ определяет двоичный вектор $a(x^1), a(x^2), \ldots, a(x^l)(1)$, где $a(x^i) = 0$, если $a(x^i) = y^i$, и $a(x^i) = 1$, если $a(x^i) \neq y^i$, $i = \overline{1,l}$. Число различных двоичных векторов вида (1) — способов разделения выборки $\widetilde{X}^l$ на два класса всеми правилами $a \in A$ — обозначается $\Delta^A(\widetilde{X}^l)$.

Определение 9. *Функцией роста* класса решающих правил A называется

$$m^A(l) = \max_{\widetilde{X}^l \in X^i \times Y} \Delta^A(\widetilde{X}^l),$$

где максимум берется по всевозможным выборкам $\widetilde{X}^l$ длины l.

Доказано [1], что функция роста либо тождественна 2^l, либо удовлетворяет неравенству

$$m_A(l) < 1,5\frac{l^{h_A}}{h_A!}$$

при $l > h_A$. Число h_A называют *емкостью* или VCD-*размерностью* класса A решающих правил.

Другими словами, емкость класса решающих правил определяет наибольшую длину выборки $l = h_A$, которую можно разделить на два класса всеми возможными 2^{h_A} способами всеми правилами $a \in A$, взятую по всем возможным выборкам $\widetilde{X}^l$ длины l для любого l. При этом, если $h_A = 2^l$, то говорят, что класс имеет бесконечную емкость.

Интерес к емкости как характеристике класса решающих правил объясняется следующим неравенством [1]

$$P(\sup_{a\in A} \mid P(a) - \nu(a) \mid > \varepsilon) < 9\frac{(2l)^{h_A}}{h_A!}e^{\frac{-\varepsilon^2 l}{4}},$$

где a — произвольное решающее правило, выбранное из класса A в процессе обучения, $\nu(a)$ — частота ошибок правила a на случайно извлеченной эмпирической выборке, $P(a)$ — неизвестная вероятность ошибки правила a на генеральной совокупности объектов, $\varepsilon > 0$. При этом конечность емкости h_A класса решающих правил A является достаточным условием равномерной сходимости частот к вероятностям при $l \to \infty$: т.е., если $h_A < \infty$, то справедливо

$$\forall\varepsilon > 0 \Rightarrow \lim_{l\to\infty} P(\sup_{a\in A} \mid P(a) - \nu(a) \mid > \varepsilon) \to 0.$$

При этом, чем меньше емкость h_A класса A, тем «быстрее» равномерная сходимость по классу A.

2. Оценка сложности образа k-параметрическкго шаблона $t[+,\cdot]$

Пусть $\mathfrak{F}$ некоторое поле с операциями сложения $+$ и умножения $\cdot$. В качестве множества объектов X будем рассматривать множество $X = \mathfrak{F}^n$, множество ответов Y — множество $Y = \mathfrak{F}$ и в качестве множества параметров — множество $\Theta = \mathfrak{F}^k$. Рассмотрим множество $T_{[+,\cdot]}$ всех шаблонов множества $\mathfrak{A}$ вида $t[+,\cdot]$ относительно параметрического множества Θ. Очевидно, что каждый шаблон $t \in T$ можно представить в виде

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

Выражение $\mathfrak{S} = x_{i_1^j} \cdot \dots \cdot x_{i_{r_j}^j}$ будем называть *свойством* ранга r_j множества X порождаемое шаблоном $t[+,\cdot]$. С учетом этого, произвольный шаблон $t \in T$ можно представить

$$t(\theta_1, \dots, \theta_k) = \sum_{j=1}^{k} \theta_k \cdot \mathfrak{S}_k,$$

где $\theta_j, j = \overline{1,k}$ степень выраженности свойства $\mathfrak{S}_k$. Таким образом, k-параметрический шаблон может определить степень выраженности у объекта ровно k свойств.

Обозначим через $P(r)$ число всех свойств ранга не больше r множества X. Тогда мощность множества T будет равна числу сочетаний из $P(r)$ свойств по k, которые шаблон t может различать. Т.е. число различных шаблонов t множества T будет равно $C^k_{P(r)}$.

Теорема 1. *Пусть $\mathfrak{F}$ некоторое поле с операциями сложения $+$ и умножения $\cdot$ и t — k-параметрический простой шаблон вида $t = t[+,\cdot]$ множества $\mathfrak{A} = \{f : \mathfrak{F}^n \to \mathfrak{F}\}$ относительно параметрического множества $\Theta = \mathfrak{F}^k$. Тогда сложность h_{Δ_t} образа Δ_t шаблона t равна k.*

Доказательство. Как уже отмечалось, емкость $h(A)$ класса решающих правил (функций) A — это наибольшая длина $l = h_s$ выборки, которую можно разделить на два класса с помощью правил семейства A всеми 2^{h_A} способами, взятую по всевозможным выборкам длины l при любом l.

Предположим, что нам дана произвольная выборка $X^l = \{(x^i \in \mathfrak{F}^n; y^i \in \mathfrak{F})\}_{i=1}^l$ длины l, где $x^i = (x^i_1, \dots, x^i_n), x^i_j \in \mathfrak{F}, j = \overline{1,n}$. Докажем утверждение теоремы.

Будем считать, что решающее правило $f = t(\theta)$ относит объект $(x^i; y^i)_{i=1}^l$ к j-ому классу, если $I(f(x^i),\ y^i) = j, j \in \{0,1\}$, где I — бинарный индикатор ошибки вида

$$I(f(x^i), y^i) = \begin{cases} 0, & \text{если } f(x^i) = y^i; \\ 1, & \text{если } f(x^i) \neq y^i. \end{cases}$$

Рассмотрим вектор $w = (w_1, \dots, w_l)$, где w_i — номер класса, к которому принадлежит прецедент $(x^i; y^i)$. Покажем, что с помощью выбора параметров $\theta_1, \dots, \theta_k$ мы можем получить все возможные 2^l различных булевых векторов w. Рассмотрим следующую систему l линейных уравнений с k неизвестными:

$$\begin{cases} f(x^1) = y^1; \\ \dots \\ f(x^l) = y^l. \end{cases} \qquad (*)$$

Так как шаблон $t = t[+,\cdot]$, то каждую функцию $f \in \Delta_t$ можно записать как $f(x) = \sum\limits_{j=1}^{k} \theta_j \cdot x_{i_1^j} \cdot \dots \cdot x_{i_{r_j}^j}$, а систему $(*)$ в виде:

$$\begin{cases} \sum\limits_{j=1}^{k} \theta_j \cdot x^1_{i_1^j} \cdot \dots \cdot x^1_{i_{r_j}^j} = y^1; \\ \dots \\ \sum\limits_{j=1}^{k} \theta_j \cdot x^l_{i_1^j} \cdot \dots \cdot x^l_{i_{r_j}^j} = y^1; \end{cases} \quad \text{или} \quad \begin{cases} \sum\limits_{j=1}^{k} \theta_j \alpha^1_j = y^1; \\ \dots \\ \sum\limits_{j=1}^{k} \theta_j \alpha^l_j = y^l, \end{cases} \qquad (**)$$

где $\alpha^i_j \in \mathfrak{F}$ и $\alpha^i_j = x^j_{i_1^j} \cdot \dots \cdot x^j_{i_{r_j}^j}$, $j = \overline{1,l}$.

Для того, чтобы доказать утверждение теоремы, покажем, что с помощью выбора параметров $\theta_1, \dots, \theta_k$(выбора различных функций) можно построить все 2^l различных булевых вектора w, если $l = k$, и нельзя, — при $l > k$.

I. Пусть $l = k$.

1) Предположим, что все уравнения системы $(**)$ линейно независимые, тогда всегда можно подобрать параметры $\theta_1, \dots, \theta_k$ так, чтобы равенства с номерами $\{i_1, \dots, i_j\}$ системы $(**)$, $j = \overline{0,k}$, выполнялись. В этом случае координаты w_{i_s},$s = \overline{1,j}$, вектора w будут равны нулю, а остальные — единице. Таким образом, если рассмотреть все случаи для всех $j = \overline{0,k}$, мы получим все 2^l различных вектора w. Следовательно, для $l = k$ произвольную выборку мы можем разделить на два класса всеми возможными 2^k случаями и утверждение теоремы выполняется.

2) Пусть среди уравнений системы имеются линейно зависимые. Не ограничивая общности, пусть это будут уравнения с номерами j' и j''. Тогда, если уравнение с номером j' выполняется ($w_{j'} = 0$), то, в силу линейной зависимости, и уравнение с номером j'' обратится в равенство($w_{j_{\prime\prime}} = 0$). Аналогично, обращение $w_{j'}$ в 1 будет означать обращение и $w_{j''}$ в 1. Тогда, в этом случае мы не сможем получить двоичный вектор, у которого $w_{j'} \neq w_{j''}$ и , следовательно, не сможем получить все двоичные вектора w. Отсюда следует, что нам не удасться разбить исходную выборку всеми 2^l способами.

II. Покажем, что при $l > k$ утверждение теоремы не выполняется.

1) Предположим, что все уравнения системы $(**)$ линейно независимы, тогда, учитывая, что число неизвестных системы $(**)$ будет меньше, чем число уравнений, система $(**)$ не будет иметь ни одного решения, а значит мы не сможем построить булевый вектор w, у которого $s > k$ нулевых координат. Отсюда следует, что при $l > k$, мы не сможем получить все возможные разбиения выборки на два класса.

2) Пусть теперь система $(**)$ содержит линейно зависимые уравнения. Тогда как и в случае I.2 мы не сможем построить все 2^l различные вектора w.

Таким образом, максимальное l, при котором произвольную выборку можно разделить на два класса всеми возможными 2^l способами, ровно k и $VCD(\Delta_t) = h_{\Delta_t} = k$.

Теорема доказана.

Теорема 2. *Пусть $\mathfrak{F}$ некоторое поле с операциями сложения $+$ и умножения $\cdot$ и T — множество всех параметрических шаблонов t, где t — k-параметрический простой шаблон вида $t = t[+,\cdot]$ множества $\mathfrak{A} = \{f : \mathfrak{F}^n \to \mathfrak{F}\}$ относительно параметрического множества $\Theta = \mathfrak{F}^k$. Тогда емкость h_{Δ_T} класса решающих правил Δ_T удовлетворяет неравенству $h_{\Delta_T} \leqslant k \log_2 C^k_{P(r)}$, где $P(r)$ — число свойств множества $\mathfrak{F}^n$ ранга не больше r.*

Доказательство. Учитывая, что число различных свойств ранга r равно, по условию теоремы, $P(r)$, то число всех шаблонов множества T равно числу сочетаний из $P(r)$ по k и равно $C^k_{P(r)}$. По теореме 1 максимальная длина l выборки, которую каждый шаблон $t \in T$ может разделить всеми 2^l способами на два класса, равна k. Следовательно, каждый шаблон $t \in T$ способен построить 2^k различных булевых векторов. Таким образом, число всех возможных булевых векторов, которые могут построить все шаблоны $t \in T$, не превзойдет $2^k C^k_{P(r)}$.Отсюда следует, что длина выборки $l = h_{\Delta_T}$, которую можно разделить на два класса всеми 2^l способами всеми шаблонами $t \in T$, не превзойдет $h_{\Delta_T} \leqslant \log_2 2^k C^k_{P(r)} = k \log_2 C^k_{P(r)}$.

Теорема доказана.

Следствие 1. *Класс полиномов $P_{\mathbb{R}}(p,k)(P_{\mathbb{Q}}(p,k))$ степени p с k отличными от нуля коэффициентами над полем действительных(рациональных) чисел имеет емкость (VCD)*

$$h_p \leqslant k \log C^k_{\frac{(n+p)!}{n!p!}}$$

Доказательство. Очевидно, что класс полиномов степени p с k отличными от нуля коэффициентами над полем действительных (рациональных) чисел является образом множества T всех простых k-параметрических шаблонов $t = t[+,\cdot]$ множества $\mathfrak{A} = \{f : \mathbb{R}^n \to \mathbb{R}\}(\mathfrak{A} = \{f : \mathbb{Q}^n \to \mathbb{Q}\})$ относительно параметрического множества $\Theta = \mathbb{R}^k(\Theta = \mathbb{Q}^k)$, где $+$ и $\cdot$ — это обычное сложение и умножение. Учитывая, что степени полиномов множества $P_{\mathbb{R}}(p,k)(P_{\mathbb{Q}}(p,k))$ не превосходят p, то число всех свойств множества Δ_T будет равным $\frac{(n+p)!}{n!p!}$. Тогда из теоремы 2 следует утверждение следствия.

Следствие 2. *Класс $P_{\mathbb{B}}(p,k)$ полиномов Жегалкина* [2] *степени p с k отличными от нуля коэффициентами имеет емкость (VCD) $h_P \leqslant k \log C^k_{\sum_{i=0}^{p} C^i_n}$.*

Доказательство. Покажем, что множество $\mathbb{B} = \{0,1\}$ является полем относительно операции $\oplus$ сложения по mod 2 и конъюнкции &. Очевидно, что операции $\oplus$ и & коммутативные, ассоциативные и связаны между собой законом дистрибутивности, т.е. $\forall a, b, c \in \mathbb{B}$ справедливо:

$$a \oplus b = b \oplus a,\ a\&b = b\&a,$$
$$(a \oplus b) \oplus c = a \oplus (b \oplus c),\ (a\&b)\&c = a\&(b\&c),$$
$$a\&(b \oplus c) = a\&b \oplus a\&c.$$

Нулем поля является 0, т.к. $a \oplus 0 = a$, противоположным для a будет сам $a\,(a \oplus a = 0)$, единичным — 1 и для 1 существует обратный $a^{-1} = 1$, т.к. $a\&a^{-1} = 1\&1 = 1$.

При этом класс $P_{\mathbb{B}}(p, k)$ является образом множества T при всех простых k-параметрических шаблонов $t = t[\oplus, \&]$ множества $\mathfrak{A} = \{f\colon \mathbb{B}^n \to \mathbb{B}\}$ относительно параметрического множества $\Theta = \mathbb{B}^k$. Учитывая, что степени полиномов множества $P_{\mathbb{B}}(p, k)$ не превосходят p, то число всех свойств множества Δ_T будет равным $\sum\limits_{i=0}^{p} C_n^i$. Тогда по теореме 2 следует утверждение следствия.

Заключение

В данной работе было введено новое понятие *шаблона множества функций*, а также понятия *k-параметрического, простого, полного, корректного шаблонов*, понятия *параметрической модели функции и образа шаблона*. Найдена емкость образа k-параметрического простого шаблона вида $t = t[+, \cdot]$. Получены оценки емкости класса полиномов над полем вещественных (рациональных) чисел и класса полиномов Жегалкина. Сделана попытка двухэтапного оценивания сложности класса решающих правил.

Список литературы

1. *Вапник В.Н.* Восстановление зависимостей по эмпирическим даным. — М.: Наука, 1979. — 448с.
2. *Закревский А.Д., Торопов Н.Р.* Полиномиальная реализация частичных булевых функций и систем. Изд. 2-е, стереотипное. — М.: Едиториал УРСС, 2003. — 200с.