О ДОСТАТОЧНОЙ ДЛИНЕ ОБУЧАЮЩЕЙ ВЫБОРКИ ДЛЯ КОМИТЕТНОГО РЕШАЮЩЕГО ПРАВИЛА

М.Ю. Хачай

Институт математики и механики Уральского отделения РАН
620219, Россия, г. Екатеринбург, ГСП-384, ул. С. Ковалевской, 16
mkhachay@imm.uran.ru

ABSTRACT

In this paper some aspects of solving of the pattern recognition problem with the empirical risk minimization method in committee decision rules classes (perceptrons) are discussed. Particularly a new approach to estimate the sufficient length of a training sequence is proposed.
Keywords: committee decision rule, uniform convergence of frequency to probability, VC–dimension.

ВВЕДЕНИЕ

Классы комитетных (персептронных) решающих правил, определенные над фиксированным базовым классом, имеющим достаточно простую структуру, активно используются при решении задачи обучения распознаванию образов методом минимизации эмпирического риска (см. напр. [1-2]). Рассмотрим задачу двуклассового распознавания образов. Пусть пространство описаний объектов – *X,* а базовый класс решающих правил – $F = \{f(x;\alpha) | \alpha \in \Lambda\} \subset [X \to \{0,1\}]$ (договоримся, как и в [3] полагать, что правило $f(x,\alpha)$ относит объект с описанием x к первому классу, если $f(x;\alpha) = 1$).

Класс F для каждого нечетного q порождает класс F_q комитетных решающих правил:

$$F_q = \{f(x;\alpha_1,\ldots,\alpha_q) = \Theta(\sum_{i=1}^{q} f(x;\alpha_i) - \frac{q}{2}) | f(x;\alpha_i) \in F\},\ \text{где}\ \Theta(t) = \begin{cases} 1, & \text{если } t > 0, \\ 0, & \text{иначе.} \end{cases}$$

Как обычно, полагаем, что задано вероятностное пространство $(X \times \Omega, A, P)$, где $\Omega = \{0,1\}$, а мера P известна с точностью до конечной выборки $(x^1,\omega^1),(x^2,\omega^2),\ldots,(x^l,\omega^l)$. Допущение равнозначности ошибок классификации первого и второго рода приводит к тому, что задача обучения распознаванию образов в классе правил F_q эквивалентна задаче

$$\inf_{\beta=(\alpha_1,\ldots,\alpha_q)} \int_{X\times\Omega} (\omega - f(x;\beta))^2\, dP(x,\omega)\,.$$

В рамках метода минимизации эмпирического риска в качестве аппроксимации этой задачи используется задача

$$\min_{\beta=(\alpha_1,\ldots,\alpha_q)} \frac{1}{l}\sum_{i=1}^{l} (\omega^i - f(x^i;\beta))^2\ ,$$

которая всегда разрешима и явно не зависит от неизвестной меры *P*. Известно, (см. напр. [3]) что для корректности такого перехода достаточно существование равномерной сходимости частоты к вероятности для класса событий $\{Q_\beta = \{(x,\omega) | f(x;\beta) = f(x;\alpha_1,\ldots,\alpha_q) \neq \omega\}\}$.

Достаточные условия такой сходимости, полученные в общем случае Вапником и Червоненкисом, выражаются в терминах функции роста (или, более грубо, емкости) класса событий. Известные верхние оценки функции роста (и емкости) для класса комитетных решающих правил справедливы для более широких классов. Кроме того, например, для случая, когда базовым является класс линейных решающих правил, верхние оценки емкости асимптотически совпадают с нижними [4]. В статье предлагается несколько иной подход к выводу достаточных условий равномерной сходимости для класса событий, соответствующего классу F_q, исходя из условия существования такой сходимости для класса событий, соответствующего F.

ОЦЕНКА СКОРОСТИ РАВНОМЕРНОЙ СХОДИМОСТИ ЧАСТОТЫ К ВЕРОЯТНОСТИ ДЛЯ КЛАССА “КОМИТЕТНЫХ” СОБЫТИЙ

Пусть задано вероятностное пространство (X, A, P), в котором фиксирована система событий $S_1 = \{S_\alpha \mid \alpha \in \Lambda\} \subset A$, называемая в дальнейшем базовой. Для заданного нечетного q рассмотрим систему “комитетных” событий S_q, порожденную системой S_1, каждое событие Q_β которой, где $\beta = (\alpha_1, \alpha_2, \ldots, \alpha_q)$, определяется соотношением

$$Q_\beta = \bigcup_{L \subseteq N_q, |L| = \lceil q/2 \rceil} \prod_{i \in L} S_{\alpha_i}, \qquad S \quad q \ , \ .$$

Предполагая, что для системы событий S_1 справедлива равномерная сходимость частоты к вероятности (по вероятности), скорость которой оценивается соотношением:

$$P\{\sup_\alpha |\nu(S_\alpha) - P(S_\alpha)| > \varepsilon\} < C_1(l)\exp(-C_2(l)\varepsilon^2),$$

при добавочных предположениях получим аналогичную оценку для системы S_q.

Воспользуемся следующими обозначениями. Для выборки $x(2l) = (x^1, x^2, \ldots, x^{2l})$ и события A через $\nu_1(A)$ и $\nu_2(A)$ обозначим частоты выпадения события в первой и второй полувыборках соответственно.

ЛЕММА. *Пусть заданы события* $S_1, S_2, \ldots, S_q \subset S_1$, $q = 2s+1$ *и* $Q = Q_\beta$, *где* $\beta = (1, 2, \ldots, q)$. *Если выборка* $x(2l)$ *такова, что для каждого* $L \subset N_q$ *такого, что* $|L| > s+1$ *справедливы равенства*

$$\nu_1\Big(\prod_{i \in L} S_i \setminus \bigcup_{j \notin L} S_j\Big) - \nu_2\Big(\prod_{i \in L} S_i \setminus \bigcup_{j \notin L} S_j\Big) = 0 \ \text{и} \ \nu_1\Big(\prod_{j \notin L} S_i \setminus \bigcup_{i \in L} S_j\Big) - \nu_2\Big(\prod_{j \notin L} S_i \setminus \bigcup_{i \in L} S_j\Big) = 0, \tag{1}$$

то

$$|\nu_2(Q) - \nu_1(Q)| = \frac{1}{q} \max_{i = \overline{1,q}} |\nu_2(S_i) - \nu_1(S_i)|$$

Доказательство.

Введем обозначения. Для каждого подмножества $L \subseteq N_q$ через z_L обозначим $\nu_2(A_L) - \nu_1(A_L)$, где $A_L = \prod_{i \in L} S_i \setminus \bigcup_{j \notin L} S_j$. По условию, $z_L = 0$ для каждого L такого, что $\left| |L| - \frac{q}{2} \right| > 1$.

В выбранных обозначениях $\nu_2(S_i) - \nu_1(S_i) = \sum_{i \in L} z_L = \sum_{i \in L, |L| = s} z_L + \sum_{i \in L, |L| = s+1} z_L$, а $\nu_2(Q) - \nu_1(Q) = \sum_{|L| = s+1} z_L$.

Пусть $|\nu_2(Q) - \nu_1(Q)| = a > 0$, рассмотрим случай $\nu_2(Q) - \nu_1(Q) = a$ (в случае $\nu_2(Q) - \nu_1(Q) = -a$ ход доказательства аналогичен)

Покажем, что

$$\min_{x(2l)} \{\max_{i = \overline{1,q}} |\nu_2(S_i) - \nu_1(S_i)| : \nu_2(Q) - \nu_1(Q) = a, (1)\} = \frac{a}{q}. \tag{2}$$

В самом деле, задача (2) эквивалентна задаче линейного программирования

$$\begin{aligned}
&\min u\\
&\left.\begin{aligned}
&\sum_{i\in L,|L|=s} z_L + \sum_{i\in L,|L|=s+1} z_L + u \ge 0\\
&-\sum_{i\in L,|L|=s} z_L - \sum_{i\in L,|L|=s+1} z_L + u \ge 0
\end{aligned}\right\}\ i=\overline{1,q}\\
&\sum_{|L|=s+1} z_L = a\\
&\sum_{|L|=s} z_L + \sum_{|L|=s+1} z_L = 0
\end{aligned} \tag{3}$$

Сопоставим ей двойственную задачу:

$$\begin{aligned}
&\max a w_1\\
&\sum_{i\in L}(v_i^+ - v_i^-) + w_2 = 0 \qquad (|L|=s)\\
&\sum_{i\in L}(v_i^+ - v_i^-) + w_1 + w_2 = 0 \quad (|L|=s+1)\\
&\sum_{i=1}^{q}(v_i^+ + v_i^-) = 1\\
&v_i^+, v_i^- \ge 0 \qquad\qquad (i=\overline{1,q})
\end{aligned} \tag{3^*}$$

и рассмотрим векторы $[\bar z, \bar u]$ и $[\bar v, \bar w]$, определяемые соотношениями:

$$\overline{z_L} = \begin{cases} a\left(C_q^s\right)^{-1}, & \text{если } |L| = s+1,\\ -a\left(C_q^s\right)^{-1}, & \text{иначе,}\end{cases} \qquad \bar u = \frac{a}{q}$$

и

$$\overline{v_i^+} = 0,\ \overline{v_i^-} = \frac{1}{q} \qquad (i=\overline{1,q})$$

$$\bar w_1 = \frac{1}{q},\quad \bar w_2 = \frac{s}{q}.$$

Нетрудно убедиться, что векторы $[\bar z, \bar u]$ и $[\bar v, \bar w]$ допустимы в задачах (3) и (3^*) соответственно и значения целевых функций задач, вычисленные в них – совпадают. Следовательно, $[\bar z, \bar u]$ и $[\bar v, \bar w]$ оптимальны каждый в своей задаче и оптимальные значения задач (3), (3^*) и (2) равны $\frac{a}{q}$.

Из леммы непосредственно следует

ТЕОРЕМА. *Пусть вероятностная мера P и система событий S_1 таковы, что для каждого подмножества $L \subseteq N_q$, $|L| > s+1$ справедливы равенства:*

$$P\Big(\prod_{i\in L} S_i \setminus \bigcup_{j\notin L} S_j\Big) = P\Big(\prod_{j\notin L} S_i \setminus \bigcup_{i\in L} S_j\Big) = 0.$$

Тогда для каждого $\varepsilon \in (0,1)$ справедливы неравенства

$$P\{\sup_\beta |\nu_2(Q_\beta) - \nu_1(Q_\beta)| > \varepsilon\} \le P\{\sup_\alpha |\nu_2(S_\alpha) - \nu_1(S_\alpha)| > \frac{\varepsilon}{q}\}.$$

Далее, дважды применив основную лемму равномерной сходимости [3], получим несколько следствий.

СЛЕДСТВИЕ 1. *Пусть выполнено условие теоремы и для каждого* $\varepsilon \in (0,1)$ *справедлива оценка* $P\{\sup_{\alpha} | \nu(S_\alpha) - P(S_\alpha) | > \varepsilon\} < C_1(l)\exp(-C_2(l)\varepsilon^2)$, *где* $C_1(l)$ *и* $C_2(l)$ *не зависят от* ε. *Тогда для системы* S_q *верна оценка*:

$$P\{\sup_{\beta} | \nu(Q_\beta) - P(Q_\beta) | > \varepsilon\} < 4C_1(l)\exp(-C_2(l)\frac{\varepsilon^2}{16q^2})\} .$$

Если для системы S_1 выполнено достаточное условие равномерной сходимости, то справедливо

СЛЕДСТВИЕ 2. *Пусть выполнено условие теоремы и функция роста системы* S_1 $m^{S_1}(l) \neq 2^l$, *тогда*

$$P\{\sup_{\beta} | \nu(Q_\beta) - P(Q_\beta) | > \varepsilon\} < 6m^{S_1}(2l)\exp(-\frac{\varepsilon^2}{4q^2}(l-1))\} .$$

ОЦЕНКА ДОСТАТОЧНОЙ ДЛИНЫ ВЫБОРКИ

Возвращаясь к задаче обучения распознаванию образов, введем обозначения:

$$S_1 = \{S_\alpha = \{(x,\omega) | f(x;\alpha) \neq \omega\}, \alpha \in \Lambda\}, \quad S_q = \{Q_\beta = \{(x,\omega) | f(x;\beta) \neq \omega\}, \beta \in \Lambda^q\} .$$

Тогда следствия 1 и 2 могут быть переформулированы следующим образом.

СЛЕДСТВИЕ 3.

1. *Пусть выполнено условие следствия 1, тогда для выборки длины* l *и числа* $\eta \in (0,1)$ *с надежностью* $1-\eta$ *можно утверждать, что для всех* $\beta \in \Lambda^q$ *справедливы неравенства*
$$\nu(Q_\beta) - \frac{4q}{\sqrt{C_2(l)}}\sqrt{\ln\left(\frac{4C_1(l)}{\eta}\right)} \le P(Q_\beta) \le \nu(Q_\beta) + \frac{4q}{\sqrt{C_2(l)}}\sqrt{\ln\left(\frac{4C_1(l)}{\eta}\right)}.$$
2. *Пусть выполнено условие следствия 2, тогда при аналогичных предположениях справедливы неравенства:*
$$\nu(Q_\beta) - \frac{2q}{\sqrt{l-1}}\sqrt{\ln\left(\frac{6m^{S_1}(2l)}{\eta}\right)} \le P(Q_\beta) \le \nu(Q_\beta) + \frac{2q}{\sqrt{l-1}}\sqrt{\ln\left(\frac{6m^{S_1}(2l)}{\eta}\right)}.$$

ЗАКЛЮЧЕНИЕ

Предложенный в работе подход позволяет доказывать достаточные условия равномерной сходимости частоты к вероятности для классов “комитетных” событий при более общих условиях, нежели условие конечной емкости базового класса событий. По-прежнему открытым остается вопрос конструктивного описания классов событий, для которых справедливо условие теоремы. Тем не менее, известны нетривиальные классы событий, для которых оно справедливо.

ЛИТЕРАТУРА

1. Мазуров Вл.Д. Метод комитетов в задачах оптимизации и классификации. М.: Наука. 1990.
2. Jain A.K., Duin R.P.W., Mao J. Statistical Pattern Recognition: a review. // IEEE Trans. On Pattern Analysis and Machine Intelligence. 2000, vol.22, N.1
3. Вапник В.Н. Восстановление зависимостей по эмпирическим данным. М.: Наука. 1979.
4. Хачай М.Ю. Об оценке емкости класса комитетов линейных разделяющих функций. В сб. тез. Конф. ММРО-9. М.: ВЦ РАН. 1999. С.121-123.