Концентрация меры, неравенство Талаграна

И. О. Толстихин
iliya.tolstikhin@gmail.com

19 марта 2011 г.

Изучение концентрации случайных величин вокруг их математических ожиданий является одной из важных задач в теории статистического обучения (statistical learning theory). Зная, что случайная величина с достаточной вероятностью сосредоточена вокруг её математического ожидания, мы можем перейти от изучения распределения случайной величины к оцениванию её математического ожидания, что зачастую является куда более лёгкой задачей.

Данный текст посвящен одному из подходов к изучению концентрации меры, часто применяемому в последнее время в теории машинного обучения. Основной движущей силой этого подхода является неравенство Талаграна (M. Talagrand) [1].

Первые два раздела имеют обзорный характер. Мы попытаемся наглядно продемонстрировать метод Талаграна на примере задачи о распределении длины максимальной возрастающей подпоследовательности случайной последовательности. В третьем разделе приводится неравенство концентрации меры, полученное для функционала равномерного отклонения частот с помощью описанного метода.

Содержание

1 Неравенство Талаграна

Основная идея, на которой держится результат, представленный в этом разделе, заключается во введении нового способа измерения расстояния от точки $x \in \Omega^N$ до подмножества $A \subseteq \Omega^N$.

Введем следующим образом определенные множества:

$$U_A(x) = \bigl\{u \in \{0,1\}^N : u_i = 0 \Rightarrow \exists y \in A \colon y_i = x_i\bigr\};$$

$$V_A(x) = \mathrm{conv}\bigl(U_A(x)\bigr) \subseteq \mathbb{R}^N.$$

Очевидно, что $\mathbf{1} \in V_A(x)$, а $\mathbf{0} \in V_A(x) \Leftrightarrow x \in A$. Если мы посчитаем евклидово расстояние от точки $\mathbf{0}$ до множества $V_A(x)$, то получим *выпуклое расстояние* Талаграна:

Определение 1.1 (выпуклое расстояние Талаграна)

$$f_c(A,x) = \min_{v \in V_A(x)} \|v\| = \min_{v \in V_A(x)} \sqrt{\sum_{i=1}^N v_i^2}.$$

Следующая теорема ([1], теорема 4.1.1) является основным средством получения неравенств для концентрации меры функций, зависящих от случайных величин из Ω^N.

Теорема 1.1 (неравенство Талаграна) *Пусть $A \subseteq \Omega^N$, x выбирается случайным образом равномерно из Ω^N. Тогда*

$$\mathsf{P}(x \in A)\mathsf{E}\Bigl(\exp \frac{1}{4} f_c^2(A,x)\Bigr) \leqslant 1.$$

Если ввести определение *t-расширения* множества A:

Определение 1.2 (t-расширение множества)

$$A_t^c = \{x \in \Omega^N : f_c(A, x) \leqslant t\},$$

то мы получим следующее полезное следствие из теоремы (1.1):

Следствие 1.1

$$\mathsf{P}\big(A_t^c(A)\big) \geqslant 1 - \frac{e^{-t^2/4}}{\mathsf{P}(A)}.$$

Доказательство.

$$\mathsf{P}\big(A_t^c(A)\big) = \mathsf{P}\big(f_c(A, x) \leqslant t\big) = 1 - \mathsf{P}\big(f_c(A, x) \geqslant t\big) =$$

$$= 1 - \mathsf{P}\Big(\exp\frac{1}{4}f_c^2(A, x) \geqslant \exp t^2/4\Big) \overset{\text{н-во Маркова}}{\geqslant} 1 - \frac{\mathsf{E}\Big(\exp\frac{1}{4}f_c^2(A, x)\Big)}{\exp t^2/4} \overset{\text{Теорема (1.1)}}{\geqslant} 1 - \frac{\exp -t^2/4}{P(A)}.$$

■

Для лучшего понимания устройства множества A_t^c нам пригодится следующая лемма:

Лемма 1.1 *Следующие утверждения эквивалентны:*

$$x \in A_t^c, \text{ то есть } f_c(A, x) \leqslant t;$$
$$\forall \alpha \in \mathbb{R}^N \exists y \in A\colon \sum_{i=1}^{N} [x_i \neq y_i]\alpha_i \leqslant t\|\alpha\|.$$

2 Применение: длина максимальной возрастающей подпоследовательности

Как мы видим, теорема (1.1) описывает свойства *множеств* в пространстве, но никак не затрагивает поведение случайных величин. Тем не менее, на основе этого неравенства можно получать мощные результаты о распределении случайных величин.

В этом разделе приводится наглядный пример применения неравенства Талаграна для получения утверждений о распределении случайных величин.

Задача: Пусть $\Omega = [0, 1]$, $x = (x_1, \ldots, x_N)$, $x_i \in \Omega$. Пусть координаты вектора x распределены независимо и равномерно на Ω. Введем случайную величину $L_N(x)$ — длина максимальной возрастающей подпоследовательности координат x. Более строго,

$$L_N(x) = \max\big\{p \in \mathbb{N};\ \exists\{i_1 < \cdots < i_p\} \subset \{1, \ldots, N\}\colon x_{i_1} < x_{i_2} < \cdots < x_{i_p}\big\}.$$

Нашей ближайшей целью будет получение неравенства вида

$$\mathsf{P}\big(L_N(x) \geqslant M + t\big) \leqslant \eta(t), \tag{1}$$

где M — медиана случайной величины $L_N(x)$.

Решение: Нам уже известно что-то о мерах множеств A_t^c (t-расширения множества A) и A. Попробуем связать эти множества со значениями случайной величины $L_N(x)$. Введём следующее определение:

Определение 2.1 $S(a) = \{x \in \Omega^N : L_N(x) \leqslant (a)\}$,

и докажем утверждение ([1], лемма 7.1.1):

Утверждение 2.1 $\forall x \in \Omega^N$*:*

$$a \geqslant L_N(x) - f_c\big(S(a), x\big)\sqrt{L_n(x)}.$$

Доказательство. Пусть $L_N(x) = b$. Мы можем указать $I = \{i_1 < \dots < i_b\} \subseteq \{1, \dots, N\}$ такое, что $x_{i_1} < x_{i_2} < \dots < x_{i_b}$. Воспользуемся леммой (1.1) с вектором α: $\alpha_i = [i \in I]$, $i = 1, \dots, N$ и множеством $A = S(a)$:

$$\exists y \in S(a)\colon \sum_{i=1}^{N} [x_i \neq y_i][i \in I] \leqslant f_c\big(S(a), x\big) \sqrt{\sum_{i=1}^{N} \alpha_i^2} = f_c\big(S(a), x\big)\sqrt{b}.$$

Обозначим множество $\{i \in I\colon x_i \neq y_i\} = J$. Тогда

$$|J| \leqslant f_c\big(S(a), x\big)\sqrt{b}. \tag{2}$$

Поскольку $x_{i_1}, x_{i_2}, \dots, x_{i_b}$ — возрастающая подпоследовательность, то и y_i, $i \in I \setminus J$, является возрастающей. Но $y \in S(a)$, а значит $L_N(y) \leqslant a$, то есть

$$|I \setminus J| = |I| - |J| \leqslant a. \tag{3}$$

Объединяя (2) и (3) мы получаем:

$$a \geqslant |I| - |J| \geqslant L_N(x) - f_c\big(S(a), x\big)\sqrt{L_N(x)}.$$

■

Теперь мы можем получить искомую связь между значениями случайной величины $L_N(x)$ и расстоянием $f_c\big(S(a), x\big)$:

Следствие 2.1

$$\{x \in \Omega^N\colon L_N(x) \geqslant a + v\} \subseteq \Big\{x \in \Omega^N\colon f_c\big(S(a), x\big) \geqslant \frac{v}{\sqrt{a+v}}\Big\}.$$

Следствие очевидно, поскольку функция $\frac{x-a}{\sqrt{x}}$ возрастает при $x \geqslant a$.

Теперь мы имеем все необходимое для доказательства следующей теоремы ([1], теорема 7.1.2):

Теорема 2.1 *$\forall t > 0$:*

$$\mathsf{P}\big(L_N(x) \geqslant M + t\big) \leqslant 2e^{-\frac{t^2}{4(M+t)}},$$

где M — медиана случайной величины $L_N(x)$.

Доказательство.

$$\mathsf{P}\big(L_N(x) \geqslant M + t\big) \overset{\text{Следствие (2.1)}}{\leqslant} \mathsf{P}\Big(f_c\big(S(M), x\big) \geqslant \frac{t}{\sqrt{M+t}}\Big) =$$

$$= 1 - \mathsf{P}\Big(f_c\big(S(M), x\big) \leqslant \frac{t}{\sqrt{M+t}}\Big) \overset{\text{Следствие (1.1)}}{\leqslant} \frac{e^{-\frac{t^2}{4(M+t)}}}{\mathsf{P}\big(S(M)\big)} = 2\,e^{-\frac{t^2}{4(M+t)}}.$$

■

3 Концентрация меры функционала равномерного уклонения частот

Пусть задано упорядоченное множество объектов $\mathbb{X} = \{x_1, \dots, x_L\}$, которое мы называем *генеральной выборкой*. При случайной реализации перестановки π, выбранной равномерно из симметрической группы S_L, генеральная выборка разбивается на *обучающую выборку* $X(\pi)$ длиной ℓ и *контрольную выборку* $\bar{X}(\pi)$ длиной $k = L - \ell$:

$$X(\pi) = \{x_i\colon \pi(i) \leqslant \ell\};$$
$$\bar{X}(\pi) = \{x_i\colon \pi(i) > \ell\}.$$

Перестановка π естественным образом действует на множество $\mathbb{X}$ перестановкой его элементов. Всевозможные перестановки элементов генеральной выборки будем обозначать $\pi\mathbb{X}$, $\pi \in S_L$. Таким образом в обучающую выборку $X(\pi)$ попадают первые ℓ объектов упорядоченного множества $\pi\mathbb{X}$.

Определим случайную величину $Z(\pi)$:

$$Z(\pi) = \sup_{a\in A}\Big\{\frac{1}{k}n\big(a,\bar{X}(\pi)\big) - \frac{1}{\ell}n\big(a,X(\pi)\big)\Big\}. \tag{4}$$

Это функционал *равномерного уклонения частот на классе* A. В данном случае A — класс алгоритмов, которые мы будем отождествлять с их бинарными векторами ошибок: $a \in A$, $a \equiv \{I(a,x_i)\}_{i=1}^{L}$, где

$$I(a,x_i) = \begin{cases} 1, & \text{если алгоритм } a \text{ ошибается на объекте } x_i;\\ 0, & \text{иначе.}\end{cases}$$

Как и раньше, мы хотим получить неравенство вида (1).

Вспомним, что перестановку π можно отождествить с вектором $\pi_i = \pi(i)$, $i = 1,\dots,L$. Таким образом, $S_L \subseteq \{1,\dots,L\}^L$. Если бы S_L совпадало со всем множеством $\{1,\dots,L\}^L$, то для получения искомого результата мы могли бы воспользоваться теоремой (1.1). Однако, S_L является лишь малой частью множества $\{1,\dots,L\}^L$. К счастью, справедлива следующая теорема, аналогичная теореме (1.1):

Теорема 3.1 (неравенство Талаграна для случайных перестановок) *Пусть $A \subseteq S_L$, π выбирается случайным образом равномерно из S_L. Тогда*

$$\mathsf{P}(\pi\in A)\mathsf{E}\Big(\exp\frac{1}{16}f_c^2(A,\pi)\Big) \leqslant 1.$$

Следствие (1.1) также остаётся в силе с точностью до замены $\frac{1}{4}$ в показателе экспоненты на $\frac{1}{16}$.

Попробуем повторить шаги, сделанные в прошлом разделе. Пусть $Z(\pi) = b$.

Свяжем значение случайной величины $Z(\pi)$ с расстоянием $f_c(S(c),\pi)$. Для этого мы снова воспользуемся леммой (1.1), которая утверждает, что для любого $\alpha \in \mathbb{R}^L$ найдётся перестановка $\pi' \in S(c) = \{\tau \in S_L \colon Z(\tau) \leqslant c\}$ такая, что

$$\sum_{i=1}^{L}[\pi_i \neq \pi'_i]\alpha_i \leqslant f_c\big(S(c),\pi\big)\|\alpha\|. \tag{5}$$

Воспользуемся этой леммой для α, подобранного следующим образом:

$$\alpha_i = \begin{cases} 1, & \text{если } x_i \in \bar{X}(\pi);\\ 0, & \text{иначе.}\end{cases}$$

Тогда неравенство (5) примет виде:

$$\sum_{i\colon x_i\in\bar{X}(\pi)}[\pi_i \neq \pi'_i] \leqslant f_c\big(S(c),\pi\big)\sqrt{k}. \tag{6}$$

Заметим, что случайная величина $Z(\pi)$ (4) зависит лишь от разбиения генеральной выборки $\mathbb{X}$ на обучающую $X(\pi)$ и контрольную $\bar{X}(\pi)$ подвыборки и *не зависит от порядка элементов в подвыборках $X(\pi)$ и $\bar{X}(\pi)$*. Множество всех перестановок S_L разбивается на C_L^ℓ классов эквивалентности, в пределах которых значение случайной величины $Z(\pi)$ постоянно. Это даёт нам возможность найти перестановку $\pi'' \in S_L$ из класса эквивалентности перестановки π', удовлетворяющую следующим условиям:

$$\begin{cases} Z(\pi') = Z(\pi'');\\ \sum_{i=1}^{L}[\pi_i \neq \pi''_i] = |X(\pi)\cap\bar{X}(\pi'')| + |\bar{X}(\pi)\cap X(\pi'')|.\end{cases} \tag{7}$$

Перестановка π'' также лежит во множестве $S(c)$ и

$$\pi_i \neq \pi''_i \Leftrightarrow \left[\begin{array}{l} x_i \in X(\pi) \cap \bar{X}(\pi''); \\ x_i \in \bar{X}(\pi) \cap X(\pi''). \end{array}\right. \tag{8}$$

Более того, очевидно, что

$$\left|X(\pi) \cap \bar{X}(\pi'')\right| = \left|\bar{X}(\pi) \cap X(\pi'')\right|.$$

С учетом (7) и (8), если в неравенстве (6) перестановку π' заменить на π'', оно примет следующий вид:

$$\left|\bar{X}(\pi) \cap X(\pi'')\right| \leqslant f_c\bigl(S(c), \pi\bigr)\sqrt{k}. \tag{9}$$

Остается заметить, что для любых перестановок $\tau, \tau' \in S_L$ выполнено следующее утверждение:

$$Z(\tau) \leqslant Z(\tau') + \frac{L}{\ell k} \sup_{a \in A} \Bigl\{ n\bigl(a, \bar{X}(\tau) \cap X(\tau')\bigr) - n\bigl(a, X(\tau) \cap \bar{X}(\tau')\bigr) \Bigr\}.$$

Таким образом, мы получаем:

$$Z(\pi) \leqslant Z(\pi'') + \frac{L}{\ell k} \sup_{a \in A} \Bigl\{ n\bigl(a, \bar{X}(\pi) \cap X(\pi'')\bigr) - n\bigl(a, X(\pi) \cap \bar{X}(\pi'')\bigr) \Bigr\} \leqslant \tag{10}$$

$$\leqslant c + \frac{L}{\ell k} \left|\bar{X}(\pi) \cap X(\pi'')\right| \overset{\text{нер-во (9)}}{\leqslant} c + \frac{L\sqrt{k}}{\ell k} f_c\bigl(S(c), \pi\bigr). \tag{11}$$

Мы получили утверждение, аналогичное утверждению (2.1):

Утверждение 3.1 *Для любой перестановки $\pi \in S_L$:*

$$Z(\pi) \leqslant c + \frac{L\sqrt{k}}{\ell k} f_c\bigl(S(c), \pi\bigr).$$

В качестве следствия:

Следствие 3.1

$$\{\pi \in S_L \colon Z(\pi) \geqslant c + v\} \subseteq \Bigl\{\pi \in S_L \colon f_c\bigl(S(c), \pi\bigr) \geqslant \frac{v\ell k}{L\sqrt{k}}\Bigr\}.$$

Применение неравенства Талаграна для случайных перестановок (3.1). Теперь мы можем оценить вероятность отклонения случайной величины $Z(\pi)$ от её медианы M:

Теорема 3.2 $\forall t > 0$*:*

$$\mathsf{P}\bigl(Z(\pi) \geqslant M + t\bigr) \leqslant 2 \exp -\frac{t^2 \ell^2 k}{16 L^2},$$

где M — медиана случайной величины $Z(\pi)$.

Доказательство.

$$\mathsf{P}\bigl(Z(\pi) \geqslant M + t\bigr) \overset{\text{Следствие (3.1)}}{\leqslant} \mathsf{P}\Bigl(f_c\bigl(S(M), \pi\bigr) \geqslant \frac{t\ell k}{L\sqrt{k}}\Bigr) =$$

$$= 1 - \mathsf{P}\Bigl(f_c\bigl(S(M), \pi\bigr) \leqslant \frac{t\ell k}{L\sqrt{k}}\Bigr) \overset{\text{Следствие (1.1)}}{\leqslant} \frac{\exp -\frac{t^2\ell^2 k}{16L^2}}{\mathsf{P}\bigl(S(M)\bigr)} = 2 \exp -\frac{t^2 \ell^2 k}{16 L^2}.$$

■

Замечание. В доказательстве мы допустили весьма грубую оценку при переоде (10)–(11), оценив супремум разности мощностью первого множества. Более тонкий подход к этому переходу, возможно, даст существенно более точную оценку. Подсказка, возможно, лежит в [1], глава 6.

Список литературы

[1] *Talagrand M.* Concentration of measure and isoperimetric inequalities in product spaces // *Publications mathématiques de l'I.H.É.S.* — 1995. — Vol. 81. — Pp. 73–205. http://people.math.jussieu.fr/~talagran/preprints/ihes.dvi.