УДК 519.68

А.А. ВАГИС

ОБ ОДНОМ ИНФОРМАЦИОННОМ НЕРАВЕНСТВЕ В ТЕОРИИ СЛОЖНОСТИ ЗАДАЧ ОПТИМИЗАЦИИ И ПРОЦЕДУР ИНДУКТИВНОГО ВЫВОДА

Выведено важное неравенство относительно шенноновской взаимной информации относительно двух зависимых булевых случайных величин. Оно используется при доказательстве оценок снизу в теории сложности задач оптимизации и байесовской процедуры распознавания.

Введение. В теории сложности задач оптимизации важную роль играет одно неравенство относительно шенноновской информации между случайными величинами [1]. Аналогичное неравенство используется при оценке эффективности байесовской процедуры распознавания или обучения [2]. В работе проводится доказательство такого неравенства.

Количество информации относительно двух случайных величин z_0 и z_1 определяется соотношением

$$I(z_1, z_0) = H(z_0) + H(z_1) - H(z_1, z_0), \quad (1)$$

$H(z)$ – энтропия случайной величины z [3]. Величина $I(z_1, z_0)$ обозначает убыль энтропии случайной величины z_1 при наблюдении случайной величины z_0 и наоборот, поскольку $I(z_1, z_0) = I(z_0, z_1)$. Убыль понимают как количество информации о случайной величине z_1, полученной при наблюдении случайной величины z_0.

Теорема. Пусть z_0 и z_1 – булевы случайные величины, для которых выполняются неравенства

$$P(z_0 \neq z_1) \leq \frac{3}{8},$$

$$P(z_0 = 0) = P(z_1 = 1) = \frac{1}{2}. \quad (2)$$

Тогда выполняется соотношение

$$I(z_1, z_0) \geq C_a > 0, \tag{3}$$

где C_a – абсолютная константа.

Доказательство. Из соотношений (2) вытекают неравенства

$$P(z_1 = z_0) \geq \frac{5}{8},$$

$$P(z_1 = 0, z_0 = 0) + P(z_1 = 1, z_0 = 1) \geq \frac{5}{8},$$

$$P(z_1 = 0 | z_0 = 0) + P(z_1 = 1 | z_0 = 1) \geq \frac{5}{4},$$

$$P(z_1 = 0 | z_0 = 0) \geq \frac{1}{4}, \; P(z_1 = 1 | z_0 = 1) \geq \frac{1}{4}, \tag{4}$$

так как

$$P(z_1 = 0 | z_0 = 0) = \frac{P(z_1 = 0, z_0 = 0)}{P(z_0 = 0)},$$

$$P(z_1 = 1 | z_0 = 1) = \frac{P(z_1 = 1, z_0 = 1)}{P(z_0 = 1)},$$

$$P(z_1 = 0 | z_0 = 0) \leq 1, \quad P(z_1 = 1 | z_0 = 1) \leq 1.$$

Далее получаем цепочку неравенств

$$P(z_1 = 0, z_0 = 0) \geq \frac{1}{8},$$

$$P(z_1 = 1, z_0 = 1) \geq \frac{1}{8},$$

$$P(z_1 = 0, z_0 = 1) \leq \frac{3}{8},$$

$$P(z_1 = 1, z_0 = 0) \leq \frac{3}{8},$$

$$P(z_1 = 0) \geq \frac{1}{8},$$

$$P(z_1 = 1) \geq \frac{1}{8}. \tag{5}$$

Информация $I(z_1, z_0)$ выражается следующим образом:

$$I(z_1, z_0) = 1 - P(z_1 = 0)\log P(z_1 = 0) - P(z_1 = 1)\log P(z_1 = 1) +$$

$$+ P(z_1 = 0, z_0 = 0)\log P(z_1 = 0, z_0 = 0) + P(z_1 = 0, z_0 = 1)\log P(z_1 = 0, z_0 = 1) +$$

$$+ P(z_1 = 1, z_0 = 0)\log P(z_1 = 1, z_0 = 0) + P(z_1 = 1, z_0 = 1)\log P(z_1 = 1, z_0 = 1). \quad (6)$$

Здесь $\log$ обозначает двоичный логарифм

$$H(z_0) = -P(z_0)\log P(z_0 = 0) - P(z_0 = 1)\log P(z_0 = 1) = 1.$$

Обозначим

$$P(z_1 = 0) = a, \ P(z_1 = 0, z_0 = 0) = b.$$

Из формулы (6), учитывая соотношения

$$P(z_1 = 0, z_0 = 0) + P(z_1 = 1, z_0 = 0) = P(z_0 = 0) = \frac{1}{2},$$

$$P(z_1 = 0, z_0 = 1) + P(z_1 = 1, z_0 = 1) = P(z_0 = 1) = \frac{1}{2},$$

$$P(z_1 = 0, z_0 = 0) + P(z_1 = 0, z_0 = 1) = P(z_1 = 0) = a,$$

$$P(z_1 = 0) + P(z_1 = 1) = 1, \quad (7)$$

получаем

$$I(z_1, z_0) \equiv \varphi(a,b) = 1 - a\log a - (1-a)\log(1-a) +$$

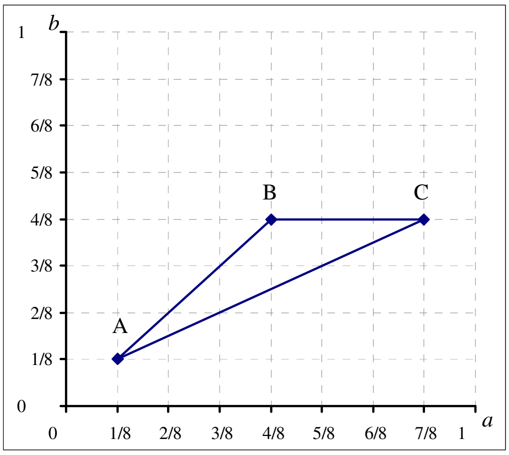
$$+ b\log b + (a-b)\log(a-b) + \left(\frac{1}{2} - b\right)\log\left(\frac{1}{2} - b\right) +$$

$$+ \left(\frac{1}{2} - a + b\right)\log\left(\frac{1}{2} - a + b\right).$$

Из формул (5), (7) заключаем, что величины a и b должны удовлетворять условиям

$$b \ge \frac{1}{8}, \ a \ge \frac{1}{8}, \ b \le \frac{1}{2}, \ b \le a, \ b \ge \frac{a}{2} + \frac{1}{16}.$$

Последнее неравенство следует из второго в (4).
Множество допустимых значений a и b лежит в треугольнике ABC (рисунок).



РИСУНОК

Имеем

$$\varphi_b'(a,b)=\log e\left[\ln b-\ln(a-b)-\ln\left(\frac{1}{2}-b\right)+\ln\left(\frac{1}{2}-a+b\right)\right]=\log\frac{b\left(\frac{1}{2}-a+b\right)}{(a-b)\left(\frac{1}{2}-b\right)}.$$

Так как $b\geq\frac{a}{2}$, то

$$\varphi_b'(a,b)\geq\log\frac{\frac{a}{2}\left(\frac{1}{2}-\frac{a}{2}\right)}{\frac{a}{2}\left(\frac{1}{2}-\frac{a}{2}\right)}=0$$

при каждом a из допустимой области. Поэтому минимальное значение функции $\varphi(a,b)$ лежит на отрезке AC, т.е. при $b=\frac{a}{2}+\frac{1}{16}$.

Таким образом,

$$I(z_1,z_0)\geq 1-a\log a-(1-a)\log(1-a)+\frac{8a+1}{16}\log\frac{8a+1}{16}+\frac{8a-1}{16}\log\frac{8a-1}{16}+$$

$$+\frac{7-8a}{16}\log\frac{7-8a}{16}+\frac{9-8a}{16}\log\frac{9-8a}{16}\equiv\varphi(a)\text{, где }\frac{1}{8}\leq a\leq\frac{7}{8}.$$

Отсюда следует, что

$$\varphi'(a) = \log e\left[-\ln a + \ln(1-a) + \frac{1}{2}\ln\frac{8a+1}{16} + \frac{1}{2}\ln\frac{8a-1}{16} - \frac{1}{2}\ln\frac{7-8a}{16} - \frac{1}{2}\ln\frac{9-8a}{16}\right] =$$

$$= \frac{1}{2}\log\frac{(1-a)^2(8a+1)(8a-1)}{a^2(7-8a)(9-8a)}.$$

Поэтому, если $a \to \frac{1}{8}a+$, то $\varphi'(a) \to -\infty$, и если $a \to \frac{7}{8}-$, то $\varphi'(a) \to \infty$. Таким образом, $\varphi'(a)$ принимает значение 0 только в одной точке $a = \frac{1}{2}$.

Действительно,

$$(1-a)^2(64a^2-1) = a^2(7-8a)(9-8a),$$

$$(1-2a+a^2)(64a^2-1) = a^2(63-128a+64a^2),$$

$$64a^2 - 1 - 128a^3 + 2a + 64a^4 - a^2 = 63a^2 - 128a^3 + 64a^4,$$

$$a = \frac{1}{2}.$$

Значит функция $\varphi(a)$ достигает минимума в точке $\frac{1}{2}$:

$$I(z_1, z_0) \geq 2\left(1 + \frac{5}{16}\log\frac{5}{16} + \frac{3}{16}\log\frac{3}{16}\right) =$$

$$= 2\left(\frac{5}{16}\log 5 + \frac{3}{16}\log 3 - 1\right) = \frac{1}{8}\left(5\log 5 + 3\log 3 - 16\right) =$$

$$= \frac{1}{8}\log\frac{5^5 \cdot 3^3}{2^{16}} > 0 \text{, так как } 5^5 \cdot 3^3 = 84375 > 65536 = 2^{16}.$$

Неравенство (3) доказано.

Результат теоремы показывает, что взаимная шенноновская информация двух случайных величин z_0 и z_1, связанных соотношениями (2), оказывается не слишком малой, т.е. при проведении опытов над этими случайными величинами приобретается конечная полезная информация.

Выведенное неравенство используется при доказательстве оценок снизу в теории сложности задач оптимизации и байесовской процедуры распознавания.

О.А. Вагіс

ПРО ОДНУ ІНФОРМАЦІЙНУ НЕРІВНІСТЬ В ТЕОРІЇ СКЛАДНОСТІ ЗАДАЧ ОПТИМІЗАЦІЇ ТА ПРОЦЕДУР ІНДУКТИВНОГО ВИВОДУ

Виведено важливу нерівність щодо шеноновської взаємної інформації стосовно двох залежних булевих випадкових величин. Воно використовується при доказі оцінок знизу в теорії складності задач оптимізації та байєсовської процедури розпізнавання.

A.A. Vagis

ABOUT ONE INFORMATIVE INEQUALITY IN THE THEORY OF COMPLICATION OF OPTIMIZATIONS TASKS AND INDUCTIVE CONCLUSION PROCEDURES

Important inequality is shown out in relation to Shennon mutual information on two dependent boole casual sizes. It is used for proof of estimations from below in the theory of complication of tasks of optimization and Bayesian pattern recognition procedures.

1. *Немировский А.С.,Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. – М.: Наука, 1979. – 383 с.
2. *Гупал А.М., Пашко С.В., Сергиенко И.В.* Эффективность байесовской процедуры классификации объектов // Кибернетика и системный анализ. – 1995. – №4. – С. 76–89.
3. *Гупал А.М., Сергиенко И.В.* Оптимальные процедуры распознавания. Обоснование процедур индуктивного вывода // Там же. – 2003. – №1. – С. 33–39.
4. *Яглом А.М.*, *Яглом И.М.* Вероятность и информация. – М.: Наука, 1973. –512 с.

Получено 14.02.2006