Доклады Академии наук СССР
1973. Том 210, № 2

УДК 519.25.8

КИБЕРНЕТИКА И ТЕОРИЯ РЕГУЛИРОВАНИЯ

Г. М. УЛАНОВ, С. В. УЛЬЯНОВ, Э. М. ХАЗЕН

ИНФОРМАЦИОННЫЕ ОЦЕНКИ ДЛЯ РИСКА В ЗАДАЧАХ ОБРАБОТКИ БОЛЬШИХ МАССИВОВ ИНФОРМАЦИИ

(Представлено академиком Б. Н. Петровым 17 VII 1972)

1. В задачах обработки больших массивов информации возникает проблема оценки потерь вследствие редукции данных, обусловленной ограничениями на «объем памяти» информационных систем (и.с.) и требованиями простоты реализации. В данной работе постановка и анализ задачи рассматривается в рамках теории оптимальных статистических решений и теории информации. При этом одновременно получаются оценки потенциальной точности принятого решения.

2. В байесовских задачах, если решение должно приниматься на основании фиксированного числа N наблюдений $(x_1, \ldots, x_N)$, при редукции к достаточной статистике $L(x_1, \ldots, x_N)$ риск не возрастает. При этом достаточность определяется по Фишеру. Для случая редукции, нарушающей достаточность, ряд информационных оценок для возрастания риска получен в работах А. Переза [1, 2].

В задачах последовательного анализа редукция наблюдаемых данных без возрастания риска имеет место в том случае, если статистика $L(x_1, \ldots, x_N)$, значениями которой определяется решение о прекращении наблюдений и выбор окончательного решения, обладает, кроме достаточности по Фишеру, свойством транзитивности [3]. В [4] получены некоторые информационные оценки для изменения риска в задачах последовательного анализа при изменении распределения вероятностей или редукции результатов наблюдений. При этом потери информации при переходе от достаточной и транзитивной многомерной* статистики $L(x_1, \ldots, x_m)$ к другой статистике $T(x_1, \ldots, x_m) = T_m$ меньшей размерности можно охарактеризовать величинами

$$J(L_m, \theta) - J(T_m, \theta) = \Delta J_m^{(1)}, \tag{1}$$

$$J(L_{m+1}, L_m) - J(L_{m+1}, T_m) = \Delta J_m^{(2)}, \tag{2}$$

где $J(x, y)$ обозначает шенноновское количество информации связи между случайными величинами x и y; θ — параметр, о котором должно приниматься решение.

Для возрастания риска вследствие нарушения свойства транзитивности статистики при редукции можно получить [4, 5] следующую оценку:

$$\Delta\rho \leqslant W_0[2\Delta J^{(2)} + 10\sqrt{\Delta J^{(2)}}]\bar{n}, \tag{3}$$

где $\bar{n}$ — среднее число наблюдений, требуемых для принятия решения последовательным решающим правилом; $\forall \Delta J^{(2)} \leqslant \Delta J_m^{2)}$; W_0 — константа, ограничивающая математическое ожидание штрафа за неверное окончательное решение.

При нарушении (вследствие редукции) свойства достаточности (по Фишеру) статистики, по которой определяется окончательное решение, но при сохранении правила остановки процесса наблюдения, возрастание

* Ее размерность может также возрастать с ростом числа наблюдений m.

байесовского риска ограничивается неравенствами

$$\rho' - \rho \leqslant \sqrt{2\rho' M_0 \Delta \bar{J}}; \tag{4}$$

$$\Delta \bar{J} \geqslant \frac{1}{W_0}\left[\rho \ln \frac{\rho}{\rho'} + (W_0 - \rho) \ln \frac{W_0 - \rho}{W_0 - \rho'}\right], \tag{5}$$

где ρ' обозначает байесовский риск в редуцированной задаче, а ρ — в исходной; $\Delta \bar{J}$ — разность шенноновского количества информации о θ, содержащейся в исходной и редуцированной последовательной выборке.

Как частный случай информационных неравенств для изменения риска при изменении законов распределения вероятностей получаются оценки для вероятностей ошибок при различении статистических гипотез. С использованием этих оценок основная теорема Шеннона о пропускной способности канала формулируется и доказывается как теорема о возможностях различения экспоненциально возрастающего (с ростом числа наблюдений) множества статистических гипотез ([6]).

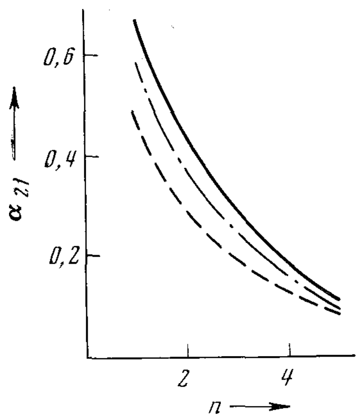

Рис. 1

3. При решении проблемы оптимизации процессов обработки и переработки больших массивов информации в и.с. на практике приходится иметь дело как с однородными, так и с неоднородными ее массивами, для которых существуют, строго говоря, различные меры информации. Следовательно, в информационных оценках для изменения риска и для вероятностей ошибок при различении многих статистических гипотез с учетом неоднородности массивов информации могут использоваться меры информации, отличные от шенноновской. Обобщенные f-энтропии (в которых вместо $(t \ln t)$ используются другие выпуклые функции $f(t)$) введены А. Реньи (см., например, ([7])), который раскрыл также их статистический смысл. При использовании такого более широкого класса информационных расхождений между вероятностными распределениями при фиксированном числе наблюдений информационные оценки для риска и вероятностей ошибок классификации испытуемых гипотез могут быть улучшены ([8]). Таким образом, учет «внутренней» структуры обрабатываемых массивов информации позволяет получить более точные оценки принятия решения, сократить необходимое число наблюдений при заданной точности вычислений, а полученные информационные оценки (как граничные условия потенциальной точности) позволяют перейти к осуществлению идеи сжатия процессов обработки информации *. В данном разделе приводятся результаты определения оценок нижних границ для вероятностей ошибок при различении многих гипотез для обобщенных мер информационных расхождений.

Обозначим α_{ij} вероятность принять гипотезу H_j при истинности H_i для данного решающего правила, которое может быть как правилом с фиксированным числом n наблюдений, так и последовательным. Пусть n обозначает момент остановки наблюдений.

Определение 1. *Обобщенным информационным расхождением* (*порядка* a *и типа* $\{\beta_i\}$) между вероятностными распределениями $p(x_1, \ldots, x_n | H_i) = p_i(x_1, \ldots, x_n)$ и $p(x_1, \ldots, x_n | H_k) =$

* Эта идея была сформулирована и поставлена перед авторами в виде специальной задачи информационной теории управления ([9]) акад. Б. Н. Петровым.

$= p_k(x_1, \ldots, x_n)$ назовем величину условного математического ожидания

$$H_a^{\{\beta_i\}}(i:k) = M\left\{\frac{(p_i(x_1,\ldots,x_n))^{a+\beta_i-1}}{(p_k(x_1,\ldots,x_n))^a}\,\middle|\, H_k\right\}. \tag{6}$$

С использованием расхождения (6) и неравенства Иенсена для функций нескольких переменных можно получить следующие оценки для вероятностей ошибок: если $\beta_i > 2$, $1-\beta_i < a < 2-\beta_i$, то

$$\begin{aligned}
&H_a^{\{\beta_i\}}(i:k) \leqslant [\alpha_{kk}^{1-a}\,\alpha_{ik}^{a+\beta_i-1} + (1-\alpha_{kk})^{1-a}(1-\alpha_{ik})^{a+\beta_i-1}];\\
&H_a^{\{\beta_i\}}(i:k) \leqslant \sum_{j=1}^{N} \alpha_{kj}^{1-a}\,\alpha_{ij}^{a+\beta_i-1};\\
&H_a^{\{\beta_i\}}(k:i) \leqslant [\alpha_{ik}^{1-a}\,\alpha_{kk}^{a+\beta_i-1} + (1-\alpha_{ik})^{1-a}(1-\alpha_{kk})^{a+\beta_i-1}];\\
&H_a^{\{\beta_i\}}(k:i) \leqslant \sum_{j=1}^{N} \alpha_{ij}^{1-a}\,\alpha_{kj}^{a+\beta_i-1}.
\end{aligned} \tag{7}$$

Неравенства (7) дополняют неравенства ([8]), полученные с использованием информационных расхождений Реньи и Кульбака. Неравенства (7) позволяют уточнить нижние границы для вероятностей ошибок при фиксированном числе наблюдений n (см. далее рис. 1).

Определение 2. Обобщенной аддитивной производящей функцией информационного расхождения порядка a типа $\{\beta_i\}$ называется функция вида

$$M_a^{\{\beta_i\}}(p_i \| p_k, m) = [H_a^{\{\beta_i\}}(i:k)]^{m/(1-a)},$$

так как

$$\frac{\partial}{\partial m} M_a^{\{\beta_i\}}(p_i \| p_k, m)\big|_{m=0} = I_a^{\{\beta_i\}}(p_i \| p_k) = \frac{1}{1-a}\ln [H_a^{\{\beta_i\}}(i:k)]. \tag{8}$$

Неаддитивной мерой информационного расхождения порядка a и типа $\{\beta_i\}$ называется мера ([10])

$$I(p_i \| p_k;\ a,\ \{\beta_i\}) = (1 - H_a^{\{\beta_i\}}(i:k))/(1-e^{1-a}). \tag{9}$$

На рис. 1 показаны результаты вычислений по формулам (6), (7). Пунктиром и штрих-пунктиром показаны результаты для мер Кульбака и Реньи соответственно. При вычислениях были приняты следующие параметры: $p_i(x) = k_i e^{-k_i x}$, $i = 1, 2$; $k_1 = 1$; $k_2 = 2$; $a = -0{,}5$; $\beta_1 = \beta_2 = 2{,}4$; $\alpha_{12} = 0{,}15$. Для меры Реньи $a = 0{,}5$.

Поступило
29 VI 1972

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

[1] A. Perez, Kybernetika, 1, № 4, 297 (1965); 3, № 1, 1 (1967); 6, № 2, 90 (1970). [2] A. Perez, In: Trans. of the IV Prague Conf., Prague, 1965, 1967, p. 55. [3] R. R. Bahadur, Ann. Math. Stat., 5, № 3, 423 (1954). [4] Э. М. Хазен, Изв. АН СССР, Техн. кибернетика, № 6, 127 (1969). [5] Э. М. Хазен, Тр. Семинара по управлению случайными процессами, июль, 1971, Киев, 1972. [6] Э. М. Хазен, Методы оптимальных статистических решений и задачи оптимального управления, 1968. [7] A. Rényi, Probability Theory, Budapest, 1970. [8] Э. М. Хазен, Изв. АН СССР, Технич. кибернетика, № 3, 184 (1971). [9] Б. Н. Петров, В. В. Петров и др., Итоги науки, Техническая кибернетика, 1968, М., 1970, стр. 221; Техническая кибернетика, 1969, М., 1971, стр. 5. [10] P. N. Rathie, Kybernetika, 7, № 2, 125 (1971); 7, № 5, 394 (1971).