

УДК 519.728

**ЭЛЕМЕНТЫ ТЕОРИИ НЕДООПРЕДЕЛЕННОЙ ИНФОРМАЦИИ<sup>1</sup>**

Л. А. Шоломов

*Институт системного анализа РАН, г. Москва, Россия***E-mail:** sholomov@isa.ru

Изложены результаты, относящиеся к информационным свойствам недоопределенных данных. Введены и изучены их информационные характеристики, для них получено обобщение ряда результатов классической теории информации, рассмотрены некоторые свойства, специфичные только для недоопределенных данных.

**Ключевые слова:** *недоопределенный источник, доопределение, энтропия, W-энтропия, принцип Шеннона, теорема кодирования, эффект Нечипорука, правило сложения энтропий.*

**Введение**

Методы эффективного использования недоопределенной информации важны для многих разделов информатики. С недоопределенными данными имеют дело в задачах распознавания, управления, принятия решений, логического синтеза схем, сжатия и передачи данных, криптографии, генетики. Разнообразные применения указывают на целесообразность изучения недоопределенных данных в качестве самостоятельного объекта подобно тому, как это делается в теории информации для полностью определенных данных. Это предполагает создание единой системы понятий, введение и исследование информационных характеристик недоопределенных данных, развитие и модификацию применительно к ним принципов, методов и результатов классической теории информации, выявление закономерностей, характерных только для недоопределенных данных, создание исчисления для формального преобразования выражений с участием разных информационных характеристик, изучение эквивалентных преобразований недоопределенных данных, разработку теоретически обоснованных алгоритмов обращения с ними.

Ниже приводятся результаты, полученные в этих направлениях. Как будет видно, некоторые результаты и методы теории информации переносятся на недоопределенные данные, некоторые модифицируются соответствующим образом, а в некоторых случаях возникают новые эффекты. Установленные факты оказались полезными при исследовании задач, связанных с синтезом схем [12], процедурами голосования [14], распознаванием текстов [15].

**1. Недоопределенные данные и их энтропия**

Под недоопределенными данными будем понимать последовательности недоопределенных символов. Каждому из них соответствует некоторое множество символов основного алфавита, одним из которых он может быть замещен (доопределен). Дадим формальные определения.

<sup>1</sup>Работа выполнена при поддержке Отделения нанотехнологий и информационных технологий РАН по программе фундаментальных исследований (проект «Теория и методы эффективного использования недоопределенных данных»).

Пусть  $M = \{0, 1, \dots, m-1\}$  — некоторое множество и каждому непустому подмножеству  $T \subseteq M$  сопоставлен символ  $a_T$ . Алфавит всех символов  $a_T$  обозначим через  $A$ , а его подалфавит  $\{a_0, a_1, \dots, a_{m-1}\}$ , символы которого соответствуют элементам множества  $M$ , — через  $A_0$ . Символы из  $A_0$  будем называть *основными*, из  $A$  — *недоопределенными*. Доопределением символа  $a_T \in A$  назовем всякий основной символ  $a_i$ ,  $i \in T$ , а доопределением последовательности в алфавите  $A$  — любую последовательность в алфавите  $A_0$ , полученную из исходной заменой всех ее символов некоторыми доопределениями. Символ  $a_M$ , доопределимый любым основным символом, играет особую роль. Его будем называть *неопределенным* и обозначать  $*$ .

Пусть имеется источник  $X$ , порождающий символы  $a_T \in A$  независимо с вероятностями  $p(a_T) \geq 0$ ,  $\sum_T p(a_T) = 1$ . Набор вероятностей  $(p(a_T), T \subseteq M)$  обозначим через  $P$  и для источника  $X$  будем использовать обозначение  $(A, P)$ . Такой источник будем называть *недоопределенным*, а при выполнении условия  $p(a_T) = 0$  для  $a_T \notin A_0$  — *полностью определенным*. В случае  $p(a_T) = 0$  для  $a_T \notin A_0 \cup \{*\}$  источник называется *частично определенным*. Подчеркнем, что мы различаем термины «недоопределенный» и «частично определенный». Иногда вместо обозначения  $(A, P)$  источника  $X$  будем использовать  $(A', P')$ , где алфавит  $A'$  получен из  $A$  удалением всех или некоторых символов  $a_T$  с  $p(a_T) = 0$ , а  $P'$  образован из  $P$  удалением соответствующих нулевых компонент. В этих обозначениях полностью определенный источник может быть записан как  $(A_0, (p_0, \dots, p_{m-1}))$ , а частично определенный — как  $(A_0 \cup \{*\}, (p_0, \dots, p_{m-1}, p_*))$ , где обозначения  $p_i$  и  $p_*$  использованы вместо  $p(a_i)$  и  $p(*) = p(a_M)$ .

Зададимся некоторым набором вероятностей  $Q = (q_i, i \in M)$  символов  $a_i \in A_0$  ( $q_i \geq 0$ ,  $q_0 + \dots + q_{m-1} = 1$ ) и введем функцию

$$\mathcal{H}(P, Q) = - \sum_{T \subseteq M} p(a_T) \log \sum_{i \in T} q_i. \quad (1)$$

Здесь и дальше логарифмы двоичные. *Энтропией источника  $X$*  назовем величину

$$\mathcal{H}(P) = \min_Q \mathcal{H}(P, Q). \quad (2)$$

Наряду с  $\mathcal{H}(P)$  будем использовать обозначение  $\mathcal{H}(X)$ . Указанная формула энтропии была предложена М. М. Бонгардом [1, с. 92] из некоторых эвристических соображений в качестве меры неопределенности задач с несколькими ответами.

Для полностью определенного источника  $(A_0, (p_0, \dots, p_{m-1}))$ , в силу известного соотношения

$$\min_{(q_0, \dots, q_{m-1})} \left( - \sum_i p_i \log q_i \right) = - \sum_i p_i \log p_i, \quad (3)$$

величина  $\mathcal{H}(P)$  совпадает с энтропией Шеннона  $H(P) = - \sum_i p_i \log p_i$ . Вместо источников можно говорить об энтропии случайных опытов с недоопределенными исходами.

## 2. Вычисление энтропии

Энтропия недоопределенного источника задана неявно, как минимум по  $Q$  функции (1). Для нахождения точек минимума полезен следующий критерий.

**Теорема 1.** Набор вероятностей  $Q$  минимизирует функцию  $\mathcal{H}(P, Q)$  тогда и только тогда, когда при каждом  $i$ ,  $i \in M$ , выполнено

$$\sum_{T: i \in T} \frac{p(a_T)}{\sum_{j \in T} q_j} \leq 1, \quad (4)$$

где строгое неравенство может иметь место лишь при тех  $i$ , для которых  $q_i = 0$ .

**Доказательство.** Вогнутая по  $Q$  функция  $-\mathcal{H}(P, Q)$  удовлетворяет условиям теоремы 4.4.1 из [3]. По этой теореме необходимым и достаточным условием ее максимума в точке  $Q$  является существование такого  $\lambda$ , что  $-\partial\mathcal{H}(P, Q)/\partial q_i \leq \lambda$ ,  $i \in M$ , где строгие неравенства могут соответствовать лишь нулевым значениям  $q_i$ . В нашем случае эти соотношения приобретают вид

$$\log e \sum_{T: i \in T} \frac{p(a_T)}{\sum_{j \in T} q_j} \leq \lambda, \quad i \in M. \quad (5)$$

Поскольку равенства могут нарушаться лишь при нулевых  $q_i$ , домножив на  $q_i$ , получаем равенства

$$\sum_{T: i \in T} \frac{p(a_T)q_i}{\sum_{j \in T} q_j} = \frac{\lambda}{\log e} q_i, \quad i \in M. \quad (6)$$

Просуммировав их по  $i \in M$  с учетом  $\sum_T p(a_T) = \sum_i q_i = 1$  и того, что

$$\sum_{i \in M} \sum_{T: i \in T} \frac{p(a_T)q_i}{\sum_{j \in T} q_j} = \sum_T p(a_T) \frac{\sum_{i \in T} q_i}{\sum_{j \in T} q_j} = \sum_T p(a_T),$$

находим, что  $\lambda = \log e$ . Подставив значение  $\lambda$  в (5), получаем требуемое утверждение. ■

На базе этой теоремы может быть указан численный метод нахождения  $\mathcal{H}(P)$  [9]. Введем оператор  $Q' = \mathcal{U}(Q)$ , сопоставляющий набору  $Q = (q_0, q_1, \dots, q_{m-1})$  набор  $Q' = (q'_0, q'_1, \dots, q'_{m-1})$ , где

$$q'_i = \sum_{T: i \in T} \frac{p(a_T)q_i}{\sum_{j \in T} q_j}, \quad i = 0, 1, \dots, m-1.$$

Нетрудно проверить непосредственно, что оператор  $\mathcal{U}$  переводит наборы вероятностей в наборы вероятностей. Домножив обе части (4) на  $q_i$  и учитывая, что строгое неравенство в (4) может иметь место лишь при  $q_i = 0$ , получаем равенства

$$\sum_{T: i \in T} \frac{p(a_T)q_i}{\sum_{j \in T} q_j} = q_i, \quad i \in M, \quad (7)$$

означающие, что минимизирующий набор  $Q$  является неподвижной точкой оператора  $\mathcal{U}$ . Приведем без доказательства утверждение из [9], дающее алгоритм вычисления  $\mathcal{H}(P)$ .

**Теорема 2.** Если  $Q^{(0)} = (q_0^{(0)}, \dots, q_{m-1}^{(0)})$  — произвольный набор вероятностей с положительными компонентами и  $Q^{(\nu)} = \mathcal{U}(Q^{(\nu-1)})$ ,  $\nu = 1, 2, \dots$ , то при  $\nu \rightarrow \infty$  последовательность  $\mathcal{H}(P, Q^{(\nu)})$  сходится к  $\mathcal{H}(P)$ .

Таким образом, нахождение численного значения энтропии трудностей не вызывает. В некоторых содержательно важных случаях может быть получено явное выражение энтропии. Это относится, например, к частично определенным источникам, энтропия которых может быть найдена на основе следующего утверждения.

**Теорема 3.** Имеет место равенство

$$\mathcal{H}(P) = (1 - p(*))\mathcal{H}(P'),$$

где  $P' = (p'(a_T), T \subseteq M)$  (включение строгое) — набор, полученный из  $P = (p(a_T), T \subseteq M)$  отбрасыванием компоненты  $p(a_M) = p(*)$  и пересчетом вероятностей  $p'(a_T) = p(a_T)/(1 - p(*))$ .

*Доказательство.* Для любого набора вероятностей  $Q = (q_i, i \in M)$  выполнено  $\log \sum_{i \in M} q_i = 0$ , поэтому

$$- \sum_{T \subseteq M} p(a_T) \log \sum_{i \in T} q_i = -(1 - p(*)) \sum_{T \subseteq M} \frac{p(a_T)}{1 - p(*)} \log \sum_{i \in T} q_i.$$

Взяв минимум по  $Q$ , получаем нужное утверждение. ■

**Следствие 1.** Энтропия частично определенного источника  $X = (A_0 \cup \{*\}, (p_0, \dots, p_{m-1}, p_*))$  задается выражением

$$\mathcal{H}(X) = (1 - p_*)H\left(\frac{p_0}{1 - p_*}, \dots, \frac{p_{m-1}}{1 - p_*}\right) = (1 - p_*) \log(1 - p_*) - \sum_{0 \leq i \leq m-1} p_i \log p_i.$$

Это вытекает из теоремы, поскольку в рассматриваемом случае набор вероятностей  $P'$  соответствует всюду определенному источнику с вероятностями символов  $\frac{p(a_i)}{1 - p(*)}$ .

Теорема 3 обобщается на случай, когда неопределенные символы не устраняются полностью, а понижается их вероятность.

**Утверждение 1.** Для любого  $r, 0 \leq r \leq p_*$ , справедливо равенство

$$\mathcal{H}(P) = (1 - r)\mathcal{H}(P^{(r)}),$$

где  $P^{(r)} = (p^{(r)}(a_T), T \subseteq M)$ ,  $p^{(r)}(a_T) = \frac{p(a_T)}{1 - r}$  для  $T \neq M$ ,  $p_*^{(r)} = \frac{p(*) - r}{1 - r}$ .

*Доказательство.* С учетом того, что для набора вероятностей  $Q = (q_i, i \in M)$  выполнено  $r \log \sum_{i \in M} q_i = 0$ , имеем

$$- \sum_{T \subseteq M} p(a_T) \log \sum_{i \in T} q_i = -(1 - r) \left\{ \sum_{T \subseteq M} \frac{p(a_T)}{1 - r} \log \sum_{i \in T} q_i + \frac{p(*) - r}{1 - r} \log \sum_{i \in M} q_i \right\}.$$

Остается взять минимум по  $Q$ . ■

### 3. Связь с принципом Шеннона

Задача сжатия недоопределенных данных состоит в том, чтобы каждой недоопределенной последовательности сопоставить двоичный код по возможности малой длины, позволяющий восстановить какое-либо ее доопределение (но не саму последовательность). Точная постановка задачи и результаты будут приведены в последующих разделах.

Эта задача может быть описана в терминах более общей задачи кодирования источников при заданной точности воспроизведения. Приведем соответствующие понятия [3, 18] применительно к дискретным (конечным) источникам. Пусть некоторый

источник  $X$  порождает символы  $b$  конечного алфавита  $B$  независимо с вероятностями  $p(b)$ , и они должны быть представлены у адресата символами  $c$  конечного алфавита  $C$ . Условия на точность воспроизведения задаются указанием множества  $W$  допустимых совместных распределений  $(p(b, c), b \in B, c \in C)$ . Теоретико-информационной мерой неопределенности источника  $X$  при точности воспроизведения  $W$  считают  $W$ -энтропию [2]

$$H_W(X) = \inf_{(p(b,c)) \in W} \sum_{b,c} p(b,c) \log \frac{p(b,c)}{p(b) \sum_{b'} p(b',c)}.$$

Она обобщает понятие  $\varepsilon$ -энтропии (скорости создания сообщений в терминологии К. Шеннона [3, 18]). Согласно принципу Шеннона,  $W$ -энтропия характеризует степень сжимаемости сообщений источника  $X$  с точностью  $W$ . Это содержательный принцип, который для многих типов источников и мер точности доказан. Формулировку, обоснование и обсуждение этого принципа (применительно к задачам квантования сообщений) можно найти в [4].

Для недоопределенных данных алфавитам  $B$  и  $C$  соответствуют  $A$  и  $A_0$ , а допустимое множество, которое в этом случае будем обозначать через  $D$ , состоит из совместных распределений  $(p(a_T, a_i), T \subseteq M, i \in M)$ ,  $\sum_i p(a_T, a_i) = p(a_T)$ , удовлетворяющих условию  $p(a_T, a_i) = 0$  для  $i \notin T$ . Соответствующую  $D$ -энтропию для недоопределенных данных будем обозначать через  $\check{\mathcal{H}}(P)$ . Поскольку  $D$  — компакт, вместо инфимума можно использовать минимум:

$$\check{\mathcal{H}}(P) = \min_D \sum_{T,i} p(a_T, a_i) \log \frac{p(a_T, a_i)}{p(a_T) \sum_U p(a_U, a_i)}. \quad (8)$$

Следующая теорема [9] показывает, что  $\mathcal{H}(P)$  и  $\check{\mathcal{H}}(P)$  являются различными представлениями одной и той же функции.

**Теорема 4.** Имеет место равенство

$$\check{\mathcal{H}}(P) = \mathcal{H}(P).$$

**Доказательство.** Пусть значение  $\check{\mathcal{H}}(P)$  достигается на совместном распределении  $(p^0(a_T, a_i)) \in D$ . Положим  $q_i = \sum_T p^0(a_T, a_i)$ .

При заданном  $T$ , воспользовавшись для выпуклой функции  $f(x) = x \log x$  неравенством Иенсена  $\sum_i \alpha_i f(x_i) \geq f(\sum_i \alpha_i x_i)$  при  $\alpha_i = q_i / \sum_{j \in T} q_j$ ,  $x_i = p^0(a_T, a_i) / (p(a_T) q_i)$ , получаем с учетом  $p^0(a_T, a_i) = 0$ ,  $i \notin T$ ,

$$\begin{aligned} & \sum_i p^0(a_T, a_i) \log \frac{p^0(a_T, a_i)}{p(a_T) q_i} = \\ & = p(a_T) \left( \sum_{j \in T} q_j \right) \sum_i \frac{q_i}{\sum_{j \in T} q_j} \frac{p^0(a_T, a_i)}{p(a_T) q_i} \log \frac{p^0(a_T, a_i)}{p(a_T) q_i} \geq p(a_T) \log \frac{1}{\sum_{j \in T} q_j}. \end{aligned}$$

Просуммировав эти неравенства по  $T$ , заключаем, что  $\check{\mathcal{H}}(P) \geq \mathcal{H}(P)$ .

Обратно, пусть величина  $\mathcal{H}(P)$  в (2) достигается на наборе  $Q^0 = (q_0^0, q_1^0, \dots, q_{m-1}^0)$ . Положим  $p(a_T, a_i) = p(a_T)q_i^0 / \sum_{j \in T} q_j^0$  при  $i \in T$  и  $p(a_T, a_i) = 0$  при  $i \notin T$ . Учитывая равенства  $\sum_T p(a_T, a_i) = q_i^0$ , вытекающие из (6), получаем

$$\begin{aligned} \check{\mathcal{H}}(P) &\leq \sum_{T,i} p(a_T, a_i) \log \frac{p(a_T, a_i)}{\sum_j p(a_T, a_j) \sum_U p(a_U, a_i)} = \\ &= \sum_{T,i} p(a_T, a_i) \log \frac{p(a_T, a_i)}{p(a_T)q_i^0} = \sum_{T,i} p(a_T, a_i) \log \frac{1}{\sum_{j \in T} q_j^0} = \mathcal{H}(P). \end{aligned}$$

Теорема доказана. ■

Из нее и теоремы 5 (см. дальше) будет следовать справедливость принципа Шеннона для недоопределенных данных.

#### 4. Комбинаторная энтропия

Будем рассматривать последовательности длины  $n$  в алфавите  $A$ . Для набора натуральных чисел  $\mathbf{l} = (l_T, T \subseteq M)$ , такого, что  $\sum_T l_T = n$ , обозначим через  $\mathcal{K}_n(\mathbf{l})$  множество всех последовательностей, в которых символ  $a_T$ ,  $T \subseteq M$ , встречается  $l_T$  раз. Скажем, что некоторое множество последовательностей в алфавите  $A_0$  *доопределяет* класс  $\mathcal{K}_n(\mathbf{l})$ , если в нем найдется доопределение для каждой последовательности из  $\mathcal{K}_n(\mathbf{l})$ . Обозначим через  $N_n(\mathbf{l})$  минимальную мощность множества, доопределяющего  $\mathcal{K}_n(\mathbf{l})$ . Величину  $\log N_n(\mathbf{l})$  назовем *комбинаторной энтропией* класса  $\mathcal{K}_n(\mathbf{l})$ . Она указывает наименьшее число двоичных символов, достаточное для кодирования последовательностей класса  $\mathcal{K}_n(\mathbf{l})$ , позволяющего восстанавливать некоторое их доопределение.

**Теорема 5.** Существует константа  $c = c(m)$ , такая, что комбинаторная энтропия класса  $\mathcal{K}_n(\mathbf{l})$  заключена в пределах

$$n\mathcal{H}(\mathbf{l}/n) - c \log n \leq \log N_n(\mathbf{l}) \leq n\mathcal{H}(\mathbf{l}/n) + c \log n.$$

*Доказательство.* Верхняя оценка. Воспользуемся методом случайного кодирования. Для этого зададимся некоторым набором вероятностей  $Q = (q_0, \dots, q_{m-1})$  и возьмем  $N \geq 1$  случайных последовательностей длины  $n$  в алфавите  $A_0$ , компоненты которых независимо с вероятностями  $q_i$  принимают значения  $a_i$  ( $i = 0, \dots, m-1$ ). Вероятность того, что случайная последовательность доопределяет фиксированную последовательность из  $\mathcal{K}_n(\mathbf{l})$ , составляет  $\prod_T (\sum_{i \in T} q_i)^{l_T}$ , а вероятность того, что ни одна из  $N$  случайных последовательностей не является ее доопределением, равна  $(1 - \prod_T (\sum_{i \in T} q_i)^{l_T})^N$ . Вероятность  $p(N)$  отсутствия доопределения хотя бы у одной последовательности из  $\mathcal{K}_n(\mathbf{l})$  не превосходит

$$|\mathcal{K}_n(\mathbf{l})| \left( 1 - \prod_T \left( \sum_{i \in T} q_i \right)^{l_T} \right)^N < 2^{mn} \left( 1 - \prod_T \left( \sum_{i \in T} q_i \right)^{l_T} \right)^N$$

( $|\cdot|$  означает мощность множества). Непосредственный подсчет с учетом соотношения  $\ln(1-x) \leq -x$  показывает, что при всяком  $N$ , удовлетворяющем условию

$$\log N \geq \log n + \log \ln m - \sum_T l_T \log \sum_{i \in T} q_i = \log n + \log \ln m + n\mathcal{H}(\mathbf{l}/n, Q),$$

выполнено  $p(N) < 1$  и потому существует доопределяющее множество мощности  $N$ . Это справедливо для любого  $Q$ , и в частности, для которого  $\mathcal{H}(\mathbf{1}/n, Q) = \mathcal{H}(\mathbf{1}/n)$ .

Нижняя оценка. Обозначим через  $t(\mathbf{1})$  максимальное число последовательностей из  $\mathcal{K}_n(\mathbf{1})$ , которое может быть доопределено одной последовательностью. Рассмотрим последовательность  $\mathbf{y}$ , являющуюся доопределением некоторой последовательности  $\mathbf{x} \in \mathcal{K}_n(\mathbf{1})$ . Пусть  $\mathbf{y}$  имеет параметры  $s_0, s_1, \dots, s_{m-1}$  ( $s_0 + \dots + s_{m-1} = n$ ). Обозначим через  $w_{Ti}$  число символов  $a_T$  последовательности  $\mathbf{x}$ , доопределенных в  $\mathbf{y}$  символом  $a_i$ . Числа  $w_{Ti}$  удовлетворяют условиям

$$\sum_T w_{Ti} = s_i \quad (i \in M), \quad \sum_i w_{Ti} = l_T \quad (T \subseteq M), \quad w_{Ti} = 0 \quad \text{для } i \notin T. \quad (9)$$

При фиксированных  $w_{Ti}$  последовательность  $\mathbf{y}$  доопределяет

$$\frac{s_0!}{\prod_T w_{T0}!} \cdots \frac{s_{m-1}!}{\prod_T w_{T,m-1}!} = \frac{\prod_i s_i!}{\prod_{T,i} w_{Ti}!}$$

последовательностей из  $\mathcal{K}_n(\mathbf{1})$ , а всего она доопределяет

$$t_{s_0 \dots s_{m-1}}(\mathbf{1}) = \sum_{w_{Ti}, (9)} \frac{\prod_i s_i!}{\prod_{T,i} w_{Ti}!}$$

последовательностей из этого класса, где сумма берется по всем наборам неотрицательных чисел  $w_{Ti}$ , удовлетворяющих условиям (9). В силу  $0 \leq w_{Ti} \leq n$  и того, что количества индексов  $i$  и множеств  $T$  ограничены константами (зависящими от  $m$ ), имеем

$$t_{s_0 \dots s_{m-1}}(\mathbf{1}) \leq n^{c_1} \max_{w_{Ti}, (9)} \frac{\prod_i s_i!}{\prod_{T,i} w_{Ti}!},$$

где  $c_1 = c_1(m)$  — константа. Отсюда

$$\begin{aligned} t(\mathbf{1}) &= \max_{s_0, \dots, s_{m-1}} \max_{w_{Ti}, (9)} t_{s_0 \dots s_{m-1}}(\mathbf{w}) \leq \\ &\leq n^{c_1} \max_{s_0, \dots, s_{m-1}} \max_{w_{Ti}, (9)} \frac{\prod_i s_i!}{\prod_{T,i} w_{Ti}!} \leq n^{c_1} \max_{w_{Ti}} \frac{\prod \left( \sum_T w_{Ti} \right)!}{\prod_{T,i} w_{Ti}!}, \end{aligned} \quad (10)$$

где для целых неотрицательных  $w_{Ti}$  выполнены условия

$$\sum_i w_{Ti} = l_T \quad (T \subseteq M), \quad w_{Ti} = 0 \quad \text{для } i \notin T. \quad (11)$$

Класс  $\mathcal{K}_n(\mathbf{1})$  содержит  $n! / \prod_T l_T!$  последовательностей. Отсюда и из (10) заключаем, что минимальная мощность  $N_n(\mathbf{1})$  доопределяющего множества для класса  $\mathcal{K}_n(\mathbf{1})$  удовлетворяет оценке

$$N_n(\mathbf{1}) \geq \frac{|\mathcal{K}_n(\mathbf{1})|}{t(\mathbf{1})} \geq n^{-c_1} \min_{(w_{Ti}), (11)} \frac{n! \prod_{T,i} w_{Ti}!}{\prod_T \left( \sum_i w_{Ti} \right)! \prod_i \left( \sum_T w_{Ti} \right)!}.$$

Из формулы Стирлинга следует, что для любых целых  $z, z_1, \dots, z_k$  ( $z \geq 2, z_1 + \dots + z_k = z$ ) выполнено

$$\log \frac{z!}{\prod_j z_j!} = z \log z - \sum_j z_j \log z_j + \theta \log z,$$

где  $-c_2 \leq \theta \leq c_2$ ,  $c_2 = c_2(k)$  — константа. С учетом этого получаем

$$\begin{aligned} \log N_n(\mathbf{1}) \geq \min_{(w_{Ti}), (11)} & \left( n \log n - \sum_T \left( \sum_i w_{Ti} \right) \log \left( \sum_i w_{Ti} \right) - \right. \\ & \left. - \sum_i \left( \sum_T w_{Ti} \right) \log \left( \sum_T w_{Ti} \right) + \sum_{T,i} w_{Ti} \log w_{Ti} \right) - c \log n. \end{aligned}$$

Минимизируемое выражение может быть преобразовано к виду

$$\begin{aligned} n \left( - \sum_T \left( \sum_i \frac{w_{Ti}}{n} \right) \log \left( \sum_i \frac{w_{Ti}}{n} \right) - \sum_i \left( \sum_T \frac{w_{Ti}}{n} \right) \log \left( \sum_T \frac{w_{Ti}}{n} \right) + \right. \\ \left. + \sum_{T,i} \frac{w_{Ti}}{n} \log \sum_{T,i} \frac{w_{Ti}}{n} \right) = n \sum_{T,i} \frac{w_{Ti}}{n} \log \frac{\frac{w_{Ti}}{n}}{\sum_U \frac{w_{Ui}}{n} \sum_j \frac{w_{Tj}}{n}}. \end{aligned}$$

В силу (11) выполнено  $w_{Ti}/n = 0, i \notin T$ , и  $\sum_i (w_{Ti}/n) = l_T/n$ , поэтому

$$\log N_n(\mathbf{1}) \geq n\check{\mathcal{H}}(\mathbf{1}/n) - c \log n.$$

Остается воспользоваться теоремой 4, по которой  $\check{\mathcal{H}}(\mathbf{1}/n) = \mathcal{H}(\mathbf{1}/n)$ . ■

Для указания параметров класса  $\mathcal{K}_n(\mathbf{1})$  (и даже просто для задания  $n$ ) требуется порядка  $\log n$  единиц информации. Будем рассматривать случай, когда энтропия  $\log N_n(\mathbf{1})$  класса  $\mathcal{K}_n(\mathbf{1})$  существенно больше  $\log n$ ; тогда  $\log N_n(\mathbf{1}) \sim n\mathcal{H}(\mathbf{1}/n)$ . Этот результат, записанный с учетом теоремы 4 в виде  $\log N_n(\mathbf{1}) \sim n\check{\mathcal{H}}(\mathbf{1}/n)$ , означает, что для частично определенных данных справедлив принцип Шеннона.

С классом  $\mathcal{K}_n(\mathbf{1})$  свяжем функционал  $h_n(\mathbf{1}) = n\mathcal{H}(\mathbf{1}/n)$ . Оценки теоремы 5 могут быть переписаны в виде

$$h_n(\mathbf{1}) - c \log n \leq \log N_n(\mathbf{1}) \leq h_n(\mathbf{1}) + c \log n. \quad (12)$$

### 5. Эффект Нечипорука

Пусть  $\mathcal{K}_n(l)$ ,  $l \leq n$ , — класс всех последовательностей длины  $n$  с  $l$  булевыми символами (0 и 1) и  $n - l$  неопределенными символами  $*$ , а  $N_n(l)$  — минимальная мощность множества, содержащего для каждой последовательности из  $\mathcal{K}_n(l)$  некоторое доопределение. Э. И. Нечипорук [7] установил, что

$$l \leq \log N_n(l) \leq l + c \log n. \quad (13)$$

Этот факт допускает следующую интерпретацию, которую будем использовать и в более общих ситуациях и называть эффектом Нечипорука. Если рассматривать двоичное кодирование недоопределенных последовательностей, при котором по коду восстанавливается какое-либо доопределение, то недоопределенные последовательности и последовательности меньшей длины, полученные из них удалением неопределенных символов, могут быть представлены кодами одинаковой с точностью до  $O(\log n)$  длины.

Теорема 5 позволяет распространить этот эффект на произвольные недоопределенные последовательности. Рассмотрим класс  $\mathcal{K}_n(\mathbf{1})$ . Класс, полученный из  $\mathcal{K}_n(\mathbf{1})$  путем удаления из его последовательностей всех символов  $*$ , обозначим  $\mathcal{K}_{n'}(\mathbf{1}')$ , где  $n' = n - l_*$ ,  $\mathbf{1}'$  — результат отбрасывания в  $\mathbf{1}$  компоненты  $l_*$ . Следующий факт обобщает (13) на случай недоопределенных данных общего вида и означает справедливость для них эффекта Нечипорука.

**Теорема 6.** Комбинаторная энтропия классов  $\mathcal{K}_n(\mathbf{1})$  и  $\mathcal{K}_{n'}(\mathbf{1}')$  удовлетворяет соотношениям

$$\log N_{n'}(\mathbf{1}') \leq \log N_n(\mathbf{1}) \leq \log N_{n'}(\mathbf{1}') + c \log n, \quad (14)$$

где  $c = c(m)$  — константа.

**Доказательство.** Применив теорему 3 при  $p(a_T) = l_T/n$ ,  $p(*) = l_*/n$ ,  $p'_T = l_T/n'$  и домножив обе части на  $n$ , приходим к равенству  $n\mathcal{H}(\mathbf{1}/n) = n'\mathcal{H}(\mathbf{1}'/n')$ , означающему  $h_n(\mathbf{1}) = h_{n'}(\mathbf{1}')$ . Отсюда в силу (12)

$$\log N_n(\mathbf{1}) \leq h_n(\mathbf{1}) + c' \log n = h_{n'}(\mathbf{1}') + c' \log n \leq \log N_{n'}(\mathbf{1}') + c \log n.$$

Чтобы убедиться в справедливости нижней оценки, достаточно заметить, что если взять доопределяющее множество для последовательностей класса  $\mathcal{K}_n(\mathbf{1})$ , содержащих символы  $*$  в  $l_*$  последних позициях, и удалить в доопределениях  $l_*$  последних символов, получится доопределяющее множество для  $\mathcal{K}_{n'}(\mathbf{1}')$ . Это дает  $N_{n'}(\mathbf{1}') \leq N_n(\mathbf{1})$ . ■

Эффект Нечипорука распространяется также на кодирование с заданным критерием верности. Пусть последовательности в конечном алфавите  $B = \{b_i, i \in I\}$  должны быть представлены последовательностями (той же длины  $n$ ) в конечном алфавите  $C = \{c_j, j \in J\}$  при выполнении некоторых условий верности воспроизведения. В общем случае можно считать, что на  $B^n \times C^n$  задано отношение допустимости  $\omega$  и  $\mathbf{b}\omega\mathbf{c}$  означает допустимость воспроизведения последовательности  $\mathbf{c} \in C^n$  вместо  $\mathbf{b} \in B^n$ . Будем полагать, что отношение  $\omega$  не зависит от нумерации разрядов последовательностей: если  $\sigma$  — некоторая перестановка разрядов  $1, \dots, n$ , а  $\sigma\mathbf{b}$  и  $\sigma\mathbf{c}$  образованы из  $\mathbf{b}$  и  $\mathbf{c}$  соответствующей перестановкой символов, то  $\mathbf{b}\omega\mathbf{c} \Rightarrow (\sigma\mathbf{b})\omega(\sigma\mathbf{c})$ . Для заданных  $\mathbf{b}$  и  $\mathbf{c}$  обозначим через  $w_{ij}(\mathbf{b}, \mathbf{c})$  число позиций последовательностей, в которых вместо символа  $b_i$  воспроизводится  $c_j$ , и положим  $\mathbf{w}(\mathbf{b}, \mathbf{c}) = (w_{ij}(\mathbf{b}, \mathbf{c}), i \in I, j \in J)$ . Отношение  $\omega$  может быть задано множеством  $W$  допустимых наборов  $\mathbf{w} = (w_{ij}, i \in I, j \in J)$

соотношением  $\mathbf{b}\omega\mathbf{c} \Leftrightarrow \mathbf{w}(\mathbf{b}, \mathbf{c}) \in W$ . При выполнении условия  $\mathbf{w}(\mathbf{b}, \mathbf{c}) \in W$  последовательность  $\mathbf{c}$  будем называть  $W$ -допустимой для  $\mathbf{b}$ , а если необходимо указать значение  $\mathbf{w} = \mathbf{w}(\mathbf{b}, \mathbf{c})$ , то  $\mathbf{w}$ -допустимой.

Пусть задан набор натуральных чисел  $\mathbf{l} = (l_i, i \in I)$ ,  $\sum_i l_i = n$ , и ему соответствует класс  $\mathcal{K}_n(\mathbf{l})$  всех последовательностей длины  $n$  в алфавите  $B$ , в которых символ  $b_i, i \in I$ , встречается  $l_i$  раз. Пусть, кроме того, задано некоторое множество  $W$  допустимых наборов  $\mathbf{w} = (w_{ij}, i \in I, j \in J)$ , согласованных с классом  $\mathcal{K}_n(\mathbf{l})$ , т.е. удовлетворяющих условиям  $\sum_j w_{ij} = l_i$ . Обозначим через  $N_{n,W}(\mathbf{l})$  минимальную мощность множества последовательностей длины  $n$  в алфавите  $C$ , в котором для каждой последовательности из  $\mathcal{K}_n(\mathbf{l})$  имеется  $W$ -допустимая последовательность. Величину  $\log N_{n,W}(\mathbf{l})$  назовем комбинаторной  $W$ -энтропией класса  $\mathcal{K}_n(\mathbf{l})$ .

Для заданных  $\mathbf{l} = (l_i, i \in I)$  и  $\mathbf{w} = (w_{ij}, i \in I, j \in J)$ ,  $\sum_j w_{ij} = l_i$ , положим

$$h_n(\mathbf{l}, \mathbf{w}) = \sum_{i,j} w_{ij} \log \frac{nw_{ij}}{l_i \sum_{i'} w_{i'j}}.$$

С классом  $\mathcal{K}_n(\mathbf{l})$  и заданным множеством  $W$  допустимых наборов для этого класса свяжем функционал

$$h_{n,W}(\mathbf{l}) = \min_{\mathbf{w} \in W} h_n(\mathbf{l}, \mathbf{w}). \quad (15)$$

Нетрудно видеть, что он может быть записан в терминах  $W$ -энтропии в виде

$$h_{n,W}(\mathbf{l}) = nH_{W/n}(\mathbf{l}/n), \quad (16)$$

где множество допустимых распределений  $W/n$  образовано всеми  $\mathbf{w}/n, \mathbf{w} \in W$ .

**Теорема 7.** Комбинаторная  $W$ -энтропия класса  $\mathcal{K}_n(\mathbf{l})$  удовлетворяет оценкам

$$h_{n,W}(\mathbf{l}) - c \log n \leq \log N_{n,W}(\mathbf{l}) \leq h_{n,W}(\mathbf{l}) + c \log n,$$

где  $c = c(|I|, |J|)$  — константа.

**Доказательство.** Верхняя оценка. Пусть минимум в (15) достигается на наборе  $\mathbf{w} = (w_{ij}, i \in I, j \in J)$ . Положим  $s_j = \sum_i w_{ij}$ ,  $\mathbf{s} = (s_j, j \in J)$  и занумеруем последовательности  $\mathbf{b} \in \mathcal{K}_n(\mathbf{l})$  и  $\mathbf{c} \in \mathcal{K}_n(\mathbf{s})$  индексами  $r = 1, 2, \dots, \frac{n!}{\prod_i l_i!}$  и  $q = 1, 2, \dots, \frac{n!}{\prod_j s_j!}$

соответственно. Образует таблицу  $\|a_{qr}\|$ , строки которой соответствуют последовательностям  $\mathbf{c}_q$ , столбцы — последовательностям  $\mathbf{b}_r$ , положив

$$a_{qr} = \begin{cases} 1, & \text{если } \mathbf{c}_q \text{ } \mathbf{w}\text{-допустима для } \mathbf{b}_r, \\ 0, & \text{в противном случае.} \end{cases}$$

В матрице из 0 и 1 с  $u$  строками и  $v$  столбцами, имеющей  $s$  единиц в каждом столбце, можно выделить множество из не более  $\frac{u}{s} \left( \ln \frac{vs}{u} + 1 \right) + 1$  строк, содержащих 1 в каждом столбце [8]. Применив этот результат при

$$u = \frac{n!}{\prod_j s_j!}, \quad v = \frac{n!}{\prod_i l_i!}, \quad s = \frac{\prod_i l_i!}{\prod_{i,j} w_{ij!}},$$

получаем оценку

$$N_{n,W}(\mathbf{1}) \leq c'n \frac{n! \prod_{i,j} w_{ij}!}{\prod_i l_i! \prod_j s_j!}.$$

Из нее преобразованиями, подобными использованным при доказательстве нижней оценки теоремы 5, приходим к неравенству

$$\log N_{n,W}(\mathbf{1}) \leq h_n(\mathbf{1}, \mathbf{w}) + c \log n,$$

из которого с учетом того, что на  $\mathbf{w}$  достигается (15), получаем требуемую оценку.

Нижняя оценка доказывается аналогично нижней оценке теоремы 5 с тем отличием, что вместо доопределяющих последовательностей следует рассматривать допустимые. ■

Распространим понятие «неопределенного символа» на кодирование с заданным критерием верности, связывая с этим понятием несущественность (для отношения допустимости) того, какими символами алфавита  $C$  воспроизводятся неопределенные символы. Символ  $b_s \in B$  назовем *неопределенным* (при отношении допустимости  $\omega$ ), если для любых  $\mathbf{b}$ ,  $\mathbf{c}$  и  $\hat{\mathbf{c}}$ , таких, что  $\hat{\mathbf{c}}$  и  $\mathbf{c}$  отличаются лишь в разрядах, где  $\mathbf{b}$  содержит символы  $b_s$ , выполнено  $\mathbf{b}\omega\mathbf{c} \Leftrightarrow \mathbf{b}\omega\hat{\mathbf{c}}$ . Поскольку разные неопределенные символы можно не различать, считаем, что такой символ в  $B$  единствен, и для его обозначения будем использовать  $*$ . Укороченным для набора  $\mathbf{w} = (w_{ij}, i \in I, j \in J)$  назовем набор  $\mathbf{w}'$ , образованный из  $\mathbf{w}$  удалением всех компонент  $w_{*j} = w_{sj}$ . Легко видеть, что допустимость последовательностей  $\mathbf{c}$  и наборов  $\mathbf{w}$  полностью определяется укороченными наборами: если  $\mathbf{w}'(\mathbf{b}_1, \mathbf{c}_1) = \mathbf{w}'(\mathbf{b}_2, \mathbf{c}_2)$ , то  $\mathbf{b}_1\omega\mathbf{c}_1 \Leftrightarrow \mathbf{b}_2\omega\mathbf{c}_2$  и  $\mathbf{w}(\mathbf{b}_1, \mathbf{c}_1) \in W \Leftrightarrow \mathbf{w}(\mathbf{b}_2, \mathbf{c}_2) \in W$ .

Рассмотрим класс  $\mathcal{K}_{n'}(\mathbf{I}')$ , последовательности которого получены из последовательностей класса  $\mathcal{K}_n(\mathbf{1})$  удалением символов  $*$ . Для него  $n' = n - l_*$ , набор  $\mathbf{I}'$  образован из  $\mathbf{1}$  устранением компоненты  $l_*$ . Множеством  $W'$  допустимых наборов для  $\mathcal{K}_{n'}(\mathbf{I}')$  считаем множество всех укороченных наборов  $\mathbf{w}'$  для  $\mathbf{w} \in W$ , где  $W$  — множество допустимых наборов для  $\mathcal{K}_n(\mathbf{1})$ . Следующая теорема, обобщающая (13) и (14), означает справедливость эффекта Нечипорука при кодировании с произвольным отношением допустимости, не зависящим от нумерации разрядов.

**Теорема 8.** Комбинаторная  $W$ -энтропия класса  $\mathcal{K}_n(\mathbf{1})$  и комбинаторная  $W'$ -энтропия класса  $\mathcal{K}_{n'}(\mathbf{I}')$  связаны соотношениями

$$\log N_{n',W'}(\mathbf{I}') \leq \log N_{n,W}(\mathbf{1}) \leq \log N_{n',W'}(\mathbf{I}') + c \log n, \quad (17)$$

где  $c = c(|I|, |J|)$  — константа.

**Доказательство.** В последующих выкладках считаем, что  $i \in I$ ,  $j \in J$ ,  $u \in I \setminus \{s\}$ , где  $b_s = *$ .

Пусть значение  $h_{n',W'}(\mathbf{I}')$  достигается на наборе  $\mathbf{w}'$ , т.е.  $h_{n',W'}(\mathbf{I}') = h_{n'}(\mathbf{I}', \mathbf{w}')$ ,  $\mathbf{w}' = (w_{uj})$ . Положим  $m'_j = \sum_u w_{uj}$  и образуем набор  $\mathbf{w} = (w_{ij})$ , дополнив набор  $\mathbf{w}'$  компонентами  $w_{*j} = \frac{l_* m'_j}{n'}$ . Для него

$$m_j = \sum_i w_{ij} = \sum_u w_{uj} + w_{*j} = m'_j + \frac{l_* m'_j}{n'} = \frac{m'_j n}{n'}.$$

С учетом этого вычислим  $h_n(\mathbf{l}, \mathbf{w})$ , игнорируя тот факт, что значения  $w_{*j}$  могут быть нецелыми:

$$h_n(\mathbf{l}, \mathbf{w}) = \sum_{i,j} w_{ij} \log \frac{nw_{ij}}{l_i m_j} = \sum_{u,j} w_{uj} \log \frac{nw_{uj}}{l_u \frac{m'_j n}{n'}} + \sum_j \frac{l_* m'_j}{n'} \log \frac{\frac{l_* m'_j}{n'} n}{\frac{m'_j n}{l_* \frac{n'}{n'}}}.$$

Это выражение преобразуется к виду

$$\sum_u \frac{n' w_{uj}}{l_u m'_j} + \sum_j \frac{l_* m'_j}{n'} \log 1 = h_{n'}(\mathbf{l}', \mathbf{w}'),$$

откуда  $h_n(\mathbf{l}, \mathbf{w}) = h_{n'}(\mathbf{l}', \mathbf{w}') = h_{n', W'}(\mathbf{l}')$ .

Округлим компоненты  $w_{*j}$  набора  $\mathbf{w}$  до ближайших целых (больших или меньших)  $\hat{w}_{*j}$  с соблюдением условия  $\sum_j \hat{w}_{*j} = l_*$ , в результате чего  $\mathbf{w}$  переходит в допустимый набор  $\hat{\mathbf{w}}$ . Выражение для  $h_n(\mathbf{l}, \mathbf{w})$  может быть представлено как сумма и разность ограниченного числа членов вида  $v \log v$ ,  $v \leq n$  (см. доказательство нижней оценки в теореме 5). При переходе от  $\mathbf{w}$  к  $\hat{\mathbf{w}}$  величины  $v$  изменятся не более чем на константу, а значение  $v \log v$  — не более чем на  $c_1 \log v \leq c_1 \log n$ . Отсюда  $h_n(\mathbf{l}, \hat{\mathbf{w}}) \leq h_n(\mathbf{l}, \mathbf{w}) + c_2 \log n$ . С учетом полученных соотношений, используя теорему 7, находим

$$\begin{aligned} \log N_{n,W}(\mathbf{l}) &\leq h_{n,W}(\mathbf{l}) + c_3 \log n \leq h_n(\mathbf{l}, \hat{\mathbf{w}}) + c_3 \log n \leq h_n(\mathbf{l}, \mathbf{w}) + c_4 \log n = \\ &= h_{n', W'}(\mathbf{l}') + c_4 \log n \leq \log N_{n', W'}(\mathbf{l}') + c_5 \log n. \end{aligned}$$

Это дает верхнюю оценку теоремы.

Нижняя доказывается аналогично нижней оценке в (14). ■

Теорему 3 можно рассматривать как теоретико-информационный аналог теоремы 6. Подобный аналог имеется и у теоремы 8.

Пусть источник  $X$  порождает символы  $b_i$  конечного алфавита  $B = \{b_i, i \in I\}$  независимо с вероятностями  $p(b_i)$ , и они должны быть представлены у адресата символами  $c_j$  конечного алфавита  $C = \{c_j, j \in J\}$ . Условия точности задаются указанием множества  $W$  допустимых совместных распределений  $p(b_i, c_j)$ ,  $i \in I, j \in J$ . Пусть в алфавите  $B$  имеется неопределенный символ  $b_s = *$ . Наряду с  $X$  рассмотрим источник  $X' = (B', P')$ , порождающий символы  $b_u$  алфавита  $B' = B \setminus \{*\}$  с вероятностями  $p'(b_u) = \frac{p(b_u)}{1 - p(*)}$ ,  $u \in I \setminus \{s\}$ . В качестве  $W'$  для него возьмем множество распределений  $p'(b_u, c_j) = \frac{p(b_u, c_j)}{1 - p(*)}$ ,  $u \in I \setminus \{s\}, j \in J$ , соответствующих распределениям  $p(b_i, c_j)$ ,  $i \in I, j \in J$ , из множества  $W$ .

Для  $W$ -энтропии имеет место соотношение

$$\mathcal{H}_W(P) = (1 - p(*)) \mathcal{H}_{W'}(P'), \quad (18)$$

доказательство которого опускаем. Оно является теоретико-информационным аналогом теоремы 8, ибо влечет равенство  $h_{n,W}(\mathbf{l}) = h_{n', W'}(\mathbf{l}')$ . Эффект Нечипорука проявляется и для более сложных информационных характеристик (см. дальше утверждение 6).

## 6. Свойства энтропии

Рассмотрим некоторые свойства энтропии  $\mathcal{H}(P)$  и сравним их со свойствами энтропии Шеннона  $H(P)$ .

**Теорема 9.** Энтропия  $\mathcal{H}(P)$  неотрицательна, причем  $\mathcal{H}(P) = 0$  тогда и только тогда, когда пересечение всех  $T$ , для которых  $p(a_T) > 0$ , непусто.

*Доказательство.* Неотрицательность очевидна. Пусть минимум  $\mathcal{H}(P, Q)$  в (2) достигается на наборе  $Q^0$ . Положим  $T^0 = \{i \in M \mid q_i^0 > 0\}$ . Если  $\mathcal{H}(P) = \mathcal{H}(P, Q^0) = 0$ , то для любого  $T$  с  $p(a_T) > 0$  выполнено  $\sum_{t \in T} q_t^0 = 1$ , а потому  $T$  содержит  $T^0$  и пересечение всех таких  $T$  непусто. Обратно, если пересечение непусто, то, назначив  $q_i = 1$  для некоторого  $i$  из этого пересечения и  $q_j = 0$  для всех  $j \neq i$ , получим набор  $Q$ , для которого  $\mathcal{H}(P, Q) = 0$ . ■

Таким образом, энтропия  $\mathcal{H}(X)$  недоопределенного источника  $X$  равна 0, лишь если порождаемые им последовательности могут быть доопределены до последовательностей из одинаковых символов. Этот факт обобщает известный результат для полностью определенного источника, энтропия которого равна 0, лишь если он порождает последовательности, образованные одинаковыми символами.

Укажем верхнюю границу энтропии источника  $X = (A, P)$  в функции от распределения  $(p(t), 1 \leq t \leq m)$  числа  $t$  доопределений символов источника:

$$p(t) = \sum_{T: |T|=t} p(a_T).$$

**Теорема 10.** Справедлива оценка

$$\mathcal{H}(P) \leq \log m - \sum_{1 \leq t \leq m} p(t) \log t,$$

достижимая для любого распределения  $(p(t), 1 \leq t \leq m)$ .

*Доказательство.* Эту оценку получим, вычислив  $\mathcal{H}(P, Q)$  на наборе  $Q = (1/m, \dots, 1/m)$ ,

$$\mathcal{H}(P) \leq - \sum_T p(a_T) \log \frac{|T|}{m} = \log m - \sum_t \sum_{|T|=t} p(a_T) \log t = \log m - \sum_t p(t) \log t.$$

Оценка достигается на наборе  $P$ , в котором всем  $t$ -элементным множествам  $T$  соответствуют равные вероятности  $p(t) / \binom{n}{t}$ ,  $1 \leq t \leq m$ . Из соображений симметрии [3] следует, что в этом случае  $\mathcal{H}(P, Q)$  минимизируется набором вероятностей  $q_0 = \dots = q_{m-1} = 1/m$ . ■

Если источник полностью определен, то  $p(1) = 1$ ,  $p(t) = 0$  для  $t \geq 2$ , и оценка теоремы превращается в известную оценку  $H(P) \leq \log m$ .

**Теорема 11.** Функция  $\mathcal{H}(P)$  вогнута, т. е. для любых  $P = (p(a_T), T \subseteq M)$ ,  $P' = (p'(a_T), T \subseteq M)$  и числа  $\alpha$ ,  $0 \leq \alpha \leq 1$ , выполнено

$$\mathcal{H}(\alpha P + (1 - \alpha)P') \geq \alpha \mathcal{H}(P) + (1 - \alpha) \mathcal{H}(P').$$

**Доказательство.** Пусть минимум функции  $\mathcal{H}(\alpha P + (1 - \alpha)P', Q)$  достигается на наборе  $Q$ . Тогда

$$\begin{aligned} \mathcal{H}(\alpha P + (1 - \alpha)P') &= -\alpha \sum_T p(a_T) \log \sum_{i \in T} q_i - (1 - \alpha) \sum_T p'(a_T) \log \sum_{i \in T} q_i \geq \\ &\geq \alpha \mathcal{H}(P) + (1 - \alpha) \mathcal{H}(P'), \end{aligned}$$

что и требовалось. ■

В отличие от обычной энтропии  $H(P)$ , функция  $\mathcal{H}(P)$  не является строго вогнутой. Из доказательства видно, что при  $\alpha \neq 0, 1$  равенство в утверждении теоремы имеет место лишь тогда, когда существует  $Q$ , минимизирующее  $\mathcal{H}(P, Q)$  и  $\mathcal{H}(P', Q)$  одновременно. Поскольку минимум в (3) достигается только при  $Q = P$ , для полностью определенных источников условием равенства является  $P = P'$  и функция  $H(P)$  строго вогнута.

Произведение  $XX'$  недоопределенных источников  $X = (A, P)$  и  $X' = (A', P')$  представляет собой источник, порождающий пары  $(a_T, a'_{T'})$  с некоторыми вероятностями  $p(a_T, a'_{T'})$ . При этом выполнены условия согласования

$$\sum_{T'} p(a_T, a'_{T'}) = p(a_T), \quad \sum_T p(a_T, a'_{T'}) = p(a'_{T'}). \quad (19)$$

Доопределением символа  $(a_T, a'_{T'})$  источника  $XX'$  считается всякий символ  $(a_i, a'_{i'})$ ,  $i \in T, i' \in T'$ . Источники *статистически независимы*, если  $p(a_T, a'_{T'}) = p(a_T)p(a'_{T'})$ . Энтропия  $\mathcal{H}(XX')$  определяется аналогично (1)–(2)

$$\mathcal{H}(XX') = \min_Q \left\{ - \sum_{T, T'} p(a_T, a'_{T'}) \log \sum_{i \in T, i' \in T'} q_{ii'} \right\},$$

где  $Q = (q_{ii'}, i \in M, i' \in M')$ ,  $q_{ii'} \geq 0$ ,  $\sum_{i, i'} q_{ii'} = 1$ .

**Теорема 12.** Для любых недоопределенных источников  $X$  и  $X'$

$$\mathcal{H}(XX') \leq \mathcal{H}(X) + \mathcal{H}(X'),$$

а если  $X$  и  $X'$  статистически независимы, то

$$\mathcal{H}(XX') = \mathcal{H}(X) + \mathcal{H}(X').$$

**Доказательство.** Пусть величины  $\mathcal{H}(X)$  и  $\mathcal{H}(X')$  в (2) достигаются на наборах  $Q = (q_i, i \in M)$  и  $Q' = (q'_{i'}, i' \in M')$ . Для всех  $i$  и  $i'$  положим  $q_{ii'} = q_i q'_{i'}$ . Тогда

$$\begin{aligned} \mathcal{H}(XX') &\leq - \sum_{T, T'} p(a_T, a'_{T'}) \log \left( \sum_{i \in T, i' \in T'} q_i q'_{i'} \right) = - \sum_{T, T'} p(a_T, a'_{T'}) \log \left( \sum_{i \in T} q_i \sum_{i' \in T'} q'_{i'} \right) = \\ &= - \sum_T p(a_T) \log \sum_{i \in T} q_i - \sum_{T'} p(a'_{T'}) \log \sum_{i' \in T'} q'_{i'} = \mathcal{H}(X) + \mathcal{H}(X'). \end{aligned} \quad (20)$$

Если  $X$  и  $X'$  статистически независимы, то  $p(a_T, a'_{T'}) = p(a_T)p(a'_{T'})$ , и в силу теоремы 1, примененной к  $X$  и  $X'$ , для всех пар  $(i, i')$  выполнено

$$\sum_{T \ni i, T' \ni i'} \frac{p(a_T, a'_{T'})}{\sum_{j \in T, j' \in T'} q_{jj'}} = \sum_{T \ni i, T' \ni i'} \frac{p(a_T)p(a'_{T'})}{\sum_{j \in T, j' \in T'} q_j q_{j'}} = \sum_{T \ni i} \frac{p(a_T)}{\sum_{j \in T} q_j} \sum_{T' \ni i'} \frac{p(a'_{T'})}{\sum_{j' \in T'} q_{j'}} \leq 1.$$

Строгое неравенство имеет место, лишь когда оно справедливо для хотя бы одного сомножителя. В этом случае  $q_i = 0$  либо  $q_{i'} = 0$ , а потому  $q_{ii'} = 0$ . По теореме 1, примененной к источнику  $XX'$ , заключаем, что

$$\mathcal{H}(XX') = - \sum_{T, T'} p(a_T, a'_{T'}) \log \sum_{i \in T, i' \in T'} q_{ii'},$$

и неравенство (20) обращается в равенство. ■

В отличие от обычной энтропии  $H$ , независимость источников не является необходимой для равенства  $\mathcal{H}(XX') = \mathcal{H}(X) + \mathcal{H}(X')$ . В частности, оно справедливо, когда области определения источников  $X$  и  $X'$  не пересекаются, т. е.  $p(a_T, a'_{T'}) > 0$ , лишь если  $a_T = *$  или  $a'_{T'} = *$  [10].

## 7. Теорема кодирования

Существенное место в теории информации отводится результатам о сжатии данных, формулируемым в терминах кодирования последовательностей, порождаемых источниками. Согласно им, нельзя добиться, чтобы средняя длина кода, приходящаяся на символ источника, была меньше энтропии, и возможно закодировать так, чтобы средняя длина на символ превосходила энтропию сколь угодно мало [3]. Аналогичный факт оказывается верным и для недоопределенных источников рассматриваемого вида. Напомним, что в случае недоопределенных последовательностей требуется по коду восстановить не саму последовательность, а какое-либо ее доопределение.

Последовательности длины  $n$  будем называть  $n$ -блоками (либо просто блоками) и обозначать символом  $B$ . Будем рассматривать следующий способ кодирования недоопределенного источника  $X = (A, P)$ . Возьмем некоторое множество  $\mathcal{D} \subseteq (A_0)^n$ , содержащее доопределения всех блоков из  $A^n$ , и каждому блоку  $B \in A^n$  сопоставим некоторое его доопределение  $D = D_B$  из  $\mathcal{D}$  (разным  $B$  могут соответствовать одинаковые  $D_B$ ). Закодируем блоки множества  $\mathcal{D}$  двоичными наборами с соблюдением условия однозначного декодирования (разделимости кода) [17]. Кодом блока  $B$  будем считать код его доопределения  $D_B$ . Последовательность, порождаемая источником  $X$ , кодируется путем разбиения ее на  $n$ -блоки и приписывания друг к другу кодов полученных  $n$ -блоков. Кодирование источника  $X = (A, P)$  называется *универсальным*, если оно не зависит от набора вероятностей  $P$ .

Обозначим через  $l_B$  длину слова, кодирующего блок  $B$ . Качество кодирования будем характеризовать средним числом кодовых символов на символ источника

$$\bar{l}_n = \frac{1}{n} \sum_B p(B) l_B,$$

где  $p(B) = p(a_{T_1})p(a_{T_2}) \dots p(a_{T_n})$  — вероятность блока  $B = a_{T_1}a_{T_2} \dots a_{T_n}$ , а сумма берется по всем  $n$ -блокам. Ставится задача оценки минимального значения  $\bar{l}_n$  по всем кодированиям источника  $X$ .

Следующий результат обобщает на недоопределенные источники теорему кодирования полностью определенных источников.

### Теорема 13.

1. При любом  $n$  и любом способе кодирования

$$\bar{l}_n \geq \mathcal{H}(X).$$

2. Существует универсальное кодирование, для которого

$$\bar{l}_n \leq \mathcal{H}(X) + O\left(\frac{\log n}{n}\right).$$

Из результата Р. Е. Кричевского [6] следует, что при универсальном кодировании понизить порядок остаточного члена в п. 2 нельзя.

**Доказательство.** Нижняя оценка. Блоки ( $n$ -блоки)  $a_{i_1} \dots a_{i_n}$  в алфавите  $A_0$  будем обозначать через  $D_{\mathbf{i}}$ ,  $\mathbf{i} = (i_1, \dots, i_n)$ , а блоки  $a_{T_1} \dots a_{T_n}$  в алфавите  $A$  — через  $D_{\mathbf{T}}$ ,  $\mathbf{T} = T_1 \times \dots \times T_n$ . Для вероятности блока  $D_{\mathbf{T}}$  будем использовать обозначение  $p_{\mathbf{T}}$ .

Каждому блоку  $D_{\mathbf{T}}$  соответствует единственный блок  $D_{\mathbf{i}(\mathbf{T})}$ , взятый в качестве его доопределения. Положим для  $\mathbf{i} \in M^n$  и  $\mathbf{T} \subseteq M^n$

$$\delta_{\mathbf{T}\mathbf{i}} = \begin{cases} 1, & \mathbf{i} = \mathbf{i}(\mathbf{T}), \\ 0, & \mathbf{i} \neq \mathbf{i}(\mathbf{T}) \end{cases} \quad (21)$$

и введем величины

$$q_{\mathbf{i}} = \sum_{\mathbf{T}} p_{\mathbf{T}} \delta_{\mathbf{T}\mathbf{i}}, \quad \mathbf{i} \in M^n. \quad (22)$$

Для них выполнено

$$\sum_{\mathbf{i}} q_{\mathbf{i}} = \sum_{\mathbf{T}} p_{\mathbf{T}} \sum_{\mathbf{i}} \delta_{\mathbf{T}\mathbf{i}} = \sum_{\mathbf{T}} p_{\mathbf{T}} = 1. \quad (23)$$

Пусть  $\mathbf{I} = \{\mathbf{i} \mid q_{\mathbf{i}} > 0\}$ , тогда

$$\sum_{\mathbf{i} \in \mathbf{I}} q_{\mathbf{i}} = 1. \quad (24)$$

Через  $l_{\mathbf{i}}$  обозначим длину кода для  $D_{\mathbf{i}}$ . Согласно определению и (21),

$$l_{D_{\mathbf{T}}} = l_{\mathbf{i}(\mathbf{T})} = \sum_{\mathbf{i} \in \mathbf{T}} \delta_{\mathbf{T}\mathbf{i}} l_{\mathbf{i}}.$$

Отсюда и из (22) с учетом введенных обозначений получаем

$$n\bar{l}_n = \sum_{\mathbf{T}} p_{\mathbf{T}} l_{D_{\mathbf{T}}} = \sum_{\mathbf{T}} p_{\mathbf{T}} \sum_{\mathbf{i} \in \mathbf{I}} \delta_{\mathbf{T}\mathbf{i}} l_{\mathbf{i}} = \sum_{\mathbf{i} \in \mathbf{I}} q_{\mathbf{i}} l_{\mathbf{i}}. \quad (25)$$

Набор длин  $l_{\mathbf{i}}$ ,  $\mathbf{i} \in \mathbf{I}$ , удовлетворяет неравенству Макмиллана — Крафта [17]  $\sum_{\mathbf{i}} 2^{-l_{\mathbf{i}}} \leq 1$ , и по свойству (3) энтропийной функции выполнено

$$\begin{aligned} \sum_{\mathbf{i} \in \mathbf{I}} q_{\mathbf{i}} l_{\mathbf{i}} &= - \sum_{\mathbf{i} \in \mathbf{I}} q_{\mathbf{i}} \log 2^{-l_{\mathbf{i}}} \geq - \sum_{\mathbf{i} \in \mathbf{I}} q_{\mathbf{i}} \log q_{\mathbf{i}} = - \sum_{\mathbf{T}, \mathbf{i} \in \mathbf{I}} p_{\mathbf{T}} \delta_{\mathbf{T}\mathbf{i}} \log q_{\mathbf{i}} = \\ &= - \sum_{\mathbf{T}} p_{\mathbf{T}} \log q_{\mathbf{i}(\mathbf{T})} \geq - \sum_{\mathbf{T}} p_{\mathbf{T}} \log \sum_{\mathbf{i} \in \mathbf{T}} q_{\mathbf{i}}. \end{aligned} \quad (26)$$

Используя определение энтропии применительно к источнику  $X^n$  и утверждение 2, заключаем, что правая часть в (26) не превосходит  $\mathcal{H}(X^n) = n\mathcal{H}(X)$ . Отсюда, из (25) и (26) получаем утверждение п. 1 теоремы.

Верхняя оценка. Для каждого класса  $\mathcal{K}_n(\mathbf{I})$  возьмем доопределяющее множество мощности  $N_n(\mathbf{I})$  и в соответствии с теоремой 5 занумеруем входящие в него блоки двоичными наборами  $\tilde{\alpha}$  длины  $n\mathcal{H}(\mathbf{I}/n) + O(\log n)$ . Сами классы  $\mathcal{K}_n$  также занумеруем

двоичными наборами  $\tilde{\beta}$  одинаковой длины, которую можно взять равной  $O(\log n)$ . Блоку  $B$  припишем кодовое слово  $\tilde{\beta}\tilde{\alpha}$ , где  $\tilde{\beta}$  соответствует классу  $\mathcal{K}_n(\mathbf{1})$ , содержащему  $B$ , а  $\tilde{\alpha}$  — номер некоторого доопределения блока  $B$ . Тогда

$$l_B = l(\tilde{\beta}\tilde{\alpha}) \leq n\mathcal{H}(\mathbf{1}/n) + O(\log n). \quad (27)$$

Нетрудно видеть, что построенное кодовое множество является префиксным и, следовательно, разделимым [17].

Обозначим через  $p(P, \mathbf{1}, n)$  суммарную вероятность блоков класса  $\mathcal{K}_n(\mathbf{1})$ , равную  $(n! / \prod_T l_T!) \prod_T p_T^{l_T}$ . Поскольку  $\sum_{\mathbf{1}} p(P, \mathbf{1}, n) = 1$ , из (27) получаем

$$\bar{l}_n = \frac{1}{n} \sum_B p(B) l_B \leq \sum_{\mathbf{1}} p(P, \mathbf{1}, n) \mathcal{H}\left(\frac{\mathbf{1}}{n}\right) + O\left(\frac{\log n}{n}\right). \quad (28)$$

По свойству полиномиального распределения [2] выполнено

$$\sum_{\mathbf{1}} p(P, \mathbf{1}, n) \frac{l_T}{n} = p_T, \quad (T \subseteq M), \quad (29)$$

где  $l_T$  и  $p_T$  — компоненты наборов  $\mathbf{1}$  и  $P$ . Применив к вогнутой функции  $\mathcal{H}$  (теорема 11) неравенство Йенсена, получим с учетом (29)

$$\sum_{\mathbf{1}} p(P, \mathbf{1}, n) \mathcal{H}\left(\frac{\mathbf{1}}{n}\right) \leq \mathcal{H}\left(\sum_{\mathbf{1}} p(P, \mathbf{1}, n) \frac{\mathbf{1}}{n}\right) = \mathcal{H}(P).$$

Отсюда и из (28) следует утверждение п. 2. ■

## 8. Лучшее доопределение

Доопределение  $\dot{X}$  источника  $X$  представляет собой полностью определенный источник, который строится по  $X$  применением операции доопределения (см. ниже). Чтобы различать символы  $a_i$  источников  $X$  и  $\dot{X}$ , символ  $a_i$  в доопределении  $\dot{X}$  будем обозначать через  $\dot{a}_i$ . Операция доопределения задается набором переходных вероятностей  $p(\dot{a}_i | a_T)$ ,  $T \subseteq M$ ,  $i \in M$ , где  $p(\dot{a}_i | a_T) = 0$  для  $i \notin T$ . Ее применение к  $X$  дает источник  $\dot{X}$ , порождающий символы  $\dot{a}_i$  с вероятностями  $p(\dot{a}_i) = \sum_T p(a_T) p(\dot{a}_i | a_T)$ . Пара  $(a_T, \dot{a}_i)$  имеет в  $X\dot{X}$  вероятность  $p(a_T, \dot{a}_i) = p(a_T) p(\dot{a}_i | a_T)$ . Дальше считаем  $p(\dot{a}_i) > 0$ , ибо символы  $\dot{a}_i$  с  $p(\dot{a}_i) = 0$  не участвуют в доопределении и их можно исключить из рассмотрения.

Функция  $\check{\mathcal{H}}$  из (8) может быть записана в виде

$$\check{\mathcal{H}}(X) = \min_{\dot{X}} I(X, \dot{X}),$$

где

$$I(X, \dot{X}) = \sum_{T,i} p(a_T) p(\dot{a}_i | a_T) \log \frac{p(\dot{a}_i | a_T)}{p(\dot{a}_i)}$$

— величина взаимной информации [3], а минимум берется по всем доопределениям  $\dot{X}$  источника  $X$ . Отсюда и из теоремы 4 следует

**Утверждение 2.** Для любого доопределения  $\dot{X}$  источника  $X$

$$I(X, \dot{X}) \geq \mathcal{H}(X),$$

и существует доопределение  $\hat{X}$ , для которого

$$I(X, \hat{X}) = \mathcal{H}(X). \quad (30)$$

Доопределение  $\hat{X}$ , удовлетворяющее условию (30), назовем *лучшим*. Этот термин возник из следующей сложностной интерпретации. Пусть недоопределенный источник  $X$  имеет единственное лучшее доопределение  $\hat{X}$  (такая ситуация типична — см. дальше). Рассмотрим кодирование «достаточно длинных» последовательностей в алфавите  $A$ , содержащих символы  $a_T$  с частотами  $p(a_T)$ . Можно доказать, что если средняя длина кода асимптотически минимальна, то для почти всех последовательностей символы  $a_T$  замещаются в доопределениях символами  $\hat{a}_i$  с частотами, асимптотически равными вероятностям пар  $(a_T, \hat{a}_i)$  в произведении  $X\hat{X}$ .

Следующая теорема указывает параметры лучшего доопределения.

**Теорема 14.** Доопределение  $\hat{X}$  является лучшим тогда и только тогда, когда оно задается переходными вероятностями

$$p(\hat{a}_i | a_T) = \frac{\hat{q}_i}{\sum_{j \in T} \hat{q}_j}, \quad i \in T, \quad (31)$$

где  $\hat{Q} = (\hat{q}_i, i \in M)$  — один из наборов, на которых достигается  $\mathcal{H}(X)$ .

**Доказательство.** Рассмотрим произвольное доопределение  $\dot{X}$ . Положим  $\dot{P} = (p(\dot{a}_j), j \in M)$  и  $J^+ = \{j \mid p(\dot{a}_j) > 0\}$ . При заданном  $T$  введем для  $i \in T \cap J^+$  величины

$$\alpha_i^{(T)} = \frac{p(\dot{a}_i)}{\sum_{j \in T} p(\dot{a}_j)}, \quad x_i^{(T)} = \frac{p(\dot{a}_i | a_T)}{p(\dot{a}_i)}. \quad (32)$$

Поскольку  $\sum_{i \in T \cap J^+} p(\dot{a}_i) = \sum_{i \in T} p(\dot{a}_i)$ , выполнено  $\sum_{i \in T \cap J^+} \alpha_i^{(T)} = 1$ . Воспользуемся для выпуклой функции  $f(x) = x \log x$  неравенством Иенсена

$$\sum_{i \in T \cap J^+} \alpha_i^{(T)} f(x_i^{(T)}) \geq f\left(\sum_{i \in T \cap J^+} \alpha_i^{(T)} x_i^{(T)}\right). \quad (33)$$

Учитывая, что

$$\sum_{i \in T \cap J^+} p(\dot{a}_i | a_T) = \sum_{i \in T} p(\dot{a}_i | a_T) = 1,$$

находим

$$\sum_{i \in T \cap J^+} \alpha_i^{(T)} x_i^{(T)} = \sum_{i \in T \cap J^+} \frac{p(\dot{a}_i | a_T)}{\sum_{j \in T} p(\dot{a}_j)} = \frac{1}{\sum_{j \in T} p(\dot{a}_j)}.$$

Подстановка этого значения в (33) дает неравенство

$$\sum_{i \in T \cap J^+} \frac{p(\dot{a}_i | a_T)}{\sum_{j \in T} p(\dot{a}_j)} \log \frac{p(\dot{a}_i | a_T)}{p(\dot{a}_i)} \geq \frac{1}{\sum_{j \in T} p(\dot{a}_j)} \log \frac{1}{\sum_{j \in T} p(\dot{a}_j)}.$$

Принимая во внимание, что  $p(\hat{a}_i|a_T) = 0$  для  $i \notin J^+$ , заменим в этом неравенстве суммирование по  $i \in T \cap J^+$  суммированием по  $i \in T$ . После этого, домножив обе части на  $p(a_T) \sum_{j \in T} p(\hat{a}_j)$  и просуммировав по  $T$ , получим

$$I(X, \dot{X}) \geq \mathcal{H}(P, \dot{P}) \geq \mathcal{H}(X).$$

Отсюда видно, что равенство (30) возможно, лишь если на наборе  $\hat{P}$ , соответствующем доопределению  $\hat{X}$ , достигается энтропия  $\mathcal{H}(X)$ . Пусть  $\hat{Q}$  — произвольная точка минимума функции  $\mathcal{H}(P, Q)$ ,  $J^+ = \{j \mid \hat{q}_j > 0\}$  и доопределение  $\hat{X}$  таково, что  $\hat{P} = \hat{Q}$ . В этом случае равенство (30) имеет место тогда и только тогда, когда при каждом  $T$  неравенство (33) обращается в равенство. Поскольку для  $i \in T \cap J^+$  все  $\alpha_i^{(T)}$  положительны, условием равенства является совпадение всех  $x_i^{(T)}$ ,  $i \in T \cap J^+$ . Обозначив их величину через  $\gamma_T$ , получаем  $p(\hat{a}_i|a_T) = \gamma_T \hat{q}_i$ . Это равенство справедливо для всех  $i \in T$ , ибо при  $i \notin J^+$  обе части обращаются в 0. Отсюда и из  $\sum_{j \in T} p(\hat{a}_j|a_T) = 1$  находим

$$\gamma_T = \frac{1}{\sum_{j \in T} \hat{q}_j},$$

что приводит к (31). Чтобы можно было назначить  $p(\hat{a}_i|a_T)$  указанным образом, необходимо выполнение равенств  $\sum_T p(a_T) p(\hat{a}_i|a_T) = p(\hat{a}_i)$ ,  $i \in M$ . Их справедливость следует из того, что при условиях (31) они совпадают с (7). ■

Недоопределенный источник назовем *категоричным*, если он имеет единственное лучшее доопределение.

**Следствие 2.** Источник категоричен тогда и только тогда, когда его энтропия достигается в единственной точке.

Частично определенный источник  $(A_0 \cup \{*\}, P)$ ,  $P = (p_0, \dots, p_{m-1}, p_*)$  назовем *нетривиальным*, если  $p_* < 1$ .

**Следствие 3.** Всякий нетривиальный частично определенный источник категоричен и для него  $\hat{Q} = P/(1 - p_*)$ .

Действительно, с учетом  $\hat{q}_0 + \dots + \hat{q}_{m-1} = 1$  неравенства (4) в рассматриваемом случае имеют вид  $p_i/\hat{q}_i + p_* \leq 1$ . Если  $p_i = 0$ , то имеет место строгое неравенство и  $\hat{q}_i = 0$ . В случае  $p_i > 0$  выполнено  $\hat{q}_i \geq p_i/(1 - p_*) > 0$ , неравенство заменяется равенством и дает  $\hat{q}_i = p_i/(1 - p_*)$ . ■

Приведем без доказательства достаточные условия для того, чтобы энтропия  $\mathcal{H}(X)$  недоопределенного источника  $X$  достигалась в единственной точке. Обозначим через  $\mathcal{T}$  систему множеств, состоящую из всех  $T$  с  $p(a_T) > 0$  и множества  $M$ . Положим  $\chi_{i,T}$  равным 1 при  $i \in T$  и равным 0 при  $i \notin T$  и образуем наборы  $\chi_i = (\chi_{i,T}, T \in \mathcal{T})$ ,  $i \in M$ . Рассмотрим выпуклый многогранник  $\sum_{i \in M} \mu_i \chi_i$ ,  $0 \leq \mu_i \leq 1$ ,  $\sum_i \mu_i = 1$ , натянутый на точки  $\chi_i$ . Пусть  $M'$  — множество всех  $i$ , для которых  $\chi_i$  принадлежат границе многогранника.

**Утверждение 3.** Если  $\chi_i$ ,  $i \in M'$ , линейно независимы, то энтропия  $\mathcal{H}(P)$  достигается на единственном наборе  $Q$ .

«Типичным» следует считать случай, когда лежащие на границе многогранника не более  $m$  точек  $\chi_i$  (размерность которых может доходить до  $2^m$ ) линейно независимы. Это обеспечивает единственность набора  $Q$ , на котором достигается энтропия  $\mathcal{H}(X)$ , и, следовательно, — категоричность источника  $X$ .

### 9. Условная энтропия недоопределенных данных

Пусть произведение  $XY$  недоопределенных источников с алфавитами  $A = \{a_T, T \subseteq M\}$  и  $B = \{b_U, U \subseteq L\}$ ,  $L = \{0, 1, \dots, l-1\}$ , задано совместным распределением  $(p(a_T, b_U), T \subseteq M, U \subseteq L)$ .

Если, как обычно, считать меру информации  $\mathcal{I}(X, Y)$  в  $X$  о  $Y$  связанной с условной энтропией  $\mathcal{H}(Y|X)$  соотношением

$$\mathcal{I}(X, Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X),$$

то для введения меры информации достаточно дать определение условной энтропии. В теории информации важную роль играет правило сложения энтропий  $H(XY) = H(X) + H(Y|X)$ . Оно (в более слабом варианте) включено К. Шенноном в число свойств, аксиоматически определяющих вид энтропийной функции. Поэтому при введении условной энтропии  $\mathcal{H}(Y|X)$  будем существенное внимание уделять этому свойству.

Рассмотрим вначале случай, когда  $X$  полностью определен, т. е. совместное распределение имеет вид  $p(a_i, b_U)$ ,  $i \in M$ ,  $U \subseteq L$ . Тогда

$$p(a_i) = \sum_U p(a_i, b_U), \quad p(b_U|a_i) = \frac{p(a_i, b_U)}{p(a_i)} \quad (34)$$

(считаем  $p(a_i) > 0$ , ибо при  $p(a_i) = 0$  символ  $a_i$  может быть исключен из рассмотрения). Условную энтропию введем обычным равенством [3]

$$\mathcal{H}(Y|X) = \sum_i p(a_i) \mathcal{H}(Y|a_i),$$

где величины  $\mathcal{H}(Y|a_i)$  находятся подобно (1)–(2) с заменой вероятностей условными вероятностями

$$\mathcal{H}(Y|a_i) = \min_Q \left\{ - \sum_{U \subseteq L} p(b_U|a_i) \log \sum_{j \in U} q_j \right\}, \quad (35)$$

$Q = (q_j, j \in L)$ ,  $q_j \geq 0$ ,  $\sum_j q_j = 1$ . Применительно к набору  $Q^{(i)}$ , на котором достигается минимум, условия (4) теоремы 1 приобретают вид

$$\sum_{U: j \in U} \frac{p(b_U|a_i)}{\sum_{u \in U} q_u^{(i)}} \leq 1, \quad j \in L. \quad (36)$$

**Теорема 15.** В случае полностью определенного источника  $X$  введенная указанным образом условная энтропия удовлетворяет правилу сложения энтропий

$$\mathcal{H}(X) + \mathcal{H}(Y|X) = \mathcal{H}(XY).$$

*Доказательство.* Положим  $q_{ij} = p(a_i)q_j^{(i)}$  и  $\tilde{Q} = (q_{ij}, i \in M, j \in L)$ . Выполнено

$$\sum_{i,j} q_{ij} = \sum_i p(a_i) \sum_j q_j^{(i)} = 1.$$

Убедимся, что набор  $\tilde{Q}$  удовлетворяет аналогам неравенств (4) для  $XY$ . Имеем для  $i \in M, j \in L$  с учетом (36)

$$\sum_{U: j \in U} \frac{p(a_i, b_U)}{\sum_{u \in U} q_{iu}} = \sum_{U: j \in U} \frac{p(a_i)p(b_U|a_i)}{\sum_{u \in U} p(a_i)q_u^{(i)}} = \sum_{U: j \in U} \frac{p(b_U|a_i)}{\sum_{u \in U} q_u^{(i)}} \leq 1.$$

Если выполнено строгое неравенство, то строгим является и неравенство (36), а потому  $q_j^{(i)} = 0$  и  $q_{ij} = p(a_i)q_j^{(i)} = 0$ .

Это означает, что  $\mathcal{H}(XY)$  достигается в точке  $\tilde{Q}$  и

$$\begin{aligned} \mathcal{H}(XY) &= - \sum_{i,U} p(a_i, b_U) \log \sum_{j \in U} q_{ij} = - \sum_{i,U} p(a_i, b_U) \log \sum_{j \in U} p(a_i)q_j^{(i)} = \\ &= - \sum_{i,U} p(a_i, b_U) \log p(a_i) - \sum_i p(a_i) \sum_U p(b_U|a_i) \log \sum_{j \in U} q_j^{(i)} = \\ &= - \sum_i p(a_i) \log p(a_i) + \sum_i p(a_i) \mathcal{H}(Y|a_i) = \mathcal{H}(X) + \mathcal{H}(Y|X). \end{aligned}$$

Теорема доказана. ■

Введенное понятие  $\mathcal{H}(Y|X)$  для полностью определенного  $X$  оказывается согласованным с колмогоровской интерпретацией [5] условной энтропии в терминах относительной сложности (подробнее см. в [11, 13]). Оно служит основой для определения условной энтропии в общем случае.

Если опыт  $X$ , участвующий в произведении  $XY$ , недоопределен, применим к  $X$  некоторое преобразование доопределения. Переходные вероятности  $p(\dot{a}_i|a_T)$ , задающие доопределение  $\dot{X}$ , позволяют найти совместные вероятности  $p(\dot{a}_i, b_U) = \sum_T p(a_T, b_U)p(\dot{a}_i|a_T)$  пар  $(\dot{a}_i, b_U)$  в произведении  $\dot{X}Y$ . Поскольку  $\dot{X}$  полностью определен, может быть вычислена условная энтропия  $\mathcal{H}(Y|\dot{X})$ . Эта величина зависит от доопределения  $\dot{X}$ . Выделим содержательно важный случай, когда такой зависимости нет.

Будем говорить, что *символ  $a_T$  конкретней  $b_U$* , если  $a_T$  является основным (т. е. принадлежит  $A_0$ ) либо  $b_U$  является неопределенным (т. е. совпадает с  $*$  =  $b_L$ ), и что *опыт  $X$  конкретней  $Y$* , если  $p(a_T, b_U) > 0$ , лишь когда  $a_T$  конкретней  $b_U$ . В частности, полностью определенный опыт  $X$  конкретней (не строго) любого опыта  $Y$ . Имеет место следующий факт.

**Утверждение 4.** Если опыт  $X$  конкретней  $Y$ , то величина  $\mathcal{H}(Y|\dot{X})$  не зависит от выбора доопределения  $\dot{X}$ .

*Доказательство.* Пусть  $X$  конкретней  $Y$ . Учитывая лишь компоненты  $p(a_i, b_U)$  набора совместных вероятностей для  $XY$ , соответствующие одноэлементным множествам  $\{T\}$ , введем  $p(b_U|a_i)$  и  $\mathcal{H}(Y|a_i)$  теми же выражениями (34), (35), что и для опыта  $X$  с полностью определенными исходами, и положим

$$\mathcal{H}_X(Y) = \sum_i p(a_i) \mathcal{H}(Y|a_i).$$

Отметим, что  $\mathcal{H}_X(Y)$  не является математическим ожиданием величин  $\mathcal{H}(Y|a_i)$ , поскольку  $\sum_i p(a_i)$ , вообще говоря, меньше 1. Покажем, что при любом доопределении  $\dot{X}$  имеет место равенство  $\mathcal{H}(Y|\dot{X}) = \mathcal{H}_X(Y)$ . Этим утверждение будет доказано.

Так как  $X$  конкретней  $Y$ , то  $p(\dot{a}_i, b_U) = p(a_i, b_U)$  при  $U \neq L$  и  $p(\dot{a}_i, *) = p(a_i, *) + p(\dot{a}_i) - p(a_i)$ . С учетом этого, введя обозначение  $r_i = 1 - \frac{p(a_i)}{p(\dot{a}_i)}$ , получаем  $p(b_U|a_i) = \frac{p(b_U|\dot{a}_i)}{1 - r_i}$  при  $U \neq L$  и  $p(*|a_i) = \frac{p(*|\dot{a}_i) - r_i}{1 - r_i}$ . По утверждению 3 заключаем, что  $(1 - r_i)\mathcal{H}(Y|a_i) = \mathcal{H}(Y|\dot{a}_i)$ . Это эквивалентно соотношению  $p(a_i)\mathcal{H}(Y|a_i) = p(\dot{a}_i)\mathcal{H}(Y|\dot{a}_i)$ , суммированием которого по  $i$  получаем требуемое равенство. ■

Поскольку в рассматриваемом случае величина  $\mathcal{H}(Y|\dot{X})$  не зависит от доопределения  $\dot{X}$ , под условной энтропией  $\mathcal{H}(Y|X)$  можно понимать  $\mathcal{H}(Y|\dot{X})$  при произвольном  $\dot{X}$ . Явное выражение  $\mathcal{H}(Y|X)$  через параметры произведения  $XU$  исходных источников задается формулой для  $\mathcal{H}_X(Y)$ .

Имеет место следующий факт, доказательство которого приведено в [11, 13].

**Утверждение 5.** Если источник  $X$  конкретней  $Y$ , справедливо правило сложения энтропий

$$\mathcal{H}(XY) = \mathcal{H}(X) + \mathcal{H}(Y|X).$$

Отметим также, что если источник  $X$  конкретней  $Y$ , то условная энтропия  $\mathcal{H}(Y|X)$  допускает колмогоровскую интерпретацию [5] (модифицированную применительно к недоопределенным данным) в терминах относительной сложности [11, 13].

В общем случае величина  $\mathcal{H}(Y|\dot{X})$  зависит от  $\dot{X}$ . Будем считать, что источник  $X$  категоричен, т. е. имеет единственное лучшее доопределение  $\hat{X}$  (такая ситуация «типична» — см. утверждение 3). Положим по определению  $\mathcal{H}(Y|X) = \mathcal{H}(Y|\hat{X})$ .

Содержательные соображения в пользу выбора лучшего доопределения для вычисления условной энтропии поясним на простом примере. Пусть в произведении  $XU$  источников с алфавитами  $\{0, 1, *\}$  и  $\{0, 1\}$  пары  $(0, 0)$  и  $(*, 1)$  порождаются с ненулевыми вероятностями, а вероятности остальных пар равны 0. Энтропия  $\mathcal{H}(X) = 0$  достигается на лучшем доопределении  $\hat{X}$ , в котором все  $*$  заменены символом 0. Для него  $\mathcal{H}(Y|\hat{X}) = \mathcal{H}(Y)$ . Если же доопределить все  $*$  символом 1, получим источник  $\dot{X}$ , совпадающий с  $Y$ . В этом случае  $\mathcal{H}(Y|\dot{X}) = 0$ . Таким образом, существует возможность уменьшения величины  $\mathcal{H}(Y|\dot{X})$  за счет «усложнения» доопределения  $\dot{X}$ . Правило сложения энтропий имеет дело с суммой  $\mathcal{H}(X) + \mathcal{H}(Y|X)$  и вычисление слагаемых на основе разных доопределений может привести к «искусственному» занижению суммы (в данном примере можно получить нулевую сумму). Поскольку  $\mathcal{H}(X)$  вычисляется на основе лучшего доопределения, то и в  $\mathcal{H}(Y|X)$  следует использовать лучшее доопределение.

## 10. Обобщенное правило сложения энтропий

*Продолжением доопределения  $\dot{X}$  на  $XU$*  называется полностью определенный источник  $\dot{X}Y'$ , построенный по  $XU\dot{X}$  применением некоторого набора переходных вероятностей  $p(b'_j|a_T, b_U, \dot{a}_i)$  и порождающий пары  $(\dot{a}_i, b'_j)$  с вероятностями

$$p(\dot{a}_i, b'_j) = \sum_{T,U} p(a_T, b_U, \dot{a}_i) p(b'_j|a_T, b_U, \dot{a}_i).$$

Подчеркнем разницу между доопределением источника  $\dot{X}Y$  и продолжением доопределения  $\dot{X}$  на  $XU$ . Доопределение строится по  $\dot{X}Y$  и не использует информацию о  $X$ , в то время как продолжение строится по  $XU\dot{X}$  и тем самым предоставляет более широкие возможности.

Величину

$$\hat{\mathcal{H}}(XY) = \min_{\hat{X}Y'} I(XY; \hat{X}Y'),$$

где  $I(\cdot, \cdot)$  — величина взаимной информации и минимум берется по продолжениям  $\hat{X}Y'$  лучшего доопределения  $\hat{X}$ , назовем *энтропией произведения  $XY$  при лучшем доопределении источника  $X$* . В общем случае недоопределенных данных правило сложения энтропий заменяется более сложным соотношением [16].

**Теорема 16.** Имеет место следующее обобщенное правило сложения энтропий

$$\mathcal{H}(X) + \mathcal{H}(Y|X) = \hat{\mathcal{H}}(XY). \quad (37)$$

Из теоремы 14 следует, что

$$\mathcal{H}(XY) = \min_{X'Y'} I(XY; X'Y'),$$

где минимум по всем доопределениям  $X'Y'$  произведения  $XY$ . В случае, когда минимум по  $X'Y' = \hat{X}Y'$  совпадает с минимумом по всем доопределениям  $X'Y'$ , обобщенное правило сложения энтропий превращается в обычное. Это имеет место, например, если  $X$  конкретной  $Y$ .

Величины, входящие в обобщенное правило, заданы неявно как минимумы некоторых выражений, в связи с чем его доказательство в общем виде требует громоздких выкладок (эскиз доказательства имеется в [16]). Но в содержательно важном случае — для частично определенных данных (когда используются лишь полностью определенные и неопределенные символы) — эти величины представимы в явном виде, что позволяет дать для этого случая более простое прямое доказательство. Ограничимся доказательством обобщенного правила для частично определенных данных.

Пусть произведение  $XY$  частично определенных источников  $X$  и  $Y$  с алфавитами  $\{a_0, \dots, a_{m-1}, *\}$  и  $\{b_0, \dots, b_{l-1}, *\}$  задано совместным распределением  $p(a_i, b_j) = p_{ij}$ ,  $p(*, b_j) = p_{*j}$ ,  $p(a_i, *) = p_{i*}$  и  $p(*, *) = p_{**}$ . Тогда  $p(a_i) = \sum_j p_{ij} = p_i$ ,  $p(*) = \sum_j p_{*j} = p_*$ . В силу следствия 3  $\mathcal{H}(X)$  достигается на единственном наборе  $\hat{Q} = (\hat{q}_i, i \in M)$ ,  $\hat{q}_i = \frac{p_i}{1 - p_*}$  и

$$\mathcal{H}(X) = - \sum_i p_i \log \hat{q}_i = (1 - p_*) \log(1 - p_*) - \sum_i p_i \log p_i. \quad (38)$$

По теореме 14 лучшее доопределение  $\hat{X}$  задается переходными вероятностями  $p(\hat{a}_i|a_i) = 1$ ,  $p(\hat{a}_i|*) = \hat{q}_i$ , и произведению  $\hat{X}Y$  соответствует совместное распределение  $p(\hat{a}_i, b_j) = p_{ij} + \hat{q}_i p_{*j} = \pi_{ij}$ ,  $p(\hat{a}_i, *) = p_{i*} + \hat{q}_i p_{**} = \pi_{i*}$ . С учетом того, что  $p(b_j|\hat{a}_i) = \pi_{ij}/\hat{q}_i$  и источник  $Y|\hat{a}_i$  частично определен, используя следствие 1, находим

$$\mathcal{H}(Y|X) = \sum_i \hat{q}_i \mathcal{H}(Y|\hat{a}_i) = - \sum_i \hat{q}_i \sum_j \frac{\pi_{ij}}{\hat{q}_i} \log \frac{\pi_{ij}}{\sum_v \pi_{iv}} = - \sum_{i,j} \pi_{ij} \log \frac{\pi_{ij}}{\sum_v \pi_{iv}}. \quad (39)$$

Для вычисления  $\hat{\mathcal{H}}(XY)$  понадобится утверждение, которое возникает как более сложное проявление эффекта Нечипорука, рассмотренного в разд. 13 (ср. с теоремой 3 и соотношением (18)). Оно справедливо для недоопределенных источников общего вида. Пусть в произведении  $XY$  пара  $(*, *)$  встречается с вероятностью  $p(*, *) > 0$ . Рассмотрим произведение  $X^0Y^0$ , в котором  $p^0(*, *) = 0$  и  $p^0(a_T, b_U) = \frac{p(a_T, b_U)}{1 - p(*, *)}$ .

**Утверждение 6.** Имеют место равенства

$$\begin{aligned}\mathcal{H}(X) &= (1 - p(*, *))\mathcal{H}(X^0), & \mathcal{H}(Y|X) &= (1 - p(*, *))\mathcal{H}(Y^0|X^0), \\ \hat{\mathcal{H}}(XY) &= (1 - p(*, *))\hat{\mathcal{H}}(X^0Y^0).\end{aligned}$$

В силу этих соотношений, при доказательстве обобщенного правила сложения энтропий можно ограничиться случаем, когда неопределенных пар  $(*, *)$  нет. Вернемся к частично определенным источникам и будем считать, что  $p_{**} = 0$ .

Рассмотрим продолжение  $\hat{X}Y'$  лучшего доопределения  $\hat{X}$  на  $XY$ . Пусть оно соответствует переходным вероятностям  $p(b'_j|a_i, *, \hat{a}_i) = x_{i,j}$ . Тогда  $p(a_i, b_j, \hat{a}_i, b'_j) = p_{ij}$ ,  $p(a_i, *, \hat{a}_i, b'_j) = p_{i*}x_{i,j}$ ,  $p(*, b_j, \hat{a}_i, b'_j) = \hat{q}_i p_{*j}$ ,  $p(\hat{a}_i, b'_j) = p_{ij} + p_{i*}x_{i,j} + \hat{q}_i p_{*j} = \pi_{ij} + p_{i*}x_{i,j}$ . Отметим, что  $\sum_{j \in J} x_{i,j} = 1$ .

Вычислим

$$\begin{aligned}I(XY, \hat{X}Y') &= \sum_{i,j} \left( p(a_i, b_j, \hat{a}_i, b'_j) \log \frac{p(a_i, b_j, \hat{a}_i, b'_j)}{p(a_i, b_j)p(\hat{a}_i, b'_j)} + \right. \\ &\left. + p(*, b_j, \hat{a}_i, b'_j) \log \frac{p(*, b_j, \hat{a}_i, b'_j)}{p(*, b_j)p(\hat{a}_i, b'_j)} + p(a_i, *, \hat{a}_i, b'_j) \log \frac{p(a_i, *, \hat{a}_i, b'_j)}{p(a_i, *)p(\hat{a}_i, b'_j)} \right).\end{aligned}$$

Подставив указанные значения и осуществив сокращения, получаем  $I(XY, \hat{X}Y') = \sum_i F_i$ , где

$$\begin{aligned}F_i &= \sum_j \left( p_{ij} \log \frac{1}{\pi_{ij} + p_{i*}x_{i,j}} + p_{i*}x_{i,j} \log \frac{x_{i,j}}{\pi_{ij} + p_{i*}x_{i,j}} + \hat{q}_i p_{*j} \log \frac{\hat{q}_i}{\pi_{ij} + p_{i*}x_{i,j}} \right) = \\ &= \sum_j \left( -(\pi_{ij} + p_{i*}x_{i,j}) \log(\pi_{ij} + p_{i*}x_{i,j}) + p_{i*}x_{i,j} \log x_{i,j} \right) + \hat{q}_i p_{*j} \log \hat{q}_i.\end{aligned}$$

Для разных  $i$  функции  $F_i$  зависят от разных переменных, поэтому минимум  $I(XY, \hat{X}Y')$  достигается минимизацией каждой из функций  $F_i$ . Отметим, что декомпозиция задачи стала возможной благодаря сведению к случаю  $p_{**} = 0$  применением утверждения 6.

При  $x \geq 0$  рассмотрим функцию  $f(x) = -(a + bx) \log(a + bx) + bx \log x$ ,  $a, b \geq 0$ . Она выпукла, ибо  $f''(x) = \frac{ab \log e}{x(a + bx)} > 0$  при  $a, b, x > 0$ , а при  $a = 0$  или  $b = 0$  функция  $f(x)$  линейна. Следовательно, функция  $F_i$  от переменных  $x_{i,j}$ ,  $j \in J$ , также выпукла. Осуществив ее дифференцирование и выполнив ряд преобразований, находим

$$\frac{\partial F_i}{\partial x_{i,j}} = p_{i*} \log \frac{x_{i,j}}{\pi_{ij} + p_{i*}x_{i,j}}.$$

Рассмотрим точку  $\alpha^{(i)} = (\alpha_{ij}, j \in J)$ , где  $\alpha_{ij} = \frac{\pi_{ij}}{\sum_v \pi_{iv}}$ , и вычислим  $\frac{\partial F_i(\alpha^{(i)})}{\partial x_{i,j}}$ . После некоторых преобразований имеем

$$\frac{\partial F_i(\alpha^{(i)})}{\partial x_{i,j}} = -p_{i*} \log \left( \sum_v \pi_{iv} + p_{i*} \right).$$

Это значение не зависит от  $j$ ; обозначим его через  $\lambda_i$ . Поскольку  $\sum_v \pi_{iv} + p_{i*} = \hat{q}_i < 1$ , выполнено  $\lambda_i > 0$ . Отсюда в силу теоремы 4.4.1 из [2] заключаем, что

минимум выпуклой функции  $F_i$  на множестве векторов вероятностей  $(x_{i,j}, j \in J)$  достигается в точке  $\alpha^{(i)}$ . Учитывая, что в этой точке

$$\pi_{ij} + p_{i*}x_{i,j} = \frac{\pi_{ij}}{\sum_v \pi_{iv}} (\sum_v \pi_{iv} + p_{i*}) = \hat{q}_i \frac{\pi_{ij}}{\sum_v \pi_{iv}},$$

после некоторых преобразований приводим  $F_i(\alpha^{(i)})$  к виду

$$F_i(\alpha^{(i)}) = \sum_j (-\hat{q}_i \frac{\pi_{ij}}{\sum_v \pi_{iv}} + \hat{q}_i p_{*j}) \log \hat{q}_i - \sum_j \pi_{ij} \log \frac{\pi_{ij}}{\sum_v \pi_{iv}}.$$

Преобразовав первую сумму к  $-(\hat{q}_i - \hat{q}_i \sum_j p_{*j}) \log \hat{q}_i = -p_i \log \hat{q}_i$ , получаем  $F_i(\alpha^{(i)}) = -p_i \log \hat{q}_i - \sum_j \pi_{ij} \log \frac{\pi_{ij}}{\sum_v \pi_{iv}}$ . Это в силу (38) и (39) дает  $\hat{\mathcal{H}}(XY) = \sum_i F_i(\alpha^{(i)}) = \mathcal{H}(X) + \mathcal{H}(Y|X)$ . Тем самым обобщенное правило сложения энтропий для частично определенных источников доказано.

#### ЛИТЕРАТУРА

1. Бонгард М. М. О понятии «полезная информация» // Проблемы кибернетики. Вып. 9. М.: Физматгиз, 1963. С. 71–102.
2. Вероятность и математическая статистика. Энциклопедия. М.: Большая Российская Энциклопедия, 1999.
3. Галлагер Р. Теория информации и надежная связь. М.: Сов. радио, 1974.
4. Добрушин Р. Л. Единые способы оптимального квантования сообщений // Проблемы кибернетики. Вып. 22. М.: Наука, 1970. С. 107–156.
5. Колмогоров А. Н. Алгоритм, информация, сложность. М.: Знание, 1991.
6. Кричевский Р. Е. Сжатие и поиск информации. М.: Радио и связь, 1989.
7. Нечипорук Э. И. О сложности вентильных схем, реализующих булевские матрицы с неопределенными элементами // ДАН СССР. 1965. Т. 163. № 1. С. 40–42.
8. Сапоженко А. А., Асратян А. С., Кузюрин Н. Н. Обзор некоторых результатов по задачам о покрытии // Методы дискретного анализа в решении комбинаторных задач. Вып. 30. Новосибирск: ИМ СО АН СССР, 1977. С. 46–75.
9. Шоломов Л. А. Информационные свойства функционалов сложности для систем недоопределенных булевых функций // Проблемы кибернетики. Вып. 34. М.: Наука, 1978. С. 133–150.
10. Шоломов Л. А. Сжатие частично определенной информации // Нелинейная динамика и управление. Вып. 4. М.: Физматлит, 2004. С. 385–399.
11. Шоломов Л. А. О мере информации нечетких и частично-определенных данных // Докл. Академии наук. 2006. Т. 410. № 1. С. 321–325.
12. Шоломов Л. А. О сложности последовательной реализации частичных булевых функций схемами // Дискрет. анализ и исслед. опер. Сер. 1. 2007. Т. 12. № 3. С. 110–139.
13. Шоломов Л. А. Информационные свойства недоопределенных данных // Дискретная математика и ее приложения: сб. лекций молодежных научных школ. Вып. IV. М.: ИПМ РАН, 2007. С. 26–50.
14. Шоломов Л. А. Исследование одного класса динамических процедур коллективного выбора // Нелинейная динамика и управление. Вып. 5. М.: Физматлит, 2007. С. 287–308.
15. Шоломов Л. А. О собственной информации нечетких текстов // Нелинейная динамика и управление. Вып. 6. М.: Физматлит, 2008. С. 305–314.
16. Шоломов Л. А. Обобщенное правило сложения энтропий для недоопределенных данных // Докл. Академии наук. 2009. Т. 427. № 1. С. 28–31.
17. Яблонский С. В. Введение в дискретную математику. М.: Высшая школа, 2006.
18. Berger T. Rate distortion theory. A mathematical basis for data compression. New Jersey: Prentice-Hall, 1971.