УДК 519.816

ЯВНЫЙ ВИД ИНФОРМАЦИОННЫХ ХАРАКТЕРИСТИК ЧАСТИЧНО ОПРЕДЕЛЕННЫХ ДАННЫХ[1]

Л. А. Шоломов

Пусть $M = \{0, 1, \ldots, m-1\}$ и каждому непустому подмножеству $T \subseteq M$ сопоставлен символ a_T. Алфавит символов a_T обозначим через A, а его подалфавит $\{a_0, a_1, \ldots, a_{m-1}\}$, 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, $i \in T$. Символ a_M, доопределимый любым основным символом, будем называть *неопределенным* и обозначать $*$.

Пусть имеется источник X, порождающий символы $a_T \in A$ независимо с вероятностями $p_T \geqslant 0$, $\sum_T p_T = 1$. Такой источник будем называть *недоопределенным*, при выполнении условия $p_T = 0$ для $a_T \notin A_0$ — *полностью определенным*, а в случае $p_T = 0$ для $a_T \notin A_0 \cup \{*\}$ — *частично определенным*.

Основные информационные характеристики недоопределенных данных заданы неявно как результат оптимизации некоторых выражений. Так, например, *энтропия источника* X вводится формулой

$$\mathcal{H}(X) = \min_Q \Big\{ -\sum_{T \subseteq M} p_T \log \sum_{i \in T} q_i \Big\},$$

где минимум берется по наборам $Q = (q_i,\ i \in M)$, $q_i \geqslant 0$, $\sum_i q_i = 1$ [1] (здесь и дальше логарифмы двоичные). Неявный вид информационных характеристик существенно затрудняет их исследование и нахождение соотношений между ними. Однако для важного и наиболее часто встречающегося типа недоопределенных данных — частично определенных удается решить соответствующие оптимизационные задачи и найти явные представления информационных характеристик.

Явное выражение энтропии частично определенного источника X с алфавитом $A = \{a_0, \ldots, a_{m-1}, *\}$ и набором вероятностей $p_0, \ldots, p_{m-1}, p_*$ имеет вид

$$\mathcal{H}(X) = (1 - p_*) \log(1 - p_*) - \sum_{0 \leqslant i \leqslant m-1} p_i \log p_i.$$

Если произведение XY частично определенных источников X и Y с алфавитами $A = \{a_0, \ldots, a_{m-1}, *\}$ и $B = \{b_0, \ldots, b_{l-1}, *\}$ задано совместным распределением p_{ij}, p_{i*}, p_{*j}, p_{**} ($i \in \{0, \ldots, m-1\}$, $j \in \{0, \ldots, l-1\}$), то условная энтропия $\mathcal{H}(Y|X)$ (определение см. в [2]) вычисляется следующим образом. Положим $p_i = \sum_j p_{ij} + p_{i*}$, $p_* = \sum_j p_{*j} + p_{**}$, $q_i = p_i/(1 - p_*)$, $\pi_{ij} = p_{ij} + q_i p_{*j}$, $\pi_i = \sum_j \pi_{ij}$. Тогда

$$\mathcal{H}(Y|X) = \sum_i \pi_i \log \pi_i - \sum_{i,j} \pi_{ij} \log \pi_{ij}.$$

Правило сложения энтропий $\mathcal{H}(X) + \mathcal{H}(Y|X) = \mathcal{H}(XY)$, играющее важную роль в теории информации, в случае недоопределенных данных заменяется некоторым более сложным соотношением — обобщенным правилом (условия, при которых обобщенное правило совпадает с обычным, приведены в [2]). Доказательство обобщенного

[1]Работа выполнена при поддержке Отделения нанотехнологий и информационных технологий РАН по программе фундаментальных исследований (проект 1-1 «Теория и методы эффективного использования недоопределенных данных»).

правила в общем случае достаточно громоздко. Для частично определенных данных можно дать более простое прямое доказательство, используя приведенные выше явные представления.

Количество информации $\mathcal{I}(X,Y)$ в X о Y находится из соотношения $\mathcal{I}(X,Y) = \mathcal{H}(Y) - \mathcal{H}(Y|X)$ и для частично определенных данных выразимо в явном виде. Рассмотрим пример. Пусть выход полностью определенного источника X, порождающего символы 0 и 1 с вероятностями p_0 и p_1, подается на вход канала, где символы стираются (заменяются на $*$) с вероятностью ε. Требуется вычислить информацию $\mathcal{I}(Y,X)$ в выходе Y канала о его входе X и информацию $\mathcal{I}(X,Y)$ во входе X о выходе Y. Используя приведенные выше формулы, получаем

$$\mathcal{I}(Y,X) = H(p_0,p_1) - p_0 H(\varepsilon p_1, 1-\varepsilon p_1) - p_1 H(\varepsilon p_0, 1-\varepsilon p_0),$$
$$\mathcal{I}(X,Y) = (1-\varepsilon) H(p_0,p_1),$$

где $H(x_0,x_1) = -x_0 \log x_0 - x_1 \log x_1$.

ЛИТЕРАТУРА

1. *Шоломов Л. А.* Сжатие частично определенной информации // Нелинейная динамика и управление. М.: Физматлит, 2004. Вып. 4. С. 385–399.

2. *Шоломов Л. А.* Правило сложения энтропий для недоопределенных данных // Материалы XVII Межгосударственной школы-семинара «Синтез и сложность управляющих систем». Новосибирск: ИМ СО РАН, 2008. С. 193–196.

УДК 519.7

ON QUATERNARY AND BINARY BENT FUNCTIONS[1]

P. Solé, N. N. Tokareva

In this paper direct links between Boolean bent functions (Rothaus, [1], 1976), generalized Boolean bent functions (Schmidt, [2], 2006) and quaternary bent functions (Kumar, Scholtz, Welch, [3], 1985) are explored. We also study Gray images of bent functions and notions of generalized nonlinearity for Boolean functions.

Let n, q be integers, $q \geqslant 2$. We consider the following mappings:

1) $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ — **Boolean function** in n variables. Its *sign function* is $F := (-1)^f$. The *Walsh Hadamard transform* (WHT) of f is $\widehat{F}(x) := \sum_{y\in\mathbb{Z}_2^n}(-1)^{f(y)+x.y} = \sum_{y\in\mathbb{Z}_2^n} F_y(-1)^{x.y}$. Here $x.y$ is a usual inner product of vectors. A Boolean function f is said to be *bent*, iff $|\widehat{F}(x)| = 2^{n/2}$ for all $x \in \mathbb{Z}_2^n$. It is *near bent* iff $\widehat{F}(x) \in \{0, \pm 2^{(n+1)/2}\}$. Note that Boolean bent (resp. near bent) functions exist only if the number of variables, n, is even (resp. odd).

2) $f : \mathbb{Z}_2^n \to \mathbb{Z}_q$ — **generalized Boolean function** in n variables. Its *sign function* is $F := \omega^f$, with ω a primitive complex root of unity of order q, i. e. $\omega = e^{2\pi i/q}$. When $q = 4$, we write $\omega = i$. Its WHT is given as $\widehat{F}(x) := \sum_{y\in\mathbb{Z}_2^n}\omega^{f(y)}(-1)^{x.y} = \sum_{y\in\mathbb{Z}_2^n} F_y(-1)^{x.y}$. As above, a generalized Boolean function f is *bent*, iff $|\widehat{F}(x)| = 2^{n/2}$ for all $x \in \mathbb{Z}_2^n$. In comparison to the previous case it not follows that n should be even if f is bent. Such functions for $q = 4$ were studied in [2]. Here we consider $q = 4$ only.

[1]The first author was partially supported by ANR grant NUGET. The second author was supported by the Russian Science Support Foundation and by the Russian Foundation for Basic Research (grants 07-01-00248, 08-01-00671, 09-01-00528).