

# Теоретико-информационные границы точности кодирования сообщений и распознавания образов по ансамблям данных

М.М. Ланге<sup>1</sup>, А.М. Ланге<sup>1</sup>

<sup>1</sup> Федеральный исследовательский центр «Информатика и управление» РАН,  
119333, Россия, г. Москва, ул. Вавилова, д. 42

## Аннотация

Исследуются вероятностные модели кодирования дискретных сообщений и распознавания образов (классификации объектов) по ансамблям данных различной модальности. Для рассматриваемых моделей построены аналитические зависимости наименьшей средней взаимной информации между ансамблем данных и множеством возможных решений от допустимой вероятности ошибки в форме монотонно убывающих функций. Приводятся примеры таких функций для схемы кодирования независимых символов конечного алфавита, представленных парами значений с возможными искажениями, и для схемы классификации составных объектов, заданных изображениями лица и подписи. Обращения полученных функций дают нижние границы вероятности ошибки при заданном количестве обрабатываемой информации. Полученные соотношения представляют двухфакторные критерии качества принимаемых решений в задачах кодирования и классификации и являются обобщениями известной в теории информации функции «скорость-погрешность» (rate distortion function).

**Ключевые слова:** кодирование источника, ансамбль данных, энтропия, классификация объектов, вероятность ошибки, взаимная информация, функция «скорость-погрешность».

**Цитирование:** Ланге, М.М. Теоретико-информационные границы точности кодирования сообщений и распознавания образов по ансамблям данных / М.М. Ланге, А.М. Ланге // Компьютерная оптика. – 2024. – Т. 48, № 3. – С. 460–470. – DOI: 10.18287/2412-6179-CO-1362.

**Citation:** Lange MM, Lange AM. Information-theoretic bounds for accuracy of letter encoding and pattern recognition via ensembles of datasets. Computer Optics 2024; 48(3): 460-470. DOI: 10.18287/2412-6179-CO-1362.

## Введение

В задачах кодирования источников дискретных сообщений с допустимой погрешностью и задачах распознавания образов (классификации) важной характеристикой качества принимаемых решений является вероятность ошибки, которая зависит от количества обрабатываемой информации. Известное в теории информации соотношение между количеством информации и вероятностью ошибки определяется функцией «скорость-погрешность» [1], которая дает наименьшее количество информации при заданной вероятности ошибки либо наименьшую вероятность ошибки при заданном количестве информации. Такое соотношение не зависит от алгоритмов кодирования и является двухфакторным критерием качества кодирования дискретных сообщений с погрешностью, измеряемой в метрике Хемминга. В настоящей работе предлагается обобщение указанного соотношения для кодирования дискретных сообщений, переданных по каналу с шумом.

В теории классификации известны методы оптимизации решающих алгоритмов, которые позволяют минимизировать вероятность ошибки [2, 3]. Однако эти методы ориентированы на конкретные алгоритмы и не дают общего подхода для нахождения потенциально достижимой точности при заданном количестве обрабатываемой информации. Поскольку теоретико-информационная модель классификации данных

схожа с моделью кодирования сообщений, переданных по каналу с шумом, естественно найти соотношение между количеством обрабатываемой информации и потенциально достижимой вероятностью ошибки классификации, которое не зависело бы от алгоритмов принятия решения.

Построение соотношений вероятности ошибки и количества обрабатываемой информации для указанных моделей кодирования и классификации базируется на рассмотренной в [4] схеме кодирования непрерывных сообщений с квадратичной погрешностью. В отличие от традиционной схемы кодирования источника с допустимой погрешностью [1, 5], в такой схеме кодирование выполняется на выходе стохастического канала наблюдения, который вносит неустраняемую погрешность.

В работе [6] получены зависимости наименьшего количества обрабатываемой информации от допустимых значений вероятности ошибки для моделей кодирования и классификации, в которых выходы каналов наблюдения представлены данными одной модальности. Цель настоящей работы состоит в построении таких зависимостей для моделей кодирования и классификации с составными каналами наблюдения, выходы которых представлены данными различной модальности. Зависимости наименьшего количества обрабатываемой информации от вероятности ошибки строятся в форме монотонно убывающих функций. Поэтому обращения таких функций дают значения

потенциально достижимой вероятности ошибки при различных значениях количества используемой информации. Такие соотношения могут быть полезны для анализа эффективности алгоритмов кодирования и классификации, поскольку позволяют оценить избыточность вероятности ошибки алгоритмов относительно потенциально возможных значений при любом количестве обрабатываемой информации.

Структура работы включает формализацию моделей кодирования и классификации (параграф 1) и построение для рассматриваемых моделей аналитических нижних границ вероятности ошибки при заданном количестве обрабатываемой информации (параграф 2). В параграфе 3 приводятся примеры таких границ для схемы кодирования статистически независимых сообщений с двумя каналами наблюдения и для схемы классификации изображений лиц и подписей по составным объектам «лицо-подпись».

**1. Вероятностные модели кодирования и классификации**

Модель кодирования дискретных сообщений, переданных по каналу с искажениями

Рассматривается модель кодирования дискретных сообщений с допустимой погрешностью в терминах понятий, принятых в теории информации [1]. Сообщениями источника являются статистически независимые символы алфавита  $U$  размера  $q \geq 2$ , на котором задано распределение вероятностей

$$P_U = \{P_U(u)\}.$$

Символы алфавита  $U$  передаются по составному каналу наблюдения, выход которого образован ансамблем  $V^M = V_1 \dots V_M$   $q$ -ичных алфавитов  $V_m$ ,  $m = 1, \dots, M$ . Каждому входному символу  $u \in U$  составного канала соответствуют столбцы выходных векторов  $v^M = (v_1, \dots, v_M)^T$ ,  $v_m \in V_m$ ,  $m = 1, \dots, M$ , с условными вероятностями

$$P_{V^M|U} = \{P_{V^M|U}(v^M | u)\}.$$

Символы алфавита  $U$  воспроизводятся по сообщениям из ансамбля  $V^M$  безошибочно либо с погрешностью символами алфавита  $\hat{U} = U$ . На алфавите  $\hat{U}$  вводятся свободные условные распределения

$$Q_{\hat{U}|V^M} = \{Q_{\hat{U}|V^M}(\hat{u} | v^M)\},$$

которые образуют канал воспроизведения.

Рассмотрим схему кодирования блоков  $u^N = (u_1, \dots, u_N)$  символов  $u_n \in U$ ,  $n = 1, \dots, N$ , в блоки  $\hat{u}^N = (\hat{u}_1, \dots, \hat{u}_N)$  символов  $\hat{u}_n \in \hat{U}$ ,  $n = 1, \dots, N$ . Каждому  $u^N$  соответствует блок  $v^{MN} = (v_1^M, \dots, v_N^M)$  столбцов  $v_n^M \in V^M$ ,  $n = 1, \dots, N$ , на выходе составного канала наблюдения, и блок  $\hat{u}^N$  строится по предъявляемому блоку  $v^{MN}$ . Указанные блоки образуют множества

$U^N = \{u^N\}$ ,  $V^{MN} = \{v^{MN}\}$  и  $\hat{U}^N = \{\hat{u}^N\}$ . При этом независимость символов источника и использование канала наблюдения и канала воспроизведения без памяти порождает на указанных множествах следующие распределения

$$P_{U^N} = \left\{ P_{U^N}(u^N) = \prod_{n=1}^N P_U(u_n) \right\}, \tag{1}$$

$$P_{V^{MN}|U^N} = \left\{ P_{V^{MN}|U^N}(v^{MN} | u^N) = \prod_{n=1}^N P_{V^M|U}(v_n^M | u_n) \right\}, \tag{2}$$

$$Q_{\hat{U}^N|V^{MN}} = \left\{ Q_{\hat{U}^N|V^{MN}}(\hat{u}^N | v^{MN}) = \prod_{n=1}^N Q_{\hat{U}|V^M}(\hat{u}_n | v_n^M) \right\}. \tag{3}$$

Множества  $U^N, V^{MN}, \hat{U}^N$  с распределениями (1) – (3) образуют пару последовательных стохастических преобразований

$$U^N \xrightarrow{P_{V^{MN}|U^N}} V^{MN} \xrightarrow{Q_{\hat{U}^N|V^{MN}}} \hat{U}^N. \tag{4}$$

Используя побуквенную меру погрешности Хемминга между любой парой блоков  $u^N \in U^N$  и  $\hat{u}^N \in \hat{U}^N$ , для схемы (4) вводится средняя на символ источника погрешность

$$\begin{aligned} E_{Q_{\hat{U}^N|V^{MN}}}(V^{MN}; \hat{U}^N) &= \\ &= (1/N) \sum_{v^{MN}} P_{V^{MN}}(v^{MN}) \sum_{\hat{u}^N} Q_{\hat{U}^N|V^{MN}}(\hat{u}^N | v^{MN}) \times \\ &\times \sum_{u^N} P_{U^N|V^{MN}}(u^N | v^{MN}) \sum_{n=1}^N d(u_n, \hat{u}_n) \end{aligned} \tag{5}$$

и средняя на символ взаимная информация

$$\begin{aligned} I_{Q_{\hat{U}^N|V^{MN}}}(V^{MN}; \hat{U}^N) &= \\ &= (1/N) \sum_{v^{MN}} P_{V^{MN}}(v^{MN}) \sum_{\hat{u}^N} Q_{\hat{U}^N|V^{MN}}(\hat{u}^N | v^{MN}) \times \\ &\times \ln \frac{Q_{\hat{U}^N|V^{MN}}(\hat{u}^N | v^{MN})}{Q_{\hat{U}^N}(\hat{u}^N)}, \end{aligned} \tag{6}$$

где

$$d(u_n, \hat{u}_n) = \begin{cases} 1, & u_n \neq \hat{u}_n \\ 0, & u_n = \hat{u}_n \end{cases},$$

$$P_{U^N|V^{MN}}(u^N | v^{MN}) = \frac{P_{U^N}(u^N) P_{V^{MN}|U^N}(v^{MN} | u^N)}{P_{V^{MN}}(v^{MN})},$$

$$P_{V^{MN}}(v^{MN}) = \sum_{u^N} P_{U^N}(u^N) P_{V^{MN}|U^N}(v^{MN} | u^N),$$

$$Q_{\hat{U}^N}(\hat{u}^N) = \sum_{v^{MN}} P_{V^{MN}}(v^{MN}) Q_{\hat{U}^N|V^{MN}}(\hat{u}^N | v^{MN}).$$

Поскольку в (5) сумма по блокам  $u^N$  в пересчете на символ источника дает вероятность ошибки на блоке воспроизведения  $\hat{u}^N$  по наблюдаемому блоку  $v^{MN}$ , средняя на символ погрешность вида (5) соответствует вероятности ошибки на множестве блоков воспроизведения  $\hat{U}^N$  по множеству наблюдаемых блоков  $V^{MN}$ .

Зависимость функционалов (5) и (6) от свободных условных распределений вида (3) позволяет при заданных значениях  $\varepsilon > 0$  ввести функцию «взаимная информация-вероятность ошибки»

$$R_{MN}(\varepsilon) = \min_{\substack{Q_{\hat{\Omega}^N | \mathbf{P}^{MN}}: \\ E_{Q_{\hat{\Omega}^N | \mathbf{P}^{MN}}}(V^{MN}; \hat{U}^N) \leq \varepsilon}} I_{Q_{\hat{\Omega}^N | \mathbf{P}^{MN}}}(V^{MN}; \hat{U}^N) \quad (7)$$

порядка  $N$ , где минимум берется по всевозможным условным распределениям  $Q_{\hat{\Omega}^N | \mathbf{P}^{MN}}$  при  $\varepsilon$ -ограничении вероятности ошибки. Учитывая, что распределения вида (1)–(3) заданы в форме произведений, функционалы (5) и (6) эквивалентны значениям  $E_{Q_{\hat{\Omega}^M | \mathbf{P}^M}}(V^M; \hat{U})$  и  $I_{Q_{\hat{\Omega}^M | \mathbf{P}^M}}(V^M; \hat{U})$  при  $N=1$ . В этом случае функция (7) не зависит от  $N$  и может быть представлена в форме

$$R_M(\varepsilon) = \min_{\substack{Q_{\hat{\Omega}^M | \mathbf{P}^M}: \\ E_{Q_{\hat{\Omega}^M | \mathbf{P}^M}}(V^M; \hat{U}) \leq \varepsilon}} I_{Q_{\hat{\Omega}^M | \mathbf{P}^M}}(V^M; \hat{U}). \quad (8)$$

Задача состоит в получении для функции (8) монотонно убывающей неотрицательной нижней границы  $\underline{R}_M(\varepsilon) \leq R_M(\varepsilon)$ . Тогда обратная функция  $\underline{R}_M^{-1}(I)$  дает нижнюю границу вероятности ошибки кодирования источника при фиксированных значениях средней взаимной информации  $I$ .

#### Модель классификации составных объектов

Рассматриваемая модель классификации аналогична рассмотренной модели кодирования и формулируется в терминах понятий, принятых в теории классификации [7]. Классы представлены множеством меток  $\Omega = \{\omega_i, i=1, \dots, c\}$   $\Omega = \{\omega_i, i=1, \dots, c\}$ ,  $c \geq 2$ , с априорным распределением вероятностей

$$P_\Omega = \{P_\Omega(\omega_i)\}_{i=1}^c. \quad (9)$$

Классифицируемые составные объекты образуют ансамбль множеств  $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ , в котором  $\mathbf{X}_m$ ,  $m=1, \dots, M$  – множество объектов, порождаемых  $m$ -м источником. Каждый составной объект задан столбцом объектов  $\mathbf{x}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M) \in \mathbf{X}^M$  одного класса, элементы которого  $\mathbf{x}_m \in \mathbf{X}_m$ ,  $m=1, \dots, M$ , являются объектами соответствующих источников. Будем считать, что на ансамбле  $\mathbf{X}^M$  заданы условные по классам распределения

$$P_{\mathbf{X}^M | \Omega} = \{P_{\mathbf{X}^M | \Omega}(\mathbf{x}^M | \omega_i)\}_{i=1}^c, \quad (10)$$

которые образуют составной канал наблюдения. Решения о классах составных объектов образуют множество оценок  $\hat{\Omega} = \{\omega_j, j=1, \dots, c\}$  меток классов.

В общем случае решение принимается по блоку составных объектов одного класса  $\mathbf{x}^{MN} = (\mathbf{x}_1^M, \dots, \mathbf{x}_N^M)$ , в котором  $\mathbf{x}_n^M \in \mathbf{X}^M$ ,  $n=1, \dots, N$ . Блоки составных объектов образуют множество  $\mathbf{X}^{MN} = \{\mathbf{x}^{MN}\}$ . Для составного канала наблюдения без памяти условные по

классам распределения вероятностей блоков составных объектов с учетом (10) имеют вид

$$P_{\mathbf{X}^{MN} | \Omega} = \left\{ P_{\mathbf{X}^{MN} | \Omega}(\mathbf{x}^{MN} | \omega_i) = \prod_{n=1}^N P_{\mathbf{X}_n^M | \Omega}(\mathbf{x}_n^M | \omega_i) \right\}. \quad (11)$$

На множестве оценок классов  $\hat{\Omega}$  вводятся свободные условные по блокам составных объектов распределения вероятностей

$$Q_{\hat{\Omega} | \mathbf{X}^{MN}} = \{Q_{\hat{\Omega} | \mathbf{X}^{MN}}(\omega_j | \mathbf{x}^{MN})\}. \quad (12)$$

Множества  $\Omega$ ,  $\mathbf{X}^{MN}$  и  $\hat{\Omega}$  совместно с распределениями (9), (11) и (12) образуют последовательность стохастических преобразований

$$\Omega \xrightarrow{P_{\mathbf{X}^{MN} | \Omega}} \mathbf{X}^{MN} \xrightarrow{Q_{\hat{\Omega} | \mathbf{X}^{MN}}} \hat{\Omega}. \quad (13)$$

Используя меру погрешности Хемминга между истинными классами и их оценками в схеме (13), вводится средняя вероятность ошибки

$$E_{Q_{\hat{\Omega} | \mathbf{X}^{MN}}}(\mathbf{X}^{MN}; \hat{\Omega}) = \sum_{j=1}^c Q(\omega_j | \mathbf{x}^{MN}) \sum_{i=1}^c P(\omega_i | \mathbf{x}^{MN}) d(\omega_i, \omega_j) \quad (14)$$

и средняя взаимная информация

$$I_{Q_{\hat{\Omega} | \mathbf{X}^{MN}}}(\mathbf{X}^{MN}; \hat{\Omega}) = \sum_{\mathbf{x}^{MN}} P_{\mathbf{X}^{MN}}(\mathbf{x}^{MN}) \sum_{j=1}^c Q_{\hat{\Omega} | \mathbf{X}^{MN}}(\omega_j | \mathbf{x}^{MN}) \times \ln \frac{Q_{\hat{\Omega} | \mathbf{X}^{MN}}(\omega_j | \mathbf{x}^{MN})}{Q_\Omega(\omega_j)}, \quad (15)$$

где

$$d(\omega_i, \omega_j) = \begin{cases} 1, & \omega_i \neq \omega_j \\ 0, & \omega_i = \omega_j \end{cases},$$

$$P_{\hat{\Omega} | \mathbf{X}^{MN}}(\omega_i | \mathbf{x}^{MN}) = \frac{P_\Omega(\omega_i) P_{\mathbf{X}^{MN} | \Omega}(\mathbf{x}^{MN} | \omega_i)}{P_{\mathbf{X}^{MN}}(\mathbf{x}^{MN})},$$

$$P_{\mathbf{X}^{MN}}(\mathbf{x}^{MN}) = \sum_{i=1}^c P_\Omega(\omega_i) P_{\mathbf{X}^{MN} | \Omega}(\mathbf{x}^{MN} | \omega_i),$$

$$Q_{\hat{\Omega}}(\omega_j) = \sum_{\mathbf{x}^{MN} \in \mathbf{X}^{MN}} P_{\mathbf{X}^{MN}}(\mathbf{x}^{MN}) Q_{\hat{\Omega} | \mathbf{X}^{MN}}(\omega_j | \mathbf{x}^{MN}).$$

Зависимость функционалов (14) и (15) от свободных условных распределений (12) позволяет при значениях  $\varepsilon > 0$  ввести соотношение «взаимная информация-вероятность ошибки» в форме условного минимума

$$\tilde{R}_M(\varepsilon) = \min_N \min_{\substack{Q_{\hat{\Omega} | \mathbf{X}^{MN}}: \\ E_{Q_{\hat{\Omega} | \mathbf{X}^{MN}}}(\mathbf{X}^{MN}; \hat{\Omega}) \leq \varepsilon}} I_{Q_{\hat{\Omega} | \mathbf{X}^{MN}}}(\mathbf{X}^{MN}; \hat{\Omega}). \quad (16)$$

Внутренний минимум в (16) берется по всевозможным распределениям  $Q_{\hat{\Omega} | \mathbf{X}^{MN}}$  при заданном  $\varepsilon$ -ограничении средней вероятности ошибки, а внеш-

ний – по длине блоков  $N$ . Для модели классификации, как и для модели кодирования, задача заключается в построении монотонно убывающей неотрицательной нижней границы  $\underline{R}_M(\varepsilon) \leq \bar{R}_M(\varepsilon)$ . Тогда обратная функция  $\tilde{R}_M^{-1}(I)$  дает нижнюю границу вероятности ошибки при заданных значениях средней взаимной информации  $I$ .

## 2. Вычисление функций

### «взаимная информация-вероятность ошибки»

#### Нижняя граница функции $R_M(\varepsilon)$

Построение нижней границы  $\underline{R}_M(\varepsilon) \leq R_M(\varepsilon)$  для функции (9) базируется на дискретной модификации подхода, предложенного в работе [4] для вычисления функции «скорость-погрешность» в схеме кодирования непрерывных сообщений, переданных по каналу с шумом. Согласно такому подходу, вводится последовательность преобразований  $V^M \rightarrow U^* \rightarrow \hat{U}$ , где  $U^* = \hat{U} = U$ .

Преобразование  $V^M \rightarrow U^*$  является детерминированным по критерию максимальной апостериорной вероятности  $u^* = \arg \max_{u \in U} P(u | v^M)$ . Этот критерий обеспечивает наименьшую вероятность ошибки  $\varepsilon_{M \min}^{(q)}$  и вероятность выхода  $P_{U^*}(u^*) = \sum_{v^M \in V^M} P_{V^M}(v^M)$  для подмножества  $V_*^M \subset V^M$  векторов, отображаемых в символ  $u^* \in U^*$ . Преобразование  $U^* \rightarrow \hat{U}$  является стохастическим, в котором переходные вероятности  $Q_{\hat{U}|U^*}(\hat{u} | u^*)$  совпадают с вероятностями  $Q_{\hat{U}|V^M}(\hat{u} | v^M)$  для векторов  $v^M \in V_*^M$  и дают условные распределения

$$Q_{\hat{U}|U^*} = \{Q_{\hat{U}|U^*}(\hat{u} | u^*)\}. \quad (17)$$

С учетом неравенства треугольника  $d(u, \hat{u}) \leq d(u, u^*) + d(u^*, \hat{u})$  и распределений (17) средняя вероятность ошибки (5) при значении  $N=1$  удовлетворяет неравенству

$$\begin{aligned} E_{Q_{\hat{U}|V^M}}(V^M; \hat{U}) &\leq \varepsilon_{M \min}^{(q)} + E_{Q_{\hat{U}|U^*}}(U^*; \hat{U}) = \\ &= \varepsilon_{M \min}^{(q)} + \sum_{u^* \in U^*} P_{U^*}(u^*) \sum_{\hat{u} \in \hat{U}} Q_{\hat{U}|U^*}(\hat{u} | u^*) d(u^*, \hat{u}). \end{aligned} \quad (18)$$

Благодаря детерминированному преобразованию  $V^M \rightarrow U^*$ , средняя взаимная информация (6) в случае  $N=1$  принимает вид

$$\begin{aligned} I_{Q_{\hat{U}|V^M}}(V^M; \hat{U}) &= I_{Q_{\hat{U}|U^*}}(U^*; \hat{U}) = \\ &= \sum_{u^* \in U^*} P_{U^*}(u^*) \sum_{\hat{u} \in \hat{U}} Q_{\hat{U}|U^*}(\hat{u} | u^*) \ln \frac{Q_{\hat{U}|U^*}(\hat{u} | u^*)}{Q_{\hat{U}}(\hat{u})}. \end{aligned} \quad (19)$$

Соотношения (18) и (19) преобразуют функцию (8) к виду

$$R_M(\varepsilon) = \min_{Q_{\hat{U}|U^*}: E_{Q_{\hat{U}|U^*}}(U^*; \hat{U}) \leq \varepsilon - \varepsilon_{M \min}^{(q)}} I_{Q_{\hat{U}|U^*}}(U^*; \hat{U}), \quad (20)$$

где минимум берется по всевозможным условным распределениям (17) при ограничении средней вероятности ошибки преобразования  $U^* \rightarrow \hat{U}$  величиной  $\varepsilon - \varepsilon_{M \min}^{(q)} \geq 0$ . Тогда нижняя граница функции (20) следует из границы Шеннона для функции «скорость-погрешность» в схеме кодирования  $U^* \rightarrow \hat{U}$  с погрешностью в метрике Хемминга [5]. Полученная граница сформулирована в следующей теореме.

**Теорема 1.** Функция  $R_M(\varepsilon)$ , определенная в (7), имеет монотонно убывающую нижнюю границу

$$\begin{aligned} \underline{R}_M(\varepsilon) &= I(U; V^M) - h(\varepsilon - \varepsilon_{M \min}^{(q)}) - \\ &- (\varepsilon - \varepsilon_{M \min}^{(q)}) \ln(q-1) \end{aligned} \quad (21)$$

на отрезке  $\varepsilon_{M \min}^{(q)} \leq \varepsilon \leq \varepsilon_{M \max}^{(q)}$ , где  $I(U; V^M)$  – средняя взаимная информация между  $U$  и  $V^M$ ,  $h(z) = -z \ln z - (1-z) \ln(1-z)$ ,  $\underline{R}_M(\varepsilon_{M \min}^{(q)}) = I(U; V^M)$  и  $\underline{R}_M(\varepsilon_{M \max}^{(q)}) = 0$ .

**Доказательство.** Согласно [5] нижняя граница Шеннона для функции  $R_M(\varepsilon)$ , переопределенной в форме (20), имеет вид

$$\underline{R}_M(\varepsilon) = H(U^*) - H_s(\hat{U} | U^*), \quad (22)$$

где

$$H(U^*) = - \sum_{u^* \in U^*} P_{U^*}(u^*) \ln P_{U^*}(u^*)$$

– безусловная энтропия на множестве решений  $U^*$  и

$$\begin{aligned} H_s(\hat{U} | U^*) &= \\ &= - \sum_{u^* \in U^*} P_{U^*}(u^*) \sum_{\hat{u} \in \hat{U}} Q_{\hat{U}|U^*}^{(s)}(\hat{u} | u^*) \ln Q_{\hat{U}|U^*}^{(s)}(\hat{u} | u^*) \end{aligned}$$

– условная энтропия по распределениям

$$Q_{\hat{U}|U^*}^{(s)} = \left\{ Q_{\hat{U}|U^*}^{(s)}(\hat{u} | u^*) = \frac{\exp(-sd(u^*, \hat{u}))}{\sum_{\hat{u} \in \hat{U}} \exp(-sd(u^*, \hat{u}))} \right\} \quad (23)$$

с параметром  $s > 0$ .

Поскольку формула (8) эквивалентна известной функции «скорость-погрешность», которая ограничивает снизу скорость любого кода с заданной погрешностью [1], энтропия выхода детерминированного преобразования  $V^M \rightarrow U^*$  удовлетворяет неравенству  $H(U^*) \geq R_M(\varepsilon_{M \min}^{(q)})$ . В этом случае  $R_M(\varepsilon_{M \min}^{(q)})$  равно минимальной средней взаимной информации в (8), которая достигается на апостериорных распределениях, обеспечивая  $R_M(\varepsilon_{M \min}^{(q)}) = I(U; V^M)$  и оценку

$$H(U^*) \geq I(U; V^M). \quad (24)$$

Для нахождения параметра  $s$  в распределениях (23) используем равенство

$$\sum_{u^* \in U^*} P_{U^*}(u^*) \sum_{\hat{u} \in \hat{U}} Q_{\hat{U}|U^*}^{(s)}(\hat{u} | u^*) d(u^*, \hat{u}) = \varepsilon - \varepsilon_{M \min}^{(q)},$$

которое дает  $e^{-s} = (\varepsilon - \varepsilon_{M \min}^{(q)}) / (q-1) (1 - (\varepsilon - \varepsilon_{M \min}^{(q)}))$ ,

$$Q_{U|U^*}^{(s)}(\hat{u} | u^*) = \begin{cases} 1 - (\varepsilon - \varepsilon_{M \min}^{(q)}), & \hat{u} = u^* \\ (\varepsilon - \varepsilon_{M \min}^{(q)}) / (q - 1), & \hat{u} \neq u^* \end{cases}$$

и условную энтропию

$$H_s(U | U^*) = h(\varepsilon - \varepsilon_{M \min}^{(q)}) + (\varepsilon - \varepsilon_{M \min}^{(q)}) \ln(q - 1). \quad (25)$$

Последующие замены условной и безусловной энтропии в (22) правыми частями соотношений (24) и (25) завершают доказательство теоремы 1.

Граница (21) является обобщением нижней границы Шеннона для функции «скорость-погрешность» в схеме кодирования независимых дискретных сообщений с допустимой средней погрешностью в метрике Хемминга [1]. В границе Шеннона используется бесшумный канал наблюдения, который обеспечивает  $H(U|V^M) = 0$  и  $I(U; V^M) = H(U)$ . При этом  $\underline{R}_M(\varepsilon_{M \min}^{(q)}) = H(U)$  и  $\varepsilon_{M \min}^{(q)} = 0$ . В общем случае нулевое значение  $\underline{R}_M(\varepsilon_{M \max}^{(q)}) = 0$  достигается в точке  $\varepsilon_{M \max}^{(q)} = (q - 1) \min_{u \in U} P(u)$ .

Минимальная вероятность ошибки  $\varepsilon_{M \min}^{(q)}$  зависит от величины условной энтропии  $H(U|V^M) \geq 0$ . Для нахождения этой зависимости воспользуемся разностью границы Шеннона и границы (21) в точке  $\varepsilon = \varepsilon_{M \max}^{(q)}$ .

Указанная разность дает равенство

$$h(\varepsilon_{M \max}^{(q)}) - h(\varepsilon_{M \max}^{(q)} - \varepsilon_{M \min}^{(q)}) + \varepsilon_{M \min}^{(q)} \ln(q - 1) = H(U | V^M),$$

которое с учетом разложения в ряд Тейлора

$$h(\varepsilon_{M \max}^{(q)}) - h(\varepsilon_{M \max}^{(q)} - \varepsilon_{M \min}^{(q)}) = h'(\varepsilon_{M \max}^{(q)})\varepsilon_{M \min}^{(q)} + \frac{1}{2} h''(\varepsilon_{M \max}^{(q)}) (\varepsilon_{M \min}^{(q)})^2 + O(\varepsilon_{M \min}^{(q)})^3$$

при малых значениях  $\varepsilon_{M \min}^{(q)}$  принимает вид

$$\frac{1}{2} h''(\varepsilon_{M \max}^{(q)}) (\varepsilon_{M \min}^{(q)})^2 + (h'(\varepsilon_{M \max}^{(q)}) + \ln(q - 1)) \varepsilon_{M \min}^{(q)} = H(U | V^M). \quad (26)$$

Решение уравнения (26) с производными

$$h'(\varepsilon_{M \max}^{(q)}) = \ln \frac{1 - \varepsilon_{M \max}^{(q)}}{\varepsilon_{M \max}^{(q)}},$$

$$|h''(\varepsilon_{M \max}^{(q)})| = \left( \varepsilon_{M \max}^{(q)} (1 - \varepsilon_{M \max}^{(q)}) \right)^{-1}$$

дает асимптотическую оценку

$$\varepsilon_{M \min}^{(q)} = \varepsilon_{M \max}^{(q)} (1 - \varepsilon_{M \max}^{(q)}) \times \left( \ln^2 \left( (q - 1) \frac{(1 - \varepsilon_{M \max}^{(q)})}{\varepsilon_{M \max}^{(q)}} \right) + 2 \frac{H(U | V^M)}{\varepsilon_{M \max}^{(q)} (1 - \varepsilon_{M \max}^{(q)})} \right)^{1/2} - \varepsilon_{M \max}^{(q)} (1 - \varepsilon_{M \max}^{(q)}) \times \ln \left( (q - 1) \frac{(1 - \varepsilon_{M \max}^{(q)})}{\varepsilon_{M \max}^{(q)}} \right). \quad (27)$$

В случае равновероятных символов алфавита  $U$  максимальная вероятность ошибки равна  $\varepsilon_{M \max}^{(q)} = (q - 1) / q$  и оценка (27) преобразуется к виду

$$\varepsilon_{M \min}^{(q)} = \frac{q - 1}{q} \left( \frac{2H(U | V^M)}{q - 1} \right)^{1/2}. \quad (28)$$

Оценки (27) и (28) демонстрируют уменьшение минимальной вероятности ошибки  $\varepsilon_{M \min}^{(q)}$  с уменьшением условной энтропии  $H(U|V^M)$  и, следовательно, с ростом средней взаимной информации  $I(U; V^M) = H(U) - H(U|V^M)$ . Этот эффект проявляется с увеличением числа каналов  $M$ . При отсутствии шума в составном канале наблюдения имеем  $H(U|V^M) = 0$  и  $\varepsilon_{M \min}^{(q)} = 0$ . В этом случае  $I(U; V^M) = H(U)$  и граница (21) совпадает с границей Шеннона, приведенной в [1].

Иллюстрация множества  $U$  и ансамбля  $V_1 V_2$ , а также характер функции  $\underline{R}_M(\varepsilon)$  при значениях  $M=1$  и  $M>1$  (сплошные кривые) показаны на рис. 1. Пунктирная кривая демонстрирует поведение нижней границы Шеннона. Множество  $U$  представлено круговой областью, множества  $V_1$  и  $V_2$  – эллиптическими областями. Указанные множества демонстрируют увеличение с ростом  $M$  средней взаимной информации  $I(U; V^M)$  как теоретико-информационной меры пересечения  $U \cap V^M$  (серая область) и уменьшение условной энтропии  $H(U|V^M)$  как теоретико-информационной меры разности  $U \setminus V^M$  (темная область), что приводит к уменьшению вероятности ошибки  $\varepsilon_{M \min}^{(q)}$ .

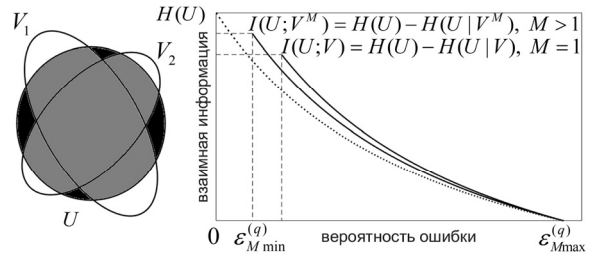


Рис. 1. Интерпретация средней взаимной информации  $I(U; V^M)$  и поведение функции  $\underline{R}_M(\varepsilon)$

Необходимо отметить, что в практике кодирования  $q$ -ичных сообщений со скоростью  $R^{(q)}$  и вероятностью ошибки  $E^{(q)}$  граница  $\underline{R}_M(\varepsilon)$  позволяет вычислить избыточность  $r(R^{(q)} | E^{(q)}) = R^{(q)} - \underline{R}_M(E^{(q)})$  скорости кода относительно потенциально возможного значения при заданной вероятности ошибки. Использование обратной функции от  $\underline{R}_M(\varepsilon)$  дает избыточность

$$r(E^{(q)} | R^{(q)}) = \begin{cases} E^{(q)} - \underline{R}_M^{-1}(R^{(q)}), & R^{(q)} \leq I(U; V^M) \\ E^{(q)} - \varepsilon_{M \min}^{(q)}, & R^{(q)} > I(U; V^M) \end{cases}$$

вероятности ошибки при заданной скорости.

#### Нижняя граница функции $\tilde{R}_M(\varepsilon)$

Методика построения нижней границы  $\tilde{R}_M(\varepsilon) \leq \underline{R}_M(\varepsilon)$  функции (16) аналогична методике,

использованной для получения границы (21). Пусть  $\Omega^* = \{\omega_k, k=1, \dots, c\}$  – множество меток классов, получаемых на детерминированном преобразовании  $\mathbf{X}^{MN^*} \rightarrow \Omega^*$  с наименьшей вероятностью ошибки  $\varepsilon_{M \min}^{(c)}$  при длине блоков  $N^* \geq 1$ . Такое преобразование реализуется на решающем правиле по максимуму апостериорной вероятности классов и порождает разбиение множества  $\mathbf{X}^{MN^*}$  на непересекающиеся подмножества  $\mathbf{X}_k^{MN^*} \rightarrow \omega_k, k=1, \dots, c,$  с вероятностями  $P(\omega_k) = \sum_{\mathbf{x}^{MN^*} \in \mathbf{X}_k^{MN^*}} P(\mathbf{x}^{MN^*}), k=1, \dots, c.$  В этом случае условные по блокам  $\mathbf{x}^{MN^*} \in \mathbf{X}_k^{MN^*}$  вероятности  $Q_{\Omega|\Omega^*}(\omega_j | \mathbf{x}^{MN^*}) = Q_{\Omega|\Omega^*}(\omega_j | \omega_k)$  дают условные распределения

$$Q_{\Omega|\Omega^*} = \{Q_{\Omega|\Omega^*}(\omega_j | \omega_k)\}. \quad (29)$$

С учетом неравенства треугольника  $d(\omega_i, \omega_j) \leq d(\omega_i, \omega_k) + d(\omega_k, \omega_j)$  и распределений (29) средняя вероятность ошибки (14) удовлетворяет неравенству

$$\begin{aligned} E_{Q_{\Omega|\Omega^*}}(\mathbf{X}^{MN^*}, \hat{\Omega}) &\leq \varepsilon_{M \min}^{(c)} + E_{Q_{\Omega|\Omega^*}}(\Omega^*, \hat{\Omega}) = \\ &= \varepsilon_{M \min}^{(c)} + \sum_{k=1}^c P_{\Omega^*}(\omega_k) \sum_{j=1}^c Q_{\Omega|\Omega^*}(\omega_j | \omega_k) d(\omega_k, \omega_j). \end{aligned} \quad (30)$$

При этом средняя взаимная информация (15) принимает вид

$$\begin{aligned} I_{Q_{\Omega|\Omega^*}}(\mathbf{X}^{MN^*}; \hat{\Omega}) &= I_{Q_{\Omega|\Omega^*}}(\Omega^*; \hat{\Omega}) = \\ &= \sum_{k=1}^c P_{\Omega^*}(\omega_k) \sum_{j=1}^c Q_{\Omega|\Omega^*}(\omega_j | \omega_k) \ln \frac{Q_{\Omega|\Omega^*}(\omega_j | \omega_k)}{Q_{\hat{\Omega}}(\omega_j)}. \end{aligned} \quad (31)$$

Соотношения (30) и (31) преобразуют функцию (17) при значении  $N=N^*$  к виду

$$\tilde{R}_M(\varepsilon) = \min_{Q_{\Omega|\Omega^*}: E_{Q_{\Omega|\Omega^*}}(\Omega^*, \hat{\Omega}) \leq \varepsilon - \varepsilon_{M \min}^{(c)}} I_{Q_{\Omega|\Omega^*}}(\Omega^*; \hat{\Omega}), \quad (32)$$

где минимум берется по всевозможным условным распределениям (29) при условии, что второй член в правой части (30) не превосходит величины  $\varepsilon - \varepsilon_{M \min}^{(c)} \geq 0$ . Тогда нижняя граница функции (32) следует из нижней границы Шеннона для функции «скорость-погрешность» в схеме кодирования  $\Omega^* \rightarrow \hat{\Omega}$  с погрешностью в метрике Хемминга, когда средняя погрешность не превосходит  $\varepsilon - \varepsilon_{M \min}^{(c)}$  [5].

**Теорема 2.** Функция  $\tilde{R}_M(\varepsilon)$ , определенная в (16), имеет монотонно убывающую нижнюю границу

$$\tilde{R}_M(\varepsilon) = I(\Omega; \mathbf{X}^M) - h(\varepsilon - \varepsilon_{M \min}^{(c)}) - (\varepsilon - \varepsilon_{M \min}^{(c)}) \ln(c-1) \quad (33)$$

на отрезке  $\varepsilon_{M \min}^{(c)} \leq \varepsilon \leq \varepsilon_{M \max}^{(c)}$ , где  $I(\Omega; \mathbf{X}^M)$  – средняя взаимная информация между множествами  $\Omega$  и  $\mathbf{X}^M$ ,  $h(z) = -z \ln z - (1-z) \ln(1-z)$ ,  $\tilde{R}_M(\varepsilon_{M \min}^{(c)}) = I(\Omega; \mathbf{X}^M)$  и  $\tilde{R}_M(\varepsilon_{M \max}^{(c)}) = 0$ .

**Доказательство.** Аналогично доказательству теоремы 1. Для функции  $\tilde{R}_M(\varepsilon)$  в форме (32) принимаем нижнюю границу [5]

$$\tilde{R}_M(\varepsilon) = H(\Omega^*) - H_s(\hat{\Omega} | \Omega^*), \quad (34)$$

где

$$H(\Omega^*) = -\sum_{k=1}^c P_{\Omega^*}(\omega_k) \ln P_{\Omega^*}(\omega_k)$$

– безусловная энтропия на множестве решений  $\Omega^*$  и

$$\begin{aligned} H_s(\hat{\Omega} | \Omega^*) &= \\ &= -\sum_{k=1}^c P_{\Omega^*}(\omega_k) \sum_{j=1}^c Q_{\Omega|\Omega^*}^{(s)}(\omega_j | \omega_k) \ln Q_{\Omega|\Omega^*}^{(s)}(\omega_j | \omega_k) \end{aligned}$$

– условная энтропия по распределениям

$$\begin{aligned} Q_{\Omega|\Omega^*}^{(s)} &= \left\{ Q_{\Omega|\Omega^*}^{(s)}(\omega_j | \omega_k) = \right. \\ &= \left. \frac{\exp(-sd(\omega_k, \omega_j))}{\sum_{i=1}^c \exp(-sd(\omega_k, \omega_i))} \right\} \end{aligned} \quad (35)$$

с параметром  $s > 0$ .

Согласно теореме кодирования [1], энтропия выхода преобразования  $\mathbf{X}^{MN^*} \rightarrow \Omega^*$  удовлетворяет неравенству  $H(\Omega^*) \geq I(\Omega; \mathbf{X}^{MN^*})$ . Кроме того, при любой конечной длине блоков  $N$  для ансамбля  $\mathbf{X}^{MN} = \mathbf{X}_1^M \dots \mathbf{X}_N^M$  справедливо неравенство  $I(\Omega; \mathbf{X}^{MN}) \geq I(\Omega; \mathbf{X}_n^M)$ , которое в случае тождественных множеств  $\mathbf{X}_n^M = \mathbf{X}^M, n=1, \dots, N$ , выполняется со знаком равенства [1]. Тогда  $I(\Omega; \mathbf{X}^{MN^*}) = I(\Omega; \mathbf{X}^M)$  и

$$H(\Omega^*) \geq I(\Omega; \mathbf{X}^{MN^*}) = I(\Omega; \mathbf{X}^M). \quad (36)$$

Для нахождения параметра  $s$  в распределениях (35) воспользуемся равенством

$$\sum_{k=1}^c P_{\Omega^*}(\omega_k) \sum_{j=1}^c Q_{\Omega|\Omega^*}^{(s)}(\omega_j | \omega_k) d(\omega_k, \omega_j) = \varepsilon - \varepsilon_{M \min}^{(c)},$$

откуда следует  $e^{-s} = (\varepsilon - \varepsilon_{M \min}^{(c)}) / (c-1) (1 - (\varepsilon - \varepsilon_{M \min}^{(c)}))$ ,

$$Q_{\Omega|\Omega^*}^{(s)}(\omega_j | \omega_k) = \begin{cases} 1 - (\varepsilon - \varepsilon_{M \min}^{(c)}), & j = k, \\ (\varepsilon - \varepsilon_{M \min}^{(c)}) / (c-1), & j \neq k, \end{cases}$$

и условная энтропия

$$H_s(\Omega | \Omega^*) = h(\varepsilon - \varepsilon_{M \min}^{(c)}) + (\varepsilon - \varepsilon_{M \min}^{(c)}) \ln(c-1). \quad (37)$$

Последующие замены условной и безусловной энтропии в (34) правыми частями соотношений (36) и (37) завершают доказательство теоремы 2.

В общем случае область определения границы (33) задается минимальной  $\varepsilon_{M \min}^{(c)} \geq 0$  и максимальной  $\varepsilon_{M \max}^{(c)} = (c-1) \min_{i=1}^c P(\omega_i)$  вероятностями ошибки. При этом, как и в схеме кодирования (4), минимальная вероятность ошибки в схеме классификации (13)

должна уменьшаться с ростом средней взаимной информации между входом и выходом канала наблюдения. Для демонстрации этого эффекта полезно привести асимптотические оценки минимальной вероятности ошибки.

Применяя технику, аналогичную используемой для вычисления асимптотики минимальной вероятности ошибки в модели кодирования, для малых значений  $\varepsilon_{M \min}^{(c)}$  получаем асимптотическую оценку

$$\begin{aligned} \varepsilon_{M \min}^{(c)} = & \varepsilon_{M \max}^{(c)} (1 - \varepsilon_{M \max}^{(c)}) \times \left( \ln^2 \left( (c-1) \frac{(1 - \varepsilon_{M \max}^{(c)})}{\varepsilon_{M \max}^{(c)}} \right) + \right. \\ & \left. + 2 \frac{H(\Omega | \mathbf{X}^M)}{\varepsilon_{M \max}^{(c)} (1 - \varepsilon_{M \max}^{(c)})} \right)^{1/2} - \\ & - \varepsilon_{M \max}^{(c)} (1 - \varepsilon_{M \max}^{(c)}) \ln \left( (c-1) \frac{(1 - \varepsilon_{M \max}^{(c)})}{\varepsilon_{M \max}^{(c)}} \right), \end{aligned} \quad (38)$$

которая при равновероятных классах принимает вид

$$\varepsilon_{M \min}^{(c)} = \frac{c-1}{c} \left( \frac{2H(\Omega | \mathbf{X}^M)}{c-1} \right)^{1/2}. \quad (39)$$

Оценки (38) и (39) демонстрируют уменьшение минимальной вероятности ошибки  $\varepsilon_{M \min}^{(c)}$  с уменьшением условной энтропии  $H(\Omega | \mathbf{X}^M)$  и, следовательно, с увеличением средней взаимной информации  $I(\Omega; \mathbf{X}^M) = H(\Omega) - H(\Omega | \mathbf{X}^M)$ . Уменьшение энтропии  $H(\Omega | \mathbf{X}^M)$  и, соответственно, минимальной вероятности ошибки  $\varepsilon_{M \min}^{(c)}$  может быть достигнуто путем увеличения числа  $M$  источников в ансамбле.

При нулевом значении условной энтропии  $H(\Omega | \mathbf{X}^M) = 0$ , которое дает наибольшую среднюю взаимную информацию  $I(\Omega; \mathbf{X}^M) = H(\Omega)$ , граница (33) совпадает с нижней границей Шеннона для функции «скорость-погрешность» в схеме кодирования независимых дискретных сообщений с допустимой средней погрешностью в метрике Хемминга [1]. В этом случае  $\varepsilon_{M \min}^{(c)} = 0$  и  $\tilde{R}_M(\varepsilon_{M \min}^{(c)}) = H(\Omega)$ . Следует отметить, что граница Шеннона соответствует модели классификации, в которой каждый объект априори принадлежит одному классу [8]. Формально это означает, что апостериорные вероятности классов по каждому объекту (простому или составному) принимают значения 1 или 0, обеспечивая  $H(\Omega | \mathbf{X}^M) = 0$ .

Рис. 2 иллюстрирует множество классов  $\Omega$ , ансамбль  $\mathbf{X}_1 \mathbf{X}_2$  и поведение функции  $\tilde{R}_M(\varepsilon)$  при  $M=1$  и  $M>1$  (сплошные кривые) и границы Шеннона (пунктирная кривая). Множество  $\Omega$  представлено круговой областью, множества  $\mathbf{X}_1$  и  $\mathbf{X}_2$  – эллиптическими областями. Из рисунка следует, что с увеличением  $M$  средняя взаимная информация  $I(\Omega; \mathbf{X}^M)$  увеличивается как мера пересечения  $\Omega \cap \mathbf{X}^M$  (серая область), а условная энтропия  $H(\Omega | \mathbf{X}^M)$  уменьшается как мера разности  $\Omega \setminus \mathbf{X}^M$  (темная область), что приводит к уменьшению минимальной вероятности ошибки  $\varepsilon_{M \min}^{(c)}$ .

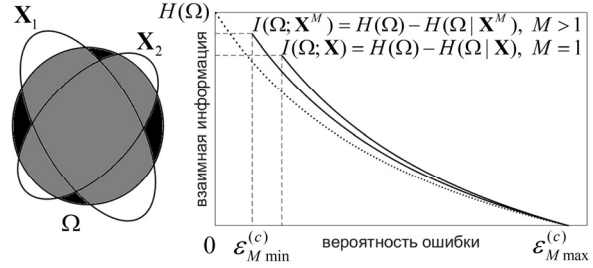


Рис. 2. Интерпретация средней взаимной информации  $I(\Omega; \mathbf{X}^M)$  и поведение функции  $\tilde{R}_M(\varepsilon)$

Практическая значимость границы  $\tilde{R}_M(\varepsilon)$  состоит в возможности использования обратной функции для вычисления избыточности

$$r(E^{(c)} | H^{(c)}) = \begin{cases} E^{(c)} - \tilde{R}_M^{-1}(H^{(c)}), & H^{(c)} \leq I(\Omega; \mathbf{X}^M), \\ E^{(c)} - \varepsilon_{M \min}^{(c)}, & H^{(c)} > I(\Omega; \mathbf{X}^M), \end{cases}$$

вероятности ошибки  $E^{(c)}$  любого решающего алгоритма относительно потенциально возможного значения при заданной энтропии  $H^{(c)}$  решений о классах предъявляемых объектов. Избыточность  $r(E^{(c)} | H^{(c)})$  может быть вычислена для классификаторов с различными разделяющими функциями и, в частности, для многоклассовых SVM классификаторов с разделяющими функциями сигмоидного типа [9]. Кроме того, теоретико-информационный критерий минимизации вероятности ошибки при фиксированных значениях средней взаимной информации между обучающим множеством данных и множеством решений об их классах может найти применение в процедурах оптимизации отбора признаков [10] и, в частности, при обучении нейронных сетей [11].

### 3. Примеры нижних границ

#### Схема кодирования независимых равновероятных сообщений с двумя каналами наблюдения

Рассмотрим вычисление границы  $\tilde{R}_M(\varepsilon)$  вида (21) для схемы, в которой составной канал наблюдения задан  $M>1$  независимыми каналами. Значение  $M=1$  соответствует схеме с одиночным каналом наблюдения. Множества переходных вероятностей  $\mathbf{P}_{V_m|U}$ ,  $m=1, \dots, M$ , таких каналов порождают на ансамбле  $V^M$  множество условных распределений

$$\mathbf{P}_{V^M|U} = \left\{ P_{V^M|U}(\mathbf{v}^M | u) = \prod_{m=1}^M P_{V_m|U}(v_m | u) \right\}. \quad (40)$$

С учетом распределений (40) средняя взаимная информация  $I(U; V^M)$  в (21) согласно [1] выражается через соответствующие энтропии в форме

$$I(U; V^M) = H(V^M) - \sum_{m=1}^M H(V_m | U), \quad (41)$$

где

$$H(V^M) = - \sum_{\mathbf{v}^M} P_{V^M}(\mathbf{v}^M) \ln P_{V^M}(\mathbf{v}^M),$$

$$H(V_m | U) = -\sum_U P_U(u) \sum_{V_m} P_{V_m|U}(v_m | u) \ln P_{V_m|U}(v_m | u),$$

$$P_{V^M}(v^M) = \sum_U P_U(u) P_{V^M|U}(v^M | u).$$

В случае  $M=1$  формула (41) дает значение средней взаимной информации между входом и выходом одиночного канала наблюдения.

В приведенном ниже примере используется алфавит источника  $U$  размера  $q \geq 2$  с одинаковыми вероятностями  $P_U(u)=1/q$  и составной канал наблюдения, который содержит  $M=2$  симметричных канала с выходными алфавитами  $V_m=U$ ,  $m=1,2$ . При вероятности  $0 < \delta_m < 1$  ошибочной передачи символа источника такие каналы имеют условные вероятности выходов

$$P_{V_m|U}(v_m | u) = \begin{cases} 1 - \delta_m, & v_m = u, \\ \delta_m / (q - 1), & v_m \neq u. \end{cases}$$

Вычисление энтропий в (41) при значении  $M=2$  дает среднюю взаимную информацию

$$I(U; V_1 V_2) = \ln q - h(\delta_1) - \delta_1 \ln(q - 1) + h(\delta_1 + \delta_2 - \delta_1 \delta_2 q / (q - 1)) - h(\delta_2) + (\delta_1 - \delta_1 \delta_2 q / (q - 1)) \ln(q - 1). \quad (42)$$

Правая часть в (42) представлена суммой средней взаимной информации

$$I(U; V_1) = \ln q - h(\delta_1) - \delta_1 \ln(q - 1)$$

и условной средней взаимной информации

$$I(U; V_2 | V_1) = h(\delta_1 + \delta_2 - \delta_1 \delta_2 q / (q - 1)) - h(\delta_2) + (\delta_1 - \delta_1 \delta_2 q / (q - 1)) \ln(q - 1).$$

В случае  $\delta_m > 0$ ,  $m=1,2$ , полученные средние взаимные информации строго положительны и, следовательно,  $I(U; V_1 V_2) > I(U; V_1)$ . Поэтому наличие второго канала наблюдения обеспечивает уменьшение условной энтропии в формуле (28) и, следовательно, приводит к уменьшению минимальной вероятности ошибки  $\varepsilon_{\min}^{(q)}$ . В частном случае, когда  $\delta_1 = 0$  или  $\delta_2 = 0$ , имеем  $I(U; V_1 V_2) = H(U) = \ln q$ ,  $\varepsilon_{\min}^{(q)} = 0$ , и нижняя граница (21) совпадает с границей Шеннона. В этом случае  $M=1$  и канал наблюдения задается одним бесшумным каналом с параметром  $\delta = 0$ .

#### Схема классификации составных объектов, заданных изображениями лиц и подписей

Рассматривается вычисление границы  $\tilde{R}_M(\varepsilon)$  вида (33) для схемы, содержащей  $M$  независимых каналов наблюдения с множествами переходных вероятностей  $P_{X_m|\Omega}$ ,  $m=1, \dots, M$ . Такие каналы образуют составной канал с множеством

$$P_{X^M|\Omega} = \left\{ P_{X^M|\Omega}(\mathbf{x}^M | \omega_i) = \prod_{m=1}^M P_{X_m|\Omega}(\mathbf{x}_m | \omega_i) \right\}_{i=1}^c \quad (43)$$

условных по классам вероятностей на ансамбле  $\mathbf{X}^M$ . Вероятности вида (43) совместно с априорными вероятностями классов, заданными в (9), позволяют вычислить среднюю взаимную информацию  $I(\Omega; \mathbf{X}^M)$  в форме, аналогичной (41):

$$I(\Omega; \mathbf{X}^M) = H(\mathbf{X}^M) - \sum_{m=1}^M H(\mathbf{X}_m | \Omega). \quad (44)$$

Численные реализации границы  $\tilde{R}_M(\varepsilon)$  получены для множества  $\Omega$ , содержащего  $c=25$  классов и ансамбля  $\mathbf{X}_1 \mathbf{X}_2$ , заданного множествами изображений лиц  $\mathbf{X}_1$  и подписей  $\mathbf{X}_2$  от 25 персон, по 40 изображений от каждой персоны. Априорные вероятности классов считались одинаковыми. Для описания информативных объектов (лица или подписи) использовались их древовидные представления наборами эллиптических примитивов [12]. Представления получены на основе дихотомического разбиения объектов на непересекающиеся сегменты и аппроксимации сегментов эллиптическими примитивами, которым присвоены средние яркости пикселей соответствующих сегментов. Центры и радиусы примитивов вычислены с использованием центральных и осевых моментов инерции аппроксимируемых сегментов. Структура таких представлений схожа с иерархической структурой, предложенной в [13]. Различие объектов на множестве их древовидных представлений измерялось расстоянием, введенным в [14].

Примеры древовидных представлений лица и подписи, заданных полутоновыми изображениями размера  $256 \times 256$  с 256 уровнями яркости, даны на рис. 3. В приведенных примерах показаны многоуровневые представления, которые содержат по  $2^l$  эллиптических примитивов на уровнях  $l=0, 1, \dots, 8$ . В таких представлениях примитивы соответствуют вершинам  $l$ -го уровня в бинарном дереве. Параметры примитива нулевого уровня используются для нормировки параметров примитивов последующих уровней. Эллиптические примитивы всех уровней строятся в собственных координатных осях представляемого объекта, а их параметры нормируются относительно параметров примитива нулевого уровня. Указанные операции обеспечивают инвариантность таких представлений к сдвигу, повороту, масштабу и яркости объектов. Для кодирования параметров примитивов (координат центров, радиусов и уровней яркости) использован 8-разрядный двоичный код.

На множествах лиц и подписей  $\mathbf{X}_m$ ,  $m=1, 2$ , введены условные по классам вероятности

$$P_{X_m|\Omega}(\mathbf{x}_m | \omega_i) = \frac{\exp(-v_{im} D^2(\mathbf{x}_m, \mathbf{x}_{im}))}{\sum_{\mathbf{x}_m \in \mathbf{X}_m} \exp(-v_{im} D^2(\mathbf{x}_m, \mathbf{x}_{im}))}, i=1, \dots, c, \quad (45)$$

которые соответствуют переходным вероятностям независимых каналов наблюдения в (43). Здесь  $D^2(\mathbf{x}_m, \mathbf{x}_{im})$  – квадрат расстояния между представля-



емым объектом  $\mathbf{x}_m \in \mathbf{X}_m$  и «центральным» представителем»

$$\mathbf{x}_{im} = \arg \min_{\hat{\mathbf{x}}_m \in \mathbf{X}_{im}} D^2(\mathbf{x}_m, \hat{\mathbf{x}}_m)$$

подмножества  $\mathbf{X}_{im} \subset \mathbf{X}_m$   $i$ -го класса, а  $v_{im} > 0$  – свободный параметр. В общем случае расстояния  $D(\mathbf{x}_m, \mathbf{x}_{im})$  вычисляются в выбранном пространстве представлений объектов. Для древовидных представлений множества лиц  $\mathbf{X}_1$  и множества подписей  $\mathbf{X}_2$  использовались расстояния с квадратичным ядром, значения которых заданы матрицами смежности, размещенными на ресурсах [15] и [16].

Условные по классам вероятности вида (45) порождают на ансамбле  $\mathbf{X}_1\mathbf{X}_2$  условные вероятности

$$P_{\mathbf{X}_1\mathbf{X}_2|\Omega}(\mathbf{x}_1, \mathbf{x}_2 | \omega_i) = P_{\mathbf{X}_1|\Omega}(\mathbf{x}_1 | \omega_i)P_{\mathbf{X}_2|\Omega}(\mathbf{x}_2 | \omega_i)$$

и безусловные вероятности

$$P_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) = (1/c) \sum_{i=1}^c P_{\mathbf{X}_1\mathbf{X}_2|\Omega}(\mathbf{x}_1, \mathbf{x}_2 | \omega_i)$$

для любой пары объектов  $\mathbf{x}_1 \in \mathbf{X}_1$  и  $\mathbf{x}_2 \in \mathbf{X}_2$ . Указанные вероятности дают энтропии

$$H(\mathbf{X}_1\mathbf{X}_2) = - \sum_{\mathbf{X}_1} \sum_{\mathbf{X}_2} P_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2) \ln P_{\mathbf{X}_1\mathbf{X}_2}(\mathbf{x}_1, \mathbf{x}_2),$$

$$H(\mathbf{X}_m | \Omega) = -(1/c) \sum_{i=1}^c \sum_{\mathbf{X}_m} P_{\mathbf{X}_m|\Omega}(\mathbf{x}_m | \omega_i) \ln P_{\mathbf{X}_m|\Omega}(\mathbf{x}_m | \omega_i)$$

для вычисления средней взаимной информации (44) при  $M=1$  и  $M=2$ . Параметры  $v_{im} > 0$ ,  $i=1, \dots, c$ , в распределениях (45) на множествах  $\mathbf{X}_m$ ,  $m=1, 2$ , выбраны из условия  $\tilde{R}(\varepsilon_{\max}^{(c)}) = 0$  в точке  $\varepsilon_{\max}^{(c)} = (c-1)/c$ , когда  $M=1$ . В случае  $M \geq 1$  средней взаимной информации  $I(\Omega; \mathbf{X}^M)$  соответствует минимальная вероятность ошибки (39), вычисляемая при значениях условной энтропии  $H(\Omega | \mathbf{X}^M) = \ln c - I(\Omega; \mathbf{X}^M)$ .

Численные реализации границы  $\tilde{R}_M(\varepsilon)$  вида (33) на множествах  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  и на ансамбле  $\mathbf{X}_1\mathbf{X}_2$  представлены сплошными кривыми на рис. 4. Полученные границы демонстрируют уменьшение потенциально возможной вероятности ошибки классификации на ансамбле по сравнению с аналогичными значениями вероятности ошибки для объектов одной модальности. Для сравнения пунктиром дана реализация нижней границы Шеннона.

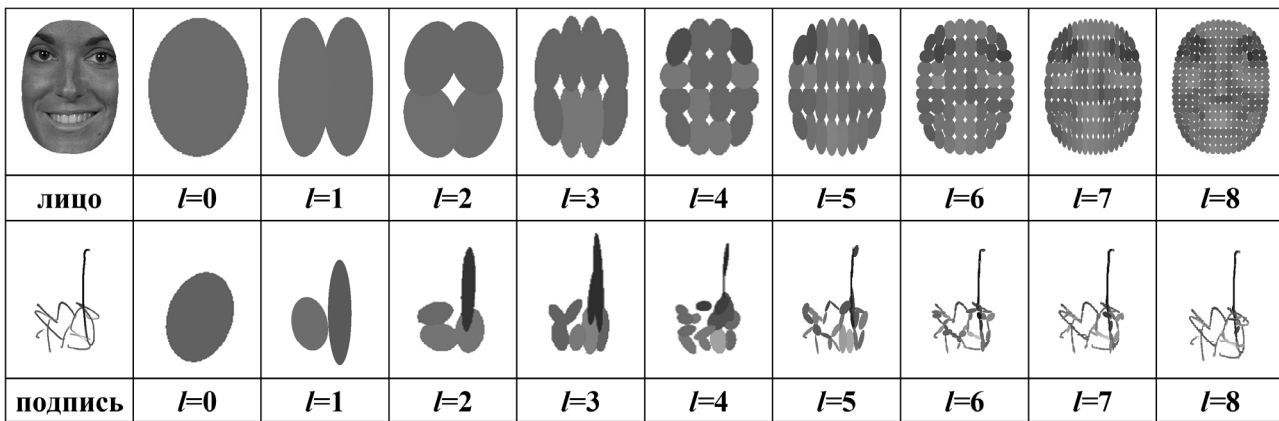


Рис. 3. Примеры представлений лица и подписи бинарными деревьями эллиптических примитивов

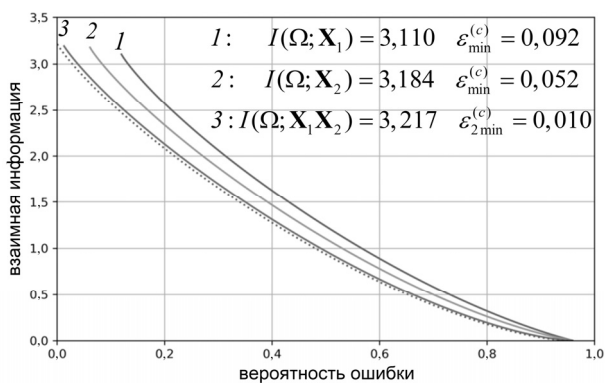


Рис. 4. Реализации границы  $\tilde{R}_M(\varepsilon)$  на множествах лиц (1), подписей (2) и на ансамбле (3)

Следует отметить, что параметры эллиптических примитивов входят в расстояния, которые используются в распределениях вида (45). Поэтому значения средней взаимной информации и минимальной веро-

ятности ошибки в границе (33) зависят от длины кодовых описаний указанных параметров.

Предложенный подход позволяет получить реализации функции  $\tilde{R}_M(\varepsilon)$  на множествах объектов в виде сигналов или изображений, для которых используются различные представления с известными метриками. Допускается объединение в ансамбль размера  $M \geq 2$  множеств объектов с различными представлениями. Для выбранного пространства представлений обратная функция от  $\tilde{R}_M(\varepsilon)$  дает нижнюю границу вероятности ошибки распознавания образов, порождаемых отдельными источниками ( $M=1$ ) или ансамблем источников ( $M>1$ ), при фиксированном количестве анализируемой информации.

### Заключение

Для модели кодирования дискретных сообщений, переданных по составному каналу с шумом, и модели распознавания образов, заданных наборами объектов

различной модальности, получены зависимости наименьшей средней взаимной информации между предъявляемыми данными и возможными решениями от заданной допустимой вероятности ошибки. Найденные зависимости сформулированы в форме монотонно убывающих функций «взаимная информация-вероятность ошибки», которые являются обобщениями известной в теории информации функции «скорость-погрешность» (rate distortion function), когда погрешность измеряется в метрике Хемминга.

Приведенные соотношения получены для схем с составными каналами наблюдения, выходные данные которых образуют ансамбли. Для рассмотренных моделей кодирования и классификации продемонстрировано уменьшение минимальной вероятности ошибки с ростом средней взаимной информации между входом и выходом канала наблюдения при увеличении размера ансамбля предъявляемых данных. Этот результат проиллюстрирован примерами функции «взаимная информация-вероятность ошибки» для схемы кодирования независимых равновероятных сообщений, в которой составной канал наблюдения задан парой симметричных каналов без памяти. Аналогичный эффект продемонстрировали численные реализации функции «взаимная информация-вероятность ошибки» для схемы распознавания изображений лиц и подписей в пространствах их древовидных представлений и на ансамбле таких представлений. Необходимо отметить, что предложенный подход допускает вычисление теоретико-информационных нижних границ вероятности ошибки классификации в пространстве векторных представлений предъявляемых данных, которые применяются в нейронных сетях.

Независимость полученных соотношений от решающих алгоритмов позволяет использовать их для анализа эффективности алгоритмов кодирования и классификации в терминах избыточности вероятности ошибки алгоритма относительно граничных значений при фиксированных значениях скорости кода или энтропии оценок классов. Планируется исследовать избыточность вероятности ошибки для алгоритмов кодирования дискретных источников и алгоритмов распознавания образов с разделяющими функциями различного вида.

### *Благодарности*

Работа выполнена при поддержке Министерства науки и высшего образования в рамках выполнения работ по Государственному заданию Федерального исследовательского центра «Информатика и управление» РАН.

### *References*

- [1] Gallager RG. Information theory and reliable communication. New York: Wiley & Sons; 1968. ISBN: 0471-29048-3.
- [2] Lam L, Suen CY. Application of majority voting to pattern recognition: An analysis of its behavior and performance. IEEE Trans Syst Man Cybern A Syst 1997; 27(5): 553-568. DOI: 10.1109/3468.618255.
- [3] Kuncheva LI, Whitaker CJ, Shipp CA, Duin RPW. Limits on the majority vote accuracy in classifier fusion. Pattern Anal Appl 2003; 6(1): 22-31. DOI: 10.1007/s10044-002-0173-7.
- [4] Dobrushin RL, Tsybakov BS. Information transmission with additional noise. IRE Trans Inf Theory 1962; 8(5): 293-304. DOI: 10.1109/TIT.1962.1057738.
- [5] Berger T. Rate distortion theory: A mathematical basis for data compression. New Jersey: Prentice-Hall Inc, Englewood Cliffs; 1971. ISBN: 013-753103-6.
- [6] Lange MM, Lange AM. Information-theoretic lower bounds to error probability for the models of noisy discrete source coding and object classification. Pattern Recogn Image Anal 2022; 32(3): 570-574. DOI: 10.1134/S105466182203021X.
- [7] Duda RO, Hart PE, Stork DG. Pattern classification. 2nd ed. New York: Wiley & Sons; 2001. ISBN: 978-0471056690.
- [8] Djukova EV, Zhuravlev YuI, Prokofjev PA. Logical correctors in the problem of classification by precedents. Comput Math Math Phys 2017; 57(11): 1866-1886. DOI: 10.1134/S0965542517110057.
- [9] Sueno HT, Gerardo BD, Medina RP. Medina multi-class document classification using Support Vector Machine (SVM) based on improved Naïve Bayes Vectorization technique. Int J Adv Trends Comput Sci Eng 2020; 9(3): 3937-3944. DOI: 10.30534/ijatcse/2020/216932020.
- [10] Brown G, Pocock A, Zhao MJ, Luján M. Conditional likelihood maximization: A unifying framework for information theoretic feature selection. J Mach Learn Res 2012; 13(8): 27-66.
- [11] Xu X, Huang SL, Zheng L, Wornell GW. An information theoretic interpretation to deep neural networks. Entropy 2022; 24(1): 135. DOI: 10.3390/e24010135.
- [12] Lange MM, Ganebnykh SN. On fusion schemes for multiclass classification with reject in a given ensemble of sources. J Phys Conf Ser 2018; 1096: 012048. DOI: 10.1088/1742-6596/1096/1/012048.
- [13] Denisova AY, Sergeev VV. Algorithms for calculating multichannel image histogram using hierarchical data structures. Computer Optics 2016; 40(4): 535-542. DOI: 10.18287/2412-6179-2016-40-4-535-542.
- [14] Lange AM, Lange MM, Paramonov SV. Tradeoff Relation between Mutual Information and Error Probability in Data Classification Problems. Comput Math Math Phys 2021; 61(7): 1181-1193. DOI: 10.1134/S0965542521070113.
- [15] Distance matrices for face dataset. 2020. Source: <<http://sourceforge.net/projects/distance-matrices-face/>>.
- [16] Distance matrices for signature dataset. 2020. Source: <<http://sourceforge.net/projects/distance-matrices-signature/>>.

### *Сведения об авторах*

**Ланге Михаил Михайлович**, 1945 года рождения, в 1968 году окончил Московский энергетический институт по специальности «Измерительная техника». В 1981 году получил степень кандидата технических наук по специальности «Теория информации и техническая кибернетика» в Институте проблем передачи информации РАН. В настоящее время работает ведущим научным сотрудником в Федеральном исследовательском центре «Информатика и управление» РАН.

тика и управление» РАН. Область научных интересов: теория информации, кодирование источников, распознавание образов, обработка изображений и анализ данных. E-mail: [lange\\_mm@mail.ru](mailto:lange_mm@mail.ru)

**Ланге Андрей Михайлович**, 1979 года рождения, в 2002 году окончил Московский государственный технический университет имени Н.Э. Баумана (МГТУ) по специальности «Прикладная математика». В 2008 году получил степень кандидата физико-математических наук по специальности «Математическое моделирование, численные методы и комплексы программ» в МГТУ. В настоящее время работает научным сотрудником в Федеральном исследовательском центре «Информатика и управление» РАН. Область научных интересов: теория ветвящихся процессов, машинное обучение, распознавание образов, анализ данных. E-mail: [lange\\_am@mail.ru](mailto:lange_am@mail.ru)

---

*ГРНТИ: 28.23.15*

*Поступила в редакцию 30 мая 2023 г. Окончательный вариант – 3 ноября 2023 г.*

---

---

# Information-theoretic bounds for accuracy of letter encoding and pattern recognition via ensembles of datasets

M.M. Lange<sup>1</sup>, A.M. Lange<sup>1</sup>

<sup>1</sup> Federal Research Center "Computer Science and Control" RAS,  
119333, Moscow, Russia, Vavilova 42

## Abstract

In this paper, we study stochastic models for discrete letter encoding and object classification via ensembles of different modality datasets. For these models, the minimal values of the average mutual information between a given ensemble of datasets and the corresponding set of possible decisions are constructed as the appropriate monotonic decreasing functions of a given admissible error probability. We present examples of such functions constructed for a scheme of coding independent letters represented by pairs of observation values with possible errors as well as for a scheme of classifying composite objects given by pairs of face and signature images. The inversions of the obtained functions yield the lower bounds for the error probability for any amount of processed information. So, these functions can be considered as the appropriate bifactor fidelity criteria for source coding and object classification decisions. Moreover, the obtained functions are similar to the rate distortion function known in the information theory.

**Keywords:** source coding, ensemble of datasets, entropy, object classification, error probability, mutual information, rate distortion function.

**Citation:** Lange MM, Lange AM. Information-theoretic bounds for accuracy of letter encoding and pattern recognition via ensembles of datasets. *Computer Optics* 2024; 48(3): 460-470. DOI: 10.18287/2412-6179-CO-1362.

**Acknowledgements:** This work was supported by the RF Ministry of Science and Higher Education within the State assignment of the FRC "Computer Science and Control" RAS.

---

## Authors' information

**Mikhail Mikhailovich Lange** (b.1945) graduated from Moscow State Power Institute in 1968 majoring in Instrumentation Technology. He received PhD in Information Theory and Technical Cybernetics from the Institute for Information Transmission Problems of RAS in 1981. Currently, he is the leading researcher in Federal Research Center "Computer Science and Control" of RAS. Scientific interests: information theory, source coding, pattern recognition, image processing and data analysis. E-mail: [lange\\_mm@mail.ru](mailto:lange_mm@mail.ru)

**Andrey Mikhailovich Lange** (b.1979) graduated from Bauman Moscow State Technical University (BMSTU) in 2002 majoring in Applied Mathematics. He received PhD in Mathematical Modeling and Numerical Methods from BMSTU in 2008. Currently, he is the research assistant in Federal Research Center "Computer Science and Control" of RAS. Scientific interests: branching processes, machine learning, pattern recognition and data analysis. E-mail: [lange\\_am@mail.ru](mailto:lange_am@mail.ru)

---

Received May 30, 2023. The final version – November 3, 2023.

---