

## АНАЛИЗ И СИНТЕЗ СИГНАЛОВ И ИЗОБРАЖЕНИЙ

УДК 004.93

СХЕМЫ КОМПЛЕКСИРОВАНИЯ РАЗДЕЛЯЮЩИХ ФУНКЦИЙ  
ДЛЯ ПОВЫШЕНИЯ ТОЧНОСТИ КЛАССИФИКАЦИИ ДАННЫХ  
ОТ АНСАМБЛЯ ИСТОЧНИКОВ

© М. М. Ланге, С. В. Парамонов

*Федеральный исследовательский центр «Информатика и управление» РАН,  
119333, Москва, ул. Вавилова, 42  
E-mail: lange-mm@mail.ru, psvbox@gmail.com*

Исследуется точность классификации данных в терминах соотношения вероятности ошибки и количества обрабатываемой информации для различных схем комплексирования. Рассматриваются схемы комплексирования слабых разделяющих функций на множестве данных одной модальности и на ансамбле данных от источников различной модальности. Для предлагаемых способов комплексирования оценивается избыточность вероятности ошибки относительно теоретико-информационной нижней границы в форме модифицированной функции «скорость—погрешность» с мерой погрешности Хэмминга. Экспериментальные оценки, полученные на множествах изображений лиц и подписей, демонстрируют уменьшение вероятности ошибки и её избыточности относительно нижней границы с увеличением количества обрабатываемой информации за счёт комплексирования слабых разделяющих функций.

*Ключевые слова:* классификация, ансамбль источников, схема комплексирования, вероятность ошибки, взаимная информация, мера погрешности Хэмминга, функция «скорость—погрешность», разделяющая функция, энтропия, избыточность.

DOI: 10.15372/AUT20230401

**Введение.** В большинстве известных работ по классификации данных критерием качества решающих алгоритмов является точность, определяемая вероятностью ошибки. Этот же критерий применяется к алгоритмам, строящимся на основе схем комплексирования [1]. Поскольку вероятность ошибки должна уменьшаться с ростом количества обрабатываемой информации, целесообразно использовать двухфакторный критерий качества принимаемых решений, который задаётся вероятностью ошибки и количеством информации. Применение такого критерия позволяет оценить избыточность вероятности ошибки любого решающего алгоритма относительно теоретико-информационной нижней границы при фиксированном количестве обрабатываемой информации [2]. При этом нижняя граница определяется модифицированной функцией «скорость—погрешность» (rate distortion function) для кодирования дискретных сообщений с допустимой погрешностью по мере Хэмминга [3].

Указанная теоретико-информационная граница задаётся неотрицательной монотонно убывающей функцией наименьшего количества информации от вероятности ошибки. Наибольшее значение такой функции равно средней взаимной информации между множеством классов и множеством классифицируемых данных и соответствует наименьшей вероятности ошибки, а наименьшее значение равно нулю и соответствует наибольшей вероятности ошибки. Обращение такой функции даёт нижнюю границу вероятности ошибки при заданном количестве обрабатываемой информации. Поэтому известные методы отбора признаков, строящиеся по критерию максимизации средней взаимной информации в множестве

данных относительно множества классов [4, 5], обеспечивают минимизацию потенциально возможной вероятности ошибки. Схожий теоретико-информационный критерий минимизации вариации информации рассмотрен в работе [6] для повышения точности алгоритмов сегментации изображений.

Исследование качества решающих алгоритмов в терминах двухфакторного критерия выполняется с использованием разделяющих функций [7], которые позволяют получить оценки вероятности ошибки и количества обрабатываемой информации в форме энтропии множества решений о классах.

В представленной работе на множестве объектов одной модальности рассматриваются слабые наборы разделяющих функций, обеспечивающие достаточно низкую точность, и способ комплексирования слабых наборов, который позволяет уменьшить вероятность ошибки за счёт увеличения количества обрабатываемой информации. Демонстрируется возможность существенного повышения точности классификации на ансамбле источников путём комплексирования разделяющих функций для объектов различной модальности. Предлагаемый подход позволяет оценить избыточность вероятности ошибки классификации относительно наименьшего значения при заданном количестве информации. Такая оценка может быть построена при любом способе комплексирования данных и, в частности, при комплексировании изображений от оптических источников дистанционного зондирования Земли (ДЗЗ) с различными пространственным и спектральным разрешениями [8].

**Формализация исследуемой задачи.** Рассматривается схема классификации

$$\Omega \rightarrow \mathbf{X}^M \rightarrow \hat{\Omega}, \quad (1)$$

в которой  $\Omega = \{\omega_i\}_{i=1}^c$  и  $\hat{\Omega} = \{\omega_j\}_{j=1}^c$  — множества классов и их оценок, где  $c \geq 2$ , а  $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$  — ансамбль множеств различной модальности, в котором каждое множество  $\mathbf{X}_m$ ,  $m = 1, \dots, M$ , содержит объекты одной модальности. В схеме (1) любой составной объект  $\mathbf{x}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M)^t \in \mathbf{X}^M$  представлен столбцом объектов одного класса, взятых по одному  $\mathbf{x}_m \in \mathbf{X}_m$ ,  $m = 1, \dots, M$ , от каждого источника.

Пусть на множествах объектов источников заданы наборы разделяющих функций

$$G(\mathbf{X}_m) = \{g_j(\mathbf{x}_m), \quad \mathbf{x}_m \in \mathbf{X}_m\}_{j=1}^c, \quad m = 1, \dots, M, \quad (2)$$

которые определяют «меру правдоподобия» оценок классов по объектам источников. Комплексирование функций (2) позволяет сформировать набор разделяющих функций на ансамбле источников

$$G(\mathbf{X}^M) = \{g_j(\mathbf{x}^M), \quad \mathbf{x}^M \in \mathbf{X}^M\}_{j=1}^c. \quad (3)$$

Нормирование функций (3) даёт оценки апостериорных вероятностей классов

$$\left\{ Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M) = \frac{g_j(\mathbf{x}^M)}{\sum_{i=1}^c g_i(\mathbf{x}^M)} \right\}_{j=1}^c, \quad (4)$$

которые при  $M = 1$  соответствуют оценкам апостериорных вероятностей по объектам источников  $\mathbf{x}_m \in \mathbf{X}_m$ ,  $m = 1, \dots, M$ .

Заданные априорные вероятности классов  $\{P_{\Omega}(\omega_i)\}_{i=1}^c$  и условные по классам вероятности  $\{P_{\mathbf{X}^M|\Omega}(\mathbf{x}^M | \omega_i)\}_{i=1}^c$  дают вероятности составных объектов

$$P_{\mathbf{X}^M}(\mathbf{x}^M) = \sum_{i=1}^c P_{\Omega}(\omega_i) P_{\mathbf{X}^M|\Omega}(\mathbf{x}^M | \omega_i), \quad \mathbf{x}^M \in \mathbf{X}^M. \quad (5)$$

Тогда для решающего алгоритма по критерию максимума оценок апостериорных вероятностей классов распределения (4) и (5) позволяют вычислить вероятность ошибки

$$E_G(\mathbf{X}^M; \hat{\Omega}) = 1 - \sum_{\mathbf{x}^M \in \mathbf{X}^M} P_{\mathbf{X}^M}(\mathbf{x}^M) \max_{j=1}^c Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M) \quad (6)$$

и энтропию

$$H_G(\hat{\Omega}) = - \sum_{j=1}^c P_{\hat{\Omega}}(\omega_j) \ln P_{\hat{\Omega}}(\omega_j) \quad (7)$$

множества решений  $\hat{\Omega}$ , где  $P_{\hat{\Omega}}(\omega_j)$  — вероятность подмножества  $\mathbf{X}_j^M \subset \mathbf{X}^M$  составных объектов, по которым принимается решение  $\omega_j \in \hat{\Omega}$ .

Нижняя граница минимума средней взаимной информации  $I(\mathbf{X}^M; \hat{\Omega})$  при средней вероятности ошибки  $E(\mathbf{X}^M; \Omega) \leq \varepsilon$  имеет вид монотонно убывающей функции [2]

$$R_M(\varepsilon) = I(\Omega; \mathbf{X}^M) - h(\varepsilon - \varepsilon_{M \min}) - (\varepsilon - \varepsilon_{M \min}) \ln(c - 1) \quad (8)$$

на отрезке  $\varepsilon_{M \min} \leq \varepsilon \leq \varepsilon_{M \max}$ , где  $I(\Omega; \mathbf{X}^M)$  — средняя взаимная информация между  $\Omega$  и  $\mathbf{X}^M$ ,  $h(z) = -z \ln z - (1-z) \ln(1-z)$ ,  $R_M(\varepsilon_{M \min}) = I(\mathbf{X}^M; \Omega)$  и  $R_M(\varepsilon_{M \max}) = 0$ . В случае равновероятных классов, когда  $P_{\Omega}(\omega_i) = 1/c$ ,  $i = 1, \dots, c$ , максимальное и минимальное значения вероятности ошибки соответственно равны  $\varepsilon_{M \max} = (c - 1)/c$  и

$$\varepsilon_{M \min} = \frac{c - 1}{c} \sqrt{2 \frac{H(\Omega) - I(\Omega; \mathbf{X}^M)}{c - 1}}, \quad (9)$$

где  $H(\Omega) = \ln c$  — энтропия множества классов.

Для заданного ансамбля данных  $\mathbf{X}^M$ , на котором выполняется обучение и тестирование исследуемых классификаторов, обращение функции (8) даёт наименьшую вероятность ошибки при фиксированном количестве обрабатываемой информации. Соотношение (9) демонстрирует уменьшение величины  $\varepsilon_{M \min}$  с ростом средней взаимной информации  $I(\Omega; \mathbf{X}^M) = H(\Omega) - H(\Omega | \mathbf{X}^M) \leq H(\Omega)$  [3]. При любом распределении классов равенство  $I(\Omega; \mathbf{X}^M) = H(\Omega)$  обеспечивает  $\varepsilon_{M \min} = 0$ . В этом случае функция (8) совпадает с соотношением «скорость—погрешность» в схеме кодирования независимых дискретных сообщений со средней погрешностью, измеряемой в метрике Хэмминга [3].

Необходимо отметить, что при отсутствии полной корреляции множеств, образующих ансамбль  $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ , средняя взаимная информация  $I(\mathbf{X}^M; \Omega)$  увеличивается с ростом размера  $M$  ансамбля [2]. Это утверждение следует из соотношения

$$I(\Omega; \mathbf{X}^M) = I(\Omega; \mathbf{X}_1) + \sum_{m=2}^M I(\Omega; \mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1),$$

в котором условные средние взаимные информации  $I(\Omega; \mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) \geq 0$  неотрицательны и равны нулю при полной корреляции множеств  $\mathbf{X}_1, \dots, \mathbf{X}_M$ . Неполная корреляция указанных множеств обеспечивает положительность условных взаимных информаций и, следовательно, увеличение  $I(\Omega; \mathbf{X}^M)$  с ростом  $M$ . Отсутствие корреляции даёт  $I(\Omega; \mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) = I(\Omega; \mathbf{X}_m)$ .

Для любой пары характеристик, заданных соотношениями (6) и (7), функция (8) позволяет определить избыточность вероятности ошибки  $E_G$  относительно потенциально возможного значения при количестве информации  $H_G$ . Указанная избыточность определяется величиной

$$r_G = \begin{cases} E_G - R_M^{-1}(H_G), & H_G \leq I(\mathbf{X}^M; \Omega); \\ E_G - \varepsilon_{M \min}^{(c)}, & H_G > I(\mathbf{X}^M; \Omega), \end{cases} \quad (10)$$

где  $R_M^{-1}(H_G)$  — значение обращения функции (8), когда  $R_M(\varepsilon) = H_G$ .

Пусть для всех источников заданы слабые наборы (2) разделяющих функций  $\{G^{(k)}(\mathbf{X}_m)\}_{k=1}^K$ ,  $m = 1, \dots, M$ . Задача состоит в нахождении способа комплексирования слабых разделяющих функций в наборы  $G(\mathbf{X}_m)$ ,  $m = 1, \dots, M$ , которые обеспечивают вероятность ошибки  $E_G(\mathbf{X}_m; \hat{\Omega}) < \min_{k=1}^K E_{G^{(k)}}(\mathbf{X}_m; \hat{\Omega})$  и избыточность  $r_G < \min_{k=1}^K r_{G^{(k)}}$ , определённые в (6) и (10) при  $M = 1$ . Необходимо также предложить способ комплексирования разделяющих функций (2) в функции (3), который приводит к дополнительному снижению вероятности ошибки и избыточности.

**Основные результаты.** Для множества  $\mathbf{X}_m$  объектов  $m$ -го источника с разделяющими функциями, принимающими значения на отрезке  $[0, 1]$ , предложен способ комплексирования функций (2) по схеме

$$g_j(\mathbf{x}_m) = \begin{cases} \max_{k=1}^K g_j^{(k)}(\mathbf{x}_m), & \exists g_j^{(k)}(\mathbf{x}_m) \geq \Delta; \\ \min_{k=1}^K g_j^{(k)}(\mathbf{x}_m), & \forall g_j^{(k)}(\mathbf{x}_m) < \Delta, \end{cases} \quad (11)$$

где  $\Delta$  — порог, выбираемый из условия оптимизации. Разделяющие функции (3) на ансамбле источников  $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$  заданы произведениями

$$g_j(\mathbf{x}^M) = \prod_{m=1}^M g_j(\mathbf{x}_m). \quad (12)$$

Экспериментальные оценки характеристик качества схем комплексирования (11) и (12) получены на множествах изображений лиц [9] и подписей [10] в пространстве их древовидных представлений с заданным расстоянием в квадратичной метрике. Структура указанных представлений подробно описана в работе [2] и аналогична иерархической структуре данных, используемой в [11] для вычисления гистограмм многоканальных изображений. Источники содержали по 1000 объектов от  $c = 25$  персон, по 40 объектов в равновероятных классах. Примеры древовидных представлений лица и подписи даны на рис. 1.

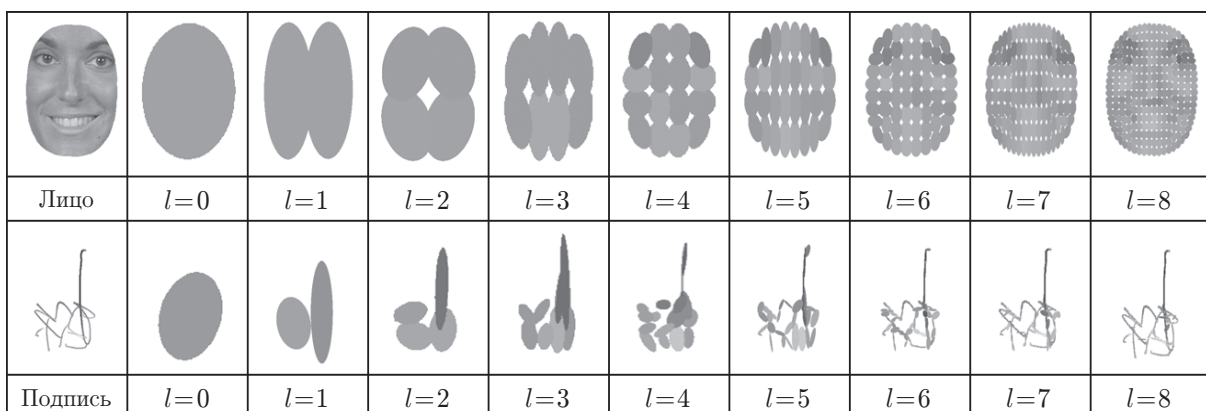


Рис. 1. Многоуровневые представления лица и подписи эллиптическими примитивами

Представления имеют информативные уровни разрешения  $l = 1, \dots, 8$ , которые содержат по  $2^l$  эллиптических примитивов. Параметры примитива нулевого уровня используются для нормирования параметров примитивов последующих уровней. Нормированные примитивы всех уровней задаются в собственных координатных осях примитива нулевого уровня и имеют номера соответствующих им вершин бинарного дерева. Построение примитивов в собственных осях нулевого уровня и нормировка параметров примитивов обеспечивают инвариантность представлений к сдвигу, повороту, масштабу и уровню яркости объектов. Аналогичные представления могут быть построены для любых объектов, заданных изображениями и имеющими идентифицируемые собственные координаты.

На множествах древовидных представлений объектов с заданным расстоянием в квадратичной метрике [2] введены условные по классам вероятности объектов, которые совместно с априорными вероятностями классов позволяют вычислить значения средней взаимной информации  $I(\Omega; \mathbf{X}_1)$ ,  $I(\Omega; \mathbf{X}_2)$  и  $I(\Omega; \mathbf{X}_1\mathbf{X}_2)$  на множестве лиц  $\mathbf{X}_1$ , множестве подписей  $\mathbf{X}_2$  и на ансамбле  $\mathbf{X}_1\mathbf{X}_2$ . Используя указанные значения взаимной информации и методику, изложенную в работе [12], вычислены теоретико-информационные границы вида (8) на множествах  $\mathbf{X}_1$ ,  $\mathbf{X}_2$  и на ансамбле  $\mathbf{X}_1\mathbf{X}_2$ .

Графики численных реализаций указанных границ даны в виде сплошных кривых на рис. 2. Пунктирная кривая соответствует функции «скорость—погрешность», когда  $I(\Omega; \mathbf{X}_1\mathbf{X}_2) = H(\Omega)$ . Для рассматриваемых данных показаны значения наибольшей средней взаимной информации и наименьшей вероятности ошибки.

В экспериментах с решающим алгоритмом по критерию максимума оценок апостериорных вероятностей классов использовались экспоненциальные разделяющие функции

$$\left\{ g_j^{(k)}(\mathbf{x}_m) = \exp \left( -v_{jm}^{(k)} d^2(\mathbf{x}_m, \mathbf{x}_{jm}^{(k)}) \right) \right\}_{j=1}^c, \quad k = 1, 2, \quad (13)$$

где  $v_{jm}^{(k)} > 0$  — свободный параметр,  $\mathbf{x}_{jm}^{(k)}$  — представитель класса,  $d(\mathbf{x}_m, \mathbf{x}_{jm}^{(k)}) \geq 0$  — расстояние между  $\mathbf{x}_m$  и  $\mathbf{x}_{jm}^{(k)}$ . Номера  $m = 1$  и  $m = 2$  соответствуют источникам лиц и

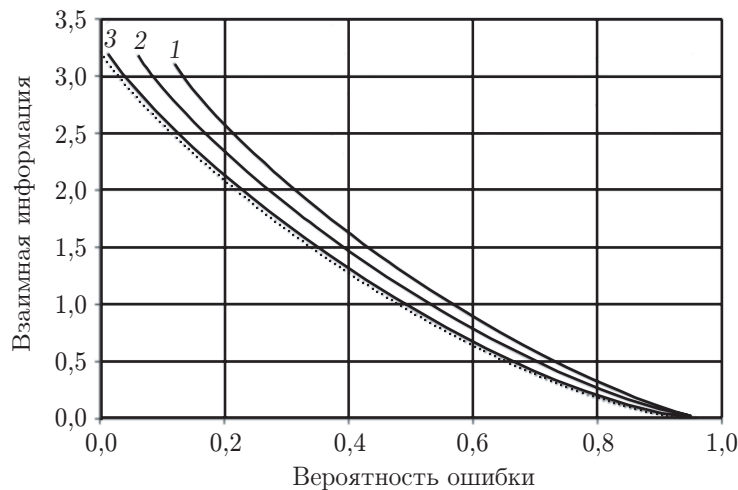


Рис. 2. Реализации зависимости наименьшей средней взаимной информации от вероятности ошибки на множестве лиц (1), множестве подписей (2) и на ансамбле (3): 1 —  $I(\Omega; \mathbf{X}_1) = 3,110$ ,  $\varepsilon_{\min} = 0,092$ ; 2 —  $I(\Omega; \mathbf{X}_2) = 3,184$ ,  $\varepsilon_{\min} = 0,052$ ; 3 —  $I(\Omega; \mathbf{X}_1\mathbf{X}_2) = 3,217$ ,  $\varepsilon_{2\min} = 0,010$

Таблица 1

Характеристики качества решающего алгоритма с различными способами комплексирования разделяющих функций			
Характеристики	$G^{(1)}$	$G^{(2)}$	$G$
Лица $\mathbf{X}_1$			
$H_G$	3,159	3,139	3,162
$E_G$	0,168	0,188	0,155
$r_G$	0,076	0,096	0,063
Подписи $\mathbf{X}_2$			
$H_G$	3,217	3,216	3,219
$E_G$	0,139	0,185	0,106
$r_G$	0,087	0,133	0,054
Ансамбль $\mathbf{X}_1\mathbf{X}_2$			
$H_G$	3,215	3,211	3,217
$E_G$	0,021	0,030	0,018
$r_G$	0,010	0,018	0,008

подписей. В качестве представителей  $\mathbf{x}_{jm}^{(k)}$  выбраны ближайшие ( $k = 1$ ) и вторые ближайшие ( $k = 2$ ) объекты к представителям классов, относительно которых сумма квадратов внутриклассовых расстояний минимальна. Параметры  $v_{jm}^{(k)}$  вычислялись с использованием средних значений и дисперсий внутриклассовых расстояний [2, 12].

В табл. 1 даны численные оценки характеристик качества классификации, полученные в режиме скользящего контроля на множествах изображений лиц и подписей, и на ансамбле этих множеств при пороге  $\Delta = 0,05$ . Оценки для отдельных источников демонстрируют уменьшение вероятности ошибки  $E_G$  и избыточности  $r_G$  с ростом энтропии  $H_G$  за счёт комплексирования наборов  $G^{(1)}$  и  $G^{(2)}$  и дополнительное уменьшение значений  $E_G$  и  $r_G$  с увеличением энтропии  $H_G$  на ансамбле источников. Ожидается усиление этого эффекта при увеличении числа наборов разделяющих функций на множествах данных одной модальности и с ростом числа источников данных различной модальности, которые образуют ансамбль.

Для используемых наборов разделяющих функций (2) и (3) полезно привести значения средней взаимной информации между множествами классифицируемых объектов и множеством оценок классов с апостериорными вероятностями (4), а также соответствующие средние значения вероятности ошибки. Для ансамбля  $\mathbf{X}^M$  и множества оценок классов  $\hat{\Omega}$  указанные усреднённые характеристики определяются функционалами

$$I_G(\mathbf{X}^M; \hat{\Omega}) = \sum_{\mathbf{x}^M \in \mathbf{X}^M} P_{\mathbf{X}^M}(\mathbf{x}^M) \sum_{j=1}^c Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M) \ln \frac{Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M)}{Q_{\hat{\Omega}}(\omega_j)}, \quad (14)$$

$$E_G^*(\mathbf{X}^M; \hat{\Omega}) = \sum_{\mathbf{x}^M \in \mathbf{X}^M} P_{\mathbf{X}^M}(\mathbf{x}^M) \sum_{j=1}^c Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M) (1 - Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M)), \quad (15)$$

где  $Q_{\hat{\Omega}}(\omega_j) = \sum_{\mathbf{x}^M \in \mathbf{X}^M} P_{\mathbf{X}^M}(\mathbf{x}^M) Q_{\hat{\Omega}|\mathbf{X}^M}(\omega_j | \mathbf{x}^M)$ . При значении  $M = 1$  функционалы (14) и (15) дают усреднённые характеристики на множествах  $\mathbf{X}_1$  и  $\mathbf{X}_2$ . Учитывая,

Таблица 2

**Средние характеристики качества разделяющих функций  
при различных способах комплексирования**

Характеристики	$G^{(1)}$	$G^{(2)}$	$G$
<b>Лица <math>X_1</math></b>			
$I_G$	2,614	2,556	2,667
$E_G^*$	0,246	0,271	0,224
$r_G^*$	0,081	0,097	0,069
<b>Подписи <math>X_2</math></b>			
$I_G$	2,565	2,414	2,716
$E_G^*$	0,236	0,299	0,179
$r_G^*$	0,089	0,124	0,059
<b>Ансамбль <math>X_1 X_2</math></b>			
$I_G$	3,146	3,116	3,151
$E_G^*$	0,031	0,043	0,028
$r_G^*$	0,013	0,021	0,011

что  $I_G(\mathbf{X}^M; \hat{\Omega}) \leq I(\mathbf{X}^M; \Omega)$ , избыточность средней вероятности ошибки определяется величиной

$$r_G^* = E_G^* - R_M^{-1}(I_G). \quad (16)$$

Табл. 2 содержит численные оценки характеристик (14), (15) и (16), полученные в режиме скользящего контроля для наборов разделяющих функций, которые использованы при вычислении показателей, приведённых в табл. 1. Соотношения  $E_G^* > E_G$  и  $I_G < H_G$  обеспечивают смещение характеристик  $E_G^*$ ,  $I_G$  относительно  $E_G$ ,  $H_G$  влево-вниз и соответственно некоторое увеличение избыточности  $r_G^* > r_G$ .

Согласно данным табл. 1 и 2 понижение значений вероятности ошибки  $E_G$  и  $E_G^*$  и соответствующих значений избыточности  $r_G$  и  $r_G^*$  может быть достигнуто за счёт применения дополнительных разделяющих функций на множествах объектов источников, а также за счёт расширения ансамбля источников биометрических данных. В частности, представляет интерес комплексирование экспоненциальных разделяющих функций с функциями сигмоидного типа, которые используются в SVM-алгоритмах, обучаемых на основе метода опорных векторов [13].

**Заключение.** Предложены способы комплексирования разделяющих функций на множестве объектов одной модальности и на ансамбле множеств различной модальности для повышения точности классификации данных. Эффективность подхода исследована в терминах вероятности ошибки при фиксированных значениях количества обрабатываемой информации. Продемонстрировано увеличение количества обрабатываемой информации за счёт применения предложенных способов комплексирования и, как следствие, уменьшение вероятности ошибки. Оценена избыточность вероятности ошибки относительно нижней границы для биометрических данных, заданных изображениями лиц и подписей, и для ансамбля этих источников. Экспериментально показано уменьшение вероятности ошибки принимаемых решений и её избыточности относительно нижней границы в связи с увеличением энтропии принимаемых решений, а также с ростом средней взаимной информации между классифицируемыми данными и оценками их классов. Предложенный подход позволяет повысить точность классификации за счёт увеличения числа наборов разделяющих функций по отдельным источникам и числа источников в ансамбле.

## СПИСОК ЛИТЕРАТУРЫ

1. **Kuncheva L. I., Whitaker C. J., Shipp C. A., Duin R. P. W.** Limits on the majority vote accuracy in classifier fusion // Pattern Analysis and Applications. 2003. **6**. P. 22–31. DOI: 10.1007/s10044-002-0173-7.
2. **Ланге А. М., Ланге М. М., Парамонов С. В.** О соотношении взаимной информации и вероятности ошибки в задаче классификации данных // Журнал вычислительной математики и математической физики. 2021. **61**, № 7. С. 119–1205. DOI: 10.31857/S0044466921070115.
3. **Gallager R. G.** Information Theory and Reliable Communication. New York: Wiley, 1968. 588 p.
4. **Brown G., Pockock A., Zhao M. J., Luj'an M.** Conditional likelihood maximization: A unifying framework for information theoretic feature selection // Journ. Machine Learning Research. 2012. N 13. P. 2–66.
5. **Fleuret F.** Fast binary feature selection with conditional mutual information // Journ. Machine Learning Research. 2004. N 5. P. 1531–1555.
6. **Мурашов Д. М.** Метод комбинирования карт сегментации изображений на основе минимизации информационной избыточности и вариации информации // Автометрия. 2022. **58**, № 5. С. 34–42. DOI: 10.15372/AUT20220504.
7. **Duda R. O., Hart P. E., Stork D. G.** Pattern Classification: 2nd ed. New York: Wiley, 2001. 738 p.
8. **Белов А. М., Денисова А. Ю.** Классификация изображений ДЗЗ с использованием алгоритма комплексования данных различных сенсоров // Компьютерная оптика. 2020. **44**, № 4. С. 627–635. DOI: 10.18287/2412-6179-CO-735.
9. **Distance matrices for face dataset.** 2020. URL: <http://sourceforge.net/projects/distance-matrices-face> (дата обращения: 30.05.2023).
10. **Distance matrices for signature dataset.** 2020. URL: <http://sourceforge.net/projects/distance-matrices-signature> (дата обращения: 30.05.2023).
11. **Denisova A. Yu., Sergeev V. V.** Algorithms for calculating multichannel image histograms using hierarchical data structures // Comput. Opt. 2016. **40**, N 4. P. 535–542. DOI: 10.18287/2412-6179-2016-40-4-535-542.
12. **Lange M. M., Paramonov S. V.** On a lower bound to classification error probability in an ensemble of data sources // Proc. of the VII Int. Conf. on Information Technology and Nanotechnology (ITNT 2021). Samara, Russia, 20-24 Sept., 2021. P. 1–6. DOI: 10.31857/S0044466921070115.
13. **Sueno H. T., Gerardo B. D., Medina R. P.** Multi-class document classification using support vector machine (SVM) based on improved naïve Bayes vectorization technique // Int. Journ. Advanced Trends in Computer Science and Engineering. 2020. **9**, N 3. P. 3937–3944. DOI: 10.30534/ijatcse/2020/216932020.

*Поступила в редакцию 30.05.2023*

*После доработки 22.06.2023*

*Принята к публикации 22.06.2023*