

О теоретико-информационной модели классификации данных*

М. М. Ланге, А. М. Ланге

lange_mm@ccas.ru; lange_am@mail.ru

Федеральный исследовательский центр "Информатика и управление" РАН, ул. Вавилова, 40, Москва, 119333, Россия

Исследуется модель классификации в терминах зависимости средней взаимной информации между множеством классифицируемых объектов и множеством решений от допустимой вероятности ошибки. Оптимизация модели состоит в минимизации средней взаимной информации по распределениям вероятностей решений при заданном ограничении на вероятность ошибки. Задача оптимизации сводится к вычислению функции скорость-погрешность в схеме кодирования номера класса по объекту, переданному по непрерывному каналу наблюдения с заданными условными по классам плотностями распределения. Построены нижние границы функций скорость-погрешность в схемах классификации без отказа и с отказом, которые позволяют сравнить значения потенциально достижимой точности при использовании различных данных и каналов наблюдения. Теоретические результаты подтверждаются экспериментальными оценками вероятности ошибки распознавания лиц по декоррелированным компонентам RGB изображений.

Ключевые слова: модель классификации; опция отказа; условная по классу плотность; разделяющая функция; средняя взаимная информация; вероятность ошибки; функция скорость-погрешность; ансамбль источников.

DOI: 10.21469/22233792.4.3.03

1 Введение

В теории кодирования источников с допустимой погрешностью Шенноном введена функция скорость-погрешность [1], которая при заданной погрешности дает нижнюю границу скорости кодирования, либо при заданной скорости определяет нижнюю границу погрешности для всевозможных способов (алгоритмов) кодирования. Функция скорость-погрешность определяется параметрами источника и метрикой погрешности и не зависит от выбранного алгоритма кодирования. В отечественной литературе эта функция известна как эpsilon-энтропия Колмогорова [2]. Поэтому качество любого конкретного алгоритма кодирования может быть оценено величиной отклонения скорости кода от нижней границы при фиксированной погрешности, либо величиной отклонения погрешности кода от нижней границы при фиксированной скорости. В теории кодирования источников указанные отклонения получили термин избыточности кода.

Следуя указанным результатам теории кодирования источников, для оценки эффективности алгоритмов классификации целесообразно найти аналогичную функцию в виде зависимости наименьшей вычислительной сложности от вероятности ошибки. Поскольку задача классификации сводится к отображению множества объектов на множество классов, то интерес представляет средняя вычислительная сложность такого отображения, которая пропорциональна средней взаимной информации между множеством объектов

*Работа выполнена при финансовой поддержке РФФИ, проект №18-07-01231.

и множеством классов [3]. Поэтому, искомая функция может быть определена в терминах зависимости наименьшей средней взаимной информации от допустимой вероятности ошибки классификации.

Среди известных работ отсутствуют публикации по исследованию эффективности схем классификации в терминах зависимости взаимной информации между объектами и классами от вероятности ошибки. Существуют работы, в которых приводятся границы точности классификации (вероятности ошибки) для ансамбля многоклассовых классификаторов, объединенных на основе голосования решений, не предусматривающих отказа [4, 5]. Экспериментальные оценки вероятности ошибки для некоторых типов многоклассовых классификаторов с опцией отказа получены в [6]. Однако эти результаты носят частный характер и не содержат общего подхода к анализу потенциально достижимой точности для заданной модели классификации.

Для исследования точности классификации предлагается использовать теоретико-информационную модель на основе известной схемы кодирования источника непрерывных сообщений с заданной погрешностью при наличии канала наблюдения с аддитивным шумом [7]. В такой схеме источник задается множеством сообщений Ω с некоторой плотностью распределения. Сообщения источника передаются по каналу наблюдения, выходом которого является множество \mathbf{X} , а выход и вход канала связаны известной плотностью распределения шума. Кодирование состоит в поиске преобразования $\mathbf{X} \rightarrow \hat{\Omega}$, где $\hat{\Omega}$ — множество всевозможных аппроксимаций сообщений из множества Ω со средней погрешностью, измеряемой в евклидовой метрике и не превосходящей допустимой величины. В случае $\Omega = \hat{\Omega}$ для такой схемы в [7] определена зависимость наименьшей средней взаимной информации $I(\mathbf{X}; \hat{\Omega})$ от величины допустимой погрешности и доказано, что эта зависимость ограничивает снизу соотношение скорости и средней погрешности, получаемые на любом алгоритме кодирования.

Для построения модели классификации предлагается следующая модификация рассмотренной схемы кодирования. Пусть множество сообщений Ω представляет конечное число классов с некоторым априорным распределением, множество \mathbf{X} образовано всевозможными объектами с условными по классам плотностями распределения, а множество $\hat{\Omega}$ — всевозможными решениями о классах предъявляемых объектов с погрешностью, вычисляемой в метрике Хемминга. Такая модификация эквивалентна вероятностной схеме классификации, которая в настоящей работе используется для построения теоретико-информационной модели многоклассовой классификации с решающими правилами без отказа и с отказом. В обоих случаях оптимизация модели сводится к получению зависимости наименьшей средней взаимной информации $I(\mathbf{X}; \hat{\Omega})$ от допустимой вероятности ошибки классификации. Для схемы без отказа такая зависимость анонсирована в [8], а в [9] исследовано соотношение средней взаимной информации между множеством классов и ансамблем данных различной модальности для различных схем комплексирования.

В настоящей работе для схем классификации без отказа и с отказом строятся зависимости наименьшей средней взаимной информации между множеством объектов и множеством решений от вероятности ошибки. Такие соотношения показывают снижение наименьшей вероятности ошибки с увеличением средней взаимной информации между множеством классов и множеством классифицируемых объектов. В общем случае демонстрируется возможность уменьшения вероятности ошибки на ансамбле источников различной модальности за счет увеличения средней взаимной информации между ансамблем и множеством классов с ростом размера (числа источников) ансамбля. Приводятся экспериментальные результаты распознавания лиц по декоррелированным компонентам изображений

с применением многоклассовых классификаторов с разделяющими функциями на основе метода ближайшего соседа (NN) и метода опорных векторов (SVM). Экспериментальные оценки вероятности ошибки классификации без отказа и с отказом подтверждают повышение качества распознавания на ансамбле изображений по сравнению с распознаванием по индивидуальным компонентам ансамбля. Сравнительные оценки вероятностей ошибки на ансамбле данных получены для традиционной схемы взвешенного голосования решений по компонентам ансамбля [10], и для схемы комплексирования компонент в составные объекты, по которым принимаются групповые решения [11, 12].

2 Формализация задачи исследования

Пусть множество классов представлено алфавитом имен $\Omega^c = \{\omega_1, \dots, \omega_c\}$, $c \geq 2$, с априорными вероятностями $P(\omega_i) > 0$, $i = 1, \dots, c$. Имена классов рассматриваются как входные сообщения некоторого канала наблюдения. Выходные данные этого канала представлены множеством объектов \mathbf{X} с условными по классам плотностями распределения $\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}$, $\mathbf{x} \in \mathbf{X}$, $\omega_i \in \Omega^c$. Будем считать, что множество решений $\Omega = \Omega^c$ по объектам из \mathbf{X} формируется на выходе тест-канала с условными вероятностями решений $\mathbf{Q} = \{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}$, $\omega_j \in \Omega$, $\mathbf{x} \in \mathbf{X}$. Множества Ω^c , \mathbf{X} , Ω , совместно с условными по классам плотностями распределения объектов и условными вероятностями решений по предъявляемым объектам порождают следующую схему классификации

$$\Omega^c \Rightarrow \boxed{\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}} \Rightarrow \mathbf{X} \Rightarrow \boxed{\{Q(\omega_j|\mathbf{x}), j = 1, \dots, c\}} \Rightarrow \Omega$$

Рис. 1 Схема классификации объектов множества на множестве классов

Вероятностные характеристики схемы на рис. 1 позволяют определить среднюю взаимную информацию $I_{\mathbf{Q}}(\mathbf{X}; \Omega) \geq 0$ и вероятность ошибки $E_{\mathbf{Q}}(\mathbf{X}, \Omega) > 0$ как функционалы от тест-канала \mathbf{Q} .

Согласно [1] средняя взаимная информация определяется функционалом

$$\begin{aligned} I_{\mathbf{Q}}(\mathbf{X}; \Omega) &= \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}) \ln(Q(\omega_j|\mathbf{x})/Q(\omega_j)) \\ &= - \sum_{j=1}^c Q(\omega_j|\mathbf{x}) \ln(Q(\omega_j)) + \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}) \ln(Q(\omega_j|\mathbf{x})) \\ &= H(\Omega) - H(\Omega|\mathbf{X}), \end{aligned} \tag{1}$$

где $H(\Omega)$ и $H(\Omega|\mathbf{X})$ — безусловная и условная энтропии соответственно. Функционал полной вероятности ошибки определяется средним значением индикатора различий в метрике Хемминга $[\omega_i \neq \omega_j]$ и имеет вид

$$\begin{aligned} E_{\mathbf{Q}}(\mathbf{X}, \Omega) &= \sum_{i=1}^c P(\omega_i) \int_{\mathbf{X}} p(\mathbf{x}|\omega_i) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}) [\omega_i \neq \omega_j] \\ &= \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}) \sum_{i=1}^c P(\omega_i|\mathbf{x}) [\omega_i \neq \omega_j] \\ &= \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}) \rho(\mathbf{x}, \omega_j), \end{aligned} \tag{2}$$

где $\rho(\mathbf{x}, \omega_j)$ — вероятность ошибки по объекту \mathbf{x} на решении $\omega_j \in \Omega$. Функционалы (1) и (2) позволяют ввести невозрастающую функцию

$$R(\varepsilon) = \min_{\mathbf{Q}: E_{\mathbf{Q}}(\mathbf{X}, \Omega) \leq \varepsilon} I_{\mathbf{Q}}(\mathbf{X}; \Omega), \quad (3)$$

где минимум берется по всем тест-каналам \mathbf{Q} при заданном ограничении ε на вероятность ошибки. Функция (3) аналогична введенной в [7] функции скорость-погрешность для множества непрерывных сообщений и погрешности, вычисляемой в евклидовой метрике. В указанной работе доказано, что значение аргумента ε функции скорость-погрешность соответствует наименьшей погрешности при скорости кодирования, равной $R(\varepsilon)$. Поскольку средняя взаимная информация $I_{\mathbf{Q}}(\mathbf{X}; \Omega)$ вида (1) пропорциональна вычислительной сложности поиска решений на множестве Ω по объектам множества \mathbf{X} [3], аргумент функции (3) дает наименьшую вероятность ошибки ε при вычислительной сложности $R(\varepsilon)$.

Задача состоит в нахождении нижней границы $R_L(\varepsilon) \leq R(\varepsilon)$ функции (3) для модели классификации без отказа и в построении обобщенной функции $\tilde{R}(\varepsilon_\theta)$ для модели классификации с опцией отказа и нахождении ее нижней границы $\tilde{R}_L(\varepsilon_\theta) = \tilde{R}(\varepsilon)$. В модели с отказом параметр $0 \leq \theta \leq 1$ определяет вероятность множества Ω^c , а $1 - \theta$ соответствует вероятности класса $\omega_0 \in \Omega^c$, интерпретируемого как отказ. В этом случае аргумент функции скорость-погрешность соответствует наименьшей вероятности ошибки ε_θ , которая может быть достигнута при вычислительной сложности $\tilde{R}(\varepsilon_\theta)$. При значении $\theta = 1$ модель с отказом эквивалентна модели без отказа. Задача состоит в построении нижней границы $R_L(\varepsilon)$ при значениях $\theta \rightarrow 1$. Используя границы $R_L(\varepsilon)$ и $\tilde{R}_L(\varepsilon_\theta)$, необходимо показать уменьшение вероятности ошибки на ансамбле источников $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ размера $M \geq 2$, который порождает составные объекты в виде наборов объектов $\mathbf{x}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M)$, $\mathbf{x}_m \in \mathbf{X}_m$, $m = 1, \dots, M$, по одному от каждого источника, принадлежащих одному и тому же классу.

3 Основные результаты

3.1 Нижняя граница функции $R(\varepsilon)$

Согласно [7] вычисление функции вида (3) эквивалентно нахождению условного минимума

$$R(\varepsilon) = \min_{\mathbf{Q}_s: E_{\mathbf{Q}_s}(\Omega^*, \Omega) \leq \varepsilon - \varepsilon_{\min}} I_{\mathbf{Q}_s}(\Omega^*; \Omega) \quad (4)$$

средней взаимной информации $I_{\mathbf{Q}_s}(\Omega^*; \Omega)$ при указанном ограничении сверху на вероятность ошибки $E_{\mathbf{Q}_s}(\Omega^*, \Omega)$, где $\Omega^* = \{\omega_k, k = 1, \dots, c\}$ — множество оценок классов, получаемых на преобразовании $F: \mathbf{X} \rightarrow \Omega^*$ с минимальной вероятностью ошибки $\varepsilon_{\min} > 0$. Минимум в (4) берется по тест-каналам $\mathbf{Q}_s = \{Q_s(\omega_j | \omega_k), j = 1, \dots, c\}$, $\omega_j \in \Omega$, с условным распределением

$$Q_s(\omega_j | \omega_k) = \frac{\exp(-s[\omega_k \neq \omega_j])}{\sum_{i=1}^c \exp(-s[\omega_k \neq \omega_i])}, \quad j = 1, \dots, c, \quad (5)$$

где $\omega_k \in \Omega^*$ и $s > 0$ — свободный параметр. Соотношения (4) и (5) сводят нижнюю границу $R_L(\varepsilon)$ к известной нижней границе Шеннона [1] для схемы кодирования множества сообщений из множества Ω^* с погрешностью, измеряемой в метрике Хемминга значением индикатора $[\omega_k \neq \omega_j]$ и при условии, что средняя погрешность не превосходит величины $\varepsilon - \varepsilon_{\min} \geq 0$.

Теорема. Нижняя граница функции $R(\varepsilon)$ имеет вид

$$R_L(\varepsilon) = I(\mathbf{X}; \Omega^c) - h(\varepsilon - \varepsilon_{\min}) - (\varepsilon - \varepsilon_{\min}) \ln(c - 1), \quad \varepsilon_{\min} \leq \varepsilon \leq \varepsilon_{\max},$$

где $h(z) = -z \ln z - (1 - z) \ln(1 - z)$, $R_L(\varepsilon_{\min}) = I(\mathbf{X}; \Omega^c)$ и $R_L(\varepsilon_{\max}) = 0$.

Доказательство. Пусть распределение на множестве Ω^* задано вероятностями $P^*(\omega_k) > 0$, $k = 1, \dots, c$. Тогда согласно [1] для функции (4) справедлива нижняя граница Шеннона в форме неравенства

$$R(\varepsilon) \geq H(\Omega^*) - H_s(\Omega|\Omega^*), \quad (6)$$

где

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

$$H_s(\Omega|\Omega^*) = - \sum_{k=1}^c P^*(\omega_k) \sum_{j=1}^c Q_s(\omega_j|\omega_k) \ln Q_s(\omega_j|\omega_k),$$

а параметр s является решением уравнения

$$\sum_{k=1}^c P^*(\omega_k) \sum_{j=1}^c Q_s(\omega_j|\omega_k) [\omega_k \neq \omega_j] = \varepsilon - \varepsilon_{\min}. \quad (7)$$

С учетом (5) из (7) имеем $e^{-s} = (\varepsilon - \varepsilon_{\min}) / (c - 1)(1 - (\varepsilon - \varepsilon_{\min}))$. Последнее соотношение дает решение для s в виде функции от разности $\varepsilon - \varepsilon_{\min}$. Подстановка этого решения в (5) приводит к тест-каналу

$$Q_s(\omega_j|\omega_k) = \begin{cases} (\varepsilon - \varepsilon_{\min}) / (c - 1), & j \neq k, \\ 1 - (\varepsilon - \varepsilon_{\min}), & j = k, \end{cases}$$

на котором достигается условная энтропия

$$H_s(\Omega|\Omega^*) = h(\varepsilon - \varepsilon_{\min}) + (\varepsilon - \varepsilon_{\min}) \ln(c - 1). \quad (8)$$

При $\varepsilon = \varepsilon_{\min}$ имеем $H_s(\Omega|\Omega^*) = 0$ и, следовательно, правая часть в (6) равна $H(\Omega^*)$. Согласно теореме кодирования для источника [1], энтропия $H(\Omega^*)$ выхода преобразования $F : \mathbf{X} \rightarrow \Omega^*$ не меньше значения $R(\varepsilon_{\min})$ функции (3) в точке $\varepsilon = \varepsilon_{\min}$. Минимальная вероятность ошибки ε_{\min} достигается на байесовском решающем правиле в форме максимума апостериорной вероятности [13] и равна

$$\varepsilon_{\min} = 1 - \int_{\mathbf{X}} (\max_{1 \leq j \leq c} P(\omega_j|\mathbf{x})) p(\mathbf{x}) d\mathbf{x}.$$

Поэтому, тест-канал

$$\mathbf{Q} = \left\{ P(\omega_j|\mathbf{x}) = \frac{P(\omega_j)p(\mathbf{x}|\omega_j)}{\sum_{i=1}^c P(\omega_i)p(\mathbf{x}|\omega_i)}, \quad j = 1, \dots, c \right\}$$

в (3) дает $R(\varepsilon_{\min}) = I_{\mathbf{Q}}(\mathbf{X}; \Omega) = I(\mathbf{X}; \Omega^c)$ и, следовательно,

$$H(\Omega^*) \geq R(\varepsilon_{\min}) = I(\mathbf{X}; \Omega^c). \quad (9)$$

Замены энтропий в (6) правыми частями соотношений (8) и (9) завершают доказательство теоремы.

3.2 Обобщение для схемы классификации с опцией отказа

Пусть объекты из множества \mathbf{X} принадлежат множеству классов $\Omega^{c+1} = \{\omega_i, i = 0, \dots, c\}$, $c \geq 1$, с априорными вероятностями $P(\omega_i) > 0$: $\sum_{i=0}^c P(\omega_i) = 1$. В этом случае $\Omega^{c+1} = \Omega^c \cup \omega_0$, $c \geq 1$, где $\Omega^c = \{\omega_i, i = 1, \dots, c\}$ – множество положительных классов и ω_0 – нулевой класс, к которому должны быть отнесены объекты, не принадлежащие множеству Ω^c . Априорные вероятности классов считаются неизвестными и порождают неизвестные параметры $\theta = \sum_{i=1}^c P(\omega_i)$ и $1 - \theta = P(\omega_0)$. На множестве \mathbf{X} заданы условные по классам плотности распределения $\{p(\mathbf{x}|\omega_i), i = 1, \dots, c\}$, плотность $p(\mathbf{x}|\omega_0)$ считается неизвестной. Предполагается, что решения по объектам множества \mathbf{X} образуют множество $\Omega = \Omega^{c+1}$ с вероятностями $\{Q(\omega_j|\mathbf{x}), j = 0, \dots, c\}$, которые удовлетворяют условиям $\sum_{j=1}^c Q(\omega_j|\mathbf{x}) = \theta$ и $Q(\omega_0|\mathbf{x}) = 1 - \theta$. Множества Ω^{c+1} , \mathbf{X} , Ω , образуют схему классификации, приведенную на рис. 2. Вероятностные характеристики приведенной схемы позволяют вычислить среднюю

$$\Omega^{c+1} \Rightarrow \boxed{\{p(\mathbf{x}|\omega_i), i = 0, 1, \dots, c\}} \Rightarrow \mathbf{X} \Rightarrow \boxed{\{Q(\omega_j|\mathbf{x}), j = 0, 1, \dots, c\}} \Rightarrow \Omega$$

Рис. 2 Схема классификации объектов множества \mathbf{X} с отказом в класс ω_0

взаимную информацию $I_{Q,\theta}(\mathbf{X}; \Omega)$ и вероятность ошибки $E_{Q,\theta}(\mathbf{X}; \Omega)$, зависящие от тест-канала $Q = \{Q(\omega_j|\mathbf{x}, \Omega^c), j = 1, \dots, c\}$ и параметра θ . С учетом принятых допущений функционал средней взаимной информации имеет вид

$$\begin{aligned} I_{Q,\theta}(\mathbf{X}; \Omega) &= \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=0}^c Q(\omega_j|\mathbf{x}) \ln \frac{Q(\omega_j|\mathbf{x})}{Q(\omega_j)} \\ &= \theta \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}, \Omega^c) \ln \frac{Q(\omega_j|\mathbf{x}, \Omega^c)}{Q(\omega_j|\Omega^c)} \\ &\quad + (1 - \theta) \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \ln \frac{Q(\omega_0|\mathbf{x})}{Q(\omega_0)} \\ &= \theta I_Q(\mathbf{X}; \Omega^c) + (1 - \theta) I(\mathbf{X}; \omega_0), \end{aligned} \tag{10}$$

где

$$\begin{aligned} Q(\omega_j|\mathbf{x}, \Omega^c) &= \frac{Q(\omega_j|\mathbf{x})}{\sum_{i=1}^c Q(\omega_i|\mathbf{x})} = \frac{Q(\omega_j|\mathbf{x})}{\theta}, \quad j = 1, \dots, c, \\ Q(\omega_j|\Omega^c) &= \int_{\mathbf{X}} p(\mathbf{x}) Q(\omega_j|\mathbf{x}, \Omega^c) d\mathbf{x} = \frac{Q(\omega_j)}{\theta}, \quad j = 1, \dots, c, \\ Q(\omega_0|\mathbf{x}) &= 1 - \sum_{j=1}^c Q(\omega_j|\mathbf{x}) = 1 - \theta, \\ Q(\omega_0) &= 1 - \sum_{j=1}^c Q(\omega_j) = 1 - \theta, \end{aligned}$$

и, следовательно, $I(\mathbf{X}; \omega_0) = 0$ в (10). Функционал полной вероятности ошибки определяется средним значением индикатора ошибочных решений и имеет вид

$$\begin{aligned} E_{Q,\theta}(\mathbf{X}, \Omega) &= \sum_{i=0}^c P(\omega_i) \int_{\mathbf{X}} p(\mathbf{x}|\omega_i) d\mathbf{x} \sum_{j=0}^c Q(\omega_j|\mathbf{x}) [\omega_i \neq \omega_j] \\ &= \int_{\mathbf{X}} p(\mathbf{x}) d\mathbf{x} \sum_{j=0}^c Q(\omega_j|\mathbf{x}) \sum_{i=0}^c P(\omega_i|\mathbf{x}) [\omega_i \neq \omega_j] \\ &= \theta E_Q(\mathbf{X}|\Omega^c) + (1 - \theta) E(\mathbf{X}|\omega_0), \end{aligned} \quad (11)$$

где $E_Q(\mathbf{X}|\Omega^c)$ и $E(\mathbf{X}|\omega_0)$ — условные вероятности ошибок соответственно по объектам из множества классов Ω^c и по объектам из класса ω_0 . С учетом соотношений

$$\begin{aligned} P(\omega_i|\mathbf{x}, \Omega^c) &= \frac{P(\omega_i|\mathbf{x})}{\sum_{k=1}^c P(\omega_k|\mathbf{x})} = \frac{P(\omega_i|\mathbf{x})}{P(\Omega^c|\mathbf{x})}, \quad i = 1, \dots, c, \\ \rho(\mathbf{x}, \omega_j) &= \sum_{i=1}^c P(\omega_i|\mathbf{x}, \Omega^c) [\omega_i \neq \omega_j], \quad j = 1, \dots, c, \\ p(\mathbf{x})P(\Omega^c|\mathbf{x}) &= p(\mathbf{x}|\Omega^c)P(\Omega^c) = p(\mathbf{x}|\Omega^c) \sum_{i=1}^c P(\omega_i) = p(\mathbf{x}|\Omega^c)\theta, \end{aligned}$$

условные вероятности ошибок в (11) соответственно равны

$$\begin{aligned} E_Q(\mathbf{X}|\Omega^c) &= \theta \int_{\mathbf{X}} p(\mathbf{x}|\Omega^c) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}, \Omega^c) \rho(\mathbf{x}, \omega_j) \\ &\quad + (1 - \theta) \int_{\mathbf{X}} p(\mathbf{x}|\Omega^c) d\mathbf{x} \sum_{i=1}^c Q(\omega_i|\mathbf{x}, \Omega^c) [\omega_i \neq \omega_0] \\ &= \theta \int_{\mathbf{X}} p(\mathbf{x}|\Omega^c) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}, \Omega^c) \rho(\mathbf{x}, \omega_j) + (1 - \theta), \end{aligned} \quad (12)$$

$$E(\mathbf{X}|\omega_0) = \theta \int_{\mathbf{X}} p(\mathbf{x}|\omega_0) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}, \Omega^c) [\omega_0 \neq \omega_j] = \theta. \quad (13)$$

Последующие подстановки (12) и (13) в (11) приводят полную вероятность ошибки к виду

$$E_{Q,\theta}(\mathbf{X}, \Omega) = \theta^2 \int_{\mathbf{X}} p(\mathbf{x}|\Omega^c) d\mathbf{x} \sum_{j=1}^c Q(\omega_j|\mathbf{x}, \Omega^c) \rho(\mathbf{x}, \omega_j) + 2\theta(1 - \theta). \quad (14)$$

Функционалы (10) и (14) и значение допустимой вероятности ошибки ε_θ позволяют определить функцию скорость-погрешность

$$\tilde{R}_L(\varepsilon_\theta) = \min_{Q: E_{Q,\theta}(\mathbf{X}, \Omega) \leq \varepsilon_\theta} I_{Q,\theta}(\mathbf{X}; \Omega), \quad (15)$$

которая при $\theta = 1$ тождественна функции (3). При значениях $\theta \rightarrow 1$ имеет место асимптотическое равенство

$$p(\mathbf{x}) = \theta \sum_{i=1}^c P(\omega_i|\Omega^c)p(\mathbf{x}|\omega_i) + (1 - \theta)p(\mathbf{x}|\omega_0) \approx \sum_{i=1}^c P(\omega_i|\Omega^c)p(\mathbf{x}|\omega_i) = p(\mathbf{x}|\Omega^c),$$

где $P(\omega_i|\Omega^c) = P(\omega_i)/\sum_{k=1}^c P(\omega_k) = P(\omega_i)/\theta$. Тогда, если априорное распределение $P(\omega_i|\Omega^c)$, $i = 1, \dots, c$, и тест-канал $Q(\omega_j|\mathbf{x}, \Omega^c)$, $j = 1, \dots, c$, одинаковы в схемах с отказом и без отказа, значения функций (3) и (15) и их аргументов при $\theta \rightarrow 1$ удовлетворяют соотношениям

$$\begin{cases} \tilde{R}(\varepsilon_\theta) = \theta R(\varepsilon), \\ \varepsilon_\theta = \theta^2\varepsilon + 2\theta(1 - \theta) \geq \varepsilon. \end{cases} \quad (16)$$

Из (16) следует, что при фиксированной средней взаимной информации, потенциально достижимая вероятность ошибки в схеме с отказом превосходит вероятность ошибки в схеме без отказа. Формулы (16) дают также параметрические соотношения для нижних границ функций скорость-погрешность, определенных в (3) и (15).

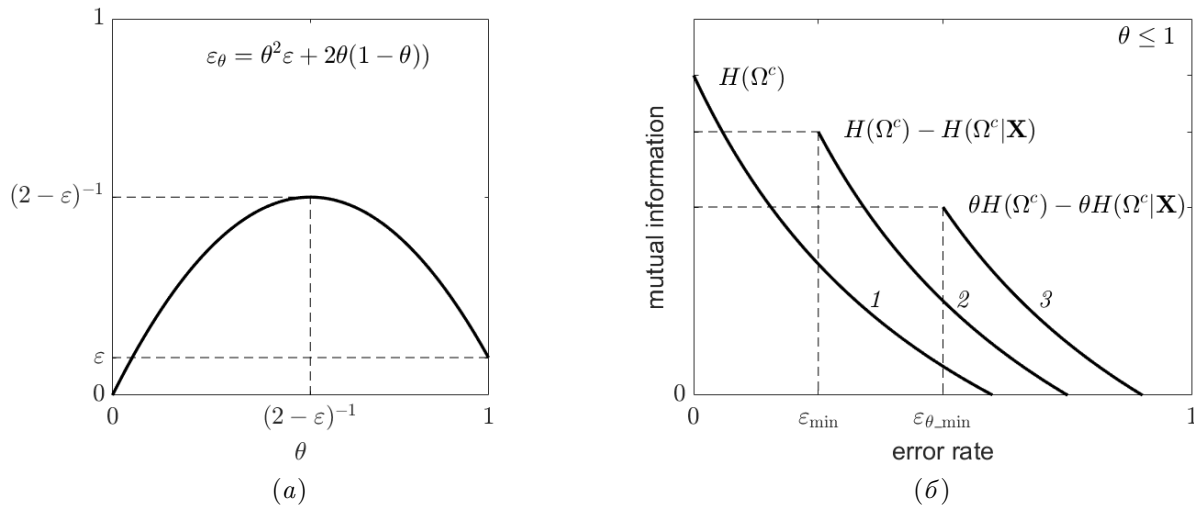


Рис. 3 Вероятности ошибки (а) и нижние границы функции скорость-погрешность (б) для схем классификации без отказа и с опцией отказа

График зависимости вероятности ошибки ε_θ в схеме с отказом от параметра θ представлен на рис. 3а. Характер нижних границ $R_L(\varepsilon)$ и $\tilde{R}_L(\varepsilon_\theta)$ как функций переменной ε иллюстрируют кривые 2 и 3 на рис. 3б. В случае $\theta \rightarrow 1$ кривая 3 приближается к кривой 2. Для сравнения дана нижняя граница Шеннона 1 для функции скорость-погрешность в схеме кодирования алфавита сообщений Ω^c с допустимой вероятностью ошибки ε в случае «бесшумного» канала наблюдения. Граница 1 следует из границы 2, когда апостериорные вероятности $P(\omega_j|\mathbf{x}) = P(\omega_j)p(\mathbf{x}|\omega_j)/\sum_{i=1}^c P(\omega_i)p(\mathbf{x}|\omega_i)$, $j = 1, \dots, c$, принимают значения 0 или 1. В этом случае условная энтропия $H(\Omega^c|\mathbf{X}) = 0$, а средняя взаимная информация $I(\mathbf{X}; \Omega^c) = H(\Omega^c) - H(\Omega^c|\mathbf{X})$ достигает наибольшего значения $H(\Omega^c)$, которое соответствует нулевой вероятности ошибки $\varepsilon_{\min} = 0$.

3.3 Вероятность ошибки на ансамбле источников

Графики нижних границ функций скорость-погрешность, представленные на рис.3б, показывают, что наименьшие вероятности ошибки в схемах классификации без отказа и с отказом зависят от средней взаимной информации $I(\mathbf{X}; \Omega^c)$ и уменьшаются с понижением условной энтропии $H(\Omega^c|\mathbf{X})$ и, следовательно, с увеличением $I(\mathbf{X}; \Omega^c)$. Последнее может быть достигнуто на ансамбле источников $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ размера $M \geq 2$, в котором множество \mathbf{X}_m , $m = 1, \dots, M$, представляет объекты m -го источника. Ансамбль порождает составные объекты в виде наборов $\mathbf{x}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M) \in \mathbf{X}^M$, в которых объекты $\mathbf{x}_m \in \mathbf{X}_m$, $m = 1, \dots, M$, принадлежат одному классу.

Средняя взаимная информация между \mathbf{X}^M и Ω^c равна $I(\mathbf{X}^M; \Omega^c) = H(\Omega^c) - H(\Omega^c|\mathbf{X}_1 \dots \mathbf{X}_M)$, где $H(\Omega^c|\mathbf{X}_1 \dots \mathbf{X}_M) \leq H(\Omega^c|\mathbf{X}_m)$, $m = 1, \dots, M$ [1]. В случае декоррелированных источников эти неравенства выполняются со знаком «меньше» и величина теоретико-информационного пересечения $I(\mathbf{X}^M; \Omega^c)$ между ансамблем \mathbf{X}^M и множеством классов Ω^c превосходит величину пересечения $I(\mathbf{X}_m; \Omega^c)$ между \mathbf{X}_m и Ω^c , $m = 1, \dots, M$. Последнее приводит к снижению минимальной вероятности ошибки ε_{θ_min} , достижимой на ансамбле, относительно аналогичных значений $\varepsilon_{\theta_min}^m$, $m = 1, \dots, M$, для индивидуальных источников, т.е. $\varepsilon_{\theta_min} < \varepsilon_{\theta_min}^m$ при $\theta \leq 1$. Справедливость этого утверждения иллюстрируется приведенной на рис. 4а схемой пересечения ансамбля $\mathbf{X}_1\mathbf{X}_2$ размера $M = 2$ с множеством классов Ω^c (множества \mathbf{X}_1 и \mathbf{X}_2 ограничены эллипсами, множество Ω^c представлено кругом). Характер нижних границ функций скорость-погрешность для источника \mathbf{X}_m и ансамбля \mathbf{X}^M показан на рис. 4б. При одинаковых значениях средней взаимной информации эти границы демонстрируют уменьшение вероятности ошибки на ансамбле по сравнению с индивидуальными источниками.

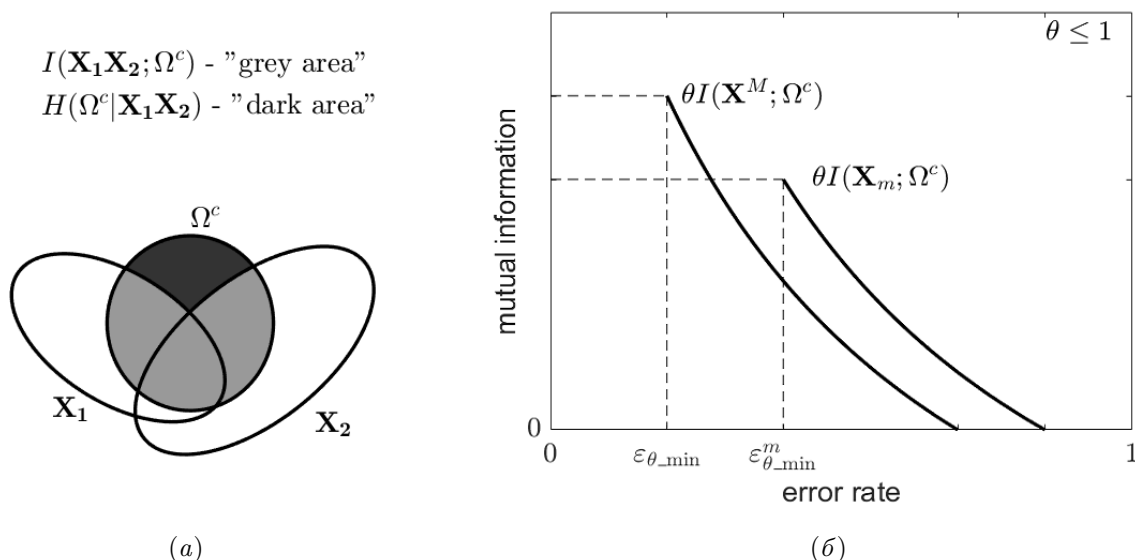


Рис. 4 Иллюстрация уменьшения вероятности ошибки на ансамбле источников

Повышение качества распознавания на ансамбле источников подтверждается приведенными в следующем разделе экспериментальными оценками вероятности ошибки для классификаторов с решающими правилами без отказа и с отказом.

4 Экспериментальные результаты

Эксперименты выполнены на цветных RGB изображениях лиц [14]. Множество исходных изображений содержит 1000 объектов от 25 персон, по 40 изображений от каждой персоны, полученных при различном ракурсе съемки. Персоны образуют множество Ω^c , содержащее $c = 25$ классов с равномерным априорным распределением $P(\omega_i|\Omega^c) = 1/c$, $i = 1, \dots, 25$. Множества декоррелированных изображений $I_1(\text{RGB})$, $I_2(\text{RGB})$, $I_3(\text{RGB})$ образуют множества объектов источников $\mathbf{X}_1 = \{I_1\}$, $\mathbf{X}_2 = \{I_2\}$, $\mathbf{X}_3 = \{I_3\}$ и ансамбль $\mathbf{X}^3 = \mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$ мощности $\|\mathbf{X}_1\| = \|\mathbf{X}_2\| = \|\mathbf{X}_3\| = \|\mathbf{X}^3\| = 1000$. Декорреляция выполнена с помощью функции `decorrstretch` в MATLAB [<https://www.mathworks.com/help/images/ref/decorrstretch.html>]. Информативные объекты источников представлены наборами эллиптических примитивов, структурированных в бинарные деревья глубины $L = 8$ [15].

На множестве древовидных представлений введена универсальная для всех источников мера различия объектов $\mathbf{x}_m \in \mathbf{X}_m$, $\hat{\mathbf{x}}_m \in \mathbf{X}_m$:

$$d(\mathbf{x}_m, \hat{\mathbf{x}}_m) \geq 0, \quad m = 1, 2, 3, \quad (17)$$

и обобщенная взвешенная мера различия составных объектов $\mathbf{x}^3 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbf{X}^3$, $\hat{\mathbf{x}}^3 = (\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \hat{\mathbf{x}}_3) \in \mathbf{X}^3$:

$$D(\mathbf{x}^3, \hat{\mathbf{x}}^3) = w_1d(\mathbf{x}_1, \hat{\mathbf{x}}_1) + w_2d(\mathbf{x}_2, \hat{\mathbf{x}}_2) + w_3d(\mathbf{x}_3, \hat{\mathbf{x}}_3) \quad (18)$$

с весовыми коэффициентами $w_m > 0$, $m = 1, 2, 3$. Меры различия (17) и (18) порождают функции сходства объектов источников и составных объектов ансамбля с классами, соответственно $0 < s_i^d(\mathbf{x}_m) < 1$ и $0 < s_i^D(\mathbf{x}^3) < 1$, $i = 1, \dots, c$ [11, 12]. Функции сходства совместно с априорными вероятностями классов дают разделяющие функции вида

$$\begin{aligned} g_i(\mathbf{x}_m) &= P(\omega_i|\Omega^c)s_i^d(\mathbf{x}_m), \quad m = 1, 2, 3, \\ g_i(\mathbf{x}^3) &= P(\omega_i|\Omega^c)s_i^D(\mathbf{x}^3) \end{aligned} \quad (19)$$

соответственно для объектов источников и составных объектов ансамбля. Функции (19) использованы в решающих правилах по объектам источников и по составным объектам ансамбля.

При любом множестве объектов \mathbf{X} , порождаемых индивидуальными источниками или ансамблем источников, использованы следующие решающие правила по предъявляемому объекту $x \in \mathbf{X}$. Для классификатора $F : \mathbf{X} \rightarrow \Omega^c$ без отказа номер класса определяется правилом

$$j(\mathbf{x}) = \arg \max_{1 \leq i \leq c} g_i(\mathbf{x}) \quad (20)$$

Для классификатора $F : \mathbf{X} \rightarrow \Omega^{c+1}$ с отказом в нулевой класс $\omega_0 \notin \Omega^c$ номер класса определяется правилом

$$j(\mathbf{x}) = \max_{1 \leq i \leq c} [g_i(\mathbf{x}) \geq \Delta_i] \arg \max_{1 \leq i \leq c} (g_i(\mathbf{x}) [g_i(\mathbf{x}) \geq \Delta_i]), \quad (21)$$

где $\{\Delta_i \geq 0, i = 1, \dots, c\}$ — множество порогов. Правило (21) дает класс ω_0 , если $g_i(\mathbf{x}) < \Delta_i$ при $i = 1, \dots, c$.

Эксперименты проведены по схеме скользящего контроля [16] для многоклассовых классификаторов типа NN и SVM, в которых функции сходства объекта с классами вычисляются на основе метода ближайшего соседа и метода опорных векторов. В указанных классификаторах использованы решающие правила (20) и (21). Для классификаторов с опцией отказа нулевой класс эмулирован объектами одного из классов, выбираемого

в скользящем режиме. Многоклассовые NN и SVM классификаторы построены по схеме «класс против всех». В разделяющих функциях NN классификаторов использованы функции сходства $s_i^d(\mathbf{x}_m)$ или $s_i^D(\mathbf{x}^3)$ экспоненциального типа, аргументом которых является значение меры различия (17) или (18) между предъявляемым объектом \mathbf{x}_m или \mathbf{x}^3 и соответствующим ближайшим представителем i -го класса обучающей выборки. Разделяющие функции SVM классификаторов используют функции сходства $s_i^d(\mathbf{x}_m)$ или $s_i^D(\mathbf{x}^3)$ сигмоидного типа. Аргументом сигмоида является алгебраическое расстояние от \mathbf{x}_m или \mathbf{x}^3 до гиперплоскости, отделяющей объекты i -го класса в обучающей выборке от объектов других классов. Алгебраическое расстояние вычисляется в векторном пространстве вторичных признаков [17], в котором компоненты вектора признаков объекта \mathbf{x}_m или \mathbf{x}^3 определяются значениями меры различия (17) или (18) между заданным объектом и всеми объектами обучающей выборки. Формальные определения функций $s_i^d(\mathbf{x}_m)$ или $s_i^D(\mathbf{x}^3)$ и даны в работе [11].

Параметры меры различия (17) для объектов индивидуальных источников и весовые коэффициенты обобщенной меры (18) для составных объектов оценивались на этапе обучения. На этапе обучения для обоих типов классификаторов с опцией отказа найдены оптимальные пороги в решающем правиле (21). Используя оценки весов источников, реализованы две схемы комплексирования: схема WMV на основе взвешенного голосования решений по индивидуальным источникам (Weighted Majority Vote [10]) и схема GDM на основе композиции источников по обобщенной мере различия составных объектов (General Dissimilarity Measure [12]). В схеме WMV решение по составному объекту $\mathbf{x}^3 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbf{X}^3$ формируется на взвешенной композиции решений (20) по объектам $\mathbf{x}_1 \in \mathbf{X}_1, \mathbf{x}_2 \in \mathbf{X}_2, \mathbf{x}_3 \in \mathbf{X}_3$. Такое решение имеет вид

$$j_{\text{WMV}}(\mathbf{x}^3) = \arg \max_{1 \leq i \leq c} (w_1[j(\mathbf{x}_1) = i] + w_2[j(\mathbf{x}_2) = i] + w_3[j(\mathbf{x}_3) = i]).$$

В схеме GDM решение по составному объекту $\mathbf{x}^3 = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3) \in \mathbf{X}^3$ принимается в соответствии с правилом (21) и имеет вид

$$i_{\text{GDM}}(\mathbf{x}^3) = \max_{1 \leq i \leq c} [g_i(\mathbf{x}^3) \geq \Delta_i] \arg \max_{1 \leq i \leq c} (g_i(\mathbf{x}^3)[g_i(\mathbf{x}^3) \geq \Delta_i]).$$

В этом случае весовые коэффициенты источников учтены в разделяющих функциях $g_i(\mathbf{x}^3)$, $i = 1, \dots, c$, вычисляемых по обобщенной взвешенной мере вида (18).

Таблица 1 Вероятности ошибки NN и SVM классификаторов без отказа ($\theta = 1$)

Классификаторы	Источники			Схемы комплексирования	
	\mathbf{X}_1	\mathbf{X}_2	\mathbf{X}_3	WMV	GDM
NN	0.0042	0.0010	0.0016	0.00004	0.00002
SVM	0.0024	0.0027	0.0025	0.00020	0.00001

Экспериментальные оценки вероятностей ошибки распознавания лиц по декоррелированным компонентам $\mathbf{X}_1, \mathbf{X}_2, \mathbf{X}_3$ и по их ансамблю $\mathbf{X}_1\mathbf{X}_2\mathbf{X}_3$ с использованием схем комплексирования WMV и GDM даны в таблицах 1 и 2. Оценки получены для NN и SVM классификаторов без отказа ($\theta = 1$) и с отказом ($\theta = 0.5$). Результаты экспериментов подтверждают повышение вероятности ошибки классификации с отказом по сравнению с

Таблица 2 Вероятности ошибки NN и SVM классификаторов с отказом ($\theta = 0.5$)

Классификаторы	Источники			Схемы комплексирования	
	X_1	X_2	X_3	WMV	GDM
NN	0.0055	0.0145	0.0295	0.0035	0.0030
SVM	0.0190	0.0500	0.0540	0.0080	0.0025

классификацией без отказа. Во всех случаях приведенные оценки демонстрируют уменьшение вероятности ошибки на ансамбле источников и преимущество схемы комплексирования GDM по сравнению со схемой WMV.

5 Заключение

Предложена теоретико-информационная модель классификации данных, которая позволяет определить наименьшую среднюю взаимную информацию между множеством классифицируемых объектов и множеством оценок классов этих объектов как убывающую функцию от допустимой вероятности ошибки. Полученная функция не зависит от конкретного алгоритма классификации и при фиксированном значении средней взаимной информации ее аргумент дает потенциально достижимую точность классификации. Поскольку средняя взаимная информация между парой множеств является мерой вычислительной сложности отображения одного множества в другое (средняя длина программы по Колмогорову), то построенная функция связывает потенциально достижимые характеристики вычислительной сложности и точности классификации. Прототипом этой функции является известная в теории информации функция скорость-погрешность для схемы кодирования источника сообщений с допустимой погрешностью при наличии канала наблюдения с шумом. В исследуемой модели классификации канал наблюдения задан условными по классам плотностями распределения признаков. Для схем классификации без отказа и с отказом получены нижние границы зависимости минимальной средней взаимной информации от вероятности ошибки. Установлено, что при фиксированной средней взаимной информации вероятность ошибки в схеме с отказом превосходит вероятность ошибки в схеме без отказа. Показано, что для указанных схем наименьшие вероятности ошибки достигаются при наибольшей средней взаимной информации, определяемой вероятностными характеристиками входных и выходных данных канала наблюдения. Продемонстрирована возможность снижения вероятности ошибки в случае составного канала наблюдения, который порождает ансамбль выходных данных (источников), обладающий большей средней взаимной информацией о множестве классов по сравнению со средней взаимной информацией от каждого источника. Полученные теоретические результаты подтверждены сравнительными экспериментальными оценками вероятности ошибки распознавания лиц по декоррелированным компонентам цветных изображений и по ансамблю компонент. Оценки на ансамбле получены с использованием композиции решений по индивидуальным компонентам (источникам) и с использованием композиции источников в форме обобщенной меры различия цветных изображений. Эксперименты продемонстрировали более высокое качество схемы комплексирования источников по сравнению со схемой комплексирования решений по данным индивидуальных источников.

Литература

- [1] *Gallager, R. G. Information Theory and Reliable Communication. Wiley and Sons, 1968.*

- [2] *Колмогоров А. Н., Тихомиров В. М.* Эpsilon-энтропия и эpsilon-емкость множеств в функциональных пространствах // В кн. Теория информации и теория алгоритмов. М.: Наука, 1987. С. 119–198.
- [3] *Колмогоров А. Н.* Три подхода к определению понятия «количество информации» // Проблемы передачи информации, 1965. 1(1), 3–11.
- [4] *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, 2–31.
- [5] *Lam L., Suen C. Y.* Application of majority voting to pattern recognition: An analysis of its behavior and performance // IEEE Transactions on Systems, Man, and Cybernetics, 1997, 27(5), 553–568.
- [6] *Tax D. M. J., Duin R. P. W.* Growing a multiclass classifier with a rejection option // Pattern Recognition Letters, 2008, 29, 1565–1570.
- [7] *Добрушин Р. Л., Цыбаков Б. С.* Передача информации с дополнительным шумом // Проблемы передачи информации, 1963, 14, 21–42.
- [8] *Ланге М. М.* Информационный критерий для сравнения классификаторов на ансамбле источников // Тезисы докладов 11-й Международной конференций «Интеллектуализация обработки информации». — Москва–Барселона - М.: Торус Пресс, 2016. С. 32.
- [9] *Lange M. M.* Information Criterion of Efficiency for Metric Classifiers in Ensemble of Sources // Communications in Computer and Information Science, 2017 (в печати).
- [10] *Kuncheva L. I.* Combining Pattern Classifiers, Methods and Algorithms, Second Edition. Wiley and Sons, 2014.
- [11] *Ганебных С. Н., Ланге М. М.* Об эффективности схем комплексирования данных для распознавания образов по ансамблю изображений // Машинное обучение и анализ данных, 2017, 3(1), С. 76–89.
- [12] *Ланге М. М., Ганебных С. Н.* Схемы коллективных решений для классификации с отказами на ансамбле источников изображений // Труды Международной конференции «Информационные технологии и нанотехнологии», ИТНТ-2018. — РФ, Самара, С. 1075–1089.
- [13] *Duda R. O., Hart P. E., Stork D. G.* Pattern Classification. 2nd ed., Wiley and Sons, 2001.
- [14] База цветных изображений лиц. URL: <http://sourceforge.net/projects/colorfaces/> (создана 04.07.2017).
- [15] *Ganebnykh S. N., Lange M. M., Stepanov D. Y.* Metric Classifier Using Multilevel Network of Templates // Pattern Recognition and Image Analysis, 2012, 22(2), С. 265–277.
- [16] *Bishop, C. M.* Pattern recognition and machine learning. Springer, 2006.
- [17] *Duin R. P. W., Ridder D., Tax D. M. J.* Experiments with a featureless approach to pattern recognition // Pattern Recognition Letters, 1997, 18, С. 1159–1166.

Поступила в редакцию 19.12.2018

On Information Theoretical Model for Data Classification*

M. M. Lange, A. M. Lange

lange_mm@ccas.ru; lange_am@mail.ru

Moscow, Federal Research Center “Computer Science and Control” of RAS

*The research was supported by the Russian Foundation for Basic Research (grant 18-07-01231).

A data classification model based on the average mutual information between a set of objects under classification and a set of decisions about classes of the objects is developed. An optimization of the model consists in minimization of the average mutual information over conditional distributions for the decisions subject to a given error rate. Finding this minimum is equivalent to calculation of the rate-distortion function in a scheme of coding the random discrete class labels that are transformed into the appropriate objects by a continuous observation channel with the known class-conditional probability densities. For the classification schemes by the decision rules without and with a reject option, the lower bounds to the rate-distortion functions are calculated. These bounds allow us to compare the potential attained error rates using the different sets of submitted objects and the different observation channels. The theoretical results are supported by experimental error rates for face recognition within the decorrelated components of RGB images.

Keywords: *classification model; reject option; class-conditional density; discriminant function; average mutual information; error rate; rate-distortion function; ensemble of sources*

DOI: 10.21469/22233792.4.3.03

References

- [1] Gallager, R. G. *Information Theory and Reliable Communication*. Wiley and Sons, 1968.
- [2] Kolmogorov A. N., Tikhomirov V. M. ε -entropy and ε -capacity of Sets in Functional Spaces // *AMS Transl.*, 1965, 1(1), C. 3–11.
- [3] Kolmogorov A. N. Tri podhoda k opredeleniyu ponyatiya "kolichestvo informatyi" // *Problems Inform. Transmission*, 1965. 1(1), 3–11.
- [4] 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, 2–31.
- [5] Lam L., Suen C. Y. Application of majority voting to pattern recognition: An analysis of its behavior and performance // *IEEE Transactions on Systems, Man, and Cybernetics*, 1997, 27(5), 553–568.
- [6] Tax D. M. J., Duin R. P. W. Growing a multiclass classifier with a rejection option // *Pattern Recognition Letters*, 2008, 29, 1565–1570.
- [7] Dobrushin R., Tsybakov B. Information transmission with additional noise // *IRE Trans. Information Theory*, 1962, 8(5), 293–304.
- [8] Lange M. Information criterion for comparison of metric classifiers in ensemble of sources // *2016 Book of abstracts of the 11th International Conference on Intelligent Data Processing: Theory and Applications (IDP-2016)*. — Torus Press, 2016. p. 33.
- [9] Lange M. M. Information Criterion of Efficiency for Metric Classifiers in Ensemble of Sources // *Communications in Computer and Information Science*, 2017 (accepted).
- [10] Kuncheva L. I. *Combining Pattern Classifiers, Methods and Algorithms, Second Edition*. Wiley and Sons, 2014.
- [11] Ganebnykh S. N., Lange M. M. On Efficiency of Fusion Schemes for Pattern Recognition in Ensemble of Images // *Machine Learning and Data Analysis*, 2017, 3(1), C. 76–89.
- [12] Lange M., Ganebnykh S. Group Decision Schemes for Classification with Reject in Ensemble of Image Sources // *Proceedings of the IV International conference on information technology and nanotechnology (ITNT-2018)*, Russia, Samara, 2018, 1075–1089.
- [13] Duda R. O., Hart P. E., Stork D. G. *Pattern Classification*. 2nd ed., Wiley and Sons, 2001.
- [14] Collection of color faces. URL: <http://sourceforge.net/projects/colorfaces/> (accessed 04.07.2017).

- [15] Ganebnykh S.N., Lange M.M., Stepanov D.Y. Metric Classifier Using Multilevel Network of Templates // *Pattern Recognition and Image Analysis*, 2012, 22(2), С. 265–277.
- [16] Bishop, C.M. *Pattern recognition and machine learning*. Springer, 2006.
- [17] Duin R. P. W., Ridder D., Tax D. M. J. Experiments with a featureless approach to pattern recognition // *Pattern Recognition Letters*, 1997, 18, С. 1159–1166.

Received December 19, 2018