

УДК 519.72

О СООТНОШЕНИИ ВЗАИМНОЙ ИНФОРМАЦИИ И ВЕРОЯТНОСТИ ОШИБКИ В ЗАДАЧЕ КЛАССИФИКАЦИИ ДАННЫХ¹⁾

© 2021 г. А. М. Ланге^{1,**}, М. М. Ланге^{1,*}, С. В. Парамонов^{1,***}

¹ 119333 Москва, ул. Вавилова, 40, ФИЦ ИУ РАН, Россия

*e-mail: lange_am@mail.ru

**e-mail: lange_mm@ccas.ru

***e-mail: psvpobox@gmail.com

Поступила в редакцию 26.11.2020 г.
Переработанный вариант 26.11.2020 г.
Принята к публикации 11.03.2021 г.

Исследуется модель классификации данных на основе зависимости средней взаимной информации между предъявляемыми объектами и принимаемыми решениями от вероятности ошибки. Оптимизация модели заключается в нахождении обменного соотношения “взаимная информация–вероятность ошибки” между наименьшей средней взаимной информацией и вероятностью ошибки, которое аналогично известной функции “скорость–погрешность” (rate distortion function) для модели кодирования сообщений с допустимой погрешностью, переданных по каналу с искажениями. Строится нижняя граница введенного соотношения, которая дает нижнюю оценку вероятности ошибки классификации на заданном множестве объектов при любом фиксированном значении средней взаимной информации. Приводится обобщение соотношения “взаимная информация–вероятность ошибки” и его нижней границы для ансамбля источников. Полученные границы полезны для оценивания избыточности вероятности ошибки решающих алгоритмов с заданными наборами разделяющих функций. Библ. 11. Фиг. 4.

Ключевые слова: классификация, ансамбль источников, вероятность ошибки, взаимная информация, соотношение “взаимная информация–вероятность ошибки”, нижняя граница, разделяющая функция, решающий алгоритм, избыточность вероятности ошибки.

DOI: 10.31857/S0044466921070115

1. ВВЕДЕНИЕ

В теории кодирования источников с допустимой погрешностью Шенноном введена функция “скорость–погрешность” (rate distortion function) [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. Поэтому качество любого алгоритма кодирования может быть оценено избыточностью скорости кода относительно нижней границы при заданной погрешности, либо избыточностью погрешности кода относительно нижней границы при заданной скорости.

Следуя результатам теории кодирования источников, для модели классификации данных целесообразно найти аналогичную функцию “взаимная информация–вероятность ошибки” в виде зависимости наименьшей средней взаимной информации между множеством классифицируемых объектов и множеством решений о классах этих объектов от заданной допустимой вероятности ошибки. Такая функция дает нижнюю границу вероятности ошибки на заданном множестве объектов при фиксированной средней взаимной информации и, следовательно, позволяет оценить избыточность вероятности ошибки решающего алгоритма, реализуемого на заданном наборе разделяющих функций.

¹⁾ Работа выполнена при частичной финансовой поддержке РФФИ (код проекта 18-07-01231).

Известны работы, в которых приводятся теоретические оценки точности для классификаторов с различными решающими правилами [2], [3]. Однако эти работы не содержат общего подхода к построению нижней границы вероятности ошибки, не зависящей от конкретных решающих алгоритмов. Поскольку многие современные модели анализа данных и машинного обучения широко используют методы теории информации [4], естественно предложить новый подход для построения нижней границы вероятности ошибки классификации данных с использованием средней взаимной информации между множеством классифицируемых данных и множеством возможных решений о классах предъявляемых данных. Цель такого подхода заключается в нахождении наименьшей средней взаимной информации как убывающей функции заданной допустимой вероятности ошибки. Очевидно, что при фиксированной средней взаимной информации такая функция дает наименьшую вероятность ошибки.

Для получения нижней границы вероятности ошибки классификации на заданном множестве объектов предлагается использовать теоретико-информационную модель на основе известной схемы кодирования источника с допустимой погрешностью при наличии канала наблюдения с шумом [5]. В предлагаемой модели метки классов и классифицируемые объекты рассматриваются как входные и выходные данные канала наблюдения, вероятностные характеристики которого определяются условными по классам вероятностями или их аппроксимациями, построенными с использованием метрики на множестве объектов. Средняя погрешность между исходными метками классов и решениями измеряется в метрике Хемминга и эквивалентна вероятности ошибки. Для такой модели вводится функция “взаимная информация—вероятность ошибки” в форме зависимости наименьшей средней взаимной информации, содержащейся в множестве объектов относительно множества принимаемых решений, от вероятности ошибки и строится нижняя граница этой функции. Предлагаемая граница анонсирована в работах [6], [7] в форме обобщения нижней границы Шеннона для функции “скорость—погрешность” в схеме кодирования независимых символов конечного алфавита с допустимой средней погрешностью в метрике Хемминга. В настоящей работе дается строгое доказательство предложенной нижней границы для обменного соотношения “взаимная информация—вероятность ошибки”.

В настоящей работе получены нижние границы соотношения “взаимная информация—вероятность ошибки” для множества объектов с заданной метрикой и для ансамбля множеств различной модальности. Показана возможность уменьшения вероятности ошибки за счет увеличения средней взаимной информации между ансамблем данных и множеством классов с увеличением числа источников в ансамбле. Построенные границы достигаются на байесовском решающем алгоритме с разделяющими функциями, заданными апостериорными вероятностями классов [8]. Приведены численные реализации нижних границ функции “взаимная информация—вероятность ошибки” для множества подписей, множества лиц и для множества составных объектов, образованных парами “лицо, подпись”.

2. ТЕОРЕТИКО-ИНФОРМАЦИОННАЯ МОДЕЛЬ КЛАССИФИКАЦИИ И ЗАДАЧА ИССЛЕДОВАНИЯ

Пусть $\Omega = \{\omega_1, \dots, \omega_c\}$, $c \geq 2$, — множество классов с априорными вероятностями $P(\omega_i)$, $i = 1, \dots, c$, и \mathbf{X} — множество объектов с условными по классам вероятностями $P(\mathbf{x}|\omega_i)$ $\forall \mathbf{x} \in \mathbf{X}$, $i = 1, \dots, c$. Будем считать, что множества Ω и \mathbf{X} являются входными и выходными данными преобразования $\Omega \rightarrow \mathbf{X}$. Множество объектов \mathbf{X} и множество решений $\hat{\Omega} = \{\omega_j = 1, \dots, c\}$ о классах этих объектов являются входом и выходом преобразования $\mathbf{X} \rightarrow \hat{\Omega}$. Пусть $\mathbf{X}^N = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ — блок объектов $\mathbf{x}_n \in \mathbf{X}$, $n = 1, \dots, N$, и \mathbf{X}^N — множество всевозможных блоков длины N . Условные по классам вероятности блоков $\mathbf{X}^N \in \mathbf{X}^N$ являются характеристиками канала наблюдения $\Omega \rightarrow \mathbf{X}^N$ и образуют множество распределений

$$P = \left\{ P(\mathbf{X}^N | \omega_i) = \prod_{n=1}^N P(\mathbf{x}_n | \omega_i) : \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N | \omega_i) = 1, i = 1, \dots, c \right\},$$

а условные вероятности решений $\omega_j \in \hat{\Omega}$ для каждого предъявляемого блока являются характеристиками тест-канала $\mathbf{X}^N \rightarrow \hat{\Omega}$ и образуют множество распределений

$$Q = \left\{ Q(\omega_j | \mathbf{X}^N) : \sum_{j=1}^c Q(\omega_j | \mathbf{X}^N) = 1, \forall \mathbf{X}^N \in \mathbf{X}^N \right\}.$$

В принятых обозначениях множества Ω и $\hat{\Omega}$ состоят их одних и тех же элементов, однако вероятности элементов множества $\hat{\Omega}$ могут отличаться от априорных вероятностей соответствующих элементов множества Ω . Множества $\Omega, \mathbf{X}^N, \hat{\Omega}$ совместно с распределениями P и Q порождают схему классификации

$$\Omega \xrightarrow{P} \mathbf{X}^N \xrightarrow{Q} \hat{\Omega}, \quad (1)$$

в которой распределения P считаются известными, а распределения Q подлежат оптимизации.

Для оптимизации множества распределений Q в (1) введем функционалы средней взаимной информации $I_Q(\mathbf{X}^N; \hat{\Omega}) \geq 0$ и вероятности ошибки $E_Q(\mathbf{X}^N, \hat{\Omega}) \geq 0$, зависящие от Q . Согласно [1], средняя взаимная информация имеет вид

$$\begin{aligned} I_Q(\mathbf{X}^N; \hat{\Omega}) &= \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^N) \ln \frac{Q(\omega_j | \mathbf{X}^N)}{Q(\omega_j)} = - \sum_{j=1}^c Q(\omega_j) \ln Q(\omega_j) + \\ &+ \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^N) \ln Q(\omega_j | \mathbf{X}^N) = H(\hat{\Omega}) - H(\hat{\Omega} | \mathbf{X}^N), \end{aligned} \quad (2)$$

где

$$\begin{aligned} P(\mathbf{X}^N) &= \sum_{i=1}^c P(\omega_i) P(\mathbf{X}^N | \omega_i), \\ Q(\omega_j) &= \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N) Q(\omega_j | \mathbf{X}^N) \end{aligned}$$

соответственно безусловные вероятности блоков $\mathbf{X}^N \in \mathbf{X}^N$ и решений $\omega_j \in \hat{\Omega}$, а $H(\hat{\Omega})$ и $H(\hat{\Omega} | \mathbf{X}^N)$ – соответственно безусловная и условная энтропии на множестве решений $\hat{\Omega}$, причем $H(\hat{\Omega}) \geq H(\hat{\Omega} | \mathbf{X}^N)$. Средняя вероятность ошибки определяется средним значением индикатора $[\omega_j \neq \omega_i]$ и имеет вид

$$\begin{aligned} E_Q(\mathbf{X}^N, \hat{\Omega}) &= \sum_{i=1}^c P(\omega_i) \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N | \omega_i) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^N) [\omega_j \neq \omega_i] = \\ &= \sum_{\mathbf{X}^N \in \mathbf{X}^N} P(\mathbf{X}^N) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^N) \sum_{i=1}^c P(\omega_i | \mathbf{X}^N) [\omega_j \neq \omega_i], \end{aligned} \quad (3)$$

где

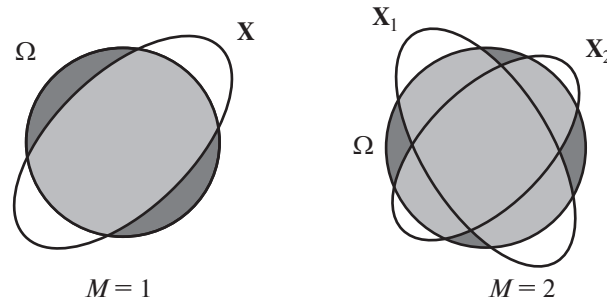
$$\sum_{i=1}^c P(\omega_i | \mathbf{X}^N) [\omega_j \neq \omega_i] = 1 - P(\omega_j | \mathbf{X}^N)$$

есть вероятность ошибки на решении ω_j по блоку объектов \mathbf{X}^N .

Используя функционалы (2) и (3), введем функцию “взаимная информация–вероятность ошибки”

$$R(\epsilon) = \min_N \min_{Q: E_Q(\mathbf{X}^N, \hat{\Omega}) \leq \epsilon} I_Q(\mathbf{X}^N; \hat{\Omega}), \quad (4)$$

где внутренний минимум берется по всевозможным множествам распределений Q при $\epsilon > 0$ и $N \geq 1$. Задача состоит в нахождении нижней границы функции $R(\epsilon)$ для источника с заданным



Фиг. 1. Интерпретация уменьшения минимальной вероятности ошибки с увеличением числа источников.

множеством объектов X и для ансамбля множеств $X^M = X_1 \dots X_M$, порождаемых $M \geq 2$ источниками различной модальности.

Следует отметить, что функция $R(\epsilon)$ убывает с ростом ϵ и, следовательно, при любом фиксированном значении дает наименьшую вероятность ошибки, которая может быть достигнута при заданной средней взаимной информации между входом и выходом классификатора. В общем случае минимальная вероятность ошибки соответствует наибольшему значению функции вида (1), которая может быть введена для $M \geq 1$ источников. Наибольшее значение такой функции должно определяться средней взаимной информацией $I(X^M; \Omega)$ между ансамблем X^M и множеством классов Ω . Иллюстрация зависимости минимальной вероятности ошибки от числа источников M дана на фиг. 1. Поскольку средняя взаимная информация $I(X^M; \Omega) = H(\Omega) - H(\Omega | X^M)$ является теоретико-информационной мерой пересечения $\Omega \cap X^M$ (“серая область”), а условная энтропия $H(\Omega | X^M)$ – теоретико-информационной мерой разности $\Omega \setminus X^M$ (“черная область”), то наименьшая вероятность ошибки должна уменьшаться с уменьшением $H(\Omega | X^M)$ и, соответственно, с ростом $I(X^M; \Omega)$. С увеличением числа M декоррелированных источников “серая область” увеличивается, а “черная область” уменьшается, что должно привести к снижению наименьшей вероятности ошибки. Формальное обоснование зависимости минимальной вероятности ошибки от энтропии $H(\Omega | X^M)$ или взаимной информации $I(X^M; \Omega)$ будет дано в разд. 3 и 4.

3. НИЖНЯЯ ГРАНИЦА ФУНКЦИИ $R(\epsilon)$ ДЛЯ ЗАДАННОГО ИСТОЧНИКА

Построение нижней границы $\underline{R}(\epsilon) \leq R(\epsilon)$ базируется на технике, предложенной в работе [5]. Пусть $\Omega^* = \{\omega_k, k = 1, \dots, c\}$ – множество решений о классах, получаемых на преобразовании $X^{N^*} \rightarrow \Omega^*$ с наименьшей вероятностью ошибки ϵ_{\min} , которая реализуется на байесовском решающем правиле для блоков $X^{N^*} \in X^{N^*}$ длины N^* (см. [8]). Введенное преобразование порождает разбиение множества X^{N^*} на непересекающиеся области решений $X_k^{N^*} \rightarrow \omega_k, k = 1, \dots, c$, которые соответствуют классам множества Ω^* . В полученном разбиении k -я область имеет вероятность

$$P^*(\omega_k) = \sum_{X^{N^*} \in X_k^{N^*}} P(X^{N^*})$$

и для любого блока $X^{N^*} \in X_k^{N^*}$ условная вероятность решения $\omega_j \in \hat{\Omega}$ одинакова и равна $Q(\omega_j | X^{N^*}) = Q^*(\omega_j | \omega_k)$. С учетом условных распределений

$$Q^* = \left\{ Q^*(\omega_j | \omega_k): \sum_{j=1}^c Q^*(\omega_j | \omega_k) = 1, k = 1, \dots, c \right\}$$

и $N = N^*$ средняя взаимная информация (2) преобразуется к виду

$$I_Q(\mathbf{X}^{N^*}; \hat{\Omega}) = I_{Q^*}(\Omega^*; \hat{\Omega}) = \sum_{k=1}^c P^*(\omega_k) \sum_{j=1}^c Q^*(\omega_j | \omega_k) \ln \frac{Q^*(\omega_j | \omega_k)}{Q(\omega_j)}, \quad (5)$$

а вероятность ошибки (3) удовлетворяет неравенству

$$E_Q(\mathbf{X}^{N^*}, \hat{\Omega}) \leq E(\mathbf{X}^{N^*}, \Omega^*) + E_{Q^*}(\Omega^*, \hat{\Omega}) = \varepsilon_{\min} + \sum_{k=1}^c P^*(\omega_k) \sum_{j=1}^c Q^*(\omega_j | \omega_k) [\omega_j \neq \omega_k]. \quad (6)$$

Соотношения (5) и (6) сводят нахождение границы $\underline{R}(\varepsilon)$ к вычислению минимума

$$\min_{Q^*: E_{Q^*}(\Omega^*, \hat{\Omega}) \leq \varepsilon - \varepsilon_{\min}} I_{Q^*}(\Omega^*; \hat{\Omega}). \quad (7)$$

Минимум (7) следует из известной нижней границы Шеннона [1] в схеме кодирования $\Omega^* \rightarrow \hat{\Omega}$ с допустимой погрешностью в метрике Хемминга $[\omega_k \neq \omega_j]$ при условии, что средняя погрешность не превосходит величины $\varepsilon - \varepsilon_{\min} \geq 0$. Полученная граница сформулирована в следующей теореме.

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

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

где $h(z) = -z \ln z - (1 - z) \ln(1 - z)$, $\underline{R}(\varepsilon_{\min}) = I(\mathbf{X}; \Omega)$, $\underline{R}(\varepsilon_{\max}) = 0$ и $I(\mathbf{X}; \Omega)$ – средняя взаимная информация между множествами \mathbf{X} и Ω .

Доказательство. Применение нижней границы Шеннона для минимума в (7) дает

$$R_L(\varepsilon) = H(\Omega^*) - H_s(\hat{\Omega} | \Omega^*), \quad (8)$$

где

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

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

а условные распределения имеют параметрическую форму

$$Q_s^* = \left\{ Q_s^*(\omega_j | \omega_k) = \frac{e^{-s[\omega_j \neq \omega_k]}}{\sum_{i=1}^c e^{-s[\omega_i \neq \omega_k]}}, j = 1, \dots, c; k = 1, \dots, c \right\}$$

с параметром $s > 0$. Значение параметра следует из уравнения

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

которое дает решение $e^{-s} = (\varepsilon - \varepsilon_{\min}) / (c - 1)(1 - (\varepsilon - \varepsilon_{\min}))$ и условные вероятности

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

на которых достигается условная энтропия

$$H_s(\hat{\Omega} | \Omega^*) = h(\varepsilon - \varepsilon_{\min}) + (\varepsilon - \varepsilon_{\min}) \ln(c - 1). \quad (9)$$

В точке $\varepsilon = \varepsilon_{\min}$ соотношение (9) дает $H_s(\hat{\Omega} | \Omega^*) = 0$ и правая часть в (8) равна $H(\Omega^*)$.

Согласно теореме кодирования для источника [1], энтропия выхода преобразования $\mathbf{X}^{N^*} \rightarrow \Omega^*$ удовлетворяет неравенству $H(\Omega^*) \geq R(\varepsilon_{\min})$. Из (4) имеем $R(\varepsilon_{\min}) = I_Q(\mathbf{X}^{N^*}; \Omega^*)$, где Q – множество распределений апостериорных вероятностей (см. [8])

$$Q(\omega_k | \mathbf{X}^{N^*}) = P(\omega_k) P(\mathbf{X}^{N^*} | \omega_k) / P(\mathbf{X}^{N^*}), \quad k = 1, \dots, c, \quad \forall \mathbf{X}^{N^*} \in \mathbf{X}^{N^*}.$$

В этом случае $I_Q(\mathbf{X}^{N^*}; \Omega^*) = I(\mathbf{X}^{N^*}; \Omega)$, а $I(\mathbf{X}^{N^*}; \Omega)$ – средняя взаимная информации между выходом и входом канала наблюдения в схеме (1). Поскольку для любого ансамбля $\mathbf{X}^N = \mathbf{X}_1 \dots \mathbf{X}_N$ справедливо неравенство $\max_{n=1}^N I(\mathbf{X}_n; \Omega) \leq I(\mathbf{X}^N; \Omega)$, которое в случае тождественных множеств $\mathbf{X}_n = \mathbf{X}$, $n = 1, \dots, N$, выполняется со знаком равенства, имеем $I(\mathbf{X}^{N^*}; \Omega) = I(\mathbf{X}; \Omega)$. С учетом сделанных замечаний получаем для энтропии $H(\Omega^*)$ следующую нижнюю оценку:

$$H(\Omega^*) \geq R(\epsilon_{\min}) = I(\mathbf{X}; \Omega). \tag{10}$$

Замены условной и безусловной энтропий в (8) правыми частями соотношений (9) и (10) завершают доказательство теоремы.

Граница $\underline{R}(\epsilon)$ убывает с увеличением ϵ и достигает наибольшего значения $I(\mathbf{X}; \Omega) = H(\Omega) - H(\Omega|\mathbf{X})$ в точке $\epsilon = \epsilon_{\min}$. Здесь

$$H(\Omega) = -\sum_{i=1}^c P(\omega_i) \ln P(\omega_i)$$

есть энтропия множества Ω и

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

есть условная энтропия множества Ω при заданном множестве \mathbf{X} , причем $H(\Omega|\mathbf{X}) \leq H(\Omega)$.

Следует отметить, что сформулированная в теореме граница $\underline{R}(\epsilon)$ является обобщением нижней границы Шеннона в схеме кодирования независимых дискретных сообщений с ограниченной средней погрешностью, измеряемой в метрике Хемминга [1]. В схеме Шеннона условные вероятности $P(\mathbf{x}|\omega_i) = [\mathbf{x} = \mathbf{x}_i]$, $i = 1, \dots, c$, принимают значения 1 и 0. В этом случае согласно формуле Байеса $Q(\omega_j|\mathbf{x}) = [\omega_j = \omega_i]$ и, следовательно, $H(\Omega|\mathbf{X}) = 0$ и $I(\mathbf{X}; \Omega) = H(\Omega)$, что дает $\epsilon_{\min} = 0$ и $\underline{R}(\epsilon_{\min}) = H(\Omega)$. В общем случае $\epsilon_{\min} \geq 0$, а наибольшая вероятность ошибки равна

$$\epsilon_{\max} = (c - 1) \min_{i=1}^c P(\omega_i)$$

и преобразуется к виду $\epsilon_{\max} = (c - 1)/c$ при равномерном априорном распределении.

Минимальная вероятность ошибки ϵ_{\min} зависит от величины условной энтропии $H(\Omega|\mathbf{X}) \geq 0$. Поскольку при значениях $\epsilon_{\min} \geq 0$ граница $\underline{R}(\epsilon)$ достигает нулевого значения в точке $\epsilon = \epsilon_{\max}$, справедливо равенство

$$h(\epsilon_{\max}) - h(\epsilon_{\max} - \epsilon_{\min}) + \epsilon_{\min} \ln(c - 1) = H(\Omega|\mathbf{X}),$$

которое с учетом разложения

$$h(\epsilon_{\max}) - h(\epsilon_{\max} - \epsilon_{\min}) = h'(\epsilon_{\max})\epsilon_{\min} + \frac{1}{2}|h''(\epsilon_{\max})|\epsilon_{\min}^2 + O(\epsilon_{\min}^3)$$

при малых значениях ϵ_{\min} преобразуется к виду

$$\frac{1}{2}|h''(\epsilon_{\max})|\epsilon_{\min}^2 + (h'(\epsilon_{\max}) + \ln(c - 1))\epsilon_{\min} = H(\Omega|\mathbf{X}).$$

Решение полученного уравнения относительно ϵ_{\min} дает при значениях производных $h'(\epsilon_{\max}) = \ln((1 - \epsilon_{\max})/\epsilon_{\max})$ и $|h''(\epsilon_{\max})| = (\epsilon_{\max}(1 - \epsilon_{\max}))^{-1}$ асимптотическую оценку

$$\begin{aligned} \epsilon_{\min} = \epsilon_{\max}(1 - \epsilon_{\max}) & \left(\ln^2 \left((c - 1) \frac{(1 - \epsilon_{\max})}{\epsilon_{\max}} \right) + 2 \frac{H(\Omega|\mathbf{X})}{\epsilon_{\max}(1 - \epsilon_{\max})} \right)^{1/2} - \\ & - \epsilon_{\max}(1 - \epsilon_{\max}) \ln \left((c - 1) \frac{(1 - \epsilon_{\max})}{\epsilon_{\max}} \right), \end{aligned} \tag{11}$$

которая в случае равновероятных классов имеет вид

$$\epsilon_{\min} = \frac{(c-1)}{c} \left(2 \frac{H(\Omega|\mathbf{X})}{(c-1)} \right)^{1/2}. \tag{12}$$

Оценки (11) и (12) демонстрируют уменьшение вероятности ошибки ϵ_{\min} с уменьшением условной энтропии $H(\Omega|\mathbf{X})$ и, следовательно, с увеличением средней взаимной информации $I(\mathbf{X};\Omega)$ при фиксированной энтропии $H(\Omega)$. При этом $H(\Omega|\mathbf{X}) = 0$ обеспечивает $\epsilon_{\min} = 0$.

4. ОБОБЩЕНИЕ ДЛЯ АНСАМБЛЯ ИСТОЧНИКОВ

Будем считать, что ансамбль $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ порождает составные объекты $\mathbf{x}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M)^t$, содержащие M объектов одного класса, по одному от каждого источника $\mathbf{x}_m \in \mathbf{X}_m, m = 1, \dots, M$, и каждый составной объект \mathbf{x}^M образует столбец размера M . Набор из N составных объектов образует блок столбцов $\mathbf{X}^{MN} = (\mathbf{x}_1^M, \dots, \mathbf{x}_N^M), \mathbf{x}_n^M \in \mathbf{X}^M, n = 1, \dots, N$; всевозможные блоки \mathbf{X}^{MN} размера $M \times N$ образуют множество \mathbf{X}^{MN} . В этом случае канал наблюдения $\Omega \rightarrow \mathbf{X}^{MN}$ задается условными распределениями вероятностей блоков \mathbf{X}^{MN} :

$$\mathbf{P}^M = \left\{ P(\mathbf{X}^{MN} | \omega_i) = \prod_{n=1}^N P(\mathbf{x}_n^M | \omega_i) : \sum_{\mathbf{X}^{MN} \in \mathbf{X}^{MN}} P(\mathbf{X}^{MN} | \omega_i) = 1, i = 1, \dots, c \right\},$$

а тест-канал $\mathbf{X}^{MN} \rightarrow \hat{\Omega}$ определяется условными распределениями вероятностей решений:

$$\mathbf{Q}^M = \left\{ Q(\omega_j | \mathbf{X}^{MN}) : \sum_{j=1}^c Q(\omega_j | \mathbf{X}^{MN}) = 1, \forall \mathbf{X}^{MN} \in \mathbf{X}^{MN} \right\}$$

по блокам составных объектов.

Множества $\Omega, \mathbf{X}^{MN}, \hat{\Omega}$ совместно с распределениями \mathbf{P}^M и \mathbf{Q}^M порождают схему классификации

$$\Omega \xrightarrow{\mathbf{P}^M} \mathbf{X}^{MN} \xrightarrow{\mathbf{Q}^M} \hat{\Omega},$$

которая аналогична схеме (1). Распределения \mathbf{P}^M и \mathbf{Q}^M дают среднюю взаимную информацию $I_{\mathbf{Q}^M}(\mathbf{X}^{MN}; \hat{\Omega})$ и вероятность ошибки $E_{\mathbf{Q}^M}(\mathbf{X}^{MN}, \hat{\Omega})$ в форме функционалов (2) и (3). Для заданного значения $\epsilon > 0$ эти функционалы позволяют ввести невозрастающую функцию

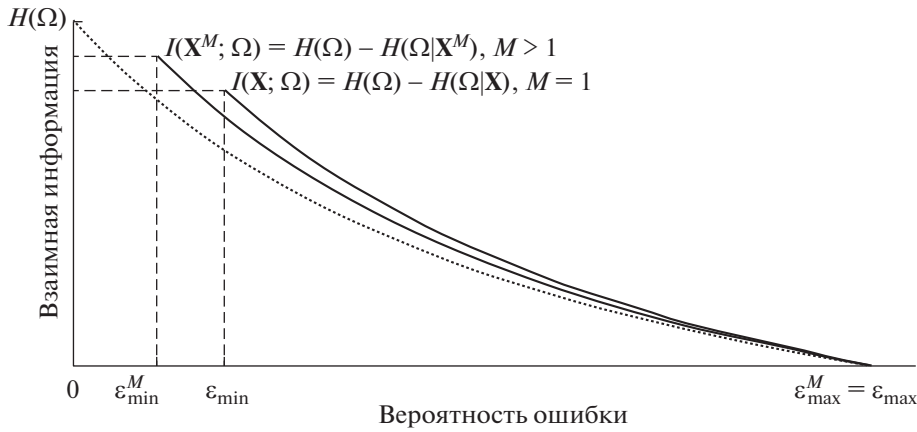
$$R_M(\epsilon) = \min_N \min_{\mathbf{Q}^M: E_{\mathbf{Q}^M}(\mathbf{X}^{MN}, \hat{\Omega}) \leq \epsilon} I_{\mathbf{Q}^M}(\mathbf{X}^{MN}; \hat{\Omega}) \tag{13}$$

в форме, аналогичной (4), причем $R_M(\epsilon) = R(\epsilon)$, когда $M = 1$. Применение техники, использованной в разд. 3 для построения границы $\underline{R}(\epsilon) \leq R(\epsilon)$, дает для функции (13) нижнюю границу

$$\underline{R}_M(\epsilon) = I(\mathbf{X}^M; \Omega) - h(\epsilon - \epsilon_{\min}^M) - (\epsilon - \epsilon_{\min}^M) \ln(c-1), \quad \epsilon_{\min}^M \leq \epsilon \leq \epsilon_{\max}^M, \tag{14}$$

которая по форме аналогична границе $\underline{R}(\epsilon)$, приведенной в теореме. В границе (14) $I(\mathbf{X}^M; \Omega)$ – средняя взаимная информация между ансамблем \mathbf{X}^M и множеством Ω , $\underline{R}_M(\epsilon_{\min}^M) = I(\mathbf{X}^M; \Omega)$ и $\underline{R}_M(\epsilon_{\max}^M) = 0$. Для минимальной вероятности ошибки ϵ_{\min}^M сохраняются асимптотические оценки вида (11) и (12) с заменой энтропии $H(\Omega|\mathbf{X})$ величиной $H(\Omega|\mathbf{X}^M)$ и заменой вероятности ошибки ϵ_{\max} величиной ϵ_{\max}^M . При любом $M \geq 1$ значение максимальной вероятности ошибки равно $\epsilon_{\max}^M = (c-1) \min_{i=1}^c P(\omega_i)$.

Уменьшение минимальной вероятности ошибки ϵ_{\min}^M с ростом числа декоррелированных множеств $\mathbf{X}_m, m = 1, \dots, M$, имеет формальное объяснение. Пусть множество \mathbf{X}_1 обеспечивает наибольшую среднюю взаимную информацию $I(\mathbf{X}_1; \Omega) = \max_{m=1}^M I(\mathbf{X}_m; \Omega)$.



Фиг. 2. Характер границ $R_M(\epsilon)$ для одного и нескольких источников.

С учетом симметрии средней взаимной информации имеем

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

где

$$I(\Omega; \mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) = H(\mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) - H(\mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1, \Omega) \geq 0,$$

$$H(\mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) = - \sum_{\mathbf{x}_1 \in \mathbf{X}_1} \dots \sum_{\mathbf{x}_m \in \mathbf{X}_m} P(\mathbf{x}_1, \dots, \mathbf{x}_m) \ln P(\mathbf{x}_m | \mathbf{x}_{m-1}, \dots, \mathbf{x}_1),$$

$$H(\mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1, \Omega) = - \sum_{i=1}^c P(\omega_i) \sum_{\mathbf{x}_1 \in \mathbf{X}_1} \dots \sum_{\mathbf{x}_m \in \mathbf{X}_m} P(\mathbf{x}_1, \dots, \mathbf{x}_m | \omega_i) \ln P(\mathbf{x}_m | \mathbf{x}_{m-1}, \dots, \mathbf{x}_1, \omega_i).$$

Декорреляция источников предполагает строгую положительность условной средней взаимной информации $I(\Omega; \mathbf{X}_m | \mathbf{X}_{m-1} \dots \mathbf{X}_1) > 0, m = 2, \dots, M$, и, следовательно, увеличение средней взаимной информации $I(\mathbf{X}^M; \Omega)$ и уменьшение условной энтропии $H(\Omega | \mathbf{X}^M) = H(\Omega) - I(\mathbf{X}^M; \Omega)$ с ростом M . Поэтому из (11) и (12) следует, что увеличение числа декоррелированных источников приводит к уменьшению ϵ_{\min}^M .

Характер границ $R_M(\epsilon)$ при значениях $M \geq 1$ показан на фиг. 2 сплошными кривыми. Для сравнения дана точечная кривая нижней границы Шеннона.

5. УСЛОВНЫЕ ПО КЛАССАМ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ ОБЪЕКТОВ

Вычисление средней взаимной информации $I(\mathbf{X}; \Omega)$, которая используется в нижней границе $R_M(\epsilon)$, требует знания совместного распределения вероятностей $\{P(\omega_i, \mathbf{x}) = P(\omega_i)P(\mathbf{x} | \omega_i), \forall \mathbf{x} \in \mathbf{X}, i = 1, \dots, c\}$ на произведении $\Omega \times \mathbf{X}$. Априорные вероятности считаются известными, а условные по классам вероятности могут быть определены с помощью метрики $d(\mathbf{x}, \hat{\mathbf{x}}) \geq 0$ в $L_p, p \geq 1$, для любой пары объектов $\mathbf{x} \in \mathbf{X}, \hat{\mathbf{x}} \in \mathbf{X}$.

Полагая, что $\mathbf{x}_i \in \mathbf{X}_i, i = 1, \dots, c$, являются “центрами” непересекающихся кластеров $\mathbf{X}_i : \bigcup_{i=1}^c \mathbf{X}_i = \mathbf{X}$ и объекты каждого класса обладают свойством компактности, определим условные по классам вероятности

$$P(\mathbf{x} | \omega_i) = \frac{e^{-v_i d^p(\mathbf{x}, \mathbf{x}_i)}}{\sum_{\mathbf{x} \in \mathbf{X}} e^{-v_i d^p(\mathbf{x}, \mathbf{x}_i)}}, \quad i = 1, \dots, c, \tag{15}$$

с параметрами $v_i > 0, i = 1, \dots, c$. Параметры распределений (15) могут быть оценены с использованием расстояний между объектами из \mathbf{X} и “центрами”

$$\mathbf{x}_i = \arg \min_{\hat{\mathbf{x}} \in \mathbf{X}_i} \sum_{\mathbf{x} \in \mathbf{X}_i} d^p(\mathbf{x}, \hat{\mathbf{x}}), \quad i = 1, \dots, c, \tag{16}$$

которые обеспечивают наименьшие рассеяния объектов внутри кластеров.

Для нахождения параметров $v_i, i = 1, \dots, c$, воспользуемся плотностями распределения

$$p(\theta_i) = \frac{v_i^{1/p} e^{-v_i \theta_i^p}}{\Gamma(1 + 1/p)}, \quad i = 1, \dots, c,$$

случайных величин $\theta_i = d(\mathbf{x}, \mathbf{x}_i)$. Используя плотность $p(\theta_i)$ и требуя максимума вероятности

$$\Pr\{|\theta_i - \mu_i| \leq \alpha_i \mu_i\} = \int_{\mu_i(1-\alpha_i)}^{\mu_i(1+\alpha_i)} p(\theta_i) d\theta_i = \frac{1}{\Gamma(1/p)} \left(\Gamma(1/p, v_i \mu_i^p (1 - \alpha_i)^p) - \Gamma(1/p, v_i \mu_i^p (1 + \alpha_i)^p) \right) \rightarrow \max_{v_i}$$

значений θ_i в α_i -окрестности среднего значения $\mu_i > 0$ (указанная вероятность выражена в терминах неполных гамма-функций), получаем

$$v_i = \frac{1}{\mu_i^p \left((1 + \alpha_i)^p - (1 - \alpha_i)^p \right)} \ln \frac{1 + \alpha_i}{1 - \alpha_i}, \quad i = 1, \dots, c, \tag{17}$$

где $0 < \alpha_i < 1$.

Параметры μ_i и α_i в (17) определяются статистиками внутриклассовых расстояний $d(\mathbf{x}, \mathbf{x}_i)$ в виде выборочных средних

$$\mu_i = \frac{1}{\|\mathbf{X}_i\|} \sum_{\mathbf{x} \in \mathbf{X}_i} d(\mathbf{x}, \mathbf{x}_i), \quad i = 1, \dots, c,$$

и выборочных дисперсий

$$\sigma_i^2 = \frac{1}{\|\mathbf{X}_i\|} \sum_{\mathbf{x} \in \mathbf{X}_i} |d(\mathbf{x}, \mathbf{x}_i) - \mu_i|^2, \quad i = 1, \dots, c.$$

Используя величины

$$\alpha_i^* = 2\mu_i / \left(\mu_i + (\mu_i^2 + \sigma_i^2)^{1/2} \right) < 1, \quad i = 1, \dots, c, \quad \delta = 1 - c^{-1} \sum_{i=1}^c \alpha_i^*,$$



















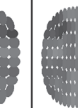
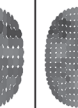
вычисляются значения $\alpha_i = \alpha_i^* + 2(1 - \alpha_i^*)\tau_s - (1 - \alpha_i^*)\tau_s^2 < 1$ в точках $\tau_s = \sum_{k=1}^s 2^{-k}, s = 1, 2, \dots$, и соответствующие растущие значения параметра v_i вида (17) и $\underline{R}(\epsilon_{\max}) < 0$ до выполнения условия $|\underline{R}(\epsilon_{\max})| \leq \delta$ при достаточно малом пороге δ , когда $\mu_i \geq \sigma_i$. Для вычисления значений $\underline{R}(\epsilon_{\max})$ используются асимптотические оценки ϵ_{\min} вида (11). Истинное значение ϵ_{\min} вычисляется коррекцией оценки ϵ_{\min} при наибольшем значении v_i путем сдвига вправо до достижения равенства $\underline{R}(\epsilon_{\max}) = 0$.

Полученные на множествах объектов от различных источников вероятности (15) с учетом оценок (16) и (17) порождают условные по классам распределения

$$\left\{ P(\mathbf{x}_m | \omega_i) = \frac{e^{-v_i d^p(\mathbf{x}_m, \mathbf{x}_{im})}}{\sum_{\mathbf{x}_m \in \mathbf{X}_m} e^{-v_i d^p(\mathbf{x}_m, \mathbf{x}_{im})}}; \sum_{\mathbf{x}_m \in \mathbf{X}_m} P(\mathbf{x}_m | \omega_i) = 1, i = 1, \dots, c \right\}, \quad m = 1, \dots, M, \tag{18}$$

на множествах $\mathbf{X}_m, m = 1, \dots, M$, и условные по классам распределения

$$\left\{ P(\mathbf{x}^M | \omega_i) = \prod_{m=1}^M P(\mathbf{x}_m | \omega_i); \sum_{\mathbf{x}^M \in \mathbf{X}^M} P(\mathbf{x}^M | \omega_i) = 1, i = 1, \dots, c \right\} \tag{19}$$

									
подпись	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$	$l = 7$	$l = 8$
									
лицо	$l = 0$	$l = 1$	$l = 2$	$l = 3$	$l = 4$	$l = 5$	$l = 6$	$l = 7$	$l = 8$

Фиг. 3. Примеры древовидных представлений подписи и лица.

на ансамбле $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$. Условные распределения (18) и (19) совместно с априорным распределением классов позволяют вычислить среднюю взаимную информацию $I(\mathbf{X}_m; \Omega)$ для каждого источника $\mathbf{X}_m, m = 1, \dots, M$, и среднюю взаимную информацию $I(\mathbf{X}^M; \Omega)$ для ансамбля источников $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ и, следовательно, получить численные реализации нижней границы вида (14) при значениях $M \geq 1$.

6. ЭКСПЕРИМЕНТАЛЬНЫЕ СООТНОШЕНИЯ “ВЗАИМНАЯ ИНФОРМАЦИЯ– ВЕРОЯТНОСТЬ ОШИБКИ” ДЛЯ МНОЖЕСТВА ПОДПИСЕЙ, МНОЖЕСТВА ЛИЦ И ДЛЯ АНСАМБЛЯ ЭТИХ ИСТОЧНИКОВ

В численном эксперименте использованы множества подписей и лиц, заданные полутоновыми изображениями размера 256×256 с 8-битовым кодированием элементов. Каждое изображение содержит один информативный объект (лицо или подпись). Множества лиц и подписей содержат по 1000 объектов от 25 персон ($c = 25$), по 40 реализаций от каждой персоны. Априорное распределение классов принято равномерным. Информативные объекты заданы древовидными представлениями в виде структурированных наборов эллиптических примитивов [9]. Различие любой пары объектов на множестве их древовидных представлений определяется расстоянием в метрике $L_p, p \geq 1$, которое является модификацией расстояния в метрике L_1 , введенного в [9]. Цель вычислительного эксперимента состояла в получении реализаций границы $R(\epsilon)$ на множестве лиц и множестве подписей с использованием матриц расстояний, заданных ресурсами [10] и [11], а также в вычислении границы $R_M(\epsilon)$ на ансамбле этих источников.

Примеры древовидных представлений подписи и лица даны на фиг. 3. Информативные уровни $l = 1, \dots, 8$ представлений содержат по 2^l эллиптических примитивов. Параметры примитива нулевого уровня используются для нормировки 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. Приведенные представления могут быть построены для любого объекта, заданного на изображении односвязным или многосвязным набором пикселей, который имеет идентифицируемые собственные оси.

В соответствии с выбранным способом представления, объект $\mathbf{x} \in \mathbf{X}$ преобразуется в набор эллиптических примитивов

$$\mathbf{x}^L = \{E_n = (\mathbf{r}_n, \mathbf{u}_n, \mathbf{v}_n, z_n)\}, \tag{20}$$

образующих бинарное дерево глубины L . Глубина дерева определяется наибольшим уровнем разрешения; эллиптический примитив E_n соответствует n -й вершине дерева и определяется вектором центра \mathbf{r}_n , векторами большой и малой полуосей $\mathbf{u}_n, \mathbf{v}_n$ и средней яркостью z_n фрагмента, аппроксимируемого примитивом. Номера вершин формируются согласно правилу: каждая промежуточная вершина с номером n порождает пару вершин следующего уровня с номерами $2n + 1$,

$2n + 2$; вершина нулевого уровня имеет номер $n = 0$. При указанной нумерации примитив E_n с номером n находится в дереве на уровне $l = \lfloor \log_2(n + 1) \rfloor$.

Используя представления вида (20), введем на множестве \mathbf{X} расстояние в метрике $L_p, p \geq 1$, между любой парой объектов $\mathbf{x} \in \mathbf{X}$ и $\hat{\mathbf{x}} \in \hat{\mathbf{X}}$. Примитивы $E_n \in \mathbf{x}^L$ и $\hat{E}_n \in \hat{\mathbf{x}}^L$ с одинаковыми номерами будем считать соответственными, так что множество пар (E_n, \hat{E}_n) образует пересечение $\mathbf{x}^L \cap \hat{\mathbf{x}}^L$. На пересечении любой пары представлений \mathbf{x}^L и $\hat{\mathbf{x}}^L$ определим их p -степенные различия по векторам центров $(\mathbf{r}_n, \hat{\mathbf{r}}_n)$, векторам полуосей $(\mathbf{u}_n, \hat{\mathbf{u}}_n)$, $(\mathbf{v}_n, \hat{\mathbf{v}}_n)$ и уровням яркости (z_n, \hat{z}_n) :

$$\begin{aligned} \rho_r^p(\mathbf{x}^L, \hat{\mathbf{x}}^L) &= \sum_{n: (E_n, \hat{E}_n) \in \mathbf{x}^L \cap \hat{\mathbf{x}}^L} \lambda_n \|\mathbf{r}_n - \hat{\mathbf{r}}_n\|^p, \\ \rho_{uv}^p(\mathbf{x}^L, \hat{\mathbf{x}}^L) &= \sum_{n: (E_n, \hat{E}_n) \in \mathbf{x}^L \cap \hat{\mathbf{x}}^L} \lambda_n (\|\mathbf{u}_n - \hat{\mathbf{u}}_n\|^p + \|\mathbf{v}_n - \hat{\mathbf{v}}_n\|^p), \\ \rho_z^p(\mathbf{x}^L, \hat{\mathbf{x}}^L) &= \sum_{n: (E_n, \hat{E}_n) \in \mathbf{x}^L \cap \hat{\mathbf{x}}^L} \lambda_n \|z_n - \hat{z}_n\|^p, \end{aligned}$$

где

$$\lambda_n = \frac{\lfloor \log_2(n + 1) \rfloor 2^{-\lfloor \log_2(n + 1) \rfloor}}{\sum_{n: (E_n, \hat{E}_n) \in \mathbf{x}^L \cap \hat{\mathbf{x}}^L} \lfloor \log_2(n + 1) \rfloor 2^{-\lfloor \log_2(n + 1) \rfloor}}$$

есть вес n -й пары соответственных примитивов. Введенные различия дают средние значения

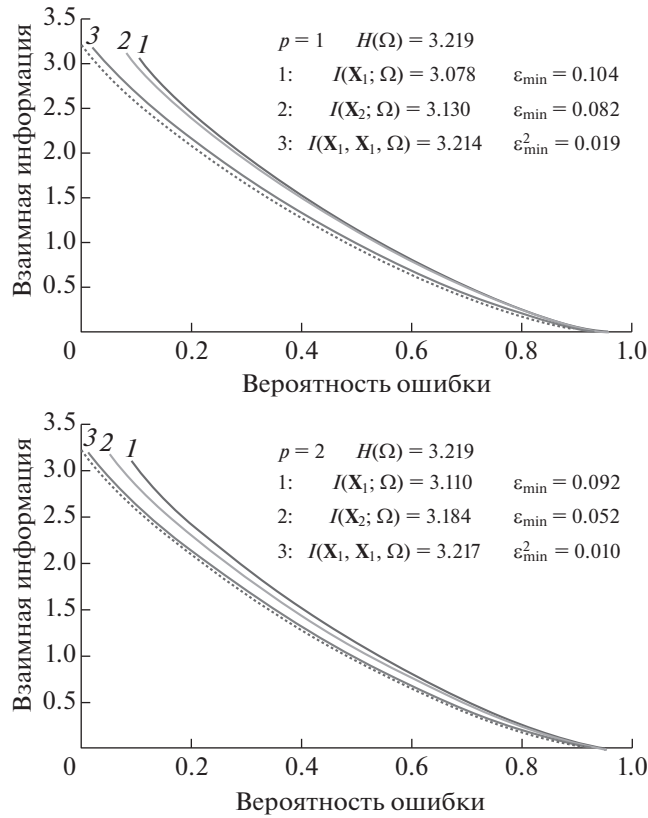
$$\begin{aligned} \sigma_r^p(\hat{\mathbf{x}}^L) &= \frac{1}{\|\mathbf{X}\|} \sum_{\mathbf{x}^L: \mathbf{x} \in \mathbf{X}} \rho_r^p(\mathbf{x}^L, \hat{\mathbf{x}}^L), \\ \sigma_{uv}^p(\hat{\mathbf{x}}^L) &= \frac{1}{\|\mathbf{X}\|} \sum_{\mathbf{x}^L: \mathbf{x} \in \mathbf{X}} \rho_{uv}^p(\mathbf{x}^L, \hat{\mathbf{x}}^L), \\ \sigma_z^p(\hat{\mathbf{x}}^L) &= \frac{1}{\|\mathbf{X}\|} \sum_{\mathbf{x}^L: \mathbf{x} \in \mathbf{X}} \rho_z^p(\mathbf{x}^L, \hat{\mathbf{x}}^L), \end{aligned}$$

на множестве объектов \mathbf{X} относительно объекта $\hat{\mathbf{x}} \in \mathbf{X}$. Указанные характеристики позволяют ввести расстояние в метрике $L_p, p \geq 1$, между объектами \mathbf{x} и $\hat{\mathbf{x}}$ в форме

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \left(\frac{\rho_r^p(\mathbf{x}^L, \hat{\mathbf{x}}^L)}{\sigma_r^p(\hat{\mathbf{x}}^L)} + \frac{\rho_{uv}^p(\mathbf{x}^L, \hat{\mathbf{x}}^L)}{\sigma_{uv}^p(\hat{\mathbf{x}}^L)} + \frac{\rho_z^p(\mathbf{x}^L, \hat{\mathbf{x}}^L)}{\sigma_z^p(\hat{\mathbf{x}}^L)} \right)^{1/p}. \tag{21}$$

Расстояния вида (21) на множествах $\mathbf{X}_m, m = 1, \dots, M$, совместно с представителями классов (16) и оценками параметров (17) полностью определяют для указанных источников условные по классам распределения вида (18), а также условные по классам распределения вида (19) для ансамбля этих источников. В экспериментах использованы расстояния в метрике L_p с параметром $p = 1$ и $p = 2$.

Графики границ $R_M(\epsilon)$, вычисленных на множестве лиц \mathbf{X}_1 и множестве подписей \mathbf{X}_2 , а также на ансамбле $\mathbf{X}_1\mathbf{X}_2$, даны на фиг. 4. Для сравнения приведена точечная кривая нижней границы Шеннона. Графики демонстрируют существенное уменьшение минимальной вероятности ошибки на ансамбле по сравнению с аналогичными вероятностями ошибки на множествах лиц или подписей. Необходимо отметить, что для рассматриваемых источников расстояние в метрике L_2 обеспечивает возможность достижения большей точности по сравнению с расстоянием в метрике L_1 .



Фиг. 4. Реализации нижних границ функции “взаимная информация–вероятность ошибки” для множества лиц X_1 (кривая 1), множества подписей X_2 (кривая 2) и ансамбля X_1X_2 (кривая 3).

7. ИЗБЫТОЧНОСТЬ ВЕРОЯТНОСТИ ОШИБКИ РЕШАЮЩЕГО АЛГОРИТМА

Полученная нижняя граница $\underline{R}_M(\varepsilon)$ позволяет оценить избыточность вероятности ошибки алгоритма, принимающего решения по блокам X^{MN} из $N \geq 1$ объектов ($M \geq 1$) с использованием набора разделяющих функций

$$G = \{g_j(X^{MN}), X^{MN} \in X^{MN}; j = 1, \dots, c\}.$$

Разделяющие функции дают значения правдоподобия решений $\omega_j \in \hat{\Omega}, j = 1, \dots, c$, по предъявляемому блоку X^{MN} и позволяют минимизировать вероятность ошибки при использовании решающего правила $j^* = \arg \max_{j=1}^c g_j(X^{MN})$. В общем случае разделяющие функции отличаются от апостериорных вероятностей, поэтому реализуемая на них вероятность ошибки превосходит минимальную вероятность ошибки байесовского алгоритма.

Решающее правило на наборе разделяющих функций G порождает разбиение множества X^{MN} на непересекающиеся области решений $X_j^{MN} \rightarrow \omega_j, j = 1, \dots, c$, которые имеют вероятности

$$Q_G(\omega_j) = \sum_{X^{MN} \in X_j^{MN}} P(X^{MN}), \quad j = 1, \dots, c.$$

Тогда набор G дает энтропию множества решений

$$H_G = - \sum_{j=1}^c Q_G(\omega_j) \ln Q_G(\omega_j)$$

и вероятность ошибки

$$\varepsilon_G = 1 - \sum_{X^{MN} \in \mathcal{X}^{MN}} P(X^{MN}) Q(\omega_{j^*} | X^{MN}) [j^* = \arg \max_{j=1}^c g_j(X^{MN})],$$

где

$$Q(\omega_j | X^{MN}) = \frac{g_j(X^{MN})}{\sum_{i=1}^c g_i(X^{MN})}$$

есть оценка апостериорной вероятности решения ω_j по блоку составных объектов X^{MN} .

Согласно теореме кодирования [1] справедливо неравенство $H_G \geq R_M(\varepsilon_G)$. Тогда пара H_G, ε_G позволяет ввести избыточность $r_G = \varepsilon_G - \varepsilon$ вероятности ошибки ε_G относительно значения

$$\varepsilon = \underline{R}_M^{-1}(H_G)[H_G < I(\mathbf{X}^M; \Omega)] + \varepsilon_{\min}^M[H_G \geq I(\mathbf{X}^M; \Omega)],$$

где $\underline{R}_M^{-1}(H_G)$ – значение обратной функции от нижней границы $\underline{R}_M(\varepsilon) = H_G$.

В случае принятия решений по блокам объектов длины $N \geq 1$, наименьшая вероятность ошибки ε_G реализуется на байесовском алгоритме, разделяющие функции которого дают апостериорные вероятности $P(\omega_j | X^{MN})$, $j = 1, \dots, c$ (см. [8]). В этом случае $H_G \geq I(\mathbf{X}^M; \Omega)$ и избыточность $r_G = \varepsilon_G - \varepsilon_{\min}^M \geq 0$. При оптимальной длине блоков $N = N^*$ байесовский алгоритм достигает минимальной вероятности ошибки ε_{\min}^M и обеспечивает нулевую избыточность.

8. ЗАКЛЮЧЕНИЕ

Исследована теоретико-информационная модель классификации данных, для которой введено обменное соотношение между наименьшей средней взаимной информацией множества классифицируемых объектов относительно множества решений по классам и заданной допустимой вероятностью ошибки. Соотношение “взаимная информация–вероятность ошибки” определено для заданного источника данных и для ансамбля источников, и является аналогом известной в теории информации функции “скорость–погрешность” (rate distortion function) для модели кодирования с заданной точностью при наличии канала наблюдения с шумом. Построена нижняя граница функции “взаимная информация–вероятность ошибки”, которая является обобщением нижней границы Шеннона. Полученная граница зависит от априорного распределения классов и условных по классам распределений на заданном множестве объектов и не зависит от решающего алгоритма. При любом фиксированном значении средней взаимной информации нижняя граница функции “взаимная информация–вероятность ошибки” дает потенциально достижимую вероятность ошибки классификации по объектам источника или по составным объектам от ансамбля источников. Показана возможность оценивания избыточности вероятности ошибки относительно нижней границы для решающего алгоритма с заданным набором разделяющих функций. Предложенная методика вычисления избыточности позволяет оценить качество любых разделяющих функций, построенных независимо от априорных вероятностей классов и условных по классам распределений предъявляемых объектов.

СПИСОК ЛИТЕРАТУРЫ

1. Gallager R.G. Information Theory and Reliable Communication. New York: Wiley & Sons, Inc. 1968.
2. Kancheva L.I., Whitaker C.J., Shipp C.A., Duin R.P.W. Limits on the majority vote accuracy in classifier fusion // Pattern Analysis and Applicat. 2003. Vol. 6. P. 22–31. <https://doi.org/10.1007/s10044-002-0173-7>
3. 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. Vol. 27(5). P. 553–568. <https://doi.org/10.1109/3468.618255>

4. *MacKay D.J.C.* Information Theory, Inference, and Learning Algorithms. C.U.P. 2003.
5. *Dobrushin R.L., Tsybakov B.S.* Information transmission with additional noise // IRE Trans. Information Theory. 1962. Vol. 8(5). P. 293–304.
<https://doi.org/10.1109/TIT.1962.1057738>
6. *Lange M., Ganebnykh S., Lange A.* An Information Approach to Accuracy Comparison for Classification Schemes in an Ensemble of Data Sources // Communications in Comput. and Informat. Sci. CCIS: Springer. 2019. Vol. 794. P. 28–43.
https://doi.org/10.1007/978-3-030-35400-8_3
7. *Lange M.M., Ganebnykh S.N., Lange A.M.* On an information-theoretical lower bound to a classification error probability // Math. Methods for Pattern Recognition: Book of abstracts of the 19th Russian Conference. Moscow: MMPR-2019. P. 59–61.
8. *Duda R.O., Hart P.E., Stork D.G.* Pattern Classification, 2nd ed. New York: Wiley & Sons, Inc. 2001.
9. *Lange M.M., Ganebnykh S.N.* On fusion schemes for multiclass object classification with reject in a given ensemble of sources // J. of Physics: Conference Series. 2018. Vol. 1096. No. 012048.
<https://doi.org/10.1088/1742-6596/1096/1/012048>
10. Матрицы расстояний изображений подписей [Электронный ресурс]. Режим доступа: <http://sourceforge.net/projects/distance-matrices-signature> (Июнь, 2020).
11. Матрицы расстояний изображений лиц [Электронный ресурс]. Режим доступа: <http://sourceforge.net/projects/distance-matrices-face> (Июнь, 2020).