

Теоретико-информационные нижние границы вероятности ошибки классификации объектов в метрических пространствах

М.М. Ланге
Федеральный исследовательский центр
«Информатика и управление» РАН
Москва, Россия
lange_mm@mail.ru

С.В. Парамонов
Федеральный исследовательский центр
«Информатика и управление» РАН
Москва, Россия
psvpobox@gmail.com

Аннотация — На множествах представлений объектов с известными расстояниями исследуется нижняя граница вероятности ошибки классификации при заданном количестве обрабатываемой информации. Исследуемая граница определяется монотонно убывающей функцией наименьшей средней взаимной информации между объектами и оценками их классов от вероятности ошибки. Для древовидных и векторных представлений получены численные реализации функции «взаимная информация-вероятность ошибки». Демонстрируется более низкая граница вероятности ошибки в пространстве векторных представлений. Отмечается возможность понижения границы путем комплексирования представлений с различными метриками.

Ключевые слова — классификация, представление объектов, расстояние, вероятность ошибки, взаимная информация.

1. ВВЕДЕНИЕ

В задаче классификации объектов в заданном пространстве их описаний (представлений) критерием качества решающих алгоритмов является вероятность ошибки, которая должна уменьшаться с ростом количества информации, используемой для принятия решения. Для вероятностной модели классификации в работе [1] построена нижняя граница вероятности ошибки как функция средней взаимной информации между множеством объектов и множеством возможных решений об их классах. Граница получена в форме модификации соотношения, известного как «gate-distortion function» для модели кодирования дискретных сообщений с погрешностью по мере Хэмминга [2].

Существенно, что в предлагаемой границе наименьшее значение вероятности ошибки зависит от средней взаимной информации между множеством объектов и множеством классов и понижается с ее ростом. Поэтому такая граница является теоретическим обоснованием максимизации указанной средней взаимной информации при отборе признаков [3]

Ранее в [4] использовалась заниженная граница, которая следует из границы, предложенной в [1], когда средняя взаимная информация стремится к энтропии множества классов, и совпадает с известной границей Шеннона [2] для кодирования с потерями.

В настоящей работе предложена методика вычисления нижней границы для вероятности ошибки классификации на множестве представлений объектов с заданным расстоянием. Методика использует условные по классам вероятности объектов в форме экспоненциально убывающих функций от квадратов расстояний между предъявляемыми объектами и "центрами" классов. Для объектов, заданных полутоновыми изображениями, рассматриваются древовидные представления наборами эллиптических

примитивов [1] и представления векторами расстояний [5] от предъявляемых объектов до представителей классов. При этом расстояния, которые образуют компоненты векторов, задаются в пространстве древовидных представлений.

Цель работы – продемонстрировать различие граничных значений вероятности ошибки классификации объектов при фиксированном количестве используемой информации в выбранных пространствах представлений и указать на возможность повышения точности путем комплексирования представлений с различными метриками.

2. ФОРМАЛИЗАЦИЯ ЗАДАЧИ

Рассматривается вероятностная модель

$$\Omega \xrightarrow{P_{X^N|\Omega}} X^N \xrightarrow{Q_{\hat{\Omega}|X^N}} \hat{\Omega}, \quad (1)$$

где $\Omega = \{\omega_i\}_{i=1}^c$ и $\hat{\Omega} = \{\omega_j\}_{j=1}^c$ – множества классов и их оценок, $c \geq 2$, X^N – множество блоков из $N \geq 1$ объектов. Множества $\Omega, X^N, \hat{\Omega}$ удовлетворяют свойству марковости и $P_{X^N|\Omega}, Q_{\hat{\Omega}|X^N}$ – соответственно множества условных по классам распределений блоков объектов и условных по блокам распределений оценок классов. На множестве Ω заданы априорные вероятности классов.

Используя среднюю взаимную информацию $I_{Q_{\hat{\Omega}|X^N}}(X^N; \hat{\Omega})$ и среднюю вероятность ошибки $E_{Q_{\hat{\Omega}|X^N}}(X^N; \hat{\Omega})$ в форме средней погрешности по мере Хэмминга как функционалы, зависящие от $Q_{\hat{\Omega}|X^N}$, в [1] для значений $\varepsilon > 0$ введена функция «взаимная информация-вероятность ошибки»

$$R(\varepsilon) = \min_N \min_{Q_{\hat{\Omega}|X^N}: E_{Q_{\hat{\Omega}|X^N}}(X^N; \hat{\Omega}) \leq \varepsilon} I_{Q_{\hat{\Omega}|X^N}}(X^N; \hat{\Omega}). \quad (2)$$

Здесь внутренний минимум берется по всевозможным $Q_{\hat{\Omega}|X^N}$ при ε -ограничении вероятности ошибки, а внешний – по длине блоков N . Для функции (2) найдена нижняя граница $\underline{R}(\varepsilon) \leq R(\varepsilon)$ в форме

$$\underline{R}(\varepsilon) = I(\Omega; X) - h(\varepsilon - \varepsilon_{\min}) - (\varepsilon - \varepsilon_{\min}) \ln(c-1), \quad (3)$$

где $I(\Omega; X)$ – средняя взаимная информация между Ω и X , $h(z) = -z \ln z - (1-z) \ln(1-z)$ – двоичная энтропия. Функция $\underline{R}(\varepsilon)$ монотонно убывает на отрезке $\varepsilon_{\min} \leq \varepsilon \leq \varepsilon_{\max}$, $\underline{R}(\varepsilon_{\min}) = I(\Omega; X)$ и $\underline{R}(\varepsilon_{\max}) = 0$. Наименьшая вероятность ошибки $\varepsilon_{\min}$ уменьшается с ростом $I(\Omega; X)$, а $\varepsilon_{\max} = (c-1) \min_{i=1}^c P(\omega_i)$, где $P(\omega_i)$ – априорная вероятность класса $\omega_i \in \Omega$. В случае $I(\Omega; X) = H(\Omega)$, $\varepsilon_{\min} = 0$ и $\underline{R}(\varepsilon)$ совпадает с границей Шеннона [2].

Вычисление характеристик $I(\Omega; \mathbf{X})$ и $\varepsilon_{\min}$ в (3) требует знания совместного распределения на множествах Ω и $\mathbf{X}$. Априорное распределение классов на множестве Ω считается равномерным с вероятностями $1/c$, а условные по классам вероятности объектов определяются с использованием расстояний $d(\mathbf{x}^{(k)}, \mathbf{x}_i^{(k)}) \geq 0, i=1, \dots, c$, между объектом $\mathbf{x} \in \mathbf{X}$ и представителем класса $\mathbf{x}_i \in \mathbf{X}$ в пространствах представлений $\mathbf{X}^{(k)}, k=1, \dots, K, K \geq 1$, множества $\mathbf{X}$.

С учетом представлений $\mathbf{x}^{(k)} \in \mathbf{X}^{(k)}, \mathbf{x}_i^{(k)} \in \mathbf{X}^{(k)}$ условные по классам вероятности объектов имеют вид

$$P_{\mathbf{x}^{(k)}|\Omega}(\mathbf{x}^{(k)} | \omega_i) = \frac{\exp(-\nu_i^{(k)} d^2(\mathbf{x}^{(k)}, \mathbf{x}_i^{(k)}))}{\sum_{\mathbf{x}^{(k)} \in \mathbf{X}^{(k)}} \exp(-\nu_i^{(k)} d^2(\mathbf{x}^{(k)}, \mathbf{x}_i^{(k)}))}, \quad (4)$$

$i=1, \dots, c$, где $\nu_i^{(k)} > 0$ – свободный параметр. В качестве представителей классов выбираются «центральные» объекты с представлениями

$$\mathbf{x}_i^{(k)} = \arg \min_{\hat{\mathbf{x}}^{(k)} \in \mathbf{X}_i^{(k)}} \sum_{\mathbf{x}^{(k)} \in \mathbf{X}_i^{(k)}} d^2(\mathbf{x}^{(k)}, \hat{\mathbf{x}}^{(k)}). \quad (5)$$

в подмножествах $\mathbf{X}_i^{(k)} \subset \mathbf{X}^{(k)}, i=1, \dots, c$.

Распределения (4) с представителями классов (5) позволяют вычислить значения $I(\Omega; \mathbf{X}) = I(\Omega; \mathbf{X}^{(k)})$ и $\varepsilon_{\min} = \varepsilon_{\min}^{(k)}$ в границе (3). Реализации границы строятся на множествах изображений лиц $\mathbf{X}_f$ и подписей $\mathbf{X}_s$ от $c = 25$ персон, по 40 объектов от каждой персоны. Для указанных данных рассматриваются множества древовидных представлений лиц $\mathbf{X}_f^{(1)}$ и подписей $\mathbf{X}_s^{(1)}$ наборами эллиптических примитивов [1] и множества представлений $\mathbf{X}_f^{(2)}, \mathbf{X}_s^{(2)}$ векторами расстояний [5] объектов до «центральных» представителей классов в пространстве древовидных представлений. На множествах древовидных представлений используется метрика с квадратичным ядром [1]. На множествах векторных представлений применяется среднее квадратичное расстояние между векторами.

3. ЧИСЛЕННЫЕ РЕАЛИЗАЦИИ ГРАНИЦЫ $\underline{R}(\varepsilon)$

Вычисление характеристик $I(\Omega; \mathbf{X})$ и $\varepsilon_{\min}$ выполнено с использованием матриц расстояний на множестве $\mathbf{X}_f^{(1)}$ [6] и на множестве $\mathbf{X}_s^{(1)}$ [7]. Численные реализации границы $\underline{R}(\varepsilon)$ показаны на рис. 1 сплошными кривыми. Пунктирная кривая соответствует границе Шеннона, когда $I(\Omega; \mathbf{X}) = H(\Omega) = \ln c = 3,219$ и $\varepsilon_{\min} = 0$.

Кривые на верхнем рисунке вычислены на множествах древовидных представлений $\mathbf{X}_f^{(1)}$ и $\mathbf{X}_s^{(1)}$, кривые на нижнем рисунке – на множествах векторных представлений $\mathbf{X}_f^{(2)}$ и $\mathbf{X}_s^{(2)}$. Поскольку для множеств лиц и подписей выполняется соотношение $I(\Omega; \mathbf{X}^{(2)}) > I(\Omega; \mathbf{X}^{(1)})$, которое обеспечивает $\varepsilon_{\min}^{(2)} < \varepsilon_{\min}^{(1)}$, реализации границы $\underline{R}(\varepsilon)$ в векторном пространстве представлений проходят ближе к границе Шеннона по сравнению с соответствующими реализациями границы в пространстве древовидных представлений.

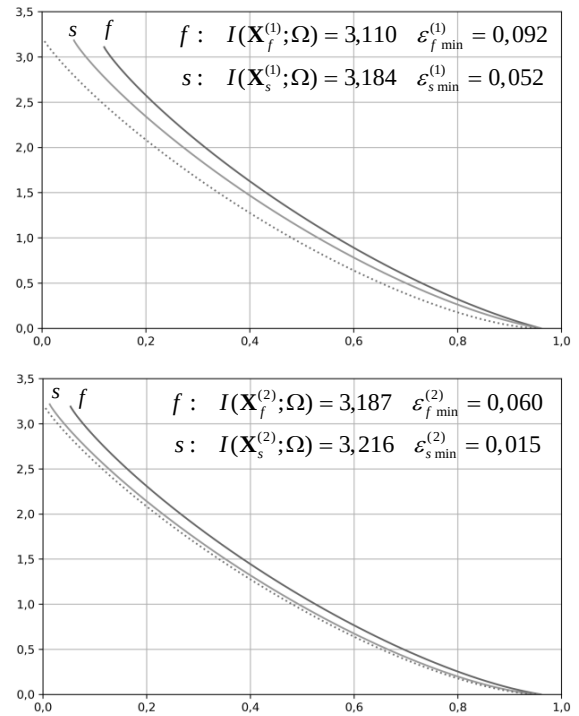


Рис.1. Реализации границы $\underline{R}(\varepsilon)$ на множествах древовидных ($k=1$) и векторных ($k=2$) представлений изображений лиц и подписей

4. ЗАКЛЮЧЕНИЕ

Предложена методика вычисления функции «взаимная информация-вероятность ошибки» на множествах представлений классифицируемых данных с заданными метриками. Благодаря монотонности такой функции, ее обращение дает нижнюю границу вероятности ошибки классификации в выбранном пространстве представлений при фиксированном количестве обрабатываемой информации. Предложенная методика допускает комплексирование представлений с различными метриками для понижения нижней границы вероятности ошибки.

ЛИТЕРАТУРА

- [1] Lange, A.M. Tradeoff Relation between Mutual Information and Error Probability in Data Classification Problems / A.M. Lange, M.M. Lange, S.V. Paramonov // Computational Mathematics and Mathematical Physics. – 2021. – Vol. 61(7). – P. 1181-1193.
- [2] Gallager, R.G. Information Theory and Reliable Communication – New York: Wiley & Sons, 1968. – 608 p.
- [3] Brown, G. Conditional Likelihood Maximization: A Unifying Framework for Information Theoretic Feature Selection / G. Brown, A. Pocock, M.J. Zhao, M. Luján // Journal of Machine Learning Research. – 2012. – Vol. 13(8). – P. 27-66.
- [4] Rigau, J. An information theoretic framework for image segmentation / J. Rigau, M. Feixas, M. Sbert // International Conference on Image Processing, ICIP. – 2004. – P. 8436089.
- [5] Duin, R.P.W. Experiments with a featureless approach to pattern recognition / R.P.W. Duin, D. de Ridder, D.M.J. Tax // Pattern Recognition Letters. – 1997. – Vol. 18. – P. 1159-1166.
- [6] Distance matrices for face dataset. [Electronic resource]. — Access mode: <http://sourceforge.net/projects/distance-matrices-face> (01.03.2024).
- [7] Distance matrices for signature dataset. [Electronic resource]. — Access mode: <http://sourceforge.net/projects/distance-matrices-signature> (01.03.2024).