

О теоретико-информационной нижней границе вероятности ошибки классификации

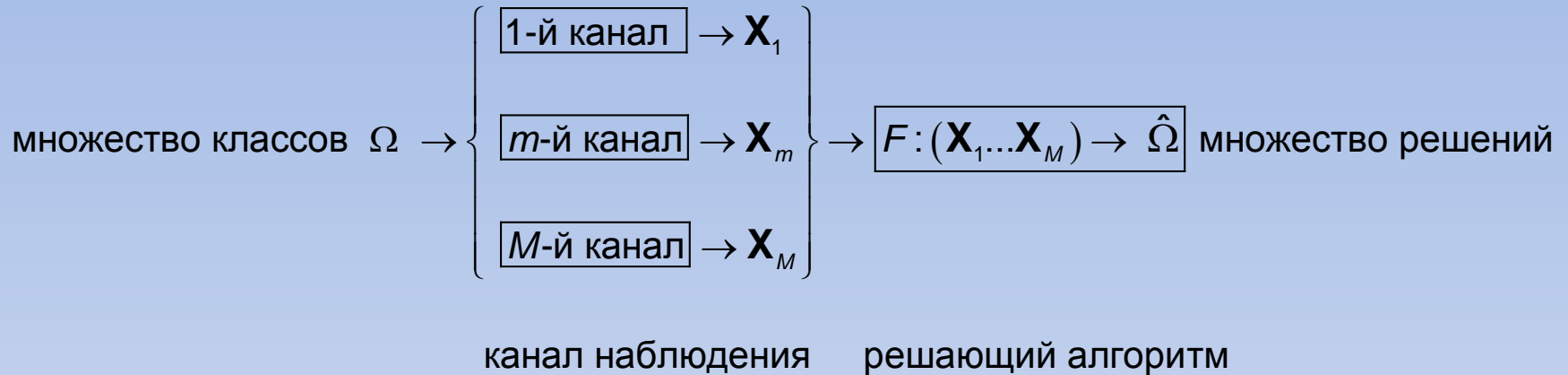
Ланге М.М., Ганебных С.Н., Ланге А.М.

Москва, Федеральный исследовательский центр
"Информатика и управление" РАН

19-я Всероссийская конференция «Математические методы распознавания
образов» ММРО-2019, 26-29 ноября, Москва

2. Предварительные замечания

- Схема классификации данных, порождаемых ансамблем источников $\mathbf{X}_1 \dots \mathbf{X}_M$, $M \geq 1$



- $\Omega \equiv \hat{\Omega}$
- $I_F(\mathbf{X}_1 \dots \mathbf{X}_M; \hat{\Omega}) \geq 0$ – средняя взаимная информация между $\mathbf{X}_1 \dots \mathbf{X}_M$ и $\hat{\Omega}$
- $P_F(\text{error}) > 0$ – вероятность ошибки алгоритма F
- Цель – найти $\min_{F: P_F(\text{error}) \leq \varepsilon} I_F(\mathbf{X}_1 \dots \mathbf{X}_M; \hat{\Omega})$ как функцию допустимой вероятности ошибки ε

3. Формальная постановка задачи

Основные определения

- $\Omega = \{\omega_i\}_{i=1}^c, c \geq 2$ – множество классов с априорными вероятностями $P(\omega_i), i = 1, \dots, c$
- $\mathbf{X}^{MN} = \{\mathbf{X}^{MN} = [\mathbf{x}_{mn}]\}$ – множество $M * N$ матриц с условными вероятностями $P(\mathbf{X}^{MN} | \omega_i), i = 1, \dots, c$
- $\mathbf{X}_m^N = (\mathbf{x}_{m1}, \dots, \mathbf{x}_{mN})$ – m -я строка: N объектов одного класса от источника \mathbf{X}_m
- $\mathbf{X}_n^M = (\mathbf{x}_{1n}, \dots, \mathbf{x}_{Mn})^T$ – n -й столбец: M объектов одного класса от ансамбля $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$
- $\mathbf{X}^{MN} = (\mathbf{X}_1^M, \dots, \mathbf{X}_N^M)$ – N составных объектов (столбцов) от ансамбля $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$
- $\hat{\Omega} = \{\omega_j\}_{j=1}^c$ – множество решений для $\mathbf{X}^{MN} \in \mathbf{X}^{MN}$ с вероятностями $Q(\omega_j | \mathbf{X}^{MN}), j = 1, \dots, c$

Теоретико - информационная модель классификации

Схема* $\Omega \Rightarrow \boxed{P = \{P(\mathbf{X}^{MN} | \omega_i), i = 1, \dots, c\}} \Rightarrow \mathbf{X}^{MN} \Rightarrow \boxed{Q = \{Q(\omega_j | \mathbf{X}^{MN}), j = 1, \dots, c\}} \Rightarrow \hat{\Omega}$

Средняя взаимная информация: $I_Q(\mathbf{X}^{MN}; \hat{\Omega}) = \sum_{\mathbf{X}^{MN} \in \mathbf{X}^{MN}} P(\mathbf{X}^{MN}) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^{MN}) \ln (Q(\omega_j | \mathbf{X}^{MN}) / Q(\omega_j))$

Средняя вероятность ошибки: $E_Q(\mathbf{X}^{MN}, \Omega) = \sum_{\mathbf{X}^{MN} \in \mathbf{X}^{MN}} P(\mathbf{X}^{MN}) \sum_{j=1}^c Q(\omega_j | \mathbf{X}^{MN}) \sum_{i=1}^c P(\omega_i | \mathbf{X}^{MN}) [\omega_i \neq \omega_j]$

Для $M \geq 1$ и $\varepsilon > 0$, вводится функция:

$$R^M(\varepsilon) = \min_N \min_{Q: E_Q(\mathbf{X}^{MN}, \Omega) \leq \varepsilon} I_Q(\mathbf{X}^{MN}; \Omega)$$

* Dobrushin R.L, Tsybakov B.S. Information transmission with additional noise // *IRE Trans. Information Theory*, 1962, **8**(5), 293 – 304

4. Вычисление функции $R^M(\varepsilon)$

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

где $I(\mathbf{X}^M; \Omega) = H(\Omega) - H(\Omega | \mathbf{X}^M)$ – средняя взаимная информация между \mathbf{X}^M и Ω ;

$H(\Omega) = -\sum_{i=1}^c P(\omega_i) \ln P(\omega_i)$ – энтропия множества классов Ω ;

$H(\Omega | \mathbf{X}^M) = -\sum_{\mathbf{X}^M \in \mathbf{X}^M} P(\mathbf{X}^M) \sum_{i=1}^c P(\omega_i | \mathbf{X}^M) \ln P(\omega_i | \mathbf{X}^M)$ – условная энтропия;

$h(z) = -z \ln z - (1-z) \ln(1-z)$ – двоичная энтропия Шеннона;

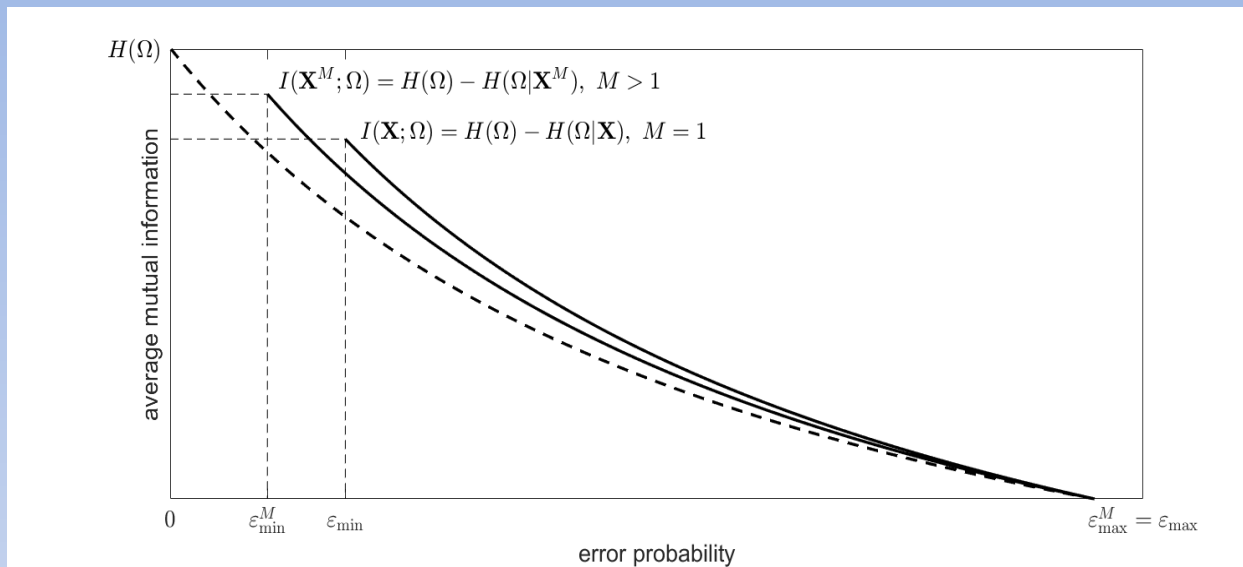
$$R_L^M(\varepsilon_{\min}^M) = I(\mathbf{X}^M; \Omega), \quad R_L^M(\varepsilon_{\max}^M) = 0$$

Для равномерного априорного распределения: $P(\omega_i) = 1/c$, $i = 1, \dots, c$

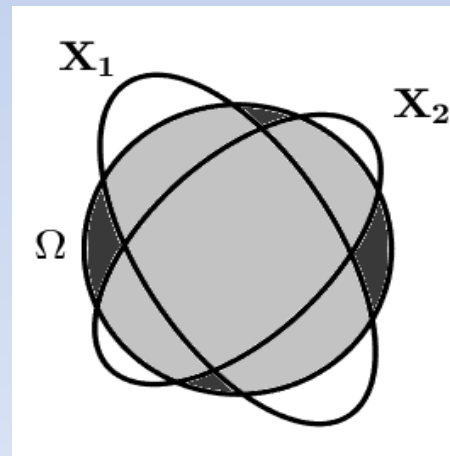
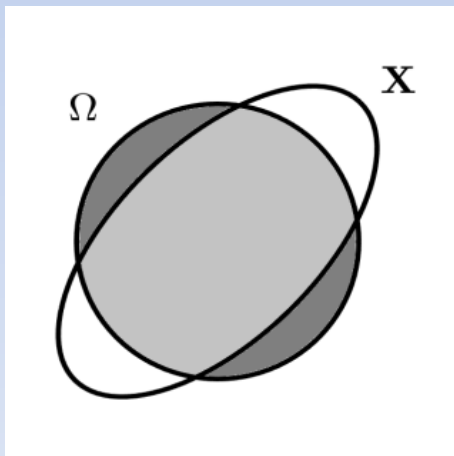
$$\varepsilon_{\min}^M = \frac{c-1}{c} \sqrt{\frac{2}{c-1} H(\Omega | \mathbf{X}^M)}, \quad \varepsilon_{\max}^M = \frac{c-1}{c}$$

5. Графическая интерпретация $R_L^M(\varepsilon)$

Графики нижних границ



Информационная мера вероятности ошибки



$I(\mathbf{X}^M; \Omega)$ – мера пересечения $\mathbf{X}^M \cap \Omega$; $H(\Omega | \mathbf{X}^M)$ – мера разности $\Omega \setminus \mathbf{X}^M$
 $H(\Omega | \mathbf{X}^M) = H(\Omega) - I(\mathbf{X}^M; \Omega)$

6. Метрики и условные по классам распределения

Мера различия объектов на множестве \mathbf{X}_m , $m = 1, \dots, M$

$$d(\mathbf{x}_m, \hat{\mathbf{x}}_m) \geq 0, \mathbf{x}_m \in \mathbf{X}_m, \hat{\mathbf{x}}_m \in \mathbf{X}_m$$

Условные по классам вероятности объектов $\mathbf{x}_m \in \mathbf{X}_m$, $m = 1, \dots, M$

$$P(\mathbf{x}_m | \omega_i) = \frac{e^{-v_m d(\mathbf{x}_m, \mathbf{x}_{im})}}{\sum_{\mathbf{x}_m \in \mathbf{X}_m} e^{-v_m d(\mathbf{x}_m, \mathbf{x}_{im})}}, v_m > 0, \mathbf{x}_{im} \in \mathbf{X}_{im} : \bigcup_{i=1}^c \mathbf{X}_{im} = \mathbf{X}_m, i = 1, \dots, c$$

Мера различия составных объектов на ансамбле $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$

$$D(\mathbf{X}^M, \hat{\mathbf{X}}^M) = \sum_{m=1}^M v_m d(\mathbf{x}_m, \hat{\mathbf{x}}_m), \mathbf{X}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M) \in \mathbf{X}^M, \hat{\mathbf{X}}^M = (\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_M) \in \mathbf{X}^M$$

Условные по классам вероятности составных объектов $\mathbf{X}^M = (\mathbf{x}_1, \dots, \mathbf{x}_M) \in \mathbf{X}^M$

$$P(\mathbf{X}^M | \omega_i) = \frac{e^{-D(\mathbf{X}^M, \mathbf{X}_i^M)}}{\sum_{\mathbf{X}^M \in \mathbf{X}^M} e^{-D(\mathbf{X}^M, \mathbf{X}_i^M)}} = \prod_{m=1}^M P(\mathbf{x}_m | \omega_i), i = 1, \dots, c$$

7. Оценки параметров канала наблюдения

Представители кластеров $\mathbf{X}_{im}, i = 1, \dots, c$ на множестве $\mathbf{X}_m = \bigcup_{i=1}^c \mathbf{X}_{im}, m = 1, \dots, M$:

$$\mathbf{x}_{im} = \arg \min_{\hat{\mathbf{x}}_m \in \mathbf{X}_{im}} \sum_{\mathbf{x}_m \in \mathbf{X}_{im}} d(\mathbf{x}_m, \hat{\mathbf{x}}_m) \in \mathbf{X}_{im}, i = 1, \dots, c$$

Параметры $v_m > 0, m = 1, \dots, M$ условных по классам распределений:

Для величин $\theta = d(\mathbf{x}_m, \mathbf{x}_{im})$ используются плотности $p_{v_m}(\theta) = v_m e^{-v_m \theta}, i = 1, \dots, c,$

которые при $0 < \alpha_m < 1$ и условии $\Pr[|\theta - \mu_m| \leq \alpha_m \mu_m] = \int_{\mu_m(1-\alpha_m)}^{\mu_m(1+\alpha_m)} p_{v_m}(\theta) d\theta \rightarrow \max_{v_m}$

дают

$$v_m = \frac{1}{2\alpha_m \mu_m} \ln \frac{1 + \alpha_m}{1 - \alpha_m}, m = 1, \dots, M$$

где μ_m и α_m определяются статистиками множества расстояний


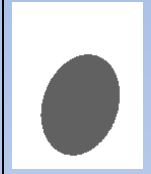
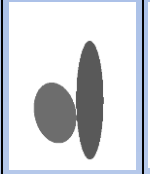
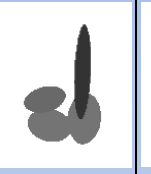



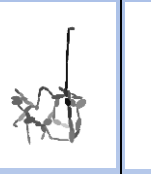
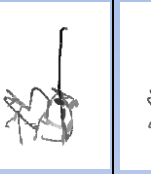


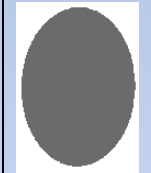
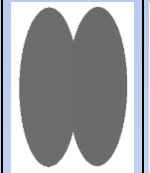
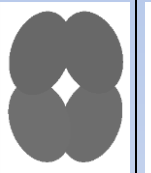

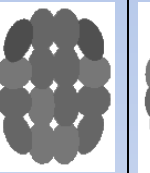
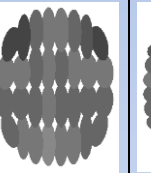
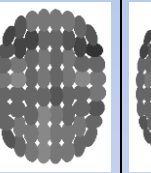
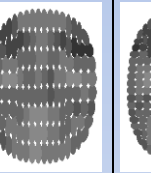
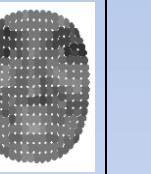
$$\{d(\mathbf{x}_m, \mathbf{x}_{im}), \forall \mathbf{x}_m \in \mathbf{X}_m, i = 1, \dots, c\}$$

8. Экспериментальные результаты

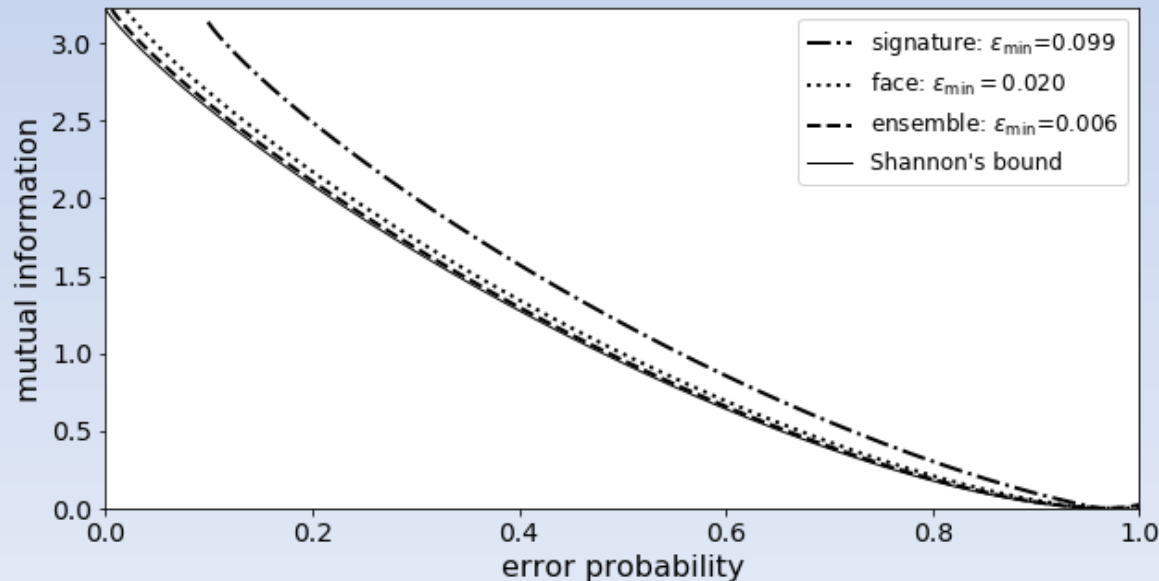
Исходные данные

Множество изображений подписей X_1 , множество изображений лиц X_2 и ансамбль X_1X_2
Изображения содержат объекты от 25 персон (классов) по 40 реализаций в классе

Примеры древовидных представлений объектов из множества X_1 и множества X_2

									
signature	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$
									
face	$l=0$	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$	$l=6$	$l=7$	$l=8$

Численные реализации границ $R_L^M(\varepsilon)$ для X_1 и X_2 ($M=1$) и для ансамбля X_1X_2 ($M=2$)



9. Избыточность вероятности ошибки алгоритмов

Разделяющие функции решающего алгоритма на ансамбле $\mathbf{X}^M = \mathbf{X}_1 \dots \mathbf{X}_M$ размера $M \geq 1$

$$G_s^M = \{g_j^s(\mathbf{X}^M), j = 1, \dots, c\}, \mathbf{X}^M \in \mathbf{X}^M, 0 \leq s \leq 1$$

Условное распределение решений

$$Q_s^M = \left\{ Q_s(\omega_j | \mathbf{X}^M) = \frac{g_j^s(\mathbf{X}^M)}{\sum_{i=1}^c g_j^s(\mathbf{X}^M)}, j = 1, \dots, c \right\}$$

Средняя взаимная информация между множеством объектов и множеством решений

$$I_{Q_s^M}(\mathbf{X}^M; \hat{\Omega}) = \sum_{\mathbf{X}^M \in \mathbf{X}^M} P(\mathbf{X}^M) \sum_{j=1}^c Q_s(\omega_j | \mathbf{X}^M) \ln(Q_s(\omega_j | \mathbf{X}^M) / Q(\omega_j)) = R_{G_s^M}$$

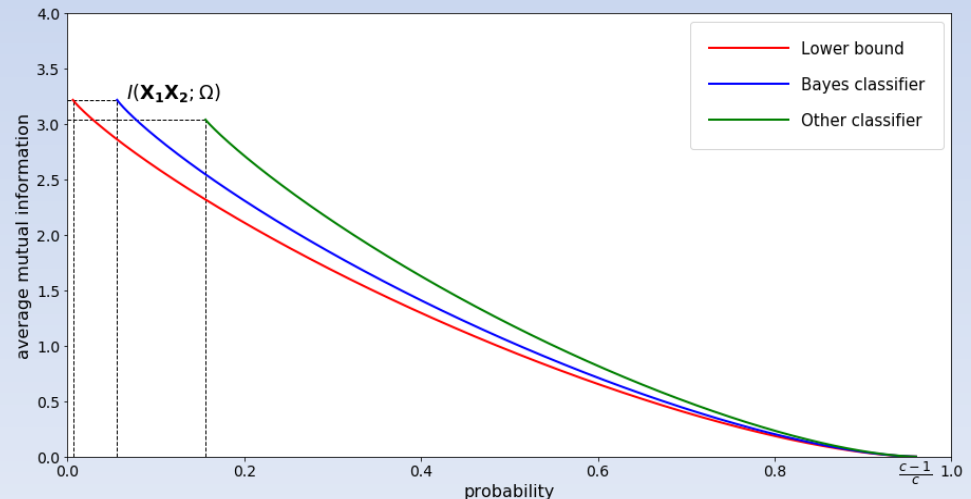
Вероятность ошибки решающего алгоритма

$$E_{Q_s^M}(\mathbf{X}^M, \hat{\Omega}) = 1 - \sum_{\mathbf{X}^M \in \mathbf{X}^M} P(\mathbf{X}^M) \max_{j=1}^c Q_s(\omega_j | \mathbf{X}^M) = \varepsilon_{G_s^M}$$

Избыточность вероятности ошибки алгоритма относительно нижней границы

$$\Delta \varepsilon = \varepsilon_{G_s^M} - \varepsilon$$

при условии $R_{G_s^M}(\varepsilon_{G_s^M}) = R_L^M(\varepsilon)$



10. Основные результаты и выводы

- Предложено обменное соотношение между количеством информации и вероятностью ошибки, которое дает нижнюю границу вероятности ошибки классификации для заданного множества объектов.
- Соотношение не зависит от решающего алгоритма и аналогично функции «скорость погрешность» (Rate Distortion Function) для схемы кодирования сообщений с допустимой погрешностью при наличии канала наблюдения с шумом.
- Построена нижняя граница для соотношения «количество информации вероятность ошибки», которая является обобщением нижней границы Шеннона в случае бесшумного канала наблюдения.
- Используя метрики на множествах древовидных представлений лиц и подписей, построены условные по классам распределения этих объектов и численные реализации нижних границ для соотношений «количество информации вероятность ошибки».
- Используя обобщенную метрику на ансамбле данных от источников различной модальности, построены условные по классам распределения составных объектов «лицо – подпись» и численная реализация нижней границы на ансамбле лиц и подписей.
- Нижняя граница для соотношения «количество информации вероятность ошибки» позволяет оценить избыточность вероятности ошибки алгоритмов с различными разделяющими функциями.