Информационный критерий для сравнения метрических классификаторов на ансамбле источников

М.М. Ланге

Федеральный исследовательский центр "Информатика и управление" РАН

11- я Международная конференция "Интеллектуализация обработки информации"
(ИОИ-11, Испания, г. Барселона, 10-14 октября 2016 г.)

1. Содержание доклада

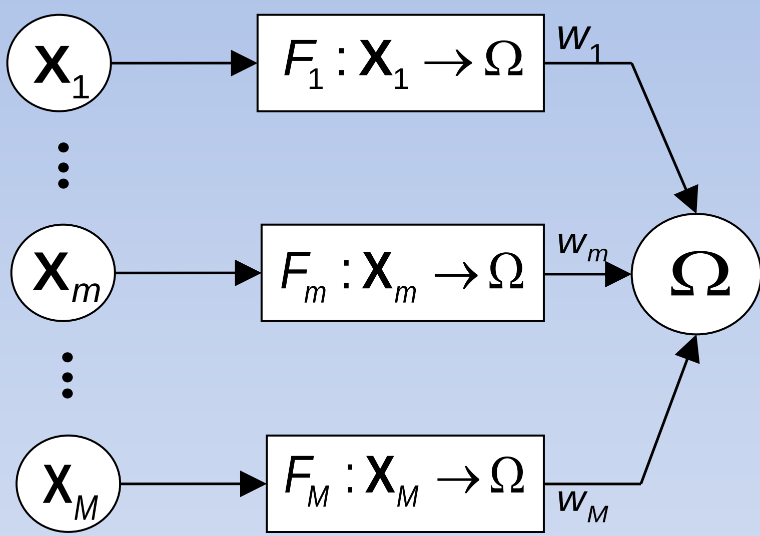
- Схемы классификации на ансамбле источников с использованием голосования решений по источникам (**Majority Voting**) и обобщенной меры (**General Measure**)
- Средняя взаимная информация ансамбля источников относительно множества классов как характеристика эффективности классификатора
- Связь с задачей кодирования источников с допустимой погрешностью при наличии канала наблюдения с шумом
- Меры различия на множествах объектов отдельных источников и на ансамбле. Условные по классам плотности распределения
- Функционалы средней взаимной информации и их оценки для MV и GM классификаторов
- Основной результат - соотношение средних взаимных информаций для MV и GM классификаторов
- Экспериментальные результаты распознавания лиц по цветным изображениям в трехкомпонентной модели HSI (на ансамбле из трех источников)

2. Схемы классификаторов на ансамбле источников

Множество классов: $\Omega = \{\omega_1, \dots, \omega_c\}, c \geq 2$ **с вероятностями** $p(\omega_i): \sum_{i=1}^{c} p(\omega_i) = 1$

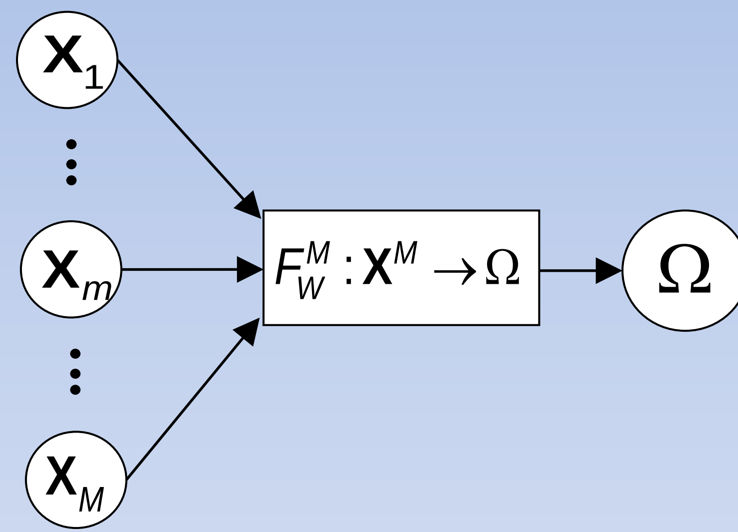
Ансамбль источников : $\mathbf{X}^M = (\mathbf{X}_1, \dots, \mathbf{X}_M)$ **с весами** $W = \{w_m > 0, m = 1, \dots, M\}$

MV классификатор



Голосование решений
по отдельным источникам *

GM классификатор



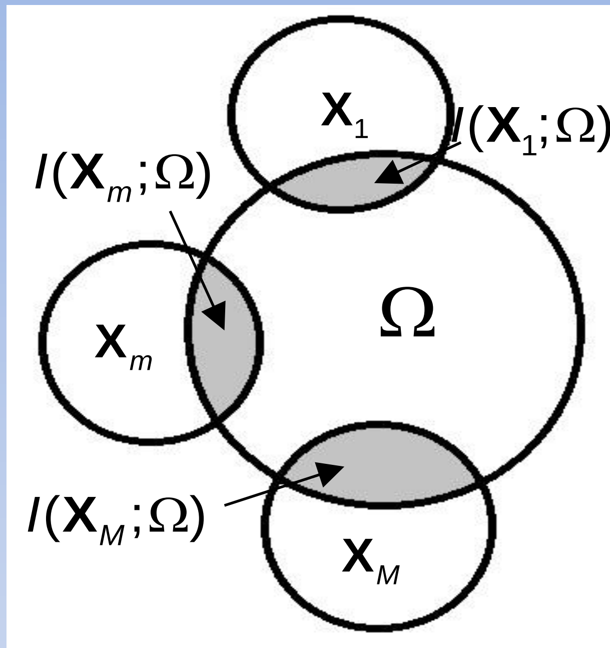
Комплексирование источников
по обобщенной мере на ансамбле * *

* L.I.Kuncheva. *Combining Pattern Classifiers* // Wiley and Sons, 2004

** М.М. Ланге, С.Н. Ганебных, А.М. Ланге. *Многоклассовое распознавание образов в пространстве древовидных представлений с многоуровневым разрешением* // Машинное обучение и анализ данных, 2016, 2(1), с.70-88

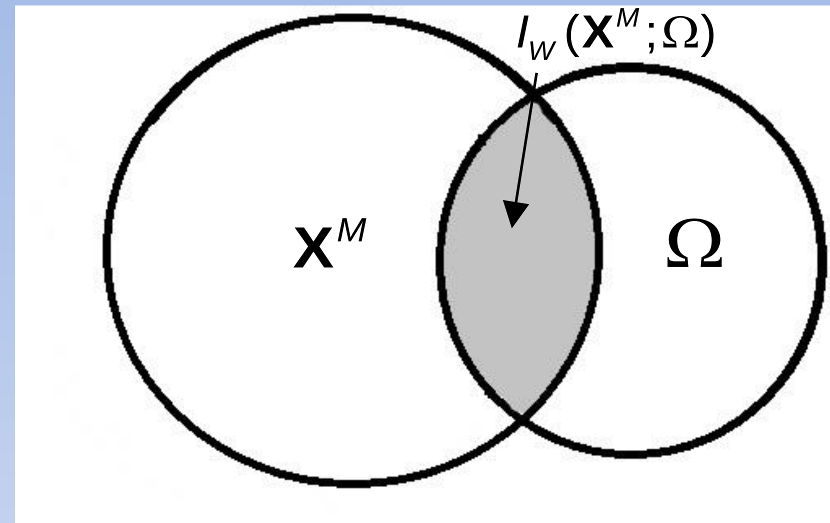
3. Средняя взаимная информация классификатора

MV классификатор



$$I_W^{\mathrm{MV}}(\mathbf{X}^M;\Omega)=\sum_{m=1}^{M} I(X_m;\Omega)\frac{w_m}{\sum_{m=1}^{M} w_m}$$

GM классификатор



$$I_W^{\mathrm{GM}}(\mathbf{X}^M;\Omega)=\frac{1}{M}\; I_W(\mathbf{X}^M;\Omega)$$

Требуется показать справедливость следующего неравенства

$$\max_W I_W^{\mathrm{MV}}(\mathbf{X}^M;\Omega)\le I_{W^*}^{\mathrm{GM}}(\mathbf{X}^M;\Omega), \qquad \text{где} \qquad W^*=\arg\max_W I_W^{\mathrm{MV}}(\mathbf{X}^M;\Omega)$$

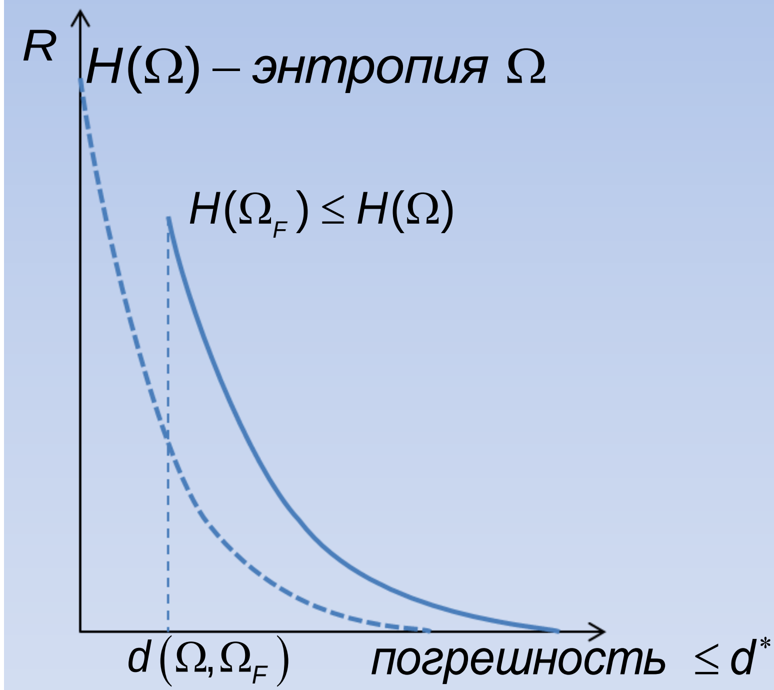
* *Information-Theoretic Models* - In book Simon Haykin. *Neural Networks. A Comprehensive Foundation* // Practice Hall International, Inc., 1998

4. Связь с задачей кодирования источников

$$\Omega \Rightarrow \boxed{\text{канал с шумом}} \Rightarrow \mathbf{X} \Rightarrow \boxed{\text{блок оценивания}} \Rightarrow \Omega_F \Rightarrow \boxed{\text{кодер}} \Rightarrow \hat{\Omega}_F$$

$\|\Omega_F\| = \|\Omega\|,\ \|\hat{\Omega}_F\| \le \|\Omega_F\|$ (кодер с погрешностью (<), кодер без погрешности (=))

Функция скорость - погрешность (*Rate - distortion function*)*



Функция Добрушина-Цыбакова (сплошная линия)

Функция Шеннона (пунктирная линия)

Средняя погрешность:

$$d(\Omega, \hat{\Omega}_F) = \frac{1}{N} \underset{\Omega, \hat{\Omega}_F}{E} \sum_{n=1}^{N} [\omega(n) \ne \hat{\omega}(n)],\ \omega \in \Omega, \hat{\omega} \in \hat{\Omega}_F$$

$$d(\Omega, \hat{\Omega}_F) \le d(\Omega, \Omega_F) + d(\Omega_F, \hat{\Omega}_F)$$

Средняя взаимная информация: $I(\mathbf{X}; \hat{\Omega}_F) \ge 0$

Функция скорость-погрешность

для допустимых значений $d^* \ge d(\Omega, \Omega_F)$:

$$R(d^*) = \min_{\hat{\Omega}_F} I(\mathbf{X}; \hat{\Omega}_F)$$

$$\hat{\Omega}_F : d(\Omega_F, \hat{\Omega}_F) \le d^* - d(\Omega, \Omega_F)$$

* Р.Л. Добрушин, Б.С.Цыбаков. *Передача информации с дополнительным шумом* // Проблемы передачи информации, 1963, 14, с. 21 - 42

5. Меры различия объектов

Мера на множестве $\mathbf{X}_m$, $\mathbf{x}_m=(x_{m1},\dots,x_{mN_m})\in\mathbf{X}_m$, $\hat{\mathbf{x}}_m=(\hat{x}_{m1},\dots,\hat{x}_{mN_m})\in\mathbf{X}_m, m=1,\dots,M$:

$$d(\mathbf{x}_m,\hat{\mathbf{x}}_m)=\sum_{n=1}^{N_m}\frac{(x_{mn}-\hat{x}_{mn})^2}{\sigma_{mn}^2}\ ,\ 0<\sigma_{mn}^2<\infty$$

Мера на ансамбле $\mathbf{X}^M=(\mathbf{X}_1,\dots,\mathbf{X}_M)$, $\mathbf{x}^M=(\mathbf{x}_1,\dots,\mathbf{x}_M)\in\mathbf{X}^M$, $\hat{\mathbf{x}}^M=(\hat{\mathbf{x}}_1,\dots,\hat{\mathbf{x}}_M)\in\mathbf{X}^M$:

$$D(\mathbf{x}^M,\hat{\mathbf{x}}^M)=\sum_{m=1}^{M}w_m d(\mathbf{x}_m,\hat{\mathbf{x}}_m)\ ,\ w_m>0$$

Ядра по мерам $d(\mathbf{x}_m,\hat{\mathbf{x}}_m)$ **и** $D(\mathbf{x}^M,\hat{\mathbf{x}}^M)$:

Простое ядро

$$K^{(d)}(\mathbf{x}_m,\hat{\mathbf{x}}_m)=\frac{e^{-d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}}{\int_{\mathbf{X}_m}e^{-d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}d\mathbf{x}_m},\qquad 0<\int_{\mathbf{X}_m}e^{-d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}d\mathbf{x}_m<\infty$$

$$K_W^{(D)}(\mathbf{x}^M,\hat{\mathbf{x}}^M)=\frac{e^{-D(\mathbf{x}^M,\hat{\mathbf{x}}^M)}}{\int_{\mathbf{X}^M}e^{-D(\mathbf{x}^M,\hat{\mathbf{x}}^M)}d\mathbf{x}^M}=\prod_{m=1}^{M}K_{w_m}^{(d)}(\mathbf{x}_m,\hat{\mathbf{x}}_m)$$

Взвешенное ядро

$$K_{w_m}^{(d)}(\mathbf{x}_m,\hat{\mathbf{x}}_m)=\frac{e^{-w_m d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}}{\int_{\mathbf{X}_m}e^{-w_m d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}d\mathbf{x}_m},\qquad 0<\int_{\mathbf{X}_m}e^{-w_m d(\mathbf{x}_m,\hat{\mathbf{x}}_m)}d\mathbf{x}_m<\infty$$

6. Условные по классам плотности

Наборы эталонов по классам (для m **- го источника)** :

$$X_{mi} = \{\mathbf{x}_{mik} \in X_m, k = 1,\dots,K_{mi}\}, i = 1,\dots,c$$

Условные по классам плотности для m **- го источника в виде смесей** :

$$g(\mathbf{x}_m \mid \omega_i) = \sum_{k=1}^{L_{mi}} \theta_{mik} K^{(d)}(\mathbf{x}_m, \mathbf{x}_{mik}),\ i = 1,\dots,c,$$

$$\theta_{mik} > 0 : \sum_{k=1}^{L_{mi}} \theta_{mik} = 1$$

Условные по классам плотности для ансамбля в виде M **- кратных смесей** :

$$g_W(\mathbf{x}^M \mid \omega_i) = \prod_{m=1}^{M} \sum_{k=1}^{L_{mi}} \theta_{mik} K^{(d)}_{w_m}(\mathbf{x}_m, \mathbf{x}_{mik})$$

$$= \prod_{m=1}^{M} g_{w_m}(\mathbf{x}_m \mid \omega_i), \quad i = 1,\dots,c$$

7. Функционал средней взаимной информации*

Априорное распределение на множестве классов Ω :

$$p(\omega_i) > 0 : \sum_{i=1}^{c} p(\omega_i) = 1$$

Взвешенная плотность распределения на множестве $\mathbf{X}_m$, $m=1,\dots,M$:

$$p_{w_m}(\mathbf{x}_m) = \sum_{i=1}^{c} p(\omega_i) g_{w_m}(\mathbf{x}_m \mid \omega_i)$$

Средняя взаимная информация множества $\mathbf{X}_m$ **относительно множества** Ω **по взвешенным плотностям** $p_{w_m}(\mathbf{x}_m)$ и $g_{w_m}(\mathbf{x}_m \mid \omega_i)$:

$$I_{w_m}(\mathbf{X}_m;\Omega) = H_{w_m}(\mathbf{X}_m) - H_{w_m}(\mathbf{X}_m \mid \Omega)$$

Дифференциальные энтропии:

$$H_{w_m}(\mathbf{X}_m) = -\frac{1}{N_m} \int_{\mathbf{X}_m} p_{w_m}(\mathbf{x}_m) \log p_{w_m}(\mathbf{x}_m) d\mathbf{x}_m$$

$$H_{w_m}(\mathbf{X}_m \mid \Omega) = -\frac{1}{N_m} \sum_{i=1}^{c} p(\omega_i) \int_{\mathbf{X}_m} g_{w_m}(\mathbf{x}_m \mid \omega_i) \log g_{w_m}(\mathbf{x}_m \mid \omega_i) d\mathbf{x}_m$$

* R.G. Gallager. *Information Theory and Reliable Communication* // Wiley and Sons, 1968.

8. Средняя взаимная информация MV и GM классификаторов

MV классификатор

$$I_W^{\mathrm{MV}}(\mathbf{X}^M;\Omega)=\sum_{m=1}^{M} I(\mathbf{X}_m;\Omega)\frac{w_m}{\sum_{m=1}^{M} w_m}\ ,$$

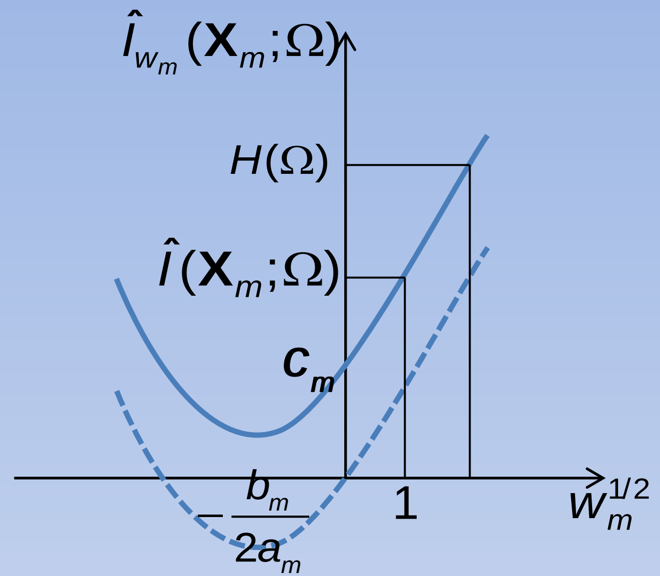
где $$I(\mathbf{X}_m;\Omega)=I_{w_m=1}(\mathbf{X}_m;\Omega),\ m=1,\dots,M$$

GM классификатор

$$I_W^{\mathrm{GM}}(\mathbf{X}^M;\Omega)=\frac{1}{M}\sum_{m=1}^{M} I_{w_m}(\mathbf{X}_m;\Omega)$$

9. Оценки функционалов взаимной информации

Оценки по источникам $(m = 1,\dots,M)$



$$I_{w_m}(\mathbf{X}_m;\Omega) \le \hat{I}_{w_m}(\mathbf{X}_m;\Omega) = a_m w_m + b_m w_m^{1/2} + c_m$$

$$I(\mathbf{X}_m;\Omega) \le \hat{I}(\mathbf{X}_m;\Omega) = a_m + b_m + c_m$$

Коэффициенты $a_m > 0, b_m > 0$ и $c_m \ge 0$ зависят от параметров гауссовых смесей по классам (наборов эталонов, дисперсий и весов компонентов смеси)

Веса источников $w_m(s,v) = (1+v)^2\, e^{s\frac{\hat{I}(\mathbf{X}_m;\Omega)}{H(\Omega)}}$ с параметрами $0 \le s \le 1, \quad v \ge 0$

Оценки средней взаимной информации по ансамблю источников

MV классификатор: $$\hat{I}_s^{\mathrm{MV}}(\mathbf{X}^M;\Omega) = \sum_{m=1}^{M} \hat{I}(\mathbf{X}_m;\Omega) \frac{w_m(s,v)}{\sum_{m=1}^{M} w_m(s,v)}$$

GM классификатор: $$\hat{I}_{s,v}^{\mathrm{GM}}(\mathbf{X}^M;\Omega) = \frac{1}{M} \sum_{m=1}^{M} \hat{I}_{s,v}(\mathbf{X}_m;\Omega)$$

10. Основной результат

Лемма. При $s \to 0$ существует значение $s^* \geq 0$, доставляющее

$$\max_s \hat{I}_s^{\mathrm{MV}}(\mathbf{X}^M;\Omega) = \hat{I}_{s^*}^{\mathrm{MV}}(\mathbf{X}^M;\Omega) \approx (\mu + s^*\Delta)H(\Omega),$$

где

$$\mu = \frac{1}{M}\sum_{m=1}^{M}\frac{\hat{I}(\mathbf{X}_m;\Omega)}{H(\Omega)} \quad \text{и} \quad \Delta = \frac{1}{M}\sum_{m=1}^{M}\left(\frac{\hat{I}(\mathbf{X}_m;\Omega)}{H(\Omega)}\right)^2 - \left(\frac{1}{M}\sum_{m=1}^{M}\frac{\hat{I}(\mathbf{X}_m;\Omega)}{H(\Omega)}\right)^2 \geq 0$$

и $s^* = 0$ при $\Delta = 0$ (при одинаковых значениях взаимной информации источников).

Теорема. При $s \to 0$ значения $s^* \geq 0$ и $v^* \geq s^*\Delta\, H(\Omega) \Big/ \frac{1}{M}\sum_{m=1}^{M}(2a_m + b_m)$,

удовлетворяющие условию $a_m w_m(s^*, v^*) + b_m w_m^{1/2}(s^*, v^*) + c_m \leq H(\Omega)$, $m = 1, \ldots, M$,

гарантируют неравенство

$$\hat{I}_{s^*}^{\mathrm{MV}}(\mathbf{X}^M;\Omega) \leq \hat{I}_{s^*,v^*}^{\mathrm{GM}}(\mathbf{X}^M;\Omega),$$

которое выполняется со знаком равенства в случае $w_m(s^*, v^*) = 1$, $m = 1, \ldots, M$.

11. Результаты распознавания лиц по цветным изображениям в модели HSI

Доли ошибок и стандартные отклонения

sources ensemble	error rate deviation	NN	MT	SVM
face: H	ε	0.015	0.008	0.006
	σ	0.005	0.006	0.003
face: S	ε	0.017	0.012	0.009
	σ	0.006	0.003	0.004
face: I	ε	0.022	0.012	0.017
	σ	0.006	0.005	0.006
face: HSI (GM)	ε	0.007	0.002	0.001
	σ	0.005	0.002	0.001
face: HSI (MV)	ε	0.010	0.005	0.005
	σ	0.005	0.003	0.005

Данные :

Источники H,S и I по 1000 объектов от каждого источника

(25 классов по 40 изображений в классе)

Схема эксперимента :

Скользящий контроль по схеме 200 кратного разбиения данных на обучающую и тестовую выборки

Разделяющие функции на основе :

NN - ближайшего эталона (nearest neighbor)

MT - смеси эталонов (mixture of templates)

SVM - опорных векторов (support vector machine)

12. Результаты и выводы

- Предложен критерий сравнения метрических классификаторов на ансамбле источников в терминах средней взаимной информации ансамбля относительно множества классов.

- Рассмотрены две схемы классификации: на основе голосования решений по отдельным источникам (Majority Voting) и на основе комплексирования источников с использованием обобщенной меры на ансамбле (General Measure).

- Для MV и GM классификаторов введены функционалы средней взаимной информации и получены их оценки.

- Показано, что при одинаковых наборах весовых коэффициентов источников максимальная величина средней взаимной информации MV классификатора не превосходит средней взаимной информации GM классификатора.

- Эксперименты на цветных изображениях лиц в модели HSI , образующей ансамбль из трех источников, продемонстрировали меньшую долю ошибок GM классификатора по сравнению с MV классификатором при использовании различных разделяющих функций.