

Мультимодальное селективное комбинирование классификаторов для выделения полезного сигнала на энцефалограмме при быстром потоковом предъявлении визуальных раздражителей

Бухонов Сергей

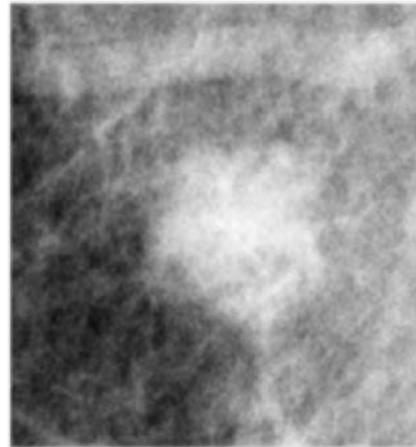
Московский физико-технический институт
Кафедра интеллектуальных систем

Научный руководитель: д.т.н. Моттль В.В.

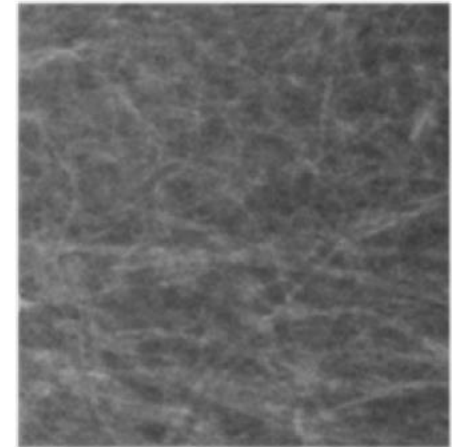
Применение ИМК для обнаружения маммограмм с патологиями



Регистрация ЭЭГ в процессе
быстро сменяющихся
изображений маммограмм



Пример целевого
объекта
(маммограмма с
патологией)



Пример нецелевого
объекта
(маммограмма без
патологией)

Постановка задачи

Исходные данные: $X = \{X_1, \dots, X_n\}$

$$Y = \{y_1, \dots, y_n\}, y_i \in \{-1, 1\}$$

$$X_i = \begin{pmatrix} x_{1,1} & \cdots & x_{1,m} \\ \vdots & \ddots & \vdots \\ x_{66,1} & \cdots & x_{66,m} \end{pmatrix}, m = 1100$$

Задача: $a: X \rightarrow Y$

Традиционный SVM

Решающее правило в виде разделяющей гиперплоскости:

$$d(\mathbf{x}; \mathbf{a}, b) = \mathbf{a}^T \mathbf{x} + b \quad \begin{cases} \geq 0 \Rightarrow \hat{y}(\mathbf{x}) = +1, \\ < 0 \Rightarrow \hat{y}(\mathbf{x}) = -1, \end{cases}$$

$$\begin{cases} \sum_{i=1}^m a_i^2 + C \sum_{j=1}^N \delta_j \rightarrow \min(\mathbf{a}, b, \boldsymbol{\delta}), \\ y_j (\sum_{i=1}^m a_i x_{ij} - b) \geq 1 - \delta_j, \quad j = 1, \dots, N, \\ \delta_j \geq 0, \quad j = 1, \dots, N, \end{cases}$$

Предварительный результат

Номер электрода	ROC score
1	0.6858
6	0.6425
9	0.5714
18	0.5844
20	0.5214
27	0.748
37	0.7203

Проблема:

избыточность информации может привести к переобучению

Цель работы

Повышение качества классификации путем:

1. Прореживание данных (сокращение признакового пространства)
2. Добавление критерия отбора наиболее информативных признаков
3. Реализация композиции классификаторов

Предварительная обработка

До предварительной обработки

$$\mathbf{x} = (x_i \in R, i = 1, \dots, m), \quad m = 1100$$

Сглаживание

$$x'_i = \frac{1}{w} \sum_{k=i-\lfloor w/2 \rfloor}^{i+\lfloor w/2 \rfloor} x_k, \quad i = \lfloor w/2 \rfloor + 1, \dots, m' - \lfloor w/2 \rfloor$$

SVM с отбором наиболее информативных отсчетов

Задача нахождения оптимальной разделяющей гиперплоскости

$$\left\{ \begin{array}{l} J_{SF SVM}(a_1, \dots, a_m, b, \delta_1, \dots, \delta_N, C, \mu) = \sum_{i=1}^m q(a_i | \mu) + C \sum_{j=1}^N \delta_j \rightarrow \min(\mathbf{a}, b, \boldsymbol{\delta}), \\ q(a_i | \mu) = \begin{cases} 2\mu |a_i| & \text{if } |a_i| \leq \mu, \\ \mu^2 + a_i^2 & \text{if } |a_i| > \mu, \end{cases} \\ y_j (\mathbf{a}^T \mathbf{x}_j + b) \geq 1 - \delta_j, \quad \delta_j \geq 0, \quad j=1, \dots, N. \end{array} \right.$$

μ - степень селективности
критерия при выборе
информативных отсчетов

Результаты

Номер электрода	Без селективности и сглаживания (ROC score)	С селективностью и сглаживанием (ROC score)
1	0.6858	0.6985
6	0.6425	0.6806
9	0.5714	0.6267
18	0.5844	0.6629
20	0.5214	0.6316
27	0.748	0.8289
37	0.7203	0.8451

Дальнейшая работа

- 1) Повышение качества классификации путем введения композиции классификаторов
- 2) Тестирование на других экспериментальных данных

Спасибо за внимание!