

Министерство науки и высшего образования
Российской Федерации

«Московский физико-технический институт
(государственный университет)»

Физтех-школа прикладной математики и информатики

Факультет управления и прикладной математики

Кафедра «Интеллектуальные системы»

Бухонов Сергей Александрович

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

Отчет о научно-исследовательской работе

Введение

Классификация сигналов энцефалограммы головного мозга применяется при решении многих задач, одна из которых - построение интерфейса “мозг-компьютер”. В настоящее время данный подход применим также и к области, исследующей обнаружение целевых объектов на изображении. Одной из задач данной области является обнаружение маммограмм с патологиями, в которой предметом анализа является мозговая активность специалиста (испытуемого) при предъявлении ему снимков маммограмм. Генерируемая мозговая активность при узнавании целевых объектов (маммограмм с патологиями) обычно содержит отклик, который не наблюдается при предъявлении нецелевых снимков.

Выявление целевых откликов сигналов ЭЭГ возможно благодаря методологии RSVP - быстрого последовательного визуального предъявления, которое формирует с каждого электрода одинаковое количество отсчетов, рассматриваемое как признаковое пространство. Однако из-за большого количества электродов (66) размер признакового пространства становится довольно большим, то приводит к низкой обобщающей способности алгоритмов распознавания.

Целью данной работы является построение оптимального классификатора для последующей селекции информативных сигналов энцефалограммы, записанных в ответ на визуальный стимул.

Постановка задачи

Имеются записи сигналов энцефалограммы головного мозга испытуемого в ответ на предъявленные визуальные стимулы. Построить комбинированный классификатор на основе SVM с различными видами регуляризации.

Описание исходных данных

Рассматриваемые объекты представляют собой записи сигналов ЭЭГ эксперта в области маммологии, зарегистрированные с 66 электродов. Каждому представленному изображению соответствует фрагмент ЭЭГ длиной 1100 отсчетов по каждому электроду.

$$(X_i, y_i), i = 1, \dots, N$$

где $X_i \in R^{1100 \times 66}$ – записи сигналов ЭЭГ, $y_i \in \{-1, 1\}$ – метка класса, к которому относится изображение (целевое или нецелевое)

Предварительная обработка

Для повышения скорости работы алгоритма осуществлено прореживание данных с шагом 11. Для уменьшения влияния шумов проведено сглаживание скользящим окном ширины 11.

Иллюстрация сигнала до обработки

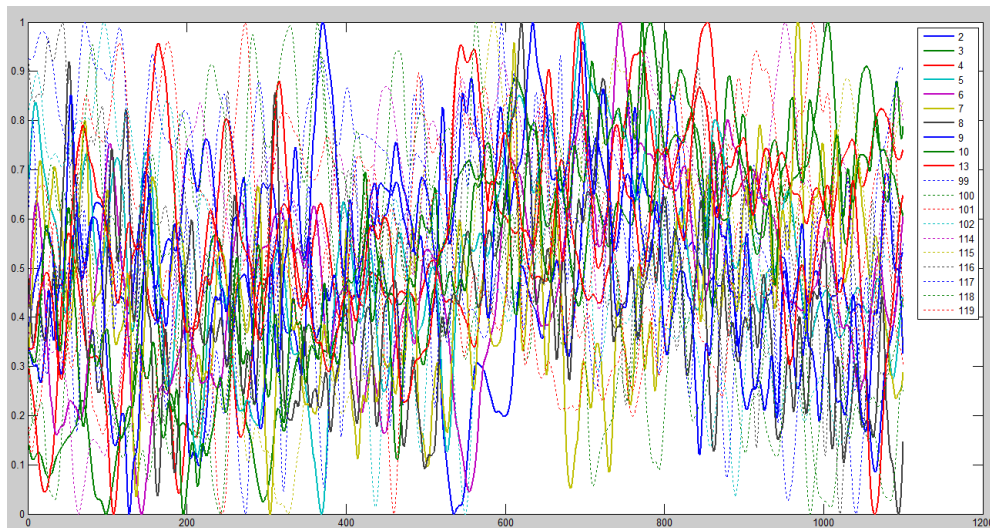
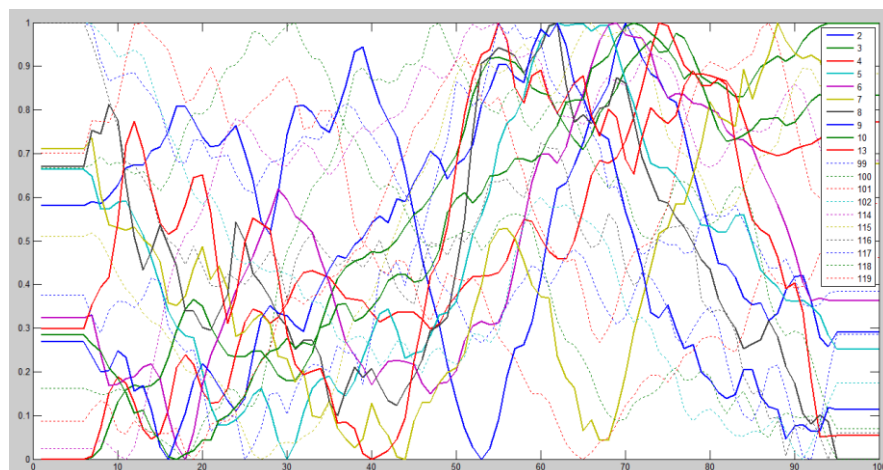


Иллюстрация сигнала после обработки



Классическая постановка задачи классификации

Для решения задачи классификации сигналов энцефалограммы был выбран SVM. Задача построения модели классификации сводится к нахождению коэффициентов разделяющей гиперплоскости, наилучшим образом разделяющей два класса.

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

Где $\mathbf{a} = [a_1, \dots, a_n]$ – направляющий вектор гиперплоскости, b – смещение гиперплоскости вдоль направляющего вектора, δ_j – величина ошибки на j -ом объекте, C – настраиваемый параметр, определяющий степень влияния ошибок.

SVM с требованием гладкости решающего правила

В предположении, что отдельные признаки объекта в составе вектора \mathbf{x} представляют собой результат упорядоченного изменения одного из свойств и что соседние признаки несут практически идентичную информацию о принадлежности объекта к тому или иному классу, разумно в процессе обучения отдавать предпочтения решающим правилам с плавным изменением коэффициентов линейной части. Для этого вводим дополнительную аддитивную составляющую

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

где γ – коэффициент гладкости решающего правила.

SVM с регулируемой селективностью

Чем больше электродов, тем больше размер признакового пространства объекта. Это может привести к проблеме переобучения. Также, не со всех электродов может поступать полезная информация, помогающая в принятии решения. Следовательно, необходим инструмент, который позволит автоматически отбирать электроды, сигналы с которых являются наиболее информативными.

Оптимизационная задача после введения данного типа регуляризации выглядит следующим образом

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

Решение данной оптимизационной задачи эквивалентно решению двойственной задачи

$$\begin{cases} L \lambda_1, \dots, \lambda_N | C, \mu = \sum_{j=1}^N \lambda_j - \sum_{i \in I} \frac{\xi_i}{2} \rightarrow \max(\lambda_1, \dots, \lambda_N), \\ \xi_i \geq 0, \quad \xi_i \geq \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \lambda_j \lambda_l - \mu^2, \quad i \in I = \{1, \dots, m\}, \\ \sum_{j=1}^N y_j \lambda_j = 0, \quad 0 \leq \lambda_j \leq \frac{C}{2}, \quad j = 1, \dots, N, \end{cases}$$

Которая может быть выражена в виде

$$\begin{cases} \hat{a}_i = \sum_{j: \hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_i(\omega_j), \\ i \in I^+ = i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l > \mu^2, \\ \hat{a}_i = \hat{\eta}_i \sum_{j: \hat{\lambda}_j > 0} y_j \hat{\lambda}_j x_i(\omega_j), \\ i \in I^0 = i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l = \mu^2, \\ \hat{a}_i = 0, \quad i \in I^- = i \in I: \sum_{j=1}^N \sum_{l=1}^N y_j y_l x_{ij} x_{il} \hat{\lambda}_j \hat{\lambda}_l < \mu^2, \end{cases}$$

Таким образом, оптимальная гиперплоскость может быть выражена в виде:

$$\sum_{j:\lambda_j>0} y_j \lambda_j \sum_{i \in I^+} x_{ij} x_i + \sum_{i \in I^0} \eta_i x_{ij} x_i + b \geq 0,$$

где числовые параметры $\{0 \leq \eta_i \leq 1, i \in I^0; b\}$ определяются решением задачи линейного программирования:

$$\begin{cases} 2\mu^2 \sum_{i \in I^0} \eta_i + C \sum_{l=1}^N \delta_j \rightarrow \min(\eta_i, i \in I^0; b; \delta_1, \dots, \delta_N), \\ \sum_{i \in I^0} \left(\sum_{l=1}^N y_j y_l x_{ij} x_{il} \lambda_l \right) \eta_i + y_j b + \delta_j \geq 1 - \sum_{i \in I^+} \left(\sum_{l=1}^N y_j y_l x_{ij} x_{il} \lambda_l \right), \\ \delta_j \geq 0, \quad j = 1, \dots, N, \quad 0 \leq \eta_i \leq 1, \quad i \in I^0. \end{cases}$$

Построение комбинированного классификатора

После построения модели классификатора ее необходимо проверить на качество классификации. В таком случае разумно построить комбинированный классификатор, состоящий из композиции классификаторов. Один из способов построения комбинированного классификатора является голосование по большинству.

Данный способ заключается в присваивании объекту такой метки класса, предпочтение которой отдали наибольшее количество классификаторов. В качестве классификаторов примем описанные ранее классический SVM, SVM с требованием гладкости решающего правила и SVM с селективным отбором признаков.

Комбинированный классификатор для случая двухклассового распознавания имеет вид.

$$y = a(x) = \text{sign} \left(\sum_{i=1}^n a_i(x) \right)$$

где $a_i(x)$ – ответ i -ого алгоритма на поданный на вход объект x .

Эксперименты

Ниже представлены значения AUC для разных электродов с учетом и без учета сглаживания и требования гладкости решающего правила.

Сглаживание					Сглаживание						
-					-						
+					+						
Требование					Требование						
гладкости					гладкости						
-					-						
+					+						
Номер электрода	1	0,6858	0,6855	0,657	0,6985	Номер электрода	34	0,5012	0,5036	0,4781	0,4944
	2	0,6431	0,6192	0,6009	0,6611		35	0,5012	0,576	0,5493	0,5558
	3	0,6644	0,6758	0,6308	0,6901		36	0,6348	0,6711	0,63	0,6985
	4	0,6683	0,6473	0,6411	0,689		37	0,7203	0,8257	0,7364	0,8451
	5	0,6582	0,6174	0,6124	0,6167		38	0,6085	0,696	0,6259	0,6997
	6	0,6425	0,6791	0,6292	0,6806		39	0,7062	0,6931	0,6811	0,7431
	7	0,6091	0,5954	0,5827	0,5816		40	0,5897	0,5857	0,5464	0,547
	8	0,6766	0,6778	0,6242	0,6873		41	0,6088	0,655	0,6291	0,6742
	9	0,5714	0,6161	0,5782	0,6267		42	0,7708	0,8584	0,7815	0,8672
	10	0,6053	0,555	0,5384	0,5746		43	0,5378	0,5366	0,5268	0,5518
	11	0,5509	0,5613	0,5724	0,576		44	0,545	0,5642	0,5664	0,5735
	12	0,5929	0,5478	0,5504	0,5593		45	0,5924	0,6715	0,6281	0,6961
	13	0,5368	0,5254	0,5541	0,5589		46	0,6428	0,7447	0,6951	0,7671
	14	0,5295	0,5405	0,5226	0,5397		47	0,5491	0,6056	0,5669	0,5759
	15	0,5595	0,6175	0,5876	0,5891		48	0,5628	0,6101	0,5775	0,5869
	16	0,5191	0,5157	0,5171	0,5119		49	0,5713	0,5866	0,5371	0,5455
	17	0,5756	0,5609	0,563	0,5891		50	0,4918	0,5419	0,5268	0,5388
	18	0,5844	0,5601	0,5798	0,6629		51	0,5852	0,603	0,575	0,6005
	19	0,6222	0,5994	0,5969	0,6692		52	0,5622	0,5515	0,5829	0,5787
	20	0,5214	0,6301	0,5672	0,6316		53	0,6943	0,7688	0,7287	0,7855
	21	0,6377	0,6699	0,6202	0,6944		54	0,5702	0,6184	0,5608	0,5666
	22	0,5103	0,508	0,5191	0,5108		55	0,5238	0,5485	0,5354	0,5248
	23	0,5471	0,5877	0,5473	0,5952		56	0,6548	0,6617	0,6275	0,6805
	24	0,616	0,6358	0,585	0,6594		57	0,6458	0,6638	0,6382	0,6146
	25	0,488	0,4853	0,5191	0,5177		58	0,7606	0,7137	0,6748	0,7351
	26	0,4931	0,4864	0,5142	0,5072		59	0,7382	0,7115	0,6733	0,6547
	27	0,748	0,8083	0,7277	0,8289		60	0,7493	0,7234	0,6854	0,7472
	28	0,7385	0,8305	0,7112	0,8351		61	0,5537	0,5049	0,5399	0,5127
	29	0,6363	0,6808	0,6716	0,663		62	0,6219	0,6625	0,5983	0,6011
	30	0,7717	0,8177	0,7439	0,8252		63	0,7336	0,7465	0,7009	0,7453
	31	0,5568	0,5786	0,5422	0,5428		64	0,6066	0,6001	0,5831	0,5783
	32	0,6661	0,6812	0,644	0,7248		65	0,4933	0,4814	0,5129	0,5021
	33	0,7652	0,8308	0,7312	0,8518		66	0,6191	0,6232	0,5492	0,5475
	все	0,764	0,805	0,815	0,811						

Далее представлены предварительные результаты работы комбинированного классификатора для классического SVM и SVM с требованием гладкости решающего правила.

Электрод	AUC	Электрод	AUC
1	0.7309	34	0.5709
2	0.7091	35	0.6436
3	0.7345	36	0.7091
4	0.6836	37	0.7964
5	0.6691	38	0.7491
6	0.7018	39	0.7527
7	0.7164	40	0.68
8	0.6764	41	0.7164
9	0.6545	42	0.8109
10	0.6036	43	0.6255
11	0.6364	44	0.6764
12	0.5818	45	0.7055
13	0.68	46	0.7309
14	0.6436	47	0.6364
15	0.6436	48	0.6836
16	0.6618	49	0.6909
17	0.6364	50	0.6582
18	0.6982	51	0.72
19	0.7018	52	0.6655
20	0.6764	53	0.7309
21	0.6764	54	0.6582
22	0.6582	55	0.6255
23	0.6909	56	0.7164
24	0.6945	57	0.6655
25	0.6291	58	0.7491
26	0.6618	59	0.7418
27	0.7491	60	0.7745
28	0.8145	61	0.6545
29	0.7673	62	0.68
30	0.7964	63	0.76
31	0.6873	64	0.6618
32	0.7636	65	0.5382
33	0.7673	66	0.6436

Дальнейшая работа:

- 1) Провести расчет AUC комбинированного классификатора с учетом SVM с селективным отбором признаков
- 2) Разработать содержание
- 3) Доработать теоретический материал.