

Задачи классификации ЭКГ сигналов

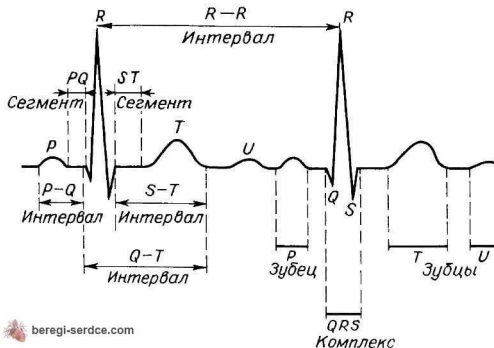
Кибитова Валерия

Спецсеминар "Алгебра над алгоритмами и эвристический поиск закономерностей"

1 декабря 2016 г.

Определение ЭКГ

Электрокардиограмма (ЭКГ) – это сигнал, отображающий электрическую активность сердца в течении определенного периода времени.



По расположенным между зубцами интервалам P-QRS-T судят о прохождении импульса по сердечным отделам.

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

BCP(HRV)

Вариабельность сердечного ритма (BCP)– это сигнал, который получается следующим образом:

- Измеряются RR интервалы (t_{RR}) секунд)
- Каждая величина преобразовывается в удар/минуту $60/t_{RR}$

Задача авторов:

Определить по кардиограмме ишемическую болезнь сердца. (Ишемическая болезнь – болезнь, которая развивается при недостаточном поступлении кислорода к сердечной мышце по коронарным артериям. Проявлениями ИБС могут быть: стенокардия, инфаркт миокарда, аритмии сердца, а также внезапная сердечная смерть.)

Данные:

10 больных и 10 здоровых пациентов, для каждого пациента 1000 кардиограмм. Частота дискретизации составляла 500 Гц.

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Предобработка сигнала

- Фильтр Баттерворта (верхних частот с частотой подавления сигнала 0,3 Гц)
- Фильтр Баттерворта (нижних частот с частотой подавления сигнала 15 Гц)
- Медианный фильтр, при помощи которого извлекается тренд. Тренд, который был получен при помощи медианного фильтра вычитается из основного сигнала.

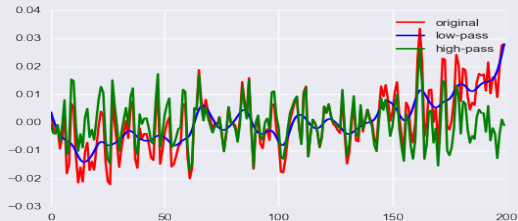
Фильтр Баттерворта

Это такие фильтры, которые проектируются так, чтобы его амплитудно-частотная характеристика (АЧХ показывает, во сколько раз амплитуда сигнала на выходе системы отличается от амплитуды на входе) была максимально гладкой на частотах полосы пропускания.

Фильтр нижних частот (верхних частот) подавляет сигналы выше(ниже) заданной частоты, и пропускает сигналы ниже(выше) заданной частоты.

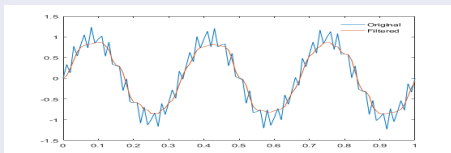
NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Пример работы фильтров Баттерворда



Медианный фильтр

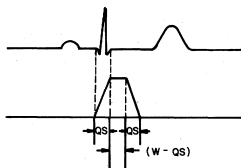
Оконный фильтр возвращающий медианное значение внутри окна;



NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Выделение R пиков с помощью алгоритма Томпкинса. Алгоритм Томпкинса состоит из:

- Фильтр нижних частот
- Фильтр верхних частот
- Операция дифференцирования (подавляет P и T, усиливает высокочастотные компоненты QRS)
- Возведение в квадрат
- Интегрирующий фильтр, который сглаживает пики в пределах одного QRS комплекса



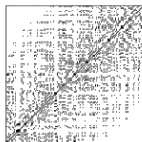
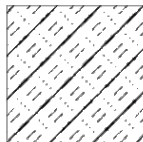
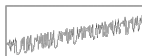
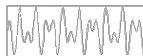
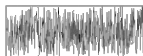
- Сканирование фрагмента сигнала, в котором ожидается R-пик на поиск максимального значения

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Признаки основанные, на Recurrence Plot

$$R(i,j) = \begin{cases} 1 & \text{if } \|x(i) - x(j)\| < \varepsilon \\ 0 & \text{otherwise} \end{cases}$$



NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Признаки основанные, на Recurrence Plot

График строится для HRV сигнала.

- $REC = \frac{1}{N^2} \sum_{i,j=0}^N R_{i,j}$ – плотность точек на графике
- $DET = \frac{\sum_{l=l_{min}}^N IP(l)}{\sum_{i,j}^N R_{i,j}}$, $P(l)$ – число диагоналей длины l , – процент точек, которые формируют диагональные линии
- $L_{mean} = \frac{\sum_{l=l_{min}}^N l P(l)}{\sum_{i,j}^N P(l)}$ – средняя длины диагонали

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Poincare Plot

Для сигнала $x_t, x_{t+1}, x_{t+2} \dots$ на графике отображаются точки $(x_t, x_{t+1}), (x_{t+1}, x_{t+2})$ и т. д. В данном случае точками являются RR интервалы.

Признаки, основанные на Poincare Plot

- Отклонение точек графика от прямой $y = x$. Описывает мгновенную вариативность RR интервалов.
- Отклонение точек графика от прямой $y = -x - 2RR_{mean}$, RR_{mean} – среднее значение RR интервалов. Описывает долговременную вариативность RR интервалов.

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Detrended fluctuation analysis

- $X(t) = \sum_{i=1}^k [X(i) - X_{mean}]$. Данные сегментируются с окном размера Δn
- На каждом сегменте для данных находится полином, который наиболее точно их представляет (обычно линейный) и находится следующая функция

$$F(\Delta n) = \sqrt{\left(\frac{1}{N} \sum_{t=1}^N [x(t) - x_{\Delta n}(t)]^2\right)}$$

Признаки, основанные на Detrended fluctuation analysis

$F(\Delta n)$ для RR интервалов, затем на логарифмических шкалах строится зависимость $\log F(\Delta n)$ от $\log(\Delta n)$.

Признак: угол наклона линии $\log F(\Delta n)$ к $\log(\Delta n)$.

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Приближенная Энтропия (Approximate entropy)

Мера, используемая для определения числа закономерностей и непредсказуемостей в сигнале.

- Необходимо зафиксировать целое m и действительное r
- $x(i) = x_i, x_{i+1}, \dots, x_{i+m-1}$
- $C_i^m = \frac{|\{x: d[x(i), x(j)] \leq r\}|}{N-m+1}$, где $d[x, x^*] = \max_a |x_a - x_a^*|$
- $\Phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \log(C_i^m(r))$
- $ApEn = \Phi^m(r) - \Phi^{m+1}(r)$

Признак

Приближенная энтропия для HRV, при $r = 0.2SD$, $m = 10$

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Выборочной энтропия

Выборочная энтропия описывает меру иррегулярности данных.

- $x_m(i) = x_i, x_{i+1}, \dots, x_{i+m-1}$
- $A = |\{x : d[x_{m+1}(i), x_{m+1}(j)] < r\}|$
- $B = |\{x : d[x_m(i), x_m(j)] < r\}|$
- $SampEn = -\log \frac{A}{B}$

Признак

Выборочная энтропия для HRV, при $r = 0.2SD$, $m = 0, 1, \dots, e - 1$, где $B(0) = N$

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Энтропия Шеннона

- Энтропия Шеннона в случае сигналов использует нормализованную спектральную плотность мощности сигнала - функцию распределения мощности сигнала в зависимости от частоты

$$\tilde{S}_{xx}(\omega) = \frac{(\Delta t)^2}{T} \left| \sum_{n+1}^N x_n e^{-i\omega n} \right|^2; \tilde{S}_{xx}^{norm}(\omega) = \frac{S_{xx}(\omega)}{\sum_{\omega} S_{xx}(\omega)}$$

, где Δt - интервал сэмплирования

$T = N * \Delta t$ - полный интервал измерения сигнала

x_n - дискретный сигнал.

Энтропия Шеннона считается как сумма всех произведений мощности определенной частоты s_{ω}^{norm} и логарифма обратной величины к этой мощности.

$$ShanEn = \sum_f s_{\omega}^{norm} \log(1/s_{\omega}^{norm})$$

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Группы признаков

Short-range scaling exponent

Аналогично Detrended fluctuation analysis, только ищется не полином, а линейная функция по методу наименьших квадратов.

Correlation dimension

- $x_m(i) = x_i, x_{i+1}, \dots, x_{i+m-1}$
- $g = |x : d[x_{m+1}(i), x_{m+1}(j)] < \varepsilon$
- $C(r) = \frac{g}{N^2}$
-

$$D2 = \lim_{r \rightarrow 0} \frac{\log C(r)}{\log(r)}$$

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

Отбор признаков

- t-тест стьюдента
- Метод главных компонент

Table 1. Range of variance significant nonlinear features for the normal and CAD groups.

	Features	Normal (Mean \pm SD)	CAD (Mean \pm SD)	p-Value
1	SD2	64.8 \pm 30.1	41.2 \pm 26.4	<0.0001
2	L _{mean}	12.7 \pm 6.67	23.0 \pm 17.5	<0.0001
3	REC	34.9 \pm 11.1	49.5 \pm 18.9	<0.0001
4	DET	98.0 \pm 1.42	99.0 \pm 1.13	<0.0001
5	ShanEn	3.22 \pm 0.388	3.64 \pm 0.560	<0.0001
6	ApEn	1.16 \pm 0.188	1.01 \pm 0.228	<0.0001
7	SampEn	1.53 \pm 0.335	1.22 \pm 0.476	<0.0001
8	α_1	1.14 \pm 0.235	0.865 \pm 0.383	<0.0001
9	D2	2.36 \pm 1.44	0.349 \pm 0.416	<0.0001

Table 2. Range of variance of six principal components obtained using 0-1 normalization and PCA accounting for 95% of variance.

Principal components	Normal (Mean \pm SD)	CAD (Mean \pm SD)
1	-0.01889 \pm 0.320712	-0.48805 \pm 0.349908
2	0.262432 \pm 0.232432	0.011774 \pm 0.225043
3	-1.34019 \pm 0.154243	-1.13456 \pm 0.266746
4	0.034841 \pm 0.19559	0.031956 \pm 0.223662
5	-0.173088 \pm 0.132831	0.255909 \pm 0.124357
6	-0.41429 \pm 0.111811	-0.45782 \pm 0.102175

NOVEL CLASSIFICATION OF CORONARY ARTERY DISEASE USING HEART RATE VARIABILITY ANALYSIS

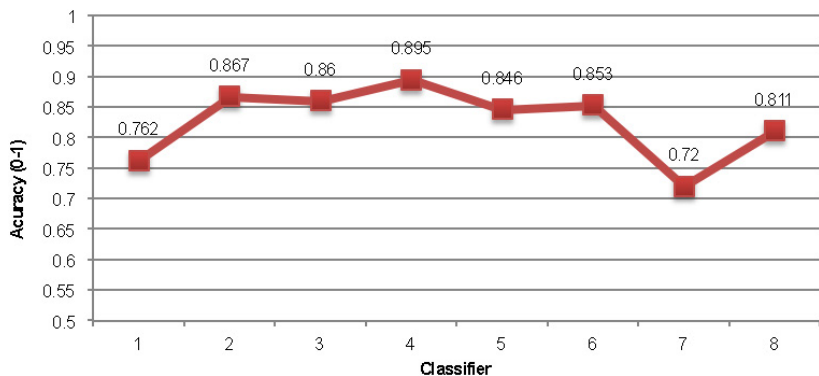


Fig. 5. Classification accuracies: 1: BN; 2: NB; 3: LibSVM; 4: ANN-MLP; 5: RBF; 6: SMO; 7: random forest; 8: CART.

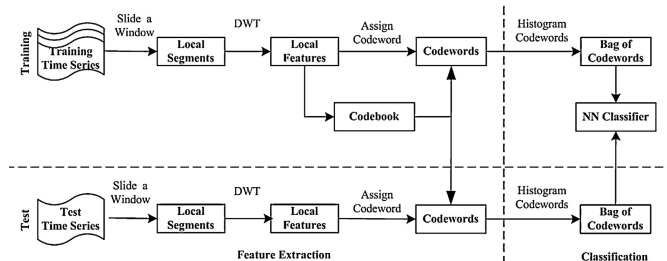
Bag-of-words representation for biomedical time series classification

Задача

Идентификация человека

Данные

- 40 человек. Частота дискретизации 250 Гц. 50 кардиограмм на каждого человека
- 15 человек. 100 кардиограмм на каждого



Локальная сегментация

Окно заданной длины p .

Нормализация каждого локального сегмента.

Для каждого локального сегмента берутся его аппроксимирующие коэффициенты дискретного вейвлет вейвлет-преобразования.

Дискретное вейвлет-преобразование

Вейвлеты – это обобщенное название временных функций, имеющих вид короткой волны той или иной формы.

Вейвлет-преобразование – разложение по базису, состоящему из вейвлетов.

$$\int_{-\infty}^{+\infty} \Psi(t) dt = 0 \quad \int_{-\infty}^{+\infty} |\Psi(t)|^2 dt < \infty$$

$$y[n] = (x * g)[n] = \sum_{k=-\infty}^{\infty} x[k]g[n - k]$$

$$y_{\text{low}}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n - k] \text{ – коэффициенты аппроксимации}$$

$$y_{\text{high}}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n - k] \text{ – коэффициенты детализации}$$

Bag-of-words representation for biomedical time series classification

- Формирование кодовых слов. Кластеризация k-means. $B = b_1, b_2, \dots, b_k$
- $c = \operatorname{argmin}_j d(b_j, x_i)$
- Признаки: мешок слов
- Классификация: 1-NN

Виды мер расстояния между гистограммами:

- Евклидова метрика

$$D_{L_2}(h, k) = \left(\sum_i |h(i) - k(i)|^2 \right)^{1/2}$$

- Расстояние гистограмм по пересечению

$$D_{HI}(h|k) = 1 - \sum_i \min(h(i), k(i))$$

h, k – нормализованные гистограммы. Чем больше гистограммы пересекаются, тем меньше будет расстояние гистограмм по пересечению.

- Хи-квадрат метрика

$$D_{\chi^2}(h, k) = \sum_i \frac{|h(i) - k(i)|^2}{h(i) + k(i) + \epsilon}$$

ϵ - малое число для избежания деления на ноль.

Хи-квадрат метрика нормирует расстояние между двумя элементами гистограммы на суммарное количество элементов в этой корзине.

Позволяет уменьшить влияние на меру часто используемых стоп-слов, различное употребление использование которых не несет много информации.

- Расстояние Дженсена-Шеннона

Гистограмму можно сравнить с помощью методов сравнения дискретных случайных - например дивергенции Кульбаха-Лейбнера.

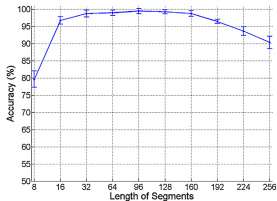
$$D_{KL}(h|k) = \sum_i h(i)(\log_2 h(i) - \log_2 k(i))$$

Расстояние Дженсена-Шеннона использует дивергенцию Кульбаха-Лейбнера и составляет симметричную относительно параметров формулу.

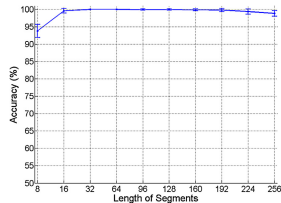
$$D_{JS} = \frac{1}{2}(D_{KL}(h|k) + D_{KL}(k|h))$$

Bag-of-words representation for biomedical time series classification

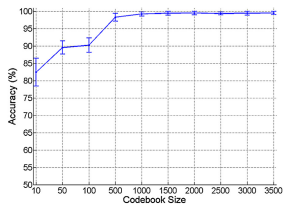
Результаты



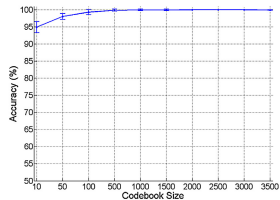
(a) ecg-15



(b) ecg-40



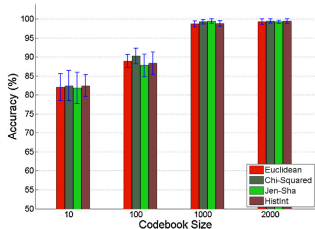
(c) ecg-15



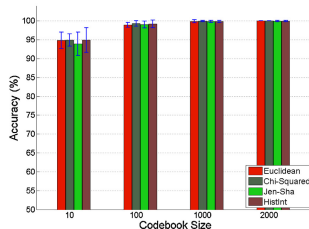
(d) ecg-40

Bag-of-words representation for biomedical time series classification

Результаты



(e) ecg-15



(f) ecg-40