

Сравнение информационного анализа электрокардиосигналов и анализа вариабельности сердечного ритма как методов диагностики заболеваний внутренних органов

Целых Влада Руслановна • МФТИ
Воронцов Константин Вячеславович • МФТИ, ВЦ РАН
Успенский Вячеслав Максимилианович • Медицинский
учебно-научный клинический центр им. П.В.Мандрыка

ММО-17, 19 – 25 сентября 2015 года

1 Анализ ВСП и информационный анализ ЭКГ-сигналов

- Анализ ВСП
- Символьное кодирование

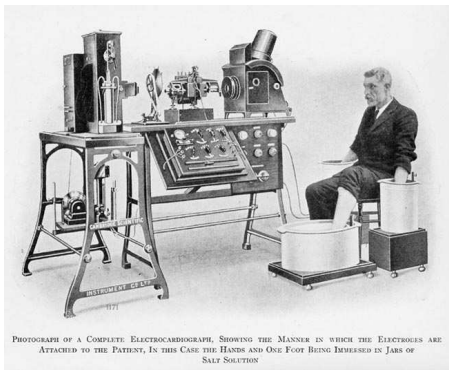
2 Линейные модели классификации

- Настройка параметров алгоритмов
- Результаты работы алгоритмов

3 Эксперименты на других данных

- РТВ
- CardioQVARK

Электрокардиография



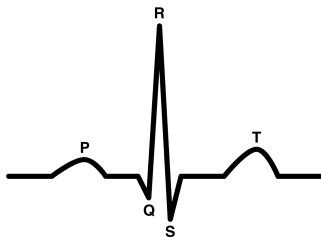
1872 — первые записи электрической активности сердца

1911 — коммерческий электрокардиограф (фото)

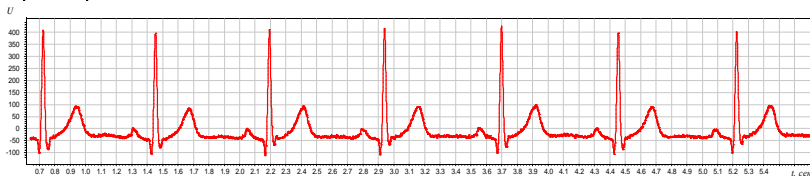
1924 — нобелевская премия по медицине, Виллем Эйнтховен

Электрокардиограмма (ЭКГ)

Пики (зубцы) ЭКГ:

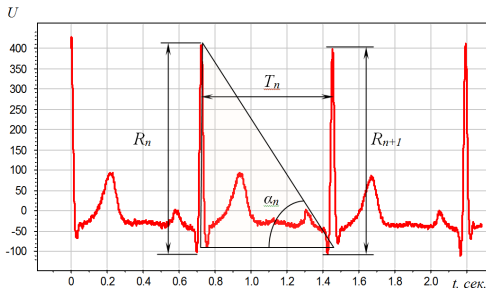


Пример ЭКГ:



Диагностика заболеваний по ЭКГ

Технология информационного анализа (В. М. Успенский, 2008) заключается в кодировании приращений амплитуд $R_{n+1} - R_n$, интервалов $T_{n+1} - T_n$ и углов $\alpha_{n+1} - \alpha_n$ ($\alpha_n = \arctg R_n/T_n$).



Анализ variability сердечного ритма (Р. М. Баевский, 2001) использует агрегированные данные об интервалах между R-пиками.

Цели исследования

Повысить качество диагностики путем подбора способов кодирования и оптимизации моделей классификации.
Сравнить информационный анализ ЭКГ-сигналов с анализом variability сердечного ритма (ВСР).

Данные

название	аббревиатура	код МКБ-10	объем
<i>абсолютно здоровые</i>	АЗ		193
анемия железододефицитная	ЖДА	D50	260
аденома простаты	ДГПЖ	N40	260
аднексит хронический	АХ	N70	276
вегетососудистая дистония	ВСД	F45.3	694
гипертоническая болезнь	ГБ	I11	1894
некроз головки бедренной кости	НГБК	M91.1	324
гастродуоденит гиперацидный	ХГ1	K29	324
гастродуоденит гипоацидный	ХГ2	K29	700
дискинезия ЖВП	ДЖВП	K83	717
желчнокаменная болезнь	ЖКБ	K80	278
ишемическая болезнь сердца	ИБС	I20	1265
мочекаменная болезнь	МКБ	N20	654
миома матки	ММ	D25	781
рак общий	РО	C00-C97	530
сахарный диабет	СД	E10-E11	871
узловой зоб щитовидный железы	УЩ	D34	748
холецистит хронический	ХХ	K81.1	340
язвенная болезнь	ЯБ	K25-K28	785

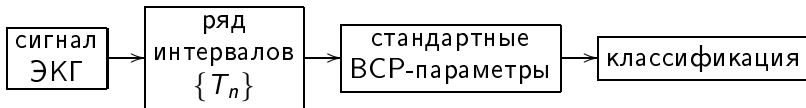
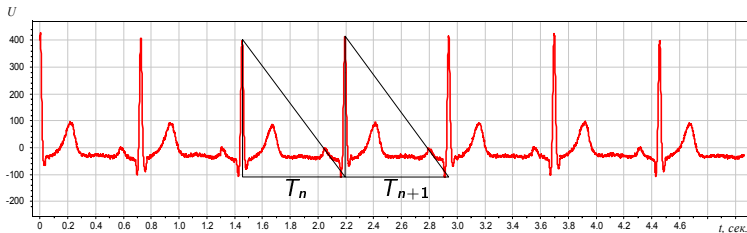
Оценивание качества классификации

Функционал качества классификации “здоровые” — “больные”

Area Under Curve (AUC) — площадь под ROC-кривой, отображающей зависимость чувствительности (доли правильно классифицируемых больных) от специфичности (доли правильно классифицируемых здоровых).

Анализ ВСР

Анализ вариабельности сердечного ритма (ВСР, HRV) — распространенный метод диагностики заболеваний по ЭКГ (ишемии, инфаркта миокарда, гипертонии и т. д.).



Стандартные параметры ВСР-анализа

Методы временной области

- $AVNN = \frac{1}{N} \sum_{n=1}^N T_n$;
- $SDNN = \sqrt{\frac{1}{N-1} \sum_{n=1}^N (T_n - \bar{T}_n)^2}$;
- $RMSSD = \sqrt{\frac{1}{N-1} \sum_{n=1}^{N-1} (T_n - T_{n+1})^2}$;
- $pNN50 = \frac{1}{N} \sum_{n=1}^{N-1} [|T_{n+1} - T_n| > 50\text{мс}]$;
- TI — длина основания треугольника, аппроксимирующего плотность распределения интервалов;
- HRVTi — триангулярный индекс ВСР.

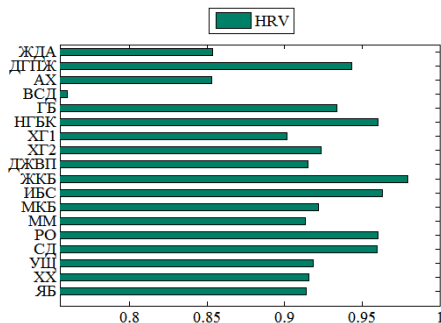
Методы частотной области

- VLF, VLF% — абсолютная и относительная мощность спектра в частотном диапазоне 0 – 0.04 Гц;
- LF, LF% — в диапазоне 0.04 – 0.15 Гц;
- HF, HF% — в диапазоне 0.15 – 0.4 Гц;
- TP — мощность спектра;
- LF/HF, $LF_{nu} = LF / (LF + HF)$;
- VLF_{peak} , LF_{peak} , HF_{peak} — частоты макс. вершин спектра.

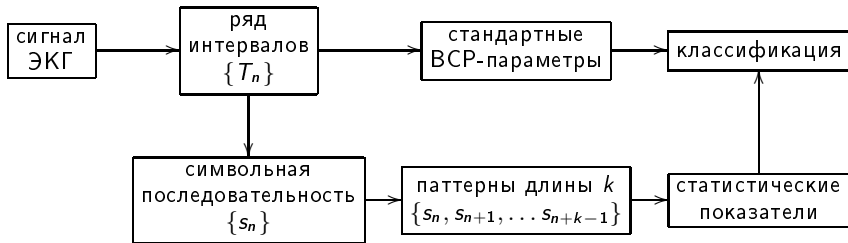
Значения AUC при кросс-валидации

Признаки: 18 параметров стандартного ВСР-анализа

Алгоритм: случайный лес



Символьное кодирование



Способы кодирования интервалов

- По значению интервала T_n :
 - Kurths (1995), Wessel (2000)

$$s_n = \begin{cases} A, & \mu < T_n \leq (1 + a)\mu \\ B, & (1 + a)\mu < T_n \leq \infty \\ C, & (1 - a)\mu < T_n \leq \mu \\ D, & 0 < T_n \leq (1 - a)\mu, \end{cases}$$

где $\mu = \bar{T}_n$, $a = 0.05$.

- Porta (2007)

$$s_n = \begin{cases} A, & \min(T_n) \leq T_n < \min(T_n) + 1 \cdot l \\ B, & \min(T_n) + 1 \cdot l \leq T_n < \min(T_n) + 2 \cdot l \\ \dots & \\ F, & \min(T_n) + 5 \cdot l \leq T_n < \max(T_n), \end{cases}$$

где $l = (\max(T_n) - \min(T_n))/6$.

Способы кодирования интервалов

- По разности соседних значений $dT_n = T_{n+1} - T_n$:
 - Betterman (2000), Yang (2003)

$$s_n = \begin{cases} A, & dT_n \geq 0 \\ B, & dT_n < 0 \end{cases}$$

- Wessel (2000)

$$s_n = \begin{cases} A, & |dT_n| \geq \delta \\ B, & |dT_n| < \delta \end{cases}$$

- Cammarota (2007)

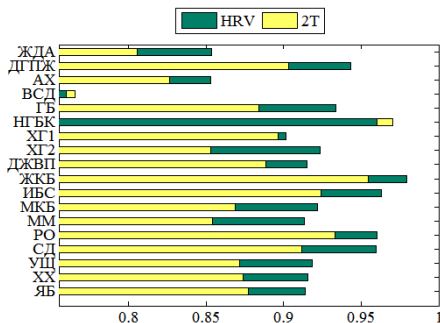
$$s_n = \begin{cases} A, & dT_n > \delta \\ B, & dT_n < -\delta \\ C, & |dT_n| \leq \delta \end{cases}$$

Обычно полагают $\delta = 10$ мс.

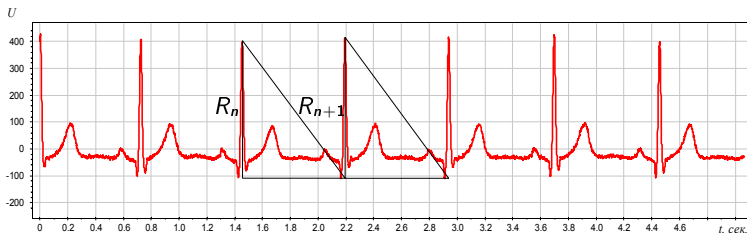
Символьное кодирование интервалов (2T)

Признаки: частоты k -грамм при кодировании ряда интервалов, $k = 6$

Алгоритм: случайный лес

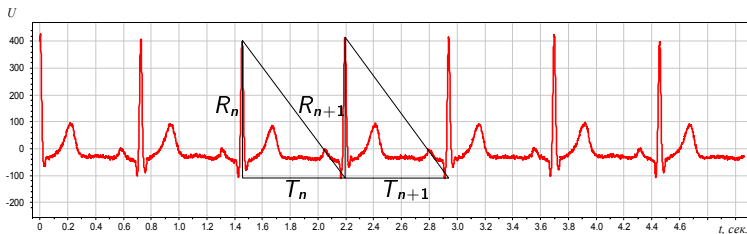


Амплитуды R -пиков



Chasal (2003): дополнительный учет амплитуд повысил точность диагностики синдрома сонного апноэ на 5%.

Кодирование интервалов и амплитуд (4RT)

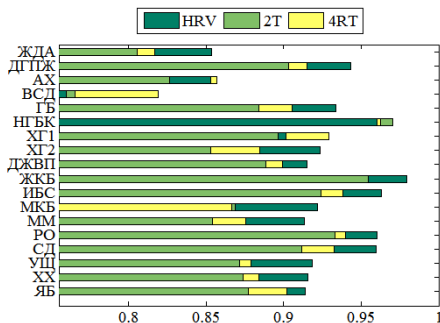


	A	B	C	D
$R_{n+1} - R_n$	+	-	+	-
$T_{n+1} - T_n$	+	-	-	+

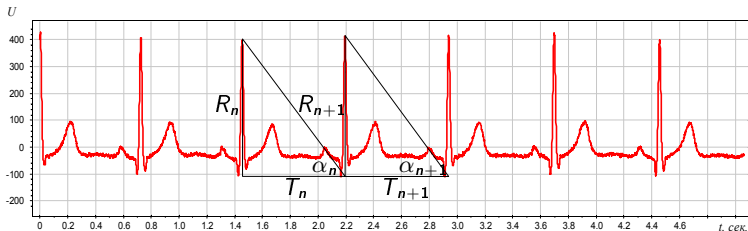
Кодирование интервалов и амплитуд (4RT)

Признаки: частоты k -грамм при кодировании амплитуд и интервалов, $k = 3$

Алгоритм: случайный лес



Кодирование интервалов, амплитуд и их отношений (информационный анализ ЭКГ-сигналов, 6RTA)



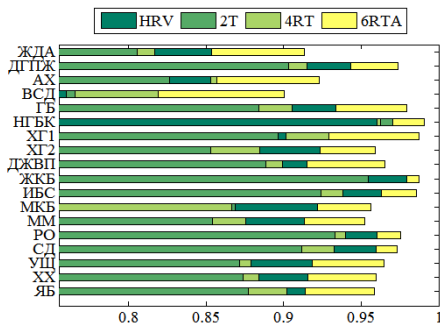
	A	B	C	D	E	F
$R_{n+1} - R_n$	+	-	+	-	+	-
$T_{n+1} - T_n$	+	-	-	+	+	-
$\alpha_{n+1} - \alpha_n$	+	+	+	-	-	-

$$\alpha_n = \arctg \frac{R_n}{T_n}$$

Кодирование интервалов, амплитуд и их отношений (6RTA)

Признаки: частоты k -грамм при кодировании амплитуд, интервалов и их отношений, $k = 2$

Алгоритм: случайный лес



Результаты на кросс-валидации, AUC (%)

название	HRV	2T	4RT	6RTA	Δ
анемия железodefицитная	85,3	80,5	81,7	91,3	6,0
аденома простаты	94,3	90,3	91,5	97,4	3,0
аднексит хронический	85,3	82,6	85,7	92,3	7,0
вегетососудистая дистония	76,0	76,5	81,9	90,1	14,1
гипертоническая болезнь	93,3	88,4	90,5	97,9	4,6
некроз головки бедренной кости	96,0	97,0	96,2	99,1	3,1
гастродуоденит гиперацидный	90,1	89,7	92,9	98,7	8,6
гастродуоденит гипоацидный	92,4	85,3	88,5	95,9	3,5
дискинезия ЖВП	91,5	88,8	89,9	96,5	5,0
желчнокаменная болезнь	98,0	95,4	95,5	98,7	0,8
ишемическая болезнь сердца	96,3	92,4	93,8	98,6	2,2
мочекаменная болезнь	92,2	86,9	86,6	95,6	3,4
миома матки	91,4	85,4	87,6	95,2	3,9
рак общий	96,0	93,3	94,0	97,5	1,5
сахарный диабет	96,0	91,1	93,3	97,3	1,4
узловой зоб щитовидный железы	91,8	87,2	87,9	96,5	4,7
холецистит хронический	91,5	87,4	88,4	95,9	4,4
язвенная болезнь	91,4	87,7	90,2	95,9	4,4

Проблема

трудная интерпретируемость модели случайного леса

Решение

линейная модель классификации

Линейные модели классификации

$\{S_i\}_{i=1}^{\ell}$ — обучающая выборка кодограмм (bRTA-кодирование),

$\{y_i\}_{i=1}^{\ell}$ — диагнозы: 0 = здоровый, 1 = больной

$f_w(S_i)$ — числовой признак, монотонно зависящий от частоты k -граммы w в кодограмме S_i

Варианты определения признаков

- $f_w(S) = p_w(S)$ — частота k -граммы w
- $f_w(S) = [p_w(S) \geq \theta]$ — встречаемость k -граммы w

Линейные модели классификации

$a(S) = [\sum_w \gamma_w f_w(S) \geq \beta]$, где γ_w — вес k -граммы w .

- Синдромный алгоритм
- Логистическая регрессия

Линейные модели классификации

Синдромный алгоритм

Задавались различные сочетания весов признаков и критерия информативности.

Логистическая регрессия

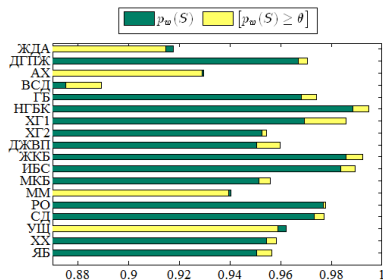
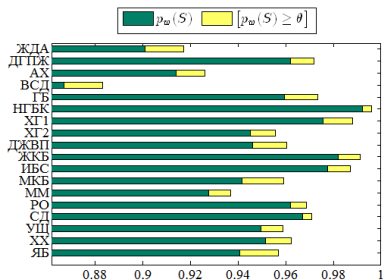
Рассматривались следующие методы снижения размерности:

- отбор K признаков с наибольшими значениями критерия информативности (средней частоты в классе больных, средней встречаемости в классе больных, ...);
- L_1 -регуляризация;
- метод главных компонент.

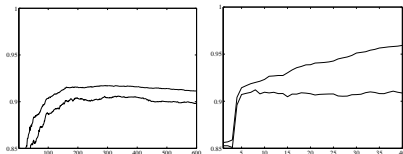
Линейные модели классификации

Синдромный алгоритм

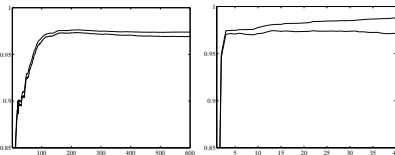
Логистическая регрессия



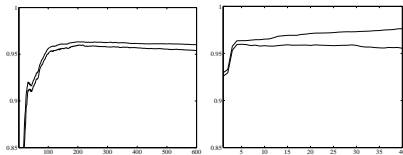
Зависимость AUC от числа признаков для SA и LR



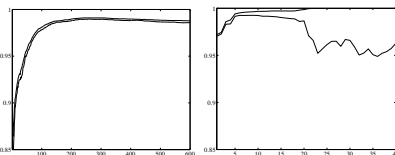
анемия



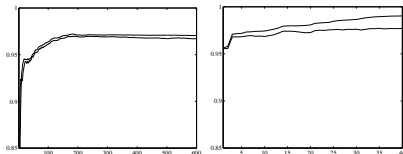
гипертония



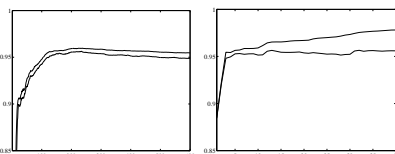
дискинезия ЖВП



желчнокаменная болезнь



сахарный диабет



язвенная болезнь

Значения AUC на контрольной выборке, %

болезнь	SA	LR	RF
некроз головки бедренной кости	98,9	99,3	99,4
желчнокаменная болезнь	98,4	97,4	98,6
ишемическая болезнь сердца	96,1	96,6	98,1
сахарный диабет	96,3	96,9	97,7
гастродуоденит гиперацидный	95,8	95,5	97,6
аденома простаты	95,7	97,0	97,4
гипертоническая болезнь	95,2	95,7	96,9
гастродуоденит гипоацидный	93,5	94,5	95,8
узловой зоб щитовидный железы	93,6	94,2	95,8
мочекаменная болезнь	93,9	93,2	95,5
рак общий	93,6	95,4	95,4
холецистит хронический	92,5	91,4	94,8
миома матки	93,7	92,4	94,6
дискинезия ЖВП	92,1	92,0	93,7
язвенная болезнь	91,9	92,8	93,2
аднексит хронический	90,3	89,7	90,2
анемия железодефицитная	84,0	85,2	89,0
вегетососудистая дистония	85,3	84,8	87,2

База данных РТВ

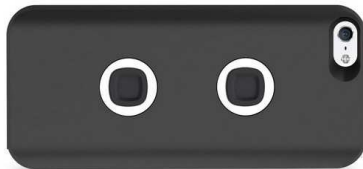
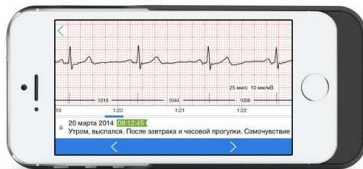
Данные:

открытая база данных по инфарктам миокарда РТВ
Национального Метрологического Института Германии,
содержащая 320 и 74 записей ЭКГ больных и здоровых
соответственно.

Результаты (AUC, %):

HRV	2T	4RT	6RTA	Δ
76.6	75.2	78.1	86.8	10.2

CardioQVARK



CardioQVARK

Данные:

база данных системы cardioQVARK; в роли болезни — признак курит/не курит. В выборке 2611 записей ЭКГ, из которых 26% принадлежат курящим людям.

Результаты (AUC, %):

HRV	2T	4RT	6RTA	Δ
86.4	81.8	87.1	87.8	1.4





При объединении признаков ВСП-анализа и частот k -грамм при 6RTA-кодировании AUC достигает 91.6%.

Основные результаты




- показана возможность диагностики заболеваний внутренних органов по ЭКГ;
- обоснован учет совместной вариабельности интервалов, амплитуд и их отношений при кодировании ЭКГ-сигнала;
- достигнуты высокие уровни чувствительности и специфичности для выбранных заболеваний.

Спасибо за внимание!

Список литературы I

-  Баевский Р. М., Иванов Г. Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения. *Ультразвуковая и функциональная диагностика*. 2001. № 3. С. 108–127.
-  Успенский В. М. Информационная функция сердца. Теория и практика диагностики заболеваний внутренних органов методом информационного анализа электрокардиосигналов. М.: Экономика и информатика, 2008. 116 с.
-  Успенский В. М. Информационная функция сердца. *Клиническая медицина*. 2008. Т. 86. № 5. С. 4–13.
-  J. Kurths, A. Voss, P. Saparin, A. Witt, H.J. Kleiner, and N.Wessel
Quantitative analysis of heart rate variability *Chaos* 1995. Vol. 5. No 1.
Pp. 88–94.

Список литературы II

-  Niels Wessel, Andreas Voss, Hagen Malberg, Christine Ziehmann, Henning U Voss, Alexander Schirdewan, Udo Meyerfeldt, Jürgen Kurths Nonlinear analysis of complex phenomena in cardiological data *Herzschrittmachertherapie und Elektrophysiologie* 2000. Vol. 11. Issue 3. Pp. 159–173.
-  Porta A., Tobaldini E., Guzzetti S., Furlan R., Montano N., Gneccchi-Ruscione T. Assessment of cardiac autonomic modulation during graded head-up tilt by symbolic analysis of heart rate variability *American Journal of Physiology - Heart and Circulatory Physiology* 2007. Vol. 293. No 1.
-  Bettermann H., Cysarz D. and Van Leeuwen P. Detecting cardiorespiratory synchronization by respiratory pattern analysis of heart period dynamics: the musical rhythm approach *International Journal of Bifurcation & Chaos* 2000. Vol. 10. No 10.

Список литературы III

-  Albert C.-C. Yang, Shu-Shya Hseu, Huey-Wen Yien, Ary L. Goldberger and C.-K. Peng Linguistic Analysis of the Human Heartbeat Using Frequency and Rank Order Statistics *Physical review letters* 2003. Vol. 90. No 10.
-  Camillo Cammarota, Enrico Rogora Time reversal, symbolic series and irreversibility of human heartbeat *Chaos, Solitons & Fractals* 2007. Vol. 32. No 5. Pp. 1649–1654.
-  Philip de Chazal, Conor Heneghan, Elaine Sheridan, Richard Reilly, Philip Nolan, and Mark O'Malley Automated Processing of the Single-Lead Electrocardiogram for the Detection of Obstructive Sleep Apnoea. *IEEE Transactions on Biomedical Engineering*. June, 2003. Vol. 50. No. 6. Pp. 686–696.
-  Boussejot R., Kreiseler D., Schnabel A. Nutzung der EKG-Signaldatenbank CARDIODAT der PTB über das Internet. *Biomedizinische Technik*. Band 40, Ergänzungsband 1. 1995. Pp. 317.
-  <http://cardioqvark.ru/>