

# ПРИМЕНЕНИЕ АНАЛИЗА ДАННЫХ ПРИ ОБРАБОТКЕ СИГНАЛА ЭКГ

Михальский А.И.

Институт проблем управления им. В.А. Трапезникова РАН, Москва, Россия  
ipuran@yandex.ru

*Аннотация. При анализе данных анализируется регистрируемый процесс – ЭКГ или ЭхоКГ сигнал без привлечения описания процесса его генерации в виде математической модели. Такой подход приводит к построению «модели» процесса. Поставленная задача решается путём анализа эмпирических сигналов, возникающих при различных патологиях.*

*Ключевые слова: экспериментальные данные, модель данных, электрокардиограмма, эхокардиограмма, оценка толщины стенки миокарда, коррекция интервала QT, диагностика не кардиологических заболеваний.*

## Введение

Метод математического моделирования широко применяется при анализе записей ЭКГ. При этом создаётся математическое описание объекта – сердца, генерирующего электрические потенциалы, регистрируемые аппаратом ЭКГ. В отличие от метода математического моделирования объекта для обработки регистрируемого сигнала ЭКГ, использование методов анализа данных позволяет обрабатывать сигнал без построения модели объекта. Речь идёт об использовании множества записей реального сигнала, полученных в различных ситуациях, и на их основании принятия решения о наличии патологии, заболевания, повышения риска развития неблагоприятного заболевания. Три примера из реальной практики иллюстрируют данный подход: оценка толщины стенки миокарда, коррекция интервала QT, диагностика не кардиологических заболеваний.

## 1. Оценка толщины миокарда по данным ЭхоКГ

Вторым, по распространенности генетически обусловленным заболеванием миокарда у детей является гипертрофическая кардиомиопатия [1]. Это заболевание приводит к снижению продолжительности жизни и к ухудшению ее качества. Ежегодная смертность от данной патологии, составляет 3–8% [2, 3]. Применение сдерживающей терапии предусматривает меры замедления увеличения толщины миокарда, и таким образом продлевает жизнь пациентам. Оценка критической толщины миокарда, при которой необходимо начать применение сдерживающей лекарственной терапии, является важной задачей, обеспечивающей успех лечения.

Современным методом обследования морфологических и функциональных изменений сердца и его клапанов является эхокардиография (ЭхоКГ). Проведение скринингового ЭхоКГ-исследования позволяет судить о толщине миокарда межжелудочковой перегородки и задней стенки левого желудочка, лечению подлежат только больные, имеющие либо обструкцию выводящего тракта левого желудочка более 30 мм рт. ст., либо глобальную систолическую дисфункцию с фракцией выброса менее 50%, либо ишемию миокарда, либо показания для антиаритмической терапии. Эти изменения отражаются на выраженности гипертрофии миокарда, изменении деформации, толщины миокардиального сегмента, при которой впервые обнаруживается сегментарная систолическая миокардиальная дисфункция. Алгоритм ранней диагностики сегментарной миокардиальной дисфункции для обнаружения первых признаков доглобального нарушения функции миокарда, позволит повысить эффективность профилактики жизнеугрожающих состояний и внезапной сердечной смерти у пациентов с гипертрофической кардиомиопатией.

### 1.1. Алгоритм решения задачи

Оценка толщины миокарда, по данным ЭхоКГ, ставилась как математическая задача поиска минимума производной функции зависимости деформации миокарда от толщины миокарда и включала этап построения полиномиальной регрессии и этап оценки точки перегиба регрессионной зависимости как решение некорректно поставленной задачи [4]. Алгоритм был реализован в виде программного модуля, сопряжённого с ультразвуковым аппаратом «Artida» фирмы Toshiba (Япония). Пользовательский интерфейс показан на рис. 1.

Для определения толщины миокарда используют изображения левого желудочка на уровне его основания и средней трети по короткой оси сердца, полученные во время диастолы, определяемой по ЭКГ. После определения центра полости левого желудочка программное обеспечение автоматически определяет границы эндокарда и эпикарда. Предусмотрена возможность ручной коррекции границы в

зависимости от конкретной толщины стенки. Затем следует полуавтоматическое разделение каждой зоны исследования: на уровне основания и средней трети левого желудочка на 6 сегментов: нижний, задний, боковой, передний, переднесептальный и септальный. Толщина миокарда измеряется в каждом из представленных сегментов миокарда в трех местах (с краев сегмента и по середине). Среднее арифметическое этих трех измерений принимается за толщину сегмента. По результатам обследования 38 детей больных гипертрофической кардиомиопатией было построено распределение оценки толщины миокарда и рассчитано перцентильное распределение, что дало возможность оценить значение критической толщины миокарда.



Рис. 1. Пользовательский интерфейс модуля оценки толщины миокарда

## 2. Коррекция оценки QT интервала с учётом ЧСС

Электрокардиограмма (ЭКГ) является графическим представлением, отражающим средние значения потенциалов действия, возникающих при работе сердца. По ЭКГ определяется частота и регулярность сердечных сокращений, различные патологические фрагменты, на основании которых ставится диагноз. Основными при анализе ЭКГ являются зубцы  $P, Q, R, S, T$  как показано на рис.2 [5].

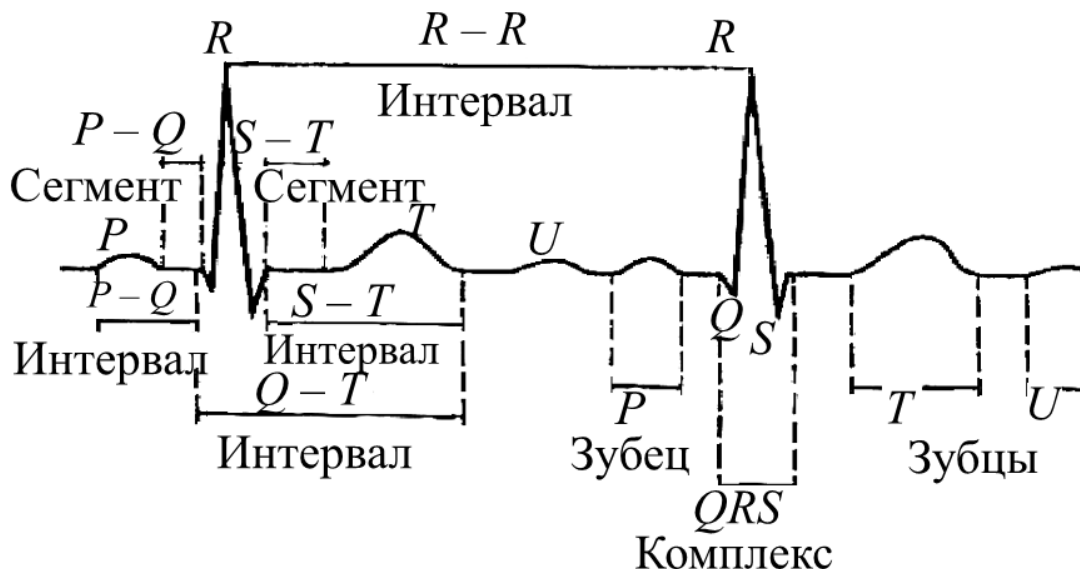


Рис. 2. Элементы кардиограммы

Интервалы и комплексы зубцов отражают особенности работы сердца, в частности интервал  $QT$  показывает сумму процессов деполяризации и реполяризации миокарда желудочков, называется электрической систолой желудочков [6].

Важность изучения интервала  $QT$  обусловлена тем, что при длинном интервале повышается риск развития фатальных нарушений ритма. Удлинение интервала  $QT$  может быть как врожденным (вследствие мутаций определенных генов), так и приобретенным вследствие электролитных нарушений, ишемии миокарда, влияния лекарственных препаратов [7]. Укороченный интервал  $QT$  характеризуется короткими рефрактерными периодами предсердий и желудочков и возможностью индукции фибрилляции желудочков во время проведения электрофизиологического исследования, ведущая к внезапной смерти. У большинства пациентов с такой патологией присутствуют явные структурные изменения сердца, однако у некоторых органическое поражение сердца не может быть идентифицировано. Клиническая значимость данного явления высока в силу того, что ему чаще подвержены молодые, в целом здоровые люди [8]. Своевременное обнаружение изменения длины интервала  $QT$  осложняется зависимостью длины от ЧСС – частоты сердечных сокращений. В медицине пользуются нормативными значениями интервалов при ЧСС равной 60 ударам в минуту, что ставит задачу коррекции длины интервала, измеренного по ЭКГ при текущем ЧСС, к ЧСС, равной 60 ударов в минуту.

В кардиологической литературе, начиная с 1920 года, предложены десятки формул для вычисления нормативного значения  $QT$ . Исторически первый из предложенных математических методов учёта зависимости интервала  $QT$  от ЧСС принадлежит Bazett [9], основанной на формуле для электрической систолы, предположив использовать коэффициент зависимости, который впоследствии заменили на измеренное по ЭКГ значение интервала  $QT$ . До настоящего времени эта формула остается наиболее популярной в клинической практике. В [10] приведен обзор других формул коррекции длины интервала  $QT$ . Недостатком этих формул является то, что они получены с использованием данных здоровых взрослых людей. Формула, ориентированная на детей в условиях быстро меняющихся физиологических процессов, отражающихся на ЧСС, отсутствует.

## 2.1. Построение формулы оценки скорректированного значения $QT$

Методология пересчета измеренного значения длины интервала  $QT$  в скорректированное значение длины при ЧСС равной 60 ударов в минуту –  $QT_c$  приведено в [10]. На рис. 3 открытыми символами показаны экспериментальные данные длины интервала  $QT$  и ЧСС у практически здоровых детей в возрасте от 0 до 18 лет (2705 мальчиков, 2682 девочек), полученные в рамках исследования «ЭКГ-скрининг детей и подростков Российской Федерации» Научно-исследовательского института педиатрии РНИМУ им. Пирогова [10].

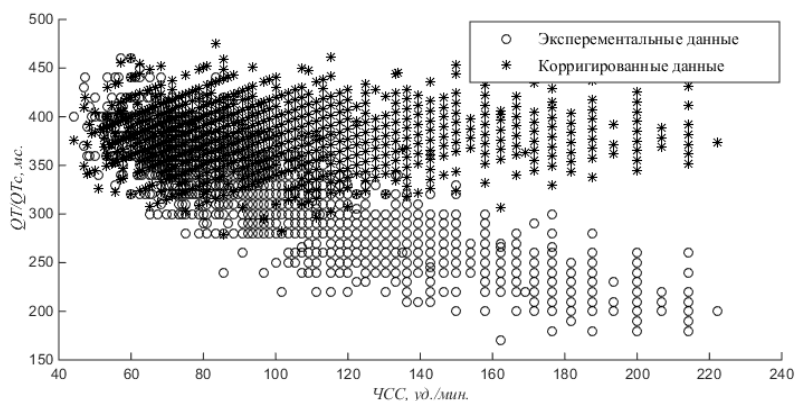


Рис. 3. Исходные значения  $QT$  и рассчитанные по формуле регуляризованного кубического полинома значения  $QT_c$

Звездами отмечены рассчитанные по этим данным скорректированные значения  $QT_c$ . Расчет скорректированных значений осуществлялся с помощью кубической регрессии с регуляризацией по формуле

$$QT_c = QT(hr) - \sum_{j=1}^k a_j (hr - 60)^j$$

здесь  $hr$  – ЧСС,  $QT(hr)$  – измеренная по ЭКГ длина интервала  $QT$ ,  $QT_c$  – скорректированная длина интервала  $QT$ ,  $a_j$  – коэффициенты регуляризованного кубического полинома. Коэффициенты  $a_j$  ( $j=1,2,3$ ) определялись путём минимизации регуляризованного функционала

$$\sum_{i=1}^n (QT_i - P(hr_i))^2 + \gamma \int_a^b P''(\tau)^2 d\tau \rightarrow \min_P$$

здесь  $P$  – кубический полином,  $\gamma$  – параметр регуляризации, задающий соотношение между точностью приближения эмпирических данных и величиной, характеризующей кривизну регрессионной зависимости, а и  $b$  границы диапазона ЧСС, на котором контролируется кривизна регрессионной зависимости. Степень полинома и величина параметра регуляризации определялись по данным [10].

Из рис. 3 видно, что скорректированные значения  $QT_c$  значительно меньше зависят от ЧСС, чем  $QT$ . Качество скорректированного значения  $QT_c$  характеризуется коэффициентом корреляции с длиной интервала  $RR$  – величиной, ЧСС. Этот коэффициент для рассчитанного по формуле регуляризованного кубического полинома  $QT_c$  равен 0.006, превосходя широко используемую формулу Bazett, для которой коэффициент корреляции равен -0.166.

### 3. Диагностика не кардиологических заболеваний по ЭКГ

Интересный подход к диагностике заболеваний методом анализа данных по ЭКГ описан в [11]. Рассматривая сердце как информационный орган [12] авторы предложили изучать не только вариабельность сердечного ритма [13], но и вариабельность высоты  $R$  зубцов. Для этого последовательность интервалов и амплитуд кардиоциклов преобразуется в вектор признаков, включающий тройки значений:  $T$  – длина интервала,  $R$  – высота зубца,  $a$  – арктангенс отношения  $R$  к  $T$ . Для уменьшения влияния сторонних факторов значения признаков заменяются знаками их приращения на соседних кардиоциклах, кодируемые 6-ю буквами  $A, B, C, D, E, F$ . Полученная последовательность символов преобразуется в набор *триграмм* – набор трех последовательно идущих символов, анализируемый далее для постановки диагноза методами машинного обучения, применяемыми при анализе текстов.

#### 3.1. Анализ триграмм

С помощью статистических тестов авторы показали наличие статистической связи между триграммами и 18-ю не кардиологическими заболеваниями, среди которых гипертоническая болезнь, желчнокаменная болезнь, ишемическая болезнь сердца, онкопатология различной локализации, сахарный диабет, язвенная болезнь. Диагностика заболевания осуществлялась несколькими методами бинарной классификации, в котором за класс 0 принимались здоровые люди, а за класс 1 – люди, страдающие конкретным заболеванием.

##### *Наивный байесовский классификатор*

В методе наивного байесовского классификатора [14] признаки считались независимыми между собой. Внутри классов признаки имели распределение Пуассона, параметры которого оценивались по обучающей выборке. В процессе обучения для каждого заболевания формировался набор из  $K$  наиболее информативных триграмм, которые использовались при постановке диагноза. Триграммы, не входящие в набор информативных, опускались.

##### *Логистическая регрессия*

Чтобы освободиться от гипотезы о независимости признаков, рассматривался алгоритм оценки апостериорной вероятности принадлежности к классу – логистическая регрессия [14]. Для повышения надёжности результата диагностики этот алгоритм дополняется методом снижения размерности путём введения  $L_1$  регуляризации (метод LASSO [14]), либо построением логистической регрессии на  $N$  главных компонентах в пространстве исходных признаков.

#### 3.2. Результаты

Расчеты показывают высокие значения чувствительности и специфичности для всех рассмотренных заболеваний [11] при точности выше 90% . Алгоритм на основе наивного Байеса практически не подвержен переобучению и очень устойчив. Изменение числа наиболее информативных триграмм  $K$  мало влияет на качество диагностики. Значение  $K$  можно подбирать по обучающей выборке, вообще не выделяя тестовую. Логистическая регрессия подвержена переобучению. Качество результата на обучении существенно отличается от результата на тесте при превышении оптимального числа главных компонент  $N$ . Оптимизация  $N$  возможна только по тестовой выборке.

#### 4. Заключение

Рассмотрены примеры анализа сигналов ЭКГ и ЭхоЭКГ при решении задач оценки толщины миокарда по данным ЭхоЭКГ, коррекции оценки интервала QT с учётом ЧСС, диагностики не кардиологических заболеваний по ЭКГ. Все задачи объединены подходом, при котором отсутствует построение модели объекта – сердца, а строится «модель» данных. При этом напрямую решается поставленная задача оценки важной характеристики или постановки диагноза, минуя решение «промежуточной» задачи – создание математической модели сердца. Если две первые задачи не выходят за традиционный круг кардиологических проблем, то последняя задача посвящена диагностике не кардиологических заболеваний и подтверждает гипотезу, что кардиограмма имеет гораздо более высокий информационный потенциал, чем просто регистрация изменения электрических потенциалов. Решение рассмотренных задач стало возможно благодаря применению современных подходов и методов анализа данных и машинного обучения.

#### Литература

1. Черных Н.Ю., Грознова О.С., Довгань М.И., Подольский В.А. Изменение деформации миокарда как ранний маркер миокардиальной дисфункции при гипертрофической кардиомиопатии у детей // Рос. вестн. перинатол. и педиатр. 2016. Т. 61, N 5. – С. 70–74.
2. Maron B.J. Prevention of sudden death with the implantable defibrillator in hypertrophic cardiomyopathy // Rev. Esp. Cardiol. – 2006. – Vol. 59, N 6. – P. 527–530.
3. Ly H.Q., Greiss I., Talakic M. et al. Sudden death and hypertrophic cardiomyopathy: a review // Can. J. Cardiol. – 2005. – Vol. 21, N 5. – P. 441–448.
4. Тихонов А.Н., Арсенин В.Я. Методы решения некорректных задач. – М: Наука, 1979. – 331 с.
5. Школьников М.А., Миклашевич И.М., Калинина Л.А. Нормативные параметры ЭКГ у детей и подростков –М.: Ассоциация детских кардиологов России, 2010. – 232 с.
6. Qiu H., Bird G.L., Qu L., et al. Evaluation of QT Interval Correction Methods in Normal Pediatric Resting ECGs // Computers in Cardiology. – 2007. – Vol. 34. P. 431–434.
7. Фурман Н.В., Шматова С.С. Клиническое значение удлинения интервалов QT и QTc на фоне приёма лекарственных препаратов // Рациональная Фармакотерапия в Кардиологии. 2013. Т.9, N 3. – С. 311–315.
8. Гукасова И.И. Синдром укороченного интервала QT (клиника, диагностика, лечение) // Анналы ритмологии. 2004. N 4. – С.17–22.
9. Bazett H.C. An analysis of the time relations of Electrocardiograms // Heart. – 1920. – Vol. 7 – P. 353-370.
10. Горлицев В.П., Михальский А.И., Калинин Л.А., Школьников М.А., Школьников В.М. Метод коррекции электрокардиографического интервала с учетом частоты сердечных сокращений // Проблемы управления. 2016. N 6 – С. 65-70.
11. Uspenskiy V.M., Vorontsov K.V., Tselykh V.R., Bunakov V.A. Information function of the heart: discrete and fuzzy encoding of the ECG-signal for multidisease diagnostic system // Advanced mathematical and computational tools in metrology and testing X (Series on Advances in Mathematics for Applied Sciences). – World Scientific Publishing Co (Singapore) – Vol. 86 – P. 387-384.
12. Успенский В.М. Информационная функция сердца. Теория и практика диагностики заболеваний внутренних органов методом информационного анализа электрокардиосигналов – М.: Экономика и информатика, 2008. – 116 с.
13. Баевский Р.М., Иванов Г.Г. Вариабельность сердечного ритма: теоретические аспекты и возможности клинического применения // Ультразвуковая и функциональная диагностика. 2001. N 3. – С. 108–127.
14. Мерков А.Б. Распознавание образов. Введение в методы статистического обучения – Едиториал УРСС, 2011 – 256 с.