

Компенсация QT-интервала на сигнале стандартной ЭКГ для повышения точности алгоритмов обработки предсердной активности

А. Н. Михайлов

Санкт-Петербургский государственный электротехнический университет
«ЛЭТИ» им. В.И. Ульянова (Ленина)

lexarobotr2d2@gmail.com

Аннотация. В статье предложен метод компенсации QT-интервала на сигнале стандартной ЭКГ, направленный на повышение точности алгоритмов обработки предсердной активности. Основной подход включает кластеризацию QT-интервалов по форме с последующим выделением среднего образа для каждого кластера и его вычитанием из соответствующих комплексов. Это позволяет минимизировать влияние желудочковой активности на сигнал ЭКГ, что способствует улучшению детекции Р-волн. Разработанный метод может быть использован для повышения эффективности алгоритмов анализа предсердной активности, что имеет важное значение для диагностики и мониторинга сердечно-сосудистых заболеваний.

Ключевые слова: ЭКГ; QT; DTW; DBSCAN; предсердная активность

I. ВВЕДЕНИЕ

Алгоритмы обработки сигналов стандартной ЭКГ, разработанные на данный момент компанией «Инкарт», реализуют обнаружение границ комплексов QRS и Т-волны с хорошей точностью, но при обнаружении Р-волны могут возникать трудности, связанные со смещением Р-волны на Т-волну, фибрилляциями и прочими подобными случаями. При этом корректное обнаружение Р-волн является важным этапом при обработке стандартной ЭКГ, так как положение Р-волн позволяет определить различное поведение предсердной активности, в том числе и всевозможные аритмии.

С целью увеличения точности работы алгоритмов в рамках данной работы приведён алгоритм, производящий операцию вычитания среднего образа построенных кластеров QT-интервалов, что позволит сохранить интересующие частотные составляющие на сигнале ЭКГ для дальнейшего анализа предсердной активности сердца.

II. КЛАСТЕРИЗАЦИЯ ИНТЕРВАЛОВ

В данной работе было принято решение компенсировать QT интервалы, а не QRS комплексы, так как точки Q и S могут быть смещены относительно друг друга сильнее, чем точки Q и конец Т-волны, так как распространено явление смещения ST-сегмента. Чем больше точки отклонены друг от друга относительно изолинии, тем сильнее будут скачки при последующей подстановке итогового сигнала на место исходного, поэтому минимизация данного отклонения способствует улучшению работы алгоритма.

А. Метрика. Dynamic Time Warping

Для того чтобы разделить рассматриваемые интервалы на кластеры, необходимо определить метрику, которой будет определяться схожесть или различие между интервалами. В качестве данной метрики была выбрана цена пути при динамической трансформации временной шкалы (Dynamic Time Warping, далее – DTW). Данный выбор объясняется тем, что эта метрика чувствительна к форме сигнала, при этом имеет низкую чувствительность к временному смещению между сигналами и их растяжению (сжатию) вдоль шкалы времени, что позволяет компенсировать неточности в детекции границ интервалов.

DTW-алгоритм строит матрицу возможных отображений одного ряда на другой, с учётом того, что отсчёты рядов могут как сдвигаться, так и изменять уровень. Путь в этой матрице с минимальным значением стоимости – и есть DTW-расстояние. [1]

Для двух сигналов Q и C создаётся матрица размером $|Q|$ на $|C|$, затем заполняется расстоянием между каждой возможной парой точек в данных двух временных рядов. Любое возможное отклонение между двумя временными рядами – это путь через матрицу. Нам нужен лучший из них. Формульно данное расстояние можно выразить следующим образом [2]:

$$DTW(Q, C) = \min \left\{ \frac{\sqrt{\sum_{k=1}^K w_k}}{K} \right\} \quad (1)$$

При этом значение для каждой клетки данной матрицы будет определяться следующим выражением [2]:

$$\gamma = d(q_i, c_i) + \min\{\gamma(i-1, j-1), \gamma(i-1, j), \gamma(i, j-1)\} \quad (2)$$

Таким образом, чем больше сигналы похожи друг на друга по форме, тем меньше будет расстояние, рассчитанное по DTW для них. При этом данная метрика хорошо подходит для сигналов, на которых расстояние между некоторыми элементами непостоянно, что особенно актуально для нестационарных биологических сигналов.

Ниже на рис. 1 приведена матрица поиска расстояния DTW, построенная для тестовых сигналов Q и C, изображённых сбоку и сверху от матрицы соответственно, по правилам, описанным выше. На рис. 2 изображено сопоставление сигналов Q и C по построенной ранее матрице.

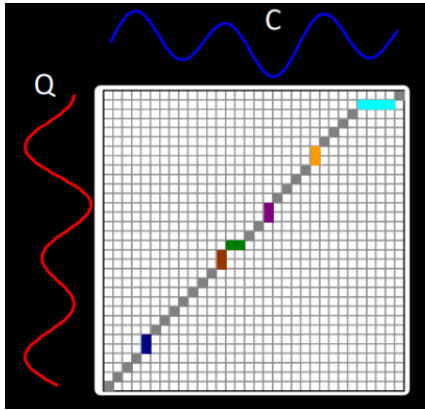


Рис. 1. Матрица DTW, построенная для тестовых сигналов Q и C [2]

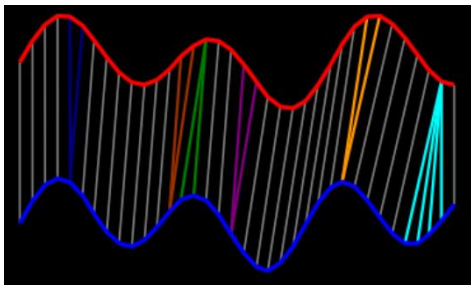


Рис. 2. Результат сопоставления двух тестовых сигналов по DTW

В. Метод кластеризации. DBSCAN

Существует большое множество алгоритмов кластеризации, позволяющих разделить набор данных по данным метрикам на группы (кластеры). Так как в выполнении данной работы заранее неизвестно количество кластеров (количество морфологий QT-интервалов в общем случае может быть ограничено, но какое количество из них присутствует в данной записи – неизвестно), то для выполнения поставленной задачи необходимо использовать метод, который не требует в качестве входного параметра количество кластеров. Таким образом, классические алгоритмы, такие как k-means и c-means, не подходят. Для выполнения поставленной задачи был выбран DBSCAN.

DBSCAN (Density-Based Spatial Clustering of Applications with Noise) – это метод кластеризации, основанный на плотности данных. Он группирует точки, которые находятся близко друг к другу в плотных областях, и отделяет шумовые точки, которые лежат в разреженных областях. Основные параметры метода: *eps* (максимальное расстояние между точками для их объединения в кластер) и *minPts* (минимальное количество точек для формирования кластера). Алгоритм работает, находя «соседей» каждой точки в радиусе *eps*, и если их количество превышает *minPts*, то эти точки образуют кластер. Процесс продолжается, пока все точки не будут обработаны [3].

Таким образом, используя алгоритм DBSCAN с метрикой расстояния между интервалами, рассчитанной по DTW, становится возможным кластеризовать сигналы точно по форме и на неопределённое количество кластеров. Но существует проблема, связанная с чувствительностью DTW к дрейфу изолинии, а также данный подход способен разделить данные одинаковой формы, но различной амплитуды. Чтобы избежать указанных ошибок, данные необходимо предварительно подготовить к кластеризации. На рис. 3 приведено

изображение пошаговой работы алгоритма, где красные точки – шумовые, для которых не найден кластер, зелёные и жёлтые – точки, образовавшие кластер.

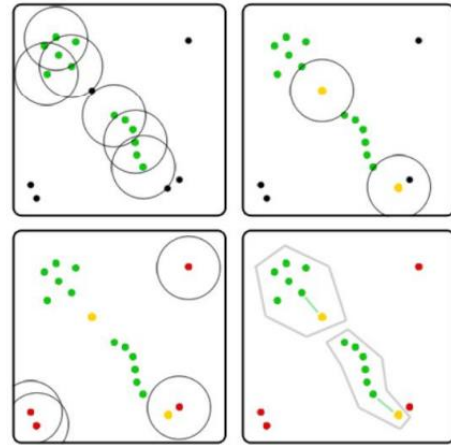


Рис. 3. Иллюстрация работы алгоритма DBSCAN

С. Подготовка сигналов

Для того чтобы избежать ошибок алгоритма DTW, связанных с дрейфом изолинии, исходный сигнал подвергается фильтрации фильтром высоких частот с частотой среза 0.5 Гц. Тем не менее, данная мера может допустить перепады высот между точками Q и концом T, поэтому полученные интервалы приводятся к нулю путём вычитания прямой, соединяющей левую и правую точки данного интервала. Это позволяет избежать ошибок алгоритма, не сильно искажая при этом форму интервала.

Также необходимо объединять в один кластер участки сигнала схожей формы, но различной амплитуды. Для этого рассчитывается среднеквадратическое отклонение для каждого интервала, а затем каждая точка интервала делится на полученное значение. Таким образом СКО каждого интервала становится равным 1. При этом важно отметить, что каждое рассчитанное значение СКО следует заранее запомнить, так как в дальнейшем его нужно будет использовать для вычитания среднего образа из исходного интервала.

На рис. 4 приведён пример исходных, ненормированных QT-интервалов, а на рис. 5 – те же интервалы, но приведённые к нулю и нормированные по СКО.

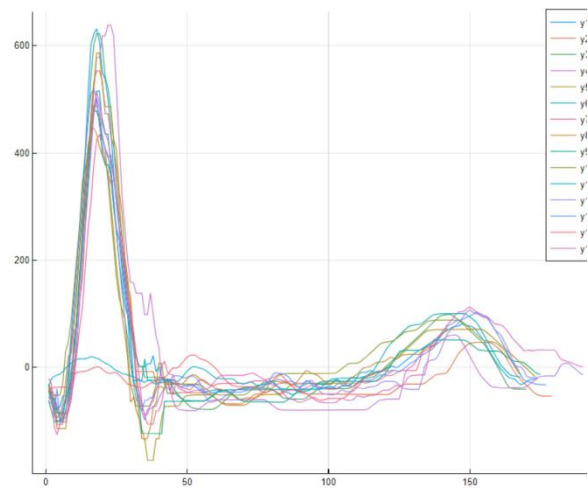


Рис. 4. Исходные интервалы QT

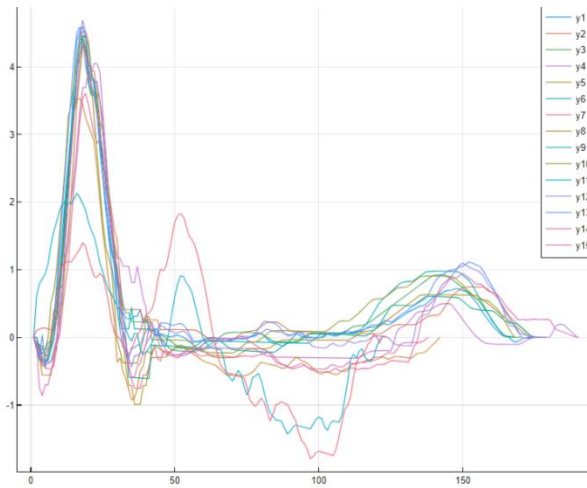


Рис. 5. Интервалы QT, нормированные по СКО и с приведёнными к нулю началом и концом интервала

D. Результаты кластеризации

Помимо описанных выше процедур стоит помнить, что каждый из рассматриваемых сигналов является 8-канальным, а значит при кластеризации необходимо рассматривать не одно отведение, а все сразу. Существует два основных способа реализовать это – суммировать расстояния по каждому отведению или же считать расстояния между каждой точкой каждого комплекса как расстояние между 8-мерными точками, где i -я координата является значением данной точки в i -м отведении. В данной работе используется второй вариант, так как он более чувствителен к синхронным изменениям сигналов, что актуально для многоканальной ЭКГ. [2]

Используя все вышеперечисленные методы, становится возможным разделить сигналы на схожие по форме группы для дальнейшей работы с ними. На рис. 6 приведён пример разделения интервалов одной записи на кластеры (на изображении приведено только одно отведение – I), где сигналы одного цвета принадлежат одному кластеру.

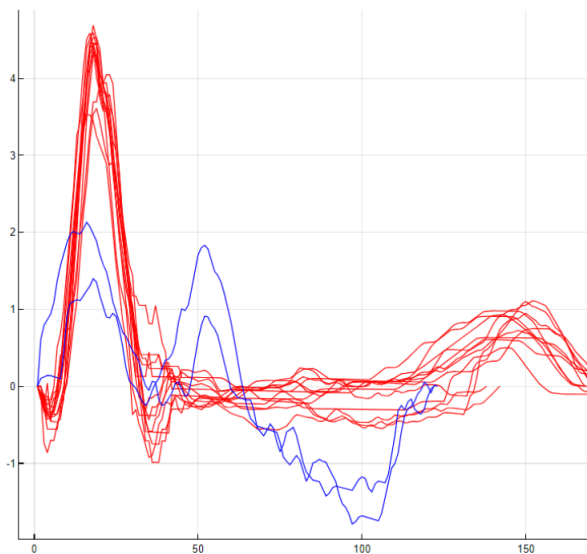


Рис. 6. Результат кластеризации методом DBSCAN с метрикой расстояния, рассчитанной по DTW

III. НАХОЖДЕНИЕ И ВЫЧИТАНИЕ СРЕДНЕГО ОБРАЗА

A. Расчёт среднего образа. DTW barycenter averaging

Теперь, после разделения интервалов на группы, необходимо для каждой из данных групп найти её средний образ. Средним образом кластера здесь и далее будем называть наиболее типичного представителя данного кластера (его геометрический центр), по характеристикам которого можно судить обо всем кластере (центр кластера). [4]

Аналогично прошлым пунктам существует множество вариантов для нахождения среднего образа кластера, но все ранее озвученные трудности остаются актуальными (DTW помогло синхронизировать интервалы по времени для расчёта метрики, но сигналы при этом не изменились), поэтому использование среднего арифметического между точками, выравненными по границам, не даст должного результата. Решением данной проблемы является использование метода усреднения барицентра по DTW.

DBA расширявается как DTW barycenter averaging. DBA – это метод усреднения, который заключается в итеративном уточнении изначально (потенциально произвольной) усредненной последовательности, чтобы минимизировать ее квадратичное расстояние (DTW) до усредненных последовательностей. [5]

Цель состоит в том, чтобы минимизировать сумму квадратов расстояний DTW от средней последовательности до набора последовательностей. Эта сумма складывается из единичных расстояний между каждой координатой средней последовательности и координаты связанных с ней последовательностей. Принцип DBA заключается в вычислении каждой координаты средней последовательности как барицентра связанных с ней координат набора последовательностей [5]. На рис. 7 приведён кластер QT интервалов и обнаруженный для него средний образ. Обращаю внимание на то, что данный сигнал взят из другой записи, нежели сигналы с рис. 4–6.

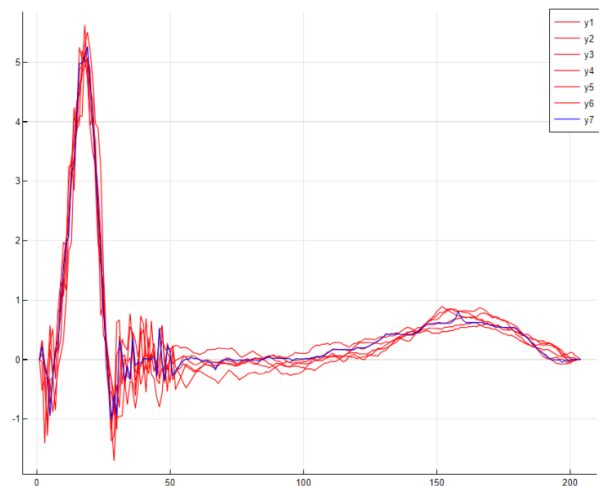


Рис. 7. Кластер интервалов (красный) и его средний образ (синий)

B. Вычитание среднего образа из сигнала

После всех проведённых процедур получены средние образы кластеров, а также известно какой сигнал к какому кластеру относится. Последнее, что осталось сделать –

провести вычитание среднего образа из исходного интервала. Данный шаг также требует индивидуального подхода для каждого интервала по двум причинам: нормировка по СКО, синхронизация во времени.

Чтобы вычесть из некоторого интервала средний образ его кластера, необходимо: привести данный интервал к нулю; умножить средний образ на заранее полученное значение СКО для данного интервала; провести сопоставление сигналов с помощью DTW; из каждой точки исходного интервала (не приведённого к нулю) вычесть среднее арифметическое всех связанных с ней точек среднего образа.

На рис. 8 проиллюстрирована связь между точками некоторого интервала и соответствующего ему среднего образа, полученная на основе матрицы DTW.

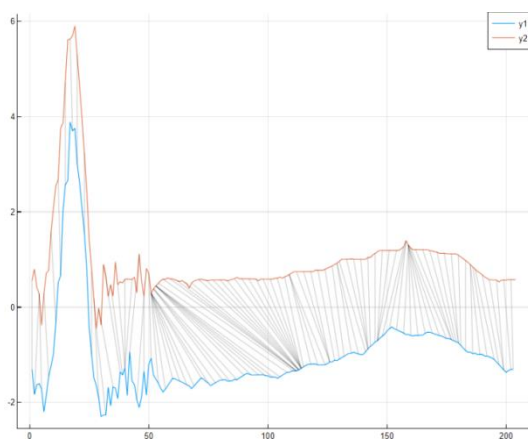


Рис. 8. Связь точек между исходным интервалом (синий) и средним образом кластера (оранж.), к которому он принадлежит, рассчитанная по DTW

IV. АНАЛИЗ ПОЛУЧЕННЫХ РЕЗУЛЬТАТОВ

На рис. 9 приведена иллюстрация работы алгоритма на одной из записей без патологий, а на рис. 10 приведён пример работы алгоритма на сигнале с ярко выраженной помехой на изолинии, на котором видно, что алгоритм позволяет компенсировать QT-интервал, при этом сохраняя чёткую форму предсердной активности.

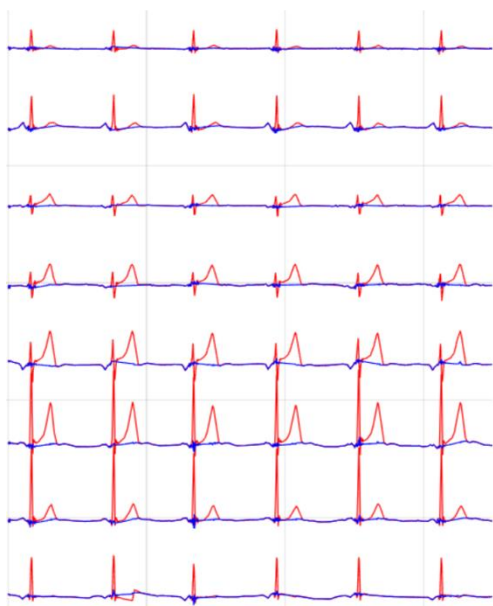


Рис. 9. Пример работы алгоритма, все отведения (красный – исходный сигнал, синий – после обработки)

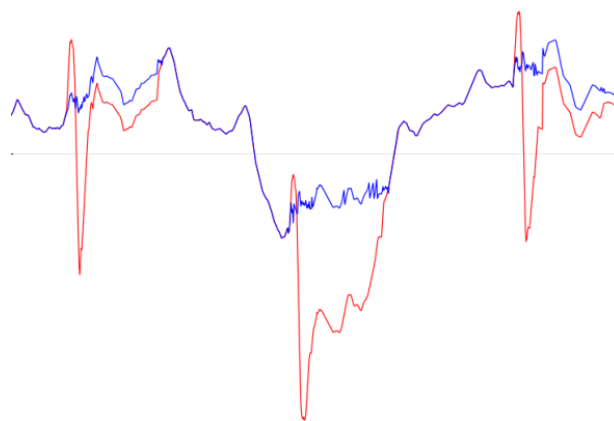


Рис. 10. Пример работы алгоритма на сигнале с помехой на изолинии (красный – исходный сигнал, синий – после обработки)

V. ЗАКЛЮЧЕНИЕ

В результате последовательного использования данных алгоритмов стало возможным сгруппировать по форме QT-интервалы исходных сигналов на кластеры, вычислить средний образ для каждого кластера и вычесть его из исходного сигнала. На полученных после данной обработки сигналах отсутствуют скачки и явно выраженные QRS-комплексы или T-волны, в то время как предсердная активность сохранена и доступна для дальнейшего анализа.

Данный алгоритм используется НАО «Инкарт» в программном обеспечении по обработке стандартной ЭКГ.

БЛАГОДАРНОСТЬ

Выражаю благодарность НАО «Инкарт» за предоставление широкой базы данных с файлами ЭКГ для обучения алгоритма, а также непосредственно руководителю отдела разработки Гуменному В.Г. за предоставление идеи данного алгоритма и полное сопровождение в процессе разработки.

СПИСОК ЛИТЕРАТУРЫ

- [1] Pavel Senin. Dynamic Time Warping Algorithm Review. Information and Computer Science Department, University of Hawaii at Manoa. Honolulu, USA, 2008. 23 p.
- [2] Муин А., Кюо Э. Дж. Extracting Optimal Performance from Dynamic Time Warping // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2016). San Francisco, California, 2016. С. 2129–2130. DOI: 10.1145/2939672.2939870.
- [3] Шуберт Э., Сандер Й., Эстер М., Кригель Х. П., Сьюй С. DBSCAN Revisited, Revisited: Why and How You Should (Still) Use DBSCAN // ACM Transactions on Database Systems. 2017. Т. 42, № 3. С. 1–21. DOI: 10.1145/3068335.
- [4] Сулов С.А. Кластерный анализ: сущность, преимущества и недостатки // Вестник НГИЭИ. 2010. С. 51–57.
- [5] Петижан Ф., Кеттерлен А., Ганкарски П. A global averaging method for dynamic time warping, with applications to clustering // Pattern Recognition. 2011. Т. 44, № 3. С. 678–693. DOI: 10.1016/j.patcog.2010.09.013.