

УДК 681.518.2

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ СЛОЖНЫХ НЕСТАЦИОНАРНЫХ СИГНАЛОВ НА ПРИМЕРЕ ЭЛЕКТРОКАРДИОГРАФИЧЕСКИХ СИГНАЛОВ

Костыгов А.М., Лузянин И.С., Артемов С.А., Гвозденко С.В.

*Пермский национальный исследовательский политехнический университет,
Пермь, e-mail: dekan@etf.pstu.ru*

В статье представлены методы интеллектуального анализа трехмерного представления электрокардиографического сигнала (ТЭКГ) с целью повышения эффективности анализа электрокардиографических сигналов (ЭКГ) с помощью извлечения дополнительных данных. Описан метод подготовки данных для последующего анализа. Разработаны алгоритмы расчета и визуализации дополнительных данных, описаны применяемые математические методы и средства интеллектуального анализа, построена модель интеллектуального анализа с целью выявления паттернов, которые могут быть использованы для сравнительной оценки точности методов. Предложены методы кластеризации и классификации данных ТЭКГ на основе алгоритмов k-средних и деревьев решений. Показана эффективность предложенных методов при решении задач анализа паттернов электрокардиографических сигналов. Данный подход дает наибольшую точность в задаче интеллектуального анализа, что приведено в данной статье.

Ключевые слова: интеллектуальный анализ, ЭКГ, кластеризация, классификация, деревья решений, алгоритм k-средних

INTELLIGENT ANALYSIS OF COMPLEX NON-STATIONARY SIGNALS ON THE EXAMPLE OF ELECTROCARDIOGRAPHICAL SIGNALS

Kostygov A.M., Luzyanin I.S., Artemov S.A., Gvozdenko S.V.

Perm National Research Polytechnic University, Perm, e-mail: dekan@etf.pstu.ru

The article presents intellectual analysis methods of three-dimensional ECG signal (3D ECG) for increasing the efficiency of the electrocardiographic signals (ECG) analysis by extracting additional data. Discloses a method of preparation of data for later analysis. Algorithms for preparing and visualizing additional data are described. Mathematical methods and tools for data analyzing are used to analyze are described. The model of predictive analysis for identifying patterns, which are used for comparative assessment accuracy of methods is constructed. Clustering and classification methods for 3D ECG based on k-means algorithm and decision trees are proposed. The efficiency of proposed methods for solving the task of ECG signals pattern analyzing is shown. This approach provides high accuracy in data mining task.

Keywords: intelligent analysis, ECG, clustering, classification, decision trees, k-means

С распространением программно-аппаратных комплексов сбора и идентификации биологических сигналов, а та же математических и статистических пакетов обработки данных, появились средства, позволяющие производить диагностику работы органов и систем человеческого организма. На основе полученных данных предсказываются нарушения в работе органа задолго до их появления, определяются тончайшие аномалии в их работе. Одним из методов предсказания аномалий является интеллектуальный анализ данных. Интеллектуальный анализ подразумевает извлечение информации из массивов данных, которые показывают определенные взаимосвязи с определенной точностью. Повышение точности является основной проблемой анализа, решением которой является поиск атрибутов, которые более конкретно описывают поведение сигнала.

Среди всего многообразия биологических сигналов наибольший интерес пред-

ставляют электрокардиографические сигналы (ЭКГ). В общем случае сигналы ЭКГ представляют собой трехмерные циклы, показывающие распространение электрического заряда по сердечной мышце с течением времени. Для упрощения задачи анализа сигналы ЭКГ обычно рассматриваются как проекция трехмерных циклов (ТЭКГ) на одну из осей координат. При этом взаимодействия между компонентами сигнала не учитываются. Основная гипотеза данной работы состоит в том, что анализ ТЭКГ даст дополнительные данные, для улучшения метода диагностирования патологий, скрытых в корреляции между сигналами по осям.

Применение паттернов в электрокардиографических сигналах

Паттерном называется закономерность в графике, повторяющаяся время от времени на определенном участке. Анализируя повторяемость некоторого отклонения, можно выявлять изменения в работе сердца,

такие как вентрикулярная тахикардия, экстрасистолия, миокардит, и другие [3].

Существует множество методов поиска закономерностей в сигналах ЭКГ, однако все они сводятся к решению задач интеллектуального анализа данных с использованием той или иной модели [2, 7].

Описание модели интеллектуального анализа поиска паттернов

В данной работе применяется следующая модель поиска паттернов ТЭКГ (рис. 1) [5, 6].

Модель состоит из следующих элементов:

1. Запись ЭКГ, непосредственно с помощью прибора.

2. Выделение отдельных сердечных сокращений и сегментация сигнала.

Сегментация сигнала необходима для выделения участков, характерных для конкретных процессов, происходящих в сердечной мышце. Изменения формы сигнала на каждом участке могут свидетельствовать о различных заболеваниях. Основная сложность при разделении сигнала ЭКГ на сегменты заключается в определении границ каждого сегмента. Поскольку электрические процессы, происходящие в сердечной мышце, носят непрерывный характер, в один момент времени могут протекать сразу два процесса. В этом случае граница между ними не является очевидной. Существует множество методов с различной точностью, такие как:

- скрытая марковская модель;
- вейвлет-декомпозиция;
- алгоритм динамической трансформации.

3. Предварительная обработка данных.

Состоит из двух этапов: фильтрация и выравнивание.

– В первом случае производится фильтрация низкочастотной (НЧ) и высокочастотной (ВЧ) составляющих исследуемого сигнала.

– Выравнивание графиков производится с помощью вейвлет-преобразования.

Для вычисления смещений сигналов относительно друг друга применяется вейвлет-преобразование [8] – трансформация временного сигнала в частотно-временной. Далее на основе вейвлет-функции рассчитываются отклонения, такие как позиция и масштаб для каждого биения.

4. Выделение параметров, поиск атрибутов.

На данном этапе производится сбор наибольшего количества информации об объекте. Здесь происходит разделение на интеллектуальный анализ временных рядов и многомерных массивов.

5. Кластеризация.

Этап кластеризации необходим для обучения модели без учителя. Так как изначально не известны классы и их количество, применяется метод устойчивой кластеризации.

6. Классификация.

Является следующим шагом после кластеризации и основана на алгоритмах интеллектуального анализа. Ее задача предсказать с определенной долей вероятности на основе входных атрибутов, класс, присвоенный предыдущим этапом кластеризации.

7. Алгоритм интеллектуального анализа

Здесь происходит подбор необходимого алгоритма, его настройка и обучение. Вы-

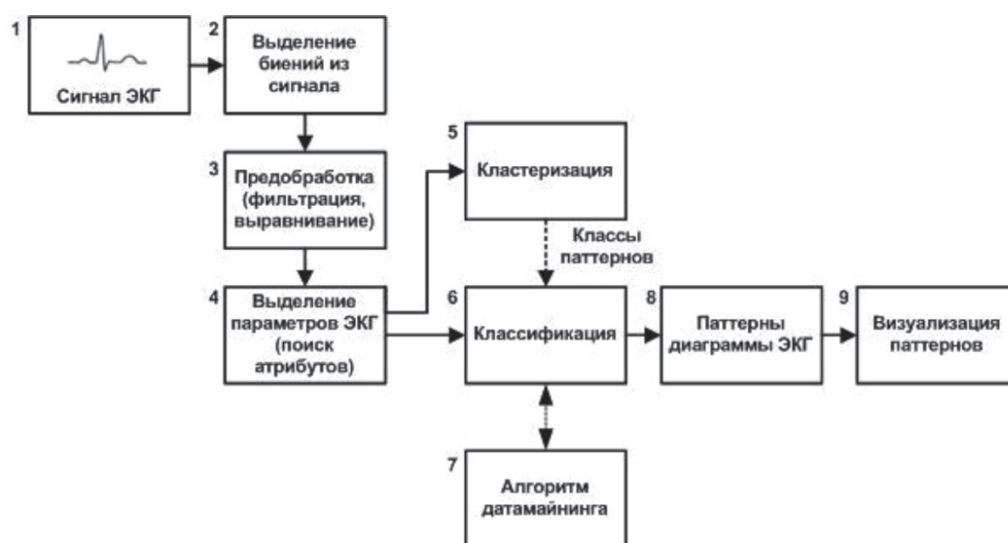


Рис. 1. Структурная схема поиска паттернов

бирается алгоритм, который имеет наилучшие результаты классификации паттернов.

8. Паттерны диаграммы

В результате анализа отдельные участки ТЭКГ объединяются в классы. Данные о классах определенных участков ТЭКГ записываются в специальный файл, на основе которого может производиться классификация других сигналов.

9. Визуализация паттернов

Для практического применения механизма определения паттернов производится визуализация паттернов. Участки с наибольшим отличием на графиках видны благодаря цветовой кодификации кластеров.

Описание и подготовка атрибутов для интеллектуального анализа данных

К основным атрибутам модели относятся непосредственно значения самих сигналов, номер точки, номер биения и номер пациента. Для улучшения точности алгоритмов и сравнения работы с временными и многомерными данными необходимы дополнительные атрибуты, характеризующие сигнал с другой стороны. Рассмотрим каждый дополнительный атрибут отдельно:

1. Евклидова дистанция

Наиболее распространенный атрибут – это измерение подобности графиков. Относительно среднего графика необходимо рассчитать подобность каждого к среднему. Расчет подобности – это расчет суммарной дистанции между графиками (рис. 2, а) [3]. В результате расчета получается массив расстояний с привязкой к номеру точки:

$$D(C, Q) = L_p(C, Q). \quad (1)$$

Такое представление разницы между графиками просто для понимания, но возникает ряд трудностей, связанных со смещением графиков относительно друг друга.

Наиболее подходящий и учитывающий данные недостатки Евклидовой дистанции

является алгоритм динамической трансформации временной оси (ДТВО) (рис. 2, б).

Результатом работы алгоритма является матрица соответствия точек одного графика к другому, на базе которого рассчитываются дистанции между соответствующими точками. ДТВО позволяет учесть растяжение и сжатие графика, его смещение и переусиление.

2. Расстояние между точками представляет собой дистанцию между точками в пространстве и рассчитывается на основе магнитуды сигнала в каждой точке.

3. Расчет отклонений представляет собой измерение евклидовой дистанции в трехмерном пространстве.

4. Положение динамического вектора электрической оси сердца (ЭОС) содержит две переменные. В полярных координатах вектор описывается углами азимута и нормали.

После подготовки данных формируется CSV-файл, который содержит атрибуты исследуемого сигнала.

Описание работы модели интеллектуального анализа данных ЭКГ и ТЭКГ

На рис. 3 приведена модель интеллектуального анализа данных, построенная в пакете IBM Modeler. На вход модели подается CSV-файл с данными. Далее производится подготовка данных: разделение данных на типы и отсеивание ненужных данных для повышения точности анализа данных. Следующими этапами являются кластеризация и классификация подготовленных данных.

Алгоритм кластеризации

Для решения задачи обучения без учителя необходимо сначала получить классы паттернов. С этой целью необходимо решить задачу кластеризации.

Наиболее простым алгоритмом кластеризации является метод *k*-средних (*k-means*) [4].

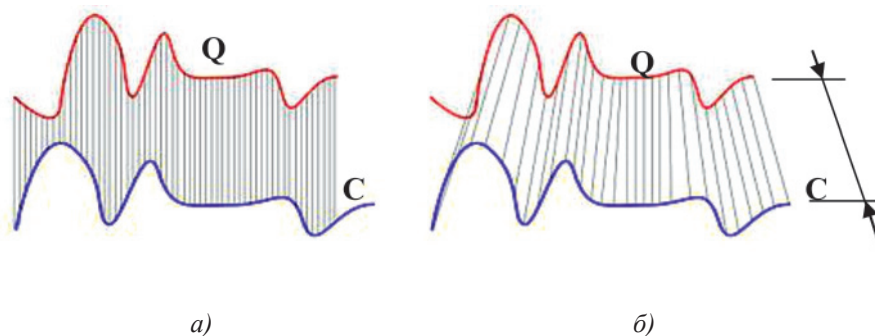


Рис. 2. Сравнение подходов расчета: а – Евклидова дистанция; б – алгоритм ДТВО

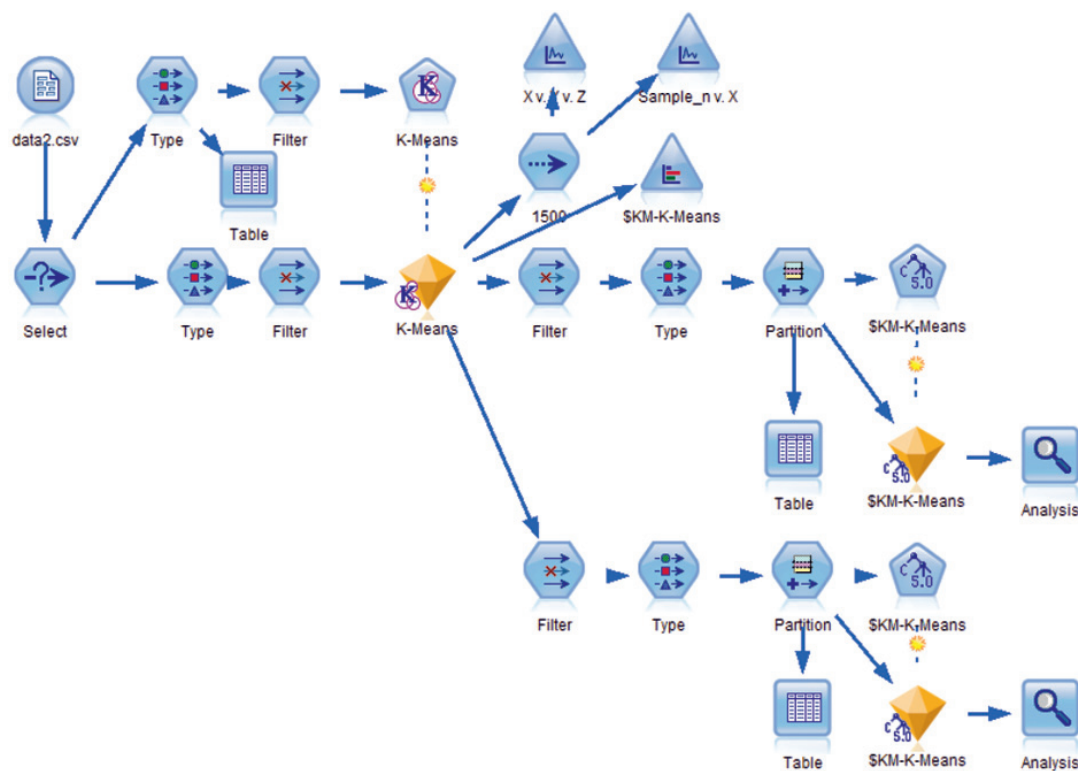


Рис. 3. Модель интеллектуального анализа сигналов ЭКГ



Рис. 4. График метода устойчивой кластеризации

Для работы алгоритма k -средних, как говорилось ранее, необходимо знать количество кластеров. Для этого применяется метод устойчивой кластеризации. На рис. 4 изображен график разделения кластеров, размеры наименьшего и наибольшего кластеров. Как видно, в точке с размерностью 15 кластеров происходит изменение размеров и разделения.

Алгоритм классификации

Наиболее точные результаты в данной работе дает дерево решений C5.0. Деревья

решений представляют собой граф, имеющий иерархическую структуру в вершинах которого стоят условия. На основе этих условий принимаются решения для перехода к нижестоящим вершинам.

Алгоритм C5.0 работает только с категориальными данными с высокой степенью устойчивости к проблемам пропуска данных и большим их количеством.

Результаты интеллектуального анализа

После обучения и классификации получаем результат, приведенный на рис. 5.

'Partition'	1_Training		2_Testing	
Correct	29 323	99,01%	12 357	96,35%
Wrong	293	0,99%	468	3,65%
Total	29 616		12 825	

'Partition'	1_Training		2_Testing	
Correct	25 384	85,71%	10 960	85,46%
Wrong	4 232	14,29%	1 865	14,54%
Total	29 616		12 825	

Рис. 5. Результат классификации ТЭКГ и ЭКГ

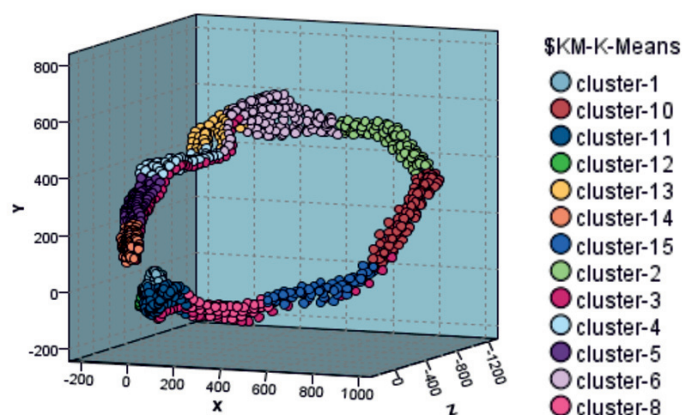


Рис. 6. Визуализация паттернов ТЭКГ

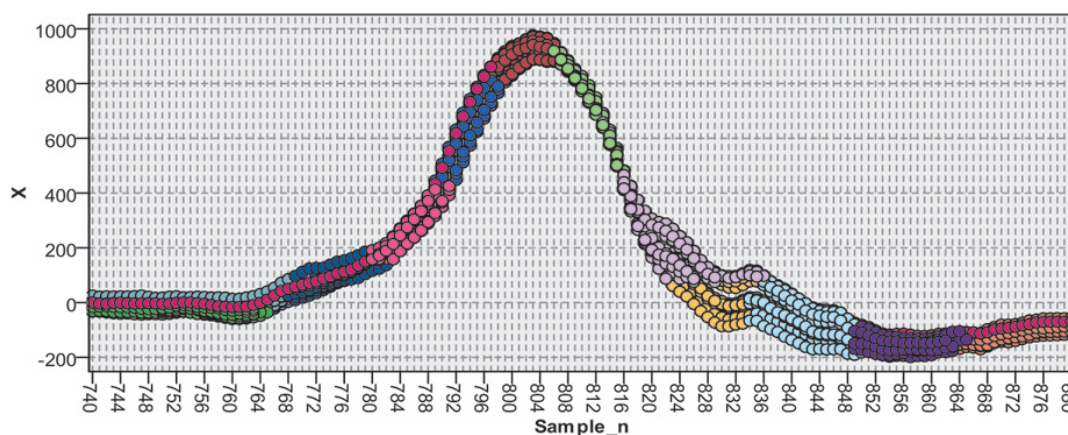


Рис. 7. График проекции паттернов на ось X

Как видно из полученных результатов, использование дополнительных атрибутов ТЭКГ, в сравнении с атрибутами используемых при анализе ЭКГ, улучшает точность классификации.

Для наглядного рассмотрения положения кластеров необходимо произвести их визуализацию. На рис. 6 и 7 отображены полученные графики.

Выводы

В статье был произведен интеллектуальный анализ данных ЭКГ и ТЭКГ с целью определения специфических паттернов внутри исследуемых сигналов и их классификации по определенным признакам. На первом этапе была проведена выборка тестовых сигналов ЭКГ. После чего проводилась

предварительная подготовка данных. После этого в программе *Matlab* была выполнена визуализация циклов ТЭК. На втором этапе была выполнена кластеризация данных. На последнем этапе выполнялась классификация данных и сравнение полученных результатов.

Список литературы

1. Артемов С.А., Кычкин А.В. Энергомониторинг подвижных объектов. //Материалы краевой научно-технической конференции. Автоматизированные системы управления и информационные технологии. – 2014. – С. 126–130.
2. Eamonn Keogh, Jessica Lin, Ada Fu, HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications // ICDM – 2005. – P. 50–60.
3. K.O. Gupta, Dr. P.N. Chatur, ECG Signal Analysis and Classification using Data Mining and Artificial Neural Networks // International Journal of Emerging Technology and Advanced Engineering – 2012. – P. 56–60.
4. Kychkin A.V., Artemov S.A. Software and hardware solution for energy monitoring system // Proceedings of 6th International conference of young scientists on solutions of applied problems in control and communications, data processing and data analysis. – Perm, PNRPU, 2015. – P. 3–11.
5. Marian Kotas, Projective Filtering of Time-Aligned ECG Beats // IEEE transactions on biomedical engineering. – 2004. – Vol. 51, № 7. – P. 1129–1140.
6. Michael Steinbach, Levent Ertöz, Vipin Kumar, The Challenges of Clustering High Dimensional Data // New Directions in Statistical Physics. – 2004. – P. 273–309.
7. M. Kotas, Robust projective filtering of time-warped ECG beats// computer methods and programs in biomedicine – 2008. – P. 161–172.

8. Paolo Buono, Aleks Arisb, Catherine Plaisant, Amir Khellab, Ben Shneidermanb, Interactive Pattern Search in Time Series // VDA – 2005. – P. 1–11.

References

1. Artemov S.A., Kychkin A.V. Jenergomonitoring podviznykh obektov. //Materialy kraevoy nauchno-tehnicheskoy konferencii. Avtomatizirovannye sistemy upravleniya i informacionnye tehnologii. 2014. pp. 126–130.
2. Eamonn Keogh, Jessica Lin, Ada Fu, HOT SAX: Finding the Most Unusual Time Series Subsequence: Algorithms and Applications // ICDM 2005. pp. 50–60.
3. K.O. Gupta, Dr. P.N. Chatur, ECG Signal Analysis and Classification using Data Mining and Artificial Neural Networks // International Journal of Emerging Technology and Advanced Engineering 2012. pp. 56–60.
4. Kychkin A.V., Artemov S.A. Software and hardware solution for energy monitoring system // Proceedings of 6th International conference of young scientists on solutions of applied problems in control and communications, data processing and data analysis. Perm, PNRPU, 2015. pp. 3–11.
5. Marian Kotas, Projective Filtering of Time-Aligned ECG Beats // IEEE transactions on biomedical engineering. 2004. Vol. 51, no. 7. pp. 1129–1140.
6. Michael Steinbach, Levent Ertöz, Vipin Kumar, The Challenges of Clustering High Dimensional Data // New Directions in Statistical Physics. 2004. pp. 273–309.
7. M. Kotas, Robust projective filtering of time-warped ECG beats// computer methods and programs in biomedicine 2008. pp. 161–172.
8. Paolo Buono, Aleks Arisb, Catherine Plaisant, Amir Khellab, Ben Shneidermanb, Interactive Pattern Search in Time Series // VDA 2005. pp. 1–11.