



OSTIS-2016

(Open Semantic Technologies for Intelligent Systems)

УДК 004.891.3

ПРИМЕНЕНИЕ СЕТЕЙ БАЙЕСА ДЛЯ АНАЛИЗА ЭКГ

Саган В.Ю., Фридман Р.

*Институт прикладного системного анализа Национального технического университета Украины
«Киевский политехнический институт», г. Киев, Украина*

Vitaliysagan14@gmail.com

Ronfridman1995@gmail.com

В данной работе описаны исследования в области автоматизированного анализа работы сердца по электрокардиограммам с применением машинного обучения. В частности описаны результаты применения байесовских сетей и наивного байесовского классификатора для определения наличия кардиологических патологий. Полученные результаты указывают на то, что данный подход применим к поставленной задаче. Наилучшие результаты были получены с использованием улучшенной байесовской сети.

Ключевые слова: электрокардиограмма; байесовский классификатор; байесовские сети; машинное обучение; генетический алгоритм; сердечнососудистые заболевания.

Введение

Проблема сердечнососудистых заболеваний (ССЗ) крайне остро стоит в современном обществе. По данным отчета всемирной организации здравоохранения ССЗ являются причиной более 30% всех смертей в мире, что делает их основной мировой причиной смертности. Таким образом, профилактика сердечнососудистых заболеваний является крайне важной. [FridmanR., SaganV., 2015]

На данный момент на смену классическому для медицины подходу, предусматривавшему в основном реагирование врачей на возникшие заболевания, приходит новая парадигма. Она называется Predictive, Preventive and Personalized Medicine (PPPM) что означает прогностическая профилактическая персонализированная медицина. Основной акцент в этой парадигме сделан на профилактике заболеваний и их лечении на ранних стадиях. В частности она предусматривает возможности внедрения профилактических и диагностических средств и методов повседневную жизнь пациентов, что с одной стороны позволит снизить нагрузку на врачей, а с другой – увеличит качество контроля здоровья пациентов.

На данный момент специалисты используют различные методы анализа состояния сердца, например УЗИ и МРТ, но базовым методом остается электрокардиография (ЭКГ). ЭКГ – это методика регистрации и исследования

электрических полей, образующихся при работе сердца. ЭКГ записывается при помощи специального устройства – электрокардиографа. Он считывает данные о разности электрических потенциалов в различных четко определенных точках тела человека и затем строит по этим данным электрокардиограмму.

Данные для электрокардиографа поступают с установленных на теле человека электродов. Традиционно они расположены на специальных присосках, которые должны быть размещены в четко определенных местах. Однако в последнее время появилось множество решений использующих носимые датчики типа Holter, которые позволяют считывать данные даже человеку, не имеющему специальной подготовки. Из этого следует, что в случае создания качественной системы, анализирующей данные полученные электрокардиографом, в ближайшее время контроль данных ЭКГ может стать ежедневной профилактической процедурой, особенно для людей, которые находятся в зоне риска.

1. Анализ ЭКГ

1.1. Схема записи данных на ЭКГ

Классическая схема записи ЭКГ подразумевает установку 12 электродов, по данным с которых формируется 12 отведений. Каждое отведение представляет собой временной ряд, который описывает некоторую картину электрической активности сердца. Таким образом, отведения

одинаковы по своей природе, однако могут содержать в себе разную информацию. На практике для анализа состояния сердца чаще всего используют 6 отведений, которые обозначаются как I, II, III, VR, VL и VF. Основным объектом анализа при обработке ЭКГ являются зубцы, которые отражают различные этапы сердечного сокращения. Признаки, используемые методами машинного обучения при обработке ЭКГ, могут быть получены непосредственно из зубцов (длина S-T интервала, высота R-зубца), однако это не единственный способ их получения (рис. 1).

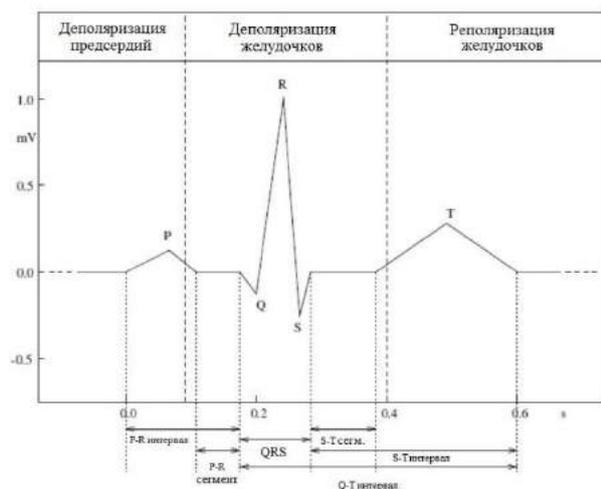


Рисунок 1 – Отображение удара сердца на ЭКГ человека

1.2. Выделение признаков из ЭКГ

Методы машинного обучения не могут быть непосредственно применены к сигналу ЭКГ и требуют преобразования исходного сигнала. В дальнейшем алгоритмы машинного обучения позволят принять решение, насколько здоров пациент. Однако на начальном этапе необходима предобработка сигнала по схеме, показанной на рисунке 2.

Авторами представлена классификация различных типов признаков: Признаки выделяемые из ЭКГ классифицируются на: интервальные признаки, морфологические признаки, основанные на изучении структуры ЭКГ сигнала, признаки, в основе которых лежит вейвлет-преобразование, признаки, полученные при переводе ЭКГ сигнала в частотную область (с использованием быстрого преобразования Фурье) и признаки, сформированные на основе формы сигнала (shapematching).



Рисунок 2 – Этапы предварительной обработки сигнала ЭКГ

Каждый класс, может содержать от нескольких десятков до нескольких сотен различных признаков, которые захватывают некоторые специфические характеристики кривой ЭКГ и содержат информацию о сердечных патологиях.

Во время фазы обучения модели выбраны

наиболее информативные признаки среди всех признаков и используются для дальнейшего обучения модели. Все остальные признаки были отброшены. Детальнее об этой фазе в 3 разделе. Во время рабочей фазы все извлеченные признаки из каждого комплекса для каждого отведения хранятся в векторах. Затем каждый заданный вектор для каждого периода времени (например, для 5 комплексов) упакован в столбец матрицы признаков (featurematrix). Матрица признаков используется в дальнейшем в качестве входа для байесовской сети. [Chereda H.O., Nikolaiev S.S., 2015]

При выполнении исследования была использована база данных, собранная специалистами из университетов BilkentUniversity и BaskentUniversity в Анкаре, Турция. В ней содержатся данные о кардиограммах 452 людей. Каждой кардиограмме соответствует 279 различных параметров. В первую очередь это основные данные о пациенте, например пол, возраст, рост и вес. В выборке не содержатся записи ЭКГ, однако в ней указаны некоторые характеристики каждой из них. Например, указаны средние значения длин некоторых интервалов, амплитуды некоторых зубцов и также присутствие или отсутствие зубцов характерной формы (например, оборванных или двухфазных). В выборке содержатся все вышеназванные характеристики каждого из 12 отведений. Помимо этого каждой записи соответствует класс, к которому она относится. Например, первый класс соответствует записям здоровых людей, второй - ишемической болезни сердца а пятый - синусовой тахикардии. Всего таких классов 12.

2. Применение байесовских сетей

2.1. Описание байесовского классификатора

В медицинском контексте, условная вероятность – это вероятность того, каким будет заключение С, учитывая некоторые свидетельства / наблюдения, Е, где существует зависимость между С и Е. Эта вероятность обозначается как

$P(C|E)$, где

$$P(C|E) = \frac{P(E|C) \times P(C)}{P(E)} \quad (1)$$

Теорема Байеса - метод нахождения апостериорной вероятности с использованием условной

$$P(E|C) = \frac{P(C|E) \times P(E)}{P(C)} = \frac{P(C, E)}{P(C)} \quad (2)$$

Этот условная связь позволяет исследователю получить информацию о вероятности С или Е с известным результатам другого. Теперь рассмотрим сложную проблему с n переменными, где связь между ними не известна для целей прогнозирования

выходной переменной одного класса (например, узел 1 на рисунке 3).

Если все переменные относились к одному общему распределению, эквивалентному всем узлам, которые являются родителями первого уровня, число возможных комбинаций переменных будет равно $(2^n - 1)$.

Для каждой комбинации, должно быть задействовано достаточное количество образцов, чтобы получить реальную оценку вероятности. Это приводит к необходимости использования очень большого количества данных. Если зависимость отношений между переменными может быть определена, то в результате независимые переменные будут удаляться, и меньше узлов будет находиться рядом с целевым узлом. Эти удаления родительских узлов приводят к значительному снижению количества комбинаций переменных, тем самым уменьшая количество необходимых данных. При наличии пропусков в данных условная вероятность определяется с помощью переменных, которые присутствуют. Способность байесовских сетей справиться с пропусками в данных и ее низкие требования к качеству информации, полученной на основе априорных переменных, являются ее основными преимуществами. [M. Wiggins and others, 2008]

Для этого исследования, были разработаны два метода сетей: первый с помощью классификатора поиска «a greedyhill-climbsearch» и второй, основанный на эволюционных вычислениях с использованием классификатора. Была измерена производительность обоих методов и построено сравнение с точностью классификации, полученной обычным байесовским классификатором.

2.2. Применение генетического алгоритма для выделения наиболее весомых признаков

При решении комбинаторной задачи исследования байесовской сети также может быть использован генетический алгоритм. Общий вид усовершенствованного генетического алгоритма показан на рисунке 3.

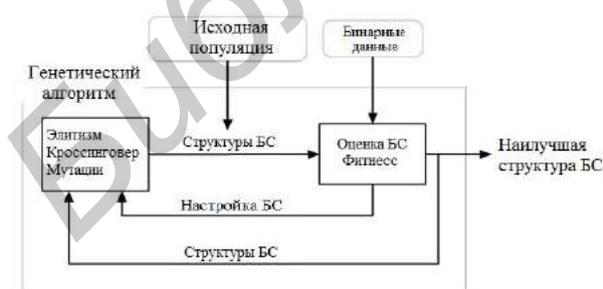


Рисунок 3 – Общий вид усовершенствованного генетического алгоритма

Алгоритм может быть использован для построения новых структур по результатам, полученным для начальной структуры. Итерационные случайные мутации и кроссоверы сетей внутри популяции проверяются и наиболее приспособленные из них сохраняются для

следующих поколений. После прохода поколения популяция эволюционирует, оставляя наилучшие структуры. Плохие структуры удаляются. Улучшение достигнуто в том числе потому, что структура полученной сети является динамической. Это позволяет реализовать метод интеллектуального построения модели без исчерпывающего поиска всех возможных комбинаций структуры узлов. [Ernest w. Lau and others, 2000]

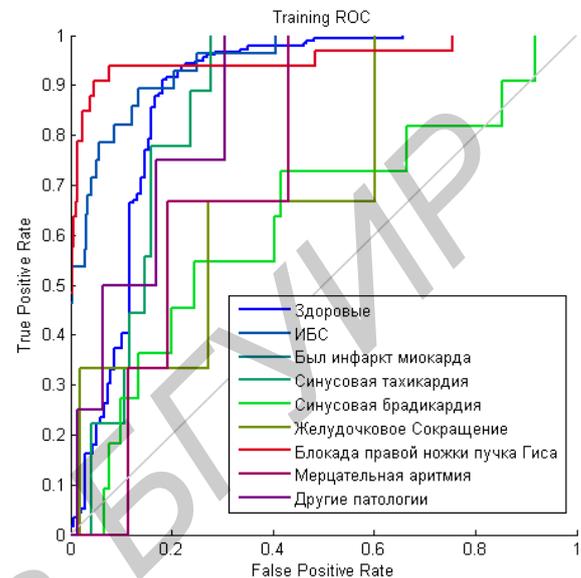


Рисунок 4- ROC-кривые, построенные для классификации патологий по ЭКГ с использованием байесовской сети и генетического алгоритма

3. Анализ полученных результатов

В ходе проведения исследования были получены достаточно интересные и многообещающие результаты. С применением генетического алгоритма была построена улучшенная Байесовская сеть. Она показала довольно хорошие результаты на некоторых из классов, представленных в выборке. Результаты применения улучшенной байесовской сети представлены на рисунке 4. Были выделены 9 классов, на которых классификатор работает лучше всего и для них были построены ROC-кривые. Наиболее приятными являются неплохие результаты при определении ишемической болезни сердца (ИБС), которая является основной причиной смертности в мире и следовательно наиболее опасным из ССЗ. Также получены хорошие результаты для блокады правой ножки пучка Гиса, синусовой тахикардии и класса здоровых людей.

Заключение

Во время проведения данного исследования была подтверждена возможность применения алгоритмов машинного обучения, а именно байесовских сетей, в автоматизированном анализе ЭКГ. В совокупности с возникновением новых технических средств, позволяющих упростить считывание ЭКГ и распространением доступных

вычислительных машин достаточно высокой мощности, это позволяет рассматривать совершенно новый подход к профилактике ССЗ.

Этот проект был разработан при поддержке
IHSTOP IASA Outsourcing Training and Production
Centre "ITSEA".

Библиографический список

[ERNEST W. LAU and others, 2000] ERNEST W. LAU. The Bayesian Approach Improves the Electrocardiographic Diagnosis of Broad Complex Tachycardia / ERNEST W. LAU [and others] // PACE 2000; 23 [Pt. I]: 1519–1526

[M. Wiggins and others, 2008] Evolving a Bayesian classifier for ECG-based age classification in medical applications / M. Wiggins [and others]; Applied Soft Computing 8 (2008) 599–608

[Парадинец, 2008] Парадинец, А. В. Генетический алгоритм диагностики сердечнососудистых заболеваний / А. В. Парадинец // Сборник работ 59-й научной конференции студентов и аспирантов БГУ 2002 год

[Fridman R., Sagan V., 2015] Fridman R., Sagan V. Approach to ischemia detection using artificial neural network/ Fridman R., Sagan V. // Kyiv-Cherkasy, Ukraine. – Cherkasy, 2015. – P. 155-156.

[Chereda H.O., Nikolaiev S.S., 2015] Chereda H.O., Nikolaiev S.S. Architecture of automatic machine learning system for heart pathologies detection using ECGs/ Chereda H.O., Nikolaiev S.S. // SAIT 2015, Kyiv, Ukraine, June 22-25, 2015. Proceedings. – ESC "IASA" NTUU "KPI", 2015. – P. 27-28.

APPLICATION OF BAYESIAN NETWORKS FOR ANALYSIS OF ECG

Fridman R., Sagan V.Y.,

*National Technical University of Ukraine "Kyiv
Polytechnic Institute"*

*Institute for Applied System Analysis,
Kyiv, Ukraine*

Vitaliysagan14@gmail.com

Ronfridman1995@gmail.com

Introduction

This paper describes the research in the field of automated analysis of the heart by electrocardiograms (ECGs) using machine learning algorithms. In particular, it describes the results of the application of Bayesian networks and naive Bayesian classifier in detection of the presence of cardiac pathologies. The results that were obtained indicate that this approach can be applied to the task. The best results were obtained by using the improved Bayesian network.

Main Part

Machine learning methods cannot be directly applied to the raw ECG signal and require specific ECG features to make a decision how healthy the patient is.

ECG signal may contain from dozens to a couple of hundreds of different features that capture some specific characteristics of the ECG complex curve and contain information about heart pathologies.

During model learning phase the most informative features are selected among all the features and are

saved for the model training. All the rest features are truncated. During working phase all extracted features from each complex for each lead are stored in the vectors. Then each given vector for each time frame (for e.g. 5 complexes) is packed as the row of the feature matrix. Then this matrix is used further as an input for machine learning methods.

More than 200 heart pathologies exist and they have different influence into ECG. Some of them (like arrhythmias) influence on temporal pattern of ECG activity and are found with temporal pipeline of the system. Other pathologies influence on complexes' parameters (influence complexes' parameters) and are detected by pattern matching pipeline.

To achieve better performance a set of models which consist of Bayesian network and a greedy hill-climb search is applied to feature matrix.

Conclusion

This study has confirmed the possibility of application of machine learning algorithms, namely, Bayesian networks, for automated ECG analysis. In conjunction with the emergence of new technical means to simplify the collection of the ECG data and the spread of affordable computing devices of sufficiently high power, it allows us to consider a new approach to prevention of cardiovascular diseases.