

9. Егорова В.Н.. Интерлейкин-2: опыт клинического применения в педиатрической практике / В.Н. Егорова, И.В. Бабаченко, М.В. Дегтярева, А.М. Попович. – СПб: Издательский дом "Новости правопорядка", 2009. - 64 с.

ИСПОЛЬЗОВАНИЕ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ АНАЛИЗА МЕДИЦИНСКИХ ДАННЫХ

М.Д. Шагарова
(г. Томск, Томский политехнический университет)
e-mail: mds1@tpu.ru

APPLYING SOFTWARE FOR INTELLIGENT ANALYSIS OF MEDICAL DATA

M.D. Shagarova
(Tomsk, Tomsk Polytechnic University)

The paper dwells on the search of Data Mining Software. This paper describes of comparing software which support Decision Tree Method, gives description of the uses results Decision Tree Method In decision support system.

Key words: software, decision tree, data mining, knowledge base, decision support system

Актуальность данной темы определяется необходимостью применения результатов интеллектуального анализа данных, полученных с помощью специального программного обеспечения для формирования базы знаний системы поддержки принятия решения диагностирования заболеваний. Целью данной работы является обзор программного обеспечения интеллектуального анализа данных, включающих метод построения деревьев решений, результаты работы которого представляются как в графическом виде, так и в виде решающих предложений.

Неотъемлемой частью системы поддержки принятия решений является интеллектуальный компонент, с помощью которого:

- осуществляется поиск функциональных и логических закономерностей в накопленных данных в базе данных информационной системы;
- на основе применяемых правил к новым исследуемым данным осуществляется соответствующий вывод предполагаемого прогноза.

В зависимости от масштаба решаемых задач может потребоваться большое количество правил для применения к данным разного характера. Поэтому необходимо такое средство интеллектуального анализа данных, которое позволит не только выявить скрытые закономерности, интерпретировать их, но и применить результаты анализа для построения интеллектуального компонента системы поддержки принятия решений.

Среди способов для прогностических систем часто используются продукционные модели. Для обработки накопленного массива данных (архивных данных) необходим такой метод интеллектуального анализа данных, с помощью которого можно выявить зависимости и на их основе построить правила логического вывода. В качестве средств решения задач интеллектуального анализа данных выбраны деревья решающих правил. Выбор метода дерева решений объясняется возможностью применения результатов для построения правил в виде предложений типа *если (условие), то (действие)* и дальнейшего их применения в системе поддержки принятия решений. Достоинством деревьев решений является возможность интерпретировать результат, то есть понять, какими решениями руководствуется модель при получении прогноза.

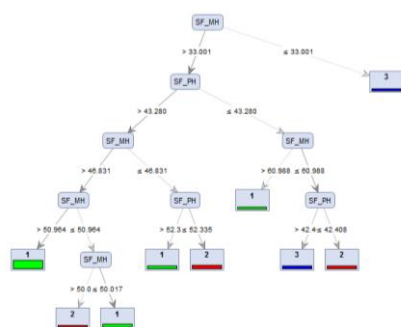
Для задач интеллектуального анализа разработано и применяется специальное программное обеспечение. Сегодня рынок программного обеспечения интеллектуального анализа данных, позволяющих выявить в них закономерности, представлен как коммерческими предложениями, так и предложениями с открытым кодом. Таким программным обеспечением является: SAS Enterprise Miner, SPSS Clementine, Weka, RapidMiner, R, STATISTICA, Data Miner, Orange Canvas, DataEngine, DBMiner, WebMiner, MARS, Datamite, GainSmart, XLMiner, IntelligentMiner, Darwin, AI Trilogy, Alice, AnswerTree, BrainMaker, JDBCMiner, Braintel, DecisionTime и т.д.

Среди критериев поиска предлагаемых решений в контексте данного озора является наличие в них возможности реализации метода построения дерева решений. Основным шагом выбора инструмента является создание хорошей модели, также наличие лучшей интерпретации результатов для построения системы поддержки принятия решений.

В результатах исследований [2] отмечено, что с помощью алгоритма дерева решений в RapidMiner легко получить такие параметры модели, как: более высокую точность классификации 92.49%, наименьшую сложность и высокую производительность с точки зрения обработки данных или скорости реализации. В отчете [3] приводится экспертная оценка и результаты тестирования Weka, RapidMiner. Такие характеристики как предпроцессинг данных, визуализация данных, удобство использования RapidMiner имеют более высокую оценку по сравнению с Weka. Также общая оценка сравниваемых характеристик выше у RapidMiner.

Учитывая проведенный обзор, в качестве инструментария для построения дерева решений и логических правил выбрана среда RapidMiner.

Интерпретация результатов в RapidMiner представлена как в графическом виде, так и в виде текста (рис. 1). На рисунке представлены данные, полученные в ходе формирования базы знаний для системы поддержки принятия решений о диагностировании заболеваний[4].



Фрагмент:

```
HADS_T ≤ 14.500
| SF_MH ≤ 60.988
| | SF_PH > 42.408: G20 {G20=6,
G35=0, G24=0, здоров=0}
| | SF_PH ≤ 42.408: G24 {G20=0,
G35=0, G24=6, здоров=0}
```

Рис. 3 Пример дерева, построенного с помощью средств RapidMiner

Пример кодирования правила (фрагмент):

```
IF HADS_T ≤ 14,500 AND SF_MH ≤ 60.988 AND SF_PM >42,408 THEN: D = G20
IF HADS_T ≤ 14,500 AND SF_MH ≤ 60.988 AND SF_PM ≤ 42,408 THEN: D = G24
```

Деревья принятия решений – один из популярных методов помощи в принятии решений. Построение деревьев решений позволяет наглядно разобраться в структуре данных, создать работающую модель классификации данных. Рассматриваемое специальное программное обеспечение позволяет исследовать зависимости в большом объеме и находить скрытые закономерности. Использование RapidMiner позволяет получить решение не только в виде графического дерева, но и в виде решающих предложений. Данные решения можно представить на языке программирования и применить их при разработке интеллектуального компонента системы поддержки принятия решений, в этом и заключается практическая значимость данного исследования.

ЛИТЕРАТУРА

1. Ida Moghimipour, Malihe Ebrahimpour. Comparing Decision Tree Method Over Three Data Mining Software // International Journal of Statistics and Probability; Vol. 3, No. 3; 2014
2. V.Ramesh, P.Parkavi, K.Ramar. Predicting Student Performance: A Statistical and Data Mining Approach // International Journal of Computer Application (0975 – 8887) Volume 64 – No.8, February 2013
3. Open Source Data Mining Software Evaluation // US Centers for Disease Control and Prevention. URL:<http://www.phiresearchlab.org/downloads/OpenSourceDataMining.pdf>
4. Марухина О.В., Мокина Е.Е., Берестнева Е.В. Применение методов Data Mining для выявления скрытых закономерностей в задачах анализа медицинских данных //Фундаментальные исследования. 2015. № 4. С. 107 -113.

СЕГМЕНТАЦИЯ МРТ-ИЗОБРАЖЕНИЙ С ПОМОЩЬЮ АЛГОРИТМА К-СРЕДНИХ: ДОСТОИНСТВА, НЕДОСТАТКИ

С.А. Эль-Хатиб

(г. Таганрог, Южный Федеральный Университет)

e-mail: samer_elkhatib@mail.ru

MRI IMAGE SEGMENTATION USING K-MEANS ALGORITHM: ADVANTAGES, DISADVANTAGES

S.A. El-Khatib

(Taganrog, Southern Federal University)

Abstract. K-means segmentation algorithm is proposed for brain MRI images segmentation. Advantages and disadvantages for using this algorithm are listed. Research results and future improvements by combining K-means and bio-inspired algorithms are proposed.

Keywords: K-means, fuzzy C-means, image segmentation, bio-inspired algorithms, MRI.

Введение. Сегментацией называется процесс разбиения изображения на области, имеющие различие по определенному признаку или признакам. Данный процесс нашел широкое применение во многих областях, в том числе в медицинской диагностике, являясь одним из предварительных этапов при классификации снимков. Медицинские снимки являются сложными изображениями.

На данный момент предложено большое количество алгоритмов сегментации. Но ни один алгоритм не является универсальным. Результативность работы каждого алгоритма зависит в первую очередь от природы обрабатываемого изображения. В данной статье приводится исследование общеизвестного алгоритма к-средних (k-means), применительно к МРТ-изображениям головного мозга. Проведено сравнение с известным алгоритмом Fuzzy C-means по времени выполнения и качеству работы на снимках хорошего качества.

Алгоритм к-средних. Наибольшее распространение среди методов кластеризации получил алгоритм к-средних, более известный как быстрый кластерный анализ[2]. В данном методе требуется предположение о наиболее вероятном числе кластеров. Описание алгоритма:

1) *Первичное распределение точек по кластерам*

Выбираются начальные К центров кластеров. У каждого кластера может быть только один центр. Выбор первичных центров может происходить одним из следующих образов: случайный выбор, выбор первых к-наблюдений, отбор к-наблюдений с целью максимизации первичного расстояния. На первом этапе каждая точка в результате относится к определенному кластеру.