

Review Paper on Application of Data Mining on Healthinformatics

Snehal H. Chaflekar^{1*}, R. Shinganjude²

^{1,2}Information technology, PBCOE,RTMNU,Nagpur, India

*Corresponding Author: snehalchaflekar@gmail.com, Tel.: 9096764935

Available online at: www.ijcseonline.org

Accepted: 12/Aug/2018, Published: 31/Aug/2018

Abstract— Data mining is a relatively new field of research whose major objective is to acquire knowledge from large amounts of data. In medical and health care areas, due to regulations and due to the availability of computers, a large amount of data is becoming available. On the one hand, practitioners are expected to use all this data in their work but, at the same time, such a large amount of data cannot be processed by humans in a short time to make diagnosis, prognosis and treatment schedules. Data mining in the medical domain works on the past experiences (data collected) and analyse them to identify the general trends and probable solutions to the present situations. This paper is concerned with the application of data mining techniques in the domain of the medical field of heart diseases/attack. We carried out extensive experiments applying different data mining techniques including Relevance analysis, Association Rules Mining and Clustering. We report the findings which are very promising.

Keywords— Diagnosis, prognosis, clustering

I. INTRODUCTION

Data mining is currently implemented in many clinical environments where relationships and patterns provide new medical knowledge. Most of these applications consist of database which holds vast amounts of data. ‘The typical data mining processes involves transferring data originally collected in production systems in a data warehouse cleaning or scrubbing the data to remove errors and check for consistency or formats, and then mining the data using the different techniques[1]. The deficiency that exists at present is that various data mining techniques are available but patterns and rules discovered are vary from one another which could resulted in different diagnoses.

Data Mining is the process of extracting useful information and patterns from enormous data. Data Mining includes collection, extraction, analysis and statistics of data. It is also known as Knowledge discovery process, Knowledge Mining from Data or data/ pattern analysis. Data Mining is a logical process of finding useful information to find out useful data.

II. DATA MINING TECHNIQUES

One of the most important task in Data Mining is to select the correct data mining technique. Data Mining technique has to be chosen based on the type of business and the type of problem your business faces. A generalized approach has to be used to improve the accuracy and cost effectiveness

of using data mining techniques. There are basically seven main Data Mining techniques which is discussed in this article. There are also a lot of other Data Mining techniques but these seven are considered more frequently used by business people.

1. Statistics
2. Clustering
3. Visualization
4. Decision Tree
5. Association Rules
6. Neural Networks
7. Classification

A) Clustering

By examining one or more attributes or classes, you can group individual pieces of data together to form a structure opinion. At a simple level, clustering is using one or more attributes as your basis for identifying a cluster of correlating results. Clustering is useful to identify different information because it correlates with other examples so you can see where the similarities and ranges agree.

Clustering can work both ways. You can assume that there is a cluster at a certain point and then use our identification criteria to see if you are correct. The graph in Figure 3 shows a good example. In this example, a sample of sales data compares the age of the customer to the size of the sale. It is not unreasonable to expect that people in their twenties (before marriage and kids), fifties, and sixties

(when the children have left home), have more disposable income.

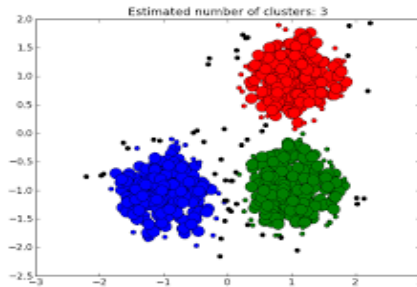


Figure-1

In the example, we can identify two clusters, one around the US\$2,000/20-30 age group, and another at the US\$7,000-8,000/50-65 age group. In this case, we've both hypothesized and proved our hypothesis with a simple graph that we can create using any suitable graphing software for a quick manual view. More complex determinations require a full analytical package, especially if you want to automatically base decisions on *nearest neighbor* information.

Plotting clustering in this way is a simplified example of so called *nearest neighbor* identity. You can identify individual customers by their literal proximity to each other on the graph. It's highly likely that customers in the same cluster also share other attributes and you can use that expectation to help drive, classify, and otherwise analyze other people from your data set.

You can also apply clustering from the opposite perspective; given certain input attributes, you can identify different artifacts. For example, a recent study of 4-digit PIN numbers found clusters between the digits in ranges 1-12 and 1-31 for the first and second pairs. By plotting these pairs, you can identify and determine clusters to relate to dates (birthdays, anniversaries).

III. HEALTH INFORMATICS IN BIG DATA MINING

According to definition the data which can't be handle or which is complex and around petabyte in size is big data. In Health informatics is data is this size is not possible, as Big data define 4 V's i.e. Volume, Value, Velocity, Variety, in data. as data in large volume, Big data is not about the it's about the value of the data ,the data is speedily generated ,the data is of different types such as structured or unstructured. Data gathered for health informatics satisfies many of these qualities. Volumes comes from the records of the patients, value comes from the important information from the data, velocity come from the high speed of the data generated, Variety comes from the different types of data of the patients. In US the data mining is health informatics saves the health around \$4500 billion each year.

IV. DATA IMPLEMENTATIONS AND PREPARATION

Data mining itself relies upon building a suitable data model and structure that can be used to process, identify, and build the information that you need. Regardless of the source data form and structure, structure and organize the information in a format that allows the data mining to take place in as efficient a model as possible.

Consider the combination of the business requirements for the data mining, the identification of the existing variables (customer, values, country) and the requirement to create new variables that you might use to analyze the data in the preparation step.

You might compose the analytical variables of data from many different sources to a single identifiable structure (for example, you might create a class of a particular grade and age of customer, or a particular error type).

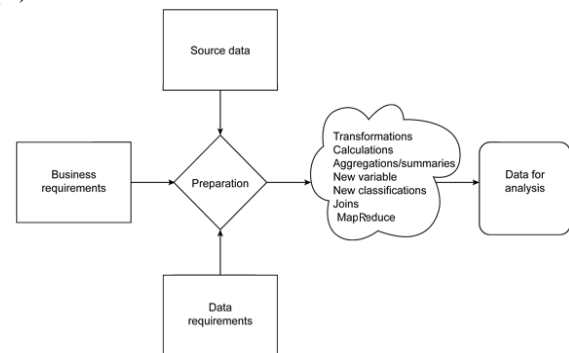


Figure-2

Depending on your data source, how you build and translate this information is an important step, regardless of the technique you use to finally analyze the data. This step also leads to a more complex process of identifying, aggregating, simplifying, or expanding the information to suit your input data.

REFERENCES

- [1] Mrs.a.vanitha, Dr.n.nagadeepa, "Analysis of current applications and issues Of data mining in healthcare", International Journal of Advanced research in Computer science engineering and information technology, 25-oct-2014
- [2] Khalid raza, "Application of data mining in bioinformatics", Indian Journal of computer science and engineering, may 2014.
- [3] Akanksha, vinod maan, "Data mining with big data in health informatics", international journal of computer science trends and technology (ijcst) – volume 5 issue 2, mar – apr 2017
- [4] Snehal Chaflekar, "Intermediate Graphical Language using SDT", International Journal of Innovative Research in Science, Engineering and Technology, Vol. 6, Issue 5, May 2017

Authors Profile

Ms. Snehal Chaflekar, Faculty of PBCOE, Nagpur

Ms. Rupali Shinganjude, Faculty of PBCOE, Nagpur