Chapter 4

Overview of Predictive Modeling Approaches in Health Care Data Mining

Sunita Soni
Bhilai Institute of Technology, India

ABSTRACT

Medical data mining has great potential for exploring the hidden pattern in the data sets of the medical domain. A predictive modeling approach of Data Mining has been systematically applied for the prognosis, diagnosis, and planning for treatment of chronic disease. For example, a classification system can assist the physician to predict if the patient is likely to have a certain disease, or by considering the output of the classification model, the physician can make a better decision on the treatment to be applied to the patient. Once the model is evaluated and verified, it may be embedded within clinical information systems. The objective of this chapter is to extensively study the various predictive data mining methods to evaluate their usage in terms of accuracy, computational time, comprehensibility of the results, ease of use of the algorithm, and advantages and disadvantages to relatively naive medical users. The research has shown that there is not a single best prediction tool, but instead, the best performing algorithm will depend on the features of the dataset to be analyzed.

1. INTRODUCTION

Data mining has successful application on various fields like e-business, marketing and retail have led to its application in KDD in other industries and sectors. Among these sectors recently discovered is healthcare. A wealth of information available in the field of Health care environment. Due to vast storage capacity of computing device a large volume of patient record is available, storing

number of attribute corresponding to different medical test and some more attribute like age and disease by which the patient is suffering. Data mining techniques can be applied to create knowledge rich healthcare environment. This chapter intends to provide a survey of Predictive (detection of disease) data mining techniques that are in use today in medical research and public health. To evaluate the usage in terms of accuracy, computational time, comprehensibility of the

results, an extensively study the various Predictive Modeling technique from basic Predictive algorithm ID3 to advanced technique like Baysian Belief Network, Neural Network etc. in the field of health care Data Mining has been given. We will also compare the different techniques with respect to their advantages, disadvantages and suitability. In section 2 we have discussed the basic definition and classification of data mining methods. In section 3 we have discussed the various predictive techniques in data mining and also the critical review of the current research relating to these predictive modeling approaches in health care data mining. In section 4 conclusion and future work is given.

2. DATA MINING METHOD AND ITS CLASSIFICATION

The KDD is responsible to transform low-level data into high-level knowledge for decision making. Data mining being one of the important steps of KDD is the nontrivial process of identifying valid, new, potentially useful, and ultimately understandable patterns in data. The two primary goals of data mining tend to be prediction and description. Prediction involves using some variables or fields in the data set to predict unknown or future values of other variables of interest. *Description,* on the other hand, focuses on finding patterns describing the data that can be interpreted by humans. Therefore, it is possible to put data-mining activities into one of two categories.

Predictive data mining: Predictive models can be used to forecast explicit values, based on patterns determined from known results. This technique is becoming an essential instrument for researchers and clinical practitioners in medicine. These methods may be applied to the construction of decision models for procedures such as prognosis, diagnosis and treatment planning, which – once evaluated and verified –may be embedded within clinical information systems. *Classification* and *prediction* are major predictive data mining task.

In clinical medicine predictive modeling is particularly useful tool for decision making using patient data. The goal is to derive models that can use patient specific information to predict the outcome of interest and to thereby support clinical decision-making. Such models may be used by clinicians and allow a prompt reaction to unfavorable situations.

Descriptive data mining: The goal of descriptive data mining is to gain an understanding of the analyzed system by uncovering patterns and relationships in large data sets. Association rule mining (ARM) is one of the most important and well researched techniques of data mining for descriptive task. The Descriptive Mining techniques include Clustering, Association Rule Discovery and Sequential Pattern Discovery, are used to find human-interpretable patterns that describe the data in the form of clusters, itemsets, association rules and sequential patterns. Descriptive modeling is not directly useful in clinical medicine for example ARM will only association among various symptoms whereas clustering methods will simply group the patient based on their disease.

3. PRELIMINARIES AND LITERATURE SURVEY OF PREDICTIVE DATA MINING METHODS

3.1. Classification Tree

A Classification trees is a powerful way to extract classification patterns from data; this is an easy to use, popular method from machine learning but infrequently used in health-care environment. Decision tree algorithm is a data mining induction techniques that recursively partitions a data set of records using depth-first greedy approach (Hunts et al, 1966) or breadth-first approach (Shafer et al,

Related Content

An Empirical Analysis of Delhi - Mumbai Sector Flight Fares

T. Godwin (2017). *International Journal of Business Analytics (pp. 60-78).*

www.irma-international.org/article/an-empirical-analysis-of-delhi---mumbai-sector-flight-fares/187209

Server Operating Environment and Business Continuity Drivers

Nijaz Bajgoric (2009). *Continuous Computing Technologies for Enhancing Business Continuity (pp. 79-102).*

www.irma-international.org/chapter/server-operating-environment-business-continuity/7134

A Model for Estimating the Benefits from Network Security Systems

Soumyo D. Moitra (2014). *International Journal of Business Analytics (pp. 1-20).*

www.irma-international.org/article/a-model-for-estimating-the-benefits-from-network-security-systems/117546

Analyzing Business Intelligence Systems Based on 7s Model of McKinsey

Amin A. Shaqrah (2018). *International Journal of Business Intelligence Research (pp. 53-63).*

www.irma-international.org/article/analyzing-business-intelligence-systems-based-on-7s-model-of-mckinsey/203657

The Impact of Data-Complexity and Team Characteristics on Performance in the Classification Model: Findings From a Collaborative Platform

Vitara Pungpapongand Prasert Kanawattanachai (2022). *International Journal of Business Analytics (pp. 1-16).*

www.irma-international.org/article/the-impact-of-data-complexity-and-team-characteristics-on-performance-in-the-classification-model/288517