Medical Domain Knowledge and Associative Classification Rules in Diagnosis

Sung Ho Ha, Kyungpook National University, Korea

ABSTRACT

Hospital information systems have been frustrated by problems that include congestion, long wait time, and delayed patient care over decades. To solve these problems, data mining techniques have been used in medical research for many years and are known to be effective. Therefore, this study examines building a hybrid data mining methodology, combining medical domain knowledge and associative classification rules. Real world emergency data are collected from a hospital and the methodology is evaluated by comparing it with other techniques. The methodology is expected to help physicians to make rapid and accurate diagnosis of chest diseases.

Keywords: Chest Diseases, Data Mining, Emergency Environment, Hybrid Model, Medical Decisions

1. INTRODUCTION

The hospital’s emergency department (ED) is a complex unit in which the fight between life and death is only a breath away. This ED has been frustrated by the problems of overcrowding, long waiting time, patient care delays, and high costs over decades. Accordingly, to solve these problems has become the hottest issue in this area. Several internal or external factors have contributed to the long processing time and patient care delay: patient characteristics, ED staffing patterns, access to health care providers, patient arrival time, management practices, and test and treatment strategies chosen (Fromm et al., 1993). Understanding these factors well is an important step to improve the efficiency of patient care in an ED.

Most hospitals today have employed certain kinds of hospital information systems to manage their healthcare or patient data. These systems typically generate vast amounts of data in the form of number, text, chart, and image. This raises an important question: “How can healthcare practitioners turn that data into useful information that would enable to make intelligent clinical decisions?” Considering the fast growth of data content, size, and diversity, researchers have focused on techniques to find useful information from collections of data.

Although its application to medical data analysis has been relatively limited until recently, the term ‘data mining’ has been increasingly used in the medical literature over
the past few years. The goal of predictive data mining in clinical medicine is to derive models that use specific patient information to support clinical decision-making. Data mining models can be applied to building of decision-making procedures such as prognosis, diagnosis, and treatment planning, which once evaluated and verified, could then be embedded within clinical information systems (Riccardo & Blaz, 2008).

Therefore, the purposes of this study are as follows: first, using data mining techniques, this study focuses on generating the association rules that help physicians to decide which lab tests patients should be tested, which can eliminate unnecessary lab tests to classify disease (e.g., chest pain) and reduce testing time and cost in the ED. Second, this study aims at building a classification scheme that supports to make a complex diagnosis, which can help physicians to formulate clinical decisions more quickly and more accurately. The organization of the paper is as follows: Section 2 explains medical data mining and its application to the emergency department. Section 3 illustrates the research methodology used in this study and section 4 applies the methodology to real-world emergency data. Section 5 evaluates the methodology and compares its performance with other techniques. Section 6 provides conclusions and future directions.

2. LITERATURE REVIEW

Medical data mining has been applied to accurate classification and rapid prediction for prognosis and diagnosis of patients in a specialized medical area (Masuda et al., 2002). It has been also used for training unspecialized doctors to solve a specific diagnostic problem (Kononenko, 2001). Among several algorithms for classification and prediction tasks, a decision tree is one of the most frequently used techniques in a medical data mining area. While it is easy to find many cases to prove the decision tree to be useful in the business domain, the decision tree enables to predict prognoses and diagnoses in the domain of medicine, using tree-structured models or in the form of ‘IF condition-based-on attribute-values THEN outcome-value’ to identify useful features of importance.

Yun (2008) utilized a C4.5 algorithm to build a decision tree in order to discover the critical causes of type II diabetes. She has learned about the illness regularity from diabetes data, and has generated a set of rules for diabetes diagnosis and prediction. Khan et al. (2009) used decision trees to extract clinical reasoning in the form of medical expert’s actions that are inherent in a large number of electronic medical records. The extracted data could be used to teach students of oral medicine a number of orderly processes for dealing with patients with different problems depending on time.

Tan et al. (2007) used the Apriori algorithm to mine the rules for the compatibility of drugs from prescriptions to cure arrhythmia in the traditional Chinese medicine database. The experimental results showed that the drug compatibility obtained by the Apriori algorithm is generally consistent with the traditional Chinese medicine for that disease. Abdullah et al. (2008) adopted an association algorithm to find the relationship between diagnosis and prescription. They stated that purchases and medical bills have much in common. Therefore, the Apriori algorithm was useful to figure out large item sets and to generate association rules in medical billing data.

Delphine et al. has presented a complementary perspective on the activities of the emergency department for specific patient groups: over 75 year old and under 75 year old patients (2008). She thought once validated, these views would be used as decision support tools for delivering better care to this population. Lin et al. (2010) found a way to raise the accuracy of triage through mining abnormal diagnostic practices in the triage. A two-stage cluster analysis (Ward’s method, K-means) and a decision tree analysis were performed on 501 abnormal diagnoses done in an emergency department. Ceglowski et al. (2007) discovered ‘treatment pathways’ through mining medical treatment procedures in the emergency department. They found that the workload in the
Language Focus for Genetics and Molecular Biology Students
[www.igi-global.com/chapter/language-focus-genetics-molecular-biology/76129?camid=4v1a](www.igi-global.com/chapter/language-focus-genetics-molecular-biology/76129?camid=4v1a)

Prediction of Protein-Protein Interactions Between Human Host and Two Mycobacterial Organisms
[www.igi-global.com/article/prediction-protein-protein-interactions-between/40968?camid=4v1a](www.igi-global.com/article/prediction-protein-protein-interactions-between/40968?camid=4v1a)