7th International Conference on Advances in Computing & Communications, ICACC-2017, 22-24 August 2017, Cochin, India

Extracting Association Rules from Medical Health Records Using Multi-Criteria Decision Analysis

Lakshmi K.S[a] [*], G.Vadivu[b]

[a]Department of Information Technology, Rajagiri School of Engineering & Technology, Kochi, Kerala, India
[b]Department of Information Technology, SRM University, SRM Nagar, Kattankulathur, Tamil Nadu, India

Abstract

Medical Health Records serves as rich knowledge sources for data mining. The availability of enormous amounts of medical data can be utilized to extract constructive facts using various data mining techniques. Numerous researches have been conducted in the field of medical data mining. In this paper, a novel approach is suggested to extract association rules from medical records by choosing the best association rule mining algorithm using multiple-criteria decision analysis. The purpose of the project is to discover correlation between diseases, diseases and symptoms, diseases and medicines.

1. Introduction

Data mining is the process of extracting hidden information from large collection of data. There are various fields where data mining has been applied successfully for discovering useful information. Financial data analysis, telecommunication industry, biological data analysis, intrusion detection are some of the major fields where data mining is widely used. It is mostly useful in medical field for effective diagnosis and opting appropriate treatment mechanisms. Massive volume of intricate data is being generated by healthcare industry about hospitals, patients, medical equipments, diseases, claims, treatment cost etc. These data require effective processing to extract useful information.

*Corresponding author. Tel.: +918547833680;
E-mail address:lakshmiks@rajagiritech.edu.in

Data mining comes up with a set of tools and techniques which when applied to this processed data, provides knowledge to healthcare professionals for making appropriate decisions and enhancing the performance of patient management tasks. Patients with similar health issues can be grouped together and effective treatment plans could be suggested based on patient's history, physical examination, diagnosis and previous treatment patterns.

This paper present a new method for the extraction of association rules from medical health records using various data mining algorithms. Careful analysis of medical health records will help in discovering valuable information. Natural Language Processing techniques are used along with Data Mining algorithms for extracting association rules. Association rules reveal correlation between keywords in textual data. Extracting association rules from medical documents helps in finding association among various diseases, diseases and symptoms, diseases and medicines.

2. Literature Review

Classification technique in data mining is widely used for mining data in medical field. Various researches have been done to focus on disease prediction and classification. Lot of prediction systems are available for heart disease detection and progression [1][7]. Kumar Y et.al proposed a new model based on rule based classification for the prediction of liver diseases [2]. In [3], a systematic review has been done on systems for the detection and prediction of apnea events using a classification model.

Clustering and association rule mining are other techniques used in medical data mining. S.Vijayarani et.al [4] developed a new clustering algorithm which is named as weight based k-means algorithm for identifying the leukaemia, inflammatory, bacterial or viral infection, HIV infection and pernicious anaemia diseases from the hemogram blood test samples data set. T.Santhanam et.al [5] has implemented a system using K-Means and Genetic Algorithm for dimensionality reduction and SVM to classify the diabetes dataset. Jung-Kyu Choi et.al [6] conducted a study on data mining techniques for foot disease prediction. Vincenza Carchiolo et.al suggested a new method for the extraction of medical information from Twitter data using NLP techniques [8].

3. System Architecture

The overall architecture of the proposed system is shown in Fig.1. The dataset consists of 309 medical records which are medical transcription files. Natural Language Processing techniques are used for the extraction of medical terms.
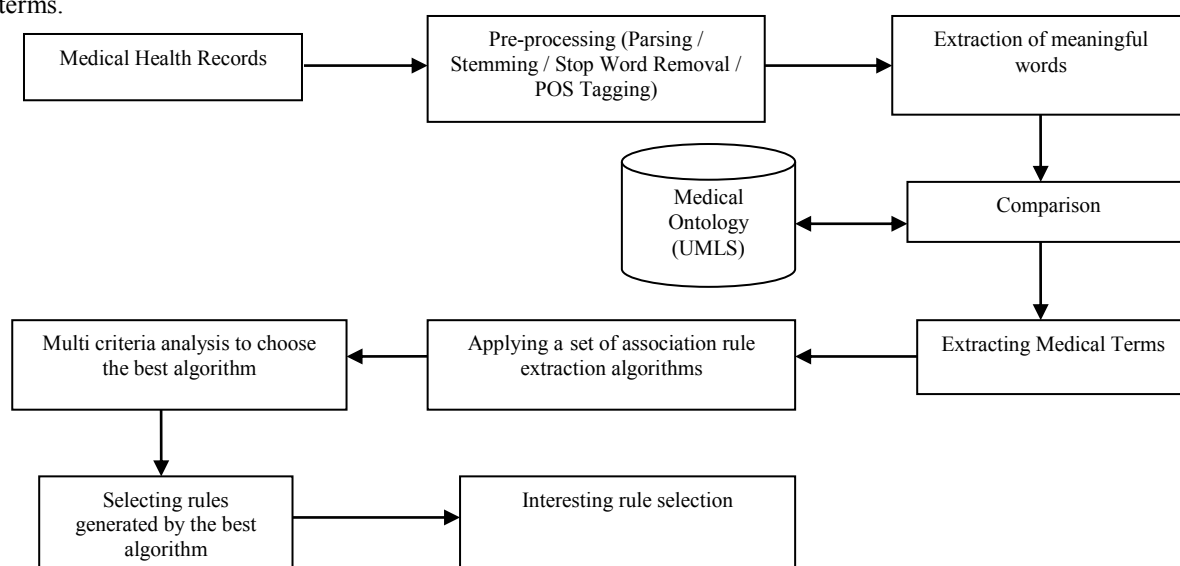

Fig.1.System Architecture