An Automatic Associated Diseases Identification Application

Ghada Mohammed Alsebayel , Sara SaeedAlzhrani, and SulafAlmagooshy

Ghada.sebayel@gmail.com , Sara_alzhrani@hotmail.com, thesulaf@yahoo.com

Information Technology Department
King Saud University, Saudi Arabia
Supervised by
Dr. GhadaBadr
ghbadr@ksu.edu.sa
Assistant Professor, Computer Science Department
King Saud University, Saudi Arabia

Abstract

Important data is usually kept in databases and data warehouses, but as the database grows it becomes significantly difficult to study and explore. One of the domains that cannot function effectively without databases is the medical facilities. They keep critical data and construct important studies that can literally save lives. However, due to the huge database size, this can be a time consuming process that would probably take much human resources. What if there is an easy way to study and analyze huge data? discover only the interesting knowledge? what if this can be done automatically in a matter of seconds?

This is achievable with the help of technology combined with a data mining approach, it can help to analyzes data and discover useful information. Hence, we got inspired to implement an application that can serve medical professionals and is based on data mining approaches so it can automatically discover useful information in datasets. Our tool aims to work on patients' data and find out the frequent patterns of diseases then conclude some rules that associate disease A with some other disease B based on how frequent they usually occur together. The application can serve as an analytical tool when used on huge data files, such as patients records, by discovering and presenting only the interesting relations between diseases that can be found in the database.

For the sake of this application, we focused on identifying associations and frequent patterns on a specific set of diseases: Stroke, heart failure, pulmonary insufficiency, intermittent claudication, atherosclerosis,renal failure, obesity, diabetes and hypertension. The application gives the flexibility of updating the dataset by adding new patients' information.

In our tool, we applied one of the most popular data mining techniques: association rules based on frequent patterns using Apriori Algorithm. The code is developed with an object oriented programming style(java).The basic element is the class patient, which saves a list of diseases infecting a patient. The second class is the combination generator, which generates all possible combinations of diseases. The application can present combinations at different levels based on the user's parameters. It can be a combination of two up to the maximum number of distinct diseases in the dataset. The tool presents the interesting association rules (i.e. the rules satisfying minimum support determined by users).

This application is useful in domains that are in need for data analysis and knowledge discovery in large datasets. For example, doctors can use it to reveal rules associating diseases together. Then they can use these rules to predict what are the diseases that would arise in the patient's future. It can also be useful in studying the most common diseases in a society and constructing a treatment plan accordingly. Thus it will be very useful for studying and researching purposes.

We presented the idea to some doctors and they suggested some enhancements that we hope to encounter in our future work , we hope to implement more user friendly interfaces, also to expand the work to cover more diseases and larger datasets.
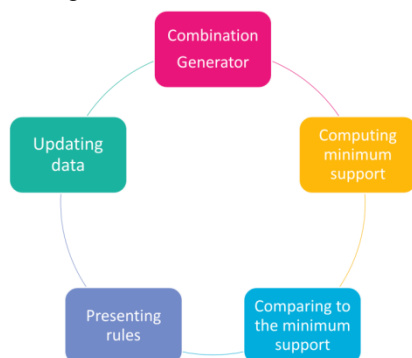
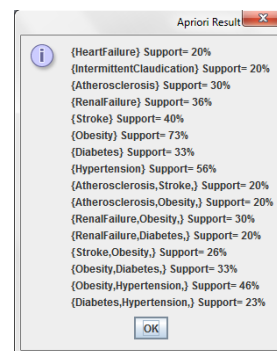Figure1 shows system a high level view of the system implementation .

Figure2 system output.