Abstract

Data mining technologies have been used extensively in the commercial retail sectors to extract data from their “big data” warehouses. In healthcare, data mining has been used as well in various aspects which we explore. The voluminous amounts of data generated by medical systems form a good basis for discovery of interesting patterns that may aid decision making and saving of lives not to mention reduction of costs in research work and possibly reduced morbidity prevalence. It is from this that we set out to implement a concept using association rule mining technology to find out any possible diagnostic associations that may have arisen in patients’ medical records spanning across multiple contacts of care. The dataset was obtained from Practice Fusion’s open research data that contained over 98,000 patient clinic visits from all American states. Using an implementation of the classical apriori algorithm, we were able to mine for patterns arising from medical diagnosis data. The diagnosis data was based on ICD-9 coding and this helped limit the set of possible diagnostic groups for the analysis. We then subjected the results to domain expert opinion. The panel of experts validated some of the most common associations that had a minimum confidence level of between 56-76% with a concurrence rate of 90% whereas others elicited debate amongst the medical practitioners. The results of our research showed that association rule mining can not only be used to confirm what is already known from health data in form of comorbidity patterns,
but also generate some very interesting disease diagnosis associations that can provide a good starting point and room for further exploration through studies by medical researchers to explain the patterns that are seemingly unknown or peculiar in the concerned populations.

References

Extracting Diagnosis Patterns in Electronic Medical Records using Association Rule Mining


Index Terms

Computer Science
Pattern Recognition

Keywords
Medical Diagnosis Patterns  Electronic Medical Records  Health Informatics  Association Rule Mining
Apriori.