Strategy of Automated Clinical Pratices for Reading Electronic Health Records Dynamically According To The Real-Time Fluctuation of The Patient's Condition

P.Kiruthika[1], R.Indra[2], R.Leelavathi[3]
[1, 2, 3] Department of Computer Science
[1, 2, 3] Shrimati Indira Gandhi College, Trichy, Tamilnadu,India

Abstract- *Reasoning is a crucial task performed by the inference engine of the clinical decision support systems, which combines medical knowledge with patient specific data and generates relevant decisions. There are different reasoning methods, suitable for different knowledge representations and application area. This paper reviews the most common methods and describes how they are used in real systems. Furthermore, it outlines the remaining weaknesses of the reasoning mechanisms and provides directions for future research and improvements. Scientists have developed many different reasoning mechanisms, which are available to be used by the inference engine within the clinical diagnosis and treatment system (CDTS). However, even decades since the CDTSs were initially introduced, there are still unresolved problems and no single method has been found to answer all questions. The main purpose of this paper is to review the different reasoning methodologies and to provide directions for further research and improvements.*

Keywords- Big data, case-based reasoning, clinical diagnosis, medical record, disease detection.

I. INTRODUCTION

Data mining is process of extracting hidden knowledge from large volumes of raw data. Data mining is used to discover knowledge out of data and presenting it in a form that is easily understand to humans Disease Prediction plays an important role in data mining. Data mining is used in pensively in the field of medicine to predict disease such as heart disease, lung cancer, breast cancer etc. This paper analyses the heart disease predictions using different classification algorithms. Medicinal data mining has high potential for exploring the unknown patterns in the data sets of medical data. Heart disease was the major cause of causalities in the world. Half of the deaths occur in the countries like India, United States are due to cardiovascular diseases. Data mining is the heart (core) step, which results in the discovery of implicit but potentially valuable knowledge from huge amount of data. Data mining technology provides the user with the methods to find new and implicit patterns from massive data. A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Most hospitals today employ some sort of hospital information systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making [4]. There is a wealth of hidden information in the4ses data is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main motivation for this paper.

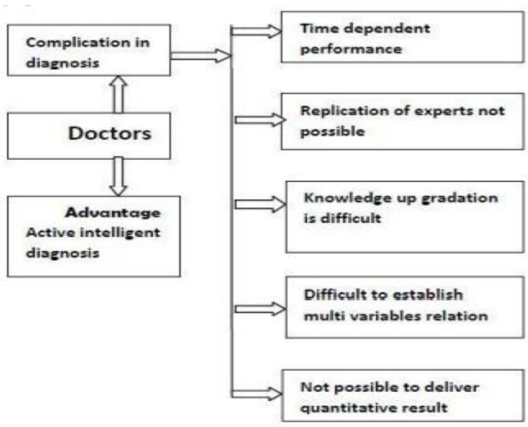

Figure 1.

II. BACKGROUND AND RELATED WORK

A. Semantic transformation model for clinical documents in big data to support healthcare analytics:

The standardized health care documents are being adopted at an exponential rate all around the world which poses several challenges about its Large scale analysis and comprehension. The health care standards are complex and difficult to understand for a health analytics expert due to its comprehensive nature. This paper proposes a semantic transformation model of the health care documents in a distributed environment to tackle the voluminous data and its variety. In this paper Hadoop is used for the semantic transformation model and clinical document architecture (CDA) standard for the case study [1]. The case study shows that the health analytics can be well supported by the transformation model as it is simple and tailor made for the situation. All the stakeholders in the healthcare industry i.e. the physicians, specialists and health insurance companies generate healthcare data in different forms and standards. These standards usually have a very complex structure which makes the healthcare data very challenging. New standards record comprehensive data for every patient. From patient's admission to discharge, many documents are generated comprising of tests, reports, previous history and doctor recommendations etc.

During this process a lot of data is generated as patients today are monitored constantly through different devices [l]. With a lot of data becoming a norm in healthcare, extracting analytics from the data is a big challenge due to performance and scalability issues.

B. Survey of Clinical Data Mining Applications on Big Data in Health Informatics:

Every piece of information learned in human health has the potential to improve the length and quality of life for patients. However, the search for more knowledge in the domain of Health Informatics has led to vast quantities of data, far more than can be easily processed by researchers. Fortunately, just as this problem of Big Data has begun to challenge advances in Health Informatics, so too can the data mining and machine learning techniques used for the general study of Big Data be brought to bear on the these problems. This paper will discuss recent studies done on Big Data Analytics in the field of Health Informatics [2], which sought to answer various clinical questions, using data acquired from the molecular, tissue, and patient levels of Health Informatics. We also consider what work remains to be done in this area. Health informatics is the collection and utilization of information from all aspects of health care, with the goal of predicting what will happen for a patient in a manner that is both efficient and accurate. The end goal of health informatics is Health Care Output (HCO), or the quality of care that the health care field can give to the end users (i.e. patients). Research can use data mining and analytics on medical data in order to devise new ways of making decisions which are faster, more efficient, more accurate, and more cost effective than current methods which do not utilize these tools. Research in health informatics uses data collected at all levels of the medical process from the molecular all the way up to entire populations [6], with the goal of improving HCO. The studies discussed in this paper use data from the molecular, tissue, and patient levels with the goal of answering clinical questions. The breakthroughs that result from these studies can potentially improve the care delivered by physicians to their patients.

C. An Integration Profile of Rule Engines for Clinical Decision Support Systems:

Rule engine has become an indispensable component for many clinical decision support systems. Due to the complexity and heterogeneity of clinical data, one big challenge for rule-based clinical applications is mapping the data from various data sources to rule variables. This paper proposed a rule engine integration profile that uses a shared ontology between the rule engine and external systems to facilitate data acquisition. Based on the integration profile, a diagnostic clinical decision support application was successfully deployed in a chines hospital. Rule engine has become an indispensable component in the practical implementation of knowledge engineering and artificial intelligence. In the healthcare domain, rule engine has been widely used to translate clinical practice guideline into clinical decision support systems (CDSS).Related work in rule engine design and application is abundant. As early as 1985,the NASA (National Aeronautics and Space Administration) agency built the CLIPS engine as a tool for building rule and object based expert systems, and CLIPS is now widely used in business, industry, and academia. After that, several rule engines have been developed. JESS is a rule engine in Java platform, and derived from the rule-based portion of CLIPS.
Fuzzy CLIPS combines fuzzy algorithm with rule based reasoning. JBoss Drools engine is another rule engine in Java platform. There are also commercialized and close sourced rule engines such as Oracle Fusion, IBM Web Sphere and Microsoft WWF Rules Engine. Above-mentioned rule engines are popular in business and industry.

D. A Streaming Ensemble Classifier with Multi-Class Imbalance Learning for Activity Recognition:

Stream multi-class imbalance learning in smart home applications is an evolving learning area that incorporates the challenges of both multi-class imbalance and stream learning. Moreover, another argument in the learning from the imbalanced multi-class distributions that cause misleading classification outcomes is the imbalanced ratio in a sensor data stream which is vigorously changing. Due to the presence of an inadequate representation of sensor data stream and class distribution skews, learning from such data entails a new algorithm to transform balanced data into a model in a stream fashion. In this paper [7], we propose a new multi-class stream imbalance ensemble method where the base learner is a Naive Bayesian classifier. In this approach, each training instance from any of the classes involved in learning based on thresholding on the median prior probability to aid in balancing the classes. Our proposed method diverges from state-of-the-art approaches with regard to being robust to outliers, retains more useful information, and is less sensitive to over-fitting. Also, it has a simple conceptual justification and is easy to implement. We illustrate the effectiveness of the proposed method on two smart home test bed datasets. Our proposed method compares favourably with state-of-the-art approaches. In many real-world applications such as diagnosis of rare diseases, fraud detection, and human activity recognition in smart homes, the data are usually distributed in an imbalanced way i.e. some classes have far more examples than other classes. In addition, in data stream learning [3], data can have a skewed class distribution. Therefore, it is often meaningless to report a high accuracy when there are imbalanced classes, because a minority class will be dominated by one or more majority classes. Learning from such data sets that contain imbalanced classes usually produces biased classifiers that have a higher predictive accuracy over the majority classes, but poorer predictive accuracy over the minority classes that are often of more interest. In such circumstances, though the accuracy seems high, the solution is useless since no minority class will be detected.

III. ALGORITHM

MACHINE LEARNING ALGORITHM

Machine Learning (ML) is coming into its own, with a growing recognition that ML can play a key role in a wide range of critical applications, such as data mining, natural language processing, image recognition, and expert systems. ML provides potential solutions in all these domains and more, and is set to be a pillar of our future civilization. The supply of able ML designers has yet to catch up to this demand. A major reason for this is that ML is just plain tricky. This tutorial introduces the basics of Machine Learning theory, laying down the common themes and concepts, making it easy to follow the logic and get comfortable with the topic. So what exactly is “machine learning” anyway? ML is actually a lot of things. The field is quite vast and is expanding rapidly, being continually partitioned and sub-partitioned ad nauseam into different sub-specialties and types of machine learning.

SEMANTIC INFERENCE ALGORITHM (SI)

Semantic Inferences are steps in reasoning, moving from premises to conclusions. Charles Sanders Peirce divided inference into three kinds: deduction, induction, and abduction. Deduction is inference deriving logical conclusions from premises known or assumed to be true, with the laws of valid inference being studied in logic. Induction is inference from particular premises to a universal conclusion. Abduction is inference to the best explanation. Human inference (i.e. how humans draw conclusions) is traditionally studied within the field of cognitive psychology; artificial intelligence researchers develop automated inference systems to emulate human inference. Statistical inference uses mathematics to draw conclusions in the presence of uncertainty. This generalizes deterministic reasoning, with the absence of uncertainty as a special case. Statistical inference uses quantitative or qualitative (categorical) data which may be subject to random variations.

SENTIMENT ANALYSIS

Sentiment analysis (also known as semantic inference) refers to the use of natural language processing, text analysis and computational linguistics to identify and extract subjective information in source materials. Sentiment analysis is widely applied to reviews and social media for a variety of applications, ranging from marketing to customer service.

CLASSIFICATION

There are two major forms of data analysis that can be used for extracting models describing important classes or to predict future data trends. These two major forms are as follows

- Classification
- Prediction

Classification models predict categorical class and prediction models predict continuous valued functions. For example, we can build classification model [5], to categorize bank loan applications as either safe or risky, or a prediction models to predict the expenditures in dollars of prospective customers on computer equipment given their income and occupation. Following are the examples of cases where the data analysis task is Classification.

Examples

- A bank loan officer wants to analyze the data in order to know which customer (loan applicant) is risky or which are safe.
- A marketing manager at a company needs to analyze a customer with a given profile, who will buy a new computer.

- In both of the above example, a model or classifier is constructed to predict the categorical labels. These labels are unsafe or safe for loan application data and sure or not for marketing data.

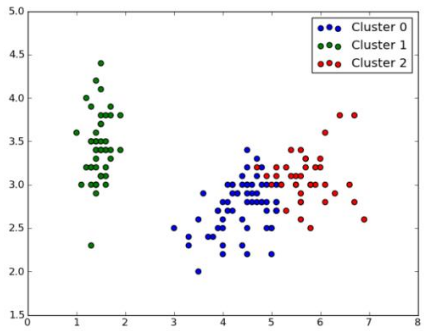

Figure 2.

IV. CONCLUSION

We have gained significant experience in the 8+ years since we first started to apply semantic web technology to electronic medical records in support of clinical research and quality-of-care measurement. The benefits of this technology have now been amply proven over multiple years of production use at the Cleveland Clinic's Heart and Vascular Institute. However, being a pioneer was not always easy. Initially the biggest challenges were due to immature or unavailable semantic web tools. This affected both initial design of Semantic and our work efforts, as we had to build much more of the infrastructure from scratch than would now be necessary.Semantic web tooling has improved greatly since the Semantic project began in 2003, so others embarking on this route will have an easier voyage.

In future work, Clinical Diagnosis and Treatment System to become a more a newly designed semantic inference algorithm supports clinicians in the decision making process as well as saving time and expense for the patients.

REFERENCES

[1] S. Hussain and S. Lee, “Semantic transformation model for clinical documents in big data to support healthcare analytics," published on Oct. 2015.

[2] M. Herland, T. M. Khoshgoftaar, and R. Wald, ``Survey of clinical data mining applications on big data in health informatics". (ICMLA) Dec. 2013.

[3] A. Bifet and R. Gavaldà, “Adaptive learning from evolving data streams," Data Anayze., Aug. 2009.

[4] E. S. Berner, Ed., Clinical Decision Support Systems: Theory and Practice, Springer, 2007.

[5] L. Breiman, J. Friedman, R. Olshen, and C. Stone, Classification and Regression Trees. Monterey, CA, USA: Wadsworth, 1994.

[6] N. Oza and S. Russell, “Online bagging and boosting," in Artificial Intelligence and Statistics. Morgan Kaufmann, 2001.

[7] A. Bifet, G. Holmes, B. Pfahringer, R. Kirkby, and R. Gavalda, “New ensemble methods for evolving data streams," in Proc. 15th ACMSIGKDDt. Conf. Knowledge. Discovery Data Mining (KDD) 2009.