Volume 5, Issue 10, October-2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Ontology Based Disease Diagonsis Algorithms

Abinaya Sambath Kumar
M Phil Research scholar,
Department of computer science,
Dr N.G.P Arts and Science College,
Tamil Nadu, India

A. Nirmala
Assistant Professor (SG),
Department of Computer Applications,
Dr N.G.P Arts and Science College,
Tamil Nadu, India

Abstract - Data mining methods are used to analyse the medical data information resources. In this paper, the data collected from the patient database facilitate the diagnosis process. The key components of the proposed cardiovascular decision support framework have been developed using an ontology driven approach. These records are necessary to analysis the intangible information which might be helpful in crisis situation and this can be achieved by various data mining algorithms.

Keywords: C4.5, K-NN, SVM, Decision Support framework, Ontology driven approach.

I. INTRODUCTION

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. Knowledge discovery is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data. Data mining tools predict behaviors and future trends, allowing businesses to make proactive, knowledge-driven decisions. Data mining tools can answer business questions that traditionally were too time consuming to resolve. They scour databases for hidden patterns, finding predictive information that experts may miss, because it lies outside their expectations. Data Mining derives its name from the similarities between searching for valuable information in a large database. The processes require either sifting through an immense amount of material, or intelligently probing it to find where the value resides. Data mining is a field at the intersection of computer science and statistics, is the process that attempts to discover patterns in large data sets. It utilizes methods at the intersection of artificial intelligence, machine learning, statistics and database systems. The overall goal of the data mining process is to extract information from a data set and transform it into an understandable structure for further use.

Some of the key benefits of data mining include,

1. It is one of the most effective services available today. With the help of data mining, one can discover precious information about the customers and their behavior for a specific set of products and evaluate and analyze, store, mine and load data related to them.
2. An analytical CRM model and strategic business related decisions can be made with the help of data mining as it helps in providing a complete synopsis of customers.
3. An endless number of organizations have installed data mining projects and it has helped them see their own companies make an unprecedented improvement in their marketing strategies.
4. Data mining is generally used by organizations with a solid customer focus. For its flexible nature as far as applicability is concerned is being used vehemently in applications to predict crucial data including industry analysis and consumer buying behaviors.
5. Fast paced and prompt access to data along with economic processing techniques has made data mining one of the most suitable services that a company seeks.

In this paper, Data mining can improve decision-making by discovering patterns and trends in large amounts of complex data. There are two primary goals of data mining-prediction and description. Prediction involves some variables or fields in the data set to predict unknown or future values of other variables of interest. On the other hand Description focuses on finding patterns describing the data that can be interpreted by humans. Many diseases have closely related symptoms which make it difficult for the doctors to predict the exact disease. It helps in predicting the disease which is nearly accurate. Even though the prediction is not very accurate it at least gives the doctor a brief idea what the disease could be. In literature various techniques are available in data mining domain. Ontology's are also created for disease diagnosis. Section 2 reviews the state of the art in the healthcare patient records for ontology driven approach. Section 3 explains our various algorithms for the development of patient records. Section 4 presents conclusions and future work.

II. DATA MINING FOR DISEASE DIAGNOSIS

2.1 Ontology Driven Approach:

Ontology performing specific actions within the ontology domain. The Ontology design incorporated different stages during the referral process through relationships among parent-subclasses. The information about patient's chest pain

type, their past family history, previous cardiovascular history and personal demographics information is also modelled through domain-specific classes within the ontology design.

The adoption of Ontologies inspired approach yields good results in terms of standardizing health care guidelines. The construction of knowledge base through an ontology inspired approach will have the benefit of problem independence. This knowledge base can be extended and reuse in a variety of different problems and therefore will have multiple mapping among knowledge base and decision models. The knowledge base will update the decision models without any costly software engineering work and maintenance will be cost effective across decision models and within the knowledge base.

2.2 Decision Support Framework:

We are developing an ontology driven decision support framework as shown in Fig 1 to solve such knowledge representation and clinical decision making issues in the cardiovascular domain**.**

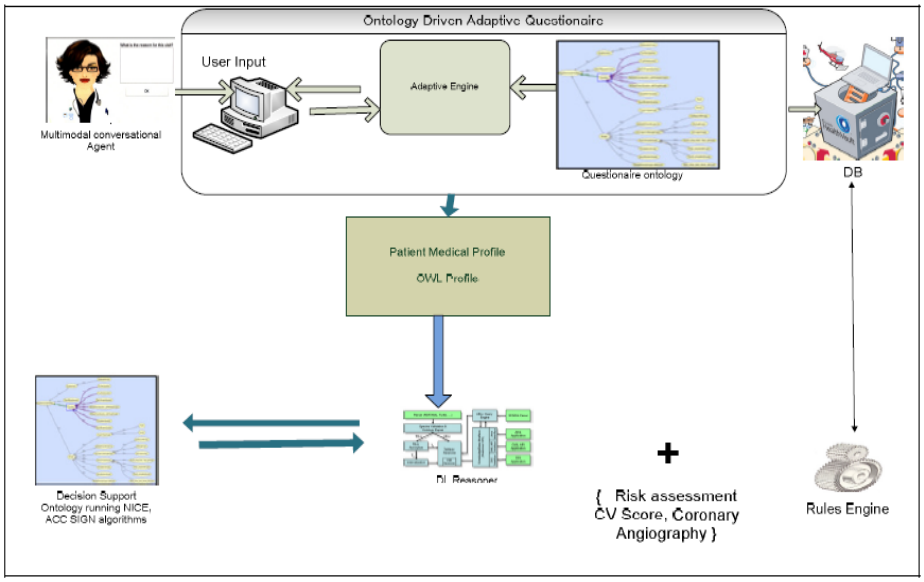

Fig. 1 Cardiovascular Decision Support Framework for Primary and Secondary Care

Patients who present with chest pain continue to present a major diagnostic challenge for both primary and secondary care physicians. This is due, in part, to the low specificity of chest pain as a symptom of significant coronary artery disease and the danger of misdiagnosis in patients at risk of major cardiac events. Furthermore, chest pain is a very common symptom, between 20% and 40% of the general population will experience chest pain in their lives with up to 1% of visits to a general practitioner due to chest pain. Chest pain is also common presenting complaint in patients attending A and E accounting for approximately 5% of visits to the emergency department. Furthermore, up to 40% of emergency hospital admissions are due to chest pain.

III. METHODOLOGY

In this section, we used several algorithms for coronary heart disease. We will discuss elaborately.

3.1 KNN (K means Nearest Neighbour):

K-means clustering algorithm thereafter K-NN classification algorithm is used to classify the data.

1. The first step is the pre-processing where the inconsistent values are removed.
2. The K-means clustering algorithm is used to identify and correct these inconsistencies.
3. The missing values are replaced by means and medians.
4. After the inconsistencies are removed the K-NN classification algorithm is used to classify the data.
5. This model is then tried for different values of *k.*

K means clustering algorithm and decision tree algorithms combined and used for coronary heart disease. We choosing the data set J48 and use the parameter measures such as Accuracy, Sensitivity and specificity and then for this data set and KNN algorithm we get accuracy as 78.17%

3.2 C4.5:

C4.5 algorithm is used as the training algorithm to show rank of heart attack with the decision tree. Finally, the heart disease database is clustered using the K-means clustering algorithm. The results showed that the system is capable of predicting the heart attack successfully.

1. The database is pre-processed successfully by eliminating identical records and providing missing values.
2. The polished data set is then collected by K-means algorithm with the K value of 2.
3. The forms are mined efficiently from the collection applicable to heart disease, using the MAFIA algorithm.

4. The model consortiums of heart attack parameters for ordinary and risk level along with their values and levels are detailed.
5. In that, ID lesser than of (#1) of weight contains the normal level of prediction and higher ID other than #1 comprise the higher risk levels.
6. A subsequent decision tree is created to show the heart attack level using C4.5 algorithm by information gain.

C4.5 proposes a classification algorithm which combines KNN and genetic algorithm, to predict heart disease of a patient. The author uses genetic search as a measure to check redundant and irrelevant attributes, and to rank the attributes which contribute towards classification. Least ranked attributes are removed, and classification algorithm is built based on evaluated attributes. The classifier classifies heart disease data set as either healthy or sick.

1. Apply genetic search on the data set
2. Attributes are ranked based on their value
3. Select the subset of higher ranked attributes
4. Apply (KNN+GA) on the subset of attributes that maximizes classification accuracy
5. Calculate accuracy of the classifier

C4.5 and KNN algorithms combined and used for coronary heart disease. For using this algorithm, the accuracy level is as higher than the KNN and Decision Tree such as 93.73%.

3.3 SVM:

Support vector machine method is based on the subsumption relations among the extracted concepts. Let $Spec(c_x, c_y)$ denote that concept c_x is a specialization (subclass) of another concept c_y. The parameter measures such as specificity, accuracy, sensitivity and it estimated as,

$$\text{Spec} = \frac{c_x}{c_x + c_y}$$

SVM classification is used to predict the patient records for affecting heart disease diagnosis. It is based on classification algorithm and it also predict the heart disease risk level easily. It gives better accuracy compare with KNN and C4.5 algorithm. The accuracy level 95%.

IV. DATA SET AND SIMULATION

4.1 Medical Data Set:

Patient's medical history encapsulates their demographics along with their past and present cardiac and non cardiac related clinical symptoms, location of the current chest pain (left side of the chest) association of the chest pain with breathing, its severity and whether or not their chest pain is spreading. After analyzing the patient's medical history generated through the Patient Semantic Profile ontology the semantics are extracted to generate electronic healthcare records. These EHRs are of utmost importance for the clinicians for the effective risk assessment operations and efficient clinical decision making.

ExampleSet (11 examples, 2 special attributes, 8 regular attributes)

Row No.	Diagnosis	Result	Age	Gender	Pain Type	Known CAD	HyperChol	High BP	Smoking	Family Hist..
1	prob	angio	71	m	typ	y	n	n	n	y
2	prob	angio	77	m	typ	y	y	n	y	y
3	prob	angio	67	m	typ	n	y	y	n	y
4	prob	angio	74	f	typ	n	n	n	n	n
5	prob	angio	52	m	atyp	n	n	n	y	y
6	prob	angio	74	f	typ	n	y	y	y	n
7	pos	ett	66	m	atyp	n	y	n	n	n
8	pos	angio	61	f	typ	n	n	y	y	n
9	prob	angio	64	f	typ	n	n	y	y	n
10	prob	medical	85	f	typ	n	n	n	n	n
11	def	admitted	66	f	typ	n	n	n	n	n

V. CONCLUSION

We have presented an ontology driven based cardiovascular decision support framework. We have discussed the development of CAD which used as several algorithms. Healthcare Records which have been developed as part of the cardiovascular decision support framework. The main advantages of the ontology driven approach are as follows: This allows cost effective maintenance of the chest pain decision support system because of a clear demarcation among knowledge base and decision support functions.

We have done survey on various data mining algorithms for disease diagnosis. Still No method is 100% even doctors diagnosis can't be 100%.Such system are useful for doctors to provide assistance. Various algorithms have been proposed for diagnosis. For multiple disease diagnosis system must provide faster diagnosis result, system must be self learning. Such system must consider various parameters such as recent medical trends, seasonal effects.

REFERENCE

[1] Jones, M., et al.: Systematic review: prognosis of angina in primary care. Family Practice 23, 520 (2006)

[2] Stern, S., et al.: Presenting symptoms, admission electrocardiogram, management, and prognosis in acute coronary syndromes: differences by age. The American Journal of Geriatric Cardiology 13, 188–196 (2004)

[3] Ruigómez, A., et al.: Chest pain in general practice: incidence, comorbidity and mortality.Family Practice 23, 167 (2006)

[4] Nandalur, K.R., et al.: Diagnostic Performance of Stress Cardiac Magnetic Resonance Imaging in the Detection of Coronary Artery Disease: A Meta-Analysis. Journal of the American College of Cardiology 50, 1343–1353 (2007)

[5] Jain, A., et al.: Impact of a Clinical Decision Support System in an Electronic Health Record to Enhance Detection of 1-Antitrypsin Deficiency. Chest 140, 198 (2011)

[6] Zheng, H.-T., Kang, B.-Y., Kim, H.-G.: An Ontology-Based Bayesian Network Approach for Representing Uncertainty in Clinical Practice Guidelines. In: da Costa,P.C.G., d'Amato, C., Fanizzi, N., Laskey, K.B., Laskey, K.J., Lukasiewicz, T., Nickles, M., Pool, M. (eds.) URSW 2005 - 2007. LNCS (LNAI), vol. 5327, pp. 161–173.Springer, Heidelberg (2008)

[7] Bouamrane, M.-M., et al.: Experience of Using OWL Ontologies for Automated Inference of Routine Pre-Operative Screening Tests. Life Sciences

[8] Bouamrane, M.-M., Rector, A.L., Hurrell, M.: Ontology-Driven Adaptive Medical Information Collection System. In: An, A., Matwin, S., Raś, Z.W., Ślęzak, D. (eds.) ISMIS 2008. LNCS (LNAI), vol. 4994, pp. 574–584. Springer, Heidelberg (2008) An Ontology Driven and Bayesian Network 41

[9] Cambria, E., et al.: Sentic PROMs: Application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Systems with Applications

[10] Cambria, E., Hussain, A.: Sentic Computing: Techniques, Tools, and Applications. In: SpringerBriefs in Cognitive Computation. Springer, Heidelberg (2012)

[11] Abidi, S.R., et al.: Ontology-based Modeling of Clinical Practice Guidelines: A Clinical Decision Support System for Breast Cancer Follow-up Interventions at Primary Care Settings Computerization of BC Follow-up CPG Development of Breast Cancer Ontology The BC ontology model. Computer

[12] Hurley, K.F., et al.: Ontology Engineering to Model Clinical Pathways: Towards the Computerization and Execution of Clinical Pathways 2. Developing a Clinical Pathway Ontology: Our Approach. Symposium A Quarterly Journal In Modern Foreign Literatures, 0–5 (2007)

[13] Farooq, K., et al.: Ontology-driven cardiovascular decision support system. In: 2011 5th International Conference on Pervasive Computing Technologies for Healthcare (PervasiveHealth), pp. 283–286 (2011)

