Advances in Computational Sciences and Technology
ISSN 0973-6107 Volume 10, Number 3 (2017) pp. 353-361
© Research India Publications
http://www.ripublication.com

Design an Disease Predication Application Using Data Mining Techniques for Effective Query Processing Results

P. Hariharan
Assistant Professor,
PG &Research Department of Computer Science and Applications,
Adhiparasakthi College of Arts & Science, G.B.Nagar, Kalavai – 632 506,
Vellore District, Tamilnadu, India.

K.Arulanandham
Assistant Professor,
PG &Research Department of Computer Science and Applications,
Government Thirumagal Mills College, Gudiyatham – 632 602,
Vellore District, Tamilnadu, India.

Abstract

In the query processing results environment, there are many different query processing techniques to satisfy the user request. All the Query processing techniques are available for satisfy the common applications. There is no available for individual applications (for example E-learning, banking and health applications etc) In the different application environment, in order to achieve a effective results of query processing techniques are required

This type of effective query processing result techniques are implementing in order to different applications at the time of developing, the respective applications of query processing results will be very effective and needful. At the time of developing the respective application will be easy. When the absence of effective query processing results techniques in any applications, the applications will have take more time for developing and not to be effective and it may be large. This paper gives the idea, what are the algorithms in data mining are well suited, and also analyze how to implementing the specific operation in different applications for effective query processing results. We also propose a framework to analyzing the healthcare support applications with adopting the suitable technique for future in use. Keywords: Service *oriented Architecture (SOA), k-means*

INTRODUCTION

We study only here the possibilities data mining techniques, for implementing the logic of developing queries in different applications. Some of the proposed applications area are E-learning and banking. These experimental results will be implementing in practically through queries and also produce the effective results in future. Specifically, design an application for health system field for diagnosis the diseases and also produce the effective results by data mining technique such as K-means algorithm. Because, the Disease prediction and diagnosis is one of the important applications where data mining tools and techniques are used to providing successful results because of significant improvements in technology. This paper identifies gaps in the research on disease prediction, diagnosis and it also proposes a model to systematically close those gaps.

This proposed work used to predicate the diseases using k-means algorithm, Large Memory Storage and Retrieval (LAMSTAR) & SOA (service oriented architecture). Whereby the system elements of diagnosis, data portal These experimental results will be implementing in practically through queries and also produce the effective results in future.

BACKGROUND AND LITERATURE SURVEY

Given a query, the various query processing techniques to satisfy the user requirements. The selections of values are from zero to hundred percentage. The found of relations will calculate the different query processing techniques being implemented.

To implement effective query processing techniques for the cost model of query processing. Traditionally, query processing does not manage individual application. We introduce the logic of individual application at the query processing environment

Shomona Gracia Jacob et al [1]. Compare the error rates and measures produced by the various classification algorithms on the Wisconsin breast tissue dataset and the effect of feature selection algorithms on improving the accuracy of classification of carcinoma in the breast tissue dataset. The existence of other tissue features like Fibro- adenoma, Mastopathy that indicate a higher risk of developing cancer in future is also classified.

Shweta Kharya [2] discussed the techniques that can be used for breast cancer classification. The predictor can be used to design a web based application to accept the predictor variables and automated system decision Tree based prediction can be implemented in remote areas etc.

M.Akhil jabbar et al [3] developed as generalization of mathematical models of human cognition or neural biology. Two models of neural network have been developed. These models are feed forward neural networks trained with back propagation and radial basis function neural networks. The two models are three layer

networks which are made up of an input that is connected with the hidden layer with aid of connection weight

DATA MINING TECHNIQUES

A) Proposed technique for effective query processing results in various applications (E-learning & Banking):

i) E-learning :

Personalized E-learning demonstrated improving the searches in the WWW. This proposed work personalized E-learning engine PEL that manage the user's preferences in the form of concepts Ontologies data**.** To locate the information in WWW, PEL classifies the logic into content concepts and location concepts. The preferences are organized in an ontology which are used to manage a personalized ranking function for rank adaption in future. We propose personalized E-learning based on profiles and domain ontology. We have to categorize the two ranges in the complete activities one offline tasks; it contains the data preparation ontology and mining the online job that represents as production. We propose an online prediction techniques for conclude to decide personalizing a query use**.** The proposed algorithm as,

1. Genetic Algorithm scheme

2. K-means clustering algorithm

ii) Banking:

The ideas to represent a protocol for secure the rules in association for horizontally distributed database. The aim is to be calculate for all rules with the help of confidence at least for minimal support size and satisfactory level and that contains in the unified database for the current leading protocol. Our protocol is depends upon on the fast Distributed Mining (FDM) algorithm, which is an unsecured version of the Apriori algorithm. The main components in the protocol are two novel secure multi-party methods one that executes the union of private subsets and that each of the interacting sites with another. One of the main components in our protocol is a novel secure multi-party protocol for computing the Union (or intersection) of private subsets that each of the communicating sites hold.

The proposed algorithm as,

1. Apriori Algorithm

2. Fast Distributed Mining (FDM).

B) Proposed technique and implementing features for effective query processing results in disease prediction application:

i) About K-means algorithm

The k-Means algorithm is a distance-based clustering algorithm that partitions the data into a specified number of clusters. Distance-based algorithms rely on a distance function to measure the similarity between cases. Cases are included to the nearby cluster respective to the distance function use The k-Means by specifying any of the following:

- Number of clusters
- Growth factor for memory allocated to hold clusters
- Convergence tolerance

The k-means algorithm well suits the application to produce the predicate of diseases. This implementation of Query processing will be producing the effective results.

ii) Diseases Predication application by using K-means algorithm :

The doctor will be identifying easy the patient diagnosis from the use of data mining techniques (such as k-means algorithm). In this features the doctor will be avoiding to take the number of lab test related to disease There are four databases will maintain for identify the diagnosis from the patient diseases/symptoms. They are,

1. Diagnosis_symptoms_Med_hist
2. Diagnosis_symptoms_lab_test
3. Patient_history
4. Patient_diag_sym _recent_trend_service

1. Diagnosis_symptoms_Med_hist : :

This database will store the details of different diseases/symptoms and their related diagnosis. These details will be collecting through the Medical theory. To Retrieving the details of specific diagnosis from entering the related diseases/symptoms and evaluating the pattern matching symptoms services.

2. Diagnosis_symptoms_lab_test:

This database will store the details of different diseases/symptoms and their related diagnosis. These details will be collecting through the lab test. To Retrieving the details of specific diagnosis from entering the related diseases/sysmptoms and evaluating the pattern matching symptoms services

3. Patient history :

This database will store the details for patient history (such as symptoms, diseases etc) Also the database maintained the details of patient record (such as what are all the treatment given by doctor by specific diagnosis from related diseases/symptoms in previously).

4. Patient_diag_sym _recent_trend_service : This database will store the details of different diseases/symptoms and their related diagnosis. The details will be collecting through recent treatment related by symptoms/diseases given by doctor of specific diagnosis. This database performs automatic updates from patient history database. (by recent treatment).To Retrieving the details of specific diagnosis from entering the related diseases/symptoms and evaluating the pattern matching symptoms services.

First the doctor retrieving the diseases/symptoms from patient history database After retrieving the diseases/symptoms, the doctor will be identifying the diagnosis from the two databases(diagnosis_symptoms_Med_hist,diagnosis_symptoms_lab_test.).The identify lists as (dlistmh1,dlistlt1) will be evaluating by pattern matching services. After the doctor retrieving the diagnoses is from patient_diag_sym_recent_ trend_ service database by pattern matching symptoms services. This diagnosis list name as dreclist1. All the three lists will be short listing by cluster (dlistmh1, dlistlt1, dreclist1). The list will be using to compute the probability of each occurrence of particular diseases from the medical data. The probability will be computing based on the distance vector. The highest priority cluster will be producing the results.(k-means Algorithm).

Moreover, it can be produces the accurate results by LAMSTAR Network for assigning weights. In this part the correct diseases sorting by the doctor. . The final report is then prepared to obtain the correct symptoms. The correct symptom thus obtained is then compared with the original symptoms entered. This information is fed to the LAMSTAR Network for applying weights. If any storing pattern information matches the given subs word within a present tolerance. The system maintains weights based to the following procedure:

$$W_{i,m}(t+1) = W_{i,m}(t) + \acute{a}i\,(X_i(t) - W_{i,m}(t)), \text{for } m: €min$$

Where,

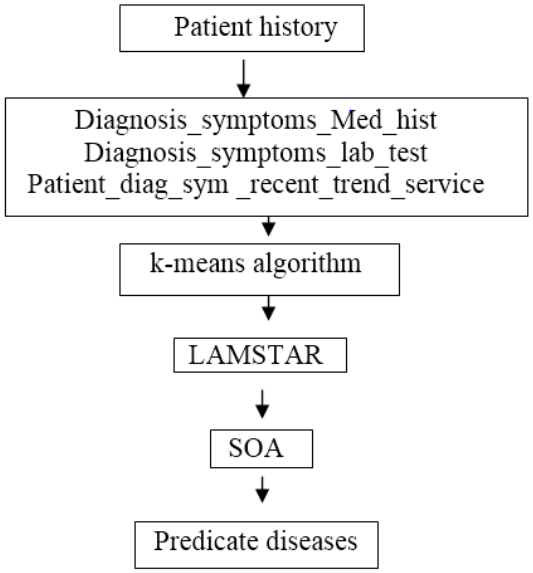
Wi,m (t+1) = Modified weights in module I - neuron m;

Ái = Learning coefficient for module I;€min = Minimum; error of all weights vectors Wi in module

The query input conditions of the attributes were evaluated and verified whether all of the conditions in a rule was satisfied. If all the conditions in a rule had been satisfied

for the given data then corresponding rule would be fired. The rules that were fired had been processed to derive the output. The predicted output which indicates the presence or absence of the disease is displayed as output to the user of the system. The results will be helping to finalize the patient diagnosis. This feature will be using for only to produce as gain the preparation results. The doctor will be taking the decision to finalize the patient diagnosis from the results and also with his knowledge. After the finalized diagnosis results will be storing on patient history database for future use.

Finally conclude that SOA was well suited to implement this concept it to improve the delivery of important information and sharing of data across the community of healthcare system. The system enforced as various services in the existing SOA, result in easy implementation, integration and scalability with existing EHRs.SOA also handles the related issues to data security and patient confidentiality.



iii) Implementation features of K-means

Algorithm :

The main motivation in this paper is based on the assumption that the instance with similar attribute values is more likely to have similar class label. Similarity is measured based on Euclidean distance. Therefore, the misclassified instances after clustering are deleted and correctly classified instances are considered for further classification using decision-tree classifier. This approach considered the result of k-means because the misclassification rate was less. The k-means algorithm takes the input parameter, k, and partitions a set of N points into k clusters, so that the resulting intracluster similarity is high but the intercluster similarity is low. Clustering is the process of grouping same elements. This technique may be used as a preprocessing step before feeding the data to the classifying model. The attribute values need to be

normalized before clustering to avoid high value attributes dominating the low value attributes.

Prediction Algorithm:

1. Configure the dataset 1-R-S

Where 1=no of input , R – no of hidden inputs, S - no of output values.

2. The no of weights are calculated based on storing table W.

3. Assume no of digits in weight from Dataset D.

4: Choose symptoms wi form population Dataset Si

Step

5: for each weighted symptoms

{

Extract weight G-I;

Keep the weight for each input for train Dataset.

Calculate the fitness Value Fi for each of symptoms from Population dataset;

}

6. Apply LAMSTAR

7. output Fi for each Wi

PATTERN MATCHING SERVICES

In the pattern matching services, symptom matching using iterative search utilizes data that is stored. The first step of the Algorithm involves selecting the symptoms shown by the patient. The algorithm gives the list of all possible diseases ranked according to the number of symptoms matched in the database. The list is generated after input of every symptom. After the first iteration for the second iteration the next list of symptoms will be shortlisted according to the disease list that was obtained in the previous iteration.

The new symptom list will contain symptoms of only those diseases that were obtained in the previous list, if *headache*, *fever* and *pain* in the *sinuses* are entered, then the weights G15, G16 and G19 will be considered. Next all the weights will be added and compared to all subclasses V1, V2, V3 and V4 is most likely the answer depending on its weight. Finally all the diseases in class V4 are considered and if sinusitis (D4) weight is closer to the sum of all the input symptoms weights, then it is possible diagnosis.

RESULTS AND DISCUSSION

The results will be showing clearly that the proposed method performs well compared to other similar methods in the literature, considering the fact that the attributes taken for analysis are not direct indicators of disease symptoms. The model to predict the class was built with training samples and validated with an independent test sample. Experts were consulted to validate the rules generated by the system and the validated rules were stored in the knowledge base. The above concepts, features and implementation will have developed in practical manner through query, it will be producing as effective query processing results

CONCLUSION AND FUTURE WORK:

This research uses LAMSTAR Network, K means algorithm to assist the doctor to perform differential diagnosis along with the possible implementation using SOA techniques. By using these techniques, it improves the overall speed and better results. The results obtained for the prediction of the multiple disease show that the system can classify the positive samples with better accuracy as compared to classification of negative classification. This research will have conducted experiments on oracle query in future. This Query application will have been producing as effective results.These results will compare the performance (both explicitly and implicitly) to existing system and it will have produces the results in effective manner

REFERENCES

[1] Z. Dou. R.song and J-R" Wen " A large Scale Evalution and analysis pf personalization Search Strategies" Proc Int'Conf.World Wide Web(WWW) .pp.581-590,2007

[2] John S. Bresse David Heckerman and Carl Kadie " Emperical Analysis of Predictive Algorithms for Colloborative Filtering" 2010.

[3] K.Fry, E-learning markets and providers: some issues and prospects", Educationand training Emerland 43(4) 2001,233-239.

[4] P.De Bra P.Brusilovsky & G.J Houben,"Adaptive hypermedia from systems to framework" ACM Computing Survery 31(4es) 1999.

[5] R.Agarwal and R. Srikant Fast algorithm for mining association rules in large databases. In VLDB pages 487-499,1994.

[6] D.Bever, S,Micali and P.Rogaway, The round complexity of secure protocols. In STOC pages 503-513,1990.

[7] A.Ben-David, N.Nisan and B.Pikas, FairplayMP- a system for secure multi-party computation. In CCS pages 257-266,2008.

[8] Shomona Gracia Jocab and R.Geetha Ramani,"Discovery of Knowelwdge Patterns in Clinical data through Data mining algorithms: Multi-class

catgorization of Breast Tissue Data",International Journal of Computer Applications(0975-8887) Volume32-o.7,October 2011.

[9] Shweta Kharya,"Using Data Mining Techniques for Diagnosis and Prognosis of Cancer Disease", International Journal of Computer Science,Engineering and Information Technology(IJCSEIT), Vol.2 No2,April 2012.

[10] M.Akhil jabbar. B.L Deekshatula and Priti Chandra," Classification of Heart Diseases using K-Neerest Neighbour and Genetic Algorithm",International Conference on Computational Intelligence: Modelling Techniques and Applications(CIMTA) 2013`