INTERNATIONAL JOURNAL OF SCIENTIFIC PROGRESS AND RESEARCH (IJSPR) ISSN: 2349-4689

An Association Rule Mining Technique to Predict Relevant Field of Question Online QA System

Heena U H[1] , Kavana H S[2], Manjushree H S[3], Sunitha.R[4]
[123]*Student, Dept of CSE,* [4]*Assistant Professor, Dept. of CSE,*
Malnad College of Engineering Hassan

Abstract -Online data prediction is a challenging task in the present mounting era of big data world. There exist many techniques to predict the web users data, among those Question-Answer system is one. In a Question Answering system, the user submits a question and waits for the answer as the response. If the system is capable of predicting the user's future interest as the next question, its performance will improve greatly. This paper predicts users' future questions based on the current interaction records of the user with the system. Their current interactions with the system show what they are interested in. These interaction records are maintained in the form of Questions log from which the user sessions are extracted. Based on the user sessions, the system predicts the next question for which the user may become interested in near future. A sample Questions Log is selected for the purpose of performing experiments. The model of Association rule mining is applied to predict the future question of the user. This proposed model prediction performance is better compared to previous existing Question Answering systems.

Keywords: Association Rule mining, QA system, Data mining, Prediction, Question log analysis

I. INTRODUCTION

Question Answering system: It is a system which provides the answers for the user's questions. There is no restriction for the type of questions and the type of users. The questions belongs to any fields and the answers for every questions are obtained. It is a computer science discipline with in the fields of information retrieval and natural language processing (NLP),which is concerned with building systems that automatically answer questions posed by humans in a natural language.

Data Mining: It is the computational process of discovering patterns in large datasets involving methods at the intersection of artificial intelligence, machine learning, statistics and database system. 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. The knowledge discovery in databases (KDD) process is commonly defined with stages are Selection, Pre- processing, Transformation, Data mining ,Interpretation/evaluations. It exists, however, in many variations on this theme ,such as the Cross Industry Standard Process for Data Mining (CRISP-DM) which defines six phases are business understanding, data understanding ,data preparation, modeling, evaluation, deployment Or a simplified process such as, pre – processing, data mining, results validation. Data mining involves, six common classes of tasks are anomaly detection, association rule learning, clustering, classification ,regression, summarization

Let X be an item set, X=>Y an association rule and T a set of transaction of a given database. Support: It is an indication of how frequently the item set appears in the database. The support of X with respect to T is defined as the proportion of transaction t in the database which contains item set X.

$$supp(X)=|\{t \in T; X \subseteq t\}| / |T|$$

Confidence is an indication of how often the rule has been found to be true. The confidence value of a rule, X=>Y, with respect to a set of transaction T,is the proportion of the contains X which also contains Y.

$$conf(X \Rightarrow Y)=supp(X \cup Y) / supp(X)$$

Association Rule Mining: This is the technique for discovering interesting relationship hidden in large data sets. The uncovered relationship can be represented in the form of association rules or set of frequent items. Association rules are created by analyzing data for frequent if/then patterns and confidence to identify the most important relationships. Support is an indication of how frequently the items appears in the database.

Confidence includes the number of times the if/then statements have been found to be true. It is useful to analyzing and predicting customer behavior. Programs use association rules to build programs capable of machine learning .Machine learning is a type of artificial intelligence(AI) that seeks to build programs with the ability to become more efficient without being explicitly programmed.

Prediction: It is the process of predicting the value of particular attribute based on the values of other attribute. The attribute to be predicted is known as the target/dependent variable and the attribute used for making the prediction are known as independent variables.

Log Analysis: Is a Record of interactions between the search engine and the user. The major objective of the technique of Log Analysis is the Information Extraction. In this newer QA system of predicting the relevant field of questions in online QA system. The technique of log analysis for information extraction and also for recording the interactions between client and server and applying association rule mining technique of data mining approach to predicting the questions of user's area of interest that the user may ask in future.

II. SYSTEM MODEL

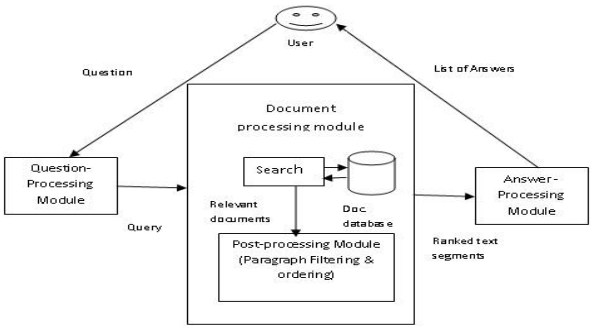

Fig.2.1 Architecture of QA system

In this architecture proposes the users future question based on their interactions maintained in the form of questions in questions log. It predicts the users next query using Association rules mining. It predicts the users future search interest on the basis of his/her interactions with the QA system.

III. PREVIOUS WORK

In Existing QA system, user submits a questions and waits for answer and the system retrieves the answer for the question and also displays related blogs and this suggestion is based on the question types inputs by the user.

There are many existing QA systems with different Approaches and Techniques,

An Educational Model based on Knowledge Discovery in Database(KDD)[1] to predict learners Behavior using Classification Techniques. In this system the Learners Behavior is predicted using Classification Technique. The system examines the students history of accessing the UniversityLearningManagementSystem(ULMS).

Classification technique classifies the data based on Knowledge Discovery in Data Base to predict learners behavior. The Students frequently interact with the LMS. Distance Education Learner make use of LMS to get mark of pass in certain course. The final rating of students based on the students history of accessing data in the university LMS. The successful completion of the program depends on how students interact with the activities posted on LMS and helps identify DE learners who needs early intervention for better academic achievements and meaningful online learning environment.

Recommending Documents for Complex Question Exploration[2] by Analyzing Collective Browsing Behavior. In this system Recommending the relevant and useful documents for the Complex questions by analyzing the user's Browsing Behavior. This also helps the user to find the right source of information whether the data is very large and stored in repositories. Novel Approach is used to recommend the document to user by analyzing users Browsers Behavior. Here the novel data set of document, browsers behavior are collected as users research complex questions. Based on this dataset, machine learning algorithm predicts the document useful and it should be recommended.

Collaborative filtering in social network a community based approach[3] are We use technique called as community based approach such as Community is considered as a group of users in which users interact with others frequently than with other groups .with use of social networks there arises a usual question that it provided a better results based on the opinions from different communities users with same taste are frequently connected with users of same group. A user may also find products based on the choices of community members.

Machine learning for question answering[4] from tabular data are In this question answer system will automatically answer natural language question in a human like manner .by moving expensive text, information system can achieve good run-time behavior with large amount of information .It uses several table to answer a users question . The main issue is that the result for the given question in which tables and fields should be searched from which -user answers should be extracts. An increasing number of patients seek medical information online. We use NLP(Natural Language Processing) for processing normal language and UMLS(Unified Medical Language System)[5] for medical terms. Now a day's diabetes continues to increase worldwide ,it becomes a serious public health problem ,collect the questions from patients and provide the health advice by professionals in perceived time constraints based on three expert-vetted online resources.

QA remains a challenging task in NLP.A general QA system contains at least two modules. Namely, a query processing module to analyze the user`s questions and retrieve the data to select appropriate answers. In this

system four components are Question analysis - Understanding the content of a question and also classifies the type of the question. Knowledge retrieval - It searches related documents using the information obtained in question analysis. Answer extraction – Selects candidates from the retrieved documents. Answer generation – Process ranks candidates and provides the most relevant answer. In this paper user can be ask restaurant relate questions like services, food ,party ,location, equipment, discount ,price ,contact, information, opening hours[6].

III.PROPOSED METHODOLOGY

Proposed system is a new system where user submits the question and waits for the answer as the response and also predicts the users future question based on the current interaction records of the user with the system and current interactions show the users area of interest. Prediction is done by the Data Mining technique called “Association Rule”

Scope and objective of the proposed system such as Proposed system is a online based application, Generates the answer for the users question ,Objective is to predict the user next question and provide services according to the users area of interest, This system predicts users future question based on the interaction records of users, Use of Data Mining technique called Association Rules to discover patterns between user questions.

Working of QA system are the user enters his question on the interface of the QA system, This question is classified by the Question classifier module, Search is done by the searcher module for the answer(s) to the questions, The interactions between the user and the system are maintained in a file termed as Question Log.

In proposed system we can use mainly 3 actors are Administrator - Administrator is a one who maintains the entire QA system. Administrator has the full accessibility of the application (QA system). Members - Member is a registered user. Member is a one who receives the services from the application. The key service given by this QA system is “prediction of next question” based on the previous questions posted by the users. Guest - Guest is a new comer to the application, Guest has only the limited accessibility.

Modules of the project are , Administrator have modules like login module, manage types [Question Types] , set special words, manage members, browse QA System QA Verification Module. Member have modules such as login module, QA module, answer module. Predict Next Question using these actions such as user session extraction, question extraction, question filter module, question pre-processor, association rules generation, next question predictor. In this module, system predicts users future questions based on the current interaction records of the user with the system. QA History [Question Log] - In this module, student can view the answers for the questions posted by themselves. Guest have these modules are view basic information, registration, view recent/trending QA module.

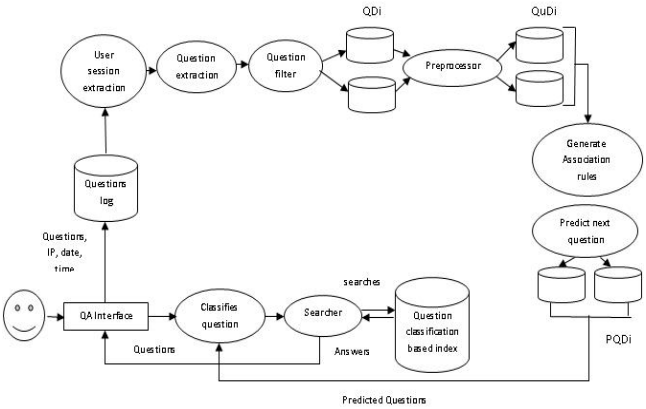

Fig.3.1 Architecture of predicts next users next questions in association rule.

IV. SIMULATION/EXPERIMENTAL RESULTS

In Proposed Model We Can Use Question Log For Selecting The Data And It Stores Frequently Itemsets By Users It Can Be Generated By Association Rule Mining. The Rules Generated Are Given As Input To The “Next Questions Predictor” That Predicts The Users Next Question. Based On Users Data Accessed Itemset It Gives 98% Efficiency.

The following figures shows the resuls of the QA system using association rule,

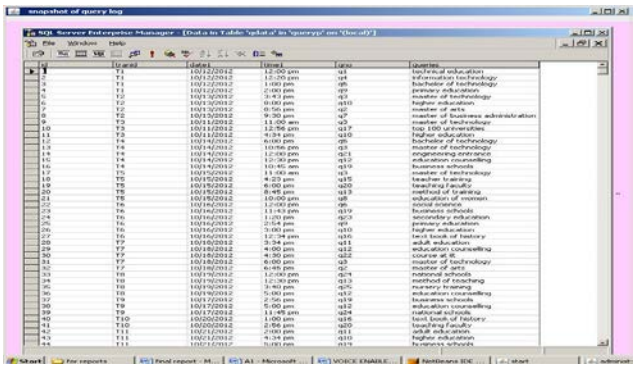

Fig.4.1It Shows question log used by the users.

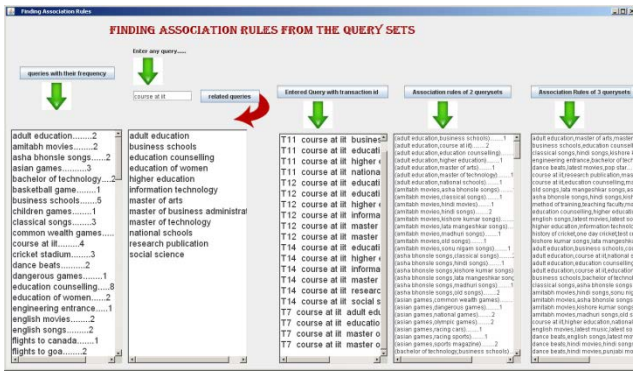

Fig.4.2This shows the frequent itemsets and association rule.

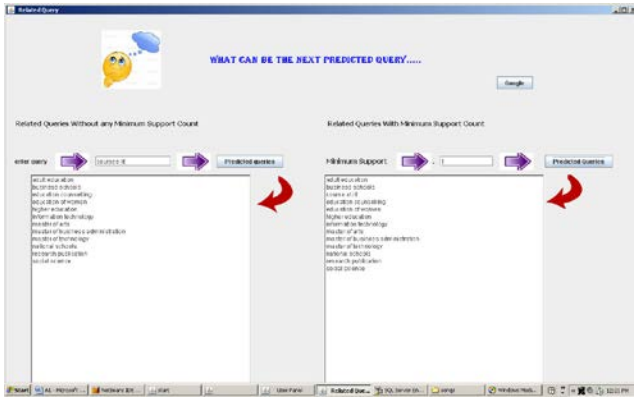

Fig.4.3(a)

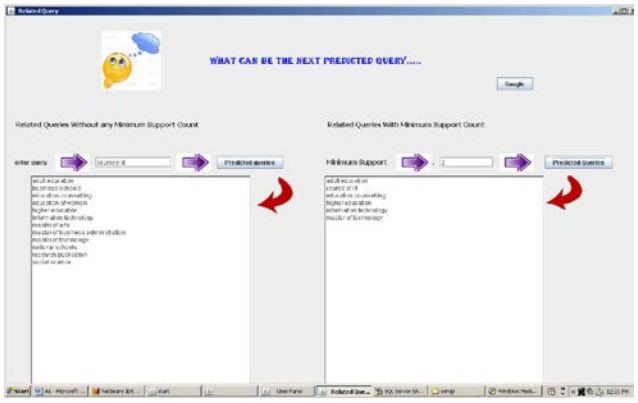

Fig.4.3(b)

Both Fig.4.3(a) and Fig.4.3(b) are shows the prediction of next question

V. CONCLUSION

User submits a question and waits for the answer as the response. If the system is capable of predicting the user's future interest as the next question, its performance will improve greatly. The result predicts users' future questions based on the current interaction records of the user with the system.

In Proposed work, the system predicts the user's next question based on their interaction between the system. Using Association rule mining can be used predicts the user's next question, this technique is suitable for predict the next question using support and confidence. This provides most accurate result to the users.

VI. FUTURE SCOPES

SMS Verification Module Can Be Include As A Future Enhancements For User Registrations. So That Fake Accounts Can Be Restricted. We Can Insert Online Discussion Forum Module As A Future Enhancements Where Users Of The QA System Can Discuss On A Topic And Share Their Ideas.

REFERENCES

[1] An Educational Model Based on Knowledge Discovery in Databases (KDD) to Predict Learner's Behavior Using Classification Techniques Benilda Eleonor V. Comendador, Lorena W. Rabago and Bartolome T. Tanguilig III Graduate Programs Office Technological Institute of the Philippines Quezon City, Philippines email: bennycomendador@yahoo.com, lwr823@yahoo.com, bttanguilig_3@yahoo.com

[2] Recommending Documents for Complex Question Exploration by Analyzing Collective Browsing Behavior Alya Asarina-MIT Lincoln Laboratory Lexington, MA, USA alya@mit.edu Olga Simek-MIT Lincoln Laboratory Lexington, MA, USA osimek@ll.mit.edu

[3]Collaborative Filtering in Social Networks: A Community-based Approach, The Anh Dang, Emmanuel Viennet L2TI – Institute Galil´ee - Universit´e Paris-Nord 99, Avenue Jean-Baptiste Cl´ement - 93430 Villetaneuse - France {theanh.dang,emmanuel.viennet}@univ-paris13.fr

[4]Machine Learning for Question Answering from Tabular Data, Mahboob Alam Khalid Valentin Jijkoun Maarten de Rijke ISLA, University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands {mahboob,jijkoun,mdr}@science.uva.nl

[5]Answering Diabetic Patients' Questions Using Expert-vetted Online Resources: A Case Study,Yuqun Zeng1,2, Xusheng Liu1*1The Second Clinical CollegeGuangzhou,University of Chinese Medicine,Guangzhou, Guangdong, P.R. China *e-mail: liuxu801@126.comLiwei Wang2, Hongfang Liu2,Yanshan Wang2*2Division of Biomedical Statistics and InformaticsMayo Clinic,Rochester, MN, USA*e-mail: Wang.Yanshan@mayo.edu

[6] Principle-based approach for semi-automatic construction of a restaurant Question Answering System from limited dataset. Ting-Hao yang, Yu-Lun H Sieh, You- Shan chung , Shih-hung Liu ,Yung-Chun chang, Wen-Lian hsn: Institute of Information Systems and Applications ,National Tsing hua University : Institute of information science,AcadamiaSinica{tinghaoyang,morphe,yschung,dapi, journey,changyc,hsu}@iis.sinica,edu.tw

[7] L. Limam, D. Coquil, H. Kosch, and L. Brunie, "Extracting user interests from search query logs: A clustering approach," DEXA '10 Proceedings of the 2010 Workshops on Database and Expert Systems Applications, 2010.

[8] B. M. Fonseca, P. B. Golgher, E. S. de Moura,, and N. Ziviani, "Using Association Rules to Discover Search Engines Related Queries," in Proceedings of the First Conference on Latin American Web Congress, pp. 66–71, 2003.

[9] Z. Zhang, and O. Nasraoui, "Mining search engine query logs for query recommendation," in Proceedings of the 15th international conference on World Wide Web, pp. 1039– 1040, 2006.

[10] Z. Cheng, B. Gao, and T. Liu, "Actively predicting diverse search intent from user browsing behaviors," in Proceedings of the 19th international conference on World wide web, pp. 221–230, 2010.

[11] D. Gupta, A. Puniya, and K.K. Bhatia, "Prediction of the Query of the Search Engine using Back propagation Algorithm," IJCSE, 2011.