International Journal for Modern Trends in Science and Technology
Volume: 03, Special Issue No: 01, February 2017
ISSN: 2455-3778
http://www.ijmtst.com

Finding of Interesting Rules from Association Mining by Ontology and Page Ranking

D.Saravanam[1] | Dr.S.Vijayalakshmi[2] | Dennis Joseph[3]

[1,3]Faculty of Operations & IT, IFHE University, IBS Hyderabad, Telangana, India.
[2]Faculty of Finance, IFHE University, IBS Hyderabad, Telangana, India.

To Cite this Article
D.Saravanam, Dr.S.Vijayalakshmi and Dennis Joseph, "Finding of Interesting Rules from Association Mining by Ontology and Page Ranking", *International Journal for Modern Trends in Science and Technology*, Vol. 03, Special Issue 01, 2017, pp. 46-50.

ABSTRACT

In Data Mining, the usefulness of association rules is strongly limited by the huge amount of delivered rules. Association rule mining is considered as one of the most important tasks in Knowledge Discovery in Databases among sets of items in transaction. To overcome this drawback, several methods were proposed in the literature such as item set concise representations, redundancy reduction, and post processing. However, being generally based on statistical information, most of these methods do not guarantee that the extracted rules are interesting for the user. Thus, it is crucial to help the decision-maker with an efficient post processing step in order to reduce the number of rules. This paper proposes a new interactive approach to prune and filter discovered rules. First, we propose to use ontology's in order to improve the integration of user knowledge in the post processing task. Second, we propose the Rule Schema formalism extending the specification language for user expectations. Furthermore, an interactive framework is designed to assist the user throughout the analyzing task. Applying our new approach over voluminous sets of rules, we were able, by integrating domain expert knowledge in the post processing step, to reduce the number of rules to several dozens or less. Moreover, by using page ranking algorithm, the required data can be mined.

I. Introduction

Today the Web has become the largest information source for people. Most Information retrieval systems on the Web consider web pages as the smallest and undividable units, but a web page as a whole may not be appropriate to represent a single semantic. A web page usually contains various contents such as navigation, decoration, interaction and contact information, which are not related to the topic of the web-page. Furthermore, a web page often contains multiple topics that are not necessarily relevant to each other. Therefore, detecting the semantic content structure of a web page could potentially improve the performance of web information retrieval. Many web applications can utilize the semantic content structures of web pages.

For example, in web information accessing, to overcome the limitations of browsing and keyword searching, some researchers have been trying to use database techniques and build wrappers to structure the web data in building wrappers, it is necessary to divide the web documents into different information chunks. Previous work uses ad hoc methods to deal with different types of web pages

1.1 Existing System

- Appropriate is the first algorithm proposed in the association rule mining field and many other algorithms were derived from it. Starting from a database, it proposes to extract all association rules satisfying minimum

thresholds of support and confidence. This causes to mine excess of rules.

- They involve another method called Support threshold, it makes intractable for a decision maker to analyze the mining result.
- Post processing methods can improve the selection of discovered rules.
- By introducing the pruning method, it avoids unexpected rules.
- Expected and unexpected results are separated and grouped.
- Finally these grouping result are visualized in graphical representation.

1.1.1 Disadvantages

- Pruning avoids unexpected results, which it causes a major drawback as because the user interestingness might be in unexpected results.
- Support threshold gives the minimum rules but it increases the association rules.
- More time consumption to reduce the rule.

1.2 . Proposed System

In order to overcome the above the above problem, we used the three methods to reduce the association rule.

1.2.1 Role Schema formulation

- It is a language based.
- Here some set of operators are implemented in role schema to reduce the rules, they are Pruning and filtering method.
- Filtering: This method involves three concepts as Confirming, Unexpectedness, and Exception.
- Confirming Method filters the rules matching the unexpected rule.
- Unexpectedness Method reduces the filtering process which is done by role schema.
- Exception Method extracts the confirming rules.

1.2.2 Filtering

Filtering in the sense it reduces the expected and unexpected rules.

- Selects only certain rules.

Filters all the couples of rule item.

1.2.3 Ontology

It concentrates the user needs in a database.

- Most appropriate representation to express the complexity of the user knowledge, and several specification languages.
- Ontology comprises of two concepts, they are 1) Attribute and 2) Direct.
  - *Attribute:* It groups the sub related concepts from unexpected rules
  - *Direct:* Grouping the facility and life quality in a discovered rule.

1.2.4 Advantages

- Computational time will be reduced.
- User can view both the expected and unexpected rules.
- User interesting rules can be easily obtained
- Dozens of association rules can be reduced

II. Web Mining

It is the application of data mining techniques to discover patterns from the Web. According to analysis targets, web mining can be divided into three different types, which are Web usage mining, Web content mining and Web structure mining.

2.1 Web usage mining

Web usage mining is the process of finding out what users are looking for on the Internet. Some users might be looking at only textual data, whereas some others might be interested in multimedia data.

2.2 Web structure mining

Web structure mining is the process of using graph theory to analyze the node and connection structure of a web site. According to the type of web structural data, web structure mining can be divided into two kinds:

1. *Extracting patterns from hyperlinks in the web:* a hyperlink is a structural component that connects the web page to a different location.

2. Mining the document structure: analysis of the tree-like structure of page structures to describe HTML or XML tag usage.

Web mining Pros and Cons == or ==Web Content Mining

2.3 Pros

Web mining essentially has many advantages which makes this technology attractive to corporations including the government agencies. This technology has enabled ecommerce to do personalized marketing, which eventually results in higher trade volumes. The government agencies are using this technology to classify threats and fight against terrorism. The predicting capability of the mining application can benefits the society by identifying criminal activities. The companies can establish better customer relationship by giving them exactly what they need. Companies can understand the needs of the customer better and they can react to customer needs faster.

2.4 Cons

Web mining, itself, doesn't create issues, but this technology when used on data of personal nature

might cause concerns. The most criticized ethical issue involving web mining is the invasion of privacy. Privacy is considered lost when information concerning an individual is obtained, used, or disseminated, especially if this occurs without their knowledge or consent The obtained data will be analyzed, and clustered to form profiles; the data will be made anonymous before clustering so that there are no personal profiles. Thus these applications de-individualize the users by judging them by their mouse clicks. De-individualization, can be defined as a tendency of judging and treating people on the basis of group characteristics instead of on their own individual characteristics and merits. These practices might be against the anti-discrimination legislation. The applications make it hard to identify the use of such controversial attributes, and there is no strong rule against the usage of such algorithms with such attributes.

III. Proposed Method

3.1 User query
3.2 Creation Domain ontology
3.3 Improving effectiveness by using interactive frame work using the filter.
3.4 Applying page ranking algorithm.

3.1 User query

Every user must need an authorization. Authorization is for only prescribed users entering the querying scheme rather than unauthorized access. Users and Server have an authorization entry. Sometimes by mistake, the user gives wrong user name and password, the server generates the warning to every mismatch inputs. If it is matched, then the user get the connection otherwise the server quits the unauthorized person.

For example:

a) to access the PDF documents the privileges are given to the lecturer and to access the WORD doc the privileges are given to the student.

b) In that case we can set the two privileges. First to the staffs that they are working in the CTS can see one type of data and the public can able to see another type of data's

3.2 Creation of domain ontology

Here when the user clicks the particular document that particular links and about that link Meta data also maintained in the databases by using the table. This database is called domain ontology database in these we are storing the user's past information and by using the association rule we are taking the consideration in domain's database also.

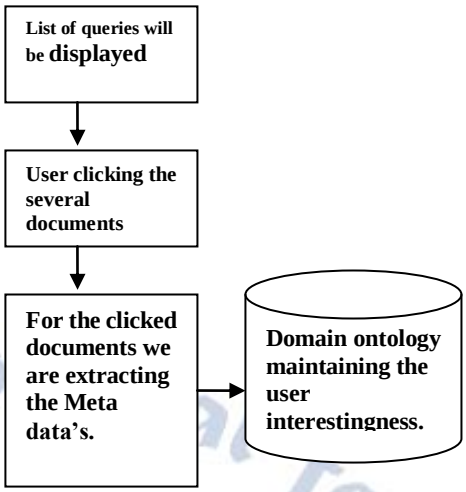

Fig 1. Creation of Domain Ontology

3.3 Improving effectiveness by using interactive frame work using the filter

In order to reduce the large number of rules delivering by rule scheme here we do the framework by using the filters mechanism. When the query is passed means first it will take two databases domain ontology database and the main database.

example

when the user giving the query java, the query will pass to the data bases now using the association rule the large no of rules are eliminated. Because in the domain ontology the user having the interestingness in oops. The results will be displayed as "java oops" because these two have relationship concepts.

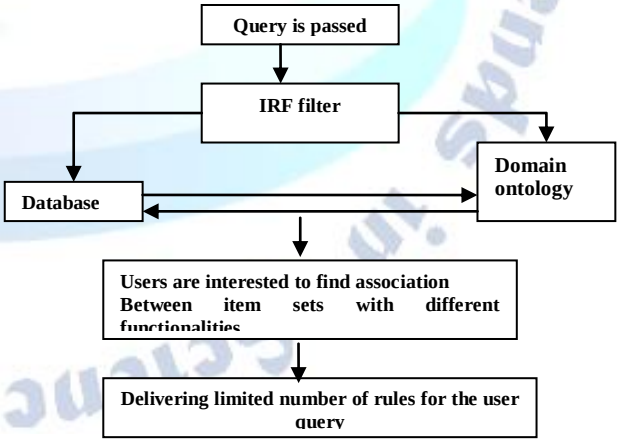

Fig 2.Iinteractive frame work using the filter.

3.4 Applying page ranking algorithm

Here we ranking algorithm defines the rank results, most of the solutions need to work on the whole annotated knowledge base. Relevance is measured as the probability that a retrieved resource actually contains those relations whose

existence was assumed by the user at the time of query definition

IV. Conclusion

This paper discusses the problem of selecting interesting association rules throughout huge volumes of discovered rules. The major contributions of our paper are stated below. First, we propose to integrate user knowledge in association rule mining using two different types of formalism: ontology's and rule schemas. On the one hand, domain ontology's improve the integration of user domain knowledge concerning the database field in the post processing step. On the other hand, we propose a new formalism, called Rule Schemas, extending the specification language proposed by Liu et al. The latter is especially used to express the user expectations and goals concerning the discovered rules. Second, a set of operators, applicable over the rule schemas, is proposed in order to guide the user throughout the post processing step. Thus, several types of actions, as pruning and filtering, are available to the user. Finally, the interactivity of our ARIPSO framework, relying on the set of rule mining operators, assists the user throughout the analyzing task and permits him/her an easier selection of interesting rules by reiterating the process of filtering rules.

By applying our new approach over a voluminous questionnaire database, we allowed the integration of domain expert knowledge in the post processing step in order to reduce the number of rules to several dozens or less. Moreover, the quality of the filtered rules was validated by the expert throughout the interactive process. And as an enhancement the queries are managed using the privileges. And the results will be displayed at the ranks basis according to the count based.

V. Future Enhancement

Organizations are taking advantage of "data-mining" techniques to leverage the vast amounts of data captured as they process routine transactions. Data-mining is the process of discovering hidden structure or patterns in data. However several of the pattern discovery methods in data mining systems have the drawbacks that they discover too many obvious or irrelevant patterns and that they do not leverage to a full extent valuable prior domain knowledge that managers have. This research addresses these drawbacks by developing ways to generate *interesting* patterns by incorporating managers' prior knowledge in the process of searching for patterns in data. Specifically we focus on providing methods that generate *unexpected* patterns with respect to managerial intuition by eliciting managers' beliefs about the domain and using these beliefs to seed the search for unexpected patterns in data. Our approach should lead to the development of decision support systems that provide managers with more relevant patterns from data and aid in effective decision making.

Objective measures such as support, confidence, interest factor, correlation, and entropy are often used to evaluate the interestingness of association patterns. However, in many situations, these measures may provide conflicting information about the interestingness of a pattern. Data mining practitioners also tend to apply an objective measure without realizing that there may be better alternatives available for their application. In this paper, we describe several key properties one should examine in order to select the right measure for a given application. A comparative study of these properties is made using twenty-one measures that were originally developed in diverse .fields such as statistics, social science, machine learning, and data mining. We show that depending on its properties, each measure is useful for some application, but not for others. We also demonstrate two scenarios in which many existing measures become consistent with each other, namely, when support-based pruning and a technique known as table standardization are applied. Finally, we present an algorithm for selecting a small set of patterns such that domain experts can find a measure that best fits their requirements by ranking this small set of patterns.

The first one is that the effectiveness s largely depends on the numbers of labeled training data. However, we may only obtain some positive documents.

The second one is that it is hard to distinguish nonintersecting topics from interesting topics. Here, we develop an ontology mining technique to overcome the above drawbacks

It is not easy to obtain the right information from the Web for a particular Web user or a group of users due to the obstacle of automatically acquiring Web user profiles

A user accessing an information system with the intention of satisfying an information need, may have to reformulate the query issued several times and sift through many results until a satisfactory, if any, answer is obtained. This is a very common

experience especially for Web searchers, due to information abundance and users.

VI. Experimental Results

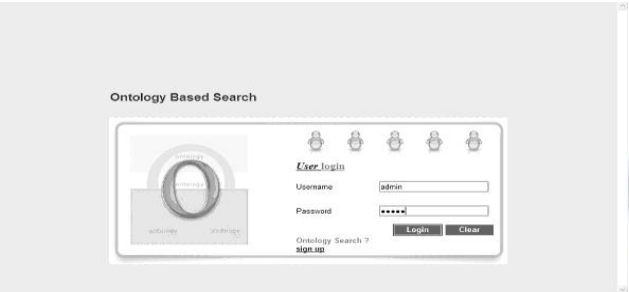

Fig 3. Ontology Based Search

Fig 4. Query by Search

Fig 5. Query Result

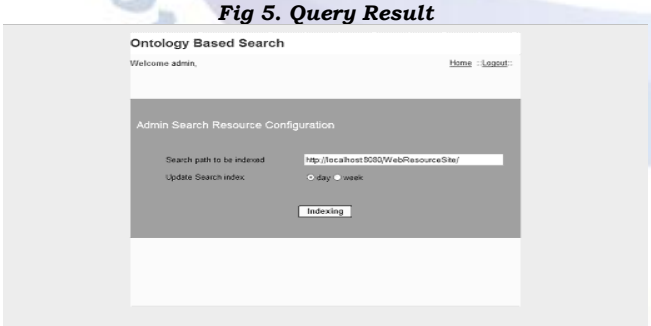

Fig 6. Admin Resource Configuration

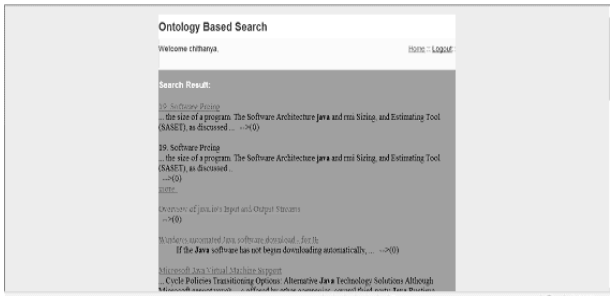

Fig 7. Search Result

REFERENCES

[1] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," Proc. ACM SIGMOD, pp. 207-216, 1993.

[2] U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, and R. Uthurusamy, Advances in Knowledge Discovery and Data Mining. AAAI/MIT Press, 1996.

[3] A. Silberschatz and A. Tuzhilin, "What Makes Patterns Interesting in Knowledge Discovery Systems," IEEE Trans. Knowledge and Data Eng. vol. 8, no. 6, pp. 970-974, Dec. 1996.

[4] M.J. Zaki and M. Ogihara, "Theoretical Foundations of Association Rules," Proc. Workshop Research Issues in Data Mining and Knowledge Discovery (DMKD '98), pp. 1-8, June 1998.

[5] D. Burdick, M. Calimlim, J. Flannick, J. Gehrke, and T. Yiu, "Mafia: A Maximal Frequent Itemset Algorithm," IEEE Trans. Knowledge and Data Eng., vol. 17, no. 11, pp. 1490-1504, Nov. 2005.

[6] J. Li, "On Optimal Rule Discovery," IEEE Trans. Knowledge and Data Eng., vol. 18, no. 4, pp. 460-471, Apr. 2006.

[7] M.J. Zaki, "Generating Non-Redundant Association Rules," Proc. Int'l Conf. Knowledge Discovery and Data Mining, pp. 34-43, 2000.

[8] N. Pasquier, Y. Bastide, R. Taouil, and L. Lakhal, "Efficient Mining of Association Rules Using Closed Itemset Lattices," Information Systems, vol. 24, pp. 25-46, 1999.

[9] H. Toivonen, M. Klemettinen, P. Ronkainen, K. Hatonen, and H. Mannila, "Pruning and Grouping of Discovered Association Rules," Proc. ECML-95 Workshop Statistics, Machine Learning, and Knowledge Discovery in Databases, pp. 47-52, 1995.