Global Journal of Pure and Applied Mathematics.
ISSN 0973-1768 Volume 12, Number 4 (2016), pp. 3577-3584

http://www.ripublication.com/gjpam.htm

Reduce convention for Large Data Base Using Mathematical Progression

D.Saravanan and A.Ronald Tony

Faculty of operations & IT, IFHE University,
IBS, Hyderabad 501203, Telangana, India.
Assistant professor, Faculty of Computing,
Sathyabama University, Chennai-600199, Tamil Nadu, India

Abstract

The convenience of association standards is firmly restricted by the gigantic measure of conveyed principles. To defeat this disadvantage, a few strategies were proposed in the literature, for example, thing set concise representations, repetition lessening, and post preparing. In any case, being for the most part in light of factual data, the greater part of these routines doesn't promise that the extricated standards are fascinating for the users. In this manner, it is pivotal to help the chief with a productive post handling venture so as to lessen the quantity of principles. This paper proposes another intelligent way to deal with prune and channel found guidelines utilizing Data Mining. To begin with, we propose to utilize ontology's keeping in mind the end goal to enhance the incorporation of user learning in the post handling errand. Second, we propose the Rule Schema formalism amplifying the detail language proposed by Liu et al. for user desires. Besides, an intuitive structure is intended to help the users all through the examining undertaking. Applying our new approach over voluminous arrangements of standards, we were capable, by coordinating domain expert learning in the post preparing venture, to diminish the quantity of guidelines to a few handfuls or less. In addition, the nature of the sifted principles can also be validated by the domain expert at different focuses in the intuitive procedure.

Key terms: Data mining, Association Rules, Pre-processing, post-processing, User Knowledge, Rule based system, Datasets.

1. Introduction

Among the items in the set of transaction databases, it can be valuable information to

the decision maker that the purpose of discovering implicative tendencies. Support and confidence of the association rule interestingness where the two sets of X and Y and X ∩ Y. Apriori association rule mining used in industry and in many ways from that of the finding strong Association Rule between the data sets.. Starting from a database, it is satisfying the minimum level of support and confidence proposes to acquire all of the association rules. It is well known that the processing algorithms can detect the size of an outlawed association rules; for example, the rules and the exchange of several dozen to several hundreds of thousands of attributes in a received message. Therefore, we support the increase in the threshold, there are more efficient methods and still found the rules are clear, and therefore, the less they are interesting to the user. As a result, it is sufficient in order to obtain valuable information to come up with low support threshold. Introduced rule mining, knowledge is considered as one of the most important tasks.

Unfortunately, the rules of the largest, is the low level of support for the study as a result of a decision maker to intractable it is. Overpasses tests show that used to be almost impossible, and it is an effective technique for reducing the number of rules is important to help the decision-maker. To cope with this drawback, several methods proposed in the literature. On the one hand, different methods no redundant rules, or using pruning techniques, closed to reduce the number of rules, the maximum or optimal item sets, and by creating a number of mechanisms were introduced to reduce the number of item sets. On the other hand, after the discovery of the processing methods can improve the selection rules. Pruning different special post processing modes, summarizing, grouping, or as visualization, is used for this. Pruning is the removal of uninteresting or redundant rules. Summarizing, concise set of rules are generated. Consolidation groups are the rules of procedure; and improves the readability of the display is adapted to a large number of rules using graphical representations. However, at present the most commonly reported post-processing methods are based on statistical information. Interestingness strongly depends on the user's knowledge and goals of the regime, since the more interesting of these methods does not guarantee that the extracted rules. For example, if the user knows the unexpected rules, rules already known to be pruned. If the user needs to pay attention to the rules specified schemas, or rules, only need to select the subset.

2. Existing System

Starting from a database, it proposes to extract all association rules satisfying minimum thresholds of support and confidence. It is very well known that mining algorithms can discover a prohibitive amount of association rules; for instance, thousands of rules are extracted from a database of several dozens of attributes and several hundreds of transactions, valuable information are often represented by those rare, low support, and unexpected association rules which are surprising to the user. So, the more we increase the support threshold, the more efficient the algorithms are and the more the discovered rules are obvious, and hence, the less they are interesting for the user. As a result, it is necessary to bring the support threshold low enough in order to extract valuable information. Unfortunately, the lower the support is, the

larger the volume of rules becomes, making it intractable for a decision-maker to analyze the mining result. Experiments show that rules become almost impossible to use when the number of rules overpasses 100. Thus, it is crucial to help the decision-maker with an efficient technique for reducing the number of rules.

2.1 Disadvantage of existing systems

- Usefulness of association rules is strongly limited by the huge amount of delivered rules.
- It is crucial to help the decision-maker with an efficient technique for reducing the number of rules.

2.2 Proposed System

This paper proposes a new approach to interactive post processing technique. ARIPSO (Association of Interactive Post-Processing Rule Schemas using the ontology) discovered prune and filter rules. In the proposed technique first, we propose the use of ontology's internet in order to strengthen the integration of user knowledge processing work. Second, we have to use the user's expectations, beliefs and ideas with ontology specification language extending the rule to introduce structural formalism. Furthermore, analysis of the structure of an interactive and iterative process designed to assist the user. In order to describe our approach to intervention measures that can be user schemas on the definition of the rule of the law relies on a set of operators. This paper is structured as follows: the notations and definitions used throughout the paper introduces. Our purposes here, we proposed the use of ontology's. A review of the research domain and its associated tasks and the framework proposed Presents and describes its components. It reported on a questionnaire dedicated to the results obtained by applying our method and finally we present the conclusions and directions for future research shows

2.2.1 Advantage of proposed system

- Reduce the number of item sets by generating closed, maximal optimal item sets, and several algorithms to reduce the number of rules, using nonredundant rules, and pruning techniques.
- Domain ontology's improve the integration of user domain knowledge concerning the database field in the post processing step.
- The integration of domain expert knowledge in the post processing step in order to reduce the number of rules to several dozens or less.

3. Experimental setup

The Association Rule Interactive post-Processing using Schemas and Ontologies contains five major module, those modules are following

3.1 Discover User Knowledge using possible data set.

This module is used to collect the information from various types of users of domain. The most significant tasks are information detection in Databases. Among sets of

items in operation databases, it aims at discover implicative tendency that can be valuable information for the decision-maker.

3.2 Rules Schema Formalism

This module contains Rule schemas, is proposed in order to guide the user throughout the post processing step. Thus, several types of actions, as pruning and the version of user opportunity is more general, and thus, filtered rules are more exciting for the user filtering, are available to the user. Pruning consists in removing uninteresting or redundant rules. These two filters can be applied over rules whenever the user needs them with the main goal of reducing the number of rules.

3.3 Finding Frequent Item set Generation

The process of capturing dependency and implication between databases items, and express the potency of the model connection. Frequent closed item sets in order to reduce the number of frequent item sets. This is a condensed demonstration of all the transactions in the database.

3.4 Redundancy Reduction

In this module we are going to reduce the redundancy among the data. For that the users have to describe as the difference between the confidences of two rules in a specification/ generalization relationship. The specific rule is pruned if the proposed measure is less than a pre-specified threshold, so the rule does not bring more information compared to the general one.

3.5 Report Generation

This module is used to get the final report of the association rule mining process.it will generate the report to analyze the knowledge sharing of the user. It will help to express the user expectations and goals concerning the discovered rules.

3.1 Experimental Outcome:

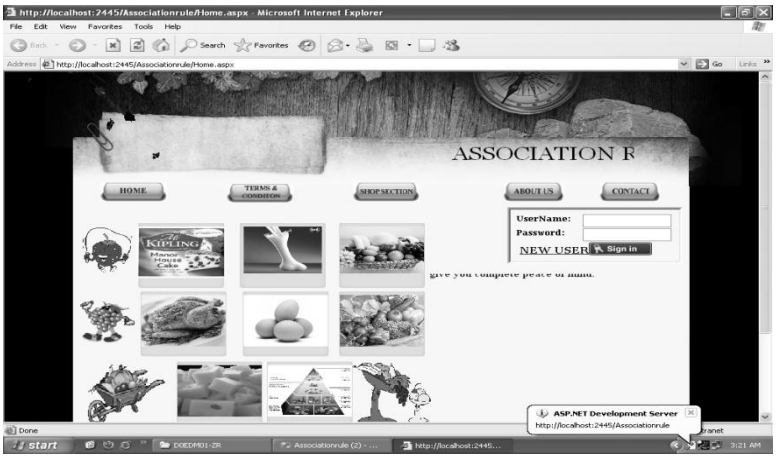

Figure 1: Users input module

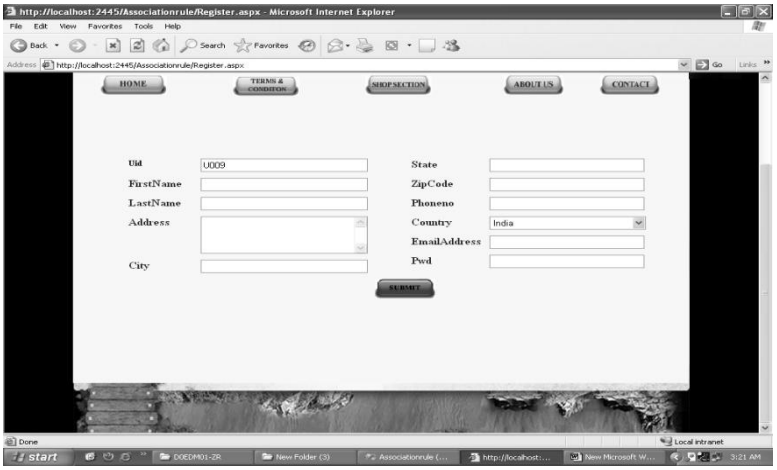

Figure 2: Discovery of knowledge given data set.

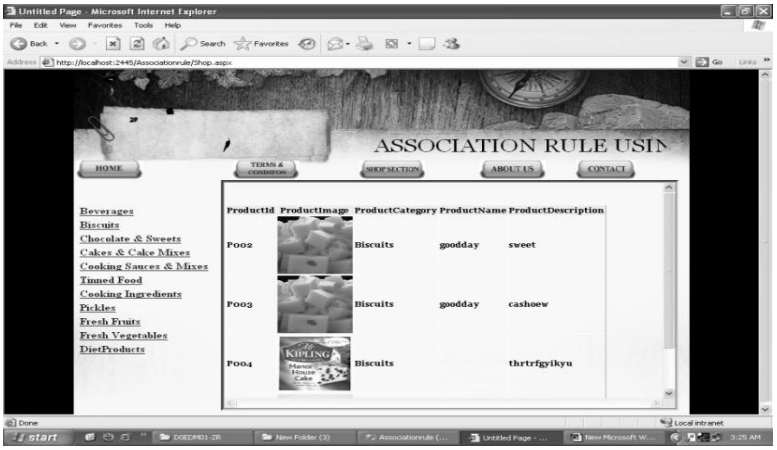

Figure 3: Association rule forming of chosen item set

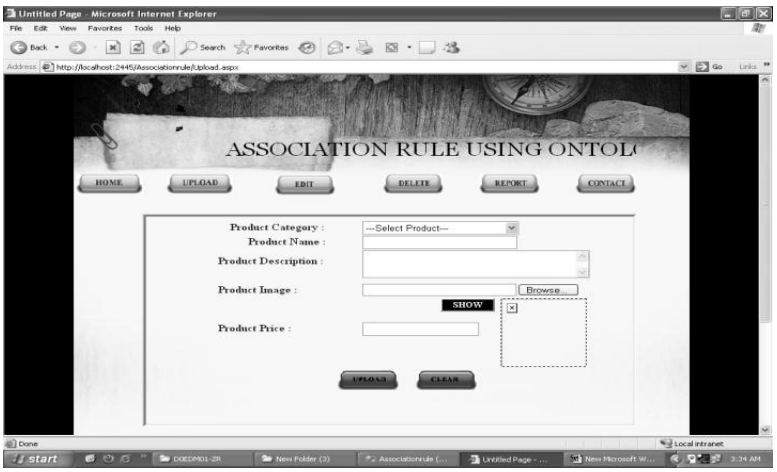

Figure 4: Association rule using ontology

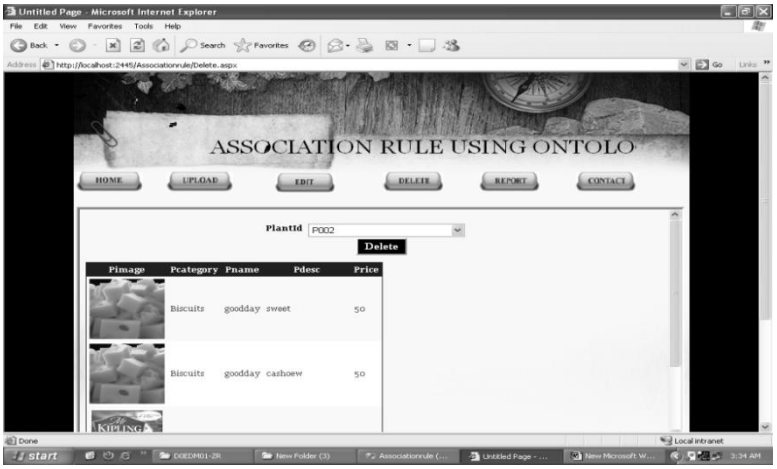

Figure 5: Association rule schema

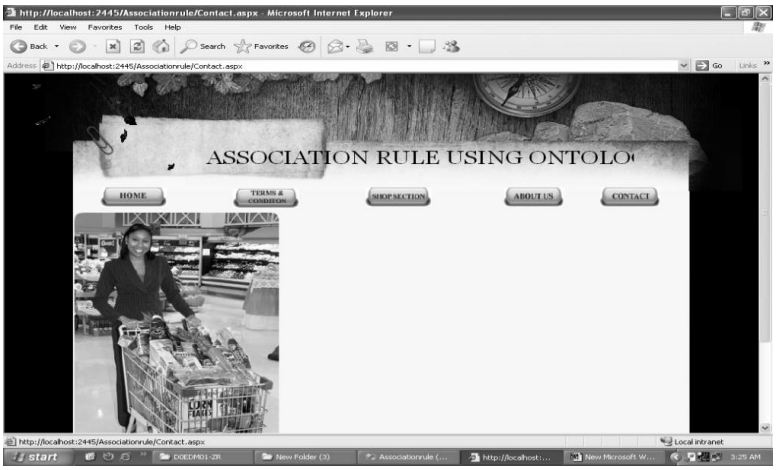

Figure 6: Redudency Reduction Schema

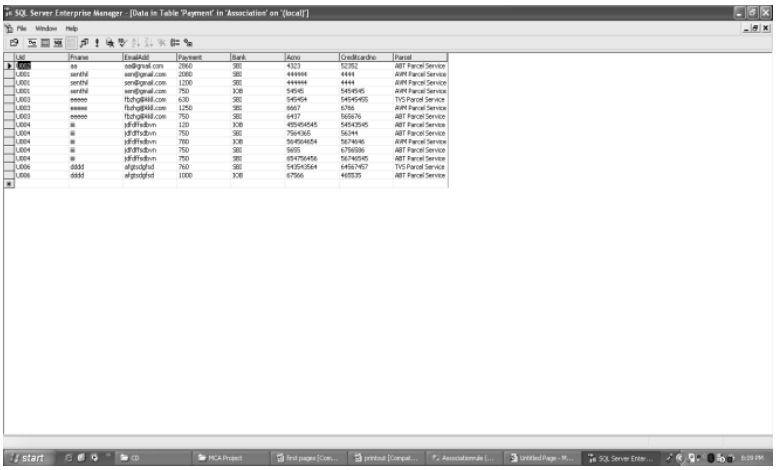

Figure 7: Report Generation

4. Conclusion and future enhancement

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: ontologies and rule schemas. On the one hand, domain ontologies improve the integration of user domain knowledge concerning the database field in the postprocessing step.

F**uture enhancement**

The latter is especially used to express the user expectations and goals concerning the discovered rules

5.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] D.Saravanan, Dr.S.Srinivasan, (2013)., Matrix Based Indexing Technique for video data, Journal of computer science, 9(5), 2013, 534-542.

[3] D.saravanan, Dr.S.Srinivasan (2012). Video image retrieval using data mining Techniques, Journal of computer applications (JCA), Vol V,Issue 01, 2012. 39-42.

[4] D.Saravanan, A.Ramesh Kumar, "ContentBased Image Retrieval using Color Histogram", International journal of computer science and information technology (IJCSIT), Volume 4(2), 2013, Pages 242-245, ISSN: 0975-9646.

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

[6] 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.

[7] D.Saravanan, Dr.S.Srinivasan(2013) "Video information retrieval using :CHEMELEON Clustering" International Journal of Emerging Trends & Technology in Computer Science (IJETTCS), Volume-02,Issue 01, January – February 2013, Pages 166-170.

[8] D.Saravanan, Dr.S.Srinivasan (2011). A proposed new Algorithm for analysis for analysis of Hierarchical clustering in video Data mining, international journal of Data mining and knowledge engineering, vol 3, no 9.

[9] D.Saravanan, Dr.S.Srinivasan (2010).Indexing ad Accessing Video Frames by Histogram Approach, In the Proc. Of International Conference on RSTSCC 2010, 196-1999.

[10] 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.

[11] 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.

[12] D.Saravanan, Effective Multimedia Content Retrieval, International Journal of Applied Environmental Sciences,Volume 10, Number 5 (2015), pp. 1771-17783.

[13] D.Saravanan, Performance Anlaysis of video data image using Clustering Technique, Indian journal of science and technology,Vol 9(10), DOI: 10.17485/ijst/2016/v9i10/79731, March 2016, ISSN (Print) : 0974-6846.

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

[15] D.Saravanan,V.Somasundaram "Matrix Based Sequential Indexing Technique for Video Data Mining "Journal of Theoretical and Applied Information Technology 30th September 2014. Vol. 67 No.3 Pages 725-731.

[16] D. Saravanan, A. Ronald Doni & A. Abisha Ajith (2013), "Image Information Hiding: An Survey", *The SIJ Transactions on Computer Science Engineering & its Applications (CSEA)*, The Standard International Journals (The SIJ), Vol. 1, No. 1, Pp. 1-5.

[17] D.Saravanan "Text information Reterival using Data mining Clustering Technique" International Journal of Applied Engg.Research" Vol.10,No 3(2015),PP 7865-7873