IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661, p- ISSN: 2278-8727Volume 16, Issue 1, Ver. VIII (Feb. 2014), PP 101-103
www.iosrjournals.org

Finding Fast Changing Patterns In the presence of hierarchy

Sahane Priya[1], Prof A.N. Nawathe [2]
[1]*(Computer Engg.,AVCOE/Pune, India)*
[2]*(Computer Engg, AVCOE/ Pune, India)*

Abstract : *History Generalized Algorithm is used in this paper which is used to find the generalized as well as non generalized itemsets .after finding those itemsets find out the particular node in the tree structure and attach those itemsets in specific order. One dynamic pattern, the history generalized pattern ,that represents the development of an itemset in successive time periods, by accounting the information about its recurrent generalizations characterized by minimal redundancy some time it becomes infrequent. Higen mining, The higen miner, that focuses on avoiding itemset mining followed by postprocessing by developing a support-driven itemset generalization .To focus the attention on the minimally redundant recurrent generalizations and reduce the amount of the generated patterns, the finding a subset of higens, namely the nonredundant higens, . Tests do on both real and synthetic datasets show the competence and the effectiveness.*
Keywords: *Data mining algorithm, Data mining Method.*

I. INTRODUCTION

Problem definition: Given an ordered set of timestamped structured datasets, a hierarchy, and a minimum support threshold that is min sup,In this paper show the problem of mining all higens. Definition of HIGEN: D is the dataset D={d_1,d_2,d_3…….d_n}is an ordered sequence with timestamp and It's hierarchy which is built in excess of data items in D_i € D this is case of frequent itemsets. Not generalized itemset and it's minimum support threshold . A HIGEN HG_{it} associated with time stamp or generalized itemsets g_1,g_2……g_n
Following are some condition:
-if sup_count(it,D_i)>=min_sup then gi=it
-else gi=git were git is an predecessor of it with T frequent in D_i & categorized by a minimum generalization level among the set of frequent predecessor of it
Related Work: Literature survey of that concept is that in this topic used different data mining algorithm this algorithm having some limitations and some drawback.To overcome that drawback they have to used higen algorithm and non-redudandant higen algorithms are to be used.
1.Frequent itemset mining Algorithm: The identification of sets of items, products, symptoms and characteristics, which occured together in the given database, is of the more fundamental tasks in Data Mining process. The aim of this algorithm is for searching frequent sets from the need to analyze also called as supermarket transaction data. We introduced mining association rules between sets of items in Large Databases the problem of mining association rules between sets of items in a large database of customer transactions. Each transaction consists of items purchased by a customer in a visit. finding those rules that have:
- Minimum transactional support s | the union of items in the consequent and antecedent of the rule is present in a minimum of s% of transactions in the database.
-Minimum con_dence c | at least c% of transactions in the database that satisfy the antecedent of the rule also satisfy the consequent of the rule.
The rules that we f have find item in the resultant and a union of any number of items in the antecedent. We solve this problem by decomposing it into two subproblems:
1. Finding all itemsets, called large itemsets, that are present in at least s% of transactions.
2. Generating from each large itemset, rules that use items from the large itemset.
steps: join and prune.
1. Join
l finding Lk, a set of candidate k-itemsets is generated by joining Lk-1 with itself
l The items within a transaction or itemset are sorted in lexicographic order
l For the (k-1) itemset: li[1]<li[2]<…<li[k-1]
l The members of Lk-1 are joinable if their first(k-2) items are in common
l Members l1, l2 of Lk-1 are joined if (l1[1]=l2[1]) and (l1[2]=l2[2]) and …
and (l1[k-2]=l2[k-2]) and (l1[k-1]<l2[k-1]) – no duplicates
l The resulting itemset formed by joining l1 and l2 is l1[1], l1[2],…, l1[k-2],l1[k-1], l2[k-1]
2. Prune
-Ck is a superset of Lk, Lk contain those candidates from Ck, which are frequent

-Scanning the database to determine the count of each candidate in Ck –heavy computation
-To reduce the size of Ck the Apriori property is used: if any (k-1) subset of a candidate k-itemset is not in Lk-1, then the candidate cannot be frequent either,so it can be removed from Ck. – subset testing (hash tree)[2]

2. fuzzy decision trees:

Many data mining techniques are available fuzzy techniques le which are discussed as below:Aassociation rule (i) a changed rule if its support or confidence in the duration is different from its support confidence in the preceding duration; (ii) a perished rule if its support and/or confidence less then the user specified thresholds in the duration; and (iii) an added rule if its support and confidence greater than or equal to the user-specified thresholds in the duration. To find the regularities managing the changes in association rules, we introduced to use linguistic variables and linguistic terms to represent the changes and to use fuzzy decision trees to discover the changes. The fuzzy decision trees can then be converted to fuzzy rules.These fuzzy rules are called fuzzy meta-rules they are rules about rules. Furthermore, the discovered fuzzy meta-rules can be used to predict any change in the association rules in the future. To evaluate the performance of our approach, we make use of a set of synthetic datasets, each of which is a set of transactions collected in a specific time period. A set of association rules is discovered in each dataset. Fuzzy decision trees are then constructed in the discovered association rules to mine the changes in these rules. The experimental results show that our approach is very promising.[6]

3. Info Miner Algorithm:

Effective mining algorithm, InfoMiner+, to simultaneously mine significant patterns and the associated subsequences.we introduced a new mining problem of partial periodic pattern with random replacement.[9]

4. Existing Chatterbot system:

Bot is an English simplification for the word robot, which is an agent which collaborates with a user or another program, simulating a human activity. The bots can be categorized in many categories, such as academics, searches, trades, etc; this deals with chatterbots or conversation robots.The goal of this type of bot is to answer and questions, in a particular way that people think they are talking to another person, instead of a computer program. The chatterbots have inside its knowledge base a set of simulated dialogs, to communicate with users in a natural language, and can be used as interfaces in a broad series of applications like electronic commerce, distance learning, among others. The chatterbots use Artificial Intelligence to simulate a dialog with a human being based in a “stimulus – response”routine: you make questions and he provides answers based on those questions. After a question is submitted in a natural language, the program queries a knowledge base and sends an answer trying to mimic the human behavior.

II. Chatter Bot System

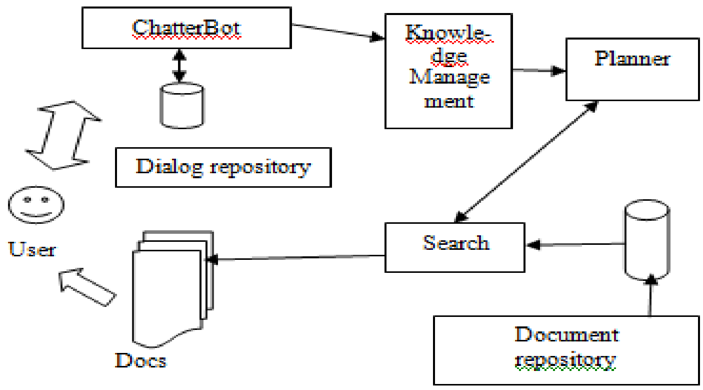

Fig 2.1 Chatter bot system

III. Conclusion

This paper discovered the problem of change mining in the case of frequent itemsets. To represent the development of itemsets in different time periods without discarding relevant but rare knowledge due to minimum support threshold enforcement, it proposes to extract generalized itemsets characterized by minimal redundancy in this case one itemset becomes infrequent in a certain time duration. To this aim, two kinds of dynamic patterns, namely the Higens and the Nonredundant higens, have been introduced.

Acknowledgements

As wise person said about the nobility of the teaching profession, Catch a fish and you feed a man his dinner, but teach a man how to catch a fish, and you feed him for life. In this context, I would like to thank our helpful project guide Prof. Nawathe A.N. who had been an incessant source of inspiration and help. Not only did she inspire me to undertake this assignments, she also advise me throughout its course and help me during my times of trouble.I would like to thank H.O.D. of Department of Computer Engineering Prof. Paikrao R. L. for motivating me.Also, I would like to thank other member of Computer Engineering department who helped me to handle this assignment efficiently, and who were always ready togo out of their were to help us during our times of need.

REFERENCES

[1] Discovering Temporal Change Patterns in the Presence of Taxonomies Luca Cagliero

[2] R. Agrawal, T. Imielinski, and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," ACM SIGMOD Record, vol. 22, pp. 207-216, 1993.

[3] R. Agrawal and G. Psaila, "Active Data Mining," Proc. First Int'l Conf. Knowledge Discovery and Data Mining, pp. 3-8, 1995.

[4] R. Agrawal and R. Srikant, "Mining Generalized Association Rules," Proc. 21th Int'l Conf. Very Large Data Bases (VLDB '95),pp. 407-419, 1995.

[5] M.L. Antonie, O.R. Zaiane, and A. Coman, "Application of Data Mining Techniques for Medical Image Classification," Proc. Second Int'l Workshop Multimedia Data Mining (MDM/KDD '01), 2001.

[6] W.-H. Au and K.C.C. Chan, "Mining Changes in Association Rules: A Fuzzy Approach," Fuzzy Sets Systems, vol. 149, pp. 87-104, Jan. 2005.

[7] E. Baralis, L. Cagliero, T. Cerquitelli, V. D'Elia, and P. Garza,"Support Driven Opportunistic Aggregation for Generalized Itemset Extraction," Proc. IEEE Fifth Int'l Conf. Intelligent Systems (IS '10), 2010.

[8] S. Baron, M. Spiliopoulou, and O. Gnther, "Efficient Monitoring of Patterns in Data Mining Environments," Advances in Databases and Information Systems, L. Kalinichenko, R. Manthey, B. Thalheim,and U. Wloka, eds., vol. 2798, pp. 253-265, Springer, 2003.

[9] M. Bo¨ ttcher, D. Nauck, D. Ruta, and M. Spott, "Towards a Framework for Change Detection in Datasets," Research andDevelopment in Intelligent Systems XXIII, M. Bramer, F. Coenen,and A. Tuson, eds., pp. 115-128, Springer, 2007.

[10] D.W.-L. Cheung, J. Han, V. Ng, and C.Y. Wong, "Maintenance of Discovered Association Rules in Large Databases: An Incremental Updating Technique," Proc. 12th Int'l Conf. Data Eng. (ICDE '96),pp. 106-114, 1996.

[11] C. Clifton, R. Cooley, and J. Rennie, "Topcat: Data Mining for Topic Identification in a Text Corpus," Proc. Third European Conf. Principles and Practice of Knowledge Discovery in Databases,2002.

[12] G. Dong, J. Han, J.M.W. Lam, J. Pei, K. Wang, and W. Zou,"Mining Constrained Gradients in Large Databases," IEEE Trans.Knowledge and Data Eng., vol. 16, no. 8, pp. 922-938, Aug. 2004.

[13] G. Dong and J. Li, "Mining Border Descriptions of Emerging Patterns from Dataset Pairs," Knowledge and Information Systems,vol. 8, pp. 178-202, Aug. 2005.

[14] B. Liu, W. Hsu, and Y. Ma, "Discovering the set of Fundamental Rule Changes," Proc. Seventh ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining, pp. 335-340, 2001.