NAVIGATION PATTERN MINING ALGORITHM FOR CONTENT BASED IMAGE RETRIEVAL

T. Selvaganapathy
M. Tech, Department of IT, Hindustan University
Chennai, India
Mail: selva.ganpy87@gmail.com.

DR. K. Ramesh kumar
Associate professor (IT), Hindustan University
Chennai, India
Mail: **rameshkumark.dr@gmail.com**

Abstract— **Frequent pattern mining algorithm is an interesting problem and most essential field in data mining, a Navigation pattern defines the user's behavior towards a particular area, in this paper we propose an algorithm for mining user navigation Patterns using frequent item sets, by using transaction Reduction method. By using Association rule mining we extract the interesting Correlation and relation between the large volumes of Transactions in the databases. In this paper we merge the navigation pattern mining algorithm with the Content based image retrieval application. Thus by merging it we obtain an effective knowledge discovery process by which it overcomes the draw backs of the existing system This work thus used to extract the navigation patterns discovered from the data bases used in the Content Based image retrieval. And also this work is used as the offline Knowledge discovery process in CBIR.**

Keywords— Association rule mining, frequent item sets, Transaction Reduction, Content based image retrieval.

Introduction

Data Mining refers to extracting or mining knowledge from large amounts of data. In 1993, Agarwal, Imielinski, and Swami introduced a class of regularities, association rules and gave an algorithm for finding such rules [1].Data mining allows users to analyze large databases to solve business decision problems. Data mining is, in some ways, an extension of statistics, with a few artificial intelligence and machine learning twists thrown in. Like statistics, data mining is not a business solution, it is just a technology. For example, consider a catalog retailer who needs to decide who should receive information about a new product. The information operated on by the data mining process is contained in a historical database of previous interactions with customers and the features associated with the customers, such as age, zip code, and their responses. The data mining software would use this historical information to
Build a model of customer behavior that could be used to predict which customers would be likely to respond to the new product. By using this information a marketing manager can select only the customers who are most likely to respond. The operational business software can then feed the results of the decision to the appropriate touch point systems, so that the right customers receive the right offers.

Frequent item sets play an essential role in many data mining tasks that try to find interesting patterns from databases, such as association rules, correlations, sequences, episodes, classifiers, clusters and many more of which the mining of association rules is one of the most popular problems. The original motivation for searching association rules came from the need to analyze so called supermarket transaction data, that is, to examine customer behaviour in terms of the purchased products. Association rules describe how often items are purchased together. For example, an association rule "beer =>chips (80%)"states that four out of five customers that bought beer also bought chips. Such rules can be useful for decisions concerning product pricing, promotions, store layout and many others.

Content-based image retrieval (CBIR) is a technique to search for images relevant to the user's query from an image collection [2]. In the last decade, the conventional CBIR schemes employing relevance feedback have achieved certain success [3]. CBIR consists of online image retrieval and Offline knowledge discovery, these navigation patterns used to predict offline knowledge discovery process. This work plan is to propose an algorithm for mining user navigation patterns using for frequent item sets for transaction Reduction and Association rule mining is to extract the interesting Correlation and relation between the large volumes of Transactions. This work merges the new navigation pattern mining algorithm with image retrieval application. Finally, this work will compare the new algorithm with the CBIR technique. the rest of the paper is organized as follows: section 2 describes the problem study in the navigation pattern mining, section 3 describes a brief of past studies in Navigation pattern mining as well as the CBIR ,followed by section 4 which describes the methods to improve the efficiency of apriori algorithm, in section 5 we propose the algorithm for the navigation pattern mining & merging the algorithm with the CBIR ,in section 6 the experimental results are discussed ,finally we conclude in section 7.

2. Problem study

2.1 Need of Frequent Item set Mining?

Studies of frequent item set (or pattern) mining is acknowledged in the data mining field because of its broad applications in mining association rules, correlations , and graph pattern constraints based on frequent patterns, sequential patterns ,and many other data mining tasks. Efficient algorithms for frequent items are crucial for mining association rules as well as for many other data mining tasks. The major challenge found infrequent pattern mining is a large number of result patterns.

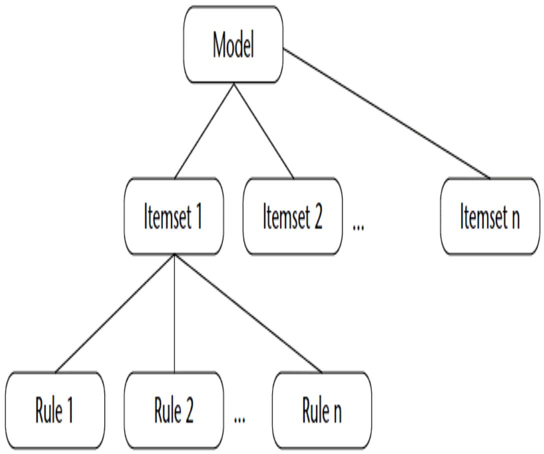


Figure 2: model for frequent item – sets mining using association rules.

As the Minimum threshold becomes lower; an exponentially large number of item sets are generated. Therefore, pruning unimportant patterns can be done effectively in mining process and that becomes one of the main topics in frequent pattern mining. Consequently, the main aim is to optimize the process of finding patterns which should be efficient, scalable and can detect the important patterns which can be used in various ways. Figure 2 describes the model for frequent item set mining, using association rules.

3. Related work

Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, Fellow, and Vincent S. Tseng [4] has developed the following concepts, Content based image retrieval (CBIR)is the efficient way for the image retrieval system, To resolve the aforementioned problems, they propose a novel method named navigation-Pattern-based Relevance Feedback (NPRF) to achieve the high retrieval quality of CBIR with RF by using the discovered navigation patterns. The expected scenario for effectiveness versus efficiency is clearly illustrated in it. This paper focuses on the discovery of relations among the users' browsing behaviors on RF. Basically the frequent patterns mined from the user logs are regarded as the useful browsing paths to optimize the search direction on RF. In their NPRF approach, the users' common interests can be represented by the discovered frequent patterns (also called frequent item sets). Through these navigation patterns, the user's intention can be precisely captured in a shorter query process. In this phase, the Apriori-like algorithm is performed to exploit navigation patterns using the transformed data.

Peng-Yeng Yin, Shin-HueiLi [5] has developed the following concepts, with the rapid development of internet technology, the transmission and access of image items have become easier and the volume of image repository is exploding. Association rule mining has been successfully applied to many business and industrial domains such as catalogue design, cross marketing, and loss-leader analysis [6]. Association rule mining finds significant associations or correlation relationships among a large repository of data items. These relationships, represented as rules, can assist the users to make their decisions. Since modern image databases experience with many users, the collective feedback from their interactions with the system can be exploited using the association rule mining. The association rules with high confidence indicate the consensus of relevance concept among multiple users. The related images of association rules with the highest rule confidence are likely to be the most relevant ones. Hence, the first display of retrieved images to a particular query can be determined by inferring image relevance from association rules and will attain a high precision rate.

Ali Selamat, Muhammad Khairi[7] Ismail developed the following concepts on CONTENT BASED IMAGE RETRIEVAL. The performance of the Content Based Image Retrieval (CBIR) can compute using similarity of the images where user can retrieve from the image database. They have randomly selected one of the 1000 images as query, and other 999 remaining images as two datasets of training samples. The dataset A contains only the most attraction place in Malaysia and the second dataset with B label contains only the images of places in Malaysian state. they have used guftan- kessel algorithm for their purpose of image retrieval.

4. Methods to improve apriori efficiency

i. Hash-based item set counting: A k-item set whose corresponding hashing bucket count is below the threshold cannot be frequent.

ii. Transaction reduction: A transaction that does not contain any frequent k-item set is useless in subsequent scans.

iii. Partitioning: Any item set that is potentially frequent in DB must be frequent in at least one of the partitions of DB.

iv. Sampling: mining on a subset of given data, lower support threshold + a method to determine the completeness.

v. Dynamic item set counting: add new candidate item sets only when all of their subsets are estimated to be frequent.

This work uses Transaction reduction Method for improving Apriori's efficiency

5. Navigation Pattern Algorithm

This algorithm uses the Transaction reduction method by which it overcomes the draw backs of Apriori algorithm.

5.1 Theorem & Algorithm

Theorem: If c ε C_{k-1} and c. support<min sup, T items< k, m = 1, then c is useless in C_{k-1} where T items is total item count in each transaction and m is the no. of combinations.

Algorithm:

INPUT : D, min sup

OUTPUT: L (D, min sup)

1) L_1 = {large l-item sets};

2) $\overline{C}_1$= Database D; (With all items not in L1 and ∀ t_i T_{items}=1 removed)

3) For (k=2; $\overline{C}_{k-1} \neq \Phi$; k++) do begin

4) C_k= Apriori-gen(L_{k-1}) // New candidates

5) $\overline{C}_k$= Φ;

6) For all c εC_k do begin

7) $\overline{C}$ =Φ;

8) T_c={t.TID | t ε $\overline{C}_{k-1}$,(c –c[k])ε t. set-of item sets Λ (c–c[k]) ε t . set-of-item sets)}

9) If $|T_c|$ ≥min sup then begin

10) $L_k = \bigcup_k \{c\}$;

11) For all p ε T_c do

12) If (T_{items}> k) then begin

13) $\overline{C} = \bigcup_k < p, c >$;

14) end

15) end

16) If $|\overline{C}|$ ≠1 then begin

17) $\overline{C}_k=\overline{C}$;

18) end

19) end

20) end

21) end

22) Ans = $\bigcup_k L_k$;

5.2 Merging New Algorithm with CBIR

The navigation pattern mining is used to discover the user navigation paths, the offline discovery of CBIR includes a new algorithm by which the frequently occurring item sets are stored in log databases, and these log data bases are useful in finding the navigation behaviour of users. Once the user logs in, he navigates through a set of images all the images are stored in the large databases. These images are assigned to an value for creating the frequent item sets, once these frequent items are created we can discover the knowledge by the navigation behaviour using association rules.

Association rules are thus used to create interesting correlation between the item sets. In our approach we take 1000 images and find to discover the correlation between them, from their browsed behaviours. Each navigation path is stored in the log data bases and used for the generation of knowledge using association rules here it uses the navigation pattern algorithm which effectively overcomes the drawbacks of apriori. This proposed algorithm is thus reduces the database scans. And thus produces the desired results. The proposed algorithm is merged with the offline discovery of image retrieval of CBIR.

6. Experimental results

The Experimental results show the knowledge discovery process using the proposed algorithm. Here we use the transaction file which consists of the transactions obtained in n*m table. Each row represents transactions and columns separated by space and represents item sets. the transactions are mapped to binary values each line represents an single transaction as of the example taken here. These 0's & 1's represent the images whether the images are visited or not, an visited image is denoted by 1 and the other by 0.here these transaction file is given as an input for the knowledge discovery process. The next input file given here is the configuration file, the configuration file consists of the following lines, they are number of items, number of transactions, minimum support. The output of these is written in a separate file. The output consists of the frequent items which are present in the transaction these are the knowledge we obtain from the data sets or data bases. These frequent items are mapped to that of the images in the data base these frequent items are thus mapped and they are displayed in the output which will be representing the knowledge that are obtained from the large data base. This process represents the offline discovery of the data base in the content based image retrieval. we also display the navigation patterns by which the images are traversed along with the knowledge discovery, by using the proposed algorithm. The algorithm using Transaction reduction thus overcomes the drawbacks in apriori algorithm and it proves to be effective than the existing methods

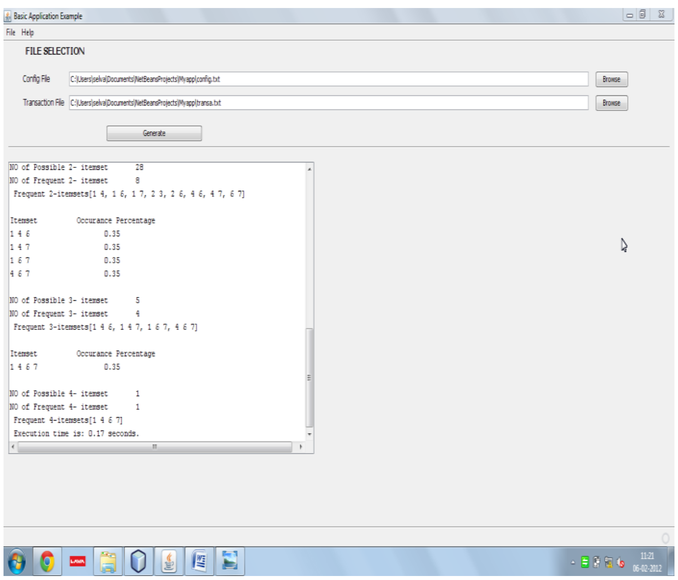


Figure 2: screenshot for generation of frequent items from CBIR data sets in net beans IDE.

Input configuration: 8 items, 10000 transactions, min sup = 0.3%

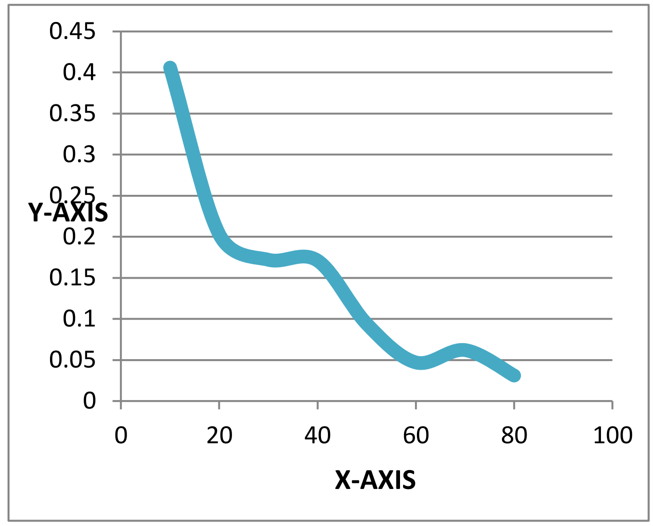
Algorithm has started...........

Item set	Occurrence Percentage
1	0.5
2	0.55
3	0.45
4	0.5
5	0.4
6	0.55
7	0.5
8	0.45

NO of Possible 1- item set 8

NO of Frequent 1- item set 8

Frequent 1-itemsets [1, 2, 3, 4, 5, 6, 7, 8]



Graph 1: run time vs. min support

Graph runtimes of various min supports

X-axis: minimum support

Y-axis: run time

Item set	Occurrence Percentage
1 4	0.35
1 6	0.4
1 7	0.35
2 3	0.3
2 6	0.3
4 6	0.4
4 7	0.4
6 7	0.35

NO of Possible 2- item set 28

No of frequent 2- item set 8

Frequent 2-item sets [1 4, 1 6, 1 7, 2 3, 2 6, 4 6, 4 7, 6 7]

Item set	Occurrence Percentage
1 4 6	0.35
1 4 7	0.35

1 6 7	0.35
4 6 7	0.35

NO of Possible 3- item set 5
NO of Frequent 3- item set 4
Frequent 3-itemsets [1 4 6, 1 4 7, 1 6 7, 4 6 7]

Item set	Occurrence Percentage
1 4 6 7	0.35

NO of Possible 4- item set 1
NO of Frequent 4- item set 1
Frequent 4-itemsets [1 4 6 7]
Execution time is: 0.265 seconds.

7. Conclusion

Thus the above proposed algorithm uses the Transaction reduction method by which it helps in reducing the data base scans. And also it eliminates the transactions in which it does not contain any frequent k-item set, which is useless in the subsequent transactions. Thus this algorithm greatly improves the image retrieval by using the proposed approach .this work can be performed for large databases with minimal transaction reduction .thus this algorithm is more effective & efficient when compared to that of the existing methods used for the offline knowledge discovery of CBIR systems. The future work will include the online image retrieval process of CBIR.

References

[1]. R. Agrawal, J.Shafer, .Parallel mining of association rules., IEEE Transactions on knowledge and Data Engineering, 8(6), December 1996.

[2]. K.A. Hua, and N. Jiang, "Improving Image Retrieval Effectiveness in Query-by-Example Environment," Proc. 2003 ACM Symp. Applied Computing, pp. 774-781, 2003.

[3]. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, Dec. 2000.

[4]. Ja-Hwung Su, Wei-Jyun Huang, Philip S. Yu, and Vincent S. Tseng, "Efficient Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patterns" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 3, MARCH 2011

[5]. Peng-Yeng Yin *, Shin-Huei Li" Content-based image retrieval using association rule mining with soft relevance feedback" J. Vis. Commun. Image R. 17 (2006) 1108–1125

[6]. J. Han, M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers, San Francisco, 2001.

[7]. Ali Selamat, Muhammad Khairi Ismail" Effectiveness of Relevance Feedback for Content Based Image Retrieval Using Gustafson-Kessel Algorithm" Proceedings of the International Conference on Computer and Communication Engineering 2008 May 13-15, 2008 Kuala Lumpur, Malaysia