Implementation of Apriori Algorithm on Electricity Billing System

Vipin Saxena
Professor, Deptt. of Computer Science
B. B. Ambedkar University
Lucknow (U.P.) 226025, INDIA

Sudhir Kumar Singh
Deptt. of Computer Science and Engineering
Bhagwant University, Ajmer, Rajasthan,INDIA

ABSTRACT

The relationships among the classes play an important role for the selection of the object-oriented database. In this context, association rules show the relationship between the two classes. As the object-oriented design for the development of the software grows, the relationships among the classes also grow towards the complex software design but one can develop optimize object-oriented design through right association rules between classes. The present paper deals to design the right association rules for the object-oriented databases taken from the classes. A real case study of Electricity Bill Deposit System (EBDS) is considered in Indian scenario and Apriori algorithm is used for finding the use of frequent data sets through the right association rules. The association rules are designed through well-known Unified Modeling Language (UML). The present work is an implementation of Apriori algorithm towards the database of EBDS.

Keywords

UML, Apriori Algorithm, Lattice Structure, Association Rules, Data Item Sets.

1. RELATED WORK

Association rules are used to find out the hidden relationship from the stored database. Association rule is a key concept of data mining technique for finding the hidden relationship. Association rules works like if and then statement. Association rules are dividing in two parts, one is antecedent other is consequent. An antecedent finds the items from the stored large data item while consequent is an item found in dataset and then combined with antecedent. Association rules are created by analyzing data for frequent item set. Support and Confidence are most important relationships in association rules. Support is an indication, how it appears in the database while Confidence shows how many times it is appearing in database.

Let us first describe some of the important references related to the concepts of data mining. Agrawal et al. [1] have described the irregularities from the large scale transaction data generated at point-of-sale in supermarkets by the use of association rules. Sergey et al. [2] have considered the problem of analyzing market-basket data and they presented some contributions for finding large item sets based on sampling. Cai et al. [3] have proposed a technique, where the items are considered as per weight. The   special promotions are given to the item which maximise the revenue of different items. Authors also described weighted association rules with the weight which presented the algorithm to explore the problem of down words closure property which support measure in the un-weighted case and the case is no longer exit. Dunkel and Sparkar [4] have described it as the main aspects namely representation organization and the next one is access of the particular data. When the loss of input and output are considered the access speed of data are significantly affect the performance. Authors also used one thing, comparison between column and row wise approaches and the data accessed through Apriori association algorithm and it is found that counting using Apriori algorithm with data accessed in column wise manner is better by reducing the degree to which the data as well as contents needs to be rapidly brought into the memory. Walter [5] has described the idea of tupple or element recording which helps to improve the low level efficiency of the algorithm. Burdick et al. [6] have described a depth first traversal of item set lattice and developed as a search as used as strategy algorithm which is the effective pruning mechanism for mining performance. A combination of vertical bitmap representation of data and efficient bitmap compression schema are described and dependent upon for implementation of support counting. Chai [7] has presented the advance Apriori algorithm which enhances the perfection of generating association rules in order to adopt new method to depth the association problem of redundant generation of rules itemsets. Taboada [8] has described the concept of data mining for extracting useful hidden knowledge from large volumes of data for their use in decision support system.Weng et al. [9] have described the relationships hidden in large data sets by association analysis for frequent item sets or association analysis which highlight the relationship to reduce support or confidence of the rules. The other important related reference on data mining is [10].

Umarani [11] has introduced identification of association rule in large databases in order to form the first step in mining of association rules in the process of frequent itemset mining. Verma [12] has described the database community mostly on the association rule of data mining. The rule seeks to correlate among attributes in a database and is important in generating frequent item sets from large database. Regular retrieval and storage of huge data have leaded the Industries to develop association rules from the database. He et al. [13] have described aggregate values of varying attributes computed by various systems by processing data in large databases and/or warehouses.

In the present work, association rules have been designed to find out hidden relationship from a large amount of data stored in database. Object-oriented database has been used for the EBDS in Indian context. Association rules are represented through Unified Modeling Language class diagram. Apriori algorithm is used for finding the frequent used database from the large data sets and represented through diagram. The present work is useful when the size of database is growing day-by-day in many business organizations.

2. UNIFIED MODELING LANGUAGE FOR ASSOCIATION RULES

Unified Modeling Language (UML) represents the graphical view of the modeling language and the important feature of UML is to design the software models. It is very useful for

analysis of any research problem and UML class diagram depicts the class name, attributes and operations. UML class diagram divides the data into three rows. First row shows the class name, second row shows the attribute name while third column shows the operations. The present work is a case study of EBDS and figure 1 shows the UML class representation of EBDS and it consists of relationship between Meter and Plan. The association is that the Meter distributed according to Plan. A customer depicts take a different type of plans depending on the condition. The UML class diagram as shown in figure 2 depicts the relationship between Meter, Plan and Meter_Reader. A customer depicts the electricity bill depending on the total reading. The UML class diagram as shown in figure 3 represents the relation between Meter, Plan, Meter_Reader and Bill. This diagram shows that customer may take any type of plan to deposit the electricity bill. However, the amount will inadvertently depend on the total reading.
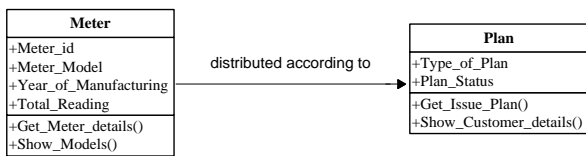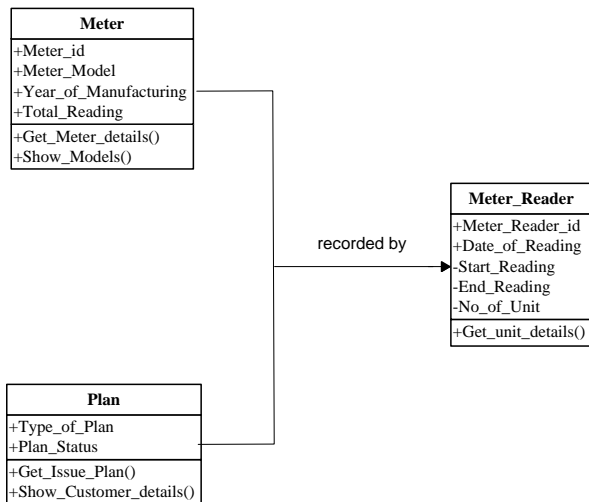
Fig 1: Representation of Binary Association

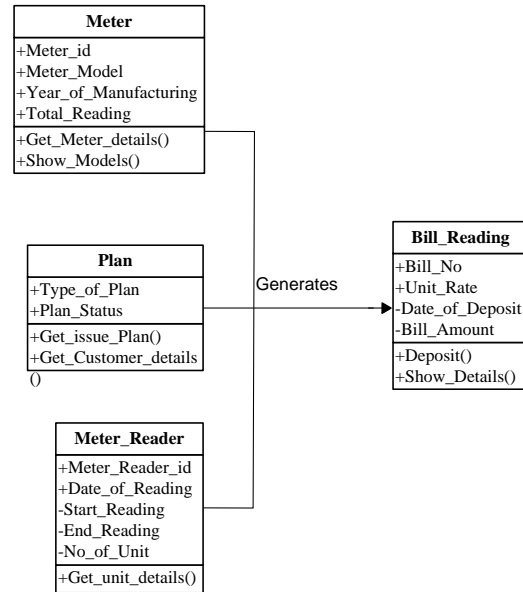**Fig 2: Representation of Ternary Association**

Fig 3: Representation of Many to One Association

3. DESIGN OF ASSOCIATION RULES

Let us design the association rules for which table t1 is created for the object-oriented database as represented in Table 1. Object-oriented database is better than the relational database because the relation of object-oriented database establishes the connection between objects and not a table. Table t1 stores different data item sets. Applying the association rules in object-oriented database generates query processing capability and finds out the hidden relationship data.

Customer records are stored in data items. Large amount of electrical bill data is stored in data items. This paper mainly deals with customers' deposit of the electricity bill and recording stored figures in data items. Many types of combinations are stored in data items. Association rule is applied in electrical bill record to find out hidden frequent items.

Table 1: Items Sets for Electrical Bill System

SNo.	Items Name
1.	{Meter_id, Type_of_Plan}
2.	{Meter_id,Type_of_Plan,Date_of_Reading, Bill_Amount}
3.	{Bill_Amount, Type_of_Plan, Date_of_ Reading,Unit_Rate}
4.	{Meter_id,Type_of_Plan,Date_of_Reading,No_of_Unit, Bill_Amount}
5.	{Date_of_Reading, Meter_id, Type_of_Plan, No_of_Unit, Unit_Rate, Bill_Amount}

Let us establish binary association which is represented by means of 1 or 0. Items can be presented in the binary form. Large numbers of data is stored in the form of item data sets. Items are stored in serial order. If items are present in the

transaction, it is represented by 1 and if items are not present in the transaction then it is reflected by 0. This type of representation does not show Meter_id, Type_of_Plan, Date_of_Reading and Bill_Amount is being presented in the first position in items data set. It means Meter_id, Type_of_Plan, Date_of_Reading and Bill_Amount have a value of 1 and while all other values are 0. Binary Representation of EBDS is shown below in Table 2.

Table 1: Binary Association of Data Items

TID	Meter_id	Type_of_Plan	No_of_Unit	Unit_Rate	Date_of_Reading	Bill_Amount
1	1	1	0	0	0	0
2	1	1	0	0	1	0
3	0	1	0	1	1	1
4	1	1	1	0	1	1
5	1	1	1	1	1	1

Now, let I= (i_1, i_2, i_3..........i_d) be the set of all data items for the electricity bill deposit system and T= (t1,t2,.......t_n) be the set of all transitions. Each transition t_i is a subset of items i. Item set is represented in the above. Every item is present as a candidate1-itemset. Low count item will be removed from the item set i.e. Items with count of less than 3 will be removed. For example Less than 3 count items are No_of_Unit and Unit_Rate. No_of_Unit and Unit_Rate will be removed in Candidate1-Itemset.

Table 3: Candidate 1-Item set

Item	Count
Meter_id	4
Type_of_Plan	5
No_of_Unit	2
Unit_Rate	2
Date_of_Reading	4
Bill_Amount	3

Items removed count lees than 3

The removed items of Candidate 1 item set do not appeared in Candidate 2 item set because their value is less than 3. Candidate 2 item set is a combination of two item sets. Apriori algorithm postulates that all the superset of the infrequent items must be infrequent and low support item will be removed.

Table 4: Candidate 2-Itemset

Item sets	Count
Meter_id, Type_of_Plan	4
Meter_id, No_of_Unit	2
Meter_id, Unit_Rate	2
Meter_id, Date_of_Reading	4
Meter_id, Bill_Amount	3

Less than 3 count items removed

Candidate 2 itemset removed items cannot appeared in Candidate3-itemset. Candidate 3-itemset is a combination of three itemsets. Less than 3 support items will be removed according to the Apriori algorithm law and low support item sets are discarded. From the table 5 it is depicted that customer may deposit the Bill_Amount according to Meter_id and Type_of_Plan with Date_of_Reading.

Table 5: Candidate 3-Itemset

Item sets	Count
Meter_id, Type_of_Plan, Date_of_Reading	3
Meter_id, Type_of_Plan, Bill_Amount	3

4. LATTICE SEQUENCE THROUGH APRIORI RULES

The Apriori rules said all data items are frequent and subset must also be frequent. For example any data items that contains {a, b, c} must here subsets {a, b} {a, c} {b, c} and {a}, {b},{c}. Let us design the lattice structure which consists of data item sets as a, b, c, d, e and f represented by Meter_id, Type_of_Plan, No_of_Unit, Unit_Rate, Date_of_Reading and Bill_Amount. The root of the tree is always Null . The lattice structure containing above by using Apriori algorithm is shown below in figure 4. The binary representation shows the combination of data items. If a is combination b, c, d, e, f then stable the relation ab, ac, ad, ae, af. b is combination c, d, e, f then stable the relation bc, bd, be, bf. c is combination d, e, f then stable the relation cd, c e, c f. d is combination e and f then stable the relation de and df. e is combination of f then shows the relation ef.
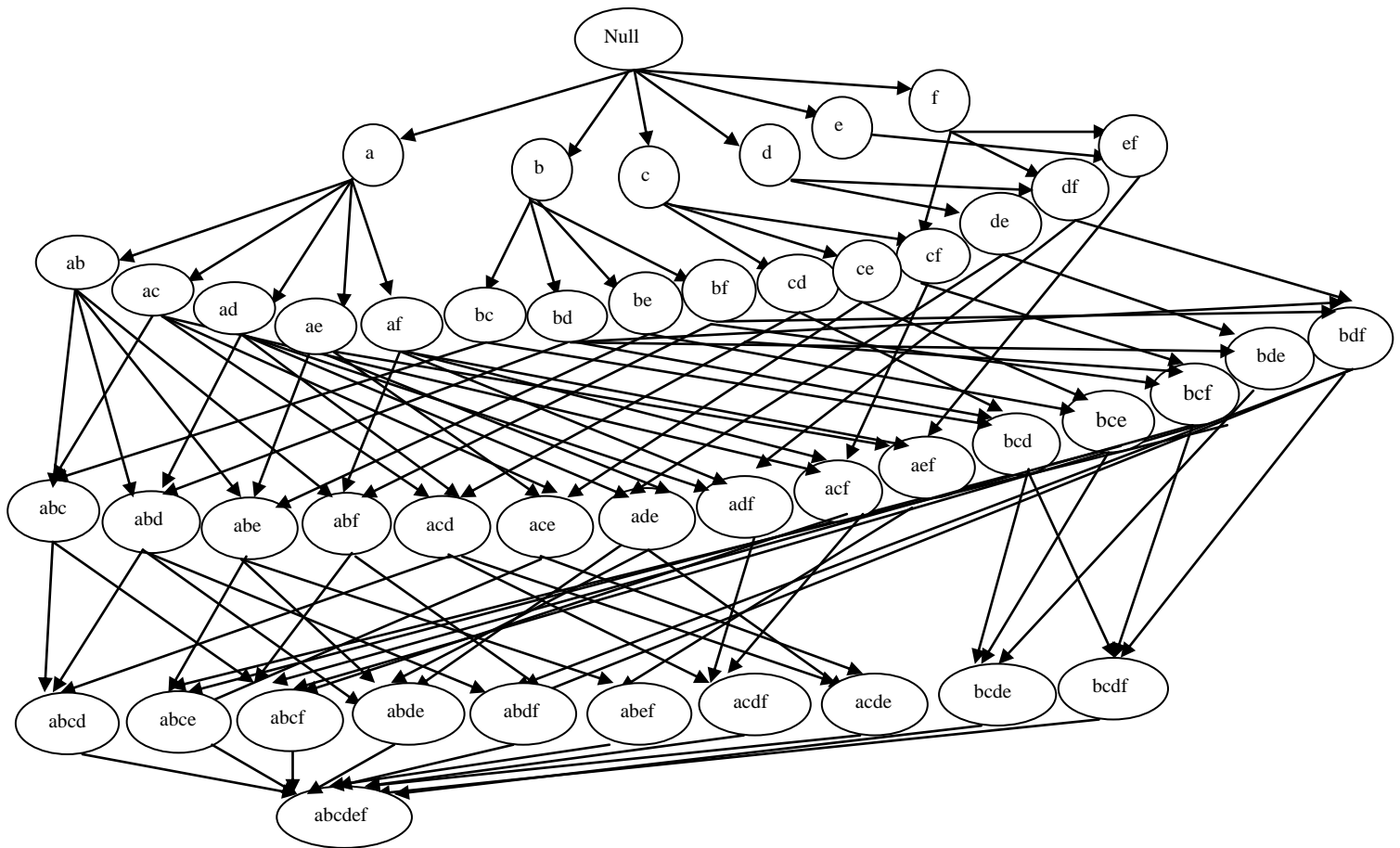
Fig.4: Lattice Structure of Data Item Sets

5. CONCLUDING REMARKS

From the above work it is concluded that UML is a powerful tool for representing the relationship among the object oriented classes by means of association. By the use of Aprori algorithm, one can get the information about the most frequent data item Set. It is implemented on the electricity bill deposit system in context of Indian scenario. The data item Sets are represented through lattice structure which is a combination of data item Sets representing the most frequent used object-oriented data items taken from the UML class diagrams. The above work can be extended by writing the Java code to display the electricity bill information on the mobile hand-held device.

6. REFERENCES

[1] Agrawal, R., Imielinski, T., Swami, A., "Mining Association Rules Between Sets of Items in Large Databases", SIGMOD, Page 207-216, 1993.

[2] Sergey, B., Motwani, R., Jeffrey D., Shalom, T., "Dynamic Itemset Counting and Implication Rules for Market Basket Data", International Conference proceedings ACM SIGMOD, Page No 255-264, May 13-15, 1997. .

[3] Cai, C.H., Fu, A.W.C., Cheng, C.H., Kwong, W.W., "Mining Association Rules with Weighted Items", International Proceedings. IDEAS'98 Database Engineering and Applications Symposium, Page No 68–77, 1998.

[4] Dunkel, B., Soparkar , N., "Data organization and access for efficient data mining ",15th International Conference on Data Engineering Proceeding ,Page No 522-529,23-26 Mar 1999.

[5] Walter A. Kosters, Elena Marchiori, and A. J. Oerlemans, "Mining Clusters with Association Rules", Proceedings of Symposium on Advances in Intelligent Data Analysis, ISBN 3-540-66332-0 IN 1999.

[6] Burdick, D., Calimlim , M., Flannick,J., Gehrke, J., Yiu , T., "MAFIA: A Maximal Frequent Itemset Algorithm," IEEE Transactions on Knowledge and Data Engineering, vol. 17, no. 11, pp. 1490-1504, Nov. 2005.

[7] Chai, S., "The Research of Improved Apriori Algorithm for Mining Association Rules", This Paper Published in Service Systems and Service Management International Conference, Page 1-4, June 2007.

[8] Taboada, k., "Association rules mining for handling continuous attributes using genetic network programming and fuzzy membership functions", This Paper Presented in IEEE, Page 2723-2729, 2007.

[9] Weng, C. C., Chen, T.S., Lo, C.H., "A Novel Algorithm for Completely Hiding Sensitive Association Rules", IEEE, Page 202-208, 2008.

[10] Tan, Ping-Ning, Steinbach, Michael, Kumar, Vipin, "Introduction to Data Mining", Pearson Education, ISBN 978-81-317-1472-0, Fourth Edition, 2009.

[11] Umarani, V., "On developing an effectual progressive sampling based approach for association rule discovery" This Paper Published in Information Management and Engineering (ICIME), Page No 8-12, 2010.

[12] Verma, G., "An Effectual Algorithm for Frequent Itemsets Generation in Generalized Data Set Using Parallel Mesh Transposition" This Paper Presenting in IEEE, Page 719-724, 2012.

[13] He, B., Hsiao ,I .H., Liu, Z., Huang, Y., Chen, Y., "Efficient iceberg Query Evaluation Using Compressed Bitmap Index ", This Paper Appear in IEEE, Page 1570-1583, 2012.