MINING USER DEMANDED DATA FROM DATA MART USING INDUCTIVE RULE MINING

Paulraj M and Sivaprakasam P
Department of Computer Science, Vinayaka Missions University
Salem, Tamil Nadu, India
Department of Computer Science, Sri Vasavi College
Erode , Tamil Nadu – 638 316, Tamil Nadu, India

ABSTRACT

A data warehouse is a database used for reporting and analyzing the data stored in the repositories. A data mart acts as the accessing form of the data warehouse situation used to obtain the data out to the users. Accessing of data in the data warehouse is a challenging approach since the user needs better understanding of the data structure stored in the repositories. To handle this issue, data mart is introduced. Data marts built separate functional data repository layers based on the requirements and applications of the corporate data applications. However function requirements of users are not easily understood by the data warehouse model. It needs efficient decision support system to extract the required user demanded data from data warehouse. Existing work only identified the functional activities of the data mart based on attribute relativity in the data mart and it does not extract the user demanded information from the repositories. To improve the decision support system to extract the required user demanded data from data mart, inductive rule mining is used. The decision support is done with inductive rule mining on functional data marts segregates of the layered data repository and extracted the required information for the user. The induced rules are proposed to be the supportive knowledge for identifying the user needed information. An experimental evaluation is conducted with benchmark datasets from UCI repository data sets and compared with existing functional behavior pattern for data mart based on attribute relativity in terms of number of decision rules, extracted data relativity, analysis of functional behavior.

Key words: Decision support system, Inductive rule mining, Inductive rules, Data mart, Data warehouse

1. INTRODUCTION

The word Decision Support (DS) is used regularly and in a mixture of frameworks linked to decision making. Recently, for instance, it is frequently declared in relation with Data Warehouses (DW) and On-Line Analytical Processing (OLAP). Another current tendency is to connect DS with Data Mining. A data warehouse sustains a copy of data from the resource operation systems. This architectural complication presents the chance to continue data history, incorporate data from manifold source systems, allowing an inner analysis across the endeavor. The data quality in DW normally improved, by presenting reliable codes and metaphors, deteriorating or yet fitting bad data. The DW presents the association's information constantly and offers a distinct general data model for all data of interest apart from the data's source.

Decision support usually engages the combination of data and knowledge organization to support human on creating efficient alternatives. In the framework of online route contented delivery scalable to well distinctive individual, decision making is on the source of stable varying requirements that entail a rapid reaction. Conventional instinctive methods of decision making are no longer sufficient to contract with such obscured situation. The role of Decision Support in Decision Making process is shown in fig 1.

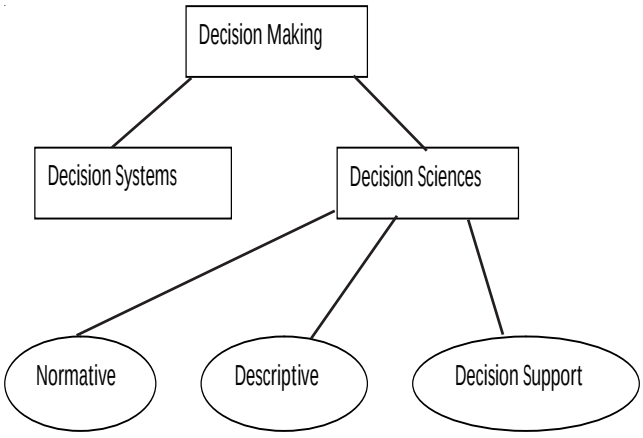

Fig 1 role of DS in Decision making process

A data mart is a decision support system including a separation of the endeavor's data paying attention on precise functions or behaviors of the endeavor. Data marts have specific trade values such as determining the collision of advertising promotions, or determining and forecasting sales presentation, or determining the collision of new product prologues on company profits, or determining and forecasting the presentation of a novel business division. Data Marts are precise trade software applications. A data warehouse, different from a data mart, contracts with numerous subject areas and is naturally executed and proscribed by an essential organizational unit such as the corporate Information Technology (IT) group.

To improve the data extraction process fr

mart, in this work, inductive rule mining algorithm is used with an extensive approach used in the data mart for an extraction of information from the data storage repositories by forming a set of decisive rules from the functional behavior pattern framework of the data mart.

2. LITERATURE REVIEW

A data mart is an uncomplicated structure of a data warehouse [12] that is alerted on a distinct issue (or functional area), such as Finance, Sales, or Marketing. Data marts [11] are regularly constructed and restricted with data mart architecture [8] by a distinct section within an organization. The author also presented a tool for formalizing the data mart with DaWaII [9]. Discretization is a serious constituent of data mining whereby uninterrupted attributes of a dataset are changed into separate ones by producing periods either before or through learning. The author presents a new discretization technique EDISC [1] which uses the entropy based standard but receives a class-tailored approach to discretization.

Inductive learning methods [2] permit the scheme exclusive to deduce a form of the pertinent occurrence of an unidentified process by mining information from tentative data. An extensive choice of inductive learning methods [5] is now obtainable, potentially ensuring diverse point of accuracy on dissimilar difficulty domains. Mining of inconsistent data in data warehouse is a challenging approach, the author in [3] presents three diverse strategies to assemble voting based on maintenance, power and majority in the bagged MLEM2 algorithm for rule induction [4].

In data warehouse, mining essential information from the database is an important concept. Induction based decision rule model used here for generating inferences and implicit hidden behavioral aspects in the web usage mining [6] which investigates at the web server and client logs. Inductive learning is a proficient method to build knowledge from the examination of a set of cases. The author [7] centers on inductive learning algorithms that are mainly used in data mining and it allows the creation of a fuzzy decision tree which characterizes a set of decision rules with rough set theory [10]. To improve the extraction of data from data mart, in this work, we are using an inductive rule mining algorithm.

3. EXTRACT USER DEMANDED DATA FROM DATA MART USING INDUCTIVE RULE MINING

The proposed work is efficiently designed for extracting the user demanded information from the data storage repositories using inductive rule mining (IRM). The proposed work of extracting user needed information follows two set of operations. The first operation describes the process of decision making process and the second operation is to describe the decision inductive rule mining for extracting the information based on inductive rules derived from the large set of repositories. The process of the complete proposed work is shown in fig 3.1.

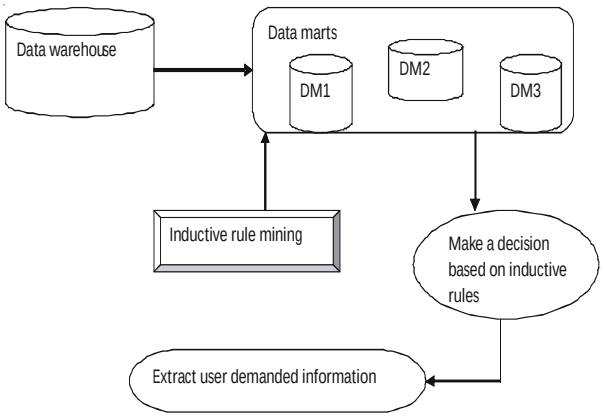

Fig 1 Architecture diagram of the proposed extracting user demanded information from IRM

At first, data mart is a subset of data warehouse. The data mart consists of set of information based on the operational goal of the attributes present in the traditional organization. After organizing the data mart, the inductive rule mining is applied to derive a set of rules based on the functional attributes of the data present. The IRM will present a set of rule and it easily makes the decision process of which set of information is going to be extracted by the user. The formation of decision rules with a set of inductive rule mining process is computed based on the processes involved.

3.1 Decision Support system in data mart

Certainly, DS is an ingredient of resolution making processes. A decision is termed as the selection of one between a number of substitutes, and Decision Making signifies to the entire procedure of building the choice, which contains:

- Evaluating the problem,
- Gathering and validating information,
- Recognizing choices,
- Predicting consequences of decisions,
- calculate decisions.

The decision making process is done with three main strategies:

1. *Intelligence:* Finding problem and analysis.
2. *Design:* Formulation of solutions, representation and simulation.
3. *Choice:* Decision making, and implementation.

In data mart, the process of identifying the functional operational goal of the system is analyzed and being processed. The problem of determining the analysis of the given validated information in the DW is fo

The representation of problem and the decision making process is implemented with solution and simulation of the information available in the DW.

3.2 Decision rule induction for extracting information

With the promising growth in data warehousing technology, a massive expensive resource of information from which we can persuade functional decision rules. But with the traditional method, the number of engendered decision rules is incredible. In this work, intend a diverse strategy of inducing definite and probable decision rules and the induction process is generated by the query. The information concerning about the users' needs is stored in a table structure and decision rules are persuaded by posing query on any attribute. Using this scheme, the relevant inductive rules are minimized. The framework of the proposed inductive rule mining for extracting the user needed information is shown in Fig. 3.2.

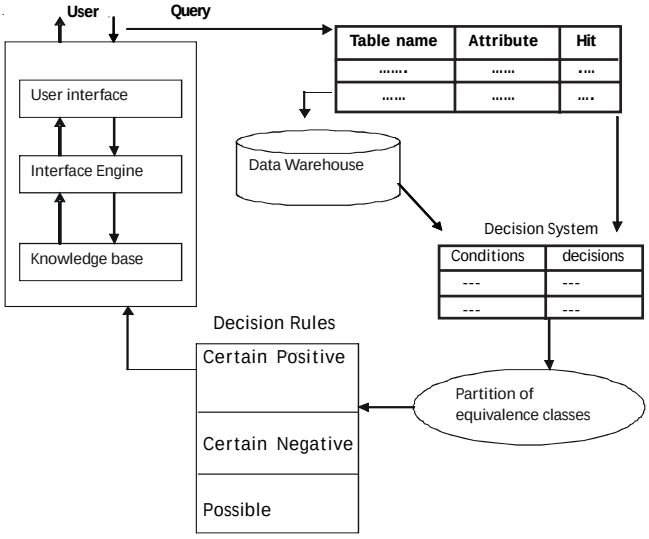

Fig 3.2 Process of Decision rule induction for extracting user information

The proposed framework of decision rule induction for extracting the users' needed information is invoked by users' query. Once the query has been processed, the supporting data structure has been restructured with the tables and attribute's name mined from the query. The column hit in the table form calculates the number of times that the attributes has been used. The counter is varied in descending order to set the most commonly used attribute in the first row. This attribute forever referred to by users' queries; hence it is importance in producing decision rules standed on this attribute value. The approach of inducing decision rules supported on the most commonly inquired attribute is described in the following algorithm.

Algorithm Decision Rule Induction

Input : User's query and a data warehouse

Output : Decision rules

Step 1 : Extract table names Ti and attribute names Aj from the query in DW

Step 2 : Process the attribute ranking (AR) table and revise the hit counter recognized by each Ti and Aj

Step 3 : Descending arrange the AR-table based on the hit value

Step 4 : Extract the top row of AR-table to achieve T1 and A1

Step 5 : Create a decision table A = <U, A, d> where d = A1, A = a set of attributes in T1,

U = a set of records in T1

Step 6 : Pre-process A by

- eliminating attributes with number of discrete values = | T1|
- discrediting attributes with genuine values

Step 7 : Partition U into similarity classes

Step 8 : Generate certain, negative, possible rules

Step 11 : Generalize all decision rules using dimension tables and hierarchical information from the data warehouse

Step 12 : Include rules into the knowledge base

Decision rules are normally used in classification and prediction of data obtained in the data mart. It is easy yet a dominant way of data representation. The models formed by decision rules are characterized and identified the possible rules of the data referred by the users' queries. Induction based decision rule algorithm is used to select the hit attribute at each attribute in the data storage repositories. The attribute with the highest hit is selected as the similarity classes for the data warehouse. This attribute reduces the information and provide the information needed by the user and categorize the attributes in the resulting partitions.

4. EXPERIMENTAL EVALUATION

The proposed extracting user demanded information is efficiently analyzed for data mart based on inductive rule mining and is implemented using Java. An experimental study is done to estimate the effectiveness and performance of the proposed extracting user demanded information for data mart using inductive rule mining. The effectiveness of the proposed extracting user demanded information for data mart using inductive rule mining is estimated on Insurance company benchmark datasets from UCI repository with varying characteristics. Information about customers consists of 86 variables and includes product usage data and socio-demographic data derived from zip area codes. But in this work, we haven ta[illegible]

10 attributes for evaluating the performance of the proposed IRM for extracting user demanded information from repositories. The attributes used here are name, age, education, income, marital status, number of houses, number of children, number of insurance policies, number of motorcycle / car policies, number of life insurances.

Based on the formation of inductive rules, the user demanded information is extracted from the data repositories stored in the data mart. The inductive rules have been formed based on the decision rules and the decision is done efficiently by choosing the attributes related with operational goal. The performance of the proposed extracting user demanded information for data mart using inductive rule mining is measured in terms of

i) Number of decision rules,

ii) Extracted data relativity,

iii) Reliability

Decision rules are a set of funcitons which draws an examination to a proper action. The decision rules plays an important role in data warehousing concepts which determine the set of rules for extracting the user demanded information from the data storage repositories. After the formation of decisive rules, the user demanded data are retrieved from the repositories and the retrieved data relativity is found out using the decisive rule induction algorithm process is termed as **extracted data relativity** process. **Reliability** is the term used to identify the reliable manner of the extraction of data retrieved from the data storage repositories.

5. RESULTS AND DISCUSSION

When compared to an existing functional behavior pattern for data mart based on attribute relativity, the proposed extracting user demanded information for data mart using inductive rule mining is effective in terms of number of decision rules, extracted data relativity and reliability of the data mart by maintaining the decision rules in the data mart in a successful manner for extracting the information. Since we are using inductive rule for an extraction of user demanded data, the functional behavioral pattern of the data mart/organization is efficiently analyzed and viewed. The experiments are conducted with Insurance company benchmark data sets to estimate the performance of the proposed extracting user demanded information for data mart using inductive rule mining. The below table and graph describes the performance of the proposed extracting user demanded information for data mart using inductive rule mining with an existing scheme.

No. of users' queries	Number of decisive rules	
	Proposed IRM	Existing FBP for data mart
5	2	6
10	7	10
15	11	14
20	13	18
25	15	22

Table 5.1 no. of users' queries vs. no. of decisive rules

Table 5.1 described process of decisive rules formed in the data mart for analyzing the functional behavior of user needed information. The outcome of the proposed IRM for extracting user demanded data is compared with an existing Functional behavioral pattern for data mart.

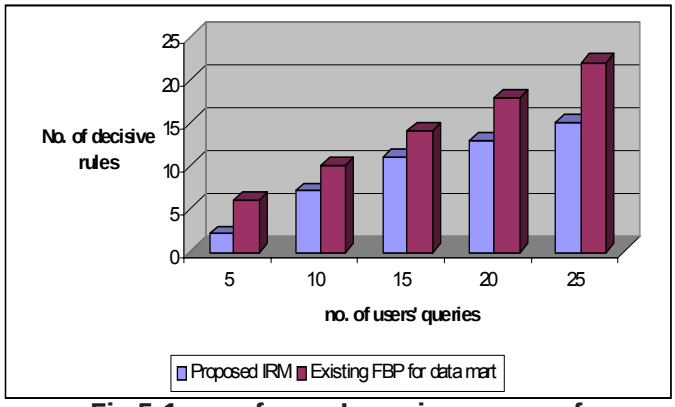

Fig 5.1 no. of users' queries vs. no. of decisive rules

Fig 5.1 describes the decisive rules formed in the data mart for analyzing the functional behavior of user needed information. The proposed Inductive rule mining used for extracting the user demanded information by deriving some rules which has been used for further experimentation. Existing Functional Behavior pattern for data mart only identified the function of organization based on the operational goal of the data mart. In the proposed IRM for data mart efficiently identified the operational goal of the data mart by deriving the inductive rules. Based on the number of users' queries, the number of decisive rules formed in the data mart is low in the proposed IRM for data mart contrast to an existing FBP in DW. Lower the decisive rules, higher the relativity of data needed by the user.

No. of decisive rules	Extracted data relativity (%)	
	Proposed IRM	Existing FBP for data mart
5	5	2
10	9	7
15	13	10
20	17	12
25	20	15

Table 5.2 no. of decisive rules vs. extracted data

Table 5.2 described the extracted data relativity of functional data attributes requested by the user present in the data mart for analyzing the functional behavior. The outcome of the proposed IRM for extracting user demanded data is compared with an existing Functional behavioral pattern for data mart.

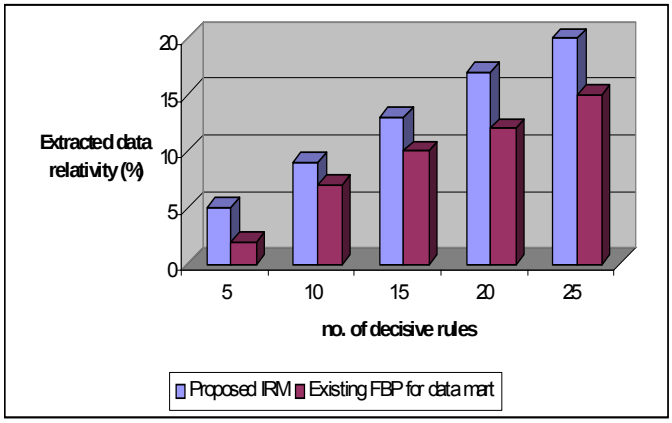

Fig 5.2 no. of decisive rules vs. extracted data relativity

Fig 5.2 describes the extracted data relativity of functional data attributes requested by the user present in the data mart for analyzing the functional behavior. The proposed Inductive rule mining chosen the relevant attributes of the data mart and derives the rules and those rules are analyzed efficiently based on the relevant attributes of the data repositories. Existing FBP for data mart only identified the process of data mart but the proposed IRM for data mart efficiently extracted user demanded information based on the number of rules formed which are closely related with the operational goal of the data mart. Based on the number of decisive rules formed by decision inductive rule algorithm, the extracted data relativity present in the data mart is high in the proposed IRM for data mart contrast to an existing Functional behavioral pattern for data mart.

No. of users	Reliability	
	Proposed IRM	Existing FBP for data mart
10	8	4
20	18	10
30	24	14
40	28	18
50	34	24

Table 5.3 no. of users vs. reliability

Table 5.3 described the reliability of IRM of the data mart and the outcome of the proposed IRM for extracting user demanded data is compared with an existing Functional behavioral pattern for data mart.

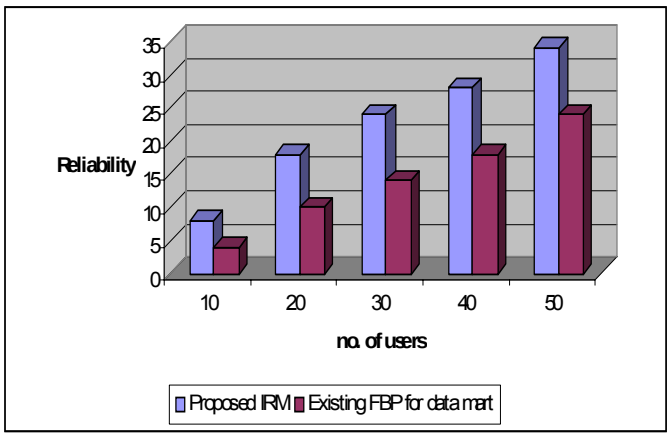

Fig 5.3 no. of users vs. reliability

Fig 5.3 describes the reliability of analysis of user demanded information of the data mart. The proposed IRM chosen the data which are necessary for the user requested process based on the operational goal of the data mart. Since the number of decisive rules formation is low, the user requested information in the data mart is also being accurate. Even though the number of users increases, the reliability of the data mart is high in the proposed IRM for extracting the user needed information in data mart contrast to an existing FBP in DW.

Finally, it is being observed that the proposed work efficiently used the inductive rule mining process for extracting the user requested information from the data storage repositories and the experiments are also efficiently done with the benchmark datasets and formed the decisive rules based on the algorithm and the demanded information done by the user is being processed in a reliable manner.

6. CONCLUSION

In existing functional behavioral pattern in data mart, the analysis of functional behavior only be analyzed and processed. The issues raised over existing Functional behavioral pattern are that it does not efficiently retrieve the user demanded information based on attribute relativity. The proposed Work used inductive rule mining for handling the user demanded information, at first, the relevant attributes of the particular data mart is used for the operational goal of the data mart. The decision support is efficiently done with inductive rule mining on functional data marts segregates of the layered data repository. The decision support system is achived in a reliable manner to extract the required user demanded data from data warehouse. An efficiently developed inductive rule mining for data mart extracted the user demanded information and the experimental results showed that the proposed Inductive rule mining for data mart analyzed efficiently and performance is measured in terms of decisive rule formation and the reliability contrast to an existing FBP for data mart based on attribute relativity.

REFERENCES

[1] Shehzad, K., "EDISC: A Class-tailored Discretization Technique for Rule-based Classification", IEEE Transactions on Knowledge and Data Engineering, 2011

[2] Alippi, C. , Braione, P. , "Classification methods and inductive learning rules: what we may learn from theory",

IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, sept 2006

[3] Cohagan, C. et. Al., "Mining Inconsistent Data with the Bagged MLEM2 Rule InductionAlgorithm", IEEE International Conference on Granular Computing (GrC), 2010

[4] Jerzy W. Grzymala-Busse Sachin Siddhaye, "Rough Set Approaches to Rule Induction from Incomplete Data", Proceedings of the IPMU'2004, the 10th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Perugia, Italy, July 4–9, 2004, vol. 2, 923–930

[5] Nittaya Kerdprasop, Narin Muenrat et. Al., "Decision Rule Induction in a Learning Content

Management System", International Journal of Computer and Information Engineering 2:2 2008

[6] Poongothai, K. and S. Sathiyabama, "Efficient Web Usage Miner Using Decisive Induction Rules", Journal of Computer Science 8 (6): 835-840, 2012, ISSN 1549-3636

[7] Bouchon-Meunier, B. , Marsala, C., "Fuzzy inductive learning: Principles and applications in data mining", 3rd International Conference on Intelligent System and Knowledge Engineering, 2008. ISKE 2008.

[8] Desmarais, M.C. , " Web log session analyzer: integrating parsing and logic programming into a data mart architecture", The 2005 IEEE/WIC/ACM International Conference on Web Intelligence, 2005. Proceedings

[9] Cabibbo, L. et. Al., "DaWall: a Tool for the Integration of Autonomous Data Marts", Proceedings of the 22nd International Conference on Data Engineering, 2006. ICDE '06.

[10] Y. Y. Yao. On the generalizing rough set theory. Proc. of the 9th Int.Conference on Rough Sets, Fuzzy Sets, Data Mining and Granular Computing (RSFDGrC'2003), Chongqing, China, Oct. 19–22, 2003, 44–51.

[11] Xinrui Liu et. Al., "The research of building production-oriented data mart for mine enterprises based on data mining", Sixth International Conference on Natural Computation (ICNC), 2010

[12] Polyzotis, N.; Skiadopoulos, S et. Al., 'Meshing Streaming Updates with Persistent Data in an Active Data Warehouse', IEEE Transactions on Knowledge and Data Engineering, 2008

ABOUT THE AUTHOR

PAULRAJ M received a Master Degree from Madurai Kamaraj University, India, Post Graduate Diploma in Business Administration from Annamalai University, India and Master of Computer Application (M.C.A) degree from Indira Gandhi National Open University, New Delhi, India, in 2002. He joined the Computer Centre , Department of Statistics, Ministry of Planning, New Delhi, and He joined the National Crime Records Bureau, Ministry of Home Affairs, New Delhi, where he was a Junior Staff Officer (System Analyst). He is currently Senior Systems Manager of Department of Information Technology at Canara Bank, Head Office, Bangalore, India, his job interests include introducing new technology and projects related to financial systems & analysis. He currently preparing a Ph.D in Computer Science, Research Scholar, thesis titled "Performance Analysis for Data Mart" in a reputed University at Tamil Nadu, India.

CO-AUTHOR

Dr. SIVAPRAKASAM P, M.Sc., M.Phil., Ph.D is associate professor, Department of Computer Science, at Sri Vasavi College, Erode, Tamil Nadu, India. He obtained Ph.D. from Periyar University, Salem, Tamil Nadu, India in Analysis of Web Caching Architecture.

***