

Design of Flexible Mining Language on Educational Analytical DataBase

Shirin Mirabedini

Department of Computer Engineering, Payame Noor University, Tehran, Iran

Corresponding author: Shirin Mirabedini

ABSTRACT: Enormous volume of data in the database is related to storage training centers. Routine use of these data, operation reports for educators and students. But today's sophisticated use of the massive volume of data stored in these warehouses can be performed to explore the use of the data. Currently, the most important analytical processing technology for effective utilization of data and its importance increases day by day, its scope encompasses a wide range of e-commerce, business, and industry and education systems. Among the factors affecting analytical processing techniques, the use of language is describe the process of exploring the possibility of analytical database and the data mining process, providing guidance to meet the objectives of the educational centers. An analytical sample database based on the language of instruction is designed to be flexible in exploring the possibilities this technology by user, there will also facilitate the necessary changes in order to explore the various methods for more flexibility in achieving the goals of the user. This article explores the language of instruction in analytical databases are described.

Keywords: Analytical Process Educational Analytic Database, Mining Query Language, Business Intelligence.

INTRODUCTION

Recently, much attention has been using business intelligence tools in teaching and learning and many researchers have studied how to apply data mining techniques, warehouse and online analytical processing on such systems. The tools specifically developed to support business intelligence in learning environments of these devices can be noted that the GISMO, CourseVis, Sinergo / CoIAT and MEATEP. Most of these tools address the students' data collection systems such as education systems, learning management systems, content management systems and etc. and integrated them into the desired form will be provided to teachers, principals and other administrators learning centers. Some of the information contained in data warehouse and integrated with the application of data mining techniques, automatically or semi-automatically provide suggestions to improve teaching and learning for teachers and students. The main difference between conventional reporting tools that provide learning management systems is that these systems in most data analysis and dynamic taking the time to being presented to reporting. Data cubes and online analytic processing tools also make it possible for managers to provide key performance indicators from different perspectives and at different levels of detail, are observed. The data warehouse also provides the possibility to analytical data warehouses and operational data are independent of each shelf, and access to them should be independent of the operating database. Figure 1 shows a general view of a business intelligence strategy.

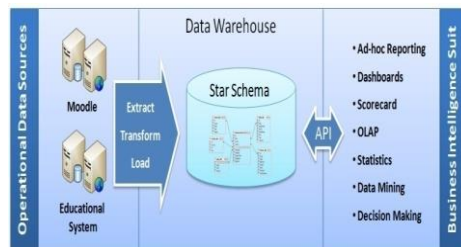


Figure 1. Overview of a Business Intelligence Strategy

As shown in the figure, data from different sources, such as learning management systems, course management systems and the process of extraction, transformation and loading is transferred into the warehouse. The warehouse is in the form of a subject-oriented structure using star schema is designed to be used in data integration and provides analytical processing tools. As is clear from the form of tools, analytical processing is data mining. Data mining is the process of analyzing training is designed to explore the training data, this process would be to look for consistent patterns or systematic relationships between variables of education, then confirm or prove the validity of these findings is done using authentic patterns. Extracting useful information from large training data, and convert them to the required knowledge from universities and educational institutions, especially in managerial decision making requires the use of new techniques in the field of Business Intelligence and Data Analysis. One of the tools and attitude data mining training in data processing environment of universities and educational institutions, recommendations automatically and semi-automatically improve teaching and learning provide for teachers and students. Given the importance of data mining, training, planning data mining language training is an important element in the creation of analytical data has special importance. Language provides data mining; on the other hand, using data mining system, educational facilities, and appropriate response to the demands of the analytical data are derived from the mass of available data. It should be noted how the structure of operation and ease of use educational data mining techniques, the main characteristics of this language are to be considered as a measure of comparison between these languages. Exploring language UnQL, WebSQL, WebOQL including data mining language is presented in this paper will be discussed. Also, a flexible language question in the analysis and search algorithms are introduced that use the grammar of the language, members of an educational system capable of extracting patterns in enterprise data are dense, so that optimal search algorithms should be chosen. Visitors will be able to use this system, thus each operation classification, clustering, and derive the dependence of the parameters that should be obtained from the generalized method. Other features of the language that is based on SQL, so the ease of learning and ease of using it inherits. Implementor of the language system with graphical user interfaces, it is able to communicate and interact with users, query they received and interpreted by the interpreter and the system implementation.

Analytical database and data analysis in the field of Business Intelligence:

In recent years, the data management of assets in competitive markets the world has become. In the business area, not far when companies, organizations and educational institutions in order to access information and perform specific operations, automated management processes (invoice management and supply chain management and enterprise resource planning, etc.) used return, but today these centers, organizations and companies to integrate and control all the information they need in order to analyze their business goals achieved in each stage. To assist organizations with the purpose of analyzing databases and analytical processes were introduced, the system of online analytical processing (OLAP) systems have evolved into an online transaction processing (OLTP) , which are in line with the rapid processing of multidimensional data analysis are employed. Using the reporting system of education, student enrollment, finances and the like acts are much faster than systems using multi-dimensional data OLTP. Online processing systems using a hierarchical structure and a powerful multi-dimensional data model provides for organizing information, and reports to simplify the calculations very quickly. The data warehouse systems are associated with online processing of multi-dimensional data cubes that provides an opportunity for rapid analysis of data from different databases. Recently, the use of databases and analytical tools are also a great boom in the education systems, many researchers have focused on the use of analytical tools to process the warehouse and in the educational systems.

Language of exploring databases and analytical training:

To achieve the goal of data analysis in the field of education-related technologies, including methods for better indexing, using filtering agents to gather specific information and recommendations are made to improve the

efficiency of search that have been presented in the form of high-level query language called the language with the goal of knowledge extraction. Query languages are categorized according to the type of database or data warehouse environment, the following requirements.

➤ **Defined on non-structured query language data**

Defined on non-structured data query languages are languages that are defined on different types of non-structured data such as HTML documentation and training content. Exploring language UnQL, WebSQL, WebOQL are placed in this category.

❖ **Query language defined on structured data**

These types of data mining query language learning and knowledge extraction are to perform search operations on databases. Here's a broader perspective, we examined the DMQL data mining language. This language education data mining capabilities in relational databases have been on several levels, flexible and interactive user interface. The language in question is intended as a summary of some data mining, data mining of association rules, classification, clustering and finding specific patterns. The language of the user interface, flexible, capable and multi-level debugging and searching facilities such as tables, charts, manuals and dynamic selection puts at their disposal. This language research team from Simon Fraser University in Canada was designed by Dr. Hahn. This language will be implemented under DBMiner system capable of efficient extraction of a variety of laws and knowledge to effectively and efficiently be available. Other features include integrated security systems; graphical user interface is very powerful and relevant tool to represent knowledge, power mining classification rules, association, and cluster, high efficiency and the powerful of responses. DMQL questions derived from the SQL language, training data can be used to analyze it. Types of knowledge that can be extracted by the language of the classification rules, clustering analysis method is dependent on the exchange of information and technology transfer is possible by language. DMQL question the decision tree classification language is implemented in the model tree, each node represents a feature value and each split test, display test output classes are the leaves of the tree. DMQL questions derived from the SQL language, it can be used to analyze training data. Types of knowledge that can be extracted by the language of the classification rules, clustering analysis method is dependent on the exchange of information and technology transfer is facilitated by the language. DMQL question the decision tree classification language is implemented in the model tree, each node represents a feature value and each split test, display test output classes are the leaves of the tree.

Weaknesses in data mining language teaching:

Languages, data mining exploration in general are at least three types of classification rules, association, clustering them. Data mining systems also search these laws do operate in different ways and with different algorithms and methods. For example, searching through the D-Tree classification rules or D - Rule is possible that each input parameter variation leads to the creation of different algorithms, such as ID3, C4. 5 and so on. The other category of association rules mining is generally inductive algorithm is derived. The algorithm also makes a variety of different input parameters S and C and various methods have been established. Clustering algorithms are also using similar methods such as K-means and that too with a variety of parameters K and production of various computing cores median and Primary Production of the algorithm is different. It should be noted that the language in question, concrete and specific use of each algorithm for exploring a variety of laws. In this respect, the user's language preferences are applied to determine the feasibility of the desired algorithm. To overcome these shortcomings, a question of language requires that these algorithms to be generalized to the users and even the addition of defined methods, the user can enable the introduction of new algorithms to search a particular type of system rules. The user can thus their algorithm to a high-level language systems and algorithms by introducing a specific name registered, they can be used in subsequent searches.

Design flexibility in exploring the question of training data:

The proposed language, language is a question that has been designed for interactive learning and flexible data mining. The language must be one of the techniques explored all search terms such as classification, clustering, association analysis supports the language commonly used data mining. On the other hand, it requires flexibility, which makes feature selection techniques provide different rules for data mining or exploring ways to their users. Thus, users of these languages well able for example to explore the classification rules and decision tree method with feature selection formula parameters and the final condition of a certain, extract association rules with inductive algorithms, with a coefficient of reliability to support its case and the clustering analysis method and the parameters K and K_Means to do a certain formula to calculate the distance. Also, the flexibility of this language, users of the

new methods introduced through a high level language, the proposed method can capture the unique name and use it in their future explorations. Designing efficient exploration of the language analysis requires a flexible and deep understanding of the strengths and weaknesses of the various mechanisms and techniques explored. The design of this language is intended to represent a query, the following components:

1. Type of data search: The first step in data analysis exploring the definition of education, specifying the data to be searched is explored. In data mining, identifying the required attributes in the database and the data is necessary to define certain dimensions. In the language we have used to facilitate the integration of simple SQL query.
2. Type of technique to search: Explore the processes of teaching analytical data, including classification rules, association and clustering, and so on. The proposed language allows exploring a variety of both classification and association rules provide for the clustering analysis is also for the classification of anyone to hold.
3. Method needed to search: After determining the type of knowledge, the flexibility of the proposed language, methods and techniques to identify user requirements is mandatory. The rest of the query language should be introduced search algorithms.
4. Parameters and thresholds: Analysis for each algorithm, a set of parameters is defined as the input parameters, determining various techniques in educational data mining methods.
5. Introducing new techniques to explore: If it is not already defined data mining techniques due to the flexibility of the language features proposed for space exploration is spent on introducing new techniques to high-level language.
6. View patterns extracted for presentation: The inclusion of queries and provide query results, we have used a graphical user interface.

Proposed language rules:

Words of this language consist of two parts: a part of the same SQL queries and other departments between high-level language programming language and spoken language is a restricted form of pseudo code is defined by the user. The first words of the language are:

Having, Cluster SQL PART = { Select , From , Where , Group By,By, Using, Extract Rules, Mine, Association Rule, Classification Rule, Clustering, Generate, Use Database ,Order By , With, For , As, > , < , = ,*, ..}

The second part of the vocabulary of the language is somewhat similar to the VB language is as follows: Algorithm PART= { Initialize, Partition, with, Begin / end, call, If / then / else/ endif, Repeat / Until, There, Exist, Not Exist, AND, OR, NOT, forall, Function, Isnull, Message, Inset, Mode, Change, Subset, Input, Output, Max, Min, Gain, Return, ", + , _ , / , * , ^ , log , exp , (,) , & , ...}

In this section, the proposed language rules of grammar in BNF, we have defined as follows:

```

<FQL >:: = < FQL _ Statement > ; {< FQL _ Statement >}
< FQL _ Statement >:: =
< Flex _ Data _ Mining _ Statement >
< Flex _ Data _ Mining _ Statement >:: =
    < Open _ DB _ Statement >
    < FIND >
    < Kind _ Of _ Pattern >
    < Limited _ SQL _ Query >
    < Kind _ Of _ Algorithm >
< Open _ DB _ Statement >:: =
    USE DATABASE < DataBase _ Name >

< Kind _ Of _ Pattern >:: =
    < Classification _ Rules > | < Association _ Rules >
                                | < Clustering >

< Classification _ Rules >:: =
    CLSSSIFICATION RULES AS < Output _ Name >
    CLASSIFY BY < Field _ Name >:
                                {< Field _ Name >}
    
```

```

< Association _ Rules >:: =
    ASSOCIATION RULE AS
                                < Output _ Name >
< Clustering >:: =
    CLUSTERING AS < Output _ Name >
< Kind _ Of _ Algorithm >:: =
    < Classification _ Rule _ Algorithm > |
    < Association _ Rule _ Algorithm > |
    < Clustering _ Algorithm > |
< Classification _ Rule _ Algorithm >:: =
    GENERATE < Pattern _ Name > USING
                                <Class_ Algorithm _ Name >
    <Class_ Parameters _ Statement >
< Class_Parameters _ Statement >:: =
    ATTRIBUTE SELECTION = < Att _ Sel _ Value >
    TERMINATION CRITERIA = < Ter _ Criteria _ Value >
< Association _ Rule _ Algorithm >:: =
    EXTRACT RULES USING
                                <Ass_ algorithm _ Name >
    WITH <Ass_ Parameters _ Statement >
< Ass_Parameters _ Statement >:: =
    SUPPORT = < Sup _ Value >
    CONFIDENCE = < Conf _ Value >
< Clustering _ Algorithm >:: =
    CLUSTER BY < Field _ Name > USING
                                < Clust_Algorithm _ Name >
    WITH <Clust_Parameters _ statement >
< Clust_Parameters _ Statement >:: =
    NO_CLUSTERS = <Number_of_Clusters>
    DISTANCE= < Top _ Limited _ Distance _ Value >

```

Based on the proposed language BNF syntax, grammar rules of the class is defined as follows.

Grammar extraction of classification rules:

```

USE DATABASE dbi
FIND CLASSIFICATION RULE AS <output>

```

```

SELECT  fi, fj, ....
FROM    tbi, ....
WHERE   Conditions, ....
GROUP BY Field name
GENERATE < D – tree / D – Rule > USING Methodi
ATTRIBUTE SELECTION = < Att _ Sel _ Value >
TERMINATION CRITERIA = < Ter _ Criteria _ Value >

```

To clarify the capabilities of grammar, a part of an organization with database schema Consider the following:

```

Student (Sno, Sname, Status, ID_no , Major, DNo, Birth-date, Address, Avg )
Course ( C_no, Title, DNo , Desc )
Department ( DNo, DTitle , Head_name , Place)

```

The following example query is to explore the classification rules:

```

USE DATABASE University
FIND CLASSIFICATION RULE AS dep-view
CLASSIFY BY Status
SELECT dno , Pno, Pname , address, Avg

```

```
FROM Student
WHERE DTitle='Computer' AND
Major = "IT " AND Status = "1 "
GENERATE D-Tree USING ID3
ATTRIBUTE SELECTION =...
TERMINATION CRITERIA =...
```

An example database open university students in the Department of Computer and Information Technology are studying in the field, based on the situation of indigenous and non-indigenous students, students' academic achievement will be classified if the output displays the ID3 decision tree algorithm using two parameters and conditions specified in the condition of the feature selection algorithms

CONCLUSION

As mentioned, the method of data analysis training is associated with various scientific domains such as online analytical processing, data mining, machine learning, psychometric and statistical techniques as well as other areas of image information and computer modeling. In this paper we explore the process of implementation in order to analyze training data, language design question; we do the act of exploring flexibility. Flexibility in the choice of algorithm, a solution is to deal with the problem of limiting users' data mining language training to a predetermined set of algorithms which try to overcome this problem have been proposed language. As mentioned, all existing query languages such as DMQL the possibility of exploring a variety of laws, including the rules of classification, clustering, association analysis can provide to the users, but users of these languages do not have any choice in the search algorithm. If they are not willing to approach the algorithm identified and fixed in the system of their choice will have no choice but to change the system language and its implementation, as in any language specific algorithms is intended to explore. Thus, considering the scale of this problem, we proposed a language designed with the aim of flexibility in our search algorithm, the new algorithm, the user can, where even the language itself is system introduced tremendous. It should be noted that changing the data type from relational to object-oriented syntax, the language will not change, it should only be turned commentator, said that another reason is the high-level language. Because of flexibility, offering a new algorithm in the field of data mining system developed is capable of, so that it can either provide a new algorithm to their users. Thus, using the data mining language training, a step toward expanding the scope of application of data mining have been taken without the need to change tools. By reading the letters and papers of many languages worldwide active analytical database, comparison criteria, such as extensibility, simplicity and ease of use, the type of data that is specific language to be able to work with them, these types of laws. Search is based on SQL and flexibility, including cases in which the efficacy query languages are used. The proposed grammar in addition to the general operations of the query language encompasses a unique flexibility and extensibility and thus has overcome some of the shortcomings existing in other languages. It is possible to define new methods for data mining with its extensive facilities in high-level training to the users that this issue of language training in the production of a data base of analytical data will make important analytical.

REFERENCES

- Berkhin P. 2002. "Survey of Clustering Data Mining Techniques", Accrue Software, CA. 2002.
- Boyer J and Frank B. 2010. "Business Intelligence Strategy; A Practical Guide for Achieving BI Excellence", 2010.
- Espejo PG, Romero C, Ventura S and Hervás C. 2005. "Induction of Classification Rules with Grammar-Based Genetic Programming". Conference on Machine Intelligence, 2005.
- Han J, Chiang Y and Chee S. 1997. "DBMiner: A System for Data Mining in Relational Databases and Data Warehouses", 1997.
- Han J and Zaiane O. 1996. "DMQL: A Data Mining Query Language for Relational Databases. Simon Fraser University", B.C. Canada, 1996.
- Malerba D, Appice A and Vacca N. 2001. "SDMOQL: An OQL-based Data Mining Query Language for Map Interpretation Tasks", Orabona 4, 70125 Bari, Italy, 2001.
- Menke R. 2002. "Bear Test Analyzer: An experiment in Bayesian Classification of text", 2002.
- Mostow J, Beck J, Cen H, Cuneo A, Gouvea E and Heiner C. 2005. "An Educational Data Mining Tool to Browse Tutor-Student Interactions: Time Will Tell! In: Proc. of Workshop on educational data mining", pp. 15-22 (2005).
- Zorrilla ME. 2009. "Data Warehouse Technology for E-Learning", Springer-Verlag Berlin Heidelberg 2009.