International Journal of
Emerging Research in Management &Technology
ISSN: 2278-9359 (Volume-4, Issue-3)

Research Article | **March 2015**

An Analytical Study of Data Mining with the Help of UML Model and its Application

Dr. P. C. Agrawal
Visiting Professor, Mewar University
Gangrar, Chittorgarh, Rajasthan, India

Aparna Chaturvedi
Research Scholar, Department of Computer Science
Mewar University, Gangrar, Chittorgarh, Rajasthan, India

Abstract-

Data Mining analytical study goes for modeling and database Design methods. In this Research Paper we discuss Knowledge discovery methods and modeling of data mining tables. When we create models in data mining for UML Models and its application we use OLAP Tool. So An Analytical Study of Data Mining with the help of UML Model and its application takes place.

Keywords--- Knowledge Discovery, Data Mining Models, Phases, Olap Tool, Making Diagrams.

I. INTRODUCTION

Generally, data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users to analyze data from many different dimensions or angles, categorize it, and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens of fields in large relational databases. Although data mining is a relatively new term, the technology is not. Companies have used powerful computers to sift through volumes of supermarket scanner data and analyze market research reports for years. However, continuous innovations in computer processing power, disk storage, and statistical software are dramatically increasing the accuracy of analysis while driving down the cost.

II. KNOWLEDGE DISCOVERY

Business understanding

This phase focuses on understanding the project objectives and requirements from a business perspective, then converting this knowledge into a DM problem definition and a preliminary plan designed to achieve the objectives.

Data understanding

The data understanding phase starts with an initial data collection and proceeds with activities in order to ***g***et familiar with the data, to identify data quality problems, to discover first insights into the data or to detect interesting subsets to form hypotheses for hidden information.

Data preparation

We propose the use of Data Mining *data* model to document the source data of the project. We could use Data Mining data model in each subtask of Data understanding or Data preparation tasks to model the transformations in the data.

III. MODELING

Build model

This subtask could be document through Data Mining algorithm and *Data Mining model* models. Data Mining technique model should be used as input for developing the Data Mining algorithm model.

We are going to define the DM-UML data mining models: data mining use case model, data mining goal model, data mining data model, data mining technique model, data mining algorithm model, data mining models model, as well as the elements required to be able to specify the data mining problem to be addressed.

Data mining use case model

We use dependencies (dashed directed lines) to relate each of the elements appearing in these diagrams, as a change in the definition of any of the model elements will be propagated to the elements to which it is related. The data mining use case is the foundation for the DM-UML models. The data mining use case describes the proposed functionality of the knowledge extracted by the data mining tasks as seen by the user in order to achieve a particular business goal. A data mining use case represents a discrete unit of interaction between a user and the knowledge. Different data mining use cases applied to the same information represent different ways of using or extracting knowledge for different or the same business goals. A data mining use case is a single unit of meaningful work; for example, rank products based on their

properties or sales, or develop a customer profile are both data mining use cases. Each data mining use case has a description that describes its functionality.

Data mining goal model
The data mining goal model represents the data mining project requirements, that is, what is expected to be gained from the data mining project in terms of knowledge rather than business. The data mining use case model explains how these data mining goals are used by actors data mining goals, depicting the data mining goal generalization hierarchies, if any, and their relationship to the use cases from which they derive. Each data mining use case has to have at least one associated data mining goal, and several data mining use cases can tackle one data mining objective.

Data mining data model
The data mining data model represents the sources of the available data for the project, with tables,
columns, data types and data relations. This model is based directly on the UML definition for data models, but has been adapted by means of stereotypes to tailor the data models to the needs of data mining projects (data integration, transformation and derivation). This model represents the physical data model, that is, the structures to be stored in the data source.

Data mining technique model
Data mining technique diagrams show the data mining techniques used to be able to achieve the data mining goals proposed in the data mining use cases. These diagrams show the data mining techniques and their possible input data related to the data mining goals that they achieve or help to achieve. They may also show the data sources that they use. Data mining technique is to be applied to output a particular business goal.

Data mining algorithm model
The data mining algorithm model shows the data mining algorithms to be used to solve the problem. The algorithms to be used can be implemented in the data mining tool that is used in the project or can be developed ad hoc. Apart from representing the algorithms, these diagrams may also include the available data sources from where the data are to be taken. As the algorithms are directly derived from the data mining techniques, the data, if included, must match the data that appear, if any, in the respective data mining techniques diagram. A data mining algorithm can also be represented by the input data it uses, the algorithm parameters and values, or both (data and parameters).

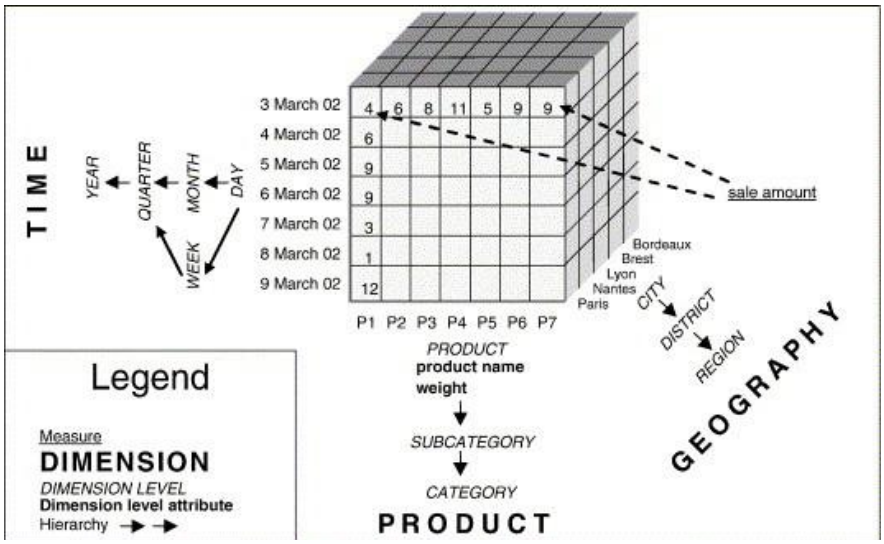
Data mining models model
This model shows which data mining models will be built and where they are stored (files) in the data mining tool used in the project. This should to enable traceability, first, among the data mining algorithms and, second, among the models and files where they are stored in the data mining tool (strictly speaking, model work spaces and files). Additionally, the definition of the files will later enable the definition of the configuration elements that will take part in project configuration management.

IV. OLAP TOOL

OLAP tools enable users to analyze multidimensional data interactively from multiple perspectives. OLAP consists of three basic analytical operations: consolidation (roll-up), drill-down, and slicing and dicing. Consolidation involves the aggregation of data that can be accumulated and computed in one or more dimensions. For example, all sales offices are rolled up to the sales department or sales division to anticipate sales trends. By contrast, the drill-down is a technique that allows users to navigate through the details. For instance, users can view the sales by individual products that make up a region's sales. Slicing and dicing is a feature whereby users can take out (slicing) a specific set of data of the OLAP cube and view (dicing) the slices from different viewpoints. Each *measure* can be thought of as having a set of *labels*, or meta-data associated with it. A *dimension* is what describes these *labels*; it provides information about the *measure*. A simple example would be a cube that contains a store's sales as a *measure*, and Date/Time as a *dimension*. Each Sale has a Date/Time *label* that describes more about that sale.

MOLAP---
This is the more traditional way of OLAP analysis. In MOLAP, data is stored in a multidimensional cube. The storage is not in the relational database, but in proprietary formats. Excellent performance: MOLAP
cubes are built for fast data retrieval, and are optimal for slicing and dicing operations can perform complex calculations: All calculations have been pre-generated when the cube is created. Hence, complex calculations are not only doable, but they return quickly.
MOLAP is a "multi-dimensional online analytical processing".'MOLAP' is the 'classic' form of OLAP and is sometimes referred to as just OLAP. MOLAP stores this data in an optimized multi-dimensional array storage, rather than in a relational database. Therefore it requires the pre-computation and storage of information in the cube - the operation known as processing. MOLAP tools generally utilize a pre-calculated data set referred to as a data cube. The data cube contains all the possible answers to a given range of questions. MOLAP tools have a very fast response time and the ability to quickly write back data into the data set.

In this type of OLAP, a cube is aggregated from the relational data source (data warehouse). When user generates a report request, the MOLAP tool can generate the create quickly because all data is already pre-aggregated within the cube.

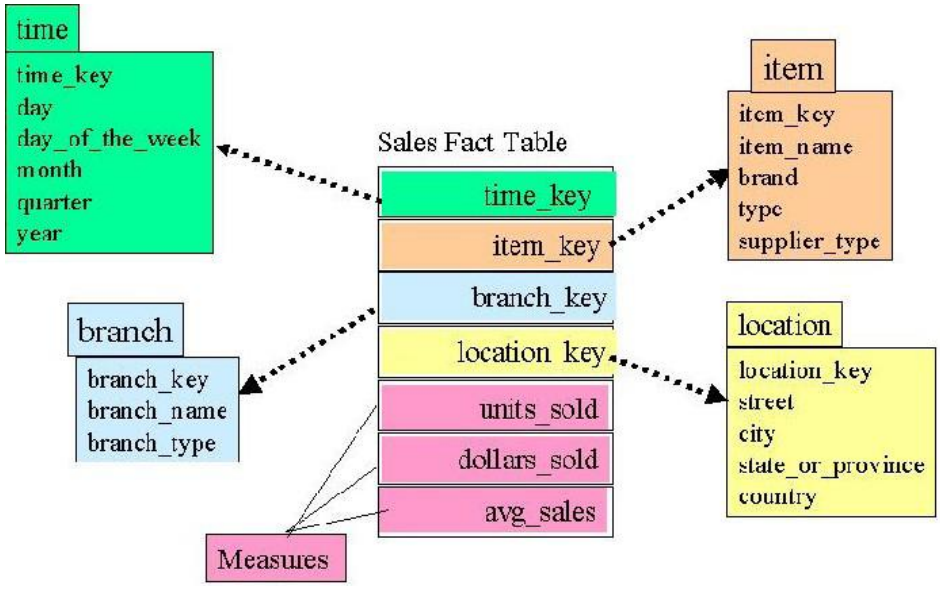


ROLAP---

This methodology relies on manipulating the data stored in the relational database to give the appearance of traditional OLAP's slicing and dicing functionality. In essence, each action of slicing and dicing is equivalent to adding a "WHERE" clause in the SQL statement can handle large amounts of data: The data size limitation of ROLAP technology is the limitation on data size of the underlying relational database.

In other words, ROLAP itself places no limitation on data amount can leverage functionalities inherent in the relational database: Often, relational database already comes with a host of functionalities. ROLAP technologies, since they sit on top of the relational database, can therefore leverage these functionalities**.**

In this type of OLAP, instead of pre-aggregating everything into a cube, the ROLAP engine essentially acts as a smart SQL generator. The ROLAP tool typically comes with a 'Designer' piece, where the data warehouse administrator can specify the relationship between the relational tables, as well as how dimensions, attributes, and hierarchies map to the underlying database tables.



HOLAP---

The HOLAP storage mode combines attributes of both MOLAP and ROLAP. HOLAP technologies attempt to combine the advantages of MOLAP and ROLAP. For summary-type information, HOLAP leveraged cube technology for faster performance. When detail information is needed, HOLAP can "drill through" from the cube into the underlying relational data. There is no clear agreement across the industry as to what constitutes "Hybrid OLAP", except that a database will divide data between relational and specialized storage. For example, for some vendors, a HOLAP database will use relational tables to hold the larger quantities of detailed data, and use specialized storage for at least some aspects of the smaller quantities of more-aggregate or less-detailed data. HOLAP addresses the shortcomings of MOLAP and ROLAP by combining the capabilities of both approaches. HOLAP tools can utilize both pre-calculated cubes and relational data sources.

Hybrid online analytical processing (HOLAP) is a combination of relational OLAP (ROLAP) and multidimensional OLAP (usually referred to simply as OLAP). HOLAP was developed to combine the greater data capacity of ROLAP with the superior processing capability of OLAP.

V. CONCLUSION

In Data Mining the Knowledge Discovery is the process of analyzing data from different perspectives and summarizing it into useful information - information that can be used to increase revenue, cuts costs, or both. Knowledge Discovery comes in many stages. Since UML Modeling is used to make many models so different type of Diagrams in CRISP DM-UML is drawn.

OLAP tool is used make all type of cubical diagrams that are generated by Data Mining tables and columns.In Analytical Study of Data Mining using UML Model and its application we make models of Data Mining based Projects.UML is used to make models in database design methods.

REFERENCE

[1] V. Saxena, D. Arora and S. Ahmad, "Object oriented distributed architecture system through UML", *IEEE International Conference on Advanced in computer vision and information Technology*, November 28-30, 2007.

[2] V. Saxena and G. A. Ansari "UML Modeling & Protection of Domain Based System", International Journal of Computer Science and Network Security (IJCSNS), Vol. 8, No. 7, pp.338-344, July 2008.

[3] V. Saxena, N.M.P. Verma and A. Pratap, "A Data Mining Technique for a Secure Electronic Payment Transaction," International Journal of Economics and Finance, Vol. 2(4), pp. 12-17, 2010.

[4] Óscar Marbán and Javier Segovia**, "**Extending UML for Modeling Data Mining Projects (DM-UML) International Journal of Information technology and Software Engineering Vol 3(2) pp.1-8,2013.

[5] Thomas H. Davenport,Jeanne G. Harris,David W. De Long and Alvin L. Jacobson, "Data to Knowledge to Result" CALIFORNIA MANAGEMENT REVIEW VOL 43, NO. 2 pp.117-136WINTER 2001.

[6] Bora Beran1, Catharine van Ingen1, Ilya Zaslavsky2 and David Valentine, " OLAP Cube Visualization of Environmental Data Catalogs",Microsoft Research, 2 San Diego Supercomputer Center Microsoft Technical Report MSR-TR-2008-70 July 14, 2008.

[7] François Pinet, André Miralles, Sandro Bimonte ,Françoise Vernier ,Nadia Carluer4, Véronique Gouy4 and Stephan Bernard, "The use of UML to design agricultural data warehouses**",** paper presented10/4 at AGENG 2020 Conference, Septembre 6-8 2010,

[8] J. Rumbaugh, I. Jacobson, and G. Booch., The Unified Modeling Language Reference Manual, 2nd ed. Addison-Wesley, Boston, 2005.

[9] Hasan H. "Using OLAP and Multidimensional data for Decision Making" IEEE International Conference on Advanced in computer vision and information Technology, Sep/oct 2001.

[10] S.Chaudhari and U.Dayal"An overview of Data Warehousing and OLAP Technology",ACM SIGMOD Record Volume 26 March 1997.

[11] P.K. Chaurasia and V. Saxena, "Mobile Based Electricity Bill Deposit System Through UML," Journal of Software Engineering and Applications, Vol. 4(3), 187-190, 2011.

[12] J. Han and M. Kamber, Data Mining: Concepts and Techniques, 2nd ed., Morgan Kaufmann Publishers, , March 2006.

