

Mining Techniques Used for Financial Organization

Sonika Jalhotra

Persuing Mtech From Subharti Institute of Technology & Engineering, Meerut

Abstract—With the increase of database applications in banking system, mining interesting information from huge database becomes of most concern. As we know, the data processed in data mining may be obtained from many sources in which different data types may be used. In this work we will concern the data types used in banks and various data mining methods. Our main focus is to make Mining effective by selection of an appropriate data mining method for different data types used in banking system. In this work four kinds of data forms for data mining and three kinds of data mining techniques are discussed. A Data type generalization process is proposed. Using the data type generalization process, the user can direct to the goal of application.

Keywords—Data mining, knowledge discovery database, data type forms, data mining technique.

I. INTRODUCTION

In recent years, Banks contain amount of historical data, transactional data and relational data. Automated data generation and gathering leads to tremendous amounts of data stored in bank's database but we have lack of ability to extract proper knowledge from the database. Data Mining is the automated discovery of non trivial, previously unknown, and potentially useful knowledge embedded in database. The types of data stored in database restrict the choice of data mining methods. In this paper we will retrieve knowledge from different type of data stored in database by using appropriate data mining techniques.

II. RELATED WORK

The fact is data is no use until and unless the knowledge is extracted out from it and data mining is the only way to discover hidden knowledge from massive amount of data. In this paper we proposed a concept of effective data mining by using appropriate data mining techniques on different types of data stored in database. Data Mining is the process of discovering new correlations, patterns, and trends by digging into large amounts of data stored in database(warehouse). In summary the whole process of data mining is known as KDD(Knowledge Discovery in Databases). KDD is a synonym of data mining which comprises of three stages.

- The understanding of business and data.
- Performing the pre-process tasks.
- Data mining and reporting.

The whole process of data mining is the so-called KDD as shown in figure 1.

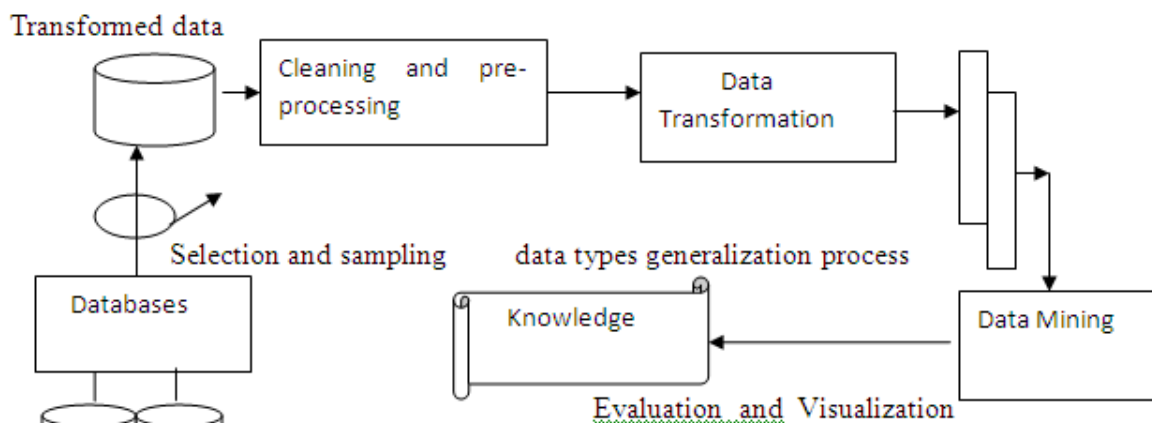


Figure 1: The KDD process and the role of data types generalization.

We solve data mining problem with data transformation. Data types transformation must based on the selected data mining method. The target of doing data transformation is to make the specified data set suitable for the organizations. The flow of solving data mining problem with doing data transformation is shown in Figure2

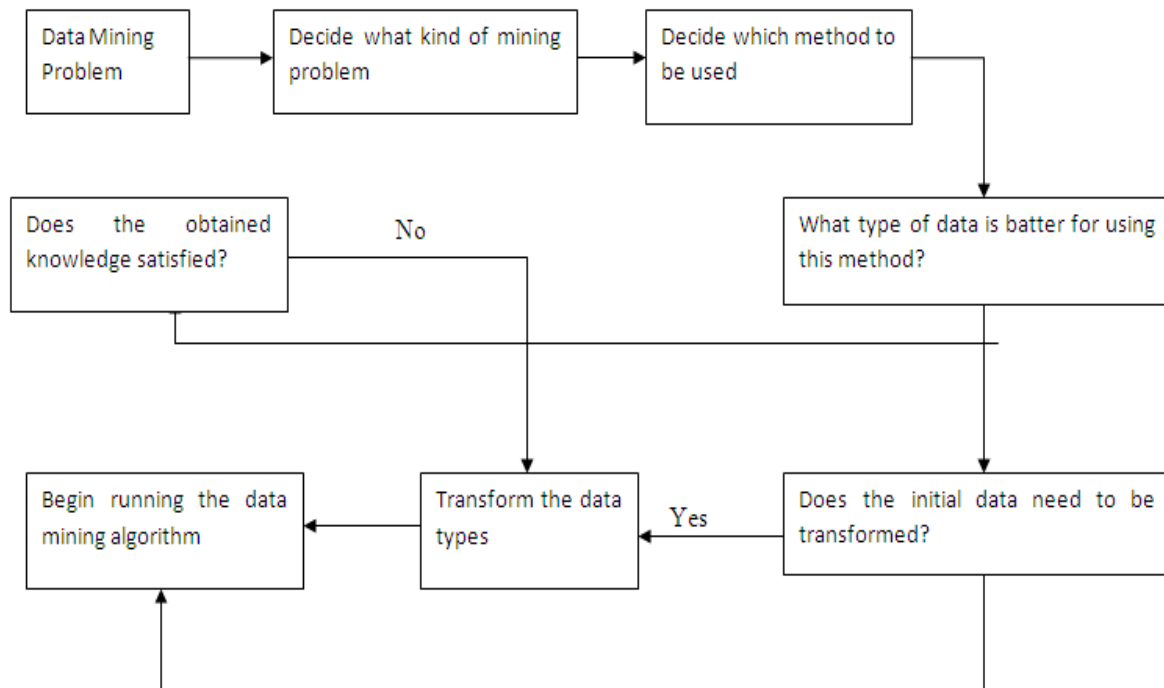


Figure 2: Solving data mining problems with data transformation data types transformation must based on the selected data mining method

III. ANALYSIS OF THE PROBLEM WITH DATA TYPES DURING THE MINING

In recent years due to the rapid growth of database applications data mining techniques become more important. So we introduced the concept of data mining methods but it is difficult to choose a data mining method to fetch data from the data warehouse without prior knowledge of data types stored in data warehouse. This depends on the characteristics of the data to be mined and the kind of knowledge to be found through the data mining process. Characteristics of the data focused on the relationship of data with other data stored in database and its attributes which contribute in mining process. In financial firms we deal with different kinds of data, each data mining technique is not suitable for the variety of data types stored in database of these firms. So this paper focus on this problem which can adversely affect on the productivity, performance of a financial firm if not considered properly. We now illustrate the analysis as follows.

Four kinds of data forms for data mining used in financial organizations

Banks or Financial firms generally deal with data forms in textual, temporal, transactional, relational forms. Different kinds of data forms are used to store different kinds of data types. We describe each kind of data forms in the following:

- (1) Textual data forms: Textual data forms are used to represent text or documents. This kind of data forms can be seen as a set of characters with huge amount.
- (2) Temporal data forms: Temporal data forms are used to represent historical data (data that varies with time).
- (3) Transactional data forms: Transactions done by user such as with draw of money, transfer of money from one account to other etc. can be stored in transactional data forms.
- (4) Relational data forms: Relational data forms are the most widely used data forms. The basic units of relational data forms are relations (tables). Relations are composed of attributes, and each of which can be different data type.

Three kinds of data mining techniques cover all the data types used in financial firms

A bank or financial firm can provide good and effective result if they use appropriate mining technique for the data stored in database. This provides quality in work and extraction of knowledge stored in database within time limit given to an organization. Effectiveness of data mining technique is based on the characteristics of the data to be mined and the knowledge user try to find out, which can be divided into three kinds of techniques for a financial firm. Each data mining technique is used for particular data type which description is given in following points:

- (1) Multilevel data generalization, summarization, and characterization :

Generalization is a process which abstract large set of relevant data in a database from a low conceptual level to relatively high ones.

For example Data and objects in databases often contain detailed information at primitive concept levels. For example, the customer relation in bank database may contain attributes describing low-level customer information such as customer name, account number, address, customer age, customer address etc. It is useful to be able to summarize a large set of data and present it at a high conceptual level. This is also shown in figure 3

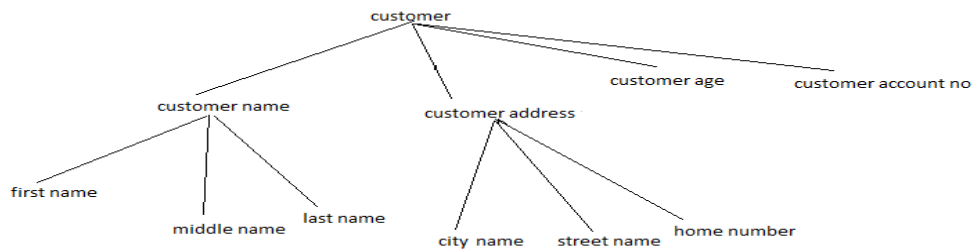


Figure 3 represents generalization of customer relation at conceptual level

Knowledge to be provided: This technique observes the data stored in database with a higher view. For example certain rules made by banks such as a time period to pay installment money for a policy or for opening new account user's id proof etc.

Appropriate data type: It is applied to relational databases which contain data in the form of tables.

(2) Mining association rules

In data mining, association rule learning is used to discover interesting relations between variables in large databases. For example if a customer want to take loan from a bank then he must also has an account on the branch of that bank (customer must has account number)

This association is represented as

$$\{\text{Customer, loan}\} \rightarrow \{\text{account number}\}$$

Association rule mining does not consider the order of attributes either within a transaction or across transaction.

Example database with 4 attributes and 4 transactions

Transaction id	Loan number	id	Time period to pay	Bank name
1	0	1	0	1
2	1	0	0	1
3	1	1	0	1
4	1	1	1	1

Problem of association rule mining is defined as : let $I = \{i_1, i_2, i_3, \dots, i_n\}$ be a set of n binary attributes used in association.

Let $D = \{t_1, t_2, t_3, \dots, t_n\}$ be a set of transactions. Each transaction in D contains a unique transaction ID and contains a subset of the attributes in I. A rule is defined as the implication of the form $X \Rightarrow Y$ where $X, Y \subseteq I$ and $X \cap Y = \emptyset$. The set of items (for short item sets) X and Y are called antecedent and consequent of the rule respectively.

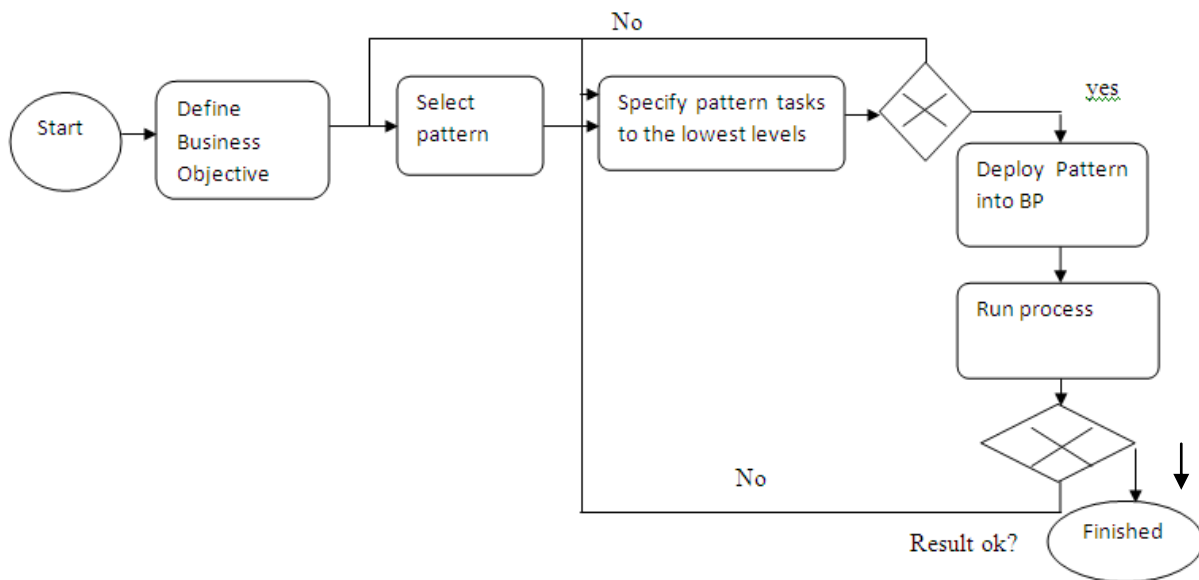
Knowledge to be provided: This technique is used to find out the associations between items from a huge amount of transactions.

Most Suitable data type: It is usually applied to transactional data forms.

(3) Pattern based approach

Here we focus on how to enable the reuse of existing solutions that have been proven to be successful using Meta process. This Meta process describes the steps needed to use a pattern for a given business process. Figure 5 visualizes the Meta process and its steps. The First task is to define the business objectives of the application. After that a data mining pattern is selected that matches with these business objectives. Then, the tasks of the selected pattern are specified to the executable level according to a given specification is chosen. If it is observed that the pattern cannot be specified as executable, the meta-process steps back to the task of choosing a new pattern.

Figure 4: The meta-process for applying a process pattern to a given business process.



Example-an employee of a company placing an order to another company which is owned by himself. This is done by computing a similarity between employees and company owners based on several features such as name, address or bank accounts. We do not go into more details of the data mining method, as this is not important for understanding our pattern approach. Basically what is needed to apply this data mining solution to a problem is to first check if the problem is a procurement fraud problem, second to specify which features to be used for the similarity, and third to connect the inputs and outputs of the data mining process. For all other steps ready-to-use code is already available.

Figure 5: An example of a data mining pattern

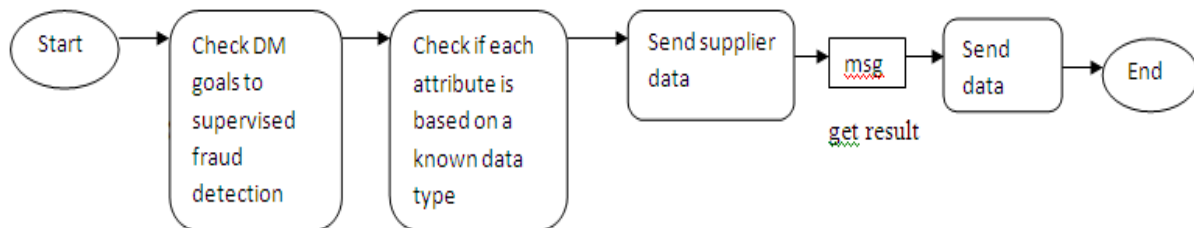


Fig. 5 shows an example of a pattern for this approach to procurement fraud detection. In the top pool Requirements the pre-requirements for applying this pattern are modeled. This includes checking the data mining goal (procurement fraud detection) and the data format as well as sending data, receiving the result and sending labeled data. The other pools Data Mining Classification, Data Mining Model Building and Data Mining contain the (partially already specified executable) tasks of the data mining process and the respective services.

Knowledge to be provided: This technique is used to find out the pattern in database similar to the pattern in hand. It can be divided into two kinds: text similarity search and time-series similarity search.

Most Suitable data type: Textual data forms are suitable for text similarity search and temporal data forms are suitable for time-series similarity search.

Concluding remarks: In this paper, we mentioned the fact that none of data mining technique can be applied to all kinds of data types. To increase the effectiveness of an financial organization we discuss the data types used in an organization and appropriate method to mine that data properly from the data warehouse of financial organization.

REFERENCES

- [1]. Hornick, M.F., Marcade, E., Venkayala, S.: Java Data Mining: Strategy, Standard, and Practice. Morgan Kaufmann, San Francisco (2006)
- [2]. Wegener, D., Røping, S.: On Integrating Data Mining into Business Processes. In: Proc. of the 13th International Conference on Business Information Systems. LNBI, vol. 47, pp. 183-194, Springer (2010)
- [3]. Codd E. F., "Providing OLAP to user-analysts : An IT mandate", Technical Report, E.F. Codd and Associates, 1993.

- [4]. Bhandari I., "Attribute Focusing: Data mining for the layman", Research Report RC 20136, IBM T.J Watson Research Center.
- [5]. Marban, O. Segovia, J., Menasalvas, E., Fernandez-Baizan, C.: Toward data mining engineering: A software engineering approach. *Information Systems* 34 (1) (2009)
- [6]. Data Modeling Fundamentals: A Practical Guide for IT Professionals
By Paulraj Ponniah
- [7]. Atwood, D.: *BPM Process Patterns: Repeatable Design for BPM Process Models*. BPTrends, May (2006)
- [8]. R•uping, S., Punko, N., G•unter, B., Grosskreutz, H.: Procurement Fraud Discovery using Similarity Measure Learning. In: *Transactions on Case-based Reasoning*, 1(1), pp. 37-46 (2008)