

RedBox-A Data Mining Approach for Improving Business Intelligence

Tapan Nayak

Research Scholar
Department. of Computer Science
CMR University, Bangalore
Karnataka – India

ABSTRACT

This main idea behind introducing RedBox concept is to introduce a new Data Mining approach using clustering methodology for extracting knowledge base information or data pattern to improve business intelligence (BI) in a business. BI is a paradigm for modern business world which could be a priority based mining concept to improve the productivity in business. A group of transacted data or informative data where date is an identifier can be passing through a clustered based data mining algorithm in different priority cluster, and the priority key identify as a priority group cluster by which information derive. The main focus area of this paper is introducing a knowledge based business intelligence using mining tools to provide better decision to business. This paper describes a new approach unlike other mining algorithm for clustering data where data are grouping of similar instances /objects and determine whether two objects are similar or dissimilar and give appropriate group in each cluster is required, which is used like BI tools. The conceptual theme of this paper is to give importance to know the objective reason of the transaction happened the date, which could be parameters for business intelligence. The main objective of this paper is to present the extension of the new clustering techniques consisting of multiple clusters and presents a novel method of data clustering and improves business intelligence. Further extension consists in a distributed approach allowing direct access to the information system by WEB tools.

Keywords:- RedBox, Cluster, Business Intelligence, Methodology, Web.

I. INTRODUCTION

Business should have a corollary to the adage “those who cannot remember the past are condemned to repeat it.” In the case of business, problems arise when we keep repeating past practices without understanding the effects of those practices [10]. RedBox is a new paradigm in business worlds and a new way of reaching customers and making money. It is an innovative concept where Business Intelligence (BI) playing a major role through mining tools like clustering method. Clustering is a fundamental task in data mining techniques which provide solutions for the extraction of new information. Data mining survey emphasis is on clustering in dataset. Such clustering is characterized by large datasets with many attributes of different types, potentially being related to each other's. Clustering in data mining was brought to life by intense developments in information retrieval, and data mining primarily works with large databases. BI is an automated application of non-user predefined

components which inform the given database where business can take immediate action. RedBox based BI has major components in improvement of business module, whereas we need to spend minimum time to take decision. In current business trend a BI based system tools can give potential customer and minimized the other unusual expenditure like promotion and advertisement for improvement of business strategy. RedBox has support for clustering objects/users using k-means clustering model. This paper designed on primary objective to identify valid date in which business was good and what is the reason behind such event. This Redbox approached the real world decision making, emphasize the concept of a decision making process that involves more factors and aspects like: the use of own and external knowledge, involvement of various, actors", aspects, etc., individual habitual domains, non-trivial rationality, different paradigms [1].

1.1 Motivation

Reasons for RedBox

- **Data Recovery.** Clustering is a division of data into groups of similar objects. Representing the data by fewer clusters necessarily loses certain fine details, but achieves simplification. It models data by its clusters. Data modeling puts clustering in a historical perspective rooted in mathematics, statistics, and numerical analysis.
- **Improvement of Intelligence.** The development of intelligence skills and potentially protect the business interest in such a way that, improving productivity and increase the level of business high.
- **Cost Management.** The development of new technology in adaption of current trend of business where business intelligence can improve the cost cutting methodology in such a way that gives profit to the business.
- **Data Integration :** The combined data sources which is the routine data collection from various data sources example from heterogeneous databases and data warehouses and combined in a suitable manner. At this step, the relevant data to the analysis process is targeted and retrieved from the data source(s).
- **Market Share.** To examine how a data model works, what are specifications of its structure and identify potential component to be used for potential business intelligence infrastructure to save business data for further analysis without intruder's interference to improve market share.
- **Training Purposes.** RedBox emphasis on module learning for the purpose of improvement in department wise development which is the main purposes may be understand the key issues of an unsuccessful decision and subsequently improve the decision policies. The implementation of technical intelligence (understand what your competitor is actually doing, versus what they say they are doing).

II. APPLICATION OF REDBOX FOR RETAILS BUSINESS

Data extraction from daily transaction data and patterns extracted from applying data mining techniques on stored data which can be used to maintain database, by improving their usability through simplifying user navigation and information accessibility and improving the content and the structure of the database in such a way that, understand the requirements of both business owner and user which will consequently increase the

overall profit of the business or the industry that the maintained data structure belongs to .

Another application field of data mining is using association rule mining to analyze market basket data. A transactions database contains information about customer's transactions, where each transaction is a collection of items. Association rule mining captures the relationships between different items. An association rule finds the probability that two different items occur together in the transactions database. Association rule mining is finding all association rules that have support and confidence values greater than or equal a user-specified minimum support (minsup) and minimum confidence (minconf) respectively. minsup and minconf are functions that measure the interestingness of an association rule. Those rules are called interesting association rules. But the interestingness of an association rule that represents a group of items can have many different meanings. For example, an interesting rule may give some information about well-sold products. On the other hand, if we have a number of non-interesting association rules, we can also use them to gain some information about bad-sold products which is also considered a valuable information that can be invested by the marketers to improve their marketing strategies [3].

Here a product designs approach of each cluster with unique product ids shown. The basic idea of this knowledge based engine is to extract meaningful knowledge from given data sets through data mining algorithm where date of transaction is the identifier of BI. In this paper a modern concept of data mining and using active mining tools as an algorithm has been highlighted. The above picture shows the transacted items derived from priority or group cluster and the business intelligence to applied to determined why this transaction occurred on this particular date and the BI interface can define the reasons of transaction happened. In this paper a design model of knowledge extraction from each transaction has been highlighted.

The new concept of mining algorithm which is proposed to develop is to take predefined business model whereas customer can be influence to take decision which will give profit to the business and subsequently they also get benefitted. RedBox is a well-defined model which shows the derivative objective of business intelligence and constructive structure of mining algorithm in the context of knowledge management shown Figure 2.

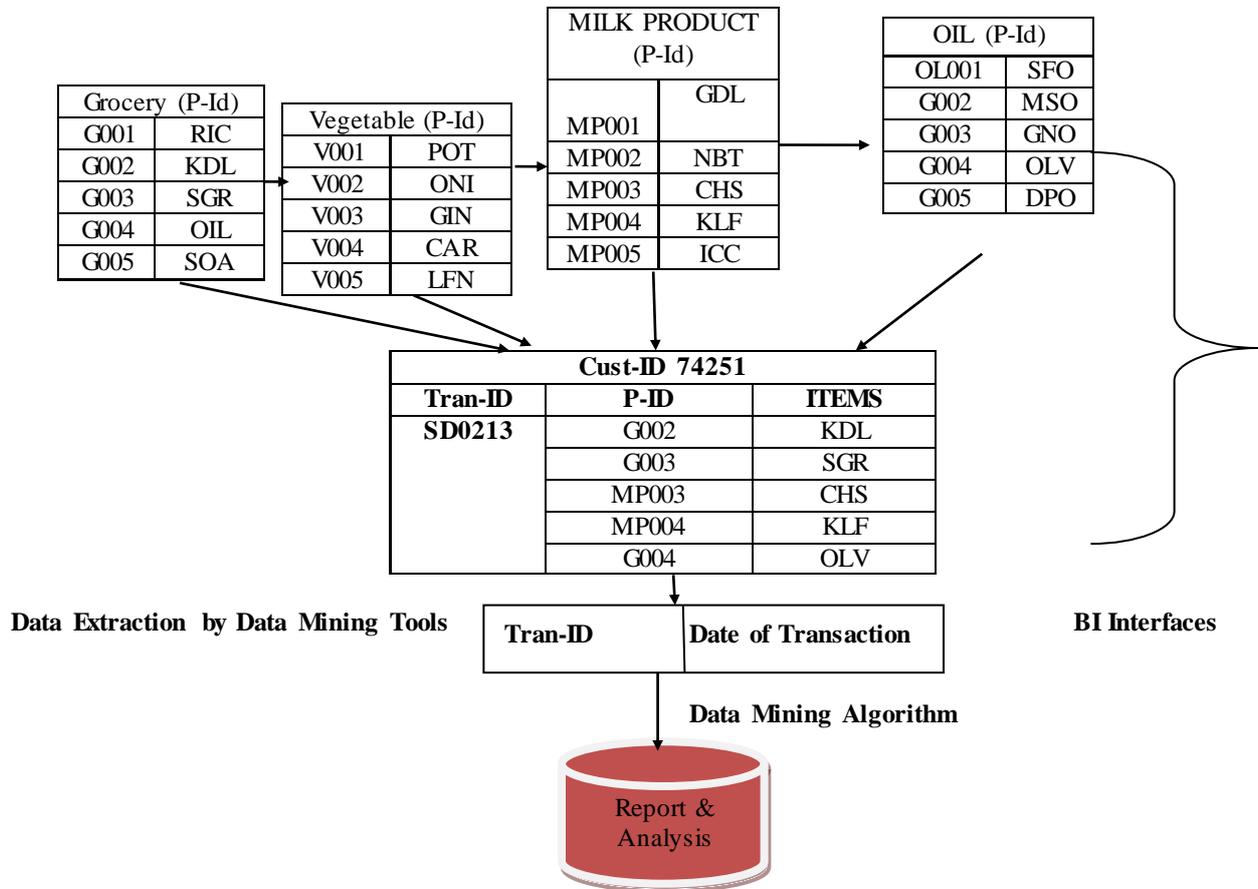


Figure 1: Data Mining Model with Different Cluster:

The proposed RedBox concept provides solution for customer oriented systems. This methodology can be verified through various experimental setups. In this work transaction dataset is used for testing the proposed data set for products purchase by the users. This method help active users find items they want to buy from a business.

III. DATA MINING TECHNIQUES FOR BUSINESS INTELLIGENCE MODEL

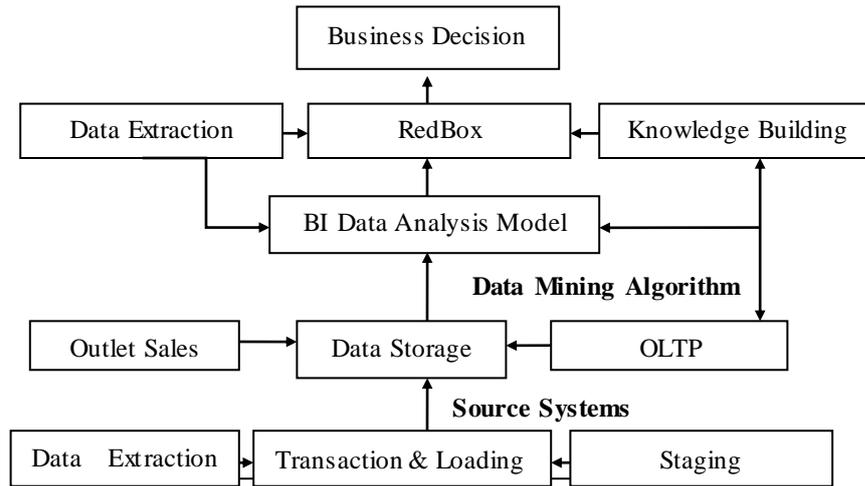
The algorithms in the Data Mining Framework are designed to work on data present within an Analytic Services Model (ASM). The design of the ASM should take into consideration the data needs for all kinds of analyses (OLAP and Data Mining) that the user is interested in customer’s performance in terms of transaction processing. Once the data is brought into the ASM environment it can then be accessed through the Data Mining Framework for predictive analytics. The Data Mining Framework uses mining tools to identify sections within the ASM to obtain input data for the algorithm as well as to write back the results. The Data Mining Framework can only take regular dimension members as mining attributes. What this implies is that

only data that is referenced through regular dimension members (not through attribute dimensions or user defined attributes) can be presented as input data to the Data Mining Framework. Accordingly, the data that is required for predictive analytics should be modeled within the standard dimensions and measures within a cube. From clustering customers into market segments and finding the characteristics of frequent flyers to learning what items are purchased with other items, key data mining methods, including classification, prediction, and affinity analysis as well as data reduction, exploration, and visualization for Business Intelligence [7].

In order to maintain the information, a new systematic way has been used such as database. In this database, there are collection of data organized in the form of tuples and attributes. In order to obtain knowledge from a collection of data, business intelligence methods are used. Data Mining is the powerful new technology with great potential that help the business environments to focus on only the essential information in their data warehouse. Using the data mining technology, it is easy for decision making by improving the business intelligence [9]. The application model of RedBox

concept has been discussed in this paper for implication of mining approaches in business intelligence based model. An advance data mining concept with the help of data mining tools can be design to give better work methodology. The methodology start from the problem definition, then data collection from transaction

database are discussed. Here data organized so there is no need of preprocessing, then we come to data mining methods that are Decision Tree, Rule Induction, and clustering followed by Association Algorithm and evaluation of result. Figure 2.



Work Methodology for RedBox

3.1 Clustering Techniques

The main data mining technique clustering, also group objects together but unlike classification algorithms, these techniques do not use a predefined set of categories. Clustering is useful for exploring our data and finding groupings implied by the data itself. Consider the sales transactions of a consumer supermarket store. Her clustered customers by sales date might have several groups emerge: customers who buy primarily milk item and baby food frequently; customers who buy fish every time; customers who buy vegetable, milk, and oil, etc.

3.2 Decision Trees

The use of decision tree [4; 5] is essentially a flow chart of questions or data points that lead to a decision. These decisions generate rules for the classification of a dataset. Decision tree systems try to create optimized paths, ordering the questions such that a decision can be made in the least number of steps. An example of a decision tree method is Classification and Regression Trees (CART). CART provides a set of rules that can be applied to a new (unclassified) dataset to predict which records will have a given outcome. Decision trees are great for situations in which a visitor comes to a website with a particular need so he can be assigned to his target need easily. Decision trees are well suited

for our approach as they help in dividing customers and products into different groups and categories with respect to different attributes which enables an easy analysis and characterization of the target data. Classification algorithms are used to predict how a person, object, transaction or event should be categorized. You might want to categorize current customers into three categories: loyal, possibly leaving, likely to leave. With data about customer characteristics, purchasing patterns, etc. and a sufficient number of examples of each type of customer, you can use classification models to help understand how each current customer should be categorized [10].

3.3 Rule Induction

The use of rule induction [6] is the extraction of useful if-then rules from data based on statistical significance. Rule induction defines the statistical correlation between the occurrences among certain items in a dataset. One of the main data mining tasks that uses rule induction method is the association rule mining by which associations between different elements of the target data can be extracted which is very helpful in our approach to understand customer buying behaviors and trends.

3.4 Association Algorithm

The use of Association Algorithm for recommendation engine that is based on a product and market based analysis. This engine suggests products to customers based on what they bought earlier. The model is built on a dataset containing identifiers. These identifiers are both for individual cases and for the items that cases contain. These groups of items in a data set are called as an item set. The algorithm traverses a data set to find items that appear in a case. MINIMUM_SUPPORT parameter is used any associated items that appear into an item set. Market basket analysis, also known as associative analysis, is useful for analyzing groups of items that commonly appear together in the data. A supermarket manager would not be surprised to find bread, peanut butter and jam frequently purchased together in US stores. Less obvious groupings can be found using market basket analysis techniques. These can help support the development of cross selling efforts, loss leader campaigns, or product placement within a physical store [10].

The proposed RedBox approach focuses on a recover of components that may affect the behavior (and, thus the comprehension) of the Business Intelligence; i.e. components like tools, input/output forms, frames, and links are modified, while interacting.

IV. CONCLUSION AND FUTURE WORK

This paper presents an introduction of RedBox technology and new clustering techniques for business data structure and also describes tool architecture to cluster from existing clustering methods, but also with the more challenging data pattern content using business intelligence model. The application of data mining algorithm also highlighted to improve the BI objectives. This will be very useful for business relevance of the data mining process where potential customer could identify and needs to be validated by further studies: the paper assumes that designing new methodology, and which is best fit for homogenous data structure applicable to multi-tasking business module.

REFERENCES

- [1] Lior Rokach, Oded Maimon, Clustering Methods, Department of Industrial Engineering, Tel-Aviv University
- [2] Pavel Berkhin, Survey of Clustering Data Mining Techniques, Accrue Software, Inc.
- [3] Omari, (2008), Data Mining for Retail Website Design and Enhanced Marketing, PP-2.
- [4] Pat Langley and Herbert A. Simon. Applications of Machine Learning and Rule Induction. Communications of the ACM, 38(11):54–64, 1995.
- [5] Agnar Aamodt and Enric Plaza. Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. AICom - Artificial Intelligence Communications, 7(1):39–59, 1994.
- [6] Pat Langley and Herbert A. Simon. Applications of Machine Learning and Rule Induction. Communications of the ACM, 38(11):54–64, 1995.
- [7] Galit Shmueli, Nitin R. Patel, Peter C. Bruce , (2010). Data Mining for Business Intelligence: Concepts, Techniques, and Applications in Microsoft Office Excel® with XLMiner®, 2nd Edition.
- [8] Janusz Kacprzyk, (2006), Web Intelligence, Business Intelligence and Decision Support Systems: A Challenge for Fuzzy Logic and Soft Computing, 3rd International IEEE Conference Intelligent Systems.
- [9] Prabhu, et al, (2014), A New Hybride Algorithm for Business Intelligence Recommender System, International Journal of Network Security & Its Applications (IJNSA), Vol.6, No.2,pp.43-52.
- [10] Dan Sullivan, (2012), Next Generation Business Intelligence: Data Mining [online] http://www.tomsitpro.com/articles/business_intelligence-data_mining_tools-data_analytics-spss-olap_analysis