

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/339760420>

Analyzing Data Mining and Its Application to Smart Business

Conference Paper · December 2019

DOI: 10.1109/TIMES-ICON47539.2019.9024369

CITATIONS

0

READS

57

6 authors, including:



[Saifun Nahar](#)

University of Missouri - St. Louis

16 PUBLICATIONS 160 CITATIONS

[SEE PROFILE](#)



[Happy Nkanta Monday](#)

Oxford Brookes College of Chengdu University of Technology

50 PUBLICATIONS 574 CITATIONS

[SEE PROFILE](#)



[Grace U. Nneji](#)

Oxford Brookes College of Chengdu University of Technology

48 PUBLICATIONS 447 CITATIONS

[SEE PROFILE](#)



[Hassan Sani Abubakar](#)

University of Electronic Science and Technology of China

16 PUBLICATIONS 79 CITATIONS

[SEE PROFILE](#)

Analyzing Data Mining and Its Application to Smart Business

* Saifun Nahar¹, Ting Zhong¹, Happy N. Monday², Grace U. Nneji¹, Michael O. Mills², Hassan S. Abubakar³

¹School of Information and Software Engineering

²School of Computer Science and Engineering

³School of Information and Communication Engineering

University of Electronic Science and Technology of China

Chengdu, Sichuan P.R China

* sainahbd@163.com

Abstract— This paper presents the main concepts of smart business (SB) and the application of smart business to provide solutions to the problems faced by a corporation. In this context, the main emphasis is on the approach of data mining including one of the tools used by SB that uses Decision Trees and Artificial Intelligence techniques along with other techniques. Moreover, algorithms such as Genetic Algorithms system and the Neural Networks are considered in its implementation. The concept of Smart Business (SB) has clarified these problems and consists of broad category technologies and application programs used to extract, store, analyze and transform high dimensional data. The approach of Smart Business is presented according to the concepts of Data Warehouse, Data Mart and Data Mining .

Keywords— smart business; data mining; artificial intelligence; neural network; decision tree

Introduction

The accelerated technological growth and consequent computerization made it achievable to make operations. As hardware became cheaper and better performing, large systems were created, enabling fast and efficient generation of high dimensional data. These are stored in the archive, which has a common characteristic: the difficulty of extraction of information. Important data are saved within the company and their access is very difficult. From this, there is no gainsaying that this gradual process of deploying systems produced multiple stocks of data, with and consolidation difficulties. The concept of Smart Business (SB) has clarified these problems and consists of broad category technologies and application programs used to extract, store, analyze and transform high dimensional data. Thus, producing an environment of knowledge, where there is systematic production of managerial information that is able to assist companies to take the better business decisions. These decisions are based on real facts, deep enough to discover the causes of a tendency or a problem. Smart Business has some main features:

- The capability to extract and integrate data from different type of sources.
- The transformation of records obtained from information useful for business knowledge.
- The worth of experience.
- The process of contextualized data.
- The search for cause of effective relationships, working with hypotheses and developing strategies for competitive actions.

Building a Smart Business environment requires data from the various computerized systems and existing techniques in the company. The data set collected is the raw material for a range of transformations, whose final product is loaded on the warehouse. The views are created for the possibility of real-time managerial decisions making. Business applications that use the information stored in a Data Warehouse, can include queries and reports, OLAP (On-Line Analytical Processing) analyzes, forecasts, statistics, and Data Mining. The concept of Data Mining arose with the objective of improving the use of information stored in a Data Warehouse, using intelligent algorithms that can select the most relevant standards for applications. Within this context, some techniques such as Decision Tree, Genetic Algorithms and Neural Networks are utilized. This paper is structured into ten sections.

I. Smart Business (SB)

The era of loyalty, customization, impressing the client and intelligence is applied to business. As such, large corporate databases began to produce variants, such as Data Warehouse, with the objective of delivering decision-makers, the information in the most accurate path and usable as possible. The data that until then were representatives of common facts such as name, address, telephone, among others, are today sophisticated in the representation of images, videos, sounds, temporal data, indicators spreadsheets, HTML pages and XML structures, accompanying the changes requested by a company now leveraged by other industries such as entertainment, communication and e-commerce. Informatics made the data and then turned them into information. Now the aim is to form those raw materials. BI stands for skill to

structure, access and exploits information, usually stored in DW / DM (Data Warehouse, Data Marts) perceptions, understandings, and knowledge, which can produce a better decision-making process. It is the approach of Data Mining (data mining) aimed at improving the use of these huge arsenals of information through the identification of the pattern of correlations that are commonly invisible conventional. Product indicators purchased from or patterns of fraud practiced by business managers, in their daily lives, to discover sinuous correlations that will certainly lead to better the gondola of your market or modify the criteria for risk process of a proposal for a loan. The era of SB-Smart Business is alive, with information machined from the data being directly applied to business [1]. As well as the management area knowledge management system (KMS), BI aims to create data stocks and facts that support key decision-making processes. While the two approaches above aim to look at within the company, there is a third one that aims to outside the company. It is the approach of CI- Competitive Intelligence. One can understand CI as an SB applied to the world outside corporate boundaries, primarily on textual and factual information that respect of market movements and competitors. The strategy goes through the neighborhoods of the concepts of KMS and is coupled with Mining. The proposal of BI coupled with those of CRM (Customer Relationship Management) and SCM (Supply Chain Management) to target on clients is something which seems definitive. SCM represents the production processes and shipment of the final product, from the client. CRM are the systems, applications, and technologies devoted to deal with the client of the company, with the objective of developing the marketing, sales, and services. A new concept emerges in the computer industry: EAI (Enterprise Application Integration), which can be translated as the integration of enterprise applications. In this work, the approach of Smart Business is presented according to the concepts of Data Warehouse [7], Data Mart [1] and Data Mining [5], covered in the next section.

II. Data Warehouse, Data Mart And Data Mining

The data that inhabit the traditional legacy systems recently implemented, ERP (Enterprise Resource Planning), or integrated management packages, which constitute the basis of the business processes of the companies, are formatted and structured in a transactional way, thus hindering their treatment informational. Thus BI should be understood as the development process of:

- Special structures for storing information such as Data Warehouse (DW), Data Mart (DM) and ODS (Operational Data Store), with the goal of assembling an information resource base capable of sustaining the intelligence layer of the organization is possible to be applied to its business, as elements differential and competitive. Together with the approach of DM, DW, OSD, the SB concept also includes the ETC Toolkit - Extraction, Treatment, and Cargo, which are fundamental for the transformation of the resource from transactional to informational data. While DW and DM

refer to dimensional data structures, refurbished with the objective of providing differentials, the concept of OSD, in turn, is similar to the storage and processing of data operations, in a consolidated manner, but without the dimensional characteristics. The ODS beyond half the way between the legacy and the DW also provides important information decision-making, because of its characteristic of consolidation and integration of various data sources.

- Special applications for data processing, such as OLAP (On-Line Analytical Processing) and Data Mining. The term OLAP (On-Line Analytical Processing), now widely used, translated for on-line analytical processing, represents this characteristic of working the data, with operator's dimensions, allowing a multiple and analysis. The concept of Data Mining by, on the other hand, it is more related to the analysis of inference than with analysis dimensional representation of data and represents a form of search based on algorithms that recognize hidden patterns in the data necessarily revealed by the other approaches such as OLAP. Figure 1 shows schematically the components of Data Warehouse, Data Mart, ODS and Tools (among which Mining), composing the mosaic of SB. In addition to the data warehouses in the form consolidated data warehouse or by subject (such as Data Marts) and ODS, ESF-Transformation and Cargo, bound for the actions of collection, cleaning, preparation, and loading of such information. The mining processes will work on an extract of data specially prepared for this form of treatment.

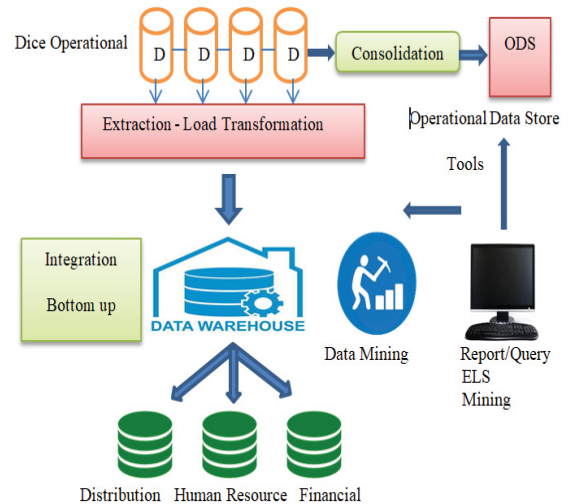


Fig 1: Components of BI- Business Environment Intelligence

III. Market And BI Tools

The tools for a Smart Business environment Intelligence can be organized as Construction, Management, Use, and Storage. Construction tools are designed to help in analyzing of extracting data from diversified sources; it's processing,

processing, and loading into the DW / DM final structures. The tools of Management aim to the storage and use of the DW / DM and the repository, where metadata information resides, responsible for defining the structures and processes desired transformation. The tools of use are, in essence, the mechanisms, through which users manipulate the data in the DW / DM and obtain the required information. They offer an arsenal of operators, among others drill-down, up, cross, through and work on Web interfaces.

Data Mining is present in this kind of tool. The Smart Business market points to quite the starting of the new millennium. With the dizzying growth of the SB concept and the strong trend of mergers and acquisitions, the SB will be fairly disputed in the coming years. A Computer Associates strengthened its SB arsenal with acquisitions of Platinum and Sterling, which in turn had already purchased other smaller companies. Now, Computer Associates should integrate them into an SB mosaic [1]. The Oracle has developed a powerful set of tools beginning with the gaining process of Oracle Express, followed by other tools in the mining area (Darwin's Thinking Machines). In the scenario of products that will be fighting for emerging SB market, two to catalyze the attention of DW and DM developers, and directly influence the processes of analysis and choice of OLAP solutions for companies: the BW- Business Information Warehouse, SAP and MS-SQL Server Analysis Services, Microsoft.

IV. Data Mining

One of the big problems is the transformation of data into information. How to do this in an automated way and in the smallest as long as possible? Another answer is a combination of conventional statistics and artificial intelligence, which results in a very commented today, Data Mining. Data mining processes are extremely complex, besides being a work that will hardly be resolved only by a Data Mining specialist. In any Data Mining project is required to exist of a professional with the knowledge of the business, since he has full mastery of the subject. To prevent result of the modeling is not of the support a decision. According to [8], the approach of Data Mining fluctuates between a form of statistics and a concept revolutionary, now applied to the market. In fact, the purpose of this technique is to find patterns, not yet discovered in data that can generate correct answers for new cases. This process of searching and interpreting is typically collective and iterative, involving the repetitive application of specific mining methods of data. In the scheme are the big blocks of the project, with the phases of Preparation, Mining, Analysis, and Application. The preparation phase consists of activities that will from the planning of a separate database to the data subject to Mining until the activity of loading the database for the Mining process. The arrangement of the available data to be considered in a project will vary from according to the chosen Mining algorithm. Depending on this algorithm, the data will be formatted in much different way. This process of data preparation is determinant in the success of Data Mining and usually time and resource. The mining phase is responsible for creating the models Data Mining, define samples or population and select data to train the model. In addition, here the formatting required by the tools. For

example, neural networks require dichotomous form (yes/no) and demand decision tree groupings such as good, medium and bad. Finally, it creates the predictors or key attributes for business analysis. Some data mining basic models can be considered in the analysis phase such as:

- Association rules: These are rules that are formed with information that exists in the database. Are relationships that in most cases are not known by the business owner
- Classification: these are processes used to define groups or classes of elements, depend on certain pre-established requirement. Several approaches are used to define reference models (networks neural, decision-based, rules-based trees) that some allow the explicit definition of the class (decision tree) and others its implicit model (neural networks).
- Sequential Standards: processes that aim at the identification of facts that imply in other facts, in different times of time.
- Aggregation: aims to obtain data that are similar in a cluster.

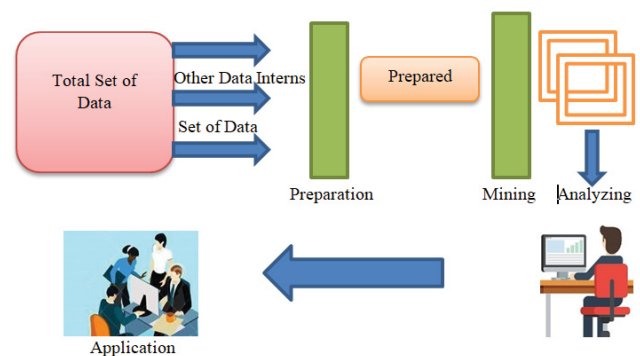


Fig 2: Data Mining Process Overview.

These data mining models also can be used in an integrated way, performing cascade analyzes, with operators applied to the results of others. Once the model has described and tested, the application, which is given by the use of those algorithms adjusted in real systems situations. Some products source code, resulting models and algorithms defined and compiled, which can be incorporated into traditional technique and to carry out the required analyzes. There is a different kind of Data Mining system that can help to find information to their decision-making. Regression Analysis technique on web-based data mining tool is detailed in [6]. Techniques such as Conglomerate Analysis, Predictors with Time Series, Decision Tree, Neural Networks and Genetic Algorithms are important. These last three are presented below. Only specialists can identify the best technique for each problem.

V. Decision Tree

The decision tree is a technique creates and organizes rules for classification and decision-making in tree diagram format, which will classify their observations or predict future results. If your data are divided into dichotomous classes, for example, infected against non-infected, tree decision-making can be built to create rules that cases or new cases, with precision. It starts with a single group that brings together all cases in the study. As the tree expands, this base of into modules representing categories of variables analyzed. Each branch of the tree is assembled by these nodules that are opening in mutually exclusive subgroups. Each node and each has a portion of the response obtained in study. The use of Decision Tree is comprehensive and following approaches:

- a) Segmentation: Identification of groups depends on the identification of common features presented by the elements. For example, market segmentation.
- b) Stratification: Determination of rules so that may assign each case to one of several existing categories, such as sort a risk-taker client high risk, medium risk, low risk.
- c) Prediction: Creation of rules for application in events futures. It can also be used to specify between predictable attributes and the worth of a continuous variable. For example, the identification of marketing actions may lead to an increase in sales.
- d) Data reduction and variable filtering: can be used to identify which variables have the most influence on the response, thus reducing the volume of variables under study.
- e) Identification of interactions: identification of interactions relevant only in certain subgroups and specifications in parametric models formal.
- f) The merit of categories and discretization of variables continuous: means the possibility of recording of categorical and continuous variables with information.

Decision Trees can be applied in several areas. Some typical examples of its application are presented in [1]. Among them, we highlight: Mala Direct; Credit Scoring (also known as credit squeeze); Market analysis; Control of Quality; Human Resources; Medical and Internal Policy Studies

VI. Neural Networks

Neural Networks are an increasingly important technology system in Data Mining. Its great advantage is their ability to learn from the experiences, not being restricted to an order pre-set sequence. They consist of algorithms and computational procedures that mimic the learning of the brain. This technique is nodules whose processing resembles that of the neurons, hence its name. It does not offer defined estimators of a neural network, with a certain mass of data, nor will always repeat itself with another. The nodes are connected as a network and function in parallel. The first stage of nodules is made of the input nodes. They receive the input of the variables provided by the database, transform it according to a function (called activation function), producing output information which will be transferred to the next node phase. Various input information will be received in turn. This phase is composed of the hidden nodes, which, in more complex neural networks, can form several layers. Finally, we

have the output nodes. They process the information received and produce a response, but don't send it to another node since it is already the end result of the network. If the network is classification, the output node is already the end. In the case of forecast models, the output node already represents a predictive value. The important note is that there is no universally better than all. The success of Data Mining relies heavily on the experience and researcher, who will have to identify the best tool to be used, according to the type of response and with the way in which they are your data. Therefore, the application examples of procedures can be repeated. The technique of Neural Networks can be used in Data Mining to solve problems related to Marketing; Predictive Models; Sales; Finance; Energy and; Production. Examples of this kind of applications are presented in [1].

VII. Genetic Algorithms

Genetic Algorithms (Gas) are methods of optimization and search motivated by the mechanisms of evolution of populations of living beings [8]. These algorithms follow the fundamental of natural selection and survival of the fittest, declared in 1859 by the English naturalist and physiologist Charles Darwin in his book, the origin of Species. Optimization is the search for the good solution for a given problem, in order to save the surplus resources, be they financial, time, etc. It consists of trying various solutions and uses information obtained in this process in order to find better solutions. The first step of a typical Genetic Algorithm is the generation of a basic population of chromosomes. It is population is made by a random set of chromosomes, which represent the possible solutions to the problem to be solved. During the evolutionary process, this population figures out and each chromosome receives a (known as aptitude), reflecting the aspect of the solution that it represents. In general, the chromosomes are selected and the least fit is discarded (Darwinism). Selected members may undergo the changes in their characteristics through crossover operators, generating offspring for the next generation. This process will continue until a satisfactory solution is found. More recent research points to the use of Hybrid Genetic Algorithms, where Genetic Algorithms are combined with other techniques such as Decision Tree, Data Mining. Debora [3]; [4] proposes to solve the problem of imbalance of classes in a database, depend on the use of a hybrid method with Algorithms Genetic and Decision Tree. In this case, the Algorithm Gene acts as a Wrapper using decision tree algorithm, to obtain the target populations.

VIII. Smart Business Applications

SPSS is a leading technology company in Data Mining and analytical applications to develop decisions. Such applications meet the objectives of the Smart Business, whose premise is to assist organizations in the management of the future through learning from the past, understanding the present and assume potential problems and possibility. The company provides all the applications necessary for process analysis, from planning, data management, analysis and distribution of results. Some of

these applications analyze information to determine reliably what will happen in the future. The decision tree is useful on Data As discussed in Section V. Each branch of the tree separates the data in subsets corresponds to a rule that produces pre-conditions for the data current and new data. The three components of decisions are useful, for example, in segmentation of marketing to identify specific groups in a bank customer data; in risk analysis to discover factors that determine whether a specific application is good or bad credit risk; management. Customer relationship management (CRM) in how clients interact with a particular company. One of the products offered by the company is the Answer Tree tool, which is software for Data Mining that enables business users to analyze data with greater precision. It uses the Decision (<http://www.spss.com/spssstd/tools.htm>). An example of its application is when an organization wants to determine which customers or group of citizens that best match their promotions and programs, marking precisely the people and eliminating the conjectures that lead to sending offers to arbitrary groups of people in your email lists. In this case, Answer Tree allows the organization to reach more effectively the right group of people. It uses scalable Decision Trees, which segments and predicts how groups will respond to and programs. Answer Tree offers four decision tree, (you can choose what you want to best applies to user data), Answer Tree offers four decision tree, (you can choose what you want to best apply to user data), shows templates allowing you to view the groups of that subject. The graphs are necessary the allow the results quickly. Data mining algorithm and technique has been proposed in [5].

IX. Conclusion

Nowadays, SB is important for decision making. In this context, it can be said that DW (Data Warehouse) has become the main tool for the work of the marketing department and that most of the information for decision-making is to get from the systems to maintain decision-making. Among the various, use of SB systems, traffic, revenue tracking, and its trend,

churn analysis and customer retention. Several techniques, tools, and concepts exist in the Marketplace. However, when a company seeks to implement an SB-Smart Business solution, the first one you should identify what your problem is, how to attack it, and mainly how to solve it. These issues are fundamental to make a good choice of the tool or technique to be used. Once we have identified the need to support decision-making when it comes to development of software, we intend to work on investigating the applicability of Smart business SB in software.

ACKNOWLEDGMENT

Work supported by University of Electronic Science and Technology of China

REFERENCES

- [1] D. R. Carvalho, "A hybrid decision tree/genetic algorithm for coping with the problem of small disjuncts in Data Mining", In: 2000 Genetic and Evolutionary Computation Conference, Las Vegas. 2000.
- [2] D. R. Carvalho, "A hybrid genetic algorithm decision tree approach for coping with unbalanced classes", In The Second International Conference on Practical Application of Knowledge Management, London. 1998.
- [3] Z. Chen, "Data Mining and Uncertain Reasoning: An Integrated Approach", John Wiley & Sons, Canada, 2001.
- [4] H. Heinrichs, Jeen-Su Lim, "Integrating web-based data mining tools with business models for knowledge management". *Decision Support System*, vol. 35, 2003, p. 103-112.
- [5] H. Cheng, C. Rong, K. Hwang, W. Wang, and Y. Li, "Secure big data storage and sharing scheme for cloud tenants," in *China Communications*, vol. 12, no. 6, pp. 106-115, June 2015. doi: 10.1109/CC.2015.7122469
- [6] X. He, Q. Ai, R. C. Qiu, W. Huang, L. Piao, and H. Liu, "A Big Data Architecture Design for Smart Grids Based on Random Matrix Theory," in *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 674-686, March 2017. doi: 10.1109/TSG.2015.2445828
- [7] E. P. Xing *et al.*, "Petuum: A New Platform for Distributed Machine Learning on Big Data," in *IEEE Transactions on Big Data*, vol. 1, no. 2, pp. 49-67, 1 June 2015. doi: 10.1109/TBDDATA.2015.2472014
- [8] W. Huang, Z. Chen, W. Dong, H. Li, B. Cao, and J. Cao, "Mobile Internet big data platform in China Unicom," in *Tsinghua Science and Technology*, vol. 19, no. 1, pp. 95-101, Feb. 2014. doi: 10.1109/TST.2014.6733212