Chinese Business Review, ISSN 1537-1506
December 2013, Vol. 12, No. 12, 834-845

A New Approach for Knowledge Discovery in Distributed Databases Using Fragmented Data Storage Model*

Masoud Pesaran Behbahani
Azad University (IR) in Oxford, Oxford, UK
Islam Choudhury, Souheil Khaddaj
Kingston University London, Londonn, UK

Since the early 1990, significant progress in database technology has provided new platform for emerging new dimensions of data engineering. New models were introduced to utilize the data sets stored in the new generations of databases. These models have a deep impact on evolving decision-support systems. But they suffer a variety of practical problems while accessing real-world data sources. Specifically a type of data storage model based on data distribution theory has been increasingly used in recent years by large-scale enterprises, while it is not compatible with existing decision-support models. This data storage model stores the data in different geographical sites where they are more regularly accessed. This leads to considerably less inter-site data transfer that can reduce data security issues in some circumstances and also significantly improve data manipulation transactions speed. The aim of this paper is to propose a new approach for supporting proactive decision-making that utilizes a workable data source management methodology. The new model can effectively organize and use complex data sources, even when they are distributed in different sites in a fragmented form. At the same time, the new model provides a very high level of intellectual management decision-support by intelligent use of the data collections through utilizing new smart methods in synthesizing useful knowledge. The results of an empirical study to evaluate the model are provided.

Keywords: data mining, decision-support system, distributed databases, knowledge discovery in database (KDD)

Introduction

Large enterprises usually encounter enormous electronic transactions upon distributed infrastructures. They use huge database management systems (DBMS) to store and utilize these transaction. The stockpiled data collections in these systems are the best source for providing necessary material for decision-making process to support managers and executives. The existing models for supporting decision-making such as Fayad Knowledge Discovery in Database (KDD), Sample, Explore, Modify, Model, and Assess (SEMMA), and Cross Industry Standard Process for Data Mining (CRISP-DM) encounter practical problems while facing many types of real-world data sources, including the data collections in distributed databases. The distributed

* Project Supported by Azad University (IR) in Oxford, Strout Court, Oxford Road, Eynsham, Oxford, OX29 4DA, UK.
Masoud Pesaran Behbahani, Ph.D. Researcher, Azad University (IR) in Oxford.
Islam Choudhury, Ph.D., Principal Lecturer, School of Computing and Information Systems, Kingston University London.
Souheil Khaddaj, Ph.D., Reader, School of Computing and Information Systems, Kingston University London.
Correspondence concerning this article should be addressed to Masoud Pesaran behbahani, Flat 16, Falconry Court, Fairfield South, Surrey, Great London, UK, Kt1 2UR. E-mail: Masoud@kingston.ac.uk.

databases usually store fragmented data in large scales. This fragmentation makes the distributed database more efficient by keeping the tuples at the sites where they are used the most to minimize data transfer. Therefore the end-users can practice a much faster transaction speed. This fragmentation also provides more data security by minimizing inter-site traffic. Unfortunately this kind of data storage causes practical problems for the existing decision-support models. The problem is more severe in heterogeneous distributed databases.

Distributed databases might be heterogeneous or homogeneous. In a heterogeneous distributed database, different sites may use different schemas, and different database management system software with different format of data storage. In a homogenous distributed relational database they use the same DBMS nevertheless still they split the conceptual relations to vertical and/or horizontal fragments. These fragments are usually stored in different geographical sites where they are more regularly accessed. Vertical fragmentation can be defined as a projection on the relation *R*. Each projected subset must include the primary key so the original relation can be reconstructed by taking the natural join of all fragments: $R = R_1 \bowtie R_2 \bowtie \bullet \bullet \bullet \bowtie R_n$. while $\bowtie$ represents natural join operator in relational algebra. A horizontal fragment can be defined as a selection on the relation *R*. Each tuple of relation must belong to at least one of the fragments, so that the original relation can be reconstructed by taking the union of all fragments: $R = R_1 \cup R_2 \cup \ldots \cup R_n$, while $\cup$ represents union operator in relational algebra. This paper aims to introduce a new approach that uses a novel strategy for data engineering and supporting management decision-making process. The paper shows how the new strategy not only can tackle the mentioned problem, but also can be significantly effective in increasing the competitive advantage of any kind of organization by intelligent use of available data. This model basically tries to synthesize useful knowledge from collections of organized data.

Literature Review and Background

Decision-support systems have evolved from two main areas of research. The theoretical studies of organizational decision-making (Simon, Cyert, March, and others) conducted at the Carnegie Institute of Technology during the late 1950s and early 1960s and the technical work (Gerrity, Ness, and others) carried out at Massachusetts Institute of Technology (MIT) in the 1960s (Shim, Warkentin, Courtney, Power, Sharda, & Cjroster, 2002). Since the early 1990, substantial advancements in database management systems have provided excellent platforms for emerging new dimensions of data engineering. Regarding the importance of the decision-support systems, academics and practitioners started to design new models for supporting decision-making using database intelligence and data mining techniques. A pioneer in this area is Fayyad who clearly defined a model for KDD process (Fayyad, Piatetsky-Shapiro, & Smyth, 1996). KDD as defined by who is the practice of using data mining methods to extract what is considered as knowledge according to the specification of measures and procedures. The KDD process is preceded by the development of an understanding of the application domain, the relevant prior knowledge, and the goals of the end-user. His work successfully followed by researchers from SAS institute Inc. by introducing SEMMA. These phrases refer to the process phases required to conduct a data mining project. The problem with SEMMA is that it is configured to help the users of the SAS Enterprise Miner software. Another framework, CRISP-DM, initially was conceived in 1996. It is a non-proprietary, documented, and freely available data mining model. CRISP-DM organizes the data mining process into six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment (Shearer, 2000), while generally, the sequence of the phases is not strict. A comparative study by Azevedo and Santos (2008) comparing KDD, SEMMA, and CRISP-DM, illustrates

that SEMMA and CRISP-DM can be viewed as implementations of the KDD framework. According to this study, five stages of the SEMMA process can be seen as a practical implementation of the five stages of the KDD process. Another useful analytical framework is the one developed by Chung-Shing (2001). He introduced the framework primarily for evaluating ecommerce business models and strategies, though it is applicable in a wider range.

In all the previous models, practical problems occur when facing real-world data sources. The problem is serious when the data source is stored in a distributed database. The new introduced model provides a solid strategy to solve the mentioned problem. It also can be significantly effective in increasing the competitive advantage of any kind of organization by intelligent use of available data. Therefore it can be used even in the situations that there is not any problem while facing the data sources. The new introduced model is developed in two versions. Multidimensional Mining Management Model (4M) is the name of first version of the artifact that can be used to create a very competent decision-support system. The model emphasizes on using multidimensional data model to provide an interactive investigative and exploratory business perception. Multidimensional Multilayer Mining Management Model (5M) is the name of the advanced version of the artifact. The latter version is more effective than 4M and other existing models because it provides organizational insight. While 4M can solve the problem with distributed data and is featured by its multidimensionality, 5M is multidimensional and multilayer. It is multidimensional because the same as 4M, it focuses on multidimensional data model concepts. It is also multilayer because it uses multilayer mining structure based on the Multilayer Mining Theory (Pesaran Behbahani, Khaddaj, & Choudhury, 2012). These mining structures provide the platform for multilayer mining algorithms to maintain proactivity for the model, rather than just adjusting to situations and waiting for problems to happen. The paper primarily presents the outline of the new model. The results of some empirical study have shown the usefulness of the theory. The model can provide the desired organizational insight thanks to its multilevel mining structures (Pesaran Behbahani, 2012). Though 5M is introduced the focus in this paper is limited to the main features of 4M. The paper also presents the results of evaluating the model by adapting it to the ebusiness discipline. This adaptation results in introducing new software which is called EBAF.

Outline of the New Introduced Model

5M is a model based on multidimensional data and multilayer mining structures designed to intelligently use available data collections in the organizational databases. This intelligent use of data makes the model a significant progress over the existing ones in many aspects. The final outline of the model mirrors the final results of a thoroughgoing empirical study. Though originally designed for distributed databases, 5M serves as a generic model for all organizations wishing to capture database intelligence to significantly enhance their administrative behavior and gain profitable performance. It encourages best practices and offers organizations the structure needed to realize better, faster results from database intelligence. An overview of the model is shown in Figure 1.

Requirement Analysis and Enhancement Plan

In order to understand which data should later be analyzed, and how, it is vital to initially understand the business structure and objectives for which they are finding a solution. The organization understanding phase starts with an initial assessing the situation of the organization and proceeds with identifying its main target.

The most challenging part starts here by trying decomposition of the main target down into measurable objectives. The phase proceeds with identifying main influencers on the objectives, translating the objectives into technical requirements, and activities in order to get familiar with organization resources. Organization understanding perhaps is the most important stage of the model.

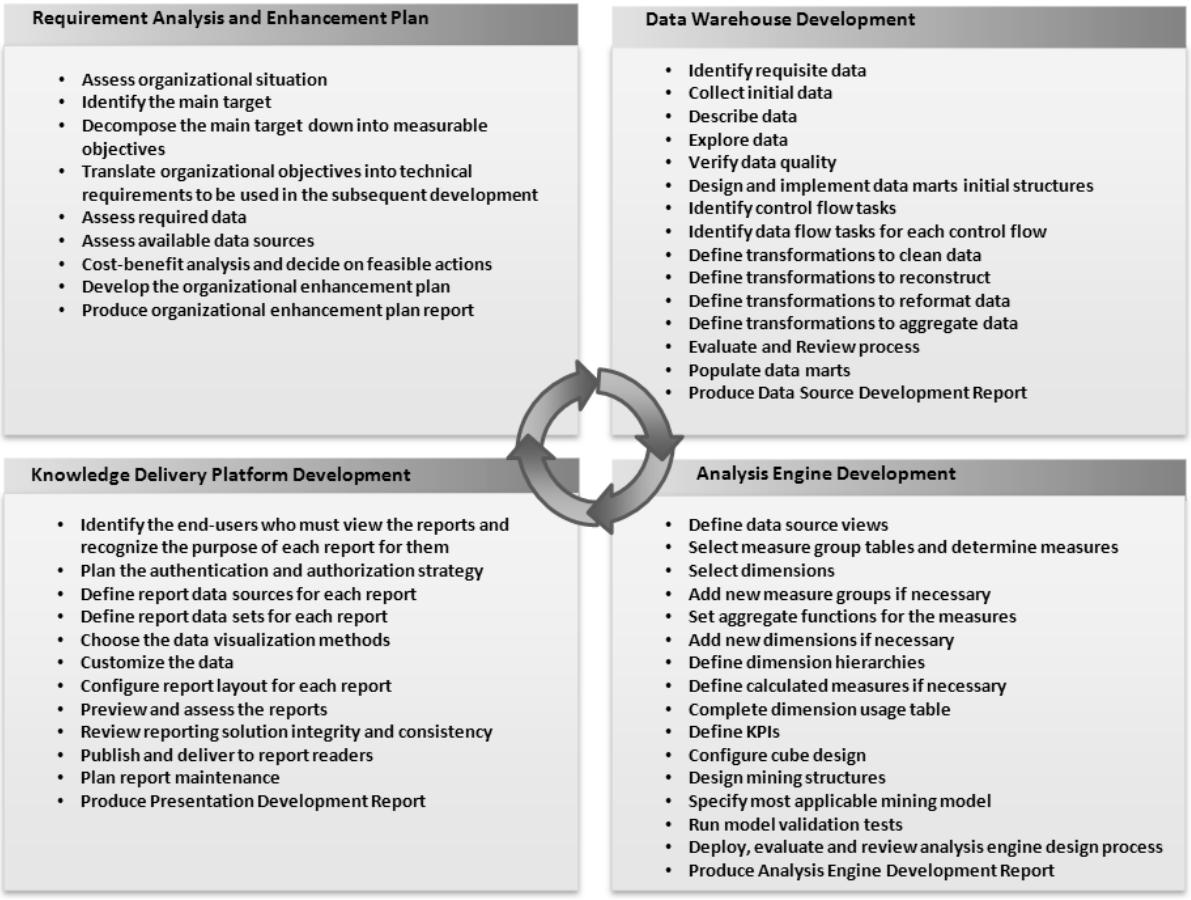

Figure 1. A summary of the new introduced model approach (4M).

Centralized Data Warehouse Development

The next phase of model tries to tackle the problem of organizing the collections of data. These collections may be dispersed or distributed in very different ways and the relations can be vertically and horizontally fragmented. These fragments are usually stored in different geographical sites where they are more regularly accessed. Even in a centralized database this data also may be in different and not appropriate formats. This phase consists of identifying required data, collecting initial data, describing data, exploring data, verifying data quality, designing data mart initial structure, identifying control flow tasks, identifying data flow tasks, cleaning data, reconstructing data, reformatting data, and carrying out aggregation calculations on data (see Figure 1). In the final step, the process would be evaluated and reviewed and then data can be populated to data marts and a related report can be produced.

Analysis Engine Development

The data marts that are developed in previous stage only can store leaf-level values, i.e., the measure values that are in intersection of all of the dimensions of a data mart. To solve the problem, some or all of the

possible data aggregates should be calculated ahead of time and stored within the multidimensional data structures. These multidimensional data structures are the best platform for building multilayer mining structures and applying proactive multilayer mining models. There are some distinct key steps in this phase that are summarized in Figure 1. The figure shows that after defining data sources, measures and measure groups, dimensions and dimension hierarchies, Key Performance Indicators (KPI), partitions, mining structures are designed. Then some model validation tests run to validate the models against the measures of accuracy, reliability, and usefulness. The last steps in this phase include evaluating and reviewing the analysis process, deploying the cube design to analysis server, and then processing the multidimensional structure in the analysis server. The final step is producing an analysis engine development report.

Knowledge Delivery Platform Development

Knowledge delivery platform stage in 5M primarily is about choosing the end-users who are in charge of decision-making and designing appropriate reports for them. Knowledge delivery platform can contain a number of report projects, and each report project in turn can contain a number of reports. A report actually is a piece of art meant to convoy a message. This message changes based on the data that drives it. Each report internally contains two distinct sets of instructions that determine what the report will contain. The first set of instructions is data definition. Data definition controls where the data for the report comes from and what information is to be selected from that data. Data definition instruction set contains two distinct parts: the data source and the dataset. The data source instruction set is needed by the report to gain access to a data source that provides data for the report. When the report is compiled, it uses the data source instructions to gain access to the data source. It then extracts information from the data source into a new format that can be used by the report. This new format is called a dataset. The second set of instructions is about the report layout. This instruction set specifies that which field of data goes into which location in the paper layout. There are some basic steps in knowledge delivery platform development phase that are summarized in Figure 1. Identification of the end-users, identification the purpose of each report, planning the authentication and authorization strategy, defining report data sources and data sets, choosing the data visualization methods, configuring reports layout, previewing and assessing the reports are main steps of this phase. It is also very important to have a plan for report maintenance. This stage includes identifying day-to-day activities that need constant monitoring and developing an efficient monitoring. A report of the process is produced at the end, including the list of components like data sources and dataset queries that can be reused.

Research Methods, Validation, and Evaluation

This section validates 5M by adapting it to a real application as a case study and assessing the results in an empirical study. To validate the results, the model has been implemented and applied to ebusiness discipline and the results are evaluated. Choosing ebusiness to apply the generic model is done based on figures that showed a fast growing rate in the ebusiness branches specially in ecommerce domain. Ecommerce originally was identified as the facilitation of electronic commercial transactions, but in recent years, data mining, data warehousing and data integration modeling techniques (Giordano, 2011), and Business Intelligence (BI) have become parts of its body. The term BI has been defined in different ways and in various contexts. Langit (2009) defined it as effective storage and presentation of key enterprise data so that authorized users can quickly and easily access and interpret it. B. Knight, D. Knight, Jorgensen, LeBlanc, & Davis (2010) considered it as a term

that encompasses the process of getting data out of the disparate systems and into a unified model, so it can be used to analyze, report, and mine the data. The approach in this adaptation is more business-driven, rather than current software-driven ones (Fernandez, 2011). Actually traditional views of business activities, like that of Kotler and Kelly (2006) have mainly focused on the physical and human aspects of the organization. The information view of them started getting conceptualized with contributions from Holland and Naude (2004), Jayachandran, Sharma, Kaufan, and Raman (2005) and Kumar Kar, Kumar Pani, and Kumar De (2010) by emphasizing on marketing activities. The instance implementation of the new model is carried out for verification purposes by using Microsoft Visual Studio 2010 and SQL Server 2008. EBAF serves as a best practice blueprint for all kind of enterprises wishing to capture business intelligence and enhance their CRM. This all will be done through creating an architecture that not only provides useful information, but also provides organizational insight. A simplified overview of EBAF is provided by Figure 2.

In Figure 2, a five-stage conversion model including awareness, contact, engagement, conversion, and retention phases is proposed to help identify mid-level mining structures in business domain. The left side of the figure shows how EBAF classifies the people to six main state groups, target audience, aware target audience, unique visitors, active unique visitors, actors, and finally the clients. The right side summarizes the influencers that affect awareness, contact, engagement, conversion, and retention efficiency factors (Pesaran Behbahani, Khaddaj, & Choudhury, 2011). These influencers shape multilayer mining structures.

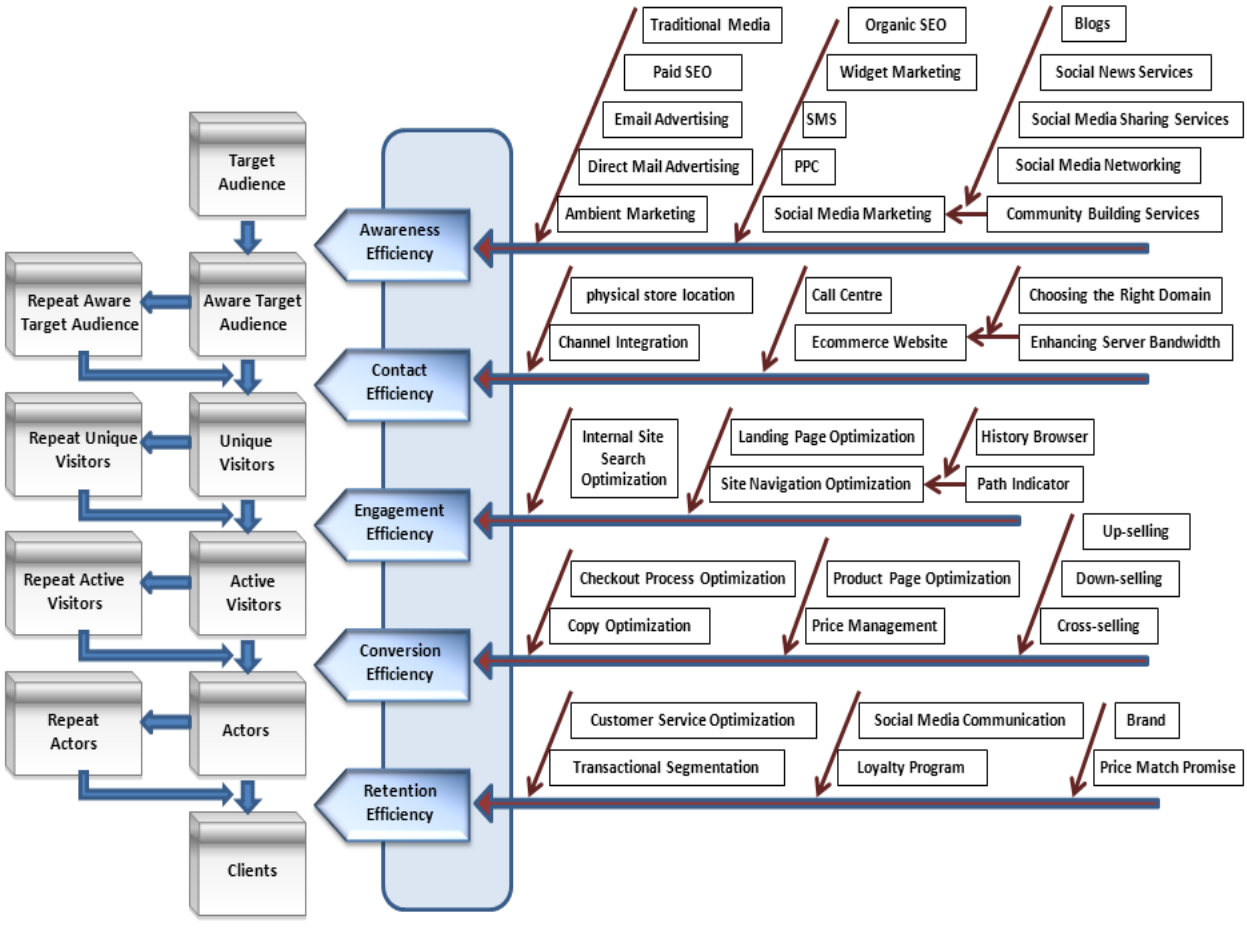

Figure 2. Adapting the new model to ebusiness discipline produces a conversion model.

We call this diagram “EBAF Conversion Model”, because it illustrates how the target audience can be converted to aware the target audience, unique visitors, active visitors, actors, and finally permanent clients. Because the right hand side of the diagram also shows a summary of the main activities that should be considered in designing mining structure layers, now we are almost ready to start data warehouse development phase.

Following the roadmap provided by 5M, results in EBAF analysis core in a shape of a trilateral analysis server. It integrates enterprise multilayer KPI analysis, multilayer multidimensional analysis, and multilayer data mining analysis. Figure 3 provides an overview of characteristics of this core. EBAF multilayer enterprise KPI analysis offers a presentational business insight of critical measures. EBAF multidimensional cube analysis delivers an interactive, investigative, and exploratory business perception. EBAF multilayer data mining analysis has a proactive role to provide discovery business vision.

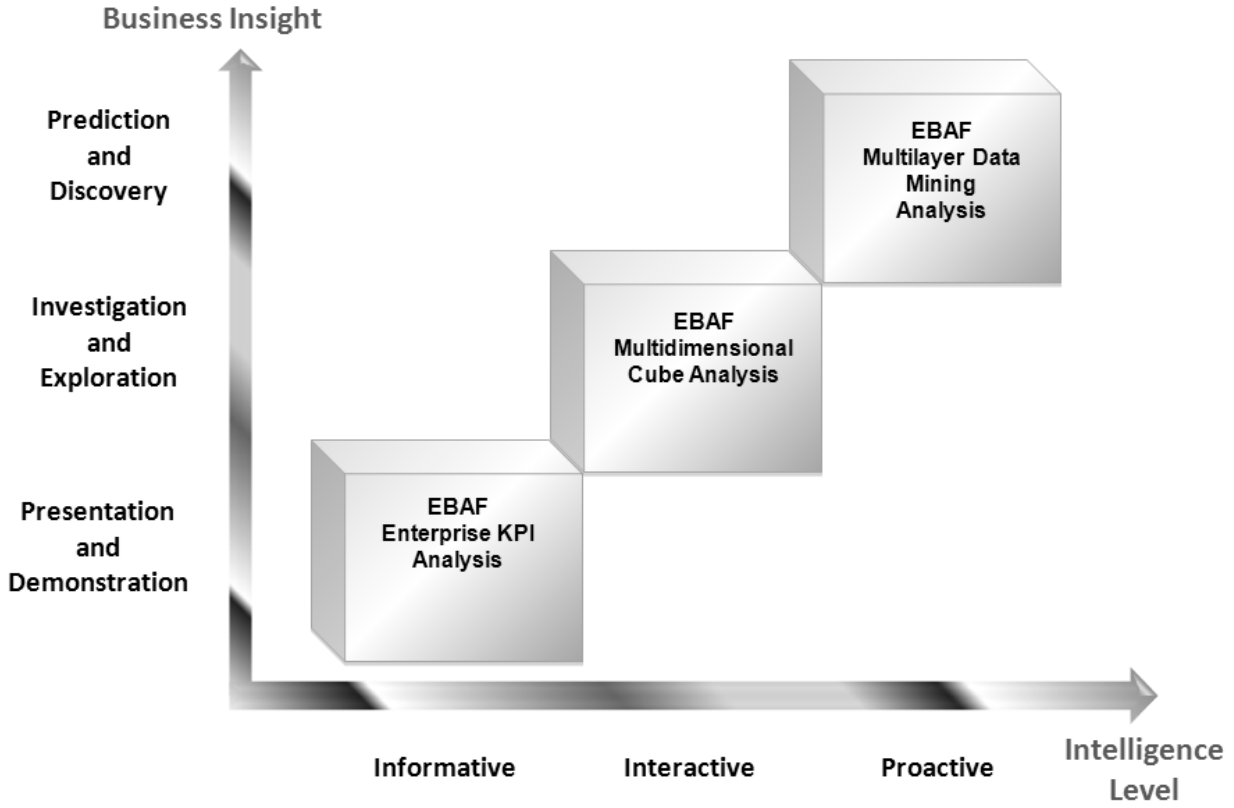

Figure 3. Analysis core characteristics at the third stage of new model.

Model Accuracy Testing and Verification

Each software project represents a complex system with its own life cycle, starting with the phases of planning and designing up to the implementation, testing, and validation. This section is about testing EBAF results and the process of assessing how the EBAF models perform against real data. Testing is an essential part of the design life-cycle of any software application. The literature on software testing and validation is huge and it includes detailed discussion of different approaches to it. But there is a big difference between testing a generic software system developed based on 5M EBAF and a normal software product. A system developed

from 5M EBAF model has several aspects. Therefore testing such a system is also multifaceted issue. One of the main features of such a system is that its multidimensional structure is constructed upon a data warehouse. Therefor it seems that data warehouse testing can be considered at least as a fundamental procedure. There are many general system testing activities that can be used in testing the underneath data warehouse of a 5M EBAF product, e.g., data backup testing, data recovery testing, on-line time response testing, and data accessibility testing. But still there are big differences between testing a system based on data warehousing and generic software systems. Golfarelli and Rizzi (2009) have spotted the differences between testing data warehouse systems and generic software systems or even transactional systems as:

- Software testing is predominantly focused on program code, while here testing is directed regarding available data and required information. As a matter of fact, the key to data warehouse testing is to know the data and what the answers to user queries are supposed to be;
- Data warehouse testing involves a huge data volume, which significantly impacts performance and productivity;
- Data warehouse testing has a broader scope than software testing because it focuses on the correctness and usefulness of the information delivered to users. In fact, data validation is one of the main goals of data warehouse testing;
- Though a generic software system may have a large number of different use scenarios, the valid combinations of those scenarios are limited. On the other hand, data warehouse systems are aimed at supporting any views of data, so the possible combinations are virtually unlimited and cannot be fully tested;
- While most testing activities are carried out before deployment in generic software systems, data warehouse testing activities still go on after system release;
- Typical software development projects are self-contained. Data warehousing projects never really come to an end. It is very difficult to anticipate future requirements for the decision-making process, so only a few requirements can be stated from the beginning. Besides, it is almost impossible to predict all the possible types of errors that will be encountered in real operational data. For this reason, regression testing is inherently involved.

Like any other software system, different types of tests can be devised for data warehouse systems. Regression test checks that the system still functions correctly after a change has occurred. Regression test is important for data warehouse systems because of their ever-evolving nature. Unit test is a white-box test performed on each individual component considered in isolation from the others. Integration test is a black-box test where the system is tested in its entirety. The peculiar characteristics of data warehouse testing and the complexity of data warehouse projects ask for a deep revision and contextualization of these test types, aimed in particular at emphasizing the relationships between testing activities on the one side, design phases and project documentation on the other side (Golfarelli & Rizzi, 2011). According to 5M approach, the issues about data warehouse verification are mainly considered in the third phase of 5M, regarding the Tanuska framework (Tanuska, Moravcik, Vazan, & Miksa, 2009).

It is also important to consider the issue for proactive aspect of the framework. Needless to say, it is important that the researcher validate these mining models generated in the framework by understanding their quality and characteristics before deploying them into a production environment. Here model accuracy testing and evaluation serve two purposes. The first purpose is the prediction of how well the final model will work in the future or even whether it should be used at all. The second purpose is to find the best model that maintains EBAF objectives. The approach to model validation in this research is partitioning data into training and testing

sets. This approach is an established method and is widely used by practitioners. In this approach, some portion of data from the training data set is reserved for testing. In Figure 4, the lift chart graphically represents the improvement that the models provide when compared against a random guess for 24.75% population percentage.

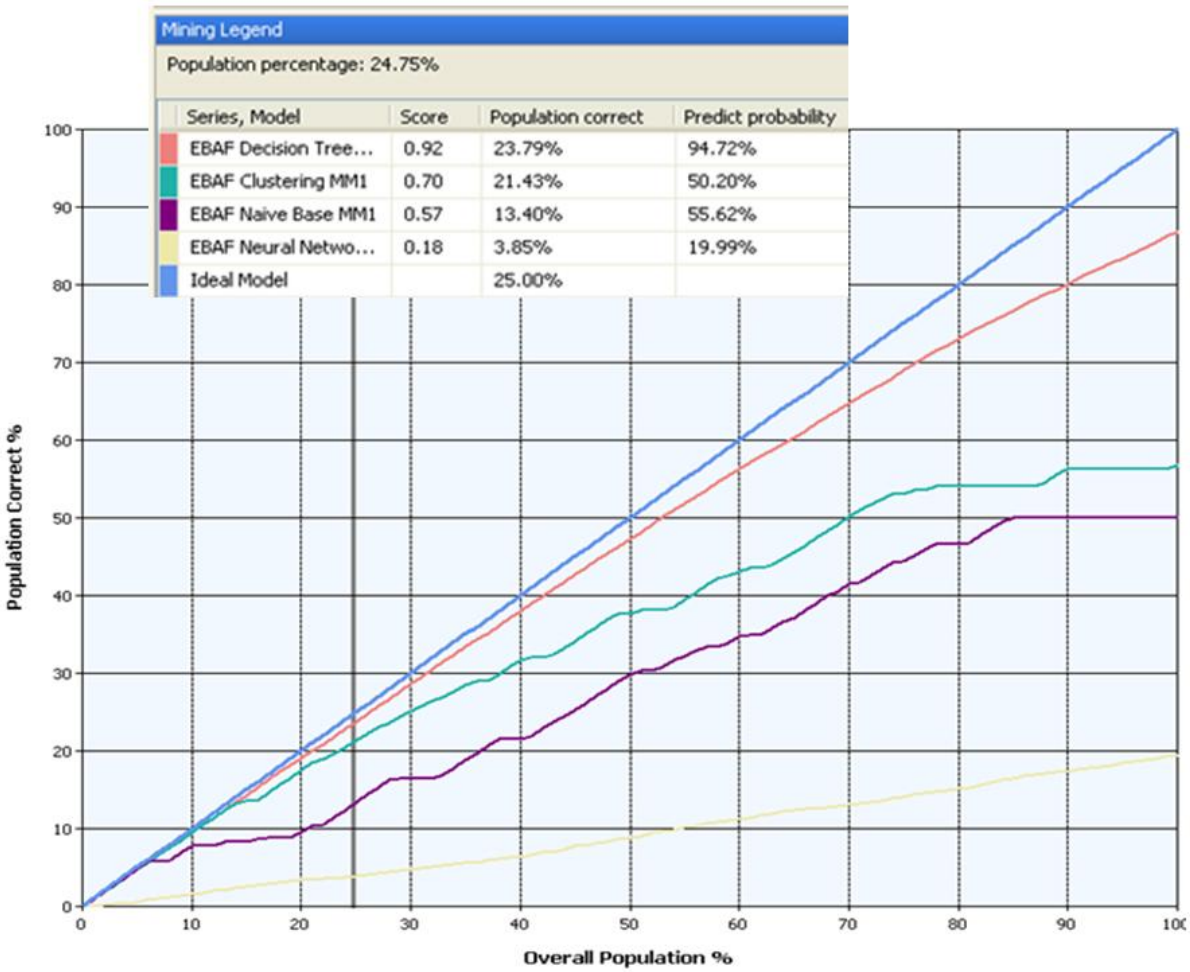

Figure 4. Lift chart shows the significant improvement against a random guess for 24.75% population percentage.

There are five lines in this graph. The bottommost line in the chart is the result of the neural network model that does not have any effect on the prediction improvement. Therefore the line can also be considered as a blind guess. The straight uppermost line is for the ideal model that each model tries to get closer to it. The three lines below the ideal model are correspondently related to "EBAF Decision Tree MM1", "EBAF Clustering MM1", and "EBAF Naïve Base MM1" models.

Figure 5 helps to link this chart to classification matrix and double check the results. The classification matrix reveals that the correct population for whole overall population is Total Correct Population / Total Population = (562 + 400 + 163 + 777 + 252) / (841 + 830 + 725 + 885 + 1031) = (2154 / 4312) = 49.95% for "EBAF Naïve Base MM1". By selecting 99% of population, the lift chart shows that population correct figure is approaching to this value.

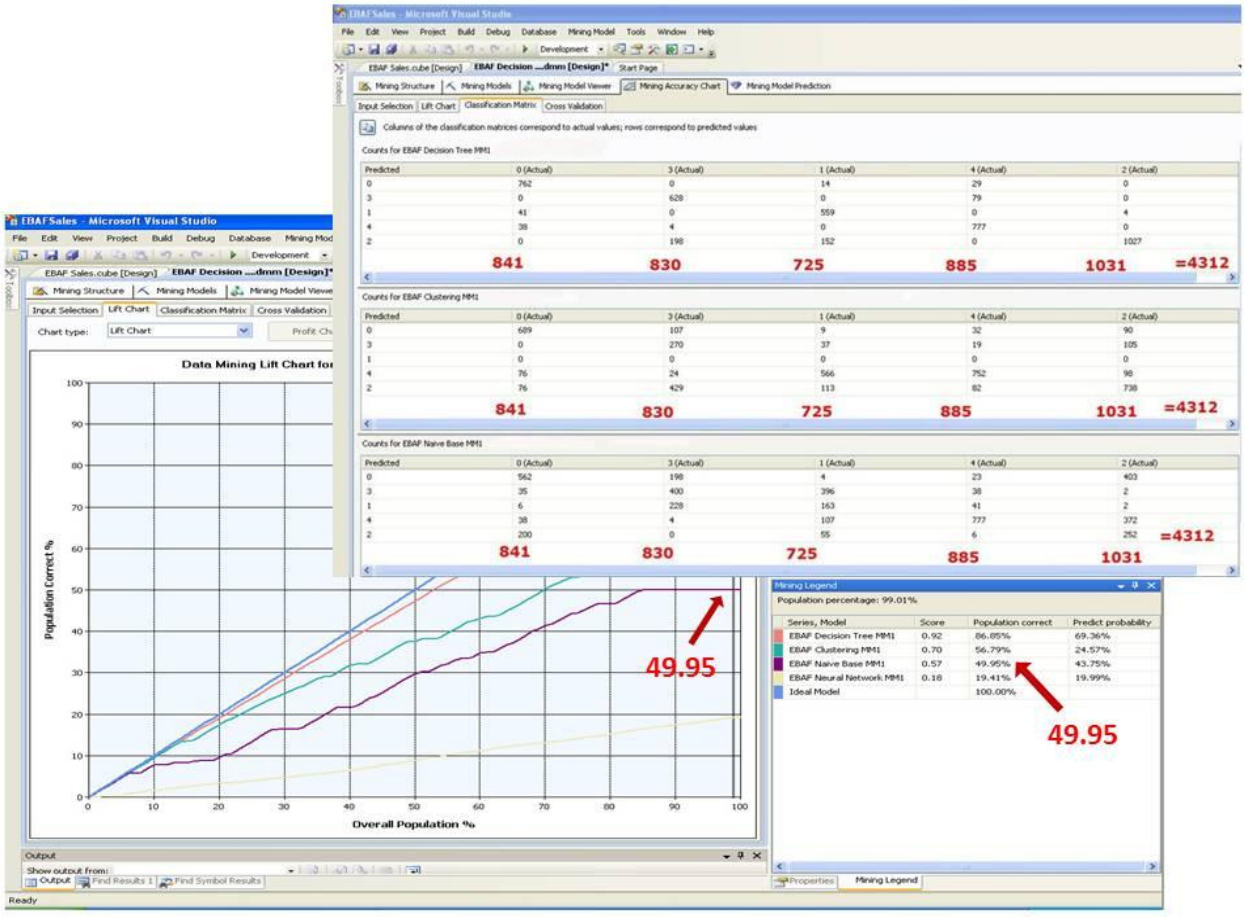

Figure 5. The figures in classification matrix verify the lift chart results

The value of the models in increasing the turnover may be better understood using the profit chart. Like a lift chart, a profit chart can be used to compare multiple models. To get a better understanding of the underlying explanations in resulted figures, the correspondent lift chart must be linked to the resulted profit chart. Figure 6 focuses on calculating the profit gained by the EBAF Decision Tree MM1 model. The figure yields £354,300 revenue by multiplication of population, population correct, and revenue per individual, i.e., 50000 × 47.24 × 15. Taking out the cost that is 80,000, will result in a £274,000 net profit.

The results are calculated based on assuming that revenue per individual is £15, individual cost is £3, and there is also a £5,000 fixed cost. The total population is assumed to be 50,000. The chart shows that applying "EBAF Neural Network MM1", does not outcome in any more profit. It reveals that applying EBAF Decision Tree MM1, EBAF Clustering MM1, and EBAF Naïve Base MM1, respectively result in £274,302, £201,771, and £143,504 more profit. The chart clearly proves that EBAF Clustering MM1 and EBAF Naïve Base MM1 cannot produce any more profit if we try to cover more than 75% of population:

Fixed Cost = 5,000

Individual Cost = 3

Revenue per Individual = 15

Population Percentage (Chosen by grey line) = 50%

Population Correct% (Extracted From Lift Chart) = 47.24%

Revenue = (50,000 × 0.4724 × 15) = 354,300
Cost = (50,000 × 0.5 × 3 + 5,000) = 80,000
Profit gained by EBAF Decision Tree MM1 model = (354,300 – 80,000) = 274,000.

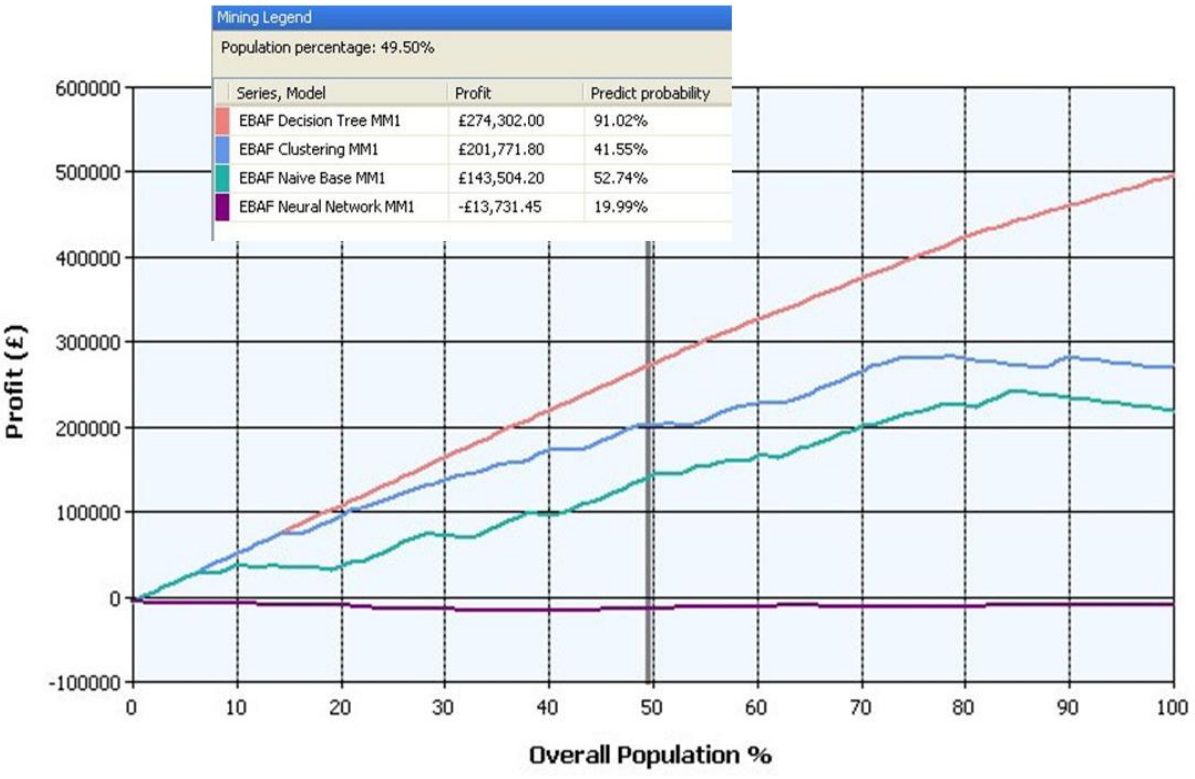

Figure 6. Calculating the profit gained by EBAF Decision Tree MM1 model covering 50% overall population.

Conclusions

Distributed databases that use fragmented data storage have practical problems in hosting existing data engineering models. A novel multidimensional approach is presented to solve the problem. The model is known as 5M. The model is evaluated and validated by adapting to ebusiness as a case study and test results show significant enhancement in decision-making process. The resulted ebusiness framework seems to be very successful in providing higher levels of business intelligence for the enterprises utilizing ecommerce as their main transactional model. There is also a debate about the expense of employing the model in enterprises. Though the model provides a solid insight for organization management, it can be costly due to providing the required data sources for multilayer influencers and developing multilayer mining structures. Therefore a simplified version of the model is also provided. This shortened version of the model can be called 4M and seems suitable for low-budget projects. The emphasis on building multilayer mining structures in 5M is not seen in 4M. EBAF is the name of a case study evaluates and validates the new model. EBAF provides a roadmap to gain incredible competitive advantages in ecommerce marketplace. EBAF offers the key ability to respond with more agility to changing business conditions using effective and corresponding actions. EBAF analysis core utilizes the EBAF Conversion Model constituents to create multilayer mining structures and finally enhances and optimizes the conversion model's efficiency factors.

References

Azevedo, A., & Santos, M. F. (2008). *KDD, SEMMA and CRISP-DM: A parallel overview*. In Proceedings of the *IADIS European Conference on Data Mining* (pp. 182-185). Amsterdam.

Chung-Shing, L. (2001). An analytical framework for evaluating e-commerce business models and strategies. *Internet Research, 11*(4), 349-359.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). *From data mining to knowledge discovery in database*. *American Association for Artificial Intelligence, 17*(3), 37-54.

Fernandez, M. T. (2011). Business strategy model. *International Journal of Innovation, Management and Technology, 2*(4), 301-308.

Giordano, A. D. (2011). *Data integration: Blueprint and modeling techniques for a scalable and sustainable*. Boston: IBM Press.

Golfarelli, M., & Rizzi, S. (2009). A comprehensive approach to data warehouse testing. Proceedings of the *ACM Twelfth International Workshop on Data Warehousing and OLAP* (pp. 17-24). New York.

Golfarelli, M., & Rizzi, S. (2011). Data warehouse testing. *International Journal of Data Warehousing and Mining, 7*(2), 26-43.

Holland, P. C., & Naude, P. (2004). The metamorphosis of marketing into nn information-handling problem. *The Journal of Business & Industrial Marketing, 19*(3), 167-178.

Jayachandran, S., Sharma, S., Kaufan, P., & Raman, P. (2005). The role of relational information processes and technology use in customer relationship management. *Journal of Marketing, 69*, 177-192.

Knight, B., Knight, D., Jorgensen, A., LeBlanc, P., & Davis, M. (2010). *Knight's microsoft business intelligence 24 hour trainer*. Indianapolis, Indiana: Wiley Publishing, Inc..

Kotler, P., & Keller, K. L. (2006). *Marketing management* (12th Ed.). New York: Prentice hall.

Kumar Kar, A., Kumar Pani, A., & Kumar De, S. (2010). A study on using business intelligence for improving marketing efforts. *Business Intelligence Journal, 3*(2), 141-150.

Langit, L. (2009). *Smart business intelligence solutions with Microsoft SQL server 2008*. Redmond, W.A.: Microsoft Press.

Pesaran Behbahani, M. (2012). A business intelligence framework to provide performance management through a holistic data mining view. Proceedings from the *UK Academy for Information System 17th Annual International Conference 2012*. Oxford, UK.

Pesaran Behbahani, M., Khaddaj, S., & Choudhury, I. (2011). A multilayer data mining approach to an optimized ebusiness analytics framework. Proceedings of *International Conference on Economics Development and Research* (pp. 66-71). Dubai, UAE.

Pesaran Behbahani, M., Khaddaj, S., & Choudhury, I. (2012). Enhancing organizational performance through a new proactive multilayer data mining methodology: An ecommerce case study. *International Journal of Innovation, Management and Technology, 3*(5), 600-607.

Shearer, C. (2000). The CRISP-DM model: The new blueprint for data mining. *Journal of Data Warehousing, 5*(4), 13-22.

Shim, J. P., Warkentin, M., Courtney, J., Power, D., Sharda, R., & Cjroster, C. (2002). Past, present, and future of decision support technology. *Journal of Decision Support Systems—Special Issue: Decision Support Systems: Directions for the Next Decade, 33*(2), 111-126.

Tanuska, P., Moravcik, O., Vazan, P., & Miksa, F. (2009). The proposal of data warehouse testing. Proceedings of the *20th Central European Conference on Information and Intelligent Systems* (pp. 7-11). Paulinska, Slovakia.