Business Intelligence: Optimization techniques for Decision Making
Mary Jeyanthi Prem * and M.Karnan *

*VELS University, Pallavaram, Chennai, Tamilnadu, India.
*Tamil Nadu College of Engineering, Coimbatore, Tamil Nadu, India.

Abstract — Business intelligence is a broad category of applications and technologies for gathering, providing access to, and analyzing data for the purpose of helping enterprise users enable the better optimal business decisions. Business Intelligence (BI) is about getting the right information, to the right decision makers, at the right time. Dynamic decision making is effectively dealt with through an instinctive approach, and require precisely based on Analytical methodologies and Mathematical models. This paper describes the basic knowledge of Business intelligence and suitable optimization techniques for the optimal and dynamic decision making in the current business world. This paper aims at analyzing Business Intelligence Systems (BI) in the context of opportunities for improving decision-making in a contemporary organization.

Keywords: Business Intelligence, Decision making, Optimization techniques , Genetic Algorithm, Ant Colony Optimization.

1. INTRODUCTION

Business intelligence, or BI, is an umbrella term that refers to a variety of software applications used to analyze an organization’s raw data for intelligent decision making for business success. BI as a discipline is made up of several related activities, including data mining, online analytical processing, querying and reporting. Techniques include multidimensional analyses, mathematical projection, modeling, ad-hoc queries and 'canned' reporting.

BI leads to:

- fact-based decision making
- “single version of the truth”

The main purpose of BI systems is to provide Decision makers with tools and methodologies that allow them to make effective and timely decisions.

Effective decisions: The application of rigorous analytical methods allows decision makers to rely on information and knowledge which are more dependable. As a result, they are able to make better decisions and devise action plans that allow their objectives to be reached in a more effective way. Indeed, turning to formal analytical method forces decision makers to explicitly describe both the criteria for evaluating alternative choices and the mechanisms regulating the problem under investigation. Furthermore, the ensuing in-depth examination and thought lead to deeper awareness and comprehension of the underlying logic of the decision making process.

Timely decisions: Enterprises operate in economic environments characterized by growing levels of competition and high dynamism. As a consequence, the ability to rapidly react to the actions of competitors and to new market conditions is a critical factor in the success or even the survival of a company. If the decision makers can rely on a business intelligence system facilitating their activity, we can expect the overall quality of the decision making process will be greatly improved.

With the help of mathematical models and algorithms, it is actually possible to analyze a large number of alternative actions, achieve more accurate conclusions and reach effective and timely decisions. We may conclude that the major advantage deriving from the adoption of the business intelligence system is found in the increased effectiveness of the decision-making process.
2. LITERATURE REVIEW:

In a 1958 article, IBM researcher Hans Peter Luhn\(^4\) used the term business intelligence. He defined intelligence as: “the ability to apprehend the interrelationships of presented facts in such a way as to guide action towards a desired goal.”

Business intelligence as it is understood today is said to have evolved from the decision support systems which began in the 1960s and developed throughout the mid-80s. Decision Support System (DSS) originated in the computer-aided models created to assist with decision making and planning. From DSS, data warehouses, Executive Information Systems, OLAP and business intelligence came into focus beginning in the late 80s.

In 1989 Howard Dresner\(^5\)(later a Gartner Group analyst) proposed "business intelligence" as an umbrella term to describe "concepts and methods to improve business decision making by using fact-based support systems." It was not until the late 1990s that this usage was widespread.

Followed by these studies, a great number of researchers involved in examining to create a huge amount of valuable information in the form of e.g. e-mails, memos, notes from call-centers, news, user groups, chats, reports, web-pages, presentations, image-files, video-files, marketing material and news etc... However, organizations often only use these documents once.

According to Merrill Lynch, more than 85 percent of all business information exists as the before-mentioned information types. These information types are called either semi-structured or unstructured data. The management of semi-structured data is recognized as a major unsolved problem in the information technology industry. According to projections from Gartner (2003)\(^6\), white collar workers will spend anywhere from 30 to 40 percent of their time searching, finding and assessing unstructured data.

2.1 DEMERITS:

There are several problems/challenges when trying to develop BI with semi-structured data, and according to (Inmon & Nesavich, 2008)\(^15\) some of those are:

1. Physically accessing unstructured textual data – unstructured data is stored in a huge variety of formats.

2. Terminology – Among researchers and analysts, there is a need to develop a standardized terminology.

3. Volume of data – As stated earlier, up to 85% of all data exists as semi-structured data. Couple that with the need for word-to-word and semantic analysis.

4. Searchability of unstructured textual data – A simple search on some data, e.g. apple, results in links where there is a reference to that precise search term. (Inmon & Nesavich, 2008) gives an example: “a search is made on the term felony. In a simple search, the term felony is used, and everywhere there is a reference to felony, a hit to an unstructured document is made. But a simple search is crude. It does not find references to crime, arson, murder, embezzlement, and such, even though these crimes are types of felonies.”

5. Existing business intelligence tools suffer from a lack of analysis and visualization capabilities and traditional result list display by search engines often overwhelms business analysts with irrelevant information.

6. They do not provide sufficient support in dealing with the upcoming challenges in BI industry, such as real-time decision making.

7. There is NO standardized enterprise-wide BI Methodologies.
3. METHODOLOGY:

The procedure or procedures used to make a system or design as effective or functional as possible, especially the mathematical techniques involved. The approaches to optimizing systems are varied and depend on the type of system involved, but the goal of all optimization procedures is to obtain the best results possible; subject to restrictions or constraints that are imposed.

The first step in modern optimization is to obtain a mathematical description of the process or the system to be optimized. A mathematical model of the process or system is then formed on the basis of this description. Depending on the application, the model complexity can range from very simple to extremely complex. System models used in optimization are classified in various ways, such as linear versus nonlinear, static versus dynamic.

Certain models lend themselves to rapid and well-developed solution algorithms, whereas other models may not. When choosing between equally valid models, therefore, those that are cast in standard optimization forms are to be preferred. That is:

3.1 Genetic Algorithm
3.2 Ant colony optimization

3.1 Genetic Algorithm (GA):

A heuristic search technique used in computing and Artificial Intelligence to find optimized solutions to search problems using techniques inspired by evolutionary biology: mutation, selection, reproduction [inheritance] and recombination. This search technique used in computing to find exact or approximate solutions to optimization and search problems.

Algorithm 1: Genetic Algorithm

<table>
<thead>
<tr>
<th>STEP 1</th>
<th>[Start] Generate random population of n chromosomes (suitable solutions for the problem)</th>
</tr>
</thead>
<tbody>
<tr>
<td>STEP 2</td>
<td>[Fitness] Evaluate the fitness f(x) of each chromosome x in the population</td>
</tr>
<tr>
<td>STEP 3</td>
<td>[New population] Create a new population by repeating following steps until the new population is complete</td>
</tr>
<tr>
<td>STEP 3.1</td>
<td>[Selection] Select two parent chromosomes from a population according to their fitness (the better fitness, the bigger chance to be selected)</td>
</tr>
<tr>
<td>STEP 3.2</td>
<td>[Crossover] With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.</td>
</tr>
<tr>
<td>STEP 3.3</td>
<td>[Mutation] With a mutation probability mutate new offspring at each locus (position in chromosome).</td>
</tr>
<tr>
<td>STEP 3.4</td>
<td>[Accepting] Place new offspring in a new population</td>
</tr>
<tr>
<td>STEP 4</td>
<td>[Replace] Use new generated population for a further run of algorithm</td>
</tr>
<tr>
<td>STEP 5</td>
<td>[Test] If the end condition is satisfied, stop, and return the best solution in current population</td>
</tr>
<tr>
<td>STEP 6</td>
<td>[Loop] Go to STEP 2</td>
</tr>
</tbody>
</table>
Each iteration of this process is called a generation. A GA is typically iterated for anywhere from 50 to 500 or more generations. The entire set of generations is called a run. At the end of a run there are often one or more highly fit chromosomes in the population. Since randomness plays a large role in each run, two runs with different random-number seeds will generally produce different detailed behaviors. GA researchers often report statistics (such as the best fitness found in a run and the generation at which the individual with that best fitness was discovered) averaged over many different runs of the GA on the same problem.

**MERITS OF GENETIC ALGORITHM:**

1. Good for “noisy” environments
2. Always an answer; answer gets better with time
3. Inherently parallel; easily distributed
4. Many ways to speed up and improve a GA-based application as knowledge about problem domain is gained
5. Easy to exploit previous or alternate solutions
6. Flexible building blocks for hybrid applications
7. Substantial history and range of use

**3.2 ANT COLONY OPTIMIZATION ALGORITHM (ACO):**

This algorithm is a probabilistic technique for solving computational problems which can be reduced to finding good paths through graphs. The original idea has since diversified to solve a wider class of numerical problems, and as a result, several problems have emerged, drawing on various aspects of the behavior of ants. A short path, by comparison, gets marched over faster, and thus the pheromone density remains high.

Ant Colony Optimization (ACO) studies artificial systems that take inspiration from the behavior of real ant colonies and which are used to solve discrete optimization problems. In 1999, the Ant Colony Optimization metaheuristic was defined by Marco Dorigo, Di Caro and Gambardella.
The ants move from vertex to vertex along the edges of the construction graph exploiting information provided by the pheromone values and in this way incrementally building a solution. Additionally, the ants deposit a certain amount of pheromone on the components, that is, either on the vertices or on the edges that they traverse. The amount $\Delta \tau$ of pheromone deposited may depend on the quality of the solution found. Subsequent ants utilize the pheromone information as a guide towards more promising regions of the search space.

The ACO metaheuristic is:

```plaintext
Set parameters, initialize pheromone trails
SCHEDULE_ACTIVITIES
    ConstructAntSolutions
    DaemonActions  {optional}
    UpdatePheromones
END_SCHEDULE_ACTIVITIES
```

Figure 3: ACO Algorithm

The metaheuristic consists of an initialization step and of three algorithmic components whose activation is regulated by the `SCHEDULE_ACTIVITIES` construct. This construct is repeated until a termination criterion is met. Typical criteria are a maximum number of iterations or a maximum CPU time.

The `SCHEDULE_ACTIVITIES` construct does not specify how the three algorithmic components are scheduled and synchronized. In most applications of ACO to NP-hard problems however, the three algorithmic components undergo a loop that consists in

(i) the construction of solutions by all ants

(ii) the (optional) improvement of these solution via the use of a local search algorithm, and

(iii) the update of the pheromones.
Ants (blind) navigate from nest to food source. Shortest path is discovered via pheromone trails each ant moves at random; pheromone is deposited on path; ants detect lead ant’s path, inclined to follow and more pheromone on path increases probability of path being followed.

1 NP-hard problems - non-deterministic polynomial-time hard), in computational complexity theory, is a class of problems that are, informally, "at least as hard as the hardest problems in NP"
ACO System:

Algorithm 2: Ant Colony Optimization Algorithm

| STEP 1: Starting node selected at random |
| STEP 2: Path selected at random |
| STEP 2.1: Based on amount of “trail” present on possible paths from starting node |
| STEP 2.2: Higher probability for paths with more “trail” |
| STEP 3: Ant reaches next node, selects next path |
| STEP 4: Continues until reaches STEP 1. |
| STEP 5: Finished “tour” is a solution |
| STEP 6: A completed tour is analyzed for optimality |
| STEP 7: “Trail” amount adjusted to favor better solutions |
| STEP 7.1: Better solutions receive more trail |
| STEP 7.2: Worse solutions receive less trail. |
| STEP 8: Higher probability of ant selecting path that is part of a better-performing tour |
| STEP 9: New cycle is performed |
| STEP 10: Repeated until most ants select the same tour on every cycle (convergence to solution) |

ALGORITHM IN PSEUDO CODE:

- Initialize Trail
- Do While (Stopping Criteria Not Satisfied) – Cycle Loop
  - Do Until (Each Ant Completes a Tour) – Tour Loop
    - Local Trail Update
    - End Do
    - Analyze Tours
    - Global Trail Update
  - End Do

3.2. IMPLEMENTATION:

The traditional linear programming software continues to be refined in both simplex method and interior point algorithms. The emphasis is on taking advantage of problem characteristics to reduce the problem size or to speed up a specific algorithmic step. The result is the ability to solve really large problems. It has also enabled the modelers to consider uncertainty in the decision situation through stochastic programming with recourse type approaches.
This level contains tools for presenting and analyzing data from previous levels. There are many graphical tools for building friendly and flexible presentations like: reports, graphics, charts builders, web pages which can be integrated into an organizational portal.

The implementation of the software tool to have the ability to provide the Optimal data sets with effective and accuracy of data and allow managers to view data in different perspective, to drill-down and roll-up to aggregate levels, to navigate and on-line query data sets in order to discover new factors that affect business process and also to anticipate and forecast changes inside and outside the organization.

Figure 8: Flow Chart Showing Genetic Programming modeling Process
4. EXPERIMENT & RESULTS:

A key step in the formulation of any optimization problem is the assignment of performance measures that are to be optimized. The success of any optimization result is critically dependent on the selection of meaningful performance measures. In many cases, the actual computational solution approach is secondary. Ways in which multiple performance measures can be incorporated in the optimization process are varied.

Table 1: Comparison of ERP Reports Vs Metaheuristic Reports

<table>
<thead>
<tr>
<th>CHARACTERISTICS</th>
<th>ERP REPORTS</th>
<th>METAHEURISTICS REPORTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objectives</td>
<td>Analyse indicators that measure current and internal activities or daily reports</td>
<td>Process Optimization, analyse key performance indicators, forecast internal and external data, internal and external focus.</td>
</tr>
<tr>
<td>Level of decisions</td>
<td>Operational / Medium</td>
<td>Strategic / High</td>
</tr>
<tr>
<td>User involved</td>
<td>Operational level of management</td>
<td>Executives, strategic level of management</td>
</tr>
<tr>
<td>Data Management</td>
<td>Relational databases Datawarehouse</td>
<td>Datawarehouse OLAP DataMining</td>
</tr>
<tr>
<td>Number of records/transaction</td>
<td>Limited</td>
<td>Huge</td>
</tr>
<tr>
<td>Data Orientation</td>
<td>Record</td>
<td>Cube</td>
</tr>
<tr>
<td>Level of detail</td>
<td>Detailed, summarised, pre-aggregate</td>
<td>Aggregate</td>
</tr>
<tr>
<td>Age of data</td>
<td>Current</td>
<td>Historical/current / prospective</td>
</tr>
</tbody>
</table>

5. PERFORMANCE AND ANALYSIS:

In the Banking & finance sectors, the Customer profitability analysis. Determinate the overall profitability of individual customer, current and long term, provide the basis for high-profit sales and relationship banking, maximize sales to high-value customers, reduce costs to low-value customers, provide the means to maximize profitability of new products and services. Establish patterns of credit problem progression by customers class and type, warn customers to avoid credit problems, to manage credit limits, evaluate of the bank’s credit portfolio, reduce credit losses. Improve customer service and account selling, facilitate cross selling, improve customer support, strengthen customer loyalty.

Efforts undertaken to develop BI systems have resulted in many business solutions that allow for effective support of manager’s work. Practice shows that the most significant business effects are obtained while using the following analyses offered by the BI systems:

- Analysis that supports cross selling and up selling - Market Basket Analysis provides knowledge on what kind of services and products should be sold together in sets or which set should be recommended to a particular customer. Using classification models to select customers who are the most susceptible to a particular offer is another practical application of the discussed solution. It allows to direct
marketing activities correctly and – as a result – to reduce costs of the campaign while simultaneously increasing its effectiveness.

• Customer segmentation and profiling - it is based on grouping customers in some homogenous segments. BI systems enable both descriptive and predictive segmentation. Within descriptive segmentation the following segmentations are carried out: demographic segmentation (on the basis of the data including customer’s income, age, sex, education, marital status, ethnic group, religion, etc.); behavioral segmentation (on the basis of the data including frequency of shopping, amount and sort of purchased products, etc.); and motivational segmentation (on the basis of variables that describe reasons of customers’ purchases – this kind of data usually come from questionnaires and surveys carried out).

• Analysis of parameters importance - The Bivariate statistical analysis, stepwise regression algorithm or artificial neuronal networks are mainly used in this case.

• Survival time analysis - The analysis describes a distribution of survival time for individuals of a given population, monitors strength of other parameters impact on the expected survival time, and additionally, it enables to compare distributions of survival time between different sub-populations. Taking advantage of this method, a company may be given an invaluable insight into customer behavior and find some ways to prolong customer’s survival time.

• Logistics optimizations – Employing advanced data mining techniques, it is possible to show the best available solution for actual and complex optimization problems. Quality of such solutions is usually much higher than the quality offered by traditional solutions of optimization methods.

• Forecasting of strategic business processes development - Abilities to understand and forecast development of strategic business processes make up a foundation of the correct planning of any business activity. That is why, modeling of multidimensional forecasts based on historical, present and anticipated data is so important. Analyses of time series make it possible to identify and analyse hidden trends and fluctuations (e.g. in marketing data or sales data).

• Credit scoring models enable to determine financial risk that is related to particular customers. It is based on the data that come from application forms provided by a customer subject to analysis. Appropriate dealing with customers who are characterized by high risk of stopping payments makes it possible to reduce losses effectively. Credit scoring finds its application in, banking (cash loans, assessment and tolerance of late payments) and in many other sectors related. Correct selection of the models depends on the analysis objective and specifics of the analyzed data:

• application scoring – used in case of new customers; information on them is available only on the basis of the completed application forms;

• behavioral scoring – paying attention to additional information on customers’ track records; it predicts customers’ future behavior; and

• profit scoring – expanding of the basic scoring model; it pays attention not only to probability of paying credits back by customers, but also helps to assess what sort of profit may be expected as a result of co-operation with a particular customer; it is a more sophisticated model because it considers several additional economic factors.

6. CONCLUSION:

Business intelligence need to provide us with feedback information that can be used to evaluate a decision. It can provide that foundational and feedback information. Key Performance Indicators (KPIs) are highly summarized measures designed to quickly relay the status of that measure. They usually reflect the most vital aspects of the organization. By bringing discipline to strategic financial modeling, facilitating the world wide operational planning and forecasting, and linking strategies with operations. By letting management, finance, and operating staff focus on analyzing information rather than gathering and processing it, such solutions provide organizations with the agility they need to capitalize on business opportunities, optimize resources, and link strategic goals to operational plans.
Contemporary organizations have faced a necessity for complex and semi-structured decision-making. Dispersion of information sources and decentralization of a decision making process result in insufficiency of present information management models.

Metaheuristics algorithm is used for both Static and Dynamic Combinatorial optimization problems. Convergence is guaranteed, although the speed is unknown. Hybrid algorithms combining solution constructed by “probabilistic constructive” with local search algorithms yield significantly improved solution. It proposes a new way of thinking the solution of the non-linear complex problems.

We are constrained by linear thinking: it is hard for us to understand how all the various parts of the system interact and add up to the whole. It is very important to understand how large scale emergent patterns and behaviors can result from the actions and interactions of the individual components of a system. It imperative to analyses strategies required for Global Optimization.

7. REFERENCES:

Science and Technology, 55, 873–891.

[40] Internet Resources:
http://inform.nu/Articles/Vol4/indexv4n4.htm
http://www.intelligententerprise.com