Data Mining as a Technique for Knowledge Management In Business Process Redesign

J. Dhilipan

Abstract

Business Process Redesign (BPR) is undertaken to achieve order-of-magnitude improvements over 'old' form of the organisation. Practitioners in the academia and business world have developed a number of methodologies to support this competitive restructuring that forms the current focus of concern, many of which have not been successful. This paper suggests the use of Data Mining (DM) as a technique to support the process of redesigning a business by extracting the much-needed knowledge hidden in large volumes of data maintained by the organization through the DM models.

Keywords: Data Mining, Knowledge Management, Business Process Redesign, Business reengineering, Artificial Neural Networks.

1. Introduction

Knowledge Managers in any organization need to integrate Information Systems Strategies with Business Strategies in order to attain their vision and mission. The dividend yields a victory over their competitors through connection and interaction with their environment. Therefore, performing surgery on management overhead does not need to be matched in a dark room instead it requires transparency as suggested by Strassmann (1995). First, one must gain acceptance from those who know how to make the organization work well. Second, the organization must elicit their cooperation in telling them where the cutting will do the least damage. Third, employees must be willing to share with the organization insights about the removal of an existing business process that will improve customer service.

The redesigning of an organization's processes is variously called business re-engineering, business process re-engineering, business process design, business redesign and so on. A useful working definition of BPR is given in Smith (1996) as the fundamental rethinking and radical redesign of an entire business - its processes, jobs, organizational structure, management systems, values and beliefs.

BPR helps rethinking a process in order to enhance its performance. Academics and business practitioners have been developing methodologies to support the application of BPR principles. However, most methodologies generally lack actual guidance on deriving a process design thereby threatening the success of BPR (Selma et al, 2003). Indeed a survey has proved that 85% of BPR projects fail or experience problems (Crowe et al, 2002).

Data Mining (DM) is a field that has recently attracted the attention of various researchers and organizations. According to Mena (1999) "Data Mining is the process of discovering actionable and meaningful patterns, profiles and trends by sniffing through your data using pattern recognition technologies such as neural networks, machine learning and genetic algorithms". DM tools can answer business questions that traditionally were high time consuming to resolve. They search databases for hidden patterns, finding predictive information that experts may miss because it lies outside their expectation.

Assistant Professor(S.G) Dept. of Computer Applications SRM University, Ramapuram Campus, Chennai
E-Mail : jd_pan@yahoo.co.in

Given the amount of management attention that has been devoted to the notion of BPR, it is not surprising that a number of tools and techniques have emerged to support it. Tools that support BPR can conveniently be categorized into two sets: those that help analyze and model the business from a process perspective, and those that help plan the workflow of the business. Any or all of these tools may be supported in software. The need to deploy Data Mining Technique to BPR was carried out, and it was discovered that the hidden knowledge generated by Data Mining tools can serve as a basis for knowledge managers in organizations to redesign the whole business process so as to suite the current business development and challenges and to remain at competitive level with other business organisations.

2. Knowledge Management (Km)

Knowledge is an expensive commodity, which if managed properly is a major asset to the company. Knowledge is a complex and fluid concept. Although KM has achieved a level of popularity among firms worldwide, it has no unique or standardized definition. For the purpose of this paper, we define KM as a systematic process of finding, selecting, organizing, distilling and presenting knowledge in a way that improves the organization's interest. A key objective of KM is to ensure that the right knowledge is available at the right time in a manner that enables timely decision-making (Hariharan, 2002).

KM encompasses the way that organizations function, communicates, analyze situations, come up with novel solutions to problems and develop new ways of doing business. It can also involve issues of culture, custom, values and skills as well as relationships with suppliers and customers.

Wiig (1997), in his work said that organizations might pursue five different knowledge management (KM) strategies:

1. KM as business strategy
2. Intellectual asset business strategy
3. Personal knowledge asset responsibility strategy
4. Knowledge creation strategy and
5. Knowledge transfers strategy.

This paper presents business organizations with data mining techniques as an approach that supports such knowledge creation, sharing and transfer mechanisms.

3. Data Mining techniques

Data Mining, the extraction of hidden predictive information from large databases, is a powerful technology with great potential to help companies focus on the most important information in their data warehouses. Data mining tools predict future trends and behaviours, allowing businesses to make proactive, knowledge-driven decisions. Most companies already collect and refine massive quantities of data.

The application areas of DM as contained in recent literatures as corroborated in Jiawei (2003) include: retail industry, telephone calling patterns, natural disaster, web log click stream, financial data analysis, bio-informatics, content-based e-mail processing systems, analyzes of data from specific experiments conducted over time, analysis of nation's census database, and so DM techniques can be implemented rapidly on existing software and hardware platforms to enhance the value of existing information resources, and can be integrated with new products and systems as they are brought on-line.

There are three groups of DM users namely, Application users, Designers and Theorists. It is usually common that the theorists based on some principal assumptions usually formulate new ideas. Therefore, some users are primarily interested in this group.

Those concerned with the application of DM such as knowledge Managers which as a direct result of their interest in DM research and design they are referred to as the 'DM researcher /designer'. Finally, the respondents concerned primarily with the using or solving problems, for which DM offered an effective approach, are referred to as the "DM application group".

The most commonly used techniques in data mining are:

1. **Artificial Neural Networks**: this is a non-linear predictive model that learns through training and resembles biological neural networks in structure.
2. **Decision trees**: tree-shaped structures that represent sets of decisions. These decisions generate rules for the classification of a dataset.
3. **Genetic Algorithms**: They are optimization techniques that use process such as genetics combination, mutation, and natural selection in a design based on concepts of evolution. It tries to mimic the way nature works. It is an adaptive heuristic search algorithm premised on the evolutionary ideas of natural selection and genetics.
4. **Rule Induction**: the extraction of useful if-then rules from data based on statistical significance.
5. **Regression Methods**: this tries to identify the best linear pattern in order to predict the value of one characteristic we are studying in relation to another.

3.1 DM tasks

Some of the tasks solved by Data Mining are:

1. **Prediction**: a task of learning a pattern from examples and using the developed model to predict future values of the target variable.
2. **Classification**: a task of finding a function that maps records into one of several discrete classes.
3. **Detection of relations**: a task of searching for the most influential independent variables for a selected target variable.
4. **Explicit modeling**: a task of finding explicit formulae describing dependencies between various variables.
5. **Clustering** a task of identifying groups of records that are similar between themselves but different from the rest of the data.
6. **Market Basket Analysis**: processing transactional data in order to find those groups of products that are sold together well
7. **Deviation Detection**: a task of determining the most significant changes in some key measures of data from previous or expected values.

3.2 BENEFITS of DM techniques to web information management

A company or an organization encompassing data mining techniques can enjoy a number of benefits; these includes understanding customers' behaviour, making a judgment on the effectiveness of the company's web site- if there is one, and benchmarking marketing campaigns (Doherty, 2000 & Mena, 1999).

3.2.1 Understanding customers' behaviour

The benefits that fall under this category are summarized below:

1. Establishing the probability of customers coming back to the company or their web site.
2. Calculating the number of new customers coming to the company or their web site.

3. Identify patterns relating either to navigation routes that customers follow or to what they buy.
4. Discover whom byes what and look for any cross-relationships between clients.

3.2.2 Understanding the web site's strong points

In this category, we can find the following benefits:

1. Developing a better layout of the company's web site.
2. Identifying popular and non-popular areas of the web site.
3. Personalizing online advertisement.

4. Business Process Redesign (Bpr)

When BPR is used carefully, it can take organisations into a new realm of competitive effectiveness. However, the redesign of individual processes will always have a limited impact unless it is implemented as part of a wider view of the organization as a whole and that wider view must take root into the corporate culture. According to Wendy (1997), this is the difference between business re-engineering and process re-engineering since the first takes this wider perspective while the second is far more focused.

The purpose of this paper is to present a data mining technique that would allow business practitioners, senior managers and decision makers in organisations to extract useful, relevant, previously hidden knowledge from the organisation's database which after careful management of this knowledge yields the much knowledge needed to actualize the Business Process Redesign (BPR).

Ascari et al, (1995) found that certain factors are common to all BPR initiatives. Common features are:

1. The need for IT solutions tailored to fit the business
2. The focus on processes
3. The intent to use a pilot project approach
4. The need for top management commitment
5. The need for the communication of plans

The importance of other factors however, varied by whether the organization was competitively successful or was in a crisis situation. Features strongly sought by those in a competitive crisis were:

1. The need for a refocusing on the customer
2. The need to create coherent incentive programme
3. An emphasis on training
4. The redefinition of jobs
5. The need for cross-functional teams
6. The move towards empowerment

Kotter (1995) identified what he saw as the eight key mistakes that organisations engaged in BPR make. They are:

1. Not establishing a great enough sense of urgency
2. Not creating a powerful enough guiding coalition.
3. Lacking a vision.
4. Under-communicating the vision by a factor of ten.
5. Not removing obstacles to the new vision.
6. Not systematically planning for and creating short-term wins.
7. Not anchoring changes in the corporation's culture.

4.1 The BPR framework

The idea behind a framework is to help practitioners by identifying the topics that should be considered and how these topics are related (Alter, 1999). In this perspective the framework should identify clearly all views one should consider whenever applying a BPR implementation project.

For BPR, we suggest to use the framework described in figure 1. It is derived as a synthesis of the WCA (Work-Centered-Analysis) framework (Alter, 1999), the MOBILE workflow model (Jablonski and Bussler, 1996), the CIMOSA enterprise modeling views (Berrot and

Vemadat, 2001) and the process description classes of (Seidmann and Sundarajan, 1997). In this framework, six elements are linked as shown in figure 1.

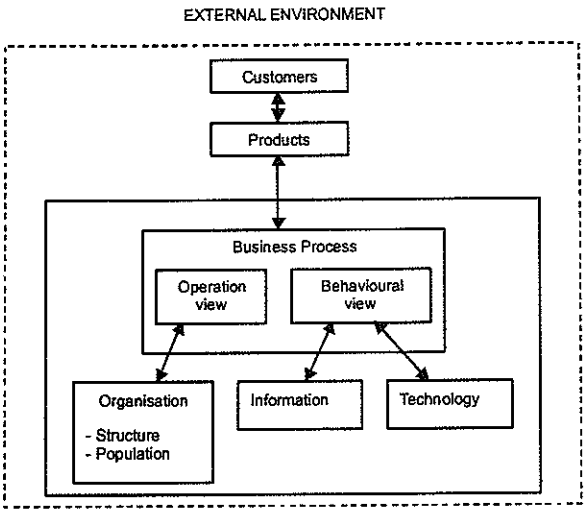

Figure 1 : Framework for BPR implementation (Adapted from Selma et al, 2003)

5. The DM/BPR framework

In order to achieve our purpose for this paper, it is very important to explain how the DM/BPR tool shown in figure 2 will extract and transfer the much-needed knowledge necessary for implementing the new business. Data on past business processes including vision, technology, management, sales, services, accountability and leadership is accumulated over time in a database. A clear understanding of this is required after which careful examination and analysis is carried out to organize the data in order to suit our purpose. The DM model is then built which could be a neural network model, genetic algorithm model, association models, decision tree models, clustering model or regression models as the case may be. The selected model is tested on the data to yield fruitful DM results previously unknown to managers and decision makers in the organization. The top managers and decision-makers take this new knowledge and implement on the BPR framework described in figure 2 to activate the new business process.

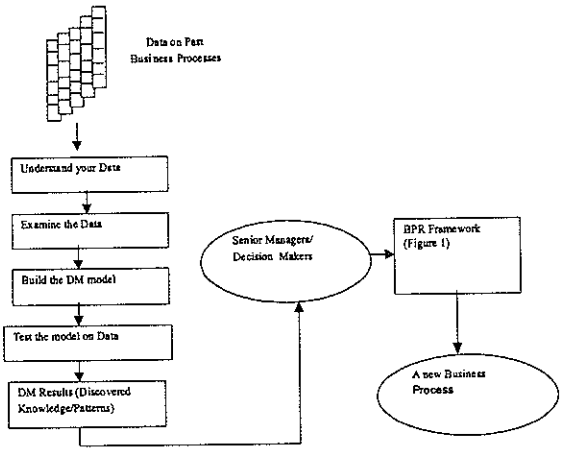

Figure 2 : DM/BPR Framework

5.1 DM as a technique for knowledge management in BPR

5.1.1 Getting the relevant data

As it was formerly stated that DM is the extraction of hidden predictive information from large databases, allowing businesses to make proactive, knowledge-driven decisions. Most companies already collect and refine massive quantities of data. Data management today is required of the ability to extract interesting patterns from large and raw data to help decision-making. The importance of collecting data that reflect your business or scientific activities to achieve competitive advantage is widely recognized now. Powerful systems for collecting data and managing it in large databases are in place in all large and mid-range companies.

However, the bottleneck of turning the data into success is the difficulty of extracting knowledge about the system you study from the collected data. For instance, there is an unprecedented growth of the use of World Wide Web for commercial and scientific purposes in the past few years, most especially in the commercial sector where people are encouraged to conduct all their transactions online. This coupled with the advances in communication technology resulted in the accumulation of data on the Internet. This data, which indicates the user's behaviour

is kept in files specially, created for that purpose called, **log files.**

There is therefore need to extract meaningful but hidden patterns from these large files through data mining techniques. In a constantly changing business environment, people or managers in various departments of industries or organizations can make their organization become much more competitive if they could get this vital information about their customer's habits.

If this vital information is gotten by managers responsible for the promotion of company's products, it would be possible to apply direct marketing techniques to every customers so that no money is wasted in vain advertisement. This could also lead to the alteration of the organization's web page layout to suit the new developments. All these could be achieved through data mining.

According to the webopedia encyclopedia of computing technology a log file is defined as "a file that lists actions that have occurred". These files are generated by servers – a computer or a device on a network that manages network resources – and contain a list of all requests made to the server by the network's users.

Information in the log files has to be written in a specific format that will facilitate the analysis of the file and instruct the computer as to how to read and use it. Log files are generated and kept on web servers; there are a variety of them in the market e.g. Apache web server (at the Apache server Website, 2000).

5.1.2 Examine the Data

In addition to the choice of the web server to take care of the user's requests, there are a wide variety of options as to how the data will be stored; that is, there are varieties of formats in existence. The common log file formed would be explained in this paper. A typical entry of this common log file format might look like the line below:

75.172.197.16 - - [23/dec/2006:09:14:23 + 0100] "GET / ~ john / business / P1625 / src / TicTacToe / docs / Intro.html HTTP / 1.0" 200 1234

The first field – 75.172.197.16 is the host making the request (The Apache Webster, 2000: Perl Tutorial website, 2000), this is a symbolic name and in the case it is not available, the IP address of the site making the request can be gotten.

The second field is the login name of the user who is making the request. Most servers may not give this for security reasons, which is the reason why a dash (–) is recorded in the log file.

The third field comprises the full name of the person who is making the request, as in the above case it is disabled in most servers and a dash (–) is recorded. Note that if a request is made for a file that is password – protected, then the user's identity should appear in this field.

The next entry in brackets comprises the data and the time the request was made and of the format dd/m/yyyy and hh:mm:ss respectively. An obvious problem here is that of the number of different time zones around the world. The time zone used is the Greenwich Mean Time (GMT).

The next in a set of double quotes after which a request is made to the server through the relevant command GET. After which we have the specific file request, which in the case of the above example is:

john / business / P1625 / src / TicTacToe / docs / Intro.html, under the directory of john.

The next entry is the protocol – in which HTTP 1.0 is used in most cases. The next entry is a three – digit code that shows the result of the request, that is, the success or

failure of the server to accommodate the request. The first digit can take five values:

Table 3: The meaning of the first digit of the status code (Adopted form Perl Tutorials website, 2000)

Value of first digit	Meaning
1	Informational
2	Success
3	Further action needed
4	Mistake from the person making the request
5	Server's failure

A number of sample codes and their meanings are shown below:

Table 4: Some status codes and their meanings (Adopted from Perl Tutorials website, 2000)

Status code	Description
200	OK
204	No content
301	Moved permanently
302	Moved temporarily
400	Bad request
401	Unauthorized
403	Forbidden
404	Not found
500	Internal server Error
501	Not implemented
503	Service Unavailable

In table 4, the code returned was 200, which means that the request was successfully completed. The final entry of the log file's format is the length of file transferred and this is 1234 bytes for the above example.

5.1.3 Analysing web log files

Many log analyzers have been developed today. Some of these can be downloaded from the Internet as commercial program, freeware or shareware. Some examples are: 123LogAnalyzer, WebTrends, AlterWind, SurfStats, FastStats, and Sawmill.

In the market place, there are a large number of ready–made programs that help us analyse the log files our server has generated. Their processes differ, some are free and some sophisticated one can be bought at very high prices. The choice depends on our organization's special needs.

5.1.4 Evaluating web log files

The step that follows the analysis stage is evaluation. Once the mining tools i.e. the log analyzers have been applied, we have some resulted–which are figures only a step away from fulfilling the initial goal of turning our raw data into usable information. These results can then be analyzed by managers or decision–makers in the organization or by experts brought in from outside the company and valuable associations and patterns previously unknown to them can be generated.

5.1.5 Building and testing DM Models to parse the log file

As discussed above, there are a large number of packages available in the market place that will perform the parsing and analysis of the log file for us. It is also possible for someone to write his/her own program to do the same task. There are two major benefits that can be enjoyed in doing that. They are:

1. **Extensibility:** this means the user can extend, add or remove components from the program rather than waiting for the next upgrade of the commercial program
2. **Customizability:** this means a user can write and perform his/her own queries and get the results of choice rather than some general statistics e.g. one might be interested in the nationality of customers that access our home page between a particular times of the day.

5.1.6 Factors to consider in writing your own program

There are two factors to bear in mind.

1. **Size of log files:** a log file of a medium sized company can be anything in size up to thirty or forty megabytes or more. Note that, we need not load the whole file into memory, as this would have some disastrous effects on the performance of the computer. Loading and manipulating one line at a time better do it.

2. **Speed of expected result:** a language that would perform these tasks quickly is needed and which will give us the opportunity to perform the tasks we want at a very high speed. For example, Perl, C++ etc are good for that kind of programs.

5.2 Consequences that may arise from the BPR through DM and KM

Consequences that may arise are:

1. **Predicting cross–sell opportunities and making recommendations:** Whether you have a traditional or web–based operation, you can help customers quickly locate products of interest to them and simultaneously increase the value of each communication with your customers.

2. **Identifying your best prospects and then retaining them as customers:** By concentrating your marketing efforts only on your best prospects your organization will save time and money; thus increasing effectiveness of your marketing operations.

3. **Segmenting Markets and personalizing communications:** It is possible now to identify distinct group of customers, patients, students or natural phenomena that require different approaches in their handling.

4. **Learning parameters influencing trends in sales and margins:** Now you can know what combination of parameters is actually influencing trends in sales and margins and general operations.

5. **Saving costs and time by:** Streamlining processes (limiting the number of departments/people involved in a single process), removing non-value adding activities and identifying where systems support is inadequate.

6. Conclusion

The process of extracting knowledge hidden from large volumes of data (DM) has proved very successful in solving many business or scientific problems to achieve competitive advantage. As suggested in the DM/BPR framework, the DM model can be deployed on the massive data collected from past business processes of the organization which then yields the much needed previously unknown knowledge and trends needed by top managers or decision makers in the organization for effective business process redesigning.

The proposed DM/BPR framework transforms the old business into a new prospect oriented business organization by carefully re-engineering the old system incorporating the new discovered knowledge which helps the manager to make wise and informed business decisions in the area of accountability, business change management expertise, business process analysis, business model design, business model implementation and others.

References

1. Alter, S. (1999) 'Information systems: a management perspective, Addison Wesley, Amsterdam.

2. Ascari, A., Rock, M. and Dutta, S. (1995) Reengineering Organisational Change, *European Management Journal*, Vol 13 No 1 pp. 1-30.

3. Berrot, G. and Vemadat, F. (2001) 'Enterprise modeling with CIMOSA: functional and organizational aspects', *Production Planning and Control*, Vol. 12, No 2; pp 128-136.

4. Campbell D.T. and Fiske D.W. (1959) Convergent and discremenant validation by the Multitrait-multimethod Matrix .*Psychological Bulletin* 56(2) ,81-105

5. Cronbach L.J. (1971) Test Validation . In Educational Measurement (Thorndike RL, Ed) .pp. 443-507 ,American Council on Education Washignton ,DC.

6. Crowe, T. J., Fong, P.M., Bauman, T.A and Zayas-Castro, J.L 'Quantitative risk level estimation of Business Process Re-engineering efforts' *Business Process Management Journal*, vol. 8 No 5 (16 October 2002) pp. 490-511, 22 MCB University Press.

7. Doherty P., (2000), Web Mining: The e–Tailers' Holy Grail, dmDirect, [online] available at http:// www.dmreview.com/editorial/dmdirect/ dmdirect_articles.cfm? EdID =1891 & issue = 012800 & record = 1.

8. Gupta, A. and McDaniel, J. (2002), 'Creating competitive advantage by effectively managing knowledge management, *Journal of knowledge Management Practice*, Vol 3, No 2; pp 40-49.

9. Hariharan, A. (2002), Knowledge Management: A strategic Tool, *Journal of knowledge Management Practice*, Vol. 3; No. 3; pp 50-59

10. Jablonski, S. and Bussler, C. (1996) 'Workflow management: modeling concepts, architecture and implementation, London, International Thompson Computer Press.

11. Jiawei, H (2003) Data Mining : Current status and Research Directions ; School of Computing Science ,Simon Fraser University , March Burnaby B.C. Canada ;August http://www.cs.sfu.ca/~han

12. Kotter, J. (1995) 'Leading Change: Why Transformation Efforts Fail, *Harvard Business Review*, Mar-Apr; pp.59-67.

13. Mena J. (1999), 'Data Mining Your Website', dpDigital Press, ISBN: 1-55558-222-2, The United States of America.

14. Neal, A., & Griffin, M. A. (1999). Developing a model of individual performance for human resource management. *Asia Pacific Journal of Human Resources, 37*, 44-59.

15. Perl Tutorial Website (2000), Working with files in Perl, [Online], Available at: http:// www.tjhsst.edu/~dhyatt/perl/ex3.html

16. Seidmann, A. and Sundararajan, A. (1997) 'The effects of task and information asymmetry on business process redesign' *International Journal of Production Economics*, Vol 50, No. 213; pp 117-128.

17. Selma, L. M., Farhi, M. and Hago A. R. (2003) 'Case-Based Reasoning as a Technique for knowledge Management in Business Process Redesign, *Electronic journal of knowledge Management, volume 1*, issue 2 113-124

18. Smith, S. (1996) Rules of Engagement, *Computer Weekly*, 14 Mar pp 36-7 Strassmann, P. (1995), The Politics of Information Management, Information Economics Press.

Authors Biography

J. Dhilipan, presently working as assistant professor in SRM University, Chennai. He has a total of 10 years of teaching experience. He has published 12 papers in national conferences and 4 papers in international conferences. He has attended various workshops and seminars.