

Data Mining and Business Intelligence Applications in Telecommunication Industry

Madhuri V. Joseph

Abstract- *Telecommunication companies today are operating in highly competitive and challenging environment. Huge volume of data is generated from various operational systems and these are used for solving many business problems that required urgent handling. These data include call detail data, customer data and network data. Data Mining methods and business intelligence technology are widely used for handling the business problems in this industry. The main application areas of BI and Data Mining in telecommunication industry include fraud detection, network fault isolation and improving market effectiveness.*

Keywords: *Data Mining, Telecommunications, Business Intelligence, Fraud Detection, Network fault Isolation, Marketing & CR.M*

I. INTRODUCTION

The concept of Data Mining has gained a widespread market acceptance. Telecommunication is one of the most data intensive industries in the world. One of the first industries to adopt Data Mining is the telecommunications. Companies in the telecom industry are making use of Data Mining technologies to improve their marketing techniques, for identification of customer fraud and for the better management of their networks. The popularity of Data Mining in the telecommunication industry can be viewed as an extension of the use of expert systems in the telecommunication industry [1]. The major problem with the expert systems is that they are expensive to develop. The four key challenges that the telecommunication industry faces today are summarized by [2] as the 4 Cs: Consolidation, Competition, Commoditization and customer service. Most of the telecom companies have realized that the huge volume of data they collect and possess could be effectively utilized for solving their business problems by converting them into information and knowledge. Data Mining can be viewed as a technique automatically generating this knowledge from the data available. One of the first industry to experience the benefits from the application if Business Intelligence (BI) and Data Mining technologies in the telecommunications industry.

II. OVERVIEW

The Data Mining & BI applications in any industry depend on two main factors: the data that are available in the industry and the business problems that are to be solved with the BI and Data Mining technologies. The business problems arise in the telecommunication industry need urgent attention.

Manuscript published on 28 February 2013.

* Correspondence Author (s)

Madhuri V. Joseph, Sr.Lecturer, Dept of Computing, Muscat College, Muscat, Sultanate of Oman.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](http://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

The key success element for Data Mining & Business Intelligence applications is the use of powerful tools and techniques for the possession of huge amount of high quality data. Telecommunication companies store the call detail records (CDR), which contains descriptive information for each phone call. In 2001 AT&T recorded 300 million call details records per day from the long distance customers. Call detail data can be used for marketing and fraud detection purposes. Telecom companies also maintain customer information like billing information or credit score information. This information can be often used to improve the results of Data Mining. Telecommunication companies generate and store the data related to their networks. Large telecommunication networks use self-diagnostic capabilities for generating status and alarm messages. These messages can be mined to support the network management functions such as fault isolation and predictions.

The telecommunication industry faces a number of challenges in Data Mining. This is because of the tremendous volume of data belonging to the companies. One issue is that the telecommunication data is in the form of transactions, which is not at the proper level for semantic Data Mining. Another issue faced by the telecommunication industry is that the data generated by the telecommunication applications like fraud detection and network fault identification need to be processed real time. The effort to address these issues has made the telecommunication industry a leader in the research area of Data Mining [3]. The final issue associated with the data in telecommunication industry involves rarity. Predicting & identifying rare events is quite difficult for many Data Mining algorithms [4].

Table 1: Data Mining applications and Business problems in Telecommunication industry.

Data Mining Applications	Business Problems Addressed
Marketing & CRM	<ul style="list-style-type: none"> Measure customer value and retain profitable customers. Generate customer profiles and mine these for marketing purposes. Acquire new customers Maximize profit. Churn analysis which includes churn prediction & churn management.
Fraud Detection	<ul style="list-style-type: none"> Identification of fraudulent users. Detection of fraudulent entry to customer accounts. Discovery of unusual patterns like busy hour, switch & route congestions patterns etc.
Network Management	<ul style="list-style-type: none"> Identifications of network fault & its predication Management of system work load and resource usage. Identification user group behavior and comparison of data traffic.



Data Mining is one of the most sophisticated data analytical techniques used in BI systems; various Data Mining applications have been set out in the telecommunication industry. The three main application area identified are: Marketing & Customer relationship Management (CRM), Fraud detection and Network Management. BI applications in the Telecommunications industry can be classified into 6 categories namely: Marketing analysis, Financial analysis, Sales analysis and Billing, Customer care and CRM, Fraud management and Network optimization [2]. These six applications areas can be grouped into the 3 main application areas. According to the business problem addressed, the BI and Data Mining applications can be further divided as shown in the table below.

III. TYPES OF TELECOMMUNICATION DATA

Useful applications cannot be developed without understanding the various data used in telecommunication industries. So the first step in the Data Mining process is to understand the data. The different kinds of data used in this industry are mainly grouped into 3 different types.

A. Call detail data

This is the information about the call, which stores as the call detail record. The number of call detail records generated is huge since every call is placed on the network, the details are stored. Call detail record includes information like originating and terminating phone numbers, date, time and duration of call. Usually these call detail records are not directly used for Data Mining. A list of features can be generated from the call detail data such as

- Average call duration
- Average number of call originated per day
- Average number of call received per day
- Percentage of no-answer calls
- Percentage of day time calls (office hours)
- Percentage of weekday calls (Monday – Friday)

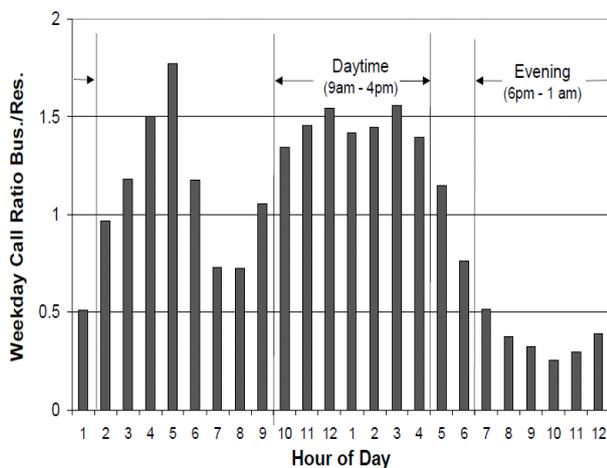


Fig 1 Comparison of Business and Residential Hourly Calling Patterns

These features can be used to generate a customer profile, which would be used to distinguish residential and business customers. The construction of these features can be done with customer sense and expert knowledge. Call detail data is generated in real time.

B. Network data

Telecommunication networks contain thousands of components, which are interconnected. These components are capable of generating error and status messages which leads to a large volume of network data. These network data are used for network management functions like fault detection. Expert systems have been developed to analysis these messages automatically, since the huge volume of network messages generated cannot be handles by technicians. Hence Data Mining technologies are used in identification of network faults by automatically extracting knowledge from network data. Network data is also generated in real time which can be accomplished by applying a time window to the data.

C. Customer data

Like any other business, telecommunication companies also have millions customers. Hence it is very much essential to have a database for storing the information about these customers. Information about the customer will include:

- Name of the customer
- Address details
- Payment history
- Service plan and so on

Group customer data is used to provide call detail data in order to identify fraud.

IV. DATA MINING AND BI APPLICATIONS

The two main factors on which Data Mining and BI applications relay on include the availability of the problem that has to be approached and solved by the Data Mining and BI technologies and the availability of Data for implementing the technologies. The main reason behind the significance of Data Mining and Business Intelligence Applications in the Telecommunications industry is the availability of tremendously large volume of data.

A. Marketing and customer relationship management (CRM)

Telecommunication companies maintain a huge volume of data about their customers and their call details. This information can be used to profile the customers and these profiles can be used for marketing and forecasting purposes. The emphasis of marketing application in telecommunication industry has moved from identifying new customers to measuring customer value and then taking steps to return the profitable customers. This shift has happened because it is expensive to acquire new customers than retaining the existing ones. A numerous Data Mining methods can be used to generate the customer life time value (the total net income a company can expect from a customer over time) for telecommunication customers. Different Data Mining techniques are used to model customer life time value for telecommunication customers [5], [6]. The key element of modeling the life time value for a telecommunication customer is to estimate how long he/she will remain with their current network. It will help the company to predict when a customer is likely leave and to take proactive steps to retain the customer.

One of the serious issues that the telecommunications industries face is the customer churn. The process that a customer leaving a company is referred to as churn and churn analysis can be done through numerous systems and methods [7], [8] & [9]. Customer churn is a significant problem because of the loss of revenue and high cost of attracting new customers. Data Mining techniques are widely used for churn analysis. These rules were generated using SAS Enterprise Miner, a sophisticated Data Mining package that supports multiple Data Mining techniques For example to classify a customer as a residential or a business customer can be given as below.

Rule 1: If $< 43\%$ of calls last 0 – 10 seconds and $< 13.5\%$ of calls occur during weekend and as of neural network algorithms based on time of day call distribution then business customer.

Rule 2: If calls received over two – month period from at most 3 unique area codes and $< 56.6\%$ of calls last 0-10 seconds then residential customer & Neural network & decision. Tree algorithms are used to predict the probability that a customer is a residential or a business customer. An evaluation indicates that the accuracy of Rule 1 is 88% and that of Rule 2 is 70%. Data Mining and business Intelligence methods and techniques could be applied for defining product and channel profitability. Association rules and sequential patterns are used to promote telecommunications services. Data Mining techniques like classification, clustering and association rule are used for solving telecommunication business problems. Data Mining techniques are used for outlining four CRM dimensions namely customer identification, attraction, retention and development. Customer information and call details can be used to determine customer behavior and identify the opportunities to support customer base expansion and customer churn reduction. Data Mining can be very useful in marketing programs development. For example if a service provider has a goal to increase the number of customers paying bills online or to increase revenue from advertisers then Data Mining techniques can be utilized.

B. Network Fault Isolation & Prediction

Telecommunication networks are comprised of highly complex configurations of hardware and software. Since the industry requires optimum network efficiency and reliability, most of the network elements have the capability of self-diagnosis and generating status and alarm messages. Expert systems were developed to handle alarms [10]. Network fault isolation in the telecommunication industry is a quiet tedious task because of the following reasons.

- Huge volume of data
- A single fault can generate different unrelated alarms.

Hence alarm correlation has an important role in predicting network faults. A proactive rapid response is very much essential for maintaining the reliability of the network. Data Mining techniques like classification, neural network and sequence analysis can be used for identifying network faults.

The telecommunication Alarm Sequence Analysis (TASA) is a Data Mining tool which support fault identification by searching for recurrent patterns of algorithms [11]. This information can be used to generate a rule based alarm correlation system, which can be used for identifying faults in real time. [Genetic algorithm is another method to predict the telecommunication switch failures

[12]. Time weaver is a genetic algorithm which has the capability to operate directly on the raw network level time series data [12]. This algorithm will identify patterns that will successfully predict the target event. Bayesian Belief Networks can also be used to identify the network faults [13]. Standard classification tools can be used to generate rules to predict future failures but it has several draw backs [14]. Most importance drawback of this is that some information will be lost in reformulation process.

C. Fraud Detection

Fraud is very serious issue that the telecommunication industry faces since it leads to the loss of revenue by billions of dollars. As provided by Gosset & Hyland 1999, the telecommunication fraud can be defined as “any activity by which telecommunication service is obtained without intention of paying”. Telecommunication fraud can be classified into two categories namely Subscription fraud & Superimposition fraud [15]. Subscription fraud occurs when a customer opens an account with the intention of never paying. Telecommunication companies consider Superimposition frauds are the most significant problems which occurs when a perpetrator gains illegal access to the account of a legitimate customer. Both subscriptions fraud and Superimposition fraud should be detected immediately and customer account should be deactivated. Cellular cloning was a very serious issue in 1990’s. This was eliminated with the Authentication methods.

Deviation detection and Anomaly detection are the most common techniques used for detecting superimposed fraud. Combined use of customer signatures dynamic clustering and pattern recognition are some other methods which are recently applied in this area. Absolute analysis and differential analysis are considered as the two main sub categories of approaches for fraud detection. According to [16], the most often used techniques for fraud detection in telecommunication include statistical modeling, Bayesian rules, visualization methods, clustering, rule discovery, neural network, Markov models as well as combinations of more than one method. Customer data can also be used for detecting fraud. For example price plan and credit rating information can be in cooperated into the fraud analysis [17].

Another common method for fraud detection is to create a profile of customer’s calling behavior and compare activity against this behavior. This calling behavior can be generated by briefing the call detail records for a particular customer. Fraud can be identified immediately after it happens, only if the call details records are updated in real time. Fraud detection system works at the customer level, not at the individual call level. Fraud detection application involves predicting a relatively rare event where the class distributions involved is highly twisted.

V. FUTURE TRENDS AND CHALLENGES FOR DATA MINING AND BI IN TELECOMMUNICATIONS

Since the quality of the data is the primary concern in telecommunication industry, Data Mining has an important role. More complex data pre-processing is needed since the telecommunication data is needed in the form of transactions or events.

Scalability is a major concern because the industry is handling very large databases. Fraud detection and network fault isolation in the telecommunication increase the significance of real-time operation. Data Mining applications should always consider the privacy issues. This is especially true in telecommunication industry because the companies maintain highly private information such as whom each customer calls. It is clear that the new Data Mining applications will be developed and deployed which will eliminate some of the current applications.

Conference on Knowledge Discovery and Data Mining (pp. 409-413). New York: ACM Press.

VI. CONCLUSION

Data Mining and BI applications play a significant role in the telecommunication industry due to the availability of large volume of data and the rigorous competition in the sector. The primary application areas include marketing and Customer Relationship Management, Fraud detection and Network Management. The recent developments in the Data Mining and BI fields and the implementation and enhancement of existing techniques and methods ensure the continuous growth and compatibility of telecommunication companies that make use of them.

REFERENCES

- [1]. Liebowitz, J. (1988). Expert System Applications to Telecommunications. New York, NY: John Wiley & Sons.
- [2]. Pareek, D.: Business Intelligence for Telecommunications. Auerbach Publications, Taylor & Francis Group LLC. (2007).
- [3]. Aggarwal, C. (Ed.). (2007). Data Streams: Models and Algorithms. New York: Springer.
- [4]. Weiss, G. (2004). Mining with rarity: A unifying framework. SIGKDD Explorations, 6(1), 7-19.
- [5]. Rosset, S., Neumann, E., Eick, U., & Vatrik (2003). Customer lifetime value models for decision support. Data Mining and Knowledge Discovery, 7(3), 321-339.
- [6]. Freeman, E., & Melli, G. (2006). Championing of an LTV model at LTC. SIGKDD Explorations, 8(1), 27-32.
- [7]. Wei, C., Chiu, I.: Turning Telecommunications call details to Churn Predictions: A Data Mining Approach. In Expert Systems with Applications 23, pp. 103-112. (2002).
- [8]. Mozer, M., Wolniewicz, R., Grimes, D., Johnson, E., & Kaushansky, H. (2000). Predicting subscriber dissatisfaction and improving retention in the wireless telecommunication industry. IEEE Transactions on Neural Networks, 11, 690-696.
- [9]. Mani, D., Drew, J., Betz, A., & Datta, P (1999). Statistics and Data Mining techniques for lifetime value modeling. Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 94-103). New York, NY: ACM Press.
- [10]. Weiss, G., Ros, J., & Singhal, A. (1998). ANSWER: Network monitoring using object-oriented rule. Proceedings of the Tenth Conference on Innovative Applications of Artificial Intelligence (pp. 1087-1093). Menlo Park: AAAI Press.
- [11]. Klemettinen, M., Mannila, H., & Toivonen, H. (1999). Rule discovery in telecommunication alarm data. Journal of Network and Systems Management, 7(4), 395-423.
- [12]. Weiss, G., & Hirsh, H (1998). Learning to predict rare events in event sequences. In R. Agrawal & P. Stolorz (Eds.), Proceedings of the Fourth International Conference on Knowledge Discovery and Data Mining (pp. 359-363). Menlo Park, CA: AAAI Press.
- [13]. Sterritt, R., Adamson, K., Shapcott, C., & Curran, E. (2000). Parallel Data Mining of Bayesian networks from telecommunication network data. Proceedings of the 14th International Parallel and Distributed Processing Symposium, IEEE Computer Society.
- [14]. Sasisekharan, R., Seshadri, V., Weiss, S. Data Mining and forecasting in large-scale telecommunication networks. IEEE Expert 1996; 11(1):37-43.
- [15]. Fawcett, T., & Provost, F. (2002). Fraud Detection. In W. Klossgen & J. Zytow (Eds.), Handbook of Data Mining and Knowledge Discovery (pp. 726-731). New York: Oxford University Press.
- [16]. Shortland, R., Scarfe, R.: Data Mining Applications In BI. In BT Technology Journal, Vol 25, Nos 3 & 4, July/October 2007.
- [17]. Rosset, S., Murad, U., Neumann, E., Idan, Y., & Gadi, P. (1999). Discovery of fraud rules for telecommunications—challenges and solutions. Proceedings of the Fifth ACM SIGKDD International