Business Intelligence Applications and Data Mining Methods in Telecommunications: A Literature Review

Dorina Kabakchieva

Sofia University „St. Kl. Ohridski"
125 Tzarigradsko shosse Blvd., 1113 Sofia, Bulgaria
dorina@fmi.uni-sofia.bg

Abstract. Telecommunication companies are operating today in an extremely challenging business environment. The Telecommunication industry is in possession of large quantities of data, generated from numerous operational systems, and is confronted with many business problems that need urgent handling. Naturally, it has been among the first to adopt Business Intelligence (BI) and Data Mining technologies. The main purpose of this paper is to present a literature review related to BI and Data Mining in Telecommunications, from business perspective - defining the main areas of BI and Data Mining applications, and from research perspective - identifying the most common Data Mining techniques and methods used.

Keywords: Business Intelligence, Data Mining, Telecommunications.

1 Introduction

Telecommunications industry has undergone intensive growth and development during the last decade. Telecom operators and carriers are operating today in an extremely challenging business environment, due to the increasing customer dissatisfaction with existing services; market uncertainty; bandwidth commoditization; limited market capital; large, expensive, and inflexible IT infrastructures. The four key challenges that the Telecommunications industry faces today are summarized by Pareek (2007) in [1] as the 4 Cs: consolidation, competition, commoditization and customer service. They are also thoroughly discussed by Sumathi and Sivanandam (2006) in [2], Mattison (1997) in [3], Weiss in [4], [5], and briefly mentioned in many of the other referenced literature sources. In order to survive and remain competitive, Telecommunications companies of all sizes are aggressively moving from a product-strategy-based to a customer-strategy-based business model, placing customer in the central focus of their activities. They have also realized that the enormous quantities of data they collect and possess, could be effectively used to support them in solving some of the important business problems and for getting competitive advantage, by turning it into useful information and knowledge – the most important company assets in the knowledge-based society. The Telecommunications industry was one of the first to experience the benefits from the application of Business Intelligence (BI) solutions and the adoption of the Data Mining technologies.

This paper presents a literature review related to BI and Data Mining research issues in Telecommunications. The idea is to do that from both, the business perspective - defining the main areas of BI and Data Mining applications, and the research perspective - describing the most often used Data Mining methods and techniques.

2 BI and Data Mining Applications in Telecommunications

The BI and Data Mining applications in any industry depend on two main factors: the availability of business problems that could be successfully approached and solved with the help of BI and Data Mining technologies, and the availability of data for the implementation of such technologies. As mentioned in most of the analyzed literature sources, these two main prerequisites are entirely fulfilled for the Telecommunications sector. The Telecommunications industry is confronted with many business problems (briefly discussed above) that need urgent handling by using innovative powerful methods and tools, and is in possession of large quantities of high quality data that is a key success factor for BI and Data Mining applications.

Telecommunications data is generated in the many different operational systems used and could be generally classified as three main types: (1) customer contractual data – personal data about the customers, including name and address, service plan, contact information, credit score, family income and payment history; (2) call detail data –detailed call records, including the originating and terminating phone numbers, the date and time of the call and the duration of the call customer transactions, and billing data, from which knowledge could be extracted at customer level, about the customers calling behaviour; (3) network data – data resulting from the operation of the extremely complex configurations of equipment, minimally including a timestamp, a string that uniquely identifies the hardware or software component generating the message and a code that explains why the message is being generated. The identified three types of Telecommunications data are generally described by Weiss in [4], [5]. The data multidimensionality is discussed by Sumathi et al. in [2], and is considered as one of the most important factors for the wide variety of BI and Data Mining applications.

The availability of tremendously large volumes of Telecommunications data, described above, is a very important reason for the recent very high interests in BI and Data Mining in the Telecommunications industry. Telecom BI applications are discussed in only one of the analyzed literature sources [1], the others are focused on Data Mining applications. However, since Data Mining is one of the most sophisticated data analytical techniques used in BI systems, the application areas can be generalized.

There are three main application areas identified as a result of the analysis of the selected literature sources: (1) marketing, sales and customer relationship management (CRM); (2) fraud detection; and (3) network management.

Pareek in [1] divides Telecom BI applications in six categories: fraud management (2), financial analysis (1), marketing analysis (1), network optimization (3), sales analysis and billing (1), and customer care and analytical CRM (1). However, they could easily be classified in the above defined three main application areas (as shown by the numbers provided in the brackets). Sumathi

et al. in [2] talks about "multidimential analysis of Telecommunications data" that could be referred to the third main application area – network management (3), "fraudulent pattern analysis and the identification of unusual patters" which is actually the second main application area – fraud detection (2), and "multidimensional association and sequential pattern analysis" that is associated with the first main application area – marketing, sales and CRM (1). Weiss in [4], [5] also specifies three categories of Data Mining applications in Telecommunications, including marketing, fraud detection, and network fault isolation and prediction. Several case studies, related to "fault diagnosis" (3), "fraud" (2), "marketing" (1), "credit assessment" (1), and "litigation assessment" (1) are provided in [17]. Most of the other analyzed literature sources are focused on describing Data Mining applications in the marketing, sales and CRM area (1) and fraud detection area (2). Various applications, related to customer scoring and segmentation, marketing campaign assignment and management, customer churn management, risk analysis and fraud detection, are discussed in [6]. Ngai at al. provide a literature review and classification of CRM applications in their recent paper [7]. Churn management, which is a very important CRM and marketing problem, is a very hot topic presented in [10], [14], [15]. The marketing and CRM related problems concerning customer classification and segmentations are considered in [11], [12], [18]. The quality of the provided Telecommunication services, which is also a topic within the marketing, sales and CRM application area (1), is considered in [9]. Fraud detection (2) is the research topic in [13], [16].

The BI and Data Mining applications within the identified three main areas could be further classified according to the specific business problem they address. Based on that, BI and Data Mining applications are summarized in Table 1.

Table 1. BI and Data Mining Applications in Telecommunications.

BI and Data Mining Application Areas	Business Problems Addressed

Marketing, Sales and CRM	• Generating customer profiles from call detail records and mining these profiles for marketing purposes
	• Measuring customer value and retaining profitable customers
	• Maximizing the profit obtained from each customer
	• Discovering association and sequential patterns to promote Telecommunication services
	• Acquiring new customers
	• Churn Analysis
	  ° Churn prediction: predicting whether a particular customer will churn and when it will happen
	  ° Churn management: understanding why particular customers churn and applying efforts to retain them

Fraud Detection	• Identification of potentially fraudulent users and their atypical usage patterns (subscription fraud) • Detecting attempts to gain fraudulent entry to customer accounts (superimposed fraud) • Discovering unusual patterns that may need special attention such as busy-hour, frustrated call attempts, switch and route congestion patterns, etc.
Network Management	• Network fault identification ° Alarm correlation (for relating multiple alarms to a single fault) • Network fault prediction • Identifying and comparing data traffic • System work load management • Resource usage management • User group behaviour

3 Data Mining Methods Applied for Solving Business Problems in Telecommunications

This section focuses on the results of the literature review, concerning the implementation of different Data Mining functions and methods for solving significant Telecommunications business problems in the identified three main areas of application – marketing, sales and CRM; fraud detection and network management.

3.1 Marketing, Sales and CRM

The huge amounts of customer contractual and call detail data, collected in the operational systems of Telecommunication companies, provide valuable opportunities for creating customer profiles and obtaining customer scoring, segmentation, and classification by using a variety of Data Mining methods and techniques. The discovered patterns and models in data could then be successfully used for marketing, sales and CRM purposes, e.g. for campaign assignment and management, channel efficiency improvement, loyalty programs development, up-sell and cross-sell analysis, predicting product sales and customer trends, churn prediction and management, better customer care and more efficient analytical CRM, etc. All these activities will substantially contribute to increasing the profits of Telecommunication companies, and to improving their financial performance and compatibility.

Customer segmentation could be based on mapping the company products against a customer life-cycle model, thus reflecting the thinking that customers require different products and services at various points in their lives. BI and Data Mining methods and techniques could also be applied for defining customer, product and channel profitability. All these issues are discussed in [1]. The discovery of association and sequential patterns that could be used to promote Telecommunications services, the use of visualization tools, and churn analysis, are considered in [2]. Some of the most common Data Mining techniques used for solving Telecommunications business problems and appearing in many of the analyzed literature sources include classification, clustering and association rule mining. A clustering method based on genetic algorithm for Telecommunications customer subdivision is presented in [11]. The authors in [18] have focused on solving the multi-criteria clustering problem for obtaining a typology of the customers of a large Telecommunications company, by using two different modeling approaches - multiple factor analysis and prior transformation of data. The experience, concerning the application of factor analysis, clustering and quantitative association rules, for finding the service adoption patterns of segmented groups of mobile Telecommunication subscribers in Korea, is presented in [12].

The extremely important topics concerning CRM are thoroughly discussed in [7], where the authors provide a literature review and classification of Data Mining techniques used for the outlined four CRM dimensions – customer identification, attraction, retention and development. The study reveals that association, classification, clustering, forecasting, regression, sequence discovery and visualization, are the mostly often applied Data Mining techniques, while association rules, decision trees, genetic algorithms, neural

networks, K-Nearest Neighbour and linear/logistic regression are the most widely used Data Mining methods.

Churn prediction and management is a very important CRM aspect that has become of great concern to Telecommunications companies, because it has been established that the cost of retaining an existing subscriber is generally much lower than the cost of acquiring a new customer. This topic is discussed in many of the analyzed literature sources, and is the basic research focus in [10], [14], [15]. The authors propose in [10] a churn-prediction technique based on subscriber contractual information and call pattern changes extracted from call details, using a multi-classifier class-combiner approach. The study in [14] compares various Data Mining techniques that can assign a "propensity-to-churn" score periodically to each subscriber of a mobile operator, concluding that decision trees and neural networks are very effective Data Mining methods for churn predictive models development. The factors influencing customer churn in the Korean mobile Telecommunications service market are studied in [15].

3.2 Fraud detection

One of the definitions of Telecommunication fraud is provided by Gosset & Hyland (1999) and cited by Hilas in [13] as "any activity by which Telecommunications service is obtained without intention of paying". Fraud is a serious problem for Telecommunication companies, leading not only to revenue lost for telecom providers, but sometimes to burden for their customers as well. The most common methods used for fraud detection are based on the analysis of users' activities, comparing new behaviour against a profile of old behaviour. Two types of Telecommunications data are basically used – customer contractual data and call detail data.

Weiss classifies Telecommunications fraud in [4], [5] as subscription fraud, occurring when a customer opens an account with the intention of never paying it, and superimposition fraud, occurring when a perpetrator gains illicit access to the account of a legitimate customer. Superimposed fraud is considered more significant problem and therefore it is often the basic topic of research efforts. Deviation detection and anomaly detection techniques are very often applied for superimposed fraud detection [4], [5]. Some recently applied methods also mentioned in [4], [5] include combined use of customer signatures, dynamic clustering and deviation detection; pattern recognition by using visualization tools for recognizing unusual patterns.

Hilas considers in [13], [16] two main subcategories of approaches for fraud detection – the absolute analysis, searching for limits between legal and fraudulent behaviour, and differential analysis, trying to detect extreme changes in the user's behaviour. The author provides in [16] a comprehensive review of research in Telecommunications fraud detection, stating that the most often used techniques come from the area of statistical modeling, including rule discovery, clustering, Bayesian rules, visualization methods, Markov models, neural network classification, as well as combinations of more than one method. He is also discussing in [16] the possibility for applying both, supervised and unsupervised learning methods for fraud detection.

3.3 Network Management

Growing and maintaining profit margins in the Telecommunications industry requires optimum network efficiency and ensuring high network reliability. BI and Data Mining analytical tools have proven to be very effective for comparing a wide range of metrics across network operations, creating real-time reports for identifying problems that need immediate attention, and generating alerts for instant notification of emergency situations requiring rapid response. These issues are considered in [1], [2], [4], [5], [9], [17]. Sequence analysis, time-series data analysis, and classification, are the Data Mining techniques considered in [4], [5]. Quality analysis of a digital mobile Telecommunications network is the research topic in [9], focusing on the application of Data Mining methods (rough sets, classification trees, self-organizing maps) for detecting the changes in service quality based on the network measurements generated in the operational systems. Machine learning classification techniques, e.g. neural networks, are used in the case study describing fault diagnosis in [17].

The identified Data Mining techniques and methods, used in the defined three main areas of application in the Telecommunication industry, are summarized in Table 2.

Table 2. Data Mining Techniques and Methods Applied in Telecommunications.

BI and Data Mining Application Areas	Data Mining Techniques	Data Mining Methods
Marketing, Sales and CRM	Association Classification Clustering Forecasting Regression Sequence discovery Visualization	Association rules Decision trees Genetic algorithms Multiple factor analysis Neural networks K-Nearest Neighbour Linear/logistic regression Visualization methods
Fraud Detection	Outliers detection Deviation detection Statistical modelling Dynamic clustering Classification Visualization for pattern recognition	Anomaly detection techniques Rule discovery Clustering algorithms Bayesian rules Visualization methods for recognizing unusual patterns Markov models Neural networks
Network Management	Classification Prediction Sequence analysis Time-series data analysis	Genetic algorithms Bayesian Belief Networks Rough sets Classification trees Self-organizing maps

4 BI and Data Mining Challenges in Telecommunications

The Telecommunications industry faces a number of BI and Data Mining challenges [1], [4], [5]. Data quality is always of primary concern and a prerequisite for achieving reliable and actionable results from the BI and Data Mining efforts. The acquiring of cross-functional or holistic view of data is extremely difficult for Telecommunication companies. Another important

issue is the more complex data pre-processing needed, and especially data aggregation to the appropriate semantic level, because of the availability of Telecommunications data in the form of transactions/events. Scalability is also a key concern because of the very large databases handled and maintained by Telecommunication companies. Real-time operation is of great importance as far as fraud detection and network fault detection are concerned. The prediction and identification of rare events is also a challenge for many Data Mining algorithms.

5 Conclusions

BI solutions and Data Mining play an important and increasing role in the Telecommunications industry due to the severe competition in the sector, the availability of large quantities of data generated and possessed by Telecommunication companies, and the existing technological capabilities for transforming that data into actionable knowledge to support strategic decision making. The main application areas include marketing, sales and CRM, fraud detection and network management. The implementation and enhancement of existing Data Mining techniques and methods, and the recent advances in the BI and Data Mining fields, especially in mining data streams, handling sequential and temporal data, classifying and predicting rare events, reveal their great potential for ensuring sustainable growth and compatibility of Telecommunication companies that are using them.

References

1. Pareek, D.: Business Intelligence for Telecommunications. Auerbach Publications, Taylor & Francis Group LLC (2007)
2. Sumathi, S., Sivanandam, S.: Introduction to Data Mining and its Applications. Springer-Verlag Berlin Heidelberg (2006)
3. Mattison, R.: Data Warehousing and Data Mining for Telecommunications. Artech House Inc. (1997)
4. Weiss, G.: Data Mining in Telecommunications. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers, pp1189-1201. Kluwer Academic Publishers (2005)
5. Weiss, G.: Data Mining in the Telecommunications Industry. In Wang J. (ed.), Encyclopedia of Data Warehousing and Mining, Second Edition, Information Science Publishing, Volume 1, pp.486-491 (2008)
6. Ngai, E., Xiu, L., Chau, D.: Application of Data Mining Techniques in Customer Relationship management: A Literature Review and Classification. In: Expert Systems with Applications 36, pp.2592-2602 (2009)
7. Costea, A.: The Analysis of the Telecommunications Sector by the Means of Data Mining Techniques. Journal of Applied Quantitative Methods, Vol.1, No.2, pp.144-150 (2006)
8. Vehvilainen, P., Hatonen, K., Kumpulainen, P.: Data Mining in Quality Analysis of Digital Mobile Telecommunications Network. In: Proceedings, XVII IMEKO World Congress, June 22-27, 2003, Dubrovnik, Croatia, pp.684-689 (2003)
9. Wei, C., Chiu, I.: Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach. In: Expert Systems with Applications 23, pp.103-112 (2002)
10. Hong, R., Yan, Z., Ye-rong, W.: Clustering Analysis of Telecommunication

Customers. In: The Journal of China Universities of Posts and Telecommunications, April 2009, 16(2), pp.114-116 (2009)

11. Sohn, S., Kim, Y.: Searching Customer Patterns of Mobile Service Using Clustering and Quantitative Association Rule. In: Expert Systems and Applications 34, pp.1070-1077 (2008)

12. Hilas, C.: Designing and Expert System for Fraud Detection in Private Telecommunications Networks. In: Expert Systems and Applications 36, pp.11559-11569 (2009)

13. Hung, S., Yen, D., Wang, H.: Applying Data Mining to Telecom Churn Management. In: Expert Systems and Applications 31, pp.515-524 (2006)

14. Ahn, J., Han, S., Lee, Y.: Customer Churn Analysis: Churn Determinants and Mediation Effects of Partial Defection in the Korean Mobile Telecommunications Service Industry. In: Telecommunications Policy 30, pp.552-568 (2006)

15. Hilas, C., Mastorocostas, P.: An Application of Supervised and Unsupervised learning Approaches to Telecommunications Fraud Detection. In: Knowledge-Based Systems 21, pp.721-726 (2008)

16. Shortland, R., Scarfe, R.: Data Mining Applications in BI. In: BT Technology Journal, Vol.25, Nos 3&4, July/October (2007)

17. Abascal, E., Lautre, I., Mallor, F.: Data Mining in a bicriteria clustering problem. In: European Journal of Operational Research 173, pp.705-716 (2006)

18. Berry, M., Linoff, G.: Data Mining Techniques: For Marketing, Sales and Customer Relationship Management. Wiley Publishing Inc. (2004)