IJITKM 0973-4414 Volume-9, Number-1 Jun-Dec 2015 pp. 20-23 **Impact Factor 2.8** available online at www.csjournalss.com

Data Mining Techniques for Fault Detection in Mobile Communication Networks

Poonam Chaudhary & Vikram Singh
Department of Computer Science & Applications
Chaudhary Devi Lal University, Sirsa (INDIA)

ABSTRACT: The rapid advancements in mobile communication technologies have introduced cheap, low-power and multifunctional devices. Analysis of data pertaining to mobile communication has oftenly been used as a ground for data mining research, such as mining frequent patterns and clusters on data streams, collaborative filtering, and analysis of social networks. Mobile communication network faults may cause malfunctions and outages, thereby putting the mobile service providers at a disadvantages vis-à-vis competitors. In this light, detecting network faults becomes paramount for mobile operators. This article presents a review of prevalent data mining techniques for fault detection in mobile communication networks.
Keywords: Mobile communication, fault detection, and data mining.

1. INTRODUCTION

Mobile communication data analysis has been often used as a background application to motivate many research problems in data mining. Although, very few data mining researchers have had a chance to see a working data mining system on real mobile communication data (Chang, L. et al. 2008).

Data mining is the process of automatically discovering useful information in large data repositories. Many times, traditional data analysis tools and techniques cannot be used because of the massive volume of data gathered by automated collection tools such as point-of-sale data, Web logs from e-commerce portals, earth observation data from satellites, genomic data. From time to time the non-traditional nature of the data implies that ordinary data analysis techniques are not applicable.

The data mining applications for any industry depend on two factors: the data that are available and the business problems facing the industry. This provides background information about the data maintained by mobile communications companies. The challenges associated with mining mobile communication data are also described.

Mobile communication companies maintain data about the phone calls that traverse their networks in the form of call detail records that contain descriptive information for each phone call. In 2001, long distance customers of AT&T network generated over 300 million call records per day (Cortes & Pregibon, 2001). The fact that such call records are kept for several months means billions of call records are readily available for data mining research. The call detail data are useful for marketing and fraud detection research and applications.

Mobile communication companies also maintain extensive customer information, like billing information, as well as information obtained from outside parties, such as credit score information. That information can be quite useful and often is combined with mobile communication-specific data to improve the results of data mining, e.g, while call detail data can be used to identify suspicious calling patterns, as a customer's credit score is often incorporated into the analysis before determining the likelihood that fraud is actually taking place.

Mobile communications companies also generate and store an extensive amount of data related to the operation of their networks. The network elements in these large mobile communication networks have some self-diagnostic capabilities that permit them to generate both status and alarm messages. All these streams of messages can be mined in order to support network management functions that are namely fault isolation and prediction.

Thus, the scalability of data mining methods is a key concern. A second issue is that mobile communication data is often in the form of transactions events and is not at the proper semantic level for data mining, e. g, one typically wants to mine call detail data at the customer (i.e. phone line) level but the raw data represents individual phone calls. Therefore it is often necessary to aggregate data to the appropriate semantic level before mining the data. Another is to utilize a data mining method that can operate on the transactional data directly and extract sequential or temporal patterns. Another issue arises because much of the mobile communications data is generated in real-time and many mobile re-communication applications, like fraud identification and network fault detection, need to operate in real-time. Because of its efforts to address this issue, the mobile communications industry has been a

leader in the research area of mining data streams (Aggarwal, 2007). One way to handle data streams is to maintain a signature of the data, which is a summary description of the data that can be updated quickly and incrementally.
A final issue with mobile communication data and the associated applications involves rarity. For example, both mobile communication fraud and network equipment failures are relatively rare. For predicting and identifying rare events has been shown to be quite difficult for many data mining algorithms (Weiss, 2004) and therefore this issue must be handled carefully in order to ensure reasonably good results.

2. NETWORK FAULT DETECTION

Telecommunication networks are extremely complex configurations of hardware and software. Almost the network elements are capable of at least limited self-diagnosis. These elements may collectively generate millions of status and alarm messages each month (Fawcett & Provost, 1997). Alarms must be analyzed automatically in order to identify network faults in a timely manner for effectively manage the network—or before they occur and degrade network performance. The proactive response is essential to maintaining the reliability of the network. So, volume of the data, and because a single fault may cause various different, seemingly unrelated, alarms to be generated, so the task of network fault isolation is quite difficult. The data mining has a role to play in generating rules for identifying faults (Han, J. et al., 2002).

2.1 Mobile network faults

Mobile network fault can be defined as an abnormal operation or defect at the component, equipment, or sub-system level that is significantly degrades performance of an active entity in the network or disrupts communication. Each and every error is not faults as protocols can mostly handle them. Mostly faults may be indicated by an abnormally high error rate. The fault can be defined as an inability of an item to perform a required function (a set of processes defined for purpose of achieving a specified objective), excluding that inability due to preventive maintenance, lack of external resources, or planned actions.

Characteristics: There is lack of a generally accepted definition of what constitutes behaviour of a normal mobile network fault (Hajji, et al., 2001; Hajji & Far, 2001; Lin & Druzdzel, 1997). Therefore, it is very difficult to characterize the mobile network faults accurately. However, there are estimations (based on statistics of the network traffic) as to what characterize a mobile network fault. The mobile network faults are characterized by transient performance degradation, high error rates, loss of service provision to the customers (i.e., loss of signal loss of connection, etc), delay in delivery of services and getting connectivity.

Causes: The main causes of network faults differ from network to network. Managing complex hardware and software systems has always been a difficult task. The Internet and the proliferation of web-based services have increased the importance of this task, while aggravating the problem (faults) in at least four ways (Meira, 1997; Thottan & Ji, 1998; Hood & Ji, 1997; Lazar et al., 1992).

- Mainly Speed of software development and release means less reliable and more frequently updated software.
- The multi-tier and distributed software architectures increase the complexity of the mobile network environment and obscure causes of both functional and performance problems.
- The internet style service construction implies more dynamic dependencies among the distributed software elements of the overall services making it difficult to construct and maintain accurate system models.
- The internet scale deployments increase the number of service elements under a particular administrator's responsibility.
- Many heterogeneous networks.
- The new innovations means interoperation of different networks must be kept to some level leading to faults.
- The overloading of power supply gadgets, natural disasters, etc.

3. DATA MINING FOR FAULT DETECTION

Data mining is an expanding area of research in artificial intelligence and information management whose objective is to extract relevant information from large databases (David, J. et al, 2002). Typical data mining and analysis tasks include classification, regression, and clustering of data, aiming at determining parameter or data dependencies and finding various anomalies from the data.

3.1 Grid Computing: Grid computing has been proposed as a novel computational model, distinguished from conventional distributed computing by its focus on large-scale resource sharing, innovative applications, and, in few cases, high-performance orientation. Nowadays grids can be used as effective infrastructures for distributed high-performance computing and data processing. A grid is a geographically distributed computation infrastructure composed of a set of heterogeneous machines that users can access via a single interface. Grids therefore, provide common resource-access technology and operational services across widely distributed virtual organizations composed of institutions or individuals that share resources.

3.2 Self-Organizing Map: SOM is an important unsupervised competitive learning algorithm, being able to extract statistical regularities from the input data vectors and encode them in the weights without supervision (Feher, K., 1995). Such a learning machine will then be used to build a compact internal representation of the mobile network, in the sense that the data vectors representing its behavior are projected onto a reduced number of prototype vectors (each representing a given cluster of data), which can be further analyzed in search of hidden data structures. The main advantages of their solution are the limited storage and computing costs. However, SOM requires processing time which increases with the size of input data.

3.3 Discrete Wavelet Transform: Discrete Wavelet Transform (DWT) is used to reduce the input data size, features of the data can be extracted without losing the significant data can be used for anomaly detection. Wavelets have been extensively employed for anomaly (Aquino & Barria, 2011) and fault detection (Yadaiah & Nagireddy, 2007). DWT has also been integrated with SOM to detect system faults (Xu & Zhao, 2002). In particular, feature vectors of the faults have been constructed using DWT, sliding windows and a statistical analysis. DWT is a mathematical transform that separates the data signal into fine-scale information known as detail coefficients, and rough-scale information known as approximate coefficients. Its major advantage is the multi-resolution representation and time-frequency localization property for signals. Usually, the sketch of the original time series can be recovered using only the low-pass-cut off decomposition coefficients; the details can be modeled from the middle-level decomposition coefficients; the rest is usually regarded as noises or irregularities.

3.4 CDLC and MIDP: The Connected, Limited Device Configuration (CLDC) and the Mobile Information Device Profile (MIDP) have emerged as J2ME standards for mobile phone applications development which are used with DMS services. The role of CLDC and MIDP component is to apply Data Mining Services in mobile.

Connected, Limited Device Configuration (CLDC): The Connected, Limited Device Configuration is a configuration specification in J2ME. The CLDC specifies the APIs for devices with less than 512 KB of RAM available for the Java system and a periodic (limited) network connection. It specifies a Java virtual machine with essential features, called the KVM, as well as several APIs for fundamental application services. Configurations provide core functionality and a way to provide greater flexibility but no services for driving the user interface, for managing the application life-cycle, for maintaining and updating continual data on the device or for secure access to information stored on a network server (Ortiz, 2004).

Mobile Information Device Profile (MIDP): Several networks have conducted a survey on users' watching behavior [29] which reflects that user behavior pattern recognition is not so easy task; we can achieve this by CLDC and MIDP component. Instead of replacing existing TV service, mobile services should be complementary, and offer more interactive means for users to watch their chosen content. The CLDC component specifies the connection between the MIDP profile and the connecting components with the server. The Mobile Information Device Profile (MIDP) is a specification for a J2ME profile. It is layered above the CLDC and adds APIs for application life cycle, user interface, networking, and continual storage.

4. CONCLUSION

Mobile communication services are a fast growing industry. Very limited patterns could be found from real data by human analysts thereby paving way for avenues of data mining research for pattern hunting in mobile communication data sets. Various data mining techniques are discussed for fault detection in mobile communication and further new technique will be introduced for fault detection.

References:

1. Feher, K., *"Wireless Digital Communications: Modulation and Spread Spectrum Applications*". Upper Saddle River, NJ: Prentice Hall, 1995.
2. Chang, L., Wang, T., Yang, D. and Luan, H.,*"Seqstream: Mining closed sequential patterns over stream sliding windows*". In *Proceeding of ICDM'08*, Pisa, Italy, December 2008.
3. Cortes, C., Pregibon, D., *"Signature-based methods for data streams.*" Data Mining and Knowledge Discovery 2001; 5(3):167-182,
4. Aggarwal, C. (Ed.). (2007), "*Data Streams: Models and Algorithms".* New York: Springer. Weiss, G. M., Provost, F., *"Learning when training data are costly: The effect of class distribution on tree induction*". Journal of Artificial Intelligence Research 2003; 19:315- 354.
5. Fawcett, T., Provost, F. ,*"Adaptive fraud detection. Data Mining and Knowledge Discovery*" 1997; 1(3):291-316.
6. Han, J., Altman, R. B., Kumar, V., Mannila, H., Pregibon,*" D. Emerging scientific applications in data mining*". Communications of the ACM 2002; 45(8): 54-58.
7. Hajji, B. & Far, B. H. (2001), *"Continuous Network Monitoring for Fast Detection of Performance Problems", Proceedings of 2001 International Symposium on PerformanceEvaluation of Computer and Telecommunication Systems*, July 2001.
8. Hajji, B.; Far, B. H. & Cheng, J. (2001), *"Detection of Network Faults and Performance Problems*", *Proceedings of the Internet Conference*, Osaka, Japan, Nov. 2001.
9. Lin, Y. &Druzdzel, M. J. (1997), *"Computational Advantages of Relevance Reasoning in Bayesian Belief Networks", Proceedings of the Thirteenth Annual Conference inUncertainty in Artificial Intelligence (UAI-97),* pp. 342-350, Morgan Kaufmann Publishers, Inc., San Francisco, CA, 1997.
10. Meira, D. M. (1997), *"A Model for Alarm Correlation in Telecommunications Networks",PhD Thesis*, Federal University of Minas Gerais, Belo Horizonte, Brazil, Nov. 1997.
11. Thottan, M. & Ji, C. (1998), *"Proactive Anomaly Detection Using Distributed Intelligent Agents"IEEE Network*, Sept./Oct. 1998.
12. Hood, C. S. & Ji, C. (1997), *"Proactive Network Fault Detection", Proceedings of the IEEE INFOCOM*, pp. 1139-1146, Kobe, Japan, April 1997.
13. Lazar, A.; Wang, W. & Deng, R. (1992), " *Models and algorithms for network fault detection and identification: A review", Proceedings of IEEE ICC, Singapore, pp.999-1003, November 1992.*
14. David J. Hand, H. Mannila, and P. Smyth, *"Principles of data mining"*, MIT Press, 2001
15. V.A. Aquino and J.A. Barria, *"Anomaly detection in communication Networks using wavelets,"IEEE Proc. in Commun*ications, vol.148, no.6, pp. 355-362, Dec. 2001
16. Yadaiah, N. and Nagireddy, R., *"Fault detection techniques for power transformers,"*Industrial & Commercial *Power Systems Technical Conf*., pp. 1- 9, 2007.
17. Xu, Z. and Zhao, Q., *"A novel approach to fault detection and isolation based on wavelet analysis and neural network,"Electrical and Computer Engineering,* vol. 1, pp. 572–577, May. 2002.
18. Ortiz, 2004, *"A Survey of J2ME Today*", Sun Developer Network (SDN)