

Comparative Study of Pattern Mining Techniques for Network Management System Logs for Convergent Network

Bodhisattwa Gangopadhyay^{1,2}, Artur Arsenio^{1,2}, Claudia Antunes¹

¹ Instituto Superior Técnico,
Av. Rovisco Pais, 1, 1049-001 Lisboa

² Nokia Siemens Networks Portugal S.A.
Rua Irmãos Siemens 1-1A, 2720-093 Portugal
{Bodhisattwa.Gangopadhyay, Artur.Arsenio, Claudia.Antunes}@ist.utl.pt
{Bodhisattwa.Gangopadhyay, Artur.Arsenio}@nsn.com

Abstract.¹ The concept of Pattern Mining has obtained significant focus in Telecommunications Network Management System (NMS). A big volume of literature has been dedicated to this field and valuable progress was also observed. Both sequential and structured pattern mining techniques were observed to be considered in NMS. In particular NMS logs (Performance and Alarm) pose several interesting issues for pattern mining. Pattern mining can help in various NMS activities such as alarm correlation, alarm associations, self-healing or pro-active fault management. In this review article, we present an overview of the different pattern mining techniques used in NMSs, compare them and finally select the best that can be beneficial to NMS for Radio over Fiber (RoF) like convergent networks. The pattern mining technique will be one of the basic steps that will be needed to implement various data processing functionalities (such as using sequential pattern mining to extract episode rules from network system alarms) of an intelligent NMS.

Keywords: Pattern Mining, Radio-over-Fiber, Network Management Systems.

1 Introduction

Communication networks with several radio access technologies, such as in the FUTON [1] infrastructure, have increasing complexity in terms of various network elements and also the network data generated, in a single infrastructure [1] [2]. This generated data in large quantity related to faults, warnings and KPIs must be

¹ This work is supported in part by the European Commission, in the context of the project FUTON “Fibre Optic Networks for Distributed, Extendible Heterogeneous Radio Architectures and Service Provisioning”, grant agreement FP7 ICT-2007-215533
Bodhisattwa Gangopadhyay wishes to thank Fundação para a Ciência e a Tecnologia, Portugal, for support under grant SFRH/BDE/33799/2009.

monitored for a rapid decision making process to enable a network management with self-healing and pro-active fault management capacities.

Network management systems are aimed to possess the capability of pro-actively managing network configurations, traffic routing, fault and performance management in real time, thus looking towards optimizing traffic throughput in the network [3].

Data mining technology is imperative to the requirements mentioned here for extracting the hidden knowledge about the behavior of the network. The knowledge discovery in databases has different parts which involve data collection and cleaning, pattern discovery, post-processing of the discovered knowledge and finally using this discovered knowledge [4]. The authors in [5] argue that it is highly desirable that networks have the ability for detecting symptoms of the network exceptional conditions and enable proactive network management.

Pattern mining is the data mining technique which is employed to find the existing patterns in the available databases. The word "pattern" in the term "pattern mining" is also interpreted as the association rules. Though the original motivation behind this was related to analyze of supermarket data, afterwards several works have been undertaken to employ pattern mining in telecommunication network management systems. History of pattern mining dates back to 1993 in the published article of Agarwal et.al. related to the market basket analysis in the form of association rule mining [6]. Work related to application of pattern mining in telecommunications is found in [7] where the authors propose a solution procedure by employing constrained based sequential pattern mining for mining sequential alarms from the alarm data of a Global System for Mobile (GSM) system. Observing the features of the alarm data, data is cleaned without compromising the quality of the obtained sequential alarm patterns. In telecommunication alarm correlation analysis for large alarm databases, mining of association rules is one of the primary methods while the efficiency of the algorithms is very important for these large datasets [8]. The work in [8] proposes an algorithm based on layered frequent pattern tree for mining frequent patterns and demonstrated a better time and space efficiency.

Call detail data, network data and customer data are the types of telecommunication data. Our focus will be on network data (alarms, KPIs) where NMS acts on.

This paper concentrates on the study of the different pattern mining techniques aimed towards data originating from telecommunication network management systems. Specifically, the focus is on data related to the alarms and KPIs originating from the communication networks. The goal of this work is to study and find the best pattern mining technique/algorithm which is available for convergent networks (wireless and fixed) and the work falls in the scope of designing the pro-active network management system that can suit the best for heterogeneous network architectures such as FUTON. Pattern mining algorithms have attracted a lot of interest from researchers in recent times related to communication network data. Selecting the wrong pattern mining technique can result in lower detection rates or an increased number of false alarms [9]. This work studies and compares the different state-of-the-art pattern mining techniques/ algorithms and evaluates the best for a heterogeneous environment. As the alarm data have time related characteristics,

sequential pattern mining algorithm is best suited for such kind of data. This paper presents an overview of state-of-the-art research work (related to data mining) relevant for RoF management system, which is still a scientific field lacking significant previous research work applied to RoF scenarios.

2 Pattern Mining and NMS

Knowledge Discovery in Databases (KDD) is used to find hidden knowledge from existing large databases and NMS alarm or performance data are perfect example of such databases. An example of an operation in network management is alarm correlation where the KDD method is often applied. The KDD method is comprised of various steps such as: knowledge requirement setting, data selection, data mining, result interpretation, and knowledge incorporation [10]. In the context of NMS, data mining techniques are restricted to pattern extraction or pattern mining in the available data [9].

The data mining process is also considered to be an iterative process which is comprised of a number of steps.

- Pre-processing
- Choice of analysis methods
- Analysis
- Interpretation and evaluation of results [9].

A pioneer work in this field can be considered the Telecommunication Alarm Sequence Analyser (TASA) system [9], which exhibited the usage of knowledge discovery methods for almost the first time in the industry and was used for the knowledge extraction from the network alarm databases [9]. The pattern discovery used in TASA could be briefly mentioned as follows [4]: “Given a sequence S of alarms, a set E of alarm predicates, a class \square of episodes built from the predicates of E , a frequency threshold c , and window widths W and W' , find the confidences of all rules (\square, e, W, W') whose frequency is at least c .” TASA discovers episodic rules between alarms based on the temporal relationship existing between them [4]. An example of a rule can be cited of the following: “if alarms of type link alarm and link failure occur within five seconds, then an alarm of type high fault rate occurs within 60 seconds with probability 0.7”[11].

Also taking into account the fact that single fault conditions trigger multiple correlated alarms, fault management was proposed in [12] which took care of correlated alarms. In this scope they developed “Intelligent search of interesting patterns in sequences” to find correlated alarms. Alarm patterns on the basis of topology model were presented to the network operator for further decision related to matching the alarm pattern to a single triggering event.

The work done in [13] proposes a mining algorithm that used time constraints for restricting the time between the alarms. They also focused on a method for discovering sequential alarm patterns. But this work was limited to cleaning alarm

events and finding sequential alarm patterns but did not provide correlated alarms for further processing.

3 Telecommunication alarm databases

It is well known that the first step in pattern mining lies in the understanding of the data. In case the generated data is not suitable for pattern mining, then possible steps are needed to change the raw data into a format which makes it suitable for pattern mining [10]

Before going to telecommunication alarm databases, it is worth mentioning the telecommunication network which is comprised of a network of links and nodes where messages are passed from one part of the network to another over multiple links and through various nodes. Switches, exchanges, transmission equipments are example of some of the components which form the network [4]. Whenever there is an abnormal condition in the network, alarms are generated by each of the components, sub-components and software modules. These alarms are normally generated in form of messages and contains the time-stamp, the address of the node/component from where the alarm was generated, the severity of the alarm and the status of the alarm (whether cleared or not). The NMS at the operation and maintenance centre of the network receives these alarms and store them in the database. Also, the alarms are reported to the network operator for taking appropriate actions to eradicate the cause and bring the network back to normal.

Processing of this huge amount of alarms are done in steps such as alarm filtering, alarm correlation and so on. To perform alarm filtering and correlation, apriori knowledge is required. The alarm filtering and correlation can be seen as processes to reduce the number of alarms to be presented to the network operators. But requirement still stays to predict such faults depending on the Key Performance Indicators (KPIs) regularly generated by the network. Pro-active fault management can be considered to be an example of this.

4 Trends of Pattern Mining usage in NMS

This section tabulates the comparative evaluation of all the eight algorithms described in the previous stage and TASA.

The a priori knowledge is required for the pattern mining algorithms for telecommunication data and this includes domain knowledge as well as detailed structural information [9]. With propagation of time, the nature of the system might change and this forces a change in the a priori knowledge. As studied by [9] and the functional part of TASA system, there are certain requirements needed by the data mining (knowledge discovery methods) for the telecommunication data which are listed below [7]: “

- Application domain terminology and semantics used in user interface

- Immediate, accurate and understandable results
- Easy-to-use methods
- Interfaces and integrability towards legacy tools
- Adaptability to process information
- Use of process information
- Efficiency and appropriate execution time
- Reduced iterations per task
- Easy to learn
- Increases efficiency of domain experts by reducing time spent per task”

It is evident from the studies made that, pattern mining in transactional databases, mining classification rules for classifying data dependent on interesting features are some of the various pattern mining applications. Pattern mining in sequential patterns is the knowledge discovery being done on any ordered data. But it can be inferred that, important information can be extracted from alarm data generated by telecommunication system by applying pattern mining techniques related to sequential data. The important factor for pattern mining for telecommunication alarm data is the time difference between two events is an important factor. This factor puts forward the relevance of

The algorithm described in [14] is based also on an a priori approach. They introduced new criteria for the selection of sequence where sequence plausibility and coherence is evaluated in the network topology context. This algorithm is based on an assumption that depends on the closeness of the alarm-generating elements. They assume that closer the alarm generating elements a much greater relationship is plausible between the alarms. A priori knowledge related to the network configuration is not totally relied upon. The topographical information encoded within the alarms is used instead. It is evident from this work that, usage of the topographical information encoded within the alarms gives a much better performance of the pattern mining and hence the data mining algorithms.

WINEPI algorithm described in [15, 16] uses a priori knowledge and aims toward mining temporal association rules in sequential data by employing sliding time window by employing a sliding time window. The maximum and minimum sequence duration constraint of the time window is specified by the user. This algorithm is capable of efficient identification of event sequences in sequential input data. But as the WINEPI algorithm have a constraint on the amount of data processing and time (both time and space), a formulation was done to reduce the number of non-effective windows, known as the EW-WINEPI algorithm. This work could bring a inference that, the number of episode rules which could be discovered depends on the time window, and with increase in the minimum Support and minimum confidence, the number of episode rules become smaller.

In the work presented in [7], it is observed that data cleaning is desirable before proceeding with pattern mining. As mentioned before, the work proposed in [7] also utilized the time constraints for restricting time difference between two alarm events. Also the quality of knowledge discovered was evaluated and the data cleaning was found to be an important measure for improving the functioning of the pattern mining

algorithms. The quality of evaluation was determined by finding the ratio between difference in execution time before and after cleaning with execution time before cleaning. Thus differing from normal pattern mining algorithms, it becomes a fact that telecommunication alarm data contains not only valuable information but also lots of dummy information. Also for telecommunication alarm data, the reduction time in execution of pattern mining algorithm is an added value with pre-cleaning of the data.

The work in [17, 18], instead of acting on the time sequences as other works, concentrates on discovery of the structural data as the data collected from telecommunication network management systems is found to have explicit or implicit structural components. They used a computationally-constrained beam search.

Also, [19] looked into pattern mining of patterns which have mutual dependency. Telecommunication alarm data has exhibited a tendency to have mutual dependencies in alarm patterns. The mutual dependencies result from physical dependencies manifested as a set of events when an event is triggered. Thus mutual patterns can be used for constructing signatures for problematic situations. Also redundancy in alarm data generated by monitoring elements results in mutual alarm patterns, which makes it worth to inspect the mutual pattern mining.

[20] predicts telecommunication equipment failures from the network alarm sequences. This work is aligned with normal sequential pattern mining algorithms.

The work in [21,22] looked into the distributed pattern mining where the alarm logs are divided into smaller logs according to topological location. This is for specific implication that pattern mining is done for alarms generated from the locally generated alarms. This marks a major difference from state-of-the-art sequential pattern mining algorithms. The distributed pattern mining also results in scalability issue and this also justifies the fault detection and resolution system. This also contributes to the resilience of the telecommunication system.

Overall, it can be concluded that pattern mining for telecommunication data follows sequential pattern mining techniques with slight variations, such as noticing the time stamps, the time difference between the data generated. In certain cases, distributed processing of the data from the telecommunication data are answers to scalability and resilience issues. It was also noticed that, directly running the pattern mining algorithms on telecommunication data is not the best method. A data pre-processing gives much better result in terms of time and accuracy when compared with the previous. This is because raw telecommunication alarm data contains dummy information which is many times replicated. Distributed processing does not fit for convergent radio-over-fiber infrastructures (such as case of FUTON) as a centralized processing is desired with all the computing facility stationed in the Central Unit overlooking the network and the alarm generating elements are very simple with least computational power.

5 Requirements for RoF like networks

In the previous sections we had an overview related to the state-of-the-art pattern mining techniques applicable to the alarm handling in different network management systems. Fault or performance management in Convergent Networks implies the management of all those networks which forms portion of the integrated mobile and fixed wireless networks. This survey makes us conclude that, almost no work was done related to the application of data/pattern mining in fault management or alarm handling specific to convergent networks. But it is also true that, the proposals and results obtained by the researchers in using data/pattern mining in fault management replicates many scenarios obtained in convergent networks. The specific area of interest in our case lies in RoF like convergent networks where alarms are generated both from fixed and wireless networks.

The RoF part of the FUTON system is not prone to large number of alarms. For instance, a passive optical network (PON) is susceptible to fewer errors and alarms when compared to wide area network (WAN). This situation is similar to [20] where alarms are not generated in abundance and the training data is hard to be found. On the other hand there are new challenges: the correct identification of the faulty source, the impact on deployed services of such faults or performance impairments, and assurance of service levels over heterogeneous mobile networks are similar problems faced in all kinds of networks. The work proposed in [21] is a better solution for a high speed network as in FUTON but the distributed processing of the faults oppose the idea of central processing of FUTON RoF Manager based on the fact that a centralized network management system is required to manage a highly distributed network for faults, congestion, and to assure adequate QoS across the whole network.

The Fault Management module of the FUTON RoF Manager requires functionalities such as fault localization, alarm correlation, pro-active fault management and so on. The referred literature in this survey has definitely shown that usage of pattern mining/ data mining is sure to add value to the system in terms of intelligence for achieving the required fault management functionalities. It is definite that frequent sequential pattern is the kind of patterns that will be found in the alarms generated from the RoF infrastructure. This is similar to alarms from other networks. The algorithm that offers the best match for such a scenario is yet to be decided and out of the scope of this work. Apriori algorithm can be a solution as the number of candidates would not be very large. But it is certain that the algorithm needs to process numerical data rather than symbolic data.

Pattern mining for telecommunication alarm data stays limited to the steps such cleaning of alarm events and discovering sequential alarm patterns, and does thus not provide correlated alarms adopted for further processing. Further, data processing algorithms are thus a requirement for pro-active fault management.

6 Conclusions

This paper presents a survey on the usage of pattern mining in the area of network management systems. Specifically, interest is shown at fault management systems. Brief descriptions about pattern mining methodologies are stated at the beginning followed by the application of pattern mining techniques in Network Management Systems. The most famous and initial work in network management systems involving pattern mining techniques, TASA is also analyzed. We also briefly describe the kinds of data specific to telecommunication alarms on which pattern mining needs to be applied. We put forward eight state-of-the-art works besides TASA, from the field of pattern mining as applied on alarms originating from telecommunication networks. At the end a critical analysis is done on the requirements of the pattern mining technique needed for FUTON RoF like convergent networks. This work highlighted some of the issues that affect the ability of pattern mining such as big volume of alarms generated from some part of the networks (e.g. wireless) or very rare events (such as alarms generated from PONs) or ability to work in real-time (using distributed data mining).

Application of pattern mining techniques in network management systems (especially fault management) is one of the oldest examples of usage of data mining because of the huge data volume that is collected. We have critically analyzed the requirement of pattern mining techniques for convergent networks case specific scenario. Pattern mining usage is sure to accelerate for giving operators competitive advantage for performing alarm correlation or fault localization.

References

1. Pato S., Pedro J., Santos J., Arsénio A., Inácio P., Monteiro P., "On Building a Distributed Antenna System with Joint Signal Processing for Next Generation Wireless Access Networks: The FUTON Approach". 7th Conference on Telecommunications, Portugal.
2. I.Santiago C., Gangopadhyay B., Arsenio A., Ramkumar M.V., Prasad Neeli R., "Next Generation Radio over Fiber Network Management for a Distributed Antenna System", Wireless Vitae 2009, Aalborg, Denmark, 19 May, 2009
3. I.Burn-Thornton, K.E., Garibaldi J., Mahdi A.E., "Pro-active Network Management Using Data Mining", Global Telecommunications Conference 1998, Globecom'98, Vol.2, Page(s): 1208-1211
4. Toivonen H., Ronkainen P., Mannila H., Klemettinen M., Hättönen K., "Knowledge Discovery from Telecommunication Network Alarm Databases"
5. I.Kulkarni, P.G., McClean, S.I., Parr, G.P., Black, M.M., "Deploying MIB Data Mining for Proactive Network Management", 3rd International IEEE Conference on Intelligent Systems, Page(s): 506-511, September 2006
6. I.Agarwal, R., Imielinski, T., Swami, A., "Mining Association Rules Between Sets of Items in Large Databases", SIGMOD Conference 1993, Page(s): 207-216
7. I.Ouh, Jain-Zhi., Wu, Pei-Hsin., Chen, Ming-Syan., "Experimental Results on a Constrained Based Sequential Pattern Mining for Telecommunication Alarm Data",

Second International Conference on Web Information Systems Engineering Vol. 2, 2001

8. Li, Tong-yan., Li, Xing-Ming., "A LFP-tree based method for association rules mining in telecommunication alarm correlation analysis", The Journal of China Universities of Posts and Telecommunications, July 2007
9. Hätönen K., "Data mining for telecommunication network log analysis", PhD Thesis, Series of Publications A, Report A-2009-1
10. Pekko Vehviläinen, Kimmo Hätönen, and Pekka Kumpulainen. "Data mining in quality analysis of digital mobile telecommunications network." In Proceedings of XVII IMEKO World Congress, pages 684 – 689, Dubrovnik, Croatia, June 2003. IMEKO.
11. Weiss, G.M., "Data Mining in Telecommunications", Dept. of Computer and Information Science, Fordham University
12. Tuchs, K.D., Jobmann, K., "Intelligent Search for Correlated Alarms Events in Databases", International Symposium on Integrated Network Management Proceedings, Page(s): 285-288, 2001
13. Jain-Zhi O., Pei-Hsin, W., Ming-Syan, C., "Experimental Results on a Constrained based Sequential Pattern Mining for telecommunication alarm data", Proceedings of the Web Information Systems, December 2001
14. Devitt A., Duffin J, Moloney R, "Topographical Proximity for Mining Network Alarm Data", SIGCOMM'05 Workshops, 2005
15. H. Mannila, H. Toivonen, and A. I. Verkamo, "Discovery of frequent episodes in event sequences." Data Mining and Knowledge Discovery, Page(s):259–289, 1997.
16. Hou S., Zhang X., "Alarms Association Rules Based on Sequential Pattern Mining Algorithm", Fifth International Conference on Fuzzy Systems and Knowledge Discovery, 2008
17. Baritchi A., Cook D.J., Holder L.B., "Discovering Structural Patterns in Telecommunication Data", FLAIRS'00 Proceedings, American Association for Artificial Intelligence, 2000
18. Cook D.J., Holder L.B., Djoko S., "Scalable discovery of informative structural concepts using domain knowledge.", IEEE Expert, 11(5), 1996.
19. Sheng M., Hellerstein J.L., "Mining Mutually Dependent Patterns for System Management", IEEE Journal on Selected Areas in Communication, Vol. 20, No.4, Page(s): 726-736, 2002
20. Weiss G.M., "Predicting Telecommunication Equipment Failures from Sequences of Network Alarms", Handbook of Knowledge Discovery and Data Mining, Oxford University Press
21. Gardner R.D., Harle D.A., "Fault Resolution and Alarm Correlation in High Speed Networks using Database Mining Techniques", International Conference on Information, Communications and Signal Processing, Page(s): 1423-1428, Singapore, 1997
22. Manilla H., Toivonen H., Verkamo A.I., "Discovery frequent episodes in sequences", First International Conference on Knowledge Discovery and Data Mining, Page(s): 210-215, Canada, 1995