ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/285336538

A Survey of Network Faults Classification Using Clustering Techniques

Article · October 2013

CITATIONS
2

READS
230

2 authors:

Karwan Qader
Cihan University-Erbil
12 PUBLICATIONS **45** CITATIONS

SEE PROFILE

Mo Adda
University of Portsmouth
115 PUBLICATIONS **759** CITATIONS

SEE PROFILE

All content following this page was uploaded by Karwan Qader on 01 December 2015.

The user has requested enhancement of the downloaded file.

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

A Survey of Network Faults Classification Using Clustering Techniques

Karwan Qader[1], and Mo Adda[2]

University of Portsmouth, School of Computing

Buckingham Building, Lion Terrace, PO1 3HE Portsmouth, Great Britain

Abstract. The last decade has witnessed an increasing usage of data mining techniques such as clustering into many applications including network faults classification. In communication networks often large volume of network faults are generated. Even a single fault may result in large number of alarms causing information redundancy. Providing a coherent classification scheme would certainly help network management process, avoid system breakdown, by isolating faults earlier, and predict peculiar events. In this paper, we survey different clustering algorithms to classify network faults and provide an evaluation based on the speed, accuracy, efficiency and cost. From accuracy point of view some of the algorithms yield high accuracy as indicated in the literature. Out of them Fuzzy Cluster-Means is considered more suitable for network faults classification based on the context in which they are applied.

Index Terms Clustering, classification, network faults, artificial intelligence, neural networks

I. INTRODUCTION

Computer networks are witnessing unprecedented growth in different forms such as wired and wireless networks. With the growing use of computer networks, the topics of network fault management and data mining techniques have risen to prominence. In essence, any single fault may trigger a significant amount of network faults which emphasises the need for clustering algorithms that aim to isolate faults in advance, predict the behaviour of particular events and overall, aid in the network management process.

Over the recent years, a plethora of clustering techniques has emerged in the existing body of research. While each of the techniques has its advantages and disadvantages, the on-going academic debate has failed to result in a general consensus regarding the most suited approach towards network faults classification. However, this paper attempts to identify suitable clustering algorithms for the mentioned task. Various clustering algorithms are reviewed and compared in terms of speed, accuracy, efficiency and performance. The speed indicates the time taken to perform faults classification. Accuracy denotes the correctness of classifications without false positives. Efficiency represents the level of resource consumption while the performance denotes the improvement of speed and accuracy when compared with other algorithms.

The presented paper therefore builds on the on-going academic debate and aims to provide a critical survey of network faults classification using clustering techniques. The key contribution of this study can be found in the enhancement of the existing body of knowledge by providing a comparative analysis of the available methods and techniques.

The paper is organised into six main sections. Following the brief introduction to the studied topic, the second section outlines the underlying theoretical background for the study of data mining and clustering techniques as well as the types of faults frequently occurring. Building on the conclusions drawn in the second section, the third section provides an overview of the key studies in the existing literature which provides further insight into various approaches used in network faults classification. The fourth section compares various algorithms and techniques used for network fault classification and the fifth section provides a critical evaluation of their effectiveness and limitations. The sixth and the final section of the presented paper summarises the key findings and draws wider conclusions.

II. RELATED WORKS

Many techniques have been proposed for network faults detection and classification. Though there has been considerable research in this area, it is still considered a challenging task to detect network faults and classify them. This is due to ever growing network flows and the types of faults.

The wireless networks are of many types such as Wireless Sensor Networks (WSNs) [1], Mobile Ad Hoc Networks (MANETs) [2], Vehicular Ad Hoc Networks (VANETs) [3] and Wireless Mesh Networks (WMNs) [4]. In all such networks there are a set of functions meant for detecting, isolating and correcting network malfunctions. This phenomenon is known as network wireless fault management which is an important part of network management discipline. The heuristic algorithm proposed in the paper is capable of providing near optimum

description of the faults. The paper in [5] provides an overview of network fault management activities. These are automated activities relying on artificial intelligence techniques such as Content Based Reasoning (CBR) and Neural Networks (NNs) as the traditional human actions for fault classification is not feasible. As explored in [6] the network faults can be of hardware faults or software faults that are to be handled separately. The hardware faults occur due to the incorrect design of hardware components while the software faults can occur due to the bugs in the software elements which are unknown. Both kinds of faults need certain steps [5] to be followed in order to manage faults efficiently. They compared their algorithm with the similar one proposed in [7] where a general methodology is presented for fault identification. The main drawback of [7] is that the algorithm depends on prior information that needs to be either obtained experimentally or simply guessed. Zheng Rong Yang et al. [3] analyzed faults related analogue integrated circuits as integrated circuits are widely used. Identification and classification of such faults is the problem they solved using neural networks and statistical methods. Their proposed neural network structure is used to cluster the faults detected with different means and variances. This technique known as robust heteroscedastic probabilistic neural network outperformed general classification methods. Irene Katzela and Mischa Schwartz [8] studied network faults pertaining to telecommunications domain and understood that a single fault can result in large number of faults that make it difficult to know the source of failure and its makes it worse when there are multiple faults. For fault diagnosis they proposed a heuristic algorithm which correlates alarms and localize faults. They have achieved it by taking into account dependencies among objects in the network. The algorithm was capable of providing near-optimum explanation of the alarms received.

Wei Fu et al. [9] proposed classification and pattern recognition techniques for diagnosing faults in case of electromechanical systems. They used principal component analysis (PCA) and classification technique. Though it is not directly related to network faults, the usage of pattern recognition is a significant step forward in the research. Xingyan Li and Lynne E. Parker [10] proposed an approach named SAFDetection meant for sensor analysis based fault detection for monitoring multi-robot team tasks. This has significance as there is constant increase in the usage of robotic application in the real world. Each deviation of robot from expected behaviour is considered a fault. Three techniques are used to detect faults based on the sensor data. They have improved the SAFDetection approach by using PCA along with these three techniques which are Principal Component Analysis (PCA), time invariant state transition probabilities, and Probabilistic Clustering Algorithm for best fault detection results. Mouhammd Al-Kasassbeh and Mo Adda [11] studied Mobile Agents (MA) with intelligence and capabilities to move to various nodes in the network and gather information known as Management Information Base (MIB) for analysis. Their proposed approach for network fault detection is based on Weiner filter that analyze abnormalities in MIB variables. This approach takes advantage of the correlation matrix to detect abnormality in the traffic. The correlation matrix holds data of cross – correlation with required MIB variables and input MIB variables. They showed how to detect faults with respect to server crash, link failure, broadcast storm, and babbling node with both light and heavy scenarios.

In [12] a fault diagnosis method is proposed which is known as Input Output Classification Mapping (IOCM). It also uses Radial Basis Function Neural Network (RBF) [9] and maps input parameters to the output parameters of RBF. From input signals, input parameters are extracted and then decomposed using wavelet decomposition. Afterwards the faults are classified using IOCM. To detect and identify faults, the experiments are done using a mechanical system called Tennessee Eastman Challenge Process (TECP). Pattern recognition and classification techniques are used in [9] for fault diagnosis. The experiments are done on electrical power vehicles in transportation industry to demonstrate the proof of concept of fault detection and classification. There are many techniques for detecting faults in a system such as qualitative reasoning based on neural networks, numerical modelling, analytical modelling, and fuzzy reasoning [13]. However, in [9] phenomenological modelling is used for fault diagnosis.

Fuzzy Cluster Means (FCM) is used in [14] for senor fault detection. This algorithm plays an important role in solving problems in many areas including fuzzy intelligent control, pattern classification, and to classify fault patterns. FCM is most widely used algorithm for classifying faults and pattern classification. It can classify network faults into two or more clusters thus making the management of network faults easy which it makes use of Euclidean Distance for measuring similarity. Clustering algorithms can also be used for quality analysis and detect fault prone module in a software product. It does mean that clustering algorithms take a software product with different programming constructs, identify and classify the faults predicted, as demonstrated by Sandhu, Kaur & Kaur [15] who propose a density-based clustering algorithm for detecting faulty module in the software in the early stages to avoid maintenance problems. It used different metrics for finding the faults in software systems. Out of many techniques such as neural network, fuzzy logic, kalman filter, Bayesian, least square and fuzzy cluster-means, the FCM is used for intended solution to detect faults. Artificial Ant Cluttering technique is used in [16] for detection and diagnosis of faults. The faults are electrical and mechanical in nature in induction motors. This technique is an intelligent approach that detects faults and diagnoses effectively. It is an unsupervised technique which is inspired by ants in the real world.

Mohammad Mokhtare et al. [17] used yet another clustering technique for fault detection. The technique is known as Multistage Gath-Geva Clustering. The author stated that FCM and Fuzzy Gustafson-Kessel (FGK) algorithms are very popular for clustering. However they used Gath-Geva Clustering for the task of fault detection and isolation. The problem of fault detection and isolation

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

is considered by the algorithm as a pattern classification problem.

E. Denise W. Gürer, Irfan Khan, Richard Ogier [5] discussed many AI (Artificial Intelligence) techniques for network fault management and then proposed a hybrid AI solution based on both case based reasoning techniques and neural networks. According to them the drawbacks in the traditional clustering techniques led to the usage of AI technologies, Neural Networks (NNs), Bayesian Belief Networks (BBNs) and Case Based Reasoning (CBR) systems. The hybrid system proposed by them is as shown in fig. 1.

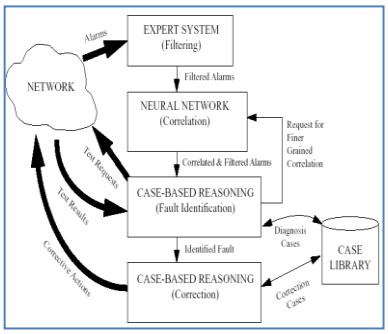

Fig. 1. Hybrid Approach to Network Fault Management [5]

As seen in figure 1, an Expert System takes network alarms as input and filters them through suppression, count, compression and generalization. Filtered alarms are given to NNs for correlation that returns correlated and filtered alarms. The output is fed to case based reasoning for further testing and analysis. The CBR makes use of prior experience too to make decisions on taking more data solving problems and running data through NNs again etc. Finally the hybrid system results in classification of faults. Thus the hybrid system helps in classifying faults as part of the whole automated fault management process as shown by Al-Kasassbeh and Adda [11].

In [18] a neural network is explored for analogue fault detection and classification. The technique is named as Heteroscedastic Probabilistic Neural Network (RHPNN). As stated in [19] neural networks have artificial intelligence which they are of different kinds such as Back Propagation Neural Network (BPNN), Probabilistic Neural Network (PNN), Self-Organizing Mapping (SOM) and Radial Basis Function Neural Network (RBF). Dongsheng Wu et al [20] present an algorithm for classification of industrial system faults. The algorithm is based on two techniques namely Probabilistic Neural Network (PNN) and K-Means clustering. The K-Means is used to make clusters while the PNN is used for diagnosing faults. According to Dongsheng Wu et al. K-Means is widely used for clustering algorithm due to its simplicity and performance and with respect to Artificial Neural Networks (ANNs), they prefer PNN which makes use of supervised learning in feed forward approach. The PNN's advantages include rapid training speed, guaranteed convergence, and support for incremental training. To avoid misclassification problem, both K-Means and PNN are used [20].

III. COMPARISONS OF TECHNIQUES FOR SYSTEM FAULTS CLASSIFICATION

This section provides a comparative study of the available techniques or algorithms for network faults classification. In the literature many approaches are found that include techniques based on artificial intelligence, pattern recognition, classification, Fuzzy Cluster-Means, Multistage Gath-Geva Clustering, neural networks, and Artificial Ant Clustering etc. In [18] Robust Heteroscedastic Probabilistic Neural Network (RHPNN) is used for detection and classification of analogue faults. The PNN works classified given data objects based on the conditional probability density functions explored in [21] and [22]. However, the PNN is more difficult to train. To solve this problem RHPNN is introduced in [18] which take fewer steps for convergence, and less testing time. It needs only 20 steps for convergence while BPNN needs 10,000 times. The training time taken by RHPNN is only 0:00:59 seconds while PNN and BPNN take 00:1:15 and 1:17:00 respectively. The testing time taken by RHPNN is only 00:06 seconds while PNN and BPNN take 01:38 and 03:07 minutes respectively. In the case of rapid transit vehicles, pattern recognition approach is used and it is proved to be very adequate for fault detection as proposed in [9]. The classification accuracy of this approach is reasonably more as the clusters conformed to the known behavioural patterns in vehicles. This is achieved even without performing extensive feature selection. However, the experiments are carried out on relatively small dataset. SAFDetection system proposed by Xingyan Li and Lynne E. Parker [10] improved version that uses the Principal Component Analysis (PCA), probabilistic clustering algorithm and time-variant. The state transition probabilities have been changed in order to define the operation of the robot system more precisely. The experiments are done to detect faults in tightly-coupled multi-root team tasks. They compared three algorithms namely K-Means, Soft K-Means and FCM. With respect to accuracy, the Soft K-Means and FCM have provided effectively equivalent performance. PCA with Soft K-Means yielded best results in fault classification.

In [12], fault detection Input Output Classification Mapping (IOCM) is used which is applied to a mechanical system. Rather than using IOCM alone, adding Discrete Wavelet Decomposition (DWD) improved performance and accuracy of fault detection. The classification of faults is carried out using three kinds of input in three different experiments. The tested samples include 40 belt fault signals, 40 gear fault signals, and 40 bearing fault signals. The accuracy percentage of all these three types of faults when DWD and IOCM are combined is 100%. In [15], Density Based Clustering technique is used for early

detection of fault prone modules in software development. Based on the metrics available with respect to software coding, the DBC technique works well. This algorithm achieved a probability detection of 92.8%. In [16], Artificial Ant Clustering technique is used to detect and diagnose faults pertaining in induction motors. The faults are of electronic and mechanical in nature. The AAC is an unsupervised classification method. The experimental results reveal that it produces best results when compared with other techniques such as Multi-Layered Feed Forward Network (MLFF) and Mountain Clustering/ANFIS. The AAC has shown a 1.33 % classification error rate while the others show 2% and 2.67% respectively. When PCA is used along with AAC, the classification error rate is reduced to 0%.

Gath–Geva Clustering is used in [17] for fault detection and isolation. It is considered to be a pattern classification problem. The GGC algorithm is used along with other techniques known as PCA and GA. The PCA and GA are used for coping with the curse of dimensionality while the GGC is used for isolating faults. The validation of fault detection is done in terms of specificity and sensitivity. The GGC algorithm achieved 80% sensitivity and 99.6% specificity. The experiments are done on Visbreaker process unit which is used in an oil refinery. As proved in [20], the K-Means clustering and PNN when used independently do not yield good results in terms of accuracy of classifying faults. However, the algorithm which combines both K-Means and PNN is capable of classifying faults with higher accuracy. The faults were generated by Tennessee Eastman Process, a bench mark problem in process engineering. The experiments are done with 25 variables and two types of known faults. From this, it can be concluded that the K-Means alone cannot provide the required accuracy and the same case with PNN. However, the combined approach that includes both of the techniques provides higher accuracy. In case of Wireless Sensor Networks (WSNs) [23], faults occur more frequently. Efficient fault detection and recover are essential in such networks. Clustering in this case is used for clustering nodes for control, management, energy efficiency and routing. The approach used for fault detection is Cluster-based and Cellular Approach. In [14], Fuzzy Cluster-Means is used for sensor fault detection. In both [14] and [23] fault detection is done in WSN. However, the FCM algorithm is used in [14] which are efficient for signal separation. This algorithm is proved to be accurate and with higher performance in terms of fault classification.

IV. EVALUATIONS

This section evaluates the various techniques described in related works section in terms of speed, accuracy of fault classification, efficiency and performance. However, the literature available reveals the accuracy explicitly while the efficiency speed and cost are implicitly given. Table 1 shows summary of the findings.

Table 1. Evaluation of Clustering Techniques

Technique/Method	Speed	Accuracy	efficiency	Performance
RHPNN [18]	High	High	High	High
PNN [21]	Less than RHPNN	Less	N/A	Low
BPNN [22]	Less than RHPNN	N/A	N/A	Low
FCM [14] [10]	N/A	High	N/A	N/A
K-Means [10] [20]	High	Low	N/A	N/A
K-Means with PNN [20]	High	High	N/A	High
Soft K-Means [10]	High	High	N/A	High
PCA [10]	N/A	Low	N/A	N/A
AAC [16]	N/A	High	N/A	High
PCA with AAC [16]	N/A	High	N/A	High
IOCM [12]	N/A	Low	N/A	Less
IOCM with DWD	N/A	100%	N/A	High
DBC [15]	N/A	High	N/A	N/A
CCA [23]	N/A	N/A	N/A	N/A
GGC [17]	N/A	High	N/A	N/A

It is evident from table 1 that high level of accuracy in fault classification is achieved by algorithms RHPNN, FCM, K-Means with PNN, Soft K-Means, AAC, AAC with PCA, IOCM with DWD, DBC and GGC. As the accuracy is measured explicitly in most of the review papers, this can be considered for evaluation. The other metrics are implied to some extent though there is no explicit mentioning of them. However, the speed of RHPNN, PNN and BPNN are given explicitly. Perfect evaluation seems difficult as the methods in the literature are not meant for classifying faults of same kind. Even they are in different fields and no benchmark datasets are used consistently. Table 1 shows the results of various algorithms or techniques used for fault classification in terms of speed, accuracy, efficiency and performance. However an important observation here is that the experiments are conducted in different networks or environments.

The study conducted by Kalilian and Mustapha [24] uncovered further limitations which can be associated with particular algorithms commonly used. The authors make an example of the comparison between K-Means and distribution-based clustering. While the former suffers from a key limitation of incorrectly cut borders between individual clusters, the latter requires extra effort from the user. In line with the outlined shortcoming of different networks or environments used in the conducted experiments, the inability to objectively compare or contrast the particular limitations of individual clustering techniques suggests that no one-fits-all solution can be advised. Instead they are to be considered based on the context in which methods are employed.

V. CONCLUSIONS

In this paper algorithms and techniques used in the literature has been studied for network fault classification. We conclude that there are many techniques pertaining to data mining, artificial intelligence and other kinds. Out of all those Fuzzy Cluster-Means is widely used for clustering solutions in various applications including network fault classification. Other clustering algorithm found suitable is Soft K-Means algorithm. There are many algorithms providing high accuracy in fault classification. However, after considering the context in which they are applied, the FCM is an iteratively optimal algorithm which

ISSN (Print) : 2319-5940
ISSN (Online) : 2278-1021

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 2, Issue 10, October 2013

is considered more flexible especially for IP traffic analysis and network fault classification as it can treat such data naturally.

REFERENCES

1. Lewis, F.L.: Wireless Sensor Networks: Cook D.J., Das S.K. (eds.) To Appear in Smart Environments: Technologies, Protocols, and Applications 2004. John Wiley, New York (2004)
2. Iyer, S.: Mobile Ad Hoc Networks. IIT Bombay (2000)
3. Kargl, F.: Vehicular Communications and VANETs. CCC Ulm, Ulm University (2006)
4. Richardson, M.: Wireless Mesh Networking. Mountain States Networking (2006)
5. Gürer, D.W., Khan I., Ogier, R.: An Artificial Intelligence Approach to Network Fault Management, pp.1--10. Sprint, Kansas (n.d.)
6. Wang, Z.: Model of network faults. Meandzija, B., Westcott, J. (eds.) Integrated Network Management, Elsevier Science Publishers B.V., I. North Holland (1989)
7. Bouloutas, A., Calo, S., Finkel, A.: Alarm correlation and fault identification in communication networks. In: IEEE Transactions on Communications. vol. 42, pp. 523--533. IEEE Press, New York (1994)
8. Katzela, I., Schwartz, M.: Fault in Communication Networks Schemes for Identification. IEEE. 3(6), 753--764 (1995)
9. Fu,W., Li, K. F., Neville, S., Gregson, D.: Fault Diagnosis for Rapid Transit Using Pattern Recognition and Classification Techniques . IEEE. 356--340 (2003)
10. Li, X., Parker, L. E.: Design and Performance Improvements for Fault Detection in Tightly-Coupled Multi-Robot Team Tasks. IEEE. 198--204 (2008)
11. Al-Kasassbeh, M. Adda, Mo.: Network fault detection with Wiener filter-based agent. Elsevier. 32 (1), 824--833(2009)
12. Barakat, M, Lefebvre, D., Khalil, M., Mustapha, O., Druaux, F.: Input - Output Classification Mapping for the Fault Detection, Identification and Accommodation, IEEE. 403--411(2010)
13. Pouliezos, A.D., Stavrakakis, G.S.: Real Time fault monitoring Of Industrial Process. Kulwer, Dordrecht (1994)
14. ElMadbouly, E. E., Abdalla, A. E., ElBanby, G. M.: Sensor Fusion and Sensor Fault Detection with Fuzzy Clustering. IEEE. 265--268 (2010)
15. Sandhu, P. S., Kaur, M., Kaur, A.: A Density Based Clustering Approach for Early Detection of Fault Prone Modules. In: International Conference on Electronics and Information Engineering (ICEIE 2010), 2, pp. 52--530. IEEE (2010)
16. Soualhi, A., Clerc, G, Razik, H.: Detection and Diagnosis of faults in Induction Motor Using an Improved Artificial Ant Clustering Technique. In: IEEE Transactions on Communications. vol. 60(9), pp. 4053 - 4062. IEEE, (2011)
17. Mokhtare, M., Vahed, S. H., Shoorehdeli, M. A., Fatehi, A.: Fault Detection and Isolation of Vis breaker Unit in Oil Refinery using Multistage Gath-Geva Clustering. In: 20th Iranian Conference on Electrical Engineering (ICEE2012), pp. 1018--1022. IEEE (2012)
18. Zheng Rong Yang et al. Applying a Robust Heteroscedastic Probabilistic Neural Network to Analog Fault Detection and Classification. IEEE. 19 (1), pp.142--152 (2000).
19. Kohonen, T.: Self-Organization and Associative Memory (3ed.). Springer-Verlag, Berlin (1988)
20. Wu, D., Yang, Q., Tian, F., Zhang, D. X.: Fault Diagnosis Based on K-Means Clustering and PNN. IEEE. pp.173--177 (2010)
21. Specht, D. F.: Probabilistic Neural Networks for Classification, Mapping, or Associative Memory. In: Proc. Int. Conf. Neural Networks, vol. 1, pp. 525--530. Neural Networks (1988)
22. Specht, D. F.: Probabilistic Neural Networks. Neural Networks. vol. 3, pp. 109--118 (1990)
23. Akbari, A., Beikmahdavi, N.: Cluster-Based and Cellular Approach to Fault Detection And Recovery in Wireless Sensor Networks. IEEE. pp. 148--153 (2010)
24. Kalilian, M., Mustapha, N.: Data stream clustering: Challenges and issues. In: Proceedings of the International MultiConference of Engineers and Computer Scientists, vol. 1. IMECS, Hong Kong (2010)

View publication stats