Classification and Visualization of Alarm Data Based on Heterogeneous Distance

Boxu Zhao, Tsinghua University, Beijing, China
Guiming Luo, Tsinghua University, Beijing, China

ABSTRACT

Alarm classification and visualization of historical data is significant and sophisticated in the area of smart management in telecom network due to alarm flood and propagation. In this article, we propose a heterogeneous distance to compute the similarity distance matrix of alarms, which is applied to alarm classification. By using Multidimensional Scaling, alarm data in high dimension is translated into a 2-dimensional graph in alarm windows. Then alarm attention and relation are clearly shown by comparing current and past alarms. Experiments show MDS based on the heterogeneous distance has a better classifying effect than other distance measures. The case study demonstrates the method can show alarm correlation easily and help to locate faults when applied to the analysis of telecom alarm data.

KEYWORDS

Alarm Classification, Alarm Visualization Tools, Dimensionality Reduction, Heterogeneous Distance

1. INTRODUCTION

The Network Management System (NMS) of telecom network has made it possible to deal with everyday events. There are a large number of alarms collected by network devices in NMS, which is the basis of follow-up fault diagnosis. Ideally, every alarm should draw the attention of an operator, but many alarms may occur almost simultaneously to overwhelm experienced operators. The situation is well known as alarm floods or alarm showers. Alarms do not always contain accurate information about network faults clearly, that is, the fault may trigger a flood of alarms produced by other devices connected to the device when a fault occurs, and not all of these alarms can instruct the root cause of failure, resulting in that critical alarms are buried under unnecessary ones. Therefore, alarm management has also become a focus in industrial systems of different areas. Stern etc. proposed a method of reducing the size of alarm sets by taking advantage of redundant information and applied it to alarm data of three large Natural Gas Processing Plants (Soares, Pinto, & de Souza, 2016). Jacobs and Dagnino (2016) introduced a novel graph-based data mining approach that can be used to analyze industrial alarm data and used the experiment on a power generation station and oil refinery datasets to demonstrate scalability and efficiency of the method. Allan etc. proposed a set of practices for monitor alarm reduction without undermining patient safety in intensive care unit to improve patient safety by optimizing alarm systems. (Allan, Doyle, Sapirstein, & Cvach, 2017) Alarm analysis has also been recognized as an important problem in system monitoring and faults detection. (Kondaveeti, Shah, & Izadi, 2009) (Yang, Shah, Xiao, & Chen, 2012) (Akatsuka, Noda, & Sugimoto, 2013)

Researchers have focused on ‘alarm correlation’ to reduce the number of alarms and at the same time find out all possible faults. A typical thought of alarm correlation is to establish an alarm

DOI: 10.4018/IJDWM.2018040104

Copyright © 2018, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
knowledge base and determine the maximum possible failure or faults. A few alarm correlation algorithms have been proposed, including Rule-Based Correlation, (Cronk, Callahan, & Bernstein, 1988) Case-Based Reasoning, (Slade, 1991) Fuzzy Logic, (Zadeh, 1996) Bayesian Networks, (Heckerman, Mamdani, & Wellman, 1995) and Neural Networks. (Gurer, Khan, Ogier, & Keffer, 1996) The above methods satisfy small-scale network and it is not enough to meet the needs of network maintenance only through expert knowledge. So, some researchers (Manganaris, Christensen, Zerkle, & Hermiz, 2000) (Julisch, 2002) focused on data mining to analyze alarm sequence and dig related rules, confidence is calculated to obtain the correlation rules. Weiss described how data mining can be used to uncover useful information buried within nuisance alarms, present several data mining applications and demonstrated that data mining can be used to identify network abnormal situations. (Weiss, 2005) Mannila presented the efficient algorithm (WINEPI) for detection of frequent episodes of a given number of episodes. (Mannila, Toivonen, & Verkamo, 1997) This algorithm is applied in telecommunication alarm management in TASA. The data mining method focuses on related rules to find out faults in the network but fails to show visual display to operators. Taniar proposed exceptions rules mining algorithm to generate candidate exception based on the knowledge about negative association rules in the database, which can also be applied to find out abnormal alarms and diagnose faults. (Taniar, Rahayu, Lee, & Daly, 2008) (Daly, & Taniar, 2004)

In recent years, some researchers committed themselves to visualizing correlation information using graphical tools. Xu presented Alarm Association Algorithms based on Spectral Graph Theory (AAASG), which can analyze the transformation of alarm feature spectrum and dynamically display alarm mode visually to detect and predict the faults. (Xu, & Guo, 2009) Knodaveeti present High-Density Alarm Plot (HDAP) and Alarm Similarity Color Map (ASCM) to effectively identify nuisance alarms such as chattering and related alarms in industrial alarm systems. (Knodaveeti et al., 2009) Yang used a Gaussian kernel method to generate pseudo-continuous time series, weakening the effect of missed, false and chatting alarms when visualizing information. (Yang et al., 2012) Singular value decomposition technique is also used to find out redundant tags. Akatsuka introduced Levenshtein Distance and proposed a new method of analyzing similarities between sequential alarms, which is able to correctly distinguish similarities between sequential alarms. (Akatsuka et al., 2013)

These methods mainly specialize in occurrence times and tag names of alarms and analyze the correlation information between different alarm tags. However, alarm data in telecom usually includes some other attributes except for tag name and time stamp, like IP address generating the alarm, alarm severity and alarm summary etc. These alarm attributes are different types—categorical, numerical and textual, which are not often considered but may have important information. (Table 1) For example, in text data field “alarm summary” in our case study, there are about 1400 different language pattern, possibly including text interpretation for alarm appearance. Besides, an alarm tag often means a collection of alarms of same alarm type, so the analysis is based on correlation

<table>
<thead>
<tr>
<th>Attribute name</th>
<th>Instance</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Device</td>
<td>hqma_cmhqmail</td>
<td>Device name</td>
</tr>
<tr>
<td>IP</td>
<td>10.1.28.9</td>
<td>IP address of host or network device</td>
</tr>
<tr>
<td>Severity</td>
<td>5</td>
<td>Alarm severity: a quantitative index</td>
</tr>
<tr>
<td>Summary</td>
<td>Process opt/ sunspu/amgrnot in operation</td>
<td>Text interpretation why alarm appears</td>
</tr>
<tr>
<td>Time</td>
<td>2012/5/11 15:50</td>
<td>Time stamp when alarm occurs</td>
</tr>
<tr>
<td>Type</td>
<td>Ping alarm</td>
<td>Alarm type or tag</td>
</tr>
<tr>
<td>Processing status</td>
<td>closed/open</td>
<td>Alarm status: open or closed</td>
</tr>
</tbody>
</table>
Related Content

Combining kNN Imputation and Bootstrap Calibrated: Empirical Likelihood for Incomplete Data Analysis
[www.igi-global.com/article/combining-knn-imputation-bootstrap-calibrated/46943?camid=4v1a](www.igi-global.com/article/combining-knn-imputation-bootstrap-calibrated/46943?camid=4v1a)

A Particle Filtering Based Approach for Gear Prognostics
[www.igi-global.com/chapter/particle-filtering-based-approach-gear/73449?camid=4v1a](www.igi-global.com/chapter/particle-filtering-based-approach-gear/73449?camid=4v1a)
Using Data Mining Techniques to Discover Patterns in an Airline’s Flight Hours Assignments
www.igi-global.com/article/using-data-mining-techniques-to-discover-patterns-in-an-airlines-flight-hours-assignments/181883?camid=4v1a

Cost-Sensitive Learning in Medicine
Alberto Freitas, Pavel Brazdil and Altamiro Costa-Pereira (2009). *Data Mining and Medical Knowledge Management: Cases and Applications* (pp. 57-75).
www.igi-global.com/chapter/cost-sensitive-learning-medicine/7527?camid=4v1a