Alarm Oracle: A Graph Neural Network for Causal Structure Analysis of Alarm Logs

Luke J. Miller
University of Missouri, Kansas City

Semir Hot
University of Missouri, Kansas City

Abstract

This paper presents Alarm Oracle, a sophisticated graph neural network engineered for conducting root cause analysis in the complex milieu of alarm cascade scenarios. Employing an unsupervised approach complemented by the Hawkes process, Alarm Oracle excels in the precise identification of initiating events in alarm cascades. Its development is catalyzed by both historical disasters and the contemporary challenges faced by network administrators in managing expansive alarm logs. The system, trained and tested on a synthetic dataset that emulates Huawei's network characteristics, showcases its prowess through a novel Hawkes Process-based causal discovery algorithm. This algorithm, further refined with an Expectation-Maximization framework and a hill climbing technique, ensures high precision in determining causal relationships. The findings from this study illuminate the system's potential in network management and disaster prevention, reinforcing the critical demand for advanced analytical tools in the realm of complex systems management.

1 Introduction

Alarm Oracle represents an application of graph neural networks, meticulously crafted for the root cause analysis of alarm cascades. This system, operating on an unsupervised framework and harnessing the Hawkes process, is adept at iteratively pinpointing root causes with a high degree of confidence. Designed with a focus on precision rather than sensitivity, Alarm Oracle aims for the utmost accuracy and reliability in its identified causal correlations. Its practical utility is realized through collaboration with domain experts, where the system aids network administrators in unraveling the intricate structures behind alarm cascades, enabling timely and effective interventions.

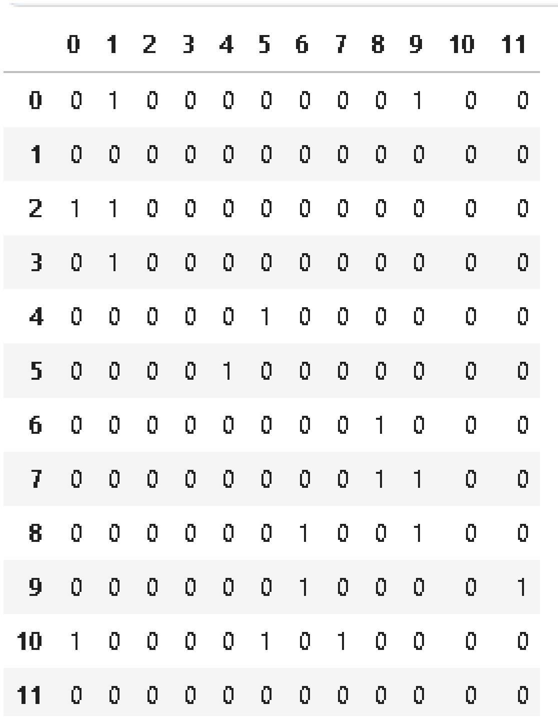
2 Novelty and Motivation

The impetus for developing Alarm Oracle extends beyond the realm of network administration, drawing lessons from historical disasters such as the Piper Alpha Oil disaster and the Three Mile Island Nuclear Accident [3]. These calamitous events highlight the dire consequences of unmanaged alarm cascades and the indispensable role of efficient root cause analysis tools in high-stakes settings. Despite advancements in technology, the challenge of managing complex alarm systems persists in various modern scenarios [5], underscoring the necessity of advanced tools like Alarm Oracle for proactive and remedial measures in network management.

3 Data

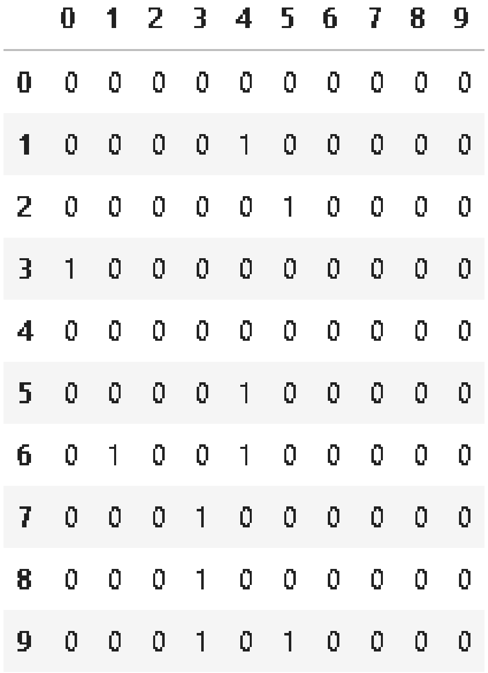
The dataset [1] employed in this study was initially sourced from Huawei. However, it lacked explicit ground truth information necessary for determining the causality of events. To address this limitation, a synthetic dataset was created, mirroring the characteristics of the Huawei dataset. This involved the construction of a network topology comprising various devices, each assigned a probabilistic set of alarms. The methodology for generating this synthetic data included the establishment of a network topology, defined by the number of devices and their interconnectedness. For each device, a specific set of potential alarms was designated, with assigned probabilities dictating the likelihood of these alarms occurring within a given timeframe.

A causality matrix was also developed to illustrate the potential for one alarm to trigger another, adding a layer of complexity to the simulation. In this model, the occurrence of an initial alarm could probabilistically instigate a subsequent alarm on the same or an adjacent device, effectively simulating a realistic alarm cascade. This framework was utilized to generate four distinct datasets, each varying in duration, network topology, alarm density, and the probabilities associated with triggering alarm cascades. These datasets provided the basis for training, with network topologies and alarm logs



	0	1	2	3	4	5	6	7	8	9	10	11
0	0	1	0	0	0	0	0	0	0	1	0	0
1	0	0	0	0	0	0	0	0	0	0	0	0
2	1	1	0	0	0	0	0	0	0	0	0	0
3	0	1	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	1	0	0	0
7	0	0	0	0	0	0	0	0	1	1	0	0
8	0	0	0	0	0	0	1	0	0	1	0	0
9	0	0	0	0	0	0	1	0	0	0	0	1
10	1	0	0	0	0	1	0	1	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1: An example Network Topology Matrix



	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0
6	0	1	0	0	1	0	0	0	0	0
7	0	0	0	1	0	0	0	0	0	0
8	0	0	0	1	0	0	0	0	0	0
9	0	0	0	1	0	1	0	0	0	0

Figure 2: An Example Alarm Causality Matrix

serving as primary inputs, while the causality matrix was reserved for post-training verification within the unsupervised learning model. To facilitate this process, two key classes were developed: AlarmDataPreprocessor and AlarmDataGenerator. These classes were instrumental in both cleaning the data and creating synthetic alarm data. The procedure encompassed a detailed configuration of parameters and methodologies, ensuring the generation of representative alarm logs and adjacency matrices for effective training. This approach underscores the meticulous and systematic nature of the data preparation phase, crucial for the robustness and reliability of the resulting analysis.

4 Algorithm

The Alarm Oracle's algorithm is an intricate causal structure learning algorithm, fundamentally based on the principles of Topological Hawkes processes [2], and is specifically designed for analyzing spatio-temporal event sequences in complex networks. At its core, the algorithm employs a combination of Poisson and Hawkes processes. The former lays the groundwork for modeling random events within the network, particularly when no historical data is available, while the latter extends this model by incorporating the memory of past events to enhance the prediction of future event intensities.

In terms of its operation, the algorithm begins with the construction of a network topology using a given adjacency matrix. This topology is pivotal in mapping out the relationships and interactions between various nodes (or events) within the network. The algorithm then embarks on a detailed learning process, analyzing the temporal sequences of events across different nodes, and seeking to establish causal connections among them.

One of the most notable aspects is its iterative approach to refining the causal relationships within the network. This is achieved through a combination of optimization techniques, including gradient descent and hill climbing, which are used to optimize hyperparameters and causal weights within the model. These steps are crucial in fine-tuning the algorithm's ability to accurately model and predict the complex interdependencies and causal structures inherent in network data.

Another key feature is its use of a hill-climbing algorithm to search for the most probable causal graph, which is then used to generate a causal matrix. This matrix is a critical component of the model, as it represents the intensity and direction of causal effects between events, offering insights into the underlying dynamics of the network.

The algorithm also incorporates a series of stopping criteria, including the Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC), which are used to determine the optimal point for halting iterations. This ensures that the model achieves a balance between accuracy and computational efficiency.

In essence, the algorithm within Alarm Oracle is a sophisticated and nuanced approach to understanding and modeling complex causal relationships in networked environments. It intelligently integrates various computational and statistical methods to offer a comprehensive view of how events influence one another over time and space within a network.

5 Results

The analysis of the results, particularly with a focus on the four confusion matrices generated figures 1 through 4, underscores the high precision of the system in accurately identifying true causal correlations within the network. A significant aspect of these results is the complete absence of false positives, which is a testament to the system's effectiveness in directing operators' attention towards relevant alarms, thereby minimizing the risk of misdirection. Balancing precision and recall are often dependent upon the nature of the problem [4], and in this instance, it was determined that precision was more important in real-world scenarios where the emphasis is on reducing the search-space. Impressively, the system achieved a precision rate of 100 percent, successfully identifying about 50 percent of potential alarm cascades in the evaluated dataset.

It is important, however, to contextualize these findings within the nature of the dataset and the algorithm's operational parameters. The potential for an alarm cascade in the dataset does not inherently imply that such a cascade actually transpired within the evaluated period. This distinction is crucial, as the system's ability to predict potential cascades may not always correlate with actual events. Therefore, further analytical scrutiny is warranted, specifically geared towards discerning which of these cascades were merely probabilistically generated as opposed to those that actually occurred. Such an in-depth analysis would provide a more nuanced understanding of the system's predictive capabilities and its practical utility in real-world scenarios, where distinguishing between probable and actual alarm cascades is vital for efficient and accurate network management.

6 Conclusion

The relevance of Alarm Oracle is underlined by both historical precedents and current challenges in network management. Through the processing of synthetic data mirroring Huawei's network, Alarm Oracle has demonstrated its potential to be a pivotal tool in real-world scenarios, equipping operators with crucial insights for prompt and effective decision-making. The prospective enhancements in algorithmic efficiency and adaptability to real-world data signal a promising future for Alarm Oracle in crucial network management and disaster prevention roles.

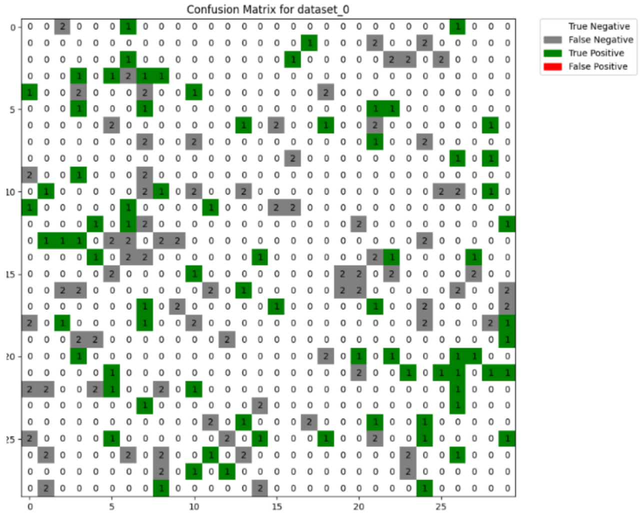


Figure 3: Confusion Matrix: Dataset 0

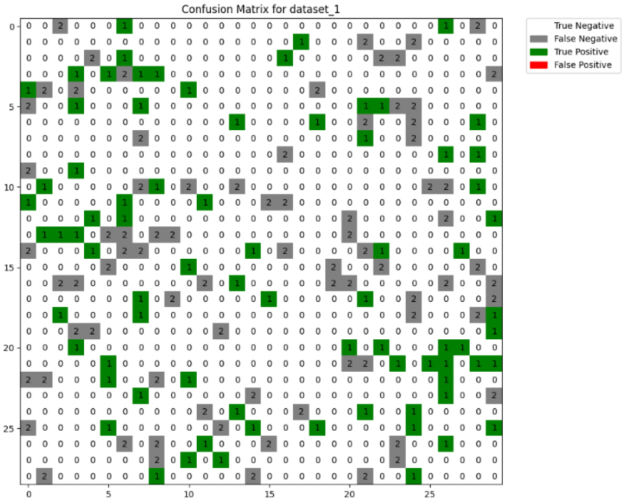


Figure 4: Confusion Matrix: Dataset 1

References

[1] Pcic 2021: Causal discovery, Aug 2021. https://competition.huaweicloud.com/information/1000041487/dataset.

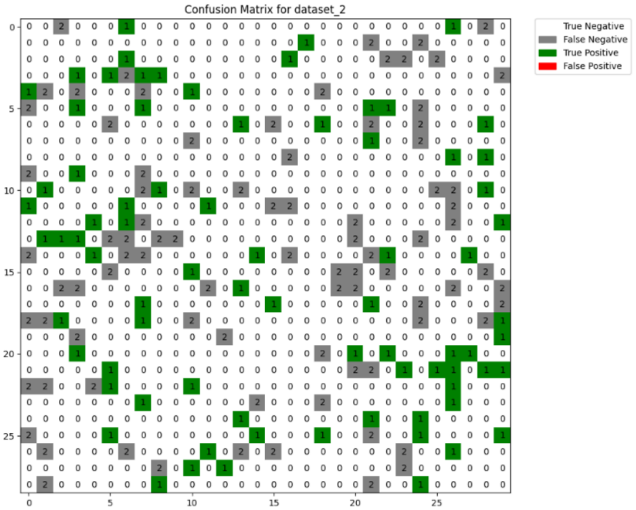


Figure 5: Confusion Matrix: Dataset 2

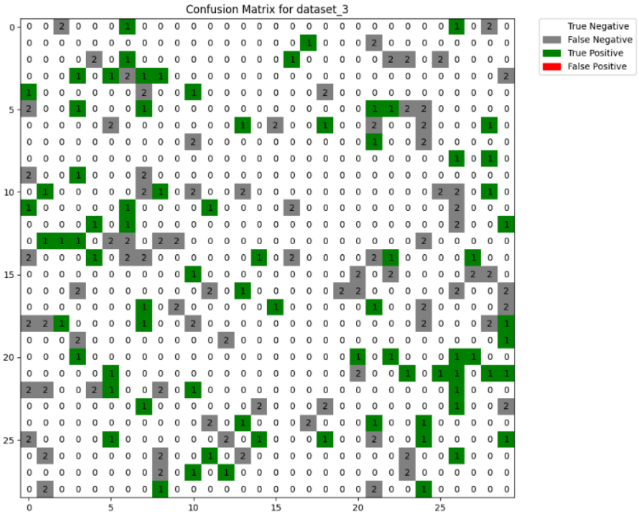


Figure 6: Confusion Matrix: Dataset 3

[2] Cai, R., Wu, S., Qiao, J., Hao, Z., Zhang, K., and Zhang, X. Thps: Topological hawkes processes for learning causal structure on event sequences. *IEEE Transactions on Neural Networks and Learning Systems* (2022), 1–15.

[3] Gray, M., and Rosen, I. *The Warning: Accident at Three Mile Island*. WW Norton & Company, 2003.

[4] Oyen, D., Niculescu-Mizil, A., Ostroff, R., Stewart, A., and Clark, V. P. Controlling the precision-recall tradeoff in differential dependency network analysis, 2013.

[5] Wang, J., He, C., Liu, Y., Tian, G., Peng, I., Xing, J., Ruan, X., Xie, H., and Wang, F. L. Efficient alarm behavior analytics for telecom networks. *Information Sciences 402* (2017), 1–14.