Anomaly Detection based on Alarms Data

Michel Kamel, Anis Hoayek, Mireille Batton-Hubert

ANOMALY DETECTION BASED ON ALARMS DATA

Michel Kamel, Anis Hoayek and Mireille Batton-Hubert

Mathematics and industrial engineering department, Ecole des Mines de Saint-Etienne, University Clermont Auvergne, CNRS UMR 6158 LIMOS

ABSTRACT

Alarms data is a very important source of information for network operation center (NOC) teams to aggregate and display alarming events occurring within a network element. However, on a large network, a long list of alarms is generated almost continuously. Intelligent analytical reporting of these alarms is needed to help the NOC team to eliminate noise and focus on primary events. Hence, there is a need for an anomaly detection model to learn from and use historical alarms data to achieve this. It is also important to indicate the root cause of anomalies so that immediate corrective action can be taken. In this paper, we aim to design an anomaly detection model in the context of alarms data (categorical data) in the field of telecommunication and that can be used as a first step for further root cause analysis. To do this, we introduce a new algorithm to derive four features based on historical data and aggregate them to generate a final score that is optimized through supervised labels for greater accuracy. These four features reflect the likelihood of occurrence of events, the sequence of events and the importance of relatively new events not seen in the historical data. Certain assumptions are tested on the data using the relevant statistical tests. After validating these assumptions, we measure the accuracy on labelled data, revealing that the proposed algorithm performs with a high anomaly detection accuracy

KEYWORDS

Alarms, Anomaly detection, Events data, Probabilistic scoring distribution.

1. INTRODUCTION

Anomaly detection is an aspect of data mining that has been the subject of research in many fields, such as telecommunications, information technology and finance.

There are several definitions of anomaly in the literature. Hawkins [1] defines an anomaly/outlier as an observation, which deviates considerably from the remaining observations, as if generated by a different process. Dunning and Friedman [2] state that anomaly detection involves modelling what is normal in order to discover what is not. In general, anomalies are events with a special behaviour that is dissimilar to that of normal events, and it is expected that this behaviour would be detected by analysing underlying data. Therefore, there is an urgent need for intelligent algorithms to identify such abnormal behaviour.

Anomaly detection improves data quality by deleting or replacing abnormal data. However, in certain cases, anomalies reflect an extreme event and can provide useful new knowledge. For example, the detection of such anomalies can prevent material damage and encourage predictive maintenance in the industrial field. It also has applications in several other areas such as health [3], cybersecurity [4], finance [5], natural disaster [6], and telecommunication [7].

David C. Wyld et al. (Eds): AI, AIMLNET, BIOS, BINLP, CSTY, MaVaS, SIGI - 2022
pp. 103-114, 2022. CS & IT - CSCP 2022 DOI: 10.5121/csit.2022.121810

Several methods have been proposed for detecting anomalies, each of which has its own strengths and weaknesses. Patcha and Park [8] reviewed all the known methods used for anomaly detection. Additionally, an overview of existing techniques covering several approaches is presented in [9] and [10].

Despite the large volume of literature on anomaly detection for numeric data e.g., time series, there is limited knowledge on the problem of abnormal behaviour in the context of categorical and structured textual data.

In this paper, we aim to design an anomaly detection algorithm in the context of alarms data (categorical data) in the field of telecommunication. In other words, in a given period of time, each network element of a telecommunication network generates a set of Key Performance Indicators (KPIs) and alarms that describe its behaviour. Alarms are typically categorical data with different characteristics (i.e., name, description, severity of the event, start time, end time), triggered to indicate a certain event occurring on the network element. Based on this information, those intervals of time are detected that have a high probability/score of displaying abnormal behaviour. Alarms data is important in a real-world context when KPIs are unavailable and cannot be calculated or extracted. It should be noted that alarms are events that can start popping up on a certain network element at any time. Therefore, each alarm can be considered to be equivalent to a binary random variable that can appear at any time with a certain probability.

Here, we propose an approach that introduces two new, innovative aspects. First, four features are calculated and aggregated to define events data during a certain interval of time; this includes the number of alarms, occurrence time, inter arrival time, transition frequency (Markovian model) and historical frequency. By combining this information, we compute an abnormality score which is, to the best of our knowledge, the first time that an anomaly detection algorithm has incorporated all the attributes of an event. In fact, in the majority of prior influential studies only a few of the previously cited attributes were considered. [11] consider just the Markovian component; in [12], a feature selection step is proposed prior to anomaly detection, which is a process that is associated with a high risk of loss of key information and requires significant effort for data labelling; [13] consider categorical data to be textual and vectorize it before the anomaly detection phase which is also associated with a high risk of loss of information. Second, the proposed algorithm enables users to extract local and focused information about one of the previously discussed features which may provide greater insight into the root cause of the anomaly (also known as anomaly fingerprint).

In this paper, we first describe the methodology used to build the abnormality score. We then present an application of the algorithm and analyse the results.

2. METHODOLOGY

We propose a semi-parametric scoring system that reflects the different behavioural aspects of a component of a network during a given interval of time using alarms data generated for that component. These aspects are (2.1) the number of alarms, (2.2) the inter-arrival time between alarms, (2.3) the transition probability of two consecutive alarms, and (2.4) the historical frequency of an alarm. The calculation of the final score is demonstrated in Subsection 2.5 and the optimization of the model weights is shown in Subsection 2.6. Because alarms are generated from each network component, of which there are different types, we group these components by type when drawing inferences from the data to reduce volatility and heterogeneity in the calculated statistics.

2.1. Number of Alarms

It is a common practice in parametric statistics to assume a Poisson distribution while modelling the number of occurrences of a certain event during a fixed period. Therefore, under this assumption, we begin by estimating the rate parameter λ of the Poisson distribution by calculating the arithmetic average of the number of alarms across all the intervals for each different component type of the network. Therefore, if we have L different types of components in the network, L different rate parameters $\lambda 1 ,\dots , \lambda L$ are estimated.

Now, let Nl ,l = 1, …, L denote the random variables (r.v.) indicating the number of alarms generated by a component of type $l\in$ {1, … , L} over an interval of time. Based on the previous assumption, Nl follows a Poisson distribution with rate parameter λl . Note that $\mathbb{E}[Nl\ ] = \lambda l$ and it can easily be shown that the proposed estimator is a minimum variance unbiased estimator (MVUE) of λ. Hence, if n denotes the observed number of alarms in a fixed interval for a component of type l, the associated probability can be computed as shown in Equation (1).

$$\mathcal{P}_1^l = \mathbb{P}[N_l = n] = \frac{e^{-\lambda_l}{\lambda_l}^n}{n!} \qquad (1)$$

Hence, in order to standardize this probability and transform it into a score that reflects the number of alarms, and the fact that a higher-than-average score indicates a higher probability of abnormal behaviour, S 1 l can be defined as:

$$S_1^{\ l} = \begin{cases} \frac{\mathbb{P}[N_l = int(\lambda_l)] - \mathcal{P}_1^l}{\mathbb{P}[N_l = int(\lambda_l)] - \min\limits_{\text{over all data}} (\mathcal{P}_1^l)}, & \text{if } n \geq int(\lambda_l) \\ 0 \qquad\qquad , \text{otherwise} \end{cases} \qquad (2)$$

where $int(.\ )$ denotes the integer part of a real number. Then, a value of S 1 l close to one implies that the number of alarms indicates abnormal behaviour in the specified interval. Note that $int(\lambda l$) represents the mode of a Poisson distribution of parameter λl .

2.2. Inter-Arrival Time

Using the same logic, we consider the intervals of time during which at least two alarms were detected for each type of component L. We also define the r.v.Yl representing the time between two consecutive alarms that occurred within the same interval of time. It is common to model such r.v. by an exponential distribution with rate parameter μl , which is estimated by calculating the inverse of the arithmetic average of the time between two consecutive alarms during all the intervals for each different component type of the network. Note that under the previous assumption, $\mathbb{E}[Yl\ ]$ = 1 $\mu l/$. Hence, if the number of alarms during an interval for a component of type $l\in$ {1, … ,L} is $n \geq 2$, an associated probability can be computed as shown in Equation (3).

$$\mathcal{P}_2^l = \mathbb{P}\left[Y_l \leq \frac{\sum_{j=1}^{n-1} y_j}{n-1}\right] = 1 - e^{-(1/\mu_l)\frac{\sum_{j=1}^{n-1} y_j}{n-1}} \qquad (3)$$

where yj ,j = 1,… , n − 1 denotes the time between alarms j and j + 1. Similarly, this probability can be standardized and transformed into a score to reflect that alarms occurring consecutively

within a very short span of time are more likely to indicate abnormal behaviour, as shown in Equation (4).

$$S_2^l = \frac{\max_{\text{over all data}} (\mathcal{P}_2^l) - \mathcal{P}_2^l}{\max_{\text{over all data}} (\mathcal{P}_2^l) - \min_{\text{over all data}} (\mathcal{P}_2^l)} \qquad (4)$$

Here, a value of S_2^l close to one implies that the time between consecutive alarms indicates abnormal behaviour in the specified interval.

2.3.Transition Probability

In the same context as that of the inter-arrival time score, and based on all the observed alarms during all the intervals for a component of type l, we define the state space of alarms as $El = \{a1, \ldots, aK\}$, where K denotes the number of unique observed alarms in component l. Subsequently, we empirically compute the transition probabilities, $\forall i,j \in \{1, \ldots K\}$, as shown in Equation (5).

$$p_{a_i a_j} = \text{probability of observing } a_j \text{ after } a_i \qquad (5)$$

Hence, we obtain a transition matrix in a similar manner to a Markov chain, that summarizes all the historical transitions that have occurred for each type of component. Then, to highlight abnormal behaviour during a given interval, we identify the occurrence of transitions that are historically uncommon. Practically, if the number of alarms during an interval for a component of type l is $n \geq 2$, where these alarms are elements of the state space E^l denoted by $x_1, \ldots, x_n$, an associated probability can be computed as shown in Equation (6).

$$S_3^l = \frac{\max_{\text{over all data}} (\mathcal{P}_3^l) - \mathcal{P}_3^l}{\max_{\text{over all data}} (\mathcal{P}_3^l) - \min_{\text{over all data}} (\mathcal{P}_3^l)} \qquad (7)$$

As described previously, the probability is standardized and transformed into a score to reflect that the alarms that occur consecutively and that have not occurred one after the other frequently in the past are more likely to be displaying abnormal behaviour. This score is obtained as shown in Equation (7).

$$S_3^l = \frac{\max_{\text{over all data}} (\mathcal{P}_3^l) - \mathcal{P}_3^l}{\max_{\text{over all data}} (\mathcal{P}_3^l) - \min_{\text{over all data}} (\mathcal{P}_3^l)} \qquad (7)$$

Here, a value of $S\ 3\ l$ close to one implies that during this interval, a non-frequent transition is occurring, which is likely to be abnormal behaviour.

2.4. Historical Frequency

Now, we consider the historical frequency of the alarms occurring during an interval. In other words, an alarm of a certain type that is historically infrequent is considered to be more critical and should be highlighted. In real world scenarios, given that access to big data can be limited, this attribute helps in identifying infrequent or non-occurring events in the network, especially

high impact events that occur rarely. Then, for a component of type l we consider the state spaceof alarms $E^l = \{a_1, \ldots, a_K\}$, and the historical frequency of each of these alarms is computed and denoted by $f_i i \in \{1, \ldots K\}$.

Further, to highlight abnormal behaviour during a given interval, we focus on the alarm with the lowest historical frequency among those that occurred during this interval, which are denoted by $x_1, \ldots, x_n \in E^l$, with $n \geq 1$. P_4^l is first defined as shown in equation (8)

$$P_4^l = \max_{k=1,\ldots,n} \frac{1}{f_k} \qquad (8)$$

This is derived using all the available historical intervals data. This is followed by standardization, where P_4^l is transformed into a score quantity as shown in Equation (9).

$$S_4^l = \frac{P_4^l - \min_{\text{over all data}} (\mathcal{P}_4^l)}{\max_{\text{over all data}} (\mathcal{P}_4^l) - \min_{\text{over all data}} (\mathcal{P}_4^l)} \qquad (9)$$

Here, a value of S_4^l close to one implies that a non-frequent alarm occurs during this interval, which indicates abnormal behaviour.

2.5. Final score and individual contributions

To obtain a final abnormality score for a given interval of time and for a particular component of the network of type $l \in \{1, \ldots L\}$, the previously computed scores are aggregated as weighted average measures as shown in Equation (10).

$$S^l = \sum_{i=1}^{4} w_i S_j^l \qquad (10)$$

with $0 < w_i < 1$ and $\sum_{i=1}^{4} w_i = 1$.

The values of different weights are determined based on interactions with subject matter experts (SMEs) and a supervised grid search approach that will be discussed later.

A value of S^l close to one indicates that abnormal behaviour is being displayed during the specified interval and addressing this should be considered to be a priority for the SMEs. In addition, diagnostic information can be extracted from the four individual scores which may provide a starting point for the SMEs to analyse the root cause of the detected anomalies. This additional information is used to explain the derived score by specifying which alarms occurred, which interarrivals are low, which transitions are rare and which alarms have the lowest occurrence historically.

2.6. Validation And Optimization

After assigning a score to each of the intervals across all the components of the network, we validate the results by comparing our labels to the ones given by the SMEs (labels are determined by manual inspection of the data to identify occurrences of anomalies). From the scores obtained

in Subsection 2.5, the labels are determined based on a predefined fixed threshold denoted by s, such that:

$$S^{l}_{\text{labeled}} = \begin{cases} 1 & \text{if } S^{l} > s \\ 0 & \text{Otherwise} \end{cases} \qquad (11)$$

The values of the weights in Subsection 2.5 and the value of the threshold s are determined based on a grid search process [14], where several scenarios/combinations of the underlying parameters are considered. The selected combination is the one with the best performance based on the accuracy of the confusion matrix that shows the degree of similarity between our labels and the SMEs labels, and the value of the Area Under the ROC Curve (AUC) [15]. Such optimization makes the algorithm similar to supervised ML models with the aim of maximizing the correlation between labels and features. This is a unique supervision method to replicate and learn human decisions.

3. APPLICATION

The methodology described in Section 2 is applied on real world data obtained from a virtual telecommunication network to identify intervals with a high probability of displaying abnormal behaviour. Data description, results and analyses, and the advantages of the proposed algorithm will be presented in Subsections 3.1, 3.2 and 3.3 respectively.

3.1. Data Description

From a virtual telecommunication network, and for a given period, we consider alarms occurring on the different network elements over 30-minute intervals with a sliding window step of 5 minutes. The concept of the sliding window is introduced to consider events (i.e., alarms) that overlap between two consecutive intervals. In addition, to assure that the different aspects of the methodology of Section 2 are applicable, we estimate the parameters of the underlying distributions—Poisson for counting alarms and exponential for inter-arrival time—separately on the three types of components that are present in the network and are indexed as $l \in \{1,2,3\}$.

These alarms (categorical data) with their different levels of severity, e.g., critical, minor and major, occurring during a given interval indicate the occurrence of abnormal behaviour. Based on the abnormality score that is computed by the proposed algorithm, SMEs should prioritize intervention in such cases.

3.2. Results And Analysis

We first estimate the parameters of the Poisson distribution λl ,$l \in \{1,2,3\}$ and the exponential distribution μl , for each type of component by applying the methods described in Subsections 2.1 and 2.2. For each type of component, all the available alarm occurrences across all the intervals are used. Note that only those intervals with at least two alarms are considered for the estimation of μl because the exponential distribution models the time between two consecutive alarms (in minutes). Moreover, goodness-of-fit tests are conducted to test the feasibility of the assumption that the number of alarms and the time between two consecutive alarms are governed respectively by Poisson and exponential distributions. Pearson's chi-squared test [16] is used for the goodness-offit test. The estimation results and the p-values of the statistical tests are represented in Tables 1 and 2 respectively.

Table 1. Parameter estimation.

Network element components	λ_l	μ_l
i=1	0.312	0.166
i=2	0.158	0.201
i=3	7.055	0.191

Table 2. Goodness of fit tests.

	p-values	
Network element components	**Poisson**	**Exponential**
i=1	0.98	0.84
i=2	0.231	0.279
i=3	0.785	0.871

Table 2 shows that when the different types of components are considered separately, the assumption about the underlying distribution appears to be reasonable. Therefore, the parametric approach defined in Subsections 2.1 and 2.2 can be relied upon to compute scores $S1\ l$ and $S2\ l$ for each interval for the different network elements.

The third type of score is based on a transition matrix of the probabilities of different descriptions of alarms for a given type of component. To compute such a matrix, the empirical approach described in Subsection 2.3 is applied. Table 3 shows the transition matrix of alarm descriptions for the component of type l = 3.

Table 3. Transition matrix.

Alarm Description	**3(a)**	**3(b)**	**3(c)**	**3(d)**	**3(e)**	**3(f)**	**3(g)**	**3(h)**
3(a)	1	0	0	0	0	0	0	0
3(b)	0	0.2	0	0	0	0	0	0.8
3(c)	0	0	1	0	0	0	0	0
3(d)	0	0	0	1	0	0	0	0
3(e)	0	0	0	0	0	0	1	0
3(f)	0	0	0	0	0	1	0	0
3(g)	0	0	0	0	0	0	1	0
3(h)	0	0.3	0	0	0	0	0	0.7

As an example of how to read Table 3, we can say that for network element of type l = 3, when at least two alarms occur during an interval, alarms of description 3(b) are followed by alarms of the same description in 20% of cases and alarms of description 3(h) in 80% of cases.

Using these matrices and applying the method described in Subsection 2.3, one can obtain the third score $S_3{}^l$ for each interval for the different network elements.

The next step is to compute the historical frequency of each alarm description for a given type of component and to use these frequencies, as described in Subsection 2.4, to calculate the fourth score $S4\ l$ for each interval for different network elements. An example of these frequencies is shown in Table 4 for the network element of type l = 1. Then, when an alarm of description 1(b) occurs during an interval and is observed to have a low historical frequency, then the interval is suspected to be displaying abnormal behaviour.

Table 4. Frequency table.

Alarm Description	Frequency
1(a)	2248
1(b)	10
1(c)	8608
1(d)	2324
1(e)	862
1(f)	17684
1(g)	29
1(h)	253
1(i)	441
1(j)	1348

Now, for each component type $l \in \{1,2,3\}$ and for all the intervals, the final abnormality score S^l can be computed by applying Equation (10) and by setting the initial values for the weights $w_i, i = 1,2,3,4$ (e.g., $w_i = \frac{1}{4} \forall i$). In addition, to apply Equation (11), a threshold s needs to be determined to label all the intervals with 1 if abnormal behaviour is taking place and 0 otherwise.

To optimize the choice of the underlying weights and threshold, the SMEs label a parallel and independent abnormal behaviour based on the same data for the same intervals. Then, based on a random grid search process, the parameters of the algorithm are optimized, for each component type, on two levels. First, among all the tested combinations of weights w_i verifying $0 < w_i < 1$ and $\sum_{i=1}^{4} w_i = 1$ we select the one with the highest AUC. Second, among all the threshold used to draw the optimal ROC curve, we select the one with the highest accuracy in terms of true positives and true negatives, i.e., we maximize the sum of the diagonal terms of the confusion matrix. Hence, we begin the grid search process by considering the following equation:

$$w^* = \underset{w \in \mathfrak{D}}{\operatorname{argmax}} AUC(w) \qquad (12)$$

where w is a vector of weights $w_i, i = 1,2,3,4$ and $\mathfrak{D}$ is the set of all the considered combinations of weights during the first level of the grid search process. Once the optimal combination of weights w^* is selected, we select optimal threshold by applying the following equation:

$$s^* = \underset{s \in \mathcal{T}}{\operatorname{argmax}}\big(TP(s) + TN(s)\big) \qquad (13)$$

where $TP(s)$ and $TN(s)$ denote the true positive and true negative labels respectively when the threshold s is fixed. $\mathcal{T}$ represents the set of all the considered thresholds during the second level of the grid search process. The construction of the sets $\mathfrak{D}$ and $\mathcal{T}$ is done with collaboration and validation by the SMEs. Furthermore, we are not concerned by the phenomenon of overfitting because we are using the latter grid search process merely to optimize the selection of the underlying weights of different scores and the threshold based on a matching method with labels fixed by the SMEs.

Considering network element of type $l = 1$, Fig. 1 shows the optimal ROC curve with a maximum AUC of 0.975 corresponding to the vector of weights $w^* = (w_1^* = 0.41, w_2^* = 0.29, w_3^* = 0.2, w_4^* = 0.1)$ for intervals with at least two alarms (i.e., S_2^l and S_3^l are computable) and a vector $w^* = (w_1^* = 0.8, w_2^* = 0, w_3^* = 0, w_4^* = 0.2)$ for intervals with less than two alarms.

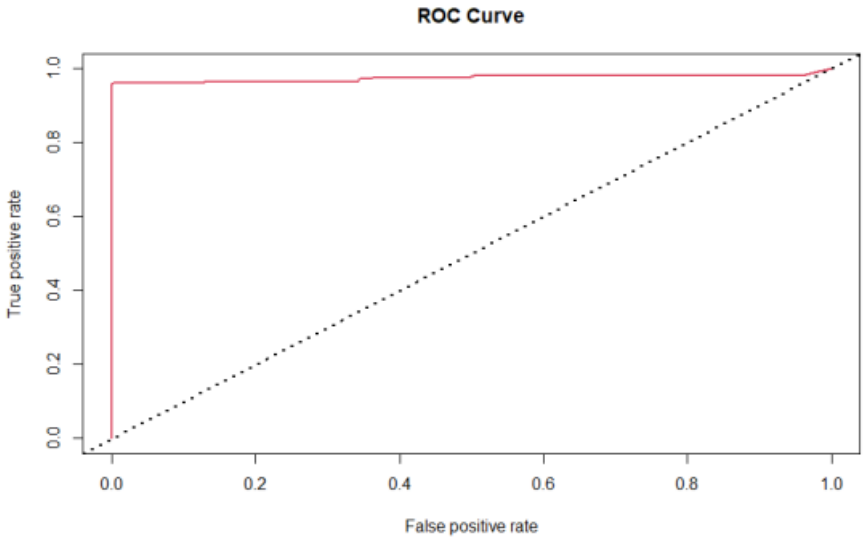

Figure 1. Optimal ROC curve

Table 5 shows the optimal confusion matrix corresponding to a threshold of 0.58. In other words, for components of type $l = 1$, we will apply the following rule to label anomalous behaviour across all the intervals:

$$S_{l\text{abeled}}^l = \begin{cases} 1 & \text{if } S^l > 0.58 \\ 0 & \text{Otherwise} \end{cases}.$$

In the following certain interpretations and metrics are discussed based on the confusion matrix:

- True positive: 2521 intervals; False positive: 531 intervals.
- True negative: 199513 intervals; False negative: 103 intervals.

- Sensitivity: proportion of true positive among SMEs abnormal intervals: $2521/(2521 + 103) = 0.961.$
- Specificity: proportion of true negative among SMEs non abnormal intervals: $199513/(199513 + 531) = 0.997.$
- Accuracy: $(2521 + 199513)/(2521 + 531 + 103 + 199513) = 0.997.$

Table 5. Confusion matrix.

Predicted Labels \ SMEs Labels	1	0
1	2521	531
0	103	199513

Based on all the previous metrics computed after interactions with the SMEs, it is evident that the proposed algorithm is performing well with a high accuracy, and that we can rely on it to detect abnormal behaviour in future intervals. Furthermore, the algorithm is applied on online arriving alarms data and intervals that have been declared as anomalous and validated by experts. Here, approximately 1% of the intervals under control were behaving in an abnormal way, which is very reasonable in practice and is commonly encountered by SMEs. In addition, the algorithm presented in this paper has several advantages when compared to the classical anomaly detection approach. Most of these advantages will be enumerated in the next subsection.

3.3. Advantages Of The Proposed Algorithm

Compared to popular anomaly detection models, the proposed algorithm has four main advantages:

- Our algorithm is already adapted to be an online anomaly detection model applied directly to new arrivals for the purpose of highlighting abnormal behaviour. Hence, there is no need to train such a model on a sample and to test it on another because such a model has no risk of overfitting.
- To the best of our knowledge, this is the first time that an anomaly detection algorithm, based solely on alarms categorical data, has successfully extracted diagnostic information from the four different components of the global score (i.e., $S_1^l, S_2^l, S_3^l \text{and} S_4^l$) to help SMEs initiate root cause analysis of the detected anomalies.
- The proposed algorithm has the ability to generate abnormality scores based solely on alarms data without any additional information about numeric KPIs, which is uncommon in the field of anomaly detection for telecommunication networks.
- The interpretability of this model adds great value and is important for both developers and users.

3.4. Test 0f Independence Between Different Type of Alarms

The algorithm proposed in this paper assumes the existence of one family of alarms. In fact, if other families of alarms are available, our model can easily be generalized by proposing a weighted anomaly score for the different families of alarms. Additionally, we can consider the same optimisation process proposed in Subsections 2.6 and 3.2 to determine the values of the different weights.

Further, to ensure the statistical independence between different families of alarms in terms of occurrence time we suggest an independence test. This is essentially a uniform distribution goodness-of-fit test using classical chi-squared test. Therefore, for alarms of family $\mathcal{A}$, we test whether the occurrence times of such alarms, between two alarms of another family $\mathcal{B}$, are uniformly distributed. It is important to note that in order to apply such an approach, the intervals of time separating the occurrence of two alarms of the same family need to be normalized.

4. CONCLUSION AND PERSPECTIVES

In this paper, an innovative anomaly detection algorithm that solely uses structured alarms (categorical data) has been presented. The proposed model takes into consideration four different attributes extracted from alarms occurrence data to compute a global anomaly score. This can then be used to extract diagnostic information that helps SMEs in performing root cause analysis. Our algorithm is shown to be more advantageous than other existing anomaly detection models, when applied in the same context.

Moreover, we applied the algorithm to real data in the field of telecommunication. The results were then validated by SMEs who provided positive feedback and found that our algorithm outperforms the previously used classical approaches. Users of such a model are also convinced by its output because it relies on the behaviour of historical data and generates real-time ranking of events occurring on a network component in terms of abnormality.

A first perspective of this work is to mathematically formalize a model/algorithm using the extracted information from different sub-scores in order to enhance existing root cause analysis methods based solely on alarms data. A second perspective is to combine alarms data with other type of non-numeric features, e.g., textual data, to build a more complete anomaly detection approach that covers novel aspects that have not been addressed before. Such pioneering work can be initiated by drawing inspiration from [17].

ACKNOWLEDGMENTS

The authors thank the data science team of B-Yond and the SMEs for useful discussions and for their support.

REFERENCES

[1] D. M. Hawkins,(1980)Identification of outliers, vol. 11,Springer.
[2] T. Dunning &E. Friedman (2014) Practical machine learning: a new look at anomaly detection, O'Reilly Media Inc.
[3] A. Ukil, S. Bandyoapdhyay, C. Puri, & A. Pal,(2016) "IoT healthcare analytics: The importance of anomaly detection", IEEE 30th international Conference on advanced information networking and applications (AINA), pp 994-997.
[4] D. A. Bierbrauer, A. Chang, W. Kritzer, & N. D. Bastian,(2021) "Anomaly detection in cybersecurity: Unsupervised, graph-based and supervised learning methods in adversarial environments", arXiv preprint arXiv:2105.06742.

[5] M. Sekar,(2022) “Fraud and anomaly detection”, Machine Learning for Auditors, pp. 193–202, Springer.
[6] S. Miau & W.-H. Hung,(2020) “River flooding forecasting and anomaly detection based on deep learning”, IEEE Access, vol. 8, pp. 198384–198402.
[7] M. Kamel, A. Hoayek & M. B. Hubert, (2022) “Probabilistic approach for anomaly detection with geometric dynamics”, unpublished.
[8] A. Patcha&J.-M. Park,(2007) “An overview of anomaly detection techniques: Existing solutions and latest technological trends”, Computer networks, vol. 51, no. 12, pp. 3448–3470.
[9] C. C. Aggarwal,(2017) “An introduction to outlier analysis”, Outlier analysis, pp. 1–34, Springer.
[10] V. Chandola, A. Banerjee, &V. Kumar,(2009) “Anomaly detection: A survey”,ACM computing surveys (CSUR), vol. 41, no. 3, pp. 1–58.
[11] H. Ren, Z. Ye, & Z. Li, (2017) “Anomaly detection based on a dynamic Markov model”, Information Sciences, vol. 411, pp. 52–65.
[12] Y. Liu, H. Xu, H. Yi, Z. Lin, J. Kang, W. Xia, Q. Shi, Y. Liao, &Y. Ying, (2017) “Network anomaly detection based on dynamic hierarchical clustering of cross domain data”,IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 200–204.
[13] B. Nie, J. Xu, J. Alter, H. Chen, &E. Smirni, (2020) “Mining multivariate discrete event sequences for knowledge discovery and anomaly detection”, 50th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 552–563.
[14] M. Claesen & B. De Moor,(2015) “Hyperparameter search in machine learning”, arXiv preprint arXiv:1502.02127.
[15] T. Fawcett,(2006) “An introduction to roc analysis”, Pattern recognition letters, vol. 27, no. 8, pp. 861–874.
[16] K. Pearson,(1900) “X. on the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling”, The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, vol. 50, no. 302, pp. 157–175.
[17] K. Ezukwoke, H. Toubakh, A. Hoayek, M. Batton-Hubert, X. Boucher, &P. Gounet, (2021) “Intelligent fault analysis decision flow in semiconductor industry 4.0 using natural language processing with deep clustering”, IEEE 17th International Conference on Automation Science and Engineering (CASE), pp. 429–436.

AUTHORS

Michel KAMEL, twelve years of experience in data science and machine learning model development and currently working for B-yond handling the data science function and practices, in parallel actively researching in the field of anomaly detection within the telecommunication industry. PhD student at Ecole des Mines de SaintEtienne.

Anis HOAYEK, Associate professor in the field of probability and statistics at Ecole des Mines de Saint-Etienne. Member of mathematics and industrial engineering department.

Mireille BATTON-HUBERT, Full Professor in the field of probability and statistics at Ecole des Mines de Saint Etienne. Head of mathematics and industrial engineering department.