2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)

Active Prewarning Technology for Smart Grid Based on Unsupervised Learning

Ruiting Qu
State Grid Liaoning Information and Communication Company
Shenyang, China
1316383827@qq.com

Xiaolei Tian*
State Grid Liaoning Information and Communication Company
Shenyang, China
*2415438625@qq.com

Shuo Chen
State Grid Liaoning Information and Communication Company
Shenyang, China
258098970@qq.com

2024 4th International Conference on Machine Learning and Intelligent Systems Engineering (MLISE) | 979-8-3503-7507-7/24/$31.00 ©2024 IEEE | DOI: 10.1109/MLISE62164.2024.10674102

***Abstract*—This paper proposes an early warning technology for power grid cloud platform microservice architecture alarm information based on unsupervised learning, discovers potential abnormal patterns and rules, sends out early warning signals in a timely manner, and improves the ability to identify and process power grid cloud platform microservice architecture alarms. According to the above steps, based on the establishment of the power grid business microservice operation alarm log alarm information knowledge base, this article will study the cloud platform operation monitoring data scoring algorithm, classify the cloud platform monitoring information, and then filter out the suspected abnormal data. For different business types of complex scenarios, research on abnormal information classification mining technology based on statistical models, proximity, clustering and classification, apply improved K-means clustering[1] and LOF outlier detection technology[2] to establish a combined detection model, and detect anomalies The data is clustered and outliers detected to achieve active early warning technology of unsupervised learning.**

Keywords-smart grid; unsupervised learning; K-means clustering; LOF

I. INTRODUCTION

Due to the numerous and interdependent services within the microservice architecture of large-scale power grid business systems, a fault in one service can easily affect the stability and availability of the entire system. For example, in a smart grid big data business application system, which involves multiple services such as marketing business applications, electricity information collection, and distribution, a delay or cessation in response from the smart grid big data service makes it difficult to determine which specific business system has encountered a fault. This paper will first investigate proactive fault warning technology for various associated microservices in large-scale power grid business systems. Dynamic detection of cloud platform microservice faults involves monitoring and detecting services during system operation to promptly identify and handle faults. Dynamic detection can analyze cloud platform service logs and metrics data or use monitoring tools to assess system performance and availability. In recent years, with advancements in machine learning and artificial intelligence, more research has started using these technologies to improve the accuracy and efficiency of microservice fault detection. Machine learning-based fault detection methods leverage these technologies to enhance the accuracy and efficiency of cloud platform fault detection. By training models to recognize fault patterns and predict the probability of faults occurring, these methods can automatically learn and adapt to different system environments and fault patterns, improving the robustness and automation of power grid cloud platform systems.

Therefore, this paper will focus on proactive fault warning technology for various associated microservices in large-scale power grid business systems using machine learning methods. It will utilize collected fault warning information from microservice operation log monitoring systems and establish a knowledge base of warning information through supervised learning. Then, it will classify operational monitoring information using unsupervised learning methods to achieve proactive fault warning in system operations.

II. RELATED WORK

Existing fault prediction methods can be mainly divided into those based on mechanism models, knowledge-based methods, and data-driven methods. Liu et al.[3] proposed a fault prediction method for wind turbine gearboxes that combines locality preserving projections (LPP), kernel extreme learning machine (KELM), and information entropy. Li et al.[4] introduced an online early fault prediction method for inter-turn short circuits in the excitation windings of synchronous generators using deep belief networks (DBN). Liu et al.[5] combined kernel principal component analysis (KPCA) with gradient boosting decision tree (GBDT) algorithms to construct a fault prediction model for rolling bearings in aviation generators. Wei et al.[6] applied an XGBoost algorithm optimized by the Bayesian optimization algorithm (BOA) to build a temperature prediction model for the front bearing of wind turbine generators, thereby allowing for early monitoring of abnormal signals.

Mechanism and knowledge-based fault prediction methods have their own drawbacks, such as model failure due to complex environmental factors and the difficulty in rationalizing experts' practical experience. Data-driven fault prediction methods, on the other hand, do not require consideration of the system's internal complex mechanisms. They directly build fault prediction models using the system's historical data, thus offering higher generality, adaptability, and accuracy.

Currently, there is relatively little research on fault prediction for cloud data centers in power grids. Xiao et al. [7] used density-based spatial clustering of applications with noise (DBSCAN) to partition historical operation data and adopted random forest (RF) classification methods to establish a classification prediction model. Zhu et al. [8] developed a fault prediction model based on thermodynamic mechanisms and data mining methods. Chen et al. [9] used an evidence-based k-nearest neighbor algorithm for early fault warning of equipment. These methods can achieve fault prediction for cloud data centers, but the accuracy and timeliness of the prediction models still need improvement.

III. METHODS

A. ULM

Anomaly scoring algorithm and quality classification of power grid cloud platform microservice operation alarm data, and screening method for suspected abnormal information: This topic divides the results of abnormal information detection into two different types: abnormal labels and abnormal scores. The abnormal label means that after the corresponding algorithm is executed, a label will be assigned to each sample in the power grid cloud platform microservice data set to indicate whether the sample is an abnormal object, which is equivalent to the output of a two-classification algorithm. Anomaly score means that each sample will be assigned an anomaly score. Generally, the larger the score, the higher the abnormality of the corresponding sample. The algorithm calculates or sorts the abnormality threshold based on the abnormality score of each data sample to be detected. The top k objects can be selected as abnormal data according to specific needs, or a specified threshold can be used to segment all objects into abnormal and normal.

This topic's alarm information early warning technology based on unsupervised learning for the microservice architecture of the power grid cloud platform can take the following steps:

- Data preprocessing: Preprocess the collected power grid cloud platform microservice alarm business data, including data cleaning, standardization, normalization and other operations to remove noise and outliers and ensure the quality and consistency of the data.
- Feature extraction: Extract relevant features from the preprocessed power grid cloud platform microservice data. These features can include time series data, system performance indicators, network traffic data, etc. For time series data, common feature extraction methods include statistical features, frequency domain features and time domain features. For cloud platform system performance indicators, key indicators related to faults can be extracted, such as CPU utilization, memory usage, disk I/O, etc.
- Cluster analysis: Use clustering algorithms in unsupervised learning, such as K-Means, DBSCAN, etc., to perform cluster analysis on the data characteristics of the microservices of the power grid cloud platform. The results of cluster analysis can help identify potential unusual patterns and data structures. By grouping similar power grid cloud platform microservice alarm data together, abnormal events with similar characteristics can be discovered, thus providing clues and insights for cloud platform fault detection.
- Anomaly detection: Based on the cluster analysis of microservice data on the power grid cloud platform, anomaly detection algorithms are used to identify abnormal samples, that is, data points or events that are significantly different from normal samples. Common anomaly detection algorithms include statistics-based anomaly detection algorithms (such as box plots), distance-based anomaly detection algorithms (such as Isolation Forest), etc.
- Rule mining: Mining the corresponding rules for the detected abnormal samples of microservices on the power grid cloud platform. Through the association rule mining method, the correlation and dependency between different alarms can be discovered. For example, if network fault A occurs, network fault B is likely to occur. This rule reveals the correlation between fault A and fault B, which can be used as the basis for subsequent warnings. When fault A is detected, the cloud platform system can provide an early warning of possible fault B so that timely measures can be taken to prevent further faults.
- Model evaluation and optimization

This study categorizes the results of anomaly detection into two different types: anomaly labels and anomaly scores. Anomaly labels refer to the process where the respective algorithm assigns a label to each sample in the power grid cloud platform microservice dataset upon completion, indicating whether the sample is an anomaly, akin to the output of a binary classification algorithm. Anomaly scores, on the other hand, assign an anomaly score to each sample, where a higher score typically indicates a higher degree of anomaly for the corresponding sample. The algorithm can calculate anomaly thresholds or sort based on the anomaly scores of each data sample to be detected. Depending on specific requirements, it can select the top k objects as anomaly data or use a designated threshold to classify all objects into anomalous and normal categories. The detailed approach is as follows:

Assume the power grid cloud platform microservice dataset contains P true anomaly samples and N true normal samples. For a specified data analysis, in the anomaly detection results:

- The number of samples that are actually anomalies and are predicted as anomalies by the algorithm is defined as True Positive (TP).
- The number of samples that are actually anomalies but are predicted as normal by the algorithm is defined as False Negative (FN).
- The number of samples that are actually normal but are predicted as anomalies by the algorithm is defined as False Positive (FP).
- The number of samples that are actually normal and are predicted as normal by the algorithm is defined as True Negative (TN).

Thus, we can define the following:

$$\mathrm{P} = \mathrm{TP} + \mathrm{FP} \tag{1}$$

$$N = TN + FN \quad (2)$$

For data that clearly exhibits anomaly threshold characteristics based on the aforementioned scoring algorithm, it is directly classified as an anomalous data type. For suspected anomalous data objects, this project implements anomaly diagnosis techniques for the power grid cloud platform's operational monitoring data based on statistical models, proximity, clustering, and classification. These techniques will be used to detect and evaluate anomalies in the power grid cloud platform monitoring data, and depending on different data, requirements, and objectives, an improved K-means clustering and Local Outlier Factor (LOF) outlier detection technique will be applied to establish a combined detection model to achieve proactive warning prediction of cloud platform operational information.

To validate the effectiveness of the Unsupervised Learning Model (ULM) methods, the following three networks were selected for comparative experiments. Detailed descriptions of them are as follows:

B. *VAMPIRE*

VAMPIRE[10] is a lightweight pre-training framework designed for effective text classification when data and computational resources are limited. It pre-trains a unigram document statistical model as a Variational Autoencoder (VAE) using domain-specific unlabelled data, and then employs its internal states as features in a downstream classifier. The structural diagram is shown in Figure 1.

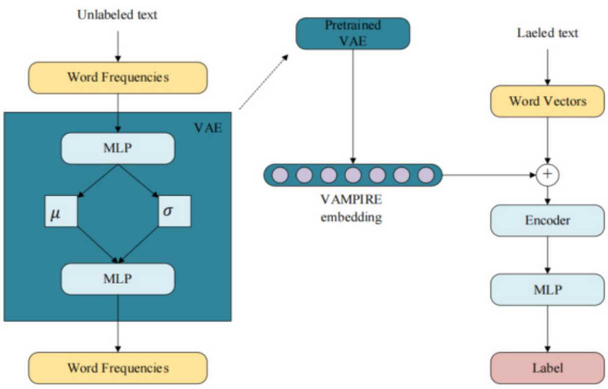

Figure 1. VAMPIRE network structure diagram

C. *BERT*

Unlike other language representation models, BERT[11] (Bidirectional Encoder Representations from Transformers) is designed to pre-train deep bidirectional representations from unlabelled text by jointly conditioning on both left and right context in all layers. In this experiment, a pre-trained uncased BERT model was fine-tuned for classification tasks. Average pooling was applied to the output of the BERT encoder. The network structure is illustrated in Figure 2.

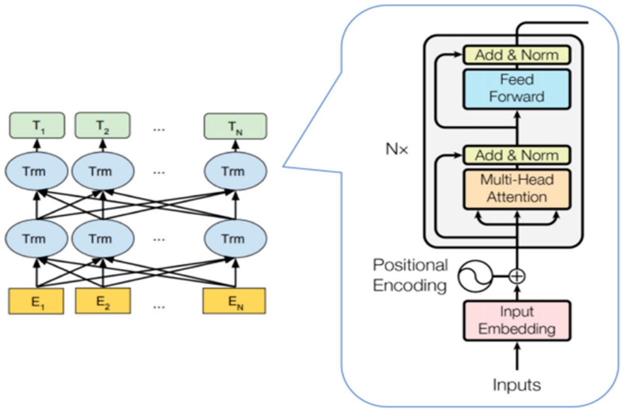

Figure 2. Bert network structure diagram

D. *MixText*

MixText[12] is a semi-supervised learning method for text classification that uses the TMix data augmentation technique. It inserts text in the hidden layers to create a large number of augmented training samples and performs label guessing on unlabelled data. By mixing labeled, unlabeled, and augmented data, MixText demonstrates significant performance advantages over current pre-training and fine-tuning models. The network structure is illustrated in Figure 3.

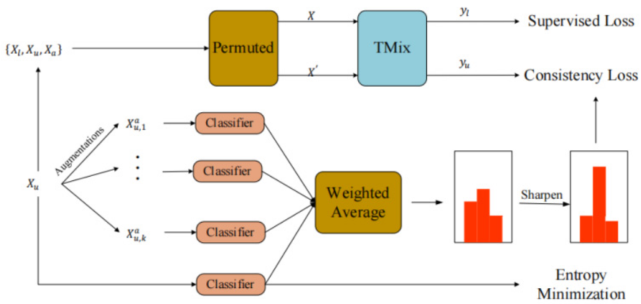

Figure 3. MixText network structure diagram

For the parameter a in the Beta distribution, Figure 4 illustrates the distribution imagesfor different values of a. Generally, when the labeled data for each class is fewer than 100, ais set to 0.4. According to the formula for calculating , a smaller a is more likely to producea y around 0.1. When the number of labeled samples is very low, the model's initialprediction capability is poor, so the influence of the unsupervised learning loss functionshould be weakened to balance the impact of unlabeled data on the model. As the modellearns and its prediction capability improves, setting a to 2 is more likely to produce a yaround 0.5, thereby increasing the influence of unsupervised learning on the model. whenthe number of labeled samples for each class exceeds 200, the model has a decentclassification ability even in the early stages, so setting a to 16, which increases theprobability of producing a y around 0.5, is more appropriate. In this paper, the value of a is set to 16. The Beta distribution imagesare shown in Fiqure 4.

Regarding the hyperparameter T in the Softmax function, when T decreases, the probabilitydifferences between the classes in the Softmax output become larger. Conversely, when

Tincreases, the prediction results across classes tend to be smoother. Therefore, in thisexperiment, the hyperparameter T is set to 0.6.

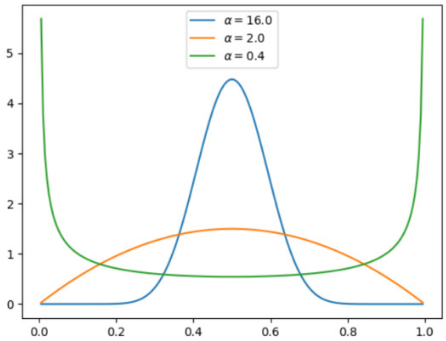

Figure 4. Distribution of beta when a takes different values

IV. EXPERIMENTS RESULTS

A. Datasets

This experiment uses the power grid cloud platform information system fault and alarm data set (hereinafter referred to as the FAD data set), with a total of 10,700 pieces of data.

B. Ressults

Table 1 classified by each model

Datasets	model	Precision(%)		
		10	200	2500
FAD	VAMPIRE	53.8	70.1	75.3
	BERT	57.4	76.9	80.2
	MixText	67.1	79.7	81.8
	ULM	70.3	81.0	82.9
SST-2[13]	VAMPIRE	53.3	73.9	78.5
	BERT	56.8	76.6	80.3
	MixText	70.0	81.5	83.0
	ULM	70.5	81.3	84.6
SubJ	VAMPIRE	66.2	84.6	88.5
	BERT	71.4	88.3	91.5
	MixText	80.1	91.4	92.7
	ULM	79.7	92.1	94.8
IMDB	VAMPIRE	64.3	82.2	85.8
	BERT	67.5	86.9	89.8
	MixText	78.7	89.4	91.3
	ULM	77.4	90.4	92.1

From Table 1, it can be observed that when the number of labeled samples is severely limited, models trained using unsupervised learning networks can achieve good experimental results. For example, when each class in the dataset contains only 10 labeled samples, ULM achieves the highest classification accuracy at 70.3%, while the lowest accuracy is 53.8% for VAMPIRE. Employing supervised learning in such scenarios would lead to severe overfitting. When the number of labels is sufficient, the maximum difference in accuracy between ULM and supervised learning is only 4.9%, as compared to existing supervised learning experimental results. Moreover, for imbalanced datasets, compared to using LTA techniques alone, ULM networks can achieve results similar to supervised learning, and in some cases even outperform it.

Failure Accuracy Rate Of Power Grid Cloud Platform Similarly, for imbalanced datasets, compared to using LTA techniques alone, LTA-SSL networks can achieve results similar to supervised learning, and in some cases even outperform it. Particularly, in the IMDB dataset related to the power grid cloud service, the experimental results of LTA-SSL are much better than LTA alone. Numerically, LTA-SSL achieves the highest classification accuracy at 92.1%, whereas with LTA alone, the model's best performance is only 81.6%. This is due to the relatively long average sentence length in the IMDB dataset, resulting in a large amount of redundant information. Even with the application of LTA techniques, the dataset exhibits significant class imbalance, making supervised learning networks prone to overfitting. On the other hand, when the amount of labeled data is severely insufficient, LTA-SSL networks can still achieve good experimental results, offering broader application scenarios and practical significance. Furthermore, compared to VAMPIRE, the other models consistently demonstrate better performance. In most experiments, the ULM model's classification accuracy is higher than that of other models, confirming the effectiveness of the ULM method.

Additionally, for the FAD and IMDB datasets, more batches of labeled data were selected for training, with each batch containing 5000 unlabeled samples. The experimental results are shown in Figure 5 and Figure 6. From the figures, it can be observed that, compared to VAMPIRE, the other three models exhibit better and more stable performance.

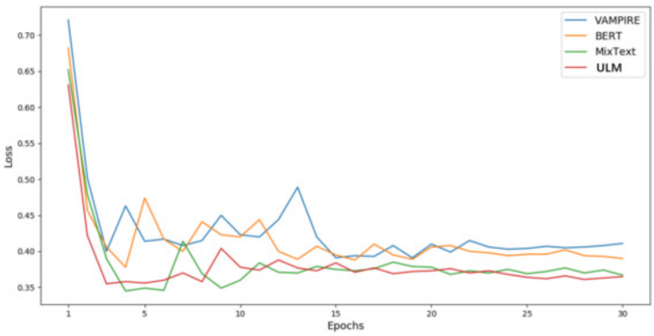

Figure 5. Loss curve on power grid cloud platform information system failure and alarm data

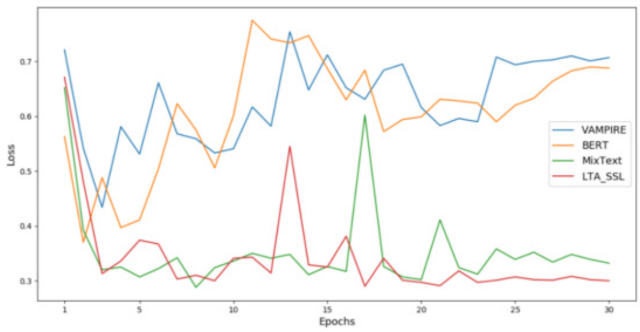

Figure 6. Loss curve on power grid cloud platform information system failure and alarm data

V. CONCLUSIONS

This article uses the above steps to implement the power grid cloud platform microservice architecture alarm information early warning technology based on unsupervised learning, discovers potential abnormal patterns and rules, issues early warning signals in a timely manner, and improves the ability to identify and process power grid cloud platform microservice architecture alarms. According to the above steps, based on the establishment of the power grid business microservice operation alarm log alarm information knowledge base, this topic will study the cloud platform operation monitoring data scoring algorithm, classify the cloud platform monitoring information, and then screen out the suspected abnormal data , for different business types of complex scenarios, research on abnormal information classification mining technology based on statistical models, proximity, clustering and classification, and apply improved K-means clustering and LOF outlier detection technology to establish a combined detection model, Perform clustering and outlier detection on abnormal data to realize active early warning technology of unsupervised learning.

ACKNOWLEDGMENT

In the process of experiment and writing, the contributors get a lot of writing suggestions and work guidance from editors and reviewers, which helps us to make the content more rigorous and easier to understand. Here, we would like to express our most sincere thanks to you. Again, this work is supported by the State Grid Corporation Science and Technology Project Funded “Key technology and product design research and development of power grid data pocket book”(5700-202228195A-1-1-ZN).

REFERENCES

[1] Ahmed M, Seraj R, Islam S M S. The k-means algorithm: A comprehensive survey and performance evaluation[J]. Electronics, 2020, 9(8): 1295.

[2] Cheng Z, Zou C, Dong J. Outlier detection using isolation forest and local outlier factor[C]//Proceedings of the conference on research in adaptive and convergent systems. 2019: 161-168.

[3] LIU S, LIU C L, ZENG H Q. Research on fault warning for wind turbine gearbox based on kernel extreme learning machine[J]. China Measurement & Test, 2019, 45(2): 121-127.

[4] LI J Q, CHEN Y T, LI S X. Early warning of inter-turn short circuit fault in excitation winding of synchronous generator based on deep confidence network[J]. Electric Power Automation Equipment, 2021, 41(2): 153-158.

[5] LIU Y, WANG C, ZHOU P. A warning method for rolling bearing fault of civil aero-engine[J]. Journal of Propulsion Technology, 2021(2): 289-298.

[6] WEI L, HU X D, YIN S. Optimized-XGBoost early warning of generator front bearing fault[J]. Journal of System Simulation, 2021(10): 2335-2343.

[7] XIAO L, LUO J, OUYANG C M. Research on coal mill fault prediction based on semi-supervised learning method[J]. Thermal Power Generation, 2019, 48(4): 121-127.

[8] ZHU P C, QIAN H, JIANG C. Coal mill early warning system based on the combination of thermodynamic mechanism and data mining[J]. Journal of Harbin University of Science and Technology, 2020, 25(1): 43-50.

[9] CHEN X, WANG P, HAO Y. Evidential KNN-based condition monitoring and early warning method with application in power plant[J]. Neurocomputing, 2018, 315: 18-32.

[10] Gururangan S, Dang T, Card D, et al. Variational pretraining for semi-supervised text classification[J]. arXiv preprint arXiv:1906.02242, 2019.

[11] Bert: Pre-training of deep bidirectional transformers for language understanding[J]. arXiv preprint arXiv:1810.04805, 2018.

[12] Chen J, Yang Z, Yang D. Mixtext: Linguistically-informed interpolation of hidden space for semi-supervised text classification[J]. arXiv preprint arXiv:2004.12239, 2020.

[13] Quteineh H, Samothrakis S, Sutcliffe R. Textual data augmentation for efficient active learning on tiny datasets[C]//Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020: 7400-7410.