INTERNATIONAL JOURNAL OF TECHNO-ENGINEERING
IJTE
ISSN: 2057-5688

ANOMALY DETECTION IN SELF-ORGANIZING NETWORKS: CONVENTIONAL VERSUS CONTEMPORARY MACHINE LEARNING

[1]B Vasantha, [2]B Krithika, [3]G Satwika, [4]J Sreeja, [5]M Tejashwini

[1]Assistant Professor, Department of CSE(DS), Malla Reddy Engineering College for Women (Autonomous Institution – UGC, Govt. of India), Hyderabad, INDIA.

[2345]UG,Department of CSE(DS), Malla Reddy Engineering College for Women (Autonomous Institution – UGC, Govt. of India), Hyderabad , INDIA.

ABSTRACT

This paper presents a comparison of conventional and modern machine (deep) learning within the framework of anomaly detection in self-organizing networks. While deep learning has gained significant traction, especially in application scenarios where large volumes of data can be collected and processed, conventional methods may yet offer strong statistical alternatives, especially when using proper learning representations. For instance, support vector machines have previously demonstrated state-ofthe-art potential in many binary classification applications and can be further exploited with different representations, such as one-class learning and data augmentation. We demonstrate for the first time, on a previously published and publicly available dataset, that conventional machine learning can outperform the previous state-of-the-art using deep learning by 15% on average across four different application scenarios. Our results further indicate that with nearly two orders of magnitude improvement in computational speed and an order of magnitude reduction in trainable parameters, conventional machine learning provides a robust alternative for 5G self-organizing networks especially when the execution and detection times are critical.

INDEX TERMS Anomaly detection, deep learning, machine learning, mobile network communications, self organizing networks.

I INTRODUCTION

Cellular networks induce a large amount of data and 5G networks further increases the data and it is expected to become even more complex. In order to address upcoming challenges, SONs approach is used which leads to a number of limitations and the effectiveness in a 5G scenario would not be sufficient to achieve the global network optimization and operational cost reduction goals. Thus, in the future we can opt for Big Data Empowered SONs (BES) as the most effective solution to autonomic network management for 5G systems. In this enormous data there could be anomalies during calling activities due to large amount of users. These anomalies can be detected using advanced technologies like CADM using CDR. The solution to this problem is to apply knowledge based anomaly detection methods and set rule policies depending on network behaviour. Here, we present one variant of knowledge based technique, a rule-based technique, for detecting network anomalies for users traveling from one city to another. The method is flexible as well as robust for the detection of anomalies. We use an approach for anomaly detection by analysing call-detail records in combination with recent Big Data analytical tools (Hadoop (HDFS, Map-Reduce)). Since this algorithm affects the performance and security of the users, we can use other wide range of algorithms that can be applied for detecting anomalies with the most efficient one being clustering techniques. The most important, unsupervised learning processes such as K-means clustering algorithm which is a very simple algorithm for finding useful patterns. Its inability to escape from local optima can be overcome by combining with the PSO algorithm. PSO is a high efficient heuristic technique having the capability to escape from local optima and with low computational complexity. The anomaly detection technique combines the Kmeans and the PSO algorithm and it is performed by comparing real traffic and clusters centroids thus enhancing the performance and securing the privacy of the users.

II RELATED WORK

SOURCES OF INFORMATION WITHIN A MOBILE NETWORK: The information that we mentioned in the introduction comes from various network elements, such as base stations, mobile terminals, gateways and management entities, and these can be categorised as follows: 1. Control information related to regular short-term network operation, covering

ISSN: 2057-5688

functionalities such as call/session set-up, release and maintenance, security, QoS, idle and connected mode mobility, and radio resource control. 2. Control information related to SON functions associated with optimization and maintenance of cellular systems, and which cover aspects such as radio link failure statistics, inter-cell interference and cell load signalling, etc. 3. Management information covering long-term network operation functionalities, such as Fault, Configuration, Accounting, Performance and Security management (FCAPS), as well as customer and terminal management. An example of such information is that defined for Operation and Management (OAM), which consists of aggregated statistics on network performance, such as number of active users, active bearers, successful/failed handover events, etc. per base station, as well as information gathered by means of active probing 4. Authentication, Authorization and Accounting (AAA) information, including for example Charging Data Records (CDRs). 5. Customer relationship information, e.g., complaints about bad service quality, churn information, etc.

By leveraging all the information available within a mobile network, a Big-data Empowered SONs (BES) approach can bring mobile networks to improve the operational processes and culture of the organisation to meet the expectations of the customers. Some of the practical examples are mentioned below: A BES would autonomously recognize the cause of the problem on basis of data which was recorded previously and restore it with little or no engineer interference, where as in simple threshold based network fault detection systems that needs an engineer interference to run measurements to identify the source of the problem.

By analysing the correlation among conflicting performance goals of various SON functions, and dynamically specifying operating point that gives the best performance trade-off, BES would solve the coordination between SON functions, which is the current open challenge. Traditionally data is inaccessible for network management purposes, such as CDRs, it could be leveraged by a BES, for example to determine the users typical mobility patterns , and proactively get the network optimization actions, such as taking informed handover decision, concentrating spectrum resources in location with larger number of users, and when base stations are foreseen to be not needed, turn off the base stations.

III IMPLEMENTATION

All implementations were performed using Python. We tested both one-class and binary SVM models on all datasets. The SMOTE algorithm was used with the binary SVM model to adjust for the imbalanced datasets. The datasets were preprocessed, where the time, UserID, and location features were not included in the training process for fairness. Approximately 10 % of the normal and anomalous samples were separated for testing. Both the training and testing samples were normalized to between (0,1]. The one-class SVM model was trained with only normal samples using Gaussian RBF kernels. After training, we generated the SVM probability outputs with test samples, including both normal and anomalous samples, to obtain receiver operating characteristic (ROC) curves along with area-underthe-curve (AUC) scores as performance metrics. The binary SVM model is trained in exactly the same fashion except that in addition to the above process, the anomaly samples are oversampled with SMOTE to generate balanced datasets prior to training and testing the algorithm.

EVALUATION METRICS

In this study, we evaluated performance by looking at ROC curves and AUC scores. The ROC is a probability curve that shows the model's ability to identify the positive class appropriately. It is plotted with the true positive rate (TPR) on the y-axis and false positive rate (FPR) on the x-axis, where TPR is the percentage of correctly classified positive outputs and FPR is the percentage of incorrectly classified positive

IV RESULTS & DISCUSSION

The ROC curves for each of the four datasets are shown in Fig. 3. The red curve in each figure represents the latest state-of-the-art reported on this dataset using a deep autoencoder [8]. For comparison, different SVM implementations are represented in different colors, including binary and one-class combinations with and without augmentation using SMOTE. The results

ISSN: 2057-5688

are generally consistent except in the case of dataset 4, where all SVM combinations outperformed the deep autoencoder at all levels of TPR & FPR. In the case of the first three datasets, the deep autoencoder outperformed the SVM implementations without augmentation. However, when SMOTE is used, both one-class and binary modalities of the SVM demonstrate higher performance compared to the original paper, and in some cases significantly so.

It is important to discuss the nonsignificant or in some cases negative impact of SMOTE on the performance of one-class SVM topology. SMOTE works by generating samples based on the nearest neighbor similarities of intraclass samples and the differences of interclass samples. This method works best when training a binary classifier where one class may be less represented than another class as evident by the performance boost observed in binary SVM training. However, in one-class learning representation, only the majority class is used in the training which would mean that SMOTE could only have an indirect effect on the number and quality of the samples generated for the normal class and subsequently does not have a direct role in boosting performance. The drop in performance in some cases can similarly be linked to the quality of the anomaly class samples being generated (and thus effecting testing performance) not making up for the additional information which now cannot be used in the training process.

V CONCLUSION

In this study, we explored the premise of conventional machine learning when compared to deep learning for anomaly detection in SONs. Anomaly detection was a popular application area of deep learning for cell outages in communication networks. However, as in other domains, conventional methods can still provide strong statistical alternatives to the right learning representations. In this paper, we focused on SVMs with one-class and binary learning scenarios on a previously published and publicly available dataset. We found that while deep learning was highly competitive, standard SVMs using RBF kernels, can be trained to outperform a deep autoencoder approach. Both one-class

ISSN: 2057-5688

and binary classification can benefit immensely from synthetic augmentation of the dataset using SMOTE with improvements in detection accuracy by as much as 15% on average over four different application scenarios. Future work will study the impact of augmentation on other learning algorithms, specifically statistical deep learning, such as variational auto-encoders. Work presented in this paper can further be extended to other applications beyond anomaly or outage detection. Specifically, there has been increased attention to modulation detection in next generation mobile wireless networks where fast, robust, and light machine learning models could enable time-critical applications in signal classification and modulation detection. Improvements in speed can be realized both at the algorithm level and data preprocessing stages using techniques such as principal component analysis to identify the most relevant features for classification and detection. Finally, statistical learning algorithms, such as Gaussian Process Regression, which have gained immense popularity as alternatives to deep learning can be applied to different scenarios especially when data is not present in sufficiently large volumes to properly train DL models with many parameters.

```
[ ] rf=RandomForestClassifier()
    rf.fit(X_transformed,y)
    rf_pred=rf.predict(X_transformed)

[ ] rf_df=pd.DataFrame()
    rf_df['actual']=y
    rf_df['pred']=rf_pred
    rf_df.head()

[ ] print(accuracy_score(y, rf_pred))

[ ] confusion_matrix(y, rf_pred)

[ ] print(classification_report(y, rf_pred))
```

Fig 1: Model building

REFERENCE:

[1] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski,''Five disruptive technology directions for 5G,'' IEEE Commun. Mag., vol. 52, no. 2, pp. 74–80, Feb. 2014.

[2] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, ''What will 5G be?'' IEEE J. Sel. Areas Commun., vol. 32,no. 6, pp. 1065–1082, Jun. 2014.

[3] A.-S. Bana, E. de Carvalho, B. Soret, T. Abrão, J. C. Marinello, E. G. Larsson, and P. Popovski, ''Massive MIMO for Internet of Things (IoT) connectivity,'' Phys. Commun., vol. 37, Dec. 2019, Art. no. 100859.

ISSN: 2057-5688

[4] D. Ciuonzo, P. S. Rossi, and S. Dey, ‘‘Massive MIMO channel-aware decision fusion,’’ IEEE Trans. Signal Process., vol. 63, no. 3, pp. 604–619,Feb. 2015.

[5] Airhop. Accessed: Jun. 18, 2021. [Online]. Available: http://www. airhopcomm.com

[6] I. de-la-Bandera, R. Barco, P. Muñoz, and I. Serrano, ‘‘Cell outage detection based on handover statistics,’’ IEEE Commun. Lett., vol. 19, no. 7,pp. 1189–1192, Jul. 2015.

[7] Ericson. Network Outage. Accessed: Jun. 18, 2021. [Online]. Available: http://telecoms.com/494091/

[8] M. Z. Asghar, M. Abbas, K. Zeeshan, P. Kotilainen, and T. Hämäläinen,‘‘Assessment of deep learning methodology for self-organizing 5G networks,’’ Appl. Sci., vol. 9, no. 15, p. 2975, Jul. 2019.

[9] S. Rajendran, W. Meert, V. Lenders, and S. Pollin, ‘‘Unsupervised wireless spectrum anomaly detection with interpretable features,’’ IEEE Trans. Cognit. Commun. Netw., vol. 5, no. 3, pp. 637–647, Sep. 2019.

[10] J. Burgueño, I. de-la-Bandera, J. Mendoza, D. Palacios, C. Morillas, and R. Barco, ‘‘Online anomaly detection system for mobile networks,’’ Sensors, vol. 20, no. 24, p. 7232, Dec. 2020.