A GNN-based Few-shot learning model on the Credit Card Fraud detection

Rongrong Jing
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
jingrongrong2019@ia.ac.cn

Hu Tian
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
tianhu2018@ia.ac.cn

Gang Zhou
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
zhougang2020@ia.ac.cn

Xingwei Zhang
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
zhangxingwei2019@ia.ac.cn

Xiaolong Zheng*
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
xiaolong.zheng@ia.ac.cn

Daniel Dajun Zeng
Institute of Automation, Chinese Academy of Sciences
School of Artificial Intelligence, Chinese Academy of Sciences
Beijing, China
dajun.zeng@ia.ac.cn

2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI) | 978-1-6654-3337-2/21/$31.00 ©2021 IEEE | DOI: 10.1109/DTPI52967.2021.9540093

***Abstract*—In the era of big data, large-scale data can be very effective in improving model performance. However, in the real world, high-quality data is usually difficult to acquire due to privacy or cost. Especially when it comes to credit card fraud, the fraud samples are quite rare. Detecting card fraud with few samples is a meaningful task. Graph neural network (GNN) is a good way to deal with few samples because an advantage of GNN is that information can be disseminated through connections between nodes. However, the data structure of credit cards cannot be applied by the GNN-based method directly. In this paper, we proposed a GNN-based few-shot learning method which can detect credit card fraud with few samples effectively. We constructed a learnable parametric adjacency matrix method relying on the similarity of features to pass messages and utilized the GCN layer to extract node features. We compared our method with classical machine learning algorithms and other graph neural networks on the real-world data set. Our experimental results show that our proposed model can perform better extremely with fewer training samples than baselines.**

Keywords—few-shot learning, card fraud detection, graph neural network

I. INTRODUCTION

Credit card payment is a common way of payment in our daily life, which is widely used on a variety of occasions, such as eating, shopping, travel and so on. However, with the popularity of credit cards, some people violate the credit card management laws and regulations and use credit cards to commit fraud and defraud property. Fraudulent transactions by users may cause serious losses to cardholders and credit card companies Therefore, using transaction information to detect fraud is helpful for card issuers to realize anti-fraud and protect cardholders' property security. However, due to the privacy of users and trade secrets, it is difficult to obtain real credit card transaction information. Besides, the quality of data is not good enough. Due to confidentiality issues, those existing data sets cannot provide the original functionality and more background information about the data. This kind of data problem can not only exist in the credit cards scene but in all areas of life. In the data-driven era, data is becoming more and more valuable, and data sharing and protection have become a serious issue.

Deep learning-based models rely on large amounts of training data. A severe lack of data will lead to model overfitting. However, Tagged samples are very rare actually in the real world due to the acquisition costs and labeling costs. Although there are lots of solutions to solving these problems like data augmentation and regularization, there are still lots of problems needed to be solved. At the same time, few-shot learning has attracted the attention of scholars with the development of machine learning and deep learning. Few-shot learning for classification problems has been used in many fields like computer vision and natural language progressing and applied to a variety of fields widely like medical fields, financial fields.

The concept of the few-shot emerged from the field of computer version for image processing tasks like image classification, recognition, and segmentation. Then, few-shot learning has made some progress in the field of natural language processing. The most excellent models are applied in these two fields. Among them, Graph neural network (GNN) is a good way to deal with few samples because an advantage of GNN is that information can be disseminated through connections between nodes. However, the data structure which is different from the image cannot be applied by the GNN-based few-shot method directly.

To address these limitations, we proposed a GNN-based few-shot learning end-to-end model which can detect fraud with few samples effectively. We construct a learnable parametric adjacency matrix method relying on the similarity of features to pass messages and utilize the GCN layer to extract node features. We compare our method with classical machine learning algorithms and other graph neural networks on the real-world data set. Our experimental results show that common classifiers which depend on huge amounts of data cannot predict accurately and our proposed model performs better with fewer training samples. Our model not only can be applied in this task but also can be extended to solve other tasks.

The remainder of this paper is designed as follows. We introduce the related work about few-shot with graph neural network in the section II. And we introduce the framework of few-shot problem and describe our method in section III.

* Corresponding author

Section IV displays the experiment results. Finally, we conclude our work in Section V.

II. Related Work

A. Few-shot on fraud card detection

Credit card frauds can be made in many ways such as simple theft, application fraud, counterfeit cards, never received issue (NRI) and online fraud (where the cardholder is not present) [1]. It is an extremely difficult but common problem that needs to be solved. Dataset may exist lots of problems, such as label imbalance and data missing [2]. In particular, the success of credit card fraud detection models largely depends on a large amount of training data. Recently, more researches focus on the few samples of card fraud detection. Meta-learning received a lot of attention. Stolfo et al. [3] applied some machine learning models with meta-learning strategies on the real-world credit card dataset and explored the relationship between the data distribution and evaluation indicators. Pun et al. [4] constructed the meta-learning classifier by combining the basic classifiers and their predictions to detection fraud using fewer training samples. Recently, few-shot learning algorithms have been applied widely. Zheng et al. [5] proposed the federated framework to make banks train models on their own local data using a meta-learning-based approach proposed classifier on K negative samples in each mini-batch. Zhou et al. [6] regard the multi-label classification as a semi-supervised few-shot learning problem using the prototypical network. Kavitha et al. [7] proposed the few-shot learning model to solving the unbalanced data problem with few samples.

B. Few-shot learning

Few-shot learning is firstly applied in the field of computer vision and the development of the application of natural language process is slower. Few-shot learning can be applied to various tasks where there are few samples. Zhao et al. [8] divided the few-shot models according to the learning framework into three categories, metric-learning-based, meta-learning-based, GNN-based. The structure of the graph has the advantages of strong expressiveness and intuitive display. With the rise of machine learning in recent years, machine learning has been gradually applied to graph analysis. Graph neural network is a model for processing graph domain information based on deep learning. Due to its good performance and interpretability, it has become a widely used graph analysis method recently.

Garcia et al. [9] proposed the GNN-based few-shot learning framework to classify images. Kim et al. [10] focus on the edge-labeling graph and proposed the GNN-based few-shot model that can predict whether two connected nodes belong the same class. Gidaris et al. [11] reconstructed the classification weights using Denoising Autoencoder network on the GNN-based few-shot model. GNN-based model is significant to be explored widely because of its interpretability and good performance, especially can be expanded into other areas.

III. Methods

In this section, we demonstrated the framework of few-shot learning task and introduced our GNN-based few-shot model.

A. Problem Definition

The purpose of few-shot learning is to build a model that can be learned from a small number of samples. C-way K-shot problem indicates that the model need to distinguish C classes by learning from $C \times K$ samples. Meanwhile, K, a pretty small number, indicates that only Ksamples are participated in every training episode. The total training process is divided into lots of meta-task and $C \times K$ samples from the training set will be selected as the support set to predict the label of a sample selected randomly in each meta-task.

Given the credit card fraud data $T = \{(x_1, l_1), \dots, (x_s, l_s), (\tilde{x}_1, l_1), \dots, (\tilde{x}_t, l_t)\}$, where s is the size of training set, t is the size of testing set, and $l \in \{0,1\}$ represents the label of sample whether is fraud. As mentioned above, we need to classify one sample at each task. The input for each training meta-task is $T_i = \{(x_1, l_1), \dots, (x_k, l_k), (x_{k+1}, l_{k+1}), \dots, (x_{2k}, l_{2k}), (x_p, l_p)\}$, where $l_1, \cdots, l_k = \{0\}$ and $l_{k+1}, \cdots, l_{2k} = \{1\}$, (x_p, l_p) which needs to be classified is selected from training set at random in the training process and it will be replaced by $(\tilde{x}_{test}, l_{test})$ from the testing set when we test the performance of our model. As shown in Fig.1, the i_{th} sample x_i and its label l_i can be spliced together as an initial input vector to the GNN model. So, the input for each task contains $2K \times C + 1$ samples. The object function is as follows.

$$\min_{\theta} \frac{1}{L} \sum_{i \le L} l(\phi(T_i; \theta), Y_i) + R(\Theta) \quad (1)$$

where L means the number of meta-tasks, and $R(\theta)$ is the regularization to avoid overfitting. Due to the card fraud detection task is a classification task, we choose the Cross entropy loss function.

$$l(\phi(T; \theta), Y) = -\sum_{k} y_k \log P(Y_* = y_k | T) \quad (2)$$

where $k \in \{0,1\}$, The object function can be simplified as follows.

$$L_\Theta = \min_{\Theta} \frac{1}{L} \sum_{i \le L} \sum_{k \in \{0,1\}} y_k \, log \, P(Y_* = y_{ik} | T_i) + R(\Theta) \quad (3)$$

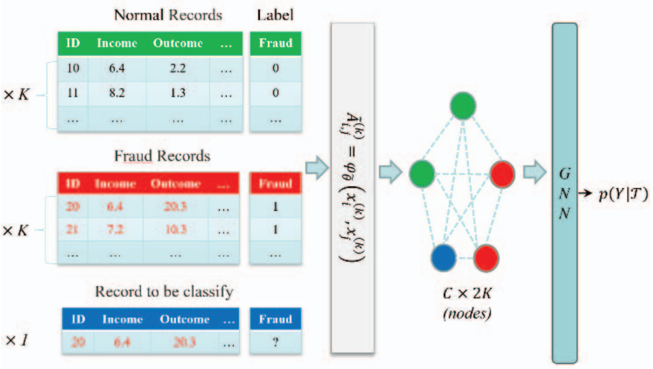

Fig. 1. The framework of few-shot learning on the credit card detection task.

B. Graph neural network

We regard the fraud detection task as a node classification task where the label information can be propagated from the samples labeled to the unlabeled node, which can assist in improving the prediction performance on a few samples. Each sample set T_i in each meta-task i is regarded as a graph

$G_i(X, E)$, where X contains those labeled samples and unlabeled sample x_i which needs to be classified. The difference between the training and testing process is where the x_i comes from.

For the input signal $F \in R^{V \times d}$ calculating from the given the weighted graph, we use the adjacency matrix A and graph convolution layer $Gc(\cdot)$ whose basic architecture is designed as Garcia et al. [9] introduced is setting to learn implicit information of graph.

$$f_l^{(k+1)} = Gc(x^{(k)}) = \rho\left(\sum_{B \in A} Bf^{(k)}\theta_{B,l}^{(k)}\right) \tag{4}$$

where

$$l = 1, \ldots, d_{k+1},$$

$$\theta_{B,l}^{(k)} \in \mathbb{R}^{d_k \times d_{k+1}},$$

$$\theta = \left\{\theta_1^{(k)}, \ldots \theta_{|A|}^{(k)}\right\}_k,$$

$$A = \{\tilde{A}^{(k)}, 1\}$$

They are the parameters need to be trained and $\rho(\cdot)$ selected to be LeakyReLU is a point-wise non-linearity function, where $a \in (0,1)$.

$$\text{LeakyReLU}(x) = \begin{cases} x, & x \geq 0 \\ ax, & x < 0 \end{cases} \tag{5}$$

Considering characteristic of the card fraud data, we defined the adjacency matrix A as follows,

$$\tilde{A}_{i,j}^{k} = \phi_\theta\left(f_i^{(k)}, f_j^{(k)}\right) \tag{6}$$

$$\phi_\theta\left(f_i^{(k)}, f_j^{(k)}\right) = \theta^T abs(f_i - f_j) \tag{7}$$

where ϕ_θ is a parametrized function that can learn the difference between two nodes through the training process. ϕ_θ is a symmetric function utilizing the distance property to measuring similarity, therefore, $\phi(a, b) = \phi(b, a)$ and $\phi(a, a) = 0$.

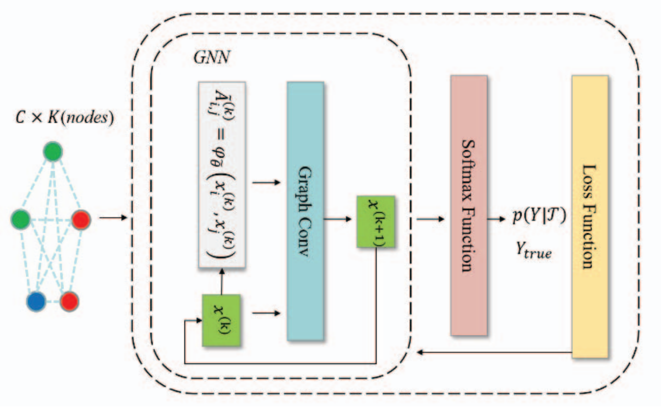

Fig. 2. The architecture of the GNN-based few-shot model.

As shown in Fig.2, the sample vectors with their labels vector are input to the graph neural network to construct the edge between two nodes according to the Eq. 6 - 7. Then, node vectors are updated through graph convolution, and edge vectors are constantly updated through node vectors, which constitutes a deep graph neural network. The last layer of GNN is a Softmax function that can convert the node vectors to the predicted probability of the unclassified node, so that we can calculate the loss by Eq. 3 at each iteration. Then the parameters can be updated by back propagation. After repeated iterations, the accurate credit card detection model with few-shot has been constructed.

IV. Experiment

A. Data

The data set, called German Credit data, was collated by Professor Dr. Hans Hofman (https://archive.ics.uci.edu/ml/machine-learning-databases/statlog/german/). This data contains 1000 real credit records from banks. Each record includes 24 categorial or symbolic characteristics such as job, sex, age and we encode the non-numeric attributes with the one-hot encoding approach. This data set has two advantages over other data sets. Firstly, the proportion (30%) of fraudulent data to the total data is pretty large to ensure that there are enough different negative samples in the training process. Secondly, the data is complete, and there is no missing value. This can avoid additional influencing factors on the model caused by unbalanced sample labels and missing data. The training set, the valid set, and the test set are divided according to the ratio of 7:2:1.

B. Experiment setting

We designed two groups of comparative experiments to show the performance of machine learning and deep learning algorithms. In the first experiment, we compared our model with some practical and classical algorithms to understand the learning process of the model with a small number of training samples. Then the performance of the proposed GNN-based few-shot model under different conditions is analyzed in the second experiment. The baselines are descripted as follows.

a) Three machine learning models are selected, includingLogistic Regression [12], Support Vector Machine [13], and XGBoost [14]. They are widely used in mathematical statistics, typical machine learning and decision tree respectively. They have demonstrated their excellent performance in a variety of different tasks.

b) We adopt GraphSage [15] which is an inductive node embedding framework among the Graph neural network models. Due to the key idea of GraphSage is aggregating characteristic information from the neighbors of a node to learn an embedded function, GraphSage may be suitable for few-shot learning by constructing the graph with the similarity of the node features in order to improve the performance of label propagation. At the same time, GraphSage and our proposed model are both semi-supervised learning models and they all take advantage of graph structure to propagate node features.

C. Result

In the first experiment, we compared our model with baselines. We set up five few-shot learning tasks shown in Fig 1. In order to simulate the condition of few-shot learning for the practical and classical machine learning algorithms, We set up different numbers of training samples. For example, if the few-shot learning task is 2-way 2-shot and we set the number of training iterations is 5, then the number of training samples for baselines is the number of all the samples used in our model. So we set the input size as 25 samples for each experiment. In particular, for GraphSage, we designed a semi-supervised task where the input of the model is the adjacency matrix A calculating by Eq. 6 - 7. Similar to our model, there are also the same number of unlabeled samples to predict. We compared the performance of classifiers under the different few-shot learning tasks.

To avoid randomness, we conducted 100 experiments and when the training process fails to converge, the experiment will be terminated and restarted. We found our model will converge after 20 training iterations through testing. So we set the number of iterations 20. Table 1 shows the average of the accuracies of all effective experiments.

TABLE I. THE ACCURACY UNDER DIFFERENT FEW-SHOT TASKS

Classifier	LR	SVM	XGBoost	GraphSage	Ours
2-way 1-shot	0.51	0.49	0.51	0.50	**0.53**
2-way 2-shot	0.52	0.49	0.50	0.61	**0.70**
2-way 5-shot	0.59	0.53	0.54	0.69	**0.73**
2-way 20-shot	0.81	0.67	0.78	0.79	0.80
2-way 50-shot	**0.90**	**0.72**	**0.92**	**0.91**	**0.88**

As Table. I shown, common classifiers which depend on huge amounts of data cannot predict accurately when the number of iterations is small. The performances of classifiers are not stable due to the shifty distribution of training data. When the amount of training data is small, the classification accuracy is about 0.5 which means that models do not learn the ability to predict data. During the experiment, we observed that when the sample size was small, almost more than half of the experiments failed to converge. On the contrary, our model shows good performance when training data is less. It is proved that our model does have a great effect on few-shot learning. As the training data increases, the prediction accuracy of baselines has been greatly improved. These baselines show their original excellent performances.

TABLE II. THE PERFORMANCE WITH DIFFERENT LABELS

Classifier	At Risk	Precision	Recall	Accuracy
2-way 1-shot	0	0.51	0.46	0.53
	1	0.48	0.54	
2-way 2-shot	0	0.63	0.67	0.70
	1	0.71	0.75	
2-way 5-shot	0	0.78	0.62	0.73
	1	0.51	0.75	

In the second experiment, we explored the performance of the model on different types of samples of each training task after training 20 iterations. As shown in Table. II, it is obvious that the larger the scale of training data at each meta-task, the better the prediction accuracy. At the same time, we can find that there is no significant difference in the prediction effect because the labels at each task are balanced.

V. CONCLUSIONS

In this paper, we proposed a GNN-based few-shot learning end-to-end model which can detect fraud with few samples effectively. We constructed a learnable parametric adjacency matrix method relying on the similarity of features to pass messages and utilize the GCN layer to extract node features. Our experimental results show that common classifiers which depend on huge amounts of data cannot predict accurately and our proposed model performs better with fewer training samples. Our framework provides a novel perspective to regard few-shot learning as a supervised message passing task using the GNN layer and our model can be applied to explore other tasks. And it is worth exploring when the labels are imbalanced extremely in the future work.

ACKNOWLEDGMENT

We want to thank Runsong Jiang, and Zizhen Deng for many discussions. This work is supported the Ministry of Science and Technology of China under Grant Nos: 2020AAA0108401, 2019QY(Y)0101, the Ministry of Health of China under Grant No. 2017ZX10303401-002 and 2017YFC1200302，and the Natural Science Foundation of China under Grant No. 71602184 and 71621002.

REFERENCES

[1] Y. Sahin, and E. Duman, "Detecting credit card fraud by ANN and logistic regression." pp. 315-319.

[2] R. Jing, H. Xian, Y. Li, X. Zhang, X. Zheng, Z. Zhang, and D. Zeng, "Improving the Data Quality for Credit Card Fraud Detection." pp. 1-6.

[3] S. Stolfo, D. W. Fan, W. Lee, A. Prodromidis, and P. Chan, "Credit card fraud detection using meta-learning: Issues and initial results." pp. 83-90.

[4] J. Pun, and Y. Lawryshyn, “Improving credit card fraud detection using a meta-classification strategy,” International Journal of Computer Applications, vol. 56, no. 10, 2012.

[5] W. Zheng, L. Yan, C. Gou, and F.-Y. Wang, "Federated Meta-Learning for Fraudulent Credit Card Detection."

[6] F. Zhou, X. Qi, C. Xiao, and J. Wang, “MetaRisk: Semi-supervised few-shot operational risk classification in banking industry,” Information Sciences, vol. 552, pp. 1-16, 2021.

[7] M. Kavitha, and M. Suriakala, "Real time credit card fraud detection on huge imbalanced data using meta-classifiers." pp. 881-887.

[8] K. Zhao, X. Jin, and Y. Wang, “Survey on few-shot learning,” Journal of Software, vol. 32(2), pp. 349-369, 2021.

[9] V. Garcia, and J. Bruna, “Few-shot learning with graph neural networks,” arXiv preprint arXiv:1711.04043, 2017.

[10] J. Kim, T. Kim, S. Kim, and C. D. Yoo, "Edge-labeling graph neural network for few-shot learning." pp. 11-20.

[11] S. Gidaris, and N. Komodakis, "Generating classification weights with gnn denoising autoencoders for few-shot learning." pp. 21-30.

[12] D. W. Hosmer Jr, S. Lemeshow, and R. X. Sturdivant, Applied logistic regression: John Wiley & Sons, 2013.

[13] T. Joachims, Making large-scale SVM learning practical, Technical report, 1998.

[14] T. Chen, and C. Guestrin, "Xgboost: A scalable tree boosting system." pp. 785-794.

[15] W. L. Hamilton, R. Ying, and J. Leskovec, “Inductive representation learning on large graphs,” arXiv preprint arXiv:1706.02216, 2017.