Data-Driven Approach for Credit Card Fraud Detection with Autoencoder and One-Class Classification Techniques

Abdoul-Fatao Ouedraogo, Cédric Heuchenne, Quoc-Thông Nguyen, Hien Tran

Data-driven approach for Credit Card Fraud Detection with Autoencoder and One-Class Classification techniques

Abdoul-Fatao Ouedraogo[1], Cédric Heuchenne[2], Quoc-Thông Nguyen[3], and Hien Tran[1]

[1] Institute of Artificial Intelligence and Data Science, Dong A University, Da Nang, Vietnam
[2] HEC Management School, University of Liège, Liège 4000, Belgium
[3] Université de Lille, ENSAIT, GEMTEX, F-59000 Lille, France

Abstract. With the development of e-commerce, payment by credit card has become an essential means for the purchases of goods and services online. Especially, the Manufacturing Sector faces a high risk of fraud online payment. Its high turnover is the reason making this sector is lucrative with fraud. This gave rise to fraudulent activity on the accounts of private users, banks, and other services. For this reason, in recent years, many studies have been carried out using machine learning techniques to detect and block fraudulent transactions. This article aims to present a new approach based on real-time data combining two methods for the detection of credit card fraud. We first use the variational autoencoder(VAE) to obtain representations of normal transactions, and then we train a support vector data description (SVDD) model with these representations. The advantage of the representation learned automatically by the variational autoencoder is that it makes the data smoother, which makes it possible to increase the detection performance of one-class classification methods. The performance evaluation of the proposed model is done on real data from European credit cardholders. Our experiments show that our approach has obtained good results with a very high fraud detection rate.

Keywords: Anomalies Detection · Outliers · Autoencoder · Variational Autoencoder · One Class Classification · Credit Card Fraud.

1 Introduction

In recent years with the development of e-commerce, we observe an increase in the volumes of electronic transactions leading to an increase in credit card fraud. Since then,fraud detection has become a topic that reaches out to all industries, such as financial industries, banks, government agencies and insurance, etc. Sectors that often process a large of the transaction, for example, the Manufacturing industry, have been a victim of fraud easily. Especially, in the era

when the relation between finance and manufacturing is concrete [6, 17]. Despite the efforts of struggling organizations, millions of dollars are wasted each year due to fraud. For that, many techniques [5] have been developed in recent years for cybersecurity and reducing fraudulent transactions [13, 19]. Using developed data mining tools such as machine learning through algorithms such as support vector machine, random forest, neural networks, impending models can be produced to detect these fraudulent transactions. These anticipation models can help focus resources in the most efficient way to recover or recover losses due to fraud. Although these methods overcome the deficiency of knowledge acquisition in traditional rule-based expert systems. They still have some deficiencies, especially these methods are based on the idea of supervised learning, which needs a balanced dataset of both normal transactions and fraudulent transactions. So these methods do not work well in case of fraud detection since fraudulent transactions are much fewer than normal transactions.

Therefore, we propose a new approach for credit card fraud detection using a variational autoencoder and one-class classification techniques. In contrast with traditional classification methods which are focused on classifying samples of two or more classes, one-class classification methods try on to learn a model on samples of one class and distinguish them from samples of the other classes. In our work, we focus on distinguishing fraudulent transactions from normal transactions, but we use, normal transactions to train our model to distinguish the other class (fraudulent transactions).

The rest of the paper is structured as follows. Section 2 presents the related work, Section 3 explains our approach with Variational Autoencoder (VAE) and SVDD. Then the Section 4 presents the implementation and results analysis. Finally, the conclusion is given in Section 5.

2 Related works

With the development of e-commerce, the credit card has become an essential payment method for online purchases of goods and services, and since then fraudsters have taken advantage of it to carry out unhealthy activities and steal users money. Due to these problems, several researches on the detection of the credit card fraud have been conducted in order to reduce losses. Many techniques for detecting credit card fraud have been presented in recent years[5]. In [16], the authors presented a new approach for automatic detection of frauds in credit card transactions based on non-linear signal processing and it can be applied to several datasets using parameters derived from key performance indicators of business. Bahnsen et al. [2] proposed a cost the sensitive method based on Bayes minimum risk to represent realistically the monetary gains and losses due to fraud detection.

Hegazy et al.[11] developed Frequent Pattern based on customer's previous transaction activity as Legal or Fraud transactions introducing using Rough Set and Decision Tree Technique clustering algorithm in Enhanced Fraud Miner algorithm which attained good improvement in finding the false alarm rate when

compared to other models. Dai et al. [7] proposed a method that combined the supervised and unsupervised approach by fusing various models to train and record the spending behavior for each cardholder based on their previous transactions and for every new transaction the fraud score is computed from the fraud pattern. A hybrid and adaptive method is presented by Batani [3] to detect fraud using cardholder's financial status, social status, and OTP by assigning weights using Artificial Neural Networks to produce cardholder's social status and Hidden Markov Model to extract financial profile from bank database.

Awoyemi et al. [1] compared Naïve Bayes, K-nearest neighbor, and Logistic regression models and concluded that K-NN performs well. Tran et al. [18] proposed two real-time data-driven approaches, one-class support vector machine (OCSVM) and T2 control chart which attained good accuracy and low false positives. Dal Pozzolo et al. [9] compared Random Forests (RF) with Neural Network (NN) and Support Vector Machine (SVM) where the Random Forests performed well as expected and suggested the accuracy can be improved by increasing the training data size. Carneiro et al. [4] applied 10-fold cross-validation to Random Forests, SVM, and Logistic regression and tested with balanced and unbalanced data then concluded that Random Forests attained the best performance.

Fu, Cheng, and Tu [10] captured the essential characteristics of frauds by using a model based on the convolutional neural network (CNN). Jurgovsky et al. [12] proposed LSTM (Long Short Term Memory) to aggregate the previous purchase pattern of the cardholder and to improve the accuracy of the incoming transaction, compared sequence learner LSTM and static learner (Random Forest) where LSTM is prone to overfitting even with few nodes, hence suggested increasing the size of data. In [15], 10-layer deep Variational Auto-Encoder (VAE) was applied and compared with Decision Tree, SVM, and Ensemble Classifier (AdaBoost algorithm) where AdaBoost achieved high Precision and recall for VAE. Pumsirirat and Yan [14] proposed Auto-Encoder and Restricted Boltzmann and confirmed supervised learning is appropriate for the historical transaction in credit card fraud detection.

3 Proposed approach description

In this Section, we propose our solution which is an approach combining the variational autoencoder(VAE) and the support vector data description(SVDD). Initially, the model will be trained on the VAE through which we will obtain residuals, which will then be used to train our SVDD model. The Fig. 1 shows the structure of our proposed approach.

3.1 Description of Variational Autoencoder

A variational autoencoder is an autoencoder whose encoding distribution is regularized during training to avoid over-fitting and to ensure that the latent space has good properties allowing us to generate new data. Statistically, the variational autoencoder is a technique based on Bayesian learning. Unlike a traditional

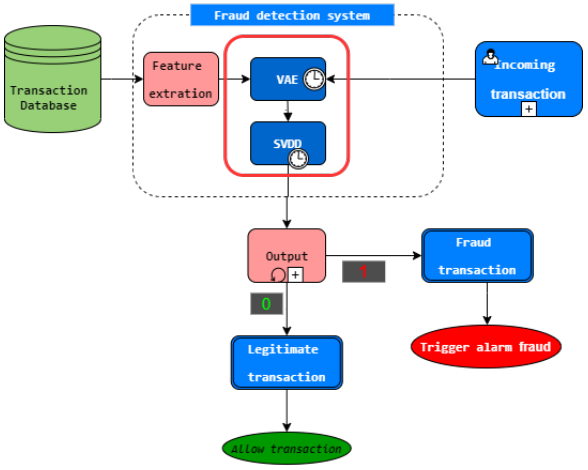

Fig. 1. Approach to the proposed solution

autoencoder, the VAE represents the input as a probability distribution with a mean and standard deviation rather than a set of numbers. We then sample from the latent distribution and get some numbers. We feed those numbers through decoding. We retrieve an example that looks like something from the original dataset, except that it was newly created by the model. The model is structured as follows: Firstly, the input is encoded as a distribution over the latent space. Secondly, a point in latent space is sampled from this distribution. In the third step, the sampled point is decoded and the reconstruction error can be calculated. finally, the reconstruction error is downgraded via the network. The structure of original VAE is shown in Fig. 2.

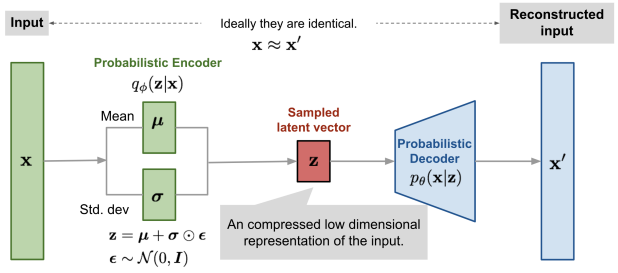

Fig. 2. Illustration of variationnel autoencoder.

3.2 Support Vector Data Description

Support vector data description (SVDD) is a technique related to One-class SVM [20]. Instead of finding a maximum margin hyperplane in feature space that best separates the mapped data from the origin as One-class SVM, the object of SVDD is to find the smallest hypersphere with center $\mathbf{c}$ and radius R that covers the normal instances in the training data-set. The SVDD is a useful method for outlier detection and has been applied to a variety of applications. Denote $\mathbf{x}_i \in \mathbb{R}^n, i = 1 \ldots, N$ as a set of training data. The SVDD is equivalent to solving the following *primal optimization*:

$$\begin{aligned} &\underset{R,\mathbf{a},\xi}{\text{Minimize}}\ R^2 + C\sum_{k=i}^{N}\xi_i \\ &\text{subject to } (x_i - a)^T(x_i - a) \leq R^2 + \xi_i\ \forall \xi_i \geq 0 \end{aligned} \tag{1}$$

the parameter $C > 0$ is used to control the influence of the slack variables ξ_i. After the optimization problem is solved, a hyper-spherical model is characterized by the center $\mathbf{a}$ and the radius R. By incorporating the constraint into (1), the optimization problem is solved using Lagrange's method:

$$L(R, a, \alpha, \xi_i) = R^2 + C\sum_i \xi_i - \sum_i \alpha\{R^2 + \xi_i - (x_i^2 - 2ax_i + a^2)\} - \sum_i \gamma_i\xi_i \tag{2}$$

With the Lagrange multiplier $\alpha_i \geq 0$ and $\gamma_i \geq 0$. By setting the partial derivatives to 0, new constraints are obtained:

$$\begin{aligned} &\sum_i \alpha_i, \\ a = \frac{\sum_i \alpha_i x_i}{\sum_i \alpha_i} &= \sum_i \alpha_i x_i, \\ &C - \alpha_i - \gamma_i = 0 \forall_i \end{aligned} \tag{3}$$

We can remove the variable γ_i from equation (3) since $\alpha_i \geq 0$, then let us use the constraints $0 \leq \alpha_i \leq C, \forall_i$. By rewriting (2) and replacing in (3) we have:

$$L = \sum_i \alpha_i < x_i, x_i > - \sum_{i,j} \alpha_i\alpha_j < x_i, x_j > \tag{4}$$

with constraints $0 \leq \alpha_i \leq C, \sum_i \alpha_i = 1$.

To determine if a z test point is in the sphere, the distance to the center of the sphere must be calculated. A test object z is accepted when the distance is less than the radius, that is, when $(z - a)^T(z - a) \leq R^2$. Expressing the center of the sphere in terms of support vectors, we accept the objects when:

$$< z, z > -2\sum_i \alpha_i < z, x_i > + \sum_{i,j} \alpha_i\alpha_j < x_i, x_j > \leq R^2. \tag{5}$$

Using kernel functions leads to a better compact representation of the training data. The SVDD formation with kernel functions leads to the flexible data description. The Gaussian kernel function used in this paper is defined as

$$K(x_i, x_j) = \exp(-||x_i - x_j||_2^2/s^2), \tag{6}$$

where s is Gaussian bandwidth parameter.

4 Implementation and results

4.1 Dataset

Our data comes from online e-commerce transactions of European credit cards provided in [8]. Data sets contain credit card transactions over a two-day collection period in September 2013 by European cardholders. There is a total of 284315 transactions without fraudulent transactions and 492 fraudulent transactions. The dataset contains numerical variables which are the result of a principal component analysis (PCA).

In the first part of the application of the variational autoencoder for the extraction of the characteristics we used 70% of the normal data for the training and 30% of the normal + the 492 of the fraudulent data for the test. In the second part of the application of SVDD we were forced to decrease the data due to a memory problem of our machine to facilitate the process of reading the data. For this, we have created 20 training sets each containing 10,000 data belonging to the normal data and 3000 for the test set. We then evaluate the model on each training data and then calculate the average.

4.2 Performance analysis

In binary classification, a machine learning model can make two types of errors when testing. It can either falsely predict sample data from the positive class as negative or sample data from the negative class as positive. The metric evaluation in the case of an unbalanced dataset requires taking into account the true positive rate and the false positive rate. The metrics we used for evaluation are: The AUC(Area Under Curve-Receiver) score, Recall, Precision and F1_score.

4.3 Results and interpretation

We compare our VAE + SVDD approach with AE + SVDD. We used the AUC and F1 measurement ratio to judge the accuracy of the method. We trained our SVDD model using the Gaussian bandwidth **s** parameter. We used the values from the following set: s = [0.001, 0.04, 0.1, 0.5]. The table 1 shows the results of the two approaches with different values of the Gaussian kernel **s** parameter. We notice an increase in the number of **supports vectors** as we increase the value of **s** and the noticeable change in performance of the VAE + SVDD and AE + SVDD algorithms. For each value of **s** given we notice that the performance of

the model in terms of the F1 score of VAE + SVDD is superior to AE + SVDD. We can say that this is due to the fact that in the VAE the data is smoother than the classic autoencoder (AE).

s	method	nSVs	Precision	Recall	F1	AUC
0.001	VAE+SVDD	8	99.96	97.07	93.07	87.74
	AE+SVDD	10	99	86.68	92.85	86.80
0.04	VAE+SVDD	81	99.36	96.69	97.93	95.68
	AE+SVDD	130	98.52	96.78	97.21	95.10
0.1	VAE+SVDD	118	98.80	96.75	97.77	95.71
	AE+SVDD	310	96.29	97.33	96.81	95.60
0.5	VAE+SVDD	739	91.34	97.15	94.16	94.84
	AE+SVDD	2813	64.14	98.94	80	96.56

Table 1. Comparison of accuracy (in %) with different value of s.

5 Conclusion

In this paper, we proposed a new approach for real time data-driven fraud detection using variational autoencoder and support vector data description. Thirst the model is trained with the variational autoencoder to learn the representation of normal transaction and then we this representation to train our SVDD. The advantage of this technique is that the representation learns by the VAE permit increasing the performance of the SVDD. The overall results in terms of value AUC, precision, F1 show good accuracy for our model. In the future, we will try to tune the value of the kernel bandwidth parameter s, since that the performance of the SVDD depends on this parameter.

References

1. Awoyemi, J.O., Adetunmbi, A.O., Oluwadare, S.A.: Credit card fraud detection using machine learning techniques: A comparative analysis. In: 2017 International Conference on Computing Networking and Informatics (ICCNI). pp. 1–9. IEEE (2017)
2. Bahnsen, A.C., Stojanovic, A., Aouada, D., Ottersten, B.: Cost sensitive credit card fraud detection using bayes minimum risk. In: 2013 12th international conference on machine learning and applications. vol. 1, pp. 333–338. IEEE (2013)
3. Batani, J.: An adaptive and real-time fraud detection algorithm in online transactions. International Journal of Computer Science and Business Informatics **17**(2), 1–12 (2017)
4. Carneiro, N., Figueira, G., Costa, M.: A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems **95**, 91–101 (2017)
5. Chaudhary, K., Yadav, J., Mallick, B.: A review of fraud detection techniques: Credit card. International Journal of Computer Applications **45**(1), 39–44 (2012)

6. Cheng, Y., Wu, D.D., Olson, D.L., Dolgui, A.: Financing the newsvendor with preferential credit: bank vs. manufacturer. International Journal of Production Research pp. 1–20 (2020)
7. Dai, Y., Yan, J., Tang, X., Zhao, H., Guo, M.: Online credit card fraud detection: A hybrid framework with big data technologies. In: 2016 IEEE Trustcom/BigDataSE/ISPA. pp. 1644–1651. IEEE (2016)
8. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G.: Credit card fraud detection and concept-drift adaptation with delayed supervised information. In: 2015 international joint conference on Neural networks (IJCNN). pp. 1–8. IEEE (2015)
9. Dal Pozzolo, A., Caelen, O., Le Borgne, Y.A., Waterschoot, S., Bontempi, G.: Learned lessons in credit card fraud detection from a practitioner perspective. Expert systems with applications **41**(10), 4915–4928 (2014)
10. Fu, K., Cheng, D., Tu, Y., Zhang, L.: Credit card fraud detection using convolutional neural networks. In: International Conference on Neural Information Processing. pp. 483–490. Springer (2016)
11. Hegazy, M., Madian, A., Ragaie, M.: Enhanced fraud miner: credit card fraud detection using clustering data mining techniques. Egyptian Computer Science Journal (ISSN: 1110–2586) **40**(03) (2016)
12. Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.E., He-Guelton, L., Caelen, O.: Sequence classification for credit-card fraud detection. Expert Systems with Applications **100**, 234–245 (2018)
13. Nguyen, Q.T., Tran, K.P., Castagliola, P., Huong, T.T., Nguyen, M.K., Lardjane, S.: Nested one-class support vector machines for network intrusion detection. In: 2018 IEEE Seventh International Conference on Communications and Electronics (ICCE). pp. 7–12. IEEE (2018)
14. Pumsirirat, A., Yan, L.: Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. International Journal of advanced computer science and applications **9**(1), 18–25 (2018)
15. Raza, M., Qayyum, U.: Classical and deep learning classifiers for anomaly detection. In: 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST). pp. 614–618. IEEE (2019)
16. Salazar, A., Safont, G., Soriano, A., Vergara, L.: Automatic credit card fraud detection based on non-linear signal processing. In: 2012 IEEE International Carnahan Conference on Security Technology (ICCST). pp. 207–212. IEEE (2012)
17. Tran, P.H., Rakitzis, A., Nguyen, H., Nguyen, Q.T., Tran, H., Tran, K.P., Heuchenne, C.: New methods for anomaly detection: Run rules multivariate coefficient of variation control charts. In: 2020 International Conference on Advanced Technologies for Communications (ATC). pp. 40–44. IEEE (2020)
18. Tran, P.H., Tran, K.P., Huong, T.T., Heuchenne, C., HienTran, P., Le, T.M.H.: Real time data-driven approaches for credit card fraud detection. In: Proceedings of the 2018 international conference on e-business and applications. pp. 6–9 (2018)
19. Truong, T.H., Ta, P.B., Nguyen, Q.T., Du Nguyen, H., Tran, K.P.: A data-driven approach for network intrusion detection and monitoring based on kernel null space. In: International Conference on Industrial Networks and Intelligent Systems. pp. 130–140. Springer (2019)
20. Vapnik, V.: The nature of statistical learning theory springer new york google scholar. New York (1995)