 CMC, vol.60, no.1, pp.179-192, 2019

A Scalable Approach for Fraud Detection in Online E-Commerce Transactions with Big Data Analytics

Hangjun Zhou[1, 2, *], Guang Sun[1, 3], Sha Fu[1], Wangdong Jiang[1] and Juan Xue[1]

Abstract: With the rapid development of mobile Internet and finance technology, online e-commerce transactions have been increasing and expanding very fast, which globally brings a lot of convenience and availability to our life, but meanwhile, chances of committing frauds also come in all shapes and sizes. Moreover, fraud detection in online e-commerce transactions is not totally the same to that in the existing areas due to the massive amounts of data generated in e-commerce, which makes the fraudulent transactions more covertly scattered with genuine transactions than before. In this article, a novel scalable and comprehensive approach for fraud detection in online e-commerce transactions is proposed with majorly four logical modules, which uses big data analytics and machine learning algorithms to parallelize the processing of the data from a Chinese e-commerce company. Groups of experimental results show that the approach is more accurate and efficient to detect frauds in online e-commerce transactions and scalable for big data processing to obtain real-time property.

Keywords: Big data analytics, machine learning, online e-commerce transactions, fraud detection, scalable processing.

1 Introduction

In the present digitalization era, the improvements in technology like mobile Internet, e-commerce, Internet finance, big data, etc., have brought up a revolution for the convenience and availability around the globe. Almost all activities are now capable of going online and e-commerce online transactions such as online shopping of goods or online trading various other products have increased to a great extent all of the world [Turban, Outland, King et al. (2017)]. To deal with the online e-commerce transactions, today most of people shopping or trading online possess an online account that usually binds with one or more credit cards or debit cards to widely make online purchases. Furthermore, in China almost everyone from young people to elderly ones has his/her own online e-commerce account on Alipay or Wechat to pay the money through mobile Internet, which is now becoming nearly the most popular mode of payment in China for

[1] Hunan University of Finance and Economy, Changsha, 410205, China.

[2] Nanjing University of Science and Technology, Nanjing, 210094, China.

[3] College of Engineering, The University of Alabama, Box 870200, Tuscaloosa, Alabama, USA.

[*] Corresponding Author: Hangjun Zhou. Email: zhjnudt@gmail.com.

almost all online e-commerce transactions and everyday offline transactions [Chen, Tao,Wang et al. (2015)]. On the other hand, these e-commerce applications handle many sensitive data like financial transaction records, personal information, behavior habits, locations, etc., which is very critical and important to be kept safe because it's also an attractive target for fraudsters to steal. For example, fraudsters can commit frauds by making an identity theft. If the frauds happen, it may cause huge financial loss and hurt the reputation of e-commerce sites. So although the use of the online e-commerce transactions has made life more convenient and productive, it has also opened a variety of threats at the same time. New fraud risk can emerge in all shapes and sizes and the chances of committing frauds have greatly increased with the progression of new technology. As a result, fraud detection in online e-commerce transactions has become an important issue to be concentrated on.

However, in terms of transaction time response, methods, behaviors, characteristics, etc., the fraud detection in online e-commerce transactions is not totally the same to that in the existing areas, such as credit card transaction. Particularly, the volume, variety, velocity and value of the massive amounts of data generated in online e-commerce transactions is big data level, which makes the fraudulent transactions more covertly scattered with genuine transactions than before. Moreover, the fraud detection in online e-commerce transactions is required to accomplish not only with high accuracy but also along with real-time responsiveness due to the characteristics of online e-commerce transactions. Therefore, this article proposed a novel scalable and comprehensive approach for fraud detection in online e-commerce transactions, which is designed and implemented majorly with four logical modules using big data analytics on Apache Spark and Hadoop to process data in parallel and detect fraud effectively and efficiently with real-time property.

The rest of the article is organized as follows. Literature of related works is described in Section 2. Section 3 introduces the preliminaries of relevant machine learning algorithms. A scalable approach for fraud detection in online e-commerce transactions with big data analytics is proposed in Section 4. In Section 5, groups of experiments are implemented to evaluate the efficiency of the proposed approach. Conclusions and future work are summarized in Section 6.

2 Related works

Fraud detection has been researched for a long time and almost any technological system that involves money and services, such as online e-commerce transaction, credit card, insurance, telecommunication, Internet marketing, etc., must supervise and implement anti-fraud measures to decrease the losses. There are so many machine learning models used for fraud detection in credit card area, which gives the references for fraud detection in online e-commerce transactions [Abdallah, Maarof and Zainal (2016)]. A neural network is firstly used to establish a fraud detection system on a large sample of labelled credit card account transactions and applied in Mellon Bank in 1994 [Ghosh and Reilly (1994)]. A decision tree and boolean logic method is proposed to differentiate the normal transactions and fraud transactions [Kokkinaki (1997)]. In paper Maes et al. [Maes, Tuyls, Vanschoenwinkel et al. (2002)], Bayesian belief networks and artificial neural networks have been also introduced for fraud detection. Based on real-life credit card data of

transactions, support vector machines and random forests are evaluated for credit card fraud detection [Bhattacharyya, Jha, Tharakunnel et al. (2011)]. Paper Halvaiee et al. [Halvaiee and Akbari (2014)] introduces a fraud detection model based on Artificial Immune Systems to increase the accuracy, reduce the cost and decrease system response time. In Mule et al. [Mule and Kulkarni (2014)], a credit card fraud detection mechanism using hidden Markov model is presented to work on human behavior while doing online shopping. With the use of big data technology, Convolutional Neural Network (CNN) is becoming prevailed in classification. A CNN-based fraud detection framework is designed and implemented to capture the intrinsic patterns of fraud behaviors learned from labeled data [Fu, Cheng, Tu et al. (2016)]. Paper Modi et al. [Modi and Dayma (2017)] provides comparative study of different techniques, such as decision tree, rule based mining, neural network, fuzzy clustering approach, hidden Markov model, etc., to detect frauds in credit card transactions.

With the increasingly huge amount of data generated from credit card system being too massive to be processed in a single computer, scalable machine learning algorithms based on big data techniques are proposed in recent years. In Hormozi et al. [Hormozi, Akbari, Hormozi et al. (2013)], the negative selection algorithm is implemented in parallel on the Apache Hadoop platform to decrease the training time of models for credit card fraud detection and evaluate the performances on real world financial data. A web service framework for credit card fraud detection in nearly real-time is described in Tselykh et al. [Tselykh and Petukhov (2015)]. Paper Phulari et al. [Phulari, Lamture, Madage et al. (2016)] introduces a big data architecture based on Flume, Hadoop and HDFS for fraud detection. The fraud risk management using real-time big data processing and intelligent risk models at Alibaba is introduced in Chen et al. [Chen, Tao, Wang et al. (2015)]. In Carcillo et al. [Carcillo, Pozzolo, Leborgne et al. (2018)], a scalable real-time credit card fraud finder with big data tools and machine learning approach is proposed. The experiment is conducted on a massive dataset of real credit card transactions with scalable and accurate results.

Based on the experiences of fraud detection in credit card transactions, some existing model algorithms are expanded to detect fraudulent online e-commerce transactions. A total order-based model for logical graph of behavior profiles for each user is proposed with an information entropy-based diversity coefficient and a state transition probability matrix to detect the fraud in online shopping transactions [Zheng, Liu, Yan et al. (2018)]. Wang et al. [Wang, Liu, Gao et al. (2017)] presents a deep-learning-based transaction fraud detection system CLUE deployed at JD.com to capture detailed information and models sequences on users' click actions with the recurrent neural network. In Shrivastava et al. [Shrivastava and Pateriya (2018)], a framework for fraud detection with machine learning techniques is described to make e-commerce sites more secure and efficient. However, the credit card transaction and online e-commerce transaction are not totally the same in terms of transaction time response, methods, behaviors, characteristics, etc. [Turban, Outland, King et al. (2017)]. Especially, the volume, variety, velocity and value of the massive amounts of data generated in online e-commerce transactions make the fraudulent transactions more covertly scattered with genuine transactions than before. Therefore, existing fraud detection techniques for online e-commerce transactions, usually without big data analytics on Spark and Hadoop, are not often sufficiently

capable and fast enough to detect these frauds accurately. Implementation of scalable, comprehensive, faster and efficient fraud detection systems with big data analytics in online e-commerce transactions has thus become crucial for all e-commerce issues to minimize the losses.

3 Machine learning algorithms

A variety of machine learning algorithms are applied in the fraud detection of online transaction in recent years. In this article, Convolutional Neural Network, Decision Tree and Support Vector Machine are mainly introduced and applied in our approach.

3.1 Convolutional neural network

Convolutional Neural Network (CNN) is one kind of deep machine learning algorithms based on artificial neural network. With the use of local connection and weight sharing, a CNN can maintain the deep structure of the neural network and meanwhile decrease the number of network parameters, which brings about good generalization ability, easy training advantage and better classification effect. As depicted in Fig. 1, a CNN network model is usually composed of input layer, convolution layer, pooling layer, dense (fully connected) layer and output layer.

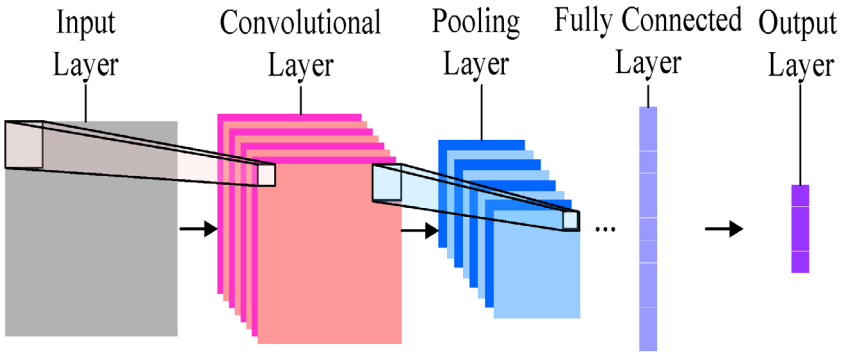

Figure 1: A CNN network model

In a CNN network model, convolutional layer is one of the core layers and always has the highest cost of calculation and complexity. However, by using local connectivity, each neuron node in one layer is merely connected to neuron nodes in the previous and adjacent layer so that the scale of parameters of CNN is reduced greatly. Moreover, the implementation of shared weights also decreases the amount of parameters of CNN model. The convolution of feature maps in previous layer is calculated through the convolution kernel and then the output feature map is computed and obtained with activation functions. The formula of convolutional layer is as follows:

$$x_j^l = f\left(\sum_{i \in M_j} x_i^{l-1} * k_{ij}^l + b_j^l\right)$$

where x_j^l represents the j-th feature graph in l-th convolution layer, the symbol "*"

represents the operation of convolution, M_j represents the set of all the input feature maps, k_{ij}^l represents the weight of the convolution kernel *j* in the layer *l*, b_j^l represents the bias, f represents the activation function.

Pooling layer usually has two kind of functions: Max Pooling and Average Pooling, which are the methods to perform down-sampling operation and decrease the complexity. The formula of pooling operation is:

$$s_j = \frac{1}{|R_j|}\sum_{i \in R_j} a_i$$

where s_j represents the pooling value of *j*-th pooling region R_j, a_i represents the value in the pooling region outputted from activation function of *i*-th activation value.

Given an input or set of inputs, the activation function is used to define the output of a neuron and improve the nonlinear processing ability of a CNN network. Practically, there are more than a dozen kinds of activation functions and the typical ones are Sigmod, ReLU, TanH and so forth.

3.2 Decision tree

Based on the probability of occurred events, a decision tree learning method is to predict and select the relative optimum solution by comparing the solutions to be evaluated with probability calculation and tree-like graph. Generally, decision tree learning algorithms, such as ID3 or C4.5, use Entropy to measure Information Gain, and in some cases prune the tree based on Entropy to obtain better classification results. The Formula of the Entropy is as follows:

$$H(x) = -\sum_{i=1}^{n} p(x_i) log_2 p(x_i)$$

where *x* represents a random variable that takes on values from the set $\{x_1, x_2, \ldots, x_i, \ldots, x_n\}$, p_i represents the probability of class *i*, and $\sum_{i=1}^{n} p_i = 1$.

3.3 Support vector machine

In machine learning field, a Support Vector Machine (SVM) performs classification that analyzes data by finding the hyperplane that maximizes the margin between the two classes. For datasets that are linear separable, perhaps with a few exceptions or some noise, a SVM has capability to achieve good classification by finding optimal hyperplane. Suppose the linear discriminant function is $f(x) = w^T x + b$, the SVM formula is as follows:

$$\begin{cases} \min \frac{1}{2}\|w\|^2 \\ y_i[w^T x_i + b] - 1 \geq 0, i = 1,2,\cdots,n \end{cases}$$

where *x* represents to a training or test pattern, w^T represents the weight vector, *b* represents the value of the bias term, *n* represents the number of samples, $y_i[w^T x_i + b] - 1 \geq 0$ represents the constraint condition.

For the nonlinear datasets, a SVM can map the original feature space to some higher-dimensional feature space to solve the classification problem. With the use of SVM kernel functions, nonlinear classification calculation can be implemented more simply and efficiently. The nonlinear classifier formula is as follows:

$$f(x) = sign\left(\sum_{i=1}^{n} a_i y_i K(x_i, x) + b\right)$$

where a_i represents the Lagrange multiplier, y_i represents the output, $K(x_i, x)$ is the kernel function.

4 A scalable approach for fraud detection in online e-commerce transactions

In order to enhance the capability to handle the increasingly complicated online transaction fraud cases, a novel scalable approach for fraud detection in online e-commerce transactions using big data analytics effectively and efficiently has been proposed in this section. This approach uses Apache Spark on Yarn as the infrastructure to distributedly implement the machine learning algorithms with big data so as to improve the efficiency of fraud detection. As we can see in Fig. 2, at first the Hadoop HDFS is initiated on the cluster of data nodes where the dataset is distributedly stored. Then Spark environment is created and client node uses SparkContext to transform the processing request into Directed Acyclic Graph (DAG) in driver program. The DAG is analyzed into stage tasks and sent to the Resource Manager that has initiated a Node Manager on each Spark worker node. Each Node Manager receives one or several computing tasks and initiates Executor containers to run the tasks, so that the whole data processing can be implemented in parallel on Spark cluster and the general run-time is reduced to obtain real-time property.

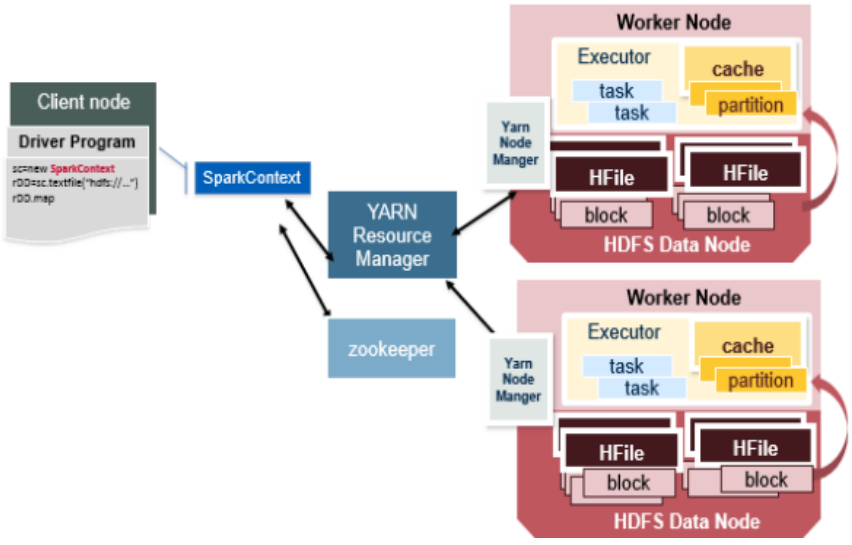

Figure 2: Apache Spark architecture

Based on the Apache Spark infrastructure, the proposed approach is designed to provide

real-time anti-fraud measures for online e-commerce transaction and be easy to scale out. It majorly includes four logical modules: transaction data preprocessing, expert systems with knowledge bases and rules, fraud detection with machine learning algorithms and deviant behavior pattern discovery. The framework of the approach is out-lined in Fig. 3. The big data of online e-commerce transactions, such as trading data and payment data, is firstly reprocessed in transaction data preprocessing module and transmitted to rule-based expert system module, which is realized with Spark streaming and distributed platform Apache Kafka, to detect whether the online transaction behaviors are abnormal. Then, the normal data out of expert system module is used to run behavioral analysis with CNN, Decision Tree and SVM algorithms in machine learning fraud detection module. Based on the result of machine learning classification, comprehensive voting strategy is applied to obtain the outcome of fraud behavior detection. Furthermore, the normal data examined from the previous modules is analyzed in the deviant behavior discovery module with clustering algorithms in case of new fraudulent pattern of online transactions.

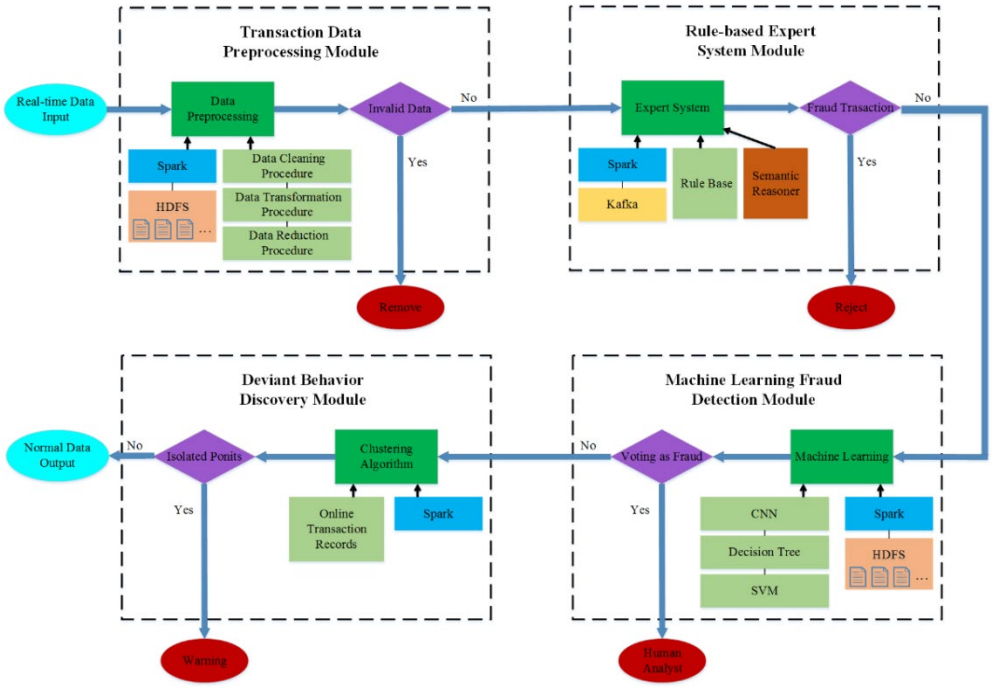

Figure 3: Framework of the approach

4.1 Transaction data preprocessing module

The raw data of online e-commerce transactions is not suitable for the data mining in big data set so that it has to be preprocessed at first. The transaction data preprocessing is implemented on Spark and HDFS framework. Data cleaning procedure smooth the noise data, corrects the wrong data, cleans the duplicated data and unifies the data format. Data transformation procedure transforms the online transaction data into the forms suitable for big data mining through data generalization, data smoothing and aggregation, data normalization. Data reduction procedure obtains the reduced data representation similar

to the original data set with much smaller volume and maintaining the data integrity for data analysis.

4.2 Rule-based expert system module

The preprocessed valid data is transmitted to expert system that majorly consists of real-time streaming pipeline, rule base and semantic reasoner. For the very large volume of big data set of online e-commerce transactions, the design of real-time streaming pipeline with Spark and Kafka technique is to ensure that the online transaction data set can be handled with high throughput and low latency in millisecond level, so as to meet the real-time processing requirement of online e-commerce transaction. Rule base is a set of rules that come from real fraud cases and are defined by relevant industry experts. These rules extract the behavioral features of online transaction and reject the transaction if it satisfies the fraud rules. After the rule base is established, there may be emerging cases of new fraud behaviors, thus the rule base can be effectively updated with new rules corresponding to the new features of fraud behaviors, and meanwhile without the affection to the regular running of the whole system. Semantic reasoner is used to infer logical consequences based on the interaction of online transaction data and the rule base.

4.3 Machine learning fraud detection module

The normal data from expert system is stored in HDFS and read in memory for Spark processing. Before the machine learning model can be used to detect the fraudulent online transactions, the features such as customer age, profession, gender, transaction records, transaction amounts, credits, etc., are extracted and stored in vectors by Spark as the input for the model training of CNN, Decision Tree and SVM. The model training is running based on Map-Reduce framework. In Map stage, the Spark mapper function reads each online transaction record in parallel from sample data as supervised input, extracts sample features and fits the model correlation coefficient according to the requirements of CNN, Decision Tree and SVM, and output the training result in key-value structure. In Reduce stage, Spark reducer function distributedly collects the output datasets of Map stage and merges the records with the same key. Moreover, the reducer program computes the correlation coefficients of the different models and stores them into HDFS for the real-time fraud detection of online e-commerce transactions.

When the data input is from the test dataset or the running-time data streaming of online transactions, Spark respectively uses trained models of CNN, Decision Tree and SVM to compute the results of fraud detection, which are further applied to archive the classification outcome through voting strategy. The formula of voting strategy is as follows:

$$r = \begin{cases} 1, & if\ r_{cnn} + r_{dt} + r_{svm} \geq 2 \\ 0, & if\ r_{cnn} + r_{dt} + r_{svm} < 2 \end{cases}$$

where r represents the voting result, r_{cnn}, r_{dt} and r_{svm} respectively represents the classification result of fraud detection corresponding to each model. If the classification result is a fraudulent online transaction, the value will be 1, otherwise it is 0.

With the help of machine learning algorithms on Spark and HDFS, the fraud detection of online e-commerce transactions can be implemented in parallel so as to archive the real-time processing efficiency and offer the support to a human analyst.

4.4 Deviant behavior discovery module

To the massive data transmitted from the previous modules, a distributed K-Means clustering algorithm is implemented on Spark framework to analyze and discover deviant behaviors that maybe are new and undiscovered fraudulent online transactions. The idea is features are extracted based on the records, such as consuming times, amounts, locations, behaviors in different periods and personal information, etc., and the distributed K-Means clustering algorithm is used to analyze and cluster the online e-commerce transaction customers or their records. If there are obvious isolated points in the clustering result, they will be marked as potential deviant transactions and warnings will be given to human analysts to supervise these points with special concentration. In addition, the big data techniques of Spark can accelerate the velocity of distributed clustering algorithm and improve the clustering efficiency on massive data.

5 Experiments and evaluations

Groups of experiments are implemented on a cluster consisting of 12 machines, each with 8 cores and 32 GB of RAM. The operating system is CentOS 7 with Java Development Kit 10.0.2. Apache Spark stable release 2.3.1 is run on top of the cluster resource negotiator Hadoop Yarn and HDFS.

In this article, the experiment results are evaluated in terms of Recall, Precision, Accuracy, and F1-Score, which are defined in Tab. 1. The true positive (TP) means the number of correctly predicted fraud transactions among all the true fraud transactions, the false positive (FP) means the number of normal transactions which are incorrectly predicted as fraud transactions, the true negative (TN) means the number of correctly predicted normal transactions among all the true normal transactions, and the false negative (FN) means the number of fraud transactions which are incorrectly predicted as normal transactions. The Precision is the synonym for the positive predictive value. The Recall is synonym for the true positive rate. The Accuracy indicates the general fraud detection performance. F1-Score is the harmonic mean of precision and recall.

Table 1: Performance metrics

Performance metrics	Formulas
Precision	$\frac{TP}{TP + FP}$
Recall	$\frac{TP}{TP + FN}$
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$
F1-Score	$2 \times \frac{Precision \times Recall}{Precision + Recall}$

To evaluate the model of proposed approach, the original experimental dataset is obtained from B2C online transaction records of a Chinese e-commerce company. It

contains over 6 million records of online e-commerce transactions with a time span of 5 months and each transaction record has 41 dimensions. The ratio of legitimate transactions over fraud transactions is approximately 200:1. To relieve the problem of the imbalanced dataset, different balanced experimental sample datasets of training data and test data are constructed based on the original dataset. Before the experiments, the routine data preprocessing of data cleaning, data transformation and data reduction has been implemented and feature engineering is realized by an automated method to generate features. For the reason to maintain data confidentiality of sensitive information, not all the data dimensions are mentioned.

In the groups of experiments, the proposed approach is compared with traditional CNN, Decision Tree and SVM models respectively on 5 different balanced experimental sample datasets, and each comparative test is performed on the same dataset. The experimental results of precision comparison on different sample datasets are demonstrated in Fig. 4. For all the models, the precision test results of fraud detection for online e-commerce transactions on 5 datasets are above 60% at least. The average precision rate of the proposed approach is over 92% and the highest precision result is close to 95% on dataset 1. Although the precision rate of the proposed approach is a little lower than 90% on dataset 3, it still performs better than those of other 3 models. On 5 sample datasets, the precision results of fraud detection of CNN, Decision Tree and SVM are lower than that of the proposed approach because more false positives are classified as fraudulent e-commerce online transactions by them.

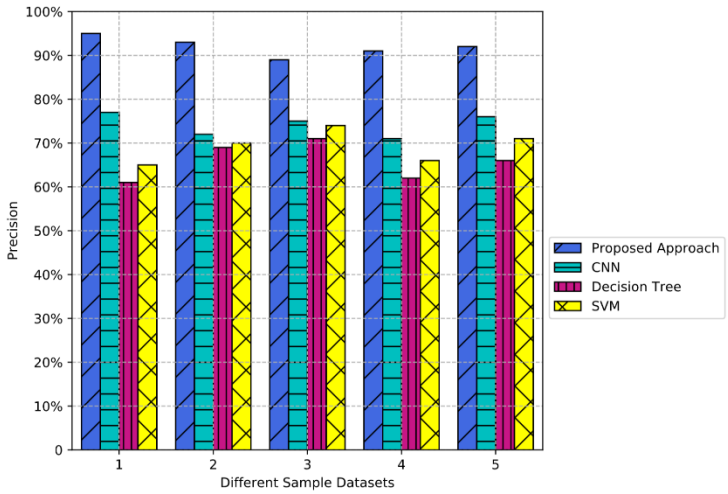

Figure 4: Precision results on different sample datasets

In Fig. 5, the comparative results of recall tests are described. Generally, the recall results of 4 models are above 57% at least and the proposed approach and CNN are better than the other two with about 20% higher on average. The performance of CNN is close to that of the proposed approach on 5 datasets with its average recall rate nearly reaching 90%. The recall rate of the proposed approach is stabilized over 90% and the highest rate is almost 94%.

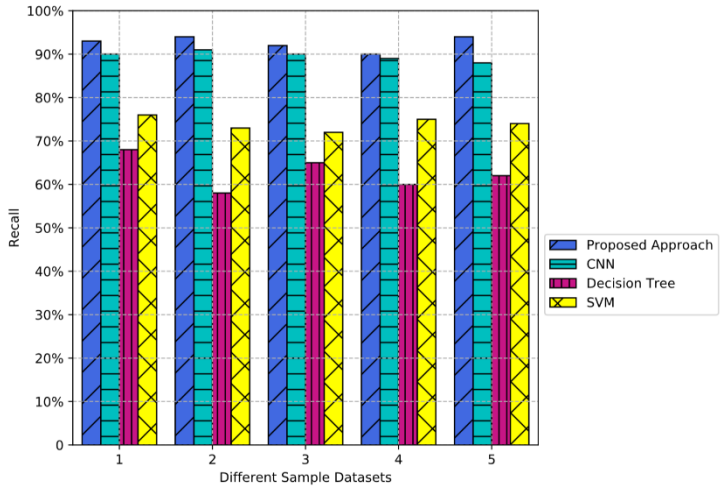

Figure 5: Recall results on different sample datasets

The Accuracy rate shows the overall fraud detection performance, which is depicted in Fig. 6. As can be seen, the accuracy rates of all models are at least above 67% and the accuracy of the proposed approach is about 96%. The CNN also has the close accuracy rates to those of the proposed approach on 5 datasets, however, in front of big data sets, the proposed approach is faster than CNN on training and testing.

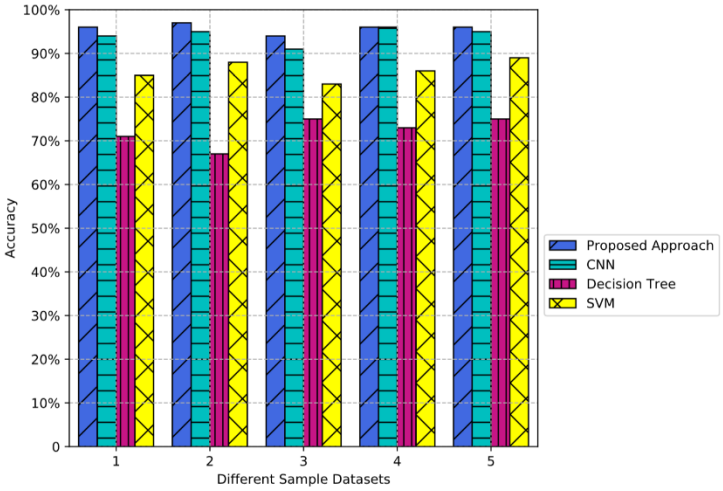

Figure 6: Accuracy results on different sample datasets

F1-Score conveys the balance between the precision and the recall by a weighted average way. As shown in Fig. 7, the F1-Scores of each model on 5 sample datasets are not lower than 56%. In general, the F1-Scores of the proposed approach are around 90% and the average value of them is almost 92% with the highest F1-Score reaching 95%, which is better than the other models in performance.

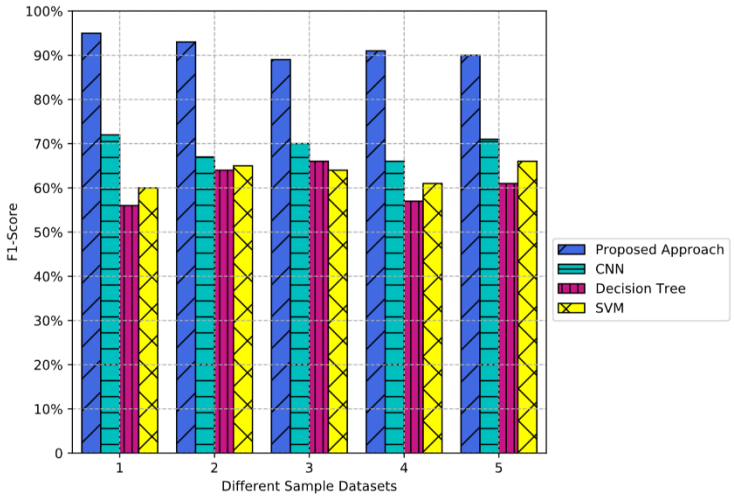

Figure 7: F1-Score results on different sample datasets

6 Conclusions and future work

With the rapid development of wired and mobile payment applications on the Internet, an increasing amount of attention is currently paid to the big data analytics for the fraud detection of online e-commerce transactions. In this article, a scalable approach is proposed to enhance the efficiency of anti-fraud measures, which uses Apache Spark on Yarn as the infrastructure to process the big data sets of a Chinese e-commerce company. It majorly includes four logical modules: transaction data preprocessing module, rule-based expert system module, machine learning fraud detection module and deviant behavior discovery module. The groups of experimental results show that the proposed approach performs better with a higher level of detection precision rate, recall rate, accuracy and F1-score than the existing relative models. Moreover, while the approach improves the effectiveness of fraud detection, it implements the big data techniques of Spark to speed up the data processing in parallel so as to meet the real-time requirement of online e-commerce transactions. In future work, the focus will be concentrated on the application of the proposed approach in the industrial big data environment, such as Cloud computing environment in a data center, to verify the efficiency and robustness of fraud detection for online e-commerce transactions.

Acknowledgements: This research work is supported by Hunan Provincial Education Science 13th Five-Year Plan (Grant No. XJK016BXX001), Social Science Foundation of Hunan Province (Grant No. 17YBA049), Hunan Provincial Natural Science Foundation of China (Grant No. 2017JJ2016), 2017 Hunan Provincial Higher Education Teaching Re-form Research Project (Grant No. 564). The work is also supported by Open foundation for University Innovation Platform from Hunan Province, China (Grand No. 16K013) and the 2011 Collaborative Innovation Center of Big Data for Financial and Economical Asset Development and Utility in Universities of Hunan Province. We also thank the anonymous reviewers for their valuable comments and insightful suggestions.

References

Abdallah, A.; Maarof, M. A.; Zainal, A. (2016): Fraud detection system: a survey. *Journal of Network and Computer Applications*, vol. 68, no. 6, pp. 90-113.

Bhattacharyya, S.; Jha, S.; Tharakunnel, K; Westland, J. C. (2011): Data mining for credit card fraud: a comparative study. *Decision Support Systems*, vol. 50, no. 3, pp. 602-613.

Carcillo, F.; Pozzolo, A. D.; Leborgne, Y. A.; Caelen, O.; Mazzer, Y. (2018): SCARFF: a scalable framework for streaming credit card fraud detection with spark. *Information Fusion*, vol. 41, no. 5, pp. 182-194.

Chen, J.; Tao, Y.; Wang, H.; Chen, T. (2015): Big data based fraud risk management at Alibaba. *The Journal of Finance and Data Science*, vol. 1, no. 1, pp. 1-10.

Fu, K.; Cheng, D.; Tu, Y.; Zhang, L. (2016): Credit card fraud detection using convolutional neural networks. *International Conference on Neural Information Processing*, pp. 483-490.

Ghosh, S; Reilly, D. L. (1994): Credit card fraud detection with a neural-network. In *IEEE System Sciences Proceedings of the Twenty-Seventh Hawaii International Conference*, pp. 621-630.

Halvaiee, N. S.; Akbari, M. K. (2014): A novel model for credit card fraud detection using artificial immune systems. *Applied Soft Computing*, vol. 24, no. 11, pp. 40-49.

Hormozi, H.; Akbari, M. K.; Hormozi, E.; Javan, M. S. (2013): Credit cards fraud detection by negative selection algorithm on hadoop (to reduce the training time). *5th Conference Proceedings of Information and Knowledge Technology*, pp. 40-43.

Kokkinaki, A. I. (1997): On a typical database transactions: identification of probable frauds using machine learning for user profiling. *IEEE Proceedings of Knowledge and Data Engineering Exchange Workshop*, pp. 107-113.

Maes, S.; Tuyls, K.; Vanschoenwinkel, B.; Manderick, B. (2002): Credit card fraud detection using bayesian and neural networks. *Proceedings of the 1st International Naiso Congress on Neuro Fuzzy Technologies*, pp. 261-270.

Modi, K.; Dayma, R. (2017): Review on fraud detection methods in credit card transactions. *International Conference Proceedings on Intelligent Computing and Control*, pp. 1-5.

Mule, K.; Kulkarni, M. (2014): Credit card fraud detection using hidden markov model. *International Journal of Innovative Technology & Adaptive Management*, vol. 1, no. 11, pp. 13-17.

Phulari, S.; Lamture, U. S.; Madage, S. V.; Bhandari, T. K. (2016): Pattern analysis and fraud detection using hadoop framework. *International Journal of Engineering Science and Innovative Technology*, vol. 5, no. 1, pp. 92-100.

Shrivastava, S.; Pateriya, R. K. (2018): Secure framework for e-commerce applications in cloud environment. *Proceedings of Improving E-Commerce Web Applications Through Business Intelligence Techniques*, pp. 82-109.

Tselykh, A.; Petukhov, D. (2015): Web service for detecting credit card fraud in near real-time. *Proceedings of the 8th International Conference on Security of Information*

and Networks, pp. 114-117.

Turban, E.; Outland, J.; King, D.; Lee, J. K.; Liang, T. P. (2017): *Electronic Commerce 2018: A Managerial and Social Networks Perspective*. Springer.

Wang, S.; Liu, C.; Gao, X.; Qu, H.; Xu, W. (2017): Session-based fraud detection in online e-commerce transactions using recurrent neural networks. *Joint European Conference Proceedings of Machine Learning and Knowledge Discovery in Databases*, pp. 241-252.

Zheng, L.; Liu, G.; Yan, C.; Jiang, C. (2018): Transaction fraud detection based on total order relation and behavior diversity. *IEEE Transactions on Computational Social Systems*, vol. 99, no. 8, pp. 1-11.

.