Information Sciences 557 (2021) 302–316

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture

Xinwei Zhang [a,b], Yaoci Han [a], Wei Xu [a,*], Qili Wang [a]

[a] School of Information, Renmin University of China, Beijing, 100872, China
[b] Department of Statistics, Rutgers University - New Brunswick, NJ, 08904, United States

ARTICLE INFO

Article history:
Received 19 December 2017
Received in revised form 31 March 2019
Accepted 10 May 2019
Available online 16 May 2019

Keywords:
Credit card fraud
Deep learning
Feature engineering
Fraud detection
Behavior analysis

ABSTRACT

Credit card transaction fraud costs billions of dollars to card issuers every year. A well-developed fraud detection system with a state-of-the-art fraud detection model is regarded as essential to reducing fraud losses. The main contribution of our work is the development of a fraud detection system that employs a deep learning architecture together with an advanced feature engineering process based on homogeneity-oriented behavior analysis (HOBA). Based on a real-life dataset from one of the largest commercial banks in China, we conduct a comparative study to assess the effectiveness of the proposed framework. The experimental results illustrate that our proposed methodology is an effective and feasible mechanism for credit card fraud detection. From a practical perspective, our proposed method can identify relatively more fraudulent transactions than the benchmark methods under an acceptable false positive rate. The managerial implication of our work is that credit card issuers can apply the proposed methodology to efficiently identify fraudulent transactions to protect customers' interests and reduce fraud losses and regulatory costs.

1. Introduction

In recent years, the volume of credit card transactions has been dramatically increasing due to the popularization of credit cards and the rapid development of e-services, including e-commerce, e-finance and mobile payments. The large-scale utilization of credit cards and the various transaction scenarios without rigid verification and supervision inevitably lead to billion-dollar losses caused by credit card fraud [8]. Although an accurate estimate of the loss is difficult to obtain since card issuers are often reluctant to release the statistics [5], there is some publicly available data that show the severity of credit card fraud. According to the Nilson Report (2015), the losses due to credit and debit card fraud reached $16.31 billion in 2014, up 19%. Cybersource (2013) reported that online fraud caused a $3.5 billion dollar estimated loss in 2012 [28].

Credit card fraud can be categorized into two types: application fraud [32] and behavior fraud [5]. Application fraud refers to situations in which an application for a credit card is fraudulent. It occurs when a fraudster applies for a new credit card using counterfeit identity information and the card issuer approves this application. Behavior fraud occurs after a credit card is approved and issued. It refers to credit card transactions that reflect fraudulent behavior. Fraud detection and fraud prevention have always been important concerns for card issuers and important research topics for researchers since detecting and preventing even a small portion of fraudulent activity would save millions of dollars. In this paper, our focus is on the

* Corresponding author.
E-mail address: weixu@ruc.edu.cn (W. Xu).

https://doi.org/10.1016/j.ins.2019.05.023
0020-0255/

problem of fraudulent behavior detection. We are mainly concerned with how to identify fraudulent transactions as soon as possible after they occur. In the rest of this paper fraud detection refers to detecting fraudulent behavior unless otherwise noted.

Due to the large volume of credit card transactions, manually verifying every transaction to detect fraud is infeasible for credit card issuers. Therefore, machine/statistical learning methods are commonly used to detect fraudulent transactions automatically. There are several good reviews regarding machine learning methods in fraud detection [1,5]. For supervised learning, previous research works mainly investigated statistical models with shallow architectures, for example, neural networks with a single hidden layer, support vector machines, hidden Markov chains and logistic regression. Here, a model with a shallow architecture refers to a model that contains one single layer of nonlinear transformation. The nonlinear transformation maps the raw input data from its original space into a feature space. As natural extensions to shallow architecture models, deep architecture models, known as deep learning, employ multiple layers of nonlinear transformation and stack these nonlinear transformation layers in a hierarchical structure where the output of a lower layer is treated as the input to the next layer. In recent years, deep architecture models have been reported to outperform traditional machine learning methods for pattern recognition and representation learning in several fields, including image and speech coding, image and speech recognition and information retrieval [2,17,22,25,26,36]. Therefore, we also wondered whether deep learning could be successfully applied to fraud detection.

A fraud detection model can be regarded as a predictive model of transactional behavior in which past transaction behaviors are used to predict the legitimacy of current transaction behavior. Feature engineering in fraud detection is expected to construct feature variables that properly summarize and represent the transaction behavior information from raw transaction records. Good feature engineering is the foundation of an effective fraud detection model and hence important to the success of a fraud detection system. Van Vlasselaer [39] noted that previous feature engineering frameworks in fraud detection research fit the recency, frequency, monetary (RFM) framework. The RFM framework is a marketing technique used to analyze the past behavior of consumers [20]. Recency concerns how recently a customer has made purchases and frequency indicates how often a customer makes purchases, while monetary value indicates how much a customer spends. However, we believe that the RFM framework overlooked the intrinsic heterogeneity in credit card transactions and needs to be improved since different types of transactions entail intrinsically different behaviors. For example, a cash withdrawal is a different transaction behavior from a dining payment. The former is usually riskier since fraudsters are more motivated to withdraw cash from a stolen card.

In this paper, we propose a novel feature engineering framework with deep learning models for developing a credit card fraud detection system. A feature engineering framework based on homogeneity-oriented behavior analysis (HOBA) is proposed to generate feature variables for the fraud detection model. Then, deep learning techniques are incorporated into the fraud detection system to deliver good fraud detection performance. We conduct empirical experiments on a real-world dataset from one of the largest commercial banks in China to assess the effectiveness of the framework. The experimental results illustrate the efficacy of the proposed methodology as a feasible mechanism for credit card fraud detection.

The main contributions of our study are threefold. First, we propose a homogeneity-oriented behavior analysis (HOBA), a novel behavior analysis framework for credit card transactions. A feature engineering framework based on the HOBA is employed to provide better feature variables for predicting the fraudulent behavior of credit card transactions compared to the recency-frequency-monetary (RFM) framework. Second, we employ deep learning techniques to more effectively model transaction behaviors and achieve better performance in fraud detection. We conduct comparative experiments to compare the performance of several deep learning algorithms with that of traditional machine learning methods using different feature sets. Third, we evaluate the credit card fraud detection model from a practical perspective and investigate the model performance under certain false positive rates. To the best of our knowledge, this is the first successful study of applying HOBA to analyze the past behaviors of card holders to detect fraudulent credit card transactions. The managerial implication of our work is that credit card issuers can apply the proposed methodology to efficiently detect fraudulent transactions among the massive amounts of transactions to protect the interests of customers and reduce fraud losses and regulatory costs.

The remainder of the paper is organized as follows. The second section provides a literature review of credit card fraud and deep learning. The third section presents our proposed method, and in the fourth section, the results of an empirical analysis are reported based on real-life datasets. Finally, the conclusions and future work are summarized in the last section.

2. Literature review

2.1. Credit card fraud

As we have described, credit card fraud is divided into two types. One is application fraud, and the other is behavior fraud. Behavior fraud mainly includes the theft/stolen-card fraud, counterfeit-card fraud and card-not-present fraud. Theft/stolen-card fraud is the most common type of fraud. Typically, fraudsters will spend as much as possible after stealing a card or obtaining a lost card. The time intervals between the transactions are typically short when fraudsters perpetrate theft/stolen card fraud. Similarly, counterfeit-card fraud occurs when fraudsters steal the credit card information and make a fake one. The victims still have their cards and make legitimate transactions, and the fraudsters use the fake cards to perpetrate fraud-

ulent transactions. Fake cards are often used only a few times and are abandoned before the victims realize the misappropriation of their card information. The third type of behavior fraud is card-not-present fraud, which occurs when transactions are made remotely. In this case, only the card information, such as card number, holder name and expiration date, is used to make a transaction. The difference between counterfeit-card fraud and card-not-present fraud is that fraudsters will use a physical card rather than only the card information in the former case.

Although the annually growing amount of credit card fraud losses raises the demand for state-of-the-art fraud detection techniques, the difficulty in obtaining credit card transaction datasets is a high barrier for researchers to overcome in attempting to develop innovative solutions in the fraud detection field [21]. There are several good discussions about the main issues, techniques and challenges in the fraud detection field [1,5]. In academia, researchers mainly focus on two aspects of fraud detection: statistical modeling methods and feature engineering methods.

From the perspective of statistical methods, both supervised and unsupervised learning methods are considered to construct an effective fraud detection system. For supervised learning methods, a historical two-class (fraudulent/legitimate) labeled transaction dataset is used to construct a classification model. Researchers have studied the performance of numerous techniques for credit card fraud detection, such as artificial neural networks (ANN) [3,6,16], Bayesian belief networks [31], decision trees [35], random forest [4,9] and hidden Markov models [38]. Unsupervised methods do not require a labeled dataset. Instead, they group customer characteristics (or transaction characteristics) into different clusters to find unusual behaviors. When an incoming behavior shows a strong departure from the normal behavior groups, it is suspected to be a fraud. Unsupervised methods in fraud detection include peer group analysis [5] and self-organizing maps [33,43]. Whether supervised or unsupervised, a statistical model for fraud detection will produce a suspicion score (or fraudulent likelihood value) for a transaction or an account. When implemented in a fraud detection system, a cutoff threshold is set to determine whether a transaction is approved, rejected or investigated. Setting reasonable cutoff thresholds is a practical challenge.

There are some challenging issues for supervised learning and unsupervised learning in fraud detection. The first problem for supervised learning is that it relies on an accurate labeled dataset. However, in practice, some historical transactions may be mislabeled. Second, a credit card transaction dataset is extremely unbalanced, where only less than 0.1% of the transactions are fraudulent. The unbalanced dataset also poses a big challenge for supervised learning. Either a good sampling procedure or an adjustment to the cost function is needed. The unsupervised methods produce higher false alarm rates because unusual behavior is often found to correspond to a legitimate transaction [24]. In practice, a false alarm rate usually represents the cost of capturing fraudulent behavior, and a lower false alarm rate means less interference to normal accounts. Obviously, a model with a lower false alarm rate is preferred for fraud detection system users, so most studies in the fraud detection field focus on supervised learning methods.

Fraud detection modeling usually involves a domain-based feature engineering process. Good feature variables will increase the performance of machine learning methods. In the fraud detection field, directly using all of the historical transaction records of each account is impractical because the number of transactions of each account is very high and is different. However, a transaction-level classification that only uses the information of a single transaction effectively causes an information loss of the historical transactions for the associated account. Thus, transaction aggregation strategies are proposed to generate features to ameliorate the weakness of transaction-level classification [41]. Although the choice of the period of transaction aggregation is still a problem [21], the concept of a transaction aggregation strategy provides a reliable way to derive the input feature variables for fraud detection.

2.2. Deep learning

Deep learning has recently been a highly focused area in the field of machine learning [42]. Deep learning methods, including convolutional networks, deep belief networks and deep autoencoders, are a class of hierarchical learning architectures in which there are many layers of information processing for representation learning or pattern classification [27]. The origin of the deep learning concept can be traced to studies of artificial neural networks. The typical method of training the weights of neural networks is the back-propagation algorithm [18]. However, the effectiveness of the back-propagation algorithm decreases dramatically when the depth of the neural networks increases, which may be caused by problems such as poor local optima and error dilution. The success of deep architectures should be attributed to the key idea of unsupervised layer-by layer greedy learning, which empirically eliminates the optimization difficulty in the training parameters of deep architecture. The training procedure of the deep belief network can be regarded as the first successful case of effective deep architecture training [19].

Apart from image and audio areas, recent research has shown that deep learning techniques can also be employed in the financial field and achieve better predictive results. Fischer and Krauss [13] used long short-term memory networks to predict time series and indicated that deep learning has great potential in financial market prediction. Kraus and Feuerriegel [23] applied deep learning techniques and transfer learning to analyze financial disclosures and use the predictive results to facilitate financial decision-making. Moreover, a few pioneering studies have tried to introduce deep learning in financial fraud detection. For example, Fiore et al. [12] proposed a strategy to generate synthetic examples based on generative adversarial networks to improve the classification performance in credit card fraud detection by dealing with the problem of class imbalance. These remarkable studies have inspired researchers to incorporate deep learning techniques to improve prediction accuracy.

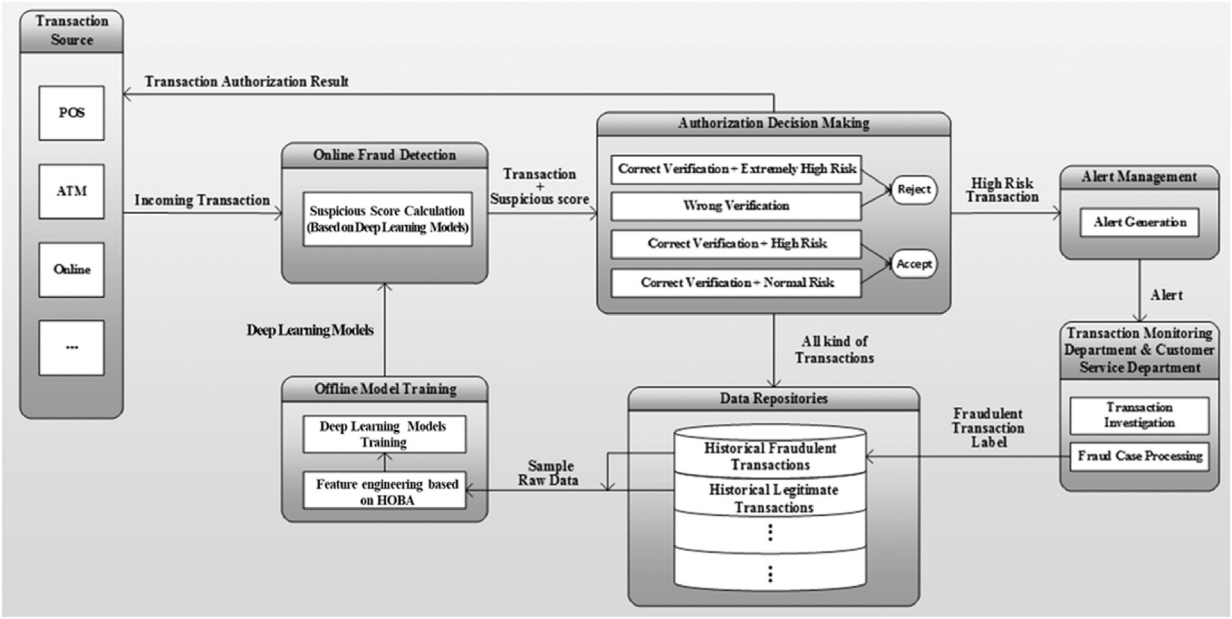

Fig. 1. The fraud detection process.

3. The proposed framework

3.1. Overview

In this section, we describe our proposed feature engineering framework with a deep learning architecture for credit card fraud detection. The fraud detection process of our framework is shown in Fig. 1.

As can be seen from Fig. 1, we briefly introduce the fraud detection process here. Suppose a transaction comes to the credit card issuer as an authorization request. First, the transaction is input into the Online Fraud Detection module, and the corresponding suspicion score is calculated based on the historical transaction information of the associated account. The transaction and its suspicion score are input into the Authorization Decision Making module, and the authorization decision is made based on the acceptance verification result and level of fraud risk. The acceptance verification mainly concerns whether the open-to-buy amount of the associated account is sufficient for the transaction, and sometimes it verifies the personal identification information (i.e., PIN). A transaction with a suspicion score higher than a certain cut-off value is defined as a "high risk" transaction. The cut-off value is usually determined according to the tolerance of the false positive rate (FPR) of the model. "Extremely high risk" transactions are defined based on a higher cut-off value in the same way.

Two types of transactions will be rejected at the authorization step. One is a transaction with a correct verification result but an extremely high fraud risk, and the other is a transaction with an incorrect verification result. Transactions with correct verification results and normal fraud risk are approved. In the Alert Management module, the high-risk transactions will trigger alerts to the transaction monitoring department. The transaction monitoring department will alert the credit card users to the occurrence of these high fraud risk transactions through text, email and or mobile app notifications or sometimes carry out a transaction investigation telephonically. The fraud cases reported by the transaction monitoring department and the customer service department are collected, and the corresponding transactions are flagged as fraudulent transactions in the database.

In this paper, we mainly focus on the Offline Model Training module. A training dataset is sampled from historical transactions over a certain period (one year/half year) in the data repository. Feature engineering based on the HOBA is carried out to obtain the feature variables for the statistical model. Then, a deep learning model is constructed for credit card fraud detection. After training, the model is placed into the Online Fraud Detection module to calculate the suspicion scores for credit card transactions in real time.

3.2. Credit card data description

We list the essential attributes of a raw transaction record in Table 1 to show the main frame of the credit card transaction table. Although the entire structure of the transaction information table may be slightly different among card issuers, the essential attributes listed should be contained in the database of card issuers and are obtainable for fraud detection modeling.

Table 1
Essential attributes for credit card transaction data.

Attribute Name	Description	Information Level
Account number	Associated account number	Account Status
Open to buy	The available balance	
Credit Limit	The maximum amount of credit of the associated account	
Card number	Credit card number	Transaction Detail
Transaction Date	Date of the transaction	
Transaction Time	Time of the transaction	
Transaction Amount	The transaction amount submitted by the merchant	
Transaction Type	Transaction types, such as a cash withdrawal or purchase	
Currency Code	The currency code	
Merchant Category Code	The merchant business type code	
Merchant Number	The merchant reference number	
Transaction Country	The country where the transaction takes place	
Transaction City	The city where the transaction takes place	
Entry Mode	A code that describes how the cardholder account information was entered into the terminal	
Approval Code	The response to the authorization request, i.e., approve or reject	

From the fraud detection modeling aspect, a credit card transaction record contains the transaction detail and the account status. Most attributes related to the transaction detail can be regarded as multilevel categorical variables that encode the characteristics of a credit card transaction. Here, the characteristics of a transaction is a broader concept than the type of transaction. The type of transaction is typically a field in the transaction record table, and it is used to indicate whether a transaction record is a purchase, cash withdrawal or another type of transaction. A characteristic of a transaction refers to any intrinsic property of the corresponding transaction behavior. A characteristic can be a category based on a specific attribute, such as the transaction channel type, transaction country or others. Furthermore, it can also be a category based on manmade rules. For example, if the POS entry mode of a transaction shows that it is entered via a magnetic stripe, this transaction has a characteristic of the magnetic stripe entry mode. The behavior of transactions with the magnetic stripe entry mode characteristic is worth analyzing separately because this type of transaction is more likely to be fraudulent than chip and PIN transactions. Another example is that we may define a transaction that occurs between 1 am and 5 am as an abnormal time transaction, the abnormality of the transaction time is a possible transaction characteristic.

3.3. Feature engineering framework based on HOBA analysis

As we have mentioned, feature engineering has a great influence on fraud detection performance. Previous studies have mainly considered transaction aggregation strategies in the feature engineering process. Every time a transaction occurs, the feature variables are calculated based on the incoming transaction and past transactions within the aggregation period. Van Vlasselaer [39] noted that the principle of this type of feature engineering fits in the RFM framework. However, we note that using the RFM framework as a guiding principle is not proper for a behavior analysis of credit card transactions. The reason is that credit card transaction records are intrinsically heterogeneous and different transaction characteristics correspond to different behavior patterns [7]. For example, credit card transactions include purchases, refunds, cash withdrawals and balance enquiry transactions. In most cases, the frequency of cash withdrawal transactions is lower than that of purchase transactions. Similarly, there is a large difference in the frequency (recency/monetary) between any two types of card transactions. Therefore, indiscriminately analyzing how often (how frequently/ how much) a credit card holder makes all these types of transactions is less meaningful.

Furthermore, in addition to overlooking the heterogeneity, the RFM analysis does not take the transaction location into account. It is quite understandable because the transaction location appears not to be very helpful in promoting sales. However, analyzing where credit card holders make a transaction and whether there is an abnormality in transaction locations is very helpful in credit card fraud detection. In many cases, fraudsters make transactions in an unusual location that is far from the card holders' residence or common transaction places.

Therefore, we propose a feature engineering framework based on homogeneity-oriented behavior analysis. The homogeneity-oriented behavior analysis indicates that we carry out a behavior analyses on a group of intrinsically homogeneous historical transactions (i.e., with the same transaction characteristic). The rationale is that a credit card transaction can be regarded as a point in a space consisting of transaction characteristics, time, geographic space, and monetary value. With different transaction characteristics, the transaction distribution is different, which leads to the heterogeneity of credit card transactions. For a more reasonable behavior analysis, we advocate extracting information based on transactions in the same subspace of the hyperspace where a certain characteristic is fixed.

In our proposed framework, we first need to determine a set of characteristics worth investigating for fraud detection. Then, for each of these characteristics, four types of customer behavior analysis will be carried out:

Recency- How recently did a credit card user make a transaction with a given **characteristic**?
Frequency- How often did a credit card user make transactions with a given **characteristic**?

Monetary value- How much did a credit card user spend in transactions with a given **characteristic**?
Location- Where did a credit card user make transactions with a given **characteristic**?

Two strategies, a transaction aggregation strategy and a rule-based strategy, are applied to fulfill the above homogeneity-oriented behavior analyses and extract the feature variables on historical transaction data. The primary strategy is the transaction aggregation strategy. It can be decomposed into four components: the aggregation characteristic, aggregation period, transaction behavior measure and aggregation statistics.

- The aggregation characteristic determines the category of the transactions that are aggregated within the aggregation period.
- The aggregation period determines the time range of the transactions that need to be aggregated. In the literature, there are four common aggregation periods: the last hour, the last day, the last week and the last month.
- The transaction behavior measure determines a numerical measure to quantify the behavior of a transaction and is associated with the above four types of behavior analysis. For example, the transaction amount is a common behavior measure in previous work. It quantifies how much is spent in a particular transaction. The transaction time interval between two sequential transactions is also a measure to quantify how recently a customer has made a transaction. In our work, we want to identify two more helpful measures that are not considered in previous works. The first one is the geographical distance between the locations of two sequential transactions, which quantifies the variation of transaction locations. The other is the utilization percentage of the open-to-buy amount (or the available balance), namely, the ratio between the transaction amount and the open-to-buy amount. Compared to the transaction amount, the utilization percentage of open-to-buy is a relative monetary measure. It avoids the magnitude problem in the transaction amount due to the difference in economic resources among credit card holders. In addition, fraudsters are often likely to make full use of the open-to-buy amount within a short period. Relative monetary information is helpful in detecting this kind of behavior.
- The aggregation statistics determine the statistics we apply to the sequence of numerical values extracted by the behavior measures on the transactions. Common statistics are the count, average, sum and standard deviation.

Once the above four elements are given, we can determine a feature variable. For example, if the aggregation period, aggregation characteristic, transaction behavior measure and aggregation statistics are the last week, purchase transaction, transaction time interval and average, respectively, the variable we obtain is the average transaction time interval between two successive purchase transactions within the last week. For every transaction, the aggregation strategy first extracts historical transactions with the characteristic of purchase, i.e., transaction records of which the binary indicator value for the purchase characteristic is 1, within the past week. Then, the time interval between two sequential transactions is calculated based on these extracted transactions. The average value of these time intervals is the value of the variable (average transaction time interval of purchase transactions within the last week). A schematic diagram of the aggregation strategy is shown in Fig. 2.

In addition to using the transaction aggregation strategy, we also apply a rule-based strategy to generate some categorical feature variables that are directly treated as categorical feature variables. Adopting the rule-based strategy aims at investigating some complex characteristics of credit card transactions. For example, we create a binary feature to indicate whether a transaction is the second highest value (suppose that it is greater than $500) transaction at an abnormal time in a specific high-risk country within one hour. This type of binary variable is usually created by fraud system users with their domain knowledge and experience. Different fraud system users could analyze their own credit card fraud cases and design different binary variables to flag transaction characteristics that they think are related to a certain type of fraudulent behavior. These features could be adjusted whenever a new type of fraud behavior becomes prevalent or an old type of fraud behavior loses efficacy.

The problem of feature selection arises with the invention of the above framework. There are infinitely many possible combinations of the four elements, the aggregation period, the aggregation characteristic, transaction behavior measure and the aggregation statistics. From experience in previous work, there are four common choices of the aggregation period, five choices of aggregation characteristics, two behavior measures and four statistics that result in 160 possible compositions to create 160 corresponding feature variables. The number of features will increase in our work because we propose more choices in the aggregation strategy and some categorical binary variables for a more comprehensive behavior analysis.

However, there is no agreement regarding which type of feature variables should be used. As far as we are concerned, the optimal selection of feature variables is highly related to the dataset. Rather than using a small set of generally designed features, selecting appropriate features from a large set of candidate features is preferred in practice to achieve the best fraud detection performance. However, traditional feature selection methods are usually time-consuming and sometimes cause overfitting [40]. In addition, variations in the number and implementation of selected features will make updating the model difficult for a real-time fraud detection system. One of the best advantages of deep leaning is hierarchical feature selection along the successive layers of increased abstraction in detecting patterns, which is the key reason why we apply deep learning techniques to fraud detection. The deep learning techniques can effectively learn useful data representations from the complete dataset without losing much information. Furthermore, we only need to adjust the weights during model updating. Thus, in our work, we employ deep learning methods on the entire set of feature variables and do not perform feature selection.

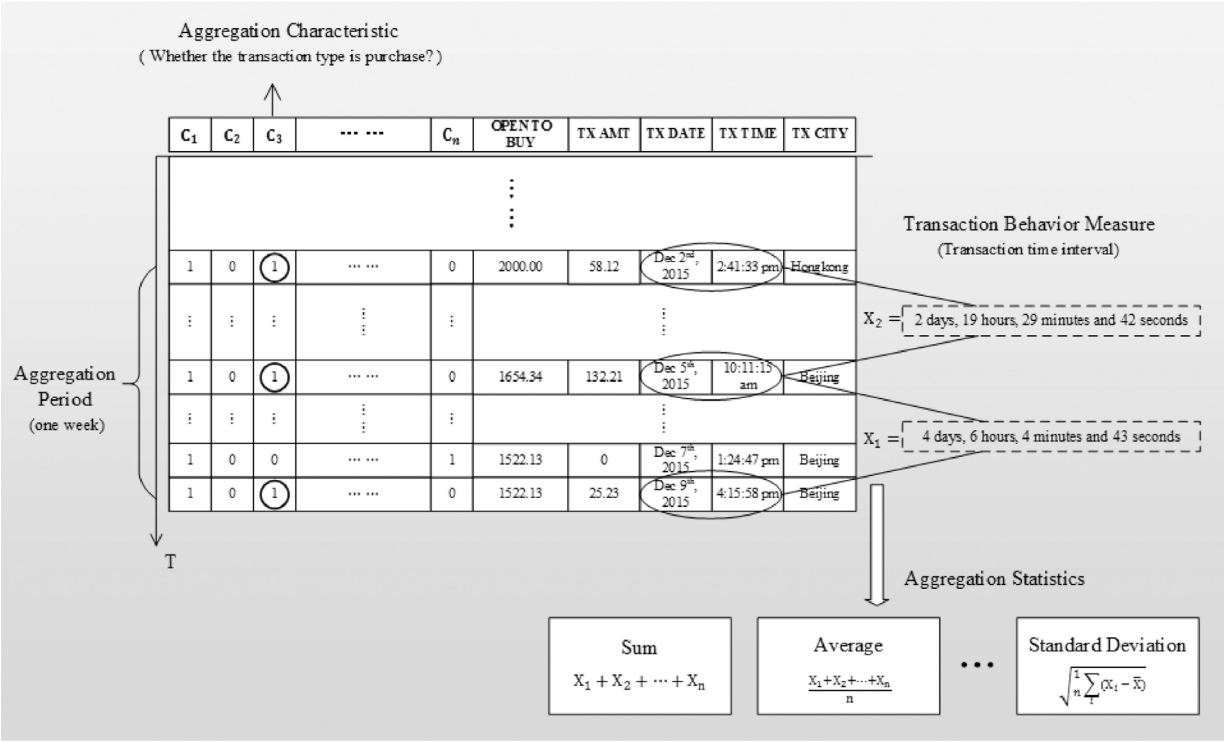

Fig. 2. Schematic diagram of applying the transaction aggregation strategy with the HOBA principle.

3.4. Deep learning for fraud detection

3.4.1. Deep belief networks

A deep belief network (DBN) can be constructed by stacking many restricted Boltzmann machines (RBMs) [15]. Typically, for binary input features, a Bernoulli-Bernoulli RBM is constructed as the first layer, and a Gaussian-Bernoulli RBM is constructed as the first layer for continuous value input features. After training the first layer of the RBM, the second layer Bernoulli-Bernoulli RBM is constructed by treating the hidden layer of the first RBM as the visible input layer. Similarly, more RBMs can be used layer-by layer to construct a DBN, where the hidden layer of a lower RBM can be used as the visible layer of the RBM of the next layer. An important advantage of this layer by layer greedy training strategy is that it requires no class labeling. We plot a schematic plot of the training process of the DBN in Fig. 3.

For a classification problem, a final layer, which represents the class labels of the training data, is added to the DBN. A discriminate learning method, typically the error backpropagation method, can be added to fine-tune the weights after generative pretraining. Then, the final training weights are used as the weights of a traditional feed-forward neural network with the same structure. This pretraining method has been shown to effectively overcome the error dilution problem in traditional stochastic gradient descent training of multilayer neural networks.

3.4.2. Convolutional neural networks

A convolutional neural network (CNN) consists of an input layer, an output layer and several hidden layers, including convolutional layers, pooling layers, normalization layers and fully connected layers. The CNN model is well suited for learning a large amount of data and has achieved great success in various fields, such as image recognition, natural language processing and recommender systems [11]. To apply the CNN model to credit card fraud detection, feature transformation is performed to generate feature matrices based on extracted one-dimensional features [14], as shown in Fig. 4. The rows of a generated feature matrix contain the original features arranged by feature type, while the columns represent recent transactions in order. Then, these feature matrices, each of which corresponds to a transaction record and reflects recent trading habits, are used to train a CNN classifier.

3.4.3. Recurrent neural networks

The most prominent characteristic of a recurrent neural network (RNN) is its recurrent structure, where connections between internal nodes form a directed graph along a successive sequence, which facilitates remembering past sequences of inputs. Depending on their internal memory capacity, RNNs have been proven well suited for learning from experience

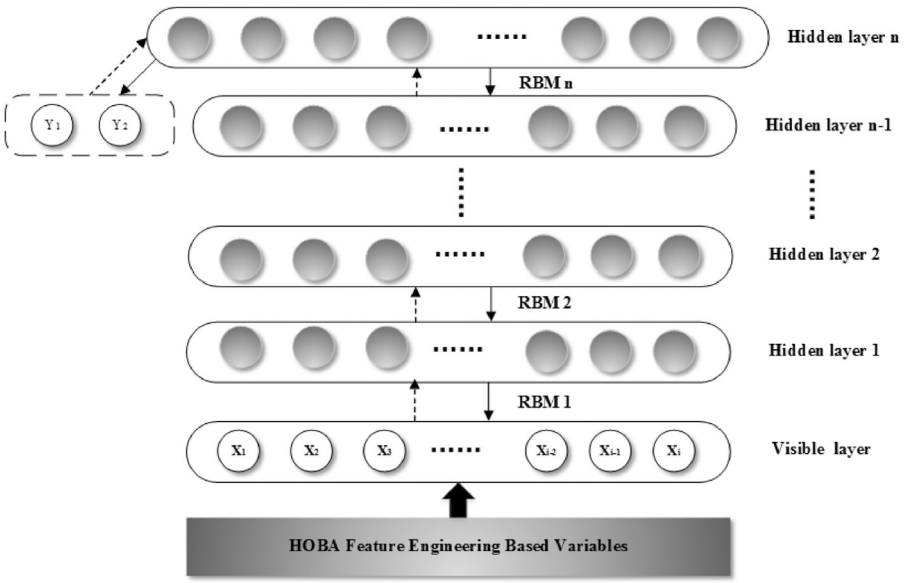

Fig. 3. Training a deep belief network.

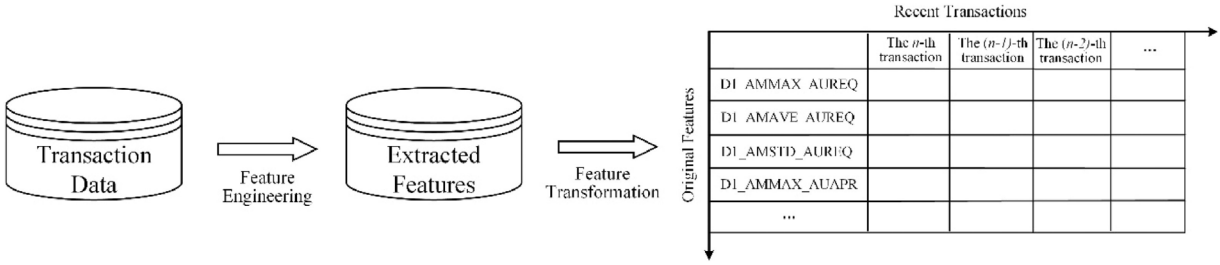

Fig. 4. Illustration of feature transformation for CNN modeling.

to process, classify and predict sequences and have achieved outstanding results in fields such as speech recognition and spam detection [34]. The architecture of the RNN model used in this study is the Elman network [10], which is presented in Fig. 5.

The chronological transaction records of a card holder constitutes a trading behavior sequence in which his hidden trading habits can be reflected. Each transaction denotes an element in the sequence. Then, the RNN model is employed to detect suspicious transaction activities associated with credit card fraud based on input sequential data.

4. Empirical analysis

4.1. Data description

In this study, we use a real-life dataset from one of the largest commercial banks in China. The fraudulent transactions are labeled according to records from the fraud investigation department. The dataset contains 153,685 transaction records and is highly imbalanced (only 2046 (<1.5%) fraudulent transactions). We preprocess the dataset to rule out duplicates, outliers and some transactions that were not launched by the customers themselves, for example, reversal transactions. We obtain 114,779 transactions in total to perform the experiment. The dataset is divided into a training-validation dataset and a test dataset. The training dataset contains 61,735 transactions, and the testing dataset contains 53,044 transactions. The numbers of fraud transactions in the training set and the testing set are 866 and 1176, respectively.

4.2. Experimental setup

The main objective of this paper is to examine the performance of our proposed fraud detection method which is based on deep learning architecture together with the efficacy of a HOBA analysis. We compare three popular deep learning models,

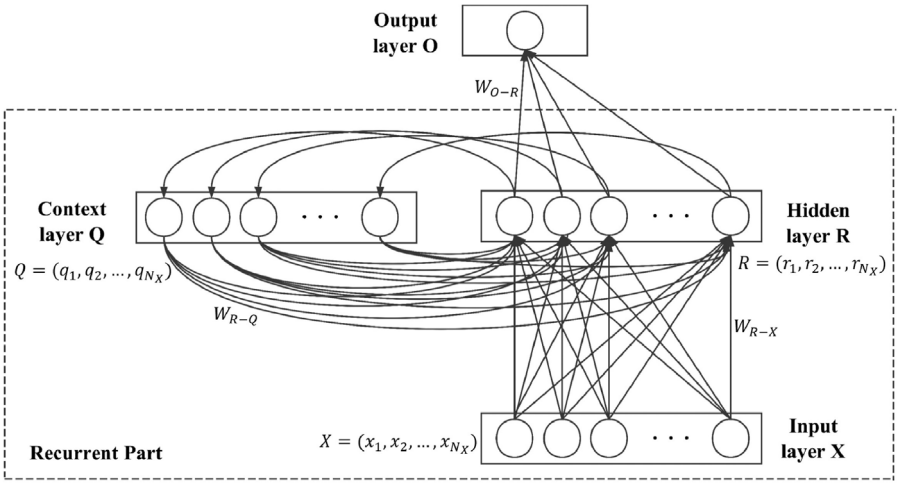

Fig. 5. The architecture of the recurrent neural network.

including the DBN, CNN and RNN, with the three most widely used data mining techniques in fraud detection, i.e., the random forest (RF) and support vector machine (SVM), together with traditional shallow-structured artificial neural network (denoted as BPNN). More specifically, we primarily carry out experiments to examine the performance of our deep-learning-based method, which exploits the feature variables generated by a HOBA analysis as the model input. The detection performance of three traditional methods using variables generated by the RFM analysis framework are used as the benchmarks.

Under the HOBA feature engineering framework, we use the transaction aggregation strategy and rule-based strategy to generate 1410 variables. For the transaction aggregation strategy, we choose a subset of all possible compositions of 13 aggregation characteristics, 9 aggregation periods, 4 transaction behavior measures (transaction amount, time interval between two transactions, geographical distance interval between two transactions and usage of open-to-buy), and 4 aggregation statistics (mean, standard deviation, max, and sum). The 9 aggregation periods we choose, including the past one day, the past three days, the past five days, the current transaction, the past five transactions, the past seven transactions, the past fifteen transactions, the past sixth to tenth transactions and the past eighth to fifteenth transactions, can be categorized into short, medium and long periods to provide a summary of the behavior information from short term to long term. The aggregation characteristics we choose are purchase, cash withdrawal, abnormal time, foreign transaction country, high risk MCC, online transaction and some other characteristics relevant to identifying fraudulent behavior. For comparison, the traditional machine learning methods use the variables under the RFM framework as suggested in the literature [39]. We similarly calculated three types of RFM variables under five levels of aggregation to construct 60 variables for the RFM analysis, and we also added some categorical variables as suggested. Therefore, the RFM variable set contains 87 variables in total.

The popular 10-fold cross-validation approach is used for model evaluation and model selection to avoid overfitting classifiers [37]. Ten equal-sized subsets of the training-validation set are randomly generated. Then, nine subsets are used to train a classifier, while the remaining subset is used to test it. This procedure is repeated 10 times, with each subset used nine times for training and once for testing. Automated tuning with a grid search in parameter space is employed for fine tuning the important parameters to ensure the fairness and objectivity of the results for the comparative experiments.

4.3. Evaluation criteria

For the classification problem, we obtain the predicted probability (suspicion score in the fraud detection) for each observation, which ranges from 0 to 1 and represents the probability of being the positive class. When we set a certain cutoff value (set as 0.5 by default) to classify each observation into the positive or negative class, the confusion matrix below is obtained.

Table 2
The definition of the confusion matrix.

		Predicted class	
		Positive	Negative
Actual class	Positive	TP	FN
	Negative	FP	TN

Table 3
Performance of classifiers using different feature sets.

Classifiers	RFM Features				
	F1-Measure	Precision	Recall	Accuracy	AUC
BPNN	0.282	28.02%	28.40%	96.80%	0.929
SVM	0.257	29.79%	22.62%	97.10%	0.929
RF	0.311	29.56%	32.82%	96.90%	0.935
DBN	0.373	34.94%	39.97%	97.02%	0.958
CNN	0.340	36.25%	32.06%	97.24%	0.957
RNN	0.328	35.82%	30.19%	97.25%	0.956
Classifiers	HOBA Features				
	F1-Measure	Precision	Recall	Accuracy	AUC
BPNN	0.418	41.44%	42.18%	97.40%	0.936
SVM	0.358	40.10%	32.40%	97.43%	0.941
RF	0.437	44.59%	42.77%	97.55%	0.949
DBN	0.568	62.60%	51.96%	98.25%	0.976
CNN	0.518	55.73%	48.38%	98.00%	0.968
RNN	0.540	58.28%	50.26%	98.10%	0.971

Several commonly used classification performance measures based on the confusion matrix, shown in Table 2, are employed in this paper to evaluate the fraud detection performance:

$$\text{Accuracy}: \ (TP + TN)/(TP + FP + TN + FN) \tag{1}$$

$$\text{Recall (or Sensitive/True positive rate)}: \ TP/(TP + FN) \tag{2}$$

$$\text{Precision}: \ TP/(TP + FP) \tag{3}$$

$$\text{F1-measure}: \ 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \tag{4}$$

$$\text{False positive rate}: \ FP/(FP + TN) \tag{5}$$

However, researchers all agree on the inadequacy of the above performance measures in fraud detection. Because of the significant imbalance in the dataset, classifying every transaction as a legitimate transaction will produce an overall accuracy that even a state-of-the-art model cannot beat. In addition, the four classes of classification results in the confusion matrix will incur different misclassification costs in credit card fraud detection. The false negatives, which we erroneously predict as negative, will often cause larger costs than the false positives, since the later class only requires the cost of an investigation.

Therefore, in addition to the above measures, we adopt the AUC (the area under the ROC curve) value as an overall performance measure. The ROC curve (Receiver Operating Characteristic) is a graphical plot. It plots the true positive rate (TPR) against the false positive rate (FPR) at different possible thresholds. The AUC is considered to be a better overall performance measure than accuracy because it is independent of a cutoff value. The closer the AUC value comes to 1, the better the overall performance a model has.

4.4. Experimental results

There are two main questions that are answered through empirical experiments. The first question is whether the deep learning techniques can provide better fraud detection performance. The second is whether our proposed feature engineering framework can provide better variables for machine learning methods in fraud detection. The first experiment group (presented in Subsection 4.4.1) aims to investigate these two questions. We evaluate the fraud detection performance of both the deep learning models and traditional machine learning methods using different feature sets. Moreover, to evaluate the fraud detection models from the practical aspect, we conduct the second experiment group (presented in Subsection 4.4.2) to investigate the performances of the models under a certain tolerance to the false positive rate.

4.4.1. Overall fraud detection performance

The evaluation results of the fraud detection performance of both the deep learning techniques and traditional machine learning models using different feature sets are shown in Table 3. Table 4 presents the experimental results of McNemar's test [30], which compares the classification performance of the different classifiers. The F1-measure and AUC values can reflect the overall performance of the models. The recall metric represents the proportion of the actual fraudulent transactions that have been correctly predicted, while the precision metric denotes the proportion of the correctly predicted fraudulent transactions to the predicted fraudulent transactions. Both recall and precision are important evaluation metrics. An anti-fraud system with higher recall can protect the interests of customers, while a system with higher precision can reduce

Table 4
McNemar's test results on classifier performance.

Classifiers	Compared with the classifier (H_0: Classification results are equal)		
	BPNN+RFM	SVM+RFM	RF+RFM
BPNN+HOBA	0.00***	0.00***	0.00***
SVM+HOBA	0.00***	0.00***	0.00***
RF+HOBA	0.00***	0.00***	0.00***
DBN+RFM	0.00***	0.00***	0.00***
CNN+RFM	0.00***	0.00***	0.00***
RNN+RFM	0.00***	0.00***	0.00***
DBN+HOBA	0.00***	0.00***	0.00***
CNN+HOBA	0.00***	0.00***	0.00***
RNN+HOBA	0.00***	0.00***	0.00***

*p-Value < 0.1, **p-Value < 0.05, ***p-Value < 0.01.

regulatory costs to some extent. By comparing the predictive results of the deep learning models with those of traditional machine learning methods in Table 3, we can see that the application of deep learning techniques dramatically improves the credit card fraud detection capabilities. Deep learning models (DBN, CNN and RNN) outperform all three baseline machine learning methods (BPNN, SVM, RF) using the RFM feature set in the F1-measure value, precision and AUC value and achieve much better results on all the evaluation metrics when using the HOBA feature set. As denoted in Table 4, the improvements in classification performance over the benchmark models (BPNN+RFM, SVM+RFM and RF+RFM) are all statistically significant. The comparative results indicate that deep learning improves fraud detection performance by extracting useful data representations from the input data. In addition, the RF outperforms the other two traditional data mining methods, which is consistent with the results of many past studies.

Specifically, from the aspect of the F1-measure and AUC values, which measure the overall performance of classifiers, the DBN achieves the highest F1-measure score of 0.373 and the highest AUC value of 0.958 when using the RFM feature set. Meanwhile, the highest F1-measure score and AUC value when using the proposed HOBA feature set are also attained by the DBN classifier. On average, the deep learning models have an improvement over the baseline machine learning approaches of 22.6% on the F1-measure score and 2.8% on the AUC value when using the RFM feature set. The deep learning models with HOBA have much better AUC values and F1-measure scores than baseline models, outperforming them by nearly 34.2% on the F1-measure score and 3.2% on the AUC value on average. The improvements of the deep learning techniques under different feature sets demonstrate that deep learning is more powerful in representation learning with a larger feature set (the HOBA feature set). As a conclusion of the above analyses, the deep learning approaches dramatically outperform all the other baseline models regardless of the RFM feature set or the HOBA feature set. The DBN classifier with HOBA variables obtains the best performance and significantly outperforms the benchmark models.

Then, we compare the fraud detection performance of the classifiers in the upper and lower parts of Table 3 to evaluate the effectiveness of the proposed HOBA feature engineering framework. The classifiers using the HOBA feature set achieve considerably better F1-measure scores, precision, recall and AUC values than those using the RFM feature set, outperforming them by nearly 0.158 on the F1-measure score, 18.06% on precision, 13.65% on recall and 0.013 on AUC value. It is obvious that feature engineering based on HOBA provides extra benefits to all the data mining methods in detecting suspicious transactions. These results indicate that the application of deep learning techniques and HOBA feature engineering leads to a significant improvement in detection performance.

We plot the ROC curves of the classifiers associated with different variable sets in Fig. 6, from which we can directly compare the overall performances of the anti-fraud models under different circumstances.

As can be seen from Fig. 6, the curves of the deep learning models using the HOBA feature set are located closest to the top left corner of the figures, suggesting that the proposed feature engineering methodology for credit card fraud detection with a deep learning architecture has achieved better performance. These figures also demonstrate the utility of deep learning techniques and HOBA feature engineering. Overall, these results illustrate the efficacy of the proposed methodology.

4.4.2. Fraud detection performance under differing false positive rates

To make a further evaluation, we test model performance under different false positive rates (FPRs) and incorporate the business-related information to assess the effectiveness of the proposed method [29]. From a practical perspective, the fraud detection department does not welcome a model with a high false positive rate. As we can see from the formulas of FPR and recall, a higher recall value means that we can capture more fraudulent transactions. However, as we detect more fraudulent transactions, we may misclassify more legitimate transactions. As a result, the false positive rate may also increase, i.e., more false positives. In practice, these false positives not only increase the cost of fraud investigations but also incur a reputation loss for the credit card issuer. If the FPR of a fraud detection model is too high, it causes excessive interference in legitimate transactions, which may finally annoy costumers and lead to customer turnover. Customer turnover will cause a large loss to card issuers and even offset the benefit bought by the fraud detection model. Therefore, the fraud detection department is more interested in the model performance under an acceptable FPR tolerance. Considering the massive volume of credit card transactions, the acceptable FPR value is very small.

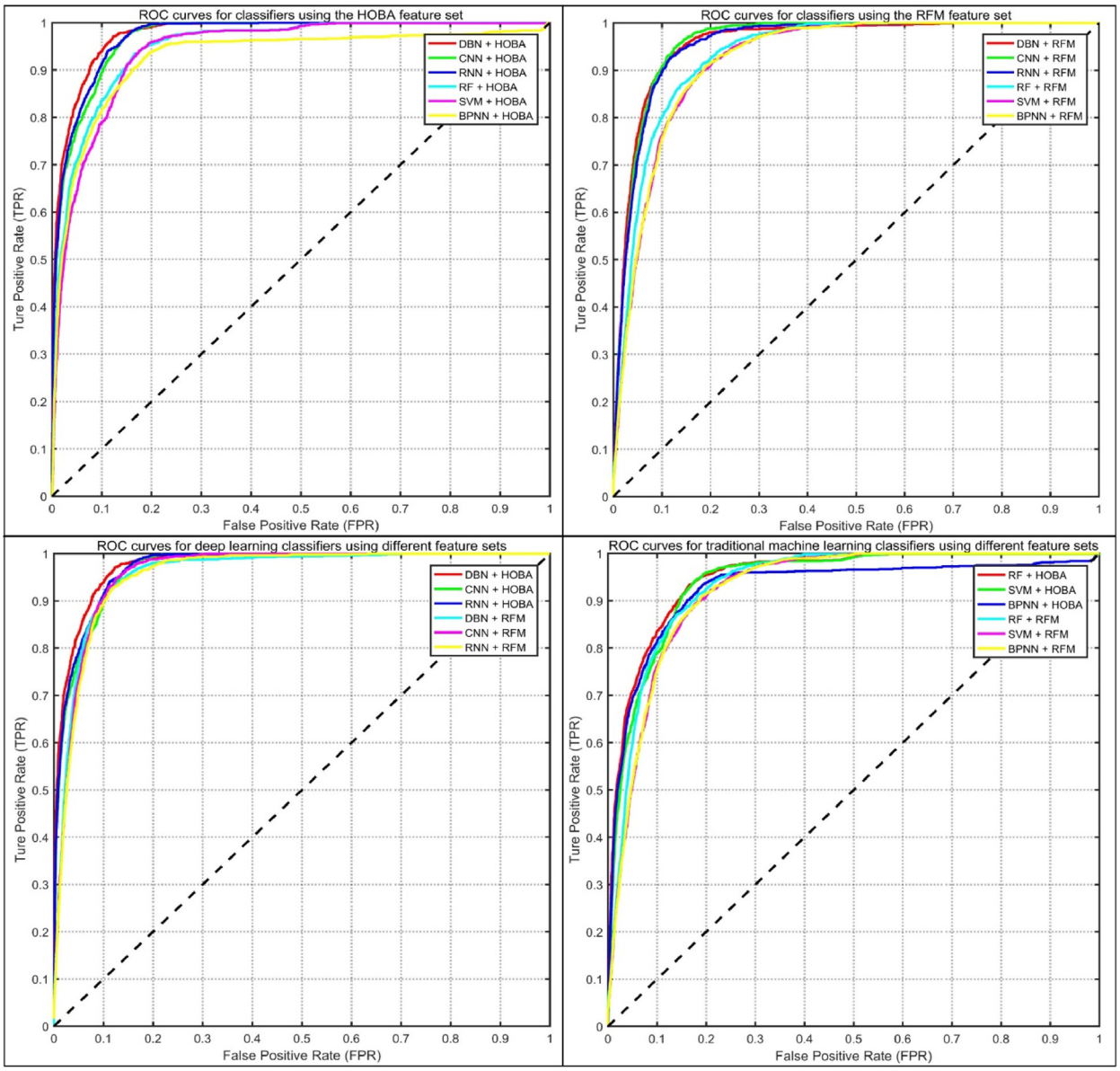

Fig. 6. ROC curves for different approaches.

According to the fraud analysts at the commercial bank, a fraud detection model with an FPR lower than 1% is just perfect, and 3% is the maximum acceptable FPR rate. Therefore, in our work, we set two different tolerance values of FPR as 1% and 3%. Then, we compare the performance of the different approaches. Table 5a shows the performance measures under the tolerance of 1% FPR. The DBN model using the HOBA feature set obtains the highest F1-measure score of 0.577, still outperforming all the other approaches on the overall performance measure. It also has the highest recall of 58.33%, which indicates that the model can detect more than half of the suspicious transaction activities under an expected low false positive rate. The deep learning models still perform much better than the traditional machine learning methods with different feature sets. In addition, the F1-measure scores of the classifiers using the HOBA feature set range from 0.341 to 0.577, while those of classifiers using the RFM feature set range from 0.234 to 0.326. Thus, again, it shows some evidence that the feature engineering based on HOBA provides a better feature variable set.

We list the performance results under the tolerance of 3% FPR in Table 5b. The combination of the deep learning models and the HOBA feature set continues to provide much better experimental results. All these classifiers obtain much better recall since we relax the false positive rate tolerance. More fraudulent transactions can be identified. However, the precision inevitably decreases as a sacrifice. The overall F1-measure scores and accuracy are lower under such an FPR tolerance of FPR

Table 5a
Performance measures at a 1% maximum false positive rate.

Classifiers	RFM Features				
	FPR	F1-Measure	Precision	Recall	Accuracy
BPNN	1%	0.245	31.42%	20.07%	97.26%
SVM	1%	0.234	30.28%	19.13%	97.23%
RF	1%	0.256	32.42%	21.09%	97.28%
DBN	1%	0.326	39.19%	27.89%	97.44%
CNN	1%	0.314	37.89%	26.87%	97.40%
RNN	1%	0.307	37.21%	26.11%	97.39%
Classifiers	HOBA Features				
	FPR	F1-Measure	Precision	Recall	Accuracy
BPNN	1%	0.391	44.35%	35.03%	97.59%
SVM	1%	0.341	40.23%	29.59%	97.46%
RF	1%	0.417	46.27%	37.93%	97.65%
DBN	1%	0.577	57.12%	58.33%	98.11%
CNN	1%	0.521	53.65%	50.60%	97.94%
RNN	1%	0.535	54.51%	52.47%	97.98%

Table 5b
Performance measures at a 3% maximum false positive rate.

Classifiers	RFM Features				
	FPR	F1-Measure	Precision	Recall	Accuracy
BPNN	3%	0.313	24.59%	43.11%	95.81%
SVM	3%	0.322	25.27%	44.56%	95.85%
RF	3%	0.344	26.77%	48.21%	95.93%
DBN	3%	0.411	31.23%	60.03%	96.18%
CNN	3%	0.387	29.65%	55.53%	96.09%
RNN	3%	0.381	29.23%	54.59%	96.06%
Classifiers	HOBA Features				
	FPR	F1-Measure	Precision	Recall	Accuracy
BPNN	3%	0.394	30.08%	56.89%	96.11%
SVM	3%	0.380	29.20%	54.42%	96.06%
RF	3%	0.412	31.32%	60.12%	96.19%
DBN	3%	0.488	36.18%	75.00%	96.51%
CNN	3%	0.463	34.58%	69.90%	96.40%
RNN	3%	0.473	35.24%	71.68%	96.45%

as well. The recall of the deep learning classifiers using the HOBA feature set ranges from 69.90% to 75.00%, which suggests that the proposed anti-fraud system can identify most of the fraudulent transactions under an acceptable false positive rate. These performance gains indicate that the proposed feature engineering framework with a deep learning architecture has a stronger fraud detection capability from a practical perspective.

5. Conclusions and future work

In summary, this paper proposes a novel feature engineering framework with a deep learning architecture for credit card fraud detection. A feature engineering framework based on a homogeneity-oriented behavior analysis (HOBA) is proposed to generate the feature variables representing the behavior information for fraud detection models. Compared to previous works, our feature engineering framework takes the heterogeneity of credit card transactions into consideration and carries out a behavior analysis only on homogenous transactions. We also enrich the previous transaction aggregation strategy with the behavioral measures of geographical location distance and relative monetary amount. Moreover, we employ deep learning techniques for more effective fraud detection modeling based on a vast number of extracted features. We conduct comparative experiments to compare the performances of several deep learning algorithms and traditional machine learning methods based on different feature sets.

The empirical study is carried out on a real-world dataset from one of the largest commercial banks in China. The results show that the deep learning approaches dramatically outperform all other baselines regardless of the RFM feature set or the HOBA feature set. The DBN classifier with the HOBA variables obtains a significantly better performance than the benchmark models. The experimental results also reveal that feature engineering based on HOBA benefits all the data mining methods in detecting fraudulent transactions. These comparative experimental results suggest that the performance improvement of the proposed method is attributable to the application of deep learning techniques and HOBA feature engineering. Additionally,

we provide a practical evaluation of the credit card fraud detection model and investigate the model performance under a certain tolerance of the false positive rate. The empirical analysis results provide evidence to support that our proposed method also has a significant advantage in credit card fraud detection capabilities from a practical perspective.

To the best of our knowledge, this is the first successful study of applying homogeneity-oriented behavior analysis to analyze the transaction behaviors of credit card holders for fraud detection. The managerial implication of our work is that card issuers can apply the proposed methodology to efficiently and effectively identify fraudulent credit card transactions to protect the interests of customers and reduce fraud losses and regulatory costs.

There are still some limitations in our work. For example, we did not assess the computational cost of our proposed HOBA feature engineering framework, which constructs a much larger variable set than RFM. Therefore, in future work, we want to carry out further research from two aspects. The first focuses on researching the computational demand of a real-time fraud detection system. The second is to explore the application of more advanced machine learning methods and possible combinations of deep learning methods and traditional data mining methods in fraud detection.

Conflict of interest

We declare that we have no financial and personal relationship with other people or organizations that can inappropiately influence our work.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grant No. 71771212, U1711262), Humanities and Social Sciences Foundation of the Ministry of Education (No. 14YJA630075, 15YJA630068), and Fundamental Research Funds for the Central Universities, and Research Funds of Renmin University of China (No. 15XNLQ08).

References

[1] A. Abdallah, M.A. Maarof, A. Zainal, Fraud detection system: a survey, J. Netw. Comput. Appl. 68 (2016) 90–113.
[2] O. Abdel-Hamid, A.R. Mohamed, H. Jiang, L. Deng, G. Penn, D. Yu, Convolutional neural networks for speech recognition, IEEE-ACM Trans. Audio. Sp. 22 (10) (2014) 1533–1545.
[3] E. Aleskerov, B. Freisleben, B. Rao, Cardwatch: a neural network based database mining system for credit card fraud detection, in: Proceedings of the IEEE/IAFE Computational Intelligence for Financial Engineering (CIFEr), IEEE, 1997, pp. 220–226.
[4] S. Bhattacharyya, S. Jha, K. Tharakunnel, J.C. Westland, Data mining for credit card fraud: a comparative study, Decis. Support Syst. 50 (3) (2011) 602–613.
[5] R.J. Bolton, D.J. Hand, Statistical fraud detection: a review, Stat. Sci. 17 (3) (2002) 235–249.
[6] R. Brause, T. Langsdorf, M. Hepp, Neural data mining for credit card fraud detection, in: international Conference on Tools with Artificial Intelligence, IEEE Comput. Soc. (1999) 103.
[7] C. Bravo, L.C. Thomas, R. Weber, Improving credit scoring by differentiating defaulter behaviour, J. Oper. Res. Soc. 66 (5) (2015) 771–781.
[8] N. Carneiro, G. Figueira, M. Costa, A data mining based system for credit-card fraud detection in e-tail, Decis. Support Syst. 95 (2017) 91–101.
[9] A. Dal Pozzolo, O. Caelen, Y.A. Le Borgne, S. Waterschoot, G. Bontempi, Learned lessons in credit card fraud detection from a practitioner perspective, Expert Syst. Appl 41 (10) (2014) 4915–4928.
[10] J.L. Elman, Finding structure in time, Cognitive Sci 14 (2) (1990) 179–211.
[11] M.J. Er, Y. Zhang, N. Wang, M. Pratama, Attention pooling-based convolutional neural network for sentence modelling, Inform. Sci. 373 (2016) 388–403.
[12] U. Fiore, A.D. Santis, F. Perla, P. Zanetti, F. Palmieri, Using generative adversarial networks for improving classification effectiveness in credit card fraud detection, Inf. Sci. 479 (2019) 448–455.
[13] T. Fischer, C. Krauss, Deep learning with long short-term memory networks for financial market predictions, Eur. J. Oper. Res. 270 (2) (2018) 654–669.
[14] K. Fu, D. Cheng, Y. Tu, L. Zhang, Credit card fraud detection using convolutional neural networks, in: International Conference on Neural Information Processing, Springer, 2016, pp. 483–490.
[15] Z. Geng, Z. Li, Y. Han, A new deep belief network based on RBM with glial chains, Inf. Sci. 463 (2018) 294–306.
[16] S. Ghosh, D.L. Reilly, Credit card fraud detection with a neural-network, in: Proceedings of the 27th Hawaii International Conference on System Sciences, IEEE, 1994, pp. 621–630.
[17] M. Hayat, M. Bennamoun, S. An, Learning non-linear reconstruction models for image set classification, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2014, pp. 156–163.
[18] R. Hecht-Nielsen, Theory of the backpropagation neural network, Neural Netw. 1 (1) (1988) 65–93.
[19] G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets, Neural Comput 18 (7) (2006) 1527–1554.
[20] A.M. Hughes, Strategic Database Marketing, McGraw-Hill, New York, 2000.
[21] S. Jha, M. Guillen, J.C. Westland, Employing transaction aggregation strategy to detect credit card fraud, Expert Syst. Appl. 39 (16) (2012) 12650–12657.
[22] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, F.F. Li, Large-scale video classification with convolutional neural networks, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society, 2014, pp. 1725–1732.
[23] M. Kraus, S. Feuerriegel, Decision support from financial disclosures with deep neural networks and transfer learning, Decis. Support Syst. 104 (2017) 38–48.
[24] M. Krivko, A hybrid model for plastic card fraud detection systems, Expert Syst. Appl. 37 (8) (2010) 6070–6076.
[25] H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations, in: International Conference on Machine Learning, ACM, 2009, pp. 609–616.
[26] H. Lee, P.T. Pham, L. Yan, A.Y. Ng, Unsupervised feature learning for audio classification using convolutional deep belief networks, in: Advances in Neural Information Processing Systems, 2009, pp. 1096–1104.
[27] A.L.D. Loureiro, V.L. Miguéis, L.F. da Silva, Exploring the use of deep neural networks for sales forecasting in fashion retail, Decis. Support Syst. 114 (2018) 81–93.
[28] N. Mahmoudi, E. Duman, Detecting credit card fraud by modified fisher discriminant analysis, Expert Syst. Appl. 42 (5) (2015) 2510–2516.
[29] S. Maldonado, C. Bravo, J. Lopez, J. Perez, Integrated framework for profit-based feature selection and SVM classification in credit scoring, Decis. Support Syst. 104 (2017) 113–121.

[30] Q. Mcnemar, Note on the sampling error of the difference between correlated proportions or percentages, Psychometrika 12 (2) (1947) 153–157.
[31] S. Panigrahi, A. Kundu, S. Sural, A.K. Majumdar, Credit card fraud detection: a fusion approach using Dempster–Shafer theory and Bayesian learning, Inf. Fusion 10 (4) (2009) 354–363.
[32] C. Phua, R. Gayler, V. Lee, K. Smith-Miles, On the communal analysis suspicion scoring for identity crime in streaming credit applications, Eur. J. Oper. Res. 195 (2) (2009) 595–612.
[33] J.T.S. Quah, M. Sriganesh, Real-time credit card fraud detection using computational intelligence, Expert Syst. Appl. 35 (4) (2008) 1721–1732.
[34] Y. Ren, D. Ji, Neural networks for deceptive opinion spam detection: an empirical study, Inform. Sciences 385 (2017) 213–224.
[35] A. Shen, R. Tong, Y. Deng, Application of classification models on credit card fraud detection, in: International Conference on Service Systems and Service Management, IEEE, 2007, pp. 1–4.
[36] Y. Shen, X. He, J. Gao, L. Deng, G. Mesnil, A latent semantic model with convolutional-pooling structure for information retrieval, in: Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management, ACM, 2014, pp. 101–110.
[37] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Inf. Process. Manag. 45 (4) (2009) 427–437.
[38] A. Srivastava, A. Kundu, S. Sural, A. Majumdar, Credit card fraud detection using hidden Markov model, IEEE Trans. Depend. Secure 5 (1) (2008) 37–48.
[39] V. Van Vlasselaer, C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, B. Baesens, APATE: a novel approach for automated credit card transaction fraud detection using network-based extensions, Decis. Support Syst. 75 (2015) 38–48.
[40] A. Wang, N. An, G. Chen, L. Li, G. Alterovitz, Accelerating wrapper-based feature selection with k-nearest-neighbor, Knowl-based. Syst. 83 (C) (2015) 81–91.
[41] C. Whitrow, D.J. Hand, P. Juszczak, D.J. Weston, N.M. Adams, Transaction aggregation as a strategy for credit card fraud detection, Data Min. Knowl. Disc. 18 (1) (2009) 30–55.
[42] W. Xu, Q. Wang, R. Chen, Spatio-temporal prediction of crop disease severity for agricultural emergency management based on recurrent neural networks, Geoinformatica 22 (2) (2018) 363–381.
[43] V. Zaslavsky, A. Strizhak, Credit card fraud detection using self-organizing maps, Inf. Secur. 18 (1) (2006) 48–63.