A hybrid deep learning model for consumer credit scoring

Conference Paper · May 2018

DOI: 10.1109/ICAIBD.2018.8396195

4 authors, including:

Bing Zhu
Sichuan University

A Hybrid Deep Learning Model for Consumer Credit Scoring

Bing Zhu, Wenchuan Yang, Huaxuan Wang, Yuan Yuan*
Business School
Sichuan University
Chengdu, 610065, China
e-mail: 1533293533@qq.com

***Abstract*—Consumer credit scoring is an essential part of credit risk management in the fast-growing consumer finance industry and various data mining techniques have been proposed and used on it. Recently, deep learning techniques have gained significant popularity and shown excellent performance in many fields such as image recognition, computer vision and so on. In this paper, we try to take the advantage of deep learning and introduce it into consumer credit scoring. We propose a hybrid model that combines the well-known convolutional neural network with the feature selection algorithm Relief. Experiments are carried on a real-world dataset from a Chinese consumer finance company, and the results show that the proposed model gets superior performance in comparison with other benchmark models such as logistic regression and random forest.**

Keywords-consumer credit scoring; hybrid model; deep learning; convolutional neural network; relief algorithm

I. INTRODUCTION

Nowadays, due to the rapid development of information technology, consumer credit business, especially the online one, grows vigorously worldwide. The scale of online consumer finance transactions in China grows from 6 billion in 2013 to 436.7 billion in 2016. Meanwhile, the amount of consumer credit held by banks reached $1132 billion in 2013 in US [1]. The increasing demand for consumer credit provides great opportunities as well as risks. This results in credit scoring being developed as an indispensable part of credit risk management.

Credit scoring is often treated as a binary classification task. The idea of credit scoring model is trying to use characteristics of consumers like age, gender, saving amount, employment status and so on, to determine whether customers are credit-worthy or not [2]. A broad range of techniques has been applied to solve the credit scoring problem. Basically, those methods can be divided into two groups: statistical methods (e.g. logistic regression, discriminant analysis) and machine learning techniques (e.g. support vector machine, k-nearest neighbor, decision tree, neural network) [3]. However, the researchers haven't come across any conclusive proof that one method is irrefutably superior over another. During the past few years, deep learning techniques emerge along with the evolution of the computing power. Deep learning has achieved successes in various areas, such as computer vision, pattern recognition, speech recognition, emotion recognition and natural language processing, where they get superior performance compared to traditional machine learning and statistical techniques [4].

In this paper, we introduce convolutional neural network (CNN) which is a representative technique in deep learning to the consumer credit scoring. Convolutional neural network first appeared in the paper of Yann Lecun, Leon Bottou, Yoshua Bengio, and Patrick Haffner [5], which is designed to handle the variability of data in 2D shape. Since then, CNN has been widely applied to many image processing tasks. Considering its outstanding ability, we decide to apply it to consumer credit scoring to see whether it still works well. We propose a hybrid credit scoring model that combines the convolutional neural network with the well-known feature selection algorithm Relief. To our knowledge, this is the first attempt to apply the convolutional neural network to consumer credit scoring. In the hybrid model, the application of Relief algorithm is meant to reduce the computational burden of the convolutional neural network. In order to verify the performance of the proposed hybrid model, we carry out an empirical experiment and compare it with logistic regression and random forest on a dataset collected from a Chinese consumer finance company. The results show that our model outperforms both logistic regression and random forest. The rest of our paper is organized as follows. Section 2 briefly reviews the existing techniques used in credit scoring. Section 3 introduces the proposed hybrid model. Section 4 presents the empirical experiment procedures and results. Section 5 gives the conclusion and directions for our future work.

II. RELATED WORK

Credit scoring, proposed by Durand [6] about 70 years ago, has become an essential part for credit risk management. During the past few years, many classification models are developed to tackle with the credit scoring problem. Logistic regression [7] and decision trees [8] are the most widely-used models in credit scoring. Both two models are proven to be simple and efficient. More complicate machine learning techniques appear such as support vector machine (SVM) and neural network, also broadly applied to credit scoring [1]. Moreover, ensemble methods which combine advantages of various single classifiers are developing fast recently. For example, Maher Alaraj and Maysam F. Abbod [9] developed the multiple classifier system that employs neural networks, support vector machine, decision trees and naive Bayes as base models. The ensemble approach improved the prediction performance against other base classifiers validated over five real-world credit scoring datasets.

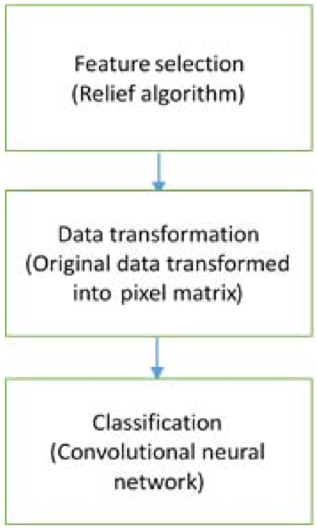

Figure 1. The stages of Relief-CNN.

Many empirical studies have compared the performance of different classification models on credit scoring. For instance, West [10] compared neural network with traditional machine learning techniques. The results showed that neural network has better performance, and logistic regression is also an efficient model. Belloti and Crook [11] concluded that support vector machine improves the accuracy efficiently against logistic regression, linear discriminant analysis and KNN using a dataset provided by a major financial institution. Stefan Lessmann, BartBaesens, Hsin-Vonn Seow, and Lyn C. Thomas [1] presented a comprehensive comparison which involved logistic regression, neural network, Bayesian network, decision trees, and support vector machine. The final results showed that simple models and advanced ones have pretty similar performance. To sum up, there is no consensus about the best models on credit scoring. Researchers all over the world are still searching for better techniques.

In the last two or three years, attracted by the remarkable ability of deep learning, some studies have begun to apply deep learning algorithms to managing credit risk. Niimi [12] developed a deep neural network for credit-card data analysis, and found out that it exhibits a similar performance as commonly used algorithms. Cuicui Luo, Desheng Wu and Dexiang Wu [13] designed a deep belief network (DBN) for credit scoring and compared it with SVM, logistic regression and multilayer perceptron on credit default swaps dataset. The results revealed that DBN yields the best performance. Khiem Tran and Thanh Duong [14] proposed a boosted deep network which combines genetic programming and deep learning. The proposed model got the best accuracy against other traditional machine learning methods. The application of deep learning in credit scoring is still at the early stage. No researcher has ever applied convolutional neural network to credit scoring.

III. METHODOLOGY

In this paper, we propose a hybrid model named "Relief-CNN". It's a combination of two techniques: convolutional neural network and Relief algorithm. It consists of three key stages, which is presented in the figure 1.

The hybrid model first selects important features from the original dataset by using Relief algorithm. The utilization of relief algorithm can efficiently reduce the size of pixel matrix, which can reduce the computational burden of the convolutional neural network. Then, based on the selected feature subset, every instance in the dataset is transformed into a pixel matrix that can be seen as a simple gray image so that CNN can process. Finally, the convolutional neural network receives the input image and performs the classification task. More details about each step are described as follows.

A. Feature Selection

Feature selection is the technique for selecting the subset of highly relevant features by removing redundant, irrelevant and noisy features from original dataset. Quite lots of datasets hold hundreds of features, which may bring lots of problem to the modeling process. The operation is meant to reduce the dimensionality of dataset, which helps the learning algorithm to work faster and more efficient.

Relief is a well-known algorithm for feature selection. It was first proposed by Kira [15] in 1992, served as a simple and efficient method to feature importance weighting. The main idea of relief algorithm is based on nearest neighbor rule.

Given a training data S= $\{X_1, X_2, \dots\dots, X_n\}$ that has n instances, and each instance X_i can be seen as a p-dimensional vector $\{x_{i1}, x_{i2}, \dots, x_{ip}\}$ from the feature set F= $\{f_1, f_2, \dots, f_p\}$. Starting for the initial feature weight vector W_0, Relief algorithm picks a random instances X_i together with its Near-hit instance $nearHit_i$ (the close neighboring sample that belongs to the same class) and Near-miss instance $nearMiss_i$ (the close neighboring sample that belongs to the opposite class) . Relief then uses the following formula to update the feature importance weights:

$$W_i = W_{i-1} - diff(X_i - nearHit_i)^2 + diff(X_i - nearMiss_i)^2 \quad (1)$$

Relief repeats the procedure m times and divide each element of the final weight vector by m. This becomes the relevance vector. Features are selected if their relevance is greater than a threshold τ.

B. Data Transformation

We design a method to transfer every instance into a pixel matrix, which can be seen as a gray image. For each observation $X_i = \{x_{i1}, x_{i2}, \cdots, x_{is}\}$, we first discretize the continuous variables into categorical ones with k values. Then we reshape every x_{ij} into a binary value vector $\{l_1, l_2, \cdots, l_k\}$.

$$l_i = \begin{cases} 1, & when\ x_{ij}\ fall\ into\ the\ i^{th}\ category \\ 0, & otherwise \end{cases} \tag{2}$$

After the transformation, every observation correspond to a labeled gray image with the dimension of $k\times s$.

$$y_i + \begin{pmatrix} 0 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 0 \end{pmatrix}_{k\times s} \tag{3}$$

C. Convolutional Neural Networks Category

Convolutional neural network is a powerful deep learning architecture for processing two-dimensional image data. The basic structure of CNN is shown in Figure 2. As shown in Figure 2, CNN contains two special types of layers called convolutional layer and pooling layer. Convolutional layer is the core building block of the CNN. It consists of a set of learnable filters that slide over the image to extract features. Pooling layer is set to reduce the spatial size of representation as well as the number of parameters and the amount of computation in the network, hence improve the model efficiency and control overfitting.

Compared to traditional neural network, convolutional neural network replaces general matrix multiplication with convolution, which reduces the number of weights used in the network and allows the image to be imported directly. Another important characteristic of CNN is parameter sharing. The basic idea of parameter sharing is to learn one set of parameters through the whole process instead of learning different parameters sets at each location. This unique feature improves the efficiency of whole network. In our paper, we decide to use the CNN with the structure proposed by Alex et al. [16], which included the dropout technique to reduce complex co-adaptations of neurons and prevent overfitting.

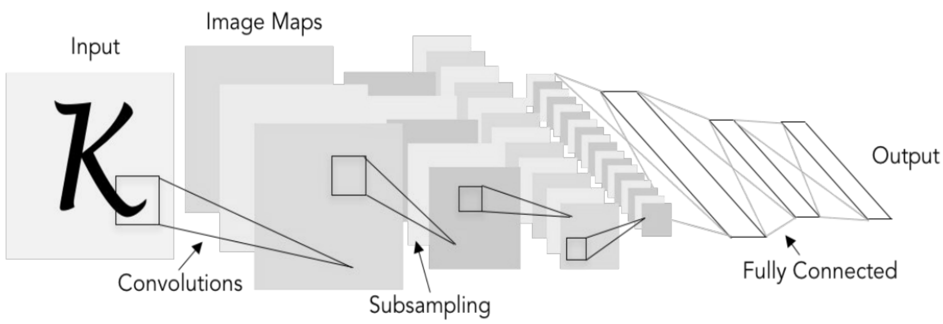

Figure 2. The basic structure of CNN.

IV. EMPIRICAL EXPERIMENTS

The consumer credit data used in the experiment is collected from a Chinese consumer finance company. There are 24,837 customer credit records in dataset. The dataset contains 576 numeric attributes that describe consumer information such as daily payments, micro-credit repayment, credit card payment, current account balance, etc. The dataset includes 2098 un-credit-worthy instances and 22,739 credit worthy instances.

Since the dataset may contain many redundant features that can increase the calculation dramatically and affect the model performance, we first use Relief algorithm to choose the 50 most important features. Then follow the proposed procedure mentioned before, a simple 16×50 pixel matrix (gray image) with a class label is obtained for each consumer. After that a convolutional neural network using the adam optimizer along with following parameter settings is constructed. (Table 1)

Our hybrid model is trained with 12 epochs and the batch size of 128. The dataset is sliced into training and validate sets with the ratio of 70/30. To ensure the reliability of outcome, the data partitioning process is repeated 10 times and we take the average as the final result. Besides, we also compared our model with random forest and logistic regression. In this experiment, Keras, which is a high-level neural networks API, is used to run the convolutional neural network.

TABLE I. PARAMETERS SETTINGS OF RELIEF-CNN

Layer	***Parameters***
Convolutional Layer#1	Filters 32, kernel size 3×3
Convolutional Layer#2	Filters 64, kernel size 3×3
Pooling Layer	Pool size 2×2
Dropout Layer#1	Dropout rate 0.25
Dropout Layer#2	Dropout rate 0.5
Dense Layer	128 units

TABLE II. EXPERIMENTAL RESULTS

Method	***AUC***	***K-S statistic***	***Accuracy***
Relief-CNN	**0.6989**	**0.312**	**0.9164**
Random forest	0.601	0.235	0.914
Logistic regression	0.5221	0.064	0.8581

Table 2 summarizes the results of our experiments. For each model, AUC, K-S statistic and accuracy are displayed. AUC refers to the area under the Receiver Operating

Characteristics curve, and it measures the distinguishing ability of the classification model. The K-S statistic is calculated as the maximum difference between the curves generated by the true positive and false positive rates. It is a commonly used metric of classifier performance in the credit-scoring application domain.

From the results, we can see that our hybrid model Relief-CNN gets a much better AUC value of 0.6989 than other two models, random forest (0.601) and logistic regression (0.5221). For K-S statistic, the Relief-CNN gets a value of 0.312, followed by random forest (0.235) and logistic regression (0.064). It indicates that the hybrid deep learning model can better distinguish the credit-worthy instances from un-credit-worthy instances. Besides AUC and K-S statistic, it can be seen that the accuracy rate of Relief-CNN (91.6%) is slightly higher than random forest (91.4%), and the logistic regression with 85.8%. Based on all the three measures, we conclude that Relief-CNN yields significantly better results than other two benchmark classification techniques.

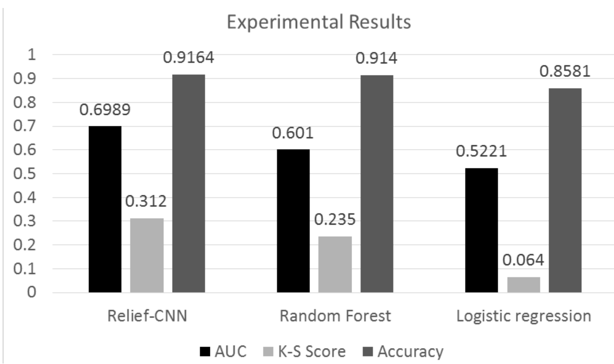

Figure 3. The results of AUC, K-S statistic and accuracy.

V. CONCLUSION

In the consumer finance industry, credit risk now is a necessary factor to succeed. Credit scoring plays an important role in managing credit risk. In this paper, we develop a new tool for credit scoring by proposing a hybrid deep learning model which combines the convolutional neural network (CNN) with Relief algorithm. Our experiments compared the hybrid model with logistic regression and random forest on a real-world dataset from a Chinese consumer finance company. The results clearly show that our hybrid model Relief-CNN is superior to the benchmark algorithms. We strongly believe that deep learning techniques can provide strong support for credit scoring. As a direction of further research, we will explore other ways to reshape data, and work on the structure optimization of the convolutional neural network.

ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (Grant No. 71401115) and funded by Sichuan University (Grant No. skqy201742).

REFERENCES

[1] Lessmann, S., Baesens, B., Seow, H.V., and Thomas, L.C., 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research 247, 1, 124-136.

[2] Crook, J.N., Edelman, D.B., and Thomas, L.C., 2007. Recent developments in consumer credit risk assessment. European Journal of Operational Research 183, 3, 1447-1465.

[3] Hooman, A., Marthandan, G., Wan, F.W.Y., OMID, M., and Karamizadeh, S., 2016. Statistical and data mining methods in credit scoring. Journal of Developing Areas 50 (5), 371-381.

[4] Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., and Alsaadi, F.E., 2016. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11-26.

[5] Lecun, Y., Bottou, L., Bengio, Y., and Haffner, P., 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE 86, 11, 2278-2324.

[6] Durand, D., 1941. Risk Elements in Consumer Instalment Financing. NBER Books.

[7] Bensic, M., Sarlija, N., and Zekic-Susac, M., 2005. Modelling small-business credit scoring by using logistic regression, neural networks and decision trees. Intelligent Systems in Accounting Finance & Management 13, 3, 133–150.

[8] Nie, G., Wei, R., Zhang, L., Tian, Y., and Shi, Y., 2011. Credit card churn forecasting by logistic regression and decision tree. Expert Systems with Applications 38, 12, 15273-15285.

[9] Alaraj, M. and Abbod, M.F., 2016. Classifiers consensus system approach for credit scoring. Knowledge-Based Systems. 104, 89-105.

[10] West, D., 2000. Neural network credit scoring models. Computers & Operations Research 27, 11, 1131-1152.

[11] Bellotti, T. and Crook, J., 2009. Support vector machines for credit scoring and discovery of significant features. Expert Systems with Applications 36, 2, 3302-3308.

[12] Niimi, A., 2015. Deep learning for credit card data analysis. In Proceeding of World Congress on Internet Security, 73-77.

[13] Luo, C., Wu, D., and Wu, D., 2016. A deep learning approach for credit scoring using credit default swaps. Engineering Applications of Artificial Intelligence 65, 465-470.

[14] Tran, K., Duong, T., and Ho, Q., 2017. Credit scoring model: A combination of genetic programming and deep learning. In Proceeding of Future Technologies Conference, 145-149.

[15] Kira, Kenji, Rendell, and A, L., 1992. A practical approach to feature selection. In Proceedings of the Ninth International Conference on Machine learning. 249-256.

[16] Alex, Sutskever, I., and Hinton, G.E., 2012. ImageNet classification with deep convolutional neural networks. In Proceeding of International Conference on Neural Information Processing Systems, 1097-1105.