Credit Scoring with Deep Learning

Conference Paper · February 2018

3 authors:

Lkhagvadorj Munkhdalai
Chungbuk National University

Oyun-Erdene Namsrai
National University of Mongolia

Keun Ho Ryu
Chungbuk National University

Credit Scoring with Deep Learning

Lkhagvadorj Munkhdalai[1], Oyun-Erdene Namsrai[2], Keun Ho Ryu[1*]
[1] Database and Bioinformatics Laboratory, School of Electrical and Computer Engineering,
Chungbuk National University, South Korea
[1] {lhagii, khryu}@dblab.chungbuk.ac.kr
[2] National University of Mongolia, Mongolia
[2] oyunerdene@num.edu.mn

Abstract - **Deep learning is the most accurate state-of-the-arts approach in artificial intelligence. Many supervised and unsupervised problems have been successfully resolved by deep learning. Credit scoring is one of the important challenges to the decision-making process for the lending institutions. This paper applies the deep learning approach for credit scoring based on the most common four datasets. We determine an optimal number of nodes in hidden layers, an effective batch size and epoch number using the grid search method. The results from this study show that our proposed algorithm performs very well in a credit scoring application.**

Keywords-deep learning; grid search; credit scoring;

I. INTRODUCTION

In recent years, many types of internet-based financial service have been growing rapidly. For instance, a customer can take deposit and lending using the internet and without any human involvement. Therefore, it is imperative to develop a credit scoring model in constantly changing financial and operating environments. Although many machine learning methods have been investigated the credit scoring problem over previous years. Unfortunately, very few studies have been conducted based on the deep learning approach [1-3]. The aim of this paper is to improve the accuracy of credit scoring model based on deep learning using most common datasets for credit scoring.

Deep learning is an innovative upgrade of the artificial neural network approach and there are several types of deep learning architecture such as a multilayer neural networks, a convolutional network, and a recurrent neural network [4]. Figure 1 shows the two hidden layers neural network architecture.

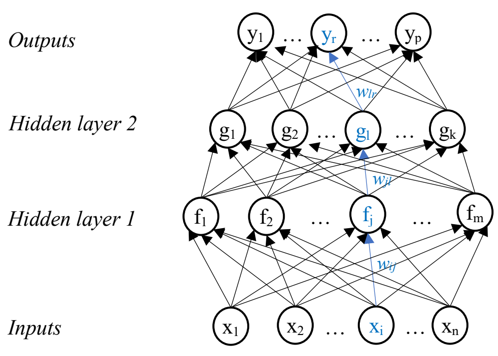

Figure 1. Two hidden layers neural network. The *n* inputs, *m* and *k* nodes of each hidden layer and *p* outputs.

We apply one of deep learning architecture as known as the multilayer neural network to credit scoring problem [5]. Moreover, a grid search algorithm is used to define an optimal number of nodes in each hidden layer. Our proposed algorithm demonstrates promising result compared with the recent highest-performing papers which used mutual datasets [6, 7].

This paper consists five sections. Section II briefly introduces the related work. Section III describes our proposed algorithms. Section IV shows datasets and experimental results from the proposed algorithm. Section V concludes this research and gives possible future research idea in this field.

II. RELATED WORK

This section presents a comparative analysis of the literature on deep learning and other recent models with the highest performance in credit scoring.

Five neural network architectures were investigated for credit scoring in [8]. Namely, the traditional MLP network, a mixture of experts (MOE), radial basis function (RBF), learning vector quantization (LVQ), and fuzzy adaptive resonance (FAR). They utilized the most common datasets such as Australia and German. MOE architecture gave the promising results which an error of German is equal 0.2434 (accuracy - 0.7566) and error of Australia is equal 0.1332 (accuracy – 0.8668) compared to other architectures.

Deep learning with random forest feature importance approach was proposed for credit scoring by [3]. This paper also evaluated their method on German (accuracy – 0.7468) and Australia (accuracy – 0.8624) datasets. Recent proposals for credit scoring with deep learning studies used another dataset [1, 2]. The first paper trained on credit default swaps dataset using MLP architecture and showed promising result [1]. The second paper considered the multi-period mortgage risk based on five hidden layers deep learning architecture [2].

[6, 7] studies achieved the highest performance mutual datasets in this study. These studies introduced ensemble classification approaches such as classifiers consensus system and boosted decision tree for credit scoring. The classifiers consensus system gave the highest accuracy at 0.7772, 0.8798 and 0.8788 for German, Australia and Japanese datasets respectively. The boosted decision tree also showed the good performance such as accuracy at 0.7734, 0.8792 and 0.6936 for German, Australia and Taiwan datasets respectively.

III. OUR PROPOSED ALGORITHM

The first deep learning that has been utilized is multilayer neural network (MLP) [9]. We do selection because MLPs are pass information forth to output by using feedforward links. The loss function correlates to the number of links (weight parameters). Accordingly, the performance of deep learning model depends on the loss function and optimization method. If the number of hidden layers and its nodes are large, our model will overfit. Especially, it is more improper to learning from given small dataset [10]. Therefore, we use a grid search with cross-validation in order to define the optimal number of nodes in three hidden layers neural network for each dataset.

Figure 2 indicates the framework of our proposed algorithm to build the model. Furthermore, it is shown in algorithm I. It consists two stages. First stage, according to mentioned above we define the optimal number of nodes in each hidden layer. Second stage, in order to improve the performance of our model we search effective batch size and epoch number using grid search based on the selected networks [11].

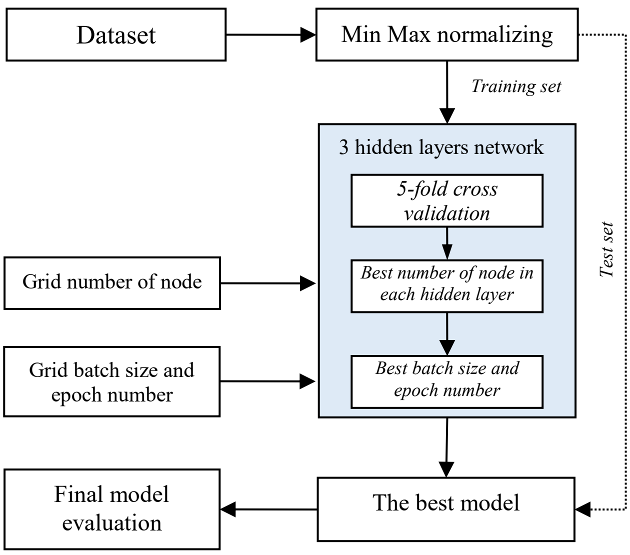

Figure 2. The framework of our proposed algorithm

Finally, we evaluate our constructed model using accuracy and area under the Receiver operating characteristic curve (AUC).

Algorithm I. Proposed algorithm for credit scoring with deep learning.

Require: Training set $\{x_i, y_i\}_{i=1}^{n}$
Require: Range of nodes *{N}*
Require: Initial batch size and epoch number $\{B_0, P_0\}$
Require: Range of batch size and epoch number $\{B, P\}$
for *i ←1 to N* ***do***
  for *j ←1 to N* ***do***
    for *k ←1* to N **do**
      $Y_{train} \leftarrow sigm\left(\theta^{N_k}, sigm\left(\theta^{N_j}, sigm(\theta^{N_i})\right)\right)$
      {3 hidden layers neural network using sigmoid function, batch size B_0, and epoch number P_0}
      Store N_i', N_j', N_k' with the best accuracy *{5-fold cross validation}*
    end for
  end for
end for
for *m ←1 to K* ***do***
  for *n ←1 to K* ***do***
    $Y_{train} \leftarrow sigm\left(\theta^{N_k'}, sigm\left(\theta^{N_j'}, sigm(\theta^{N_i'})\right)\right)$ with B_m, and P_n
    Store B_m', P_n' with the best accuracy *{5-fold cross validation}*
  end for
end for
return $\{N_i', N_j', N_k'$ *and* $B_m', P_n'\}$

IV. EXPERIMENTAL RESULT

Dataset and Experimental setup are described in the first part of this section. Second part demonstrated experimental results and performance comparison of other recent studies with higher performance.

A. Dataset and Experimental setup

We select most common four datasets in order to evaluate the performance of our proposed algorithm against previous studies. These datasets retrieved from the UCI repository [12], namely German, Australia, Japanese and Taiwan.

TABLE I. DESCRIPTION OF DATASETS

Dataset	Samples	Features	Good/Bad	Training	Test
German	1000	21	700/300	700	300
Australian	690	15	307/383	483	207
Japanese	690	16	307/383	483	207
Taiwan	6000	23	3000/3000	4200	1800

As shown in Table I, the sample size of all the datasets is small. Deep learning approach works much better on the bigger dataset. However, we trained our proposed algorithm using these datasets. Perhaps if there will be available public dataset with numerous samples for credit scoring, our proposed algorithm works very well.

In experimental setup, we applied 3 hidden layers neural network for credit scoring. This case, grid search space of node's number is chosen by [4, 8, 16, 32, 64] in each hidden layer. This step also needs fixed batch size (32) and epoch number (200).

Thereafter, we searched the effective batch size and epoch number from [1, 16, 32, 64] and [150, 200,250, 300], respectively. All experiments for this study are performed in Python language using the 'Keras' deep learning library [13] and parameter optimizing method is used Adam optimizer [14].

B. Experiment results and performance comparison

This section presented the result of this study step by step. Also, we compared our result with other recent studies.

Firstly, the selected best number of nodes, batch size and epoch are shown in Table II.

TABLE II.THE OPTIMAL NUMBER OF NODES, BATCH SIZE AND EPOCH NUMBER.

Dataset	Nodes 1	Nodes 2	Nodes 3	Batch size	Epoch
German	16	64	8	64	250
Australian	32	64	16	32	150
Japanese	16	32	4	16	150
Taiwan	32	4	4	8	200

The grid search algorithm provided the 125 results with a different number of nodes for each dataset as shown in Figure 3. We chose the number of nodes which with the highest accuracy and then taken into next step. The effective batch size and epoch number were selected in this step based on the optimal number of nodes.

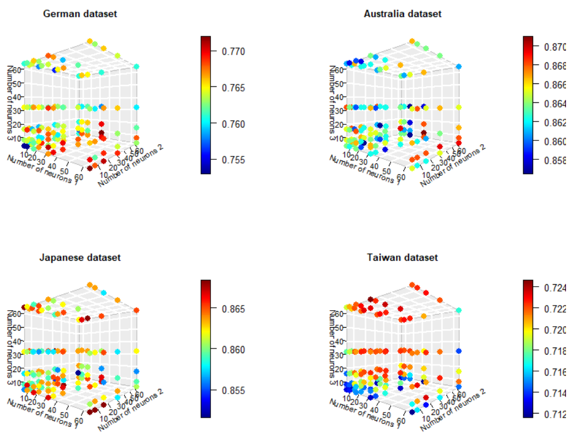

Figure 3. The result of grid search algorithm. The accuracy used to color the graphics.

Finally, we evaluated the built model using the optimal number of nodes, batch size, and epoch number based on the test dataset. The experiments are looped 5 times to the evaluation criteria for these experiments are averaged to compare with other studies.

Table III showed the results of our model and performance comparison.

TABLE III. THE RESULTS AND PERFORMANCE COMPARISON.

No	Year	Study	Dataset	Accuracy	AUC	Splitting technique
1	2000	[8]	German	0.7566	-	10-fold cross validation
			Australian	0.8668	-	
2	2016	[3]	German	0.7468	-	5-fold cross validation
			Australian	0.8624	-	
3	2016	[6]	German	**0.7772**	**0.8023**	5-fold cross validation
			Australian	**0.8798**	**0.9328**	
			Japanese	**0.8788**	**0.9328**	
4	2017	[7]	German	0.7734	-	100 times looped (random sampling)
			Australian	0.8792	-	
			Taiwan	0.6936	-	
5	Our algorithm		German	0.7710	0.7811	5 times looped (random sampling)
			Australian	0.8681	0.9289	
			Japanese	0.8653	0.9226	
			Taiwan	**0.6990**	**0.7227**	

For German, Australia and Japanese datasets, the deep learning model achieved the accuracy 0.771, 0.8681, and 0.8653, respectively. Although our proposed algorithm showed better performance for these datasets [8, 3], it was weak compared to the performance of [6, 7]. It is probably due to the small sample size.

For Taiwan dataset, the deep learning model showed the best performance. Taiwan dataset contains 6000 instances and the sample size is much bigger than other datasets. For this reason, the model outperformed the best previous model by 0.22% [7].

However, many supervised classification models show very well performed result in credit scoring application. This paper provided the deep learning with bigger data sample can demonstrate better performance than other models.

V. CONCLUSION

This paper applied deep learning approach with grid search method for credit scoring. We trained 4 different datasets on 3-hidden layers neural network. Our proposed algorithm demonstrated promising results. Especially, bigger sample dataset was given the best performance by deep learning. To this end, the authors anticipate potential future work in this area that includes developing deep learning model for credit scoring based on the bigger dataset.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2017-2013-0-00881) supervised by the IITP (Institute for Information & communications Technology Promotion) and supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (No.2017R1A2B4010826).

REFERENCES

[1] Luo, Cuicui, Desheng Wu, and Dexiang Wu. "A deep learning approach for credit scoring using credit default swaps." Engineering Applications of Artificial Intelligence (2016).

[2] Giesecke, Kay, J. Sirignano, and A. Sadhwani. Deep learning for mortgage risk. Working paper, Stanford University, 2016.

[3] Van-Sang, Ha, and Nguyen Ha-Nam. "Credit scoring with a feature selection approach based deep learning." MATEC Web of Conferences. Vol. 54. EDP Sciences, 2016.

[4] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. "Deep learning." Nature 521.7553 (2015): 436-444.

[5] Haykin, Simon. Neural networks: a comprehensive foundation. Prentice Hall PTR, 1994.

[6] Ala'raj, Maher, and Maysam F. Abbod. "Classifiers consensus system approach for credit scoring." Knowledge-Based Systems 104 (2016): 89-105.

[7] Xia, Yufei, et al. "A boosted decision tree approach using Bayesian hyper-parameter optimization for credit scoring." Expert Systems with Applications 78 (2017): 225-241.

[8] West, David. "Neural network credit scoring models." Computers & Operations Research 27.11 (2000): 1131-1152.

[9] Haykin, Simon S., et al. Neural networks and learning machines. Vol. 3. Upper Saddle River, NJ, USA:: Pearson, 2009.

[10] Ngiam, Jiquan, et al. "On optimization methods for deep learning." Proceedings of the 28th international conference on machine learning (ICML-11). 2011.

[11] Keskar, Nitish Shirish, et al. "On large-batch training for deep learning: Generalization gap and sharp minima." arXiv preprint arXiv:1609.04836 (2016).

[12] Asuncion, Arthur, and David Newman. "UCI machine learning repository." (2007).

[13] Chollet, François. "Keras: Deep learning library for theano and tensorflow." URL: https://keras. io/k (2015).

[14] Kingma, Diederik, and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980(2014).