Credit Card Fraud Detection using Autoencoders

Anurag Mitra[1], Mukul Siddhant[2]

BCA Students, Department of CSE, Galgotias University, Greater Noida (UP), INDIA

Dr.Gururama Senthilvel. P. Associate Professor, Department of CSE, Galgotias University, Greater Noida (UP), INDIA

Abstract- In today's life or economy credit card plays a veryimportant role. Credit card becomes a necessary part of business, household and bank transactions. Using of credit card carefully and responsibly gives a enormous benefit to the user, fraudulent activities happen and give financial damage to the user or card holder. The growth of E-commerce industry led to use of credit card or many platform for online purchase or different transaction because of this the fraudulent activities was also increased. Bank facing many issues for detecting the credit card fraudulent transactions. For finding the fraud of credit card machine learning plays a vital role. For predicting these fraud detection of credit card we use many machine learning methods or algorithms, past data we collect and analyze it and make a machine learning model to detect the fraudulent activities which going to happen. The performance of fraud detecting in credit card transaction is greatly affected by sampling approach on data-set, selection of variable and detection technique used. This project objective to use of efficient approach to detect automatic fraud related to banks or insurance company using deep learning algorithm called autoencoder. We are using the European card holder real-time dataset of September 2013. The data was unbalanced in the dataset so for this autoencoder is perfect to provide the accurate results. We can reconstruct the normal data through the autoencoder and anomalies was detected at the time of reconstruction error threshold and consider the anomalies.

Keyword: fraud detection, unsupervised learning, autoencoder, credit card

1. Introduction

Without the information of the owner of credit card, use of credit card comes under credit card fraudulent. Credit card fraud comes under two groups one was behavioral and the second one was application. Application fraud are going to happen when the new card issuing by banks and the fraudster using fraud documents. Multiple application was submitted by fraudster with the one user details only.

The four primary principal of behavioral fraud was mail burglary, stole/lost card, card holder not present and fake card. Mail burglary fraud occur only when the credit card fraudster get the credit card details through mail before receiving the physical credit card to the real owner. Stolen/lost credit card fraud occur only when the fraudster stole the credit card to the owner and gain admittance to a lost card. In both card holder not present and fake card fraud occur only when the credit card is acquired by the fraudster without the knowledge the card holder. In today's world the exchanges directly utilizing the credit card through the mail, phone or the web. As for the reference the high fraud of the credit card country is Ukraine is the top country in terms of credit card fraud with 19% and with 18.3% Indonesia have credit card fraud rate. Yugoslavia have the 17.8% and the most dangerous country in terms of fraud.

2. Literature Survey

The two general classification are grouped for credit card fraud recognition strategies: Client behavior analysis and fraud analysis.

The task at transaction level of primary gathering of procedures manages supervised classification. In these strategies, exchanges are marked as fake or ordinary in light of past authentic data. The dataset is used in the paper to make the classification models which help to see the state (normal or fraud) new records.

To copy the working of human mind the artificial neural network is a collection of connected nodes. Each node has some different node with weighted association in a adjoining layer. Single nodes consider the input and get the associated node and usage the nodes along with a basic figures to capacity output values.

Inspired by natural growth, The John Holland were initially presented the genetic algorithm (GA). GA finds the best solutions with a populace of applicant solutions that are customarily addressed as binary strings called chromosomes. A Hidden Markov Model is a twofold em-had relation with a stochastic cycle that applied on demonstrate considerably more convoluted stochastic cycles is compared with a traditional Markov model. The hidden states have the framework which is hidden in Markov cycle. In less complex Markov models, states with pure change chances are just unclear parameters.

A supervised learning model have the support vector machines with a learning algorithm which can analyze and identify the patterns for regression and classification tasks. The important thought of support vector machine was to

Catch an optimal hyperplane that can isolate instances of having two classes, linearly. The gap between a few minimal instance hyper-plane thought to be located are called support vectors.

Conditional dependencies was Bayesian network which is a graphical model among arbitrary variables. A coordinated acyclic graph is a fundamental graphical model. For finding the unknown probabilities the Bayesian network was very valuable from the sight of ambiguity.

3. Autoencoder

Autoencoder is the type of neural network. With the help of numerous layers the neural network have been made, the main part of autoencoder is the as much data input layer contain the same data contain the output layer. The almost same number of units have precisely the purpose are the input layer and output layer.

The autoencoder plans to imitate the input data. It constructs the output of duplicate data after analyzing it and reconstruct in unsupervised style.

By going through the network, the autoencoder remade by each dimension of networks. It may be sometimes it appear unimportant to involve in the neural networks, hence in replication cycle the size of the input data diminished in a smaller illustration. When the middle layer of the neural network contrasted with input and output layer the have a lots of number units. Accordingly, the diminished representation of the input was grip by middle layer. The reduced representation of the input is remade from the output.

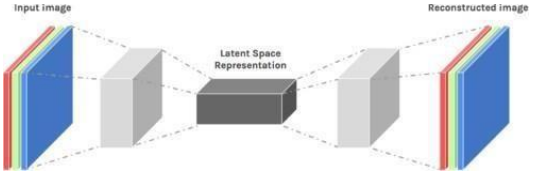

Fig 1. Autoencoders

The autoencoder have two main parts that are encoder and decoder, that is defined in terms of transition ϕ and ψ, such that:

$\phi: X \rightarrow F$

$\psi: F \rightarrow X$

$\phi, \psi = arg \min_{\phi,\psi} ||X - (\psi \circ \phi)||^2$

We have to set 4 hyper-parameter prior to training an autoencoder:

- **Code size**: Code size address the nodes which is present in the middle layer. More compression bring by more modest size.
- **Number of layers:** Autoencoder comprise as much layers as we need.
- **Nodes per layer:** It increases in each layer of decoder and reduced in each layer of encoder.
- **Loss function:** We can utilize the binary cross entropy, or the mean squared error. From the range [0 1] the input values, we can utilize the cross entropy and mean squared error.

In autoencoder there are two stages of training: fine-tuning network weights or unsupervised learning. The activation function is the main factor of the first stage.

We pick a blend of relu and elu beginning functions. In forward engendering is useful on every input and working for output. Afterwards the deviation of x and x will be intended. On the last stage, the error will be back propogated from a network for updating weights. For upgrading network stages normally gradient descent algorithm and standard learning methods for changing the each layer are used. However, slowest optimizers are in this algorithm. Here we utilized AdamOptimizer which is utilized for the most part in deep learning. It utilizes parameters (momentum) from the moving averages which assists Adam algorithm with utilizing greater effective steps with out any need of fine-tuning. High computational cost is the main disadvantages of this algorithm.

4. Proposed Method

The dataset has 284807 transaction and out of these 492 transaction was fraudulent and rest all are actual transaction. Original feature not in hand for privacy of the user and it contains the 29 features which shows the PCA mapping function and add two uncharted features which is named as time and amount of thetransaction.

The aim of the research to find the fraudulent transaction from a actual transaction in an unsupervised manner. We train our autoencoder model setting by parameter in Table 1:

name	values
Number of neurons	Input Layer:29 Hidden Layer 1st:14 Hidden Layer 2nd:7 Hidden Layer 3rd:7 Hidden Layer 4th:14 Output Layer:29
Learning rate	0.1
Optimizer	AdamOptimizer
Number of epochs	100
Activation functions	ReLU, tanh

Table 1. Autoencoder Parameters

After training the autoencoder, the model will be saved and loaded, the autoencoder takes incoming transaction and gives the output.

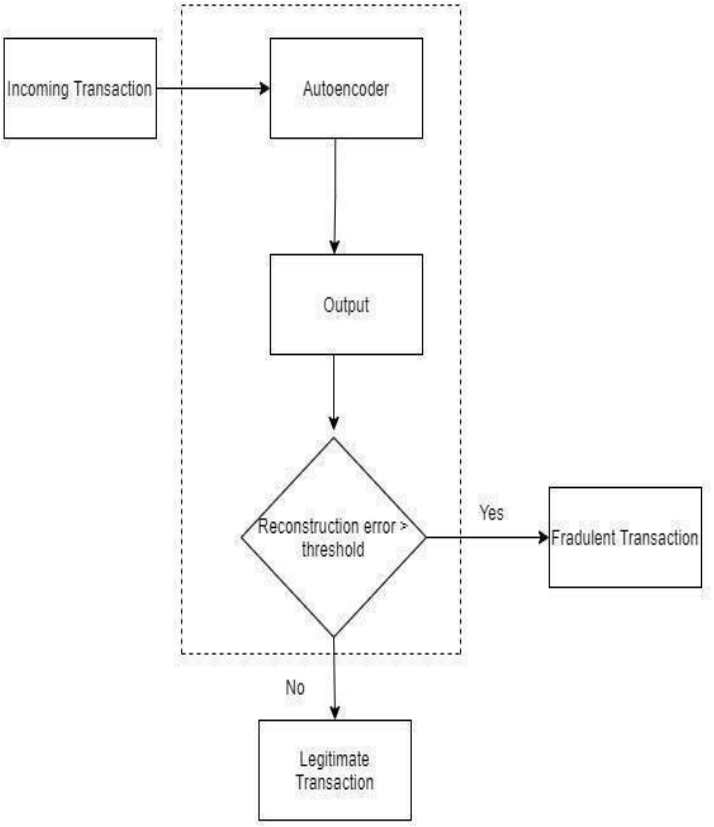

Fig 2.System Architecture

6. Evaluation Metrics

There are many measures for many algorithm and those measures are developed to calculate different things. So, there should be limitations for assessment of various proposed methods. True positive (TP), True negative (TN), False positive (FP) and False negative (FN) the connection between them are very much accurate so usually researchers frequently adopted for fraud detection of credit card to compare the accuracy.

Measures	Formula
Accuracy (ACC)/Detection rate	TN + TP/TP + FP + FN + TN
Precision/Hit rate	TP/TP + FP
True positive rate/Sensitivity	TP/TP + FN
True negative rate /Specificity	TN/TN + FP
False positive rate (FPR)	FP/FP+TN
ROC	False positive rate was opposite of true negative rate
Cost	Cost = 100 * FN + 10 * (FP +TP)
F1-measure	2 × (Precision ×Recall)/(Precision +Recall)

Table 2. Network Evaluation metrics

In data mining for measuringabove metrices. We use confusion matrix fig.3.

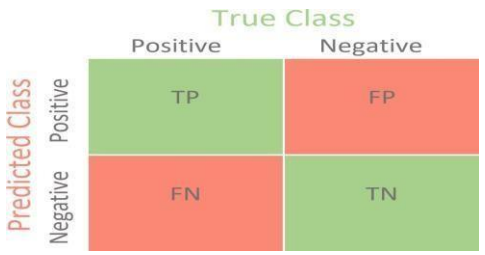

Fig 3.Confusion Matrix

6. Results

According to Fig. 3 after 100 epochs running a network, you can see the training loss and validation loss.

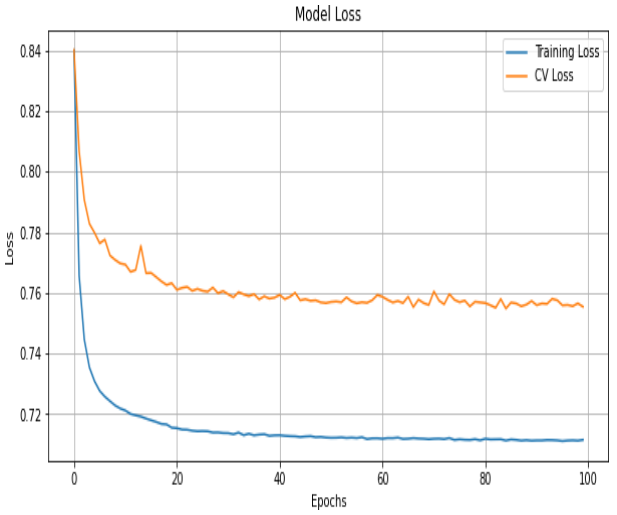

Fig 4. Loss Values

On table.3 the reconstruction error on testing data is shown. The fraudulent transaction was 97 and actual transaction was 56864.

	0	1
count	56864.0	97.0
mean	0.706328	31.63478
std	2.575618	46.635970
min	0.061731	0.185452
25%	0.273163	4.269731
50%	0.424408	11.268653
75%	0.640054	52.293395
max	157.272851	264.172610

Table 3. Comparison of Reconstruction error for fraud and non-fraud transactions

The confusion matrix of threshold 2 and 3.1 was shown is fig5 and fig.6 and classification table was shown in table.4 and table.5.

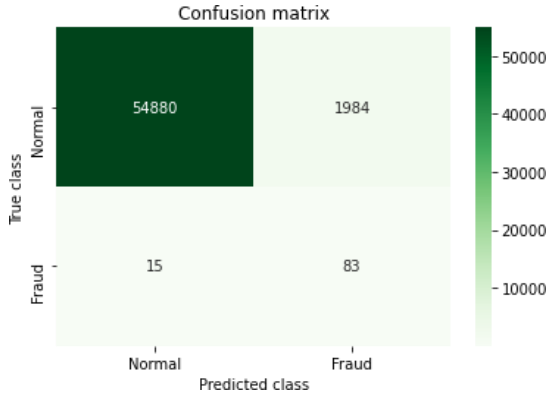

Fig 5. Threshold=2 from confusion matrix

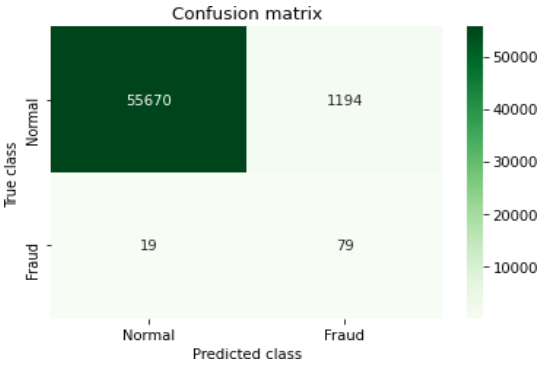

Fig 6. Threshold=3.1 from confusion matrix

	precision	recall	f1-score	support
0	1.00	0.98	0.97	56764
1	0.05	0.86	0.08	98
accuracy			0.97	56862
macro avg.	0.52	0.91	0.54	56862
weighted ave.	1.00	0.97	0.98	56862

Table 4. Classification Report for Threshold=2

Recall has the power to find the relevant cases in other words the number of true positive divided by true positive number added false negative numbers.

Precision informs that true positive was divided by true positive and added the false positive numbers.

	precision	recall	f1-score	support
0	1.00	0.98	0.97	56764
1	0.05	0.86	0.11	97
accuracy			0.97	56862
macro avg	0.55	0.90	0.55	56862
weighted avg	1.00	0.97	0.98	56862

Table 5. Classification Report for Threshold=3.1

7. Conclusion

In this paper we are using a real time model with real dataset for credit card fraud detection. We are using autoencoder which comes in neural network. Autoencoder gives very accurate result with F1 score. Hence, future studies will focus

On deep learning for the real time dataset classification problems. This model help credit card holder to detect the unusual behavior or the fraud on their transaction or credit card details.

References

1. KhyatiChaudhary, JyotiYadav, BhawnaMallick, “A review of Fraud Detection Techniques: Credit Card”, Volume 45– No.1 2012.

2. Linda Delamaire, Hussein Abdou, John Pointon, “Credit card fraud and detection techniques: a review”, Volume 4, Issue 2, 2009.

3. S. Ghosh and D. L. Reilly, “Credit card fraud detection with a neural-network”, 1994.IEEE Computer Society Press

4. Holland, J. H. “Adaptation in natural and artificial systems.” Ann Arbor: The University of Michigan Press. (1975).