CREDIT CARD FRAUD DETECTION USING HYBRID DEEP NEURAL NETWORK

PRAKHAR PATIDAR, SAHIL PARYANI, PRANJAL SONGARA

Abstract: As we are entering the digital world the use of plastic currency and online transactions is increasing, hence the credit card application data of customers is accumulating day by day as well as the ratio of fraudulent applications which are passing by undetected. Application fraud is increasing significantly with the development of modern technology, resulting in the loss of billions of dollars worldwide each year. The purpose of this paper is to detect credit card fraud by applying deep neural networks. Using self-organizing maps (SOM) combined with supervised learning to find optimal parameters, the network is trained to reach stability and be optimal so that an appropriate model can be found to detect whether the application submitted is valid or fraud. This problem of detecting fraud application can be regarded as a classification problem. The optimal network to perform classification on credit card data set is explored and implemented in an open source deep learning library, Keras which is based on top of two libraries, namely TensorFlow and Theano. The primary goal and contribution of this review paper are to present the overview of deep learning and how it works and provide a view on some deep learning algorithms and provide a salutary addition to it that can be used to achieve better fraud detection.

Keywords: Application fraud, neural network, deep learning, SOM, fraud detection

I. INTRODUCTION

The advent of new technologies as Automated Teller Machines (ATMs) and credit card systems have amplified the amount of fraud loss for many banks. Performing the analysis for all of the transaction data manually is literally impossible, also analyzing every transaction is legitimate or not could be very expensive, automation might be solution to this problem. Fraud detection is a continuously evolving discipline. Whenever criminals detect one detection method that's in place, they will adapt their strategies and try others. Many major banks have traditionally relied on old rules-based expert systems to catch fraud, but these systems have proved all too easy to beat; the financial services industry is relying on increasing complex fraud detection algorithms and are trying to improve their systems. The Nilson Report shows that U.S. card fraud (credit, debt, etc.) was reportedly $8 billion in 2016 and expected to increase to $12 billion by 2020. An ideal fraud detection system should be able to identify the fraud transaction accurately and should make the detection possible in real-time transactions. Many in the financial services industry have updated their fraud detection and are shifting to Artificial Intelligence to get better solution for their problems. Financial companies, such as PayPal, have already started to modify their algorithms for fraud detection so that they would be more accurate with the addition of more advance Artificial Intelligence techniques.

Machine learning is one subfield of artificial intelligence, which presents human intelligence by machines. Deep learning as one branch of machine learning which has risen as one of the most popular research hotspots currently. In this paper we are focusing on Artificial Neural Networks (ANN), that is a major area based on deep learning. ANN are the computational imitations of the biological human brain neural structure. ANN, like the brain, is flexible and self-learning system that can be used to solve complex problems that are difficult when described using a mathematical model, such as pattern recognition and classification, function approximation and control.

II. LITERATURE SURVEY

Credit card frauds happen on a daily basis and are increasing as we are progressing towards the digital age .A lot of losses are incurred due to these fraudulent transactions from invalid accounts .Nowadays credit cards have been rendered ubiquitous ,they are used

everywhere for example in shopping ,movies ,petrol pump ,ticket booking and in all possible stores .As the credit cards are used on a daily basis fraudsters use these large data of normal transactions and cleverly hide the fraudulent transactions in between ,so that they are hard to find by any person .For now in this paper we are going to focus on filtering out fraudulent applications that people fill in by adding wrong details that seems normal and gets them a credit card they can exploit and never return the money .As these behavioral nuances can be caught by human eye but is a exhaustive process with high chances of error , a trained artificial neural network that has objectivity ,stability and trained through a dataset full of these applications that were found in the past would be a much better option in finding a fraud application among normal ones in the future .In the last decade a lot of work has been done in order to filter out these fraud applications by keeping track of what transactions are happening those accounts .Many methods that have been showcased in the previous years to filter out frauds ,we will review some of them briefly over here.

Firstly ,the K-nearest neighbor (KNN) method uses algorithms that are based on supervised learning and filter out fraud transactions by first calculating a nearest point and then if any transaction occurs near that particular point then it is classified as a fraud transaction .The nearest point can be understood by keeping it as the point of abnormality which is carefully calculated by KNN .KNN algorithm is many times confused with another algorithm by the name of K-means ,here k stands for clustering and this method is a unsupervised learning algorithm and works by first filtering out the data into groups before performing any other actions ,K-means tries to recognize patterns after the data group division is done .Whereas KNN calculates a nearest point based on the data of the previous transactions fed into it .

Decision tree induction is a supervised learning method where we create a decision tree to reach to a particular solution. A decision tree contains some internal nodes and every node represents a test on a particular attribute and all branches in decision tree represent an outcome of test and each leaf node will represent output. For a new transaction with unknown class it is passed down the decision tree from root node till output node and transaction values are checked against the nodes of the decision tree .

The Hidden Markov Model (HMM) mechanism is a double embedded stochastic process .here if any input transaction is not accepted by the system with enough probability then it would be considered as a fraudulent transaction .In this model the fraud detection system (FDS) is first trained with the normal spending pattern of a credit card holder .every transaction is then is checked in the FDS and labelled genuine or fraud .

Another algorithm that we come across is outlier detection which can be implemented through both supervised and unsupervised learning .Supervised outlier detection uses a training dataset to classify the given input outlier whereas unsupervised outlier detection is similar to clustering .It is concluded that unsupervised outlier is more favorable as this algorithm learns to discriminate between fraud and normal transactions on its own by studying a large array of normal transactions and doesn't require previously defined rules to make the classification .

The current research is recognizing deep learning and neural network models as the hot topic for gaining the solution for solving these type of classification problems, such as the credit card fraud detection.

III. DEEP LEARNING

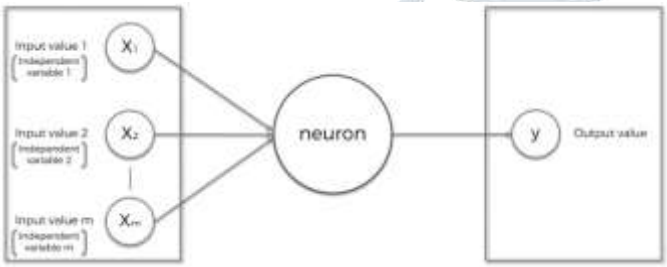
Artificial Intelligence is the science of getting computers to act without being explicitly programmed. Machine learning system, a subset of AI, builds the learning model that effectively "learns" how to estimate from a particular enlisted training data and then uses that learnt model to apply prediction analysis on the new data in future. In this new era, Machine learning is mostly in use to demonstrate the promise of producing consistently accurate estimates without needing multiple manual edits to the program. The two types of models in machine learning are basically, the supervised learning model and the unsupervised learning model. An Hybrid model is the one which combines both supervised and unsupervised model to form a specific model to achieve a distinct function.

Deep learning can be considered as a subset or a deeper level within the machine learning domain. The machine learning models do become progressively better at whatever their function is, but they still need some guidance whenever it returns an inaccurate prediction. But with a deep learning model, the algorithms can determine on their own if a prediction is accurate or not using a hidden layer of intelligence developed by itself by learning from previous data. To achieve this and to get a computer to function like a brain we make the use of artificial neural networks (ANN).

Artificial Neuron

An artificial neuron is a digital construct of biological neurons created using mathematical function. These neurons are the elementary units on which an Artificial Neural Network works. An artificial neuron is composed of a set of weighted inputs, along with a

transition function and an activation function. The neuron function receives one or more weighted inputs and sums them to produce an output. These neurons are connected in layers to form a system, these connections are characterised as synaptic weights. The synaptic weights change as new data is received and processed, and this is how learning occurs. Each neuron has two modes of operation; the training mode and the usage mode. During training mode, the neuron are trained to for a particular set of input patterns. During usage mode, when the taught input pattern is detected, it generates the associated output. As number of signals passing through different number of layers increases, the artificial brain extracts more information until it can distinguish each input being passed. This process of fine tuning the system is known as deep learning



Deep Neural Network

An Artificial Neural Network is an information processing paradigm which is inspired by the way biological nervous systems, such as how the brain processes an information. It comes under the category of supervised learning models. The key element of this network is the novel structure of the information processing system. Neural networks consist of a series of algorithms which are modeled after the human brain, which constantly try to recognize patterns and categorize and classify information. A number os neural nodes are connected to each other forming a single layer of computational network. Whenever we provide such network with inputs, the data is paased through a hidden layer which tries to learn and recognize the pattern so that when new data arrives it can accurately classify the data in the output.

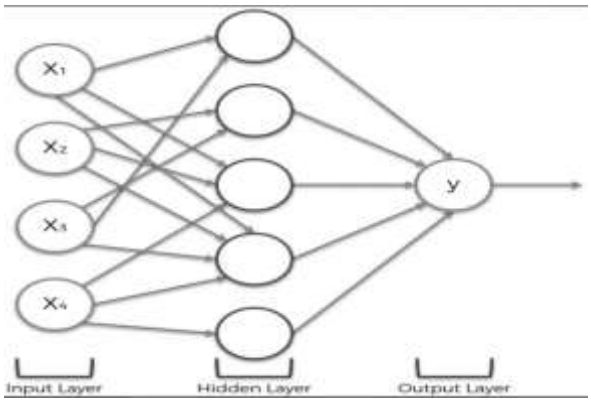


Fig: Single hidden layer neural network

When deep learning methods are used in unison with the neural network, the number of layers increase which increases the system's capacity to learn and this concept of layered learning is sometimes referred to just as "Deep Neural Network", referring to the number of different layers involved. The different layers of such a system could be seen as a nested hierarchy of related concepts or decision trees. Deep neural network systems need large quantities of data in order to be trained as the systems learns from exposure to huge number of data points. Google Brain learning to recognize cats after being shown over ten million images can be seen as an early example of this.

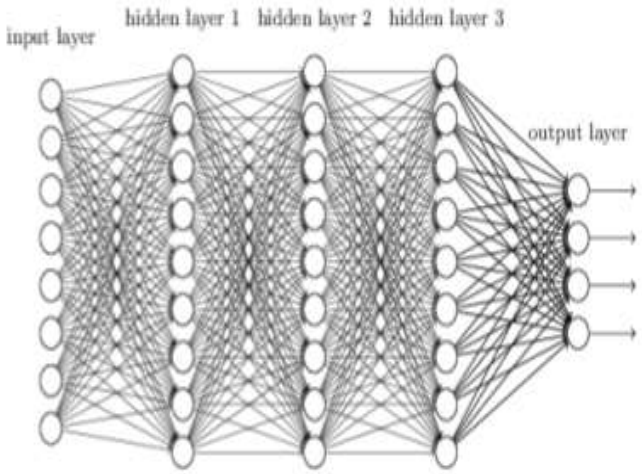
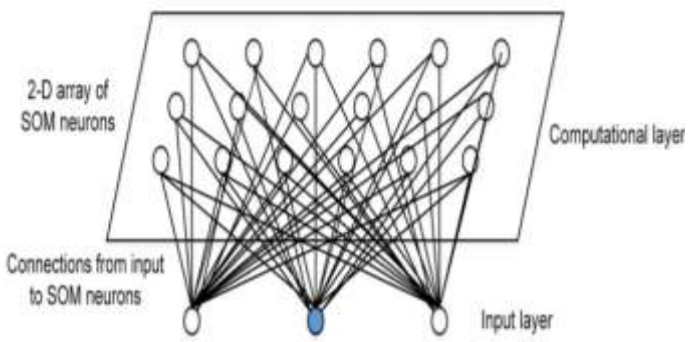


Fig: Multi- layered Deep Neural Network

Self-Organizing Maps (SOM)

The self-organizing maps come under the category of unsupervised learning model. In such models the networks learn from their own classifications of training data without external help. In a particular class of unsupervised learning, the output neurons compete amongst themselves to be activated, with the result that only one is activated at any one time, such class is known as competitive learning. The

The self- organizing maps are a sheet like artificial neural network, the cells of which become specifically tuned to various input signal patterns or classes of patterns through an unsupervised learning process. A self-organizing map configures its neuron according to the topological structure of input data, this process iteratively tuned the weight of neurons and in result produces clustering of input data.



(Taken from [12])

Most applications use two-dimensional and rectangular grid, many applications also use hexagonal grids, and some one, three or more dimensional spaces. SOMs produce low-dimensional projection images of high-dimensional data distributions while preserving the similarity relations between the data items.

IV. PROPOSED METHOD

In this paper we aim to enhance the already existing deep learning model of the self-organizing maps (SOM) by introducing a supervised learning model with it to get better organized results. The SOM technique is already a successfully proved technique to find credit card fraud but the main problem with it is that its results are a little ambiguous and unorganized. We focus on trying to improve these limitations by combining the SOM unsupervised learning model with a supervised learning neural network model which could find the probability ranking of fraud for each customer by using the output of SOM as its input.

For analyzing fraud application, first the SOM technique is applied on the credit card application dataset which contains all credit card details of the different customer in the bank. We get a map of neural nodes, i.e. neurons, some of the nodes are the normal nodes that follow the general rule and are non-potential frauds and a part of them are expected to form a cluster which contains potential frauds. The cluster contains the outline neurons which we choose to select to work with. The outline neuron is characterized by high MID (Mean inter-neuron distance) and it contains both category of customers, the ones who got their application approved and the ones who didn't. We ideally assume that the widest threshold neurons are the best option to go with for getting the most accurate list. Hence, the coordinates of these outline neurons are used to get list of potential fraud applicant.

Usually this list is forwarded to the bank authorities for further investigation, but it still has some deficiencies as the list is not sorted and the investigators do not know which one to prioritize. Hence, we suggest an augmentation of a supervised learning model which will allow us to find probabilities of fraud for each customer and would organize them in form of rankings so that it is easy for the investigators to choose the order in which they need to explore.

The main problem here is to switch from unsupervised learning model to a supervised learning model as the supervised learning requires a dependent variable which is a binary variable that is needed the train the model and no such variable is present in the unsupervised learning model. To solve this, we propose an approach in which we use the results of SOM as the input for supervised learning model. To do so we make use of a matrix of features which contain all columns of relevant information from the dataset. Now, to get value for dependent variable we first initialize a vector function which contains a negative value for each customer record in the dataset. This vector is checked with the SOM result list and the vector values are updated to positive for the record found in the list. These values form the dependent variable which can be used with out supervised learning model to train the model and provide us with best possible rankings of fraud probabilities of each customer.

As the data increases the model gets trained better and the prediction becomes more accurate. The complexity and size of supervised model depends on the amount of data which the model would be processing. To create this hybrid deep learning model, we aim to use Keras as a high-level neural network API implemented by python which works on two basic machine learning libraries, namely TensorFlow and Theano.

V. CONCLUSION

With the advancement in technology, the use of credit card is increasing and so is the rate of credit card fraud. Banks are trying to implement different algorithms and techniques to overcome this problem. Most techniques were first based on fraud transactions but later the focus shifted to the fraud credit card applications which help them counter fraud at the budding level itself. Such techniques which help to analyze the previous credit card application data and predict the possibility of fraud are very useful to banks as they help them to predict if an application is potentially valid or fraudulent. If a fraud is caught at the application level, they can direct reject and report it which saves them a lot of trouble of later going through the fraudulent transactions made using card issued on that application and then rectifying them. There are many different ways being studied and tested which can be used to tackle this. This paper works in the same direction and proposes a new approach with SOM to achieve better and organized fraud detection.

VI. REFERENCES

[1] Navanshu Khare and Saad Yunus Sait ," Credit Card Fraud Detection Using Machine Learning Models and Collating Machine Learning Models", International Journal of Pure and Applied Mathematics Volume 118 No. 20 2018, 825-838

[2] Apapan Pumsirirat, Liu Yan," Credit Card Fraud Detection using Deep Learning based on Auto-Encoder and Restricted Boltzmann Machine", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 1, 2018

[3] Mr Yogesh M Narekar , Mr Sushil Kumar Chavan," A Review On Credit Card Fraud Detection Using BLAST-SSAHA Method", International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 11, November 2015

[4] Krishna Kumar Tripathi , Mahesh A. Pavaskar," Survey on Credit Card Fraud Detection Methods", International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, Volume 2, Issue 11, November 2012)

[5] Teuvo Kohonen ,"The Self-Organizing Map" ,PROCEEDINGS OF IEEE,VOL. 78,NO. 9 ,SEPTEMBER 1990

[6] MITALI Bansal ,Suman, " Survey Paper on Credit Card Fraud Detection", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 3, March 2014

[7] Richard J. Bolton and David J. Hand," Statistical Fraud Detection: A Review", available : http://metalab.uniten.edu.my/~abdrahim/ntl/Statistical%20Fraud%20Detection%20A%20Review.pdf

[8] Esam Kamal MARIED, Mansour Abdalla ELDALI ,Osama Omar ZIADA," A Literature Survey of Deep learning and its application in Digital Image Processing", UNIVERSITY OF TURKISH AERONAUTICAL ASSOCIATION DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING : (EEE 573),available : https://www.researchgate.net/publication/317346042_A_Literature_Study_of_Deep_learning_and_its_application_in_Digital_Image_Processing

[9] Kumar, P., Singh, Y. A study on software reliability prediction models using soft computing techniques (2013) International Journal of Information and Communication Technology, 5 (2), pp. 187-204. DOI: 10.1504/IJICT.2013.053119

[10] John A. Bullinaria,," Self Organizing Maps: Fundamentals", Introduction to Neural Networks : Lecture 16,2004

[11] Chandrahas Mishra , Dharmendra Lal Gupta , Raghuraj Singh," Credit Card Fraud Identification Using Artificial Neural Networks", International Journal of Computer Systems (ISSN: 2394-1065), Volume 04– Issue 07, July, 2017

[12] Dubravko Miljković," Brief Review of Self-Organizing Maps", MIPRO 2017/CTS