Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 127 (2018) 133–138

Procedia Computer Science

www.elsevier.com/locate/procedia

The First International Conference On Intelligent Computing in Data Sciences

ConvNets for Fraud Detection analysis

Alae Chouiekh[a,*] , EL Hassane Ibn EL Haj[b]

[a]*Laboratory of Multimedia ,signal and Communications Systems, National Institute of Posts and Telecommunications, Rabat,Morocco*
[b]*Laboratory of Multimedia ,Signal and Communications Sytems,National Institute of Posts and Telecommunications,Rabat,Morocco*

Abstract

Fraud activity is a big concern for telecom companies. The advances in technology and system information have significantly increased fraud activities, which can have negative impacts on revenue gains and services quality. Therefore, there is an urgent need for telecom companies to develop efficient algorithms that detect early potential frauds and/or prevent them. In this paper, we used deep learning techniques as an effective method to detect fraudsters in mobile communications. Fraud datasets from the customer details records (CDR) of a real mobile communication carrier were used and learning features were extracted and classified to fraudulent and non-fraudulent events activity. Different experiments were performed to evaluate the performance of our proposed model. We found that deep convolution neural networks (DCNN) technique outperformed other traditional machine learning algorithms (Support vector machines, Random Forest and Gradient Boosting Classifier) in term of accuracy (82%) and training duration. Thus, the use of this model can reduce the cost related to illegal use of services without payment.

*Keywords:*Fraud; Machine learning; Deep learning; CNN

1. Introduction

Due to the increased development of new technologies recently, many fraudulent activities have been arising including online banking, [2], e-commerce credit card transactions frauds [3], in addition to the telecommunication fraud [5], leading to multi-billion losses worldwide each year [6].

In fact, Fraud is very costly to all telecom carriers in term of capacity and income lost. According to the report

* Corresponding author. Tel.: +212-661-305-476.
*E-mail address:*alae.chkh@gmail.com

1877-0509
10.1016/j.procs.2018.01.107

published recently by Neural Technologies in 2016, the average loss of telecom industry was estimated to $249 billion dollars USD due to fraud activities. Fraud can be defined as an illegal use of telecom infrastructure like mobile communications with an intention for not paying services, misuse of voice calls (or data, SMS, MMS), cheating in subscriptions and using illegally services in the networks of telecom providers [1].

Telecom companies generate a huge amount of raw data including voice calls, SMS, recharge events, subscriptions and other services. The high volume of data collected and available in each day, which needs to be processed and manipulated, constitutes a big challenge in the telecom industry, and as a consequence, many fraudulent events can take place at any service, leading to a considerable loss of revenue to companies. Therefore, designing an accurate machine-learning model is essential to improve service usage monitoring and show significant revenue protection, in order to detect fraudulent events and activities in time.

The proposed Fraud detection methods in the telecom industry can be related to data mining techniques [7] or machine learning algorithms as suggested by different research papers [8,9,10]. Our main contribution in this research is to propose a new approach based on deep architecture models, taking benefit from deep learning success in different classification tasks such as object detection [11], image recognition [12], natural language processing [13] and dimensionality reduction [14]. Deep learning algorithms use deep architectures of multiple layers to extract features from raw data through a hierarchical progression of learned features using different layers from bottom navigating upwards without prior knowledge of any rule, which it becomes more challenging when dealing with a huge amount of data, thus deep learning helps significantly to avoid feature engineering process which is time and resource consuming.

In this paper we present a real time fraud detection system which is based on deep convolution neural networks (DCNN) model. Our main goal was to analyze behavioral patterns of customers in order to detect fraudulent events. In more details, a given subscriber or customer is classified as fraudster based on his features derived from the CDRs (customer details records) representing logs of the following event during a specific timestamp:

- ✓ Call Events including: number of outgoing/incoming calls, number of international calls, with corresponding charging units, max calls duration, and number of calls per day.
- ✓ Recharge Events representing the number of recharges per day with corresponding amount units.
- ✓ Subscriptions including number of subscriber offers per day with their charging units.
- ✓ Data Events which include the downloading volume with corresponding charging units.
- ✓ SMS Events including the number of SMS per day with corresponding charging units.

We consider that each user has accumulated a specific profile during his historical behavior explaining what is the most expected values of his daily consumption of services in operator's network. More specifically, the fraud detection is based on measures describing the changes on customer's profile and the transition from a normal user to a fraudster one.

Convolution neural networks (CNN) has become recently a focus to many researchers in studying image classifications tasks due to its efficiency on learning more meaningful and useful representations that yielded optimal results, outperforming other conventional machine learning algorithms. Using this method in the analysis of Fraud detection in the telecom industry would be worth investigating further. In our work, we investigated the effectiveness of CNN to uncover more insights into the fraud detection use case in the telecom industry. Each customer is represented by an image describing his behaviour (calls, sms, etc.) in the network. This paper proceeds as follows: section 2 describes how the data set is manipulated and presents our proposed Deep convolution neural networks architecture. Sections 3 evaluate the results of our model and compare it with the traditional machine learning algorithms using real telecom data set. Section 4 summarizes most of our conclusions.

2. Proposed CNN architecture

Our datasets are represented as artificial images describing the profile history of different subscribers during two months of behaviour; this historical data will be used as indicator capturing any change from a normal behaviour to fraud. We generated 18000 images representing the profile of 300 users during the period of 60 days. Fig.1 describes how the data is collected and passed through our deep convolution neural network.