Detection of Mobile Phone Fraud Using Possibilistic Fuzzy C-Means Clustering and Hidden Markov Model

Sharmila Subudhi, Veer Surendra Sai University of Technology, Sambalpur, India

Suvasini Panigrahi, Department of CSE and IT, Veer Surendra Sai University of Technology, Sambalpur, India

Tanmay Kumar Behera, Veer Surendra Sai University of Technology, Sambalpur, India

ABSTRACT

This paper presents a novel approach for fraud detection in mobile phone networks by using a combination of Possibilistic Fuzzy C-Means clustering and Hidden Markov Model (HMM). The clustering technique is first applied on two calling features extracted from the past call records of a subscriber generating a behavioral profile for the user. The HMM parameters are computed from the profile, which are used to generate some profile sequences for training. The trained HMM model is then applied for detecting fraudulent activities on incoming call sequences. A calling instance is detected as forged when the new sequence is not accepted by the trained model with sufficiently high probability. The efficacy of the proposed system is demonstrated by extensive experiments carried out with Reality Mining dataset. Furthermore, the comparative analysis performed with other clustering methods and another approach recently proposed in the literature justifies the effectiveness of the proposed algorithm.

KEYWORDS

Call Detail Records, Fraud Detection, Hidden Markov Model, Observation Symbol, Possibilistic Fuzzy C-Means, Profile Sequence

1. INTRODUCTION

The affordability of the mobile phone technology has resulted in exponential growth in mobile phone subscriptions. Due to this hike, the number of telecom fraud incidents is also increasing day-by-day. Fraud in mobile telecommunication indicates the deceptive methods adopted by a person (fraudster) for successfully obtaining the telephony services free of charge or at a reduced rate (Gosset and Hyland, 1999). The mobile communication fraud has always been a centre of attraction for fraudsters as it is very much easier to get a subscription via fake identifications. Besides, calling from a mobile phone is not bound to a specific location. Moreover, in today's digital era, the rising popularity of mobile commerce facilities has also given more opportunities to the imposters in gaining access and exploiting sensitive information of genuine users for carrying out fraud. Huge monetary losses are thus incurred by the subscribers as well as the telecom service provider companies due to such illegal usage of mobile phones and its services.

A survey done by US-based Communications Fraud Control Association (CFCA) presented that the fraudsters using telecom service without paying the bills has cost the global telecom industry around $46.3 billion in 2013 (Stokes, 2013). According to a study conducted by the organization

DOI: 10.4018/IJSE.2016070102

Financial Fraud Action United Kingdom (FFA UK) in 2014, around £23.9 million was lost in UK due to phone scam, which is three times more than in 2013 (Kosmides, 2014). The figures in the study show the increased trend of losses occurring as a consequence of fraudulent activities in telecom industries. In addition to the financial losses incurred due to increasing number of fraud cases, the subscriber's trust on the mobile phone service providing company also decreases, thus affecting the reputation of the organizations (Hoath, 1998). Therefore, there is a need to develop a robust mobile phone Fraud Detection System (FDS) for minimizing the financial losses.

Telecommunication fraud can be categorized into several types (Burge, 2000). One among them is the superimposed fraud, which can be described as the exploitation of a genuine user's account to make some amount of illegitimate activities by the fraudster. The genuine subscriber can stay unaware of this type of fraud for a long time as the number of fraudulent calls may be relatively small in the overall call volume (Cox et al., 1997). The aim of this work is to detect the superimposed fraud as it poses a bigger problem for the telecommunication industry.

The fundamental concept behind identifying the superimposed fraud is a thorough analysis of a subscriber's Call Data Records (CDRs). The CDR may be described as a metadata consisting of the caller's and callee's contact numbers, the time period and duration of a call, types of calls made and many more. The problem of detecting fraud in telecommunication networks heavily relies on the modelling of user profile by analyzing past CDRs of the subscriber. Mining the required information and getting the appropriate knowledge and pattern from the call record database helps in revealing certain behavioral characteristics of the respective subscriber hidden inside the CDRs.

The building of normal calling profiles are usually done by applying various supervised and unsupervised learning methods (Laskov et al., 2005) on user's past call records. However, due to public unavailability of labeled fraud data, the normal patterns can be extracted from the CDRs by using unsupervised learning techniques. These methods group similar call patterns together in a cluster, which provides relevant insights on the user's calling behavior. Once the normal calling profile is built, these patterns can be used by a mobile phone FDS for analyzing a user's current calling behaviour against the established profile to discriminate the normal user activity from the fraudulent ones. This subsequently helps the companies in improving their services and reducing the losses.

In this research, a novel anomaly-based hybrid fraud detection model has been proposed by integrating two unsupervised learning methods - Possibilistic Fuzzy C-Means (PFCM) clustering algorithm and Hidden Markov Model (HMM). Initially, PFCM is applied for building normal profile according to subscriber's past calling patterns. Thereafter, some parameter values essential for training an HMM are estimated from the user behavioral profiles. After the training of HMM is over, a profile sequence is generated. For every incoming call record, a new observation symbol is generated by PFCM clustering technique and given to the trained HMM model. This model finally classifies the new call pattern as normal or anomalous.

The application of hard clustering is inappropriate in this type of real world problem as it is unable to deal with the overlapping clusters. Therefore, fuzzy clustering has been used for accurately capturing the uncertain calling behavior of the subscribers so that individual calling instances may belong to more than one cluster. Furthermore, HMM has the capability of detecting fraudulent behavior using subscriber's calling patterns only.

The rest of the paper is organized as follows: Section 2 briefly introduces the related research carried out in this field as well as various application specific fraud detection problems. Section 3 sheds some light into the background study of HMM and PFCM techniques. Section 4 focuses on the proposed fraud detection model. The experimentation and comparative performance analysis are provided in Section 5 to demonstrate the effectiveness of the proposed approach. Finally, Section 6 concludes the paper by providing a brief summary about the contributions along with some future directions.

Runtime Verification on Robotics Systems

Zhijiang Dong, Yujian Fu and Yue Fu (2015). *International Journal of Robotics Applications and Technologies (pp. 23-40).*

www.igi-global.com/article/runtime-verification-on-robotics-systems/134032?camid=4v1a