Industrial Engineering Journal
ISSN: 0970-2555
Volume : 49, Issue 4, No. 1, April : 2020

A Modified Machine learning based technique for network call deception

DASARI LAKSHMI NARAYANA REDDY

Associate Professor
lakshmi1217@gmail.com

KRISHNA SWATHI

Assistant professor
Swathi.1992h@gmail.com

SOMARLA LOKESH KUMAR

Assistant Professor
lokeshkumaralts@gmail.com

Abstract

Detecting fraudulent calls on a network in real time and blocking them automatically. The call history on the network is collected for a certain period of time, along with call topographies for each of the collected call history by recipient number. Machine learning is then used to make decisions regarding whether or not a recipient number and a call to that recipient number may be fraudulent. It is possible to implement the choice model on the network in order to detect and prevent fraudulent calls.

Introduction

The field of telecommunications is confronted with a great deal of difficulty in terms of the communication across networks about fraudulent calls. According to a variety of survey tools, it is estimated that the annual losses caused all over the globe by fraudulent calls or deceptive actions have increased to $40 billion. In the small and medium-sized telecom business, the losses are growing at a rate that is greater than the earnings. It is an expense for both the government and non-government entities to maintain a full-time monitoring system for fraudulent operations. In order to overcome this difficulty, an effort has been made in this work to handle the issue by generating decision trees utilizing clustering analysis and machine learning modules in a manner that is both cost effective and efficient.

Related Work

In his article "2015," Niall J. Conroy presents a misleading and counterfeit news indicator system that uses linguistic and network analysis methodologies. Linguistic techniques, in which the text of misleading posts is mined and analyzed to correlate linguistic patterns with deception, and network techniques, in which likewisepost metadata or structured network enquiries are performed, will be paired to create collective deception activities.

Anton Wiens [2014] makes the suggestion that user profiles should be used in order to train for detecting deception calls based on the values of each profile. Because there is a lack of labeling data, only a select few techniques can be employed for deception call sensing when using supervised approaches.

The work of Iulia Lefter [2010] recommends the enunciation of each step in the development of an emotion identification system from the existing databases, the sentiment-specific topographies that are relevant for emotion detection, and the machine learning algorithms that are used. Classifiers based on Support Vector Machines (SVM) are used for indalogue sentiment recognition. SVM will regulate a hyperplane that will fully distinguish two classes. A hyperplane exploits the border between two datasets, and the trials that sit on the boundary are referred to as support vectors. This helps to produce an oratorically sovereign crossjustification framework.

Gideon Mendels[] makes the suggestion that one may frequently detect deceit from discourse. CXD took use of a large-scale corpus of both deceptive and non-deceptive dialog in order to train and evaluate spectral, lexical, and acoustic-prosodic feature sets using a number of different machine learning modules. Create a single hybrid deep model that takes into account both acoustic and lexical topographies, and then train them together so that they may obtain more advanced results on the CXD corpus.

During conference conversations, Larcker, D. [2012] presents estimated linguistically-based categorization modules of fraudulent calls. The psychological and linguistic research conducted in the past in relation to deception helps to identify certain word clusters, which are then used to construct prediction modules.

The model is developed by word kinds associated towards deceit and by traditional arithmetical evaluations, as was done in Baohua Wang's [2011] study, which evaluated classification modules of fraudulent calls made during conference calls. Despite this, their effectiveness ranges from 60% to 70%, and linguistic topographies may be used to identify false calls.

Graaff AJ makes the suggestion that discovering fraudulent calls from the average set of calls period and prolonged call period over calls made by the customer that are short and are equivalent to a defined threshold. For the purpose of training the best potential threshold values, machine learning is utilized as a technique. Therefore, every customer has their own set of criteria for prediction in order to detect misleading calls.

The project "Life cycle of a phone fraud" intends to develop machine learning modules that can identify the real kind of equipment or device that was used to make a call, as well as the geographic location that a call may be determined to be from based on its phoneprint.

The software used by Tata Communications technology monitors to detect and prevent deceptive behavior. The program addresses many anti-fraud technologies, such as automated reporting, machine learning, crowd sourcing, real-time monitoring, subscriber alerts, and big data analytics. When the anti-fraud software detects a fraudulent call, the network as a whole is immediately blocked from receiving calls from that number. This prevents any future fraudulent behavior.

David Lary [2010] proposed automated detection and reporting of online auction seller deception risk by using call through API interface and web site GUI data harvesting application feedback collection application fordata cleaning & analysis module to get cleaned data applying machine-learning algorithm and decision support system. This would be accomplished by using call through API interface and web site GUI data harvesting application feedback collection application.

Proposed System

Detecting fraudulent activity on a network through the accumulation of call histories on the network over a predetermined period of time, each call history including a number of call topographies for a call made to a recipient number, call topographies from each of the poised call histories organized in accordance with the recipient number, which results in a combination of call features for each recipient number. The resultant data points are derived from a series of data points that have been transformed into call features via the use of dimension reduction, with each data point denoting a specific call characteristic for the receiver number.

The execution of a clustering analysis that is generated out of a collection of data points into two or more clusters and the labeling of call attributes as being deceptive or non-deceptive based on the cluster that each individual data point belongs to. The execution of a supervised learning module on each tagged call feature as a trained information to create at least one or more decision modules to distinguish misleading calls is performed. Recognizing fraudulent calls made over the network by using at least one decision module and initiating a predetermined action in response to the detection of the fraudulent call.

It is suggested to use two different kinds of features: arithmetical features and categorical features, with arithmetical columns carrying numerical data and categorical columns including various forms of data than numerical data. It is possible for the misleading module to mislead the machine learning module into believing that a one-hot transformation, which converts a categorical call characteristic into an arithmetic number, and that the two are equivalent. The decision module for determining if a call is deceptive is obtained by the deceptive analysis module by applying a supervised learning module to the whole dataset. The decision tree is an example of one kind of decision module. A cross-validation model is used by the deceptive analysis module to collect training data points in order to generate a decision tree module for use in anticipating misleading calls. One of the most important advantages of the decision tree module is the forecasting module, which graphically displays the mechanism behind decision making.

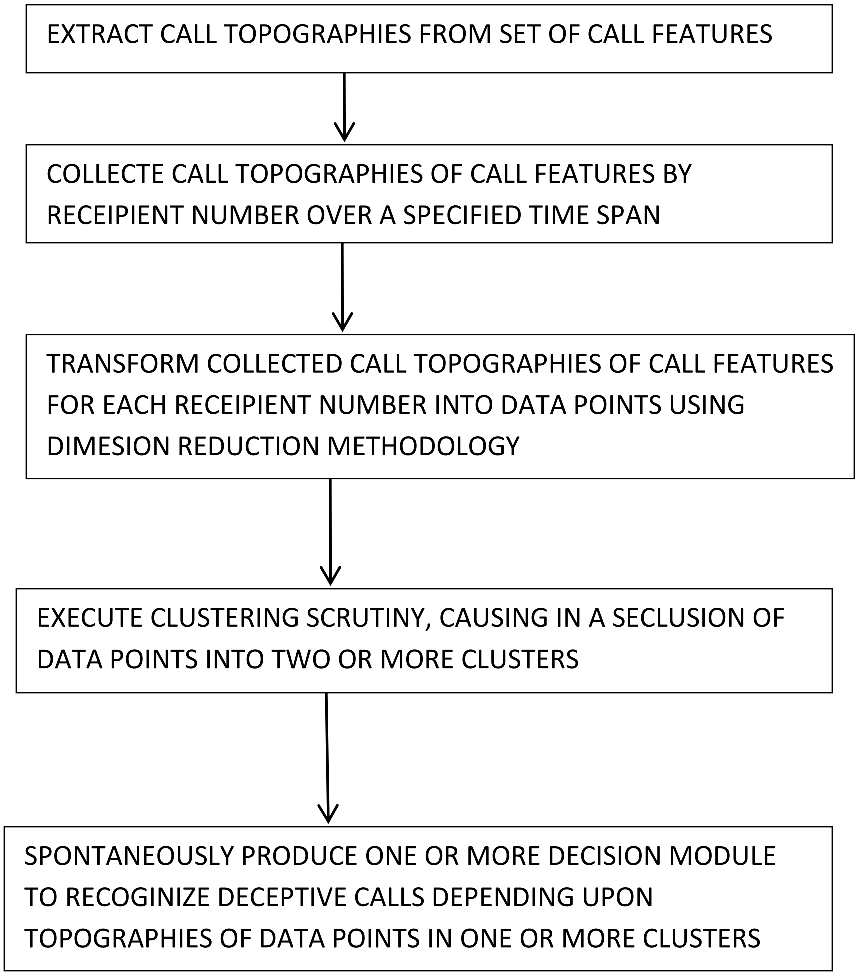
EXTRACT CALL TOPOGRAPHIES FROM SET OF CALL FEATURES
COLLECTE CALL TOPOGRAPHIES OF CALL FEATURES BY RECEIPIENT NUMBER OVER A SPECIFIED TIME SPAN
TRANSFORM COLLECTED CALL TOPOGRAPHIES OF CALL FEATURES FOR EACH RECEIPIENT NUMBER INTO DATA POINTS USING DIMESION REDUCTION METHODOLOGY
EXECUTE CLUSTERING SCRUTINY, CAUSING IN A SECLUSION OF DATA POINTS INTO TWO OR MORE CLUSTERS
SPONTANEOUSLY PRODUCE ONE OR MORE DECISION MODULE TO RECOGINIZE DECEPTIVE CALLS DEPENDING UPON TOPOGRAPHIES OF DATA POINTS IN ONE OR MORE CLUSTERS

FIG: Representation of Deceptive Calls Recognition Approach

Conclusion

The traditional research methods for the finding of fraudulent calls are somewhat unreliable. It is expected that trained specialists would be able to detect such moves. According to the findings of the research, human beings as untruth indicators are not much better than machines in terms of detecting lies in communication. One of the most significant drawbacks of these study results is that they are dependent on the trust of arithmetic, and they neglected the dishonest communication that accompanies those who are dishonest. The purpose of the suggested system is to use a method based on machine learning in order to identify dishonest calls in a network.

References

1. Niall J. Conroy, Victoria L. Rubin, and Yimin Chen “Automatic Deception Detection: Methods for Finding Fake News “,ASIST 2015, November 6-10, 2015, St. Louis, MO, USA.
2. Anton Wiens, Torsten Wiens and Michael Massoth’ “A new Unsupervised User Profiling Approach for Detecting Toll Fraud in VoIP Networks “,AICT2014 : The Tenth Advanced International Conference on Telecommunications.
3. Iulia Lefter,Leon J. M. Rothkrantz,David. A. van Leeuwen,PascalWiggers “Automatic Stress Detection in Emergency (Telephone) Calls“,Int. J. of Intelligent Defence Support Systems, Vol. x, No. x, xxxx,2010 Inderscience Enterprises Ltd.
4. Gideon Mendels, Sarah ItaLevitan, Kai-Zhan Lee, Julia Hirschberg “Hybrid Acoustic-Lexical Deep Learning Approach for Deception Detection”, Columbia University, USA.
5. Larcker, D., Zakolyukina, A.“Detecting Deceptive Discussions in Conference Calls”, Journal of Accounting Research, 50(2), 495540.2012.
6. BaohuaWang,Xiaolong Wang “Deceptive Financial Reporting Detection: A Hierarchical Clustering Approach Based on Linguistic Features “,2012 International Workshop on Information and Electronics Engineering (IWIEE),2011 Published by Elsevier Ltd.
7. Graaff AJ, Engelbrecht AP “An Overview of Models to Detect and Analyze Fraud in the Telecommunications Environment” School of Information Technology, University of Pretoria, South Africa.
8. MyleOtt, Yejin Choi, Claire Cardie,Jeffrey T. Hancock”Finding Deceptive Opinion Spam by Any Stretch of the Imagination”Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics, pages 309–319,Portland, Oregon, June 19-24, 2011. File name : P11-1032.pdf
9. Lifecycle of a Phone Fraudster: Exposing Fraud Activity from Account Reconnaissance to Takeover using Graph Analysis and Acoustical Anomalies.
10. “Fraud Protection Toolkit”,2014 Tata Communications Ltd.

11. David Lary, Alexey N. Nikitkov and Dan N. Stone “Which Machine-Learning Models Best Predict Online Auction Seller Deception Risk?”,National Aeronautics and Space Administration (NASA) Goddard Space Flight Center,February 14, 2010.