A Modified Method for Network Call Deception Based On Machine Learning

[1]Dr.Subramaniyan Saravanan, [2]Naveen Kumar Arem, [3]Pusukuri Prasanna Lakshmi
[1]Associate Professor, Department of CSE, Narsimha Reddy Engineering College, Secunderabad, Telangana
[2,3]Assistant Professor, Department of CSE, Narsimha Reddy Engineering College, Secunderabad, Telangana

Abstract

The real-time identification and automated prevention of fraudulent calls on a network. The network gathers call records over a certain timeframe, including call topographies associated with each recorded call history based on the recipient's phone number. Machine learning is then used to determine the legitimacy of a receiver number and the associated call, with the aim of identifying potential fraudulent activity. The implementation of a choice model on the network has the potential to effectively identify and mitigate fraudulent calls.

1.Introduction

The telecoms industry has significant challenges in effectively communicating and addressing fraudulent calls across networks. Based on a range of survey instruments, it has been projected that the global yearly financial losses resulting from fraudulent calls or misleading acts have risen to $40 billion. Within the context of the small and medium-sized telecommunications industry, it is seen that the losses incurred are seeing a growth rate surpassing that of the profits. Maintaining a comprehensive monitoring system for fraudulent activities incurs costs for both governmental and non-governmental institutions. To address this challenge, this study attempts to tackle the problem by using clustering analysis and machine learning modules to generate decision trees in a way that is both cost-effective and efficient.

2.Related Work

Niall J. Conroy's essay titled "2015" introduces a news indicator system that employs linguistic and network analysis approaches, which is said to be deceptive and fraudulent. The utilization of linguistic methods involves the extraction and examination of misleading posts' textual content to establish connections between language patterns and deceptive behaviour. Similarly, network techniques include the analysis of post metadata or structured network inquiries. These two approaches will be used to facilitate the execution of collective deception activities.

According to Anton Wiens (2014), it is proposed that the use of user profiles may be employed to facilitate the training process for identifying deception calls, by considering the values associated with each profile. Due to the limited availability of labelled data, the use of supervised algorithms for detecting deceit in phone calls is constrained to a limited number of strategies.

According to the research conducted by Iulia Lefter in 2010, it is suggested that a comprehensive approach should be used in the construction of an emotion identification system. This approach involves clearly articulating each stage involved in the system's creation, including the utilization of existing databases, the discovery of sentiment-specific patterns that are pertinent to emotion recognition, and the implementation of appropriate machine learning algorithms. Support Vector Machines (SVM) classifiers are often used for the task of sentiment recognition in indalogue. Support Vector Machines (SVM) are capable of effectively determining a hyperplane that can completely separate two distinct classes. A hyperplane is used to delineate the boundary between two datasets, with the data points that lie on this boundary being referred to as support vectors. This framework aids in the development of a crossjustification framework that is characterized by its oratorical sovereignty.

According to Gideon Mendels, it is possible to regularly identify instances of deception via dialogue. The study conducted by CXD used a substantial corpus of deceptive and non-deceptive dialogues to train and assess several feature sets, including spectral, lexical, and acoustic-prosodic features, using multiple machine learning modules for this purpose. Develop a unified hybrid deep model that incorporates both acoustic and lexical features, and afterwards train the model jointly to achieve enhanced performance on the CXD corpus.

In the context of conference discussions, Larcker (2012) proposes linguistically-based classification modules for estimating fraudulent calls. Previous psychological and linguistic research has been undertaken to explore deception, aiming to find specific word clusters that may be used to develop prediction modules.

The model was constructed using word categories related to deception and conventional numerical assessments, similar to the approach used in Baohua Wang's [2011] research. Wang's work focused on evaluating classification modules for identifying fraudulent calls made during conference calls. However, the efficacy of these methods varies between 60% and 70%, and language topographies may be used to detect erroneous predictions.

Graaff AJ proposes the identification of fake calls by comparing the average duration of calls during a certain time to the duration of lengthy calls, while also considering calls made by customers that are brief and fall below a predetermined threshold. Machine learning is used as a method in order to train the most optimal threshold values. Hence, each individual consumer has a unique set of parameters for the purpose of predicting and identifying deceptive phone calls.

The objective of the project titled "Life cycle of a phone fraud" is to create machine learning modules capable of accurately classifying the specific kind of equipment or device used for initiating a call, along with determining the geographical origin of the call based on its phoneprint.

Tata Communications employs software technologies to effectively monitor and identify dishonest behaviour, hence enabling proactive prevention measures. The initiative encompasses a range of anti-fraud technologies, including automated reporting, machine learning, crowd sourcing, real-time monitoring, subscriber notifications, and big data analytics. Once the anti-fraud software identifies a fake call, it promptly imposes a network-wide ban on incoming calls originating from the identified number. This measure serves to mitigate potential instances of fraudulent conduct in the future.

In a study conducted by David Lary in 2010, a method for automated identification and reporting of potential deception risk among online auction sellers was proposed. This method involved utilizing an API interface and a web-based graphical user interface (GUI) data harvesting application, along with a feedback collection application, for the purpose of data cleaning and analysis. The cleaned data was then subjected to a machine-learning algorithm and a decision support system to facilitate the detection and reporting process. This objective may be achieved by using an API interface for call routing, as well as a web-based graphical user interface (GUI) for data extraction, and an application dedicated to collecting feedback.

3. Proposed System

The identification of fraudulent behaviour on a network can be achieved by analyzing call histories accumulated over a predetermined time period. Each call history contains multiple call topographies associated with a specific recipient number. By organizing the call topographies from all the collected call histories based on the recipient number, a set of combined call features can be obtained for each recipient number. The resulting data points are obtained by transforming a sequence of data points into call characteristics via the use of dimension reduction. Each data point represents a distinct call characteristic associated with the recipient's phone number.

The process involves doing a clustering analysis on a dataset, resulting in the formation of two or more distinct clusters. Subsequently, the call characteristics are assigned labels indicating whether they are deceptive or non-deceptive, depending on the cluster membership of each individual data point. The process involves applying a supervised learning algorithm to each labelled call characteristic, using the learned data to generate one or more decision modules that can effectively differentiate between deceptive and non-deceptive calls. The identification and reaction to fraudulent calls made over the network include the use of a decision module to identify such calls and then trigger a specified action.

It is recommended to use two distinct types of features in data analysis: arithmetical features and categorical features. Arithmetical features consist of columns containing numerical data, while categorical features include a range of data formats other than numerical. There exists a potential for the misleading module to deceive the machine learning module by creating the false impression that a one-hot transformation, which involves converting a categorical attribute into a numerical representation, is identical to the original attribute. The determination of call deception is facilitated by the decision module, which is derived from the deceptive analysis module by the application of a supervised learning algorithm to the whole dataset. The decision tree may be seen as an illustrative instance of a decision module. The deceptive analysis module employs a cross-validation technique to gather training data points, enabling the generation of a decision tree module. This decision tree module is then used to anticipate fraudulent phone calls. One of the primary benefits of the decision tree module is in its forecasting capabilities, since it visually illustrates the underlying process of decision-making.

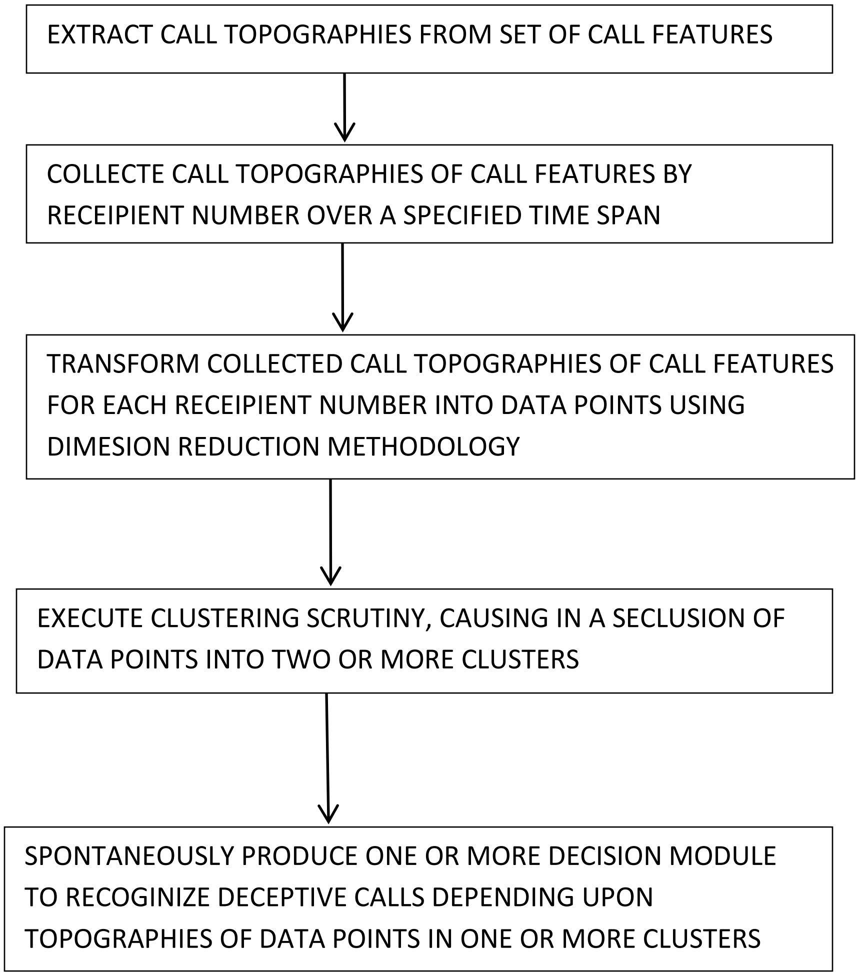
EXTRACT CALL TOPOGRAPHIES FROM SET OF CALL FEATURES
COLLECTE CALL TOPOGRAPHIES OF CALL FEATURES BY RECEIPIENT NUMBER OVER A SPECIFIED TIME SPAN
TRANSFORM COLLECTED CALL TOPOGRAPHIES OF CALL FEATURES FOR EACH RECEIPIENT NUMBER INTO DATA POINTS USING DIMESION REDUCTION METHODOLOGY
EXECUTE CLUSTERING SCRUTINY, 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

4.Conclusion

The conventional research methodologies used for the detection of fraudulent calls exhibit a degree of unreliability. It is anticipated that those who have received specialized training will possess the capability to identify such manoeuvres. Based on the study results, it can be concluded that humans, as discerners of falsehoods, exhibit a comparable level of proficiency to robots when it comes to detecting deceit in interpersonal communication. One notable limitation of these research findings is their reliance on the assumption of mathematical accuracy, while overlooking the presence of deceptive communication shown by individuals with dishonest tendencies. The objective of the proposed system is to use a machine learning approach to detect fraudulent calls inside a network.

5. 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.