Fraud detection in telephone conversations for financial services using linguistic features

Nikesh Bajaj, Tracy Goodluck Constance, Marvin Rajwadi, Julie Wall, Mansour Moniri
Intelligent Systems Research Group, University of East London, UK
{n.bajaj, t.goodluckconstance, m.rajwadi, j.wall, m.moniri}@uel.ac.uk

Cornelius Glackin, Nigel Cannings
Intelligent Voice Ltd., London, UK
{neil.glackin, nigel.cannings}@intelligentvoice.com

Chris Woodruff, James Laird
Strenuus Ltd., London, UK
{chris.woodruff, james.laird}@strenuusltd.com

Abstract

Detecting the elements of deception in a conversation is one of the most challenging problems for the AI community. It becomes even more difficult to design a transparent system, which is fully explainable and satisfies the need for financial and legal services to be deployed. This paper presents an approach for fraud detection in transcribed telephone conversations using linguistic features. The proposed approach exploits the syntactic and semantic information of the transcription to extract both the linguistic markers and the sentiment of the customer's response. We demonstrate the results on real-world financial services data using simple, robust and explainable classifiers such as Naive Bayes, Decision Tree, Nearest Neighbours, and Support Vector Machines.

1 Introduction

With the rapid increases in technological development, fraud is a major concern for safe and trustworthy digital communications and transactions. To deal with this, several data mining and machine learning systems have been developed [1]. The techniques available in the literature have mostly focused on financial fraud, such as insurance and credit card fraud [2], analysing the anomalous behaviour of the customer in financial transactions [3]. However, fraud is not only limited to financial transactions, it also involves being deceptive by providing false statements, such as in loan applications and when trying to uncover the account status of others, etc.

Compared to other fraud detection techniques, we have not yet found other research in the literature which reports the analysis of telephone conversations between customers and service providers, such as insurance companies or banks. In fact, a telephone conversation is usually the first point of contact. Analysis of this data-rich communication can reveal potential fraudulent cues at a very early stage to prevent fraud. To achieve this, a linguistic based approach using Natural Language Processing (NLP) techniques [4] can be used.

The requirement for financial and legal services to deploy any advanced Artificial Intelligence (AI) algorithm is that it must be a transparent and fully explainable system. Financial institutions are required to explain and justify the decision taken for customers or clients, for example, whether they will pay out on an insurance claim or not. This limits the use of sophisticated and complex systems

AI for Social Good workshop at NeurIPS (2019), Vancouver, Canada.

Table 1: Linguistic Markers

Marker	Example
Causation: Providing a certain level of concreteness to an explanation. [10, 13]	Because, Effect, Hence
Negation: Avoiding to provide a direct response [14]	No, Not, Can't, Didn't
Hedging: Describes words which meaning implicitly involves fuzziness [15]	May be, I guess, Sort of
Qualified assertions: Unveils questionable actions [15]	Needed, Attempted
Temporal Lacunae: Unexplained lapses of time [15]	Later that day, Afterwards
Overzealous expression: Expresses some level of uncertainty [15]	I swear to God, Honestly
Memory loss: Feigning memory loss [15]	I forget, Can't remember
Third person plural pronouns: Possessive determiners to refer to things or people other than the speaker [10]	They, Them, Theirs
Pronouns: Possessive determiners to refer to the speaker by overemphasising their physical presence [10, 16]	I, Me, Mine
Negative emotion: Negative expressions in word choice [17, 10, 18]	Afraid, Sad, Hate, Abandon, Hurt
Negative sentiment: Negative emotional effect [18]	Abominable, Anger, Anxious, Bad
Positive emotion: Positive expressions in word choice [10, 18]	Happy, Brave, Love, Nice, Sweet
Positive sentiment: Positive emotional effect [18]	Admire, Amazing, Assure, Charm
Disfluencies: Interruption in the flow of speech [10]	Uh, Um, You know, Er, Ah
Self reference words: Deceivers tend to use fewer self-referencing expressions [15]	I, My, Mine
Nominalised verbs: Nouns derived from verbs. Nominalisations tend to hide the real action. [19]	Education, Arrangement

such as Deep Neural Networks (DNN) in this area. However, the most recent advancements are tending towards achieving full Explainable AI [5–7].

In this paper, we propose an approach to use linguistic features by exploiting the syntactic and semantic information in the transcripts of telephone conversations. We demonstrate the results of this approach on real-world data, collected from two financial services institutions. We trained simple, robust and explainable classifiers to achieve an explanation of the decision process while revealing the importance of features responsible for the decision.

2 Proposed Approach

The proposed approach is designed to analyse the transcription of a telephone conversation, generated using state-of-the-art Automatic Speech Recognition technology in real-time, which allows our deception detection approach to also work in real-time. This approach extracts the linguistic features of the transcription and trains the explainable classifiers to analyse and validate the decision process.

Language is a medium of communication, where the choice of words can reflect the emotional and cognitive state of the speaker. Only training and rehearsal can allow a speaker to control their vocabulary to not leak any emotional state [8]. Psychologists suggest that speakers often have no control over their choice of words and can reveal their emotions involuntarily [9]. Deceptive speech is considered to be a deliberate choice to mislead and the language used can reveal an underlying cognitive state. Many studies have shown that linguistic cues can indicate the elements of deception in language [10, 11]. This work considers two types of linguistic features. One is extracted from the syntax of the language, focusing primarily on the words used, called *Linguistic Markers*. The second focuses on the overall sentence structure, its semantics, and the sentiment of the dialogue.

2.1 Linguistic Markers

The observation of linguistic cues to detect deceptive speech focuses mainly on word/phrases either in oral or written form. There is an exhaustive list of linguistic markers provided by [10, 12], and we have adapted a subset of these linguistic markers for our study, provided in Table 1.

2.2 Sentiment

Many studies support the presence of negative words and sentiment in deceptive speech [20–22], which can be spotted by linguistic markers. However, linguistic markers analyse only the syntactic information of the dialogue text and are prone to miss the overall sentiment. In cases where the sentence has a negative sentiment but with no existing negative words, linguistic markers have limited capability. To overcome this, we use sentiment as a feature, which reflects the polarity of the speaker's feelings in a dimension from negative to positive. Therefore, sentiment can be used to detect the elements of deception in speech. To extract the sentiment of dialogue in telephone conversations, we used a DNN which was trained on the IMDB movie review dataset [23]. Full information on the development, training, and evaluation of this sentiment-analysis based DNN can be found in our previous work [24]. The use of a trained model on IMDB's dataset is considered to be effective, assuming that the set of domain-dependent words is very small [25]. However, an efficient domain adaptation using transfer learning can be used to extract the sentiment for given context [26].

3 Experiments & Results

3.1 Dataset

To evaluate our proposed approach, we employed real-world data collected from financial services institutions. This dataset contains the transcriptions of 56 telephone conversations, collected from two different financial institutions. Ideally, a larger dataset would be desirable, however, this data is limited volume due to legal & ethical constraints. The transcription of each conversation includes the operator's questions and the customer's responses. From the dataset of 56 calls, 32 are known fraudulent calls and 24 are non-fraudulent. As this is real-world data, the timing of the calls varies widely. The average number of responses are 19 ± 15. The shortest conversation in the dataset has only four responses from the customer; while the longest has 101 responses.

3.2 Feature Extraction and Modeling

From each telephone conversation, only the customer's responses are used for the feature extraction. Two types of linguistic features are extracted: namely *Linguistic Markers* and *Sentiment*, as explained in Section 2. For linguistic markers, the frequency of each of the 16 markers from Table 1 present in the customer's response is computed. Then the sentiment of each customer response is estimated using the DNN and scaled from -1 to 1. As there are a different number of responses in each conversation, we computed statistical measures from each individual response within the conversations. In total, 11 sentiment-related features were extracted, namely: mean, standard deviation (SD), minimum (Min), maximum (Max), median, interquartile range (IQR), Kurtosis, Skewness, positive energy (pE), negative energy (nE), and total number of responses (tR).

Finally, we trained the models (Naive Bayes, Decision Tree (DTree), k-Nearest Neighbors (kNN), and Support Vector Machines (SVM) with the individual features (e.g. marker, sentiment) and then we combined the features and re-trained. The parameters for the models are as follows: DTree with $depth = 3$, kNN with $k = 3$, and SVM with linear kernel and $C = 1$. The choice of models and their parameters were restricted by three properties: simplicity, robustness, and explainability. Since the data size is small, we trained and tested each model with K-Fold cross-validation, with K=10. The mean and SD of the training and testing accuracies for each model are tabulated in Table 2 and plotted in Figure 2.

3.3 Discussion

It can be observed that the highest testing accuracy achieved with solely the linguistic markers is 65.5% using kNN, whereas, with sentiment features, it is only 62% using an SVM. However combining both features, improves the testing accuracy to 69% with only 0.1 deviation for the SVM. One of the decision trees built with combined features is shown in Figure 1. The decision process is very visible from the tree and shows the importance of the features. For example, the most important feature is - Median of sentiment value with threshold 0. Another important feature is 'third person plural pronoun', which is indication of deception, reflecting a customer's attempt to discuss third person, while the call is about his/her own financial account. It can also be noticed that qualified

assertion, negative emotion, causation, and nominalised verb are also important linguistic features. Interestingly, a variation in sentiment values (SD) of responses, is also an important feature, indicating the too much change in the language of customer.

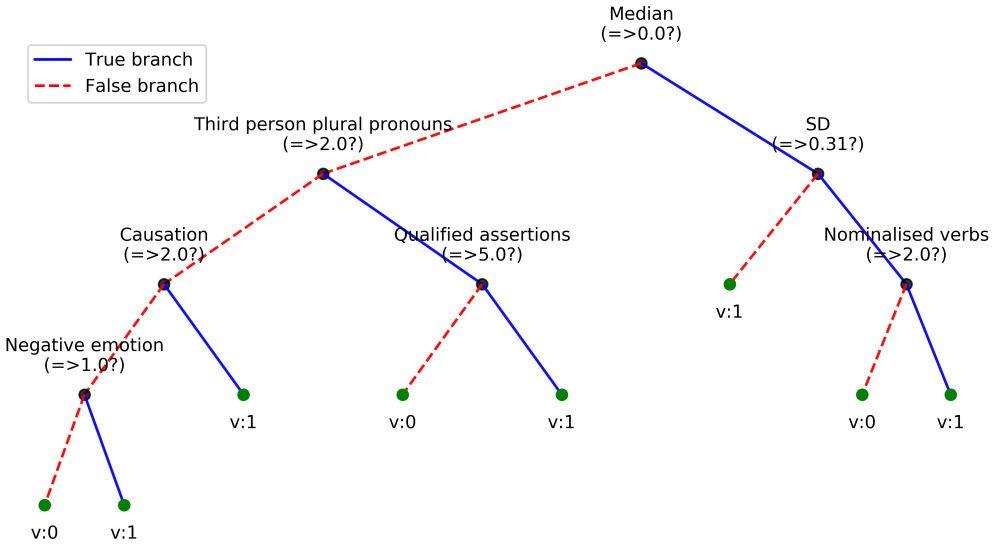

Figure 1: A Decision Tree for fraud detection. Leaf node v:0 - Non-Fraud, v:1 - Fraud

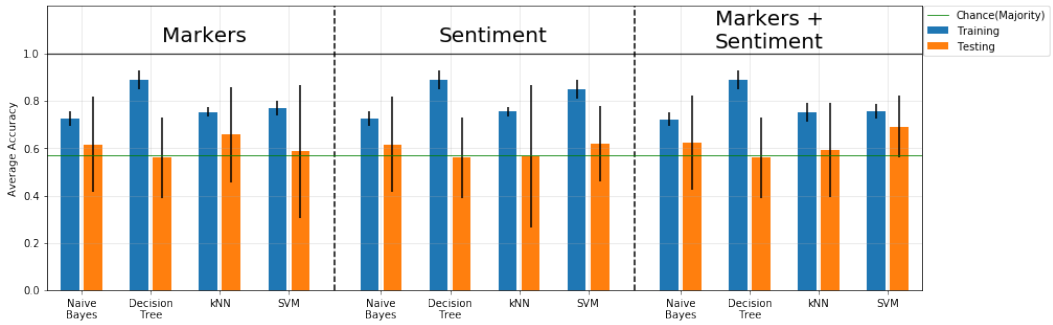

Figure 2: Average performance of K-Fold(K=10) for different models

Table 2: Results of modeling with K-Fold (K=10)

Features	Accuracy	Model			
		Naive Bayes	DTree (d=3)	kNN(k=3)	SVM(Linear)
Markers	Training	0.7241 ± 0.03	0.8871 ± 0.04	0.7521 ± 0.02	0.7679 ± 0.03
	Testing	0.6167 ± 0.20	0.5600 ± 0.17	0.6567 ± 0.20	0.5867 ± 0.28
Sentiment	Training	0.7241 ± 0.03	0.8871 ± 0.04	0.7540 ± 0.02	0.8491 ± 0.04
	Testing	0.6167 ± 0.20	0.5600 ± 0.17	0.5667 ± 0.30	0.6200 ± 0.16
Markers + Sentiment	Training	0.7222 ± 0.03	0.8871 ± 0.04	0.7500 ± 0.04	0.7560 ± 0.03
	Testing	0.6233 ± 0.20	0.5600 ± 0.17	0.5933 ± 0.20	0.6900 ± 0.13

4 Conclusions and future work

The proposed approach for fraud detection in financial services telephone conversations has employed two different types of linguistic features, namely markers and sentiment. While markers exploit the

syntactic information of the conversation, sentiment uses semantic information. The results presented in the paper show that combining these features produces the highest average accuracy. In order to achieve transparency in the decision process, with the limited dataset size, the choice of models were kept simple, robust and explainable. The financial and legal services are required to explain the decision process made with any mode of operation. For this same reason, an example of a decision tree from these experiments is shown in Figure 1. With a small decision tree, it is easy to explain the procedure producing outcome. Future work plans to extend the presented work for different scenarios including legal and insurance services, with the aim to again employ real-world data.

References

[1] Eric WT Ngai, Yong Hu, Yiu Hing Wong, Yijun Chen, and Xin Sun. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, 50(3):559–569, 2011.

[2] Adrian Bănărescu. Detecting and preventing fraud with data analytics. *Procedia economics and finance*, 32:1827–1836, 2015.

[3] Masoumeh Zareapoor, Pourya Shamsolmoali, et al. Application of credit card fraud detection: Based on bagging ensemble classifier. *Procedia computer science*, 48(2015):679–685, 2015.

[4] Christopher D Manning, Christopher D Manning, and Hinrich Schütze. *Foundations of statistical natural language processing*. MIT press, 1999.

[5] David Gunning. Explainable artificial intelligence (xai). *Defense Advanced Research Projects Agency (DARPA), nd Web*, 2, 2017.

[6] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. Why should i trust you?: Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1135–1144. ACM, 2016.

[7] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems*, pages 4765–4774, 2017.

[8] Jules Henry. The linguistic expression of emotion. *American Anthropologist*, 38(2):250–256, 1936.

[9] John R Schafer. *Grammatical differences between truthful and deceptive written narratives*. Citeseer, 2007.

[10] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001.

[11] Katie Cohen, Fredrik Johansson, Lisa Kaati, and Jonas Clausen Mork. Detecting linguistic markers for radical violence in social media. *Terrorism and Political Violence*, 26(1):246–256, 2014.

[12] Sean L Humpherys. A system of deception and fraud detection using reliable linguistic cues including hedging, disfluencies, and repeated phrases. 2010.

[13] Jeffrey T Hancock, Lauren Curry, Saurabh Goorha, and Michael Woodworth. Automated linguistic analysis of deceptive and truthful synchronous computer-mediated communication. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, pages 22c–22c. IEEE, 2005.

[14] Susan H Adams. *Communication under stress: indicators of veracity and deception in written narratives*. PhD thesis, Virginia Tech, 2002.

[15] Joan Bachenko, Eileen Fitzpatrick, and Michael Schonwetter. Verification and implementation of language-based deception indicators in civil and criminal narratives. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 41–48. Association for Computational Linguistics, 2008.

[16] Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1566–1576, 2014.

[17] David B Skillicorn and Ayron Little. Patterns of word use for deception in testimony. In *Security Informatics*, pages 25–39. Springer, 2010.

[18] Dan Jurafsky. *Lexicons for Sentiment, Affect, and Connotation*. Pearson Education India, 2000.

[19] Maria Lapata. The disambiguation of nominalizations. *Computational Linguistics*, 28(3):357–388, 2002.

[20] Wendell C Rudacille. *Identifying lies in disguise*. Kendall/Hunt, 1994.

[21] Kittie Wells Watson. Oral and written linguistic indices of deception during employment interviews. 1981.

[22] Elizabeth Wade. *Communicating about narrative (in) accuracy*. Stanford University, 1993.

[23] Andrew L Maas, Raymond E Daly, Peter T Pham, Dan Huang, Andrew Y Ng, and Christopher Potts. Learning word vectors for sentiment analysis. In *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1*, pages 142–150. Association for Computational Linguistics, 2011.

[24] Marvin Rajwadi, Cornelius Glackin, Julie Wall, Gérard Chollet, and Nigel Cannings. Explaining sentiment classification. *Proc. Interspeech 2019*, pages 56–60, 2019.

[25] Yasuhisa Yoshida, Tsutomu Hirao, Tomoharu Iwata, Masaaki Nagata, and Yuji Matsumoto. Transfer learning for multiple-domain sentiment analysis—identifying domain dependent/independent word polarity. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

[26] Pedro Henrique Calais Guerra, Adriano Veloso, Wagner Meira Jr, and Virgílio Almeida. From bias to opinion: a transfer-learning approach to real-time sentiment analysis. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 150–158. ACM, 2011.