

# Fraud detection in telephone conversations for financial services using linguistic features

———读书笔记

姓名：曾昕菟 学号：17211170

## 一、论文背景及目的

目前电话欺诈的行为屡见不鲜，而得益机器学习发展迅速，该论文提出了一种利用语言的某些特征，从而使得在电话中检测出欺诈的方法。应对的欺诈主要是金融欺诈，例如保险和信用卡的欺诈，该方法考虑到转录的句法和语义信息，进而提取语言特征和客户反应，从而可  
进行判断。

## 二、论文研究思路

考虑到目前在金融方面，电话往往是最常用来沟通的途径，因此基于使用自然语言处理（NLP）可以在早期就揭示可能存在的欺诈行为，从而减少欺诈的发生。但是值得注意的是，能使用的技术应该尽可能透明化并且能够证明对客户无害，否则在这种敏感的行业，会受到处罚，因此有些技术，例如深度神经网络技术难以采用。

该论文认为词语的选择可以反映说话者的情感和认知状态。只有训练才能使说话者控制自己的词汇，以免泄漏任何情绪状态，因此在存在欺骗性词语的话中，很有可能就代表着可能正在进行带有欺诈性的行为，为了提取出带有欺诈性的特征，主要考虑两个方面，其一是用词，其二是句子的结构，也就是带有的感情色彩。

从局部看，先考虑用词，用词主要是看的某些能够反应特定信息的词汇，比如因果关系往往能够提供某种解释，因此需要注意这类词汇，又比如指代自己的代词，这个是考虑到欺诈方往往不太会指代自己，因此这种用词也一定程度可以反应是否存在潜在的欺诈行为。此外还有很多的具有代表性的用词。更具体可以观看下图。

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

从整体看，第二个注意点就是句子结构，句子结构在英语中与表达的感情色彩息息相关，而研究表明，在欺诈行为中，消极感情会往往被欺诈方突出表现，该方法基于机器学习，在 IMDB 电影库中使得机器得到学习，从而能够有效判断出言语中带的感情色彩是否消极。

### 三、论文结果

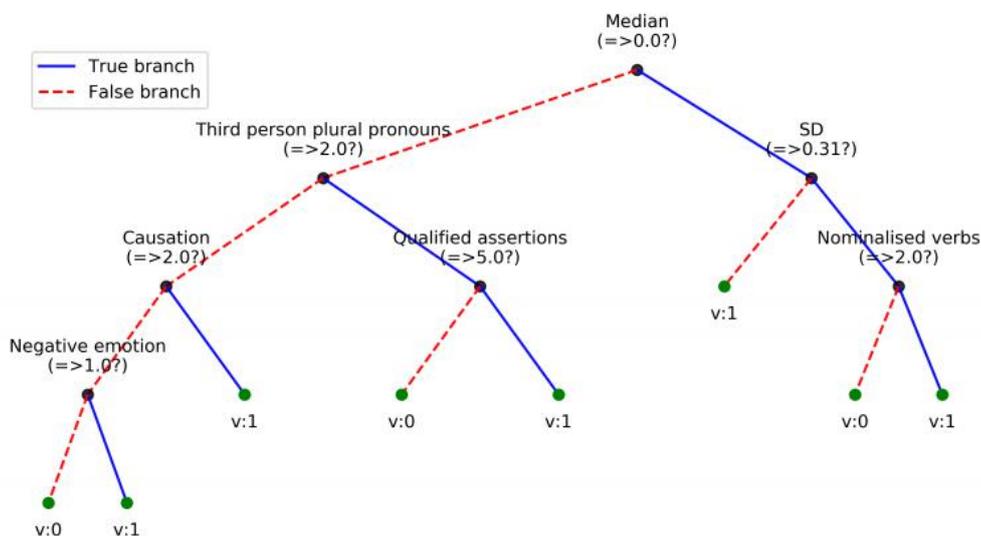
论文作者从一堆电话对话中人为地判断了 56 个对话，这 56 个对话中有 32 个欺诈电话，总体的平均客户回答次数为 19 次，最少的只有 4 次，而最多的有 101 次。

基于以上真实数据，作者将这 56 个对话套用以上提出的理论方法，并对客户的回答用以上思路进行分析，即分析用词和整体态度，在提取完特征后采取四种方法分别进行欺诈分析。结果如下图所示。

| 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 |

在实际应用可以有一种如下的决策树判断，可以看出比较重要的判断因素是感情数值化后的中值大小、第三人称代词等，这些正如前面思路分析时候提到的一样，与欺诈行为息息相

关，因此可以很好的反应出是否有欺诈倾向。在最后进入  $v=0$  说明无欺诈， $v=1$  说明有欺诈。而事实也说明该决策树的判断准确率达到六成以上。



#### 四、个人感想

该论文产生的主要背景是当前日益严重的电话欺诈，论文从实际出发，人为考虑出电话欺诈中表现出的典型特征，并将其数值化，从而可以判断出是否有欺诈倾向，尤其是现在机器学习发展的越来越如火如荼，因此是一种判断欺诈行为的很好方法，而这种思想究其根本是一个特别有用的思想，从一定量的具体事物中提取特征统计，而根据这种统计信息往往又能反应某些具体事物的倾向，这种思想其实用的很广泛，在无线通信中，均衡就是如此，均衡有一种盲均衡的方法，该方法正是将发送信息的特征提取出来，并根据特征信息反过来确定信道情况，从而调整抽头系数。因此可以说，虽然该篇论文面向的是一个特定的场景，但是将思想概括出来，可以适用于许多场景。