An AI-Based Lie Detector for Call Center Conversations

Originally published at https://www.unite.ai/an-ai-based-lie-detector-for-call-center-conversations/

Researchers in Germany have used machine learning to create an audio analysis system intended primarily to act as an AI-based lie detector for customers in audio communications with call center and support staff.

The system uses a specially-created dataset of audio recordings by 40 students and teachers during debates on contentious subjects, including the morality of the death penalty and tuition fees. The model was trained on an architecture that uses Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM), and achieved a reported accuracy rate of 98%.

Though the stated intent of the work cites customer communications, the researchers concede that it effectively operates as a general purpose lie-detector:

> *'The findings are applicable to a wide range of service processes and specifically useful for all customer interactions that take place via telephone. The algorithm presented can be applied in any situation where it is helpful for the agent to know whether a customer is speaking to her/his conviction.*
>
> *'This could, for example, lead to a reduction in doubtful insurance claims, or untruthful statements in job interviews. This would not only reduce operational losses for service companies, but also encourage customers to be more truthful.'*

Dataset Generation

In the absence of a suitable publicly available dataset in the German language, the researchers – from Neu-Ulm University of Applied Sciences (HNU) – created their own source material. Fliers were posted at the university and at local schools, with 40 volunteers selected with a minimum age of 16. Volunteers were paid with a 10 euro Amazon voucher.

The sessions were conducted on a debate club model designed to polarize opinion and arouse strong responses around incendiary topics, effectively modeling the stress that can occur in problematic customer conversations on the phone.

The topics on which the volunteers had to speak freely for three minutes in public were:

– *Should the death penalty and public executions be reintroduced in Germany?*
– *Should cost-covering tuition fees be charged in Germany?*
– *Should the use of hard drugs such as heroin and crystal meth be legalized in Germany?*
– *Should restaurant chains serving unhealthy fast food, such as McDonald's or Burger King, be banned in Germany?*

Pre-Processing

The project favored the analysis of acoustic speech features in an Automatic Speech Recognition (ASR) approach over an NLP approach (where speech is analyzed at a linguistic level, and the 'temperature' of the discourse is inferred directly from use of language).

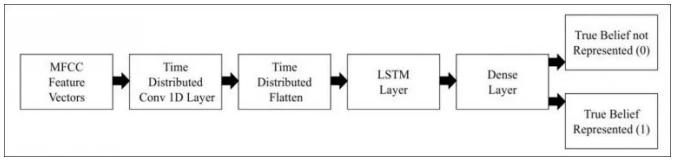
The pre-processed extracted samples were analyzed initially via Mel-frequency Cepstral Coefficients (MFCCs), a reliable, older method still very popular in speech analysis. Since the method was first proposed in 1980, it is notably frugal with computing resources in terms of recognizing recurrent patterns in speech, and is resilient to various levels of audio capture quality. Because the sessions were undertaken over VOIP platforms in lock-down conditions in December of 2020, it was important to have a recording framework that could account for poor quality audio when necessary.

It's interesting to note that the two aforementioned technical limitations (limited CPU resources in the early 1980s and the eccentricities of VOIP connectivity in a congested network context) combine here to create what is effectively a 'technically sparse' model that is (apparently) unusually robust in the absence of ideal working conditions and high-level resources – mimicking the target arena for the resulting algorithm.

Thereafter a Fast Fourier Transform (FFT) algorithm was applied against the audio segments to supply a spectral profile of each 'audio frame', before final mapping to the Mel Scale.

Training, Results and Limitations

During training, the extracted feature vectors are passed to a time-distributed convolutional network layer, flattened and then passed to an LSTM layer.



Architecture of the training process for the AI truth detector. Source: https://arxiv.org/ftp/arxiv/papers/2107/2107.11175.pdf

Finally, all the neurons are connected to each other in order to generate a binary prediction for whether or not the speaker is saying things that they believe to be true.

In tests after training, the system achieved an accuracy level of up to 98.91% in terms of intent discernment (where the spoken content may not reflect the intent). The researchers consider that the work empirically demonstrates conviction identification based on voice patterns, and that this can be achieved without NLP-style deconstruction of language.

In terms of limitations, the researchers concede that the test sample is small. Though the paper does not explicitly state it, low-volume test data can reduce later applicability in the event that the presumptions, architected features and the general training process are over-fit to the data. The paper notes that six of the eight models constructed throughout the project were over-fitted at some point in the learning process, and that there is further work to be done in generalizing the applicability of the parameters set for the model.

Further, research of this nature must account for national characteristics, and the paper notes that the German subjects involved in the generation of the data may have communications patterns which are not directly replicable across cultures – a situation that would likely arise in any such study in any nation.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.
Portfolio site: martinanderson.ai
Contact: martin@martinanderson.ai

Discover More