Data & Technology for Fact-checking

Lucas Fagan, Ethan Holland, Caroline Wang; Professor Jun Yang

Because politicians often repeat previously fact-checked claims, our goal is to automate fact-checking and increase exposure for fact-checking efforts by matching sentences from live feeds against a compilation of existing fact-checks.

We divide our problem into two subproblems:

- Given an audio source, identify and extract check-worthy factual sentences.
- Given a check-worthy factual sentence, find relevant fact-checks in a database.

We define a fact-check as relevant to a spoken claim if the fact-check either entails or contradicts the spoken claim. In NLP, sentence X *entails* sentence Y if a human assuming X to be true would infer that Y is also true. Sentence X *contradicts* sentence Y if a human assuming X to be true would infer that Y cannot be true. The problem of identifying entailments and contradictions is known as **natural language inference (NLI)**.

Premise	Label	Hypothesis
Fiction		
The Old One always comforted Ca'daan, except today.	*neutral*	Ca'daan knew the Old One very well.
Telephone Speech		
yes now you know if if everybody like in August when everybody's on vacation or something we can dress a little more casual or	*contradiction*	August is a black out month for vacations in the company.
9/11 Report		
At the other end of Pennsylvania Avenue, people began to line up for a White House tour.	*entailment*	People formed a line at the end of Pennsylvania Avenue.

Data Pipeline, with State of the Union as example

395 sentences

370 well-transcribed sentences

199 fact-checkable claims

30 matches

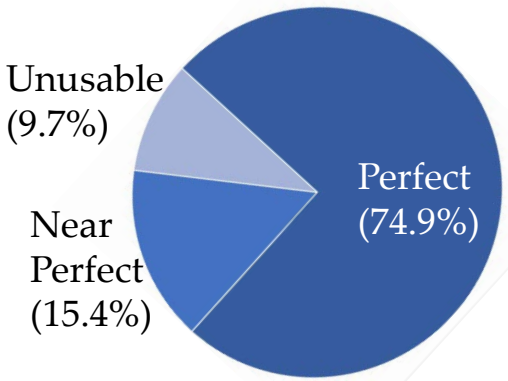
Audio Collection
We collect audio from 10 television channels, such as CNN, MSNBC and Fox, as well as from the microphone. In this example, we demonstrate the process of passing Trump's 2018 State of the Union through our pipeline.

Transcription
We use the Google Cloud Speech-to-Text API to transcribe audio. We filter out sentences with a transcription confidence of less than 0.9 (~5% of sentences).

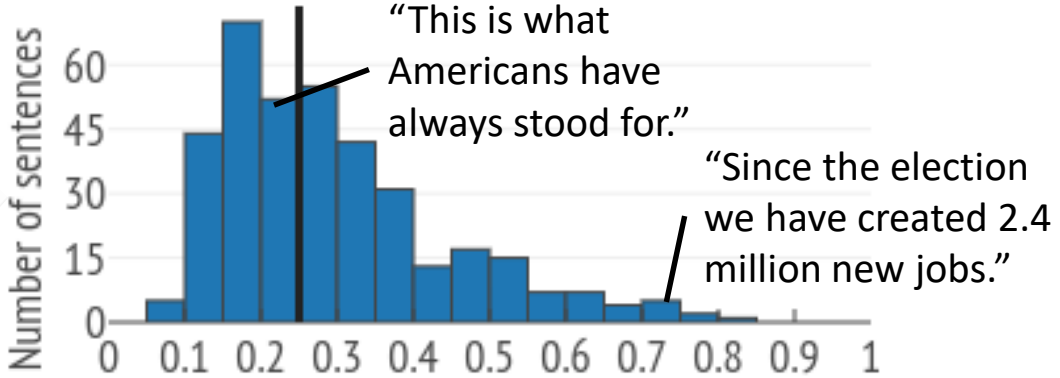
Claim Filtering
We use the ClaimBuster API with a threshold of 0.25 to filter out sentences which are not check-worthy factual claims.

Claim Matching
We find related fact-checks by comparing each claim against the Share the Facts database. Our model is described in detail below.

User Interface
We built a simple interface to present matches to users for a selected input source.



Transcription quality



ClaimBuster score distribution for Trump's State of the Union.

Matching Model

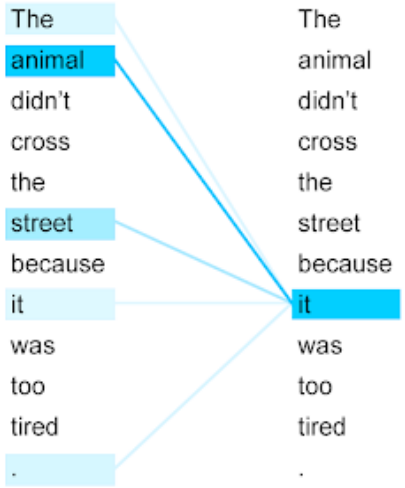
Our matching model is an adaptation of a model developed by OpenAI and released in June 2018. This model attempts to solve two problems which have plagued traditional NLI approaches:

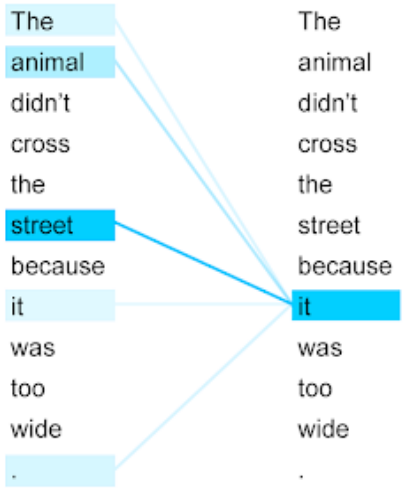
1. The difficulty of identifying word dependencies in English
2. The limited quantity of labeled NLI data.

The first problem is solved by the transformer framework, a type of neural network created by Google researchers in Dec 2017. The transformer improves upon previous architectures (CNN, RNN, RNN w/LSTM) through use of attention mechanisms which determine meaning through a weighted average of dependency on all other words in the sentence.

The second problem is solved by an unsupervised preprocessing step. OpenAI used a large corpus of books to "teach" the transformer English, and with minimal adaptation was able to transfer this unsupervised learning to beat the state-of-the-art result on 12 of 15 common NLP datasets.

Using OpenAI's weights from the unsupervised learning step, we were able to achieve near state-of-the-art results on the Multi-Genre NLI dataset (MNLI) (81.6% compared to 82.1%). When the model does not attempt to distinguish between entailment and contradiction, the accuracy increases to 85.8%.





The transformer correctly identifies that "it" refers to "animal" in the first sentence and "street" in the second.