Stance detection in news articles using Machine Learning

Shardul Naithani Rohit Satish Shinde Aravind Sivaramakrishnan

Department of Computer Science

Rutgers, The State University of New Jersey

Spring 2017

Outline

The Big Picture

- Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts.

The Big Picture

- Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts.
- In complex Machine Learning problems, the process can be broken down into stages that are easier to model and solve.

The Big Picture

- Assessing the veracity of a news story is a complex and cumbersome task, even for trained experts.
- In complex Machine Learning problems, the process can be broken down into stages that are easier to model and solve.
- Similarly, a helpful first step towards identifying fake news is to understand what other news organizations are publishing on the same topic (**stance detection**).

Stance Detection

- Given a headline and a body of text - either from the same news article or from different news articles, we aim to classify the stance of the body text relative to the claim made in the headline into one of four categories: *Agrees*, *Disagrees*, *Discusses*, or *Unrelated*.

Stance Detection

- Given a headline and a body of text - either from the same news article or from different news articles, we aim to classify the stance of the body text relative to the claim made in the headline into one of four categories: *Agrees*, *Disagrees*, *Discusses*, or *Unrelated*.
- We aim to implement some novel machine learning algorithms on the Fake News Challenge Stage 1: Stance Detection (FNC-1) dataset.

Dataset Description

- **Input:** A *headline* and a *body* text - either from the same news article, or from different articles.
- **Output:** The stance of the body text relative to the claim made in the headline into one of four categories:

Dataset Description

- **Input:** A *headline* and a *body* text - either from the same news article, or from different articles.
- **Output:** The stance of the body text relative to the claim made in the headline into one of four categories:
  - **Agrees:** The body text agrees with the headline.
  - **Disagrees:** The body text disagrees with the headline.
  - **Discusses:** The body text discusses the same topic as the headline, but does not take a position
  - **Unrelated:** The body text discusses a different topic than the headline.

Example data points

Example headline

Example data points

EXAMPLE HEADLINE

"Robert Plant Ripped up $800M Led Zeppelin Reunion Contract"

Example headline

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

CORRECT CLASSIFICATION: AGREE

"... No, Robert Plant did not rip up an $800 million deal to get Led Zeppelin back together. ..."

CORRECT CLASSIFICATION: DISAGREE

"... Robert Plant reportedly tore up an $800 million Led Zeppelin reunion deal. ..."

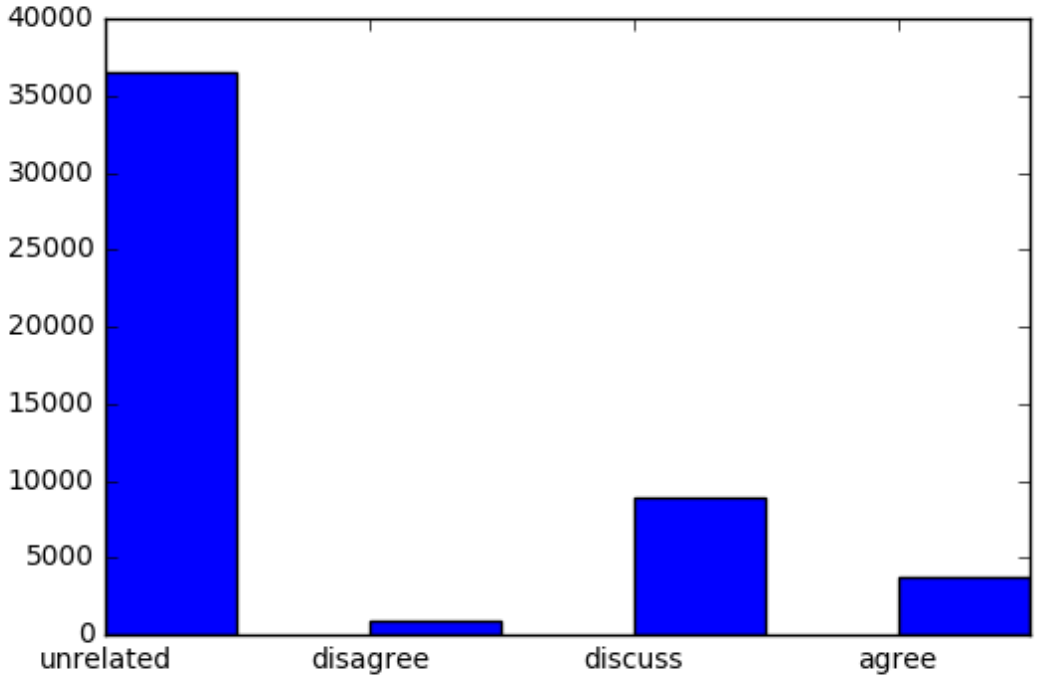
CORRECT CLASSIFICATION: DISCUSSES

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED

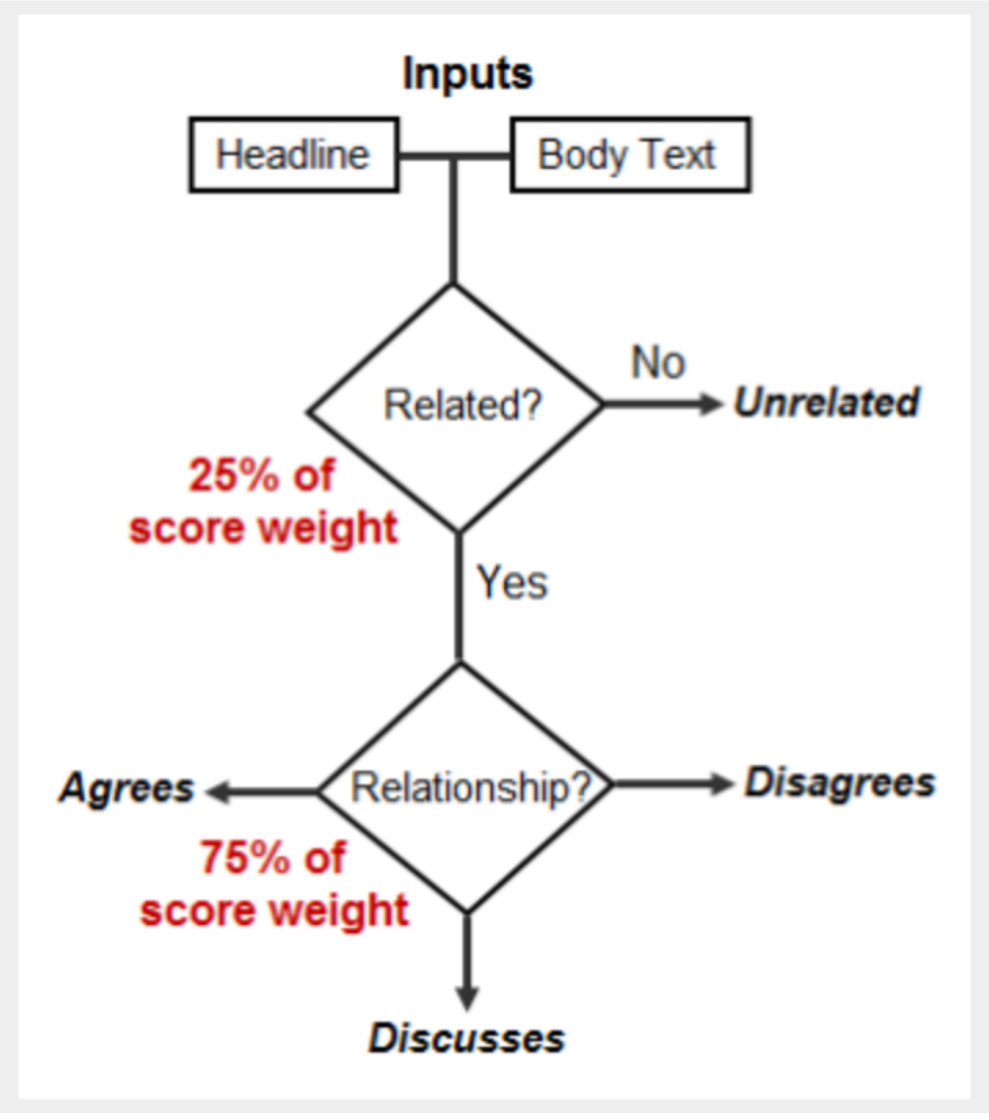
Example body snippets and their correct classification

Dataset details



Distribution of labels in the dataset

Evaluation scheme



Proposed model - Level 1 Classifier

- The Level 1 task can be thought of as determining whether a pair of body and headline text are semantically similar or not.

Proposed model - Level 1 Classifier

- The Level 1 task can be thought of as determining whether a pair of body and headline text are semantically similar or not.
- Parts of Speech tags (Nouns, Verbs, Adjectives) are an effective way of determining the synctactic role of a token in a text.

Proposed model - Level 1 Classifier

- The Level 1 task can be thought of as determining whether a pair of body and headline text are semantically similar or not.
- Parts of Speech tags (Nouns, Verbs, Adjectives) are an effective way of determining the synctactic role of a token in a text.
- We extract features based on overlap of tokens belonging to a particular tag class (Verb, Noun, Adverb, Adjective) between body text and headline.

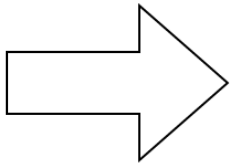
Level 1 - feature extraction

EXAMPLE HEADLINE

"Robert Plant Ripped up $800M Led Zeppelin Reunion Contract"

"... Led Zeppelin's Robert Plant turned down £500 MILLION to reform supergroup. ..."

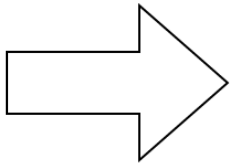
CORRECT CLASSIFICATION: AGREE



0.66 0.12 0.43

"... No, Robert Plant did not rip up an $800 million deal to get Led Zeppelin back together. ..."

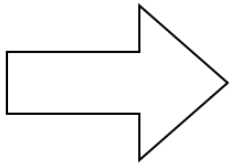
CORRECT CLASSIFICATION: DISAGREE



0.52 0.32 0.16

"... Robert Plant reportedly tore up an $800 million Led Zeppelin reunion deal. ..."

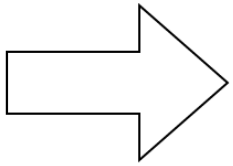
CORRECT CLASSIFICATION: DISCUSSES



0.12 0.25 0.22

"... Richard Branson's Virgin Galactic is set to launch SpaceShipTwo today. ..."

CORRECT CLASSIFICATION: UNRELATED



0.0 0.05 0.14

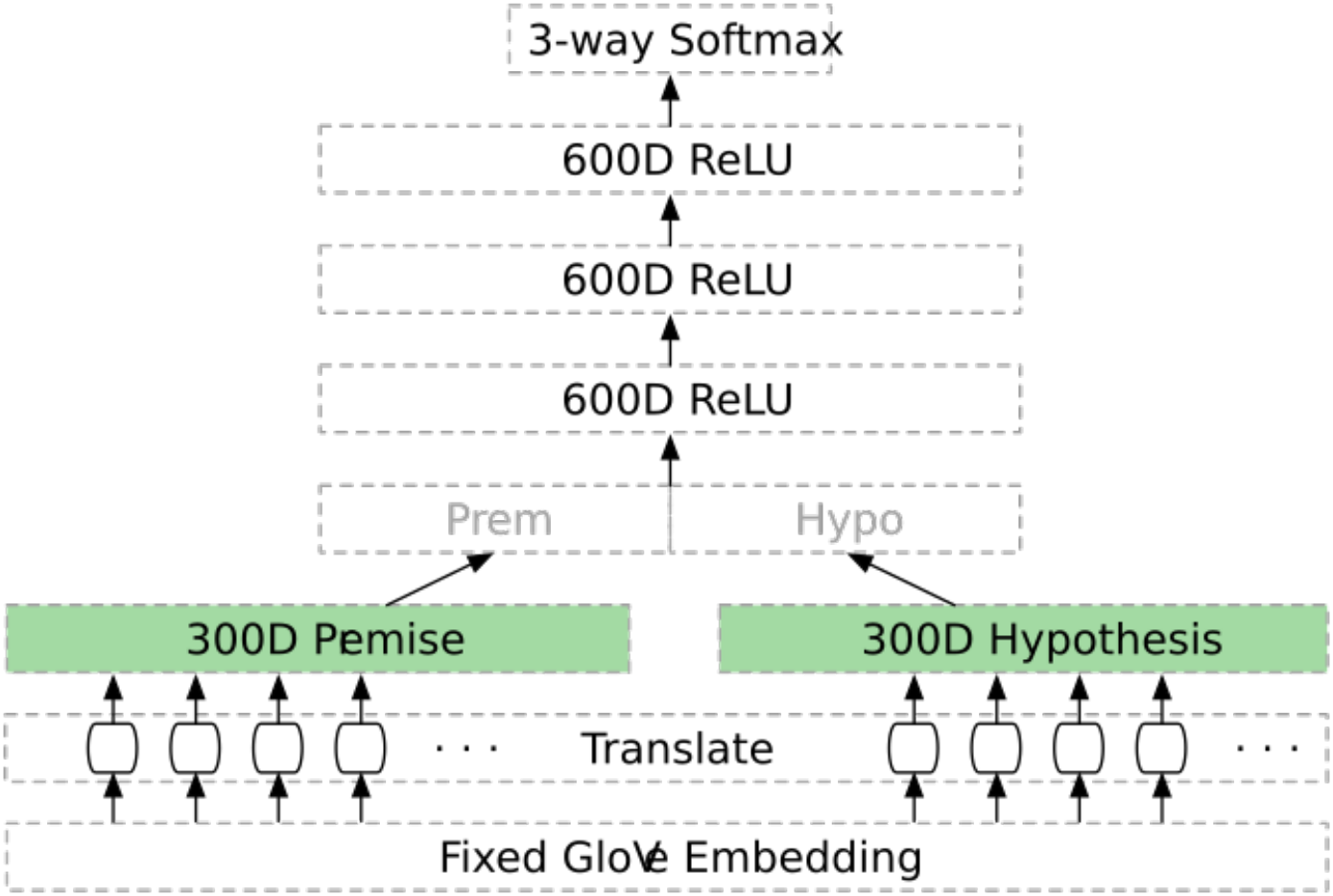
Proposed Model - Level 2 Classifier

- The task of determining whether a body text (premise) agrees, disagrees or discusses a particular headline (hypothesis) is very similar to the Stanford Natural Language Inference (SNLI) corpus.

Proposed Model - Level 2 Classifier

- The task of determining whether a body text (premise) agrees, disagrees or discusses a particular headline (hypothesis) is very similar to the Stanford Natural Language Inference (SNLI) corpus.
- We train a few simple models inspired by Bowman et al. (2015) and Bowman et al. (2016) to tackle the second stage classifier.

Model Architecture



The "translation" layer is an encoder such as summation, GRU or LSTM.

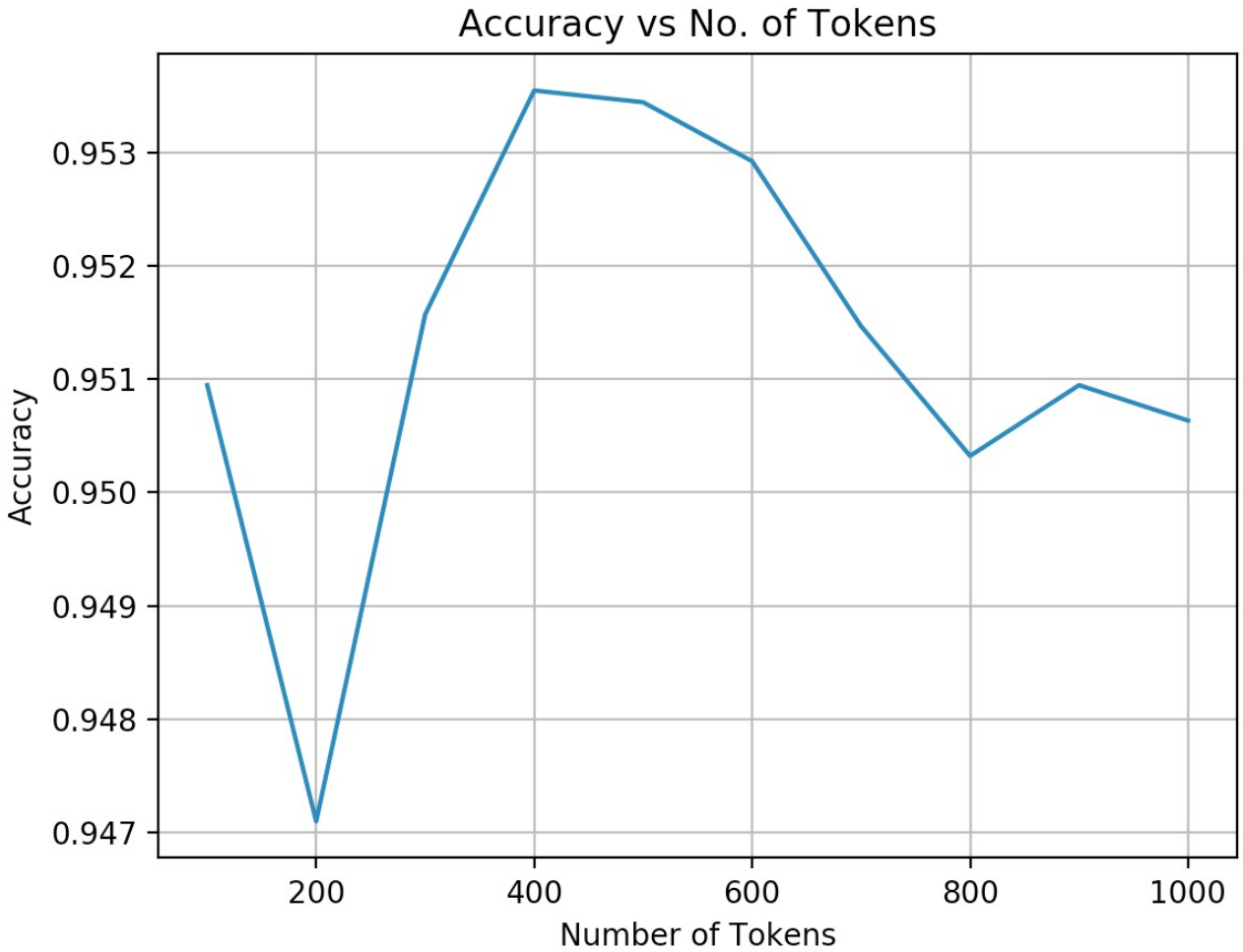
5 Fold Validation Results - Level 1

Fold	1	2	3	4	5
Logistic	0.96	0.95	0.95	0.95	**0.97**
SVM	0.96	0.95	0.95	0.95	**0.96**
1-Layer NN	0.96	0.95	0.95	0.95	**0.97**

Level 2: Gory Details

- We used Keras to implement our model.
- We use a dropout rate of 0.2, batch normalization, and RMSProp for training.
- Following Liu et al. (2016), we do not update the GloVe embeddings during training.
- Following Munkhdalai & Yu (2016), the out of vocabulary embeddings are set to zero.
- Due to lack of compute, we restrict ourselves to the first few lines of each body text while training/testing.

How many words do you need for good classification?



Accuracy of Level 1 on test data when only first n words are seen

5 Fold Validation Results - Level 2

Fold	1	2	3	4	5
Sum	**0.78**	0.76	0.73	0.69	0.77
LSTM	0.72	0.71	**0.75**	0.73	0.71
GRU	0.69	0.72	0.72	0.77	**0.78**

Overall pipeline accuracies

Model	Score
Logistic Regression + SumRNN	83.73%
SVM + SumRNN	83.47%
NN + SumRNN	83.44%
Logistic Regression + LSTM	82.09%
SVM + LSTM	82.07%
NN + LSTM	82.05%
Logistic Regression + GRU	81.69%
NN + GRU	81.64%
SVM + GRU	81.67%
Baseline	**79.53%**

Questions?