DeepNewsNet: Automated Fake News Classification

Alaukik Aggarwal
alaukika@Stanford.edu
https://youtu.be/qYqBCtlEaBM

Introduction

Fake news has affected everyone. It has been at core of several instances of inciting mob, causing riots, influencing elections, picking our leaders, and causing failure of justice.

To identify whether a statement is fake or not, Different approaches have been explored in this effort:

- Different models – FC NN, LSTM, Hybrid
- Text vectorization techniques - TF-IDF and GloVe [2] word embeddings
- Use of sentiment and statement context

Dataset

LIAR dataset [1] is used, which has 12,805 short political statements with six classes of output labels, and having text and non-text attributes:

- Text: statement, subject, speaker, speaker's job title, state info, party affiliation, location
- Not-text: history of count of statements by label type for speaker

The dataset is randomly split 80-10–10 between train, validation, and test. Label types are uniformly split between all labels (17-19%), except 'pants-on-fire' label (~10%).

Input Features

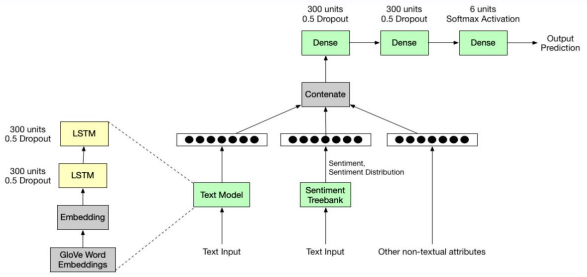
Different combination of features have been compared:

- Statement + All text attributes
- Statement + All text attributes + Sentiment feature
- Statement + All text attributes + Sentiment feature + All non-text attributes

Distribution of sentiments are generated using Sentiment Treebank [3]

Architecture

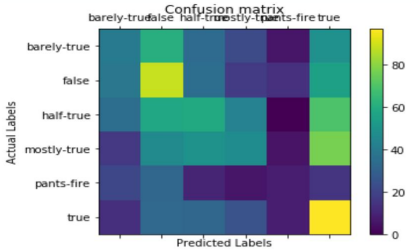
- Deep fully-connected neural network having 9 fully-connected layers with 300, 300, 300, 200, 200, 200, 100, 100, and 6 hidden units.
- LSTM model having two LSTM layers having 300 hidden units with dropout regularization, followed by a dense layer having 6 hidden units.
- Hybrid architecture that combines LSTM model for representing text features and non-textual attributes like sentiment distribution and speaker history using fully-connected layers



Results and Discussion

Model	Features	Accuracy
Existing LIAR Dataset Benchmark (CNN)	Text + non-text attributes	0.274
Shallow fully-connected (baseline)	Text (TF-IDF) only	0.25
Deep fully-connected	Text (TF-IDF) only	0.252
Deep fully-connected	Text (GloVe) only	0.258
LSTM	Text (GloVe) only	0.255
Hybrid	Text (GloVe) + Sentiment	0.294
Hybrid	Text (GloVe) + Sentiment + non-text attributes	0.402

Hybrid model performs significantly (~50%) better that existing LIAR dataset benchmark. Improvement of model performance with addition of sentiment features highlights the aspect of strong emotions in fake news. Further, significant boost in accuracy by adding history of speeches for a given speaker show the potential of context-aware models.



'true' and 'false' are classified with high confidence, however it gets fuzzy with other labels, especially for similar labels like mostly-true and half-true.

Future Work

With more time and hardware resources, it would be interesting to explore using higher dimension of word embeddings, using deeper network, and constructing a pipeline model to train separately on each text and non-textual attributes and derive predictions from combination of them.

References

1. Wang, William Yang. *""Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection."* ACL (2017).
2. Jeffrey Pennington, Richard Socher, and Christopher D. Manning. *"GloVe: Global Vectors for Word Representation."* EMNLP (2014).
3. Socher, Richard, et al. *"Recursive deep models for semantic compositionality over a sentiment treebank."* Proceedings of the 2013 conference on empirical methods in natural language processing (2013).