Fact Checking Fake News

Author: Bart Melman, Supervisors: Marcus Tomalin, Bill Byrne

MPhil Thesis for MPhil Machine Learning and Machine Intelligence, Cambridge

Introduction

The increasing concern with regard to "Fake News" in politics and social media has stimulated the development of automated fact checking systems. As a result, the FEVER challenge is setup to encourage and compare fact checking systems. My research is focused a improving on the current literature and investigate algorithms that are scalable and quick.

Fever Challenge

Data

Corpus 6 million Wikipedia pages
Training Set 200 thousand claims

Objective

1. **Label each claim as *CORRECT*, *INCORRECT* or *NOT ENOUGH INFO***
2. **Provide the correct evidence sentence(s) from the corpus**

References

- Chen, Q., Zhu, X., Ling, Z., Wei, S., Jiang, H., and Inkpen, D. (2016). Enhanced lstm for natural language inference. *arXiv preprint arXiv:1609.06038*.
- Malon, C. (2019). Team papelo: Transformer networks at fever. *arXiv preprint arXiv:1901.02534*.
- Nie, Y., Chen, H., and Bansal, M. (2018). Combining fact extraction and verification with neural semantic matching networks. *arXiv preprint arXiv:1811.07039*.
- Yoneda, T., Mitchell, J., Welbl, J., Stenetorp, P., and Riedel, S. (2018). Ucl machine reading group: Four factor framework for fact finding (hexaf). In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 97–102.

Stage 1: Document Retrieval

Objective

"Select K Wikipedia pages from the corpus which are closest to the claim"

Current Models

1. **TF-IDF** with bigrams [Malon, 2019]
2. **Logistic regression model** using the features in the claim and and title of a document [Yoneda et al., 2018]
3. **Exact matching** between the title and part of the claim [Nie et al., 2018]

Challenges

1. **Iterating through documents is computationally expensive**
   $>$ 5 million iterations per claim
2. **Wikipedia pages do not always have representative titles**
   Extremely short titles as 'Steve_Wozniak'
3. **The systems are not robust to different methods of claim generation**
   For the FEVER Challenge, claims are generated by adapting sentences directly from the text and closely match the sentences/titles.

Proposed Method

1. **Store list occurances per n-gram**
   factor 100 less iterations
2. **Robust to unrepresentative titles**
   The system uses the text of the pages itself instead of the titles and assigns a higher score to words that occur less in other documents.
3. **Entity matching using lemmatization**

Stage 2: Sentence Retrieval

Objective

"Select L sentences from K documents"

Current Models

1. **Logistic regression model** using the features in the claim and and title of a document [Yoneda et al., 2018]
2. **Neural Semantic Matching Network** [Nie et al., 2018]

Challenges

1. **There might not be L good potential evidence sentences.** Adding uninformative sentences to the last stage adds noise.

Proposed Method

1. **Make L dependent on a threshold and the scores of the sentences**
   Make the number of selected sentences dependent on a threshold

Example Claim

Claim: 'Mount Rushmore was created only by Abraham Lincoln.'
Label: INCORRECT
Evidence: Sculptor Gutzon Borglum created the sculpture 's design and oversaw the project 's execution from 1927-1941 with the help of his son , Lincoln Borglum

Stage 3: Score Retrieval

Objective

"Predict claim label from L sentences using natural language inference (NLI)"

Current Models

1. Adapted versions of the **Enhanced Sequential Inference Model (ESIM)**, [Chen et al., 2016]

Proposed Method

1. **Hybrid Inference Model (HIM)** by [Chen et al., 2016] using **tree-LSTM** with a constituency tree.
2. Using all sentences from K documents to make predictions

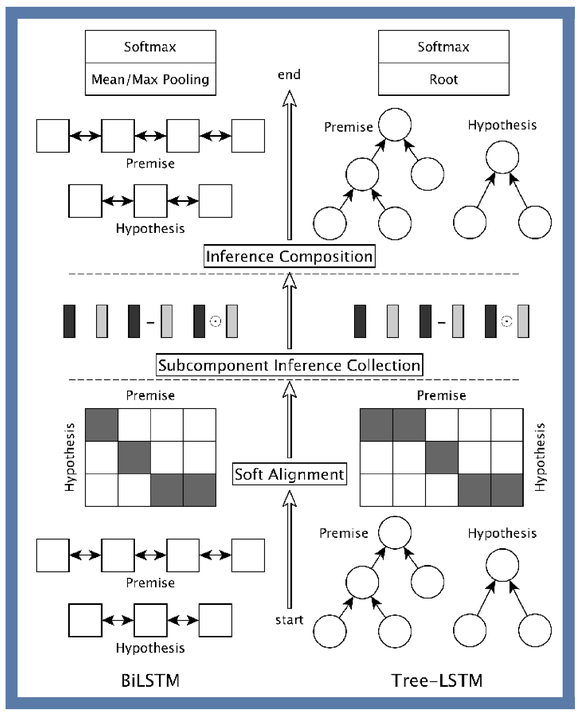

Figure 1: Hybrid Inference Model using tree-LSTM