

The Case for Multi-Task Learning: Using Natural Language Inference to enhance performance on Fake News Detection



Introduction

This academic poster concerns itself with our research into to the case for Multi-Task Learning (MTL) in the context of Fake News Detection (FND) and Natural Language Inference (NLI). Specifically, the research question under investigation is whether training and evaluating a deep neural network (DNN) model on NLI as an *auxiliary* task enhances performance on FND as the main task.

Model

The model we designed and implemented can be reviewed in Fig. 1, and draws inspiration from the Hierarchical Attention Network (HAN) as introduced by Yang et al. (2016), as well as from the Multi-Task Deep Neural Network (MT-DNN) as introduced by Liu et al. (2019).

As such, our model boasts deep, contextual word embeddings and an attentive sentence encoder, which are shared for both tasks. The sentence embeddings produced by this encoder are used as input for either the NLI classifier for our *auxiliary* task, or the attentive document encoder used in the FND classifier.

Both the attentive sentence and document encoder use a bidirectional LSTM (Hochreiter & Schmidhuber, 1997), combined with an attention mechanism at the word- and sentence level, respectively, to produce their embeddings.

Platform	Fake / real	# of articles
Gossipcop	Fake	5,323
	Real	17,817
PolitiFact	Fake	432
	Real	624
	Total	23,196

Table 1. FNN dataset statistics

Experiments

Datasets

For the FND task, we used the FakeNewsNet dataset as published by Shu et al. (2018), some statistics of which can be reviewed in Tab. 1. This data-

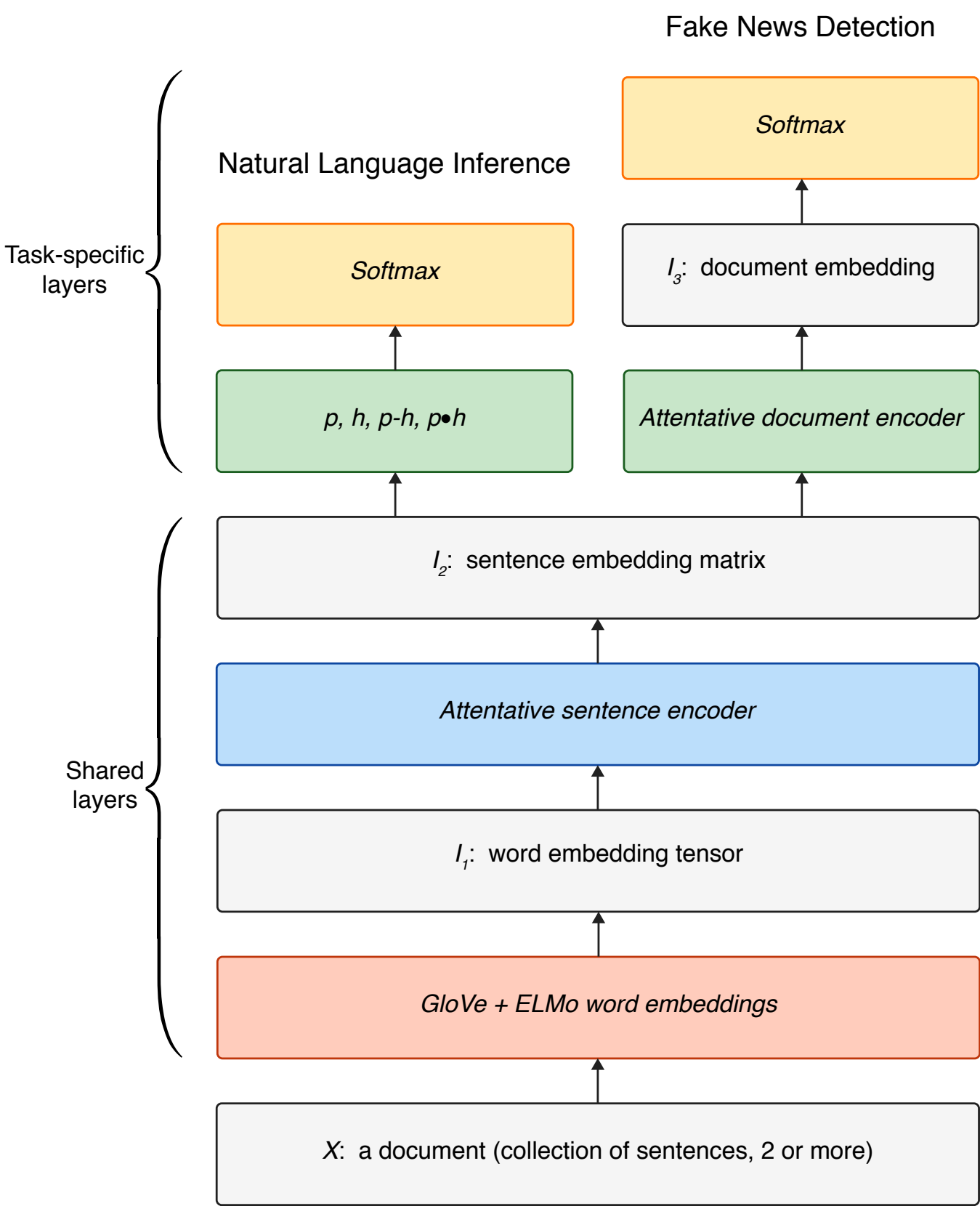


Figure 1. Our model architecture set did not come with the articles themselves, but instead with the article titles and URLs (as well as tweets, which we did not use), as well as software to scrape them.

After scraping, the data collected was quite noisy. The scraper ran into some paywalls and privacy notices, for example. Another issue with the data, was that some documents were extremely long (in terms of the number of sentences), as can be reviewed in Fig. 2. As such, we performed data cleaning to remove any such articles that either ran into paywalls / privacy notices or were too long (> 40 sentences), leaving 20,467 articles with 375k sentences in total.

For the NLI task, we used the Stanford NLI (SNLI) corpus, as introduced by Bowman et al. (2015), which contains 570k sentence-pairs with entailment labels.

Training procedure

In order to investigate our research question, we trained our model as introduced and described pre-

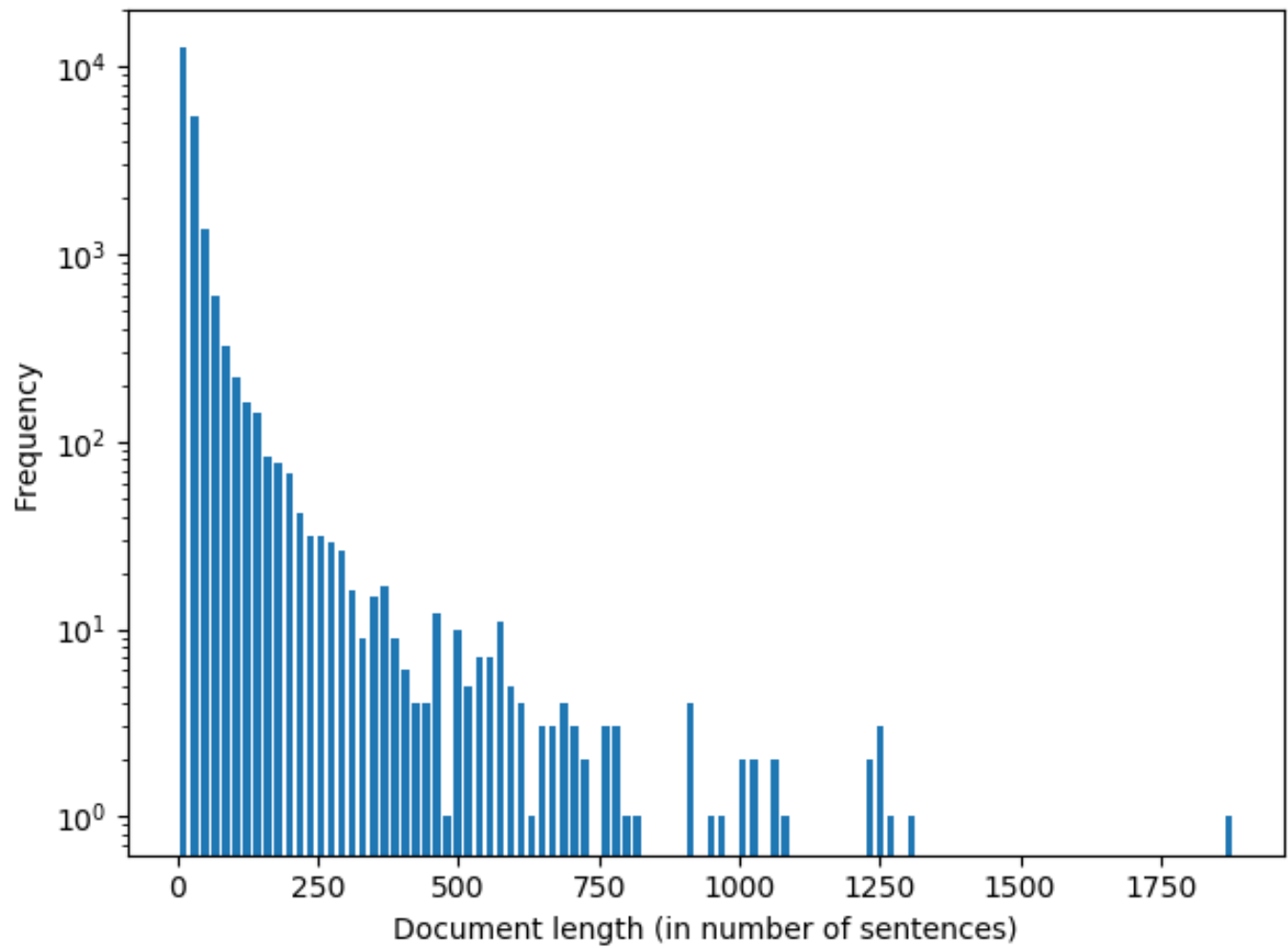


Figure 2. Plot of document length frequencies in FNN viously twice: Once with our Single-Task Learning (STL) objective, where we just trained and evaluated on our FND task, and once with our MTL objective, where we trained and evaluated our model on both the FND task, as well as the NLI task. In both cases, we trained our models for 10 epochs.

For the MTL objective, in any given epoch, the task to be trained was sampled uniformly at random, with a probability of XX. Furthermore, the combined loss of the FND task and NLI task looks as follows:

XX

Baselines

The first and most obvious baseline for our MTL learning objective, is that of the STL objective, as the only difference between the two is the addition of the *auxiliary* NLI task during training, thus allowing for direct comparisons between STL and MTL for FND.

A second baseline would be XX

Results

After training our model for the STL objective for 10 epochs, the following results were achieved:

XX

Unfortunately, due to time constraints, we have

not yet been able to train our model for the MTL objective for 10 (or any) epochs. As such, we cannot present any results for this objective yet.

Analysis

XX

References

Bowman, S. R., Angeli, G., Potts, C., & Manning, C. D. (2015). A large annotated corpus for learning natural language inference.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8), 1735-1780.

Liu, X., He, P., Chen, W., & Gao, J. (2019). Multi-Task Deep Neural Networks for Natural Language Understanding.

Shu, K., Mahudeswaran, D., Wang, S., Lee, D., & Liu, H. (2018). Fake-newsnet: A data repository with news content, social context and dynamic information for studying fake news on social media.

Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., & Hovy, E. (2016). Hierarchical attention networks for document classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 1480-1489).