Group 44 Authors: Fan Mo and Zixiao Nong

EVIDENCE DETECTION FOR CLAIM–EVIDENCE PAIRS: LSTM– ATTENTION AND DEBERTA APPROACHES

Introduction

We present two neural network approaches for the task of evidence detection, determining whether evidence is relevant to a given claim:

1. First solution implements a BiLSTM architecture with cross-attention mechanisms inspired by ESIM [1]
2. Second solution combines DeBERTa [2] transformers with a forward LSTM and self-attention layers [3] to improve classification performance.

Both solutions trained on around 21k claim-evidence text pairs, and evaluated by around 5k pairs that haven't occurred in training data. To show performance improvement of advanced model architecture, we evaluate them on weighted macro metrics plus Matthews correlation coefficient, comparing against simpler baselines.

Data Preprocessing

Solution1
For the LSTM-based model, we implemented a Part-of-Speech (POS) tokenizer based on NLTK package. Comparing with a simple traditional tokenizer, it improved 4% accuracy in average for the same model. A vocabulary was build based on training data and this tokenizer. Each claim/evidence is then converted to indices, padded, and fed to the model.

Solution 2
In the DeBERTa approach, we rely on the DeBERTa-v3-base tokenizer from Hugging Face [4] to create input IDs, attention masks, and token type IDs. To align the input of the DeBERTa model all truncated or padded to a fixed length.

Solution 1 Methodogy

Lstm–Attention

Figure 1 shows the architecture of the model. This model implements a variant of the Enhanced Sequential Inference Model (ESIM) [1], which extend the ESIM model from solving Natural Language Inference (NLI) task to the Evidence Detection (ED) task. The model processes claim and evidence text separately through embedding and BiLSTM encoding layers, applies cross-attention between them, and enhances the representation with element-wise difference and multiplication operations. A second BiLSTM composition layer and pooling operations prepare the final representation for classification. To improve the model's ability to classify difficult samples, we used Focal Loss [6] to optimise the model parameters.

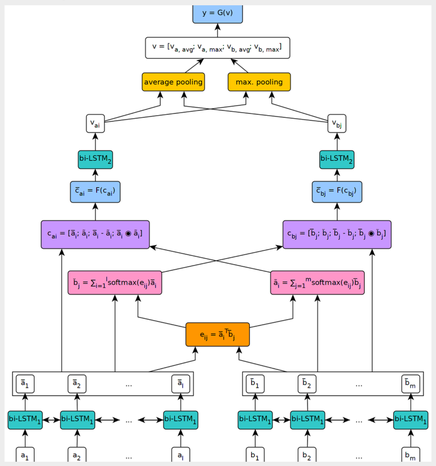

Figure 1: ESIM model Architecture [5]

Solution 2 Methodogy

Deberta-v3 and Lstm–Attention

This model fine-tunes DeBERTa-v3-base [4] on the training evidenc-claim texts, but compared with a traditional BERT fine-tune process for text sequence classification, which only a simple fully-connected classification layer, this model adds a bidirectional LSTM followed by a self-attention layer to discover more hidden information from DeBERTa outputs. The architecture processes claim-evidence pairs through the transformer encoder, captures sequential dependencies with the LSTM, generating a final context vector through self-attention, which is passed through a classification head. The attention mechanism employs a learnable attention vector to weight token representations, merge token-level features into a sentence-level feature vector [3]. *Figure 2* shows the model Architecture.

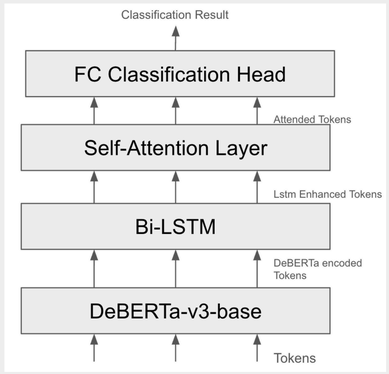

Figure 2: Solution 2 Model Architecture

Solution 1 Result

Compared with baseline bi-Lstm model implemented by us, which uses a simple Bi-Lstm encoder followed by a classification head, our model achieves a significant increase in all metrics:

Model / %	Precision	Recall	F1	MCC
Simple Lstm	0.7621	0.7320	0.7467	0.4621
Our Model	0.8284	0.8331	0.8307	0.5691

Solution 2 Result

Compared with traditional BERT fine-tune process for text sequence classification, our model achieves a little increase in all metrics:

Model / %	Precision	Recall	F1	MCC
Finetune DeBERTa	0.8901	0.8817	0.8858	0.7293
Our Model	0.8937	0.8913	0.8922	0.7338

Conclusion

Our experiments confirm that attention-based architectures significantly enhance claim-evidence matching. The LSTM-attention method outperforms a simpler BiLSTM, while our DeBERTa-LSTM model achieves further gains in F1 and MCC compared to a standard fine-tuned DeBERTa. These outcomes highlight the synergy of combining pre-trained contextual embeddings with specialized sequence modeling and cross-attention operations. Overall, the proposed methods prove effective at distilling relevant content from evidential text to support or refute claims.

Reflection

Solution 1

- We only use the development set to test the model performance, which could leads the model overfit to the development set and reduce some general ability.
- The ESIM architecture [1] is first proposed for Natural Languange Inference (NLI) task, not Evidence Detection. Therefore, there may be some performance loss when migrating the ESIM architecture to the dataset of this task.
- We only tried one pre-trained word embeddings (word2vec-google-news-300). So the solution might limited by the quality and coverage of the pre-trained word embeddings. Other pre-trained or self-trained embeddings might have potential in improving model performance.

Solution 2

- We only use the development set to test the model performance, which could leads the model overfit to the development set and reduce some general ability.
- Limited by the maximum sequence length (256 tokens). Model performance may decrease for very long claim-evidence pairs due to truncation. May inherit biases present in the pre-trained DeBERTa model.
- The forward Lstm-attention did not significantly improve the model performance, but increased the model training time and required computing resources. Therefore, the effectiveness of this architecture deserves further study on other tasks.

References

[1] Chen et al. (2016). Enhanced LSTM for Natural Language Inference.
[2] He et al. (2020). DeBERTa: Decoding-enhanced BERT with Disentangled Attention.
[3] Zhou et al. (2016). Attention-based Bidirectional LSTM Networks for Relation Classification.
[4] https://huggingface.co/microsoft/deberta-v3-base
[5] https://github.com/coetaur0/ESIM
[6] Lin et al. (2017). Focal Loss for Dense Object Detection