Evidence Detection: Pairwise Sequence Classification

With SVM(TF-IDF) and BiGRU(GloVe)

DetectiveZ
Yaqi Zhao, Ruoqing Zheng

INTRODUCTION

Task: Evidence Detection (ED) : Given an evidence, determine whether the claim it relavent to it. It is a pairwise sequece classification.

Approach: 1) Traditional learning method: Support Vector Machine (SVM) with term frequency-inverse document frequency (TF-IDF). 2) Deep learning based method: Bi-directional Gated Recurrent Units model (BiGRU) with Global Vectors (GloVe)

METHODOLOGY

SVM is a supervised learning method, which is effective in high dimensional spaces.

TF-IDF is a measure of importance of a word to a document.

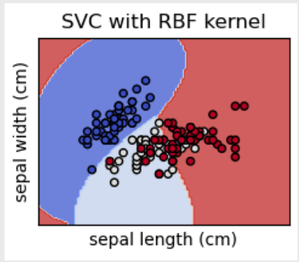

Figure 1. SVM

GRUs include reset and update gates.

A **Bi-GRU** contains two GRU layers and takes the input in both forward and backwards directions.

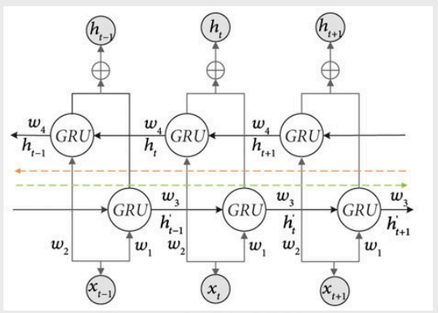

Figure 2. Bi-GRU

GloVe is an unsupervised learning algorithm for obtaining vector representations for words.

MODEL STRUCTURE

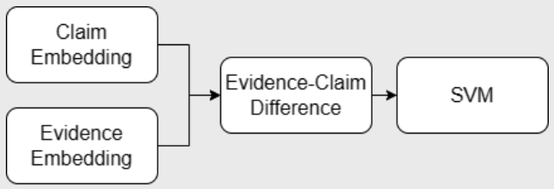

Figure 3. SVM Model Pipeline

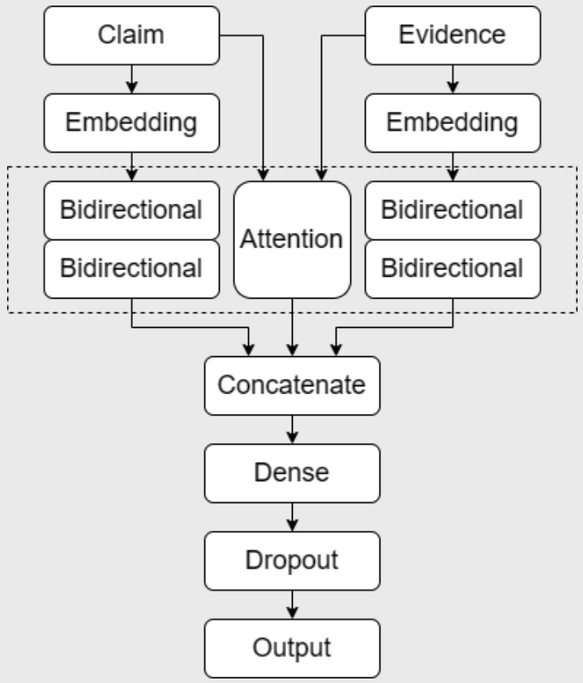

Figure 4. BiGRU Model Structure

For this task, we use the absolute of the difference between claim and evidence embeddings as the input in SVM model. In BiGRU, we add an attention layer to highlight the interesting parts of the claim and evidence. And then, we concatenate it with BiGru layer output.

RESULTS

Model	Accuracy	F1
LR(tfidf)	0.818	0.749
LR(w2v)	0.747	0.538
Naive Bayes(tfidf)	0.769	0.593
Naive Bayes(w2v)	0.730	0.422
SVM(tfidf, rbf)	0.821	0.752
SVM(w2v, rbf)	0.759	0.552
SVM(tfidf, linear)	0.817	0.754
SVM(w2v, linear)	0.732	0.427

Table 1. Comparison of traditional model performance

Metric	SVM(tfidf, rbf)	BiGRU(GloVe)
Accuracy	0.834	0.831
Precision	0.801	0.798
Recall	0.756	0.750
F1	0.774	0.768
ROC AUC	0.756	0.750

Table 3. Evaluation of final models

Model	Accuracy
GRU	0.771
BiGRU	0.792
BiGRU(GloVe)	0.831
BiLSTM(GloVe)	0.818

Table 2. Process of building BiGRU model

CONCLUSIONS

In the first approach, SVM(tfidf,rbf) becomes the final winner.🏆

In the second approach, both GRU and LSTM are implemented. BiGRU acchieves a higher accuracy compared to the baseline BiLSTM.

FUTURE WORK

- Data augmentation technique can be used to increase the data diversity, produce more robust and generalized classification results.
- Techniques like oversampling and downsampling can be applied to handle the mild inbalance in the dataset.
- Bert embeddings including sequence embeddings can be implemented.

REFERENCES

M. A. Hearst. "Support vector machines," in IEEE Intelligent Systems and their Applications, July-Aug. 1998, doi: 10.1109/5254.708428.

Dhingra N. BGT-Net: Bidirectional GRU transformer network for scene graph generation. IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021