Stance Detection using Hierarchical LSTMs

Kesha Bodawala, Sayema Mashhadi

University of Waterloo

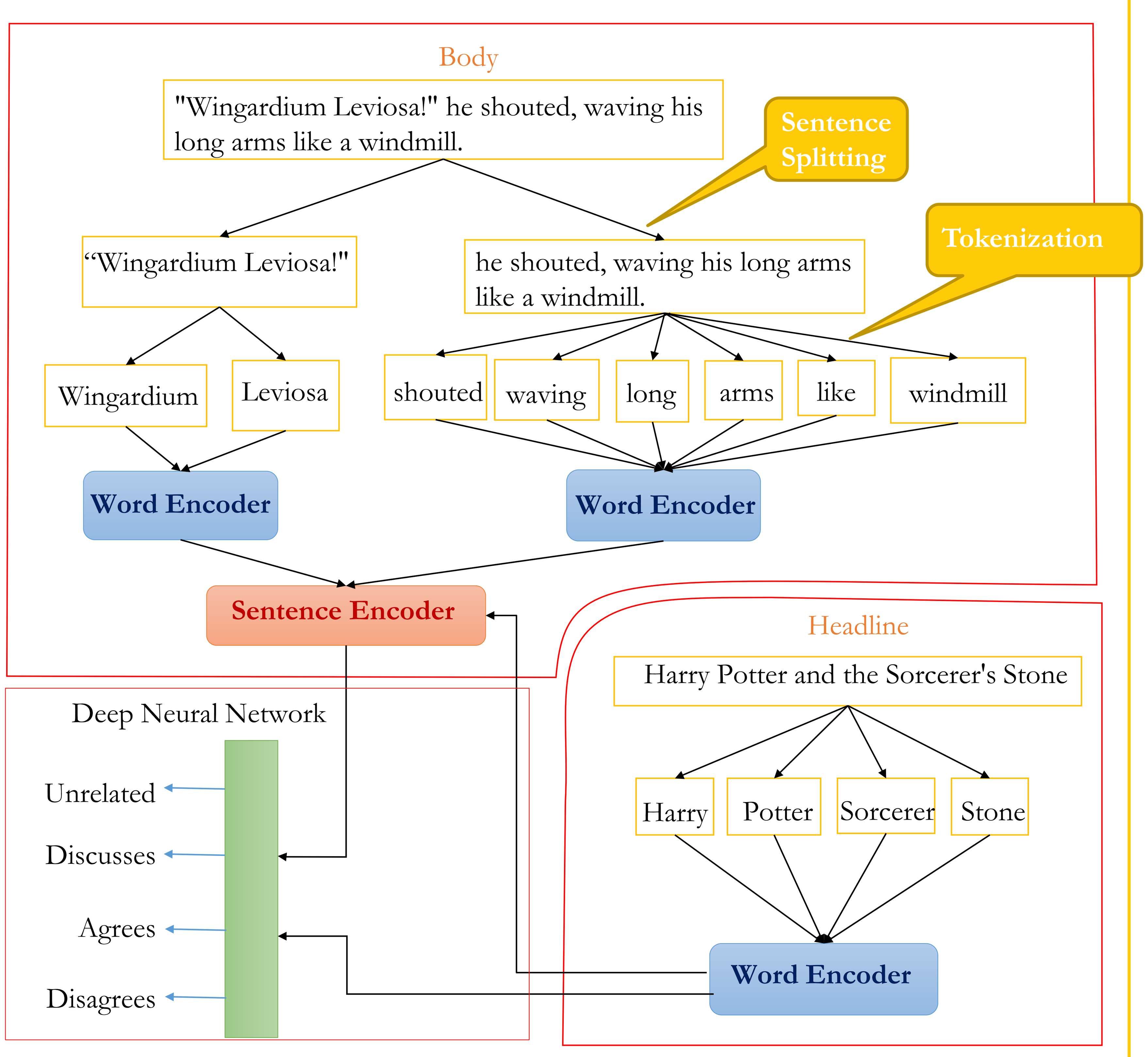
INTRODUCTION

Stance detection refers to detecting whether a body(paragraph) and a headline(title) are related or not. The proposed method is inspired from the paper 'Hierarchical Attention Networks for Document Classification'.

PRE-PROCESSING

- Body and headline are tokenized using nltk tokenizer
- Stopwords and punctuations are removed
- Vocabulary consists of the most common 20,000 words
- Words not in vocabulary are assigned 'UNK' token
- All the numbers are converted into a special 'NUM' token
- Bodies with equal number of sentences are grouped together to form batches

PROPOSED MODEL

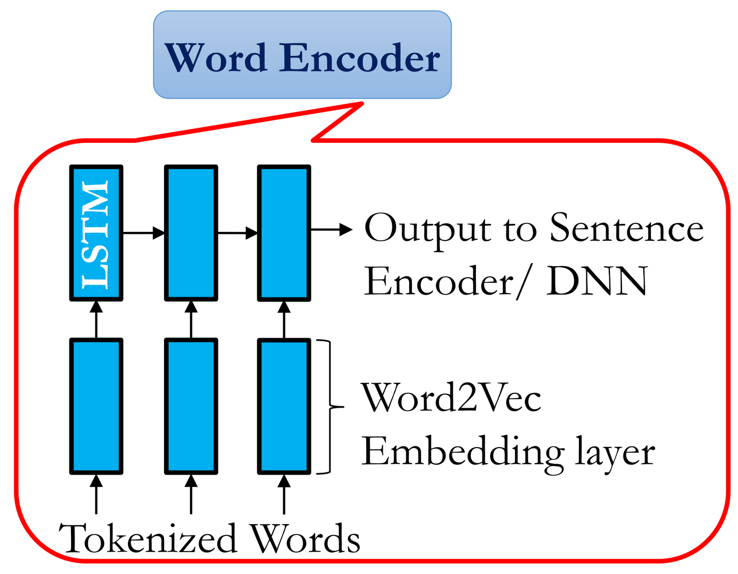


A "Word Encoder", which is similar to a *Sequence-to-Sequence* LSTM encoder, encodes headline into a vector representation. Since bodies contain multiple sentences, they are encoded using hierarchical LSTM encoders. Another "Word Encoder" encodes individual sentences of a body, while "Sentence Encoder" takes encoded sentences as input and outputs a vector representation for the entire body. To provide interaction between headline and body, hidden state of "Sentence Encoder" is initialized using output hidden state of headline's "Encoder Encoder".

Encodings of headline and body are concatenated and passed through Deep Neural Network Classifier to detect the correct class: unrelated, discusses, agrees, disagrees.

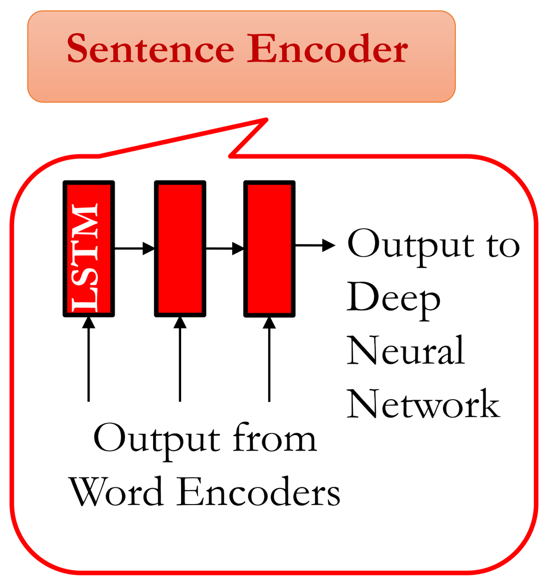
IMPLEMENTATION DETAILS

- L2 regularization and dropout in linear layer as well as in LSTM networks are used to prevent overfitting
- The dataset exhibits class imbalance. To overcome this problem, examples from smaller classes are given higher weights while calculating loss.
- "Warm Restart" learning rate schedule is used.



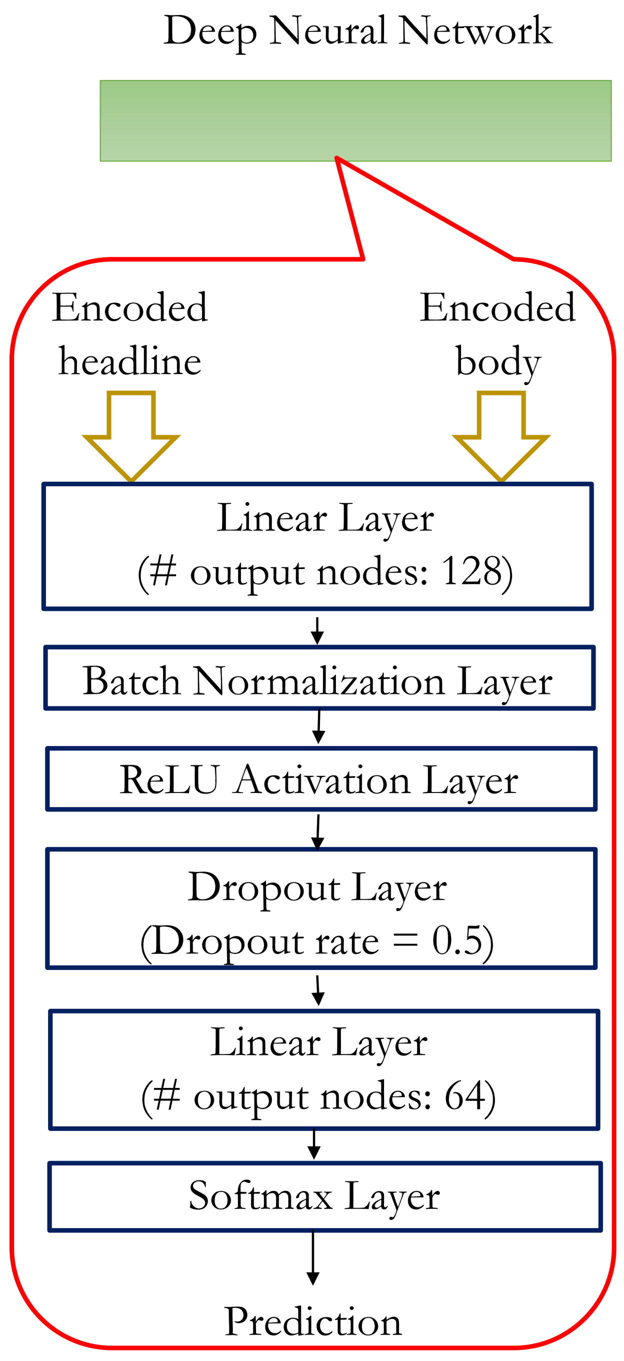
"WORD ENCODER":

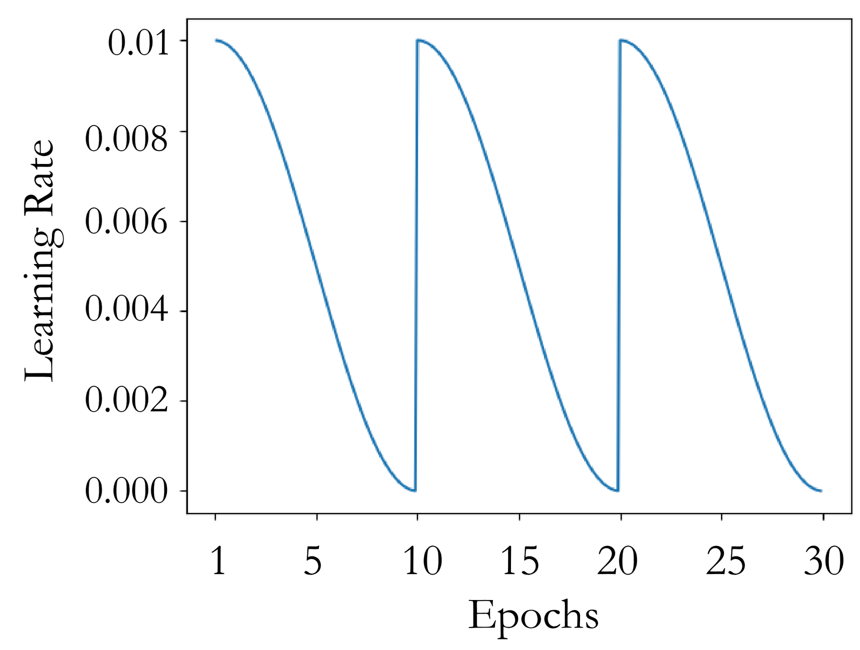
- 1-Layer LSTM encoder
- Hidden state size : 256 for Body, 128 for Headline
- Dropout rate : 0.5



"SENTENCE ENCODER":

- 1-Layer LSTM encoder
- Hidden state size : 128
- Dropout rate : 0.5





Warm Restart Learning Rate Schedule

DISCUSSION

Confusion Matrix	Class Size (%)	Predicted (%)			
		AGR	DAGR	DIS	UNR
Actual (%) AGR	5.83	50.00	24.65	15.49	9.86
DAGR	3.00	21.92	38.36	10.96	28.77
DIS	19.53	5.88	2.31	88.24	3.57
UNR	71.64	2.35	2.12	5.78	89.75

AGR-Agree, DAGR-Disagree, DIS-Discuss, UNR-Unrelated

- Validation accuracy achieved: 87%
- Smaller classes have higher misclassification rates. This shows that providing class weights is not enough to overcome class imbalance problem.

- Feeding hidden state of headline LSTM to body LSTM increases accuracy by ~2%
- Hierarchical structure of body network ensures that network learns efficiently without loss of information with the help of strategic batches of data.
- Using Word2Vec gave better results than using GloVe.

FUTURE DIRECTIONS

- Further features can be added in the layer before the dense layer, for example, polarity of body and headline.
- Attention mechanism can be used in both body and headline networks.
- Doc2Vec can be used for getting feature vectors of body text.