Quote Attribution using DNN.

Kaito Nakashima s1270120 **Supervised by Prof. MARKOV Konstantin**

Abstract

In recent years, applications and websites that post multiple news texts every day have appeared. At the same time, the technology of natural language processing has advanced, and research on various natural language processing tasks using news texts is being conducted. Among various tasks, I focused on Named-entity Recognition (NER). NER is an effort to mechanically extract named entities by adding dates, quantities, etc. to proper nouns such as organizations, places, and product names from input natural sentences. While I often see papers on extracting companies and people in news texts, I have not seen many papers on speakers and quotations in news texts, and I am interested in quotations in texts, so I conducted this study. I was looking for a paper about quotations in news texts and speakers and found an article introducing a dataset for direct quotation extraction and attribution [1]. This article serves as the inspiration for this research. Quotation extraction and Speaker attribution is a difficult task, which involves determining the span of a quotation and attributing each quotation to a speaker. The purpose of this study is to identify the quotation in a text and relate it to the speaker who uttered it. This can be applied to a variety of use cases. For example, in the case of news texts, identifying the speaker can increase the credibility of the assertion, enhance the credibility of the news, and improve the persuasiveness of the news. In the reference paper, a neural network-based speaker identification method is proposed for conversational sentences in English news texts. Using the same data as in the reference paper, I tried two alternative methods, BiLSTM and BERT + BiLSTM, and compare the results with those of the model used in the reference paper.

1 Introduction

A quotation is a speech, thought, or statement in a text. It can express an opinion, or statement, attributed to the speaker. As shown in Figure 1, there are three types of quotations: direct quotations, indirect quotations, and mixed quotations. The characteristics of each citation are explained below.

- **Direct Quotation** Repeat or quote what you hear. When using direct speech in a document, the words are placed directly between quotation marks (" ") without alteration.
- **Indirect Quotation** It is generally used to talk about the past, so it changes the tense of the words heard. Use the transitive verbs 'say', 'tell', 'ask', or use 'that' to guide the words you report. Do not use quotation marks.
- **Mixed Quotation** It is a method of psychological description, used mainly in novels, in which the author represents the feelings of the characters in the work from a third-person standpoint.

Direct quotation is important among the different types of citations. By attributing direct quotations to the speaker, the credibility and fairness of the news can be improved, and the persuasive power of the news can be improved. Direct quotations in the news are very important to ensure accountability and accountability. Since the data used in this study are labeled only for direct quotations and their speakers, only direct quotations and their speakers are extracted.

NEWS: COVID-19 PANDEMIC GETTING WORSE

There is such a voice in American society: **"Make America great again!"** In response, Trump said that **we will defeat the virus**. He shouted: **"America first and only America first!"** However, Dr. Fauci argued: **"Wear your mask."** The virus is called **"COVID-19"**. He stressed **the importance of "epidemic prevention and control"** and …

● **Direct Quotation.** ● **Mixed Quotation.** ● **Indirect Quotation.** ● **Non-Quotation.**

Figure 1: Examples of Quotation Types.

2 Model

I used the following sequence labeling methods and components. All model weights are fine-tuned to maximize the log-likelihood of the output corresponding to the ground-truth label using the cross-entropy loss. Neither BERT nor LSTM can handle text that is too long, so each paragraph in the entire text is determined as a context window to be input into the model. The following is a description of the model and its architecture.

2.1.1 Long Short Term Memory(LSTM)

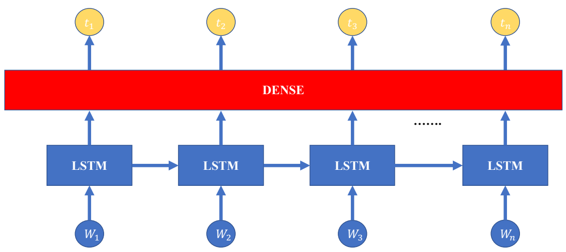

Figure 2: Architecture of LSTM

LSTM is a type of Recurrent Neural Network (RNN) that can learn long-term dependencies. Since its introduction by (Hochreiter & Schmidhuber, 1997) [2], LSTM has been shown capable of storing and accessing information over very long timespans in varied sequence labeling tasks such as POS tagging.

2.1.2 Bidirectional LSTM(BiLSTM)

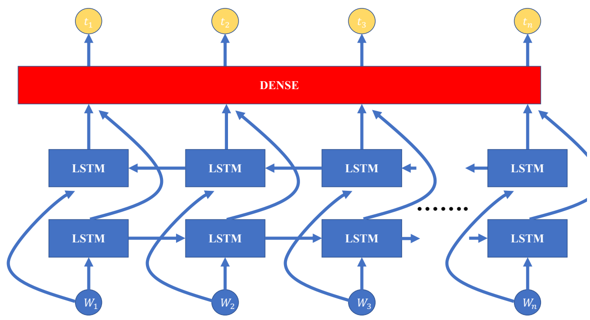

Figure 3: Architecture of BiLSTM

An LSTM is a network that learns in the order of the oldest time series and propagates the output values of the middle layer forward to the next state. In contrast, Bidirectional LSTM (BiLSTM) is a network that propagates the output of the middle layer in both directions, forward to the future and backward to the past.

2.1.3 Bidirectional Encoder Representations from Transformers (BERT)

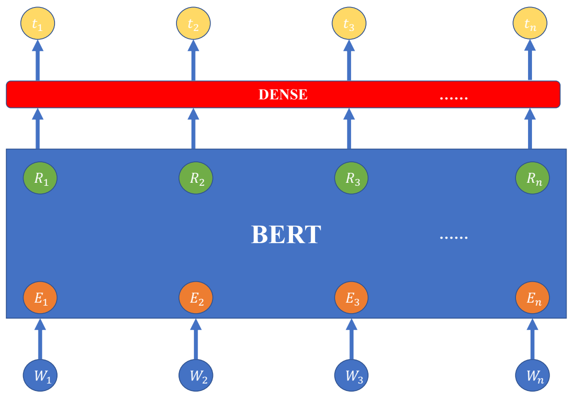

Figure 4: Architecture of BERT

BERT is a natural language processing model published by (Devlin, 2018) [3] in the paper. It can easily produce superior results for various natural language processing tasks such as translation, document classification, and question answering by fine-tuning a pre-trained model. In this study, a "bert-base-uncased" pre-training model is used for fine-tuning.

2.1.4 BERT-BiLSTM

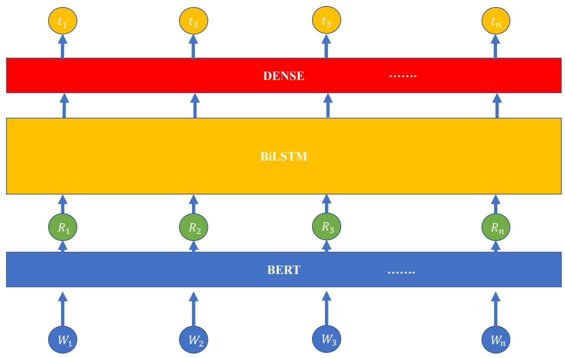

Figure 5: Architecture of BERT - BiLSTM

BERT - BiLSTM was introduced [4] when I was reading a paper researching Named Entity Recognition (NER) tasks using BIO format data, so I decided to use it at our laboratory. It consists of three parts as shown in Figure 5. First, a semantic representation of the input sentence is obtained by a language model pretrained by BERT, and then a vector representation of each word in the sentence is obtained. Then, the word vector sequence is input to BiLSTM, and finally, the tag sequence with the maximum probability is output. Compared to the BERT model, the addition of a BiLSTM layer would increase expressivity.

2.2 Beam Search Decoder

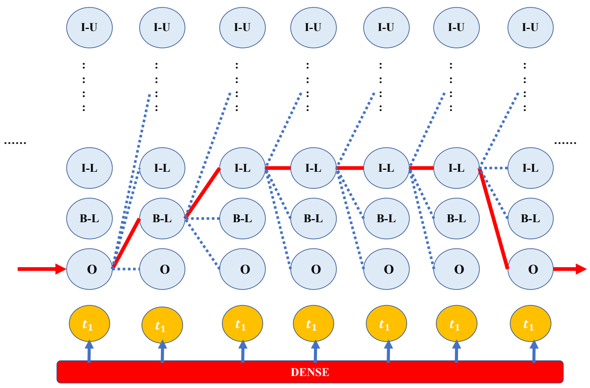

Figure 6: Transition example of Beam Search Decoder.

When predicting labels, if I use the greedy method, which selects the label with the highest score for each token (Table 1: NO), the result may not be as per the rules of the BIO format. Since this problem is caused by the fact that the scores of each token are determined independently of each other, a reasonable prediction can be obtained by using the prediction of the label sequence that has the highest sum of the scores of the labels for each token among the label sequences that follow the rules of the BIO format. The solution is to use the beam search decoder which is based on Viterbi algorithm [5]. The Viterbi algorithm solves one of the problems in Hidden Markov Models, which is to find the most likely label sequence among the observed label sequences. Figure 6 shows an example of decoding with beam search. Since the Viterbi algorithm is dynamic programming, it can record all the most probable previous steps and run them again to find the most probable order. This time, I use an approach that penalizes pairs of consecutive labels that do not follow the rules. This penalty is subtracted from the total score and the label column with the highest value is selected. As shown in Table 1, before using the Beam search Decoder, "I-L" comes after "O", which is a result that does not follow the rules. With the Beam search Decoder, "B-L" is predicted before "I-L" and the result is in accordance with the rule. Implement the Beam search Decoder after the output of the Dense layer.

Table 1: Example of Beam Search Decoder.

	Beam Search Decoder	
True Labels	NO	YES
O	O	O
O	O	O
O	O	O
B-L	O	B-L
I-L	O	I-L
I-L	I-L	I-L
I-L	I-L	I-L
I-L	I-L	I-L
I-L	I-L	I-L
O	O	O
O	O	O

3 Experiments

3.1 Data

The data used in this study is in BIO format, labeled with words tokenized by whitespace tokenizer in multiple news raw texts representative across the political spectrum, including 13 well-known online news media from five major English-speaking countries [1]. As shown in Figure 7, the model not only outputs the span of the quotation and speaker but also determines the correspondence between the quotation and speaker by predicting the speaker's direction relative to the quotation. In addition, there are four possible cases when determining the correspondence between citations and speakers. The first is when the speaker is sandwiched left and right between quotations. The second is when the quote is to the right of the speaker, the third is when the quote is to the left of the speaker, and the fourth is when there is a quote but no speaker. In this case, the quote has no speaker.

- **BIO format**:
  - **B:** 'Beginning' meaning the starting point of the unique expression
  - **I:** 'Inside' meaning inside of the unique expression.
  - **O:** 'Outside' means outside of the unique expression

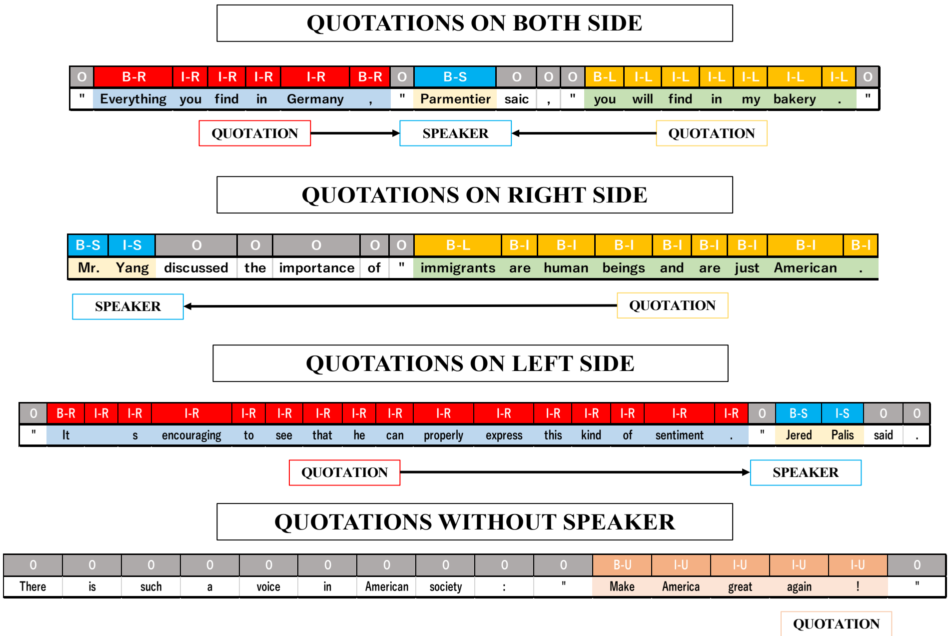

Figure 7: Relative position of the quotation and the speaker and the corresponding sequence.

- ## Types of Labels:
  - **L:** Direct quotation with speaker on the left side.
  - **R:** Direct quotation with speaker on the right.
  - **S:** Speaker.
  - **N(O):** No quotation and No speaker.
  - **U:** Quotations where the speaker is unknown

Speakers in the data have many third person (he, she) and few occurrences of names of people and proper nouns, so speakers with many occurrences in the data are replaced with speakers with the fewest random occurrences so that the model does not overfit the frequently occurring personal names. In 19,706 context windows, 10,279 quotations are obtained, of which 8,831 can be attributed to a specific speaker, and the remaining 1,522 have no explicit valid speakers. Data were split into 70:15:15 and training, evaluation and testing, respectively.

3.2 Evaluation

The evaluation of the implemented model is represented by precision, recall, and F1 score. The values are determined by the correctness of the extracted labels and the correctness of the span length, respectively. A case of the confusion matrix is often utilized with two classes, one designated the positive class and the other the negative class.

In this context, the four cells of the matrix are designated as true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN), as indicated in Table 2. The following three are defined in terms of these four classification outcomes.

Table 2: Confusion Matrix [6].

		Assigned Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

- **Recall** is calculated as the ratio between the numbers of Positive samples correctly classified as Positive to the total number of Positive samples.

$$\text{Recall} = \frac{\text{TP}}{(\text{TP+FN})}$$

- **Precision** is defined as the ratio of correctly classified positive samples (True Positive) to a total number of classified positive samples (either correctly or incorrectly).

$$\text{Precision} = \frac{\text{TP}}{(\text{TP+FP})}$$

- **F1** is defined as the harmonic mean of precision and recall. In the F1 score, we compute the average of precision and recall. They are both rates, which makes it a logical choice to use the harmonic mean.

$$F1 = 2* \frac{(\text{Precision}*\text{Recall})}{(\text{Precision} + \text{Recall})}$$

4 Result

Accuracy, recall, and F1 scores for the implemented models are shown in Table 7. The label values for each word are noted in Tables 3, 4,5, and 6. From Table 7 Results are presented in three categories: speaker, quotation, and overall. Accuracy, recall, and F1 scores for all models overall and citation are above 70%, while accuracy, recall, and F1 scores for speakers of models LSTM and BiLSTM are mostly below about 70%. features can be effectively extracted, the BERT model has stronger accuracy and generalization ability, and its speaker absolute accuracy and reproducibility are better than those of other models. Below are the results of comparing LSTM with BiLSTM and BERT with BERT-BiLSTM.

4.1 LSTM & BiLSTM

Comparing the results of the Bi-LSTM and LSTM models in Table 7, Precision is about 6% higher for speakers, about 12% higher for Recall, and about 9% higher for F1, while Precision is about 9% higher for quotations, and about 3% higher for F1. And comparing the results for each tag (Table3, Table4), it can see that the numbers for “Right Speaker” and “Unknown” tags have increased significantly. Given the above Bi-LSTM can predict future information and add it to the processing, and thus processes from both sides of the time axis, which may have led to the higher recognition rate for Bi-LSTM.

4.2 BERT & BERT-BiLSTM

Comparing the results of the BERT and BERT -BiLSTM models in Table 7. It can see that the values of Precision, Recall, and F1 have increased overall by 2~3% each. And comparing the results for each tag (Table 5, Table 6), The F1 values for the “B-Speaker”, “I-Speaker”, and “B-Unknown” tags increased significantly compared to the total. The addition of the BILSTM layer may have increased the expressive results.

Table 3: Label Result of LSTM

	precision	recall	f1-score
B-LeftSpeaker	0.78	0.92	0.85
B-RightSpeaker	0.68	0.98	0.80
B-Speaker	0.76	0.58	0.66
B-Unknown	0.66	0.39	0.49
I-LeftSpeaker	0.86	0.95	0.90
I-RightSpeaker	0.70	0.98	0.82
I-Speaker	0.68	0.51	0.58
I-Unknown	0.76	0.43	0.55
O	0.99	0.93	0.96

Table 4: Label Result of BiLSTM

	precision	recall	f1-score
B-LeftSpeaker	0.86	0.94	0.90
B-RightSpeaker	0.83	0.99	0.90
B-Speaker	0.79	0.64	0.71
B-Unknown	0.84	0.72	0.77
I-LeftSpeaker	0.89	0.97	0.93
I-RightSpeaker	0.84	0.99	0.91
I-Speaker	0.76	0.59	0.66
I-Unknown	0.87	0.74	0.80
O	0.98	0.95	0.97

4.3 BERT-BiLSTM and other models

Comparing the extraction results of BERT-BiLSTM and other models, BERT-BiLSTM often outputs correct results when the correct answer labels are all "O", but the output of other models is often wrong. An example is given in Table 8 shows two example2. In the first example, LSTM and BiLSTM extract the wrong speaker and quotations, failing to identify the speaker. This is a very confusing case because two sentences appear between Quote ("), but "Pelosi" did not speak, but wrote on Twitter. LSTM-based models may extract citations based primarily on quotes, as they misidentify content as citations. The BERT output extracts only the starting "B-" tag for quotes compared to the LSTM-based model. BERT's output is only "B-" tags, probably because BERT is more syntax sensitive. With BERT-BiLSTM, I was able to predict all labels as "O" because of the increase in expressiveness due to the addition of the BiLSTM layer. The other difference is the result of speaker extraction. An example is shown below in Table 8. The speaker in this example is "Vristow", and the LSTM model seems to misidentify the speaker's preceding word "major Republican donor" as the speaker. Since the BERT model is syntactically sensitive, it recognizes "Vristow" as a speaker, but also identifies "donor" as a speaker. BERT-BiLSTM extracts only "Vristow" as a speaker. From the above, it seems that BERT-BiLSTM is the best among the models used in this research.

Table 5: Label Result of BERT

	precision	recall	f1-score
B-LeftSpeaker	0.87	0.94	0.90
B-RightSpeaker	0.82	0.98	0.89
B-Speaker	0.85	0.94	0.89
B-Unknown	0.83	0.82	0.82
I-LeftSpeaker	0.91	0.97	0.94
I-RightSpeaker	0.85	0.98	0.91
I-Speaker	0.92	0.91	0.91
I-Unknown	0.86	0.88	0.87
O	0.99	0.95	0.97

Table 6: Label Result of BERT-BiLSTM

	precision	recall	f1-score
B-LeftSpeaker	0.88	0.95	0.91
B-RightSpeaker	0.84	0.98	0.90
B-Speaker	0.88	0.94	0.91
B-Unknown	0.84	0.84	0.84
I-LeftSpeaker	0.90	0.97	0.94
I-RightSpeaker	0.85	0.99	0.92
I-Speaker	0.94	0.93	0.93
I-Unknown	0.88	0.88	0.88
O	0.99	0.95	0.97

4.4 Comparison with Other Study

The results are compared with the results (Table 7) from [1]. Comparing the results of LSTM, the result of "Speaker" is much lower than the result of the reference paper. "Speaker" precision is down about 9%, Recall is down about 40% and F1 is down about 25%. Quotations and overall results are not much different. Since the recall of "Speaker" is 40% lower, the F1 value is also 25% lower. As a possible cause, the reference paper states that "the data is augmented by randomly replacing the speaker's name with another person's name so that the model does not overfit the frequently occurring person's name" as a data extension. I think that the difference is in the result because it is set to "change the data by replacing the speaker with another person's name at random" (**3.1**). The results of the BERT model are almost the same overall, and BERT-BiLSTM has a 6% "Precision" of "Speaker", The F1 value has increased by 3%.

5 Discussion

I found that the BERT model and BERT-BiLSTM model give high values, but there are problems in extracting quotations and speakers. A Beam search decoder can predict labels according to the rules of the BIO format, but when looking at the prediction output as a whole, there are many unnatural outputs. For example, even if the speaker is extracted, no quotations are extracted from either the left or the right. Conversely, there are unnatural predictions such as extracting quotations but not extracting speakers are observed. These phenomena are often observed when all correct labels are "O". Table 9 shows an output example of BERT-BiLSTM. In this example sentence, "Foxhall" is enclosed in quotes (" ") and the text is added to the title report. The prediction is extracting "Foxhall" as the speaker. However, no citations have been extracted. The

Table 7：Results from this study and Other Study.

Method / Label	Speaker			Quotation			All		
	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall	F1
LSTM	0.6732	0.5089	0.5796	0.7742	0.9575	0.8562	0.7344	0.7806	0.7471
LSTM [1]	0.7368	0.8286	0.7801	0.771	0.8534	0.8102	0.7700	0.8525	0.8092
BiLSTM	0.7132	0.5702	0.6337	0.8407	0.9337	0.8848	0.7904	0.7903	0.7857
BERT	0.8210	0.9308	0.8724	0.8089	0.9282	0.8645	0.8147	0.9294	0.8645
BERT [1]	0.8090	0.9402	0.8697	0.8169	0.9354	0.8721	0.8164	0.9356	0.8720
BERT – BiLSTM	0.8608	0.9435	0.9003	0.8386	0.9382	0.8856	0.8493	0.9408	0.8927

Table 8: Example of model extraction and imputation results. The blue line indicates "Speaker", the red line indicates "Right-Speaker", and the yellow line indicates "Left-Speaker".

Input	" Governors are crying out for help and Congress must act, " Pelosi wrote on Twitter. " Our state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they n eed an immediate infusion of funds to prevent the collapse of essential services. "
Truth	" Governors are crying out for help and Congress must act, " Pelosi wrote on Twitter. " Our state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they need an immediate infusion of funds to prevent the collapse of essential services. "
LSTM	" **Governors are crying out for help and Congress must act,** " **Pelosi** wrote on Twitter. " **Our state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they need an immediate infusion of funds to prevent the collapse of essential services.** "
BiLSTM	" **Governors are crying out for help and Congress must act,** " **Pelosi** wrote on Twitter. " **Our state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they need an immediate infusion of funds to prevent the collapse of essential services** "
BERT	" **Governors** are crying out for help and Congress must act, " **Pelosi** wrote on Twitter. " **Our** state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they need an immediate infusion of funds to prevent the collapse of essential services. "
BERT-BiLSTM	" Governors are crying out for help and Congress must act, " Pelosi wrote on Twitter. " Our state and local governments are in crisis, and between emergency expenses and rising unemployment sapping revenue, they need an immediate infusion of funds to prevent the collapse of essential services. "

Input	" Of course Trump is playing defense," major Republican donor vristow tells CBS News. "He's fighting to protect states he should be able to ignore." An adviser to Doug Collins'Georgia Senate bid, Eberhart calls do wn-ballot boosts " a side effect " of new ad spending in Georgia, " but not the campaign's purpose. "
Truth	" **Of course Trump is playing defense,**" major Republican donor **vristow** tells CBS News. "**He's fighting to protect states he should be able to ignore.**" An adviser to Doug Collins'Georgia Senate bid, Eberhart calls down-ballot boosts " a side effect " of new ad spending in Georgia, " but not the campaign's purpose. "
LSTM	" **Of course Trump is playing defense,**" **major Republican donor vristow tells CBS** News. "**He's fighting to protect states he should be able to ignore.**" An adviser to Doug Collins'Georgia Senate bid, Eberhart calls down-ballot boosts " a side effect " of new ad spending in Georgia, " **but not the campaign's purpose.** "
BiLSTM	" **Of course Trump is playing defense,**" **major** Republican donor vristow tells CBS News. "**He's fighting to protect states he should be able to ignore.**" An adviser to Doug Collins'Georgia Senate bid, Eberhart calls down-ballot boosts " a side effect " of new ad spending in Georgia, " but not the campaign's purpose. "
BERT	" **Of course Trump is playing defense,**" major Republican **donor vristow** tells CBS News. "**He's fighting to protect states he should be able to ignore.**" An adviser to Doug Collins'Georgia Senat bid, Eberhart calls down-ballot boosts " a side effect " of new ad spending in Georgia, " but not the campaign's purpose. "
BERT-BiLSTM	" **Of course Trump is playing defense,**" major Republican donor **vristow** tells CBS News. "**He's fighting to protect states he should be able to ignore.**" An adviser to Doug Collins'Georgia Senate bid, Eberhart calls down-ballot boosts " a side effect " of new ad spending in Georgia, " but not the campaign's purpose. "

problem in Table 9 may be solved if it is possible to say that if the speaker is extracted, the probability that the quotation exists in the sentence is high, and if the quotation is extracted, the probability that it exists in the sentence is high.

6 Conclusion

In this paper, I tested two new models, BiLSTM and BERT-BiLSTM, which have not been tested in reference papers, for extracting direct citations and corresponding speakers from the news. As a result, BiLSTM gave better results than the LSTM model tested in this study, but it was lower than the LSTM results of the reference paper. The cause seems to be the difference in "Extension of data". The results of the BERT model are the same as the reference paper, and the BERT-BiLSTM is better than the results of any model. There are several issues to be addressed in the future. One is how to expand the data. Since the extraction results of "Speaker" for LSTM and BiLSTM implemented in this study were considerably lower than those of the LSTM model in the reference paper, I need to consider how to expand the data for the "Speaker" label. I would like to change the data augmentation method for the "Speaker" label. Second is the implementation of the CRF layer. In this study, the Beam search decoder (2.2) was implemented, but the CRF layer was implemented in the reference paper [1][4],

Table 9: Example of Unnatural Extraction. The blue line indicates "Speaker".

Input	Some of the false accusations made against Browder had also appeared in newspaper articles and court documents in the UK, Foxhall added, in a report entitled " Russian 'Black PR': Examining the Practice of Ruining Reputations ".
Truth	Some of the false accusations made against Browder had also appeared in newspaper articles and court documents in the UK, Foxhall added, in a report entitled " Russian 'Black PR': Examining the Practice of Ruining Reputations ".
BERT-BiLSTM	Some of the false accusations made against Browder had also appeared in newspaper articles and court documents in the UK, **Foxhall** added, in a report entitled " Russian 'Black PR': Examining the Practice of Ruining Reputations ".

and since there was not enough time for this study, I will implement the CRF layer and compare the results. Finally, there is the issue of unnatural extraction mentioned in the Discussion. I would like to solve this issue, although I have not yet found a concrete solution. I would like to resolve the above issues in the future.

References

[1] Zhang, Yuanchi, and Yang Liu. "DirectQuote: A Dataset for Direct Quotation Extraction and Attribution in News Articles." *arXiv preprint arXiv:2110.07827* (2021).

[2] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.

[4] Jiang, Shaohua, et al. "A BERT-BiLSTM-CRF model for Chinese electronic medical records named entity recognition." *2019 12th International Conference on Intelligent Computation Technology and Automation (ICICTA)*. IEEE, 2019.

[5] Liu, Yue. "Investigation of Viterbi Algorithm Performance on Part-of-Speech Tagger of Natural Language Processing." *2017 International Conference on Computer Systems, Electronics and Control (ICCSEC)*. IEEE, 2017. T

[6] *Ting, K.M. (2011). Confusion Matrix. In: Sammut, C., Webb, G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA.*