Search Term Specific Sentiment Classification for Business News

Team 22
Yuwei Wang, Yi Xu, Yunya Wang, Isabel Zhou

Motivation

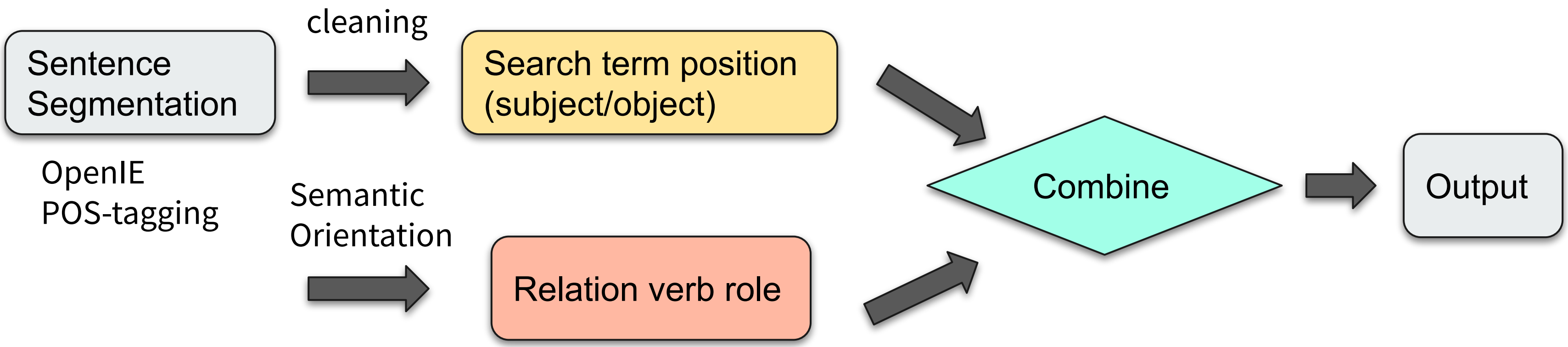
The sentiment label for our search term can be completely shifted given different positions of the term appearing in the sentence.

News Headline	Sent. Target
Fidelity replaces J.P.Morgan as sub-advisor on 2 strategies	Negative
J.P.Morgan replaces Fidelity as sub-advisor on 2 strategies	Positive
Mercer drops J.P.Morgan as managers of bond fund	Negative
J.P.Morgan drops Mercer as managers of bond fund	Positive

Data

- Existing datasets are not search term specific
- Source: Headlines from Google News, Pension & Investment
- Web-scraped, Cleaned criteria, labeled (human evaluation)
- Data Size (~1000): 60% Positive, 40% Negative
- Binary labels: more sensitive to negative news(risk management)
    - positive + neutral → positive
    - negative

Initial Idea
(Sentence Segmentation + Semantic Orientation)



Semantics Orientation powerfully takes seeds of words in opposite directions of any given dimension, we would be able to distinguish the different function labels of a verb towards subject and object respectively (https://dl.acm.org/doi/10.1145/944012.944013)

Headline	sub_indicator	obj_indicator	cleaned_rel_headline	rel_sub	rel_obj	output
Prudential replaces J.P.Morgan as subadviser on 2 strategies	0	1	replace	1	-1	-1
District of Columbia Retirement Board terminates J.P.Morgan	0	1	terminate	1	-1	-1

Initial Idea
(Sentence Segmentation + Semantic Orientation)

Limitation:

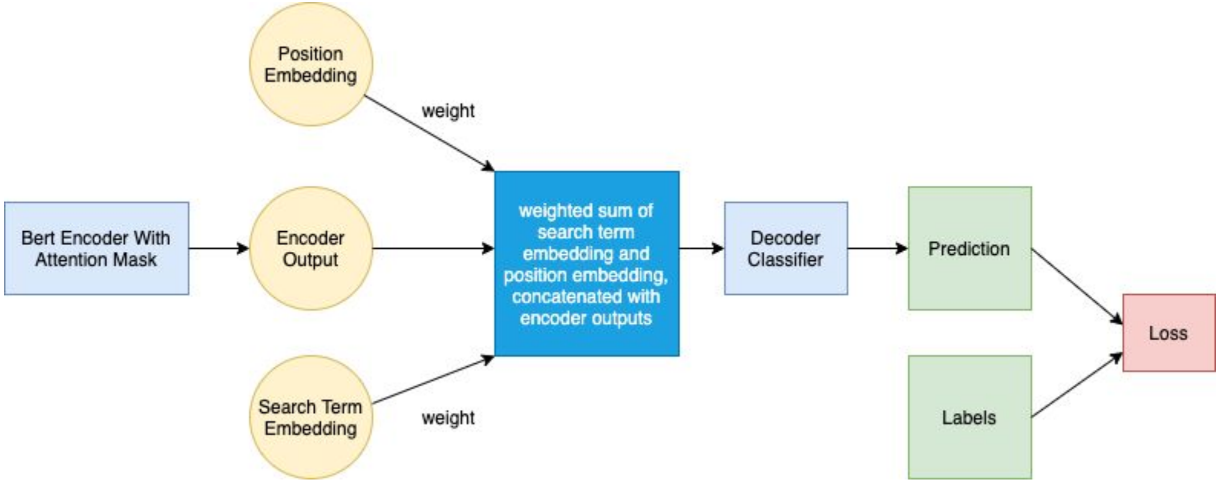
- Extraction of search term and relation verb is not always reliable, sometimes produces empty outputs
- Training Semantic Orientation (detecting relation verb's contribution to subject/object) is tricky. It's highly dependent on input of seed words, which causes great variance of accuracy.
- Performance: 736/1000 observations have outputs, among valid outputs: validation accuracy 68.7%

Next steps:

- Keep detection of search term role (subject, object)
- Adopt a NN framework for better training, but incorporate the role of search term into embeddings.

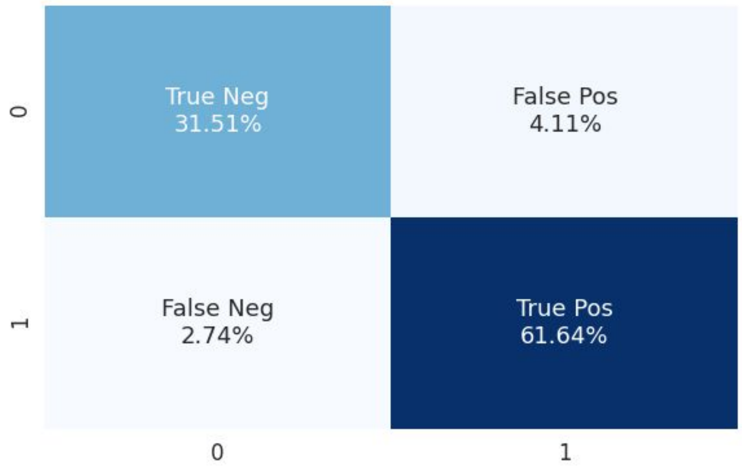
BERT related methodology

1. Fine-tuned BERT (BertForSequenceClassification)
2. BERT + self-defined Classifier
3. BERT + Search Term Embedding + self-defined Classifier
4. BERT Encoder with Attention Mask; Decoder with Concat(Weighted Sum of Search Term Embedding and Positional Embedding, Encoder Output) + Self-defined Classifier



Experiments and Results

BERT Encoder + Attention Mask + Decoder with Concat(Weighted Sum of Search Term Embedding and Positional Embedding, Encoder Output) + Self-defined Classifier



Experiments and Results

Models	Test Accuracy	False Positive Rate
Fine-tuned BERT	0.86	8.22%
BERT + self-defined classifier	0.87	8.22%
BERT + self-defined classifier + search term embedding	0.85	10.96%
BERT Encoder + Attention Mask + Decoder with Concat (Weighted Sum of Search Term Embedding and Positional Embedding, Encoder Output) + Self-defined Classifier	**0.95**	**4.11%**

Conclusions

1. Neural Network is very powerful comparing to other constructed methods, e.g sentence segmentation + Semantic Orientation.

2. Incorporating more information like object subject relationship and search term help training and significantly improve the model performance.

3. Incorporation of the state of the art BERT model help overcome the limited dataset constraint and still achieve the best performance.