TMUNLP at the NTCIR-16 FinNum-3 Task
Multi-task Learning on BERT for Claim Detection and Numeral Category Classification

Tzu-Ying Chen [1,2]
m946110012@tmu.edu.tw

Yu-Wen Chiu [1,2]
linda037@tmu.edu.tw

Hui-Lun Lin [1]
m946110005@tmu.edu.tw

Chia-Tzu Lin [1,2]
m946110009@tmu.edu.tw

Yung-Chung Chang [1,2,*]
changyc@tmu.edu.tw

Chun-Wei Tung [3]
cwtung@nhri.edu.tw

[1] Graduate Institute of Data Science, Taipei Medical University, Taiwan
[2] TMUNLP Lab, Taipei Medical University, Taiwan
[3] Institute of Biotechnology and Pharmaceutical Research, National Health Research Institutes, Taiwan
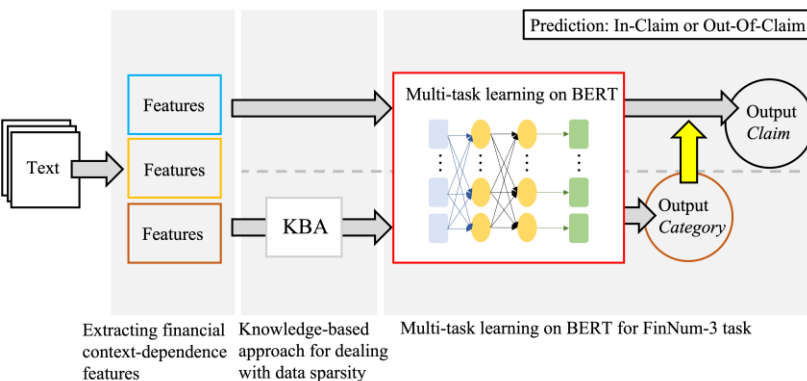
Background

In financial documents, numerals often contain important information in addition to textual data. As a result, understanding the semantics and relations between numerals and words is of great interest and benefit. The main task of the NTCIR-16 FinNum-3 is fine-grained claim detection, uncovering meaning of numerals in financial reports through claim detection and numeral category classification.

Method

Overall architecture

We developed a system that can carry out the two tasks simultaneously. Through combining different designated features and target numerals, several different inputs are passed through the model, and the output results, categories, and claims can be jointly predicted by our system.

Multi-task learning on BERT for FinNum-3 task

Extracting financial context-dependent features

We discover the following financial context-dependent (FCD) features through literature research and theoretical analysis. For word units, the features we create may have the following forms:

(a) The last two characters of the target numeral.
(b) The target numeral and two characters before and after it.
(c) From the target numeral to the start of the sentence it exists.
Finally, we enter two feature combinations FCD 1 and FCD 2 into the model, where FCD 1 is (b)+(c), and FCD 2 is (a)+(b)+(c).

Example of financial context-dependent (FCD) features

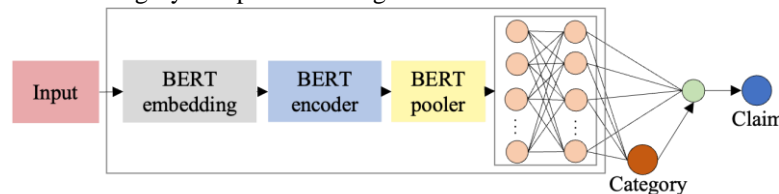| Examples	……，EPS1.05 元創下歷史新高。我們預估 2018 年營收為 289.6 億元，毛利率 35%，淨利 *59.63* 億元，EPS3.64 元。(……, EPS1.05 dollar achieved an all-time high. We estimate that revenue in 2018 will be 289.6 billion, with a gross profit margin of 35%, net income of *59.63* billion, EPS3.64 dollar.)
Target numeral	**59.63**
feature (a)	億元 (billion)
feature (b)	淨利 59.63 億元 (net income of 59.63 billion)
feature (c)	我們預估 2018 年營收為 289.6 億元，毛利率 35%，淨利 (We estimate that revenue in 2018 will be 289.6 billion, with a gross profit margin of 35%, net income of)

Knowledge-based approach for dealing with data sparsity

For the dataset provided by the organizer, Knowledge-based approach filtered out the four categories with the fewest samples at the beginning of our proposed system, namely, *Change*, *Time*, *Ranking*, and *Quantity*.

Multi-task Learning

We propose a multi-task learning approach on top of BERT for concurrent claim detection and numeral category classification. The model predicted category first, and predicted claim-or-not by predicting result of category and previous weights.

Result

Among the three proposed models, the model with the ensemble fine-tuned BERT as the pre-trained model and FCD 2 as the input feature set had the best performance, with a micro F_1-score of 94.67% for the numeral category classification task, and a micro F_1-score is 92.75% for the claim detection task.

Performance of our methods on the test set

Submission ID	Claim Detection		Numeral Category	
	micro F_1	macro F_1	micro F_1	macro F_1
Baseline	80.32%	69.19%	62.59%	20.99%
TMUNLP 1	92.82%	89.56%	94.31%	73.68%
TMUNLP 2	91.11%	87.76%	94.03%	72.99%
TMUNLP 3	92.75%	89.68%	94.67%	73.89%

Conclusion

Compared with the Capsule Neural Network models returned by organizers, the final performance can be improved greatly by our method, which demonstrated that (1) KBA is effective in dealing with the sparse data, (2) extracting keywords and key sentences from contexts as features can be beneficial, and (3) utilizing multi-task learning and ensemble learning on BERT is effective.

Keywords

BERT, Argument mining, Ensemble learning, Multi-task learning

Scan the QR-code to know us more