

Masked Sentence Model Based on BERT for Move Recognition in Medical Scientific Abstracts

Gaihong Yu^{1,2}, Zhixiong Zhang^{1,2,3†}, Huan Liu^{1,2}, Liangping Ding^{1,2}

¹National Science Library, Chinese Academy of Sciences, Beijing 100190, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³Wuhan Library, Chinese Academy of Sciences, Wuhan 430071, China

Citation: Gaihong Yu, Zhixiong Zhang, Huan Liu, Liangping Ding (2019). Masked sentence model based on BERT for move recognition in medical scientific abstracts. *Journal of Data and Information Science*, 4(4), 42–55

DOI: 10.2478/jdis-2019-0020

Received: Sep. 27, 2019

Revised: Nov. 5, 2019

Accepted: Nov. 5, 2019

Abstract

Purpose: Move recognition in scientific abstracts is an NLP task of classifying sentences of the abstracts into different types of language units. To improve the performance of move recognition in scientific abstracts, a novel model of move recognition is proposed that outperforms the BERT-based method.

Design/methodology/approach: Prevalent models based on BERT for sentence classification often classify sentences without considering the context of the sentences. In this paper, inspired by the BERT masked language model (MLM), we propose a novel model called the masked sentence model that integrates the content and contextual information of the sentences in move recognition. Experiments are conducted on the benchmark dataset PubMed 20K RCT in three steps. Then, we compare our model with HSLN-RNN, BERT-based and SciBERT using the same dataset.

Findings: Compared with the BERT-based and SciBERT models, the F1 score of our model outperforms them by 4.96% and 4.34%, respectively, which shows the feasibility and effectiveness of the novel model and the result of our model comes closest to the state-of-the-art results of HSLN-RNN at present.

Research limitations: The sequential features of move labels are not considered, which might be one of the reasons why HSLN-RNN has better performance. Our model is restricted to dealing with biomedical English literature because we use a dataset from PubMed, which is a typical biomedical database, to fine-tune our model.

Practical implications: The proposed model is better and simpler in identifying move structures in scientific abstracts and is worthy of text classification experiments for capturing contextual features of sentences.

Originality/value: The study proposes a masked sentence model based on BERT that considers the contextual features of the sentences in abstracts in a new way. The performance



of this classification model is significantly improved by rebuilding the input layer without changing the structure of neural networks.

Keywords Move recognition; BERT; Masked sentence model; Scientific abstracts

1 Introduction

The concept of move, or rhetorical move, was originally developed by Swales to functionally describe a part or section in research articles for communicative purposes (Swales et al., 2004). Authors of research papers generally need to explain the purpose, methods, results, and conclusions of their research in abstracts. Those language units are called the moves of the abstracts.

Many journals currently require authors to provide structured abstracts with explicitly annotated move labels. For example, authors usually use “Purpose” to indicate the move label and the sentences following the label to represent the aim of the study. Currently, many important journals, such as Nature and Science, still use unstructured abstracts when they publish the research articles.

Automatically recognizing moves of unstructured abstracts in research papers (move recognition in brief), which is typically a classification task, enables readers to quickly grasp the main points of research papers, and it is useful for various text-mining tasks such as information extraction, information retrieval and automatic summarization.

Many researchers have performed considerable work on move recognition. Early studies on move recognition adopted traditional machine learning methods such as naive Bayes (NB) (Teufel, 1999), conditional random fields (CRF) (Hirohata et al., 2008), support vector machine (SVM) (Ding et al., 2019; Yamamoto & Takagi, 2005), and logic regression (LR) (Fisas et al., 2016). These methods achieve good recognition performance, but they are very complicated to apply because they rely heavily on numerous carefully hand-engineered features such as lexical, semantic, structural, statistical and sequential features.

In recent years, neural networks have been widely used in NLP (natural language processing) research, including move recognition tasks (Dasigi et al., 2017; Kim, 2014; Lai et al., 2015; Ma et al., 2015; Zhang et al., 2019). Neural networks have strong nonlinear fitting ability and can automatically learn a better and deeper presentation for the input without complicated feature engineering. Methods using neural networks usually achieve better performance than traditional machine learning methods, which is one of the reasons why deep learning methods are widely used in many NLP studies.

In particular, Di Jin et al. (2018) from MIT proposed a hierarchical sequential labelling network named HSLN-RNN that used the contextual information within



surrounding sentences to help classify the current sentence. Specifically, HSLN-RNN used a Bi-LSTM layer after encoding sentence-level features to capture contextual features within sentences and a CRF layer to capture sequential features within surrounding move labels. HSLN-RNN achieved state-of-the-art results with an F1 score of 92.6% on the PubMed 20K RCT dataset.

BERT (bidirectional encoder representations from transformers) (Devlin et al., 2018), released by Google in October 2018, received widespread attention because it broke the records of 11 NLP tasks when released. After the release of BERT, some researchers performed move recognition studies based on BERT and attempted to obtain better performance. Iz Beltagy et al. (2018) fine-tuned the BERT-based model on the PubMed 20K RCT and obtained an average F1 score of 86.19%. They also released the SciBERT model, which re-pre-trained the original BERT model with a corpus from the biomedical domain. Based on SciBERT, the F1 score on PubMed 20K RCT reached 86.81%, which was better than the BERT-based model but still did not reach the highest F1 score of 92.6% based on the HSLN-RNN model (Jin & Szolovits, 2018).

By comparing the models based on BERT and HSLN-RNN, we find that the main problem of the current BERT-based models is that they only use the content of the sentences without considering the context of the sentences, in which the “content” represents the sentence, and the “context” represents the surrounding information (or the surrounding sentences) of the sentence in an abstract.

We assume that the move type of a sentence depends on not only the sentence itself but also its surrounding sentences and the context information of those sentences, which can help to improve the performance of move recognition. For instance, when we draft an abstract, a sentence of “Results” is more likely to be followed by a sentence of “Conclusions” than a sentence of “Purpose”.

In our study, we intend to integrate the content and context information of sentences in move recognition based on BERT. Inspired by BERT’s “masked language model” (MLM), we propose a “masked sentence model” (MSM) based on BERT to solve this problem. We mainly improve the move recognition task during the BERT fine-tuning procedure without changing its neural networks. The model makes full use of the content and context of the sentences.

Our key contributions are summarized as follows:

- (1) We propose a masked sentence model based on BERT that can capture not only the content features but also the contextual features of the sentences. Our model is easy to apply because it only rebuilds the input layer without any change in the structure of neural networks.



- (2) We evaluate on the public dataset for move recognition (PubMed 20K RCT) and see an improvement of approximately +4.34% F1 score compared to SciBERT, +4.96% F1 score compared to the BERT-based model, which shows the effectiveness of our masked model.

2 Methodology

2.1 Main idea

Firth (1957) proposed a distributional hypothesis in natural language processing research in which words can be identified by their context. This hypothesis has been widely used in information retrieval, topic recognition (Basili & Pennacchiotti, 2016) and other NLP studies. The BERT masked language model (MLM) is also based on this distributional hypothesis. MLM simply masks some percentage of the input tokens at random and then predicts those masked tokens based on their context.

Similarly, we propose that sentences in an abstract also follow the distributional hypothesis, and we believe that a sentence in an abstract can be identified by the contextual sentences surrounding them.

Based on this hypothesis, a novel model called the masked sentence model (MSM) is proposed. This model integrates two ideas of sentence representations for the move recognition task. Similar to traditional deep learning classifiers, it preserves using the content of the target sentence as the input of the classifier to learn the internal features of this sentence. Moreover, to capture the contextual information, it innovatively uses the whole abstract but with the target sentence masked as the input of the classifier to learn the contextual features of the target sentence.

For example, there is an abstract document from PubMed^① shown in figure 1 that contains seven sentences (from s_1 to s_7). For the second sentence (s_2), we have two representations (as shown in figure 2) for the input of a deep learning classifier: 2-a

[s_1]This survey aims at reviewing the literature related to Clinical Information Systems (CIS), Hospital Information Systems (HIS), Electronic Health Record (EHR) systems, and how collected data can be analyzed by Artificial Intelligence (AI) techniques. [s_2]*We selected the major journals (11 journals) collecting papers (more than 7,000) over the last five years from the top members of the research community, and read and analyzed the papers (more than 200) covering the topics.* [s_3]Then, we completed the analysis using search engines to also include papers from major conferences over the same five years. [s_4]We defined a taxonomy of major features and research areas of CIS, HIS, EHR systems. [s_5]We also defined a taxonomy for the use of Artificial Intelligence (AI) techniques on healthcare data. [s_6]In the light of these taxonomies, we report on the most relevant papers from the literature. [s_7]We highlighted some major research directions and issues which seem to be promising and to need further investigations over a medium- or long-term period.

Figure 1. An example of an abstract.

^① <https://www.ncbi.nlm.nih.gov/pubmed/31419820>



Research Paper

(representation based on the content of the sentence), 2-b (representation based on the context of the sentence by using the whole abstract masking the target sentence with a fixed meaningless string denoted as “[MASK]”). In the masked sentence model, we combine the two representations above to learn both the content features and contextual features of the sentence (2-c).

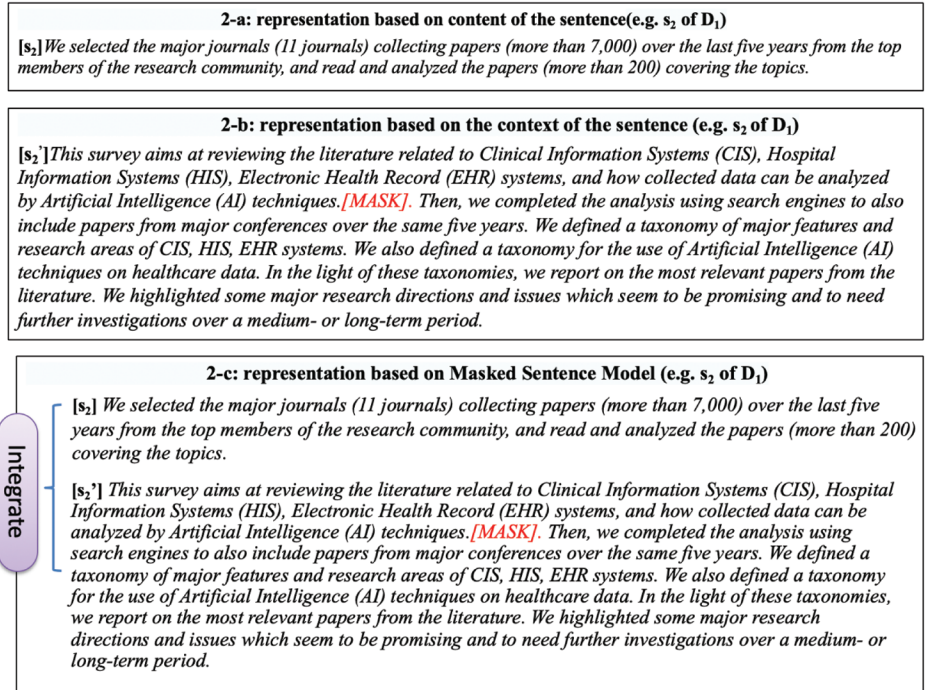


Figure 2. Sentence representations.

We use this integrated MSM representation of the sentence as input in the BERT fine-tuning procedure and conduct several experiments to verify its effectiveness.

2.2 MSM construction

Based on the main idea mentioned above, we construct the masked sentence model (as shown in figure 3) in three processing steps before BERT fine-tuning.

Step 1: Sentence information processing

In this step, for an abstract, each target sentence in the abstract of scientific papers is represented by the content of the sentence.



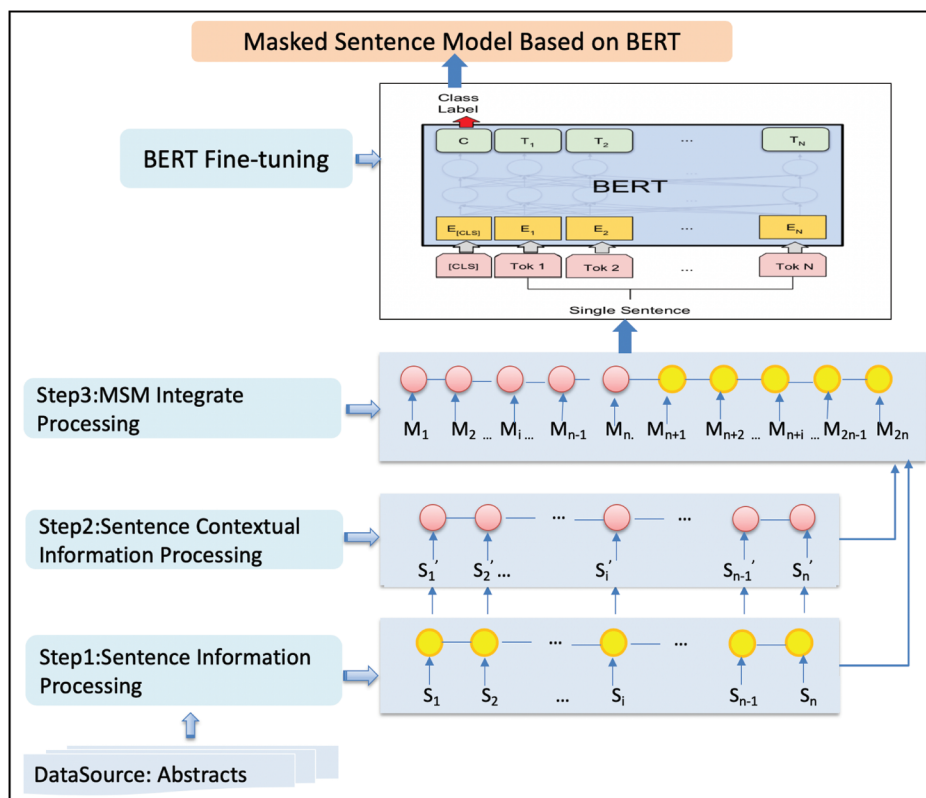


Figure 3. The architecture of the masked sentence model based on BERT.

For example, the second sentence (s_2) of the abstract shown in figure 1 is annotated with “Methods”, and the data format after this process step is shown in table 1. We use this representation to learn the internal features of the sentence.

Table 1. Data format of sentence content.

Label	The content of the sentence
Methods	We selected the major journals (11 journals) collecting papers (more than 7,000) over the last five years from the top members of the research community, and read and analyzed the papers (more than 200) covering the topics.

Step 2: Sentence contextual information processing

In step 2, we obtain the contextual information of each sentence in the abstract. Here, we adopt a new method that simply uses the whole abstract except for the target sentence, which is replaced by a [MASK] string to obtain the contextual information. In this paper, we replace each word of the target sentence with “aaa” characters to build the meaningless string of [MASK].



For sentence (s_2), the length of this sentence is 37, so the data format after this step is shown in table 2, where the second sentence is replaced with 37 “aaa” characters. We input this contextual information into the BERT fine-tuning procedure and learn the contextual features of the sentence.

Table 2. Data format of the sentence’s context.

[illegible]

Step 3: MSM integrate processing

Third, this step integrates the content and contextual information of the sentence to construct the masked sentence model. It implements the integration by inputting the above two training samples together to the BERT fine-tuning procedure to train the masked sentence model based on the BERT-base model for the move recognition task. For example, the final input representation for the second sentence (s_2) in this step is shown in table 3.

Table 3. Data format for integrating sentence content and context.

[illegible]

3 Experiments and results

3.1 Experimental design

In this paper, for different verification purposes, we design and conduct three different move recognition experiments based on BERT with the same neural network architecture and the same dataset. They only vary in the input data during the fine-tuning procedure.

Experiment 1 (or Exp1): An experiment based on the content of sentences. We fine-tune the BERT-based model for the downstream move recognition task by using the data format, as shown in table 1, which only contains the content of sentences in the fine-tuning input layer.

Experiment 2 (or Exp2): An experiment based on the context of sentences. The purpose of this experiment is to explore the rationality of our assumption based on the distributional hypothesis and verify the feasibility of the method of sentence context processing. This experiment was carried out based on the context of sentences using the data format shown in table 2 as BERT fine-tuning inputs.

Experiment 3 (or Exp3): The most important experiment based on the MSM integrated information. To verify the effectiveness of the novel MSM model proposed in this paper. This experiment uses the data format shown in table 3, which integrates the content and context of sentences as BERT fine-tuning inputs.

3.2 Datasets

Our study evaluates the MSM model on the benchmark dataset PubMed 20K RCT (Dernoncourt & Lee, 2017), which contains approximately 20,000 medical scientific abstracts for sequential sentence classification. The dataset is based on the PubMed database of biomedical literature, and each sentence of an abstract is labelled with its rhetorical role in the abstract using one of the following classes: Background, Objectives, Methods, Results, and Conclusions.

3.3 Hyper-parameters setting

Our study uses the BERT-based model (Devlin et al., 2018) with a hidden size of 768, 12 transformer blocks (Vaswani et al., 2017) and 12 self-attention heads and fine-tunes with the following settings: a batch size of 5, a max sequence length of 512, a learning rate of $3e-5$, an init_checkpoint of bert_base, train steps of 100,000 and warm-up steps of 10,000.

3.4 Evaluation metrics

For each designed experiment, our study reports the performance with the evaluation metrics of precision (P), recall (R), and F1 score on the same test set



provided by the PubMed 20K RCT dataset (29,578 sentences from 2,500 abstracts). The experimental results of each experiment are detailed in section 3.5.

3.5 Results

The results of Exp1: based on the content of sentences

Table 4 shows the results of the experiment based on the content of sentences. It can be found that just using the content of sentences can achieve effective results for move recognition at the average precision, recall, and F1 score of 86.75%, 86.61%, and 86.53%, respectively. In this experiment, promising results can be obtained for the categories of Methods and Results where both the precision and recall scores are above 91%. In the category of Conclusions, the F1 score reached a relatively good score at 83.13%. However, for the categories of Background and Objectives, the F1 scores are 69.64% and 64.20%, respectively, which cannot satisfy our expectations very well and indicates considerable room for improvement.

Table 4. The results of Exp1: based on the content of sentences.

Label	P	R	F1	Support
Background	64.37	75.85	69.64	3,077
Objectives	73.55	56.97	64.20	2,333
Methods	92.42	94.97	93.68	9,884
Results	92.08	91.09	91.58	9,713
Conclusions	84.95	81.38	83.13	4,571
Avg / Total	86.75	86.61	86.53	29,578

The results of Exp2: based on the context of sentences

Table 5 presents the results of the experiment based on the context of sentences, in which only the context of sentences is used. The performance can reach the F1 score an average of 86.09%, which indicates the validity of this method. These results prove that our basic assumption that “sentences can be identified by their context in an abstract” is reasonable. The method of using the context of sentences greatly improves the performance on the Background and Conclusions categories, whose average F1 performance improves 6.18% and 6.61%, respectively, compared with the corresponding categories based on the content of sentences. However, the performance on the Methods and Results categories decreases slightly compared to the corresponding categories based on the content of sentences. These two experiments indicate that using the content of sentences performs better in the Methods and Results categories and that using the context of sentences performs better in the Background, Conclusions and Objective categories.



Table 5. The results of Exp2: based on the context of sentences.

Label	P	R	F1	Support
Background	72.27	79.72	75.82	3,077
Objectives	70.51	60.27	64.99	2,333
Methods	90.70	89.80	90.25	9,884
Results	87.71	89.20	88.45	9,713
Conclusions	90.19	89.30	89.74	4,571
Avg / Total	86.13	86.15	86.09	29,578

The results of Exp3: based on MSM integrated information

Table 6 shows the BERT's fine-tuning experimental results based on the MSM model with integrated information of content and context of sentences. It shows that the novel model achieves the best results in the overall experiments, and the average precision, recall and F1 scores can reach 91.22%, 91.30%, and 91.15%, respectively. Compared to the first two experiments, the results of this experiment increase by 4.62% and 5.06%, respectively, in terms of F1 score. The results show that the novel masked sentence model performs better on the task of move recognition. Furthermore, the study verifies that the model can effectively improve the performance by incorporating the content and context information of sentences.

Table 6. The results of Exp3: based on MSM integrated information.

Label	P	R	F1	Support
Background	75.26	81.18	78.11	3,077
Objectives	78.08	61.98	69.10	2,333
Methods	92.98	97.48	95.17	9,884
Results	96.02	93.74	94.87	9,713
Conclusions	94.70	94.51	94.60	4,571
Avg / Total	91.22	91.30	91.15	29,578

3.6 Result analysis

Table 7 shows the comparison results of the abovementioned three experiments, which better explain the advantage of our integrated MSM model.

Table 7. Comparison of the results of the experiments.

Label	Exp1	Exp2	Exp3	Exp3-Exp1	Exp3-Exp2
	F1	F1	F1	+F1	+F1
Background	69.64	75.82	78.11	8.47	2.29
Objectives	64.20	64.99	69.10	4.9	4.11
Methods	93.68	90.25	95.17	1.49	4.92
Results	91.58	88.45	94.87	3.29	6.42
Conclusions	83.13	89.74	94.60	11.47	4.86
Avg / Total	86.53	86.09	91.15	4.62	5.06



Research Paper

From the “Exp3-Exp1”, it can be found that the MSM model greatly improves the performance on the “Conclusions” and “Background” categories, whose average F1 performance improves 11.47% and 8.47%, respectively, compared with the corresponding categories based only on the content of the sentence. Then, followed by the 4.9% improvement in the “Objectives” category. The impact on the “Methods” and “Results” categories is relatively small. Based on these comparisons, it indicates that the “Conclusions” and “Background” categories are more context-sensitive and can achieve considerable improvements by incorporating contextual information. This makes sense because the “Background” move always appears at the beginning of the abstracts, and the “Conclusions” move appears at the end of the abstract. Thus, by adding contextual information to the input, the model learns the positional information to a certain extent.

Correspondingly, from the “Exp3-Exp2”, compared with the corresponding categories based only on the context of the sentence, the MSM model obtained the highest F1 value growth in the “Results” category (6.42%), the lowest F1 value growth in the Background category (2.29%) and a relatively balanced growth in the other three categories. This indicates that the content of the sentence also plays an important role in identifying the rhetorical role of the sentence, which is used to express the author’s writing intention.

4 Comparisons & discussion

In this section, we compare the model proposed in this paper with other models and discuss the compared results. Table 8 lists our model and the other experimental models evaluated on the PubMed 20k corpus. The masked sentence model based on BERT presented in this paper is denoted by “Our Model”. “Others” contains the HSLN model, BERT-based model, and two models of SciBERT.

Table 8. PubMed 20k RCT results.

Models		F1 (PubMed 20k RCT)
Our Model	MaskedSentenceModel_BERT	91.15
Others	HSLN-RNN (Jin and Szolovits, 2018) (SOTA)	92.6
	BERT-Base (Beltagy et al., 2018)	86.19
	Sci BERT (SciVocab) (Beltagy et al., 2018)	86.80
	Sci BERT (BaseVocab) (Beltagy et al., 2018)	86.81

This shows that HSLN’s average F1 score still ranks first based on Bi-LSTM+CRF methods. Our model achieves better performance than the current models based on BERT, which outperforms 4.34 points than SciBERT (BaseVocab), and 4.96 points than the BERT-based model. However, our current model is still 1.45 points lower than HSLN. This makes sense because the sequence tag information considered in



the HSLN model is not considered in our model, so there is room for improvement based on BERT.

5 Conclusions & future work

This paper presents a novel approach to recognizing moves in scientific abstracts using the masked sentence model based on BERT. It demonstrates that integrating content and context information of sentences by using the MSM model to learn the internal and contextual features of sentences can improve the overall recognition performance. The proposed method achieves more successful results than other previous BERT-based methods. It outperforms the BERT-based and SciBERT results by 4.96% and 4.34%, respectively, on the public dataset PubMed 20k RCT. Because the model does not consider the sequential features of move labels, HSLN-RNN still has better performance.

Our MSM approach is general and easy to apply. It can achieve better performance improvement on the move recognition task only by making some optimizations in the input layer without any change on the BERT internal structure of neural networks. We believe that our model is effective for many other context-sensitive NLP tasks, including text classification and sentiment analysis.

Although our model is proven to have great performance, there are also some limitations. Our current method is still relatively simple because we simply replace each word of the sentences with a meaningless string “aaa” to generate the contextual representation of the sentences. In the future, we will improve the performance of the MSM method in the following aspects:

- (1) We will try to modify the method of generating the contextual information of the sentences using a fixed length string of 30 “aaa” to mask the target sentences and analyse its effect.
- (2) We plan to extend our MSM to cover many other important features for move recognition, such as sequential features that are not incorporated in our MSM approach.
- (3) Additionally, we would like to attempt to modify the structure of neural networks to fit the special input layer proposed in this study. In that way, this context-sensitive approach could be more efficient.

Acknowledgements

This work is supported by the project “The demonstration system of rich semantic search application in scientific literature” (Grant No. 1734) from the Chinese Academy of Sciences.



Author contributions

Zhixiong Zhang (zhangzhx@mail.las.ac.cn) and Gaihong Yu (yugh@mail.las.ac.cn) designed and produced the research. Huan Liu (liuhuan@mail.las.ac.cn) conducted the experiments. Gaihong Yu wrote the main body of the paper. Zhixiong Zhang performed many modifications and improvements and finally completed the paper. Huan Liu and Liangping Ding (dingliangping@mail.las.ac.cn) performed many paper revisions and improvement work, especially in the Introduction and Methodology Section.

References

- Amini, I., Martinez, D., & Molla, D. (2012). Overview of the ALTA 2012 shared task. In *Proceedings of the Australasian Language Technology Association Workshop 2012: ALTA 2012* (pp. 124–129). Dunedin, New Zealand.
- Badie, K., Asadi, N., & Tayefeh Mahmoudi, M. (2018). Zone identification based on features with high semantic richness and combining results of separate classifiers. *Journal of Information and Telecommunication*, 2(4), 411–427.
- Basili, R. & Pennacchiotti, M. (2010). Distributional lexical semantics: Toward uniform representation paradigms for advanced acquisition and processing tasks. *Natural Language Engineering*, 1(1), 1–12.
- Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: Pretrained contextualized embeddings for scientific text. *arXiv:1903.10676v3*.
- Dasigi, P., Burns, G.A.P.C., Hovy, E., & Waard, A. (2017). Experiment segmentation in scientific discourse as clause-level structured prediction using recurrent neural networks. *arXiv:1702.05398*.
- Devlin, J., Chang, M.W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv:1810.04805*.
- Ding, L.P., Zhang, Z.X., & Liu, H. (2019). Research on factors affecting the SVM model performance on move recognition. *Data Analysis and Knowledge Discovery*, <http://kns.cnki.net/kcms/detail/10.1478.G2.20191012.0931.002.html>.
- Firth, J.R. (1930). A synopsis of linguistic theory, 1930–1955. In: Firth, J.R., Ed., *Studies in Linguistic Analysis*, Longmans, London, 168–205.
- Fisas, B., Ronzano, F., & Saggion, H. (2016). A multi-layered annotated corpus of scientific papers. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*.
- Franck Dernoncourt & Ji Young Lee. (2017). Pubmed 200k rct: a dataset for sequential sentence classification in medical abstracts. In *Proceedings of the 8th International Joint Conference on Natural Language Processing*.
- Gerlach, M., Peixoto, T.P., Altmann, E.G., & Altmann, E.G. (2018). A network approach to topic models. *Science advances*, 4(7), eaaq1360.
- Hirohata, K., Okazaki, N., Ananiadou, S., & Mitsuru. (2018). Identifying sections in scientific abstracts using conditional random fields. In *Proceedings of the Third International Joint Conference on Natural Language Processing*.
- Ma, M.B., Huang, L., Xiang, B., & Zhou, B.W. (2015). Dependency-based convolutional neural networks for sentence embedding. *arXiv:1507.01839*.



- Peters, M.E., Neumann, M., Iyyer, M., et al. (2018). Deep contextualized word representations. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. doi: 10.18653/v1/N18-1202 arXiv:1802.05365.
- Radford, A., Narasimhan, K., Salimans, T., & Sutskever Ilya (2018). Improving language understanding by generative pre-training. [https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language understanding paper.pdf](https://s3-us-west-2.amazonaws.com/openai-assets/researchcovers/languageunsupervised/language%20understanding%20paper.pdf).
- Lai, S.W., Xu, L., Liu, K., & Zhao, J. (2015). Recurrent convolutional neural networks for text classification. In AAAI'15 Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence, pages 2267–2273.
- Swales, J.M. (2004). Research genres: Explorations and applications. Cambridge: Cambridge University Press.
- Taylor, W.L. (1953). “Cloze procedure”: A new tool for measuring readability. *Journalism & Mass Communication Quarterly*, 30(4), 415–433. doi: <https://doi.org/10.1177/107769905303000401>
- Teufel, S. (1999). Argumentative zoning: Information extraction from scientific text. Edinburgh: University of Edinburgh.
- Vaswani, A., Shazeer, N., Parmar, N., et al. (2017). Attention is all you need. arXiv:1706.03762v5.
- Yamamoto, Y. & Takagi, T. (2005). A sentence classification system for multi-document summarization in the biomedical domain. In Proceedings of International Workshop on Biomedical Data Engineering, pages 90–95.
- Yoon Kim. (2014). Convolutional neural networks for sentence classification. arXiv:1408.5882.
- Zhang, Z., Liu, H., Ding, L., et al. (2019). Moves recognition in abstract of research paper based on deep learning. In Proceedings of 2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE, pages 390–391.



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

