#381

Automatic Recognition and Classification of Future Work Sentences in Academic Articles in a Specific Domain

Wenke Hao, Yuchen Qian, Zhicheng Li, Yuzhuo Wang, Chengzhi Zhang

Department of Information Management, Nanjing University of Science and Technology, Nanjing, China

Introduction

Nowadays, various digital technologies promote the publication, diffusion, acquisition, and utilization of academic literature. Therefore, researchers, mainly those who are beginners in scientific research, are difficult to predict future research topics. The research paradigm characterized by ‘data-driven’ is emerging in the topic detection and tracking. Compared with the bibliographic information, full-text of academic papers contains more micro-semantic details. The ‘future work’ is a pivotal part of an academic paper, where the authors make suggestions for future research, point out what research is underway, or discuss the potential direction of the whole field. Taking the field of NLP as a case, this paper extracts Future Work Sentences (FWS) of articles, classifies them into different types, and then discusses the trends of future research in NLP field.

Related work

Recently, Hao et al. (2020) constructed a corpus, including FWS manually extracted from ACL papers published during 1990-2015, and constructed a classification system in accordance with grounded theory. Besides, most prior work is based on rules to identify FWS (Hu & Wan, 2015). However, manually established rules can hardly cover all the language features of FWS and are susceptible to subjective factors of experts. Therefore, we adopt machine learning methods realize automatic recognition and classification of FWS.

Method

Dataset

We download 11,952 conference papers of ACL, EMNLP, and NAACL during 2000-2019 from ACL Anthology (https://www.aclweb.org/anthology/). The manually annotated dataset consists of two parts: first, FWS of ACL during 2000-2015 are obtained from the *ACL FWS-RC* corpus (Hao et al., 2020). Second, we extract chapters related to future work in the other papers by manual reading. Subsequently, we annotate the FWS in the extracted chapters of EMNLP and NAACL according to the existing label specification and classification system. Cohen’s kappa coefficient (Cohen, 1960) is employed to measure the reliability of the labels and achieves over 0.75.

The information of annotated dataset is shown in Table 1 and Figure 1.

Table 1. Statistic information of Dataset.

Type	Count
Papers	9,508
Chapters	9,635
Sentences	62,312
FWS	10,622

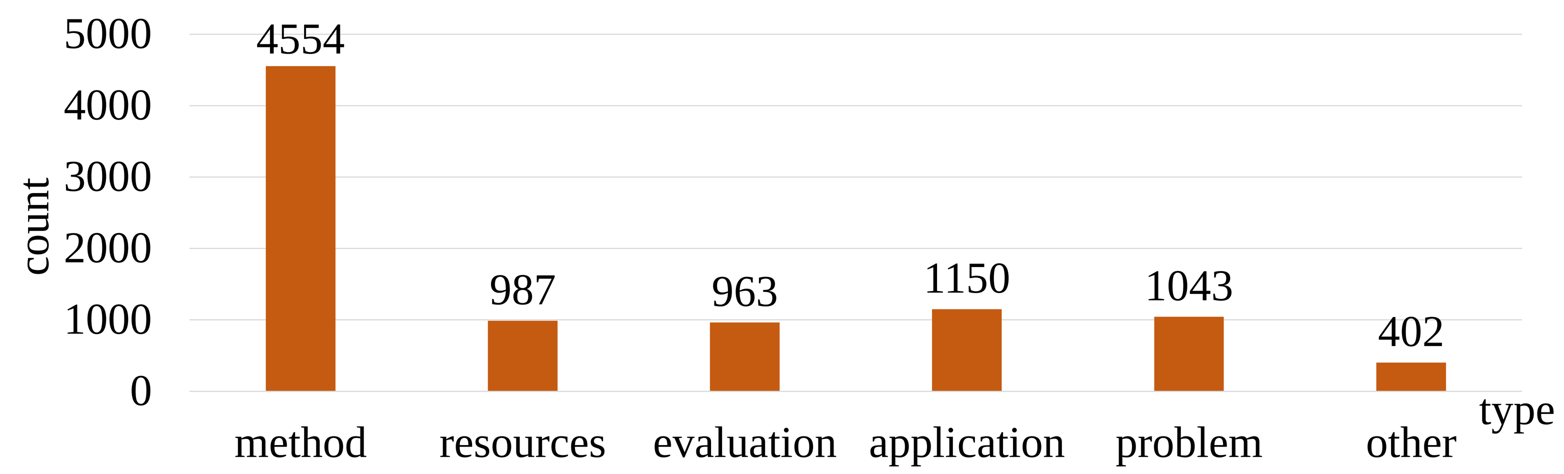

Figure 1. Number of FWS in each type.

Model Selection

For automatic recognition of FWS, we use four traditional ML models, including Naive Bayesian, Logistic Regression, Support Vector Machine and Random Forest, combining with three feature selection methods: filter, embedded, and wrapper. For automatic classification of FWS, we use the BERT pre-training model to represent the input data's features and input the acquired feature vectors into the SoftMax layer for classification. We evaluate the model performance by Precision, Recall, and F_1 score.

Result

Result of Automatic FWS recognition

The experiment result is shown in Table 2. Then we use the best model to recognize FWS from extracted chapters of ACL between 2016 and 2019. 1430 sentences are identified and we add them to our dataset. The ratio of FWS is used as an indicator, which is the share of papers that contain FWS. The result is shown in Figure 2.

Table 2. Performance of FWS recognition.

Model	Precision	Recall	F_1
LR	92.46%	81.05%	85.35%
SVM	92.39%	86.87%	89.31%
RF	92.58%	92.63%	92.08%
NB	91.22%	97.58%	**93.95%**

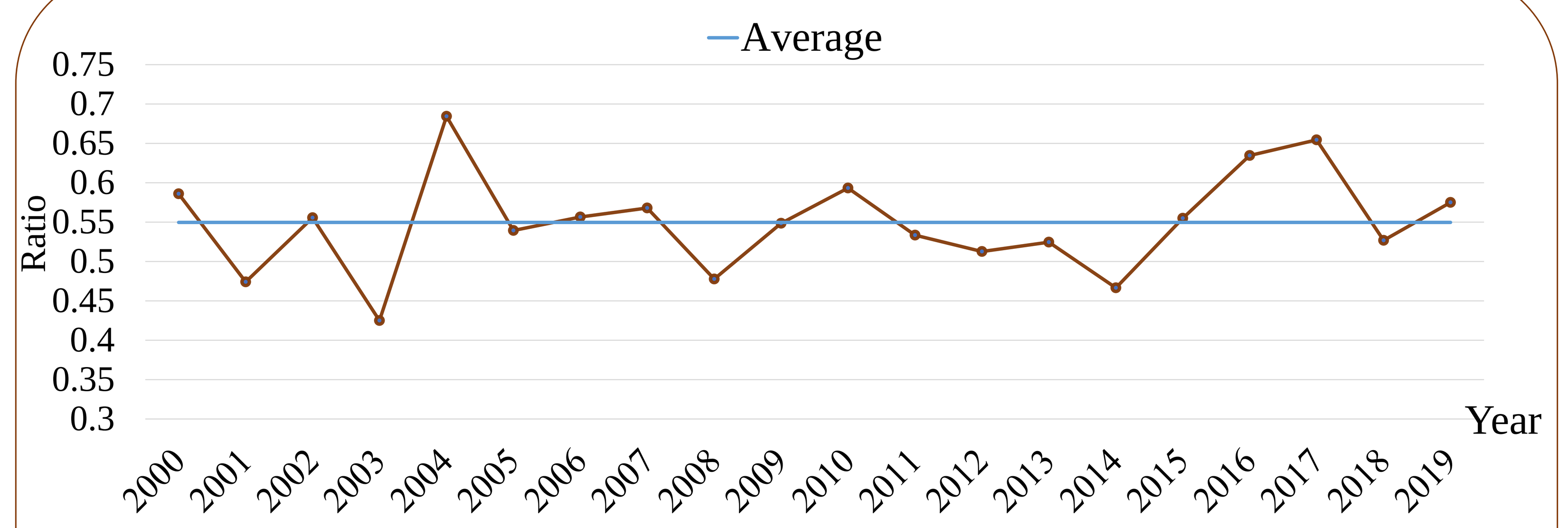

Figure 2. The ratio of FWS from 2000 to 2019.

Result of Automatic FWS classification

The experiment result is shown in Table 3. Similarly, we use the model to classify FWS without category labels extracted in the previous section. The evolution of the FWS types is presented in Figure 3.

Table 3. Performance of FWS classification.

Model	P	R	F_1
Method	89.98%	92.66%	91.30%
Resources	87.83%	82.11%	84.87%
Evaluation	91.86%	66.95%	77.45%
Application	81.32%	78.12%	80.00%
Problem	67.57%	96.15%	79.37%
Other	89.92%	63.64%	74.53%
Average	**87.20%**	**86.02%**	**85.91%**

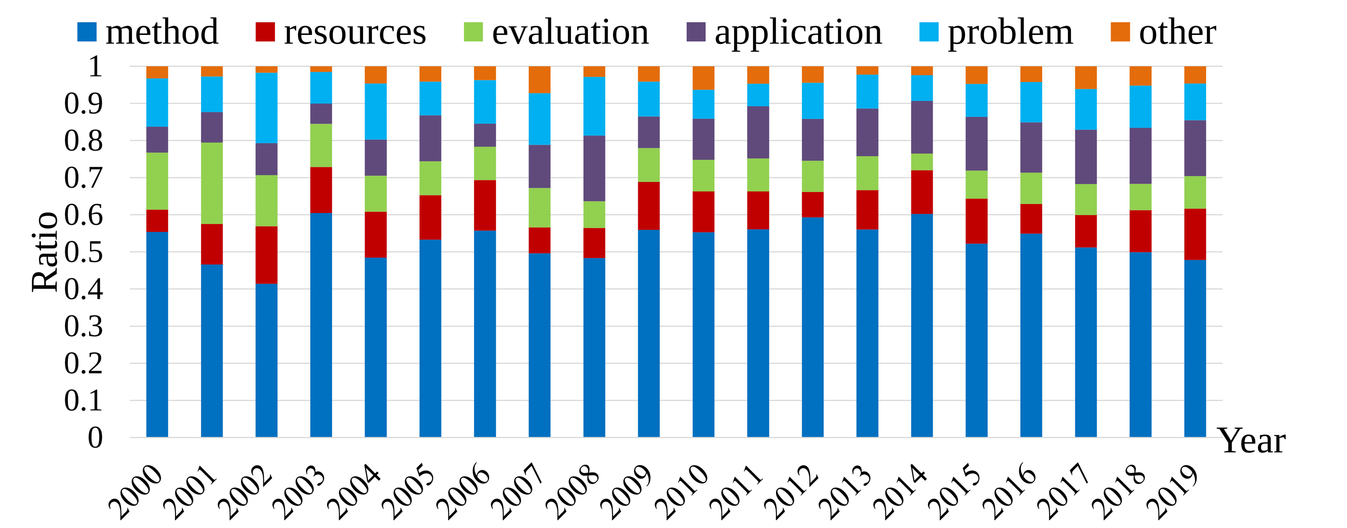

Figure 3. Ratio of FWS types from 2000 to 2019.

References

Hao, W., Li, Z., Qian, Y., Wang, Y., & Zhang, C. (2020). The ACL FWS-RC: A Dataset for Recognition and Classification of Sentence about Future Works. In Proceedings of the ACM/IEEE Joint Conference on Digital Libraries in 2020 (pp. 261-269).

Hu, Y., & Wan, X. (2015). Mining and Analyzing the Future Works in Scientific Articles. arXiv preprint arXiv:1507.02140.

Cohen, & J. (1960). A Coefficient of Agreement for Nominal Scales. Educational & Psychological Measurement, 20(1), 37-46.

Contact

Wenke Hao (haowk@njust.edu.cn)

Chengzhi Zhang (Corresponding author, zhangcz@njust.edu.cn)