Expert Knowledfge Graph and Its Application for The Transformation of Scientific and Technological Achievements

Quanzhong Wang

Wuhan University of Technology, Wuhan, 430070, china

957671408@qq.com

Abstract. In the era of knowledge economy, the transformation of scientific and technological achievements has become increasingly critical. This paper constructs expert knowledge graph according to the scenario requirements of school-enterprise cooperation projects, provides information technology support for school-enterprise cooperation, and improves the transformation efficiency of scientific and technological achievements. Firstly, this paper collects multi-source data information of experts from CNKI, Wanfang Data knowledge service platform, university official website and other databases for data integration. BERT+BiLSTM+CRF algorithm is used to extract the entity of expert scientific and technological achievements, and TF-IDF algorithm is used to generate expert labels for the scientific and technological achievements information of experts such as papers and patents. The expert knowledge graph is finally constructed by entity fusion based on neo4j graph database. The knowledge graph with expert labels constructed in this study can effectively explore the scientific research fields of experts, contribute to the transformation of scientific and technological achievements between universities and enterprises, and provide convenient conditions for enterprises to select suitable experts for cooperation.

Keywords: Knowledge graph, Transformation of scientific and technological achievements, Expert label.

1 Introduction

Although the transformation of scientific and technological achievements has received increasing attention, less than 20% of scientific and technological achievements can be transformed into productive forces in China every year, which is not only related to the lack of motivation for the transformation of scientific and technological achievements in colleges and universities, but also the difficulty of connecting enterprises with colleges and universities is the key reason for this phenomenon. In the cooperation between enterprises and universities, the key point of project initiation and achievement transformation lies in the docking of enterprise project requirements with experts and scientific and technological achievements. Due to the chaotic information of experts on

Z. Zhan et al. (eds.), *Proceedings of the 2024 10th International Conference on Humanities and Social Science Research (ICHSSR 2024)*, Advances in Social Science, Education and Humanities Research 858,
https://doi.org/10.2991/978-2-38476-277-4_50

the Internet and the scattered information of scientific research achievements, enterprises lack the preliminary understanding of experts and scholars in universities and universities and their research achievements, resulting in the failure of enterprises to select suitable target candidates for cooperation projects.

Google put forward the concept of knowledge graph in 2012. Knowledge graph can realize the unified and standardized expression of unstructured knowledge and has efficient data and knowledge retrieval ability. So far, all walks of life attach great importance to the application of knowledge graph combined with their own. Knowledge graph is divided into general knowledge graph and domain knowledge graph. The researches on general knowledge graph at home and abroad include YAGO[1], Dbpedia[2], CN-Dbpedia[3] and so on. In medicine, the application of knowledge graph is quite deep. Cheng [4] integrated the knowledge of traditional Chinese medicine through knowledge graph, and presented the knowledge in a structured series and visual way. Chai[5] takes thyroid disease as an example. Through the method of knowledge embedding, the entities and relationships in the knowledge graph are transformed into low-dimensional continuous vectors to construct the thyroid knowledge graph. Song et al. [6] proposed a deep learning method based on embedded representation to introduce time information into the knowledge graph by using gated loop units. The above studies fully show that the knowledge graph has significant value in medical fields, but at present, the research on the introduction of knowledge graph and its application into scientific and technological achievements is not extensive enough and is still in the initial stage. Therefore, driven by the demand for the transformation of scientific and technological achievements, this paper integrates multi-source heterogeneous data of scientific and technological achievements to build expert knowledge graph with experts as the core, and provides enterprises with suitable candidates for expert cooperation based on project requirements to fill the vacancy of practical application of knowledge graph in the transformation of scientific and technological achievements.

2 Knowledge Graph Construction

2.1 Design of Expert Knowledge Graph Schema

The construction of knowledge graph can be generally divided into two types: "top-down" and "bottom-up". The difference lies in whether the constructed knowledge itself has a fixed knowledge system. Therefore, this paper relies on the categories of expert scientific and technological achievements. Entities, attributes, and their relationships are extracted from the top-down knowledge graph generation mode, and the Schema of the expert knowledge graph is designed, as shown in Figure 1. Entities and attributes in the expert knowledge graph Schema include experts (age, working unit, people's experience), papers (abstract, keywords, topics, publication time, journals included), patents (patent type, application number, application date, claims), monographs (publication time, publication unit, publication word count), and scientific research projects (project source, project type, and project level Category), technology awards (awarding unit, award level, award type, award level) and labels. In the figure, ellipses are used to represent entity meanings, and moments represent attributes. Lines between figures are divided into two categories, with solid lines representing relationships between entities and dotted lines linking entities to their properties.

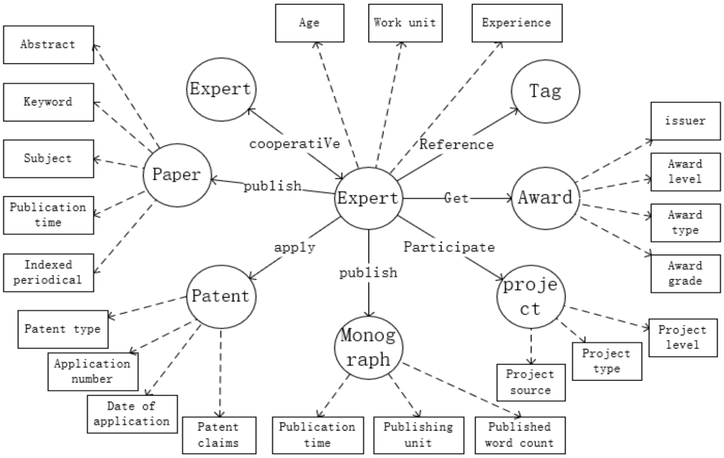

Fig. 1. Expert knowledge map schema

2.2 Data Collection and Processing

The basic information and achievement information of an in-service expert in a university on university portal websites, Baidu Encyclopedia, CNKI, Wanfang Data knowledge service platform, including papers, patents, monographs, scientific research projects, and science and technology awards, are crawled by web crawlers. Some crawlers are shown in Figure 2, and the data collected by crawlers are filtered, de-duplicated, and eliminated. A total of more than 2,300 pieces of experts' basic information data were obtained, and the total number of scientific and technological achievements was more than 52,000 pieces of data.

```
def analyze_paper(self, url):
    content = self.get_page(url)
    keyword = re.findall('target="_blank" class="">(.*?)</a></span>', content)
    if not keyword:
        print('          ')   #
        return None
    msg = dict(keyword=keyword)
    #
    title = re.findall('"{\'act_block\':\'main\',\'button_tp\':\'title\'}"\s*>\s*(\S*)\s*</a>', content)
    #
    author = re.findall('"{\'button_tp\':\'author\'}">(.*?)</a>', content)
    #
    abstract = re.findall('<p class="abstract" data-sign="">(.*?)</p>', content)
    #
    pub_time = re.findall('<p class="kw_main" data-click="{\'button_tp\':\'year\'}">\s*(\d*)\s*</p>', content)
```

Fig. 2. Data collection

2.3 Knowledge Extraction

There are many kinds of entity recognition methods, which can be divided into rule and dictionary based method, unsupervised learning method and supervised learning method from the model level. The names of scientific research projects and scientific awards cannot be matched and identified by pre-constructed dictionaries, and the basic unsupervised learning method does not consider contextual semantics in actual text processing. BERT, on the other hand, can not only learn contextual semantics, but also obtain word vectors with multiple semantic features and enhance the relationship features between words through pre-training. Therefore, BERT+BiLSTM+CRF algorithm is chosen for entity recognition in this paper. Compared with the traditional BiLSTM+CRF algorithm for named entity recognition, the Embedding layer is changed to BERT. BERT generates contextualized word vectors to improve the performance of entity recognition, as shown in Figure 3.

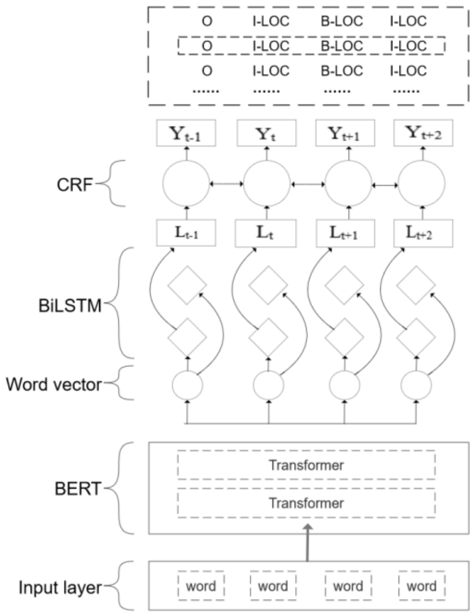

Fig. 3. Entity recognition algorithm model

Firstly, the entities in the data text of scientific research projects and scientific and technological achievements are marked by BIO labeling method. {B-,I-} are used as the prefix of the beginning word and other words of each entity text. Other irrelevant words in the text are marked with {O}. The single sentence after marking is divided into training set, test set and verification set according to the ratio of 6:1:1, and then trained by BERT+BiLSTM+CRF model, the maximum length of the input single sentence is set to 256. The sample batch batchsize of each round of experiments is set to 100. The maximum iteration number maxepoch is set to 100. The evaluation indexes of experimental performance include Precision, Recall and F1 values. Table 1 shows the identification of each entity and attribute.

Table 1. Optimal results of entity recognition experiment test set

Entity/Attribute	Precision	Recall	F1
project	98.14	99.23	98.54
Project type	98.25	98.49	98.33
Project level	83.67	97.18	89.26
Project source	76.94	87.47	81.43
Award	99.03	99.67	99.31
issuer	86.17	90.21	89.88
Award level	94.52	96.34	95.22
Award type	88.71	90.37	89.84
Award grade	95.47	96.31	95.97

2.4 Knowledge Graph Generation

After the entity, attribute and relational data identified in the previous article are organized into triplet data, python's third-party library py2neo is used to automatically import triplet data into graph database neo4j through Cypher statement, and the visualization function of neo4j can be used to display the expert visual knowledge map. Figure 4 is a partial screenshot of the expert knowledge graph data after visualization. The visualization of the graph is convenient to understand the relationship between experts and scientific and technological achievements, and to present the information of experts and their achievements in a structured manner.

Fig. 4. Local visualization of expert Knowledge graph

3 Conclusions

Aiming at the transformation of scientific and technological achievements, this paper analyzes the transformation needs of scientific and technological achievements in combination with the cooperation scenarios of school-enterprise projects. From the perspective of expert scientific and technological achievements, it integrates multi-source data information from CNKI, Wanfang data knowledge service platform, Baidu Encyclopedia and university websites, and builds an expert knowledge graph with expert labels. In addition, it provides knowledge search and intelligent question and answer application services for enterprises in project cooperation, solves the problem that it is difficult for enterprises to collect information blocked from universities, and the information of expert scientific and technological achievements is scattered and lack of structure, improves the transformation efficiency of scientific and technological achievements, and optimizes service quality and efficiency. The shortcomings of this paper are that the sample data is rarely limited to a single university, and the relationship between entities is not constructed through association rules, which makes it impossible to conduct knowledge reasoning. Future research hopes to increase the sample data volume, increase the combination of intelligent graph construction technologies, and rely on scientific and technological achievements to evaluate the expert ability, and further develop an expert recommendation system for the transformation of scientific and technological achievements.

References

1. Suchanek, F. M., Kasneci, G., & Weikum, G. (2008). YAGO: A Large Ontology from Wikipedia and WordNet. Journal of Web Semantics, 6(3), 203-217. doi: 10.1016/j.websem.2008.06.001.
2. Bizer, C., Lehmann, J., Kobilarov, G., Auer, S., Becker, C., Cyganiak, R., & Hellmann, S. (2009). DBpedia - A crystallization point for the Web of Data. Journal of Web Semantics, 7(3), 154-165. doi: 10.1016/j.websem.2009.07.002.
3. Xu, B., Xu, Y., Liang, J. Q., Xie, C. H., Liang, B., Cui, W. Y., & Xiao, Y. H. (2017, Jun 27-30). CN-DBpedia: A Never-Ending Chinese Knowledge Extraction System. Paper presented at the 30th International Conference on Industrial Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE), Arras, FRANCE. <Go to ISI>:// WOS: 000453052500044.
4. Cheng, B. Y., Zhang, Y., Cai, D. J., Qiu, W., & Shi, D. X. (2018, Aug 22-24). *Construction of traditional Chinese medicine Knowledge Graph using Data Mining and Expert Knowledge.* Paper presented at the 6th IEEE International Conference on Network Infrastructure and Digital Content (IEEE IC-NIDC), Guiyang, PEOPLES R CHINA.<Go to ISI>://WOS:000517807500043.
5. Chai, X. Q. (2020). Diagnosis Method of Thyroid Disease Combining Knowledge Graph and Deep Learning. Ieee Access, 8, 149787-149795. doi:10.1109/access.2020.3016676.
6. Song, F. F., Wang, B., Tang, Y. F., & Sun, J. (2020, Nov 12-14). Research of Medical Aided Diagnosis System Based on Temporal Knowledge Graph. Paper presented at the 16th International Conference on Advanced Data Mining and Applications (ADMA), Foshan, PEOPLES R CHINA. <Go to ISI>://WOS:000723588900019.