ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537

Knowledge Fusion Analysis and Research Based on Knowledge Graph

Tianqing Zhu

FiberHome Technologies Group

1257677800@qq.com

Abstract: With the rapid popularization of Internet technology, it is a new challenge to mine the hidden knowledge behind the data in an organized and directed way. Knowledge graph not only describes the content of semantic associations and entities existing in the objective world, but also uses graph structure to visually present the structured knowledge system for system users. Therefore, in the development of modern technology, knowledge graph has been widely concerned by industry and academia. In the era of big data, in order to build high-quality knowledge graph, knowledge fusion is very critical, among which entity alignment and entity link are important parts of knowledge fusion task. Starting from the field of film and television, this paper uses the entity alignment technology of NovEA model and the candidate entity ranking structure based on CNN-DSSM to deeply discuss the key technologies of knowledge fusion during the construction of multi-level film and television knowledge graph. The final experimental results prove that, compared with other models of the same task, the NovEA model studied in this paper has higher alignment accuracy, and the candidate entity ranking structure based on CNN-DSSM is better designed.

Keywords: Knowledge graph; Integration of knowledge; Solid alignment; Entity link; CNN-DSSM

1. Introduction

In the process of building a knowledge map, entities in the relative lack of information, usually a single knowledge base based on single data source to build knowledge map is easy to form the information isolated island, and inadequate knowledge coverage problem, and general knowledge map of structured knowledge source and a very wide range, led to the existence of the quality of the knowledge difference, knowledge overlapping problems between heterogeneous knowledge, Therefore, multi-source heterogeneous data should be fused into the target knowledge base through knowledge fusion to improve the accuracy and richness of knowledge graph. Entity alignment is a crucial part of knowledge fusion module. The quality of entity alignment model directly determines whether new entities can be aligned to the knowledge base, while whether new knowledge of entities can be fused to the knowledge graph indirectly affects the accuracy and richness of the knowledge graph. Therefore, the research of entity alignment method in knowledge fusion module is of great significance and value. In addition to the integration of structured entity information in the external knowledge base, there are also a lot of non-institutional data that can be used in general. It is necessary to extract entities in the text and related knowledge through information extraction technology such as named entity recognition. However, the entity names identified by entities still have the problem of ambiguity and diversification. Therefore, it is necessary to disambiguate the identified mentioned entities and the corresponding candidate entities in the knowledge base, so that the extracted knowledge can be added to the knowledge base to fill the knowledge base. Therefore, the research of entity link method in knowledge fusion module also has important significance and value.[1.2.3]

On the one hand, entity alignment techniques. Currently, the algorithms commonly used for entity alignment can be divided into the following four categories: rule-based, machine learn-based, deep learn-based, and embedded-based models. Next, it mainly analyzes the entity alignment model based on deep learning used in this topic. The entity alignment model based on deep learning can treat entity alignment as a text classification problem, treat entity attributes and their attribute values as the text representing the entity, and represent the text of the entity as a dense vector, and then judge whether two entities represent synonymous entities according to the classification results.

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537
Deep learning-based text classification mainly uses neural network structures such as RNN/CNN to extract features, improve the semantic expression ability of text, and then classify its embedded representation. Among them, the classic text classification is TextCNN proposed by Kim et al., which randomly initializes the word vectors in the text to obtain the vector representation of the input text, then extracts the text features through the convolutional layer, and finally transfers the text to the fully connected layer to obtain the output results. In addition, TextRNN proposed by Liu et al., they believe that RNN can process text with time series compared with CNN. The text sequence is input to RNN to extract sequence features, and then the node information of the last hidden layer is used as an abstract representation of the text, and finally it is transmitted to the fully connected layer to obtain output results.[4.5]

On the other hand, entity linking technology. Entity linking plays an important role in the construction of knowledge graph. The initial method is mainly to calculate the similarity, express the relevant features of the entity as feature vectors, and then calculate the cosine similarity. Honnibal et al. used Bow model to treat the entity with the highest cosine similarity score as the candidate entity. Later, researchers took into account the prior knowledge and entity categories of candidate entities, such as the popularity of entities and the relationship between referents and candidate entities. But these heuristic algorithms are difficult to capture more fine-grained semantic information and structural information. Later, the deep learning method became popular, and many scholars began to use the neural network method to learn the characteristics of entities and entity referents from different granularity. These scholars used the ranking method and the binary classification method. In this kind of method, the entity reference items and the context information of candidate entities are relatively large, but some information in the context has no strong correlation with the entity reference or candidate entities, which leads to the noise of this training model. Therefore, researchers began to combine attention mechanism with deep neural network to improve entity link model. On the basis of understanding the theory and construction process of knowledge graph, this paper proposes the entity alignment technology of knowledge graph based on NovEA model, which clarifies the process of knowledge fusion and provides an effective basis for modern technology research.

2. Method

2.1 Knowledge Graph

With the rapid development of artificial intelligence technology, the research process of knowledge graph is shown in Figure 1 below:[6.7]

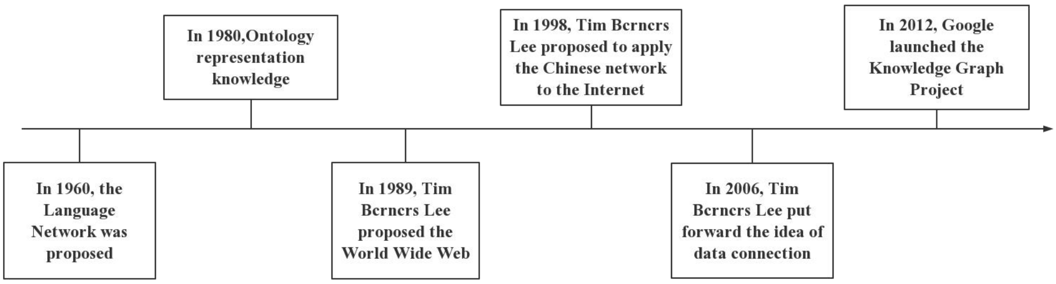

FIG. 1 The research process of knowledge graph

After putting forward the concept of Semantic Web, the concept of knowledge graph emerged after practical research. There are similarities and differences between Semantic Web and knowledge graph, among which the similarities are abstracted concepts that process the real world and form knowledge base, where nodes in the library represent things and edges represent relationships between each other. The difference is that Knowledge graph pays more attention to the relationship between entities, compared with the past, which focuses more on the relationship between concepts.

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537

The knowledge graph is mainly described in the form of graph. The attributes, entities, semantic associations and other contents in the objective world are stored in the least end member of the triplet, which contains two forms, one refers to entity A, relation and entity B, and the other refers to entity, attribute and attribute value. In order to build a knowledge graph, it is common to follow the following process as shown in Figure 2:[8]

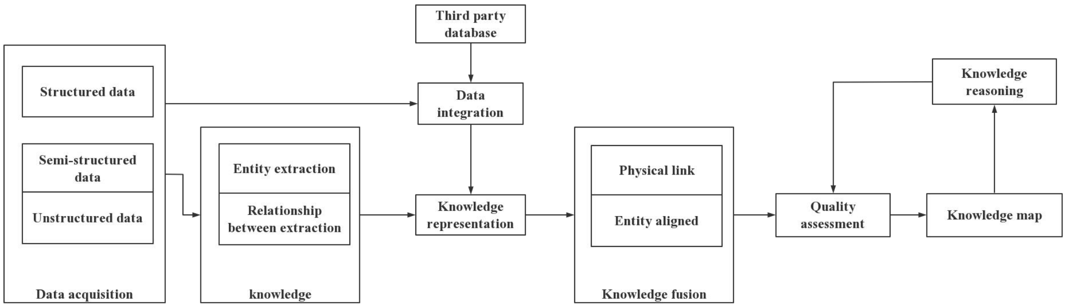

FIG. 2 Flowchart of knowledge graph construction

First, data acquisition. Abundant and sufficient data is the prerequisite for building a perfect knowledge graph. Therefore, it is necessary to obtain sufficient domain data, mainly through the crawler framework to automatically crawler the data of a certain domain, which can not only reduce the consumption of human and material resources, but also improve the efficiency of practical work.

Second, knowledge extraction. After acquiring domain data, it is not possible to construct knowledge graph directly due to their diverse forms. Therefore, this paper proposes to use dependency parsing to extract syntactic relations and collocation relations. For example, subject-predicate relation, object-predicate relation and so on.

Third, knowledge integration. The extracted knowledge is mostly scattered, repetitive and uncertain. Therefore, the fusion operation should be carried out to ensure that the extracted knowledge is correct and non-redundant before it can be applied to the knowledge graph. In this paper, two subtasks, entity alignment and entity linking, are used to complete the fusion work.

Fourth, knowledge reasoning. This part is mainly to supplement the entities and relationships that are missing in the knowledge graph to be built, and also to dig deep the deep tacit knowledge between different entities.

Fifth, quality assessment. After the preliminary construction of knowledge graph, the accuracy of knowledge graph cannot be guaranteed at this time. Therefore, to evaluate the quality of knowledge graph, knowledge reasoning can be used to connect the prediction process.[9.10]

2.2 Entity alignment method based on NovEA model

Nowadays, knowledge graph plays an increasingly significant role in search engine, information retrieval and recommendation system, and more and more industries and fields begin to construct knowledge graph in various forms. In order to form a more unified knowledge graph, researchers begin to integrate such knowledge graphs and actively search for entities that exist in different knowledge graphs but have the same meaning. This task is called entity alignment. The entity alignment method of knowledge graph based on NovEA model is shown in Figure 3 below:

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537

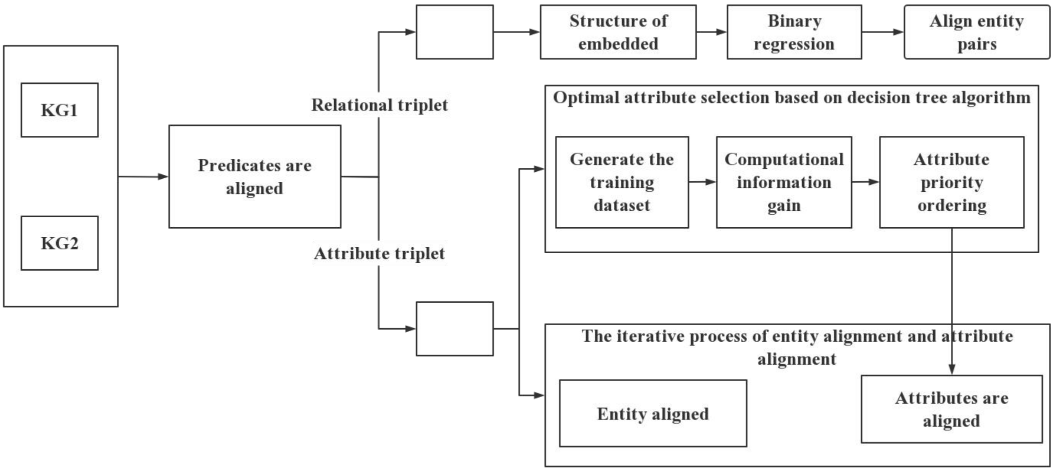

FIG. 3 Framework diagram of entity alignment method for knowledge graph based on NovEA model

Combined with the analysis of the above figure, it can be seen that this model makes full use of the attribute triplet combination relation triplet in the knowledge graph, and prioritises the attributes according to the domain characteristics in the knowledge graph. The overall operation involves the following points: First, in the predicate alignment module, the predicates in triples are aligned and named to facilitate the embedding of relations and entities into the same vector space. Secondly, relational triples and attribute triples in data sets are used to embed structure and attribute respectively. In attributes embedded in the module to choose suitable alternative, is the alignment algorithm based on public property values, and then based on the match to the entity attribute alignment, finally according to the field of knowledge map feature and the decision tree algorithm in the complete prioritization, and align attribute triples the similarity of results in weight constraints; Finally, in the entity alignment module, the binary regression model is used to integrate the results of the relational triplet combination attribute triplet, dynamically learn the corresponding weight value, and scientifically adjust the weight value between the relationship and the attribute, so as to improve the accuracy of the entity.

2.3 Candidate entity ranking structure based on CNN-DSSM

Based on the massive click exposure log of Query and Title in the search engine, I expressed Query and Title as low-dimensional semantic vectors with DNN, calculated the distance of the two semantic vectors through cosine distance, and finally trained the semantic similarity model. This model can be used not only to predict the semantic similarity of two sentences, but also to obtain the low-latitude semantic vector representation of a sentence. The overall structure of CNN-DSSM algorithm is shown in Figure 4:

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537

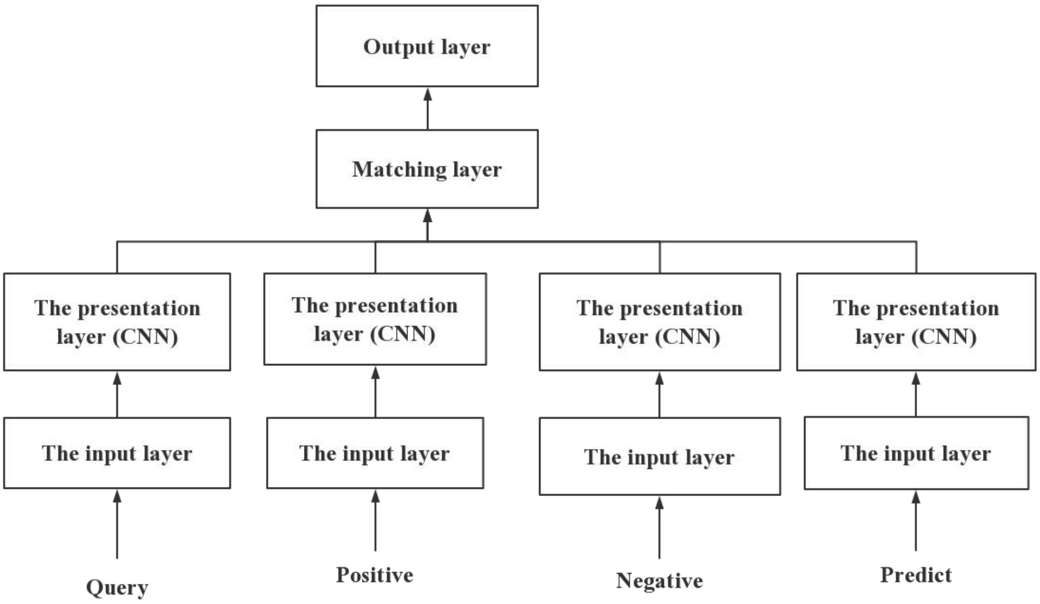

FIG. 4 Overall structure of CNN-DSSM algorithm

First, the input layer. The main method is to preprocess the corpus and treat the obtained information as the input word vector. Second, the presentation layer. This level design includes convolutional layer, pooling layer and fully connected layer. Among them, the convolution layer is to extract the context feature under the sliding window, the pooling layer is to find the global context feature for the sentence, and the fully connected layer will transform the high vector into a low wing vector. Finally, match the layers. At this level, the cosine similarity algorithm is used to calculate and analyze the correlation.

3. Result analysis

After clarifying the entity alignment technology based on NovEA model, this paper selects three real knowledge graph data sets, namely IMDB-YAGD, DBP-Yago, Movie_data, and focuses on the accuracy of model application. The experimental results of optimal attribute selection based on decision tree are shown in Table 1 below:

Table 1 Analysis of the results of optimal attribute selection

methods	IMDB-YAGO			DBP-YAGO			Movie_data		
	P	R	F	P	R	F	P	R	F
TransE	87.06	63.79	73.63	92.97	88.58	93.49	80.56	79.21	79.87
JAPE	90.01	75.58	82.17	95.21	90.23	92.65	91.52	78.65	84.62
AttrE	85.03	82.62	84.67	95.98	90.55	93.18	86.73	82.14	84.37
NovEA	94.39	80.26	86.79	97.25	94.45	95.82	92.23	79.32	85.29

Combined with the above analysis, it can be seen that the accuracy of NovEA model studied in this paper on the three data sets is improved by 7.3%, 5% and 11%, respectively. The reason is that the NovEA model uses the optimal attribute selection process of the decision tree to select the best attribute type according to the domain features of the knowledge graph, so as to ensure more accurate attribute embedding results of entities and effectively improve the final entity alignment effect. The final experimental results prove that the accuracy and value of NovEA model have been improved in several data sets selected in this paper. By comparing and analyzing different types of data combination, it can also be found that NovEA model can automatically select the optimal attributes according to data features.

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537

At the same time, the binary regression model is used to dynamically learn the weight of similarity measure from the two aspects of relation and attribute, and the combination method of the same similarity weight as shown in Table 2 below can be obtained. This proves that the effect of dynamic combination of NovEA model is stronger than that of static combination, and the dynamic combination will learn the weight value during entity alignment, and dynamically obtain the importance of domain features of knowledge graph according to the relationship between entities and attributes. In addition, compared with other entity alignment models, the experimental results show that the candidate entity ranking algorithm based on CNN-DSSM is better than other algorithms.

Table 2 Comparative analysis of static combination and dynamic combination

The data set	Static combination	Dynamic combination
DBP-YAGO	85.67	87.36
IMDB-YAGO	87.11	88.65
Movie_data	86.75	89.25

4. Conclusion

To sum up, knowledge graph is widely used in data mining, search engine and other aspects because of its powerful function. With the gradual expansion of the data scale of the Internet, scholars from all over the world began to discuss how to collect fragmented knowledge together to form a knowledge system in mining applications. Research, from the perspective of the film and television field, this paper probes into the entity aligned to NovEA model as the core technology, as an important part of the knowledge map of knowledge fusion, can from on the basis of ensuring the accuracy of the data and information and effectiveness, can be fully USES the properties of triples and the relation of knowledge map triples, so is the main topics of the current technologies to explore in our country.

Reference

[1] Liyan Shi,Lu Ye , Chuan. Tang Research Situation analysis of Integration development of 5G and artificial Intelligence in China: Based on bibliometrics and Knowledge graph [J]. World Science and Technology Research and Development, 2021, 43(6):732-749.

[2] Zhiyong Liu , Ping Chen,Zongju Zhong . Research status and trend analysis of domestic urban integration based on knowledge graph [J]. 2021(2017-2):48-67.

[3] Zihan Zhao, Qichao Wei. Dynamic Tracking and Analysis of hot spots and Frontiers in the Research of physical education Integration -- Visual Analysis Based on Knowledge Graph [J]. Contemporary Sports Science and Technology, 2022, 12(5):4.

[4] Wei Sun. Research hotspot, frontier and trend analysis of the integrated development of producer services and manufacturing industry: Based on the knowledge graph perspective [J]. Review of Industrial Economics, 2021(3):11.

[5] Yuqing Zheng ,Chujia Liu . A review of the Integrated Development of Vocational Education and Community Education in China (2000-2020) : Based on CiteSpace Knowledge Graph analysis [J]. Adult Education, 2021, 41(10):9.

[6] Li Wei ,Junhua Sun . Research on hotspots and trends of production-education integration based on knowledge graph [J]. Journal of Teachers College of Qingdao University, 2020, 008(005):60-67.

[7] Xingwu LUO ,Zhiyi Lin , Yang Liu, et al. Platform Research: Frontier Evolution and Theoretical Framework: Based on CiteSpace Ⅴ Knowledge Graph analysis [J]. Science and Technology Progress and Countermeasures, 2020, 37(22):9.

[8] Lin Gan, Shuzhong Zhou. Visual analysis of home-made education integration research based on knowledge graph [J]. Jiangsu Science and Technology Information, 2021, 38(25):6.

ISSN:2790-1688 DOI: 10.56028/aetr.2.1.537
[9] Jiafu Cheng , Qing Liu,Yaqian Zhao , et al. Review on the Integration of industry and Education in Higher Vocational Colleges: Bibliometric analysis based on Knowledge Graph [J]. Life Education, 2021(8):7.

[10] Rui Zhou ,Linyue Xing . Research on integration of urban public crisis information resources based on knowledge graph [J]. Research on Modernization of National Governance, 2021(1):19.