Computer Life (CPL) ISSN: 1819-4818
DELIVERING QUALITY SCIENCE TO THE WORLD

Research on the Construction of Knowledge Graph Data Model

Sheng Dai *

Guangdong University of Science and Technology, Department of Computer Science, Dongguan,

Guangdong 523083, China.

*dai1879424@163.com

Abstract: The knowledge graph is essentially a structured semantic knowledge base, which is now widely used in the field of artificial intelligence. This paper analyzes the three main processes of knowledge graph construction from the perspective of data model construction, introduces the RDF model and attribute graph model in the knowledge representation process, the entity alignment and ontology alignment of knowledge fusion, and briefly describes the concept and content of knowledge completion.
Keywords: Knowledge graph, Knowledge representation, Knowledge fusion, Knowledge completion.

---

1. INTRODUCTION

The concept of knowledge graph was first proposed by Google in 2012, and it has aroused widespread concern in academia and industry. Knowledge graph is an important part of artificial intelligence research, and is currently widely used in intelligent question answering and recommendation systems. This article is mainly an overview of the construction process of the knowledge graph, focusing on the analysis of the construction of the data model, and discussing the main processes and problems.

2. INTRODUCTION TO KNOWLEDGE GRAPH

Regarding the definition of the knowledge graph, it is generally accepted that the knowledge graph is essentially a semantic network. Some researchers have defined it as a structured semantic knowledge base, the meaning of the two is actually similar. Because the knowledge graph is represented by a graph structure. Knowledge graph is a graph structure used to represent the connection between entities, and it is a semantic network composed of nodes and edges. Nodes can be entities, and edges are attributes or relationships of entities.

The process of knowledge graph is generally divided into knowledge representation, knowledge acquisition, fusion, storage, and application of knowledge [1]. The construction process of knowledge graph is essentially the process of structuring a pile of data to form a data model of knowledge representation, and finally the process of applying knowledge.

3. THE PROCESS OF CONSTRUCTING KNOWLEDGE GRAPH MODEL

Three processes of constructing a model of knowledge graph: information extraction and representation, knowledge fusion, and knowledge complementation. Every aspect of this process is crucial. The main process is shown in the dotted line in figure 1.

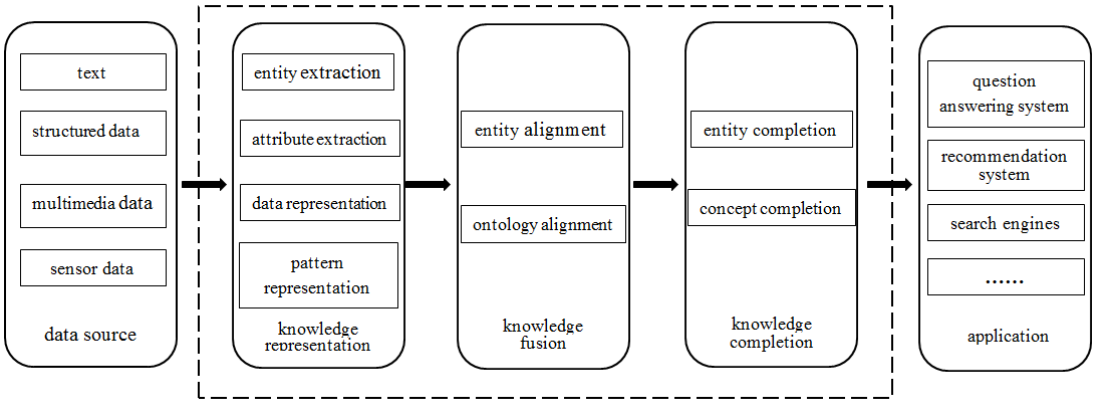

Figure 1 Model building process

4. KNOWLEDGE REPRESENTATION

Information extraction is to extract entities, attributes, relationships, and then use appropriate methods to represent these entities and the relationships between entities. Commonly used representation models are RDF and attribute graph models.

The RDF model is a standardized representation used by the knowledge graph. The RDF model defines a triple: node-edge-node, which respectively represents "entity-relationship-entity", and its semantics are subject, predicate, and object, thus forming a record. The schematic diagram is shown in Figure 2。

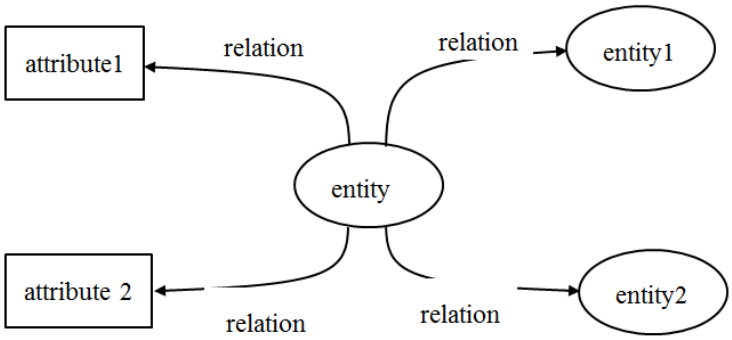

Figure 2 RDF model

Although the RDF model can express the intrinsic semantics of knowledge in a simple and clear manner, the representation of triples requires the design of a special graph algorithm to calculate the semantic and inference relationships between entities. The calculation is complicated and the scalability is poor [2].

The attribute graph model is another commonly used data model. The attribute graph model is a quintuple (V, E, ρ, λ, σ), where V represents nodes, E represents edges, ρ represents the direction between nodes, λ represents labels, and σ represents attribute values. In the attribute graph model, the relationship of nodes must be directional, and nodes, attributes, and relationships all have unique identifiers. Nodes and edges can also be labeled to indicate the type of data [3]. Neo4j is the current

mainstream application based on the attribute graph model, which can provide highly questionable data. Compared with the RDF model, the attribute graph model is not perfect in theoretical research, and the semantic expression is not strong enough.

In the process of data model construction, we first look for a similar knowledge system, based on which we collect data, and then expand and modify the data model. According to the collected entities and attributes, constantly adjust the relationship between entities and determine the attributes and the scope of attributes.

The logical structure of the data model of the knowledge graph is divided into two levels: the data layer and the pattern layer. The data layer is stored in units of fact data. For example, the RDF graph model and the attribute graph model introduced above are both data layer models.

The logical structure of the pattern layer is above the data layer, which is a deeper induction and generalization of knowledge. The model layer is actually a description of knowledge, which is equivalent to a kind of specification, which restricts and regulates concepts and attributes. Ontology is usually used to manage the pattern layer of knowledge graph. Ontology library is equivalent to the mold of knowledge base, which is used to guide how to select and express entities, attributes and relationships [4].

5. KNOWLEDGE FUSION

The knowledge graph can obtain data resources from various ways, such as documents, structured databases, multimedia data, web pages, etc. The data structure of each data source has different characteristics, and the representation method and storage method used are also very different. For these different data, it is necessary to use certain technical means for unified representation to form a unified data structure, which is convenient for the representation and storage of knowledge. This requires knowledge fusion.

Knowledge fusion includes two levels: one is model layer fusion, and the other is data layer fusion, which mainly solves the problem of data consistency and eliminates ambiguity, redundancy and errors. Knowledge fusion mainly includes two parts: entity matching and ontology alignment.

Entity matching is also called entity alignment. It mainly eliminates the problems of entity conflicts and unclear pointing. The main tasks include entity disambiguation and common finger disambiguation. Entity disambiguation is to solve the problem of different connotations of entities represented by the same name. The main technical methods include entity alignment method based on the similarity of entity attributes, space vector model alignment method, probability model alignment method, etc. Co-referential disambiguation refers to the fact that many different attributes or names refer to the same entity, and how to eliminate the influence of synonymous entities. The rapid development of machine learning methods has brought new ideas to the common finger disambiguation, and transformed the common finger disambiguation problem into a classification or clustering problem, so as to effectively realize the common finger disambiguation.

Ontology alignment is actually to solve the problem of consistency at the pattern level. Ontology is a constraint on knowledge and reflects the consensus of knowledge. The main problem of ontology alignment is to solve the management problem of consensus model changes. The current main process of solving ontology alignment: ① Calculate the parallel similarity of entities; ② Extract the

upper and lower relationship of entities; ③ Entity clustering [5]. Ontology alignment is actually the calculation of entity similarity.

6. KNOWLEDGE COMPLETION

There is a problem in the field of knowledge graph that cannot completely solve the problem of knowledge completeness, that is, it is impossible to obtain all the knowledge, and a large number of implicit relationships have not been tapped [6]. In response to this problem, it is necessary to constantly modify and perfect the knowledge graph during the construction of the data model.

Knowledge completion can be divided into concept-level completion and entity-level completion, mainly to supplement entities, attributes, and entity relationships. The main methods used at the conceptual level are logic-based, machine-based learning, content-based, and statistical relationship-based inference methods; the entity-level methods include representation-based learning completion, matrix decomposition, and translation methods [7].

7. SUMMARY

This paper takes the construction of the data model in the knowledge graph as the main line, analyzes the main process of knowledge graph construction, analyzes the RDF and attribute graph model of knowledge extraction and representation in more detail, and discusses the process of knowledge fusion and knowledge completion. Knowledge graph technology has become an important research direction of artificial intelligence. It is believed that in the near future, knowledge graph will have greater development and make artificial intelligence technology more perfect.

REFERENCES

[1] CCF.2017-2018 China Computer Science and technology development report[R]. China Machine Press, 2018.

[2] Zhiyuan Liu, Maosong Sun, Yankai Lin, et al. Knowledge Representation Learning:A Review [J]. Journal of Computer Research and Development,2016,53(02):247-261.

[3] Xin Wang, Lei Zou, Zhaokun Wang, Peng Peng, Zhiyong Feng. Research on Knowledge Graph Data Management: A Survey [J]. Journal of Software, 2019,30(07):2139-2174.

[4] Qiao Liu,Yang Li, Hong Duan, Yao Liu, Zhiguang Qin. Knowledge Graph Construction Techniques [J]. Journal of Computer Research and Development,2016,53(03):582-600.

[5] Zenglin Xu, Yongpan Sheng, Lirong He, Yafang Wang. Review on Knowledge Graph Techniques [J]. Journal of University of Electronic Science and technology of China, 2016, 45(04):589-606.

[6] Jianhui Ding, Weijia Jia. The Research Advances of Knowledge Graph completion algorithm[J].Information and Communications Technologys, 2018,12(01):56-62.

[7] Shuo Wang, Zhijuan Du, Xiaofeng Meng. Research progress of large-scale knowledge graph completion technology [J]. SCIENTIA SINICA Informationis, 2020,50(04):551-575.

Author: Sheng Dai（1992-）, Native place: Hubei, full-time teacher in the School of Computer Science and Technology of Guangdong University of Science and Technology, software engineering、big data direction.