Neural Computing and Applications (2021) 33:681–690
https://doi.org/10.1007/s00521-020-05057-5

S.I. : DPTA CONFERENCE 2019

Domain knowledge graph-based research progress of knowledge representation

Jinjiao Lin[1] · Yanze Zhao[1] · Weiyuan Huang[2] · Chunfang Liu[1] · Haitao Pu[3]

Received: 27 January 2020 / Accepted: 3 June 2020 / Published online: 21 June 2020

Abstract
Domain knowledge graph has become a research topic in the era of artificial intelligence. Knowledge representation is the key step to construct domain knowledge graph. There have been quite a few well-established general knowledge graphs. However, there are still gaps on the domain knowledge graph construction. The research introduces the related concepts of the knowledge representation and analyzes knowledge representation of knowledge graphs by category, which includes some classical general knowledge graphs and several typical domain knowledge graphs. The paper also discusses the development of knowledge representation in accordance with the difference of entities, relationships and properties. It also presents the unsolved problems and future research trends in the knowledge representation of domain knowledge graph study.

Keywords Domain knowledge graph · Knowledge representation · Entity · Relationship · Property

1 Introduction

Domain knowledge graph (industry knowledge graph or vertical knowledge graph) is based on domain-specific data. Domain knowledge graph is different from general knowledge graph which contains common sense information. Information in a domain knowledge graph is mostly suitable for a specific industry. It contains more complex knowledge and structure and plays an important role in domain information integration. With the in-depth study of knowledge graph and the progress of artificial intelligence technology, the construction of domain knowledge graph has more technical support, and studies of domain knowledge graph have gradually become a heated research topic.

At present, most researches on knowledge graph construction focus on the construction of general knowledge graph, and there are still gaps between the research of domain knowledge graph construction which contains more complex information and data. Knowledge representation, as a first step to knowledge graph construction, is the foundation of knowledge graph construction. Firstly, this research introduces knowledge representation of classical general knowledge graphs and current typical domain knowledge graphs. Secondly, it discusses the development of knowledge representation according to the difference in entities, relationships and properties.

The relevant concepts of knowledge representation are introduced in the second part. The third part discusses the knowledge representation of general knowledge graphs in different categories; the fourth part introduces the knowledge representation of typical domain knowledge graphs. And some unsolved problems in knowledge representation research have been presented in the last part.

✉ Haitao Pu
pht@sdust.edu.cn

[1] School of Management Science and Engineering, Shandong University of Finance and Economics, Jinan 250000, China

[2] School of Marxism, Shandong University of Finance and Economics, Jinan 250000, China

[3] College of Electrical Engineering and Automation, Shandong University of Science and Technology, Jinan 250000, China

2 Basic concept

2.1 Concept of knowledge representation

Knowledge representation is a set of rules to describe the world. It is the symbolization, formalization or modeling of the knowledge. Different knowledge representation methods are different formal knowledge models. In the concept of knowledge engineering representation, representation is a computer model describing the natural world, and it should meet the specific limitations of computers. Therefore, representation can be understood as a kind of data structure and a set of operations. It emphasizes the image form of the information of the natural world in a certain type of data structure in the computer and the processing methods adopted for the stored contents [1]. Knowledge representation defines the domain basic cognitive framework, which defines the basic concepts in domain and the basic semantic relationships between concepts [2].

2.2 Relevant concepts

Knowledge representation is a part of the study of knowledge graph. Wang Haofen, a professor of East China University of Science and Technology, believes that the purpose of knowledge graph is to describe various entities or concepts existing in the real world. Each entity or concept is identified with a globally unique ID, which is called identifier. Each property-value pair is used to describe the internal characteristics of an entity, and the relationship is used to connect two entities and describe their associations [3]. Therefore, knowledge representation not only involves knowledge, but also involves entities, properties and relationships.

Entities refer to something that exists in the objective world. Entities have distinction and can exist independently. An entity is a basic unit of the knowledge graph [4]. In the knowledge graph, nodes are regarded as entities [5]. A semantic class/concept is an abstract name for a set of things that have the same characteristics. A property refers to a property value from an entity, which describes the characteristics of this entity [5]. Relationships describe the contact between two or more entities [4, 5].

Knowledge is the set of all facts, concepts, rules or principles. Here, set is acquired and summarized by observing, learning and thinking about various phenomena in the objective world [3]. Knowledge includes language knowledge, commonsense knowledge, encyclopedic knowledge, domain knowledge, etc. Therefore, knowledge graph can be divided into language knowledge graph, commonsense knowledge graph, encyclopedic knowledge graph and domain knowledge graph [6]; more broadly, they can be summarized as general knowledge graph and domain knowledge graph according to their scope of application [4].

3 General knowledge graph

This section mainly introduces some general knowledge graphs and analyzes their knowledge representation. General knowledge graph includes language knowledge graph, commonsense knowledge graph and encyclopedic knowledge graph. This section selected several typical general knowledge graphs in different categories and analyzed their similarities and differences in knowledge representation. These knowledge graphs are not built successfully at one time, and each new knowledge graph is formed in the optimization of the original knowledge graph. The knowledge representation between them is different though there are also some similarities.

3.1 Language knowledge graph

Before the popularization of the Internet, lots of original expert systems and knowledge bases were constructed artificially, among which the early representative model was WordNet [6], and it can be regarded as a language knowledge graph. WordNet is the achievement of the task to develop a dictionary database undertaken by a group of psycho lexicologists and linguists at Princeton University since 1985 [7]. WordNet is regarded as an online electronic (synonymous) dictionary system, which is a proposal to combine traditional dictionary information with modern high-speed computing more effectively [8]. Its knowledge mainly comes from artificial construction, and we can simply understand WordNet as an electronic dictionary with some encyclopedic knowledge and real relationships.

WordNet assumes the role of a dictionary database, so the division of concepts or entities is not very clear. In our research, we will mainly analyze its knowledge representation through words. It takes the word as the smallest unit and builds a set of synonyms with the same meaning words. In addition, it defines a name for them and treats them as a concept. For example, *car, railcar, railway car, railroad car—(a wheeled vehicle adapted to the rails of railroad; "Three cars had jumped the rails")*—are a synonym set for streetcars; each synonym set includes a description of the concept. A word may have more than one meaning, so it may appear in more than one synonym set.

WordNet contains more relationships between words than entities. It includes the following basic relationships: synonym relationship, antonym relationship, hypernym and hyponym relationship, whole and part relationship [7–9]. Synonym relationship is a most basic relationship in WordNet, because word nodes are related by synonym relationship. It refers to the relationships between words with the same semantic meaning. Antonym relationship is mainly the relationships between the adjectives, and there are direct antonym relationship and indirect antonym relationship. Direct antonym relationship refers to the relationships between words that have opposite meaning. Indirect antonym relationship refers to the relationships between a word and the antonym of its synonym, such as Fig. 1 [8]; *good* and *bad* is the direct antonym relationship, and *great* and *bad* is the indirect antonym relationship. The hyponymy relationship (*is kind of, is a generalization of*) mainly refers to the hierarchical relation between nouns. For example, the hypernym of *dog* is the synset of *animal*, and the hypernym of *animal* is the synset of *organism*. A word may contain many hyponymy words. WordNet contains 25 basic classes (final hypernym). WordNet also includes the relationships between whole and part (*is part of*), which means a *is a component of* b, a *is members of* b or a *is substance of* b. Relationships in WordNet are generally connected by pointers, but some noun entities are also connected by relational adjectives [8].

WordNet can be regarded as an electronic dictionary with encyclopedic knowledge, and it contains some basic features and functional properties. It takes some real things in the noun set as an entity, such as robin, dog, apple, etc. Their feature properties are linked to adjectives, and function properties are linked to verbs. An entity is connected with a descriptive property by property name, such as *WEIGHT(package) = heavy*. It means that *package* has a property *WEIGHT*, and the property value is *heavy*. An entity and its functional properties are linked by pointers between noun and verb sets.

3.2 Commonsense knowledge graph

Language knowledge graph (WordNet) realized the task of adding some simple facts to knowledge network, but it lacks representation of real events. Real-life events cannot be described by one word, and general relationships and properties also cannot show more facts in the real world. So people build commonsense knowledge graph which includes Cyc [10] and ConceptNet [11]. The knowledge in Cyc is represented by higher-order logic; according to concepts and assertions, Cyc uses argumentation to achieve the reasoning of common sense knowledge based on contextual contents [12–14]. ConceptNet originated in the

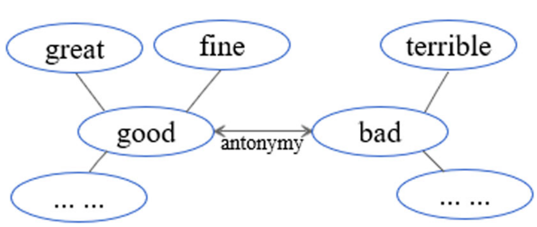

Fig. 1 Antonym relationship

Open Mind Common Sense (OMCS) project at the MIT Media Lab, which was proposed by famous artificial intelligence expert Marvin Minsky in 1999 [4]. It is a crowd sourcing knowledge base and a semantic network. ConceptNet contains a lot of common sense information, which helps computers to understand the real world. It extends the knowledge network that language knowledge graph has built. This section focuses on the knowledge representation of ConceptNet.

The node of ConceptNet is no longer a simple word, but semistructured English segments. It is usually a compound phrase or an action. ConceptNet contains more abundant knowledge of the real world than WordNet. It focuses on the common sense meaning of the natural language words (Unnamed Entity). It will connect the concepts with a lot of common sense and contains a lot of the complex relationships in real world. In addition, it can be better in context-based reasoning, whose knowledge mainly comes from OMCS project, expert knowledge base, purposeful games and other knowledge bases. It is a knowledge base automatically constructed by extracting knowledge according to certain rules [4].

The concepts and entities in ConceptNet are mainly composed with words or phrases. Najmi et al. [15] regard that entities in ConceptNet are described by noun phrase (NP). These phrases usually consist of one or more main nouns as root, with one or more other words to describe this main noun. Verb phrase (VP) is used to describe a concept sometimes. These concepts are usually extracted from the text of natural language [4]. They are more consistent with our usual expression habits. Its nodes contain not only the entity such as people, objects and regions, but also some of our actual action states, such as drink coffee, eat breakfast; these nodes help the expression of commonsense knowledge. Compared with the simple knowledge in language knowledge graph, ConceptNet passes commonsense information to the computer. For example, an entity *apple* is a fruit in WordNet and red is a property of it. But in ConceptNet, it might correspond to *Apple inc.*, rather than a fruit. The specific property and meaning must be inferred from its context.

ConceptNet5 includes 21 predefined and multilingual generic relationships (e.g., *IsA, UsedFor*, etc.) and nonformal relationships extracted from natural language texts that are closer to natural language descriptions (e.g., *on top of, caused by*, etc.) [4]. Compared with the hyponym relationship in WordNet, ConceptNet can summarize commonsense topics or categories by text and connect them with *SuperThematicKLine*, such as *buy food* and *purchase food* can be connected with *buy* by *SuperThematicKLine*. In addition to simple relationships in WordNet (*IsA, PartOf, MadeOf, SimilarTo,* etc.), ConceptNet also contains many complex relationships in reality, such as *fall* *off bicycle* and *get hurt* connected by the relationship *EffectOf*. There is no such commonsense knowledge in WordNet, nor as relationships [16]. Examples of ConceptNet are shown in Fig. 2 [17].

Najmi et al. [15] analyzed relationships and properties in ConceptNet from the upper ontology construction. They believe that properties in ConceptNet are not defined like general ontology relationships. Even if some properties are logically incorrect, it may be "meaningful" in common sense. For example, *succeed* as a property is connected with *a person* through *Desires* [18]. It is not a property in the ordinary sense. However, it conforms to our common sense, so it is expressed in such way. They also pointed out that ConceptNet also has a number of adjective phrases (AP) used to describe properties, and they are often connected with *hasProperty*. Some of the functional properties are achieved by verb phrase (VP). For example, *Movement Forward* is a verb phrase, and it can be linked to *bike* with *isCapableOf*. Liu et al. [16] pointed out that if a property appears on many nodes, and these nodes belong to one parent node, this property can be extracted to the parent node. For example, *fruit* is a parent node for *apple* and *banana*, and *sweet* is a common property for them, so it can extract (*"sweet" PropertyOf "fruit"*). In addition, some adjective phrases which have modifiability can be connected with entities as a property.

The main improvements in ConceptNet compared with WordNet are summarized as follows.

ConceptNet uses an automated approach to build knowledge graph. It includes lots of informal commonsense knowledge accumulated from human experience in the real world. On the node of entities and concepts, it is no longer a single word, but a phrase which can contain certain state information. In relationships, in addition to

Fig. 2 Example of ConceptNet

simple relationships and category relationships, it also adds fact relationships contained in the real world, such as causative relationships, causal relationships, etc. In properties, ConceptNet extends the way to extract properties. In addition, it also contains properties from commonsense.

3.3 Encyclopedic knowledge graph

Encyclopedia knowledge graph is mainly centered on the open knowledge graph supported by LOD [19] project. It mainly includes Wikidata [20], YAGO [21], Google Knowledge Graph [22], Freebase [23], etc. Wikimedia launched Wikidata in October 2012. It links pages which has same theme and allows readers to add or change data entries. Data in Wikidata are basically described by property-value pairs. For some complex information, property-value pairs are allowed to add dependencies property-value pairs. YAGO is a large semantic knowledge base conducted by the Max Planck Institute in Germany, and it has a million entities and more than five million facts [6]. It extracts facts from Wikipedia's classification system and information boxes and combines classification relationships from WordNet [24, 25]. It describes event information in more detail than Wikidata [26]. Google Knowledge Graph [22, 27] was proposed by Google in May 2012. It builds connection between entities and changes the rules of search based on keywords. It generalizes the content of the same topic and describes entities using structured fields. In addition, it clusters entities and properties based on the user's Google retrieval data. Freebase is a semantic web project started by a MetaWebin 2005. Its construction based on Wikipedia and swarm intelligence [4, 6]. It also allows to add or change data entries like Wikipedia. The knowledge representation of encyclopedia knowledge graph is relatively structured. In this section, we chose Freebase to analyze the detailed.

Compared with the previous two knowledge representations which have defined relationships, the most noticeable characteristics of Freebase are it does not control the top-level ontology very strictly [23], and visitors can create and edit the definition of classes and relationships by themselves. It can be more flexible to express different knowledge. Another notable feature of Freebase is that knowledge is stored structurally in base. Freebase is an open, shared, collaboratively built large-scale linked database [4], as well as a practical, extensible, graphical, structured database of general human knowledge. It is inspired by semantic Web research and collaborative data communities such as Wikipedia [28]. Freebase is built by community member's collaboration. Its main data sources include database such as Wikipedia and the contributions of community users. Freebase's knowledge representation framework mainly includes the following elements: Domain, Topic, Type and Property.

Each piece of information in Freebase is called a topic. Topic can be a specific and meaningful data (such as Arnold Schwarzenegger), or an abstract concept (such as PI in mathematics, Christianity) [28]. It corresponds to the node in the graph and contains information, which is unique. Each Topic corresponds to a type (category) node, and Type is equivalent to a classification of Topic. For example, Topic "Yao Ming" can correspond to Type of "Person" and "Athlete," etc. Each Type represents a unique category. However, in order to match the complex information in real life, Type can be given a different name [29, 30]. Type that belongs to the same domain can constitute a Domain. This constitutes the basic structure of Freebase: Domain -> Type -> Topic [30].

As knowledge in Freebase is structured, it uses a lightweight classification system (Type System) [29]. Therefore, it contains relationships and properties which are different from the knowledge representation of WordNet and ConceptNet. Jun [29] believes that the property is the most important concept in Freebase. Property value can be either a literal value or a relationship with other node (such as "is a parent of"). In order to show the structure of Freebase more intuitively, here is an example which is provided by Ruan Yifeng, as shown in Fig. 3 [31]. The core Topic is *Arnold Schwarzenegger*, which corresponds to several types. Though a property is connected to the node of *Arnold Schwarzenegger*, it also corresponds to a property of Type. For example, *Arnold Schwarzenegger* corresponds to Type: *Person*, the property of Person is *country of birth*, its value is *Austria*. Topic: *Arnold Schwarzenegger* and Topic: *Austria* is connected by this property, which is also a relationship between these two topics [31]. Each Type involves different properties. Therefore, Type can be regarded as a property container, which contains the most commonly used properties needed to describe a concept.

Another difference from the above two knowledge representations is that Freebase proposes a new structure to handle multiple relationships: CVT (Compound Value Types). CVT is a node that does not require an explicit name, which is used to express complex data [32]. It can be understood as a table in which multiple relationships and properties are stored, and this table is connected with node. For example, in Fig. 4 [4], CVT describes multiple relationships about Obama's tenure. When you look up Obama's tenure, there is an implicit condition for looking up the length of the tenure. They can be looked up as a whole through the CVT. The multiple relationships contain "office position," "from," "to" [4]. The structure may be more complex without the CVT.

The main differences of Freebase compared with the above two are summarized as the follows:

Freebase contains a larger scope of knowledge. It includes not only common sense and encyclopedic knowledge, but also some knowledge of popular culture, art, location information, etc. In structure, it does not have the strict ontology constraints like the above two, and its metadata are flexible to modify and add, and it can be completed by users more conveniently. In order to reflect users' different opinions and understandings, there may be conflict and contradiction in type and property [28]. It also has CVT, a compound value type for storing complex data that are not found in the above two databases. It uses a more simple structure to display knowledge.

4 Domain knowledge graph

The above are some general knowledge graphs, whose entities and concepts come from the common knowledge in real life. In some areas with strong industry knowledge background, they cannot meet the requirements fully. Therefore, researches that focus on domain knowledge graph emerged. Several typical domain knowledge graphs are introduced in this section.

4.1 Geographic information knowledge graph

GeoNames [33] is a classical knowledge graph in the field of geography, which contains over ten million pieces of geographic information (area name, location, etc.), and it is

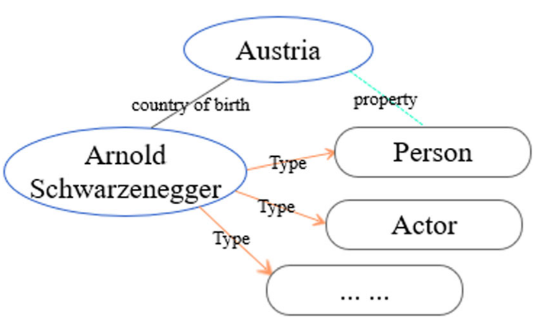

Fig. 3 Example of Freebase

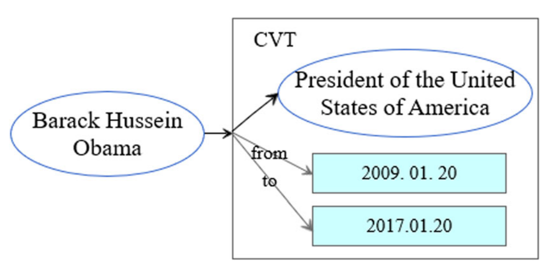

Fig. 4 Example of CVT

mainly displayed in English [34]. GeoName data were collected from the United States Geological Survey, the National Statistical Office, the National Post Office and the U.S. Army [35, 36].

GeoNames divides knowledge which contains nine feature classes, which are subdivided into 645 feature codes [37, 38]. The minimum feature set is the name, coordinates of latitude and longitude, parent regions and countries. It contains population data, aliases and links to Wikipedia, etc. [39]. It treats countries or cities as entities which corresponds to 19 pieces of information each. And some of the information fields are allowed to be empty. These pieces of information can be divided to two sorts. One is property information, such as area, population. These properties are basically geographical information related to the region. And the other is relationship information. For example, an entity can connect to feature classes by feature codes. Level 1 administrative code *admin1 code* and level 2 administrative code *admin2 code* can form hypernym and hyponym relationship [40]. Relationships contained in GeoNames are relatively simple. And these relationships are mainly based on the division of administrative regions, geographical location, attribution, geographic information etc. A simple example of GeoNames is shown in Fig. 5 [33].

The main differences of GeoNames are summarized as follows.

In terms of knowledge scope, the knowledge GeoNames contains is mainly from geographical field. Compared with the previous knowledge graphs, it covers more geographical information. On the structure, the structure of the GeoNames is relatively simple; the entities it contains have fixed and uniform properties, so there is a standard framework of knowledge representation structure. This is quite similar to encyclopedic knowledge graph such as Freebase, and it also supports user to edit data information. Because it relates to geographic information, it links to some specific map. This is not included in the general knowledge graph.

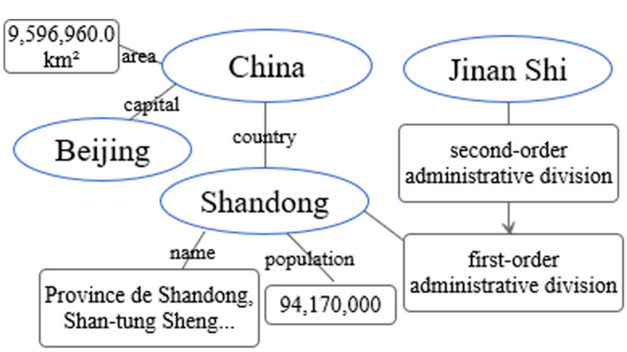

Fig. 5 Example of GeoNames

4.2 Knowledge graph of medical field

In the medical field, the construction of knowledge graphs has also been explored. There are relatively complete medical knowledge bases in this field, such as ICD-11 [41] which uses a tree structure to describe diseases and UMLS [42, 43] which use a structure form to store medical information. And Chinese scholars have also constructed knowledge graphs about traditional Chinese medicine [44–46]. These knowledge graphs contain medical concepts. Their structure is similar to the encyclopedic knowledge graph. In this section, we chose CMeKG [47] which contains more knowledge to analyze its knowledge representation.

CMeKG is mainly composed of concepts of diseases, drugs and diagnostic techniques and their relationships and properties. At present, CMeKG 2.0 [48] has 11,076 diseases, 18,471 drugs, 14,794 symptoms and 3546 treatment techniques, and it includes 1,566,494 triples to describe medical concepts, relationships and properties. It has being updated and improved constantly. Its entities include diseases, symptoms, medicines, etc. By far, the most important entity is disease. It contains treatment options, treatment drugs, diagnostic methods, symptoms, etiology and other properties. The general knowledge graph does not contain such detailed medical knowledge. CMeKG mainly includes the relationships between diseases and other entities, such as related causes, complications, related diseases, etc. In addition, it also includes relationships between symptoms and symptoms, drugs and drugs, etc. Like the general knowledge graph, it provides nodes that link to other knowledge bases. CMeKG is a simple Chinese medical knowledge graph, and its practical application

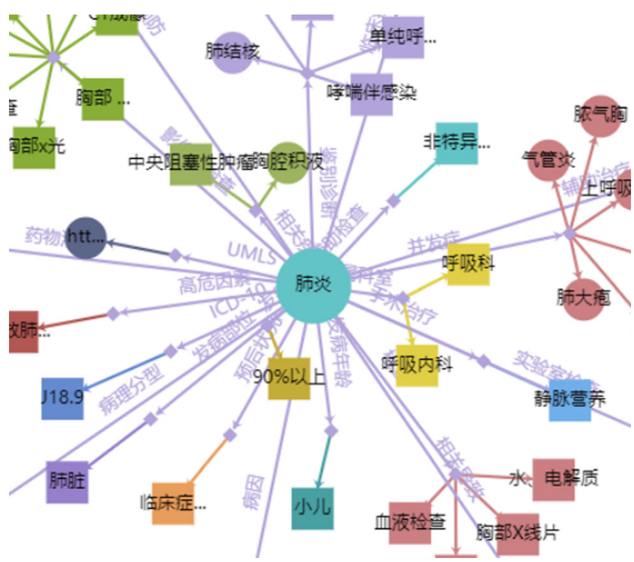

Fig. 6 An interface for CMeKG

needs more exploration. Figure 6 [48] shows an interface for CMeKG.

The main differences of CMeKG compared with the above knowledge graphs are summarized as follows:

In terms of knowledge scope, the knowledge CMeKG contains is mainly from medical field. Compared with the previous knowledge graphs, it covers more detailed medical information. On the structure, its properties and relationships are completely different from other knowledge graphs. Its properties and relationships are based on the medical field.

4.3 Knowledge graph of e-commerce field

The study on knowledge graphs of e-commerce field started earlier. The e-commerce knowledge graph is relatively mature and has been applied in various scenarios. Alibaba built the e-commerce semantic base in 2013 [49], and it includes six subsets, which are basic base, e-commerce base, entertainment base, book base, living base and miscellaneous base. It contains 33 first classes, 10 M entries and 150 relationships. The simple structure is shown in Fig. 7 [49]. This is a simple prototype of the e-commerce knowledge graph.

With the increase in data in e-commerce industry, product knowledge graph of e-commerce field is gradually established [49–52]. The data source includes e-commerce data, Web site information, industry information and encyclopedic information. In product knowledge graph, entities are the products and properties are the related features about this product. Goods belonging to different categories have different properties. For example, the food product has color, smell, shelf life and other properties, while the mobile phone product has accessories, model, battery, screen and other properties. Relationships in product knowledge graph can be roughly summarized as complement (co-buy), co-view, substitute, describe, search and IsA [52]. More broadly, it can be summarized as synonyms relationship, hypernym and hyponym relationship, holistic and partial relationship [51]; it is similar to the language knowledge graph WordNet, but relationships in product knowledge graph are more complex. Most

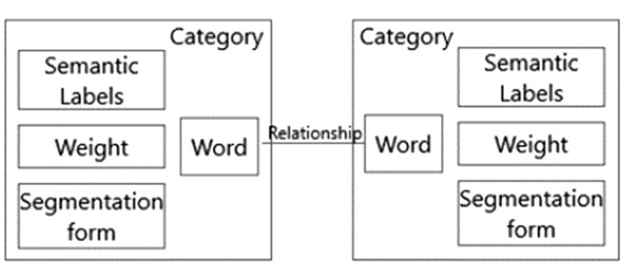

Fig. 7 Structure of e-commerce semantic base

relationships are N to N [52]. Taking mobile phones for example, *battery*, *mobile phone stents*, *audio speaker*, *charger* have *complement* relationship with *Mobile phone*. But at a finer semantic granularity, they correspond to *accessory*, *structural attachment*, *enhancement* and *add-on*. So the semantic meaning of relationships which contained in the product knowledge graph is more complicated.

With the upgrading of application scenarios, the e-commerce cognitive knowledge graph has been gradually constructed [53]. It mainly realized the function of commodity search and personalized recommendation. It includes user knowledge graph, product knowledge graph and scene knowledge graph. Through data fusion and relationships extraction, it links the three knowledge graphs to form the e-commerce cognitive knowledge graph.

Besides basic product properties, product knowledge graph includes some labels, such as *no salt*, *sugar free* or some keywords that users often search. These all are stored in the knowledge graph as properties. These properties are extracted from national regulations and user historical usage records.

The data of the user knowledge graph are derived from account information and historical usage records. In the user knowledge graph, entities are users, relationships between entities are social relationships and its properties are different from other knowledge graphs. The user knowledge graph contains general user information (*name, age, gender,* etc.), which is similar to the general user knowledge graph. Differences mainly lie in the label of the user description. It labels the user by age or some historical search data in e-commerce platform. Those labels are included in the user's properties, such as *old person*, *early pregnancy*. And the purchasing power and preferences of consumers can be inferred from their historical purchase data, and those labels also can be included in the user's properties. These properties are not included in other knowledge graphs.

In addition, the scene knowledge graph is built to connect user knowledge graph and product knowledge graph [53, 54]. Its main data source includes user's search data, product title, hot spots on the network and some industry data. It is a unique knowledge graph in e-commerce field. The scene in scene knowledge graph refers to the conceptualization of user needs, and it is a conceptual node abstracted from user's demand characteristics. The scene knowledge graph uses a short and precise phrase to describe a class of user demands. It takes the implicit user demand information as an entity and creates new nodes which are not included in other knowledge graphs, such as *Outdoor Barbecue, Breakfast for Pregnancy, Keep Warm for kids*. The name of these nodes is from the eight categories in e-commerce concept vocabulary. The eight categories are Time, Location, Object, Function, Incident, Cate/Brand, Style and IP [54]. These scene nodes have certain properties such as color, so the nodes do not have special annotated properties. Relationships in the scene knowledge graph are mainly hypernym and hyponym relationship. Figure 8 [53] is a simple example of e-commerce knowledge graph.

Compared with other knowledge graphs, the differences are summarized as follows.

In terms of knowledge scope, this knowledge graph covers more e-commerce information and contains e-commerce scenario knowledge. On the structure, it enriches the contents of entities and creates new nodes. In addition, it uses a graph as a bridge to establish the connection between the two graphs, which are different from other knowledge graphs.

4.4 Conclusion

Through the research on the knowledge representation of several domain knowledge graphs, it can be seen that the design of knowledge representation in domain knowledge graphs is mainly related to the domain business requirements and the construction of domain knowledge graph can refer to the structure and content of the general knowledge graph to some extent. The main differences between domain knowledge graph and general knowledge graph are breadth, depth and granularities [2]. General knowledge graph has a wider breadth, which covers more knowledge.

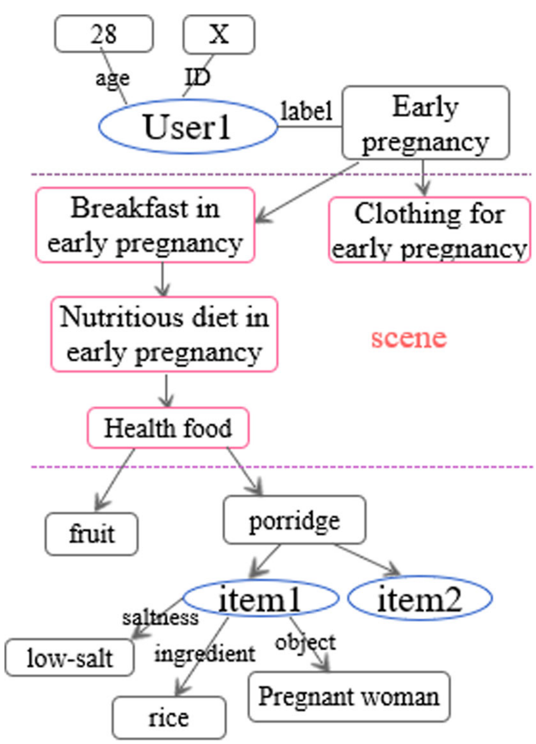

Fig. 8 Example of e-commerce knowledge graph

Domain knowledge graph which shows more detailed knowledge within the field has a deeper depth. In addition, there are some differences in the granularity of their knowledge partitioning. The main reason for the differences lies in their different knowledge backgrounds and data sources. Domain knowledge graph has more domain-specific data and knowledge, so its entities and properties may be quite different from general knowledge graph.

5 Research trends and prospects

There have been lots of researches which focus on general knowledge graph, and some researches have focused on knowledge representation learning, such as distance model SE [55], translation model TransE [56]. Liu et al. [57] summarized the method of knowledge representation learning. Research on domain knowledge graph is emerging. Knowledge representation is the first step for knowledge graph construction. Learning it helps beginners to understand the concept of knowledge graph and lays a foundation for domain knowledge graph construction.

For the research on domain knowledge graph construction, some problems should be explored.

(1) Expansion of knowledge representation

The main way to express knowledge is relational triple, whether it can be extended to multicomponent to express diverse information. For complex unstructured problems, such as the corresponding relationship between major and school in the education industry, a simple inclusion relationship cannot express it fully, whether properties and relationships can be extended?

(2) Multimode of knowledge representation

There are a lot of information resources on the network. They not only contain text, also include video and image, etc. Video and image maybe explain knowledge better than text. Therefore, how to design these nontextual resources into the structure of knowledge representation is an important problem.

(3) Knowledge representation automatic learning

Most of the knowledge representation learning methods are applicable to general knowledge graph. It is not well qualified for automatic extraction of domain knowledge with complex information. Therefore, domain knowledge automatic learning is a problem to be solved.

(4) Knowledge fusion

Compared with general knowledge, the structure of domain knowledge is more complex and it requires more data experimentation and better algorithms if we want to integrate knowledge into industry background.

(5) Data collection

The establishment of knowledge graph requires lots of data. In some specific area, such as education, a complete education knowledge graph can assist teachers to do course designing and help students collect information. There are enough data to support education knowledge graph construction. However, data collection in the education industry is still not completed. So data collecting and analyzing platform is going to be a research trend of domain knowledge graph construction.

(6) Dynamic update

The knowledge contained in the domain knowledge graph is not unchanged all the time, such as major courses every year, the grade of school or major in the education field. Therefore, how to realize the dynamic change and update of knowledge in a quick way is an important research area.

Clarifying the concept and content of knowledge representation by sorting out the development of knowledge representation is expected, which will put forward the knowledge representation about higher education in the next step and will lay a foundation for the construction of education knowledge graph as well.

Acknowledgements Teaching Reform Research Project of Undergraduate Colleges and Universities of Shandong Province (Z2016Z036), the Teaching Reform Research Project of Shandong University of Finance and Economics (jy2018062891470, jy201830, jy201810), Shandong Provincial Social Science Planning Research Project (18CHLJ08), Scientific Research Projects of Universities in Shandong Province (J18RA136), Youth Innovative on Science and Technology Project of Shandong Province (2019RWF013), SDUST Excellent Teaching Team Construction Plan (JXTD20160512 and JXTD20180510), Jinan campus of SDUST Excellent Teaching Team Construction Plan (JNJXTD201711), Teaching research project of Shandong University of Science and Technology (JNJG2017104), National Natural Science Foundation of China (61703243).

Compliance with ethical standards

Conflict of interest There are no conflicts of interests of this work.

References

1. Jue W, Xiaohong Y, Chunyi S, Jigang H (1995) Discussions on knowledge representation. Chin J Comput 18(3):212–224
2. Yanghua X (2018) Problems and measures in the implementation of domain knowledge graph. https://blog.csdn.net/xinshucredit/article/details/84852877. Accessed 14 Apr 2020
3. Shan F, Ailin L, Shu Y (2019) The concept and application of knowledge graph. Inf Commun Technol Policy 299(05):17–20
4. Development Report of Knowledge Graph in 2018 (2018) Chinese Information Processing Society of China
5. Xu/Shi/Quan et al (2017) The most comprehensive review of knowledge graph: concept and construction technology (TUG). https://mp.weixin.qq.com/s/bhk6iZdphif74HJlyUZOBQ. Accessed 05 Jan 2020
6. Jun Z (2018) Knowledge graph. Higher Education Press, Beijing
7. Miller GA (1995) WordNet: a lexical database for English. Commun Assoc Comput Mach 38(11):39–41. https://doi.org/10.1145/219717.219748
8. Miller GA, Beckwith R, Fellbaum C et al (1990) Introduction to WordNet: an on-line lexical database*. Int J Lexicogr 3(4):235–244. https://doi.org/10.1093/ijl/3.4.235
9. Tianshun Y, Li Z, Zhu G (2001) Introduction of WordNet. Appl Linguist 1:27–32
10. Lenat DB (1995) CYC: a large-scale investment in knowledge infrastructure. Commun ACM 38(11):32–38. https://doi.org/10.1145/219717.219745
11. Speer R, Havasi C (2012) Representing general relational knowledge in ConceptNet 5. In: LREC, pp 3679–3686
12. Zhao J, Sun N (2020) Government subsidies-based profits distribution pattern analysis in closed-loop supply chain using game theory. Neural Comput Appl 32:1715–1724
13. Gupta P, Sharma TK, Mehrotra D et al (2019) Knowledge building through optimized classification rule set generation using genetic based elitist multi objective approach. Neural Comput Appl 31:845–855
14. Xu Z, Zhang H, Hu C, Mei L, Xuan J, Choo K-KR, Sugumaran V, Zhu Y (2016) Building knowledge base of urban emergency events based on crowd sourcing of social media. Concurr Comput Pract Exp 28(15):4038–4052
15. Najmi E, Hashmi K, Malik Z, et al (2014) ConceptOnto: an upper ontology based on ConceptNet. In: IEEE/ACS international conference on computer systems & applications. https://doi.org/10.1109/AICCSA.2014.7073222
16. Liu H, Singh P (2004) ConceptNet: a practical commonsense reasoning tool-kit. BT Technol J 22(4):211–226. https://doi.org/10.1023/B:BTTJ.0000047600.45421.6d
17. Commonsense Computing Initiative (2020) ConceptNet. http://www.conceptnet.io/. Accessed 22 Dec 2019
18. Speer R, Havasi C (2013) ConceptNet 5: a large semantic network for relational knowledge. In: Gurevych I, Kim J (eds) The people's web meets NLP. Theory and applications of natural language processing. Springer, Berlin. https://doi.org/10.1007/978-3-642-35085-6_6
19. Europeana (2020) Linked open data. https://pro.europeana.eu/page/linked-open-data. Accessed 20 Mar 2020
20. Vrandecic D, Krtoetzsch M (2014) Wikidata: a free collaborative knowledge base. Commun ACM 57(10):78–85. https://doi.org/10.1145/2629489
21. F.M. Suchanek, G. Kasneci, G. Weikum (2007) YAGO: a core of semantic knowledge. In: Proceedings of the 16th international conference on World Wide Web, Banff, Canada, pp 697–706. https://doi.org/10.1145/1242572.1242667
22. Singhal A (2012) Introducing the knowledge graph: things, not strings. https://googleblog.blogspot.com/2012/05/introducing-knowledge-graph-things-not.html. Accessed 26 Mar 2020
23. Bollacker KD, Evans C, Paritosh P, et al (2008) Freebase: a collaboratively created graph database for structuring human knowledge. In: Proceedings of the ACM SIGMOD international conference on management of data, SIGMOD 2008, Vancouver, BC, Canada, 10–12 June, ACM. https://doi.org/10.1145/1376616.1376746
24. Hoffart J, Suchanek FM, Berberich K et al (2013) YAGO2: a spatially and temporally enhanced knowledge base from Wikipedia. Artif Intell 194:28–61. https://doi.org/10.1016/j.artint.2012.06.001

25. Suchanek FM, Kasneci G, Weikum AG (2008) Yago: a large ontology from Wikipedia and WordNet. J Web Semant 6(3):203–217. https://doi.org/10.1016/j.websem.2008.06.001
26. Ringler D, Paulheim H (2017) One knowledge graph to rule them all? Analyzing the differences between DBpedia, YAGO, Wikidata & Co. In: Kern-Isberner G, Fürnkranz J, Thimm M (eds) KI 2017: advances in artificial intelligence. KI 2017. Lecture notes in computer science, vol 10505. Springer, Cham. https://doi.org/10.1007/978-3-319-67190-1_33
27. Uyar A, Aliyu FM (2015) Evaluating search features of Google knowledge graph and bing satori entity types, list searches and query interfaces. Online Inf Rev 39(2):197–213. https://doi.org/10.1108/OIR-10-2014-0257
28. Bollacker K, Cook R, Tufts P (2007) Freebase: a shared database of structured general human knowledge. In: Proceedings of the twenty-second AAAI conference on artificial intelligence, 22–26 July 2007, Vancouver, British Columbia, Canada. https://www.aaai.org/Papers/AAAI/2007/AAAI07-355.pdf
29. Jun L (2011) Research on semantic database freebase. New Technol Libr Inf Serv 27(10):18–23
30. Bin W, Jial F (2013) Further study in freebase with property coordinate system theory. Microcomput Appl. https://doi.org/10.1145/176789.176804
31. Yifeng R (2008) Research on freebase (TUG), http://www.ruanyifeng.com/blog/2008/04/freebase_reloaded.html. Accessed 26 Dec 2019
32. Danzhuzhu (2013) Introduction to freebase (TUG). http://blog.maidou.info/?p=169. Accessed 26 Dec 2019
33. Creative Commons (2020) GeoName. http://www.geonames.org/. Accessed 30 Mar 2020
34. Bukun (2016) Introduction to the global geographic names database, GeoName. https://www.osgeo.cn/post/04b2. Accessed 30 Mar 2020
35. Wei Yong H, Danlu LX, Fei W (2016) Geographic name full text query based GeoNames and Solr. Eng Surv Mapp 25(2):28–32
36. GeoNames (2020) Datasources used by GeoNames in the GeoNames Gazettee. http://www.geonames.org/data-sources.html. Accessed 30 Mar 2020
37. Uche Ogbuji (2008) Open geographic information systems at Geonames.org. https://www.ibm.com/developerworks/cn/web/wa-realweb7/. Accessed 30 Mar 2020
38. Maltese V, Farazi F (2013) A semantic schema for GeoNames. In: INSPIRE 2013
39. Ahlers D (2013) Assessment of the accuracy of GeoNames gazetteer data. Workshop on geographic information retrieval. https://doi.org/10.1145/2533888.2533938
40. Yang YJ, Xu B, Hu JW, Tong MH, Zhang P, Zheng L (2018) Accurate and efficient method for constructing domain knowledge graph. J Softw 29(10):2931–2947. https://doi.org/10.13328/j.cnki.jos.005552
41. World Health Organization (2018) ICD-11 for mortality and morbidity statistics (ICD-11 MMS) 2018 version. https://icd.who.int/browse11/l-m/en. Accessed 1 Apr 2020
42. Zhou S, Luo P, Jain DK, Lan X, Zhang Y (2019) Double-domain imaging and adaption for person re-identification. IEEE Access 7:103336–103345
43. NML (2019) Unified medical language system. https://www.nlm.nih.gov/research/umls/index.html. Accessed 01 Apr 2020
44. Lirong J, Jing L, Tong Y et al (2015) Construction of traditional Chinese medicine knowledge graph. J Med Intell 2015(8):51–53
45. Tong R, Chenglin S, Haofen W et al (2016) Construction of traditional Chinese medicine knowledge graph and its application. J Med Inform 37(4):8–14
46. Dezheng Z, Yonghong X, Man L, Chuan S (2017) Construction of knowledge graph of traditional Chinese medicine based on the ontology. Technol Intell Eng 3(1):035–042. https://doi.org/10.3772/j.issn.95-915x.2017.01.004
47. Odmaa BYAMBASUREN, Yunfei YANG, Zhifang SUI, Damai DAI, Baobao CHANG, Sujian LI, Hongying ZAN (2019) Preliminary study on the construction of Chinese medical knowledge graph. J Chin Inf Process 33(10):1–7. https://doi.org/10.3390/info11040186
48. The Institute of Computational Linguistics, Peking University, et al (2018) CMeKG2.0. http://zstp.pcl.ac.cn:8002/. Accessed 02 Apr 2020
49. Ali (2013) E-commerce semantic base. https://wenku.baidu.com/view/1f20d94a43323968001c9254.html. Accessed 05 Apr 2020
50. Zhou S, Ke M, Luo P (2019) Multi-camera transfer GAN for person re-identification. J Vis Commun Image Represent 59:393–400
51. Shengchun D, Linlin H, Ying W (2019) Product knowledge map construction based on the e-commerce data. Data Anal Knowl Discov. https://doi.org/10.1142/S021800141951008X
52. Da X, Chuanwei R, Korpeoglu E, Kumar S, Achan K (2019) Product knowledge graph embedding for e-commerce. arXiv:1911.12481
53. Ali Technology (2018) The knowledge graph born for e-commerce, how to respond to user demand (TUG). https://mp.weixin.qq.com/s/RoiIHPCV3vXAH_sUHhTOYw. Accessed 29 Dec 2019
54. Xusheng L, Yonghua Y, Zhu KQ, Yu G, Keping Y (2018) Conceptualize and infer user needs in e-commerce. In: The 28th ACM international conference. https://doi.org/10.1145/3357384.3357812
55. Bordes A, Weston J, Collobert R, et al (2011) Learning structured embeddings of knowledge bases. In: Proceedings of AAAI. AAAI, Menlo Park, pp 301–306
56. Bordes A, Usunier N, Garcia-Duran A, et al (2013) Translating embeddings for modeling multirelational data. In: Proceedings of NIPS. MIT Press, Cambridge, pp 2787–2795
57. Zhiyuan L, Maosong S, Yankai L, Ruobing X (2016) Knowledge representation learning: a review. J Comput Res Dev 53(2):247–261