Entity Summarization in Fuzzy Knowledge Graph Based on Fuzzy Concept Analysis

Erhe Yang, Fei Hao, Jie Gao, and Doo-Soon Park

Abstract Considerable attention has recently been devoted to Knowledge Graph (KG), which has been applied in many domains. However, the information is often imprecise and vague when constructing the knowledge graph and thus the Fuzzy Knowledge Graph (FKG) emerged. Considering the increasing data in FKG, this paper firstly formulates the entity summarization in FKG and proposes an approach leveraging Fuzzy Formal Concept Analysis (FFCA). More specifically, the predicates and objects in RDF triples are deemed as attributes and objects in FFCA, respectively. Then, the fuzzy formal context can be obtained and the fuzzy concept lattice can be constructed. Finally, the concepts are ranked by the cardinality of the extent in concept lattice and the vague value of objects in RDF triples.

Keywords Entity Summarization · Fuzzy Knowledge Graph · Fuzzy Formal Concept Analysis

This research was supported by the National Natural Science Foundation of China (Grant No. 61702317), and the National Research Foundation of Korea (NRF-2020R1A2B5B01002134) and was also supported by the Natural Science Basic Research Plan in Shaanxi Province of China (2019JM-379). This is also a part of a project that has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 840922. This work reflects only the authors' view and the EU Commission is not responsible for any use that may be made of the information it contains.

E. Yang · F. Hao (✉) · J. Gao
School of Computer Science, Shaanxi Normal University, Xi'an, China
e-mail: feehao@gmail.com

F. Hao
Department of Computer Science, University of Exeter, Exeter, UK

D.-S. Park (✉)
Department of Computer Software Engineering, Soonchunhyang University, Asan, South Korea
e-mail: parkds@sch.ac.kr

J. J. Park et al. (eds.), *Advanced Multimedia and Ubiquitous Engineering*, Lecture Notes in Electrical Engineering 716, https://doi.org/10.1007/978-981-15-9309-3_3

1 Introduction

The term Knowledge Graph (KG) was coined by Google in 2012 and KG has been successfully applied in many fields, such as entity summarization [1], entity alignment [2], etc. KG describes entities and their relationships by employing the Resource Description Framework (RDF)-style triples. RDF datasets are collections of subject-predicate-object triples to describe the resources and their relationships on the web. However, the relationships between entities and their properties are often imprecise and uncertain. Besides, the descriptions of objects are accompanied by a degree of uncertainty. Thus, it is necessary to build the Fuzzy Knowledge Graph (FKG). The literature [3] attempts to construct a fuzzy knowledge base from a set of documents. Xiong et al. [4] proposed a general framework to construct fuzzy knowledge bases using feature selection. This framework can reduce the computational costs of fuzzy knowledge base construction. Nevertheless, users may be overwhelmed by the massive entity descriptions. It is hard for the end-users to search for the most important information from the lengthy FKG. As the information comes from various sources, it is often vague or ambiguous. The constructed FKG is labeled with vague value, expressing that the confidence level of the information.

Example 1 Figure 1 shows a toy example of the entity in a fuzzy knowledge graph. The entity is selected from the datasets in [5]. We assume that the objects of entity are ambiguous, thus we can add the possible values as shown in the brackets. The value within the scope of [0, 1] in the bracket represents the confidence level of the objects. People are often confused with the FKG, due to the excessive volume of descriptions of the entity. Thereby, it is essential to tackle this issue.

Nevertheless, existing works only focus on the entity summarization in KG rather than in FKG. Therefore, this paper firstly formalizes the problem of entity summarization in FKG. Formal concept analysis (FCA) is a useful mathematical theory that modeling the generalization and generalization relation between objects and attributes. As its generalization, Fuzzy Formal Concept Analysis (FFCA) [6] aims

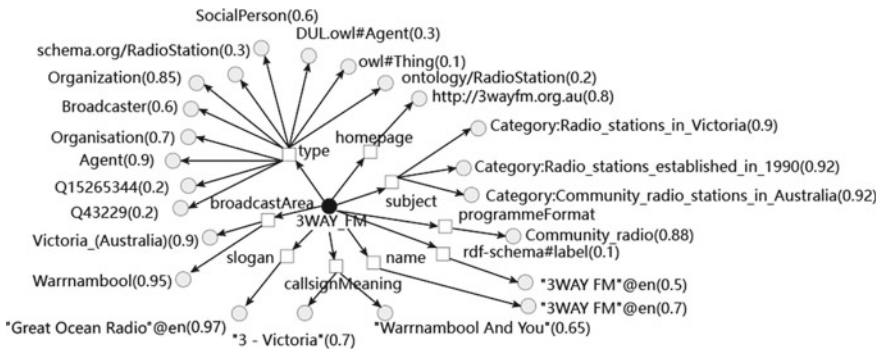


Fig. 1 The entity "3WAY_FM" in fuzzy knowledge graph

at processing fuzzy sets and representing conceptual knowledge and hierarchies. To address the novel problem of entity summarization in FKG, we present an entity summarization approach based on FFCA. Specifically, the predicates and objects of the entities in FKG are converted into attributes and objects in fuzzy formal context. Then, the fuzzy formal concepts are obtained and ranked by the weight of concepts and the vague degree of objects in RDF triples. The weight of concepts is based on the cardinality of objects in extent of concepts, which is inspired by [5]. Finally, the ranked RDF triples whose objects labeled with vague value can be obtained.

The contribution of this paper is twofold:

- Problem Formalization of Entity Summarization in FKG: We pioneer the formalization of entity summarization in FKG, which aims to select the most important and compact information of the entity in FKG. This paper focuses on the situation where the object of triples labels with vague value. The goal of this paper is to obtain the top-k RDF triples from the lengthy entity descriptions.
- FFCA-based Entity Summarization Approach in FKG: The central idea of the proposed approach is that the less the cardinality of objects in a fuzzy concept is, the more important the fuzzy concept is. Concretely, given a fuzzy knowledge graph R, we construct the fuzzy formal context K in which the tokenized objects and the predicates in fuzzy RDF triples are considered as objects and attributes while the vague degree of objects serves as the value of K. Then, the fuzzy formal concept lattice L can be built and the fuzzy concepts can be ranked by the cardinality of extent in concepts and the vague degree of objects in RDF triples. Finally, the top-k triples can be obtained.

2 Related Work

For the research on entity summarization in KG, RELIN [7] employs a variant of the random surfer model to obtain the summarized triples. Gunaratna et al. [1] presented an entity summarization approach FACES that incorporating diversity, uniqueness, and popularity of the facts. Their approach utilizes the clustering algorithm Cobweb to select the representative facts. In the literature [8], the problem of context-aware entity summarization is formalized and a Personalized Page Rank algorithm is adopted by considering user preference. With respect to the summarization approach using Formal Concept Analysis, Kim et al. [5] proposed an entity summarization approach KAFCA based on FCA. KAFCA regards the objects and predicates of entities as objects and attributes in formal context, respectively. Then it ranked the RDF triples according to the weights of extents of concepts in concept lattice. The work [9] presents a query-focused twitter summarization framework for obtaining the most representative tweets for users and summarizes them based on FFCA. However, there is no existing study focusing on entity summarization for the fuzzy knowledge graph. Hence, we present a solution for this novel problem.

3 Proposed Approach

This section firstly provides the definitions about Fuzzy Knowledge Graph and Fuzzy Concept Analysis in Sect. 3.1. Then the problem formulation is given in Sect. 3.2. Finally, Sect. 3.3 details the proposed approach.

3.1 Fuzzy Knowledge Graph

Definition 1 (Fuzzy RDF data graph) [9]. A fuzzy RDF data graph G is represented by a 6-tuple (V, E, Σ, L, μ, ρ), where V is a finite set of vertices, $E \subset V \times V$ is a set of directed edges, Σ is a set of labels, $L: V \cup E \rightarrow \Sigma$ is a function assigning labels to vertices and edges, $\mu: V \rightarrow [0, 1]$ is a fuzzy subset, and $\rho: E \rightarrow [0, 1]$ is a fuzzy relation on the fuzzy subset μ. Note that $\rho(v_i, v_j) \leq \mu(v_i) \wedge \mu(v_j)$, where $v_i, v_j \in V$.

Definition 2 (Fuzzy Formal Context) [10]. A Fuzzy Formal Context is a triple $K = (O, A, R = \varphi(O \times A))$, where O is a set of objects, A is a set of attributes, and R is a fuzzy relation in $O \times A$. Each pair $(o, a) \in R$ has a membership value $\mu(o, a)$ in [0,1].

Definition 3 (Fuzzy Formal Concept) [10]. Given a fuzzy context $K = (O, A, R = \varphi(O \times A))$, a confidence threshold T, and two sets E, I, such that $E \subseteq O$ and $I \subseteq A$, consider the dual sets E‘ and I‘, defined, respectively as follows:

$E' = \{a \in A | \mu(o, a) \geq T \; \forall o \in E\}$.

$I' = \{o \in O | \mu(o, a) \geq T \; \forall a \in I\}$.

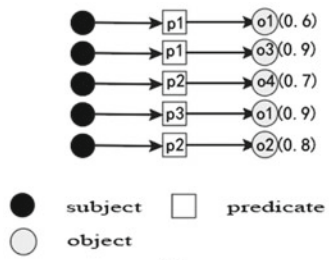
The sets E and I are referred to as the extent and the intent of the fuzzy concept, respectively.

3.2 Problem Description

Considering the increasing scale of data in FKG, it is necessary to provide a summary of the numerous descriptions of the entity. Entities in FKG are represented by RDF triples labeled with vague values. Entity summarization in FKG aims at providing a compact summary from lengthy entity descriptions. Given a fuzzy knowledge graph R, the entity summarization targets at selecting top-k RDF triples.

3.3 Proposed Approach

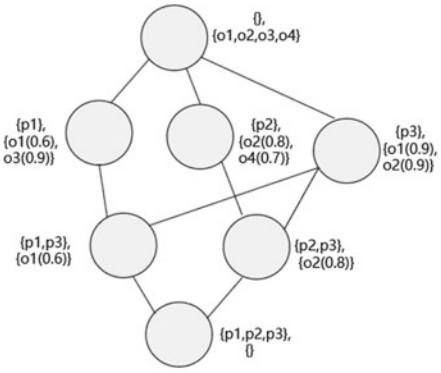
Inspired by the approach in [5], we proposed a solution to tackle the problem of entity summarization in FFG. More specifically, as shown in Fig. 2a is the RDF



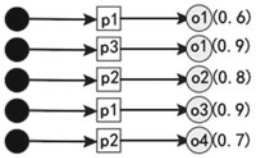
(a)

	predicate 1	predicate 2	predicate 3
tokenized object 1	0.6		0.9
tokenized object 2		0.8	0.9
tokenized object 3	0.9		
tokenized object 4		0.7	

(b)



(c)



(d)

Fig. 2 The framework of our proposed approach

triple whose objects labeled with vague values. Figure 2b is the fuzzy formal context and the corresponding fuzzy concept lattice is shown in Fig. 2c. The output RDF triples are shown in Fig. 2d. To be more precise, the ranking principle is that we firstly rank the triples according the cardinality of the extent in concepts and then by the vague values of objects when the cardinalities of objects are same. In practice, the results can be pruned according to the vague value. For example, we can remove the triples with vague value less than 0.7. Therefore, some less believable information can be trimmed (Fig. 3).

Adopting the proposed approach for the entity in Fig. 1, the following summary results can be obtained.

4 Conclusion

This paper presents an approach of entity summarization in FKG employing FFCA. Concretely, we firstly convert the predicates and objects of entity in RDF triples into attributes and objects in FFCA, respectively. Then the fuzzy formal context

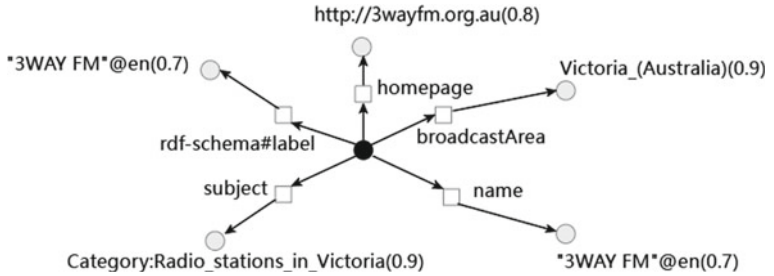


Fig. 3 The corresponding summarization of the entity "3WAY_FM" in fuzzy knowledge graph

can be obtained and the fuzzy concepts can be built. Finally, we rank the fuzzy concepts by the hierarchy of concepts in fuzzy concept lattice and the vague value of objects in RDF triples. In the future, comprehensive experiments will be conducted for evaluating the performance of the proposed approach.

References

1. Gunaratna K, Thirunarayan K, Sheth A (2015) FACES: diversity-aware entity summarization using incremental hierarchical conceptual clustering. In: Twenty-Ninth AAAI conference on artificial intelligence
2. He F, Li Z, Qiang Y, Liu A, Liu G, Zhao P, Chen Z (2019) Unsupervised entity alignment using attribute triples and relation triples. In: International conference on database systems for advanced applications. Springer, Cham, pp 367–382
3. Yarushkina N, Moshkin V, Filippov A, Guskov G (2018) Developing a fuzzy knowledge base and filling it with knowledge extracted from various documents. In: International conference on artificial intelligence and soft computing. Springer, Cham, pp 799–810
4. Xiong N, Funk P (2005) Construction of fuzzy knowledge bases incorporating feature selection. Soft Comput 10(9):796–804
5. Kim EK, Choi KS (2018) Entity summarization based on formal concept analysis. In: 1st international workshop on entity retrieval, EYRE
6. Tho QT, Hui SC, Fong ACM, Cao TH (2006) Automatic fuzzy ontology generation for semantic web. IEEE Trans Knowl Data Eng 18(6):842–856
7. Cheng G, Tran T, Qu Y (2011) Relin: relatedness and informativeness-based centrality for entity summarization. In: International semantic web conference. Springer, Berlin, Heidelberg, pp 114–129
8. De Maio C, Fenza G, Loia V, Parente M (2015) Online query-focused twitter summarizer through fuzzy lattice. In: 2015 IEEE International conference on fuzzy systems (FUZZ-IEEE). IEEE, pp 1–8
9. Li G, Yan L, Ma Z (2019) An approach for approximate subgraph matching in fuzzy RDF graph. Fuzzy Sets Syst 376:106–126
10. Formica A (2012) Semantic web search based on rough sets and fuzzy formal concept analysis. Knowl Based Syst 26:40–47