

Finding Associative Entities in Knowledge Graph by Incorporating User Behaviors

Jiangu Li

School of Information Science and Engineering, Yunnan University, Kunming, China & Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China

Peizhong Yang

School of Information Science and Engineering, Yunnan University, Kunming, China & Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China

Kun Yue

School of Information Science and Engineering, Yunnan University, Kunming, China & Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China

Liang Duan

School of Information Science and Engineering, Yunnan University, Kunming, China & Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China

Zehao Huang

School of Information Science and Engineering, Yunnan University, Kunming, China & Yunnan Key Laboratory of Intelligent Systems and Computing, Yunnan University, Kunming, China

ABSTRACT

The task of finding associative entities in knowledge graph (KG) is to provide a ranking list of entities according to their association degrees. However, many entities are not only linked in KG but also associated in terms of user behaviors, which facilitates finding associative entities accurately. This manuscript incorporates KG with user-generated data to propose the Association Entity Graph Model (AEGM) to evaluate the association degrees. They first propose the joint weighting function to evaluate the entity associations and prove its submodularity theoretically as well as the greedy algorithm to select the candidates efficiently. They define the entity association information to score the entity association and give the hill climbing search based algorithm for AEGM construction. Following, they embed AEGM to calculate the association degrees and obtain the associative entities efficiently. Extensive experiments on three datasets show that the proposed method can achieve a better performance than some state-of-the-art competitors in accurately finding associative entities.

KEYWORDS

Knowledge Graph, User Behavior, User Generated Data, Associative Entity, Submodular Set-Function

INTRODUCTION

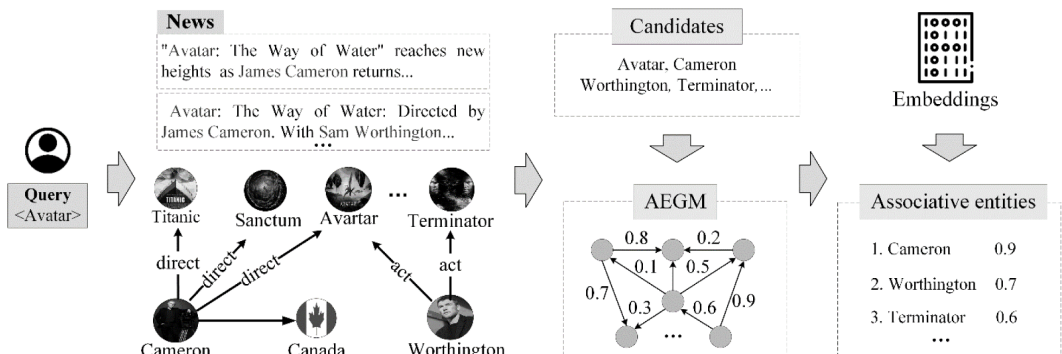
A knowledge graph (KG), such as *Freebase* (Tanon et al., 2016), *Wikipedia* (Arvola & Alamettälä, 2022), or *DBpedia* (Lehmann et al., 2015), provides structured representations of facts about real-world entities and relations (Yan et al., 2024). Given one or more query entities of a KG, the task of finding associative entities revolves around obtaining a ranked list of associative entities by calculating their association degrees. This task has gained much attention (Zhou et al., 2020) and is indispensable in

DOI: 10.4018/JDM.371751

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

search engines (Yang et al., 2019) and recommendation systems (Yang et al., 2020). Figure 1 shows the overview of the association entity graph model (AEGM).

Figure 1. Overview of the association entity graph model (AEGM)



Note. AEGM = association entity graph model.

Accurately finding associative entities is challenging due to the ambiguities in queries (Gu et al., 2019) and the inability to effectively narrow down the scope of candidate answers without information beyond the KG and query entities. For example, given the query entity *Avatar*, a user may search for the actors starring in *Avatar*, similar movies, or other associative entities from a KG, as shown in Figure 1. To resolve this ambiguity, the query-by-example method aims to return the associative entities by providing a complete example as the element that users are interested in (Ding et al., 2023). However, the associations among entities are rapidly evolving in the real world. Entities are not only linked by their relations in a KG but are also implicitly associated with user-entity interactions on online platforms, such as social media (Muppasani et al., 2024), wikis (Raganato et al., 2016), and ratings (Liu et al., 2023). In this paper, the collection of user-entity interactions is collectively referred to as user-generated data (UGD). In the context of KG applications, the associative entities may be found with limited precision and coverage by only considering the KG. As shown in Figure 1, compared to the *Titanic*, the term “Cameron” was more likely to be the ideal answer for *Avatar*, due to the frequent co-occurrence of Cameron and *Avatar* in recent news. That is, the users’ query intents could not only be inferred by a KG, but were often also hidden within users’ behaviors. In this paper, we leverage the UGD to improve the accuracy of finding associative entities by making use of the timeliness and credibility of UGD.

Intuitively, a KG and UGD involve massive entities and noisy issues (Bu et al., 2021; Wu et al., 2022), but only a portion of entities are strongly associated with each other. It is necessary to select candidates from a KG and UGD as the basis for finding the associative entities. Previous works extracted candidates w.r.t. the query entities from a KG, then calculated their association degrees from the UGD (J. Li et al., 2021a). These methods may have yielded inaccurate candidates because a KG can be incomplete, and some entities that are strongly associated with each other in the UGD may not be linked in the KG (J. Li et al., 2021b). This discrepancy may have led to the omission of association candidates during the extraction process.

Moreover, the probability of user behaviors on different entities typically reflected the strength of implicit associations between entities. Particularly, the associations between entities were directed. For example, the probability that a user likes *Avatar* and also likes the *Terminator* was 80%. Conversely, the probability that a user likes the *Terminator* and simultaneously likes *Avatar* was only 60%. This motivates us to assign weights to the directed associations among entities to find the answers accurately.

22 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/finding-associative-entities-in-knowledge-graph-by-incorporating-user-behaviors/371751

Related Content

Using Ontology Languages for Conceptual Modeling

Palash Bera, Anna Krasnoperova and Yair Wand (2012). *Cross-Disciplinary Models and Applications of Database Management: Advancing Approaches* (pp. 1-28).

www.irma-international.org/chapter/using-ontology-languages-conceptual-modeling/63660

Different Kinds of Hierarchies in Multidimensional Models

Elzbieta Malinowski (2009). *Handbook of Research on Innovations in Database Technologies and Applications: Current and Future Trends* (pp. 45-55).

www.irma-international.org/chapter/different-kinds-hierarchies-multidimensional-models/20687

FOOM - Functional and Object-Oriented Analysis and Design of Information Systems: An Integrated Methodology

Peretz Shoval and Judith Kabeli (2001). *Journal of Database Management* (pp. 15-25).

www.irma-international.org/article/foom-functional-object-oriented-analysis/3258

The Contribution of IT Governance Solutions to the Implementation of Data Warehouse Practice

Sutee Sujitparapitaya, Brian D. Janz and Mark Gillenson (2003). *Journal of Database Management* (pp. 52-69).

www.irma-international.org/article/contribution-governance-solutions-implementation-data/3295

Text-Image Retrieval With Salient Features

Xia Feng, Zhiyi Hu, Caihua Liu, W. H. Ip and Huiying Chen (2021). *Journal of Database Management* (pp. 1-13).

www.irma-international.org/article/text-image-retrieval-with-salient-features/289790