LING CAI
STYKO Lab
Department of Geography
University of California, Santa Barbara
Email: ling.cai@geog.ucsb.edu

Ling Cai, a second-year Ph.D. student at STKO Lab, Department of Geography, University of California, Santa Barbara. Her advisor is Professor Krzysztof Janowicz. She has a very rich range of interests, including temporal knowledge graphs, semantic web, recommender systems, urban computing, etc. Currently, Ling has immersed herself in developing temporally-explicit machine learning/deep learning models for downstream tasks in knowledge graphs, for instance, link prediction, conjunctive querying and reasoning. She interned in Siemens (Berkeley) summer 2019.

Does Time Matter in Reasoning over Knowledge Graphs?

Knowledge graphs (KGs), also known as knowledge bases, are widely used for storing, organizing and retrieving structured data. Typically knowledge graphs are labeled and directed multigraphs, consisting of abundant statements about the world that are in the form of triples (subject–>predicate– >object). These subjects and objects are nodes in a graph, representing people, locations, events, etc., and these predicates are relation types, conveying the relationships between two linked nodes. Due to their board coverage and highly organization, they have supported a rich range of applications such as question answering, recommender systems, conversational dialogue systems, and so on.

Albeit these successful applications, most of these studies underestimate the important role of time in these studies. Take the conjunctive query task as an example. If you look for a person who satisfies the following two statements: $(?Person- > holdPosition- > PresidentOfUSA) \wedge (?Person- > workLocation- > NewYorkCity)$, then what would be the answer? In fact, the answer varies from the concrete time you queried. Two possible scenarios would be as follows. If you queried "Find the person who **is** the president of the United States **now** and **worked** in New York City **before**?" then the answer is "Donald Trump," as the presidency of Trump started from January 2017 and he worked in New York City from 1971 to 2017. But what if you looked for a person "who **is** the president of the United States and **is working** in New York City?" The answer is "None," because Trump is the president now but he is working in White House. Without considering the validity time of statements, i.e., temporal scoping of statements, the results of queries/reasoning over knowledge graphs would be invalid. Moreover, as the world is evolving all the time, knowledge graphs are also changing rapidly. It is of great significance to study the evolving relationships among entities, for instance, the occurrence of consecutive or concurring events, from a dynamic perspective.

A variety of approaches have been proposed to successfully represent temporal information in knowledge graphs, including Temporal Descriptive Logic, Temporal Resource Descriptive Framework, Reification, Versioning, Named Graphs, etc.[1]. However, to some degree these different approaches limit the OWL reasoning capabilities, which is the key component of KGbased applications. For instance, reification, which treats the relation in a temporal statement as the object of a property, renders the OWL semantics over properties inapplicable in practice[2]. This calls for new methods that are able to support explicitly/implicitly temporal reasoning over knowledge graphs.

Recently, machine learning models, as one type of induction learning, are widely utilized in multiple KG-based downstream tasks[3], for instance, link prediction, conjunctive queries, etc., for its ability to learn implicit patterns hidden in observed data. This could be a good starting point for addressing temporal reasoning problems.

Base on the above discussion, here I propose three research directions that may contribute to temporal reasoning study:

(1) What are the different ways of time representations in knowledge graphs? Will a machine learning model a good alternative to classic reasoning for reasoning tasks? How do these representations influence the performance of machine learning models?

(2) How to integrate different time representations into machine learning models for downstream tasks, for instance, conjunctive queries?

(3) How to incorporate both spatiotemporal information into machine learning models in support of more complicated reasoning/query cases?

References

[1] Sotiris Batsakis and Euripides GM Petrakis. Sowl: a framework for handling spatio-temporal information in owl 2.0. In *International Workshop on Rules and Rule Markup Languages for the Semantic Web*, pages 242–249. Springer, 2011.

[2] Chris Welty and Richard Fikes. A reusable ontology for fluents in owl. 2006.

[3] Yankai Lin, Zhiyuan Liu, Maosong Sun, Yang Liu, and Xuan Zhu. Learning entity and relation embeddings for knowledge graph completion. In *Twenty-ninth AAAI conference on artificial intelligence*, 2015.