2021-2022-1
Final Report of Knowledge Representation and Reasoning

Major ________AI_________

Name ___Jiang Zhuoyang___

Student ID ____58119125 ____

Final Report: Appreciate TACT from my perspective

Jiang Zhuoyang
58119125
Group 9

College of AI, Southeast University, Nanjing, China
wtbzlkwtzcm@gmail.com

1 Research Paper Introduction

The paaper we researched on is Topology-Aware Correlations Between Relations for Inductive Link Predictionin Knowledge Graphs.[1]

1.1 Background

The knowledge graph structured a large amount of human knowledge, and carried out digitization. It is an important research topic in the intelligent engineering and science of modern society, and has a wide range of applications in daily life. Among them, the structural characteristics of the knowledge graph can use the old relationship to assist the establishment of the new relationship when updating the content, which is a manifestation of induction. The existing rule-based learning methods are also induction in nature. However, such inductive reasoning methods do not consider adjacent relationship triples when predicting links. If this is taken into account, it may be possible to achieve better "induction".

1.2 Main Contribution

This article clearly grasps the relationship between "semantic relevance between two relations" and "topological relationship between these two semantics on the knowledge graph".
Based on this relationship, the article proposes a new inductive reasoning method, which can make full use of this relationship, by observing and learning the intuitive knowledge graph topology, perceiving the relationship at the semantic level, so as to realize inductive reasoning.
This is TACT, which can effectively exploit Topology-Aware CorrelaTions between relations in an entity-independent manner.

1.3 Innovation

In my opinion, the most important innovation of this article is to capture the objectively existing inductive information of the "existing topological structure of the knowledge graph" and reduce the manual induction work. The article emphasizes the high correlation between graph structural features and semantic features, which is a very good starting point. Based on this, many new connections can be standardized.

1.4 Experimental Result

The author evaluated the model from two aspects: classification and classification. For these two indicators, the author compares the method proposed in this article with seven state-of-the-art methods.
From the perspective of classification, the TACT base proposed in this article is better than the inductive baseline on all data sets. TACT further improves the performance of TACT base, which is about 4% higher than GraIL on most data sets. The experimental results show that in the task of inductive link prediction, it is effective to use the TACT method to establish a topology-aware relationship model.
From the perspective of Ranking, experiments show that it is difficult for GraIL to model relational semantics when there are a large number of relations. In contrast, TACT can model complex relationship patterns by exploiting the correlations between relationships in known edge graphs. TACT-base can also significantly outperform the current state-of-the-art methods.

2 My Research Work

After reading through the paper, I discussed the research objects, research methods, and research results of the article together with other team members, and finally decided that I would introduce the overall model architecture of the TACT method. I made a logical diagram for the two major modules of the overall model architecture. Thereby forming a PPT.
After that, I was responsible for in-depth exploration of one of the two major tasks required to build the model.
I gave an in-depth introduction to the work of Modeling Correlations between Relation, starting from task motivation, to basic knowledge construction, and then to quantitative modeling of relationships.
I introduced that the motivation of this work is to grasp the relationship between "semantic relevance between two relations" and "topological relationship between these two semantics on the knowledge graph", and then introduce the correspondence between the seven topological patterns and semantic relations, which leads to RCG to RCN model construction ideas, that introduced the specific Embedding formula.

3 Relationship with the Curriculum

The method proposed in this article is actually not only related to knowledge representation, but also related to reasoning. The purpose of this method is

to make full use of the structural features of the existing knowledge graph to improve the efficiency of new content connection, and the method of improving efficiency is also in line with the induction how inference works.
The article's application of the neural network model and the proposed method based on the topological relationship of the knowledge graph itself can inspire us to follow-up learning. We can start from the graph itself and at the same time start from the attributes of the semantic relationship, explore the common points of the two, and explore The characteristics that can be quantified and connected in the two directions, and continued improvement is only the method of representation and reasoning.

References

1. Chen, J., He, H., Wu, F., Wang, J.: Topology-aware correlations between relations for inductive link prediction in knowledge graphs. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 35, pp. 6271–6278 (2021)

Final Report: from IE & QA to CSQA

Jiang Zhuoyang
58119125
Group 9

College of AI, Southeast University, Nanjing, China
wtbzlkwtzcm@gmail.com

1 Motivation

From a task-driven perspective, we understand the basic concepts of knowledge graphs, and more importantly, understand the challenges in the field of knowledge graphs.

2 Basic Technology

After that, we descended from the field of knowledge graphs to more essential technical foundations, mainly studying the technical foundations and technical applications related to "knowledge representation and knowledge graphs" in natural language processing.

2.1 IE

The first is information extraction technology. His most important role in the knowledge graph lies in the construction of the knowledge graph. IE extracts named entities, relationships and practices by traversing structured, semi-structured and unstructured text from existing encyclopedia data.
The first important technology in IE is named entity recognition, which mainly uses the underlying principles of HMM or CRF, or uses the LSTM neural network model to extract entities.
The second is event mining, which is mainly divided into event mining and event classification, using FSED and MetaLearning to realize the construction of natural language to knowledge graph.
The third is relation extraction. As one of the important components of knowledge graph triples, relations are the key technology for constructing knowledge graphs. In addition to traditional methods, we can also use MetaLearning, especially CML, to realize this process.
For the task of information extraction, we also learned how to evaluate the results of the task.

2.2 QA

The second key natural language processing task is QA. When learning natural language processing, we have learned the QA method based on knowledge graphs. In this course, we have studied QA for the application of knowledge graphs. We need to figure out the basic knowledge of QA first. There are three types of QA and two types of queries. Among the types of queries, we mainly studied the queries of complex problems and understood the challenges. After that, we learned about the two gaps that KBQA needs to solve, the lexical gap and the semantic gap for KBQA. And understand its operation process.

3 Application

Regarding the application, you can talk about the research papers of our group.[1] Today, KBQA related technology has more developments, among which, Complex Sequential Question Answering task is an important KBQA application sub-field and development direction.
In the research papers of our group, the researchers introduced the task of complex sequential question answering (CSQA), which contains a large-scale data set composed of conversations on linked QA pairs. Yamen proposed a CSQA model, which is a cross between the most advanced dialogue and QA model, and emphasized the deficiencies of this model in handling CSQA tasks.
To handle complex parsing tasks, we need to apply context to resolve co-references and ellipsis, and apply clarification to resolve ambiguous queries. So this task requires Subgraphs KG for Answer, which is Complex.
CSQA requires two systems. On the one hand, for the dialogue system, the high-level model here is the HRED model. On the other hand, for the QA system, the high-level model here is the key-value-based memory network model.
In the search engine task, KBQA involves entity-relation detection, semantic analysis and entity/relation linking;
In the map retrieval task, only entity/relation detection and entity/relation linking are involved;
In the chatbot task, dialogue management and natural language generation also need to be involved.
Maybe we can also apply basic technology and enhanced technology in more tasks related to linguistics and knowledge graphs, which requires our continuous exploration.

References

1. Saha, A., Pahuja, V., Khapra, M.M., Sankaranarayanan, K., Chandar, S.: Complex sequential question answering: Towards learning to converse over linked question answer pairs with a knowledge graph (2018)

Final Report: Our exploration of VKG

Jiang Zhuoyang
58119125
Group 9

College of AI, Southeast University, Nanjing, China
wtbzlkwtzcm@gmail.com

1 Motivation

Our topic is about Virtual Knowledge Graph(VKG) Technology. Currently, with the advancement of Semantic Technologies, large geospatial data sources have been increasingly published as Linked data on the Web. The LinkedGeoData project is one of the most prominent such projects to create a large knowledge graph from OpenStreetMap(OSM) with global coverage and interlinking of other data sources. [1]

2 My Work

After understanding the basic principles of technology and the workflow of methods, I am responsible for summarizing and exploring the evaluation and practice of the proposed methods in this paper.

2.1 Content Description

The evaluation is based on the qa task over LGD VKG Created by two frame. The sparqlify which is the previous approach and Ontop the approach this paper is introducing.They forms a kind of contrast.

Firstly there is a premise of the evaluation reproduction. This paper impliment evaluation work on these hardware. We need know that the hardware should satisfy the Datasets' demand.

After satisfication of the hardware condition, the test data set can be chosen to use. There is chosen 3 text area from OpenStreetMap, they are North-East Italy, Italy and Germany. Then the hardware will determin the dataset usage.The datasets were download from Geofabrik and loaded to PostreSQL.

Then, We can see that the OSM dataset features, which is fit the hardware condition.

The most important part is to design the query to do query answering evaluation. I divide queries into three categories.

The first category is QA through the operation of distance. Query 1 and Query 2 fell into this category.There is an instance operation of the query process.The dataset used here is NE Italy.(Actually the author has implement on all 3 dataset.) We can see that there given a class - the Amenities. And in Q1, the predefined distance is shown in the parameter form in such example.And the number of combinations of possible value is obtained and it's easy to reproduce.

Then another category is QA through the operation of Intersection. Query 3 and Query 4 fell into this category. There is an instance operation of the query process. The dataset used here is Italy. We can also see that there given a class - the Amenities. And in Q3, the given polygon is defined in such a parameter form in example. And it has the number of combinations of possible value, too.

The last category is QA through the operation of Inclusion. Query5 to Query 7 fell into this category. There is an instance operation of the query process. The dataset used here is Italy. We can also see that there given a class - the Amenities. And in Q56, the given polygon is defined in such a parameter form in example. In Q7, the given location is given by the point parameter.

2.2 Content Analysis

On all 3 dataset Record query respond time, Make comparison between Ontop and Sparqlify. We can analyse that Sparqlify can't approach all functions and in some function both can be reached by Ontop and Sparqlify, the Ontop method might work better.

All in all, ontop is built is the "Structural and Semantic" way which has the Optimization Techniques. It also has Huge Number Robustness. What's more it support more queries that combine topological and non-topological.

Finally, we can conclude that the VKG approach in LinkedGeoData is able to support GeoSPARQL queries that combine topological and non-topological operations on the database.

2.3 Future Work

Then after evaluation, the authors of this paper also gave some further research Direction depending on the Problems that can be improved and more project which can be made a usage. In addition to the improvement of data quality, we can also do some work.

Firstly, on Language tags:

- Through Sparqlify, data property and the respective language annotation can be performed at once.
- This feature is currently supported neither by R2RML nor by Ontop.
- More mappings were needed for languages ,since this.

There is a further Sparqlify feature: The declaration of IRI prefixes that term mappings can produce. A query's triple pattern can only match IRIs in one set of prefixes (such as lgd). The triple mapping produces IRIs in a different set of prefixes (such as dbr). So that, we can optimize the pruning of candidate triple mappings in the query rewriting phase

Then, for the Faceted search:

- Facilitate live exploration of reasonably sized subsets of Linked-GeoData's VKG.
- The performance was not yet sufficient for interactive purposes.
- LinkedGeoData's VKG stack together with the corresponding optimizations

Last but not least, to improve interlinking, we need know:

- "Generate using an interlinking engine and subsequently manually verify", not a scalable model.
- Wikidata community maintains links from Wikipedia to OSM, which provide a condition to make additional improvements.

3 Difficulties Encountered in Practice

We have also tried to operate actually However, I have a partial lack of knowledge of database and Linux. We have tried to get familier with the Ontop Operation/The docker Operation/the Database operation But we haven't obtained the LinkedGeoData actually. So We may do that later.

References

1. Ding, L., Xiao, G., Pano, A., Stadler, C., Calvanese, D.: Towards the next generation of the linkedgeodata project using virtual knowledge graphs. Social Science Electronic Publishing