TEST Engineering & Management

May-June 2020
ISSN: 0193-4120 Page No. 4239-4244

Query Completion using Knowledge Graph: A Semantic Approach

[1]Vidya S Dandagi, [2]Nandini Sidnal

[1,2]Department of Master of Computer Science and Engineering,
KLE DR. M.S. Sheshagiri College of Engg. & Technology, Belgaum, India
[1]vsdandagi@gmail.com, [2]sidnal.nandini@gmail.com

Article Info
Volume 83
Page Number: 4239-4244
Publication Issue:
May - June 2020

Article History
Article Received:*19 November 2019*
Revised: *27 January 2020*
Accepted: *24 February 2020*
Publication: *12 May 2020*

Abstract

The completion of the query is the first job of the search engine. Query completion is a process of suggesting a set of words or phrases Semantics is the linguistic study of the meanings in a language. It is concerned with the relationship between words, and phrases. Ontologies can be used to provide formal semantics. In this paper, we create an ontology using Resource Description Framework, where the information is symbolized as a triplet i.e. subject, predicate, and object. Knowledge Graph is a new type of Knowledge Representation. We propose a simple model that performs query completion using knowledge Graph with emphasis on object prediction by using Recurrent Neural Network. There are various graph embedding algorithms like TransE,, ComplEx, DistMult, and HolE. Embeddings are formed for these triplets using the graph embedding algorithms. Tensor factorization approach for embedding relations and entities is proposed. True triplets that are ranked in the top-N are computed by Mean Reciprocal Rank. The experimental results show how relevant the query is suggested.

Keywords: *Ontology, Query Auto completion, Knowledge Graph*

1. Introduction

Search for information by the users is rapidly increasing. The first task of a search engine is the query completion. Query completion is a collective method for search engines. Knowledge representation (KR) is the building model for a particular domain. It allows intuitive reasoning and understanding. Such formal models are called Ontology. It can be used to provide formal semantics. The association of data with proper models makes the information much more open to machine processing and interpretation. Ontology is a straight forward description of conceptualization. Ontologies are used in information retrieval to represent domain interpretation that adds a semantic layer to the system [1]. The Semantic Web is a network of data linked up in such a way that it is processed with ease by machines. It is the method of representing information on the World Wide Web. The Semantic web is an add-on to the current web, in which the information has a precise meaning [2]. Figure 1 depicts the proposed layered approach of semantic network architecture.

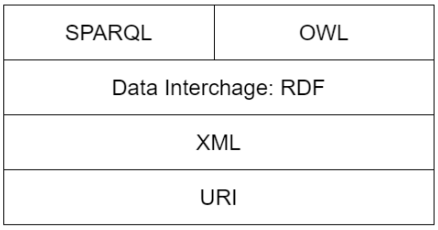


Figure 1: Architecture of Semantic web

The first layer is the Uniform Resource Identifier (URI) that is the starting point. Uniform Resource Identifier (URI) gives the identity of the resource and helps in retrieval of data.XML Schema that is the second layer is used to pinpoint common syntax in the semantic web. XML namespace identifies various markup vocabularies. RDF that is the third layer transmits a semantic model that represents data on the Web in the graph form. It is made up of triples. The head is a literal which gives the description of the resource. A predicate is the type of relationship between the head and the object. Object is a literal, which represents the "value" of the attribute. The ontology vocabulary layer is the fourth layer that defines knowledge. It relates the semantic link among the various kinds of data. It gives information about the resources. SPARQL is a demonstrative

May-June 2020
ISSN: 0193-4120 Page No. 4239-4244

language suggested by the World Wide Web Consortium. It is used to extract the data from RDF graphs. It enhances graph pattern matching for searching and information extraction [4].

Creation of Ontology:

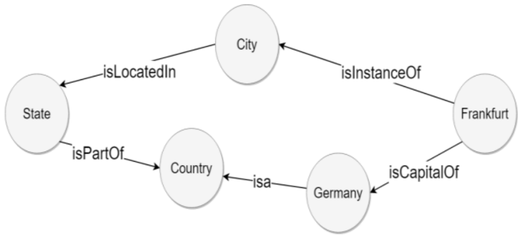


Figure 2: Ontology for tourism domain

Fig 2 depicts ontology for the tourism domain. It consists of class like a Country which has subclasses a City, State. Germany is the entity of the Country class, is a is Object property. Germany and Country class is related by the is a relationship. Frankfrut is an instance of City class. Subject as Frankfrut, is Capital Of is the predicate, which is Object Property, Germany is the object.

Query Auto-Completion is an interactive feature which helps users in formulating queries by providing completion suggestions as they type.

Features of QAC:

- Helps the user in identifying the right vocabulary to be used,
- Can minimize the typographic errors in the input,
- Speeds up the interaction,
- Saves user's time and enhances the search experience,
- Predicts the user's intended query and saves keystrokes and helps users to improve search quality.

The query completion, depends mainly on the user's click. In query completion the suggestion term that is most commonly clicked is not to the user's intended query. Multi-Armed Bandit is used to place the more relevant lower ranked completions in the top ranked completions [5]. To understand the periodical changes in the popularity of the query Discrete Fourier Transform is used [6]. The neural network language model is used for creating personalized query completions [7]. The semantic similarity between the current queries and the preceding queries is calculated by using the Word2vec model. Predicting the query intention is improved by combining the semantic similarity and query frequency [8]. Large scale databases use Synonyms and abbreviations that help in query completion. Top-most synonyms are used to carry out the completion [9].

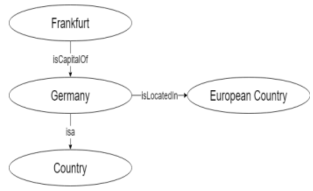
A Knowledge Graph represents the facts; it consists of entities, relationship, and semantic descriptions. Entities are objects related to real world and abstract concepts. Relationship infers the relationship between entities. Semantic description contains types and properties of entities with a well-defined meaning. The Knowledge Graph is visualized as a graph. It is been generated from a knowledge base.

(Frankfrut, is Capital Of,
Germany)

(Germany, is Located IN,
European Country)

(Germany, is a, country)

a) Factual triples in Knowledge Base



b) Entities and relations in Knowledge Graph

Figure 3: An example of Knowledge Base and Knowledge Graph

Figure 3 shows an example of a Knowledge Base and Knowledge Graph.

The basic concept of Knowledge Graph (KG) completion is to instinctive deducing missing triples by taking into consideration triples that exist. This is attained by carrying out the link prediction on the KG. A link prediction calculates the probability of an entity connected to another entity through a relation in the graph. For instance, given a query (Frankfrut, Is Capital Of, ?), it could predict that the missing entity is Germany[19]. The Reasoning is most important for developing a large Knowledge Graph and also for completion. Learning methods in Knowledge Graph reasoning are embedding, rule based reasoning There are different graph embedding model RESCAL the first matrix factorization based model , TRANSE this model uses translation based embedding. TRANSH, TRANSR, TRANSD are the extended variants of TRANSE, which project the embedding vectors of entities into various spaces. Rule learning methods are precise and learn deductive and interpretable inference rules.

2. Literature Review

Reasoning is important for building up for big Knowledge Graph completion. To increase the accuracy of link prediction various methods are put forth such as Tensor factorization and simplE. These methods use semantic features to understand the relationship among entities. The method known as probabilistic soft logic is used as a reasoning method. The tensor factorization method also helps in the link prediction problem. There are different graph embedding models like RESCAL is the first matrix factorization based model, TRANSE, this model uses translation based embedding. TRANSH, TRANSR, TRANSD are the extended variants of TRANSE, which project the embedding vectors of entities into various spaces.

Ontology carries out reasoning to work out logical results by using a group of asserted facts and axioms. Reasoning on ontology relation infers new conclusion. Different knowledge reasoning techniques are temporal and spatial reasoning. Reasoning based on time interval is known as temporal reasoning. Spatial reasoning creates general knowledge about the entity and geographical relationship between the entities [20].

Knowledge Graphs are powerful resources for many Artificial Intelligence projects, such as Question & Answering, web searching and semantic analysis. Knowledge Graph embedding is the main step of Knowledge representation. Knowledge Graphs have crypto graphed the structured data of relational facts. Knowledge Graph constitutes of triplets denoted as (h,r,t) [10]

A Knowledge Graph which creeps the entire web could be explained as a complete Semantic web. Knowledge Graph, retrieves and combines data into an ontology and uses a reasoner to derive new knowledge. Reasoning capabilities is the essential characteristics to understand new knowledge [11].

Knowledge Graphs are formed by semantic web language known as Resource Description Framework. A new method for embedding Knowledge Graph is through real valued tensors. These tensor-based embeddings retrieve relationship from the Knowledge Graph. When the average degree of an entity in a graph is high, the tensor decomposition model performs well. This achieves more improvement in predicting new facts across Knowledge Graph [13].

There are many embedding techniques such as TransE, PTransE. TransE and TransH, are easy and productive. They construct entity and relation embeddings by taking into consideration the translation from head and tail entity. In TransE model, relationships occur by explaining them on the low-dimensional embeddings of entities. TransE executes effectively in 1-to-1 relations. It has issues for modeling 1-to-N, N-to-1 and N-to-N relations. TransH and TransR, let an entity to have different representations for various relationships. A cascading embedding framework is introduced in which extraction of semantic, graph features is taken. These features are taken from knowledge embedding and graph embedding. These are the inputs for the cascade learning model. This helps in handling the challenge of unbalanced relations [12].

A simple bilinear model for Knowledge Graph completion is been proposed. Knowing the present links between the entities, link prediction helps in predicting new links for a Knowledge Graph. Tensor factorization method is useful for link prediction problems. The first tensor factorization method is Canonica Polyadic, this method computes two embedding vectors for each entity. SimplE is the improvement of Canonical Polyadic in which the two entities are learned dependently, one embedding vector for each relation and two vectors for each entity. One vector is used when the entity is head and other is used when the entity is the tail [16].

Knowledge Graph completion includes time data as one of the feature in the embedding approach. Long short term memory (LSTM) model is used to learn the knowledge graph inferred facts in which time is augmented [26].

Knowledge Graphs are not static. They are generated by adding and deleting the triples. Embedding models aim at static Knowledge Graphs. Embeddings fail to consider dynamically created Knowledge Graph. Dynamic Knowledge Graph Embedding consist of two ways of representation, knowledge embedding, contextual element embedding. Knowledge Graph embedding is learnt from scratch by an online algorithm that yields more accuracy. This model is more robust and scalable for online learning of embeddings [27].

The creation of the statistical model for relational data is known as Statistical Relational Learning. There are two models; the first depends on the latent feature model such as tensor factorization and multiway neural networks. The second model is based on extracting the observable patterns in the graph. A combination of the above has increased the power of modeling at low cost. Models on statistical methods of graphs and information extraction methods based on the text can be used in creating Knowledge Graphs. RESCAL is a relational latent feature model that triples via pair wise interaction of latent features [18].

A framework that combines embedding learning and rule learning is proposed. Embedding is studied from current triples and old triples to deduce the rules. Rules are studied by axiom injection through t-norm fuzzy logic. Embedding entities are represented as vectors and its relation are represented as matrices. This is carried out through a Linear map [21].

A new method for generating negative samples to train the Knowledge Graph embedding is known as adversarial learning. This framework depends on the policy gradient to train the generator, which creates negative triples. By generating negative samples, the discriminator model is able to learn better. A single step REINFORCE method is used to enable back propagation of error [22].

In this model, the link predictor helps in the classification process and captures the classes, labeled and unlabeled data. An algorithm named Classification Using Link Prediction (CULP) is which uses a novel formation called Label Embedded Graph and link

May-June 2020
ISSN: 0193-4120 Page No. 4239-4244

predictor to search the class of unlabeled data. A link predictor known as Compatibility Score is used as predictors for CULP. Classification using link Prediction utilizes a graph called Labeled Edge Graph that enables the use of link predictor. [14].

To enhance the exactness of link prediction a new method named Entity Link Prediction for Knowledge Graph (ELPKG) is proposed. Entity link prediction algorithm is used to know the relationship among entities. The method adopted is known as the probabilistic soft logic-based reasoning method. This method adds paths and semantic based features to understand the relationship among entities. To compute the similarity of tuples the Euclidean distance is computed. The two tuples are similar if the distance between them is the shortest. This ELPKG addresses the missing relations in the Knowledge Graph [15].

The relational learning agent named MINERVA(Meandering In networks of Entities to reach Verismiliar Answers) is developed which takes perfect steps of a decision to choose the relation edges to retrieve the correct answer. This model is strong and can understand a long chain of reasoning [25].

Multi-Label Deep Neural Network model emphasizes on relation prediction. The goal of this work is to predict the relationship among the entities in a Knowledge Graph. Link prediction among entities is important for creating huge ontologies and for Knowledge Graph completion. If the relation is predicted accurately it can be augmented to a given ontology. [17]

A novel method that combines the Convolutional Neural Network and the bidirectional long- and short-term memory is put forth. Combination of accurate triples and corrupted triples grouped to form the training information. Then a Path Ranking algorithm is adopted to get the relation paths for each training instances that are relevant to the relation r. Random walks are performed on the entire graph to understand which relation connects the source entity with its target. To understand the semantic co-relation between two entities and the path between them is done by creating a vector representation. The paths between the entities are taken up by the attention process. This model performs multistep reasoning over path representation in an embedding space. Path encoder is effective in extracting features from paths in large graphs [19].

A single high volume RNN model is introduced that allows chains of reasoning over multiple relation types. The Neural attention method is used to reason over multiple paths. To improve speed and accuracy across multiple paths, pooling is done [23].

This model predicts the whole triples for a given entity. To get triples in a Knowledge Graph as a sequence, a new model is proposed to use a Knowledge Graph specific multi-layer RNN. To predict the triples, a beam search method is implemented with a large window size [24].

3. Problem Definition

A semantic approach, to auto-complete a given query using Knowledge Graph.

4. Proposed Architecture

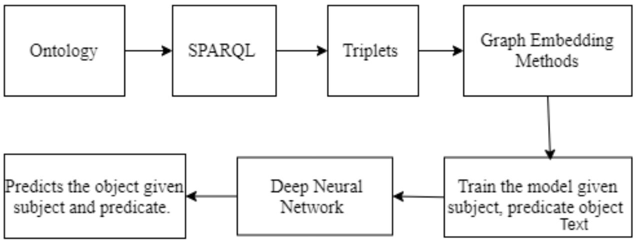


Figure 4: Model Architecture

This architecture gives an overview of query completion by predicting the object when the subject and predicate are given. Ontology is created using RDF. By using SPARQL, triplets are generated. Embeddings are created for these triplets by tensor factorization embedding model and trained by the Recurrent Neural Network model. This helps in predicting the object when subject and predicate are given.

5. Implementation

By using tensor factorization that defines two things, embedding functions for entities and relation. The values of the embeddings are learned by using the triples from the Knowledge Graph. Recurrent Neural Network is implemented using Keras.

6. Result

Mean Reciprocal Rank is computed for this model as 0.825. The model accuracy is 85%. It reflects how relevant the query is suggested.

The comparison of Mean Reciprocal Rank with the previous method, that is Compl Exembedding method is 0.692 [16].

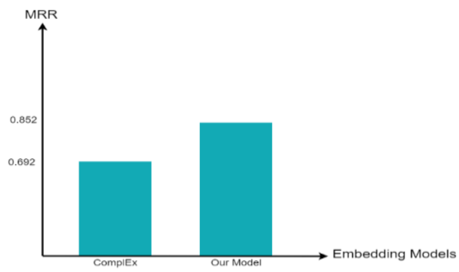


Figure 5: Comparison of our model with the ComplEx model

7. Conclusion

In the present query completion methods semantics between the terms is missing. In this article we proposed

May-June 2020
ISSN: 0193-4120 Page No. 4239-4244

a query completion using Knowledge Graph .A Knowledge Graph is represented as subject, predicate and object. Given a subject, predicate this model predicts the object. It takes into consideration semantic relationship between entities. The relevance of the query suggested is computed by accuracy.

References

[1] Bouramoul, M. Kholladi and B. Doan, "An ontology-based approach for semantics ranking of the web search engines results," *2012 International Conference on Multimedia Computing and Systems*, Tangier, pp. 797-802, 2012.

[2] Issa, Taroub. (2013). "How Web Applications Complement Search Engines?." Proceedings of the 2013 Palestinian International Conference on Information and Communication Technology, PICICT 2013. 99-106. 10.1109/PICICT.2013.26.

[3] Wang Yong-guil, "Research on Semantic Web Mining," Dept of Software. Liaoning Technical University. Huludao, Liaoning, China. yghI2000@163.net.

[4] Arias, Mario, et al. "An empirical study of real-world SPARQL queries," 1st International Workshop on Usage Analysis and the Web of Data, arXiv:1103.5043, 2011

[5] Yingfei Wang, Hua Ouyang, Hongbo Deng, and Yi Chang, "Learning online trends for interactive query auto-completion," IEEE Transactions on Knowledge and Data Engineering, 29.11, pp. 2442-2454, 2017.

[6] Danyang Jiang, Honghui Chen, and Fei Cai, "Exploiting query's temporal patterns for query autocompletion," Mathematical Problems in Engineering,. pp.1-8., 2017, 10.1155/2017/7490879.

[7] Aaron Jaech and Mari Ostendorf, "Personalized language model for query auto-completion," Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics, Volume 2, pp. 700-705, 10.18653/v1/P18-2111, 2018.

[8] TaihuaShao ,Honghui Chen and Wanyu Chen, "Query auto-completion based on word2vec semantic similarity," Journal of Physics: Conference Series, 1004, No. 1, IOP Publishing, 2018.

[9] Pengfe Xu, and Jiaheng Lu, "Top-k string auto-completion with synonyms," International Conference on Database Systems for Advanced Applications, Springer, Cham, 2017.

[10] Guan, Niannian, Dandan Song, and Lejian Liao. "Knowledge graph embedding with concepts," *Knowledge-Based Systems*, 164, pp. 38-44, 2019.

[11] Ehrlinger, Lisa, and Wolfram Wöß, "Towards a definition of knowledge graphs," *SEMANTiCS, Posters and Demos Track,Leipzing, Germany, September 13-14*, 2016.

[12] Li, Daifeng, and Andrew Madden. "Cascade embedding model for knowledge graph inference and retrieval," *Information Processing & Management, Vol.* 56.6, 102093, Nov. 2019.

[13] Padia, Ankur, et al. "Knowledge graph fact prediction via knowledge-enriched tensor factorization," *Journal of Web Semantics, Vol.* 59, 100497, 2019.

[14] Fadaee, Seyed Amin, and Maryam Amir Haeri, "Classification using link prediction," *Neurocomputing* , vol.359, pp. 395-407, 2019.

[15] Ma, Jiangtao, et al., "ELPKG: A High-accuracy link prediction approach for knowledge graph completion," *Symmetry* , vol.11.9, 1096, 2019.

[16] Kazemi, Seyed Mehran, and David Poole. "Simple embedding for link prediction in knowledge graphs," *Advances in neural information processing systems 32*,Canada, 2018.

[17] Onuki, Yohei, et al., "Relation prediction in knowledge graph by Multi-Label Deep Neural Network," *Applied Network Science*, 4.1, 20, 2019.

[18] Nickel, Maximilian, et al., "A review of relational machine learning for knowledge graphs," *Proceedings of the IEEE* 104.1, pp.11-33, 2015.

[19] Jagvaral, Batselem, et al., "Path-based reasoning approach for knowledge graph completion using CNN-BiLSTM with attention mechanism," *Expert Systems with Applications*, vol.142, 112960 , 2020.

[20] Gayathri, R., and V. Uma. "Ontology based knowledge representation technique, domain modeling languages and planners for robotic path planning: A survey," *ICT Express* 4.2, pp. 69-74, 2018.

[21] Zhang, Wen, et al., "Iteratively learning embeddings and rules for knowledge graph reasoning," *The World Wide Web Conference*, 2019.

[22] Cai, Liwei, and William Yang Wang, "Kbgan: Adversarial learning for knowledge graph embeddings," Proceedings of NAACL-HLT , Association for Computational Linguistic, New Orleans, Louisiana pp.1470–1480, June 1 - 6, 2018.

[23] Das, Rajarshi, et al., "Chains of reasoning over entities, relations, and text using recurrent neural networks," Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers, pp.132–141, Valencia, Spain, April 3-7, 2017.

[24] Guo, Lingbing, et al., "DSKG: A deep sequential model for knowledge graph completion," *China Conference on Knowledge Graph and Semantic Computing*. Springer, Singapore, 2018.

[25] Das, Rajarshi, et al., "Go for a walk and arrive at the answer: Reasoning over paths in knowledge bases using reinforcement learning," *arXiv preprint arXiv:1711.05851* (2017).

[26] García-Durán, Alberto, Sebastijan Dumančić, and Mathias Niepert. "Learning sequence encoders for temporal knowledge graph completion," Proceedings of Conference on Empirical Methods in Natural Language Processing, Association for Computational Lingustics, Brussels, Belgium pp. 4816–4821, October 31 - November 4, 2018.

[27] Wu, Tianxing, et al., "Efficiently embedding dynamic knowledge graphs," arXiv,1910. 06708v1 [cs.DB] 15 Oct 2019.