Research on Knowledge Graph Application Technology

Yong Chen¹, a Xueli Chen², b

¹Nanjing Normal University
Jiangsu Province of China
15605199883, incl. China

²Nanjing Normal University
Jiangsu Province of China
harry.cv1994@gmail.com b61021@njnu.edu.com

ABSTRACT. The development of the Internet has mainly gone through three stages. The first stage can be defined as the era of document interconnection. At this time, the main task of the Internet is to provide corresponding layout content for readers to read. The second stage can be defined as the era of data interconnection. Recently, the main task of the Internet provides an interaction that makes the user not only the reader of the web content, but also the maker of the web content [26]. The third stage can define the era of semantic interconnection, which pays more attention to the Internet. The creator and editor of network knowledge, the network truly becomes the user's understanding and provider of needs. Knowledge graph's ability is to understand its own strong semantics, as well as the openness and interconnectivity of the knowledge graph, makes the vision of the semantic connected era with circumstance as close as possible. The knowledge graph was first proposed in the project released by Google on May 17, 2012. Knowledge graph is not a new technical concept. As early as in 2006, [20], the concept of Semantic Web was mentioned. Nowadays, scholars began to call on the use of ontology model in data to express data implicit semantics. At present, intelligent information service applications are increasingly appearing in all aspects of technology and life, such as intelligent search, intelligent question and answer, personalized recommendation, and companies that have appeared in various fields and are in contact with consumers and users.

1. Introduction
The second part of the article first introduces the development of the Internet and the future trends, and leads to the application scenarios of the current knowledge graph. The concept of knowledge graph is traced back to the source. And some basic information and concepts are described.

The third part of the article describes in detail some related knowledge concepts of knowledge graphs, and systematically describes the existing technical concept application fields and research methods. The development and connection of data cleaning, data mining, knowledge extraction, knowledge acquisition, knowledge fusion, knowledge reasoning and other technologies are introduced. The fourth part of the article gives a detailed introduction to the application of knowledge graphs, and describes the different methods applied in different periods. Introduce the application of knowledge graphs in search systems, question and answer systems, recommendation systems, etc.
2. Definition and structure of knowledge graph

2.1. Definition of knowledge map

The initial definition stems from a description of the Knowledge Graph project which published by Google is defined as a bridge between Google's traditional search and smart search. In essence, the knowledge graph is a description of objective things. So far, knowledge graphs have been widely used in various large-scale knowledge bases.

2.2. Architecture of Knowledge Graph

Knowledge graph is a kind of structural data to describe objective things. Objective things are generally represented by conceptual abstractions of concepts, entities, attributes, etc. In the process of constructing knowledge graphs, common representations are triples, three. Tuples represent the two most common representations such as (entity 1 - relationship - entity 2) and (entity - attribute - attribute value). The construction of knowledge graphs is generally divided into two types. One is to extract concepts related to entity attributes from the structured data of various encyclopedia websites, and add them to the knowledge base; the other is to collect data from publicly. The resource model is extracted and added to the knowledge base through a series of screenings such as manual review. Therefore, the knowledge graph mainly involves such knowledge as data cleaning, knowledge mining, knowledge acquisition, knowledge representation, knowledge fusion, and knowledge reasoning.

3. Architecture of Knowledge Graph

3.1. Knowledge mining

Knowledge mining is the process of extracting implicit, previously unknown and potentially valuable information from a database.

In order to determine how data mining technology (DMT) and its applications are developed, Liao, S. H. et al. surveyed and classified the literature from 2000 to 2011, reviewing data mining techniques and their applications and developments. The author retrieves the required data from previous journal articles and other literatures, and analyzes the three types of knowledge, analysis, and architectures, depending on the survey and classification, and also analyzes them in different studies and Application in the field of practice.

In 2016, Banuqitah, HU et al. proposed an agent-based two-level self-supervised relationship extraction system based on the UMLS unified medical language system knowledge base [2]. The model uses a self-supervised method for relation extraction (RE). The method is to construct an enhanced training example using UMLS information with mixed text features, and the model combines Apache Spark and HBase BD technology with multi-data mining and machine learning techniques, and also with a multi-agent system (MAS).

3.2. Knowledge extraction

The KNOWITALL system is designed to automate the lengthy process of extracting a large number of facts (such as the names of scientists or politicians) from the Web in an unsupervised, domain-independent and scalable manner [8], which outlines the new architecture of KNOWITALL and design principles, and highlight its unique ability to extract information without any manual-tagged training examples.

The article raises a challenge: how can we improve the recall and extraction rate of KNOWITALL without sacrificing accuracy? Organize the full text, and then propose three different methods to deal with this challenge and evaluate their performance. Pattern learning learns domain-specific extraction rules that support additional extraction. Subclass extraction can automatically identify subclasses to facilitate recall (e.g., "chemists" and "biologists" are identified as subclasses of "scientists"). The list extracts a list of positioning class instances, learning a "wrapper" for each list. Since each method is guided from KNOWITALL's domain-independent method, these methods also exclude manual training
examples. At the end of the article, some experiments were also reported, focusing on the establishment of a list of named entities to measure the relative effectiveness of each method and to demonstrate their synergy.

Some scholars have proposed an application of a kernel method for extracting relationships from unstructured natural language sources [24] and introduced the kernel defined on the shallow analytical representation of the text and designed an efficient algorithm for computing the kernel. The author combines the support vector machine and the voting perception learning algorithm to extract the human-dependent relationship and the organization-position relationship from the text, and evaluates the proposed method, and compares it with the feature-based learning algorithm.

3.3. Knowledge Reasoning

Knowledge reasoning is mainly divided into two categories, one is graph-based reasoning and the other is logic-based reasoning. In a 2013 article, we introduced the use of neural tensor networks to reason about entities and relationships [16]. The article mainly introduces a neural tensor network of expressions, which is used to reason the relationship between two entities. Previously we always represented an entity as some discrete atomic unit, or a single entity vector. We have experimentally shown that performance can be improved when the entity representation constitutes the average of the word vectors. Finally, the authors also demonstrate that the performance of all models is improved when these word vectors are initialized with word vectors learned from unsupervised large corpora. The author evaluates the real relationship between the entities of a given subset of the knowledge base of the model by considering the prediction of additional problems. It is concluded that this model is superior to the previous model, and can classify the invisible relationships in WordNet and the accuracy of FreeBase is 86.2% and 90.0%, respectively.

Some scholars have also considered the reasoning method based on logical combination [6]. The purpose is to solve the dilemma of existing knowledge reasoning in the knowledge base of similar fields. They extend the representation of the logic of the description to generate a combination of description logic, using the similarity of concepts to relate the concepts of similar fields, thus giving the logical grammar and semantics of the combined description. The article finally gives The Tableau algorithm is used to implement this concept, and finally the feasibility of logical reasoning is confirmed by an example.

4. Application of knowledge graph

4.1. Intelligent Search System

The most important feature of the intelligent search system is that it can truly understand the needs of human beings, and can use various aspects of knowledge to analyze the needs of users and give the search answers that best meet the needs of users. The form of network ontology language was originally transformed from SHIQ and RDF to OWL. As early as 2003, scholars proposed a process of network ontology language [11]. The article introduces the necessity of OWL ontology language as a new language and also comprehensive summary the OWL's DL axioms and facts. The article also introduces its advantages and unstoppable development trend as a new language, as well as some comparisons in its development the stage. The traditional information retrieval system expresses documents and queries through keyword collections, but such search queries are not very efficient and performance-oriented.

Scholars in related fields have proposed a semantic text search method based on ontology query extension [13]. The main purpose is to use the ontology feature of the named entity and its potential related entities to perform semantic interpretation of documents and queries and to provide different advantages for semantic annotation and search with different ontology combinations.

One of the biggest problems facing intelligent search systems is the query problem. Some scholars have proposed a method to enable semantic association query in the Semantic Web. The main idea is to represent the concept of semantic association as a complex relationship between resources and entities, called ρ-Queries [1]. The main idea is to formalize the RDF data model by introducing the concept of
attribute sequences as types, and to provide a formal framework for supporting complex queries such as semantic associations. After an open source semantic retrieval system YaSemIR[5] was proposed by Davide Buscaldi et al., it also attracted great interest from experts in related fields. The query is mainly based on one or more ontology of OWL format normalized into the ontology of SKOS format for semantic query indexing out a collection of text.

4.2. Question and answer system

The question-and-answer system is similar to smart search, but it is very different. The main function of the callback system is to be able to talk to the user in the form of question and answer, and to give feedback to the answers raised by the user. Some scholars use the word embedding model to construct the question and answer system in a supervised mode [3]. They use auto-learning data and Wikipedia data for model training, and the answers to the questions are expressed in a set of triples. The author also designed a new model of word embedding to be used in the evaluation of the correctness of the triplet answer. Finally, it is confirmed by examples that the performance of the model has been greatly improved, but it uses a lot of manual tag data in the research process.

Some scholars have done an in-depth study of the question and answer system from the perspective of the figure. The study of the question and answer system based on subgraph embedding is one of them [4]. Using fewer manual attribute tags is a big advantage for them. The author is committed to learning how to answer questions under the broad subject from the knowledge base, and they use low-dimensional words and knowledge base embedding to score the answers to the answers. Finally, experiments proved that the effectiveness of the system shows good performance. Some scholars are based on the multi-column convolutional neural network question-and-answer system [7]. The author uses convolutional neural networks to learn their distribution and representation from three different aspects: answer type, answer path and answer context. A low-dimensional approach is used in entities and relationships, combined with a problem vocabulary score calculation method designed by them to assess the core vocabulary of the problem. Finally, a lot of experiments were carried out on Freebase, which shows that our system has great advantages in performance compared with other conventional systems.

Anthony Fader et al. at the University of Washington describe a new learning algorithm for open-ended question-and-answer systems [3]. The author introduces a template for a scalable learning algorithm that can lead to general problems, as well as lexical changes in entities and relationships. The biggest advantage of this algorithm is that it does not require manual annotation and can be applied to large, triad knowledge bases with no data. With the increase of the number of knowledge bases, the question and answer system of multiple databases has also emerged.

Among them, Yuanzhe Zhang et al. proposed a multi-knowledge question-and-answer system model [25]. The biggest problem we face in the process of constructing multiple knowledge bases is the alignment between the index libraries. Different alignments also fundamentally affect the interaction process. The author has an integer linearity based on this problem. The new joint model of planning unifies the two processes of alignment construction and query construction into a unified framework, which ultimately proves to be effective and can improve the performance of alignment constructs and query constructs. As shown in the figure 1, it is a simple intelligent question-and-answer system.
4.3. recommendation system

The recommendation system is the machine's knowledge of using big data, knowledge graphs, etc., comprehensively analyzing the potential needs of users, analyzing, giving conclusions, and recommending relevant content to users, thus forming a benign cycle. User-like content will be recommended. The system's data set, the recommendation system in turn recommends relevant content to the user. The key technology in the recommendation system is the construction of recommendation algorithms, and there are mainly collaborative filtering recommendation algorithms. There are three main types of recommendation systems, namely user-user (UU matrix), user-item (UV matrix), and articles - Item (VV matrix).

Jennifer Nguyen et al. proposed a matrix decomposition algorithm to improve the accuracy of the proposed algorithm [15]. The author first introduces the matrix decomposition method (MF) of collaborative filtering. The MF method is to observe the statistical user's preference through proof decomposition. The author's enhanced matrix decomposition algorithm mainly focuses on the content information directly collaborate into the matrix decomposition method, and filter together. Use forced alignment penalty to narrow the regression constraint and force potential project features to become content attributes. In addition to the matrix decomposition, there is a more popular method is the Markov
chain (MC). The MC method is to predict the next action by learning the user's recent actions. Some scholars have proposed a combination of the two methods.

A graph of personality was based on Markov chain [18]. The main features of the method are not the same as usual. Instead of using the same transformation matrix for all users, a separate transformation matrix is used for each user, which is generally a transformation cube. And the decomposition model is introduced so that each conversion can receive the effects of similar users, projects, and similar transformations.

In the field of recommendation systems, some scholars are using cyclic neural networks to enhance the construction of recommendation systems for long-term user history [10]. The main idea is to provide more accurate suggestions by modeling the entire callback, and for classics. The ranking loss function has been modified to make it more suitable for a particular problem. Labels are widely used in the web2.0 era.

At this stage, label optimization based on recommendation systems has also attracted the interest of scholars in related fields. Some scholars have extended Bayesian personalized recommendation criteria to task tag recommendation [17], and combined with stochastic gradient learning algorithm and bootstrap sampling learning algorithm to provide a linear pairwise interaction tensor decomposition model (PITF). It has a linear operation time and experimentally proves that the algorithm runs much better than the best quality method (RTF-TD) when the running time is reduced from O(k3) to O(k).

As shown in the figure 2, it is the main application of knowledge mapping.

![Figure 2. Knowledge graph is mainly applied.](image)

5. Conclusion

With the continuous development of knowledge graph, the application of future knowledge graph will become more diversified and multi-scenarios. Applications in various industries will be continuously developed and supplemented. It can be expected that in the future, you can truly realize human-machine barrier-free communication with intelligent machines. The future question-and-answer system will also be diversified and will pay more attention to the interaction of multiple rounds of dialogue. All of this can be applied to the application scenarios of the knowledge map. In the future, deep knowledge reasoning with knowledge graph and improved computational efficiency of large-scale knowledge graph require people to continually explore user needs, explore more important application scenarios, and propose more practical application algorithms. Therefore, the development of future knowledge graph requires both a rich accumulation of knowledge map technology, a keen perception of human needs, and the right application. I believe that in the future, life and work with the participation of knowledge graph will be smarter, simpler and better.
References

[1] ANYANWU, K. and SHETH, A., 2003. p-Queries: Enabling Querying for Semantic Associations on the Semantic Web. In Proceedings of the 12th International Conference on World Wide Web (2003), 690-699.

[2] BORDES, A., WESTON, J., and USUNIER, N., 2014. Open Question Answering with Weakly Supervised Embedding Models. Machine Learning and Knowledge Discovery in Databases, 165-180.

[3] BORDES, A., CHOPRA, S., and JASONWESTON, 2014. Question Answering with Subgraph Embeddings. Proceedings of EMNLP.

[4] BUSCALDI, D. and ZARGAYOUNA, H., 2013. YaSemIR: Yet another Semantic Information Retrieval System. In Proceedings of the 6th International Workshop on Exploiting Semantic Annotations in Information Retrieval, 13-16.

[5] DAOSHE, L. and SHIHAN, Y., 2012. Interdisciplinary reasoning on description logic. Application Research of Computers, 4503-4511.

[6] DONG., L., FURUWEI., ZHOU., M., and XU., K., 2015. Question Answering over Freebase with Multi-Column Convolutional Neural Networks. In Proceedings of the Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (2015), 260-269.

[7] ETZIONI, O., CAFARELLA, M., DOWNEY, D., POPESCU, A.M., SHAKED, T., SODERLAND, S., WELD, D.S., and YATES, A., 2005. Unsupervised named-entity extraction from the Web: An experimental study. Artificial Intelligence 165, 1 (Jun), 91-134. DOI= http://dx.doi.org/10.1016/j.artint.2005.03.001.

[8] HIDASI, B., KARATZOGLOU, A., BALTRUNAS, L., and TIKK, D., 2015. Session-based Recommendations with Recurrent Neural Networks. Computer Science.

[9] HORROCKS, I., PATEL-SCHNEIDER, P.F., and VAN HARMELEN, F., 2003. From SHIQ and RDF to OWL: the making of a Web Ontology Language. Web Semantics: Science, Services and Agents on the World Wide Web 1, 1, 7-26. DOI= http://dx.doi.org/10.1016/j.websem.2003.07.001.

[10] M., V., NGO., and CAO., T.H., 2010. Ontology-Based Query Expansion with Latently Related Named Entities for Semantic Text Search. In Proceedings of the 2nd Asian Conference on Intelligent Information and Database Systems (2010), 41-45.

[11] NGUYEN, J. and ZHU, M., 2014. Content-boosted Matrix Factorization Techniques for Recommender Systems.

[12] R. SOCHER, CHEN, D., MANNING, C.D., and NG, A., 2013. Reasoning with Neural Tensor Networks for Knowledge Base Completion. In Proceedings of the Neural Information Processing Systems 26. Red Hook (NY, USA2013), 926–934.

[13] RENDLE, S. and SCHMIDT-THIEME, L., 2010. Pairwise Interaction Tensor Factorization for Personalized Tag Recommendation. Association for Computing Machinery, 81–90.

[14] RENDLE., S., FREUDENTHALER., C., and SCHMIDT-THIEME., L., 2010. Factorizing Personalized Markov Chains for Next-Basket Recommendation, 811-820.

[15] SHADBOLT, N., HALL, W., and BERNERS-LEE, T., 2006. The Semantic Web revisited. IEEE Intelligent Systems 21, 3 (May-Jun), 96-101. DOI= http://dx.doi.org/Doi 10.1109/Mis.2006.62.

[16] ZELENKO, D., AONE, C., and RICHARDELLA, A., 2003. Kernel methods for relation extraction. Journal of Machine Learning Research 3, 6 (Aug 15), 1083-1106. DOI= http://dx.doi.org/Doi 10.1162/15324430332253205.

[17] ZHANG., Y., HE., S., LIU., K., and ZHAO., J., 2016. A Joint Model for Question Answering over Multiple Knowledge Bases. In Proceedings of the American Association for Artificial Intelligence (February, 12-17. 2016).
[18] ZHAO, C.L., LI, X.Y., XU, C.Y., and WANG, F., 2009. Informal learning based on web2.0. Icaie 2009: Proceedings of the 2009 International Conference on Artificial Intelligence and Education, Vols 1 And 2, 393-397.