Retraction

Retraction: Knowledge Verification Method Based on Artificial Intelligence-based Knowledge Graph Construction
(J. Phys.: Conf. Ser. 2146 012014)

Published 9 September 2022

This article has been retracted by IOP Publishing following an allegation that raises concerns this article may have been created, manipulated, and/or sold by a commercial entity. In addition, IOP Publishing has seen no evidence that reliable peer review was conducted on this article, despite the clear standards expected of and communicated to conference organisers.

The authors of the article have been given opportunity to present evidence that they were the original and genuine creators of the work, however at the time of publication of this notice, IOP Publishing has not received any response. IOP Publishing has analysed the article and agrees there are enough indicators to cause serious doubts over the legitimacy of the work and agree this article should be retracted. The authors are encouraged to contact IOP Publishing Limited if they have any comments on this retraction.

Retraction published: 9 September 2022
Knowledge Verification Method Based on Artificial Intelligence-based Knowledge Graph Construction

Ziting Yang 1,*
1Peking University, School of software & Microelectronics, Beijing, 100871, China
*Corresponding author e-mail: yangziting1ss@pku.edu.cn

Abstract. The knowledge graph connects real-world entities and concepts through their relationships, connects all different types of information to obtain a relationship network, and can analyze "relationship" issues. Creating a knowledge graph is a continuous process, and it needs to continuously learn new knowledge and update existing knowledge in the library as time and events change. However, since the accuracy of the updated new knowledge cannot be guaranteed, the new knowledge must be verified. This paper aims to study the knowledge verification method based on artificial intelligence-based knowledge graph construction. Based on the analysis of the knowledge graph construction process, the knowledge graph construction method and the knowledge verification method, knowledge verification is realized by constructing a probabilistic soft logic model. The experimental results show that the recall rate, F1 value, and AUC value of the candidate knowledge set are verified by the knowledge verification model proposed in this paper. Therefore, it can be inferred that the knowledge verification model proposed in this paper is effective.

Key words: Artificial Intelligence, Knowledge Graph, Knowledge Verification Method, Probabilistic Soft Logic Model

1. Introduction
Due to the rapid development of artificial intelligence and the rapid growth of network information in recent years, the amount of information generated every year around the world has increased exponentially. In this era of big data, traditional search engine technology can no longer meet the needs of mankind to master a large amount of information in all aspects [1-2]. The rapid development of machine learning, natural language information processing and related technologies have all made the establishment of large-scale semantic organizations possible, and the discovery of knowledge graphs has brought new ideas for improving the collection and utilization of information [3-4]. The establishment of a knowledge graph is a dynamic process. To enrich and supplement the existing knowledge graph, new knowledge points and new dynamic knowledge points must be added in a timely manner. New knowledge extraction methods often use information extraction technology to extract knowledge from multiple data sources. New knowledge extracted from traditional information extraction systems often involves certain errors. Therefore, the research on knowledge verification and reliability calculation of candidate knowledge sets has received extensive attention from domestic and foreign researchers. In foreign academic circles, some researchers first consider a
set of knowledge bases, including candidate knowledge, in order to form a high-quality knowledge base. Use weighted primary logic rules to build the credibility of the knowledge source to achieve knowledge verification based on the entity constraint of the new knowledge [5-6]; The main logic criterion of knowledge reliability does not fully consider the impact of ontology constraints on the reliability of candidate knowledge [7-8]; some researchers propose to establish a soft logic probability model and choose a more reliable knowledge set. The model they provide uses multi-source ontology and reasoning technology to infer the association with missing entities [9]; some scholars provide a knowledge verification system based on the Bayesian method, which mainly solves the problem of selecting the best knowledge to achieve more data The problem of source fusion [10]; in China, there are relatively few empirical research methods on the construction of the national knowledge graph conference, but the main research is that the knowledge empirical concept and knowledge fusion technology of data source knowledge fusion technology have been introduced at the global knowledge graph conference. Its principle and significance, therefore, it is generally believed that the Markov logic network model shows that the empirical problem of knowledge can still be solved [11-12]. The research results of predecessors provide a theoretical basis for the research of this article.

Based on consulting a large number of references related to "Knowledge Graph" and "Knowledge Verification Method", this paper combines the knowledge graph construction process, knowledge graph construction method and knowledge verification method to construct a probabilistic soft logic model to achieve knowledge verification, and finally verifies it through experiments. The model proposed in this paper is effective.

2. Research on Knowledge Verification Method Based on Artificial Intelligence-based Knowledge Graph Construction

2.1. Knowledge Graph Construction Process

(1) Data source: The amount of data on the Internet is increasing, and it contains a lot of valuable knowledge. The knowledge graph is organized by extracting valuable information from the Internet. The main data sources include encyclopedia sites, vertical sites, existing knowledge graphs and search log mining.

(2) Knowledge acquisition: how to acquire useful knowledge to complete the knowledge map. There are two main methods, one is knowledge extracted from the outside world, and the other is intermediate knowledge generated by reasoning. The investigation content includes obtaining entities, exporting relationships, exporting attributes, etc.

(3) Knowledge representation: Knowledge representation is the study of how to express and preserve knowledge graphs. The main methods are the representation of complex network graphs and the ternary representation of knowledge converted into low-dimensional knowledge graphs based on entities and relationships.

(4) Knowledge fusion: There are a lot of noise and unnecessary information in heterogeneous multi-source data, and knowledge integration is required. The data is integrated and stored with unified framework specifications such as entity alignment, knowledge processing, and knowledge update. It also includes entities. Techniques such as links and entity clarification.

(5) Knowledge reasoning: Knowledge reasoning is to discover hidden knowledge points on the basis of the existing knowledge graph, and then use the existing knowledge system to infer the possible or unknown things, and to supplement and improve the knowledge graph.

(6) Knowledge verification: In the process of dynamically constructing the knowledge graph, there is a contradiction between the acquired new knowledge and the knowledge of the original knowledge graph, and the reason for this contradiction may be the inaccuracy of the original new knowledge. The ambiguity of new knowledge indicates the need to verify new knowledge and add reliable knowledge to the knowledge graph.

2.2. Knowledge Graph Construction Method
(1) Knowledge extraction

Knowledge extraction is the main task of constructing a knowledge graph, because knowledge is an important element of the knowledge graph. Knowledge includes three elements: entity (concept), attribute and relationship, including entity extraction, relationship extraction and attribute extraction technology. When exporting entities, attribute entities, and entity relationships, entity export links use different methods for different structured data sources. The following is an overview of information technology for extracting information from various structured data sources.

1) Unstructured data

Machine learning techniques can be used to extract entities from unstructured data sources. Among them, the most commonly used technique is the bootstrapping method. The bootstrapping method only needs a semantic seed to perform entity mining. Other machine learning methods do not require corpus training sets, and are an unsupervised machine learning algorithm.

2) Semi-structured data

The semi-array data are very different from the Sass data encyclopedia. The concept organization is usually written in the form of a page containing the information in the object section. Internet feeding is often used to create content and work.

3) Structured data

The constructed data includes the existing ontology of the domain, XML files, data documents and related databases. To develop these properties, the data collected must be analyzed in a number of ways. Typically, two types of data are retrieved using OWL and XML tools. Use the toolbar to draw sections and write relationships between XML schemas. Use the GENE Toolkit to manage classes and classes in data export reports in OWL and RDF. For relational database resources, data analysis methods can be directly obtained, or invalid information can be eliminated to obtain more accurate relationships between entities.

(2) Knowledge integration

The merging of knowledge is an important step in the creation and expression of knowledge, and even the first data of human knowledge can be obtained by extracting knowledge. Because there are many types of original sources of knowledge, and quality cannot be assessed, there can be a lot of ambiguity, redundancy, and even misinformation. Therefore, the initial knowledge needs to be clarified and generalized. Knowledge merging techniques are an abstraction of the high level traditional methods of knowledge synthesis. Related technologies include entity refinement, centralized context analysis technology, multi-data fusion technology, and more.

(3) Knowledge evaluation

Assessment of the level of knowledge is an important factor related to the results of creating a knowledge map. Its goal is to increase the reliability of knowledge, eliminate cheap information, preserve reliable information and improve the quality of knowledge maps. Confidentiality refers to the reliability of data, which includes five dimensions: data source, data publisher name, data reliability, data verification and authorization.

2.3. Knowledge Verification Model

(1) Logical predicate representation

When applying feature entity division technology in the process of multi-data source fusion, such as the logical classification method that represents feature entities, the size of the entity feature set usually reaches 103 orders of magnitude. For example, when the feature set elements are described in order, the number of logical rules is It also increased exponentially. This paper divides entity attributes
into key attributes and general attributes, and fully considers the impact of each entity on the result of entity classification in the big data set. Important attributes are generally represented by the attribute name of the entity, while common attributes are represented by the attribute name.

(2) Reasoning and weight learning
The PSL model provides two most effective reasoning methods. The first concept is the maximum probability of inference (MPE), which estimates the most likely correct value of the logical rule from the existing data set. Choose the maximum likelihood parameter estimation method to learn the weights of the PSL model, and use the gradient function to estimate the weight parameters:

\[
\frac{\partial}{\partial \lambda_i} \log f(I) = -\sum_{r \in R_i} (d_r(I))^p + E\left[ \sum_{r \in R_i} (d_r(I))^p \right]
\]

(1)

(3) Probability distribution
The PSL model of entity analysis is defined through logical sentence suggestions. The probability distribution of the model is shown in equation (2). Among them, R is a basic logic rule of the PSL model; Z represents the design factor; d(r) represents the distance of the logic rule r and its satisfaction degree; the expression of the next order is introduced in the expression of P=1 The logical laws of predicates. According to the input mode definition of PSL. When checking the entity features, entity relationships, and ontology constraint values measured from the data set, the input data set is used to initialize the logic law of PSL modeling, and the weight training is completed, and then the distance value is freely defined. It can be realized if it meets the PSL modeling requirements and calculates the logic law of initialization in each system. In this way, people use the reasoning mechanism of MPE and calculate the probability that each of the three entities represents a new entity.

\[
p(I) = \frac{\exp\left[ -\sum_{r \in R} (d(r))^p \right]}{Z}
\]

(2)

3. Experiment
3.1. Experimental Data Set
In this article, we will verify the effectiveness and feasibility of the proposed knowledge verification model through actual experiments. The experimental data uses the NELL system to extract candidate knowledge. NELL is an information extraction system for web texts, which builds a knowledge base by extracting knowledge. NELL consists of four information extraction subsystems. CPL is a system that uses text boxes to export knowledge. CMC is an information retrieval system based on the knowledge word format. CSEAL refers to exporting table and list information to the network; RL uses the existing knowledge in the knowledge base to infer new knowledge. Each piece of data exported by NELL consists of six fields: entity, entity relationship, entity value, number of iterations, trust, and data source.

3.2. Evaluation Index
In order to help better evaluate the results of the test, this paper selects the accuracy rate (Precision), recall rate (Recall), F1 value, and AUC value as the test evaluation index.

4. Discussion
Figure 1. Knowledge verification experiment F1 value

Through the observation of the F1 value change curve in Figure 1, it is easy to infer that when the threshold = 0.60, the F1 value of the entity relationship, the entity tag, and both reach the maximum. Therefore, 0.60 is selected as the experimental threshold. If the candidate knowledge sets the probability of correct knowledge If the value is greater than or equal to 0.60, the knowledge block is considered to be correct and can be used as a data set to update the knowledge graph, and vice versa.

Table 1. Experimental results of knowledge verification model based on probabilistic soft logic

| Model   | AUC | F1  | Prec | Recall |
|---------|-----|-----|------|--------|
| NELL    | 0.764 | 0.672 | 0.800 | 0.579 |
| PSL     | 0.906 | 0.839 | 0.785 | 0.92  |
| PSL_Rel | 0.854 | 0.849 | 0.746 | 0.984 |
| PSL_Cat | 0.921 | 0.830 | 0.974 | 0.724 |
Figure 2. Experimental results of knowledge verification model based on probabilistic soft logic

The results of the model experiment are shown in Table 1 and Figure 2. In the chart, NELL represents the candidate knowledge set, and PSL represents the knowledge verification model proposed in this article. By comparing the experimental results, it can be seen that the recall rate, F1 value and AUC value of the candidate knowledge set are verified by the knowledge verification model proposed in this paper. Therefore, it can be concluded that the knowledge verification model proposed in this paper is effective.

5. Conclusions
In recent years, due to the rapid growth of the number of information networks and the rapid advancement of artificial intelligence technology, the retrieval quality of knowledge graphs has also been improved, and has received general attention from academia and industry. The research direction of knowledge graphs has gradually moved from basic theories. The research on knowledge verification methods for cognitive map modeling is also a recent research hotspot.

References
[1] Sang S, Yang Z, Wang L, et al. SemaTyP: a knowledge graph based literature mining method for drug discovery[J]. BMC Bioinformatics, 2018, 19(1):1-11.
[2] Wang C, Ma X, Chen J, et al. Information extraction and knowledge graph construction from geoscience literature[J]. Computers & Geosciences, 2018, 112(MAR.):112-120.
[3] Natthawut K, Ryutaro I. An Automatic Knowledge Graph Creation Framework from Natural Language Text[J]. Ieice Transactions on Information & Systems, 2018, 101(1):90-98.
[4] Zhou B, Bao J, Li J, et al. A novel knowledge graph-based optimization approach for resource allocation in discrete manufacturing workshops[J]. Robotics and Computer-Integrated Manufacturing, 2021, 71(3):102160.
[5] Wang T, Shi D, Wang Z, et al. MRP2Rec: Exploring Multiple-Step Relation Path Semantics
for Knowledge Graph-Based Recommendations[J]. IEEE Access, 2020, PP(99):1-1.

[6] Zhang S, Sun Z, Zhang W. Improve the translational distance models for knowledge graph embedding[J]. Journal of Intelligent Information Systems, 2020, 55(2):1-23.

[7] X Ma, Ma C, Wang C. A new structure for representing and tracking version information in a deep time knowledge graph[J]. Computers & Geosciences, 2020, 145(12):104620.

[8] Thrall J H, Li X, Li Q, et al. Artificial Intelligence and Machine Learning in Radiology: Opportunities, Challenges, Pitfalls, and Criteria for Success[J]. Journal of the American College of Radiology, 2018, 15(3):504-508.

[9] Polina M, Lucy O, Yury Y, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare[J]. Oncotarget, 2018, 9(5):5665-5690.

[10] Keel S, Lee P Y, Scheetz J, et al. Feasibility and patient acceptability of a novel artificial intelligence-based screening model for diabetic retinopathy at endocrinology outpatient services: a pilot study[J]. Scientific Reports, 2018, 8(1):4330.

[11] Yeung S, Downing N L, Fei-Fei L, et al. Bedside Computer Vision - Moving Artificial Intelligence from Driver Assistance to Patient Safety[J]. New England Journal of Medicine, 2018, 378(14):1271-1273.

[12] Syam N, Sharma A. Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice[J]. Industrial Marketing Management, 2018, 69(FEB.):135-146.