On the Evolution of Knowledge Graph Abroad and Its Application in Intelligent Education

Yuliu Zhang\(^1\) and Bo Zhao\(^1,2\)

\(^1\) School of Information Science and Technology, Yunnan Normal University, Kunming 650500, YN, China
zyliu97@qq.com, ykzb63@126.com

\(^2\) Key Laboratory of Educational Information for Nationalities, Ministry of Education, Yunnan Normal University, Kunming 650500, YN, China

Abstract. The development of education is more and more dependent on the intelligent learning support services in today’s information age. The intelligent education supported by artificial intelligence (AI) has drawn more attention. In particular, knowledge graph (KG) becomes the key to promote the development and innovation of education with the developing of AI. So, the evolution of the research hotspots of KG is reviewed based on the literature of Web of Science in this study. We explore the development of intelligent education on different aspects, including educational KG, cognitive diagnosis and personalized education.

Keywords: Knowledge graph · Cognitive graph · Intelligent education · Personalized learning

1 Introduction

According to the NMC’s horizon 2019 report, AI is the key of higher education to revolutionize in the next four to five years [1]. As an interdisciplinary research field, with rapid development of AI, KG will support for “AI + education” in the future, and push education towards “intelligent education”. In the realization process of intelligent education, such as semantic search of knowledge, recommendation system of personalized learning, and construction of learner portrait, most of them rely on large-scale KG. From the beginning of Google’s intelligent search engine, to now, big data analytics, chatbots, personalized education and recommendation system, are all closely related to the KG. Therefore, by comprehensively analyzing the evolution of KG, it is great significance to promote innovation in the education. From the perspective of bibliometrics, this study explores the development of intelligent education by analyzing the evolution of KG.

2 Research Design

2.1 Data Collection

In this study, we extracted data with “Knowledge Graph” or “Knowledge Visualization” from the Web of Science database, time range from 1996–2019. By eliminating the
irrelevant literature to KG, a total of 1210 residual literature were identified, which are used as the sample data source.

2.2 Research Methods

Scientific citation data visualization analysis software (CiteSpace) and Bibliographic Items Co-occurrence Matrix Builder (BICOMB) were adopted. Through the Citation Analysis and keyword co-occurrence analysis, hot topic distribution and frontier trend of KG were discussed.

3 Development Context of KG

3.1 Chronological Analysis

With the change of time, the number of literatures is an important indicator to measure the development trend of a field. In BICOMB software, “Year” is used to count the annual number of literatures, which distribution curve is significant (see Fig. 1). We can see it is to increase at a quick rate in recent years from the fitted exponential function of the curve \( y = 2.043e^{0.1321x} \) and the determination coefficient \( R^2 = 0.8286 \), the number of relating literatures.

![Fig. 1. Number of literatures on KG](image)

The foreign research about KG began in the 1950s. In 1955, Garfield [2] pioneered the idea of using citation indexes to retrieve literatures. In 1965, Price [3] proposed the citation relationship between network and scientific literature. In 1968, J. R. Quillian proposed the semantic web. In 1977, Feigenbaum put forward the concept of knowledge engineering, and deem that knowledge engineering is the application of AI, and then ontology was introduced and became the method of representing knowledge in the real world. In 1998, Tim Berners-Lee proposed the concept of Semantic Web. On this basis,
the specification of knowledge description on the World Wide Web was further proposed by the W3C. In 2006, Tim Berners-Lee [4] put forward the concept of linked data to generate large-scale networks based on the relation between entities.

In 2012, Google proposed the “knowledge Graph”, which reflects the enhanced application of large-scale KG in intelligent search engine. The typical representatives of large-scale network knowledge, such as DBpedia [5], are built on the basis of Wikipedia structured knowledge base. With the further development of AI, it has entered the stage of cognitive intelligence, which promotes the development of construction technologies of KG.

4 Research Topic Distribution of KG

4.1 Analysis of High-Frequency Keywords

The keywords of the literature are the extracted from the full text, which can objectively reflect the research hotspots in a certain field [6]. CiteSpace is used to generate the co-occurrence graphs of keywords (see Fig. 2). Co-occurrence of high-frequency keywords.

![Fig. 2. Co-occurrence of high-frequency keywords](image)

The Pathfinder Network algorithm is used to simplify the network for highlight structural features, which are mainly determined by parameter $r$. The relation of triangle inequality is defined as Eq. (1):

$$w_{ij} \leq \left( \sum_{k} w_{nk}n_{k+1}^r \right)^{1/r} \tag{1}$$

Where $w_{ij}$ represents the link weight between node $i$ and node $j$, $w_{nk}n_{k+1}$ represents the link weight between node $n_k$ and $n_{k+1}$ nodes, and $r$ is the Minkowski distance.
In addition to the “Knowledge Graph”, there are the top 10 high-frequency keywords: Ontology (52), Knowledge Visualization (52), Visualization (40), Semantic Web (33), Knowledge Graph Embedding (28), Knowledge Representation (21), Neural Network (16), Link Prediction (16), Dbpedia (15), Recommender System (14). As can be seen from the fluctuation of word frequency, which shows that it focuses on the key technologies of KG construction such as knowledge extraction, knowledge representation, knowledge reasoning and knowledge graph completion.

4.2 Analysis of Clustering

Clustering analysis of keywords can clearly reveal the hot topics in the field, cluster was extracted from the keywords of cited literature. The extraction of cluster uses a Long-likelihood-ratio algorithm, which computed as Eq. (2):

\[
LLR = \log \left\{ p(C_j/V_{ij}) / P(C_j/V_{ij}) \right\}
\] (2)

LLR is the log-likelihood ratio of the word \( W_i \) and \( C_j \), the vector \( V_{ij} (\alpha, \beta, \gamma) \) is composed of the frequency (\( \alpha \)), concentration (\( \beta \)), and dispersion (\( \gamma \)) of the word \( W_i \). The vector \( V_{ij} \) is used to determine whether \( W_i \) can be used as a feature word of category \( C_j \). \( p(C_j/V_{ij}) \) and \( P(C_j/V_{ij}) \) is the density functions of the categories \( C_j \) and \( \overline{C}_j \).

The clustering graph of keywords as shown in Fig. 3, Modularity \( Q = 0.55 \), Mean Silhouette = 0.7309, therefore, the structure of cluster is remarkable and convincing.

![Keywords clustering](image)

Fig. 3. Keywords clustering

A total of 4 clusters were generated. They are #0 (Knowledge Visualization), #1 (Knowledge Representation), #2 (Deep Learning) and #3 (Neural Networks). knowledge visualization, knowledge representation learning, deep learning and neural network are focused on.
4.3 On Research Hotspots of KG

(1) Knowledge visualization. Knowledge visualization is a graphical method to construct and transmit complex knowledge. It uses scene visualization, relationship visualization, and process visualization to create and transfer of knowledge, in order to deepen students’ “memory” of knowledge and promote learners’ cognitive processing. The visual expression has a strong “integrity” characteristic and makes people have a “global consciousness”, which can help learners better understand and learning the knowledge.

(2) Knowledge representation. Knowledge representation is the study of how to represent the knowledge of the objective world in a form that is easy for computer or machine to recognize and understand. RDF triples are mostly used to describe the relationship between entities. In recent years, with the development of deep learning, knowledge representation learning based on entities, concepts, and relationships have become mainstream [7]. Moreover, in knowledge representation, the fusions of cross-media elements and spatio-temporal dimensions are also the trend of future research.

(3) Deep learning. Deep learning (DL) is a kind of machine learning based on deep neural network, which uses statistics to model specific problems in the real world and uses trained data to solve similar problems in the field. It is a semi-theory and semi-empirical model with flexible expression. DL is derived from artificial neural networks, which can be divided into convolutional neural networks (CNN) and deep belief nets (DBN). CNN is a well-known DL architecture inspired by the natural visual perception mechanism of the living creatures [8]. DBN is probabilistic generative models that are composed of multiple layers of stochastic, latent variables [9]. In addition, Google’s TensorFlow framework can used to implement open source DL systems, which can support CNN, RNN and LSTM algorithms, and also provides a tool on the web. Which can use graphics to present the real-time characteristics of the entire network.

(4) Neural networks. Neural networks, often called artificial neural networks (ANN), are based on the basic principles of neural networks in biology, and are rooted in neuroscience, statistics, mathematics, and computer science. The neural network simulates the thinking of the human brain, and can have human-like understanding and judgment, which is a further extension of traditional logic calculations. As an indispensable part of machine learning, neural networks have been widely used in computer vision, decision optimization, image segmentation, cognitive science, and so on. At present, convolutional neural networks (CNN) and recurrent neural networks (RNN) are relatively mature. CNN is a neural network especially suitable for computer vision applications [10], and RNN is a neural network for processing sequence data [11]. The toolkit, such as tensorflow and sklearnt has been applied neural network models to solve practical problems.
5 Exploration of Intelligent Education

5.1 Educational KG

Nowadays, the construction of educational KG is regarded as the key for intelligent education. The educational KG can be constructed to provide students with a variety of knowledge services, such as knowledge query, personalized learning path, etc. When constructing the KG about a specific domain, it is necessary to discuss with the domain experts to customize the schema of the domain KG. However, the researchers mainly focus on the coded explicit knowledge, most of which do not consider the tacit knowledge. Therefore, it is urgent to construct generic and domain KG for comprehensive construction platform.

Some KG construction platforms have been constructed according to the relevant methodologies. For example, Baidu has built K-12 education KG to realize personalized learning, and providing intelligent services for students.

5.2 Cognitive Diagnosis and Cognitive Graph

In terms of education and teaching, intelligent education is faced with the problem of students’ cognitive overload. As mentioned in the Cognitive Load Theory, when the cognitive load is controlled within the range of working memory, effective learning can be happened [12]. How to reduce the cognitive load and set the gradient of students’ learning are the key. However, student’s learning paths can be found and created by cognitive diagnosis and student’s cognitive load can be reduced effectively.

On the other hand, according to Constructivism Learning Theory, learning is the process by which learners actively select and process external information based on their own experience [13]. In order to avoid the emergence of “Lost in learning” problem, it is urgent to build the cognitive graph based on KG through students’ cognitive diagnosis. It has “cognition” first, which can make cognitive diagnosis for students. Then “reasoning”, it is to provide a personalized learning path, which can promote students’ effective learning.

5.3 Personalized Education

With development of technology such as the Internet and AI, traditional educational ideas about effective learning of students have been inherited and surpassed. David Pawl Ausubel believes that students’ learning should be as meaningful as possible if it is valuable [14]. Therefore, the development of personalized education should also aim at intentional, active, real, constructive and cooperative learning.

It is urgent to create a fair, open and personalized education environment by the deep combination of AI and education, which was proposed at the international conference on AI & education in May 2019. In particular, with the coming of 5G era, unstructured data transmission can be achieved through the 5G’s wireless communication technology, which provides a more three-dimensional digital environment for intelligent education [15]. In the future, with the progress of KG and 5G, intelligent education will become the main way of online learning.
Acknowledgments. The research is supported by a National Nature Science Fund Project (No. 61967015), and Undergraduate Education and Teaching Reform Project in Colleges and Universities of Yunnan (JG2018060).

References

1. Alexander, B., Ashford-Rowe, K., et al.: EDUCAUSE Horizon Report: 2019 Higher Education Edition. Louisville, Co.: EDUCAUSE. Accessed 11 Jan 2020
2. Garfield, E.: Citation indexes for science: a new dimension in documentation through association of ideas. Science 122(3159), 108–111 (1955)
3. De Solla Price, D.J.: Networks of scientific papers. Science 149(3683), 510–515 (1965)
4. Berners-Lee T.: Linked data-designissues. http://www.w3.org/DesignIssues/LinkedData.html. Accessed 12 Jan 2020
5. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.: DBpedia: a nucleus for a web of open data. In: Aberer, K., et al. (eds.) ASWC/ISWC - 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-76298-0_52
6. Wang, X.M.: Research on the teaching quality of colleges and universities in China: track, hotspots and future trend – a Citesepace visualization based on fourteen core journals of higher education. Educ. Mon. 1, 91–103 (2018)
7. Liu, Z.Y., Sun, M.S., Lin, Y.K., et al.: Knowledge representation learning: a review. J. Comput. Res. Develop. 53(2), 247–261 (2016)
8. Gu, J.X., Wang, Z., et al.: Recent advances in convolutional neural networks. Pattern Recogn. 77, 354–377 (2018)
9. Hinton, G., E.: Deep belief networks. Scholarpedia 4(6), 5947 (2009)
10. Zeiler, Matthew D., Fergus, Rob: Visualizing and understanding convolutional networks. In: Fleet, David, Pajdla, Tomas, Schiele, Bernt, Tuytelaars, Tinne (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
11. Graves, A., Mohamed, A.R., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, Vancouver, BC, pp. 6645–6649 (2013)
12. Paas, F., Renkl, A., Sweller, J.: Cognitive load theory: instructional implications of the interaction between information structures and cognitive architecture. Learn. Sci. 32(1/2), 1–8 (2004)
13. Ültanir, E.: An epistemological glance at the constructivist approach: constructivist learning in Dewey, Piaget, and Montessori. Int. J. Instr. 5(2), 195–212 (2012)
14. Ausubel, D.P., Cook, H.: Educational psychology: a cognitive view. Am. J. Psychol. 83(2), 303 (1970)
15. Lewis, J.A.: How 5G will Shape Innovation and Security: A Primer. CSIS, Washington (2018)