Comparative Study of Knowledge Graph Models in Education Domain

Zineb Elkaimbillah[1], Maryem Rhanoui[2], Mounia Mikram[3] and Bouchra El Asri[1]
[1] *IMS Team, ADMIR Laboratory, Rabat IT Center, ENSIAS, Mohammed V University in Rabat, Morocco*
[2] *Meridian Team, LYRICA Laboratory, School of Information Sciences, Rabat, Morocco*
[3] *LRIT Laboratory, Rabat IT Center, Faculty of Sciences, Mohammed V University in Rabat, Morocco*

Keywords: Knowledge Graph, Educational Knowledge Graph, Knowledge Graph Embedding, Knowledge Graph Application

Abstract: Knowledge graph (KG) technologies are improving Artificial Intelligence. It can effectively expand the breadth of search results. Therefore, KGs continue to solve several problems in different domains, including the education field. The application of educational KGs to learning systems has recently been expanded due to increased demand in the education sector and the importance of KGs application to learning systems. In this article, we present the knowledge Graph approach, the methodology of KG development, and analyze each step. Also, we discuss the popular KG Embedding models. We provide a comparative study of KG models in the education field.

1 INTRODUCTION

The coronavirus pandemic 2019 (COVID-19) had an impact on several fields, especially the field of education. Therefore, efforts are being made to force educational systems into this vital sector. The complementarity of bottom-up (machine learning-driven) and top-down (semantic and knowledge graph) techniques is essential for representation insights from such data. Recently, the techniques of machine learning for the representation of knowledge are rapidly improving. As a result, the field of education has expanded by integrating the data models of knowledge graphs. This context is provided in a linking framework between semantic metadata, for data sharing and integration. For example, in semantic-aware Question Answering (QA) services, semantic information may be used to improve search results. (Berant et al., 2013)(Fader et al., 2014)(Yao and Van Durme, 2014), Information Retrieval(Liu et al., 2018)(Liu and Fang, 2015), and Recommender System(Bellini et al., 2017)(Wang et al., 2019).

The objective of this article is to present the general context of the Knowledge Graph and the process of its development. On the other hand, we hope to provide a comparative study of knowledge graphs applied in the education domain. We identify the specific use and application of the KG, different data sources used, and the techniques and models for integrating the KG.

We organize the article as follows. Section 2 presents definitions of knowledge graphs, followed by a discussion of various notable approaches to knowledge graph development by presenting the knowledge extraction aspect among their concepts and related approaches, and we provide popular KG embedding models. Finally, a comparative study of KG in education is presented in Section 3.

2 BACKGROUND

In this section, we will approach some definitions of Knowledge Graph (KG), then we will present the methodology of development of a KG, and then we will define what is Knowledge Graph Embedding by citing its most known models.

2.1 Knowledge Graph Definition

Before presenting the different definitions of knowledge graphs, we start with a reminder of two important concepts "knowledge" and "graph". The first one means the understanding of something such as facts, skills, or objects. According to other points of view, knowledge is acquired from many different data sources and in different ways. The second concept is mathematically defined as a structure of

Elkaimbillah, Z., Rhanoui, M., Mikram, M. and El Asri, B.
Comparative Study of Knowledge Graph Models in Education Domain.
DOI: 10.5220/0010733800003101
In *Proceedings of the 2nd International Conference on Big Data, Modelling and Machine Learning (BML 2021)*, pages 339-344
ISBN: 978-989-758-559-3
Copyright © 2022 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

objects in which pairs of objects are called vertices, nodes, or points are linked together by edges, links, or lines.

In 2012, Google proposed a Knowledge Graph, to use a Graph in their search engine. Many authors have proposed several definitions of this concept. Notable among these have already been, formally,(Tran and Takasu, 2019) defines the knowledge graph as a relation *r* between the main entity *h* and the tail entity *t*. This definition is not complete enough as it lacks the semantic level.

On the other hand, (Xiong, 2018) declared the semantic level, defining KG as a data resource that contains entities, each entity has its specific semantics and meaning.

(Vidal et al., 2019) presented the knowledge graph as a representation capable of building formal models of levels of abstraction and facts of different types.

Another effort is made to define a KG was proposed by (Bonatti et al., 2019). The authors consider it like an organized schema and several labeled relations with a formal meaning are available.

2.2 Knowledge Graph Development

Several data sources are used in the construction of knowledge graphs, such as small and large databases, text documents, web pages, and crowdsourced statements, and in either a manual, automatic, or semi-automatic way (Bonatti et al., 2019).
The methodology of creating a Knowledge Graph consists of two major components:
The top-down section: covers the process of modeling a specific field and contains the steps below : (Fensel et al., 2020):

- **Mapping step:** describes a set of predefined mapping rules on an incoming data map for a specific domain.
- **Annotation development:** This phase involves converting specified domain standards into a form interface to make manual and semi-automated knowledge acquisition procedures easier.

The bottom-up section: which explains the process of annotation in a new domain and contains the steps listed below (Fensel et al., 2020):

- **knowledge acquisition:** First, we start this process by defining the domain (e.g., education). The investigation of the domain entails knowledge extraction from data sources (structured, semi-structured, and unstructured).
- **Knowledge extraction:** describe types (Entity, Relation, Attribute), approaches, and tools aspects (Zhao et al., 2018). In terms of knowledge extraction approaches, it heavily involves NLP, text mining, and Machine learning. Examples of techniques used for entities recognition and relations extraction are Conditional Random Field (CRF)(Su et al., 2009), Machine Learning models (e.g., SVM), (BiLSTM) (Huang et al., 2015), Hidden Markov Models (HMM)(Morwal et al., 2012).
- **Knowledge fusion**: is a process that builds an ontology and assesses its quality in an iterative way (Zhao et al., 2018).
- **Knowledge storage:** This step consists of using two main types of storage, the first one, based on RDF (Resource Description Framework), and the second one, based on graphical databases. But in general, the storage is done in NoSQL databases (Zhao et al., 2018).

2.3 Knowledge Graph Embedding

Knowledge graph embeddings (KGEs) are the process of constructing propositional feature representations of entities and relations in vectors in a knowledge graph (Wang et al., 2017) to apply numeric techniques resulting in scalable besides effectiveness.
The knowledge graph embeddings are computed to fulfill specific properties; i.e., they follow given KGE models to calculate the spatial distance between two entities for the type of relationship in the low dimensional plunge vector space.

Several popular KGE models have been widely studied :

- Translating Embeddings for Modeling Multi relational Date (TransE): This is the first translation model to get the tail vector as near to the sum of the head and relation vectors as feasible. (Bordes et al., 2013).
- Knowledge Graph Embedding by Translating on Hyperplanes(TransH): The goal of transH is to reduce the model's complexity and the difficulty of training by dealing with all possible relationships (Wang et al., 2014).
- Learning Entity and Relation Embeddings for Knowledge Graph Completion(TransR): TransR can cover the 1-to-1 relationship, by separating the relational space from the entity space (Lin et al., 2015).

There are also more examples of KGE models include DistMult (Yang et al., 2014), Complex Embeddings (ComplEx) (Trouillon et al., 2016), Holographic Embeddings (HolE) (Nickel et al., 2016), Convolutional 2D KG Embeddings (ConvE) (Dettmers et al., 2018), Convolution-based model (ConvKB) (Nguyen et al., 2017). All KGE models have advantages and disadvantages, so it is reasonable to expect that a discover a more efficient way to represent knowledge.

3 COMPARATIVE STUDY

KGs have widely been used to benefit a variety of applications related to learning systems and a wealth of pedagogical data.

K12EduKG is a knowledge graphing system for K-12 educational topics. The K12EduKG system's architecture, which is made up of two components: the concept extraction and the relationship identification are both used to extract educational concepts for specific subjects. It takes Named Entity Recognition (NER) methods on educational data using the CRF model and probabilistic association rule mining (Chen et al., 2018a).

Using the object data of internal control policy documents, (Wang, 2020) proposed a knowledge graph construction method for internal control in higher education institutions.

(Chen et al., 2018b) provided a KG can be used for school topics and online courses to educate and learn. To extract instructional concepts, this study employs recurrent neural network (RNN) models on pedagogical data. To identify educational relations that interconnecting instructional concepts based on student performance, They used the probabilistic association rule mining method.

(Aliyu et al., 2020) present an automated approach based on a knowledge graph to address the problem of tertiary institution departments manually arranging course allocations, and handling the task of Question Answering to education administrators. To provide intelligent knowledge services, the system stores data in a knowledge graph in RDF/XML format.

The improvement of the quality of teaching and basic concepts of some computer science disciplines is the objective of (Qin et al., 2020). Building and visualizing educational knowledge as research content using a huge quantity of course-related information from the database as an example. It begins with obtaining database knowledge, cleaning and preparing the acquired data, and then utilizing several automatic or semi-automatic technological techniques to extract information, identify knowledge units from text datasets, and then extract entities, attributes, and relations. Finally, the collected knowledge is saved in neo4j to achieve knowledge graph visualization, producing an effective educational knowledge graph of database discipline to aid smart education.

MathGraph (Zhao et al., 2019) is a significant effort in educational KG development for automatically solving high school mathematical exercises. It is constructed using the crowdsourcing technique to represent various mathematical objects.

(Yao et al., 2020) concentrate on this issue and propose a model for embedding learning in educational knowledge graphs. They present a structural and literal embedding representation based on TransE and Bert. They use three GRUs to combine both methods, also using Knowledge Fores and Wikipedia as a data source for the construction of the educational KG.

We summarize the most important points for this comparative study in Table 1 that represent an overview of the KG approach in education, citing the models end techniques embedding used, data source, evaluation measure(s), and the limitations that can be an improvement in the future research.

4 DISCUSSION

Analysis of KG construction methodologies derived from scholarly publications in the field of education reveals a connected set of limitations and flaws related to the KG data type as several approaches have been collected as textual data and the disregard of using visual data. On the other hand, the majority of these research either did not define the method(s) utilized in the KG building, and strategies for extracting entities and relations or did not provide any information on the algorithm(s) used to create the KG.

To assess and enhance this work, we aim to respond to all the problems detected such as by including a comprehensive method for knowledge extraction, construction, and analysis of educational graphs using state-of-the-art models of natural language processing (NLP), namely named entity extraction, relation extraction, and coreference resolution and methodological extensions of KG-based multimodal (textual and visual data) extraction and analysis algorithms.

Table 1: An outline of KG models applied in the education field

Ref	KG Usage	Models	Data Source	Evaluation Measure(s)	Limitation
(Wang, 2020)	Information Retrieval and recommendation	pTransE	CNKI database, internal control policy documents	mean Silhouette	• Limited data sources, • KG embedding model not significant
(Aliyu et al., 2020)	Managing and allocating courses, Question Answering	Adhoc	Educational information system,	Case study	• Lack of proper evaluation, • limited KG resources
(Qin et al., 2020)	Improvement of the quality of teaching and basic concepts of some computer science disciplines, Question Answering	BILSTM-CRF, word2vec	Course-related information of the database	Measure Recall Precision	• Insufficient evaluation
(Yao et al., 2020)	Educational Technology and Linked prediction	TransE and BERT	Knowledge Forest, Wikipedia	Mean Rank Hits@10	• Application in a limited context, • Limited resources used for KG construction
(Zhao et al., 2019)	Solving mathematical exercises in high school	Complex, Conic and Solid	domain experts and Crowdsourcing	F1 measure	• Application in a limited context, • KG construction required a limited number of resources
(Chen et al., 2018a)	system for building knowledge graphs for K-12	probabilistic association rule mining, CRF	Chinese mathematics curriculum	AUC	• Inadequate assessment metric • A restricted application on the KG was proposed.
(Chen et al., 2018b)	Teaching and learning on school subjects and online courses institutions	RNN, mining probabilistic association rules	Pedagogical data, evaluation of learning	AUC Precision Recall F1 measure	• Limited demonstration of the KG for mathematics construction , • Limited on textual data

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

5 CONCLUSIONS

Knowledge Graph (KG) has a big impact on the education field, practice, improvement of the quality of teaching, and solving high school exercises. Overall, we conclude that KGs are capable of providing semantically organized data.

In this paper, We discussed how knowledge graphs can be used in a variety of domains, including Question Answering, Recommendation, and Information Retrieval. Also, we presented a background for the KG approach, which includes KG definitions, two methods of knowledge graph construction: top-down and bottom-up, and the presentation of KG Embeddings models. A comparison of different models of knowledge graphs utilized in the field of education was offered.

We intend to expand this research in the future by incorporating educational applications and methodological extensions of KG-based algorithms for multimodal extraction and analysis.

REFERENCES

Aliyu, I., Kana, A., and Aliyu, S. (2020). Development of knowledge graph for university courses management. *International Journal of Education and Management Engineering*, 10(2):1.

Bellini, V., Anelli, V. W., Di Noia, T., and Di Sciascio, E. (2017). Auto-encoding user ratings via knowledge graphs in recommendation scenarios. In *Proceedings of the 2nd Workshop on Deep Learning for Recommender Systems*, pages 60–66.

Berant, J., Chou, A., Frostig, R., and Liang, P. (2013). Semantic parsing on freebase from question-answer pairs. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1533–1544.

Bonatti, P. A., Decker, S., Polleres, A., and Presutti, V. (2019). Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). In *Dagstuhl Reports*, volume 8. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

Bordes, A., Usunier, N., Garcia-Duran, A., Weston, J., and Yakhnenko, O. (2013). Translating embeddings for modeling multi-relational data. In *Neural Information Processing Systems (NIPS)*, pages 1–9.

Chen, P., Lu, Y., Zheng, V. W., Chen, X., and Li, X. (2018a). An automatic knowledge graph construction system for k-12 education. In *Proceedings of the Fifth Annual ACM Conference on Learning at Scale*, pages 1–4.

Chen, P., Lu, Y., Zheng, V. W., Chen, X., and Yang, B. (2018b). Knowedu: a system to construct knowledge graph for education. *Ieee Access*, 6:31553–31563.

Dettmers, T., Minervini, P., Stenetorp, P., and Riedel, S. (2018). Convolutional 2d knowledge graph embeddings. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32.

Fader, A., Zettlemoyer, L., and Etzioni, O. (2014). Open question answering over curated and extracted knowledge bases. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1156–1165.

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., and Wahler, A. (2020). *Knowledge Graphs*. Springer.

Huang, Z., Xu, W., and Yu, K. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint arXiv:1508.01991*.

Lin, Y., Liu, Z., Sun, M., Liu, Y., and Zhu, X. (2015). Learning entity and relation embeddings for knowledge graph completion. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 29.

Liu, X. and Fang, H. (2015). Latent entity space: a novel retrieval approach for entity-bearing queries. *Information Retrieval Journal*, 18(6):473–503.

Liu, Z., Xiong, C., Sun, M., and Liu, Z. (2018). Entity-duet neural ranking: Understanding the role of knowledge graph semantics in neural information retrieval. *arXiv preprint arXiv:1805.07591*.

Morwal, S., Jahan, N., and Chopra, D. (2012). Named entity recognition using hidden markov model (hmm). *International Journal on Natural Language Computing (IJNLC)*, 1(4):15–23.

Nguyen, D. Q., Nguyen, T. D., Nguyen, D. Q., and Phung, D. (2017). A novel embedding model for knowledge base completion based on convolutional neural network. *arXiv preprint arXiv:1712.02121*.

Nickel, M., Rosasco, L., and Poggio, T. (2016). Holographic embeddings of knowledge graphs. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 30.

Qin, Y., Cao, H., and Xue, L. (2020). Research and application of knowledge graph in teaching: Take the database course as an example. In *Journal of Physics: Conference Series*, volume 1607, page 012127. IOP Publishing.

Su, K.-Y., Su, J., Wiebe, J., and Li, H. (2009). Proceedings of the joint conference of the 47th annual meeting of the acl and the 4th international joint conference on natural language processing of the afnlp. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP*.

Tran, H. N. and Takasu, A. (2019). Analyzing knowledge graph embedding methods from a multi-embedding interaction perspective. *arXiv preprint arXiv:1903.11406*.

Trouillon, T., Welbl, J., Riedel, S., Gaussier, É., and Bouchard, G. (2016). Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, pages 2071–2080. PMLR.

Vidal, M.-E., Endris, K. M., Jazashoori, S., Sakor, A., and Rivas, A. (2019). Transforming heterogeneous data into knowledge for personalized treatments—a use case. *Datenbank-Spektrum*, 19(2):95–106.

Wang, H., Zhang, F., Zhao, M., Li, W., Xie, X., and Guo, M. (2019). Multi-task feature learning for knowledge graph enhanced recommendation. In *The World Wide Web Conference*, pages 2000–2010.

Wang, J. (2020). Knowledge graph analysis of internal control field in colleges. *Tehnički vjesnik*, 27(1):67–72.

Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.

Wang, Z., Zhang, J., Feng, J., and Chen, Z. (2014). Knowledge graph embedding by translating on hyperplanes. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 28.

Xiong, C. (2018). *Text representation, retrieval, and understanding with knowledge graphs*. PhD thesis, Ph. D. Dissertation. Carnegie Mellon University.

Yang, B., Yih, W.-t., He, X., Gao, J., and Deng, L. (2014). Embedding entities and relations for learning and inference in knowledge bases. *arXiv preprint arXiv:1412.6575*.

Yao, S., Wang, R., Sun, S., Bu, D., and Liu, J. (2020). Joint embedding learning of educational knowledge graphs. In *Artificial Intelligence Supported Educational Technologies*, pages 209–224. Springer.

Yao, X. and Van Durme, B. (2014). Information extraction over structured data: Question answering with freebase. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 956–966.

Zhao, T., Chai, C., Luo, Y., Feng, J., Huang, Y., Yang, S., Yuan, H., Li, H., Li, K., Zhu, F., et al. (2019). Towards automatic mathematical exercise solving. *Data Science and Engineering*, 4(3):179–192.

Zhao, Z., Han, S.-K., and So, I.-M. (2018). Architecture of knowledge graph construction techniques. *International Journal of Pure and Applied Mathematics*, 118(19):1869–1883.

SCIENCE AND TECHNOLOGY PUBLICATIONS