Open Research Online

Editorial of the Special Issue on Deep Learning and Knowledge Graphs

Journal Item

For guidance on citations see FAQs.

oro.open.ac.uk

Editorial of the Special Issue on Deep Learning and Knowledge Graphs

Mehwish Alam [a,b,*], Davide Buscaldi [c], Michael Cochez [d], Francesco Osborne [e],
Diego Reforgiato Recupero [f] and Harald Sack [a,b]

[a] *FIZ-Karlsruhe, Leibniz Institute for Information Infrastructure, Karlsruhe, Germany*
[b] *Karlsruhe Institute of Technology, Karlsruhe, Germany*
E-mails: mehwish.alam@kit.edu, harald.sack@kit.edu
[c] *LIPN, Université Sorbonne Paris Nord, France*
E-mail: buscaldi@lipn.fr
[d] *University of Amsterdam, the Netherlands*
E-mail: m.cochez@vu.nl
[e] *KMi, The Open University, UK*
E-mail: francesco.osborne@open.ac.uk
[f] *Department of Mathematics and Computer Science, University of Cagliari, Italy*
E-mail: diego.reforgiato@unica.it

Editors: Pascal Hitzler, Kansas State University, USA; Krzysztof Janowicz, University of California, Santa Barbara, USA

1. Preface

Over the past years, there has been a rapid growth in the use and the importance of Knowledge Graphs (KGs) along with their application to many important tasks such as entity linking, recommender systems [8], etc. KGs are large networks of real-world entities described in terms of their semantic types and their relationships to each other. On the other hand, Deep Learning has also become an important area of research, achieving important breakthroughs in various research fields, especially Natural Language Processing (NLP) and Image Processing. Consequently, in recent years there have been several studies that combine Deep Learning methods with KGs. For example 1) knowledge representation learning techniques [6] aimed at embedding entities and relations in a KG into a dense and low-dimensional vector space, 2) relation extraction techniques, aimed at extracting facts and relations from the text for automatically generating KGs, 3) entity linking techniques, aimed at completing KGs, 4) using KGs as an additional prior for image recognition, etc.

2. Aims

The web contains a huge amount of information and is still expanding at a fast rate. KGs play a fundamental role in structuring and making available this information. For this reason, several big players such as Google, Facebook, and Amazon have adopted this powerful technology. A KG is a semantic network with real-world entities (organizations,

---

*Corresponding author. E-mail: mehwish.alam@kit.edu.

people, events, places) and illustrates the relationships between them. They have entity pairs that can be traversed to uncover meaningful connections which cannot be revealed in unstructured data.

Bringing KGs and machine and deep learning together [1,3–5] can systematically improve the accuracy of systems and extend the range of machine and deep learning capabilities. Leveraging KGs improves the explainability and trustworthiness of machine and deep learning models. Besides, KGs can be exploited to augment training data in cases where we have insufficient data. One more advantage provided by KGs is that they can help explain the predictions of the machine and deep learning models by mapping explanations to proper nodes in the graph.

Therefore, in order to pursue more advanced methodologies, it has become critical that the communities related to Deep Learning, KGs, and NLP join their forces in order to develop more effective algorithms and applications.

As an example, within the scholarly domain, KGs are widely adopted to support a variety of intelligent services to analyze the scientific research literature and forecast the research dynamics. Sometimes, they suffer from incompleteness such as missing affiliations, references, etc., and this led to the development of KG Embeddings and the application of different link prediction techniques [14,15]. Link prediction within a KG is the task to find a set of links between entities (nodes) of the KG which might provide more knowledge and fill potential gaps of information. The creation of KG Embeddings and novel link prediction techniques heavily rely on cutting-edge approaches to Deep Learning [19].

This special issue aims to reinforce the relationships between these communities and foster interdisciplinary research in the areas of KG, Deep Learning, and Natural Language Processing. The works that we have requested from authors should provide relevant research in the area of KGs, deep learning techniques, and/or their combinations for real and innovative applications. The topics of interest include, but were not limited to:

– New approaches for the combination of Deep Learning and KGs

  ∗ Methods for generating KG (node) embeddings Scalability issues
  ∗ Temporal KG Embeddings
  ∗ Novel approaches

– Applications of the combination of Deep Learning and KGs

  ∗ Recommender Systems leveraging KGs
  ∗ Link Prediction and completing KGs
  ∗ Ontology Learning and Matching exploiting KG-Based Embeddings
  ∗ KG-Based Sentiment Analysis
  ∗ Natural Language Understanding/Machine Reading
  ∗ Question Answering exploiting KGs and Deep Learning
  ∗ Entity Linking
  ∗ Trend Prediction based on KG Embeddings
  ∗ Domain-Specific KG (e.g., Scholarly, Biomedical, Musical)
  ∗ Applying KG embeddings to real-world scenarios.

3. Content

The special issue was able to attract 13 submissions covering relevant areas of research, i.e., deep learning and KGs. We had over 28 different affiliations of authors submitting their work. Out of which 11 papers were accepted after two rounds of reviews indicating an acceptance rate of 84%. Each paper was reviewed by 3 expert reviewers. The accepted papers include two surveys on transfer learning using KGs and on neural entity linking models based on deep learning. The other accepted papers discuss semantic table interpretation, answer selection, network representation learning method, taxonomy enrichment, identification methods for the emergence of new topics, a comparison between KG embedding for data mining and link prediction, discovering alignment relations within the biomedical domain, learning MIDI embeddings to predict music metadata, and the prediction of adverse biological effects of chemicals. All the papers above made use of KGs in different forms. In the following, we will provide a broad overview of all the accepted papers.

The first paper, "Neural Entity Linking: A Survey of Models Based on Deep Learning" [18], by Ozge Sevgili, Artem Shelmanov, Mikhail Arkhipov, Alexander Panchenko, and Chris Biemann, is a survey about neural entity linking systems proposed since 2015. The authors made a systematic comparison of their performance of them on state-of-the-art benchmarks. Moreover, the authors divided those methods by generic architectural components and grouped them by several common themes. Finally, the authors introduced a discussion of popular embedding techniques as they are often leveraged by many neural models and discussed recent use-cases of entity linking.

The second paper, "A Survey on Visual Transfer Learning using Knowledge Graphs" [11], by Sebastian Monka, Lavdim Halilaj, and Achim Rettinger is a comprehensive analysis of how the rising field of transfer learning is taking advantage of KGs. Specifically, KGs are typically used for representing auxiliary knowledge either in an underlying graph-structured schema or in a vector-based KG embedding. The survey also provides an overview of KG embedding methods and describes several joint training objectives suitable to combine them with high dimensional visual embeddings.

The third paper, "Tab2KG: Semantic Table Interpretation with Lightweight Semantic Profiles" [7] by Simon Gottschalk and Elena Demidova, presents Tab2KG, a new method for automatically inferring tabular data semantics and transforming such data into a data graph. This solution uses a one-shot learning approach that relies on lightweight semantic profiles to map a tabular dataset containing previously unseen instances to a domain ontology. Tab2KG outperforms state-of-the-art semantic table interpretation baselines on several real-world datasets from different application domains.

The fourth paper, "Answer Selection in Community Question Answering Exploiting Knowledge Graph and Context Information" [2] by Golshan Afzali Boroujenia, Heshaam Failia, and Yadollah Yaghoobzadeha present a novel answer selection method that takes advantage of the knowledge embedded in KGs. This solution uses variational autoencoders in a multi-task learning process with a classifier to produce class-specific representations of the answers. The method outperforms significantly the existing baselines on three widely used datasets.

The fifth paper, "Network representation learning method embedding linear and nonlinear network structures" [20] by Hu Zhang, Jingjing Zhou, Ru Li, and Fan Yue proposed an enhancement to the Hierarchical Graph Convolutional Networks model to learn network representations. Their method is based on unsupervised joint learning with shallow and deep learning models. The first is used to extract features from nodes and the second to obtain structural features by aggregating information from neighboring nodes. Their method improves the results previously obtained with Hierarchical Graph Convolutional Networks.

The sixth paper, "Taxonomy Enrichment with Text and Graph Vector Representations" [16], by Irina Nikishina, Mikhail Tikhomirov, Varvara Logacheva, Yuriy Nazarov, Alexander Panchenko, and Natalia Loukachevitch, targets the problem of taxonomy enrichment which aims at adding new words to the existing taxonomy. This paper provides a comprehensive study of the existing approaches to taxonomy enrichment based on word and graph vector representations. This study also explores how deep learning architectures can be used to extend the taxonomic backbones of KGs.

The seventh paper, "Analyzing the generalizability of the network-based topic emergence identification method" [9] by Sukhwan Jung, and Aviv Segev, analyzed the topic evolution method with the task to predict new topics. The method is general and can work in any collection of data where the topics are defined by their neighbors' previous relationships. Twenty sample topic networks were built within different domains such as business, materials, diseases, and computer science from the Microsoft Academic Graph dataset.

The eighth paper, "Knowledge Graph Embedding for Data Mining vs. Knowledge Graph Embedding for Link Prediction – Two Sides of the same Coin?" [17] by Jan Portisch, Nicolas Heist, and Heiko Paulheim, examines two tasks: encoding provision for data mining tasks and predicting links in a KG. The authors mentioned that the two tasks are related and showed that a set of approaches can be used for both. In particular, authors explored link-prediction based embeddings for other downstream tasks based on similarity and proposed a link prediction method based on node embedding techniques such as RDF2vec. Finally, the authors drew a list of remarks on which method should be used and in which condition.

The ninth paper, "Discovering alignment relations with Graph Convolutional Networks: a biomedical case study" [12] by Pierre Monnin, Chedy Raïssi, Amedeo Napoli, and Adrien Coulet. The article proposes to match nodes within a KG by learning node embeddings with Graph Convolutional Networks and by clustering nodes based on their embeddings, in order to suggest alignment relations between nodes of the same cluster. The authors

first applied inference rules associated with domain knowledge, independently or combined, before learning node embeddings, and measured the improvements in matching results. Then the authors observed that distances in the embedding space are coherent with the "strength" of these different relations.

The tenth paper, "MIDI2vec: Learning MIDI Embeddings for Reliable Prediction of Symbolic Music Metadata" [10] from Pasquale Lisena, Albert Merono-Penuela, and Raphael Troncy proposed a novel approach to exploit graph embedding techniques to represent MIDI files as vectors. MIDI data are thus mapped as graphs that include information about tempo, time signature, programs, and notes. Authors employed node2vec to generate embeddings and prove that the resulting vectors can be successfully leveraged to predict the musical genre and other metadata such as the composer, the instrument, or the movement.

The eleventh paper, "Prediction of Adverse Biological Effects of Chemicals Using Knowledge Graph Embeddings" [13] by Erik Bryhn Myklebust, Ernesto Jimenez-Ruiz, Jiaoyan Chen, Raoul Wolf, and Knut Erik Tollefsen. The paper proposes a KG based on major data sources used in ecotoxicological risk assessment. The authors apply this KG to an important task in risk assessment, namely chemical effect prediction. In order to do so, the authors have used nine KG embedding models for this prediction task where the authors conclude that using KG embeddings can increase the accuracy of effect prediction.

References

[1] S.B. Abbès, R. Hantach, P. Calvez, D. Buscaldi, D. Dessì, M. Dragoni, D.R. Recupero and H. Sack (eds), *Joint Proceedings of the 2nd International Workshop on Deep Learning Meets Ontologies and Natural Language Processing (DeepOntoNLP 2021) & 6th International Workshop on Explainable Sentiment Mining and Emotion Detection (X-SENTIMENT 2021) Co-Located with 18th Extended Semantic Web Conference 2021*, Hersonissos, Greece, June 6th–7th, 2021 (moved online), CEUR Workshop Proceedings, Vol. 2918, CEUR-WS.org, 2021. ISSN 1613-0073. http://ceur-ws.org/Vol-2918.

[2] G. Afzali, H. Faili and Y. Yaghoobzadeh, Answer selection in community question answering exploiting knowledge graph and context information, *Semantic Web* **13** (2022), 339–356. doi:10.3233/SW-222970.

[3] M. Alam, D. Buscaldi, M. Cochez, F. Osborne, D.R. Recupero and H. Sack (eds), *Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2019) Co-Located with the 16th Extended Semantic Web Conference 2019 (ESWC 2019)*, Portoroz, Slovenia, June 2, 2019, CEUR Workshop Proceedings, Vol. 2377, CEUR-WS.org, 2019. ISSN 1613-0073. http://ceur-ws.org/Vol-2377.

[4] M. Alam, D. Buscaldi, M. Cochez, F. Osborne, D.R. Recupero and H. Sack (eds), *Proceedings of the Workshop on Deep Learning for Knowledge Graphs (DL4KG2020) Co-Located with the 17th Extended Semantic Web Conference 2020 (ESWC 2020)*, Heraklion, Greece, June 2, 2020 (moved online), CEUR Workshop Proceedings, Vol. 2635, CEUR-WS.org, 2020. ISSN 1613-0073. http://ceur-ws.org/Vol-2635.

[5] M. Alam, A. Fensel, J.M. Gil, B. Moser, D.R. Recupero and H. Sack, Special issue on machine learning and knowledge graphs, *Future Gener. Comput. Syst.* **129** (2022), 50–53. doi:10.1016/j.future.2021.11.022.

[6] G.A. Gesese, R. Biswas, M. Alam and H. Sack, A survey on knowledge graph embeddings with literals: Which model links better literal-ly?, *Semantic Web* **12**(4) (2021), 617–647. doi:10.3233/SW-200404.

[7] S. Gottschalk and E. Demidova, Tab2KG: Semantic table interpretation with lightweight semantic profiles, *Semantic Web* **13** (2022), 571–597. doi:10.3233/SW-222993.

[8] A. Iana, M. Alam and H. Paulheim, A survey on knowledge-aware news recommender systems, *Semantic Web Journal* (2022).

[9] S. Jung and A. Segev, Analyzing the generalizability of the network-based topic emergence identification method, *Semantic Web* **13** (2022), 423–439. doi:10.3233/SW-212951.

[10] P. Lisena, A. Meroño-Pe nuela and R. Troncy, MIDI2vec: Learning MIDI embeddings for reliable prediction of symbolic music metadata, *Semantic Web* **13** (2022), 357–377. doi:10.3233/SW-210446.

[11] S. Monka, L. Halilaj and A. Rettinger, A survey on visual transfer learning using knowledge graphs, *Semantic Web* **13** (2022), 477–510. doi:10.3233/SW-212959.

[12] P. Monnin, C. Raïssi, A. Napoli and A. Coulet, Discovering alignment relations with Graph Convolutional Networks: A biomedical case study, *Semantic Web* **13** (2022), 379–398. doi:10.3233/SW-210452.

[13] E.B. Myklebust, E. Jimenez-Ruiz, J. Chen, R. Wolf and K.E. Tollefsen, Prediction of adverse biological effects of chemicals using knowledge graph embeddings, *Semantic Web* **13** (2022), 299–338. doi:10.3233/SW-222804.

[14] M. Nayyeri, G.M. Cil, S. Vahdati, F. Osborne, A. Kravchenko, S. Angioni, A. Salatino, D. Reforgiato Recupero, E. Motta and J. Lehmann, Link prediction of weighted triples for knowledge graph completion within the scholarly domain, *IEEE Access* **9** (2021), 116002–116014. doi:10.1109/ACCESS.2021.3105183.

[15] M. Nayyeri, G.M. Cil, S. Vahdati, F. Osborne, M. Rahman, S. Angioni, A.A. Salatino, D.R. Recupero, N. Vassilyeva, E. Motta and J. Lehmann, Trans4E: Link prediction on scholarly knowledge graphs, *Neurocomputing* **461** (2021), 530–542. doi:10.1016/j.neucom.2021.02.100.

[16] I. Nikishina, M. Tikhomirov, V. Logacheva, Y. Nazarov, A. Panchenko and N. Loukachevitch, Taxonomy enrichment with text and graph vector representations, *Semantic Web* **13** (2022), 441–475. doi:10.3233/SW-212955.

[17] J. Portisch, N. Heist and H. Paulheim, Knowledge graph embedding for data mining vs. knowledge graph embedding for link prediction – Two sides of the same coin?, *Semantic Web* **13** (2022), 399–422. doi:10.3233/SW-212892.

[18] L. Sevgili, A. Shelmanov, M. Arkhipov, A. Panchenko and C. Biemann, Neural entity linking: A survey of models based on deep learning, *Semantic Web* **13** (2022), 527–570. doi:10.3233/SW-222986.

[19] M. Wang, L. Qiu and X. Wang, A survey on knowledge graph embeddings for link prediction, *Symmetry* **13**(3) (2021). https://www.mdpi.com/2073-8994/13/3/485. doi:10.3390/sym13030485.

[20] H. Zhang, J. Zhou, R. Li and Y. Fan, Network representation learning method embedding linear and nonlinear network structures, *Semantic Web* **13** (2022), 511–526. doi:10.3233/SW-212968.