State-of-the-Art Instance Matching Methods for Knowledge Graphs

Alex Boyko[1,2][0000−0003−3592−4986], Siamak Farshidi[2][0000−0001−6139−921X], and Zhiming Zhao[2][0000−0002−6717−9418]

[1] Vrije Universiteit Amsterdam, The Netherlands
o.y.boyko@student.vu.nl
[2] Universiteit van Amsterdam, The Netherlands
{s.farshidi, z.zhao}@uva.nl

Abstract. Instance matching has attracted a wide range of research attentions. A systematic literature review captures knowledge regarding the state-of-the-art systematically, to analyze and report it in the form of reusable knowledge. It is difficult to compare the performance of different instance matching methods, even when the same benchmarking dataset is used.

1 Introduction

Instance matching identifies instances from different data sources that refer to the same real-world entity [7]. It is difficult to identify the state-of-the-art solution since the every instance matching approach is tailored for specific data and its properties.

2 Overview of the State of the Art Approaches

Existing knowledge graph instance matching approaches rely on embeddings to represent data in the form of vectors. They aim to capture the semantic meaning of the data by placing semantically similar inputs close together in the vector space. Figure 1 shows the output of a literacture study on instance matching approaches.

3 Conclusion

This study reviews the latest instance matching research. The approaches are often tested on different benchmarking datasets. Importantly, even when the same dataset is used, there is no single instance matching approach that performs well with every metric. Since some methods perform well on one subset and worse on the others, it is difficult to compare their performance. Future research can possibly explore ways for reducing human feedback while preserving its benefits, for example by combining supervised and self-supervised approaches.

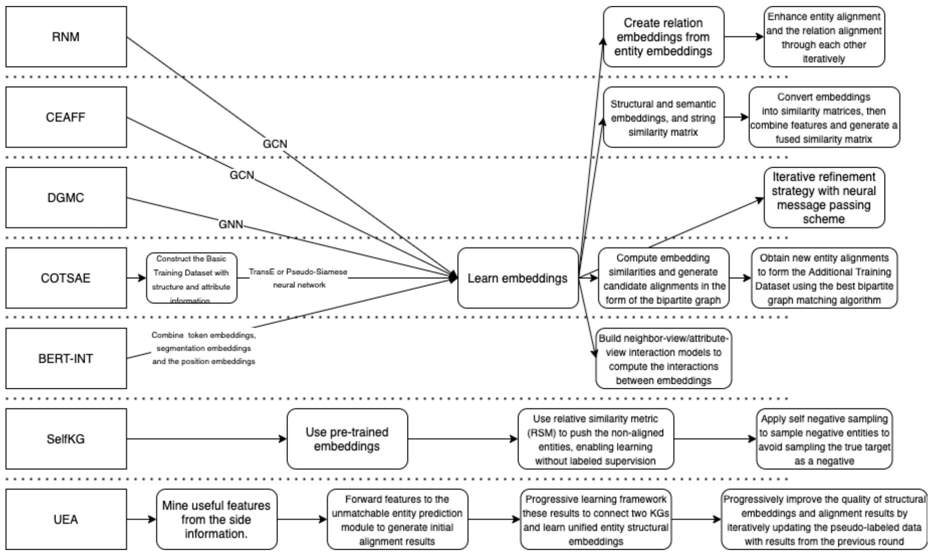

Fig. 1. Generalization of the workflow pipelines: RNM [7], CEAFF [6], DGMC [1], COTSAE [4], BERT-INT [3], SelfKG [2], UEA [5].

Acknowledgement

This work has been partially funded by the European Union's Horizon 2020 research and innovation programme, by the project of ARTICONF (825134), ENVRI-FAIR (824068) and BLUECLOUD (862409).

References

1. Fey, M., Lenssen, J.E., Morris, C., Masci, J., Kriege, N.M.: Deep graph matching consensus. ArXiv **abs/2001.09621** (2020)
2. Liu, X., Hong, H., Wang, X., Chen, Z., Kharlamov, E., Dong, Y., Tang, J.: A self-supervised method for entity alignment (2021)
3. Tang, X., Zhang, J., Chen, B., Yang, Y., Chen, H., Li, C.: Bert-int: A bert-based interaction model for knowledge graph alignment. In: IJCAI. pp. 3174–3180 (2020)
4. Yang, K., Liu, S., Zhao, J., Wang, Y., Xie, B.: Cotsae: Co-training of structure and attribute embeddings for entity alignment. In: AAAI (2020)
5. Zeng, W., Zhao, X., Tang, J., Li, X., Luo, M., Zheng, Q.: Towards entity alignment in the open world: An unsupervised approach. In: DASFAA (2021)
6. Zeng, W., Zhao, X., Tang, J., Lin, X., Groth, P.T.: Reinforcement learning based collective entity alignment with adaptive features. ArXiv **abs/2101.01353** (2021)
7. Zhu, Y., Liu, H., Wu, Z., Du, Y.: Relation-aware neighborhood matching model for entity alignment. In: AAAI (2021)