A Deep Neural Network for Link Prediction on Knowledge Graphs

Wilcke W.X.[1,2], de Boer V[1], van Harmelen F.A.H.[1], and de Kleijn M.T.M.[2]

[1] Department of Computer Science
[2] Department of Spatial Economics
VU University Amsterdam
Amsterdam, The Netherlands
{w.x.wilcke, v.de.boer, frank.van.harmelen, mtm.de.kleijn}@vu.nl

Recent years have seen the emergence of graph-based Knowledge Bases build upon *Semantic Web* technologies, known as Knowledge Graphs (KG). Popular examples are *DBpedia* and *LinkedGeoData*, which offer semantically-annotated and interconnected information extracted from *Wikipedia* and *Open Street Map*, respectively. Factual information stored in such KGs is encoded through relations (edges) between entities (vertices), both sets of which their members' semantics are strictly defined by shared ontological background knowledge. Due to these characteristics, we are able to generate new hypotheses about this information by predicting which links (i.e. relations and the entities they relate) are likely to exist given certain constraints.

Currently-popular approaches to link prediction on KGs are Inductive Logic Programming[1], logic and graph-based kernels [2, 3], and matrix and tensor factorization [4, 5]. As of late, deep Neural Networks (NN) are being considered as well, mainly due to their recent successes on solving complex learning problems. Moreover, being a latent-feature model, they are rather robust towards noise and inconsistencies, as well as able to cope with large-scale and high-dimensional data sets. Despite these useful characteristics, only a handful have yet investigated the effectiveness of deep NNs on KGs [6–8], and even fewer have taken on the challenge of exploiting the ontological background knowledge (e.g. graph features) for improved predictive performance, even though this has been proven useful [9, 10]. We endeavour to fill this gap.

Our study is investigating the effectiveness of a hybrid latent and graph-feature model capable of performing link prediction on real-world KGs. For this purpose, we are employing a deep feedforward NN that uses autoencoders for pre-training purposes. Moreover, its output consists of a vector of probabilities that correspond to all possible relations, whereas its input expects a concatenated vector with each describing any two entities. This description vector is constructed following one of several propositionalization strategies that we have developed, and uses a sparse encoding of both entities' local neighbourhood up to depth n.

Initial results with KGs up to 15 thousand unique links indicate scalability issues, which we believe are caused by the large length of the currently-used input vector, as well as the yet sub-optimized network and training algorithms. To compensate, we kept the number training epoch low and fixed the learning rate at the rather high value of 0.1. As a consequence, we will refrain from publishing evaluations, as we deem them an unreliably measure of our model's effectiveness at present.

For the near future, we are working on method to optimize our model's hyperparameters with the help of Bayesian optimization using Random Forrest as a surrogate model. To cope with the expected increase in required computational resources, we intent to run future experiments on a parallel architecture. Finally, we are working to further improve our propositionalization strategies with the aim of achieving a better representation of an entity while additionally improving scalability.

Keywords: Knowledge Graphs·Semantic Web·Deep Learning·Neural Networks·Propositionalization Strategies

References

1. Rettinger, A., Lsch, U., Tresp, V., d'Amato, C., Fanizzi, N.: Mining the semantic web. Data Mining and Knowledge Discovery **24**(3) (2012) 613–662
2. dAmato, C., Fanizzi, N., Esposito, F.: Classification and retrieval through semantic kernels. In: Knowledge-Based Intelligent Information and Engineering Systems, Springer (2008) 252–259
3. Losch, U., Bloehdorn, S., Rettinger, A.: Graph kernels for rdf data. Number 5781-5782 in Lecture notes in computer science, Lecture notes in artificial intelligence, Springer (2009)
4. Nickel, M., Tresp, V., Kriegel, H.P.: Factorizing YAGO: scalable machine learning for linked data. In: Proceedings of the 21st international conference on World Wide Web, ACM (2012) 271–280
5. Franz, T., Schultz, A., Sizov, S., Staab, S.: Triplerank: Ranking semantic web data by tensor decomposition. Springer (2009)
6. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., Potts, C.: Recursive deep models for semantic compositionality over a sentiment treebank. In: Proceedings of the conference on empirical methods in natural language processing (EMNLP). Volume 1631., Citeseer (2013) 1642
7. Huang, H., Heck, L., Ji, H.: Leveraging Deep Neural Networks and Knowledge Graphs for Entity Disambiguation. arXiv preprint arXiv:1504.07678 (2015)
8. Yu, D., Deng, L., Seide, F.: Large Vocabulary Speech Recognition Using Deep Tensor Neural Networks. In: INTERSPEECH. (2012)
9. Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K., Strohmann, T., Sun, S., Zhang, W.: Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM (2014) 601–610
10. Nickel, M., Jiang, X., Tresp, V.: Reducing the Rank in Relational Factorization Models by Including Observable Patterns. In: Advances in Neural Information Processing Systems. (2014) 1179–1187