Reproducibility Report for ACM SIGMOD 2020 Paper: "Realistic Re-evaluation of Knowledge Graph Completion Methods: An Experimental Study"

ANGELOS-CHRISTOS ANADIOTIS, École Polytechnique, Institut Polytechnique de Paris, France

The main results of the paper can be reproduced. The source code is available online at github, together with the instructions required. The execution of the experiments performed by the authors required about two weeks to finish, therefore, it is recommended to future works that need to build on top of this one, to isolate the experiments that need to be reproduced, instead of running the whole batch.

1 INTRODUCTION

The paper authors are: Farahnaz Akrami (University of Texas at Arlington), Mahammed Samiul Saeef (University of Texas at Arlington), Qingheng Zhang (Nanjing University), Wei Hu (Nanjing University), Chengkai Li (University of Texas at Arlington). The main results of the paper are reproducible; specifically, Tables 5, 6, 7, 8, 9, 10, 11, 12, and 13. Some of the tables include raw results obtained by the code and some others require some further processing. Code is provided for either case, whereas further details about the explanation of the results are already provided in the paper [1].

2 SUBMISSION

The reproducibility submission includes a Github repository where the authors explain how to execute the experiments and obtain the results included in the paper [1]. The individual models that were used are cloned by other repositories and explicitly pointed at the Github repository of the reproduced paper. There are two scripts required to obtain the results, one for executing code in Python 3 and one in Python 2. Different versions of Python were required due to compatibility issues with the individual models. Further scripts are provided by the authors to compose some of the tables that are included in the paper. Finally, the directory structure of the project is organized on a per-model basis, where the output of training on every model is written to an output directory created by the script. The only drawback is that the output of every model follows a different format, and further work is required to parse the results. The effect is not severe, given that: (*i*) there are not many final results, which makes them easily traceable at the end of the output files; (*ii*) for some tables, there are already scripts provided to re-create them.

The reproducibility submission with all material is included at the following Github repository: https://github.com/Knowledge-Graph/KG_completion.

3 HARDWARE AND SOFTWARE ENVIRONMENT

The hardware setup used in the paper and for reproducing the results are given in Table 1.

Table 1. Hardware & Software environment

	Paper	Repro Review
CPU	Intel E5-2695	AMD EPYC 7702P
GHz	2.1	2.0
RAM	256GB	950GB
GPU	2 NVIDIA Geforce GTX1080Ti	2 NVIDIA Tesla V100-PCIE-32GB

4   REPRODUCIBILITY EVALUATION

4.1   Process

The process was extremely complex, and it successfully finished only thanks to big effort put by the authors. The main issue was that the Python packages that were listed as project dependencies, could not be directly installed in any of the three different server models used for reproducing the results. After several interactions with the authors, and many changes in the list of dependencies, it became possible to deploy the training using Python 3. However, training the models used in the paper takes about 2 weeks and in the meanwhile, the server rebooted due to an external event. To avoid any similar issues, the authors setup a VPN and an account to run the whole reproducibility pipeline on their own infrastructure. All the changes in the configuration are now reflected in the master version of the project repository. The final evaluation of the results was done in collaboration with the first author, who showed all the logs, demonstrating that the results had indeed been obtained from scratch. Further, there were some small differences in the reported values with respect to what is given in the paper, showing that the experiments had not been directly copied from the server that the authors used in the first place. It worth noting that the reproducibility server was a different one than the one used in the paper, given that the authors had bought a new one in the meanwhile.

4.2   Results

The main results included in the paper, that is, Tables 5 – 13 were reproduced.

5   SUMMARY

The reproducibility process has been very complicated. However, this is not on the authors to blame, since the workflow that they followed to extract the dependencies and organize their files was correct. Unfortunately, issues related to Python version inconsistencies as well as to different operating system distribution, have introduced a lot of noise in the process. Finally, the overall duration of the experiments, that is about 2 weeks, also put another important limitation, since it is hard to allocate a server for such a long time.

REFERENCES

[1] Farahnaz Akrami, Mohammed Samiul Saeef, Qingheng Zhang, Wei Hu, and Chengkai Li. 2020. Realistic Re-Evaluation of Knowledge Graph Completion Methods: An Experimental Study. In *Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data (SIGMOD '20)*. Association for Computing Machinery, New York, NY, USA, 1995–2010. https://doi.org/10.1145/3318464.3380599