Chương 1. INTRODUCTION

Nowadays, graphs have been applied in all aspects of life. Social network graphs (e.g., Facebook [12]) illustrate how individuals are connected to each other, the places we visit, and the information we interact with. Graphs are also used as core structures in video recommendation systems (e.g., YouTube [1]), flight networks, GPS navigation systems, scientific computations, and even brain connectivity analysis. Google's Knowledge Graph [5], introduced in 2012 [6], is a notable example of how information can be structured and utilized in knowledge graphs.

Effectively exploiting knowledge graphs provides users with deeper insight into the underlying data, which can benefit many real-world applications. However, in practice, new knowledge is continuously generated, and the acquired information is often incomplete or missing. This leads to the problem of knowledge graph completion or link prediction in knowledge graphs.

Most current approaches aim to predict a new edge connecting two existing nodes. Such methods help make the graph more complete—i.e., denser—by introducing additional connecting edges. However, these approaches primarily address the problem of completion rather than the challenge of integrating new knowledge into the graph, which remains an open question. Currently, research in knowledge graph completion follows two main directions: one is optimizing an objective function to make predictions with minimal error, as in RuDiK [10], AMIE [4], and RuleN [8], which are typically used in vertex or edge classification applications. The other approach generates a ranked list of k candidate triples, where the score reflects decreasing confidence, as seen in studies such as TransE [2] and ConvKB [15], which are commonly used in recommendation systems. Our approach follows this second direction of producing a candidate list.

Within these approaches, there are two main methodologies: rule-based systems such as AnyBURL [7], and embedding-based methods such as ConvE [3], TransE [2], and ComplEx [11]. With the goal of gaining a systematic understanding of these methods, we chose to explore both directions in this thesis. For the rule-based approach, we selected AnyBURL [7], and for the graph embedding-based method, we chose KBAT [9], which employs attention mechanisms.

Our contribution in the AnyBURL method includes a Python implementation [1], along with two proposed strategies for adding new knowledge to the graph, which we term *online-to-offline* and *online-to-online*. The *online-to-offline* strategy extends AnyBURL by generating rules when a batch (set) of new knowledge is added. The *online-to-online* strategy generates rules immediately when a single new piece of knowledge (edge) is added.

For the embedding-based method, we present a review of attention mechanisms [13], their application in knowledge graphs via Graph Attention Networks (GATs) [14], and the KBAT model [9].

Our contribution in the deep learning approach includes a publicly available implementation and training process on GitHub [2], with both training code and model results openly provided.

[1] https://github.com/MinhTamPhan/mythesis

[2] https://github.com/hmthanh/GCAT

TÀI LIỆU TRÍCH DẪN

[1] Shumeet Baluja, Rohan Seth, Dharshi Sivakumar, Yushi Jing, Jay Yagnik, Shankar Kumar, Deepak Ravichandran, and Mohamed Aly. Video suggestion and discovery for youtube: taking random walks through the view graph. In *Proceedings of the 17th international conference on World Wide Web*, pages 895–904, 2008.

[2] Antoine Bordes, Nicolas Usunier, Alberto Garcia-Duran, Jason Weston, and Oksana Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in neural information processing systems*, pages 2787–2795, 2013.

[3] Tim Dettmers, Pasquale Minervini, Pontus Stenetorp, and Sebastian Riedel. Convolutional 2d knowledge graph embeddings. *arXiv preprint arXiv:1707.01476*, 2017.

[4] Luis Galárraga, Christina Teflioudi, Katja Hose, and Fabian M Suchanek. Fast rule mining in ontological knowledge bases with amie. *The VLDB Journal*, 24(6):707–730, 2015.

[5] Google. *Introducing the Knowledge Graph: things, not strings*, 2020 (truy cập ngày 27/08/2020).

[6] Shaoxiong Ji, Shirui Pan, Erik Cambria, Pekka Marttinen, and Philip S Yu. A survey on knowledge graphs: Representation, acquisition and applications. *arXiv preprint arXiv:2002.00388*, 2020.

[7] Christian Meilicke, Melisachew Wudage Chekol, Daniel Ruffinelli, and Heiner Stuckenschmidt. Anytime Bottom-Up Rule Learning for Knowledge Graph Completion, 2019.

[8] Christian Meilicke, Manuel Fink, Yanjie Wang, Daniel Ruffinelli, Rainer Gemulla, and Heiner Stuckenschmidt. Fine-grained evaluation of rule-and embedding-based systems for knowledge graph completion. In *International Semantic Web Conference*, pages 3–20. Springer, 2018.

[9] Deepak Nathani, Jatin Chauhan, Charu Sharma, and Manohar Kaul. Learning attention-based embeddings for relation prediction in knowledge graphs. *arXiv preprint arXiv:1906.01195*, 2019.

[10] Stefano Ortona, Venkata Vamsikrishna Meduri, and Paolo Papotti. Robust discovery of positive and negative rules in knowledge bases. In *2018 IEEE 34th International Conference on Data Engineering (ICDE)*, pages 1168–1179. IEEE, 2018.

[11] Théo Trouillon, Johannes Welbl, Sebastian Riedel, Éric Gaussier, and Guillaume Bouchard. Complex embeddings for simple link prediction. International Conference on Machine Learning (ICML), 2016.

[12] Johan Ugander, Brian Karrer, Lars Backstrom, and Cameron Marlow. The anatomy of the facebook social graph. *arXiv preprint arXiv:1111.4503*, 2011.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In *Advances in neural information processing systems*, pages 5998–6008, 2017.

[14] Petar Veličković, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

[15] Thanh Vu, Tu Dinh Nguyen, Dat Quoc Nguyen, Dinh Phung, et al. A capsule network-based embedding model for knowledge graph completion and search personalization. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 2180–2189, 2019.