



ĐỀ CƯƠNG KHOÁ LUẬN TỐT NGHIỆP
Dự đoán liên kết trong đồ thị phức
(Knowledge Graph Embedding for Link Prediction)

1 THÔNG TIN CHUNG

Người hướng dẫn:

– Ths. Lê Ngọc Thành (Khoa Công nghệ Thông tin)

Nhóm Sinh viên thực hiện:

1. Phan Minh Tâm (18424059)
2. Hoàng Minh Thanh (18424062)

Loại đề tài: Nghiên cứu

Thời gian thực hiện: Từ 06/2020 đến 09/2020

2 NỘI DUNG THỰC HIỆN

2.1 Giới thiệu về đề tài

Đồ thị tri thức (Knowledge Graphs-KG) là các biểu diễn cấu trúc của thông tin thế giới thực. Do khả năng mô hình hóa dữ liệu có cấu trúc, phức tạp theo cách máy tính có thể dễ dàng “hiểu được”, KG hiện đang được sử dụng rộng rãi trong nhiều lĩnh vực khác nhau, từ trả lời câu hỏi đến truy xuất thông tin và các hệ thống có thể suy luận dựa trên nội dung đã có. Việc phát triển một KG có thể

được thực hiện bằng cách trích xuất các sự kiện mới từ các nguồn bên ngoài hoặc bằng cách suy ra các sự kiện còn thiếu từ những sự kiện đã có trong KG. Phương pháp tiếp cận, được gọi là Dự đoán liên kết (Link Prediction-LP).

2.2 Mục tiêu đề tài

Cùng với nhiều kỹ thuật trí tuệ nhân tạo phát triển mạnh gần đây, đề tài tập chung nghiên cứu vào các khía cạnh của bài toán LP trên KG như đặc trưng tập dữ liệu, thuật toán, thực nghiệm đánh giá các phương pháp cũng như các kỹ thuật khác nhau cùng tìm hiểu xem liệu những đặc trưng gì của tập dữ liệu hoặc các thuật toán khác nhau ảnh hưởng tới khả năng khái quát hóa của mô hình.

2.3 Phạm vi của đề tài

LP là một lĩnh vực nghiên cứu ngày càng sôi nổi gần đây đã phát triển mạnh mẽ từ sự bùng nổ của các kỹ thuật trong trí tuệ nhân tạo như: máy học (machine learning) và kỹ thuật học sâu (deep learning). Đề tài sẽ tập trung nghiên cứu các mô hình LP sử dụng KG làm nền tảng để tìm hiểu các biểu diễn dữ liệu với số chiều thấp còn được gọi là Knowledge Graph Embeddings, sau đó sử dụng chúng để suy ra các sự kiện, quan hệ mới.

2.4 Cách tiếp cận dự kiến

Phần này nêu các phương pháp, cách tiếp cận cũng như mô hình dự kiến thực hiện trong đề tài.

Các trích dẫn từ các tài liệu sử dụng theo định dạng của tổ chức IEEE. Các ví dụ kế tiếp thể hiện trích dẫn tài liệu từ sách ([1]), từ bài báo trong tạp chí ([2]) hay từ đường dẫn đến website ([3]).

2.5 Kết quả dự kiến của đề tài

Phần này nêu mô tả dự kiến các kết quả đạt được của đề tài, bao gồm sản phẩm, các cải tiến hoặc công trình khoa học có liên quan.

2.6 Kế hoạch thực hiện

- 4/2020 - 6/2020: Tìm hiểu các kiến thức liên quan về mạng nơ-ron, KG. Đọc các tài liệu bài báo liên quan tới đề tài.
- 6/2020 - 8/2020: Hiện thực các phương pháp, thuật toán đã đề xuất và tìm hiểu.
- 8/2020 - 9/2020: Tìm hiểu các phương pháp khác thực nghiệm và so sánh kết quả với thuật toán gốc.
- 9/2020 - 10/2020: Báo cáo và bảo vệ khóa luận tốt nghiệp.

Phần này mô tả về kế hoạch thực hiện (với các mốc thời gian tương ứng) cùng với việc phân chia công việc cho các thành viên tham gia đề tài.

Tài liệu

- [1] M. Goossens, F. Mittelbach, and A. Samarin, *The L^AT_EX Companion*. Reading, Massachusetts: Addison-Wesley, 1993.
- [2] A. Einstein, “Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies],” *Annalen der Physik*, vol. 322, no. 10, pp. 891–921, 1905.
- [3] A. ROSSI, D. FIRMANI, A. MATINATA, P. MERIALDO, and D. BARBOSA, “Knowledge graph embedding for link prediction: A comparative analysis.” <https://arxiv.org/pdf/2002.00819.pdf>.
- [4] D. Q. Nguyen, T. D. Nguyen, D. Q. Nguyen, and D. Phung, “A capsule network-based embedding model for knowledge graph completion and search personalization.[on association for computational linguistics],” pp. 2180–2189, 2019.

XÁC NHẬN
CỦA NGƯỜI HƯỚNG DẪN
(Ký và ghi rõ họ tên)

TP. Hồ Chí Minh, ngày 20 tháng 09 năm 2020
NHÓM SINH VIÊN THỰC HIỆN
(Ký và ghi rõ họ tên)