



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 04/2020 đến 10/2020

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 trung 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

Hiện tại có 3 hướng nghiên cứu chính về lĩnh vực dự đoán liên kết bao gồm : Ma trận hóa (Matrix Factorization), Biến đổi hình học (Geometric), Học sâu (Deep Learning). Nhóm dự kiến tiếp cận theo hai hướng chính là sử dụng phương pháp truyền thống có tên gọi là AnyBURL [?] và kỹ thuật học sâu với mô hình KBGAT. Nhóm dự kiến sẽ đọc hiểu đề tài của các tác giả

Với mong muốn được tìm hiểu và nghiên cứu các phương pháp trí tuệ nhân tạo được áp dụng vào LG từ cổ điển đến hiện đại đề tài sẽ nghiên cứu một phương pháp cổ điển dựa trên luật (rule base [?]) và phương pháp còn lại dựa vào phương pháp học sâu (learning attention [?]). Đối với phương pháp cổ điển thuật toán cố gắng đưa các luật Horn [?] sau đó khái quát hóa chúng thành các luật tổng quát

trên tập dữ liệu huấn luyện. Phương pháp còn lại dựa vào kỹ thuật nhúng [?] đồ thị tri thức vào một không gian số chiều thấp dựa trên các láng giềng lân cận [?] sau đó học trên không gian này để đưa ra các dự đoán về các cạnh của đồ thị.

2.5 Kết quả dự kiến của đề tài

Tìm hiểu các phương pháp khác thực nghiệm và so sánh kết quả với phương pháp đang nghiên cứu với các phương pháp khác trên bốn tập dữ liệu được nghiên cứu chuẩn cho KG là FB15k [1], FB15k-273 [?], WN188 [?], WN188RR [?]

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.

Tài liệu

- [1] A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko, “Translating embeddings for modeling multi-relational data,” in *Advances in neural information processing systems*, pp. 2787–2795, 2013.

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)