

THÔNG TIN VỀ LUẬN ÁN

Tên đề tài luận án	Nghiên Cứu Mô Hình Khai Thác Mạng Thông Tin Không Đồng Nhất Và Ứng Dụng
Chuyên ngành	Công Nghệ Thông Tin
Mã số ngành	62-48-02-01
Họ tên nghiên cứu sinh	Phạm Thế Anh Phú
Người hướng dẫn khoa học	PGS.TS. Đỗ Phúc
Cơ sở đào tạo	Đại học Công nghệ Thông tin – Đại học Quốc gia Tp.HCM

1. TÓM TẮT NỘI DUNG LUẬN ÁN

Toàn bộ nội dung của luận án trình bày về các vấn đề nghiên cứu liên quan đến lĩnh vực phân tích và khai phá mạng thông tin không đồng nhất. Bao gồm các phần nội dung liên quan đến các nghiên cứu tổng quan và các cải tiến cụ thể cho bài toán tìm kiếm tương đồng trong mạng thông tin không đồng nhất (Heterogeneous Information Network – HIN). Nội dung trọng tâm của các cải tiến trong luận án là tập trung vào việc kết hợp giữa yếu tố tương đồng trong nội dung và cấu trúc liên kết, ở dạng meta-path, của các nút trong mạng HIN. Để từ đó nâng cao hiệu suất cho bài toán tính toán tương đồng. Bên cạnh đó luận án cũng đề xuất các hướng cải tiến cho bài toán dự đoán liên kết trên ngữ cảnh mạng thông tin không đồng nhất giàu nội dung. Kết quả của luận án được tổng hợp trong 4 công trình đã công bố trên các tạp chí thuộc danh mục SCIE, như sau: Expert System with Applications (1), International Journal of Machine Learning and Cybernetics (1), Intelligent Data Analysis (2). Các vấn đề cần giải quyết cũng như đề xuất cải tiến của luận án được tổ chức thành 3 bài toán chính như sau:

- **Bài Toán 1**: Trong bài toán 1, NCS trung vào việc xây dựng cơ sở lý thuyết cho vấn đề khám phá sự phân bố của chủ đề trong mạng thông tin không đồng nhất giàu nội dung (Content-based HIN, hay: C-HIN). Để từ đó kết hợp sự tương đồng trong chủ đề với mối quan hệ giữa các thực thể/nút nhằm đưa ra những mô hình cải tiến phù hợp cho việc khai phá mạng thông tin giàu nội dung. Để cụ thể cho bài toán 1 của luận án NCS đã đề xuất thuật toán W-PathSim, là một thuật toán dùng để xác định mức độ tương đồng giữa các nút trong mạng HIN thông qua việc đánh giá sự tương đồng trong cấu trúc liên kết và nội dung.
- **Bài Toán 2**: kế thừa từ các kết quả đã đạt được từ mô hình W-PathSim đã đề xuất trong bài toán 1, NCS xây dựng và phát triển mô hình học biểu diễn mạng thông tin không đồng nhất giàu nội dung, với 2 mô hình đề xuất: **W-MetaPath2Vec** và **W-MetaGraph2Vec**.

- **Bài Toán 3:** Từ các kết quả đã đạt được trong hai mô hình đề xuất tại bài toán 2, NCS áp dụng cải tiến cho bài toán dự đoán liên kết trong mạng C-HIN theo hướng tiếp cận học mô hình biểu diễn mạng thông tin không đồng nhất. Trong tâm nội dung bài toán 3 được giải quyết với mô hình đề xuất là thuật toán **W-MMP2Vec**.

2. NHỮNG KẾT QUẢ ĐÃ ĐẠT ĐƯỢC CỦA LUẬN ÁN

Thông qua việc giải quyết các thách thức liên quan đến việc phân tích và khai phá tri thức từ mạng thông tin giàu nội dung (C-HIN), NCS đã đạt được một số kết quả đáng kể ở cả hai khía cạnh học thuật và ứng dụng, bao gồm:

- **Về ý nghĩa học thuật và khoa học:** Đề xuất các mô hình cải tiến kết hợp giữa việc đánh giá sự tương quan giữa các thực thể thông qua việc đánh giá cả hai tiêu chí là các mối quan hệ giữa chúng cũng như sự tương đồng trong chủ đề sẽ hỗ trợ cho việc tăng cao hiệu suất độ chính xác của việc tìm kiếm tương đồng trên mạng thông tin cũng như kết quả tìm kiếm sẽ có ý nghĩa hơn so với việc chỉ dựa vào các liên kết giữa các thực thể trong mạng C-HIN. Bên cạnh đó, thông qua việc đề xuất độ đo tương đồng mới trên mạng thông tin không đồng nhất giàu nội dung (C-HIN), luận án cũng sử dụng các kết quả đạt được để phát triển các mô hình nhúng/biểu diễn mạng thông tin. Các mô hình này nhằm hỗ trợ cho việc chuyển đổi các thực thể trong mạng C-HIN về dạng các vector với số chiều quy định một cách hiệu quả hơn. Từ đó đóng góp cụ thể để giải quyết bài toán điển hình như tìm kiếm tương đồng và dự đoán liên kết có xét đến sự tương quan trong nội dung & chủ đề giữa các thực thể.
- **Về ý nghĩa thực tiễn/ứng dụng:** các mô hình được đề xuất trong nội dung của luận án sẽ hỗ trợ trực tiếp cho việc xây dựng các ứng dụng liên quan đến tư vấn, khuyến nghị cộng tác khoa học thông qua việc giải quyết các bài toán tìm kiếm tương đồng giữa các nút tác giả và dự đoán liên kết giữa các tác giả trong mạng thông tin học thuật. Ngoài ra các mô hình đề xuất cũng có thể được sử dụng rộng rãi cho việc giải quyết các bài toán khai phá trên các dạng mạng thông tin khác nhau như mạng xã hội và thương mại điện tử.

3. CÁC HƯỚNG PHÁT TRIỂN TRONG TƯƠNG LAI

Hướng tiếp cận của các mô hình biểu diễn mạng C-HIN được đề xuất trong luận án hiện tại chỉ áp dụng nguyên lý huấn luyện mạng nơ-ron đơn giản với một tầng ẩn, do đó hiệu suất về độ chính xác của mô hình chưa thể đạt được hiệu quả cao nhất. Do đó, một trong các hướng cải tiến

quan trọng trong tương lai của luận án là thay thế cơ chế huấn luyện bằng các kiến trúc mạng nơ-ron đa tầng của lĩnh vực học sâu (deep learning). Qua đó, có thể tăng cao hiệu suất về độ chính xác cho việc học mô hình biểu diễn mạng thông tin. Trong đó, kiến trúc mạng nơ-ron Graph Convolutional Network (GCN) đa tầng là một trong các hướng cải tiến tiềm năng.

Ngoài ra, các mô hình đề xuất trong luận án chủ yếu dựa trên nguyên lý bước đi ngẫu nhiên để mô hình hóa cấu trúc mạng thông tin ở mức độ cục bộ và tương tự (local structure) nhau giữa các nút/thực thể trong mạng thông tin. Do đó, các mô hình đề xuất trong luận án hầu như chưa thể bảo toàn một cách hiệu quả cấu trúc toàn cục (global structure) của mạng thông tin. Các cải tiến trong tương lai sẽ tập trung vào việc kết hợp bảo toàn cấu trúc của mạng thông tin ở nhiều cấp độ khác nhau (cả local structure và global structure).

CÁN BỘ HƯỚNG DẪN



PGS.TS ĐỖ PHÚC
Ngày: 05/10/2021

NGHIÊN CỨU SINH



PHẠM THẾ ANH PHÚ
Ngày: 05/10/2021