ĐẠI HỌC QUỐC GIA HÀ NỘI
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ

NGUYỄN THỊ THÙY ANH

NGHIÊN CỨU PHÁT TRIỂN CÁC KỸ THUẬT HỌC MÁY HOÀN THIỆN ĐỒ THỊ TRI THỨC

Ngành đào tạo: Hệ thống thông tin
Mã số: 9480104

TÓM TẮT LUẬN ÁN TIẾN SĨ HỆ THỐNG THÔNG TIN

Hà Nội – 2025

Công trình được hoàn thành tại: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

Người hướng dẫn khoa học:

1. PGS. TS. Hà Quang Thụy
2. PGS. TS. Phan Xuân Hiếu

Phản biện:

Phản biện:

Phản biện:

Luận án sẽ được bảo vệ trước Hội đồng cấp Đại học Quốc gia chấm luận án tiến sĩ họp tại

vào hồi giờ ngày tháng năm

Có thể tìm hiểu luận án tại:

- Thư viện Quốc gia Việt Nam
- Trung tâm Thông tin - Thư viện, Đại học Quốc gia Hà Nội

Mở đầu

Đồ thị tri thức (*ĐTTT*, Knowledge graph: *KG*) đóng vai trò quan trọng đặc biệt, cung cấp thông tin ngữ nghĩa phong phú và đã nổi lên như một động lực thúc đẩy các lĩnh vực tìm kiếm ngữ nghĩa và Trí tuệ nhân tạo (TTNT, Artificial Intelligence: *AI*) [1]. ĐTTT được biểu diễn dưới dạng một bộ ba $KG = (E, R, F)$, trong đó E là tập hữu hạn các thực thể (tập nút), R là tập hữu hạn các quan hệ (tập nhãn của các cung), còn F là một tập các thể hiện quan hệ (tập cung) $F \subseteq E \times R \times E$; một cung trong F được biểu diễn dưới dạng bộ ba (h, r, t) trong đó $h, t \in E$ và còn $r \in R$; trong cung (h, r, t), h được gọi là nút đầu, t được gọi là nút cuối và r được gọi là nhãn của cung.

Gần đây, nhu cầu *cung cấp ngữ cảnh có tính thời gian* đã trở nên hết sức cấp thiết, từ đó hình thành một kiểu ĐTTT mới - ĐTTT thời gian (Temporal Knowledge Graph: TKG) với sự bổ sung yếu tố thời gian vào tập các dữ kiện F trong ĐTTT thông thường. Thêm nữa, tìm kiếm trên Science Direct, Springer và DBLP, luận án nhận thấy rằng công bố khoa học có tiêu đề liên quan tới các cụm từ "Multi modal Knowledge Graph Completion" hoặc "Multimodal Knowledge Graph Completion" mới chỉ xuất hiện từ năm 2022 trở lại đây . Điều đó có nghĩa là nghiên cứu về Hoàn thiện ĐTTT, Hoàn thiện ĐTTT đa phương thức, Hoàn thiện ĐTTT thời gian là rất cần thiết và có tính thời sự.

Mặc dù ĐTTT có giá trị rất lớn trong nhiều ứng dụng, nhưng chúng thường được xây dựng thủ công hoặc bán tự động nên vẫn không đầy đủ. Vì vậy, bài toán Thu thập tri thức với mục tiêu là giải quyết các vấn đề về tính không đầy đủ và thưa thớt trong ĐTTT có tầm quan trọng đặc biệt. Thu thập tri thức bao gồm ba bài toán con để giải quyết ba trường hợp chưa đầy đủ điển hình của ĐTTT là: (i) Phát hiện thực thể (Entity Discovery),(ii) Trích xuất quan hệ (Relation Extraction), (iii) Hoàn thiện ĐTTT (Knowledge Graph Completion)[7].

Định hướng tới các chủ đề nghiên cứu quan trọng và xu hướng nghiên cứu về hoàn thiện ĐTTT như đã được đề cập, theo dòng nghiên cứu của các luận án Tiến sỹ trên thế giới, **luận án** đặt ra hai **câu hỏi nghiên cứu** với **mục tiêu nghiên cứu** tương ứng như sau: Câu hỏi nghiên cứu đầu tiên là về phương pháp và nguồn tài nguyên cho phép sử dụng thông tin bổ sung cho Hoàn thiện ĐTTT, tương ứng với xu hướng nghiên cứu (1). Luận án hướng mục tiêu đề xuất được các mô hình hoàn thiện ĐTTT bằng kỹ thuật học chuyển giao (transfer learning) để sử dụng thông tin bổ sung bên ngoài làm giàu thông tin bên trong cho hoàn thiện ĐTTT. Câu hỏi nghiên cứu thứ hai là về các kỹ thuật và mô hình Hoàn thiện ĐTTT trong các miền ứng dụng cụ thể có sử dụng mô hình ngôn ngữ huấn luyện trước, tương ứng với hai xu hướng nghiên cứu (3) và (4). Luận án hướng mục tiêu đề xuất được các kỹ thuật và mô hình hoàn thiện ĐTTT thời gian và hoàn thiện ĐTTT đa phương thức.

Đối tượng nghiên cứu của luận án là các mô hình hoàn thiện ĐTTT và các kỹ thuật được áp dụng trong các mô hình đó.

Phạm vi nghiên cứu của luận án tập trung tới các kỹ thuật và giải pháp cải tiến được áp dụng trong các mô hình hoàn thiện ĐTTT trên một số miền dữ liệu ĐTTT mở, ĐTTT đa phương thức và ĐTTT thời gian.

Luận án sử dụng **phương pháp nghiên cứu kết hợp** bao gồm phương pháp nghiên cứu định tính và phương pháp nghiên cứu định lượng.

Tham gia vào dòng nghiên cứu trên thế giới về hoàn thiện ĐTTT, luận án có ba đóng góp chính sau đây: (i) Đề xuất mô hình hoàn thiện ĐTTT qua học chuyển giao BERT/FastText-GRU-KGC được cải tiến từ mô hình GloVe-GRU-KGC do V. Kocijan và T. Lukasiewicz đề xuất năm 2021 [17]; (ii) Đề xuất hai mô hình hoàn thiện ĐTTT đa phương thức ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT được cải tiến từ mô hình AdaMF-MAT được Y. Zhang và cộng sự đề xuất năm 2024 [18], trong đó, ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT sử dụng mô hình nhúng ảnh Vision Transformer (ViT) và mô hình nhúng văn bản T5; (iii) Đề xuất hai mô hình hoàn thiện ĐTTT thời gian là DE-RotatE và DE-RotatE-sinc được cải tiến từ mô hình DE-SimplE do R. Goel và cộng sự đề xuất năm 2020 [19] theo ý tưởng sử dụng phép quay chiếu thay vì sử dụng phép tịnh tiến chiếu trong DE-SimplE. Các mô hình đề xuất đều được triển khai thực nghiệm đánh giá hiệu năng và được công bố tại trong các báo cáo [NCS1], [NCS4], [NCS2, NCS3].

Bố cục của luận án gồm phần Mở đầu và bốn chương nội dung, phần Kết luận và danh mục các tài liệu tham khảo.

Chương 1. Đồ thị tri thức và hoàn thiện đồ thị tri thức

1.1 Đồ thi tri thức

1.1.1 Khái niệm về đồ thị tri thức

Định nghĩa 1.1 ĐTTT [7]: ĐTTT là một bộ ba KG=(E,R,F), trong đó E là tập hữu hạn các thực thể (tập nút), R là tập hữu hạn các quan hệ (tập nhãn của các cung), còn F là một tập các thể hiện quan hệ (tập cung), $F \subseteq E \times R \times E$; một cung trong F được biểu diễn dưới dạng bộ ba (h,r,t) trong đó $h,t \in E$ và còn $r \in R$; trong cung (h,r,t),h được gọi là nút đầu, t được gọi là nút cuối và r được gọi là nhãn.

1.1.2 Phân loại ĐTTT

Theo Liang và cộng sự [23], ĐTTT có thể được phân chia làm các loại khác nhau phụ thuộc vào đặc điểm, tính chất, đặc tính của các thực thể và quan hệ trong đó, trong đó, ba loại ĐTTT điển hình là: ĐTTT mở (Open Knowledge Graph, OKG), ĐTTT đa phương thức (Multi-Modal Knowledge Graph – MMKG), ĐTTT thời gian (Temporal Knowledge Graph – TKG).

1.1.3 Khái quát về các bài toán nghiên cứu về ĐTTT

Bốn bài toán nghiên cứu về ĐTTT ở mức cao nhất là Học biểu diễn tri thức, Thu thập tri thức, ĐTTT thời gian và Các ứng dụng nhận thức tri thức dựa trên ĐTTT. Nhiệm vụ thu thập tri thức chính là việc xây dựng ĐTTT. Bên cạnh nhiệm vụ xây dựng ĐTTT thì mục đích của thu thập tri thức chính là mở rộng ĐTTT. Các bài toán mở rộng ĐTTT trong thu thập tri thức được chia làm ba loại chính: hoàn thiện ĐTTT, trích xuất quan hệ thực thể và phát hiện thực thể [5].

1.2 Hoàn thiện ĐTTT

1.2.1 Khái niệm về hoàn thiện đồ thị tri thức

Hoàn thiện ĐTTT, còn được gọi là Dự đoán liên kết (Link Prediction), nhằm mục đích tự động suy ra các dữ kiện còn thiếu cho một ĐTTT bằng cách học hỏi từ các dữ kiện hiện có trong ĐTTT đó.

1.2.2 Quy trình chung hoàn thiện ĐTTT

Quy trình chung hoàn thiện ĐTTT (sau khi ĐTTT đã được tiền xử lý) bao gồm ba phần: huấn luyện mô hình, xử lý ứng viên và xác định dữ kiện [9].

1.2.3 Hệ thống các kỹ thuật hoàn thiện ĐTTT

T. Shen và cộng sự [9] đã hệ thống hóa các kỹ thuật hoàn thiện ĐTTT bao gồm ba nhóm kỹ thuật chính hoàn thiện ĐTTT là kỹ thuật hoàn thiện ĐTTT dựa trên thông tin cấu trúc, kỹ thuật hoàn thiện ĐTTT dựa trên thông tin bổ sung và kỹ thuật hoàn thiện ĐTTT khác.

1.2.4 Một số mô hình hoàn thiện ĐTTT điển hình

Trong mục này, luận án giới thiệu các mô hình hoàn thiện ĐTTT được coi là điển hình nhất như là mô hình TransE, DisMult, RotatE, SimplE, TuckER, 5⋆E, ConvE thường được sử dụng làm nền tảng phát triển các mô hình hoàn thiện ĐTTT khác, trong đó có các mô hình được luận án nghiên cứu và cải tiến.

1.3 Hoàn thiện ĐTTT đa phương thức

Hiện nay, hoàn thiện ĐTTT đa phương thức thường tập trung vào hai hướng phát triển: tăng cường biểu diễn thực thể bằng cách dựa trên thông tin đa phương thức (hình ảnh, văn bản) của chúng; cải thiện quá trình lấy mẫu âm của các thông tin đa phương thức.

1.3.1 Tiếp cận biểu diễn thực thể với thông tin đa phương thức

Bốn mô hình hoàn thiện ĐTTT đa phương thức điển hình theo tiếp cận biểu diễn thực thể với thông tin đa phương thức là IKRL, TBKGC, TransAE, VBKGC.

Mô hình IKRL được các tác giả xây dựng theo hướng: Đầu tiên đề xuất một bộ mã hóa ảnh để tạo ra biểu diễn dựa trên hình ảnh cho từng trường hợp; tiếp theo, xây dựng biểu diễn dựa trên ảnh tổng hợp cho từng thực thể; cuối cùng, các tác giả kết hợp việc học các biểu diễn tri thức bằng các phương pháp dựa trên dịch chuyển (translation-based methods). Hàm tính điểm trong IKRL là TransE [36].

Mô hình TBKGC (Translation-Based for Knowledge Graph Completion) do H. M. Sergieh và cộng sự đề xuất [44]. Mô hình này được xây dựng bằng cách mở rộng thông tin ảnh và văn bản từ mô hình IKRL

Trong mô hình TransAE, lớp ẩn của các bộ mã hóa tự động được sử dụng như là biểu diễn của các thực thể trong mô hình TransE. Do đó, nó không chỉ mã hóa tri thức có cấu trúc mà còn mã hóa cả tri thức đa phương thức, như tri thức ảnh và tri thức văn bản trong biểu diễn cuối cùng. Hàm tính điểm tổng trong mô hình TransAE được xây dựng như sau:

$$f(h,r,t) = f_S + f_{M1} + f_{M2} + f_{SM} + f_{MS} \qquad (1\text{-}9)$$

Mô hình VBKGC (VisualBERT-enhanced Knowledge Graph Completion – VBKGC) là mô hình dựa trên nhúng và áp dụng VisualBERT như một bộ mã hóa đa phương thức để nắm bắt các đặc trưng đa phương thức hợp nhất sâu sắc của các thực thể.

1.3.2 Tiếp cận lấy mẫu âm

Phương pháp lấy mẫu âm (negative sampling) nhằm mục đích tạo ra các bộ ba âm bằng cách thay thế ngẫu nhiên các thực thể để tạo ra độ tương phản giữa mẫu dương và mẫu âm trong quá trình huấn luyện.

Phương pháp này giúp cho mô hình nhúng ĐTTT đưa ra điểm số cao hơn cho các bộ ba dương. Hai mô hình hoàn thiện ĐTTT đa phương thức điển hình theo tiếp cận lấy mẫu âm là MMRNS và MANS. Mô hình MMRNS được xây dựng để xác định các mẫu âm bằng cách sử dụng kết hợp dữ liệu đa phương thức và các mối quan hệ ĐTTT phức tạp nhằm tăng cường biểu diễn ngữ nghĩa của các thực thể. Mô hình MANS-V nhằm mục đích lấy mẫu âm các nhúng ảnh từ những cái không thuộc về thực thể hiện tại để huấn luyện mô hình giúp xác định các đặc điểm hình tương ứng với từng thực thể.

1.3.3 Tiếp cận vấn đề mất cân bằng và thiếu dữ liệu

Một số mô hình hoàn thiện ĐTTT đa phương thức theo các hướng tiếp cận khác nhau xử lý mất cân bằng và thiếu dữ liệu đó là MACO (Modality Adversarial and Contrastive Framework), NativE, MNFormer. Trong đó:

Mô hình MACO được giới thiệu bởi Zhang và cộng sự [49] với việc tạo ra các đặc trưng hình ảnh bị thiếu dựa trên thông tin cấu trúc của thực thể. Sau đó, cải thiện chất lượng của đặc trưng được tạo ra thông qua việc kết hợp các cơ chế học đối nghịch (adversarial learning) và học tương phản (constrastive learning).

Mô hình NativE được đề xuất bởi Zhang và cộng sự [51] bao gồm hai mô-đun chính: Mô-đun kết hợp thích ứng kép hướng quan hệ (Relation-guided Dual Adaptive Fusion – ReDAF) sử dụng một tập hợp các trọng thích ứng (adaptive weights) cho các phương thức khác nhau, Mô-đun huấn luyện đối kháng phương thức hợp tác (Collaborative Modality Adversarial Training – CoMAT) được thiết kế để tăng cường thông tin phương thức bị mất cân bằng.

Mô hình MNFormer nhằm giải quyết sự không đồng nhất giữa phương thức cấu trúc (structural modality) và phương thức ngữ nghĩa (semantics modality). MNFormer là một khung Transformer đa lớp mới được thiết kế để thu thập thông tin cấu trúc và ngữ nghĩa đồng thời tránh các vấn đề không đồng nhất.

1.4 Hoàn thiện ĐTTT thời gian

1.4.1 Giới thiệu về Hoàn thiện ĐTTT thời gian

Dựa trên vị trí bị thiếu trong bộ bốn (h, r, t, τ), bài toán hoàn thiện ĐTTT thời gian được chia ra thành các bài toán cụ thể hơn: dự đoán thực thể, dự đoán mối quan hệ và dự đoán thời gian.

1.4.2 Khung phân loại các phương pháp Hoàn thiện ĐTTT thời gian

J. Wang và cộng sự [54] phân loại các phương pháp TKGC dựa trên việc chúng có dự báo các dữ kiện trong tương lai hay không thành hai nhóm phương pháp gồm dựa trên nội suy (Interpolation-based TKGCs) và dựa trên ngoại suy (Extrapolation-based TKGCs).

Hai mô hình hoàn thiện ĐTTT thời gian dựa trên nhúng lịch đại DE-RotatE, DE-RotatE-sinc do luận án đề xuất và các mô hình hoàn thiện ĐTTT thời gian DE-SimplE là thuộc nhóm con các phương pháp Hàm chuyển đổi và thuộc vào nhóm các phương pháp Nội suy đặc tả thời gian.

1.5 Các độ đo đánh giá hiệu năng mô hình hoàn thiện ĐTTT

Luận án sử dụng ba độ đo đánh giá hiệu năng các mô hình hoàn thiện ĐTTT là Hits@k, Mean Rank (MR), Mean Reciprocal Rank (MRR).

1.5.1. Độ đo Hits@k

Độ đo $Hits@k$ (thường là $k = 1;\ 3;\ 10$) chỉ ra xác suất dự đoán đúng trong k bộ ứng viên hàng đầu bởi mô hình. Độ đo $Hit@k$ biểu diễn khả năng của thuật toán, mô hình dự đoán chính xác mối quan hệ giữa bộ ba.

1.5.2. Độ đo xếp hạng trung bình MR

Độ đo xếp hạng trung bình MR (Mean Rank) trong mô hình hoàn thiện ĐTTT là một độ đo hiệu năng giúp đánh giá khả năng dự đoán chính xác các quan hệ bộ ba trong ĐTTT. Chỉ số Mean Rank (MR) là giá trị trung bình của tất cả các hạng. Giá trị này càng nhỏ thì càng tốt, cho thấy mô hình dự đoán đúng với vị trí cao hơn.

1.5.3. Độ đo xếp hạng tương hỗ trung bình MRR

Độ đo xếp hạng tương hỗ trung bình MRR (Mean Reciprocal Rank) là một đô đo phổ biến dùng để đánh giá hiệu năng của các hệ thống dự đoán. Đây là giá trị trung bình của nghịch đảo thứ hạng của các dự đoán đúng đầu tiên. Giá trị MRR càng lớn thì hiệu quả dự đoán của mô hình càng tốt.

1.5.4 Các độ đo khác

Một số độ đo dựa trên ma trận nhầm lẫn trong phân lớp như độ đo chính xác thống kê (accuracy), bộ độ đo hồi tưởng, chính xác (recall, precision) và F1 cũng được sử dụng.

1.6 Xu hướng nghiên cứu hoàn thiện ĐTTT và một số luận án Tiến sỹ

1.6.1 Xu hướng nghiên cứu về hàn thiện ĐTTT

T. Shen và cộng sự [9] nêu ra tám triển vọng nghiên cứu chính: Sử dụng thông tin bổ sung, đặc biệt là thông tin bên ngoài ĐTTT; Hoàn thiện ĐTTT dựa trên luật; Sử dụng các mô hình ngôn ngữ huấn luyện trước (Pre-trained Language Model); Hoàn thiện ĐTTT cho các ĐTTT miền cụ thể; Nắm bắt tương tác giữa các ĐTTT riêng biệt; Chọn không gian phù hợp hơn mô hình hóa cho hoàn thiện ĐTTT; Sử dụng học tăng cường trong hoàn thiện ĐTTT; Học đa nhiệm cho hoàn thiện ĐTTT.

1.6.2 Một số luận án Tiến sỹ liên quan trên thế giới

Mục này giới thiệu một số luận án Tiến sỹ trên thế giới tham gia vào tám triển vọng nghiên cứu về hoàn thiện ĐTTT.

1.7 Liên hệ với nghiên cứu trong luận án

Theo xu hướng nghiên cứu trên thế giới về hoàn thiện ĐTTT, luận án tập trung vào các mô hình hoàn thiện ĐTTT dựa trên học chuyển giao (Xu hướng nghiên cứu 1), hoàn thiện ĐTTT đa phương thức (Xu hướng nghiên cứu 3, 4) và hoàn thiện ĐTTT thời gian (Xu hướng nghiên cứu 3, 4)

1.7.1. Học chuyển giao cho hoàn thiện ĐTTT trong luận án

Định nghĩa 1.6 Học chuyển giao [56]: Cho một miền nguồn D_S và một tác vụ học tương ứng $\mathcal{T}_S$, một miền đích $\mathcal{D}_T$ và tác vụ học tương ứng $\mathcal{T}_T$. Học chuyển giao là phương pháp học nhằm mục đích cải thiện hàm dự đoán đích $f_T(.)$ trong $\mathcal{D}_T$ khi sử dụng tri thức trong $\mathcal{D}_S$ và $\mathcal{T}_S$, ở đây, $\mathcal{D}_S \neq \mathcal{D}_T$ hoặc $\mathcal{T}_S \neq \mathcal{T}_T$.

Các thành phần có thể chuyển giao trong mô hình học chuyển giao cho hoàn thiện ĐTTT đó là các vec-tơ nhúng thực thể học từ đồ thị nguồn, vec-tơ nhúng của quan hệ đã học để áp dụng cho quan hệ tương tự, các tham số mô hình như trọng số của hàm tính điểm, các mô hình thần kinh, kiến thức ngữ nghĩa từ văn bản như các nhúng từ giúp đại điện tốt cho thực thể, quan hệ.

Các chiến lược học chuyển giao trong hoàn thiện ĐTTT đó là: chuyển giao không cần huấn luyện lại, huấn luyện tinh chỉnh, học đa nhiệm, học chuyển giao dựa trên module thích ứng, học siêu cấp.

1.7.2 Hoàn thiện ĐTTT đa phương thức trong luận án

Trong hoàn thiện ĐTTT đa phương thức (MMKGC), phân loại bộ ba không chỉ đánh giá độ đúng/ sai dựa trên quan hệ cấu trúc mà còn cần kết hợp dữ liệu đa phương thức như hình ảnh và văn bản mô tả. Bên cạnh đó, hoàn thiện ĐTTT đa phương thức yêu cầu mô hình có khả năng xử lý khi thiếu phương thức. Luận án nghiên cứu, phân tích chi tiết mô hình AdaMF-MAT được Y. Zhang và cộng sự đề xuất năm 2024 [18], , qua đó nhận ra được các ý tưởng cải tiến và đề xuất hai mô hình hoàn thiện ĐTTT đa phương thức ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT, trong đó, ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT tương ứng sử dụng mô hình nhúng ảnh Vision Transformer (ViT) và mô hình nhúng văn bản T5.

1.7.3 Hoàn thiện ĐTTT thời gian trong luận án

Luận án thực hiện việc phân tích chi tiết mô hình DE-SimplE, qua đó nhận ra được các ý tưởng cải tiến và đề xuất hai mô hình hoàn thiện ĐTTT thời gian là DE-RotatE và DE-RotatE-sinc được cải tiến từ mô hình DE-SimplE theo ý tưởng sử dụng phép quay chiếu thay vì sử dụng phép tịnh tiến chiếu trong DE-SimplE.

1.8 Kết luận Chương 1

Tổng hợp nội dung khảo sát trong chương 1

Chương 2. Hoàn thiện ĐTTT dựa trên học chuyển giao

2.1 Mô hình hoàn thiện ĐTTT với học chuyển giao Glove – GRU – KGC

Trong phần này, luận án trình bày về kiến trúc của mô hình KGC-TransLearning và cách thức các tác giả sử dụng một mô hình huấn luyện trước để khởi tạo mô hình cho tinh chỉnh. V. Kocijan và T. Lukasiewicz sử dụng mô hình hoàn thiện ĐTTT và các bộ mã hóa được huấn luyện trước để khởi tạo mô hình và sau đó tinh chỉnh trên tập dữ liệu nhỏ hơn. Cụ thể hơn nữa, các tham số trong mô hình hoàn thiện ĐTTT được chia sẻ giữa tất cả các bộ đầu vào và được sử dụng như một khởi tạo của các tham số giống nhau của mô hình tinh chỉnh [17].

2.1.1 Bộ mã hóa

V. Kocijan và T. Lukasiewicz [17] sử dụng hai cách tiếp cận để xây dựng ánh xạ từ một thực thể đến một không gian véc tơ với số chiều thấp đó là mô hình NoEncoder và mô hình mã hóa GRU

2.1.2 Chuyển giao giữa các tập dữ liệu

V. Kocijan và T. Lukasiewicz [17] đã chỉ ra rằng quá trình huấn luyện trước luôn luôn được thực hiện bằng cách sử dụng các bộ mã hóa GRU. Bởi vì khi chuyển giao từ bộ mã hóa NoEncoder đến bất kỳ một bộ mã hóa nào khác đều yêu cầu phải có đối sánh thực thể. Trong quá trình tinh chỉnh được thực hiện bởi bộ mã hóa GRU, các tham số của nó được khởi tạo từ bộ tham số GRU và được huấn luyện trước.

2.1.3 Mô hình hoàn thiện ĐTTT GloVe-GRU-KGC

V. Kocijan và T. Lukasiewicz [17] đề xuất một mô hình hoàn thiện ĐTTT dựa trên học chuyển giao GloVe-GRU-KGC có sử dụng ba mô hình hoàn thiện ĐTTT là ConvE [32], TuckER [33] và 5⋆E [34]. Khi đó, với mỗi bộ ba (h, r, t) qua bộ mã hóa sẽ thu được các nhúng tương ứng là $\mathbf{v}_h$, $\mathbf{v}_r$ và $\mathbf{v}_t$. Các nhúng này sẽ là dữ liệu đầu vào cho các mô hình hoàn thiện ĐTTT.

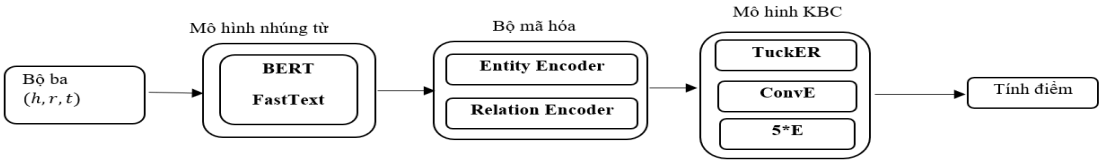
2.1.4 Ý tưởng cải tiến

Việc lựa chọn các mô hình để ánh xạ biểu diễn ĐTTT sang không gian véc-tơ đóng vai trò then chốt trong quá trình hoàn thiện đồ thị tri thức (KGC). Sự lựa chọn này quyết định chất lượng biểu diễn của thực thể và quan hệ, từ đó ảnh hưởng trực tiếp đến hiệu năng suy luận và dự đoán của mô hình. Những phép nhúng hiệu quả không chỉ giúp tạo ra các véc-tơ giàu thông tin ngữ nghĩa mà còn đặc biệt hữu ích đối với các thực thể hoặc quan hệ có kèm theo mô tả văn bản. BERT giúp trích xuất đặc trưng giàu ngữ nghĩa, hỗ trợ suy luận và dự đoán tốt hơn. Ngoài ra, trong ĐTTT có thể xuất hiện các từ mới, từ hiếm gặp hoặc tên riêng trong tập huấn luyện thì cần lựa chọn một mô hình nhúng hoạt động tốt trong ĐTTT mở với ưu điểm tốc độ tính nhúng nhanh; FastText là một lựa chọn tốt.

Bên cạnh đó, mô hình KGC-TransLearning sẽ có kết quả tốt hơn khi mô hình sử dụng bộ mã hóa GRU thay vì sử dụng bộ mã hóa NoEncoder. Từ đó, luận án đưa ra ý tưởng đề xuất để cải tiến mô hình học chuyển giao trong hoàn thiện ĐTTT như sau: Sử dụng BERT hoặc FastText để ánh xạ các từ sang không gian véc-tơ, sử dụng mô hình kết hợp GRU và kết nối đầy đủ (Fully Connected: FC) để học nhúng bộ ba. Đề xuất này vừa tận dụng sức mạnh của mô hình ngôn ngữ hiện đại để học đặc trưng ngữ nghĩa sâu, vừa khai thác khả năng mô hình hóa phụ thuộc tuần tự của GRU, từ đó tạo ra biểu diễn bộ ba giàu thông tin và nâng cao hiệu năng hoàn thiện ĐTTT.

2.2 Mô hình đề xuất BERT/FastText-GRU-KGC

Luận án chia mô hình đề xuất thành ba thành phần chính là mô hình nhúng từ, bộ mã hóa, mô hình hoàn thiện ĐTTT, xem tại Hình 2.2



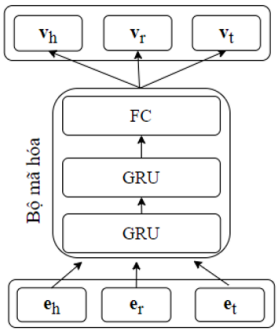
Hình 2.2. Các thành phần của mô hình đề xuất

2.2.1 Thành phần nhúng từ

Luận án đề xuất mô hình nhúng từ mới bằng cách thực nghiệm sử dụng BERT hoặc FastText thay vì sử dụng Glove.

2.2.2 Thành phần mã hóa

Mô hình đề xuất sử dụng kết hợp giữa GRU và lớp kết nối đầy đủ (Fully Connected – FC) thay vì chỉ sử dụng GRU như trong mô hình Glove-GRU-KGC gốc, chi tiết mô hình mã hóa được thể hiện tại Hình 2.4.



Hình 2.4. Kiến trúc mã hóa trong mô hình BERT/FastText-GRU-KGC

2.2.3 Thành phần hoàn thiện ĐTTT

Mô hình đề xuất cũng sử dụng ba mô hình hoàn thiện ĐTTT có hiệu năng tốt nhất đó là TuckER, ConvE và 5★E

2.3 Thực nghiệm mô hình đề xuất và đánh giá

2.3.1 Tập dữ liệu

Mô hình đề xuất cũng thực nghiệm với năm tập dữ liệu: OlpBench, ReVerb45K, ReVerb20K, FB15K237 và WN18RR. Thông tin chi tiết được thống kê tại Bảng 2.1.

Bảng 2.1. Số lượng thực thể, quan hệ và các bộ ba trong các tập dữ liệu.

	OlpBench	ReVerb20K	ReVerb45K	FB15K237	WN18RR
Thực thể	2,4M	11,1K	27K	14,5K	41,1K
Quan hệ	957K	11.1K	21,6K	237	11
Bộ ba huấn luyện	3,6M	15,5K	36K	272K	86,8K
Bộ ba xác thực	10K	1,6K	3,6K	17,5K	3K
Bộ ba kiểm thử	10K	2.4K	5.4K	20,5K	3K

2.3.2 Kịch bản thực nghiệm

1) Kịch bản thứ nhất sử dụng một mô hình được huấn luyện trước để khởi tạo các tham số và các véc-tơ nhúng, sau đó thực hiện tinh chỉnh chúng. Tương tự như trong [17], tập dữ liệu OlpBench được sử dụng để huấn luyện trước. Vì tập dữ liệu OlpBench có kích thước tương đối lớn nên luận án thực nghiệm sử dụng $1/10$ kích thước của tập dữ liệu OlpBench. Ngoài ra, các dữ liệu này được lấy một cách ngẫu nhiên.
2) Kịch bản thứ hai là huấn luyện mô hình thực nghiệm từ đầu.

Với cả hai kịch bản trên, mô hình đề xuất trong luận án đều được đánh giá trên cả bốn tập dữ liệu tiêu chuẩn trong hoàn thiện ĐTTT là ReVerb20K, ReVerb45K, FB15K237 và WN18RR.

2.3.3 Cấu hình và thiết lập thực nghiệm mô hình đề xuất

Phần này luận án sẽ trình bày về các thiết lập cho mô hình nhúng từ, mô hình mã hóa, mô hình huấn luyện trước, mô hình học chuyển giao.

- Mô hình nhúng từ: Với mô hình này sẽ sử dụng $\text{BERT}_{\text{large}}$ cho mô hình huấn luyện trước để khởi tạo các nhúng bộ ba và các véc-tơ nhúng được khởi tạo với số chiều là 100. FastText-GRU-KBC các quan hệ và thực thể được FastText huấn luyện trước sẽ được khởi tạo bằng nhúng từ tương ứng trong đó.

Các từ còn lại không trong mô hình huấn luyện trước thì được khởi tạo một cách ngẫu nhiên. Trong trường hợp này, các thực thể và các quan hệ sẽ được ánh xạ vào không gian véc tơ nhúng với kích thước là 300.

- Mô hình mã hóa: BERT-GRU-KBC: kích thước hidden_size của lớp GRU của TuckER và 5⋆E sẽ là 100, còn ConvE là 500. Lớp kết nối đầy đủ FC sẽ trả về một véc tơ nhúng đầu ra có kích thước bằng kích thước của hidden_size trong GRU. FastText-GRU-KBC: kích thước của hidden_size của thành phần cả ba mô hình hoàn thiện ĐTTT đều là 300.
- Mô hình huấn luyện trước: huấn luyện trên tập dữ liệu OlpBench với kích thước lô huấn luyện là 1024. Tốc độ học được chọn từ $\{1.10^{-4}, 3.10^{-4}\}$, tỷ lệ dropout được chọn từ $\{0.3, 0.4\}$ đối với mô hình TuckER và ConvE. Tuy nhiên, đối với mô hình 5⋆E, tỷ lệ dropout sẽ không được sử dụng, thay vào đó sử dụng chính quy hóa N3 (N3 regularization) với các giá trị được lấy trong $\{0.1, 0.3\}$ theo [64]. Với mô hình TuckER và ConvE thì số chiều của véc tơ nhúng được lựa chọn là 100 và 300 còn 5⋆E, được lựa chọn là 300 và 500. Kết quả tốt nhất của mỗi mô hình trong quá trình chạy sẽ được lưu lại và sử dụng cho việc học chuyển giao. Tất cả mô hình đều được huấn luyện với 100 epochs trên cụm DGX sử dụng Nvidia V100 GPU.
- Mô hình học chuyển giao: Tất cả quá trình học chuyển giao được thiết lập tương tự như huấn luyện trước. Tuy nhiên, các mô hình học chuyển giao được huấn luyện với số lượng epochs nhiều hơn (cỡ 200 epochs) và một số lượng lớn các siêu tham số. Ở đó, tốc độ học được lựa chọn từ $\{3.10^{-5}, 3.10^{-4}, 1.10^{-4}, 1.10^{-2}\}$. Tỷ lệ dropout được lựa chọn từ $\{0.03, 0.1, 0.2, 0.3, 0.4\}$. Hệ số chính quy hóa N3 (N3 regularization) được lựa chọn từ $\{0.03, 0.1, 0.3\}$ và kích thước lô được lựa chọn từ $\{512, 1024, 2048, 4096\}$. Số chiều của các véc tơ nhúng tương tự như số chiều của véc tơ nhúng trong mô hình huấn luyện trước. Việc học chuyển giao được thực chiện trên GPU Tesla T4.

2.3.4 Kết quả và đánh giá

I. Kết quả khi huấn luyện trước.

Bảng 2.3. So sánh giữa các mô hình huấn luyện trước của mô hình đề xuất với kết quả tốt nhất của mô hình gốc trong [17] đối với tập dữ liệu OlpBench. Với mỗi cột giá trị đánh giá thì kết quả tốt nhất được in đậm và kết quả gạch chân là tốt thứ hai.

Mô hình OKBC	Mô hình nhúng từ	MR ↓	MRR↑	H@10↑
TuckER	BERT	**4.300**	**0,100**	**0,997**
	FastText	176.400	0,005	0,011
ConvE	BERT	229.400	0,004	0,009
	FastText	157.800	0,006	0,012
5⋆E	BERT	1341000	0,010	0,001
	FastText	167400	0,004	0,011
GRU_5⋆E [17]	Glove	<u>60100</u>	<u>0,055</u>	<u>0,101</u>

Mô hình đề xuất sử dụng nhúng từ với BERT (BERT-GRU-KBC) cho kết quả tốt hơn khi sử dụng nhúng từ với FastText (FastText-GRU-KBC). Hơn nữa, mô hình đề xuất, BERT-GRU-TuckER cho kết quả vượt trội so với kết quả tốt nhất từ mô hình gốc GlovE-GRU-KBC. Theo cách đánh giá thì kết quả thu được

vẫn chưa là tốt nhất, do các thực nghiệm mới chỉ được tiến hành trên một phần của tập dữ liệu OlpBench, trong khi vẫn đánh giá trên toàn bộ tập kiểm thử gốc.

II. Kết quả khi học chuyển giao

Bảng 2.4. So sánh kết quả giữa mô hình đề xuất và mô hình gốc trên các tập dữ liệu OKBC khác nhau

Mô hình KBC	Nhúng từ	Học chuyển giao	ReVerb20K			ReVerb45K		
			MR ↓	MRR↑	H@10↑	MR ↓	MRR↑	H@10↑
TuckER	BERT	Có	239	*0,977*	*0,977*	896	**0,955**	*0,964*
		Không	**46**	**0,996**	**0,996**	727	0,246	0,245
	FastText	Có	322	0,374	0,524	1196	0,265	0,387
		Không	765	0,317	0,436	2748	0,213	0,287
ConvE	BERT	Có	622	0,305	0,427	1432	0,159	0,276
		Không	1192	0,211	0,317	1390	0,286	0,452
	FastText	Có	390	0,347	0,489	1167	0,263	0,382
		Không	417	0,338	0,472	1472	0,275	0,399
5★E	BERT	Có	187	0,124	0,974	1094	0,112	0.761
		Không	*107*	0,493	0,988	**126**	0,500	**0,995**
	FastText	Có	293	0,392	0,548	842	0,289	0,551
		Không	563	0,331	0,466	1598	0,240	0,375
GRU_5★E [17]	Glove	Có	202	0,417	0,586	*596*	0,382	0,537
GRU_TuckER [22]	Glove	Có	245	0,397	0,558	706	0,331	0,477
GRU_ConvE [17]	Glove	Có	184	0,409	0,573	600	*0,357*	0,509
GRU_5★E [65] *	Glove	Có	201	0,415	0,589	572	0,376	0,534
GRU_TuckER [65] *	Glove	Có	256	0,394	0,556	808	0,322	0,458
GRU_ConvE [65] *	Glove	Có	184	0,408	0,570	582	0,353	0,507

Qua kết quả thực nghiệm, luận án nhận thấy phần lớn các mô hình cho kết quả tốt hơn khi được tinh chỉnh với mô hình đã qua huấn luyện trước trên tập dữ liệu OlpBench. Mô hình FastText-GRU-KBC cho kết quả với kịch bản huấn luyện trước cải thiện tốt trên cả ba độ đo so với kết quả kịch bản huấn luyện từ đầu.

Mô hình BERT-GRU-KBC, kết quả kịch bản huấn luyện từ đầu tốt hơn kết quả kịch bản huấn luyện trước khi sử dụng mô hình TuckER và mô hình 5★E. Điều này xảy ra có thể được giải thích rằng do trong mô hình 5★E không chia sẻ các tham số trong quá trình học chuyển giao, mà việc này chỉ đóng vai trò là khởi tạo nhúng bộ ba.

Còn với mô hình TuckER có thể giải thích một trong những nguyên nhân chính dẫn đến hiệu năng suy giảm khi tinh chỉnh mô hình từ OLPBench sang ReVerb20K là sự khác biệt gần như tuyệt đối về tập thực thể và quan hệ giữa hai bộ dữ liệu khi có tới 99,9% mẫu kiểm thử trong ReVerb20K và 99,3% trong ReVerb45K chứa ít nhất một thực thể hoặc quan hệ không xuất hiện trong OLPBench [17].

Ngoài ra, kết quả của mô hình đề xuất BERT-GRU-KGC vẫn tốt hơn rất nhiều so với kết quả của mô hình đề xuất FastText-GRU-KGC. Cụ thể, khi so sánh kết quả tốt nhất của hai mô hình trên các tập dữ liệu (bao gồm cả kịch bản huấn luyện từ đầu và kịch bản huấn luyện trước).

Bảng 2.5. So sánh kết quả giữa các mô hình KBC trên một số ĐTTT đã chuẩn hóa.

Mô hình KBC	Nhúng từ	Học chuyển giao	FB15K237			WN18RR		
			MR↓	MRR↑	H@10↑	MR↓	MRR↑	H@10↑
TuckER	BERT	Có	1818	0,209	**0,626**	5186	0,166	0,271
		Không	1163	0,156	0,249	**615**	**0,985**	**0,985**
	FastText	Có	174	0,334	0,517	4013	0,432	0,471

		Không	187	0,328	0,526	4112	0,353	0,452
	GloVe [17]	Có	151	**0,363**	0,550	3456	0,467	0,529
ConvE	BERT	Có	704	0,181	0,288	2540	0,401	0,498
		Không	456	0,240	0,368	3897	0,389	0,442
	FastText	Có	188	0,317	0,490	4485	0,421	0,476
		Không	193	0,314	0,491	4782	0,398	0,450
	GloVe [17]	Có	200	0,325	0,510	5792	0,435	0,486
5⋆E	BERT	Có	996	0,166	0,256	2675	0,421	0,512
		Không	986	0,152	0,246	2986	0,388	0,493
	FastText	Có	168	0,336	0,521	3000	0,379	0,497
		Không	174	0,315	0,492	2880	0,386	0,502
	GloVe [17]	Có	**143**	0,357	0,544	2636	0,492	0,582

Từ bảng này, luận án nhận thấy rằng các kết quả tốt nhất thường rơi vào mô hình đã được huấn luyện trước, phần lớn kết quả của mô hình đề xuất đều tốt hơn kết quả tốt nhất của mô hình gốc của Kocijan và Lukasiewicz tại [17], kết quả của mô hình nhúng từ BERT thường tốt nhất. Ngoài ra, với từng thành phần mô hình nhúng từ và từng thành phần mô hình hoàn thiện ĐTTT, thì kết quả khi sử dụng học chuyển giao phần lớn tốt hơn kết quả khi huấn luyện mô hình từ đầu.

Ngoài ra, vào năm 2024, V. Kocijan, M. Jang và T. Lukasiewicz [65] công bố kết quả nghiên cứu mới tại tạp chí "Artificial Intelligence". Trong đó phần đầu tiên giữ nguyên cách tiếp cận học chuyển giao trong [17] và cải tiến thực nghiệm bằng cách lấy giá trị trung bình của 05 lần thực nghiệm thay vì đưa ra kết quả thực nghiệm tốt nhất như [17]. Các kết quả thực nghiệm [65] được đưa vào ba dòng cuối của **Error! Reference source not found.**. Qua so sánh, luận án nhận thấy các kết quả [65] khá tương đồng với các kết quả [17]. Tuy nhiên, các kết quả này cũng không tốt hơn so với kết quả của mô hình đề xuất trong luận án.

2.4 Kết luận Chương 2

Mục này tổng kết các kết quả nghiên cứu ở chương 2

Chương 3. Hoàn thiện ĐTTT đa phương thức

3.1 Mô hình Hoàn thiện ĐTTT đa phương thức AdaMF-MAT

3.1.1 Giới thiệu chung

Để giải quyết vấn đề mất cân bằng về thông tin phương thức trong các mô hình hoàn thiện ĐTTT đa phương thức, Zhang và cộng sự [18] đã đề xuất và xây dựng một mô hình hoàn thiện ĐTTT đa phương thức mới bằng cách kết hợp giữa việc hợp nhất đa phương thức thích ứng AdaMF (Adaptive Multi-modal Fusion - AdaMF) và huấn luyện đối kháng theo phương thức MAT (Modality Adversarial Training – MAT) để tăng cường và sử dụng hiệu quả thông tin đa phương thức. Đối với mỗi thực thể $e \in \mathcal{E}$, luận án ký hiệu các nhúng để mô tả các đặc trưng phương thức khác nhau như sau: nhúng cấu trúc ký hiệu là $\mathbf{e}_s$; nhúng ảnh ký hiệu là $\mathbf{e}_v$; nhúng văn bản ký hiệu là $\mathbf{e}_t$. Với mỗi mối quan hệ $r \in \mathcal{R}$, luận án ký hiệu nhúng mối quan hệ r này là $\mathbf{e}_r$.

3.1.2 Bộ mã hóa đặc trưng đa phương thức

Đối với ảnh, các tác giả đề xuất áp dụng bộ mã hóa ảnh được huấn luyện trước PVE (được đề xuất bởi Bao và cộng sự tại [68]) để trích rút đặc trưng ảnh $\mathbf{f}_v$ đối với mỗi thực thể e bằng cách sử dụng hàm trung bình và chiếu nó lên không gian nhúng để thu được nhúng ảnh. Cụ thể công thức dưới đây

$$f_v = \frac{1}{|\mathcal{V}_e|} \sum_{img_i \in \mathcal{V}_e} PVE(img_i) \qquad (3\text{-}1)$$
$$\boldsymbol{e}_v = \boldsymbol{W}_v . \boldsymbol{f}_v + \boldsymbol{b}_v$$

Trong đó, $\mathbf{W}_v$ và $\mathbf{b}_v$ là các tham số của lớp chiếu ảnh.

Cũng tương tự như vậy, đối với nhúng văn bản, Zhang và cộng sự cũng trích rút đặc trưng văn bản của mỗi thực thể với mô tả của nó và bộ mã hóa văn bản huấn luyện trước PTE trong BEiT (Reimers và Gurevych tại [66]). Mã CLS (mô hình BERT [53]) được áp dụng để thu được đặc trưng văn bản cấp độ câu $\mathbf{f}_t$. Sau đó, đặc trưng văn bản $\mathbf{f}_t$ được chiếu xuống không gian nhúng với các tham số lớp $\mathbf{W}_t$, $\mathbf{b}_t$ để thu được nhúng $\mathbf{e}_t$, được mô tả cụ thể bởi công thức dưới đây.

$$\boldsymbol{e}_t = \boldsymbol{W}_t . \boldsymbol{f}_t + \boldsymbol{b}_t \qquad (3\text{-}2)$$

Cuối cùng, lớp nhúng cấu trúc: Nhằm ánh xạ thực thể và quan hệ trong ĐTTT vào không gian nhúng thông qua các kỹ thuật nhúng truyền thống.

3.1.3 Hợp nhất đa phương thức thích ứng

Đối với mỗi thực thể bất kỳ thì gọi nhúng của nó là $\mathbf{e}_m$, với $m \in \mathcal{M}$, mô hình AdaMF được tính toán theo công thức dưới đây

$$\alpha_m = \frac{exp(\boldsymbol{w}_m \oplus tanh(\boldsymbol{e}_m))}{\sum_{n \in \mathcal{M}} exp(\boldsymbol{w}_m \oplus tanh(\boldsymbol{e}_n))} \qquad (3\text{-}3)$$

Trong đó α_m được gọi là trọng số phù hợp, toán tử $\oplus$ gọi là tích từng điểm (pointwise product) và $\mathbf{w}_m$ véc tơ có thể học của phương thức m. Tiếp theo, nhúng tổng hợp các phương thức ký hiệu là $\mathbf{e}_{joint}$ và được tính theo công thức dưới đây. Với mỗi thực thể khác nhau, mô hình AdaMF cho phép học các trọng số phương thức khác nhau.

$$\boldsymbol{e}_{joint} = \sum_{m \in \mathcal{M}} \alpha_m \boldsymbol{e}_m \qquad (3\text{-}4)$$

Sau đó, các tác giả sử dụng hàm tính điểm từ mô hình hoàn thiện ĐTTT RotatE được đề xuất bởi Sun và cộng sự tại [38] để đo độ hợp lý của các bộ ba. Công thức hàm tính điểm được mô tả dưới đây.

$$\mathcal{F}(h, r, t) = \left\| \boldsymbol{h}_{joint} \circ \boldsymbol{e}_r - \boldsymbol{t}_{joint} \right\| \qquad (3\text{-}5)$$

Mục đích của mô hình là tăng điểm số đối với các bộ ba dương (bộ ba đúng) và giảm điểm số của bộ ba âm (bộ ba sai) để tối thiểu điểm số của $\mathcal{L}_{kgc}$, do vậy các tác giả huấn luyện các nhúng bằng cách sử dụng hàm mất mát sigmoid được đề xuất bởi Sun và cộng sự tại [38].

$$\mathcal{L}_{kgc} = \frac{1}{|\mathcal{T}|} \sum_{(h,r,t)\in\mathcal{T}} \left(-log\, \sigma(\gamma - \mathcal{F}(h,r,t) - \sum_{i=1}^{K} p_i\, log\, \sigma(\mathcal{F}(h_i', r_i', t_i') - \gamma) \right) \quad (3\text{-}6)$$

trong đó σ là hàm sigmoid, γ là biên độ (margin) và p_i là trọng số tự đối kháng cho mỗi bộ ba âm (bộ ba tiêu cực) (h_i', r_i', t_i') được tạo ra bởi phương pháp lấy mẫu âm (dựa theo nghiên cứu của Bordes và cộng sự tại [36]). Trọng số p_i được cho bởi công thức dưới đây

$$p_i = \frac{exp\left(\beta \mathcal{F}(h_i', r_i', t_i')\right)}{\sum_{j=1}^{K} exp\left(\beta \mathcal{F}\left(h_j', r_j', t_j'\right)\right)} \quad (3\text{-}7)$$

trong đó, β là tham số và K là số lượng mẫu âm. Tác vụ chính của mô hình này là tối thiểu hàm số $\mathcal{L}_{kgc}$.

3.1.4 Huấn luyện đối kháng theo phương thức

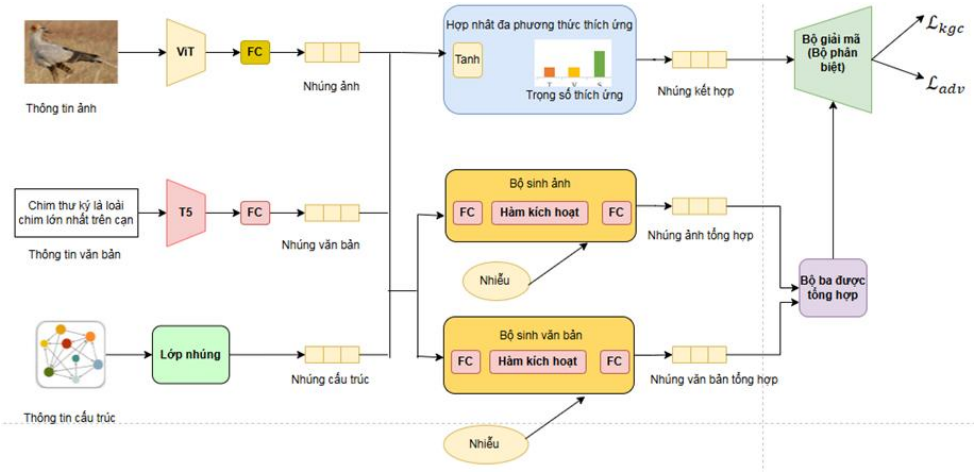
Các tác giả đã thiết kết một cơ chế huấn luyện đối kháng theo phương thức MAT để cải tiến việc mất cân bằng thông tin đa phương thức. MAT sử dụng một bộ tạo (Generator) G để tạo ra các mẫu đối nghịch và một bộ phân biệt (Discriminator) D để đo lường tính hợp lý của chúng và tuân thủ theo mô hình AT tổng quát.

Thuật toán huấn luyện của mô hình AdaMF-MAT		
Đầu vào		Một lô của bộ ba huấn luyện $\mathcal{B}$ lấy mẫu từ tập $\mathcal{T}$, thông tin đa phương thức của các thực thể, bộ phân biệt D và bộ sinh G của mô hình AdaMF
Đầu ra		Mô hình hoàn thiện ĐTTT đa phương thức với bộ phần biệt D được huấn luyện với MAT
1	**for**	Bộ ba $(h,r,t) \in \mathcal{B}$ **do**
		// Huấn luyện bộ phân biệt D
2		Lấy các nhúng tổng hợp $\mathbf{h}_{joint}$, $\mathbf{t}_{joint}$
3		Tính toán điểm số của bộ ba $\mathcal{F}(h,r,t)$ và bộ ba âm $\mathcal{F}(h_i', r_i', t_i')$
4		Tính hàm mất mát kgc $\mathcal{L}_{kgc}$
5		Tạo tập ví dụ đối kháng $\mathcal{S}$ với bộ sinh G
6		Tính toán hàm mất mát đối kháng $\mathcal{L}_{adv}$
7		Tính toán hàm mất mát tổng $\mathcal{L}_{kgc} + \lambda\mathcal{L}_{adv}$
8		Lan truyền ngược và tối ưu bộ phân biệt D
		//Huấn luyện bộ sinh G
9		Lấy các nhúng tổng hợp $\mathbf{h}_{joint}$, $\mathbf{t}_{joint}$
10		Tạo tập ví dụ đối kháng $\mathcal{S}$ với bộ sinh G
11		Tính toán hàm mất mátđối kháng $\mathcal{L}_{adv}$
12		Lan truyền ngược và tối ưu bội sinh G
13	**End**	

Hình 3.2. Thuật toán huấn luyện mô hình AdaMF-MAT [18].

3.1.5 Ý tưởng cải tiến

Đối với ĐTTT đa phương thức, việc lựa chọn mô hình học nhúng phù hợp cho mỗi phương thức sẽ ảnh hưởng trực tiếp đến hiệu năng của mô hình hoàn thiện ĐTTT vì mỗi phương thức mang một dạng thông tin riêng biệt, mô hình nhúng tốt phải bảo toàn được ngữ nghĩa của từng phương thức. Hơn nữa, việc tích hợp các véc-tơ nhúng riêng lẻ vào không gian chung đòi hỏi tính đồng nhất cao. Trong các mô hình nhúng ảnh hiện nay, có thể phân loại theo ba hướng chính: Mạng CNN truyền thống, Mạng Transformer thị giác, Mô hình học tự giám sát. Mô hình ViT (Vision Transformer) có những ưu thế vượt trội như với cơ chế tự chú ý là một cơ chế cốt lõi trong kiến trúc Transformer giúp ViT rất hiệu quả trong việc học ngữ cảnh dài và quan hệ giữa các thành phần dữ liệu, ViT cũng có ưu điểm thực tiễn: nhẹ hơn các mô hình như BEiT, dễ dùng với các ảnh đã có nên tốc độ và hiệu quả của ViT nhanh hơn, đây cũng là một yếu tố quan trọng trong huấn luyện. T5 (Text-to-Text Transfer Transformer) có kiến trúc tổng thể dạng Encoder – Decoder, điều này tạo ra nhúng mạnh ở Encoder, đồng thời sinh văn bản chất lượng ở Decoder mà các mô hình như BERT hoặc GPT không thể làm cùng lúc. Luận án nhận thấy rằng mô hình ViT và T5 có kiến trúc tương thích vì cùng dùng Transformer nên sự kết hợp giữa hai vector đầu ra dễ dàng hơn, giảm độ phức tạp và không cần adapter quá lớn nên đưa ra ý tưởng đề xuất cải tiến: thay nhúng ảnh bằng ViT, nhúng văn bản bằng T5. Cách kết hợp này chưa được khai thác nhiều trong hoàn thiện ĐTTT, đặc biệt là trong nhiệm vụ hoàn thiện đồ thị tri thức đa phương thức.



Hình 3.4. Mô hình đề xuất T5-ViT-AdaMF-MAT

3.2 Mô hình đề xuất ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT

Luận án đề xuất hai mô hình cải tiến từ mô hình AdaMF-MAT như sau: Mô hình ViT-AdaMF-MAT chỉ thay thế mô-đun nhúng ảnh của AdaMF-MAT bằng ViT; Mô hình T5-ViT-AdaMF-MAT thay thế mô-đun nhúng văn bản bằng T5 và thay thế mô-đun ảnh bằng ViT.

3.2.1 Kiến trúc mô hình T5-ViT-AdaMF-MAT đề xuất

Kiến trúc mô hình T5-ViT-AdaMF-MAT xem tại Hình 3.4; kiến trúc của ViT-AdaMF-MAT là hoàn toàn tương tự. Sau đây, luận án trình bày về các thành phần chính trong T5-ViT-AdaMF-MAT.

3.2.2 Nhúng ảnh bằng ViT

Các tác giả phân chia một ảnh thành các phần nhỏ và cung cấp một dãy các nhúng tuyến tính tương ứng cho các phần này để tạo thành đầu vào cho mô hình Transformer.

Mô hình Transformer chuẩn nhận đầu vào là dãy các nhúng token 1 chiều. Khi đó, để xử lý ảnh hai chiều, các tác giả định hình lại ảnh 2 chiều $x \in \mathbb{R}^{H\times W\times C}$ thành một chuỗi các mảng hai chiều phẳng $x_p \in \mathbb{R}^{N\times(P^2C)}$, ở đó (H, W) là độ phân giải của ảnh gốc, C là số lượng các kênh, (P, P) là độ phân giải mỗi phân mảnh ảnh, và $N = HW/P^2$ là kết quả số lượng các phân mảnh (cũng đóng vai trò là chuỗi đầu vào hiệu quả cho Transformer). Transformer sử dụng véc-tơ phẳng với kích thước D chiều thông qua tất cả các lớp của nó. Do vậy, các tác giả trải phẳng các chuỗi này ra và ánh xạ nó với phép chiếu tuyến tính D chiều được cho bởi phương trình huấn luyện dưới đây. Đầu ra của phép chiếu này là những nhúng phân mảnh.

$$\boldsymbol{z}_0 = \left[\boldsymbol{x}_{class}; \boldsymbol{x}_p^1\boldsymbol{E}; \boldsymbol{x}_p^1\boldsymbol{E}; \dots; \boldsymbol{x}_p^1\boldsymbol{E}\right] + \boldsymbol{E}_{pos} \qquad (3\text{-}11)$$
$$\boldsymbol{E} \in \mathbb{R}^{(P^2\cdot C)\times D}, \boldsymbol{E}_{pos} \in \mathbb{R}^{(N+1)\times D}$$

Tương tự như BERT, các tác giả thêm một nhúng huấn luyện vào dãy các nhúng phân mảnh ($\mathbf{z}_0^0 = \mathbf{x}_{class}$) với trạng thái của nó tại đầu ra của bộ mã hóa Transformer ($\mathbf{z}_L^0$) đóng vai trò là biểu diễn ảnh $\mathbf{y}$ (Công thức (3-12)).

$$\mathbf{y} = LN(\mathbf{z}_L^0) \qquad (3\text{-}12)$$

Cả hai quá trình huấn luyện trước và tinh chỉnh với một đầu phân loại được gắn vào $\mathbf{z}_L^0$. Đầu phân loại này được thực hiện bởi một MLP với một lớp ẩn tại thời gian huấn luyện trước và một lớp tuyến tính duy nhất tại thời điểm tinh chỉnh.

Các nhúng vị trí được thêm vào các nhúng phân mảnh để giữ lại thông tin vị trí. Các tác giả sử dụng các nhúng vị trí 1D tiêu chuẩn. Dãy kết quả của các véc tơ nhúng đóng vai trò là đầu vào cho bộ mã hóa.

Bộ mã hóa Transformer do Vaswani và cộng sự đề xuất tại [72] bao gồm các lớp xen kẽ với cơ chế tự chú ý đa đầu (Multiheaded Self-Attention - MSA) và các khối Perceptron đa tầng (Multilayer Perceptron – MLP) (Công thức 3.13 và 3.14). Chuẩn hóa lớp (Layernorm – LN) được áp dụng trước mỗi khối, và các kết nối dư sau mỗi khối. Khối Perceptron đa tầng chứa hai lớp với một hàm phi tuyến GELU (Gauss Error Linear Unit).

$$\mathbf{z}_l' = MSA\left(LN(\mathbf{z}_{l-1})\right) + \mathbf{z}_{l-1} \qquad (3\text{-}13)$$
$$l = 1 \dots L$$

$$\mathbf{z}_l = MLP\left(LN(\mathbf{z}_l')\right) + \mathbf{z}_l' \qquad (3\text{-}14)$$
$$l = 1 \dots L$$

Do chỉ có các lớp MLP là cục bộ và có tính chất bất biến, trong khi đó các lớp tự chú ý (self-attention) mang tính chất toàn cục cho nên cấu trúc lân cận hai chiều được sử dụng rất hạn chế. Vì vậy, ViT ít thiên vị quy nạp ảnh hơn hơn so với các mô hình CNN trước đó.

Trong mô hình đề xuất, đối với nhúng ảnh luận án sử dụng mô hình Vision Transformer (ViT). Khi đó, đối với mỗi thực thể $e \in \mathcal{E}$ có ảnh v, luận án sẽ trích rút đặc trưng ảnh $\mathbf{f}_v$. Do đó, Công thức (3-1) được thay đổi như sau.

$$\mathbf{f}_v = \frac{1}{|\mathcal{V}_e|} \sum_{img_i \in \mathcal{V}_e} ViT(img_i) \quad (3\text{-}15)$$

3.2.3 Nhúng văn bản bằng T5

Theo Raffel và cộng sự tại [71], ý tưởng cơ bản trong mô hình T5 là giải quyết các vấn đề xử lý văn bản như một bài toán "text-to-text", tức là đầu vào nhận một văn bản và tạo ra một văn bản mới tại đầu ra. Theo đó, Raffel và cộng sự tại triển khai mô hình dựa trên kiến trúc Transformer gốc, bao gồm: bộ giải mã và bộ mã hóa. Bộ mã hóa bao gồm một chồng các khối, mỗi khối gồm hai thành phần con: một lớp tự chú ý và một mạng truyền thẳng (feed forward network). Bộ giải mã có cấu trúc tương tự như bộ mã hóa ngoại trừ nó bao gồm một cơ chế chú ý tiêu chuẩn sau mỗi lớp tự chú ý, cơ chế này tập trung vào đầu ra của bộ giải mã. Đầu ra của khối giải mã cuối cùng được đưa vào một tầng dày đặc với đầu ra là hàm softmax, trong đó các trọng số được chia sẻ với ma trận nhúng đầu vào. Tất cả cơ chế chú ý trong Transformer được chia thành các đầu độc lập, các đầu ra từ những đầu này được ghép nối trước khi chúng tiếp tục được xử lý thêm.

C. Raffel và cộng sự [71] sử dụng một dạng đơn giản hóa của các nhúng vị trí, trong đó mỗi nhúng là một số vô hướng và được thêm vào logit tương ứng sử dụng cho tính toán các trọng số chú ý. Để tăng hiệu quả, các tác giả chia sẻ các tham số nhúng vị trí giữa tất cả các lớp trong mô hình, mặc dù trong một lớp cụ thể, mỗi đầu chú ý sử dụng một nhúng vị trí được học khác nhau.

Khi đó đối với mô hình T5-Vit-AdamMF-MAT, thành phần đặc trưng văn bản cấp độ câu $\mathbf{f}_t$ trong Công thức (3-2) thu được bằng công thức sau:

$$\mathbf{f}_t = T5(text) \quad (3\text{-}16)$$

Sau đó, đặc trưng văn bản $\mathbf{f}_t$ được chiếu xuống không gian nhúng với các tham số lớp $\mathbf{W}_t$, $\mathbf{b}_t$ để thu được nhúng $\mathbf{e}_t$ (chi tiết xem tại Công thức (3-2)).

3.3 Thực nghiệm mô hình đề xuất và đánh giá

3.3.1 Cấu hình phần cứng thực nghiệm

Phần này trình bày về cấu hình phần cứng thực nghiệm với GPU(Nvidia P100), tốc độ xử lý GPU(1,32GHz), dung lượng GPU(16 GB), dung lượng RAM(12GB), dung lượng bộ nhớ (5GB).

3.3.2 Tập dữ liệu

Giống các thực nghiệm được thực hiện như trong mô hình gốc của Y. Zhang và cộng sự [18], luận án cũng đánh giá hiệu năng của mô hình đề xuất thông qua các tập dữ liệu đa phương thức như DB15K, MKG-W và MKG-Y.

Bảng 3.2. Thông tin các tập dữ liệu

Tập dữ liệu	Số lượng thực thể	Số lượng quan hệ	Tập huấn luyện	Tập xác thực	Tập kiểm thử	Số lượng ảnh	Số lượng văn bản

DB15K	12842	279	79222	9902	9904	12818	9078
MKG-W	15000	169	34196	4276	4274	14463	14123
MKG-Y	15000	28	21310	2665	2663	14244	12305

3.3.3 Phần mềm thực nghiệm và kịch bản thực nghiệm

Luận án tiến hành hai kịch bản cài đặt nhúng văn bản:

Nhúng văn bản bằng BERT (như mô hình gốc AdaMF-MAT) cho kết quả đầu ra là một ten-xơ có chiều là (số thực thể, 768) với dữ liệu đầu vào là một mảng các chuỗi văn bản có nội dung là tên các thực thể được ánh xạ tương ứng với chỉ số của chúng trong tập dữ liệu.

Nhúng văn bản bằng T5-small [71] với đầu vào là kết quả từ tokenizer. Cụ thể, luận án dùng attention mask để chỉ định các vị trí của token thực (tức là các token không phải padding). Tiếp theo, luận án lấy đầu ra từ các bộ mã hóa của mô hình T5, rồi sử dụng attention mask để thiết lập trọng số cho các nhúng, cụ thể: 0 đối với padding và 1 đối với các token còn lại. Cuối cùng, luận án tính giá trị trung bình có trọng số đối với các token không phải padding. Từ đó, thu được nhúng văn bản của các thực thể.

Đối với mô hình nhúng ảnh, luận án sử dụng mô hình ViT (Vision Transformer [50]) với các thông số mặc định. Dữ liệu đầu vào của mô hình này là những tập ảnh khác nhau, trong đó mỗi tập ảnh sẽ tương ứng với một thực thể. Từ mỗi tập ảnh, mô hình nhúng ảnh sinh ra một tensor trung bình của các ảnh trong tập đó (Lưu ý, nếu thực thể không có ảnh thì ngầm định một tensor gồm các giá trị là 1). Khi đó, đầu ra của mô hình nhúng ảnh là các tensor với số chiều là (số thực thể, 768).

Trong quá trình sinh nhúng, cần lưu ý sao cho các véc tơ nhúng của thực thể có cùng thứ tự với nhau, nghĩa là véc tơ nhúng văn bản của thực thể phải cùng vị trí với véc tơ nhúng ảnh của thực thể đó.

Thông số cài đặt chi tiết theo từng tập dữ liệu được cho tại Bảng 3.3.

Bảng 3.3. Thông số cài đặt cho từng tập dữ liệu

Tập dữ liệu / Thông số	DB15K	MKG-W	MKG-Y
EMB_DIM	250	200	200
NUM_BATCH	1024	1024	1024
MARGIN	12	12	4
LR	$2e-5$	$2e-5$	$2e-5$
NEG_NUM	128	128	128
EPOCHS	1000	1000	1000

3.3.4 Độ đo hiệu năng

Luận án cũng đánh giá các mô hình trên các độ đo là: MRR và $Hits@n$ (với $n = 1; 3; 10$).

3.3.5 Kết quả thực nghiệm

Bảng 3.4. Kết quả thực nghiệm các mô hình. Giá trị in đậm là tốt nhất, giá trị gạch chân là tốt thứ hai theo cột

Mô hình	**Tập dữ liệu DB15K**				**Tập dữ liệu MKG-W**				**Tập dữ liệu MKG-Y**			
	MRR	$Hit@1$	$Hit@3$	$Hit@10$	MRR	$Hit@1$	$Hit@3$	$Hit@10$	MRR	$Hit@1$	$Hit@3$	$Hit@10$

AdaMF-MAT [23]	35,14	25,3	41,11	52,92	35,85	29,04	**39,01**	**48,42**	38,57	34,34	40,59	45,76
ViT-AdaMF-MAT	**36,04**	**26,39**	**41,88**	**53,53**	**36,01**	**29,75**	38,69	47,59	38,63	34,83	40,63	45,29
T5-ViT-AdaMF-MAT	35,85	26,25	41,63	53,03	35,33	28,88	38,08	47,00	**39,34**	**35,09**	**41,07**	**46,56**

3.3.6 Bàn luận

Nhìn vào bảng kết quả, luận án nhận thấy hầu hết các mô hình đề xuất đều cho kết quả tốt hơn mô hình gốc AdaMF-MAT[18]. Cụ thể:

Với tập dữ liệu DB15K, hiệu năng của hai mô hình đề xuất đều tốt hơn hiệu năng của mô hình gốc trên cả bốn chỉ số. Trong đó, mô hình ViT-AdaMF-MAT có kết quả tốt nhất trên cả bốn chỉ số so với mô hình đề xuất T5-ViT-AdaMF-MAT và mô hình gốc AdaMF-MAT.

Với tập dữ liệu MKG-W: Hiệu năng của mô hình đề xuất T5-ViT-AdaMF-MAT đều không tốt hơn hiệu năng của mô hình gốc AdaMF-MAT trên cả bốn chỉ số. Hiệu năng của mô hình đề xuất ViT-AdaMF-MAT kém hơn hiệu năng của mô hình gốc trên các bộ chỉ số $Hit@3$ và $Hit@10$. Đối với các bộ chỉ số MRR và $Hit@1$ thì hiệu năng của mô hình đề xuất ViT0-AdaMF-MAT có cải thiện so với hiệu năng của mô hình gốc.

Với tập dữ liệu MKG-Y, hiệu năng của hai mô hình đề xuất đều tốt hơn hiệu năng của mô gốc trên cả bốn chỉ số. Tuy nhiên, trong trường hợp này hiệu năng của mô hình T5-ViT-AdaMF-MAT tốt nhất khi so sánh trên cả bốn chỉ số.

Bên cạnh đó, sau khi xem xét hai tập dữ liệu MKG-W và MKG-Y, luận án nhận thấy rằng mức độ chi tiết của phần mô tả của các thực thể trong tập dữ liệu MGK-W rất hạn chế. Do vậy, mô hình T5 không phát huy được hiệu quả và thế mạnh của nó trên các văn bản dài. Ngoài ra, tỷ lệ ảnh trên các thực thể trong tập dữ liệu MKG-W ít hơn là tỷ lệ ảnh trên các thực thể trong tập dữ liệu MKG-Y. Từ đó dẫn đến hiệu năng trên các độ đo của hai mô hình đề xuất trên tập dữ liệu MKG-W không được tốt bằng hiệu năng của trên các độ đo của mô hình gốc.

Tiếp theo, luận án so sánh hiệu năng "tốt nhất" (bộ dữ liệu DB15K) và hiệu năng "tồi nhất" (bộ dữ liệu MKG-W). Theo Bảng 3.4, số lượng thực thể của bộ DB15K (với số lượng là 12842) ít hơn số lượng thực thể của bộ MKG-W (với số lượng là 15000). Tuy nhiên, số lượng thực thể không có ảnh của bộ DB15K chỉ khoảng 800 thực thể. Trong khi đó, số lượng thực thể không có ảnh của bộ MKG-W lại khá cao, khoảng 6000 thực thể. Điều đó dẫn đến tỷ lệ không có ảnh/số lượng thực thể của bộ MKG-W cao hơn hẳn tỷ lệ không có ảnh/số lượng thực thể của bộ DB15K. Cụ thể theo bảng dưới đây

	DB15K	MKG-W
Số lượng thực thể	12842	15000
Số lượng thực thể không có ảnh	800	6000

Số lượng thực thể không có ảnh/tổng số lượng thực thể	6,23%	40%

Một yếu tố khác, tỷ lệ cân bằng giữa các thực thể là người và các thực thể không phải con người trong bộ dữ liệu DB-15K tốt hơn trong bộ dữ liệu MKG-W. Tuy nhiên, khi nhìn vào các loại liên kết có trong các bộ dữ liệu, luận án nhận thấy những mối quan hệ này chủ yếu liên quan đến thực thể có tính chất con người.

3.4 Kết luận Chương 3

Mục này tổng kết các kết quả nghiên cứu ở chương 3

Chương 4. Hoàn thiện ĐTTT thời gian dựa trên nhúng lịch đại

4.1 Nhúng lịch đại ĐTTT

Trong phần này, luận án trình bày về các khái niệm của nhúng lịch đại trong hoàn thiện ĐTTT thời gian. Gọi $\mathcal{E}$ là tập hữu hạn các thực thể, $\mathcal{R}$ là tập hữu hạn các kiểu quan hệ, và $\mathcal{T}$ là tập hữu hạn dấu thời gian. Đặt $\mathcal{W} \subset \mathcal{E} \times \mathcal{R} \times \mathcal{E} \times \mathcal{T}$ là tập tất cả các dữ kiện trong thế giới thực, mô tả dưới dạng các bộ bốn $s = (h, r, t, \tau)$, ở đó h, t là các thực thể đầu và thực thể cuối, r là mối quan hệ giữa các thực thể, τ là thời gian. Đặt $\mathcal{W}^c$ là phần bù của $\mathcal{W}$. Khi đó, $s \in \mathcal{W}$ thì $s \notin \mathcal{W}^c$. Một ĐTTT thời gian $\mathcal{G}$ là một tập con của $\mathcal{W}$. Bài toán hoàn thiện ĐTTT thời gian (TKGC) là giải quyết các vấn đề thiếu liên kết trong đồ thị thời gian, nghĩa là sử dụng $\mathcal{G}$ để suy ra bộ hoàn chỉnh $\mathcal{W}$.

4.1.1 Nhúng từ lịch đại

Nhúng từ lịch đại xem xét chiều và mô hình thời gian của từ để từ đó xem nó phát triển như thế nào trong các khoảng thời gian khác nhau. Hamilton và cộng sự đã xây dựng nhúng từ lịch đại bằng cách sau: đầu tiên xây dựng các nhúng trong mỗi khoảng thời gian, tiếp theo sắp xếp chúng theo chiều thời gian, cuối cùng độ đo được sử dụng để định lượng sự thay đổi về mặt ngữ nghĩa.

4.1.2 Nhúng lịch đại ĐTTT

Định nghĩa 4.1 Nhúng lịch đại [19]:Một nhúng lịch đại DE, ký hiệu là $DEEMB : (\mathcal{E}, \mathcal{T}) \to \psi$*, là ánh xạ gán mỗi bộ đôi* (e, τ)*, trong đó* $e \in \mathcal{E}$ *và* $\tau \in \mathcal{T}$*, thành một biểu diễn ẩn trong* ψ *với* ψ *là một lớp không rỗng, chứ bộ các véc-tơ hoặc/và ma trận.*

Trong [19], R. Goel và cộng sự đã đề xuất một ánh xạ $DEEMB$ phù hợp đối với tập dữ liệu tri thức thời gian. Gọi $e \in \mathcal{E}$ là một thực thể (có thể là thực thể đầu hoặc thự thể cuối), $\boldsymbol{z}_e^\tau \in \mathbb{R}^{\boldsymbol{d}}$ là một véc-tơ thu được từ ánh xạ $DEEMB(e, \tau)$, nghĩa là $DEEMB(e, \tau) = (\dots, \boldsymbol{z}_e^\tau, \dots)$. Khi đó, $\boldsymbol{z}_e^\tau$ được xác định như công thức dưới đây.

$$\boldsymbol{z}_e^\tau[n] = \begin{cases} \boldsymbol{a}_e[n]\sigma(\boldsymbol{w}_e[n]\tau + \boldsymbol{b}_e[n]), & nếu\ 1 \le n \le \gamma d \\ \boldsymbol{a}_e[n], & nếu\ \gamma d < n \le d \end{cases} \qquad (4\text{-}1)$$

Công thức (4-1) được chia làm hai thành phần riêng biệt: thành phần 1 là nhúng tĩnh (không phụ thuộc thời gian) và thành phần 2 là nhúng động. Goel và cộng sự đã đề xuất sử dụng hàm lượng giác sin là hàm kích hoạt cho Công thức (4-1).

4.2 Mô hình hoàn thiện ĐTTT thời gian nhúng lịch đại DE-KGC

Luận án gọi mô hình hoàn thiện ĐTTT thời gian này là DE-KGC, một sự kết hợp giữa nhúng lịch đại DE và một mô hình KBC hoàn thiện ĐTTT tĩnh bất kỳ. R. Goel và cộng sự tại [19] chỉ ra rằng mô hình DE-SimplE có kết quả tốt nhất về mặt lý thuyết cũng như thực nghiệm

4.2.1 Quá trình học

Các dữ kiện trong ĐTTT $\mathcal{G}$ được phân chia vào bộ huấn luyện (training set), bộ xác thực (validation set) và bộ kiểm thử (test set). Các tham số của mô hình được học bằng cách sử dụng kỹ thuật giảm độ dốc ngẫu nhiên SGD (Stochastic Gradient Descent – SGD) với các lô nhỏ (mini-batches). Gọi B là một tập con của tập huấn luyện (nghĩa là B có thể được hiểu là một lô nhỏ). Đối với một sự kiện $f = (h, r, t, \tau) \in B$, Goel và cộng sự tại [24] đã tạo ra hai truy vấn: 1 - $(h, r, ?, \tau)$ và 2 - $(?, r, t, \tau)$. Đối với truy vấn đầu tiên, một tập câu trả lời tiềm năng được tạo ra là $C_{f,h}$ chứa h và n (được gọi là tỷ lệ âm) thực thể khác được lựa chọn ngẫu nhiên từ $\mathcal{E}$. Đối với truy vấn số hai, tương tự vậy một tập câu trả lời tiềm năng cũng được tạo ra là $C_{f,t}$. Sau đó, các tác giả tối thiểu hàm mất mát cross entropy bằng cách sử dụng kết quả từ [39], [80] để tính điểm $\mathcal{L}$ theo công thức dưới đây

$$\mathcal{L} = -\left(\sum_{f=(h,r,t,\tau)\in\mathcal{B}} \frac{\exp\big(\phi(h,r,t,\tau)\big)}{\sum_{t'\in C_{f,t}} \exp\left(\phi(h,r,t',\tau)\right)} + \frac{\exp\big(\phi(h,r,t,\tau)\big)}{\sum_{h'\in C_{f,h}} \exp\left(\phi(h',r,t,\tau)\right)} \right) \quad (4\text{-}2)$$

4.2.2 Ý tưởng cải tiến

Trong quá trình tìm hiểu và nghiên cứu về phương pháp nhúng lịch đại DE, luận án nhận thấy hiệu năng của mô hình hoàn thiện ĐTTT thời gian với sự kết hợp sẽ bị ảnh hưởng khi:

1. Thay thế mô hình hoàn thiện ĐTTT tĩnh KGC phù hợp với tập dữ liệu;
2. Thay thế hàm kích hoạt vì hàm kích hoạt là thành phần cốt lõi trong mạng học sâu nên việc lựa chọn hàm kích hoạt phù hợp giúp mạng học được quan hệ phi tuyến tốt hơn, dự đoán chính xác hơn

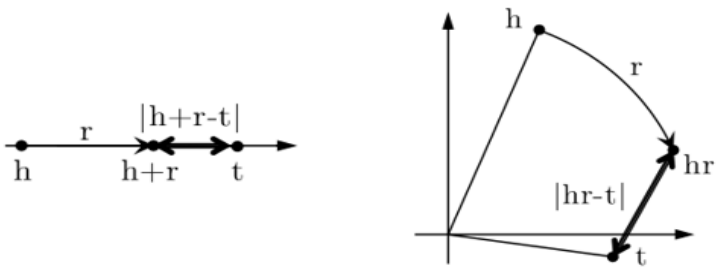
Luận án đề xuất pipeline DE-RotatE + hàm sinc có thể áp dụng cho các ĐTTT phức tạp, nhiều loại quan hệ và có tính thời gian. Bên cạnh đó, cách phân tích của luận án cung cấp nguyên tắc chọn mô hình cơ sở (base KGC model) khi kết hợp với DE đó là không phải mọi mô hình đều phù hợp, mà cần xem xét sự tương thích giữa cơ chế nhúng (vector, hộp, phép xoay) và giả định thời gian (sin, sigmoid, sinc). Đề xuất có giá trị tham khảo cho các nghiên cứu tiếp theo về hoàn thiện ĐTTT thời gian.

4.3 Đề xuất hai mô hình DE-RotatE và DE-RotatE-sinc

Hai mô hình thực nghiệm đề xuất được gọi là mô hình hoàn thiện ĐTTT thời gian DE-RotatE (thay SimplE bằng RotatE [38]) và mô hình hoàn thiện ĐTTT thời gian DE-RotatE-sinc (thay SimplE bằng RotatE [38] và cải tiến thành phần trong Công thức (4-1)).

4.3.1 Mô hình đề xuất DE-RotatE

Trong mô hình hoàn thiện ĐTTT RotatE, các mối quan hệ được định nghĩa như các phép quay trong không gian véc tơ phức. Hình 4.4 mô tả sự khác biệt giữa mô hình hoàn thiện ĐTTT TransE và mô hình hoàn thiện ĐTTT RotatE.



Hình 4.4. Mô hình hoàn thiện ĐTTT TransE [36] (trái) và mô hình hoàn thiện ĐTTT RotatE [38] (phải).

Xét một bộ ba (h, r, t), dựa trên đồng nhất thức Euler, Sun và các công sự đề xuất ánh xạ các thực thể đầu h và thực thể cuối t vào không gian nhúng phức, nghĩa là tạo thành các véc tơ nhúng $\mathbf{v}_h, \mathbf{v}_t \in \mathbb{C}^k$, ở đó $\mathbb{C}^k$ là không gian nhúng phức k chiều. Sau đó, xây dựng một ánh xạ tạo ra bởi mỗi quan hệ r như một phép quay element-wise từ thực thể đầu h đến thực thể cuối t. Cụ thể như công thức dưới đây.

$$\boldsymbol{v}_t = \boldsymbol{v}_h \circ \boldsymbol{v}_r \tag{4-3}$$

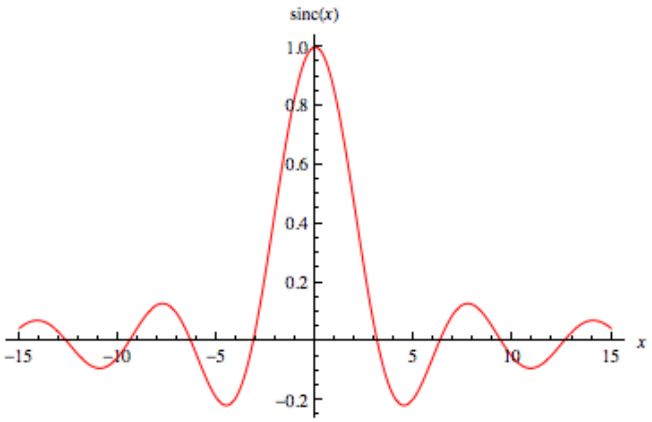
trong đó $|\mathbf{v}_{r_i}| = 1$ và $\circ$ là phép nhân Hadmard (tích element-wise).

Ngoài ra, qua thời gian thì mô hình hoàn thiện ĐTTT RotatE đã được chứng minh là mô hình có tốc độ tính toán nhanh vì sử dụng phép xoay chiều để biểu diễn quan hệ.

4.3.2 Mô hình đề xuất DE-RotatE-sinc

R. Goel và cộng sự [19] lựa chọn hàm $\sin(x)$ làm hàm kích hoạt vì tính chất tuần hoàn vô hạn của nó. Với tính chất tuần hoàn sẽ giúp mô hình nắm được tần suất của các dữ kiện có thực thể theo thời gian. Khi phân tích về ý nghĩa của các thực thể liên quan đến dữ kiện, trong thực nghiệm tiếp theo, luận án đề xuất sử dụng hàm kích hoạt $\mathrm{sinc}(x)$ thay vì sử dụng hàm kích hoạt $\sin x$ như trong [19]. Hàm $\mathrm{sinc}(x)$ và đồ thị của nó được mô ta theo công thức và hình vẽ tương ứng dưới đây.

$$sinc(x) = \begin{cases} \dfrac{\sin x}{x}, x \neq 0 \\ 1, x = 0 \end{cases} \tag{4-5}$$



Hình 0.1. Đồ thị mô tả hàm $sinc(x)$

Khi đó, công thức (4-1) trở thành

$$f(\tau) = \boldsymbol{a}_e[n]\,\mathrm{sinc}(\boldsymbol{w}_e[n]\tau + \boldsymbol{b}_e[n])$$

Đi sâu hơn thì luận án nhận thấy do tính đối xứng của hàm số $\mathrm{sinc}(x)$ qua trục tung, nên tồn tại $\tau_1, \tau_2 \in \mathcal{T}$ và $\tau_1 \neq \tau_2$ sao cho $f(\tau_1) = f(\tau_2)$, nghĩa là tồn tại hai khoảng thời gian khác nhau cho cùng một giá trị. Do vậy, công thức cho nhúng lịch đại trong mô hình DE-RotatE-sinc (cải tiến tiếp theo từ DE-RotatE) sẽ được mô tả dưới đây

$$\mathbf{v}_e^{\tau}[n] = \begin{cases} \boldsymbol{a}_e[n]\,\mathrm{sinc}(\boldsymbol{w}_e[n]\tau + \boldsymbol{b}_e[n]) + c_\tau, & nếu\ 1 \le n \le \gamma d \\ \boldsymbol{a}_e[n]\,, & nếu\ \gamma d < n \le d \end{cases} \qquad (4\text{-}6)$$

Ở đó, c_τ là véc-tơ nhúng riêng cho từng khoảng thời gian

4.4 Thực nghiệm và đánh giá hiệu năng

4.4.1 Phần mềm và phần cứng

Luận án cũng sử dụng ngôn ngữ Python và nền tảng Pytorch để thực nghiệm mô hình gốc DE-SimplE và hai mô hình đề xuất DE-RotatE và DE-RotatE-sinc.

4.4.2 Tập dữ liệu thực nghiệm mô hình

Giống như các thực nghiệm trong [19], luận án cũng thực nghiệm với ba tập dữ liệu ICEWS14, ICEWS05-15 bởi García-Durán và cộng sự [40] và GDELT bởi Leetaru [85].

Bảng 4.5. Thống kê về các tập dữ liệu ICEWS14, ICEWS05-15 và GDELT.

Datasets	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{T}\rvert$	$\lvert train\rvert$	$\lvert validation\rvert$	$\lvert test\rvert$	$\lvert\mathcal{G}\rvert$
ICEWS14	7.128	230	365	72.826	8.941	8.963	90.730
ICEWS05-15	10.488	251	4017	386.962	46.275	46.092	479.329
GDELT	500	20	366	2.735.685	341.961	341.961	3.419.607

4.4.3 Thông số cài đặt

Trong luận án này thực nghiệm hai mô hình là: mô hình hoàn thiện ĐTTT thời gian DE-RotatE và mô hình hoàn thiện ĐTTT thời gian DE-RotatE-sinc riêng biệt cùng với thông số cài đặt:

- Dữ liệu thời gian (ngày, tháng, năm) được chuẩn hóa về miền $(-1; 1)$;
- Tỷ lệ nhúng động là $0{,}32$ trên tổng số 100 chiều;
- Một số thông khác được lựa chọn như trong [19], cụ thể: Tốc độ học là $0{,}001$; dropout là $0{,}2$; batch size là 512; Số epochs được các tác giả lựa chọn là 500 cho cả ba tập dữ liệu. Do vấn đề hạn chế về phần cứng, nên luận án thay đổi số lượng epochs phụ thuộc vào từng tập dữ liệu: ICEWS14: 500 epochs; ICEWS05-15: 100 epochs; GDELT: 80 – 90 epochs.

4.4.4 Kết quả thực nghiệm

Bảng4.6. Kết quả thực nghiệm trên ĐTTT thời gian ICEWS14.

	ICEWS14				
Model	$MR\downarrow$	$MRR\uparrow$	$Hit@1\uparrow$	$Hit@3\uparrow$	$Hit@10\uparrow$
DE-SimplE [4]	-	0,526	41,8	59,2	72,5
DE-SimplE-retrain	225	0,523	41,3	59,0	72,9
DE-RotatE	*184*	*0,538*	*42,7*	*60,5*	*74,3*
DE-RotatE-sinc	**181**	**0,555**	**44,8**	**62,5**	**75,5**

Kết quả thực nghiệm trên hai ĐTTT thời gian ICEWS05-15 và GDELT được cho bởi bảng dưới đây.

Bảng 4.8. Kết quả thực nghiệm trên các ĐTTT thời gian ICEWS05-15 và GDELT

	ICEWS05-15				
Model	$MR\downarrow$	$MRR\uparrow$	$Hit@1\uparrow$	$Hit@3\uparrow$	$Hit@10\uparrow$
DE-SimplE [4]	-	*0,513*	*39,2*	*57,8*	*74,8*
DE-SimplE (retrain) với 100 epochs	104	0,499	37,9	56,4	73,6
DE-RotatE	104	0,495	37,4	55,8	73,0
DE-RotatE-sinc	**92**	**0,528**	**40,5**	**59,9**	**76,6**
	GDELT				
DE-SimplE [4]	-	**0,230**	14,1	24,8	**40.3**
DE-RotatE-sinc (80 epochs)	*66,93*	0,221	*14,12*	*23,75*	37,58
DE-RotaE-sinc (90 epochs)	**66,90**	*0,222*	**14,18**	**23,80**	*37,66*

Trong tất cả trường hợp thì kết quả của mô hình hoàn thiện ĐTTT thời gian DE-RotatE-sinc đều tốt hơn kết quả của mô hìn hoàn thiện ĐTTT thời gian DE-RotatE. Điều đó, khẳng định rằng các thành phần trong Công thức (4-1) của nhúng lịch đại DE cũng ảnh hưởng nhất định đến kết quả của mô hình hoàn thiện ĐTTT thời gian DE-KBC. Kết quả thực nghiệm càng khẳng định rằng cần phải lựa chọn hàm kích hoạt phù hợp với mô hình hoàn thiện ĐTTT tĩnh và từng tập dữ liệu ĐTTT cụ thể.

Đối với ĐTTT thời gian ICEWS05-15 (thời gian huấn luyện khoảng 172 giây/epoch). Luận án thực nghiệm với số lượng epochs nhỏ hơn so với mô hình gốc (bằng khoảng 20% so với mô hình gốc) thì kết quả của mô hình hoàn thiện ĐTTT thời gian DE-RotatE-sinc tốt hơn kết quả của mô hình hoàn thiện ĐTTT thời gian DE-SimplE trong tất cả độ đó. Nếu số lượng epochs tăng lên nữa thì luận án cho rằng kết quả sẽ còn cải thiện đáng kể.

Đối với ĐTTT thời gian GDELT, đây là cơ sơ tri thức khá lớn, thời gian huấn luyện diễn ra khá lâu (cụ thể huấn luyện khoảng 700 giây/epoch). Do vậy, luận án đưa ra kịch bản thực nghiệm và so sánh giữa mô hình hoàn thiện ĐTTT thời gian DE-RotatE-SimplE với số lượng epochs khác nhau (cụ thể là 80 và 90). Kết quả thực nghiệm chỉ ra rằng khi số lượng epochs được huấn luyện tăng lên thì kết quả của năm độ đo đều có xu hướng tăng lên.

4.5 Kết luận Chương 4

Mục này tổng kết các kết quả nghiên cứu ở chương 3

Kết luận

Kết quả chính của luận án

Luận án đã hoàn thiện mục tiêu nghiên cứu với ba đóng góp chính sau đây: (i) Đầu tiên, luận án đề xuất mô hình BERT/FastText-GRU-KBC được cải tiến từ mô hình GloVe-GRU-KBC [17] theo hai hướng: lựa chọn các mô hình ngôn ngữ BERT/FastText có ưu thế hơn GloVe trong việc chuyển giao từ dữ liệu/mô hình nguồn; sử dụng lớp kết nối đầy đủ trong nhúng từ giúp mô hình hoàn thiện ĐTTT đạt được sự cân bằng giữa khả năng học đặc trưng mạnh mẽ và tính tổng quát hóa trên toàn đồ thị tri thức. Sau đó, luận án triển khai các thực nghiệm để đánh giá hiệu năng của mô hình đề xuất và đưa ra đối sánh hiệu năng với mô hình gốc; (ii)Tiếp đó, luận án luận án đề xuất hai mô hình hoàn thiện ĐTTT đa phương thức mới dựa vào mô hình gốc AdaMF-MAT [18], trong đó, mô hình ViT-AdaMF-MAT (thay thế nhúng ảnh của mô hình gốc bằng mô hình nhúng ảnh Vision Transformer ViT) và mô hình T5-ViT-AdaMF-MAT (thay thế nhúng ảnh bằng ViT và nhúng văn bản bằng mô hình T5). Phát triển phần mềm và tiến hành các kịch bản thực nghiệm đánh giá hiệu năng mô hình đề xuất ViT-AdaMF-MAT và T5-ViT-AdaMF-MAT. Qua kết quả thực nghiệm, luận án chỉ ra

tầm ảnh hưởng của việc sử dụng các mô hình nhúng trong việc hoàn thiện ĐTTT đa phương thức; (iii) luận án đề xuất hai mô hình mới là DE-RotatE và DE-RotatE-sinc dựa trên cải tiên mô hình DE-SimplE [19], trong đó, mô hình DE-RotatE sử dụng mô hình thành phần RotatE thay cho mô hình thành phần SimplE, DE-RotatE-sinc không chỉ cải tiến như DE-RotatE mà còn cải tiến thành phần nhúng lịch đại DE thành nhúng động nhờ việc sử dụng hàm kích hoạt $\mathrm{sinc}(x)$ thay vì sử dụng hàm kích hoạt $\sin x$ như [19]. Phát triển phần mềm và tiến hành các kịch bản thực nghiệm đánh giá hiệu năng mô hình đề xuất DE-RotatE và DE-RotatE-sinc. Kết quả thực nghiệm cho thấy hiệu năng của các mô hình đề xuất đã được cải thiện.

Hạn chế của luận án

Các đề xuất chắt lọc tri thức của luận án còn chưa đạt được hiệu năng với mức độ cao.

• Luận án mới chỉ cải tiến được thành phần mô hình học chuyển giao từ mô hình hoàn thiện ĐTTT gốc dựa trên học chuyển giao. Ngoài ra, luận án mới chỉ cải tiến phần mã hóa của mô hình hoàn thiện ĐTTT gốc bằng cách sử dụng lớp kết nối đầy đủ FC vào vị trí phù hợp để tăng hiệu năng hoạt động. Cần tiến hành thêm các nghiên cứu sâu hơn về các mô hình khai thác các thông tin bổ sung cho hoàn thiện ĐTTT.

• Luận án mới cải tiến thành phần mô hình nhúng ảnh và thành phần mô hình nhúng văn bản từ mô hình hoàn thiện ĐTTT đa phương thức gốc AdaMF-MAT. Chủ đề hoàn thiện hoàn thiện ĐTTT đa phương thức mới trong thời kỳ ban đầu, tạo miền rộng mở cho các nghiên cứu tiếp theo

• Luận án mới cải tiến được thành phần mô hình hoàn thiện ĐTTT tĩnh cùng thành phần hàm kích hoạt của thành phần nhúng lịch đại DE trong mô hình hoàn thiện ĐTTT theo thời gian dựa trên nhúng lịch đại DE. Nghiên cứu phát triển nhiều hơn các kỹ thuật học máy theo hướng chú đề này cũng rất hấp dẫn.

• Một số kết quả thực nghiệm trong luận án còn chưa đạt được hiệu năng và mức độ cao như kỳ vọng. Trong đó, việc phân tích lỗi thử nghiệm chưa được tiến hành kỹ.

Định hướng nghiên cứu tiếp theo

- Tiếp tục tìm hiểu nghiên cứu kỹ thuật học chuyển giao đã đề xuất trong luận án trên các tập dữ liệu có tính chuyên biệt.
- Phân tích khung hoàn thiện ĐTTT đa phương thức CMR (Contrastive learning, Memorization, and Retrieval) gồm ba bước là Học đối chiếu tránh mâu thuẫn giữa các phương thức và đưa các hàng xóm ngữ nghĩa hữu ích đến gần nhau; Ghi nhớ lưu trữ rõ các hàng xóm ngữ nghĩa; Truy hồi tập các hàng xóm ngữ nghĩa trong giai đoạn kiểm thử để tăng cường suy luận [86]. Kết nối CRM với các tiếp cận làm giàu thông tin trong [87], [88], [89] để đề xuất ý tưởng cải tiến để xây dựng một mô hình hoàn thiện ĐTTT đa phương thức mới.
- Kết nối các kết quả nghiên cứu của luận án trong các mô hình hoàn thiện ĐTTT thời gian DE-RotatE và DE-RotatE-sinc với giải pháp tận dụng nói tích chập hai-ba chiều trong nhúng lịch đại trong [90] để nâng cấp, cải tiến các mô hình DE-RotatE và DE-RotatE-sinc.
- Khảo sát phân tích lỗi để làm cơ sở đối với những nghiên cứu tiếp theo.
- Tích hợp thành phần hoàn thiện ĐTTT vào các hệ thống chatbots.

Nghiên cứu sinh lập kế hoạch có kết quả nghiên cứu được chấp nhận công bố khoa học quốc tế vào Quý 4 năm 2025.

Danh mục công trình khoa học
của tác giả liên quan tới luận án

1. [AnhNTT1] Thuy-Anh Nguyen Thi, Thi-Hong Vuong, Thi-Hanh Le, Xuan-Hieu Phan, Thi-Thao Le, Quang-Thuy Ha. *Knowledge Base Completion with transfer learning using BERT and fastText*. KSE 2022: 1-6. **Scopus**, 01 Scopus Citation.
2. [AnhNTT2] Thuy-Anh Nguyen Thi, Viet-Phuong Ta, Xuan Hieu Phan, Quang-Thuy Ha. *An Improvement of Diachronic Embedding for Temporal Knowledge Graph Completion*. ACIIDS (2) 2023: 111-120. **Scopus**, **DBLP**, 01 Scopus Citation..
3. [AnhNTT3] Thuy-Anh Nguyen Thi, Hieu Cong Nguyen, Xuan-Hieu Phan, Quang-Thuy Ha. *Time Factor in Diachronic Embedding for Temporal Knowledge Graph Completion*. Book of Abstracts, International Joint Conference on Rough Sets (IJCRS 2023), pp. 26-28..
4. [AnhNTT4] Thuy-Anh Nguyen Thi, Cong-Hoang Le, Viet-Phuong Ta, Xuan-Hieu Phan, Quang-Thuy Ha. *The Impact of Embeddings on the Performance of Multi-modal Knowledge Graph Completion*. KSE 2024. **Scopus**.