

TRÍCH YẾU LUẬN ÁN

I. Tóm tắt mở đầu

- Tên tác giả: Nguyễn Thị Cẩm Vân
- Tên luận án: **Các mô hình học sâu kết hợp đa phương thức tiên tiến nhận diện cảm xúc trong hội thoại với thông tin không đầy đủ và mất cân bằng.**
- Ngành khoa học của luận án: Hệ thống thông tin Mã số: 9480104
- Tên đơn vị đào tạo: Trường Đại học Công nghệ, Đại học Quốc gia Hà Nội

II. Nội dung bản trích yếu

1. Mục tiêu và đối tượng nghiên cứu của luận án

- *Thứ nhất*, phát triển các mô hình học sâu tiên tiến cho bài toán nhận diện cảm xúc trong hội thoại đa phương thức (MERC), tập trung vào nâng cao hiệu quả kết hợp thông tin và mô hình hóa cấu trúc hội thoại đa người nói; các câu hỏi nghiên cứu xoay quanh việc thiết kế chiến lược kết hợp đa phương thức tận dụng tốt cả đặc trưng nội phương thức và liên phương thức ở mức phát ngôn, cũng như mô hình hóa quan hệ theo thời gian, theo người nói và cấu trúc động của hội thoại để cải thiện chất lượng nhận diện cảm xúc.
- *Thứ hai*, xây dựng các phương pháp học đa phương thức bền vững trong điều kiện dữ liệu phương thức không đầy đủ hoặc mất cân bằng; các câu hỏi nghiên cứu liên quan đến cách tái tạo đặc trưng bị thiếu và lan truyền thông tin bổ sung giữa các phương thức để duy trì biểu diễn đa phương thức ổn định, đồng thời thiết kế chiến lược học nhằm khắc phục mất cân bằng ở mức đặc trưng, mức phương thức và trong động lực học huấn luyện, tránh hiện tượng một phương thức chi phối.

2. Các phương pháp nghiên cứu đã sử dụng

Luận án sử dụng cách tiếp cận kết hợp giữa phân tích định tính và thực nghiệm định lượng. Về lý thuyết, luận án phân tích các mô hình nền tảng trong học đa phương thức, học trên đồ thị, học biểu diễn, học tự giám sát và curriculum learning, làm cơ sở đề xuất các mô hình mới. Về thực nghiệm, các mô hình này được triển khai trên những bộ dữ liệu chuẩn cho nhận diện và phân tích cảm xúc đa phương thức, với các kịch bản đánh giá được thiết kế để kiểm tra hiệu năng, độ bền vững và đóng góp của từng thành phần trong mô hình.

3. Các kết quả chính và kết luận

Tham gia vào dòng nghiên cứu trên thế giới về học máy đa phương thức, luận án đã hoàn thiện mục tiêu nghiên cứu với ba đóng góp chính sau đây:

- *Thứ nhất*, luận án đề xuất các mô hình kết hợp đa phương thức có cấu trúc cho bài toán nhận diện cảm xúc trong hội thoại. Cụ thể, luận án trình bày CORECT, một mô hình dựa trên đồ thị sử dụng mạng nơ-ron tích chập trên đồ thị quan hệ - thời gian RT-GCN để học phụ thuộc theo thời gian và theo người nói ở mức phát ngôn, cùng mô-đun tương tác liên phương thức từng cặp P-CM để mô hình hóa quan hệ đa phương thức ở mức hội thoại; đồng thời phát triển MultiDAG+CL, một mô hình kết hợp cấu trúc hội thoại dạng đồ thị có hướng không chu trình DAG với chiến lược học theo lộ trình nhằm xử lý các chuyển dịch cảm xúc và cấu trúc hội thoại phức tạp.
- *Thứ hai*, luận án đề xuất khung học đa phương thức bền vững trong trường hợp dữ liệu phương thức không đầy đủ. Luận án xây dựng Mi-CGA, một mô hình dựa trên đồ thị gồm giai đoạn biểu diễn đa phương thức không đầy đủ IMR và mạng chú ý đồ thị liên phương thức CGA-Net, cho phép vừa tái tạo các đặc trưng bị thiếu vừa mô hình hóa quan hệ nội và liên phương thức, giúp duy trì hiệu năng nhận diện cảm xúc khi một hoặc nhiều phương thức bị thiếu một phần hoặc hoàn toàn.
- *Cuối cùng*, luận án đề xuất các phương pháp đối phó với tình trạng mất cân bằng phương thức và độ khó học không đồng đều. Để xử lý chênh lệch mức đóng góp giữa các phương thức, luận án phát triển Ada2I, một kỹ thuật cân bằng động dựa trên hai cơ chế Adaptive Feature Weighting (AFW) và Adaptive Modality Weighting (AMW) nhằm điều chỉnh trọng số ở cả mức đặc trưng và mức

phương thức; đồng thời giới thiệu SPCL, một mô-đun nhẹ dạng độc lập mô hình cho phép mô hình học từ dễ đến khó dựa trên độ khó ở cấp độ phát ngôn và hội thoại, góp phần ổn định quá trình học và giảm hiện tượng một phương thức chi phối trong các tình huống cảm xúc phức tạp.

Tất cả các mô hình đề xuất đều được đánh giá qua các thực nghiệm chuẩn hóa và so sánh với các mô hình gốc hoặc mô hình tiên tiến gần nhất; kết quả cho thấy các phương pháp trong luận án đóng góp đáng kể vào việc nâng cao hiệu năng, độ tin cậy và tính bền vững của các mô hình học sâu đa phương thức cho nhận diện cảm xúc trong hội thoại, đặc biệt dưới các điều kiện dữ liệu thực tế không đầy đủ và mất cân bằng phương thức.

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

1. **Cam Van Thi Nguyen**, Tuan Mai, Son The, Dang Kieu, and Duc-Trong Le. 2023. “Conversation Understanding using Relational Temporal Graph Neural Networks with Auxiliary Cross-Modality Interaction”. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 15154–15167, Singapore. Association for Computational Linguistics <https://doi.org/10.18653/v1/2023.emnlp-main.937> (Scopus, CORE Rank A* Conference)
2. **Cam-Van Thi Nguyen**, Cao-Bach Nguyen, Duc-Trong Le, and Quang-Thuy Ha. 2024. “Curriculum Learning Meets Directed Acyclic Graph for Multimodal Emotion Recognition”. In *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, pages 4259–4265, Torino, Italia. ELRA and ICCL. <https://aclanthology.org/2024.lrec-main.380/> (Scopus, CORE Rank B Conference).
3. **Cam-Van Thi Nguyen**, The-Son Le, Anh-Tuan Mai, and Duc-Trong Le. 2024. “Ada2I: Enhancing Modality Balance for Multimodal Conversational Emotion Recognition”. In *Proceedings of the 32nd ACM International Conference on Multimedia (MM '24)*. Association for Computing Machinery, New York, NY, USA, 9330–9339. <https://doi.org/10.1145/3664647.3681648> (Scopus, CORE Rank A* Conference)
4. **Cam-Van Thi Nguyen**, Hai-Dang Kieu, Quang-Thuy Ha, Xuan-Hieu Phan, Duc-Trong Le. “Mi-CGA: Cross-modal Graph Attention Network for robust emotion recognition in the presence of incomplete modalities”, *Neurocomputing*, Volume 623, 2025, 129342, ISSN 0925-2312. <https://doi.org/10.1016/j.neucom.2025.129342> (SCIE Q1 Journal, Impact Factor: 6.5)
5. Phuong-Anh Nguyen, The-Son Le, Duc-Trong Le, **Cam-Van Thi Nguyen***. 2026. “Leveraging Self-Paced Curriculum Learning for Enhanced Modality Balance in Multimodal Conversational Emotion Recognition.” Chấp nhận đăng tại tạp chí *Neural Computing and Applications* <https://doi.org/10.1007/s00521-026-12160-6> (Scopus Q1 Journal, in press)

Hà Nội, ngày tháng năm 20...

NGHIÊN CỨU SINH

CÁN BỘ HƯỚNG DẪN

XÁC NHẬN CỦA CƠ SỞ ĐÀO TẠO