

THÔNG TIN VỀ LUẬN ÁN TIẾN SĨ

1. Họ và tên nghiên cứu sinh: Nguyễn Thị Cẩm Vân... 2. Giới tính: Nữ.....
3. Ngày sinh: 07/11/1995..... 4. Nơi sinh: Hà Nội.....
5. Quyết định công nhận NCS số 25- QĐ/CTSV ngày 14 tháng 01 năm 2022 của Hiệu trưởng trường Đại học Công nghệ, ĐHQG Hà Nội.
6. Các thay đổi trong quá trình đào tạo:
 - Bổ sung cán bộ hướng dẫn theo Quyết định số 1181/QĐ-ĐT ngày 29/11/2022 của Hiệu trưởng Trường Đại học Công nghệ. Cụ thể, bổ sung TS. Lê Đức Trọng là cán bộ hướng dẫn phụ
 - Gia hạn thời gian học tập (24 tháng) theo Quyết định số 2833/QĐ-ĐHCN ngày 03/12/2024 của Hiệu trưởng Trường Đại học Công nghệ. Cụ thể, thời gian gia hạn bắt đầu từ ngày 01/01/2025 đến ngày 01/01/2027.
 - Điều chỉnh tên đề tài luận án theo Quyết định số 391/QĐ-ĐHCN ngày 04/03/2026 của Hiệu trưởng Trường Đại học Công nghệ. Cụ thể, điều chỉnh tên đề tài luận án từ: “*Nghiên cứu các mô hình học máy tổng hợp dữ liệu đa phương thức ứng dụng trong các hệ thống thương mại điện tử*” thành “*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*”
7. Tên đề tài 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***.....
8. Ngành đào tạo: Hệ thống thông tin..... 9. Mã số: 9480104.....
10. Cán bộ hướng dẫn khoa học: PGS.TS. Hà Quang Thụy, TS. Lê Đức Trọng.....
11. Tóm tắt các **kết quả mới** của luận án:

Về kết quả chính và đóng góp mới:

- Luận án đề xuất các mô hình CORECT và MultiDAG+CL cho bài toán nhận diện và phân tích cảm xúc đa phương thức trong hội thoại; trong đó, CORECT sử dụng mạng nơ-ron tích chập trên đồ thị quan hệ - thời gian RT-GCN (Relational Temporal Graph Convolutional Network) và mô-đun tương tác liên phương thức từng cặp P-CM (Pairwise Cross-modal Feature Interaction) để cải thiện hợp nhất thông tin và mô hình hóa quan hệ liên phương thức, còn MultiDAG+CL mô hình hóa hội thoại dưới dạng đồ thị có hướng không chu trình DAG (Directed Acyclic Graph) kết hợp chiến lược học theo lộ trình (Curriculum Learning).

- Luận án đề xuất mô hình Mi-CGA, một mạng chú ý trên đồ thị liên phương thức gồm ba thành phần chính: biểu diễn đa phương thức không đầy đủ IMR (Incomplete Multimodal Representation) để mã hóa cấu trúc thiếu phương thức, mô-đun ước lượng đặc trưng FE (Feature Estimation) để tái tạo đặc trưng cho phương thức bị thiếu, và mạng chú ý đồ thị liên phương thức CGA-Net (Cross-modal Graph Attention Network) để lan truyền và hợp nhất thông tin bổ sung giữa các modality trên đồ thị hội thoại, qua đó duy trì hiệu năng mô hình trong nhiều kịch bản thiếu phương thức khác nhau..
- Luận án đề xuất hai phương pháp Ada2I và SPCL để xử lý mất cân bằng phương thức và độ khó học không đồng đều; trong đó, Ada2I điều chỉnh động trọng số đặc trưng AFW (Adaptive Feature Weighting) và trọng số phương thức AMW (Adaptive Modality Weighting), còn SPCL là mô-đun dạng độc lập mô hình (model-agnostic) cho phép mô hình học từ dễ đến khó dựa trên độ khó ở cấp độ phát ngôn và hội thoại.

Các kết quả thực nghiệm cho thấy các mô hình đề xuất đều đạt hoặc vượt các phương pháp tiên tiến hiện tại về độ chính xác, F1 và độ ổn định trên nhiều thiết lập đa phương thức, qua đó giải quyết hiệu quả các vấn đề khoa học và kỹ thuật về: (i) thiết kế chiến lược hợp nhất thông tin tận dụng tốt đặc trưng nội và liên phương thức; (ii) mô hình hóa cấu trúc động, phụ thuộc thời gian và quan hệ người nói trong hội thoại; (iii) học đa phương thức bền vững khi dữ liệu không đầy đủ và phân bố lệch.

Về ý nghĩa khoa học: Luận án xây dựng một khung tiếp cận thống nhất cho học máy đa phương thức trong nhận diện cảm xúc hội thoại, kết nối ba trục năng lực chính: mô hình hóa hội thoại đa phương thức có cấu trúc, học biểu diễn bền vững dưới điều kiện thiếu phương thức, và điều tiết động quá trình học để giảm chi phối của modality trội.

Về ý nghĩa thực tiễn: Các mô hình và kỹ thuật đề xuất có thể được áp dụng để nâng cao chất lượng các hệ thống tương tác người-máy, trợ lý ảo, phân tích cảm xúc và quan điểm người dùng trong các nền tảng mạng xã hội, hội nghị trực tuyến và dịch vụ khách hàng sử dụng dữ liệu đa phương thức, góp phần cải thiện trải nghiệm người dùng và hỗ trợ ra quyết định trong các ứng dụng phân tích hành vi và truyền thông số.

12. Khả năng ứng dụng trong thực tiễn:

Các mô hình và kỹ thuật của luận án có thể tích hợp vào các hệ thống hội thoại thông minh để nhận diện cảm xúc đa phương thức trong thời gian thực, hỗ trợ điều chỉnh chiến lược phản hồi, phân luồng chăm sóc khách hàng và cảnh báo sớm các trạng thái cảm xúc tiêu cực. Các kết quả cũng có thể được khai thác trong các nền tảng phân tích hành vi số nhằm theo dõi xu hướng cảm xúc, tối ưu nội dung và nâng cao mức độ gắn kết của người dùng.

13. Những hướng nghiên cứu tiếp theo:

Trong thời gian tới, có thể mở rộng nghiên cứu sang các ngôn ngữ và miền ứng dụng khác (ví dụ: tiếng Việt, hội thoại trong chăm sóc y tế hoặc giáo dục), cũng như tích hợp thêm các nguồn dữ liệu sinh lý hoặc hành vi để tăng độ tin cậy của mô hình. Một hướng khác là

ngiên cứu các kiến trúc nhẹ hơn và cơ chế học trực tuyến để triển khai hiệu quả trên các hệ thống thời gian thực và thiết bị tài nguyên hạn chế.

14. Các công trình đã công bố có liên quan đến 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.” *Neural Computing and Applications* <https://doi.org/10.1007/s00521-026-12160-6> (**Scopus Q1 Journal, in press**)

Ngày tháng năm 20
Xác nhận của cán bộ hướng dẫn
(Kí và ghi rõ họ tên)

Ngày tháng năm 20
Nghiên cứu sinh
(Kí và ghi rõ họ tên)

INFORMATION ON DOCTORAL THESIS

1. Full name : Nguyen Thi Cam Van..... 2. Sex: Female
3. Date of birth: 07/11/1995..... 4. Place of birth: Hanoi
5. Admission decision number: 25- QĐ/CTSV Dated 14/01/2022
6. Changes in academic process:

 - Under Decision No. 1181/QĐ-ĐT dated 29 November 2022 of the Rector of the University of Engineering and Technology on the appointment of an additional academic supervisor, Dr. Le Duc Trong was appointed as the co-supervisor.
 - Under Decision No. 2833/QĐ-ĐHCN dated 03 December 2024 of the Rector of the University of Engineering and Technology on the extension of the study period for 24 months, the extended period is from 01 January 2025 to 01 January 2027.
 - Under Decision No. 391/QĐ-ĐHCN dated 04 March 2026 of the Rector of the University of Engineering and Technology on the adjustment of the dissertation title, the dissertation title was changed from “*Research on multimodal data fusion machine learning models for e-commerce systems*” to “*Advanced Deep Multimodal Fusion Models for Emotion Recognition in Conversation Under Incomplete and Imbalanced Modalities*”.

7. Official thesis title: **Advanced Deep Multimodal Fusion Models for Emotion Recognition in Conversation Under Incomplete and Imbalanced Modalities**.....
8. Major: Information System..... 9. Code: 9480104
10. Supervisors: Assoc. Prof. Ha Quang Thuy, Dr. Le Duc Trong.....
11. Summary of the **new findings** of the thesis:

Regarding the **main results and novel contributions**:

- The dissertation proposes the CORECT and MultiDAG+CL models for multimodal emotion and sentiment recognition in conversation. CORECT employs a relational temporal graph convolutional network (RT-GCN) and a pairwise cross-modal feature interaction module (P-CM) to improve information fusion and the modeling of cross-modal relations, while MultiDAG+CL represents conversations as a directed acyclic graph (DAG) combined with a curriculum learning strategy.
- The dissertation proposes the Mi-CGA model, a cross-modal graph attention network with three main components: an incomplete multimodal representation (IMR) to encode configurations with missing modalities, a feature estimation (FE) module to reconstruct features for missing modalities, and a cross-modal graph attention network (CGA-Net) to propagate and integrate complementary information across modalities

on the conversational graph, thereby maintaining model performance under diverse missing-modality scenarios.

- The dissertation proposes two methods, Ada2I and SPCL, to address modality imbalance and uneven learning difficulty. Ada2I adaptively adjusts feature weights (AFW) and modality weights (AMW), whereas SPCL is a model-agnostic plug-in module that enables models to learn from easy to hard examples based on difficulty at both utterance and conversation levels.

Experimental results show that the proposed models achieve accuracy, F1 and stability that are competitive with or superior to state-of-the-art methods across multiple multimodal setups. These results effectively address key scientific and technical issues, including: (i) designing fusion strategies that fully exploit both intra-modal and cross-modal characteristics; (ii) modeling dynamic conversational structures, temporal dependencies and speaker relations; and (iii) developing robust multimodal learning under incomplete and distributionally skewed data conditions.

In terms of scientific significance, the dissertation establishes a unified framework for multimodal machine learning in conversational emotion recognition, integrating three main capability axes: structured modeling of multimodal conversations, robust representation learning under missing modalities, and dynamic regulation of the learning process to mitigate dominance of any single modality.

In terms of practical significance, the proposed models and techniques can be applied to improve the quality of human-computer interaction systems, virtual assistants, and user emotion and opinion analysis on social media platforms, online conferencing systems and customer service applications that rely on multimodal data, thereby enhancing user experience and supporting decision-making in behavioral analytics and digital communication scenarios.

12. Practical applicability, if any:

The models and techniques developed in this dissertation can be integrated into intelligent conversational systems to perform real-time multimodal emotion recognition, thereby supporting adaptive response strategies, customer service routing, and early alerts for negative emotional states. The results can also be exploited in digital behavior analytics platforms to track emotional trends, optimize content, and enhance user engagement.

13. Further research directions, if any:

Future work may extend the study to additional languages and application domains (for example, Vietnamese, conversational data in healthcare or education), as well as incorporate physiological or behavioral signals to further improve model reliability. Another promising direction is to design lighter architectures and online learning mechanisms for efficient deployment in real-time systems and on resource-constrained devices.

14. Thesis-related publications:

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.” *Neural Computing and Applications* <https://doi.org/10.1007/s00521-026-12160-6> (Scopus Q1 Journal, in press)

Supervisors

Date:

Signature:

Full name:

Signature:

Full name:

Doctoral Candidate

Date:

Signature:

Full name: