

THÔNG TIN LUẬN ÁN

Đề tài nghiên cứu:	Phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc
Chuyên ngành:	Khoa học máy tính
Mã số chuyên ngành:	62.48.01.01
Họ và tên NCS:	Lê Thị Thủy
Tập thể hướng dẫn:	GS. TS. Phan Thị Tươi
Cơ sở đào tạo:	Trường Đại học Bách Khoa – ĐHQG TP. HCM

1. TÓM TẮT

Phân giải đồng tham chiếu và phân tích cảm xúc là hai bài toán khá phổ biến trong xử lý ngôn ngữ tự nhiên, được rất nhiều nhà nghiên cứu quan tâm với các cách tiếp cận khác nhau. Đề bài toán phân tích cảm xúc có được kết quả đầy đủ, chi tiết mang lại lợi ích thực tiễn, luận án đề xuất giải quyết bài toán *phân giải đồng tham chiếu đối tượng* cho phân tích cảm xúc. Đây chính là mục tiêu của luận án. Kết quả của bài toán đề xuất là các bộ ba đối tượng, khía cạnh và cảm xúc của các sản phẩm hay dịch vụ trên một miền chuyên biệt. Giới hạn của luận án được áp dụng trên các văn bản cảm xúc tiếng Anh.

Để giải quyết bài toán phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc, luận án đề xuất xây dựng hệ thống với cách tiếp cận trên cơ sở tri thức - ontology cảm xúc, kết hợp các giải thuật xử lý ngôn ngữ tự nhiên và các phương pháp học máy, học sâu trên kho ngữ liệu văn bản có cảm xúc, có ngữ cảnh. Dựa trên đặc điểm của văn bản cảm xúc và mục tiêu đặt ra, luận án cần giải quyết các bài toán sau đây. **Thứ nhất**, phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc có một đối tượng. **Thứ hai**, làm giàu ontology cảm xúc hỗ trợ phân giải đồng tham chiếu đối tượng. **Thứ ba**, xác định khía cạnh ẩn hỗ trợ phân giải đồng tham chiếu đối tượng. **Thứ tư**, phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc có nhiều đối tượng. Phần thực nghiệm của luận án được thực hiện trên kho ngữ liệu các văn bản cảm xúc bình luận về các sản phẩm, dịch vụ trên trang web của *Amazon* và *YouNetMedia*. Các kết quả thực nghiệm phản ánh tính khả thi của các giải thuật, mô hình đề xuất trong luận án.

Luận án có 6 bài báo đã công bố, gồm 2 bài đăng trong danh mục tạp chí (1 bài trong nước, 1 bài quốc tế) và 4 bài đăng trong các kỷ yếu hội nghị khoa học quốc tế.

2. CÁC ĐÓNG GÓP CHÍNH CỦA LUẬN ÁN

- *Xây dựng ontology cảm xúc SO và đồ thị đồng tham chiếu CRG hỗ trợ phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc.* SO nhận dạng các thành phần của bộ ba đối tượng, khía cạnh và cảm xúc, và có thể xác định khía cạnh không tường minh cho cảm xúc; giải quyết đồng tham chiếu thực thể có tên; xác định mối quan hệ giữa các thành phần trong bộ ba. Công cụ thứ hai CRG là công cụ cuối cùng để sinh ra các bộ ba đối tượng, khía cạnh và cảm xúc.
- *Mô hình phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc cho văn bản có một đối tượng.* Từ đóng góp thứ nhất, luận án xây dựng được mô hình phân giải đồng tham chiếu đối tượng cho phân tích cảm xúc cho văn bản có một đối tượng.
- *Mô hình làm giàu ontology cảm xúc.* Làm giàu ontology cảm xúc của luận án tập trung vào việc bổ sung các thể hiện cho các lớp, cụ thể là các từ, cụm từ chỉ đối tượng, khía cạnh và cảm xúc cùng mối quan hệ giữa chúng..
- *Mô hình xác định khía cạnh ẩn trong văn bản cảm xúc.* Để xác định được khía cạnh ẩn trong văn bản cảm xúc có ngữ cảnh, luận án đề xuất phương pháp học sâu dựa trên mạng nơ-ron nhân tạo nhiều lớp và dựa trên ngữ cảnh của từ cảm xúc tham chiếu đến khía cạnh trong một kho ngữ liệu các văn bản.
- *Xây dựng mô hình phân giải đồng tham chiếu đối tượng cho văn bản cảm xúc có từ hai đối tượng trở lên.* Với văn bản có nhiều đối tượng, luận án sử dụng các văn bản có một đối tượng làm cơ sở xác định mối quan hệ giữa đối tượng, khía cạnh và cảm xúc. Cách tiếp cận của mô hình được áp dụng theo phương pháp mới hiện nay là học máy và học sâu dựa trên ngữ cảnh của văn bản và ontology cảm xúc.

3. NHỮNG VẤN ĐỀ CÒN BỎ NGÕ CẦN TIẾP TỤC NGHIÊN CỨU

Mặc dù đạt được kết quả khá tốt và được đánh giá bằng thực nghiệm, luận án vẫn còn một số vấn đề nhỏ cần quan tâm và phát triển trong tương lai.

- Phát triển và làm giàu ontology cảm xúc: xây dựng và làm giàu ontology bằng phương pháp bán tự động có nhiều hạn chế vì vậy cải tiến và sử dụng phương pháp tự động là nhu cầu cần thiết.

- Với bài toán phân giải đồng tham chiếu cho phân tích cảm xúc có nhiều đối tượng bị ảnh hưởng bởi bước tiền huấn luyện. Vì vậy nếu xuất hiện khía cạnh mới không tồn tại trong bộ từ vựng ban đầu thì cần phải thực hiện lại từ đầu, tiền huấn luyện, hoặc là không thay đổi bộ từ vựng và như vậy kết quả đạt được sẽ bị ảnh hưởng.
- Phát triển mô hình của luận án trên các văn bản cảm xúc tiếng Việt sẽ có ý nghĩa khoa học và thực tiễn cao trong lĩnh vực xử lý ngôn ngữ tự nhiên ở Việt Nam.

CÁN BỘ HƯỚNG DẪN

NGHIÊN CỨU SINH

GS.TS. PHAN THỊ TƯỞI

LÊ THỊ THỦY