

TRÍCH XUẤT THÔNG TIN SO SÁNH TRONG BÌNH LUẬN TIẾNG VIỆT DỰA TRÊN MÔ HÌNH TẠO SINH VĂN BẢN

Nguyễn Thị Thùy^{1,2}

¹21521514@gm.uit.edu.vn

²Trường Đại học Công nghệ Thông tin - ĐHQG TP.HCM

Tổng quan

1. Tóm tắt

- Đề tài tập trung vào nghiên cứu và áp dụng các mô hình tạo sinh văn bản, kết hợp với kỹ thuật Xử lý ngôn ngữ Tự nhiên, để giải quyết bài toán trích xuất nhóm ý kiến so sánh (COQE) trong tiếng Việt
- Nghiên cứu các phương pháp giải quyết bài toán COQE dựa trên các mô hình tạo sinh văn bản
- Đánh giá trên bộ dữ liệu tiếng Việt được thu thập từ cuộc thi "ComOM - Comparative Opinion Mining from Vietnamese Product Reviews".

2. Lý do chọn đề tài

- Sự bùng nổ của các nền tảng kinh doanh trực tuyến đã tạo nên sự đa dạng và phong phú của các sản phẩm và dịch vụ trong thị trường
- Trong cùng phân khúc sẽ xuất hiện nhiều sản phẩm khác nhau nên người dùng có xu hướng tìm kiếm sự so sánh giữa các sản phẩm.
- Bài toán này mong muốn giúp người tiêu dùng có thể so sánh, tổng hợp thông tin một cách thuận tiện, nhanh chóng khi lựa chọn được sản phẩm phù hợp

3. Xác định bài toán COQE

- Bài toán COQE nhận vào một câu nhận xét, nếu câu đó mang ý so sánh thì sẽ trả một tập hợp 5 phần tử gọi là "quintuple" bao gồm:
 - Subject (s): Chủ thể của sự so sánh.
 - Object (o): Khách thể được so sánh với chủ thể.
 - Comparative Aspect (ca): Khía cạnh so sánh
 - Comparative Opinion (co): Từ hoặc cụm từ so sánh thể hiện ý kiến so sánh.
 - Comparative Preference (cp): Mức độ so sánh.
- **Đầu vào:** Một câu nhận xét sản phẩm
- **Đầu ra:** Danh sách các quintuple trong câu nếu có.
 $S = \{..., (s, o, ca, co, cp), ...\}$

4. Kết quả dự kiến

- Nâng cao hiệu quả hiệu suất trên bộ dữ liệu ComOM dựa trên các mô hình tạo sinh văn bản.
- Báo cáo về các phương pháp và mô hình ngôn ngữ đã được sử dụng, cùng với các kết quả thực nghiệm, so sánh, đánh giá giữa các phương pháp.
- Viết 01 bài báo gửi đến một hội nghị trong nước.

Kiến trúc mô hình tạo sinh văn bản

