

PHÁT HIỆN ĐÁNH GIÁ GIẢ MẠO SỬ DỤNG MẠNG NGỮ NGHĨA TĂNG CƯỜNG KHÓA CẠNH

Nguyễn Hữu Đặng Nguyên - 23521045

Đặng Quốc Cường - 23520192

Trường Đại Công nghệ Thông tin,
Đại học Quốc gia TP. Hồ Chí Minh

Mục tiêu

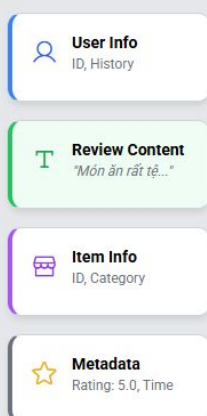
- Đề xuất mô hình phát hiện đánh giá giả dựa trên đồ thị ngữ nghĩa, tích hợp thông tin khóa cạnh (aspect) và cảm xúc.
- Khai thác đồng thời ngữ nghĩa văn bản, hành vi người dùng và mối quan hệ giữa các khóa cạnh trong cấu trúc đồ thị.
- Đánh giá hiệu quả mô hình trên các bộ dữ liệu chuẩn và phân tích đóng góp của thông tin khóa cạnh.

Bối cảnh

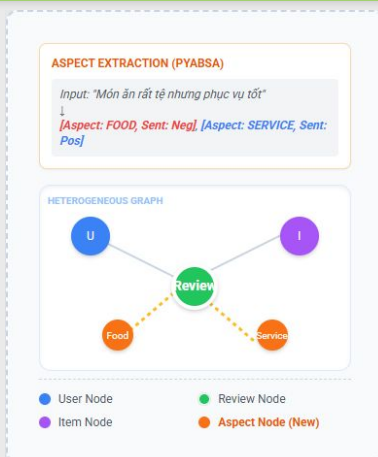
- Đánh giá trực tuyến là yếu tố định hướng quyết định người dùng và uy tín doanh nghiệp. **Đánh giá giả** càng phổ biến và tinh vi (sinh bằng AI) gây nhiều ảnh hưởng tiêu cực.
- Hiện nay, thông tin ở mức độ khóa cạnh và sự mâu thuẫn cảm xúc của đánh giá vẫn chưa được khai thác đầy đủ.
- Tích hợp **thông tin khóa cạnh** vào mô hình đồ thị được kỳ vọng giúp **cải thiện độ chính xác và tính ổn định** trong phát hiện đánh giá giả.

Tổng quan

Dữ liệu đầu vào



Xử lý & Xây dựng đồ thị



Kết quả



Giải thích: Hệ thống phát mâu thuẫn giữa cảm xúc khóa cạnh (**Aspect: Food - tệ và Rating - 5 sao**)

Mô tả phương pháp

1. Dữ liệu đầu vào

Nghiên cứu sử dụng 4 thành phần dữ liệu chính để nắm bắt cả ngữ nghĩa văn bản và hành vi đánh giá (**User, Review, Item, Metadata**) và nhãn đánh giá (**Giả / Thật**).

2. Xử lý và xây dựng đồ thị

2.1. Trích xuất khóa cạnh

Mục tiêu: Khai thác ngữ nghĩa ở mức khóa cạnh thay vì chỉ phân tích bề mặt văn bản.

Quy trình:

- Trích xuất Aspect (**VD: Food, Service**) qua Rule-based hoặc mô hình pretrain ABSA.
- Gán nhãn cảm xúc cụ thể (**Tích cực / Tiêu cực / Trung tính**) cho từng Aspect.
- Chuẩn hóa về nhóm khái niệm chung bằng các phương pháp clustering để giảm nhiễu.

2.2. Xây dựng đồ thị ngữ nghĩa di thể (Heterogeneous Graph)

- Mô hình hóa mối quan hệ đa chiều giữa 4 loại nút: **User - Review - Item - Aspect**.
- Các quan hệ:** User → Review, Review → Item, Review → Aspect, Item → Aspect...

2.3. Biểu diễn nút

- Review Node:** Embedding ngữ nghĩa từ BERT/Roberta.
- Aspect Node:** Embedding từ token (GloVe) hoặc tổng hợp từ review.
- User/Item Node:** Đặc trưng hành vi & Metadata.

3. Dự đoán & giải thích

Cơ chế hoạt động:

- Áp dụng Học sâu trên đồ thị (GNN) để lan truyền và tổng hợp thông tin.
- Embedding của review từ đồ thị rồi đưa qua lớp MLP để phân loại.

Kết quả dự kiến

Phương pháp đánh giá:

- Sử dụng các dataset chuẩn cho bài toán Fake Review Detection (**VD: YelpChi**).
- Đánh giá bằng: **Accuracy, Precision, Recall, F1-score**.
- So sánh với kết quả của các nghiên cứu liên quan và các baseline.
- Ablation study** để xác định đóng góp của các thành phần, đặc biệt là thành phần khóa cạnh trong câu.

Kết quả dự kiến: Bảng so sánh chi tiết hiệu năng mô hình dựa trên ablation study, cùng phân tích hướng mở rộng.

Được kỳ vọng có thể ứng dụng vào thực tiễn như phát hiện đánh giá giả theo thời gian thực.