

ANALYSIS OF GOOGLE RATING COMMENTS ABOUT UNIVERSITY MEDICAL CENTER HCMC ON MANY ASPECTS USING THE HSD MODEL

NGUYỄN GIANG THANH PHƯƠNG - 230201050

¹ Trường ĐH Công Nghệ Thông Tin

² University of Information Technology
HCMC, Vietnam

³ Đại học Quốc gia TPHCM

What ?

Chúng tôi áp dụng mô hình kết hợp PhoBERT và CNN cho bài phân loại các bình luận thu thập từ Google Rating theo nhiều khía cạnh, trong đó chúng tôi:

- Xây dựng được bộ dữ liệu được trích xuất từ Google Rating với 10000 bình luận.
- Triển khai, đánh giá hiệu quả mô hình kết hợp PhoBERT và CNN
- So sánh mô hình kết hợp PhoBERT và CNN với các mô hình khác về nhiệm vụ phân loại văn bản tiếng Việt.

Why ?

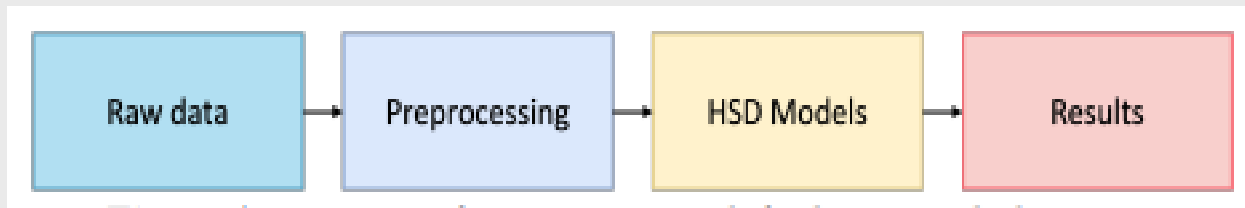
- Việc khảo sát, thu thập ý kiến của người dùng để nâng cao chất lượng dịch vụ là một yêu cầu cấp thiết .
- Nhiều nghiên cứu hỗ trợ việc khảo sát, thu thập ý kiến nhưng còn nhiều thiếu sót.
- Nhiều thí nghiệm khác nhau đã được tiến hành làm cơ sở để cho thấy mô hình PhoBERT-CNN được đề xuất vượt trội hơn

Overview



Một hệ thống xử lý ngôn ngữ tự nhiên tiên tiến, kỹ thuật phân loại những bình luận.

- Đầu vào: những bình luận tiếng Việt trên google rating.
- Đầu ra: một trong ba nhãn khác nhau được các bộ phân loại dự đoán.

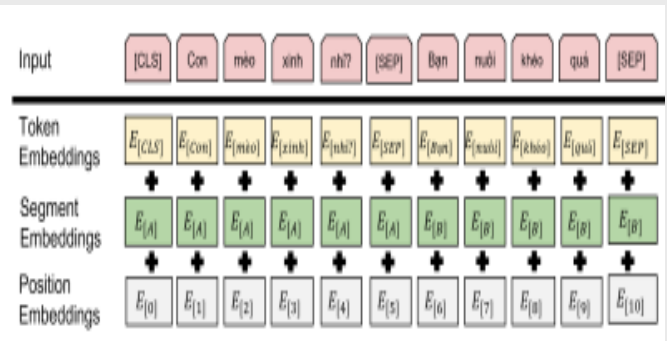


Hình 1: Tổng quan về phương pháp HSD

Description

Nội dung 1: Trình bày kiến trúc của mô hình PhoBERT [1], xây dựng bộ dữ liệu và cách thức mô hình PhoBERT được điều chỉnh để hoạt động như một lớp nhúng từ để trích xuất thông tin từ bộ dữ liệu.

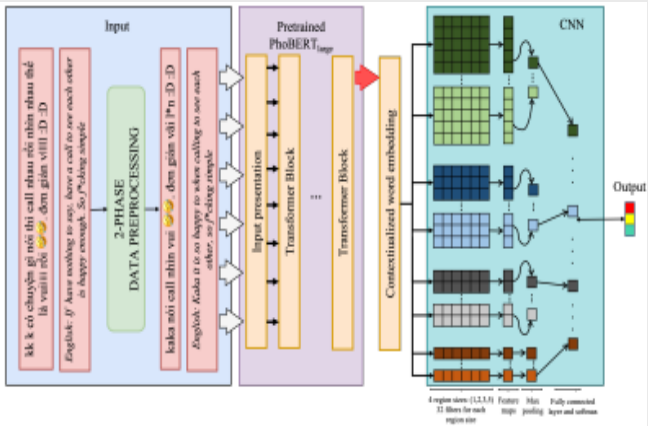
- Tìm hiểu mô hình kiến trúc của mô hình
- Xây dựng bộ dữ liệu bình luận theo tiếng Việt gồm 10000 bình luận theo google rating.
- Tìm hiểu cách xử lý, mã hóa dữ liệu khi đưa vào của mô hình PhoBERT.



Hình 2: Quá trình dữ liệu đầu vào của mô hình PhoBERT

Nội dung 2: Trình bày kiến trúc của các lớp của mô hình CNN dùng trong nghiên cứu này.

- Phân tích kỹ thuật tích chập và gộp của CNN trong việc trích xuất các khái niệm và từ khóa chính của văn bản.
- Phân tích tại sao lại sử dụng CNN là mô hình để giải quyết các nhiệm vụ phân loại văn bản ngắn.
- Tìm hiểu các thành phần kiến trúc INPUT, CONV1D, POOLING, DROPOUT và FC dùng để đánh giá trong bài toán.



Hình 3: Tổng quan mô hình HSD

Nội dung 3: Trình bày bài toán phân loại tiếng Việt áp dụng mô hình HSD là sự kết hợp giữa mô hình PhoBERT được huấn luyện trước và mô hình Text-CNN .

- Tìm hiểu cách kết hợp giữa mô hình PhoBERT có nhiệm vụ trích xuất các đặc điểm từ các câu cho đầu vào của mô hình Text-CNN và việc nhúng từ theo ngữ cảnh của các nhận xét từ PhoBERT sẽ được đưa vào mô hình Text-CNN để lấy bản đồ tính năng.
- Cài đặt, huấn luyện mô hình PhoBERT được huấn luyện trước và mô hình Text-CNN trên bộ dữ liệu tự xây dựng.