

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

NGUYỄN GIANG THANH PHƯƠNG - 230201050

Tóm tắt

- Lớp: CS2205.APR2023
- Link Github: <https://github.com/PhuongNGT/CS2205.APR2023>
- Link YouTube video: https://youtu.be/WVTEu_tz8f8

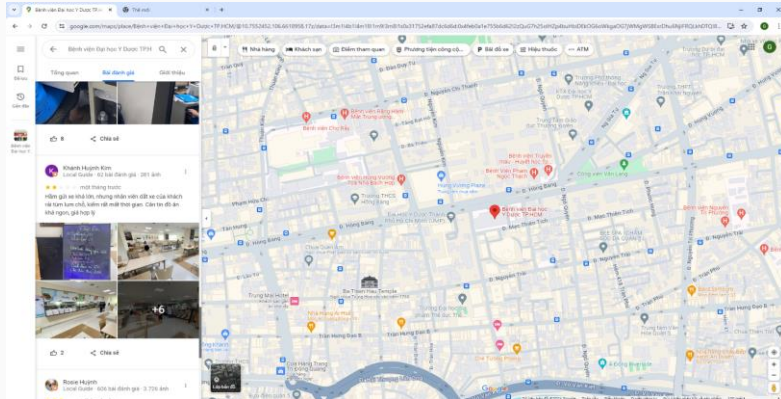


Nguyễn Giang Thanh Phương

Giới thiệu



- 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 nghiên cứu và thử nghiệm mô hình PhoBERT-CNN.



Mục tiêu

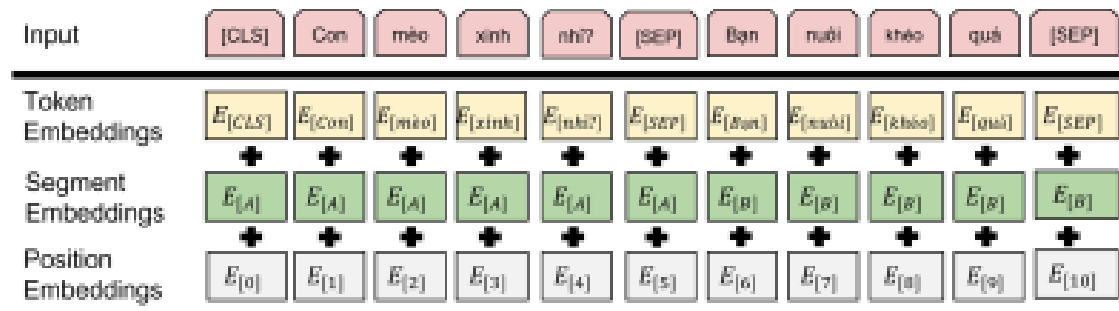


- Triển khai, đánh giá hiệu quả mô hình kết hợp PhoBERT và CNN trong việc phân loại các bình luận thu thập từ Google Rating.
- So sánh mô hình kết hợp PhoBERT và CNN với các mô hình khác về nhiệm vụ phân loại văn bản tiếng Việt.

Nội dung và Phương pháp

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.
- 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 và Phương pháp

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

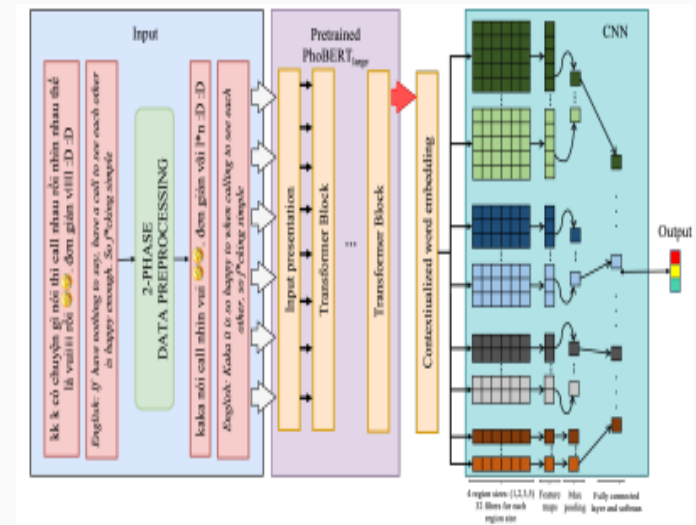
- Phân tích kỹ thuật tích chập và gộp của CNN trong việc trích xuất các khái niệm và từ khóa chính của văn bản.
- Phân tích tại sao lại sử dụng CNN là mô hình để giải quyết các nhiệm vụ phân loại văn bản ngắn.
- Tìm hiểu các thành phần kiến trúc INPUT, CONV1D, POOLING, DROPOUT và FC dùng để đánh giá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



Nội dung và Phương pháp

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.

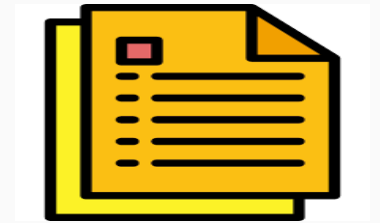


Hình 3: Tổng quan mô hình HSD

Kết quả dự kiến

Việc ứng dụng PhoBERT-CNN trong phân loại bình luận nhiều khía cạnh mang lại nhiều kết quả mong đợi bao gồm:

- Đề xuất một giải pháp tiên tiến mới cho mô hình HSD với quy trình xử lý dữ liệu gồm hai giai đoạn để làm sạch tập dữ liệu.
- Tìm ra được một cách tiếp cận hiệu quả và đơn giản để phân tích bình luận bằng tiếng Việt dựa trên mô hình kết hợp PhoBERT-CNN.
- Xây dựng thành công hệ thống phân tích bình luận bằng tiếng Việt dựa trên Google Rating theo thời gian thực.



Tài liệu tham khảo

- [1]. Nguyen, D.Q., Tuan Nguyen, A.: PhoBERT: Pre-trained language models for Vietnamese. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp. 1037–1042. Association for Computational Linguistics, Online (2020)
- [2]. He, C., Chen, S., Huang, S., Zhang, J., Song, X.: Using convolutional neural network with bert for intent determination. In: 2019 International Conference on Asian Language Processing (IALP), pp. 65–70 (2019). IEEE
- [3]. Kim, Y.: Convolutional neural networks for sentence classification. In: Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP), pp. 1746–1751. Association for Computational Linguistics, Doha, Qatar (2014).
- [4]. Pham-Hong, B.-T., Chokshi, S.: Pgsg at semeval-2020 task 12: Bert-lstm with tweets' pretrained model and noisy student training method. In: Proceedings of the Fourteenth Workshop on Semantic Evaluation, pp. 2111–2116 (2020).
- [5]. Li, X., Bing, L., Zhang, W., Lam, W.: Exploiting BERT for end-to-end aspect-based sentiment analysis. In: Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019), pp. 34–41. Association for Computational Linguistics, Hong Kong, China (2019).
- [6]. Yi, R., Hu, W.: Pre-trained bert-gru model for relation extraction. In: Proceedings of the 2019 8th International Conference on Computing and Pattern Recognition, pp. 453–457 (2019)