NÂNG CAO HIỆU SUẤT PHÁT HIỆN THƯ RÁC VỚI BERT

Nguyễn Duy Phương - 18521276

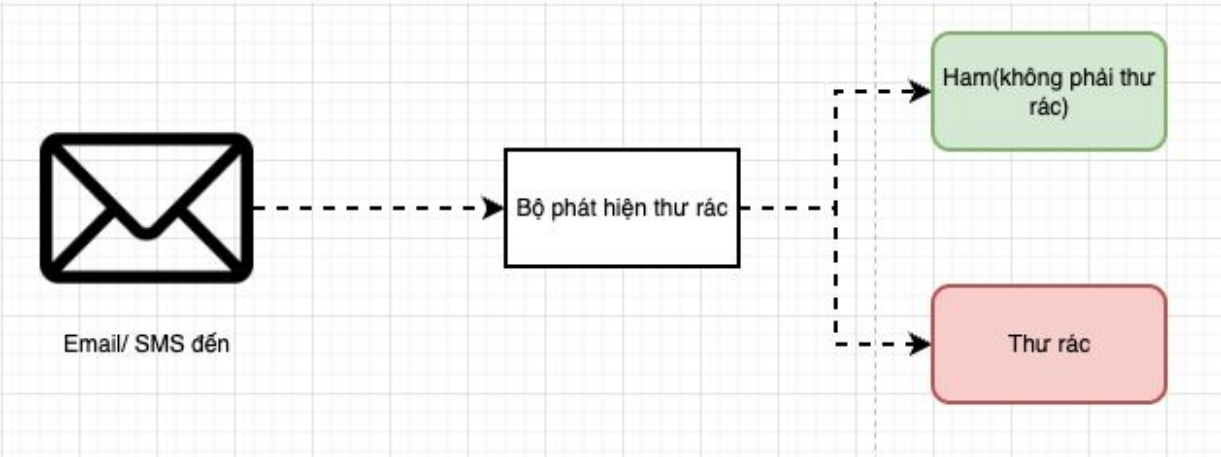
Tóm tắt

NGUYỄN DUY PHUONG
MSHV: 18521276

- Link Github: https://github.com/duyphuong0211/18521276_CS2205_Final_Project.git
- Link YouTube video: https://youtu.be/fsxHMHfPZ6I

Giới thiệu

- Email và SMS là các công cụ quan trọng trong giao tiếp hiện đại.
- Email được sử dụng rộng rãi trong nhiều lĩnh vực như kinh doanh, doanh nghiệp và người dùng cá nhân.



Giới thiệu

- Thư rác là một vấn đề phổ biến và đe dọa trải nghiệm người dùng và bảo mật thông tin cá nhân.
- Bộ phát hiện thư rác là một giải pháp quan trọng để phân loại và ngăn chặn thư điện tử không mong muốn.
- Sử dụng mô hình NLP BERT để xây dựng bộ phát hiện thư rác và SMS.

Mục tiêu

- Sử dụng BERT để phát hiện thư rác hiệu quả: Tận dụng sức mạnh của BERT, để xây dựng một hệ thống phát hiện thư rác hiệu quả và chính xác.
- Cải thiện hiệu suất phân loại hiệu quả: Sử dụng BERT context và tinh chỉnh mô hình bằng tác vụ phân loại nhị phân, mục đích là để đạt được độ chính xác và độ chính xác cao trong việc phân biệt giữa thư rác và email (ham).
- Tăng cường mã hóa và đại diện đầu vào: Mục tiêu thứ ba là cải thiện quy trình mã hóa và đại diện đầu vào cho BERT.

Nội dung và Phương pháp

- Tiền xử lý - Dữ liệu email từ SpamAssassin, SMS Spam Collection v.1, Enron corpus và Ling-Spam corpus được làm sạch bằng cách xóa URL, ký tự không mong muốn và thẻ HTML.
- Tạo tập dữ liệu đào tạo - Tập hợp dữ liệu được xử lý trước được chia thành tập dữ liệu đào tạo và đánh giá. Tập dữ liệu đào tạo được sử dụng để đào tạo mô hình phát hiện thư rác.

Nội dung và Phương pháp

- Đào tạo mô hình - Mô hình cơ sở BERT được đào tạo bằng PyTorch. Email và nhãn tương ứng được cung cấp cho mô hình theo lô 16.
- Đánh giá - Bộ dữ liệu đánh giá được sử dụng để đánh giá hiệu suất của mô hình được đào tạo.

Kết quả dự kiến

- Sử dụng BERT để phát hiện thư rác hiệu quả: Qua việc sử dụng BERT và tinh chỉnh mô hình, chúng ta đã xây dựng thành công một hệ thống phát hiện thư rác mạnh mẽ và chính xác. BERT giúp chúng ta hiểu ngữ cảnh(context) và nắm bắt được các đặc trưng quan trọng để phân loại email là thư rác hay không.
- Cải thiện hiệu suất phân loại hiệu quả: Đạt được độ chính xác và độ chính xác cao trong việc phân biệt giữa thư rác và email hợp lệ (ham).

Kết quả dự kiến

- Tăng cường mã hóa và đại diện đầu vào: Cải thiện quy trình mã hóa và đại diện đầu vào cho BERT. Việc sử dụng tokenizer BERT và thực hiện quy trình mã hóa và đại diện cho email giúp chúng ta biểu diễn hiệu quả nội dung email và nắm bắt thông tin quan trọng.

Tài liệu tham khảo

- J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding," CoRR, vol. abs/1810.04805, 2018.
- X. Liu, H. Lu and A. Nayak, "A Spam Transformer Model for SMS Spam Detection," IEEE Access, vol. 9, pp. 80253-80263, 2021.
- Safaa Magdy, Yasmine Abouelseoud, Mervat Mikhail: Efficient spam and phishing emails filtering based on deep learning. Comput. Networks 206: 108826 (2022)
- Kamaljit Kaur, Parminder Kaur: Improving BERT model for requirements classification by bidirectional LSTM-CNN deep model. Comput. Electr. Eng. 108: 108699 (2023)
- OpenAI. (2021). ChatGPT: OpenAI's Conversational AI Language Model