Improving Spam Detection Performance With BERT

Nguyễn Duy Phương

Trường Đại Học Công Nghệ Thông Tin - DHQG TP.HCM

What ?

Nghiên cứu này nhằm tận dụng sức mạnh của BERT để:

- Tận dụng sức mạnh của BERT để xây dựng một hệ thống phát hiện thư rác hiệu quả và chính xác.
- Nâng cao hiệu suất phân loại giữa thư rác và email hợp lệ bằng cách sử dụng BERT và tinh chỉnh mô hình.
- Cải thiện quy trình mã hóa và đại diện đầu vào cho BERT.

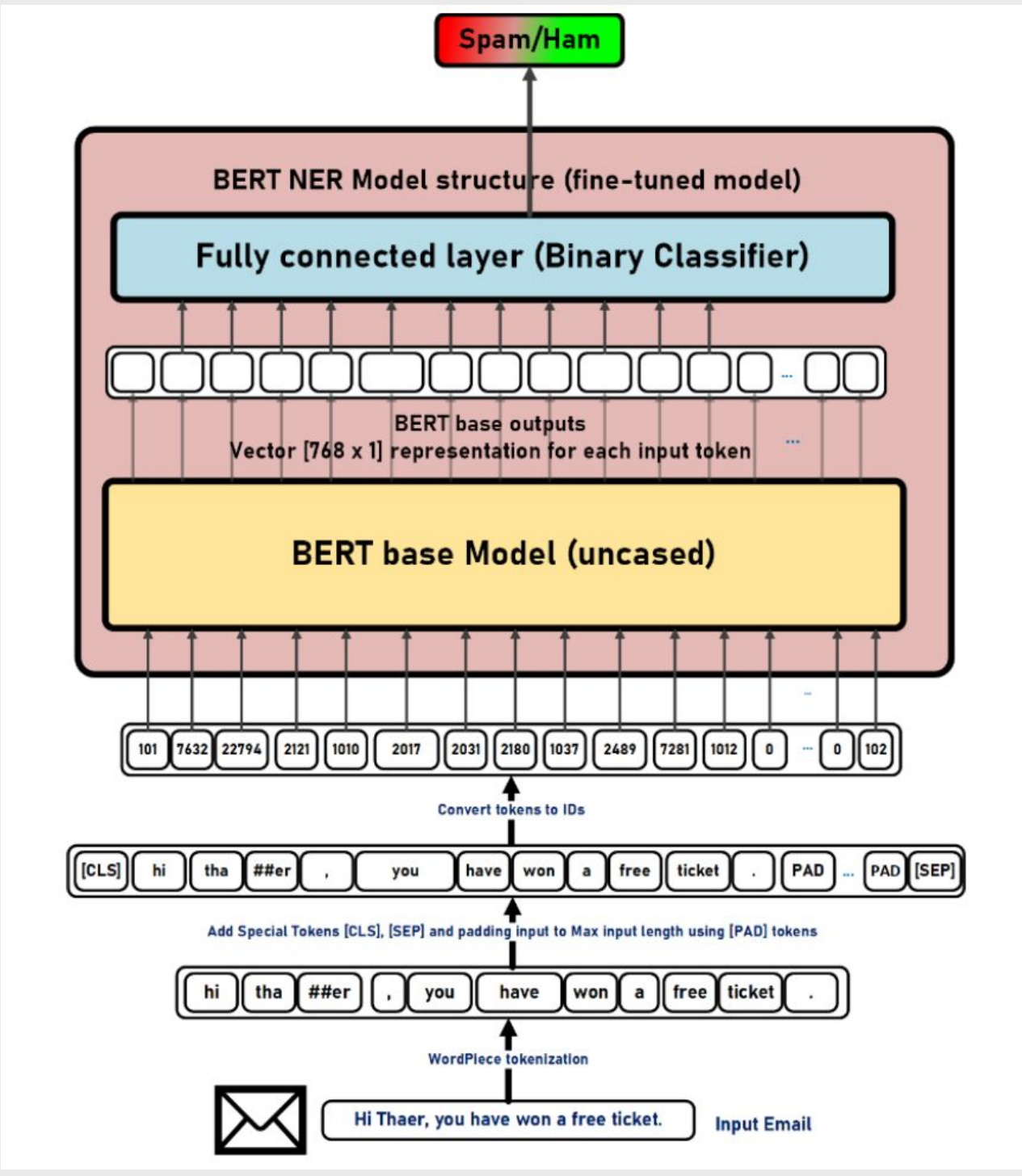
Why ?

- Phát hiện thư rác là một nhiệm vụ quan trọng vì nó bảo vệ người dùng khỏi nội dung độc hại, tiết kiệm tài nguyên mạng và chống lại tội phạm mạng. Nó ngăn chặn việc tiết lộ thông tin cá nhân và tài sản, giảm lãng phí tài nguyên băng thông và dung lượng lưu trữ không cần thiết, và ngăn chặn các hình thức tấn công và lừa đảo trực tuyến, bảo vệ người dùng và tổ chức.

Overview

Tiền xử lý dữ liệu → Tạo tập dữ liệu đào tạo → Đào tạo mô hình → Đánh giá

Description



4. Đánh giá

- Bộ dữ liệu đánh giá được sử dụng để đánh giá hiệu suất của mô hình được đào tạo.

3. 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.

2. 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.

1. Tiền xử lý dữ liệu

- 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.

NII

Nguyễn Duy Phương – Trường Đại học Công Nghệ Thông Tin - DHQG TP.HCM
TEL : 0399053869 Email : 18521276@gm.uit.edu.vn