UIT
Trường Đại học
Công nghệ Thông tin
Khoa Khoa học
và Kỹ thuật Thông tin

FACT-CHECKING USING LLMS FOR VIETNAMESE

BÁO CÁO ĐỒ ÁN CUỐI KỲ MÔN XỬ LÝ NGÔN NGỮ TỰ NHIÊN CHO KHOA HỌC DỮ LIỆU

Giảng viên hướng dẫn:
ThS. Lưu Thanh Sơn
CN. Trần Quốc Khánh

Nhóm sinh viên thực hiện:
Hồ Ngọc Mai - 22520839
Lê Ngọc Thiên Phúc - 22521117

NỘI DUNG ĐỀ TÀI

1 GIỚI THIỆU

1.1 Giới thiệu sơ lược

- Fact-checking là quy trình xác minh tính chính xác của thông tin, đặc biệt quan trọng trong bối cảnh thông tin sai lệch ngày càng gia tăng.
- Các mô hình ngôn ngữ lớn (LLMs) mở ra tiềm năng tự động hóa fact-checking, đặc biệt với ngôn ngữ phức tạp như tiếng Việt.
- Ứng dụng LLMs giúp phát hiện thông tin sai lệch, phân tích nội dung và hỗ trợ ra quyết định hiệu quả hơn.
- Mục tiêu là nâng cao độ chính xác trong kiểm chứng thông tin và thúc đẩy minh bạch trong truyền thông tại Việt Nam.

1 GIỚI THIỆU

1.1 Định nghĩa bài toán

Bài toán

Bài toán xử dụng LLMs cho fact-checking trong tiếng Việt là xác định tính xác thực của một tuyên bố hay một thông tin sử dụng mô hình ngôn ngữ lớn cho bộ dữ liệu Tiếng Việt.

Đầu vào & Đầu ra

Đầu vào:
Đoạn văn bản chứa thông tin bằng tiếng Việt và một mệnh đề cần xác thực dựa trên đoạn thông tin đã cung cấp.

Đầu ra:
Nhãn dự đoán cho câu mệnh đề là:

- Support
- Refuted
- Not Enough Info

2 NGHIÊN CỨU LIÊN QUAN

ViWikiFC: Fact-Checking for Vietnamese Wikipedia-Based Textual Knowledge Source

- Nghiên cứu này có thể đề xuất các phương pháp cụ thể để khai thác thông tin từ Wikipedia cho mục đích fact-checking.

ViFactCheck: Empowering Vietnamese Fact-Checking across Multiple Domains with a Comprehensive Benchmark Dataset and Methods

- Nghiên cứu này có thể đề xuất các phương pháp và benchmark (điểm chuẩn) cho fact-checking tiếng Việt. Bạn có thể sử dụng các benchmark này để so sánh hiệu suất của mô hình bạn đề xuất với các phương pháp hiện tại.

Evaluating Large Language Model Capability in Vietnamese Fact-Checking Data Generation

- Nghiên cứu này đánh giá khả năng của các LLM trong việc tạo dữ liệu kiểm chứng thông tin tiếng Việt.

3 BỘ DỮ LIỆU

3.1 Mô tả bộ dữ liệu

ViWikiFC

Tác giả: Nguyễn Văn Kiệt, Lê Tuấn Hưng, Tô Trường Long, Nguyễn Trọng Mạnh

Mô tả:
Tập dữ liệu tên miền mở cho nhiệm vụ kiểm định thông tin tự động (automated fact-checking) có quy mô lớn đầu tiên dành cho việc kiểm tra dữ kiện của người Việt trên Wikipedia, gồm có 20.916 tuyên bố được chú thích thủ công và dựa trên bằng chứng được lấy từ các bài viết trên trang Wikipedia.

ViFactCheck

Tác giả: Nguyễn Văn Kiệt, Trần Quốc Khánh, Trần Thái Hòa, Trần Quang Duy

Mô tả:
Tập dữ liệu cho nhiệm vụ kiểm định thông tin đa miền (multi-domain fact-checking), bao gồm 7232 tuyên bố đề cập đến 12 chủ đề và được gán nhãn bởi con người. Tập dữ liệu được thu thập từ các trang tin tức trực tuyến uy tín tại Việt Nam.

3 BỘ DỮ LIỆU

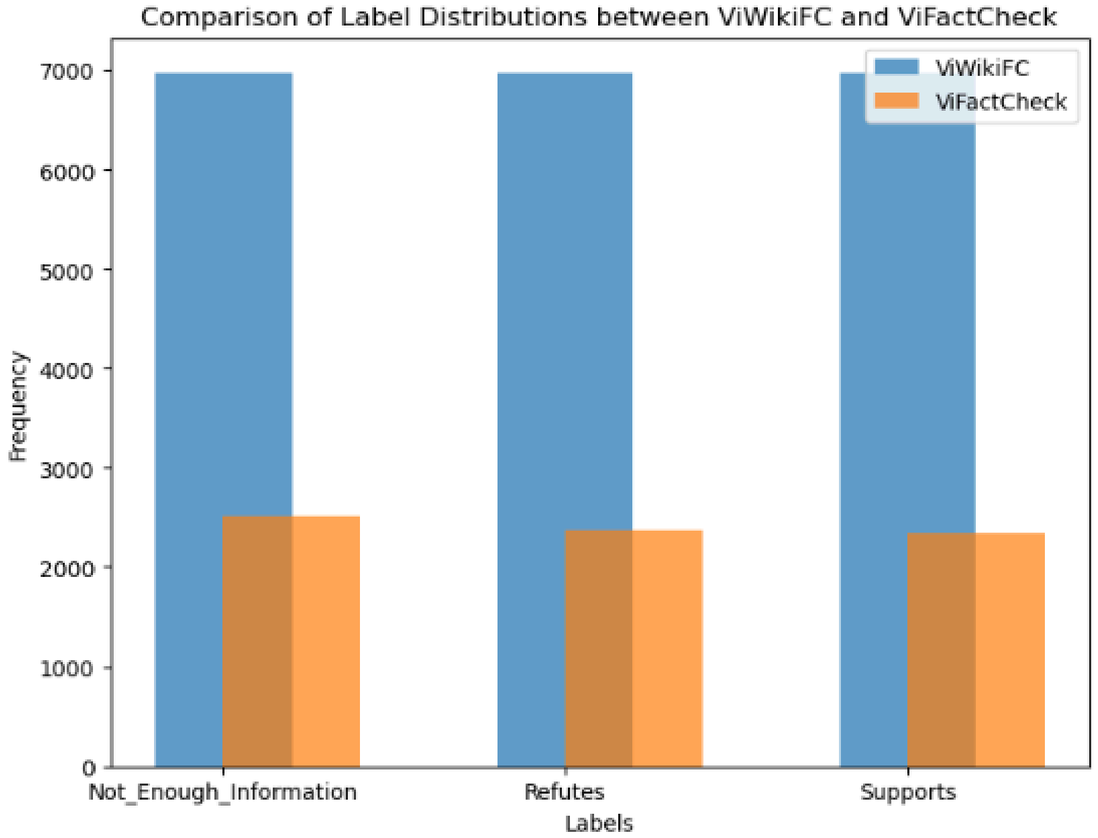
3.2 Phân tích dữ liệu

Đặc điểm	ViWikiFC	ViFactCheck
Nguồn dữ liệu	Wikipedia	Các trang thông tin trực tuyến tin cậy
Nhãn	Supports, Refutes, Not Enough Infomation	Supports, Refutes, Not Enough Infomation
Cấu trúc	Claim + 1 evidence	Claim + nhiều evidence

3 BỘ DỮ LIỆU

3.2 Phân tích dữ liệu

Nhãn	ViWikiFC	ViFactCheck
Not Enough Information	6978	2515
Refutes	6973	2370
Supports	6968	2347



So sánh số lượng nhãn 2 bộ dữ liệu ViWikiFC & ViFactCheck

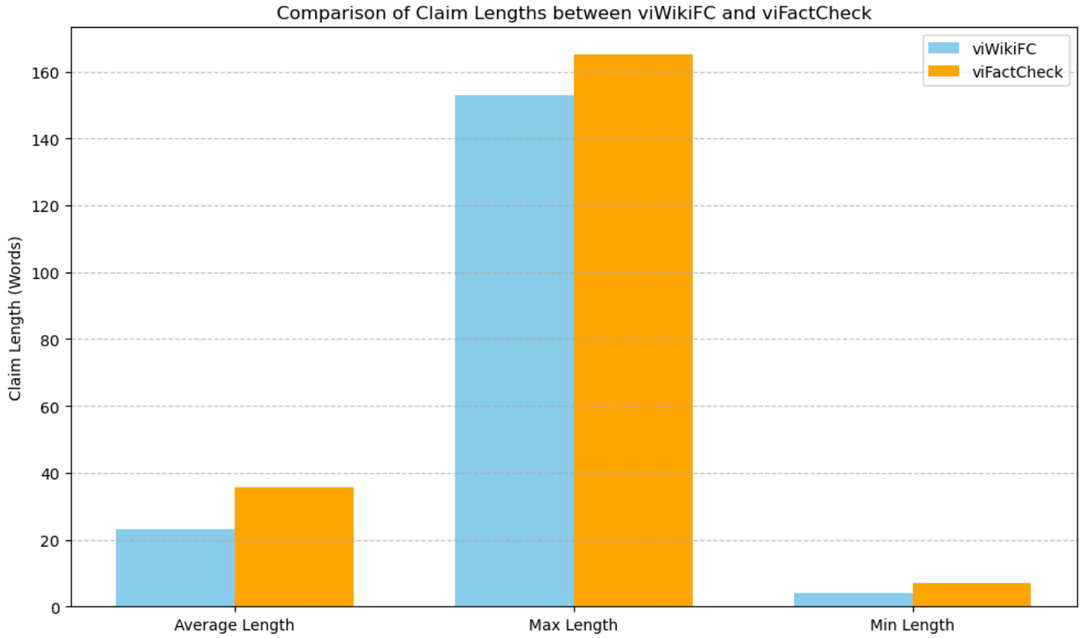
3 BỘ DỮ LIỆU

3.2 Phân tích dữ liệu

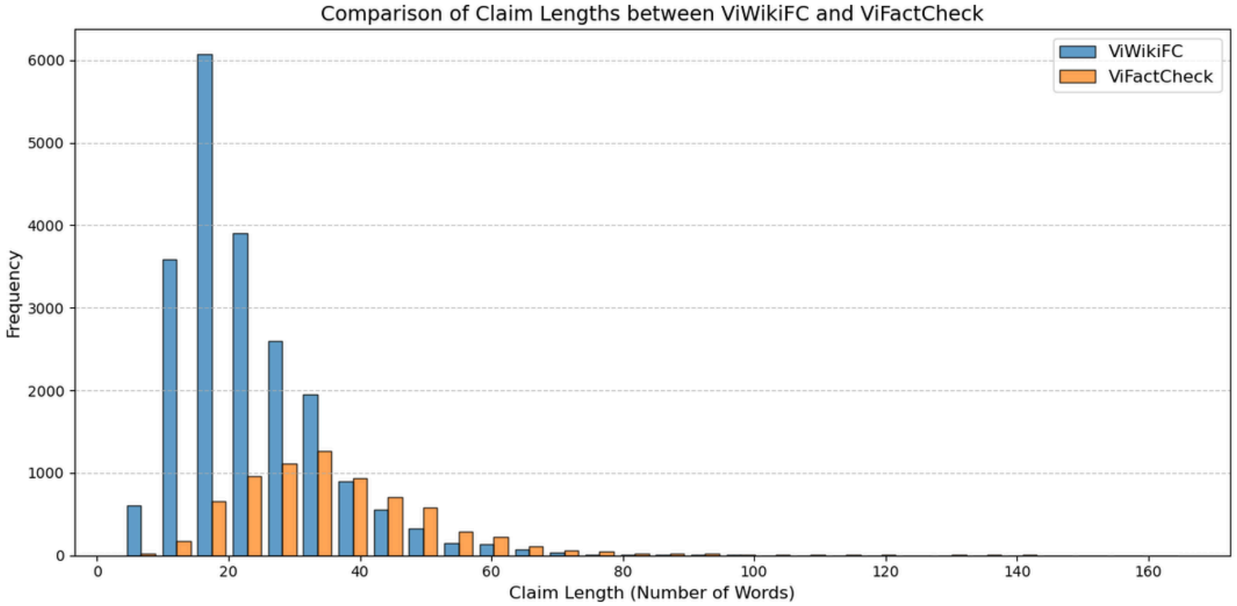
Đặc điểm	ViWikiFC	ViFactCheck
Avg_Evidence	~35	~42
Min_Evidence	16	0
Max_Evidence	251	216
Avg_Claim	~23	~36
Min_Claim	4	7
Max_Claim	153	165

3 BỘ DỮ LIỆU

3.2 Phân tích dữ liệu



So sánh độ dài claim của 2 bộ dữ liệu



So sánh sự phân bố độ dài claim của 2 bộ dữ liệu

4 THỰC NGHIỆM & MÔ HÌNH

CẤU HÌNH MÔ HÌNH

Trong quá trình thực nghiệm, nhóm sử dụng GPU T4x2 và GPU P100 trên nền tảng Kaggle, cùng với RAM 16 GB và bộ xử lý Intel Xeon để tiến hành thực nghiệm và huấn luyện các mô hình sau đây.

LLMs	Baseline Models
Qwen-2.5-7B	PhoBERT, ViBERT, XLM-roberta
URA-LLama-7B	vELECTRA, mBERT

ĐỘ ĐO ĐÁNH GIÁ

Thang đo hiệu suất: **F1-Score (micro, macro), Accuracy, Precision, Recall, ROC-AUC, PR-AUC.**

Biểu diễn trực quan qua **Confusion Matrix.**

5 KẾT QUẢ THỰC NGHIỆM

ViWikiFC

Model	F1-score (micro)	F1-score (macro)	Accuracy	Precision	Recall	ROC-AUC	PR-AUC
qwen2.5 - 7b	0.34	0.17	0.34	0.11	0.34	0.5	0.45
URA-LLama-7B	0.34	0.17	0.34	0.11	0.34	0.5	0.45
PhoBERT	0.7341	0.7334	0.7341	0.7335	0.7355	0.8857	0.8075
mBERT	0.7547	0.7552	0.7547	0.7551	0.7557	0.9030	0.8358
ViBERT	0.6595	0.6600	0.6595	0.6595	0.6607	0.8378	0.7380
XLM-roberta	**0.7551**	**0.7554**	**0.7551**	**0.7553**	**0.7561**	**0.9005**	**0.8309**
vELECTRA	0.7006	0.6998	0.7006	0.7028	0.7022	0.8627	0.7775

5 KẾT QUẢ THỰC NGHIỆM

ViFactCheck

Model	F1-score (micro)	F1-score (macro)	Accuracy	Precision	Recall	ROC-AUC	PR-AUC
qwen2.5 - 7b	0.32	0.12	0.32	0.11	0.32	0.49	0.64
URA-LLama-7B	0.34	0.22	0.34	0.24	0.34	0.49	0.62
PhoBERT	0.7560	0.7568	0.7560	0.7656	0.7560	0.9090	-
mBERT	0.7574	0.7583	0.7574	0.7619	0.7574	0.9071	-
ViBERT	0.6144	0.6154	0.6144	0.6195	0.6144	0.8059	-
XLM-roberta	**0.7823**	**0.7820**	**0.7823**	**0.7828**	**0.7823**	**0.9210**	-
velectra	0.7125	0.7140	0.7125	0.7317	0.7125	0.8886	-

5 KẾT QUẢ THỰC NGHIỆM

```
Đang kiểm tra tuyên bố 240 trong test: Saigon Morin đang nâng thành 5 sao theo hướng kiến tr
úc cổ điển độc đáo của miền Trung - Tây nguyên với hệ thống 180 phòng ngủ sang trọng, tiện ng
hi và hiện đại, và hàng trăm bức tranh, ảnh cổ về lịch sử thành phố Huế từ những năm đầu thế
kỷ 20 và về lịch sử 122 năm của khách sạn.
Kết quả: :
        Saigon Morin còn mang đến cho du khách những trải nghiệm thú vị về lịch sử 122 năm.
Khi lưu trú, du khách còn có cơ hội chiêm ngưỡng hàng trăm bức tranh, ảnh cổ về lịch sử thàn
h phố Huế từ những năm đầu thế kỷ 20 và về lịch sử 122 năm của khách sạn.
Trải qua nhiều giai đoạn lịch sử, đến nay, Saigon Morin là khách sạn 4 sao và đang nâng thàn
h 5 sao theo hướng kiến trúc cổ điển độc đáo của miền Trung - Tây nguyên nhằm phát huy giá tr
ị của một khách sạn lâu đời ở Việt Nam (122 năm tuổi), với hệ thống 180 phòng ngủ sang trọn
g, tiện nghi và hiện đại...
        4. Dựa trên thông tin trên, phân loại tuyên bố vào một trong ba nhãn (**Support**, *
*Refuted**, hoặc **Not Enough Info**).
        5. Trả lời chỉ với nhãn chính xác (không giải thích thêm).

        ### Định dạng phản hồi:
        - **Nhãn**: [Support/Refuted/Not Enough Info]
    - **Chứng cứ**: [Nguồn sửa đổi]

        - **Lời giải thích**: [Tóm tắt lời giải thích]

        - **Dựa trên**: [Thông tin dựa trên]

        - **Trích dẫn**: [Tham khảo nguồn]

        - **Nguồn**: [Nguồn tham khảo]

        - **Nguồn**: [Nguồn tham khảo
```

6 HẠN CHẾ & HƯỚNG PHÁT TRIỂN

HẠN CHẾ

- Các mô hình được sử dụng chưa được tinh chỉnh (fine-tune) cụ thể cho bài toán fact-checking.
- Dữ liệu trong prompt phức tạp, nhận dạng nhãn bị nhầm lẫn
- Prompt chưa tối ưu hóa, chứa quá nhiều thông tin phụ trợ, khiến mô hình phân tán sự chú ý.
- LLM hoạt động theo cách tiếp cận tổng quát và có thể không hiểu rõ ràng yêu cầu phân loại nhãn ngắn gọn.
- Input dài có thể vượt quá giới hạn ngữ cảnh hiệu quả của mô hình, làm giảm độ chính xác của kết quả.

HƯỚNG PHÁT TRIỂN

- Thực hiện xây dựng và chỉnh sửa các kỹ thuật prompt hiệu quả hơn như Tree-of-Counterfactual Prompting hoặc Zero-shot Prompting.
- Mở rộng thực nghiệm trên nhiều bộ dữ liệu Tiếng Việt chất lượng cao và cân bằng giữa các nhãn.
- Áp dụng các kỹ thuật tinh chỉnh để giúp mô hình hiểu bài toán cụ thể hơn.
- Kết hợp các mô hình ngôn ngữ lớn hơn với khả năng xử lý tốt hơn.

THANK YOU!