

PHÁT TRIỂN FRAMEWORK "PILOT" ĐỂ DỰ ĐOÁN KẾT QUẢ CÁC VỤ ÁN DÂN SỰ TRONG HỆ THỐNG BẢN ÁN TIẾNG VIỆT

Học viên: Nguyễn Phạm Thành Hưng - 240101047

Môn học: CS2250 - Phương pháp luận NCKH

Giảng viên: PGS.TS Lê Đình Duy

Tóm tắt

- Lớp: CS2205.FEB2025
- Link Github của nhóm:
<https://github.com/HungGradUIT19/CS2205.FEB2025>
- Link YouTube video: <https://youtu.be/O90eEzzZuE0>
- Tên thành viên: Nguyễn Phạm Thành Hưng
- MSHV: 240101047



Giới thiệu

- Bài toán dự đoán kết quả các vụ án được nghiên cứu nhằm hỗ trợ quá trình ra quyết định dựa trên dữ liệu trong các vụ án pháp lý.
- Nhiệm vụ của bài toán này là sử dụng thông tin liên quan, chẳng hạn như mô tả tình tiết và bối cảnh pháp lý thực tế của vụ án, để dự đoán kết quả cuối cùng của vụ án.

Giới thiệu

- PILOT: truy xuất thông tin về các án lệ liên quan -> mã hóa vụ án đang cần dự đoán với các vụ án có liên quan -> bộ xử lý sự thay đổi nội dung các bộ luật theo thời gian để đưa ra dự đoán kết quả cho vụ án.
- Hạn chế: Thuật toán truy xuất thông tin các án lệ chưa tốt, cần phải được cải tiến.
- Câu hỏi đặt ra: “Làm thế nào để một hệ thống truy xuất có thể truy xuất chính xác các vụ án trong tiền lệ phù hợp với vụ án đang cần được dự đoán kết quả?”

Giới thiệu

Đề xuất: Framework ViPILOT cải tiến từ PILOT. Tập trung vào việc cải tiến thuật toán truy xuất các án lệ có liên quan đến vụ án đang cần được dự đoán -> Tăng hiệu suất dự đoán của mô hình.

Mục tiêu

- Xây dựng mô hình phù hợp nhất cho việc truy xuất các bản án để tích hợp vào ViPILOT.
- Xây dựng bộ dữ liệu tiếng Việt về các vụ án dân sự với quy mô lớn cho việc đánh giá các mô hình dự đoán.
- Đánh giá ViPILOT với các phương pháp hiện có -> kỳ vọng mô hình sẽ dự đoán kết quả các vụ án tốt hơn.

Nội dung và Phương pháp

- Nghiên cứu về Text Retrieval, tìm hiểu các bước xây dựng một mô hình truy vấn văn bản, đánh giá mô hình.
- Tìm hiểu về bộ dữ liệu ECHR2023 để nắm cấu trúc của một bộ dữ liệu các bản án cho việc lựa chọn mô hình truy xuất.
- Nghiên cứu các kỹ thuật Text Retrieval truyền thống đến state-of-the-art.
- Áp dụng kỹ thuật phù hợp nhất với bài toán, có thể phát triển, tinh chỉnh mô hình để truy xuất thông tin vụ án tốt hơn.
- Xây dựng bộ dữ liệu tiếng Việt về các vụ án tại Việt Nam với quy mô lớn.

Nội dung và Phương pháp

- Đánh giá ViPILOT với các phương pháp hiện có trên 2 bộ dữ liệu: ECHR2023 và bộ dữ liệu được xây dựng.
- Sử dụng 4 chỉ số: micro-F1, micro-Jaccard, microPR-AUC và micro-ROC-AUC.
- Xây dựng bảng so sánh các mô hình cho 2 bộ dữ liệu được thử nghiệm.
- Kỳ vọng: ViPILOT có số liệu được cải thiện hơn so với các phương pháp trước trong việc dự đoán kết quả vụ án đặc biệt là trên bộ dữ liệu tiếng Việt.

Kết quả dự kiến

- Đề xuất mô hình ViPILOT đạt được hiệu suất tốt hơn so với các phương pháp hiện có.
- Chứng minh được tầm quan trọng của hệ thống truy xuất án lệ đối với hiệu suất dự đoán của mô hình.
- Mô hình có thể ứng dụng được trong thực tế hỗ trợ các thẩm phán đưa ra phán quyết cuối cùng.

Tài liệu tham khảo

- [1]. Lang Cao, Zifeng Wang, Cao Xiao, Jimeng Sun. "PILOT: Legal Case Outcome Prediction with Case Law". Association for Computational Linguistics (ACL) 2024.
- [2]. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schutze. Book: "Introduction to Information Retrieval". Cambridge University Press 2008.
- [3]. Kailash Hambarde, Hugo Proenca. "Information Retrieval: Recent Advances and Beyond". Institute of Electrical and Electronics Engineers (IEEE) 2023: 1-18.
- [4]. S. K. M. Wong, W. Ziarko, P. C. N. Wong, "Generalized vector spaces model in information retrieval," in Proc. 8th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. - SIGIR, 1985.
- [5]. J. Gao, J.-Y. Nie, G. Wu, G. Cao, "Dependence language model for information retrieval," in Proc. 27th Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Jul. 2004.
- [6]. D. Ganguly, D. Roy, M. Mitra, G. J. F. Jones, "Word embedding based generalized language model for information retrieval," in Proc. 38th Int. ACM SIGIR Conf. Res. Develop. Inf. Retr., Aug. 2015.
- [7]. J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," 2018, arXiv:1810.04805.
- [8]. A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language models are unsupervised multitask learners" OpenAI, vol. 1, no. 8, p. 9, 2019.
- [9]. Tòa án Nhân dân Tối cao Nước Cộng hòa Xã hội chủ nghĩa Việt Nam. "Trang thông tin điện tử công bố bản án, quyết định của tòa án". Có sẵn tại: <https://congbobanan.toaan.gov.vn/0tat1cvn/ban-an-quyet-dinh>
- [10]. Nguyễn Thanh Trúc Mai, Nguyễn Minh Đạt, Lưu Thanh Sơn, Nguyễn Văn Kiệt (GVHD), "ViMRHP: A Vietnamese Benchmark Dataset for Multimodal Review Helpfulness Prediction via Human-AI Collaborative Annotation". The 30th International Conference on Natural Language & Information Systems (NLDB 2025).