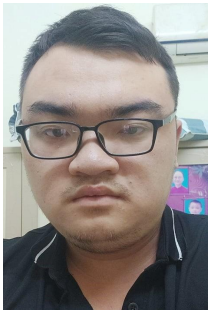
THÔNG TIN CHUNG CỦA NHÓM

- Link YouTube video của báo cáo:
  https://youtu.be/ky-eyqc82Bk
- Link slides:
  https://github.com/datthanhle/CS519.M11.KHCL/blob/main/Đồ%20án%20cuối%20kỳ/Transformer%20without%20tears%20-%20Slides.pdf

• Họ và Tên: Trần Vĩ Hào
• MSSV: 19521482
 
	• Lớp: CS519.M11.KHCL
	• Tự đánh giá (điểm tổng kết môn): 8.5/10
	• Số buổi vắng: 0
	• Số câu hỏi QT cá nhân: 3
	• Số câu hỏi QT của cả nhóm: 12
	• Link Github: https://github.com/hlhkudo/CS519.M11.KHCL
	• Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm:
	  ○ Lên ý tưởng cho đồ án
	  ○ Viết báo cáo đồ án, chỉnh sửa Slide và Poster
	  ○ Làm video YouTube

• Họ và Tên: Trương Quốc Bình • MSSV: 19521270 	• Lớp: CS519.M11.KHCL • Tự đánh giá (điểm tổng kết môn): 8.5/10 • Số buổi vắng: 1 • Số câu hỏi QT cá nhân: 3 • Số câu hỏi QT của cả nhóm: 12 • Link Github: https://github.com/noeffortnomoney/CS519.M11.KHCL • Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: ○ Làm Poster
• Họ và Tên: Lê Thành Đạt • MSSV: 17520332 	• Lớp: CS519.M11.KHCL • Tự đánh giá (điểm tổng kết môn): /10 • Số buổi vắng: • Số câu hỏi QT cá nhân: • Số câu hỏi QT của cả nhóm: • Link Github: https://github.com/datthanhle/CS519.M11.KHCL • Mô tả công việc và đóng góp của cá nhân cho kết quả của nhóm: ○ Làm Slide thuyết trình

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI
MÔ HÌNH TRANSFORMER WITHOUT TEARS: CẢI THIỆN CHUẨN HÓA CỦA KỸ THUẬT SELF-ATTENTION

TÊN ĐỀ TÀI TIẾNG ANH
TRANSFORMER WITHOUT TEARS: IMPROVING THE NORMALIZATION OF SELF-ATTENTION

TÓM TẮT

Với đề tài này, chúng tôi tập trung giải quyết ba vấn đề chính thuộc lĩnh vực *Xử lý ngôn ngữ tự nhiên*, mà cụ thể là *Dịch máy*. **Thứ nhất**, chúng tôi dựa trên mô hình Transformer truyền thống, cải tiến nó nhằm tạo ra phiên bản ***Transformer without tears***; sự điều chỉnh sẽ tập trung vào phần chuẩn hóa, từ đó phát sinh ra những bài toán con mà chúng tôi cần giải quyết có liên quan đến các kỹ thuật Postnorm, Prenorm, Scalenorm và Fixnorm, giải quyết được chúng sẽ giúp cải thiện quá trình huấn luyện của mô hình. **Thứ hai**, chúng tôi chủ động xây dựng bộ dữ liệu song ngữ Anh-Việt mới có tên gọi là BHD-EnVi trong bối cảnh nguồn dữ liệu dùng để huấn luyện cho bài toán dịch máy (Anh-Việt) còn nhiều hạn chế. **Cuối cùng**, chúng tôi xây dựng chương trình ứng dụng chạy trên nền web, giúp bài toán nặng tính lý thuyết trở thành một sản phẩm mang tính ứng dụng thực tế. Đi qua từng phần của bài báo cáo, các bạn sẽ thấy được những mục tiêu mà chúng tôi đã đề ra, những nội dung chi tiết sẽ được giải quyết ở đề tài này và những phương pháp mà chúng tôi đã áp dụng để giải quyết bài toán. Ngoài ra, kế hoạch thực hiện và phân công nhiệm vụ đã được chúng tôi vạch ra chi tiết ở phần cuối của báo cáo.

GIỚI THIỆU

Theo bảng xếp hạng chỉ số thông thạo tiếng Anh toàn cầu EF English Proficiency Index (EF EPI) được công bố vào năm 2021, Việt Nam được xếp vào nhóm thông thạo tiếng anh thấp với thứ hạng 66/112. Việc học một ngôn ngữ khác với tiếng mẹ đẻ vẫn là một việc gì đó khó khăn với dân ta. Trong bối cảnh hội nhập như hiện nay, không khó để bắt gặp những vị khách Tây, từ đường phố ở các trung tâm đô thị lớn đến các khu du lịch ở miền quê. Việc có thể thông thạo tiếng Anh trong bối cảnh này không chỉ có lợi cho bản thân chúng ta mà còn có lợi hơn cho cả nền kinh tế của một đất nước. Để giải quyết vấn đề cố hữu này nhiều người đã chọn tìm đến những công cụ hỗ trợ dịch ngôn ngữ như Google Translate,... thay vì chọn cách học tập truyền thống. Nhờ vậy mà tầm quan trọng của bài toán Dịch máy lại được tăng thêm vài bậc.

Dịch máy (Machine Translation) là một trong những ứng dụng của *Xử lý ngôn ngữ tự nhiên*. Nó được hiểu là việc thực hiện dịch một ngôn ngữ này (ngôn ngữ nguồn) sang một hoặc nhiều ngôn ngữ khác (ngôn ngữ đích) một cách tự động, không có sự can thiệp của con người trong quá trình dịch. Bài toán dịch máy đã xuất hiện từ rất lâu và sự cần thiết của nó là không phải bàn cãi khi nó được gắn với mọi lĩnh vực cuộc sống như du lịch, ngoại giao, kinh tế, văn học, v.v… và dễ dàng tiếp cận với mọi đối tượng. Hiện nay, dịch máy đã có thể thực hiện trên cả câu, rồi dùng ngữ cảnh để quyết định xem nên chọn nghĩa nào cho chính xác nhất thay vì dịch từng từ hay cụm từ như trước đây. Do đó, các nghiên cứu về nó cũng chuyển dần sang dịch máy nơ ron (Neural Machine

Translation), cách tiếp cận này mang lại những kết quả thực sự tốt và ngày càng trở nên phổ biến.
RNN, LSTM, GRU là các phương pháp tiếp cận hiện đại trong mô hình ngôn ngữ và dịch máy, từ đó khắc phục được những hạn chế gặp phải trong mạng nơ ron truyền thống. Tuy nhiên, trong nhiều bài toán về dịch thuật, việc cải thiện cũng không đáng kể. Chính vì thế kỹ thuật Attention được áp dụng để mang lại hiệu quả cao hơn. Cách tiếp cận sequence-to-sequence with attention[1] là một trong những mô hình đầu tiên áp dụng kỹ thuật này. Năm 2017, các kỹ sư của Google đã giới thiệu kỹ thuật Self-attention và đề xuất mô hình Transformer[2], cho phép thay thế hoàn toàn kiến trúc recurrent của mô hình RNN bằng các mô hình fully-connected.
Với dịch máy, không có bất kỳ mô hình nào là hoàn hảo cả. Ở mỗi thời đại khác nhau, những từ ngữ mới hay tiếng lóng trong một loại ngôn ngữ nhất định lại được sinh ra. Với việc vẫn chưa hài lòng với mô hình Transformer hiện tại, vì vậy mà trong đề tài này chúng tôi quyết định nghiên cứu và đề xuất một mô hình được cải tiến từ Transformer với tên gọi là ***Transformer without Tears***. Về cơ bản, nó giống như mô hình Transformer, chỉ khác nhau ở phần chuẩn hóa. Trong mô hình này, chúng tôi đề xuất một số phương pháp chuẩn hóa đơn giản để cải thiện quá trình huấn luyện Transformer.
Input: một câu văn bất kỳ bằng tiếng Anh hoặc tiếng Việt.
Output**:** một câu văn có nghĩa tương ứng bằng ngôn ngữ còn lại (tiếng Việt hoặc tiếng Anh).

MỤC TIÊU

- Nghiên cứu thuật toán Transformer hiện có và cải thiện hiệu suất của nó trong bài toán dịch máy nhằm tạo ra phiên bản ***Transformer without tears***. Cụ thể, chúng tôi sẽ cải thiện bằng cách thay đổi một số phương pháp chuẩn hóa nhằm tạo ra một phiên bản Transformer mới với độ đo BLEU **cao hơn**, hoặc thời gian huấn luyện **ngắn hơn**, hoặc đạt được cả hai điều trên so với phiên bản Transformer truyền thống.
- Phát triển chương trình ứng dụng minh họa từ phiên bản ***Transformer without tears*** nhưng bước đầu chỉ với hai loại ngôn ngữ chính là tiếng Anh và tiếng Việt.
- Tạo ra một bộ dữ liệu song ngữ Anh-Việt mới.

NỘI DUNG VÀ PHƯƠNG PHÁP

a. NỘI DUNG:

- Nghiên cứu thuật toán Transformer truyền thống trong bài toán dịch máy, huấn luyện mô hình để xác định độ đo BLEU trên tập dữ liệu song ngữ Anh-Việt, IWSLT'15 English-Vietnamese gồm 133,317 câu từ nhóm nghiên cứu Stanford NLP.
- Nghiên cứu và áp dụng các kỹ thuật Postnorm, Prenorm, Scalenorm và Fixnorm vào thuật toán Transformer (đây là các kỹ thuật chính trong việc cải thiện chuẩn hóa nhằm tạo ra phiên bản *Transformer without tears*), so sánh và đánh giá các mô hình ứng với từng trường hợp áp dụng các kỹ thuật nhằm tìm ra mô hình phù hợp nhất và đạt được tối thiểu một trong ba mục tiêu đã đặt ra.
- Tự xây dựng bộ dữ liệu mới BHD-EnVi gồm khoảng một triệu câu tiếng Anh lẫn tiếng Việt,

để mô hình có thể được huấn luyện từ bộ dữ liệu trên nhằm cải thiện hiệu suất và khả năng ứng dụng vào những ngữ cảnh thực tế.

- Nghiên cứu các kỹ thuật tăng cường dữ liệu (Data Augmentation) trong Xử lý ngôn ngữ tự nhiên để hỗ trợ cho việc xây dựng bộ dữ liệu mới.
- Huấn luyện mô hình Transformer without tears (phiên bản Transformer mà chúng tôi tự phát triển) sử dụng hai bộ dữ liệu IWSLT'15 English-Vietnamese và BHD-EnVi để so sánh và đánh giá các kỹ thuật đã sử dụng.
- Xây dựng chương trình ứng dụng minh họa.

b. PHƯƠNG PHÁP:

- Tìm hiểu bản đồ nhận diện cho Transformer dựa theo một số kết quả đã được thực nghiệm trước đó [3], việc sử dụng Prenorm cho cả hai module encoder và decoder ảnh hưởng gì đến tỷ lệ chia của đầu ra? Tìm câu trả lời cho câu hỏi: "Liệu việc thay thế Postnorm bằng Prenorm sẽ hiệu quả hơn trong quá trình huấn luyện so với phương pháp gốc hay không?"
- Tìm hiểu cách khởi tạo trọng số của Postnorm dựa trên ý tưởng từ Glorot và Bengio[4].
- Tìm hiểu về hai kỹ thuật Scalenorm và Fixnorm, áp dụng hai kỹ thuật này vào việc thay thế các tham số scale và shift của Layernorm.
- Tìm hiểu cách đánh giá một mô hình dịch máy, cụ thể ở đây là Transformer và Transformer without tears bằng độ đo Bilingual Evaluation Understudy (BLEU).
- Chúng tôi tạo ra một bộ dữ liệu mới tên là BHD-EnVi bằng cách tổng hợp lại từ các bộ dữ liệu đã được công khai trên mạng, ngoài ra chúng tôi cũng sẽ thu thập thêm dữ liệu mới bằng cách crawl dữ liệu từ các trang web phim song ngữ, sách song ngữ, v.v… kết hợp với các phương pháp tăng cường dữ liệu. Bộ dữ liệu dự kiến sẽ có một triệu câu tiếng Anh lẫn tiếng Việt được chia thành hai tập riêng biệt: tập En với 500,000 câu tiếng Anh, tập Vi với 500,000 câu tiếng Việt được dịch nghĩa và có thứ tự tương ứng với tập EN.
- Huấn luyện mô hình Transformer without tears với từng trường hợp cụ thể (PostLayer, PreLayer, PreFixLayer, PreFixScale) chạy trên bộ dữ liệu IWSLT'15 English-Vietnamese và BHD-EnVi, so sánh và đánh giá kết quả dựa trên độ đo BLEU.
- Xây dựng chương trình ứng dụng trên nền Web cho phép người dùng nhập đầu vào một câu tiếng Việt và xem đầu ra câu tiếng Anh được dịch tương ứng với đầu vào, hoặc ngược lại.

KẾT QUẢ MONG ĐỢI

- Báo cáo các phương pháp và kỹ thuật của mô hình Transformer without tears mà chúng tôi phát triển được sử dụng trong bài toán dịch máy. Kết quả thực nghiệm, đánh giá, so sánh các phương pháp với nhau và với mô hình Transformer truyền thống.
- Tập dữ liệu BHD-EnVi gồm một triệu câu tiếng Anh lẫn tiếng Việt sử dụng cho bài toán.
- Chương trình minh họa dịch máy, tương tự như Google Translate.

KẾ HOẠCH THỰC HIỆN

❖ Tuần 1 – 4: Tìm hiểu mô hình Transformer và các kỹ thuật Postnorm, Prenorm, Scalenorm và Fixnorm, thu thập bộ dữ liệu BHD-EnVi.

➢ Kết quả dự kiến:

- Tài liệu chi tiết cấu trúc mô hình Transformer.
- Tài liệu về các tham số và cơ chế hoạt động của các kỹ thuật Postnorm, Prenorm, Scalenorm và Fixnorm.

- Tìm ra những điểm cần điều chỉnh trong chuẩn hóa, bước đầu tạo ra phiên bản Transformer without tears với các trường hợp áp dụng kỹ thuật khác nhau.
- Tài liệu độ đo BLEU.
- Tập dữ liệu BHD-EnVi chưa hoàn chỉnh bao gồm 500000 câu tiếng Anh lẫn tiếng Việt.

❖ Tuần 5 – 8: Huấn luyện mô hình Transformer truyền thống và phiên bản Transformer without tears trên tập dữ liệu IWSLT'15 English-Vietnamese, ghi chép lại kết quả kèm đánh giá và so sánh. Đồng thời, tiếp tục thu thập phần còn lại của bộ dữ liệu BHD-EnVi.

➢ Kết quả dự kiến:

- Bảng kết quả đánh giá và theo dõi thực nghiệm của cả hai mô hình dựa trên bộ dữ liệu có sẵn IWSLT'15 English-Vietnamese.
- Kết luận lần một về các trường hợp áp dụng kỹ thuật tốt nhất cho phiên bản Transformer without tears.
- Tập dữ liệu BHD-EnVi hoàn chỉnh bao gồm một triệu câu tiếng Anh lẫn tiếng Việt được định dạng thành hai tập như đã nói ở trên.

❖ Tuần 9 – 12: Huấn luyện mô hình Transformer truyền thống và phiên bản Transformer without tears trên tập dữ liệu BHD-EnVi, ghi chép lại kết quả kèm đánh giá và so sánh.

➢ Kết quả dự kiến:

- Bảng kết quả đánh giá và theo dõi thực nghiệm của cả hai mô hình dựa trên bộ dữ liệu tự thu thập BHD-EnVi.
- Kết luận lần hai về các trường hợp áp dụng kỹ thuật tốt nhất cho phiên bản Transformer without tears và đưa ra phiên bản cuối cùng.

❖ Tuần 13 – 16: Xây dựng chương trình demo trên server thật giống như Google Translate.

➢ Kết quả dự kiến:

- Chương trình minh họa.

❖ **Phân công nhiệm vụ:**

Sinh viên thực hiện	**Nhiệm vụ**
Trần Vĩ Hào	• Phụ trách tìm hiểu các kỹ thuật Prenorm, Scalenorm và Fixnorm • Căn cứ vào các bảng số liệu, đánh giá được cung cấp, chọn ra những kỹ thuật phù hợp nhất, từ đó phát triển phiên bản Transformer without tears. • Tổng hợp tư liệu và viết báo cáo cho các kỹ thuật được giao phụ trách. • Crawl dữ liệu để xây dựng bộ dữ liệu BHD-EnVi.
Trương Quốc Bình	• Phụ trách tìm hiểu mô hình Transformer và kỹ thuật Postnorm. • Huấn luyện mô hình và đưa ra các bảng số liệu, đánh giá.

	● Tổng hợp tư liệu và viết báo cáo cho các kỹ thuật được giao phụ trách. ● Crawl dữ liệu để xây dựng bộ dữ liệu BHD-EnVi.
Lê Thành Đạt	● Xây dựng chương trình minh họa. ● Chuẩn bị slide thuyết trình. ● Crawl dữ liệu để xây dựng bộ dữ liệu BHD-EnVi.

TÀI LIỆU THAM KHẢO

[1] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, Wenyu Liu, Bin Xiao: Deep High-Resolution Representation Learning for Visual Recognition. IEEE Trans. Pattern Anal. Mach. Intell. 43(10): 3349-3364 (2021).

[2] Konstantin Sofiiuk, Ilia A. Petrov, Anton Konushin: Reviving Iterative Training with Mask Guidance for Interactive Segmentation. CoRR abs/2102.06583 (2021).

[3] saic-vul, "ritm_interactive_segmentation," 2021. [Online]. Available: https://github.com/saic-vul/ritm_interactive_segmentation.git

[4] Xavier Glorot, Yoshua Bengio: Understanding the difficulty of training deep feedforward neural networks. AISTATS 2010: 249-256.