

ỨNG DỤNG DỊCH MÁY ANH - ĐƯỢC DỰA VÀO MÔ HÌNH TRANSFORMER

Vũ Thị Minh Phương - 220104012

Tóm tắt

- Lớp: **CS2205. APR2023**
- Link Github: <https://github.com/vtmphuong/CS2205.APR2023>
- Link YouTube video:
<https://www.youtube.com/watch?v=DcmCpZr3Mks>



Vũ Thị Minh Phương
220104012

Giới thiệu

Xây dựng hệ thống dịch tiếng Anh sang tiếng Đức dựa mô hình Transformer



Input: dữ liệu ngôn ngữ tiếng Anh

Output: dữ liệu ngôn ngữ tiếng Đức tương ứng

Mục tiêu

- Nghiên cứu để hiểu rõ về cấu trúc và hoạt động của mô hình Transformer, cơ chế Self-attention
- Xây dựng bộ dữ liệu và huấn luyện dữ liệu
- Áp dụng mô hình Transformer để xây dựng được hệ thống dịch giúp tiết kiệm thời gian và cho ra kết quả chính xác cao

Nội dung

- Nghiên cứu về dịch máy và mô hình Transformer, cơ chế Self-attention
- Xây dựng tập dữ liệu huấn luyện từ tiếng Anh sang tiếng Đức
- Huấn luyện và tinh chỉnh mô hình Transformer
- Đánh giá và so sánh hiệu suất của hệ thống dịch máy với các mô hình khác
- Xây dựng chương trình minh họa

Phương pháp

- **Tìm hiểu và hệ thống các kiến thức sau:**
 - Tổng quan về dịch máy, một số các cách tiếp cận dịch máy
 - Kiến trúc mô hình Transformer, bộ mã hóa, bộ giải mã của mô hình
 - Cách thức hoạt động và tầm quan trọng của cơ chế Self-attention
- **Xử lý tập dữ liệu**
 - Khảo sát bộ dữ liệu phù hợp có sẵn tiếng Anh và tiếng Đức hoặc tự xây dựng bộ dữ liệu gồm: Tập huấn luyện, Tập đánh giá, Tập kiểm tra
 - Huấn luyện bộ dữ liệu trên

Phương pháp

- **Xây dựng chương trình thử nghiệm**

- Cài đặt thư viện torchtext, transformer,...
- Viết chương trình sử dụng mô hình transformer bằng code Python
- Chạy thử nghiệm chương trình transformer và mô hình RNN

- **Đánh giá kết quả chạy thử nghiệm**

- Dựa trên số điểm của Bilingual Evaluation Understudy Score (BLEU score)
- Đánh giá và so sánh hiệu suất của Transformer và RNN

Kết quả dự kiến

- Trau dồi thêm kiến thức về dịch máy và Transformer
- Cài đặt và thử nghiệm thành công mô hình Transformer áp dụng cho cặp ngôn ngữ Anh – Đức
- Hiệu suất dịch máy sử dụng mô hình Transformer cao hơn mô hình dịch máy sử dụng RNN
- Có thể ứng dụng được hệ thống dịch cho các cặp ngôn ngữ khác

Tài liệu tham khảo

- [1] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., N. Gomez, A., Kaiser, L., & Polosukhin, I. (2017, June 12). *[1706.03762] Attention Is All You Need*. arXiv. Retrieved May 12, 2023, from <https://arxiv.org/abs/1706.03762>
- [2] Bui, M. Q. (2021, March 29). *Tản mạn về Self Attention*. Viblo. Retrieved May 12, 2023, from <https://viblo.asia/p/tan-man-ve-self-attention-07LKXoq85V4>
- [3] Nguyen, A. V. (2020, May 1). *Transformers - "Người máy biến hình" biến đổi thế giới NLP*. Viblo. Retrieved May 12, 2023, from <https://viblo.asia/p/transformers-nguoi-may-bien-hinh-bien-doi-the-gioi-nlp-924IJPOXKPM>