TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TRẦN TRIỀU HUY - 52000**

**TRẦN THỊ NGỌC ÁNH - 52000008**

**HỆ THỐNG NHẬN DIỆN**

**TIẾNG NÓI DÂN TỘC**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**KỸ THUẬT PHẦN MỀM**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**TRẦN TRIỀU HUY - 52000**

**TRẦN THỊ NGỌC ÁNH - 52000008**

**HỆ THỐNG NHẬN DIỆN**

**TIẾNG NÓI DÂN TỘC**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**KỸ THUẬT PHẦN MỀM**

Người hướng dẫn

**PGS TS. Lê Anh Cường**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Chúng em xin chân thành cảm ơn …………………………………… …………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………………

*TP. Hồ Chí Minh, ngày 01 tháng 01 năm 2024.*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Trần Triều Huy*

*Trần Thị Ngọc Ánh*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của PGS TS. Lê Anh Cường. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 01 tháng 01 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Trần Triều Huy*

*Trần Thị Ngọc Ánh*

**HỆ THỐNG NHẬN DIỆN TIẾNG NÓI DÂN TỘC**

**TÓM TẮT**

(Time New Romans – 13)

**TITLE**

**ABSTRACT**

(Time New Romans – 13)

**MỤC LỤC**

[DANH MỤC HÌNH VẼ v](#_Toc143173056)

[DANH MỤC BẢNG BIỂU vii](#_Toc143173057)

[DANH MỤC CÁC CHỮ VIẾT TẮT viii](#_Toc143173058)

[CHƯƠNG 1. MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI 1](#_Toc143173059)

[1.1 Lý do chọn đề tài 1](#_Toc143173060)

[1.2 Mục tiêu thực hiện đề tài 1](#_Toc143173061)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 2](#_Toc143173062)

[2.1 Mạng neural hồi quy 2](#_Toc143173063)

[2.1.1 Recurrent Neural Network (RNN) 2](#_Toc143173064)

[2.1.2 Long Short-term Memory (LSTM) 2](#_Toc143173065)

[2.2 Mô hình Transformer 3](#_Toc143173066)

[2.2.1 Encoder và Decoder 3](#_Toc143173067)

[2.2.2 Attention 3](#_Toc143173068)

[CHƯƠNG 3. MÔ HÌNH ĐỀ XUẤT 4](#_Toc143173069)

[CHƯƠNG 4. THỰC NGHIỆM 4](#_Toc143173070)

[4.1 Dữ liệu thực nghiệm 4](#_Toc143173071)

[4.2 Cài đặt thực nghiệm 4](#_Toc143173072)

[CHƯƠNG 5. KẾT LUẬN 5](#_Toc143173073)

[5.1 Kết luận 5](#_Toc143173074)

[5.2 Hướng phát triển 5](#_Toc143173075)

[TÀI LIỆU THAM KHẢO 6](#_Toc143173076)

# DANH MỤC HÌNH VẼ

[Hình 2.1: Scaled Dot-Product Attention 3](#_Toc142677545)

# DANH MỤC BẢNG BIỂU

[Bảng 4.1: Thống kê kiểu thực thể trong tập VLSP 2016 4](#_Toc142677565)

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| BERT | Bidirectional Encoder Representations from Transformers |
| GEC | Grammatical Error Correction |
| MLM | Masked Language Model |
| NLP | Natural Language Processing |
| NSP | Next Sentence Prediction |

# MỞ ĐẦU VÀ TỔNG QUAN ĐỀ TÀI

## Lý do chọn đề tài

Dân tộc M’Nông là một dân tộc thiểu số ở Việt Nam, với dân số khoảng 127.334 người (theo Tổng điều tra dân số và nhà ở năm 2019), cư trú tại 51 trên tổng số 63 tỉnh, thành phố. Tiếng M’Nông là một ngôn ngữ thuộc ngữ tộc Môn-Khmer của ngữ hệ Nam Á. Tiếng M’Nông có một hệ thống ngữ âm phức tạp, với nhiều thanh điệu và phụ âm.

Hệ thống nhận diện tiếng nói dân tộc M’Nông là một hệ thống có khả năng nhận diện và phân loại tiếng nói của người M’Nông thông qua ứng dụng công nghệ thông tin. Chúng tôi hy vọng thông qua nghiên cứu ngôn ngữ, bao gồm các ngữ âm, ngữ pháp và từ vựng, hệ thống có thể được sử dụng để hỗ trợ việc học tiếng M’Nông cho người dân tộc M’Nông và cả người Kinh, cũng như ghi lại và lưu trữ tiếng nói của người M’Nông, góp phần bảo tồn nét văn hóa truyền thống của dân tộc này.

## Mục tiêu thực hiện đề tài

Mục tiêu khi chúng tôi thực hiện đề tài là xây dựng một hệ thống nhận diện tiếng nói dân tộc M’Nông có độ chính xác cao, có thể áp dụng vào thực tế và đáp ứng được nhiều yêu cầu của nhiều mục đích sử dụng khác nhau.

# CƠ SỞ LÝ THUYẾT

## Mạng neural hồi quy

### Recurrent Neural Network (RNN)

Về mặt toán học, mạng RNN có thể được mô tả như trong công thức (2.1) và (2.2). Công thứ (2.1), là vectơ đầu vào tại bước thứ . là trạng thái ẩn tại bước thứ . là một hàm phi tuyến tính (nonlinear function), thường là hàm tanh hay ReLu. là ma trận trọng số cho trạng thái ẩn ở bước trước đó , là ma trận trọng số cho đầu vào. Như vậy có thể thấy rằng, trạng thái ẩn ở bước thứ được tính dựa trên trạng thái ẩn ở bước trước đó và dữ liệu đầu vào ở bước hiện tại.

|  |  |  |
| --- | --- | --- |
|  |  | (2.1) |
|  |  | (2.2) |

Trong công thức (2.2), là ma trận trọng số cho đầu ra của mạng, là một phân phối xác suất trên từ điển tại bước thứ . Trạng thái ẩn được xem như là bộ nhớ của RNN, nó lưu trữ thông tin tính toàn được thực hiện ở các bước trước đó. Không giống như các mạng nơ-ron truyền thống, RNN chia sẽ chung bộ trọng số [, , ] cho tất cả các bước, nghĩa là các phép tính toán sẽ được thực hiện tương tự nhau cho tất cả các bước lặp chỉ các dữ liệu đầu vào. Đây là một ưu điểm của RNN giúp giảm số lượng tham số cần học cho mô hình.

### Long Short-term Memory (LSTM)

Mạng RNN chia sẽ chung một bộ trọng số giữa các bước lặp nên giảm đáng kể số lượng tham số, tuy nhiên nó vẫn là một mạng rất sâu. Trong quá trình lan truyền ngược (backward), phải thực hiện nhiều bước để có tính được đạo hàm cho những đầu vào đầu tiên của một chuỗi dữ liệu dài, do đó kết quả đạo hàm thường lớn hơn hoặc nhỏ hơn 1 đáng kể, dẫn đến giá trị của đạo hàm thường bị bùng nổ (exploding) hoặc mất mát (vanishing). Long Short-term Memory là một biến thể cải tiếng của RNN được đề xuất bởi (Hochreiter & Schmidhuber, 1997) nhằm khắc phục hai vấn đề trên.

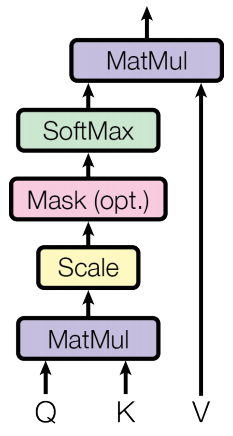
## Mô hình Transformer

### Encoder và Decoder

### Attention

Attention của mô hình Transformer bao gồm 02 kỹ thuật là Scaled Dot-Product Attention và Multi-head Attention.

#### Scaled Dot-Product Attention



Hình 2.1: Scaled Dot-Product Attention

(Nguồn: (Vaswani et al., 2023))

Sẽ có 03 ma trận , và được sử dụng trong kỹ thuật attention này, tương ứng với các khái niệm query, key và value. Các dòng trong trong ma trận và sẽ có kích thước , các dòng trong ma trận sẽ có kích thước . Các ma trận này được tạo ra từ việc cho vector embedding đầu vào đi qua 03 bộ trọng số , và .

#### Multi-head Attention

# MÔ HÌNH ĐỀ XUẤT

…

# THỰC NGHIỆM

## Dữ liệu thực nghiệm

Dữ liệu trong VLSP-2016 đã được chia sẵn thành hai tập huấn luyện và kiểm thử. Thống kê chi tiết của các kiểu thực thể và các tập dữ liệu được mô tả trong Bảng 4.1**.**

Bảng 4.1: Thống kê kiểu thực thể trong tập VLSP 2016

|  |  |  |
| --- | --- | --- |
| **Label** | **Train** | **Test** |
| **LOC** | 1210 | 1377 |
| **ORG** | 7478 | 274 |
| **PER** | 6230 | 1294 |
| **MISC** | 250 | 47 |
| **Total** | 15168 | 2992 |

## Cài đặt thực nghiệm

…

# KẾT LUẬN

## Kết luận

## Hướng phát triển

# TÀI LIỆU THAM KHẢO

Tiếng Việt

…

Tiếng Anh

Hochreiter, S., & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, *9*, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2023). *Attention Is All You Need* (arXiv:1706.03762). arXiv. https://doi.org/10.48550/arXiv.1706.03762