VIETNAM GENERAL CONFEDERATION OF LABOUR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FULL NAME - MSSV**

**TITLE**

**MIDTERM REPORT**

**INTRODUCTION TO**

**MACHINE LEARNING**

**HO CHI MINH CITY, YEAR 2023**

VIETNAM GENERAL CONFEDERATION OF LABOUR

**TON DUC THANG UNIVERSITY**

**FACULTY OF INFORMATION TECHNOLOGY**



**FULL NAME - ID**

**TITLE**

**MIDTERM REPORT**

**INTRODUCTION TO MACHINE LEARNING**

Advised by

**Assoc. Prof., Dr., Mr. Nguyen Van A**

**HO CHI MINH CITY, YEAR 2023**

**ACKNOWLEDGMENT**

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

*Ho Chi Minh City, day month year 20..*

*Author*

*(Signature and full name)*

**DECLARATION OF AUTHORSHIP**

I hereby declare that this thesis was carried out by myself under the guidance and supervision of ………………………………………..; and that the work and the results contained in it are original and have not been submitted anywhere for any previous purposes. The data and figures presented in this thesis are for analysis, comments, and evaluations from various resources by my own work and have been duly acknowledged in the reference part.

In addition, other comments, reviews and data used by other authors, and organizations have been acknowledged, and explicitly cited.

**I will take full responsibility for any fraud detected in my thesis**. Ton Duc Thang University is unrelated to any copyright infringement caused on my work (if any).

*Ho Chi Minh City, day month year 20..*

*Author*

*(Signature and full name)*

**TITLE**

**ABSTRACT**

(Time New Romans – 13)

**CONTENTS**

[LIST OF FIGURES vi](#_Toc143981663)

[LIST OF TABLES vii](#_Toc143981664)

[ABBREVIATIONS viii](#_Toc143981665)

[CHAPTER 1. NAME OF CHAPTER 1 1](#_Toc143981666)

[1.1 Section 1 of chap 1 1](#_Toc143981667)

[1.2 Section 2 of chap 1 1](#_Toc143981668)

[CHAPTER 2. NAME OF CHAPTER 2 2](#_Toc143981669)

[2.1 Name of section 1 chap 2 2](#_Toc143981670)

[2.2 Recurrent Neural Network (RNN) 2](#_Toc143981671)

[2.2.1 Long Short-term Memory (LSTM) 2](#_Toc143981672)

[2.3 Mô hình Transformer 3](#_Toc143981673)

[2.3.1 Encoder và Decoder 3](#_Toc143981674)

[2.3.2 Attention 3](#_Toc143981675)

[2.3.2.1 Scaled Dot-Product Attention 3](#_Toc143981676)

[2.3.2.2 Multi-head Attention 3](#_Toc143981677)

[CHAPTER 3. MÔ HÌNH ĐỀ XUẤT 4](#_Toc143981678)

[CHAPTER 4. THỰC NGHIỆM 4](#_Toc143981679)

[4.1 Dữ liệu thực nghiệm 4](#_Toc143981680)

[4.2 Cài đặt thực nghiệm 4](#_Toc143981681)

[CHAPTER 5. KẾT LUẬN 5](#_Toc143981682)

[5.1 Kết luận 5](#_Toc143981683)

[5.2 Hướng phát triển 5](#_Toc143981684)

[REFERENCES 6](#_Toc143981685)

# LIST OF FIGURES

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

# LIST OF TABLES

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

# ABBREVIATIONS

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

# NAME OF CHAPTER 1

## Section 1 of chap 1

Một cách tổng quát, thực thể (entity) được định nghĩa là một đối tượng được xác định rõ ràng trong một điều kiện, có tính định danh và tách biệt với các thực thể khác. Trong ngôn ngữ, thực thể là một phần quan trọng cấu thành nên văn bản, được thể hiện bằng các danh từ, cụm danh từ hoặc danh từ riêng. Bên cạnh đó, mỗi thực thể có thể được xếp vào một kiểu (type/category) tương ứng với ngữ nghĩa của nó. Thực thể và kiểu thực thể có vai trò quan trọng, giúp cung cấp các thông tin chi tiết về các đối tượng được đề cập và cho phép phân tích hiệu quả các văn bản phi cấu trúc thành các dạng dữ liệu có cấu trúc. Việc nhận dạng các thực thể và định kiểu cho chúng như trên được gọi là nhận dạng thực thể (entity recognition), và một bài toán cụ thể hơn cho các thực thể đã được định danh gọi là nhận dạng thực thể có tên (named entity recognition).

## Section 2 of chap 1

…

# NAME OF CHAPTER 2

## Name of section 2.1

Example content

## 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. Coong thuc (2.2)

## Mô hình Transformer

### Encoder và Decoder

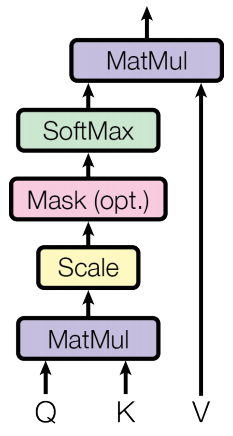
Mot cong thuc tinh tong 2 so

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

### 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

# REFERENCES

Vietnamese

…

English

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