TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO GIỮA KÌ NHẬP MÔN HỌC MÁY**

**ĐÁNH GIÁ HIỆU QUẢ CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY VÀ CONTINUAL LEARNING VÀ TEST PRODUCTION TRONG GIẢI PHÁP HỌC MÁY.**

*Người hướng dẫn*: **Thầy Lê Anh Cường**

*Người thực hiện*: **HUỲNH PHẠM TỐ NGÂN - 519H0200**

Lớp **: 19H50201**

Khóa  **: 23**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2022**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO GIỮA KÌ NHẬP MÔN HỌC MÁY**

**ĐÁNH GIÁ HIỆU QUẢ CÁC PHƯƠNG PHÁP OPTIMIZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY VÀ CONTINUAL LEARNING VÀ TEST PRODUCTION TRONG GIẢI PHÁP HỌC MÁY.**

*Người hướng dẫn*: **Thầy Lê Anh Cường**

*Người thực hiện*: **HUỲNH PHẠM TỐ NGÂN - 519H0200**

Lớp **: 19H50201**

Khóa  **: 23**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2022**

LỜI CẢM ƠN

Em xin cảm ơn thầy Lê Anh Cường– giảng viên môn “Nhập môn học máy” đã hướng dẫn, chỉ bảo để chúng em hoàn thành bài báo cáo cuối kỳ này. Chúng em mong nhận được ý kiến của thầy để cải thiện các kỹ năng về phân tích dữ liệu.

ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG

Em xin cam đoan đây là sản phẩm đồ án của riêng chúng em và được sự hướng dẫn của thầy 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 bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong đó đồ 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 em xin hoàn toàn chịu trách nhiệm về nội dung đồ án của mình.** Trường đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do chúng em gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 05 tháng 12 năm 2022*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

Ngân

*Huỳnh Phạm Tố Ngân*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của giáo viên hướng dẫn**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*TP. Hồ Chí MInh, 05 – 12 - 2022*

*(chữ kí và ghi họ tên)*

**Phần đánh giá của giáo viên chấm bài**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*TP. Hồ Chí MInh, 05 - 12- 2022*

*(chữ kí và ghi họ tên)*

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc154258291)

[ĐỒ ÁN ĐƯỢC HOÀN THÀNH TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG ii](#_Toc154258292)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc154258293)

[TABLE OF CONTENTS iv](#_Toc154258294)

[I. TÌM HIỂU, SO SÁNH CÁC PHƯƠNG PHÁP OPTIMER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY 1](#_Toc154258295)

[1. Giới thiệu 1](#_Toc154258296)

[2. So sánh 1](#_Toc154258297)

[II. TÌM HIỂU VỀ CONTINUAL LEARNING VÀ TEST PRODUCTION KHI XÂY DỰNG MỘT GIẢI PHÁP HỌC MÁY ĐỂ GIẢI QUYẾT MỘT BÀI TOÁN 3](#_Toc154258298)

[1. Giới thiệu 3](#_Toc154258299)

[2. So sánh 3](#_Toc154258300)

[3. Kết luận 4](#_Toc154258301)

[3.1 Sự quan trọng của Continual Learning 4](#_Toc154258302)

[3.2 Test Production đóng vai trò quyết định 4](#_Toc154258303)

I. TÌM HIỂU, SO SÁNH CÁC PHƯƠNG PHÁP OPTIMZER TRONG HUẤN LUYỆN MÔ HÌNH HỌC MÁY

1. Giới thiệu

Bài nghiên cứu này tập trung vào việc đánh giá, so sánh và tìm hiểu sâu rộng về các phương pháp optimizer phổ biến được sử dụng trong huấn luyện mô hình học máy. Mục tiêu là xác định ưu điểm, nhược điểm và ảnh hưởng của mỗi optimizer đối với quá trình học và hiệu suất của mô hình.

2. So sánh

|  |  |  |  |
| --- | --- | --- | --- |
| Phương pháp optimizer | Ý tưởng | Ưu điểm | Nhược điểm |
| **Stochastic Gradient Descent (SGD)** | Dựa trên việc cập nhật trọng số của mô hình bằng cách di chuyển ngược hướng theo đạo hàm của hàm mất mát. | Dễ triển khai, hoạt động tốt cho tập dữ liệu lớn. | Có thể bị nhanh chóng rơi vào các cực tiểu cục bộ. |
| **Adagrad** | Tối ưu hóa tốc độ học cho từng tham số dựa trên tần suất xuất hiện của nó. | Hiệu quả với dữ liệu thưa và biến thiên. | Có thể dẫn đến việc giảm tốc độ học quá nhanh. |
| **RMSprop** | Cải thiện Adagrad bằng cách giảm tác động của tuần suất xuất hiện của các tham số cũ trong quá trình cập nhật. | Giải quyết vấn đề giảm tốc độ học quá mạnh của Adagrad. | Có thể cần thay đổi siêu tham số. |
| **Adam (Adaptive Moment Estimation)** | Kết hợp ý tưởng của RMSSprop và Momentum. Sử dụng động moment để điều chỉnh tốc độ học của từng tham số. | Hiệu quả trên nhiều loại dữ liệu và kiến trúc mô hình. | Có thể nhạy cảm với siêu tham số. |
| **Adadelta** | Cải thiện RMSprop bằng cách giảm tốc độ chia của tần suất xuất hiện của các tham số. | Không cần siêu tham số như RMSrpop. | Hiệu suất có thể thấp trên một số tác vụ. |
| **Nadam** | Kết hợp Adam và Nesterov Momentum. | Cải thiện hiệu suất so với Adam trên một số tác vụ. | Cần thêm lớp tinh chỉnh |

Những phương pháp này đều có những ưu điểm và nhược điểm riêng, và sự lựa chọn giữa chúng thường phụ thuộc vào đặc tính của dữ liệu và kiến trúc mô hình.

II. TÌM HIỂU VỀ CONTINUAL LEARNING VÀ TEST PRODUCTION KHI XÂY DỰNG MỘT GIẢI PHÁP HỌC MÁY ĐỂ GIẢI QUYẾT MỘT BÀI TOÁN

1. Giới thiệu

Trong thời đại công nghệ 4.0, học máy đóng vai trò quan trọng trong việc giải quyết nhiều bài toán phức tạp. Để đạt được hiệu suất cao và đáng tin cậy của mô hình học máy, việc áp dụng các kỹ thuật tiên tiến như ***Continual Learning*** và ***Test Production*** trở nên ngày càng quan trọng.

2. So sánh

|  |  |  |
| --- | --- | --- |
|  | **Continual Learning** | **Test Production** |
| Ý tưởng | Là một khía cạnh của học máy mà mô hình được đào tạo liên tục trên một luồng dữ liệu không ngừng thay đổi theo thời gian. Mục tiêu là duy trì và mở rộng kiến thức của mô hình trên tập dữ liệu mới mà không quên kiến thức đã học trước đó. | Là quá trình đảm bảo rằng mô hình học máy không chỉ hoạt động tốt trên dữ liệu đào tạo mà còn trên dữ liệu mới từ thực tế. Nó liên quan đến việc kiểm tra mô hình trên các dữ liệu kiểm tra và đảm bảo rằng hiệu suất của mô hình là ổn định và đáng tin cậy khi triển khai trong môi trường thực tế. |
| Thách thức | Hiện tượng quên (catastrophic forgetting) là một vấn đề phổ biếm trong CL, nơi mô hình quên thông tin từ các tập dữ liệu trước đó khi học trên dữ liệu mới. | Môi trường thực tế có thể thay đổi và phức tạp, và dữ liệu mới có thể có đặc tính mà mô hình chưa từng gặp. |
| Giải pháp | Các kỹ thuật như Elastic Weight Consolidation (EWC), Learning without Forgetting (LwF), và Progressive Neural Networks (PNN) đã được đề xuất để giải quyết vấn đề này. | Sử dụng kỹ thuật Cross – validation để đánh giá hiệu suất trên các tập kiểm tra độc lập. Tăng cường dữ liệu và kiểm soát chất lượng dữ liệu để đảm bảo mô hình được đào tạo trên một biểu diễn tốt của dữ liệu thực tế. |

3. Kết luận

3.1 Sự quan trọng của Continual Learning

- Đóng vai trò quan trọng trong việc giữ cho mô hình học máy linh hoạt và đào tạo liên tục từ dữ liệu mới.

- Chiến lược học liên tục giúp giải quyết hiệu quả vấn đề mất kiến thức khi đối mặt với dữ liệu đa dạng

3.2 Test Production đóng vai trò quyết định

- Quá trình Test Production chính là cột mốc quyết định độ tin cậy của mô hình học máy khi triển khai trong môi trường thực tế.