TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**DỰ ÁN CUỐI KÌ MÔN NHẬP MÔN HỌC MÁY**

*Người hướng dẫn*: **Lê Anh Cường**

*Người thực hiện*: **Vũ Thị Hà Vy – 52100742**

Nhóm **: 02**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**DỰ ÁN CUỐI KÌ MÔN NHẬP MÔN HỌC MÁY**

Người hướng dẫn: **Lê Anh Cường**

Người thực hiện: **Vũ Thị Hà Vy – 52100742**

Nhóm **: 02**

Khoá  **: 25**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2023**

LỜI CẢM ƠN

Đầu tiên nhóm em xin gửi lời cảm ơn chân thành đến trường Đại học Tôn Đức Thắng, đặc biệt là khoa Công Nghệ Thông Tin đã tạo những điều kiện tốt nhất để chúng em có thể hoàn thành môn nhập môn học máy. Tiếp đến, chúng em xin bày tỏ lòng biết ơn sâu sắc đến thầy Lê Anh Cường đã tận tình truyền đạt và hướng dẫn chúng em trong suốt quá trình học tập và làm báo cáo. Chính vì vậy những lý thuyết đã được truyền đạt lại một cách dễ hiểu và có thể áp dụng trong quá trình nhận thức và hành động bản thân. Em đã có thể tiếp cận với kiến thức xã hội, quá trình nhận thức, tư duy trừu tượng kết hợp đó là nhận thức đúng đắn về sự việc, hiện tượng trong xã hội.

Trong quá trình làm bài báo cáo, em đã cố gắng nỗ lực, tuy nhiên khó tránh khỏi những sai sót bởi trình độ lý luận và kinh nghiệm thực tiễn còn hạn chế. Em rất mong nhận được sự góp ý từ thầy cô để có thêm kinh nghiệm và có thể làm tốt hơn ở những bài báo cáo tới.

Cuối cùng, em xin cảm ơn và kính chúc thầy thật nhiều sức khỏe và niềm vui trong cuộc sống và thành công trong sự nghiệp giảng dạy của mình.

**ĐỒ ÁN ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là sản phẩm đồ án của riêng chúng tôi và được sự hướng dẫn của 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 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 tôi 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 tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 16 tháng 04 năm 2023*

*Tác giả*

*(ký tên và ghi rõ họ tên)*

*Vy*

*Vũ Thị Hà Vy*

PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN

**Phần xác nhận của GV hướng dẫn**

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

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

**Phần đánh giá của GV chấm bài**

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

Tp. Hồ Chí Minh, ngày tháng năm

(kí và ghi họ tên)

MỤC LỤC

[LỜI CẢM ƠN i](#_Toc154257928)

[PHẦN XÁC NHẬN VÀ ĐÁNH GIÁ CỦA GIẢNG VIÊN iii](#_Toc154257929)

[MỤC LỤC 1](#_Toc154257930)

[CHƯƠNG 1 – NGHIÊN CỨU 2](#_Toc154257931)

[1.1 Mục tiêu và ý nghĩa của nghiên cứu: 2](#_Toc154257932)

[1.2 Tìm hiểu và so sánh các phương pháp Optimizer 2](#_Toc154257933)

[1.3 Continual Learning và Test Production 3](#_Toc154257934)

[CHƯƠNG 2 – ĐÁNH GIÁ 4](#_Toc154257935)

[2.1. Kết quả tìm hiểu và so sánh Optimizer: 4](#_Toc154257936)

[2.2. Continual Learning và Test Production trong thực tế 4](#_Toc154257937)

[2.2.1. Continual Learning 4](#_Toc154257938)

[2.2.2. Test Production 5](#_Toc154257939)

[CHƯƠNG 3 – KẾT LUẬN 5](#_Toc154257940)

[3.1. Tổng kết nghiên cứu: 5](#_Toc154257941)

[3.2. Hướng phát triển: 6](#_Toc154257942)

[3.3. Kết luận chung: 6](#_Toc154257943)

CHƯƠNG 1 – NGHIÊN CỨU

* 1. Mục tiêu và ý nghĩa của nghiên cứu:

Bài nghiên cứu sẽ giúp ta hiểu biết được ưu và nhược điểm của từng phương pháp optimizer. Từ đó lựa chọn đúng phương pháp cho phù hợp với từng bài toán cụ thể. Về Continual Learning sẽ đưa ra các thách thức và khi tìm hiểu mình sẽ có được các giải pháp cho thách thức gặp phải trong Continual Learning. Tóm lại, bài nghiên cứu này rất quan trọng cho chúng ta trong lĩnh vực học máy.

* 1. Tìm hiểu và so sánh các phương pháp Optimizer

Optimizer là thuật toán tối ưu được sử dụng để tìm giá trị tối ưu của tham số mô hình trong quá trình huấn luyện.

Các phương pháp Optimizer phổ biến bao gồm:

* Gradient descent (GD): Đây là phương pháp tối ưu nhất trong các phương pháp optimizer. GD cập nhật tham số mô hình theo hướng gradient của hàm mất mát. Gradient là vector đạo hàm riêng của hàm mất mát theo từng tham số, và nó cho biết hướng tăng nhanh nhất của hàm mất mát. Gradient Descent giảm dần giá trị của hàm mất mát bằng cách đi theo hướng ngược lại của gradient và điều chỉnh tốc độ cập nhật thông qua một tham số gọi là learning rate (tốc độ học).
* Stochastic gradient descent (SGD): Phiên bản cải tiến của GD bởi thay vì tính gradient trên tòa bộ tập dữ liệu , SGD chỉ tính gradient trên một batch dữ liệu ngẫu nhiên. Từ đó giúp giảm được thời gian tính toán và cập nhật tham số mô hình nhanh hơn
* Momentum: đây là phương pháp mở rộng của Gradient Descent bằng cách kết hợp thông tin của Gradient để cập nhật tham số mô hình và nó sử dụng tham số momemtum giúp quản lí mức độ và giảm độ dao động và tăng tốc độ hội tụ của quá trình tối ưu hóa.
* Adagrad: Phương pháp điều chỉnh tốc độ học (learning rate) theo độ dốc của hàm mất mát. Tức là Adagrad sẽ điều chỉnh tốc độ học cho từng tham số riêng biệt dựa trên lịch sử của gradient đã tính toán cho tham số đó.
* RMSprop: Phương pháp này sử dụng trung bình lũy thừa của độ dốc để điều chỉnh tốc độ học. Tuy nhiên, RMSprop khác với Adagrad bằng cách sử dụng trung bình lũy thừa chứ không dùng tổng lũy thừa của các bình phương gradient. Cũng vì vậy mà giúp cho RMSprop ổn định hơn, giảm thiểu sự giảm dần quá nhanh của tốc độ học.
* Adam: Đây là phương pháp được kết hợp các kỹ thuật của Momentum và RMSprop. Nó sử dụng thông tin về gradient và trung bình lũy thừa của gradient để cập nhật tham số mô hình. Nó cung cấp một cân bằng giữa việc duy trì động lượng của quá trình tối ưu hóa và điều chỉnh tốc độ học dựa trên thông tin gradient.
  1. Continual Learning và Test Production

Continual Learning (Học liên tục) là khi một mô hình học máy không chỉ học một lần và xong, mà nó còn tiếp tục học thêm khi có dữ liệu mới. Nghĩa là, mô hình không chỉ biết sử dụng kiến thức từ quá khứ mà còn có khả năng học thêm từ dữ liệu mới đến. Điều này giúp mô hình nắm bắt được những thay đổi trong dữ liệu và cải thiện hiệu suất của nó theo thời gian.

Test Production (Sản xuất kiểm tra) là khi chúng ta tạo ra các bộ kiểm tra để kiểm tra hiệu suất của mô hình học máy. Điều này giúp chúng ta biết được mô hình làm việc như thế nào trên các tập dữ liệu kiểm tra độc lập và đo lường độ chính xác của nó. Như vậy, chúng ta có thể đảm bảo rằng mô hình hoạt động tốt và đáng tin cậy trên các dữ liệu mà nó chưa từng gặp trước đó.

Vì vậy, Continual Learning giúp mô hình học máy tiếp tục học và cải thiện từ dữ liệu mới, trong khi Test Production giúp kiểm tra hiệu suất của mô hình trên các tập dữ liệu kiểm tra độc lập. Cả hai yếu tố này đều quan trọng trong việc phát triển và sử dụng mô hình học máy để đảm bảo rằng nó có thể thích ứng với dữ liệu mới và hoạt động tốt trên các tập dữ liệu kiểm tra.

CHƯƠNG 2 – ĐÁNH GIÁ

1. Kết quả tìm hiểu và so sánh Optimizer:

Optimizer để dẫn đến sự hội tụ nhanh chóng đến giá trị tối ưu. Optimizer sẽ giúp mô hình đạt được giá trị tối ưu trong thời gian ngắn. Ngoài ra, khả năng vượt qua điểm tối ưu cục bộ và tiếp tục tìm kiếm giá trị tối ưu khác của Optimizer. Phương pháp này nên đảm bảo sự ổn định trong quá trình tối ưu hóa. Điều này đồng nghĩa với việc các bước cập nhật tham số không gây ra sự dao động lớn và không ổn định trong quá trình huấn luyện.

1. Continual Learning và Test Production trong thực tế
2. Continual Learning

* Mô hình có khả năng tích hợp kiến thức đã học trước đó và không bị quên khi được huấn luyện với dữ liệu mới.Tính ứng dụng trong các bài toán thực tế:
* Các hệ thống nhận dạng khuôn mặt, vật thể, và đối tượng di động, nơi dữ liệu mới liên tục được thêm vào.
* Giảm thiểu ảnh hưởng của việc thêm dữ liệu mới lên hiệu suất của mô hình, đặc biệt là khi dữ liệu mới có thể có tính chất khác biệt.

1. Test Production

* Mô hình được kiểm thử trong môi trường mô phỏng thực tế với các điều kiện và biến động tương tự.
* Đảm bảo rằng mô hình được kiểm thử trong môi trường mô phỏng thực tế với các điều kiện và biến động tương tự.
* Thực hiện các kịch bản kiểm thử an toàn để đảm bảo rằng mô hình không gây ra rủi ro cho hệ thống và dữ liệu.
* Tích hợp các quy trình cập nhật liên tục để triển khai các phiên bản mới của mô hình và cải tiến hiệu suất.

CHƯƠNG 3 – KẾT LUẬN

1. Tổng kết nghiên cứu:

Về phương pháp Optimizer trong huấn luyện mô hình học máy đã đưa ra cái nhìn tổng quan về sự ảnh hưởng của chúng đối với hiệu suất và tốc độ hội tụ của mô hình. Kết quả so sánh giữa các Optimizer như Gradient Descent, SGD, Adam, và RMSprop đã được trình bày chi tiết ở trên.

Bên cạnh đó, về Continual Learning và Test Production đã cho em cái nhìn sâu sắc về cách mô hình có thể liên tục học từ dữ liệu mới mà không quên kiến thức cũ. Đồng thời, quan tâm đặc biệt đến quá trình kiểm thử và đánh giá hiệu suất của mô hình trong môi trường sản xuất thực tế. Cả hai khía cạnh này đều quan trọng để đảm bảo ổn định và hiệu quả của mô hình trong thời gian dài.

1. Hướng phát triển:

* Tối Ưu Hóa Optimizer: tối ưu hóa siêu tham số của các phương pháp Optimizer để tối ưu hóa hiệu suất cho từng loại mô hình và bài toán cụ thể
* Áp dụng các kỹ thuật mới nhất của Continual Learning để giải quyết các thách thức cụ thể trong các ứng dụng thực tế.

1. Kết luận chung:

Qua nghiên cứu trên đã giúp em hiểu rõ hơn về tầm quan trọng của việc lựa chọn Optimizer và triển khai Continual Learning trong quá trình xây dựng giải pháp học máy. Các kiến thức tìm hiểu trên đều áp dụng rộng rãi trong nhiều lĩnh vực, từ y tế đến tài chính và công nghiệp. Hướng phát triển và nghiên cứu tiếp theo sẽ đóng vai trò quan trọng trong việc nâng cao hiệu suất và khả năng linh hoạt của các mô hình học máy trong thực tế