



ĐỀ CƯƠNG KHOÁ LUẬN TỐT NGHIỆP

Huấn luyện mạng nơ-ron nhiều tầng ẩn bằng thuật toán Adam

(Training deep neural networks using Adam optimizer)

1 THÔNG TIN CHUNG

Người hướng dẫn:

– ThS. Trần Trung Kiên (Khoa Công nghệ Thông tin)

Nhóm Sinh viên thực hiện:

1. Nguyễn Ngọc Lan Như (MSSV: 1712644)

2. Hoàng Minh Quân (MSSV: 1712688)

Loại đề tài: Nghiên cứu

Thời gian thực hiện: Từ 01/2021 đến 06/2021

2 NỘI DUNG THỰC HIỆN

2.1 Giới thiệu về đề tài

Trong những năm gần đây, trí tuệ nhân tạo ngày càng nhận được nhiều sự quan tâm vì tính ứng dụng cao của nó trong thực tế. Trong đó, học máy (machine learning) là một nhánh dựa trên khả năng tự động rút trích đặc trưng và đưa ra quyết định từ dữ liệu của máy tính. Một hướng tiếp cận trong học máy đã và

đang đạt được những thành tựu vượt trội và được ứng dụng rộng rãi để giải quyết những bài toán khó khăn như xử lý ngôn ngữ tự nhiên, xử lý ảnh số và video,... là phương pháp sử dụng mạng nơ-ron nhiều tầng ẩn, hay còn gọi là mạng học sâu (deep neural network). Những mạng này bao gồm tầng nhập, các tầng ẩn và tầng xuất. Việc sử dụng mạng nơ-ron nhiều tầng ẩn tăng khả năng rút trích đặc trưng trong nhiều loại dữ liệu kể cả những loại dữ liệu không cấu trúc và không gán nhãn. Bên cạnh đó, những mạng nơ-ron này có thể học, hay được huấn luyện, để có thể tự động đưa ra những dự đoán và rút kinh nghiệm để thích ứng với dữ liệu mới. Tuy nhiên, việc huấn luyện một mạng nơ-ron nhiều tầng ẩn có thể gặp nhiều khó khăn nên hiệu suất trong thực tế thấp dẫn đến không có ý nghĩa thực tiễn. Vì thế việc khắc phục những khó khăn này là cần thiết.

Huấn luyện mạng nơ-ron nhiều tầng ẩn chủ yếu là quá trình đi tìm cực tiểu toàn cục của hàm mục tiêu thông qua việc điều chỉnh các trọng số của mạng nơ-ron nhiều tầng ẩn sao cho giá trị của hàm mục tiêu là nhỏ nhất. Hàm mục tiêu, hay còn được gọi cụ thể hơn là hàm chi phí, là hàm thể hiện độ sai sót của mạng nơ-ron nhiều tầng ẩn trên tập dữ liệu huấn luyện. Hàm chi phí có giá trị nhỏ đồng nghĩa với việc mạng nơ-ron nhiều tầng ẩn hiệu quả trên tập huấn luyện, và kỳ vọng hoạt động tốt trên dữ liệu thực. Quá trình đi tìm giá trị nhỏ nhất của hàm chi phí còn được gọi là tối ưu hóa hàm chi phí. Trong lĩnh vực học máy, việc tối ưu hóa bộ trọng số cho mạng nơ-ron nhiều tầng ẩn là một vấn đề hiện hữu trong các kiến trúc mạng nơ-ron nhiều tầng ẩn hiện nay. Bài toán được phát biểu như sau:

- Input: Hàm chi phí với các tham số là bộ trọng số của mạng nơ-ron nhiều tầng ẩn.
- Output: Bộ trọng số của mạng nơ-ron nhiều tầng ẩn đã được tối ưu sao cho giá trị của hàm chi phí là nhỏ nhất.

Giải quyết được bài toán giúp giảm đáng kể thời gian huấn luyện và tăng độ chính xác trên dữ liệu thực của mạng nơ-ron nhiều tầng ẩn. Tuy nhiên, đi tìm lời giải của bài toán gặp nhiều khó khăn. Thứ nhất, vì hàm mục tiêu có thể có nhiều

cực tiểu địa phương, điểm yên ngựa, plateau,... có thể gây cản trở quá trình tối ưu hóa. Ngoài ra, việc điều chỉnh các siêu tham số như tỷ lệ học đòi hỏi phải thử nghiệm nhiều lần và đặc trưng cho từng tập dữ liệu. Thêm nữa, khi sử dụng các hướng tiếp cận như gradient descent, stochastic gradient descent sẽ gặp khó khăn trên những tập dữ liệu không đủ thông tin để cập nhật các tham số, khi sử dụng một tỷ lệ học cố định.

Trong thời gian gần đây, một hướng tiếp cận đạt được kết quả tốt trong việc huấn luyện mạng nơ-ron nhiều tầng ẩn là sử dụng tỷ lệ học tự điều chỉnh tương ứng với từng trọng số trong quá trình huấn luyện. Đây là hướng mà chúng em sẽ tập trung tìm hiểu.

2.2 Mục tiêu đề tài

- Hiểu rõ các khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Hiểu rõ cách thức hoạt động cũng như ưu, nhược điểm của thuật toán Adam.
- Hiểu sâu về nguyên lý hoạt động, ưu và nhược điểm của các thuật toán có tỷ lệ học thay đổi trong quá trình huấn luyện.
- Cài đặt lại thuật toán và tái tạo lại các kết quả trong bài báo tương ứng, tiến hành thêm các thí nghiệm so sánh giữa các phương pháp khác để thấy rõ hơn về ưu/nhược điểm của thuật toán Adam trong việc giải quyết những khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Nếu còn thời gian sau khi đã nắm rõ các nguyên lý hoạt động thì có thể xem xét các cải tiến có thể có (chẳng hạn như sử dụng lập trình song song để tăng tốc độ xử lý)
- Rèn luyện các kỹ năng mềm như: suy nghĩ logic, lên kế hoạch, làm việc nhóm, thuyết trình,...

2.3 Phạm vi của đề tài

- Đề tài tập trung cài đặt lại thuật toán, tái tạo lại các kết quả đã đạt được trong một bài báo uy tín và giải thích vì sao thuật toán lại có hiệu quả trong việc

khắc phục những khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn; ngoài ra có thể có thêm các thí nghiệm ngoài bài báo để thấy rõ hơn ưu/nhược điểm của thuật toán trong việc giải quyết những khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn. Lý do chúng em giới hạn đề tài như vậy là vì:

1. Chúng em thấy chỉ riêng việc tìm hiểu những khó khăn trong quá trình huấn luyện mạng nơ-ron nhiều tầng ẩn, nguyên lí hoạt động của thuật toán Adam (và các kiến thức nền tảng bên dưới, cũng như các thuật toán liên quan) và có thể tái tạo lại được kết quả thí nghiệm trong bài báo đã tốn rất nhiều thời gian.
 2. Chúng em xác định là chỉ trên cơ sở nắm rõ những khó khăn trong việc huấn luyện mạng nơ-ron nhiều tầng ẩn và hiểu rõ cách thuật toán khắc phục những khó khăn đó (và các kiến thức nền tảng bên dưới) thì mới có thể có được các cải tiến thật sự trong tương lai.
- Có thể đề xuất thêm các cải tiến, tuy nhiên, đây không phải là mục tiêu chính.

2.4 Cách tiếp cận dự kiến

Dưới đây sẽ trình bày một số bài báo liên quan đến các thuật toán có tỷ lệ học thay đổi trong quá trình huấn luyện mạng nơ-ron nhiều tầng ẩn mà chúng em đã tìm hiểu được đến thời điểm hiện tại, cũng như là bài báo mà chúng em dự kiến sẽ tập trung tìm hiểu sâu và tái tạo lại kết quả thí nghiệm.

- Thuật toán đầu tiên sử dụng tỷ lệ học thay đổi trong quá trình huấn luyện mạng nơ-ron nhiều tầng ẩn là thuật toán AdaGrad được đề xuất lần đầu trong bài báo [1] được xuất bản vào năm 2011 (số lần trích dẫn bài báo tính tới thời điểm hiện tại theo trang semantic scholar là 6464). Về cơ bản thì thuật toán sử dụng một tỷ lệ học tự điều chỉnh tương ứng với từng trọng số trong quá trình huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Dựa trên nền tảng đó, nhiều thuật toán liên quan đến việc sử dụng tỷ lệ học thay đổi ra đời, như các thuật toán RMSProp [2] và Adam [3]. Trong đó nổi bật nhất là Adam của Diederik P. Kingma và Jimmy L. Ba đã tạo nên sự khác

biệt lớn về cả tốc độ huấn luyện và độ chính xác của mạng nơ-ron nhiều tầng ẩn. Có thể thấy được sự quan tâm dành cho thuật toán Adam là rất lớn khi lượt trích dẫn bài báo đã lên đến 57333 trong vòng 5 năm kể từ năm 2015 khi bài báo được đăng lần đầu ở hội nghị ICLR. Tuy nhiên, đã có sai sót trong việc chứng minh tính hội tụ của thuật toán khi mà Adam không thể tìm được đến cực tiểu toàn cục ngay cả trong các hàm lồi đơn giản. Điều đó đã được chứng minh trong bài báo [4]. Từ đó, những cải tiến như [4], [5], [6] đã được đề xuất nhằm khắc phục nhược điểm của thuật toán Adam gốc.

Với những gì đã trình bày ở trên, trong khóa luận, chúng em dự kiến sẽ tập trung tìm hiểu và cài đặt thuật toán Adam dựa trên [3] vì:

- Đầu tiên, đây là bài báo làm cơ sở cho rất nhiều thuật toán sau này và các cài đặt của nó hiện nay vẫn được sử dụng trong việc huấn luyện mạng nơ-ron nhiều tầng ẩn, minh chứng là trong hầu hết các thư viện học máy phổ biến hiện nay đều có cài đặt sẵn thuật toán Adam. Vì những lý do trên, chúng em xác định là phải hiểu rõ thuật toán Adam trong [3] thì mới có thể hiểu được các cải tiến sau này.
- Thứ hai, để giải thích tính hiệu quả của thuật toán Adam trong việc khắc phục khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn đã phủ một lượng lớn các kiến thức nền tảng (non-convex optimization, adaptive gradient descent,...) mà chúng em nghĩ để có thể hiểu rõ tường tận là một điều không dễ dàng.

2.5 Kết quả dự kiến của đề tài

- Làm rõ tính hiệu quả của thuật toán Adam trong việc khắc phục những khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Cài đặt thuật toán Adam được đề xuất trong [3] và tái tạo lại được các kết quả thực nghiệm trong bài báo.
- Có được các kết quả thí nghiệm làm rõ được các ưu/nhược điểm của thuật toán Adam trong huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Nếu có thời gian thì có thể cài đặt và thí nghiệm thêm các cải tiến.

2.6 Kế hoạch thực hiện

Công việc	Thời gian	Người thực hiện
Tìm hiểu về tình hình nghiên cứu của bài toán huấn luyện mạng nơ-ron nhiều tầng ẩn bằng thuật toán Adam	Tháng 01/2021 - Tháng 02/2021	Như, Quân
Tìm hiểu về những khó khăn trong huấn luyện mạng nơ-ron nhiều tầng ẩn và nguyên lí hoạt động của thuật toán Adam	Tháng 03/2021	Như, Quân
Cài đặt lại thuật toán để tái tạo lại kết quả đạt được trong bài báo	Tháng 04/2021	Như, Quân
Tiến hành các thí nghiệm để thấy rõ ưu/nhược điểm của thuật toán Adam trong huấn luyện mạng nơ-ron nhiều tầng ẩn	Tháng 05/2021	Như, Quân
Viết cuốn và slide báo cáo	Tháng 05/2021 - Tháng 06/2021	Như, Quân

Bảng 1: Bảng kế hoạch thực hiện khóa luận

Tài liệu

- [1] J. C. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” in J. Mach. Learn. Res., 2011.
- [2] T. Tieleman and G. Hinton, “Lecture 6.5 - rmsprop,” COURSERA: Neural Networks for Machine Learning, 2012.
- [3] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” ICLR, vol. abs/1412.6980, 2015.

- [4] S. Reddi, S. Kale, and S. Kumar, “On the convergence of adam and beyond,” ArXiv, vol. abs/1904.09237, 2018.
- [5] T. Dozat, “Incorporating nesterov momentum into adam,” 2016.
- [6] Z. Zhang, L. Ma, Z. Li, and C. Wu, “Normalized direction-preserving adam,” ArXiv, vol. abs/1709.04546, 2017.

XÁC NHẬN
CỦA NGƯỜI HƯỚNG DẪN
(Ký và ghi rõ họ tên)

TP. Hồ Chí Minh, ngày 21/02/2021
NHÓM SINH VIÊN THỰC HIỆN
(Ký và ghi rõ họ tên)