



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 rút trích đặc trưng và đưa ra quyết định từ dữ liệu của máy tính [1]. Một hướng tiếp cận trong học máy đã và đang

đạt được những thành tựu vượt trội và được ứng dụng rộng rãi là phương pháp sử dụng mạng nơ-ron nhiều tầng ẩn, hay còn gọi là mạng học sâu (deep neural network).

Để có thể sử dụng mạng nơ-ron nhiều tầng ẩn (ví dụ, ứng dụng mạng nơ-ron để dự đoán đối tượng có trong một tấm ảnh) thì ta cần huấn luyện mạng trên một tập dữ liệu, gọi là tập dữ liệu huấn luyện (ví dụ, tập dữ liệu gồm có các ảnh và nhãn đúng cho biết đối tượng có trong ảnh). Bài toán huấn luyện mạng nơ-ron nhiều tầng ẩn được phát biểu như sau:

- Input: Hàm chi phí với các tham số là bộ trọng số của mạng nơ-ron nhiều tầng ẩn. Hàm chi phí cho biết độ lỗi trung bình của mạng nơ-ron trên tập dữ liệu huấn luyện (độ lỗi là sự sai biệt giữa giá trị dự đoán của mạng với giá trị đúng). Ngoài ra, trong hàm chi phí có thể có thêm lượng "regularization" để củng cố kết quả mong muốn: bộ trọng số mà có giá trị của hàm chi phí nhỏ là bộ trọng số không chỉ có độ lỗi nhỏ trên tập huấn luyện mà còn có độ lỗi nhỏ ngoài tập dữ liệu huấn luyện.
- Output: Bộ trọng số của mạng nơ-ron nhiều tầng ẩn cho giá trị của hàm chi phí là nhỏ nhất (hoặc ít nhất là tương đối nhỏ).

Một hướng tiếp cận thường được sử dụng để giải bài toán là SGD (Stochastic Gradient Descent) đã gặp phải nhiều khó khăn. Thứ nhất, hàm mục tiêu có nhiều cực tiểu địa phương, điểm yên ngựa, plateau,... gây cản trở quá trình đi tìm cực tiểu toàn cục với một tỷ lệ học cố định. Thứ hai, 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.

Trong thời gian gần đây, một hướng tiếp cận khác 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

- Nắm rõ các khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn.

- Nắm rõ cách thức hoạt động cũng như ưu, nhược điểm của thuật toán có tỷ lệ học thay đổi trong huấn luyện mạng nơ-ron nhiều tầng ẩn.
- 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 có tỷ lệ học thay đổi 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 gốc 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 lý hoạt động của thuật toán có tỷ lệ học thay đổi (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 có tỷ lệ học thay đổi 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 [2] được xuất bản vào năm 2011 (số lần trích dẫn bài báo 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 [3] và Adam [4]. Trong đó nổi bật nhất là Adam của Diederik P. Kingma và Jimmy L. Ba xuất bản năm 2015 tại hội nghị ICLR với 57333 lượt trích dẫn đã tạo nên sự khác biệt lớn về cả tốc độ huấn luyện và độ chính xác của mạng nơ-ron nhiều tầng ẩn. Tuy nhiên, [5] chứng minh tồn tại trường hợp Adam không thể tìm được đến cực tiểu toàn cục ngay cả trong các hàm lồi đơn giản. Từ đó, những cải tiến như [5], [6], [7] đã được đề xuất nhằm khắc phục nhược điểm của thuật toán Adam.

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 [4] 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 [4] 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 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 [4] 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] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning. MIT Press, 2016.
<http://www.deeplearningbook.org>.
- [2] J. C. Duchi, E. Hazan, and Y. Singer, “Adaptive subgradient methods for online learning and stochastic optimization,” in J. Mach. Learn. Res., 2011.
- [3] T. Tieleman and G. Hinton, “Lecture 6.5 - rmsprop,” COURSERA: Neural Networks for Machine Learning, 2012.

- [4] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” ICLR, vol. abs/1412.6980, 2015.
- [5] S. Reddi, S. Kale, and S. Kumar, “On the convergence of adam and beyond,” ArXiv, vol. abs/1904.09237, 2018.
- [6] T. Dozat, “Incorporating nesterov momentum into adam,” 2016.
- [7] 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 09/03/2021
NHÓM SINH VIÊN THỰC HIỆN
(*Ký và ghi rõ họ tên*)