



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

~~Bài toán tối ưu các tham số trong mô hình mạng nơ-ron nhiều tầng ẩn được miêu tả như sau:~~

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

sao cho giá trị của hàm mục tiêu là nhỏ nhất. Hàm mục tiêu, hay ~~còn được gọi~~ cụ thể hơn là hàm lỗi chi phí, là hàm thể hiện độ sai sót của mạng nơ-ron nhân tạo trên tập dữ liệu huấn luyện. Hàm lỗi chi phí có giá trị nhỏ đồng nghĩa với việc mạng nơ-ron được coi là hiệu quả trên tập huấn luyện, và được kỳ vọng sẽ hoạt động tốt trên dữ liệu thực. Quá trình đi tìm giá trị nhỏ nhất của hàm mục tiêu còn được gọi là tối ưu hóa hàm mục tiêu. Bài toán tối ưu hóa có vai trò cốt lõi, thiết yếu trong thực tế khi nó có thể ứng dụng trong nhiều lĩnh vực khác nhau. Trong máy học, việc tối ưu hóa tham số cho mô hình học máy là một vấn đề hiện hữu trong tất cả các kiến trúc mạng nơ-ron nhiều tầng ẩn hiện nay. Bài toán được phát biểu như sau:

- Input: ~~các tham số của mạng nơ-ron được khởi tạo ngẫu nhiên hoặc khởi tạo bằng θ~~ Hàm chi phí với các tham số là các trọng số của mạng nơ-ron.
- Output: các tham số của mạng đã được tối ưu sao cho giá trị của hàm lỗi là nhỏ nhất.

~~Bài toán tối ưu được miêu tả là quá trình tìm kiếm một bộ tham số sao cho tại đó, giá trị của hàm mục tiêu là tối đa hoặc tối thiểu. Trong lĩnh vực học máy và học sâu, bài toán trở thành tối ưu các tham số của hàm mục tiêu sao cho giá trị của mô hình trên hàm lỗi là nhỏ nhất. Khi đó, mô hình sẽ được coi là hiệu quả trên dữ liệu huấn luyện, và được kỳ vọng sẽ hoạt động tốt trên dữ liệu thực. Bài toán tối ưu có vai trò cốt lõi, thiết yếu trong thực tiễn. Trong lĩnh vực học sâu, việc tối ưu tham số cho mô hình là một vấn đề hiện hữu trong tất cả các kiến trúc mạng nơ-ron nhiều tầng ẩn hiện nay. Nếu giải quyết được bài toán này thì thời gian hao tổn cho quá trình huấn luyện mạng nơ-ron sẽ được giảm đáng kể, và mô hình được kỳ vọng hoạt động tốt trên dữ liệu thực. Việc đi tìm giá trị nhỏ nhất của hàm lỗi không phải dễ dàng vì hàm mục tiêu có thể có nhiều cực tiểu địa phương, saddle point điểm yên ngựa, plateau, ... có thể gây khó khăn cho quá trình tối ưu hóa. Ngoài ra, việc điều chỉnh các hyperparameter siêu tham số như tỷ lệ học rất khó khăn, đòi hỏi phải thử nghiệm nhiều lần và đặc trưng cho từng tập dữ liệu. Với các hướng tiếp cận truyền thống như gradient descent, stochastic~~

gradient descent sẽ gặp khó khăn trên những tập dữ liệu thưa, đồng nghĩa với việc không đủ thông tin để cập nhật các trọng 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 riêng biệt cho từng tham số và thay đổi trong quá trình huấn luyện. Đây là hướng mà tụi em sẽ tập trung tìm hiểu.

2.2 Mục tiêu đề tài

- Nắm rõ cách thức hoạt động cũng như ưu, nhược điểm của thuật toán Adam.
- Nắm rõ các khó khăn khi huấn luyện mạng nơ-ron nhiều tầng ẩn.
- Hiểu sâu về nguyên lí hoạt động, ưu và nhược điểm của ~~họ các thuật toán adaptive~~ các thuật toán có tỷ lệ học thay đổi trong quá trình huấn luyện.
- Cài đặt lại thuật toán để ra được 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.
- Nếu còn thời gian sau khi đã nắm rõ các nguyên lí 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 chỉ tìm hiểu và cài đặt lại thuật toán được đề ra trong một bài báo có uy tín~~ 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 huấn luyện mạng nơ-ron nhiều tầng ẩn; ngoài ra có thể có thêm các thí nghiệm ngoài bài báo nhằm thấy rõ hơn về ưu/nhược điểm của thuật toán. Lý do chúng em giới hạn đề tài như vậy là vì:

1. Chúng em thấy chỉ riêng việc tìm hiểu ~~nguyên lí hoạt động~~ những khó khăn trong quá trình huấn luyện mạng nơ-ron nhiều tầng ẩn và nguyên lí hoạt động của thuật toán Adam nói riêng, các thuật toán thay đổi tỷ lệ học nói chung (và các kiến thức nền tảng bên dưới, cũng như các thuật toán liên quan) và có thể ~~tự cài đặt lại~~ 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 hiểu rõ thuật toán (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 họ thuật toán adaptive 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 mô phỏng lại kết quả thí nghiệm.

- ~~Bài báo~~ Thuật toán đầu tiên đề xuất ý tưởng về họ thuật toán adaptive, tên của các thuật toán có tỷ lệ học thay đổi trong quá trình huấn luyện, 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). Thuật toán AdaGrad là cơ sở cho họ thuật toán adaptive và cũng là nền tảng để có những cải tiến sau này. Về cơ bản thì thuật toán sử dụng một ~~bước nhảy riêng biệt được học trong lúc huấn luyện~~ một bước nhảy riêng biệt cho từng tham số và các bước nhảy này được tự động điều chỉnh trong quá trình huấn luyện mô hình.
- Từ 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 ~~hai tác giả~~ Diederik P. Kingma và Jimmy L. Ba đã tạo nên sự khác biệt lớn trong tốc độ và độ chính xác. 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 ~~bài toán làm ảnh hưởng đến độ hiệu quả của~~ thuật toán trong thực tế khi mà Adam không thể ~~hội tụ~~ 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 bài báo [3] vì:

- Đây là bài báo làm cơ sở cho ~~những~~ rất nhiều thuật toán sau này sau này và các cài đặt của nó vẫn được ~~áp~~ sử dụng trong thực tế, dễ thấy nhất là trong hầu hết các thư viện học sâu phổ biến hiện nay đều có cài đặt sẵn thuật toán Adam. Chúng em xác định là phải hiểu rõ thuật toán Adam trong bài báo gốc thì mới có thể hiểu được các cải tiến sau này.
- Thuật toán Adam đã phủ một lượng lớn các kiến thức (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

- Cài đặt thuật toán Adam được đề xuất trong ~~bài báo~~ [3] và mô phỏng 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.
- 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 lý thuyết của bài toán những khó khăn trong huấn luyện mạng nơ-ron nhiều tầng ẩn và nguyên lý hoạt động của thuật toán Adam nói riêng, các thuật toán Adaptive nói chung đã chọn	Tháng 03/2021	Như, Quân
Cài đặt lại thuật toán để ra được các kết quả giống như trong bài báo	Tháng 04/2021	Như, Quân
Tiến hành các thí nghiệm để thấy rõ ưu/nhược điểm của thuật toán	Tháng 05/2021	Như, Quân
Viết cuốn và slide báo cáo	Tháng 05/2021 - Tháng 07/2021	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)