TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Nguyễn Bảo Long - Cao Tất Cường

Ứng dụng của Meta Learning và Personalization Layer trong hệ thống Federated Learning

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

Tp. Hồ Chí Minh, tháng 03/2022

TRƯỜNG ĐẠI HỌC KHOA HỌC TỰ NHIÊN
KHOA CÔNG NGHỆ THÔNG TIN

Nguyễn Bảo Long - 18120201
Cao Tất Cường - 18120296

Ứng dụng của Meta Learning và Personalized Layer trong hệ thống Federated Learning

KHÓA LUẬN TỐT NGHIỆP CỬ NHÂN
CHƯƠNG TRÌNH CHÍNH QUY

GIÁO VIÊN HƯỚNG DẪN
GS.TS. Lê Hoài Bắc

Tp. Hồ Chí Minh, tháng 03/2022

Tóm tắt

Đề cương chi tiết

Thông tin chung

- Tên đề tài: Ứng dụng của Meta Learning và Personalization Layer trong hệ thống Federated Learning
- Giảng viên hướng dẫn: GS. TS. Lê Hoài Bắc
- Nhóm sinh viên thực hiện:
  - Nguyễn Bảo Long - MSSV: 18120201
  - Cao Tất Cường - MSSV: 18120296
- Thời gian thực hiện: Từ 09/2021 đến 03/2022
- Loại đề tài: Nghiên cứu

Nội dung thực hiện

Giới thiệu đề tài

Trong bối cảnh bùng nổ thông tin cũng như việc đề cao tính riêng tư dữ liệu người dùng như hiện nay, các mô hình huấn luyện tập trung truyền thống dần bộc lộ nhiều điểm yếu khiến chúng không còn phù hợp. Ba điểm yếu làm cho cách tiếp cận cũ này trở nên tốn kém, không năng suất và ảnh hưởng đến quyền riêng tư của người dùng có thể kể đến:

- Việc truyền dữ liệu từ máy người dùng về máy chủ để tiến hành huấn luyện tiềm ẩn nguy cơ lộ dữ liệu quan trọng của người dùng.
- Chi phí truyền dữ liệu từ người dùng về máy chủ để huấn luyện ngày càng lớn do lượng dữ liệu sinh ra tại thiết bị cuối ngày càng tăng cao.

- Cần một máy chủ thật mạnh mẽ để huấn luyện mô hình với lượng dữ liệu lớn như vậy.

Khái niệm federated learning (FL) được Google lần đầu giới thiệu trong nghiên cứu [10] với ý tưởng chính là huấn luyện mô hình máy học trên các tập dữ liệu riêng biệt được phân bố trên các thiết bị biên (được gọi là huấn luyện phân tán). Với ý tưởng này, việc triển khi mô hình máy học đến người dùng không còn gặp phải vấn đề về chi phí truyền tin, giúp bảo vệ quyền riêng tư dữ liệu và không đòi hỏi một máy chủ quá mạnh để huấn luyện mô hình. Tuy nhiên, dữ liệu của người dùng trong hệ thống thường không đồng nhất và có tính cá nhân hóa rất cao (tuân theo phân phối Non-IID). Điều này khiến cho hiệu suất của hệ thống FL suy giảm nghiêm trọng [16].

Một cách ngắn gọn, hiệu suất của mô hình bị giảm là do mô hình không khái quát tốt trên tập dữ liệu của người dùng. Các thuật toán meta learning (ML) được biết đến với khả năng thích ứng nhanh trên tập dữ liệu mới [6]. Điều này giải quyết chính xác vấn đề dữ liệu FL đang gặp phải. Song song với đó, để tăng thêm tính các nhân hóa mô hình cho từng người dùng, các nghiên cứu [2, 8] đề xuất sử dụng kỹ thuật personalization layer (PL), giúp tăng đáng kể hiệu suất trên hệ thống FL. Tuy nhiên, các phương pháp tối ưu kể trên vẫn tồn tại nhiều khuyết điểm và có khả năng bù trừ cho nhau. Do đó, chúng tôi tiến hành nghiên cứu về ML, PL. Từ đó, kết hợp chúng vào hệ thống FL để đạt được hiệu suất tốt hơn.

Mục tiêu đề tài

Mục tiêu chính của đề tài này bao gồm: (1) - nghiên cứu, khảo sát các thuật toán theo hướng FL, ML, PL và (2) - kết hợp cài đặt các thuật toán trên, giúp nâng cao hiệu suất của hệ thống FL khi đối mặt với dữ liệu dạng Non-IID.

Việc nghiên cứu, khảo sát các thuật toán nhằm đưa ra đánh giá về ưu, nhược điểm của từng thuật toán. Từ đó, biết cách kết hợp chúng để đạt hiệu suất tốt.

Việc kết hợp cài đặt nhằm mục đích chứng minh thực nghiệm tính hiệu quả của thuật toán đề xuất khi làm việc với dữ liệu Non-IID.

Phạm vi đề tài

Nghiên cứu [13] chỉ ra ba hướng nghiên cứu chính khi đề cập đến một hệ thống FL: (1) - Cải thiện hiệu suất của hệ thống FL, (2) - Cải thiện khả năng bảo mật của hệ thống FL, (3) - Cải thiện vấn đề về quyền riêng tư của người dùng trong hệ thống FL.

Về việc phân loại hệ thống FL, dựa trên dữ liệu đầu vào, hệ thống FL được chia thành ba loại [13]: (1) - Horizontal FL, (2) - Vertical FL, (3) - Federated transfer learning.

Về việc phân loại các kịch bản Non-IID, nghiên cứu [16] chỉ ra bốn kịch bản chính: (1) - Phân phối thuộc tính khác nhau giữa các máy khách, (2) - Phân phối nhãn khác nhau giữa các máy khách, (3) - Phân phối thời gian khác nhau giữa các máy khách, (4) - Các kịch bản khác.

Chúng tôi giới hạn phạm vi và xây dựng phương án giải quyết của đề tài dựa trên ba giả định sau:

- Loại hệ thống: Môi trường thí nghiệm (bao gồm các yếu tố như số lượng người dùng, dữ liệu, cấu hình, khả năng lưu trữ của thiết bị cuối,...) tuân theo đặc điểm của hệ thống Horizontal FL.
- Bảo mật & quyền riêng tư: Hệ thống đã đảm bảo tính bảo mật cũng như duy trì tốt quyền riêng tư của người dùng.
- Kịch bản Non-IID: Phân phối nhãn dữ liệu trên các máy khách là khác nhau.

Cách tiếp cận

Phương pháp chính.

Hệ thống FL huấn luyện trực tiếp các mô hình máy học trên các tập dữ liệu của từng người dùng, sau đó tiến hành tổng hợp tham số của mô hình thu được từ các máy khách này để thu được một mô hình toàn cục. Do đó, kiến trúc máy chủ - máy khách nghiễm nhiên trở được nghĩ đến và trở nên phổ biến. Nghiên cứu [13] nêu ra các đặc điểm chính của máy chủ và máy khách trong một hệ thống FL:

- Máy chủ: Điều phối các hoạt động huấn luyện mô hình và duy trì một bộ tham số toàn cục bằng cách tổng hợp các tham số mô hình do máy khách gửi về.

- Máy khách: Đóng vai trò là nơi huấn luyện các mô hình học theo sự chỉ đạo của máy chủ. Chúng nhận tham số toàn cục từ máy chủ, huấn luyện mô hình dựa trên bộ tham số này và gửi tham số của mô hình mới về máy chủ để tổng hợp.

Phương pháp đề xuất của chúng tôi nhằm tích hợp các thuật toán ML, PL vào hệ thống nêu trên. Trong đó, các thuật toán ML được sử dụng để tạo ra một khởi tạo tốt, giúp máy khách hội tụ mô hình nhanh chóng (chỉ sau một hoặc một vài bước huấn luyện); các thuật toán PL được thêm vào như một phương pháp giúp tăng tính cá nhân hóa của từng mô hình máy học phân bố trên máy khách.

Dữ liệu thực nghiệm. Chúng tôi sử dụng tập dữ liệu CIFAR-10 và tập dữ liệu MNIST trong tất cả các thí nghiệm. Cả hai tập dữ liệu đều sử dụng 25% số điểm dữ liệu để kiểm thử và 75% số điểm dữ liệu để huấn luyện.

Phương pháp đối sánh. Chúng tôi sử dụng thuật toán `FedAvg` làm mô hình baseline. Để so sánh công bằng, chúng tôi cho phép mô hình này fine-tune một hoặc một vài bước trên một phần tập dữ liệu kiểm thử (thuật toán `FedAvgMeta`) trước khi bước vào kiểm thử thực sự. Các kết quả của thuật toán đề xuất sẽ được đem so sánh với kết quả của `FedAvg` và `FedAvgMeta`.

Kết quả đề tài

Sau khi tiến hành nghiên cứu, chúng tôi kỳ vọng đạt được những kết quả sau:

- Nắm được ý tưởng huấn luyện mô hình của các thuật toán theo hướng FL, ML, PL. Cài đặt mô hình baseline.

- Cài đặt được hệ thống FL có tích hợp ML. So sánh độ chính xác thu được với mô hình baseline.

- Cài đặt hệ thống FL có tích hợp ML và PL. So sánh độ chính xác thu được với mô hình baseline.

Kế hoạch thực hiện

Kế hoạch thực hiện khóa luận bao gồm 4 giai đoạn, được trình bày trong bảng sau:

Bảng 1: Bảng phân chia công việc

Giai đoạn	Công việc	Người thực hiện
1 (01/09/2021 - 15/09/2021)	Tìm hiểu kiến thức nền tảng về ML	Nguyễn Bảo Long
	Tìm hiểu kiến thức nền tảng về FL	Cao Tất Cường
	Tìm hiểu kiến thức nền tảng về PL	Cao Tất Cường
	Trao đổi 2 mảng kiến thức	Cả hai
2 (16/09/2021 - 31/01/2022)	Cài đặt các thuật toán FedAvg, FedMeta(Meta-SGD)	Nguyễn Bảo Long
	Cài đặt các thuật toán FedAvg(Meta), FedAvg(MAML)	Cao Tất Cường
	Cài đặt hệ thống FL có tích hợp ML và PL	Nguyễn Bảo Long
	Phân tích và đánh giá kết quả	Cả hai
3 (01/02/2022 - 25/02/2022)	Viết luận văn	Nguyễn Bảo Long
	Làm slide thuyết trình	Cao Tất Cường
	Tập thuyết trình	Cả hai

Lời cảm ơn

Xin phép gửi lời cảm ơn chân thành nhất đến GS. TS. Lê Hoài Bắc, người đã cung cấp, chia sẻ tài liệu và kiến thức, cũng như đã trực tiếp tham gia giảng dạy, hướng dẫn chúng tôi hoàn thiện đồ án này.

Ngoài ra, chúng tôi đặc biệt cảm ơn TS. Nguyễn Tiến Huy vì những lời khuyên rất hữu ích mỗi khi chúng tôi gặp khó khăn.

Đồ án này sẽ không thể hoàn thiện được nếu không có sự hỗ trợ kiến thức đến từ chị Bùi Thị Cẩm Nhung, cựu sinh viên chương trình Cử nhân tài năng, khoa Công Nghệ Thông Tin, khóa 2017. Chúng tôi rất cảm ơn trước những đóng góp của chị.

Mục lục

Danh sách hình

Danh sách bảng

Chương 1

Giới thiệu

1.1 Đặt vấn đề & Động lực

Hiện nay, các thiết bị biên như điện thoại, máy tính bảng, thậm chí máy giặt, máy hút bụt thông minh có thể sinh ra lượng lớn dữ liệu trong quá trình hoạt động. Lượng dữ liệu này, nếu tận dụng được, có thể mang lại sự cải thiện rất lớn về độ chính xác cho các mô hình máy học hiện tại. Ví dụ, dữ liệu thu thập được từ bàn phím điện thoại có thể phục vụ tối ưu cho các mô hình ngôn ngữ; ảnh chụp được lưu trữ trong bộ nhớ điện thoại hoàn toàn có thể được sử dụng làm dữ liệu để huấn luyện cho mô hình nhận dạng ảnh; hay lịch sử duyệt web của người dùng có thể được dùng cho bài toán đề xuất sản phẩm. Những lý do trên trở thành một động lực to lớn, thúc đẩy việc tìm ra một phương pháp giúp tận dụng nguồn dữ liệu dồi dào này.

Việc ngày càng nhiều dữ liệu được sinh ra tại các thiết bị biên khiến cho phương pháp huấn luyện mô hình theo cách tiếp cận truyền thống (được gọi là huấn luyện tập trung) bộc lộ nhiều khuyết điểm. Ba điểm yếu khiến cho các tiếp cận này không còn mạnh mẽ có thể kể đến: sự vi phạm về quyền riêng tư dữ liệu, chi phí truyền tin và chi phí phần cứng máy chủ.

Sự vi phạm về quyền riêng tư dữ liệu. Phương pháp truyền thống đòi hỏi phải gửi dữ liệu người dùng về một máy chủ để tiến hành huấn luyện mô hình. Các thông tin nhạy cảm của người dùng hoàn toàn có thể bị nghe lén bởi kẻ tấn công hoặc bị khai thác khi máy chủ bị nhiễm mã độc. Điều này ảnh hưởng nghiêm trọng đến quyền riêng tư dữ liệu của người dùng - một vấn đề mà hiện nay đang nhận được rất nhiều sự quan tâm từ cả người dùng lẫn chính phủ.

Chi phí truyền tin. Dữ liệu sinh ra tại thiết bị biên đang ngày một tăng lên do văn hóa sử dụng và sự phát triển của công nghệ. Một người

dùng điện thoại thông minh giờ đây có thể thực hiện giao dịch tài chính, nghe nhạc, lướt web, xem phim ngay trên thiết bị của mình. Một máy hút bụi thông minh được trang bị các cảm biến nên hoàn toàn có thể sử dụng dữ liệu cảm biến này như một "time series". Chi phí truyền tin từ các thiết bị biên đến máy chủ để huấn luyện trở nên tốn kém và có thể gây mất thông tin, ảnh hưởng đến hiệu suất học của mô hình.

Chi phí phần cứng máy chủ. Sau khi dữ liệu được gửi về máy chủ, cần một cấu hình máy mạnh mẽ cùng khả năng lưu trữ lớn để có thể xử lý hết lượng dữ liệu khổng lồ trên trong thời gian giới hạn.

Việc các phương pháp tiếp cận máy học theo hướng truyền thống đang dần bộc lộ các nhược điểm về chi phí vận hành và bảo trì ngày càng cao, cũng như các mối nguy hiểm tiềm tàng có thể xảy ra đối với dữ liệu của người dùng, một lần nữa thúc đẩy việc nghiên cứu về một phương pháp huấn luyện giúp làm giảm chi phí phần cứng (sử dụng cho đường truyền và máy chủ), đồng thời đảm bảo tính riêng tư dữ liệu cho người dùng. Khái niệm federated learning (FL) và thuật toán `FedAvg` được đưa ra vào năm 2016 bởi Google trong nghiên cứu [10] nhằm mục đích huấn luyện mô hình máy học trên các tập dữ liệu riêng biệt được phân bố trên các thiết bị biên (được gọi là huấn luyện phân tán). Do đó, một hệ thống FL không cần một máy chủ quá mạnh để vận hành (thậm chí có thể sử dụng một máy khách để vận hành [13] , không đòi hỏi chi phí truyền tin quá lớn và đảm bảo được quyền riêng tư dữ liệu của người dùng vì không diễn ra bất cứ quá trình thu thập dữ liệu từ người dùng nào (điều mà mô hình huấn luyện tập trung bắt buộc phải làm). Dễ thấy rằng, phần lớn quá trình tính toán được chuyển đến các thiết bị biên. Tuy nhiên, khả năng lưu trữ và tính toán tại các thiết bị này ngày càng được cải thiện, khiến cho việc việc huấn luyện phân tán dần trở nên khả thi và đạt hiệu quả cao hơn.

Mặt khác, nghiên cứu [16] chỉ ra rằng, hệ thống FL hoạt động trên nền thuật toán `FedAvg` bị giảm hiệu suất nghiêm trọng khi xử lý dữ liệu đầu vào tuân theo phân phối Non-IID. Trong khi đó, dữ liệu phân bố trên máy khách là không đồng nhất và có tính cá nhân hóa tất cao. Nói cách khác, các tập dữ liệu này tuân theo phân phối Non-IID. Do đó, việc nghiên cứu và cải tiến hệ thống FL để thu được kết quả cao trên dữ liệu Non-IID là rất cần thiết. Đây chính là vấn đề mà khóa luận hướng tới giải quyết.

1.2 Phạm vi đề tài

Nghiên cứu [13] chỉ ra ba hướng nghiên cứu chính khi đề cập đến một hệ thống FL: (1) - Cải thiện hiệu suất của hệ thống FL, (2) - Cải thiện khả năng bảo mật của hệ thống FL, (3) - Cải thiện vấn đề về quyền riêng tư của người dùng trong hệ thống FL.

Về việc phân loại hệ thống FL, dựa trên dữ liệu đầu vào, hệ thống FL được chia thành ba loại [13]: (1) - Horizontal FL, (2) - Vertical FL, (3) - Federated transfer learning.

Về việc phân loại các kịch bản Non-IID, nghiên cứu [16] chỉ ra bốn kịch bản chính: (1) - Phân phối thuộc tính khác nhau giữa các máy khách, (2) - Phân phối nhãn khác nhau giữa các máy khách, (3) - Phân phối thời gian khác nhau giữa các máy khách, (4) - Các kịch bản khác.

Chúng tôi giới hạn phạm vi và xây dựng phương án giải quyết của đề tài dựa trên ba giả định sau:

- Loại hệ thống: Môi trường thí nghiệm (bao gồm các yếu tố như số lượng người dùng, dữ liệu, cấu hình, khả năng lưu trữ của thiết bị cuối,...) tuân theo đặc điểm của hệ thống Horizontal FL.
- Bảo mật & quyền riêng tư: Hệ thống đã đảm bảo tính bảo mật cũng như duy trì tốt quyền riêng tư của người dùng.
- Kịch bản Non-IID: Phân phối nhãn dữ liệu trên các máy khách là khác nhau.

1.3 Đóng góp chính

Chúng tôi chia các đóng góp chính của mình thành hai loại: đóng góp về mặt lý thuyết và đóng góp về mặt thực nghiệm.

1.3.1 Đóng góp lý thuyết

- Nghiên cứu hệ thống FL và thách thức về phân phối dữ liệu mà hệ thống Horizontal FL gặp phải.

- Khảo sát các phương pháp tối ưu hóa hệ thống Horizontal FL trên dữ liệu dạng Non-IID. Trong đó, tập trung nghiên cứu các phương pháp theo hướng Personalized Federated Averaging [5, 4] và Personalization Layer [8, 2].

- Phương pháp đề xuất của chúng tôi đã cho thấy khả năng đạt độ chính xác cao hơn trong quá trình kiểm thử với hai đối tượng người dùng (người dùng cục bộ và người dùng mới) so với các phương pháp trước đó (chỉ sử dụng `FedAvg`, chỉ sử dụng Personalized Federated Averaging, hoặc chỉ sử dụng Personalization Layer).

1.3.2 Đóng góp thực nghiệm

- Tổ chức bộ dữ liệu MNIST và CIFAR-10 theo hai hướng IID và Non-IID để tiến hành thí nghiệm.

- Cài đặt thuật toán `FedAvg`, `FedAvgMeta`, các thuật toán kết hợp giữa `FedAvg` và ML (thuật toán `FedMeta(MAML)`, `FedMeta(Meta-SGD)`).

- Cài đặt thuật toán kết hợp giữa `FedAvg` và PL (thuật toán `FedPer`, `LG-FedAvg`).

- Kết hợp các thuật toán ML và PL vào hệ thống FL.

- Fine-tune các siêu tham số như số lượng máy khách tham gia huấn luyện, số bước huấn luyện cục bộ, các siêu tham số học để mô hình đạt độ chính xác tốt nhất.

1.4 Bố cục

Trong luận văn này, chương 2 trình bày về tổng quan lý thuyết được sử dụng trong khóa luận, các lý thuyết này làm nền tảng cho nghiên cứu và đề xuất thuật toán; chương 3 đề xuất thuật toán giúp giải quyết vấn đề vừa nêu ở chương 1; chương 4 trình bày về cài đặt thực nghiệm để kiểm chứng tính hiệu quả của thuật toán; chương 5 đi vào phân tích kết quả đạt được; chương 6 nêu kết luận, những điều chưa làm được và hướng phát triển tương lai của khóa luận.

Chương 2

Tổng quan lý thuyết

2.1 Hệ thống Federated Learning

2.1.1 Định nghĩa

Định nghĩa về FL [12]: Giả sử có n máy khách, máy khách thứ i ký hiệu là c_i ($i \in [1, n]$), chứa tập dữ liệu D_i. FL là một quá trình học mà ở đó, các chủ sở hữu dữ liệu (ở đây có thể hiểu là các thiết bị biên) cùng hợp tác huấn luyện một mô hình $\mathcal{M}$ và đạt được độ chính xác f nhưng không có bất kỳ chủ sở hữu dữ liệu c_i nào chia sẻ tập dữ liệu $\mathcal{D}_i$ của chúng.

Gọi $\bar{\mathcal{M}}$ là mô hình máy học được huấn luyện trên tập dữ liệu $\mathcal{D} = \mathcal{D}_1 \cup \mathcal{D}_2 \cup ... \cup \mathcal{D}_n$ và cho độ chính xác $\bar{f}$. f và $\bar{f}$ chỉ được phép chênh lệch nhau một khoảng nhỏ. Gọi δ là một giá trị thực không âm, nếu $| f - \bar{f} | < \delta$ ta nói mô hình $\mathcal{M}$ có δ - *accuracy loss*.

Định nghĩa về tính hợp lệ [7]: Ký hiệu $\mathcal{M}_i$ là mô hình được huấn luyện trên tập dữ liệu $\mathcal{D}_i$ và cho độ chính xác f_i. Mô hình $\mathcal{M}$ được gọi là hợp lệ nếu tồn tại $i \in [1, n]$ sao cho $f > f_i$.

2.1.2 Một hệ thống Federated Learning điển hình

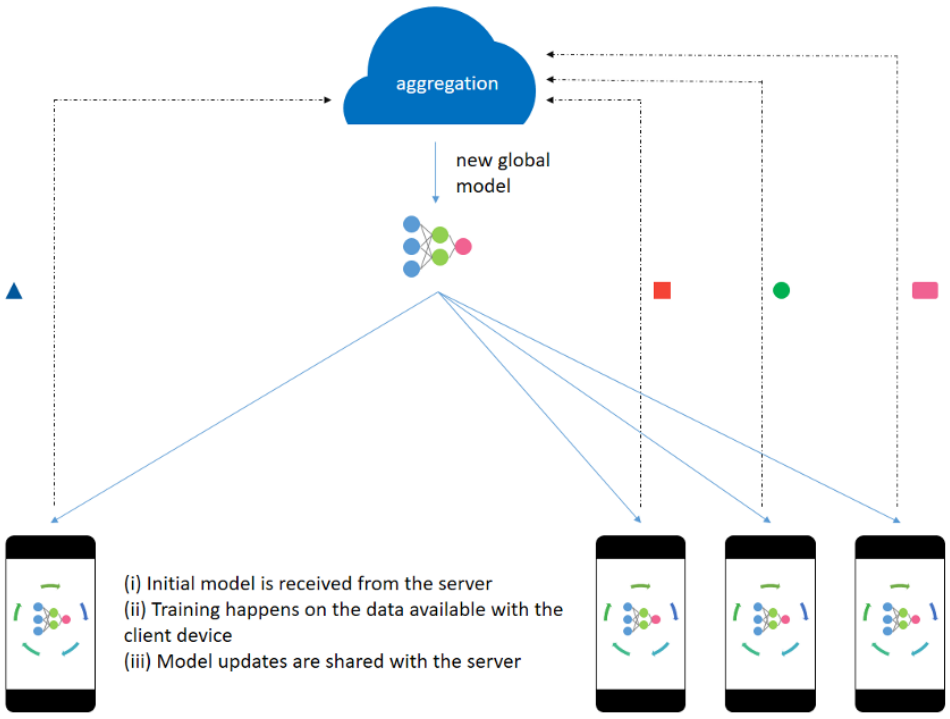
Thành phần và các tương tác trong hệ thống. Một hệ thống FL (Hình 2.1) thường bao gồm hai thành phần chính: máy chủ (đóng vai trò là đối tượng duy trì mô hình toàn cục) và máy khách (đóng vai trò là đối tượng nắm giữ dữ liệu huấn luyện). Hai thành phần này tương tác với nhau theo ba bước sau [9]:

- *Khởi tạo.* Máy chủ khởi tạo trọng số w_G^0 cho mô hình toàn cục và các siêu tham số cho quá trình huấn luyện. Thông tin này sau đó được gửi đến một tập hợp con các máy khách được chọn để tiến hành huấn luyện.

- *Huấn luyện và cập nhật mô hình cục bộ.* Tại bước huấn luyện thứ t, máy khách c_i nhận trọng số w_G^t từ máy chủ và tiến hành huấn luyện cục bộ trên tập dữ liệu D_i. Tham số θ_i^t thu được sau quá trình huấn luyện (có thể là trọng số w_i^t hoặc đạo hàm hàm lỗi g_i) được máy khách gửi về máy chủ để tổng hợp.

- *Tổng hợp và cập nhật mô hình toàn cục.* Máy chủ nhận tham số θ_i^t gửi về từ các máy khách được chọn trước đó, tiến hành tổng hợp w_G^{t+1} - trọng số mới của mô hình toàn cục và gửi trọng số này đến một tập hợp con các máy khách khác để bắt đầu bước huấn luyện toàn cục mới.

Máy chủ sẽ lặp lại bước 2 và bước 3 cho đến khi độ lỗi hội tụ hoặc độ chính xác đạt đến một ngưỡng nhất định. Khi quá trình huấn luyện kết thúc, tham số của mô hình toàn cục sẽ được phân phối đến toàn bộ máy khách trong hệ thống.



Hình 2.1: Hai thành phần chính và quá trình tương tác giữa chúng trong hệ thống FL [3]

Mục tiêu của hệ thống FL. Chúng tôi khảo sát hai mục tiêu của hệ thống FL: (1) - Mục tiêu cục bộ; (2) - Mục tiêu toàn cục.

Các máy khách trong hệ thống hướng đến việc thực hiện mục tiêu cục bộ. Ban đầu, máy khách c_i nhận một trọng số toàn cục w_G từ máy chủ. Máy khách này sau đó sẽ cố gắng tìm kiếm một trọng số w_i^* giúp cực tiểu hóa hàm lỗi cục bộ. Nói cách khác, w_i^* phải thỏa mãn thỏa mãn:

$$w_i^* = \arg\min_{w_i} f_{local}(w_i) \tag{2.1}$$

Trong đó, $f_{local}(w_i)$ là hàm lỗi trên tập dữ liệu của c_i. Với α là siêu tham số học cục bộ, $w_{i(j)}$ là trọng số tại bước huấn luyện j của c_i, lời giải của phương trình 2.1 theo phương pháp SGD có thể được viết như sau:

$$\begin{cases} w_{i(0)} = w_G \\ w_{i(j)} = w_{i(j-1)} - \alpha \nabla f_{local}(w_{i(j-1)}) \end{cases} \tag{2.2}$$

Mặt khác, mục tiêu toàn cục, cũng là mục tiêu chính của hệ thống FL, được máy chủ thực hiện bằng cách tìm kiếm một trọng số w_G^* giúp tối thiểu hóa hàm lỗi của cả hệ thống [13]:

$$\begin{aligned} w_G^* &= \arg\min_{w_G} f_{global}(w_G) \\ &= \arg\min_{w_G} \frac{1}{n} \sum_{i=1}^{n} f_{local}(w_i) \end{aligned} \tag{2.3}$$

Trong đó, $f_{global}(w_G)$ là hàm lỗi toàn cục của hệ thống. Để giải phương trình 2.3, máy chủ thực hiện tổng hợp tham số gửi về từ máy khách bằng một trong hai cách: lấy trung bình trọng số [10, 1, 14] hoặc lấy trung bình đạo hàm [4, 11].

Đặt $n_i = |\mathcal{D}_i|$ là số điểm dữ liệu của tập $\mathcal{D}_i$, $N = \sum_{i=1}^{n} n_i$ là tổng số điểm dữ liệu có trong cả hệ thống. Phương pháp lấy trung bình trọng số tính toán w_G^{t+1} từ các trọng số của máy khách như sau [10]:

$$w_G^{t+1} = \sum_{i=1}^{n} \frac{n_i}{N} w_i^t \tag{2.4}$$

Trái lại, phương pháp lấy trung bình đạo hàm đòi hỏi máy khách gửi

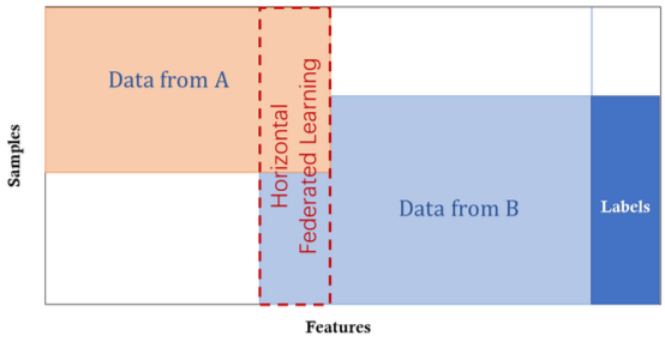
về đạo hàm hàm lỗi sau khi kết thúc quá trình huấn luyện cục bộ. Với β là siêu tham số học toàn cục, quá trình tổng hợp được biểu diễn theo công thức:

$$\begin{aligned} w_G^{t+1} &= w_G^t - \beta \left[\frac{1}{n} \sum_{i=1}^{n} \frac{n_i}{N} \nabla f_{local}(w_i) \right] \\ &= w_G^t - \beta g^t \end{aligned} \tag{2.5}$$

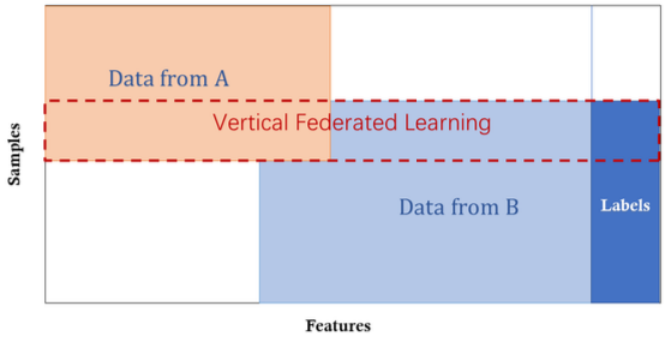
Sau khi khảo sát cả hai phương pháp tổng hợp tham số của máy chủ, nghiên cứu [13] chỉ ra rằng, việc lấy trung bình trọng số giúp hệ thống có khả năng chịu được việc mất cập nhật, nhưng không đảm bảo việc hội tụ. Trái lại, việc lấy trung bình đạo hàm giúp hệ thống đảm bảo sự hội tụ nhưng tiêu tốn nhiều chi phí truyền tin hơn. Trong nghiên cứu này, chúng tôi tổng hợp trọng số toàn cục bằng phương pháp lấy trung bình trọng số để phù hợp hơn với giới hạn về chi phí giao tiếp và lưu trữ.

Phân loại hệ thống Federated Learning. Nghiên cứu [13] đề xuất các phân loại các hệ thống FL dựa trên phân bố dữ liệu đầu vào của chúng. Theo đó, ba phân bố dữ liệu: (1) - Phân bố dữ liệu theo chiều ngang (Horizontal data partitioning), (2) - Phân bố dữ liệu theo chiều dọc (Vertical data partitioning), (3) - Phân bố dữ liệu hỗn hợp (Hybrid data partitioning) sẽ ứng với ba loại hệ thống FL (Hình 2.2): (1) - Hệ thống FL theo chiều ngang (Horizontal FL), (2) - Hệ thống FL theo chiều dọc (Vertical FL), (3) - Hệ thống học chuyển giao tri thức (Federated Transfer Learning).

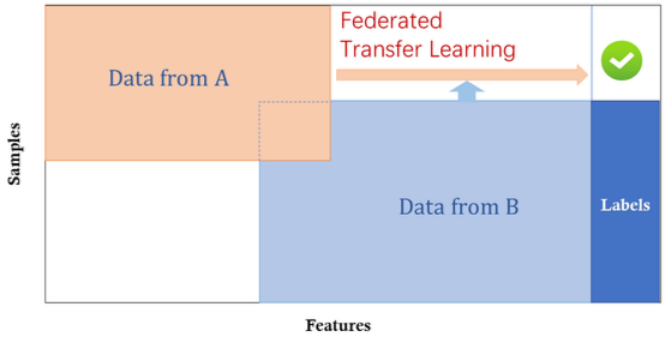
Hệ thống Horizontal FL. Phân bố dữ liệu theo chiều ngang là kiểu phân bố dữ liệu mà ở đó các bên tham gia vào hệ thống cùng sở hữu các đặc tính dữ liệu giống nhau nhưng giá trị định danh của mẫu dữ liệu của các bên là khác nhau. Ví dụ, khi các bên tham gia hệ thống là các trường đại học, họ sẽ muốn quản lý các thông tin giống nhau về sinh viên như họ và tên, mã số sinh viên,... nhưng không có một sinh viên nào tham gia hai trường đại học cùng một lúc. Kiến trúc Horizontal FL rất phù hợp để huấn luyện mô hình học tuân theo phân phối này [13].



(a) Horizontal Federated Learning



(b) Vertical Federated Learning



(c) Federated Transfer Learning

Hình 2.2: Ba loại hệ thống FL với phân bố dữ liệu tương ứng [12]

Dựa vào kiến trúc giao tiếp, có thể chia Horizontal FL ra làm hai loại: Kiến trúc client-server và kiến trúc peer-to-peer (P2P). Kiến trúc client-server, hay còn gọi là kiến trúc FL tập trung, về cơ bản sẽ thực hiện các bước huấn luyện giống như đã trình bày trong phần **Thành phần và các**

tương tác trong hệ thống. Trong khi đó, kiến trúc P2P, hay còn gọi là kiến trúc FL phân tán không có một máy chủ cố định. Tại mỗi bước huấn luyện toàn cục, một máy khách trong hệ thống được chọn làm máy chủ. Quá trình huấn luyện sau đó được thực hiện giống như kiến trúc client-server.

Một hệ thống Horizontal FL thường có số lượng máy khách rất lớn, khả năng lưu trữ và tính toán tại các máy khách không cần quá cao (ví dụ như điện thoại thông minh, máy tính bảng) và tần suất một máy khách tham gia huấn luyện là rất thấp.

Hệ thống Vertical FL. Đây là kiến trúc phù hợp với phân bố dữ liệu theo chiều dọc. Trong phân bố dữ liệu dạng này, các bên tham gia hệ thống sở hữu các đặc tính dữ liệu khác nhau nhưng giá trị định danh của mẫu dữ liệu của các bên là giống nhau. Ví dụ, khi các bên tham gia hệ thống là ngân hàng và trường đại học. Thuộc tính mà ngân hàng và trường đại học lưu trữ là rất khác nhau nhưng lại chứa thông tin của cùng một người dùng.

Hệ thống Federated Transfer Learning. Khi phân bố dữ liệu của các bên tham gia hệ thống không sở hữu chung các đặc tính dữ liệu hay giá trị định danh của từng mẫu, người ta gọi đây là phân bố dữ liệu hỗn hợp. Ví dụ, khi các bên tham gia hệ thống là một ngân hàng ở Hoa Kỳ và một trường đại học ở Việt Nam. Do cản trở địa lý và nhu cầu quản lý thông tin khác nhau, chủ sở hữu dữ liệu này sẽ không có chung thuộc tính hay giá trị định danh nào. Trong trường hợp này, FTL có thể được sử dụng để chuyển giao tri thức giữa các bên tham gia.

Dựa vào các đặc điểm phân loại nêu trên, chúng tôi xếp nghiên cứu của mình vào nhóm hệ thống Horizontal FL tập trung.

2.2 Hệ thống Federated Learning trên dữ liệu Non-IID

Dữ liệu tại các máy khách thường được sinh ra dựa trên nhu cầu của người dùng cuối. Do đó, loại dữ liệu này thường có tính cá nhân hóa cao và không đồng nhất. Nói cách khác, không có bất kỳ phân phối dữ liệu

cục bộ nào có thể đại diện cho phân phối trên toàn bộ dữ liệu, phân phối dữ liệu trên hai máy khách khác nhau là khác nhau [16]. Đây chính là ý tưởng mà thuật ngữ *dữ liệu Non-IID* muốn truyền đạt.

Mặt khác, nghiên cứu [15] chỉ ra rằng hệ thống FL có thể bị giảm hiệu quả nghiêm trọng khi đối mặt với dữ liệu Non-IID. Trong phần này, dựa trên nghiên cứu [16] chúng tôi tiến hành khảo sát các kịch bản về dữ liệu Non-IID, các hướng tối ưu hệ thống trên loại dữ liệu này. Trong đó, đi sâu vào hai hướng: Sử dụng Meta Learning và Personalization Layer trong tối ưu hệ thống FL.

2.2.1 Các kịch bản Non-IID

2.2.2 Tối ưu hệ thống Federated Learning trên dữ liệu Non-IID

Tối ưu dựa trên dữ liệu.

Meta Learning

Personalization Layer

Chương 3

Phương pháp đề xuất

Chương 4

Cài đặt thực nghiệm

Chương 5

Kết quả & Thảo luận

Chương 6

Kết luận

Tài liệu tham khảo

References

[1] Yoshinori Aono et al. "Privacy-preserving deep learning via additively homomorphic encryption". In: *IEEE Transactions on Information Forensics and Security* 13.5 (2017), pp. 1333–1345.

[2] Manoj Ghuhan Arivazhagan et al. "Federated learning with personalization layers". In: *arXiv preprint arXiv:1912.00818* (2019).

[3] Kapil Chandorikar. *Introduction to Federated Learning and Privacy Preservation*. 2020. URL: `https://towardsdatascience.com/introduction-to-federated-learning-and-privacy-preservation-75644686b559`.

[4] Fei Chen et al. "Federated meta-learning with fast convergence and efficient communication". In: *arXiv preprint arXiv:1802.07876* (2018).

[5] Alireza Fallah, Aryan Mokhtari, and Asuman Ozdaglar. "Personalized federated learning: A meta-learning approach". In: *arXiv preprint arXiv:2002.07948* (2020).

[6] Timothy Hospedales et al. "Meta-learning in neural networks: A survey". In: *arXiv preprint arXiv:2004.05439* (2020).

[7] Qinbin Li et al. "A survey on federated learning systems: vision, hype and reality for data privacy and protection". In: *IEEE Transactions on Knowledge and Data Engineering* (2021).

[8] Paul Pu Liang et al. "Think locally, act globally: Federated learning with local and global representations". In: *arXiv preprint arXiv:2001.01523* (2020).

[9] Wei Yang Bryan Lim et al. "Federated learning in mobile edge networks: A comprehensive survey". In: *IEEE Communications Surveys & Tutorials* 22.3 (2020), pp. 2031–2063.

[10] Brendan McMahan et al. “Communication-efficient learning of deep networks from decentralized data”. In: *Artificial intelligence and statistics*. PMLR. 2017, pp. 1273–1282.

[11] H Brendan McMahan et al. “Learning differentially private recurrent language models”. In: *arXiv preprint arXiv:1710.06963* (2017).

[12] Qiang Yang et al. “Federated machine learning: Concept and applications”. In: *ACM Transactions on Intelligent Systems and Technology (TIST)* 10.2 (2019), pp. 1–19.

[13] Xuefei Yin, Yanming Zhu, and Jiankun Hu. “A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions”. In: *ACM Computing Surveys (CSUR)* 54.6 (2021), pp. 1–36.

[14] Tehrim Yoon et al. “Fedmix: Approximation of mixup under mean augmented federated learning”. In: *arXiv preprint arXiv:2107.00233* (2021).

[15] Yue Zhao et al. “Federated learning with non-iid data”. In: *arXiv preprint arXiv:1806.00582* (2018).

[16] Hangyu Zhu et al. “Federated Learning on Non-IID Data: A Survey”. In: *arXiv preprint arXiv:2106.06843* (2021).