Học tập hiệu quả về giao tiếp của các mạng sâu từ dữ liệu phi tập trung

VÕ QUỐC HUY, NGUYỄN XUÂN GIANG, NGÔ HOÀNG NHẬT HUY

*1 Khoa công nghệ Thông tin, trường Đại học Công nghiệp Thành phố Hồ Chí Minh*

[*huyvo8500@gmail.com*](mailto:huyvo8500@gmail.com)*,* [*nxgiang2202@gmail.com*](mailto:nxgiang2202@gmail.com)*,* *benhathuyvip246@gmail.com*

**Tóm tắt:** Với sự bùng nổ của các thiết bị di động hiện đại, có sẵn một lượng lớn dữ liệu để huấn luyện các mô hình học sâu, từ đó có thể cải thiện đáng kể trải nghiệm người dùng. Tuy nhiên, tính nhạy cảm về quyền riêng tư và lượng dữ liệu lớn này thường gây khó khăn cho các phương pháp truyền thống mà yêu cầu tập trung dữ liệu trong một trung tâm dữ liệu để huấn luyện. Nhằm giải quyết vấn đề này, chúng tôi đề xuất một phương pháp phi tập trung được gọi là Học liên kết, cho phép dữ liệu huấn luyện phân tán trên các thiết bị di động và tổng hợp các cập nhật tính toán cục bộ để học một mô hình chung. Nghiên cứu này giới thiệu một phương pháp thực tế để đạt được hiệu quả học tập giao tiếp của các mạng sâu từ dữ liệu phi tập trung. Cách tiếp cận được xây dựng dựa trên tính trung bình của mô hình lặp và giải quyết các vấn đề phát sinh từ dữ liệu phân tán không cân bằng và không giống hệt nhau, thường gặp phải trong các cài đặt phi tập trung. Kết quả thử nghiệm của chúng tôi chứng minh tính mạnh mẽ của phương pháp được đề xuất trong việc xử lý các đặc điểm phân phối dữ liệu duy nhất của dữ liệu phi tập trung. Nghiên cứu này đóng góp vào sự tiến bộ của việc học tập hiệu quả trong giao tiếp bằng cách khai thác tiềm năng của dữ liệu phi tập trung trên thiết bị di động. Phương pháp Học tập Liên kết giải quyết hiệu quả các mối lo ngại về quyền riêng tư và tối đa hóa việc sử dụng dữ liệu phi tập trung quy mô lớn trong khi vẫn duy trì hiệu suất của mô hình. Những phát hiện này có ý nghĩa quan trọng trong việc nâng cao hiệu quả và khả năng mở rộng của các mô hình học sâu trong môi trường dữ liệu phi tập trung.

Communication-Efficient Learning of Deep Networksfrom Decentralized Data

**Abstract:** With the proliferation of modern mobile devices, there is a wealth of data available for training deep learning models, which can greatly enhance user experiences. However, the privacy sensitivity and sheer quantity of this data often make it challenging to employ conventional approaches that involve centralizing the data in a data center for training. In response, we propose a decentralized approach called Federated Learning, which enables training data to remain distributed on mobile devices while aggregating locally-computed updates to learn a shared model. This study introduces a practical method for achieving communication-efficient learning of deep networks from decentralized data. The approach is built upon iterative model averaging and tackles the issues arising from unbalanced and non-identically distributed data, commonly encountered in decentralized settings. The results of our experiments demonstrate the robustness of the proposed approach in handling the unique data distribution characteristics of decentralized data. This research contributes to the advancement of communication-efficient learning by harnessing the potential of decentralized data on mobile devices. The Federated Learning approach effectively addresses privacy concerns and maximizes the utilization of large-scale decentralized data while preserving model performance. These findings have significant implications for enhancing the efficiency and scalability of deep learning models within decentralized data environments.

**Keywords:** Federated Learning, decentralized data, communication-efficient learning

1. **Giới thiệu**

Sự phát triển nhanh chóng của các thiết bị di động hiện đại đã dẫn đến lượng dữ liệu sẵn có chưa từng có, tạo cơ hội cải thiện trải nghiệm người dùng thông qua đào tạo các mô hình học sâu. Tuy nhiên, độ nhạy cảm về quyền riêng tư và số lượng tuyệt đối của dữ liệu này thường đặt ra những thách thức khi sử dụng các phương pháp truyền thống liên quan đến việc tập trung dữ liệu vào một trung tâm dữ liệu để đào tạo. Những thách thức này đã thúc đẩy việc khám phá các phương pháp học tập phi tập trung cho phép dữ liệu đào tạo được phân phối trên các thiết bị di động.

Trong bối cảnh này, Học liên kết đã nổi lên như một cách tiếp cận đầy hứa hẹn để cho phép đào tạo các mạng sâu sử dụng dữ liệu phi tập trung. Thay vì tổng hợp dữ liệu ở một vị trí trung tâm, Học tập Liên kết nhằm mục đích đưa quá trình học tập vào chính dữ liệu đó. Bằng cách cho phép các thiết bị di động cùng nhau tính toán và cập nhật mô hình được chia sẻ, Federated Learning bảo vệ quyền riêng tư của dữ liệu và giảm bớt nhu cầu truyền dữ liệu nhạy cảm đến máy chủ trung tâm. Cách tiếp cận phi tập trung này có khả năng cách mạng hóa cách đào tạo các mô hình học sâu, cho phép những tiến bộ trong các lĩnh vực khác nhau như xử lý ngôn ngữ tự nhiên, thị giác máy tính và hệ thống đề xuất.

Trọng tâm của nghiên cứu này là học tập hiệu quả trong giao tiếp của các mạng sâu từ dữ liệu phi tập trung. Giao tiếp hiệu quả là rất quan trọng trong các tình huống học tập phi tập trung, nơi các thiết bị có tài nguyên hạn chế và kết nối mạng không liên tục. Giảm chi phí giao tiếp trong khi duy trì độ chính xác của mô hình là điều cần thiết để cho phép học tập hiệu quả trên thiết bị di động. Bằng cách giải quyết những thách thức này, chúng tôi mong muốn cải thiện khả năng mở rộng, quyền riêng tư và hiệu quả tổng thể của các mô hình học sâu được đào tạo trên dữ liệu phi tập trung.

Để đạt được hiệu quả truyền thông, chúng tôi đề xuất một phương pháp thực tế dựa trên tính trung bình của mô hình lặp. Chúng tôi đánh giá cách tiếp cận của mình thông qua các thử nghiệm thực nghiệm mở rộng, xem xét các kiến trúc mô hình và bộ dữ liệu khác nhau đại diện cho sự đa dạng của các tình huống trong thế giới thực. Các thử nghiệm của chúng tôi nhằm mục đích chứng minh tính mạnh mẽ của cách tiếp cận đối với những thách thức do dữ liệu phân tán không cân bằng và không giống hệt nhau thường gặp trong các cài đặt phi tập trung.

Những đóng góp của nghiên cứu này nằm ở việc thúc đẩy sự hiểu biết và ứng dụng học tập hiệu quả trong giao tiếp từ dữ liệu phi tập trung. Bằng cách tận dụng Học tập liên kết và khám phá các kỹ thuật đổi mới, chúng tôi mong muốn khai thác tiềm năng của dữ liệu phi tập trung đồng thời giải quyết các mối lo ngại về quyền riêng tư và cải thiện khả năng mở rộng của các mô hình học sâu. Các phần sau đây trình bày phương pháp, thiết lập thử nghiệm và kết quả được đề xuất của chúng tôi, làm sáng tỏ hiệu quả và lợi ích của việc học tập hiệu quả trong giao tiếp từ dữ liệu phi tập trung.

**Federated Learning -** Vấn đề lý tưởng cho việc học phân tán có các thuộc tính sau đây:

1. Huấn luyện trên dữ liệu thực tế từ các thiết bị di động mang lại lợi thế riêng so với việc huấn luyện trên dữ liệu ảo thông thường có sẵn trong trung tâm dữ liệu.
2. Dữ liệu này nhạy cảm về quyền riêng tư hoặc có kích thước lớn (so với kích thước của mô hình), vì vậy không nên ghi lại nó trên trung tâm dữ liệu chỉ để huấn luyện mô hình (theo nguyên tắc thu thập dữ liệu tập trung).
3. Đối với các nhiệm vụ có giám sát, nhãn trên dữ liệu có thể được suy luận một cách tự nhiên từ tương tác của người dùng.

Nhiều mô hình hỗ trợ hành vi thông minh trên thiết bị di động đáp ứng các tiêu chí trên. Hãy xem xét hai ví dụ: phân loại hình ảnh, ví dụ như dự đoán những bức ảnh có khả năng được xem nhiều lần trong tương lai hoặc được chia sẻ; và mô hình ngôn ngữ, có thể được sử dụng để cải thiện nhận dạng giọng nói và nhập văn bản trên bàn phím cảm ứng bằng cách cải thiện giải mã, dự đoán từ tiếp theo và thậm chí dự đoán toàn bộ câu trả lời. Dữ liệu huấn luyện tiềm năng cho cả hai nhiệm vụ này (tất cả các bức ảnh mà người dùng chụp và mọi thứ họ gõ trên bàn phím di động, bao gồm mật khẩu, URL, tin nhắn, v.v.) có thể nhạy cảm về quyền riêng tư. Các phân phối mà các ví dụ này được rút ra cũng có khả năng khác rất nhiều so với các bộ dữ liệu ảo dễ dàng có sẵn: việc sử dụng ngôn ngữ trong cuộc trò chuyện và tin nhắn văn bản thường khác rất nhiều so với các tập dữ liệu ngôn ngữ tiêu chuẩn, ví dụ như Wikipedia và các tài liệu web khác; những bức ảnh mà mọi người chụp bằng điện thoại di động của họ có khả năng rất khác so với những bức ảnh Flickr điển hình. Và cuối cùng, nhãn cho những vấn đề này có sẵn: văn bản được nhập vào tự nhận dạng cho việc học mô hình ngôn ngữ và nhãn cho các bức ảnh có thể được xác định bằng tương tác tự nhiên của người dùng với ứng dụng ảnh của họ (những bức ảnh nào bị xóa, chia sẻ hoặc xem).

Cả hai nhiệm vụ này đều phù hợp với việc học mạng nơ-ron. Đối với phân loại hình ảnh, mạng nơ-ron sâu feed-forward và đặc biệt là mạng tích chập đã chứng minh mang lại kết quả hàng đầu. Đối với nhiệm vụ mô hình ngôn ngữ, mạng nơ-ron tuần hoàn và đặc biệt là mạng LSTM đã đạt được kết quả hàng đầu.

**Privacy** trong Federated learning là một khía cạnh quan trọng và được coi là một ưu điểm so với việc huấn luyện trên dữ liệu được lưu trữ tại trung tâm dữ liệu.

Trong federated learning, việc truyền thông tin chỉ giới hạn ở việc cập nhật tối thiểu cần thiết để cải thiện một mô hình cụ thể. Thông tin được truyền đi là những thay đổi nhỏ nhất so với dữ liệu gốc, và nói chung không chứa nhiều thông tin hơn. Điều này đảm bảo rằng dữ liệu người dùng không bị tiết lộ hoặc phơi bày trong quá trình huấn luyện.

Hơn nữa, các cập nhật này có thể và nên được xem như tạm thời. Chúng không được lưu trữ hoặc tiếp tục tồn tại sau khi được sử dụng để cập nhật mô hình. Do đó, người dùng có thể yên tâm rằng dữ liệu cá nhân của họ không bị lưu trữ hoặc sử dụng một cách không đáng tin cậy.

Một điểm khác biệt quan trọng trong Federated learning là sự phân tán của quá trình. Dữ liệu không cần phải được gửi đến trung tâm dữ liệu tập trung để huấn luyện, mà các thiết bị địa phương sẽ gửi các cập nhật trực tiếp tới quá trình tổng hợp. Điều này giúp đảm bảo rằng dữ liệu cá nhân không bị tiết lộ trong quá trình truyền tải.

**Tối ưu hóa phân tán (Federated Optimization)**: Trong federated learning, chúng tôi gọi vấn đề tối ưu hóa ẩn chứa trong đó là tối ưu hóa phân tán, đồng thời vẽ ra mối liên hệ (và sự khác biệt) so với tối ưu hóa phân tán thông thường. Tối ưu hóa phân tán trong federated learning có một số thuộc tính quan trọng khác biệt so với vấn đề tối ưu hóa phân tán thông thường:

• **Non-IID**: Dữ liệu huấn luyện trên mỗi thiết bị người dùng thường dựa trên việc sử dụng của thiết bị di động bởi mỗi người dùng cụ thể, do đó tập dữ liệu địa phương của mỗi người dùng sẽ không đại diện cho phân phối của toàn bộ quần thể.

• **Unbalanced**: Tương tự, một số người dùng sẽ sử dụng dịch vụ hoặc ứng dụng nhiều hơn các người dùng khác, dẫn đến sự khác biệt về lượng dữ liệu huấn luyện địa phương.

• **Massively distributed**: Chúng tôi dự kiến số lượng thiết bị hoặc máy chủ tham gia vào quá trình tối ưu hóa sẽ lớn hơn nhiều so với số lượng trung bình các ví dụ trên mỗi thiết bị.

• **Limited communication**: Thiết bị di động thường offline hoặc có kết nối chậm hoặc đắt đỏ.

Trong nghiên cứu này, chúng tôi tập trung vào các thuộc tính không đồng nhất và mất cân đối của quá trình tối ưu hóa, cũng như tính quan trọng của các ràng buộc giao tiếp. Một hệ thống tối ưu hóa phân tán triển khai cũng phải giải quyết nhiều vấn đề thực tế: tập dữ liệu của các thiết bị khách thay đổi khi dữ liệu được thêm hoặc xóa; tính sẵn có của khách hàng có liên quan đến phân phối dữ liệu địa phương theo cách phức tạp (ví dụ: điện thoại từ người nói tiếng Anh Mỹ có thể được cắm vào thời gian khác so với người nói tiếng Anh Anh); và khách hàng không phản hồi hoặc gửi các cập nhật bị hỏng.

Chúng ta nghiên cứu đến vấn đề tối ưu hóa trong học máy:

A picture containing text, font, white, line

Description automatically generated

trong đó *f\_i(w) = loss(x\_i, y\_i, w)* là hàm mất mát dự đoán trên ví dụ *(x\_i, y\_i)* được thực hiện với tham số *w* . Giả sử có K máy khách (clients) mà dữ liệu được phân vùng, với P\_k là tập chỉ mục của các điểm dữ liệu trong máy khách k, với *n\_k = |P\_k|*. Do đó, công thức trên có thể được viết lại thành:

A black text on a white background

Description automatically generated with medium confidence

Nếu phân vùng *Pk* được hình thành bằng cách phân phối các ví dụ huấn luyện đều nhau ngẫu nhiên cho các khách hàng, thì chúng ta sẽ có E*Pk*[*Fk*(*w*)] = *f*(*w*) ,trong đó kỳ vọng tính toán trên tập các ví dụ được giao cho một khách hàng cố định. IID (Independent and Identically Distributed) thông thường được sử dụng trong các thuật toán tối ưu phân tán; chúng tôi đề cập đến trường hợp khi điều này không xảy ra (tức là, *Fk* có thể là một xấp xỉ tệ với f) gọi là Non-IID.

Trong tối ưu hóa trung tâm dữ liệu, chi phí giao tiếp tương đối nhỏ và chi phí tính toán là quan trọng nhất, với nhiều nỗ lực gần đây tập trung vào việc sử dụng GPU để giảm chi phí này. Trái lại, trong tối ưu hóa phân tán, chi phí giao tiếp chiếm ưu thế - thông thường chúng ta sẽ bị giới hạn bởi băng thông tải lên là 1 MB/s hoặc thấp hơn. Hơn nữa, khách hàng thông thường chỉ tự nguyện tham gia vào quá trình tối ưu hóa khi họ được sạc pin, kết nối vào mạng Wi-Fi không tính cước và có kết nối internet không giới hạn. Tuy nhiên, vì tập dữ liệu trên mỗi thiết bị là nhỏ so với tổng kích thước tập dữ liệu, và các điện thoại thông minh hiện đại có bộ xử lý tương đối nhanh (bao gồm GPU), tính toán trở nên gần như miễn phí so với chi phí giao tiếp đối với nhiều loại mô hình. Do đó, mục tiêu của chúng tôi là sử dụng tính toán bổ sung để giảm số vòng giao tiếp cần thiết để huấn luyện một mô hình. Có hai cách chính để thêm tính toán:

1) tăng đồng thời, nơi chúng tôi sử dụng nhiều khách hàng làm việc độc lập giữa mỗi vòng giao tiếp.

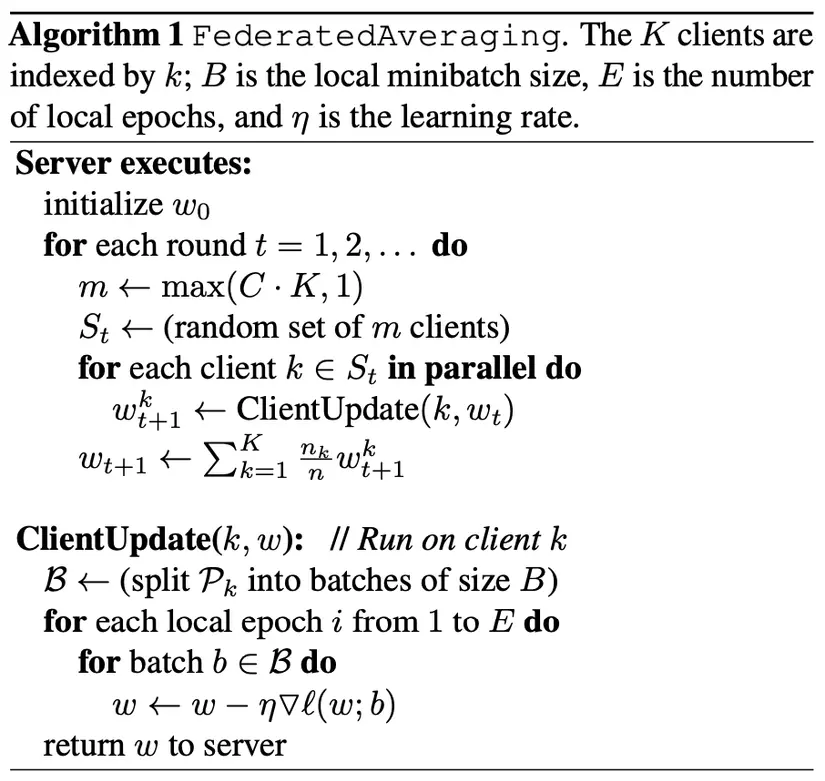
2) tăng tính toán trên mỗi khách hàng, nơi mỗi khách hàng thực hiện một tính toán phức tạp hơn giữa mỗi vòng giao tiếp thay vì chỉ tính toán đơn giản như tính đạo hàm.

**Related work:**

* Huấn luyện phân tán bằng cách lặp lại việc trung bình hóa các mô hình được huấn luyện cục bộ đã được nghiên cứu bởi McDonald et al. cho perceptron.
* Povey et al. cho mạng nơ-ron nhận dạng giọng nói
* Neverova et al. cũng thảo luận về lợi ích của việc giữ dữ liệu nhạy cảm của người dùng trên thiết bị
* Công trình của Shokri và Shmatikov có liên quan theo nhiều cách: họ tập trung vào việc huấn luyện mạng nơ-ron sâu, nhấn mạnh tầm quan trọng của quyền riêng tư, và giải quyết vấn đề chi phí giao tiếp bằng cách chỉ chia sẻ một phần tử con của các tham số trong mỗi vòng giao tiếp; tuy nhiên, họ cũng không xem xét dữ liệu không cân bằng và không đồng nhất, và đánh giá kinh nghiệm hạn chế.
* Trong tập lồi, bài toán tối ưu phân tán và ước lượng đã nhận được sự quan tâm đáng kể và một số thuật toán tập trung cụ thể vào hiệu quả truyền thông. Ngoài ra để giả định tính lồi, công việc hiện tại này thường yêu cầu rằng số lượng khách hàng nhỏ hơn nhiều so với số lượng của các ví dụ cho mỗi khách hàng, rằng dữ liệu được phân phối trên các máy khách theo kiểu IID và mỗi nút có số lượng điểm dữ liệu giống hệt nhau tất cả các giả định này bị vi phạm trong cài đặt tối ưu hóa có liên kết.

1. **The Federated Averaging (FedAvg) Algorithm**

* Chúng tôi đề xuất phương pháp Tối ưu liên kết (Federated optimization). Phương pháp này mô tả cụ thể những việc mà server và client sẽ phải thực hiện và thông tin được truyền đi giữa server và client. Độ hiệu quả của phương pháp này trực tiếp quyết định đến tính riêng tư và số vòng(round) cần update để thu được độ chính xác nhất mong muốn. Cụ thể chúng tôi giới thiệu giải thuật **FederatedAveraging**, giải thuật này thực hiện tối ưu trên client bằng stochastic gradient descent(SGD) và thực hiện lấy trung bình mô hình ở phía server. Chi tiết giải thuật được mô tả dưới đây:



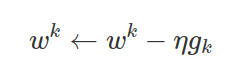
* Thuật toán được thực hiện theo các bước như sau:

1. Server: Khởi tạo tham số w\_0
2. Server: chọn ngẫu nhiên các client tham gia vào việc cải thiện mô hình (điều khiển số lượng client bằng tham số *C*, tổng số lượng client là *K*, và gửi tham số mô hình hiện tại đến client.
3. Các client: thực hiện tối ưu mô hình nhận được từ server trên tập dữ liệu cục bộ bằng SGD. (số epoch E, mini-batch size B)
4. Các client: gửi lại tham số mới của mô hình cho server
5. Server: tính trung bình các tham số mới nhận được từ các client với trọng số bằng số lượng mẫu dữ liệu tương ứng trên client.
6. Quay lại bước 2 để tiếp tục round mới.

* Với tốc độ học cố định η, mỗi máy khách k tính toán:



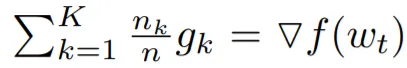
* Mỗi clients(k) huấn luyện trên dữ liệu local bằng phương pháp SGD với batch size B để thu được gradient g\_k, với learning rate η.



* Average gradient trên dữ liệu cục bộ của nó ở mô hình hiện tại w\_t , và máy chủ trung tâm tổng hợp các gradients này:



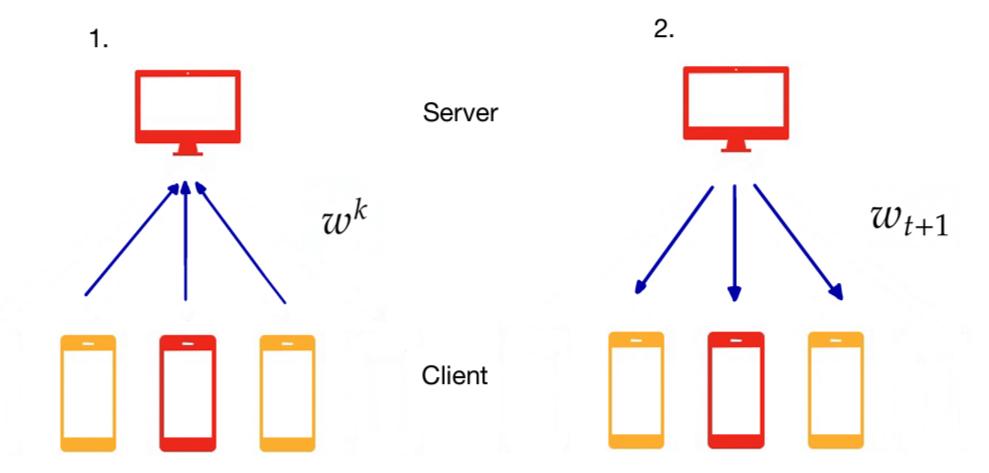
* Một bản cập nhật tương đương được đưa ra:



* Sau E epoch, các trọng số w k được cập nhật sẽ được gửi đến máy chủ. Sau đó, máy chủ kết hợp trọng số của các máy khách để có được một mô hình mới w t+1 . Điều này được thực hiện bằng cách gán trọng số cao hơn cho các máy khách có phần dữ liệu n k / K lớn hơn



* Với 3 tham số kiểm soát quá trình tính toán:
  + C: tỉ lệ client tham gia thực hiện tính toán mỗi vòng
  + E: số lần đào tạo mà mỗi client thực hiện trên tập dữ liệu cục bộ của mình trên mỗi vòng
  + B: kích thước sử dụng để cập nhật client
* Mô hình này sau đó được phân phối lại cho khách hàng để đào tạo thêm. Giải thích trên được hình dung trong hình bên dưới.



1. **Kết quả thực nghiệm**

* Thử nghiệm mô hình với bộ nhận dữ liệu **MNIST** và **CIFAR10**:

1. Một mạng neural multilayer-perceptron(MLP) với 2 lớp ẩn, mỗi lớp có 200 đơn vị sử dụng hàm kích hoạt ReLU (MNIST 2NN)
2. Một mạng neural tích chập (CNN) với hai lớp tích chập kích thước 5x5, một lớp kết nối đầy đủ với 512 đơn vị, hàm kích hoạt ReLU, và một lớp đầu ra softmax cuối cùng.

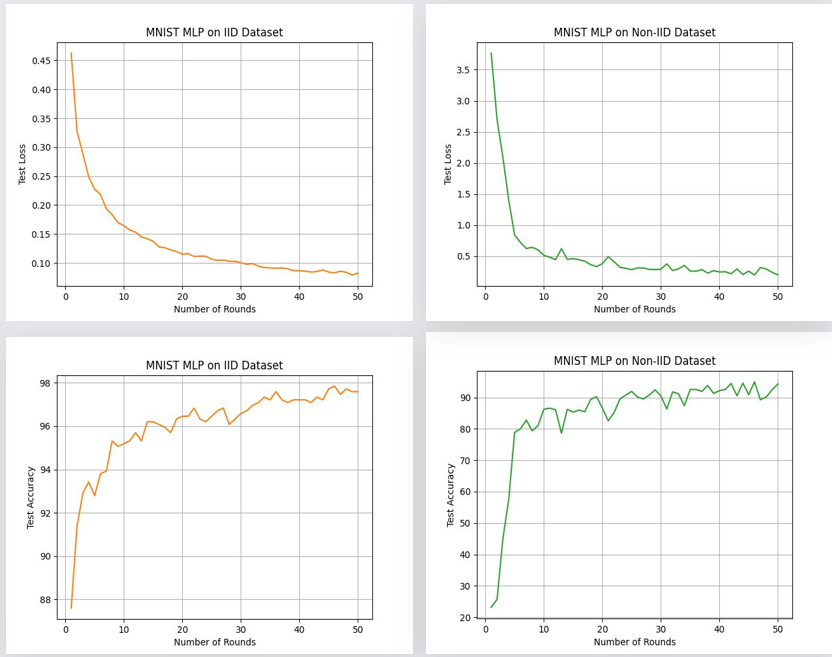
* Chúng tôi nghiên cứu hai cách phân vùng dữ liệu trên clients:

**1. IID** (Independent and Identically Distributed - độc lập và phân phối đồng nhất):

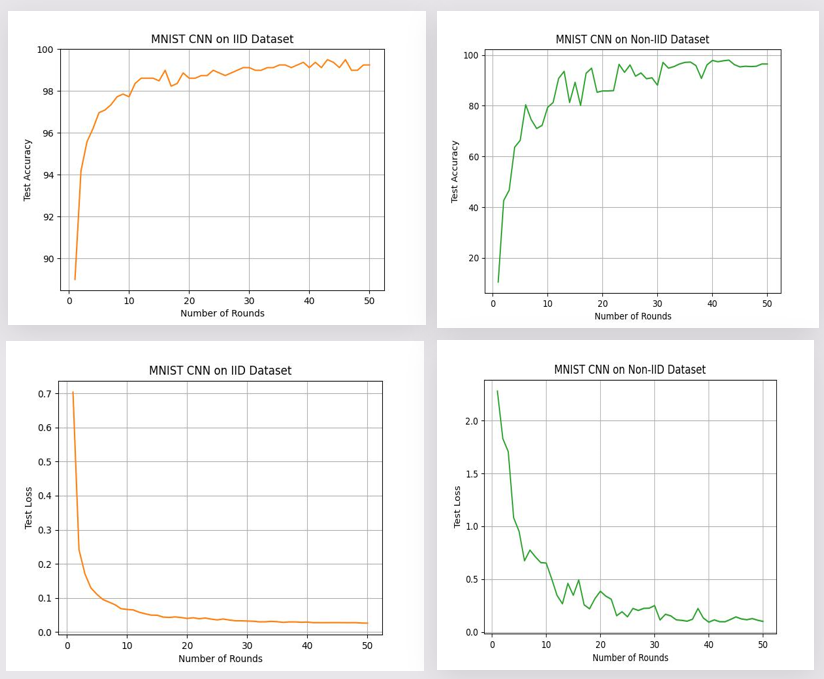
* + - Dữ liệu được xáo trộn (shuffled).
    - Sau đó, dữ liệu được chia thành 100 máy khách (clients), mỗi máy khách nhận 600 ví dụ.

**2. Non-IID** (Non-Independent and Identically Distributed - không độc lập và phân phối đồng nhất):

* + - Ban đầu, dữ liệu được sắp xếp theo nhãn chữ số (digit label).
    - Sau đó, dữ liệu được chia thành 200 phần (shards) có kích thước là 300 ví dụ mỗi phần.
    - Mỗi trong số 100 máy khách sẽ được gán 2 phần dữ liệu.
* Result training Mnist\_2NN with 50 epoch, C = 0.1, K = 100, E = 1, B = 10



* Results training Mnist\_CNN with 50 epoch, C = 0.1, K = 100, E = 1, B = 10

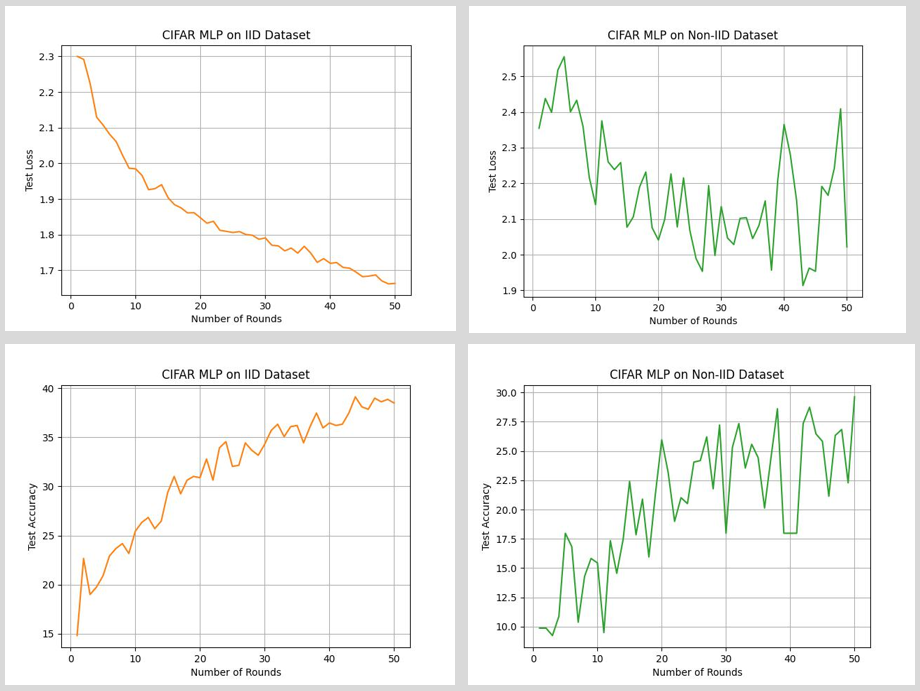


* Test Accuracy cho 2 mô hình 2NN và CNN với tập dữ liệu Mnist

|  |  |  |
| --- | --- | --- |
|  | IID | Non-IID |
| 2NN | **97.595%** | **94.304%** |
| CNN | **99.241%** | **96.456%** |

Với tập dữ liệu Mnist cả 2 mô hình 2NN và CNN đều hoạt động tốt trên IID và Non-IID

* Result training Cifar\_MLP with 50 epoch, C = 0.1, K = 100, E = 1, B = 10



* Result training Cifar\_CNN with 50 epoch, C = 0.1, K = 100, E = 1, B = 10

A picture containing text, line, handwriting, plot

Description automatically generated

* Test Accuracy cho 2 mô hình 2NN và CNN với tập dữ liệu Cifar10

|  |  |  |
| --- | --- | --- |
|  | IID | Non-IID |
| MLP | 38.481% | 29.620% |
| CNN | 60.633% | 43.797% |

* Ngoài ra chúng tôi còn thực hiện trên bộ dữ liệu Shakespeare trên mô hình LSTM với lr =0.8

A graph showing the number of rounds

Description automatically generated with low confidenceA graph with a green line

Description automatically generated with low confidence

1. **Kết luận**

* Trong bài báo này, chúng đã đưa ra mô hình Học liên kết và chứng minh được tính hiệu quả của nó trong bài toán thực tế thông qua các thí nghiệm so sánh.
* Federated learning mang lại nhiều lợi ích, bao gồm việc bảo vệ quyền riêng tư dữ liệu người dùng bởi vì dữ liệu không cần phải được chia sẻ hoàn toàn. Ngoài ra, federated learning cũng giúp giảm băng thông mạng và tải công việc trên máy chủ trung tâm. Nó cũng cho phép huấn luyện mô hình trên dữ liệu đa dạng từ các nguồn khác nhau mà không cần truyền dữ liệu về một điểm tập trung
* Về mặt hiệu quả giao tiếp giữa client và server. Những kết quả thí nghiệm cũng chỉ ra được sự cải thiện đáng kể so với các phương pháp khác.

1. **Reference**

* Martin Abadi, Andy Chu, Ian Goodfellow, Brendan  
  McMahan, Ilya Mironov, Kunal Talwar, and Li Zhang.  
  Deep learning with differential privacy. In *23rd ACM  
  Conference on Computer and Communications Security (ACM CCS)*, 2016.
* Yossi Arjevani and Ohad Shamir. Communication  
  complexity of distributed convex learning and optimization. In *Advances in Neural Information Processing Systems 28*. 2015.
* Yoshua Bengio, Rejean Ducharme, Pascal Vincent, ´  
  and Christian Janvin. A neural probabilistic language  
  model. *J. Mach. Learn. Res.*, 2003.
* Keith Bonawitz, Vladimir Ivanov, Ben Kreuter, Antonio Marcedone, H. Brendan McMahan, Sarvar Patel,  
  Daniel Ramage, Aaron Segal, and Karn Seth. Practical  
  secure aggregation for federated learning on user-held  
  data. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
* David L. Chaum. Untraceable electronic mail, return  
  addresses, and digital pseudonyms. *Commun. ACM*,  
  24(2), 1981.
* ianmin Chen, Rajat Monga, Samy Bengio, and Rafal  
  Jozefowicz. Revisiting distributed synchronous sgd.  
  In *ICLR Workshop Track*, 2016.
* Anna Choromanska, Mikael Henaff, Michael Mathieu, ¨  
  Gerard Ben Arous, and Yann LeCun. The loss surfaces ´  
  googleresearch.blogspot.com/2015/11/computer-respond-to-this-email.html,November 2015.
* Yann N. Dauphin, Razvan Pascanu, C¸ aglar Gul ¨ c¸ehre,  
  KyungHyun Cho, Surya Ganguli, and Yoshua Bengio.  
  Identifying and attacking the saddle point problem in  
  high-dimensional non-convex optimization. In *NIPS*,  
  2014.
* Jeffrey Dean, Greg S. Corrado, Rajat Monga, Kai  
  Chen, Matthieu Devin, Quoc V. Le, Mark Z. Mao,  
  Marc’Aurelio Ranzato, Andrew Senior, Paul Tucker,  
  Ke Yang, and Andrew Y. Ng. Large scale distributed  
  deep networks. In *NIPS*, 2012
* John Duchi, Michael I. Jordan, and Martin J. Wainwright. Privacy aware learning. *Journal of the Association for Computing Machinery*, 2014.
* Cynthia Dwork and Aaron Roth. *The Algorithmic  
  Foundations of Differential Privacy*. Foundations and  
  Trends in Theoretical Computer Science. Now Publishers, 2014.
* Olivier Fercoq, Zheng Qu, Peter Richtarik, and Martin ´  
  Takac. Fast distributed coordinate descent for non- ´  
  strongly convex losses. In *Machine Learning for  
  Signal Processing (MLSP), 2014 IEEE International  
  Workshop on*, 2014.
* Ian Goodfellow, Yoshua Bengio, and Aaron Courville.  
  Deep learning. Book in preparation for MIT Press,  
  2016.
* Ian J. Goodfellow, Oriol Vinyals, and Andrew M. Saxe.  
  Qualitatively characterizing neural network optimization problems. In *ICLR*, 2015.
* Slawomir Goryczka, Li Xiong, and Vaidy Sunderam.  
  Secure multiparty aggregation with differential privacy: A comparative study. In *Proceedings of the  
  Joint EDBT/ICDT 2013 Workshops*, 2013.
* Benjamin Graham. Fractional max-pooling. *CoRR*,  
  abs/1412.6071, 2014. URL http://arxiv.org/  
  abs/1412.6071.
* Sepp Hochreiter and Jurgen Schmidhuber. Long short- ¨  
  term memory. *Neural Computation*, 9(8), November  
  1997.
* Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing  
  internal covariate shift. In *ICML*, 2015.
* Yoon Kim, Yacine Jernite, David Sontag, and Alexander M. Rush. Character-aware neural language models.  
  *CoRR*, abs/1508.06615, 2015.
* Jakub Konecn ˇ y, H. Brendan McMahan, Felix X. Yu, ´  
  Peter Richtarik, Ananda Theertha Suresh, and Dave  
  Bacon. Federated learning: Strategies for improving  
  communication efficiency. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
* Alex Krizhevsky. Learning multiple layers of features  
  from tiny images. Technical report, 2009.
* Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. Imagenet classification with deep convolutional  
  neural networks. In *NIPS*. 2012.
* Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner.  
  Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 1998.
* Chenxin Ma, Virginia Smith, Martin Jaggi, Michael I  
  Jordan, Peter Richtarik, and Martin Tak ´ a´c. Adding vs. ˇ  
  averaging in distributed primal-dual optimization. In  
  *ICML*, 2015.