

THÔNG TIN CHUNG CỦA BÁO CÁO

- Họ và Tên: Phạm Xuân Vĩnh Hà
- MSSV: 230201041



- Lớp: CS2205.APR2023
- Tự đánh giá (điểm tổng kết môn): 6/10
- Số buổi vắng: 0
- Số câu hỏi QT cá nhân: 0
- Link Github:
<https://github.com/HaPXV/CS2205.APR2023/>
- Link Youtube:
<https://youtu.be/u1oW2eAvjUA>

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

HỌC HỎI TỪ CÁC THÀNH VIÊN ĐÃ XÁC NHẬN TRONG HỆ THỐNG HỌC TẬP PHÂN TÁN

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

LEARNING FROM ACKNOWLEDGEMENTS IN DISTRIBUTED LEARNING

TÓM TẮT (*Tối đa 400 từ*)

Học phân tán (Distributed Learning) là một lĩnh vực trong Học máy (Machine Learning) tập trung vào việc huấn luyện các mô hình Machine Learning trên nhiều máy tính hoặc thiết bị khác nhau, thay vì chỉ trên một máy tính duy nhất. Vấn đề này giúp giải quyết việc thiếu tài nguyên khi xử lý các tập dữ liệu lớn hoặc mô hình Machine Learning phức tạp. Thêm vào đó, Việc chia nhỏ mô hình học máy và phân tán trên nhiều máy tính giúp giảm thời gian huấn luyện mô hình, có thể dễ dàng mở rộng bằng cách thêm nhiều máy tính vào hệ thống, phù hợp cho việc xử lý các tập dữ liệu ngày càng lớn nên giúp tiết kiệm chi phí cho hệ thống. Trong đề tài này, sẽ tập trung nghiên cứu sử dụng học hỏi từ các xác nhận trong Học Phân tán" (Learning from Acknowledgements in Distributed Learning) là một cách tiếp cận mới mẻ khai thác triệt để thông tin nhận được từ các máy tính thành viên. Bằng cách phân tích những tín hiệu này, hệ thống có thể học hỏi và cải thiện hiệu quả của quá trình học tập phân tán giúp đánh giá và chứng minh tính hiệu quả của bài nghiên cứu này.

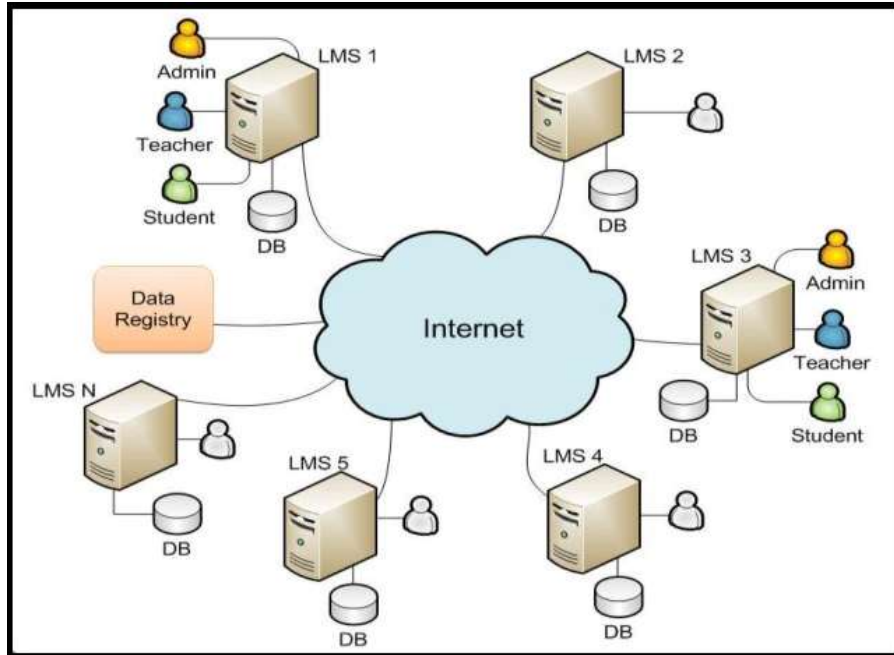
GIỚI THIỆU (*Tối đa 1 trang A4*)

Học tập phân tán (Distributed Learning) đã trở thành mấu chốt cho sự phát triển của các mô hình học máy phức tạp, cho phép phân tán việc đào tạo trên nhiều máy tính. Tuy nhiên, đảm bảo hiệu quả và độ tin cậy trong quá trình học tập phân tán vẫn là một thách thức lớn. Các hệ thống học tập phân tán hiện tại chủ yếu tập trung vào việc theo dõi tín hiệu hoàn thành tác vụ từ các máy tính thành viên (worker). Tuy nhiên, cách tiếp cận này bỏ qua một nguồn thông tin tiềm năng vô giá: các thành viên đã được Xác nhận

(Acknowledgements).

Trong đề tài này, sẽ tập trung nghiên cứu sử dụng học hỏi từ các xác nhận trong Học tập Phân tán" (Learning from Acknowledgements in Distributed Learning) là một cách tiếp cận mới mở khai thác triệt để thông tin từ các xác nhận nhận được từ các máy tính thành viên như hình minh họa dưới đây.

Bằng cách phân tích những tín hiệu này, hệ thống có thể học hỏi và cải thiện hiệu quả của quá trình học tập phân tán. Để đánh giá và chứng minh tính hiệu quả của nó.



Hình minh họa Distributed Learning

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

- Nghiên cứu và phát triển mô hình học hỏi từ các xác nhận trong Học tập Phân tán" (Learning from Acknowledgements in Distributed Learning).
- Đánh giá hiệu quả của mô hình học hỏi từ các xác nhận trong Học tập Phân tán" (Learning from Acknowledgements in Distributed Learning) nhằm tối ưu hóa toàn bộ quá trình học phân tán.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

Nội dung: Tìm hiểu cách khai thác thông tin từ các xác nhận (acknowledgements) nhận được từ các máy tính thành viên (worker machines) như:

- Thời gian hoàn thành tác vụ: Phản ánh hiệu suất của từng máy tính thành viên.
- Tình trạng lỗi (nếu có): Cho phép phát hiện sớm các sự cố tiềm ẩn và ngăn chặn lỗi nghiêm trọng.
- Các chỉ số về hiệu suất: Cung cấp thông tin chi tiết về tài nguyên sử dụng, mức độ tiêu thụ năng lượng, ... của từng máy tính. Cung cấp thông tin chi tiết về hoạt động của từng máy tính.

Phương pháp:

Thu thập xác nhận: Hệ thống thu thập các xác nhận từ các máy tính thành viên sau khi chúng hoàn thành nhiệm vụ được giao.

Trích xuất thông tin: Hệ thống trích xuất các thông tin có giá trị từ nội dung của xác nhận, chẳng hạn như thời gian hoàn thành, trạng thái lỗi, và các chỉ số về hiệu suất.

Phân tích dữ liệu: Dữ liệu được phân tích bằng các phương pháp thống kê hoặc học máy để tìm ra các mẫu hình, mối tương quan và các vấn đề tiềm ẩn.

Ra quyết định: Dựa trên kết quả phân tích, hệ thống có thể đưa ra các quyết định để cải thiện hiệu quả và độ tin cậy của quá trình học tập phân tán. Ví dụ:

- Xác định máy tính yếu kém: Nếu một máy tính thành viên thường xuyên gửi xác nhận với thời gian hoàn thành lâu hoặc báo lỗi, hệ thống có thể nhận dạng đó là máy tính yếu kém và phân bổ lại tác vụ cho các máy tính khác mạnh hơn.
- Phát hiện lỗi tiềm ẩn: Bằng cách theo dõi các mẫu trễ hoặc lỗi trong xác nhận, hệ thống có thể dự đoán các sự cố nghiêm trọng hơn trước khi chúng xảy ra.
- Tối ưu hóa phân bổ tác vụ: Hệ thống có thể phân bổ tác vụ cho các máy tính phù hợp dựa trên năng lực của từng máy tính, giúp giảm thời gian đào tạo tổng thể.

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

Mô hình học hỏi từ các xác nhận trong Học tập Phân tán" (Learning from Acknowledgements in Distributed Learning) dự kiến mang lại những kết quả như:

- *Giảm thời gian đào tạo.*
- *Nâng cao độ chính xác và tính ổn định của mô hình*
- *Cải thiện khả năng mở rộng*
- *Giảm thiểu chi phí tính toán*
- *Tăng tính minh bạch*

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

- [1]. Song, C., Kang, K. Y., Timakum, T., & Zhang Examining influential factors for acknowledgements classification using supervised learning. PLOS ONE, 2020.
- [2]. Konečný, J., McMahan, H. B., Yu, F., Pouget-Abadie, Federated learning with on-device differential privacy, 2016.
- [3]. Recht, B., Mao, X., Nugent, F., & Onnes, Gradient boosting on federated systems, 2018.
- [4] Li, T., Sahu, A. K., Talwar, A., Bao, J., & Schaul, Federated learning: Challenges, methods, and future directions, 2020 , 50-60.