CERTAIN INVESTIGATIONS ON DISTRIBUTED LEARNING PERFORMANCE ENHANCEMENTS IN FEDERATED MULTITASK LEARNING

A THESIS

Submitted by

KUMARESAN M

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

FACULTY OF SCIENCE AND HUMANITIES

ANNA UNIVERSITY

CHENNAI 600 025

NOVEMBER 2023

ABSTRACT

Federated learning enables machine learning models to learn from decentralized data without compromising privacy. The standard formulation of federated learning produces a shared model for all users. Due to statistical heterogeneity and the non-IID distribution of data across devices often by users, the local models trained solely on their private data perform better than the global shared model, which will take away their incentive to participate in the process. Several techniques have been proposed to personalize global models to work better for individual users. As solutions, personalized federated learning and Federated Multitask Learning (FMTL) have been proposed to handle the statistical diversity in FL. Personalized FL aims to build a global model, which should be an advantage to finding a personalized model that is stately for each users data. Here, the global model considered an agreed point for each user to start personalizing its local model based on its heterogeneous data distribution. FMTL directly addresses the challenge of statistical diversity in FL by learning simultaneously separate models for each user. In FMTL, the users models are separated but typically correlated, since users with similar features are likely to share similar behaviors. The relationships among the users models are to be captured by a regularization term, which is minimized to encourage the correlated users models to be mutually impacted. Constituting a large amount of disease-related data from heterogeneous devices in personalized models can be learned by using Federated Multitask Learning (FMTL). Due to system and statistical heterogeneity, a personalized model has been studied by Federated Multitask Learning (FMTL) to predict the updated infection rate of COVID-19 in the USA using a mobility-based SEIR model. Mobility-based SEIR model with an additional constraint, we can analyze the availability of beds. Here, we used real-time mobility data sets in various states of the USA during the years

2020 and 2021. We have chosen five states for the study and observe that a correlation among the number of COVID-19-infected cases exists, even though the rate of spread in each case is different.

Online FMTL framework is learning the parameter for newly joining devices dynamically in the FMTL computing environment without revisiting all the devices. But the performance of the existing multitask approaches would largely degenerate when dealing with polluted data, i.e., outliers. The proposed methodology is a robust online FMTL with an adaptive loss (OFMTL-AL) algorithm. OFMTL-AL is addressing the issue of polluted data. Finally, the numerical results show that OFMTL-AL outperforms the existing online FMTL algorithm.

Online Federated Multitask Learning (FMTL), where users perform model learning on streaming data and communicate the model to the server; however, they do not address the associated communication overhead. As a solution, the partial-sharing-based framework for online federated multitask learning, called PSO-FMTL, where users update local models from streaming data and exchange bitty fractions of the model with the server, reduce communication overhead. In contrast to classical federated multitask learning approaches, the proposed strategy provides users, who are not part of a global iteration with the freedom to update local models whenever new data arrives. In the federated implementation, the usage of the Random Fourier feature (RFF) avoids the requirement of a centralized pre-trained dictionary in the case of coherence-check. RFF is used to reduce the dictionary size. The proposed model is an event-triggered Partial Sharing-based Online FMTL (PSO-FMTL) that further reduces the computational burden of users while enhancing communication efficiency. Numerical results illustrate that PSO-FMTL has reduced communication overhead.

In federated optimization, a single owner-follower game builds an incentive framework for the federated learning and analytics system with multiple tasks. However, the convergence rate of distributed learning suffers from heterogeneous worker performance. In the proposed Optimal strategy, the owner at the upper layer decides its reward rate to incentive data owners to participate in the federated learning system. Then, based on the accuracy level of local tasks, the upper layer determines the reward rates for tasks with limited resources. The model owner should judiciously decide on the number of the data owner (e.g., mobile device or organization) at the lower layer due to the trade-off between the diversity provided by the number of data owners and the latency of completing the training. Further, we analyze the optimal strategies for the model through some numerical results.