효율적인 통신을 위한 연합 학습 환경에서의 최적화 연구: 최신 동향과 포괄적 조사

김재민, 이동현, 오준석, 원동욱, Ayalneh Bitew Wondmagegn, 조혜지, 조성래
중앙대학교
{jmkim, dhlee, jsoh, dwwon, ayalneh, hjjo}@uclab.re.kr, srcho@cau.ac.kr

Optimization in federated learning for efficient communication: Recent trends and comprehensive survey

Jaemin Kim, Donghyun Lee, Junsuk Oh, Dongwook Won, Ayalneh Bitew Wondmagegn, Hyeji Jo, Sungrae Cho
Department of Computer Science and Engineering, Chung-Ang University

요 약

온라인 학습 기반 최적화는 연합 학습의 원칙과 온라인 학습을 결합하여 동적 데이터 분포와 지속적인 학습 시나리오를 처리하는 새로운 접근 방식이다. 본 논문에서는 연합 학습 환경에서의 효율적인 통신을 위한 온라인 학습 최적화를 목표로 온라인 연합 학습의 최신 동향을 포괄적으로 조사하며, 기본 개념, 주요 방법론, 도전 과제 및 응용 분야를 다룬다.

Ⅰ. 서 론

연합 학습은 여러 엣지 장치에 분산된 데이터를 사용하여 모델을 학습하는 기계 학습 패러다임으로, 데이터 프라이버시를 보호한다. 기존의 연합 학습은 정적 환경에서 학습이 이루어지지만 실제 환경에서는 지속적으로 변화하는 데이터 스트림이 존재한다. 이러한 변화에 대응하기 위해 온라인 학습 기반의 연합 학습 최적화가 필요하다. 본 연구에서는 연합 학습을 동적 환경에 맞게 확장하여 효율적인 통신을 유지하면서 비정상적인 데이터에서도 모델 성능을 최적화하고자 한다.

Ⅱ. 본론

연합 학습(FL)은 여러 클라이언트가 데이터를 로컬에 유지하면서 협력하여 모델을 학습함으로써 데이터 프라이버시를 강화하는 방법이다. 예를 들어, 스마트폰이나 IoT 장치 등이 협력하여 데이터 공유 없이 모델을 학습한다. 온라인 학습은 새로운 데이터가 도착할 때마다 모델을 점진적으로 학습하여 시간에 따른 변화에 적응하는 방식이다[1]. 이러한 온라인 학습의 개념을 연합 학습에 적용하여 온라인 연합 학습 최적화를 가능하게 한다. 이는 엣지 장치에서 새로운 데이터를 지속적으로 학습하며 효율적인 통신을 유지하면서 비정상적인 데이터 분포를 처리할 수 있는 분산된 연합 모델을 구축하는 것을 목표로 한다.

연합 학습 환경에서 효율적인 통신을 위해서는 클라이언트 선택 전략이 매우 중요하다. 각 학습 라운드에서 적절한 클라이언트를 선택하여 중요한 데이터를 활용하고 통신 오버헤드를 최소화할 수 있다[2]. 또한 온라인 연합 학습에서는 클라이언트에서 로컬로 업데이트된 모델을 집계하는 과정이 필수적이며, 이를 위해 주로 FedAvg와 같은 집계 방법을 사용하더라도 빈번한 업데이트에 맞게 수정하는 것이 필요하다[3]. 클라이언트 간 데이터 분포가 시간에 따라 변할 수 있기 때문에, 이러한 비독립적이고 동일분포가 아닌 데이터 처리도 중요한 과제이다[4]. 이를 해결하기 위해 동적 집계 및 적응형 학습률과 같은 기법이 사용된다.

온라인 학습 기반의 연합 학습 최적화에서 중요한 도전 과제 중 하나는 통신 효율성이다. 빈번한 모델 업데이트는 상당한 통신 오버헤드를 유발할 수 있으며, 특히 클라이언트가 지속적으로 업데이트를 전송해야 할 경우 문제가 될 수 있다. 이러한 문제를 해결하기 위해 그래디언트 희소화나 주기적 집계와 같은 기법이 적용될 수 있다[5]. 또한 연합 학습은 본질적으로 프라이버시를 보호하지만, 온라인 학습의 지속적인 특성은 더 많은 취약점을 노출할 수 있다. 안전한 집계와 차등 프라이버시 기법을 적용하여 이러한 문제를 해결해야 한다[6],[7]. 데이터 분포가 시간에 따라 급격히 변할 수 있기 때문에 메타 학습 및 적응형 메커니즘과 같은 기법을 사용하여 모델의 견고성을 유지하는 것도 필요하다[8].

이러한 최적화 기법들은 다양한 응용 분야에서 활용될 수 있다. 스마트폰이나 웨어러블 장치와 같이 데이터가 동적으로 변화하는 엣지 장치에서 모델을 지속적으로 학습하는 데 온라인 학습 기반의 연합 학습 최적화가 유용하다. 자율 주행 차량의 경우에도 연합 학습을 통해 중앙 집중식 데이터 수집 없이 변화하는 환경과 조건에 적응할 수 있다. 또한 헬스케어 분야에서는 웨어러블 건강 장치를 통해 사용자 데이터 프라이버시를 보호하면서도 지속적으로 개인 맞춤형 건강 모니터링을 개선할 수 있는 온라인 학습 기반의 연합 학습이 활용될 수 있다.

Ⅲ. 결론

온라인 학습 기반의 연합 학습 최적화는 동적 데이터 환경의 문제를 해결함으로써 분산 학습 분야에서 중요한 발전을 나타낸다. 통신 효율성, 프라이버시, 비정상적인 데이터에 대한 견고성과 같은 도전 과제를 해결해야 그 잠재력을 최대한 발휘할 수 있다. 이를 위해 학습을 확장 가능하고 신뢰할 수 있도록 최적화 전략과 새로운 집계 기법을 탐구하는 지속적인 연구가 필요하다. 향후 연구 방향으로는 다음과 같은 주제가 포함된다. 첫째, 클라이언트 데이터 특성에 따라 동적으로 조정될 수 있는 적응형 집계 메커니즘의 탐구가 필요하다. 둘째, 최소한의 통신 오버헤드로 대규모 클

라이언트를 처리할 수 있는 확장성 개선 방안이 중요하다. 셋째, 빈번한 업데이트 환경에서의 보다 견고한 프라이버시 보호 알고리즘 개발이 필요하다. 이러한 연구들은 온라인 연합 학습의 성능을 향상시키고 실제 응용에서의 신뢰성을 높이는 데 기여할 것으로 기대된다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2024-RS-2022-00156353) 및 2024년도 정부(과학기술정보통신부)의 재원으로 한국연구재단(RS-2024-00453301)의 지원을 받아 수행된 연구임.

참 고 문 헌

[1] Krawczyk, Bartosz, et al. "Ensemble learning for data stream analysis: A survey." Information Fusion 37 (2017): 132-156.

[2] Haddadpour, Farzin, et al. "Federated learning with compression: Unified analysis and sharp guarantees." International Conference on Artificial Intelligence and Statistics. PMLR, 2021.

[3] Karimireddy, Sai Praneeth, et al. "Scaffold: Stochastic controlled averaging for federated learning." International conference on machine learning. PMLR, 2020.

[4] Zhu, Hangyu, et al. "Federated learning on non-IID data: A survey." Neurocomputing 465 (2021): 371-390.

[5] Xue, Ye, Liqun Su, and Vincent KN Lau. "FedOComp: Two-timescale online gradient compression for over-the-air federated learning." IEEE Internet of Things Journal 9.19 (2022): 19330-19345.

[6] Lyu, Lingjuan, et al. "Privacy and robustness in federated learning: Attacks and defenses." IEEE transactions on neural networks and learning systems (2022).

[7] Geyer, Robin C., Tassilo Klein, and Moin Nabi. "Differentially private federated learning: A client level perspective." arXiv preprint arXiv:1712.07557 (2017).

[8] Liu, Sanmin, et al. "Online active learning for drifting data streams." IEEE Transactions on Neural Networks and Learning Systems 34.1 (2021): 186-200.