

학습 데이터 평준화를 위한 선택적 연합학습 기법

김채현, 양영준, 마누르아즈말, 디팍싱, 시디카아예샤, 김동균

경북대학교

hyeon917@knu.ac.kr, youngj719@knu.ac.kr, mahnoor.ajmal@knu.ac.kr, deepak.singh@knu.ac.kr,
Asiddiq@knu.ac.kr, dongkyun@knu.ac.kr

Selective Federated Learning for Balanced Training Data Distributions

Chaehyeon Kim, Youngjoon Yang, Mahnoor Ajmal, Deepak Singh, Ayesha Siddiqua,

Dongkyun Kim

Kyungpook National University

요약

첨단 교통 기술의 발전으로 지능형 교통 시스템(Intelligent Transport System, ITS) 환경에서 실시간 데이터 처리의 중요성이 대두되고 있다. 도로변 유닛(Roadside Unit, RSU)은 각 위치에서 수집한 로컬 데이터를 활용하여 객체 인식 모델 학습을 진행한다. 다수의 RSU가 분산된 환경에서는 원본 데이터가 아닌 모델 파라미터를 서버로 전송하여 글로벌 모델 업데이트를 수행하는 연합학습 기법이 연구되고 있으며, 기존 연합학습 기법에서는 글로벌 모델 업데이트 과정에서 RSU 간 상이한 연산 성능으로 인한 집계 지연(Aggregation delay)이 증가하는 문제가 발생한다. 본 문제를 해결하기 위해 특정 RSU를 선택하여 학습에 참여시키는 선택적 연합학습이 제안되었지만, 연산 성능 중심의 RSU 선택으로 인해 특정 객체 분포에 편향된 학습과 객체별 학습 불균형 문제를 초래하였다. 본 논문에서는 RSU의 객체별 학습 정확도 및 데이터 편차를 고려하여 학습에 참여할 RSU를 선정해 객체 불균형을 완화하는 선택적 연합학습 기법을 제안한다.

I. 서론

최근 자율주행 등 첨단 교통 기술의 발전으로 인해 ITS의 중요성이 강조되고 있다. 이러한 ITS 환경에서는 교통안전 및 효율성을 보장하기 위해 저지연 및 고신뢰의 실시간 데이터 처리와 의사결정이 요구된다[1]. RSU는 ITS의 핵심 구성 요소로서 교차로의 신호등과 같은 도로변 시설에 설치되며 다양한 센서, 통신 모듈, 연산 장치가 탑재되어 있다. 차량 및 보행자를 인식하고 해당 객체의 위치, 방향, 속도, 가속도 등을 기반으로 차량 간 또는 차량과 보행자 간 충돌 가능성을 계산하고, 충돌이 예측될 경우 차량에게 실시간으로 경고 메시지를 전송한다[2].

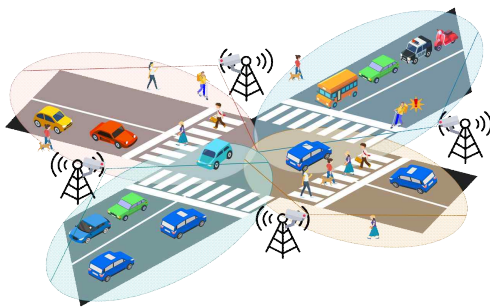


그림 1. RSU 기반 지능형 교통 시스템(ITS)

그림 1과 같이 ITS 환경에서 RSU는 교통 상황과 위치에 따라 수집하는 객체의 종류 및 비율이 상이하며 RSU 간 객체 분포가 이질적으로 나타난다. 각 RSU는 특정 도로 상황에 편향된 로컬 데이터 기반으로 학습을 수행하므로 객체 특성의 다양성이 제한되어 다양한 도로 상황을 반영하기 어렵다. 이를 해결하기 위해 RSU와 중앙 서버가 객체 인식 모델의 학습을 분산하는 연합학습 기법이 제안되었다[3]. 이 기법은 각 RSU에서 수집한 데이터를 통해 로컬 학습을 진행한 후 학습 결과물인 모델 파라미터를 중앙 서버에 전송한다. 중앙 서버는 로컬 모델 파라미터를 집계한 뒤

글로벌 모델을 업데이트하여 글로벌 모델 파라미터를 학습에 참여한 모든 RSU에게 배포한다. 이를 통해 원본 데이터를 외부로 노출시키지 않고 모델 파라미터만을 전송하여 통신 지연이 감소하고 프라이버시를 보호할 수 있다.

RSU들의 설치 시기가 다르고 여러 지역에 분산되어 운영되므로 RSU 간 하드웨어 사양, 연산 성능, 통신 환경이 이질적으로 나타난다. 최근 연구에서는 각 RSU의 연산 성능 및 자원을 고려하여 학습에 참여할 RSU를 선택하는 자원 인지 기반 선택적 연합학습을 제안하였다[4]. 해당 기법은 연산 성능이 높은 RSU 위주로 선택하고 연산 성능이 낮은 RSU를 학습에 제외하여 집계 지연을 감소시켰다. 그러나 해당 기법은 학습 참여 공정성을 고려하지 않아 특정 RSU들만 지속적으로 선택하는 결과를 초래했다.

연산 성능을 기준으로 RSU를 선택하는 구조에서는 높은 연산 성능을 가진 일부 RSU만 반복적으로 학습에 참여하게 된다. 이러한 특정 RSU의 반복 선택은 해당 RSU가 수집한 특정 객체 데이터만 과도하게 학습하며 특정 클래스의 과적합으로 인한 데이터 편향이 발생한다[5]. 또한 소수 객체 데이터를 수집한 RSU가 지속적으로 선택되지 않을 경우, 해당 객체에 대한 학습이 부족하여 글로벌 모델의 학습 불균형을 유발한다.

본 논문에서는 각 RSU 로컬 데이터의 객체 분포 및 객체별 학습 정확도를 고려하여 학습에 참여할 RSU를 선택하는 선택적 연합학습 기법을 제안한다. 본 기법은 객체별 학습 정확도를 기반으로 객체 분포 다양성을 간접적으로 반영하여 학습에 참여할 RSU 선택 기준을 설계하여 특정 객체에 대한 과도한 학습을 예방한다. 이를 통해 RSU 간 상이한 환경과 특정 데이터 과적합으로 인해 발생하는 데이터 편향을 줄이고 다양한 객체 분포를 반영한 글로벌 모델 학습을 지원하고자 한다.

II. 본론

2.1 객체 분포 및 객체별 학습 성능 기반 연합학습

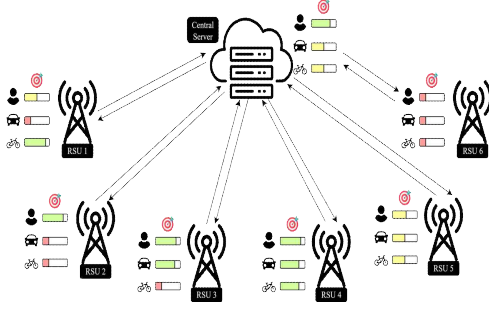


그림 2. 학습 정확도를 고려한 RSU 선택 구조

그림 2는 각 RSU의 객체별 학습 정확도를 중앙 서버와 비교하여 학습에 참여할 RSU를 선택하는 구조를 나타낸다. 본 기법은 학습에 참여할 RSU 선택 방식을 두 단계로 나뉘어 진행된다. 먼저, 객체별 학습 정확도를 고려한 1차 선정 절차를 진행한다. 글로벌 모델의 객체별 학습 정확도와 학습에 참여하는 RSU들의 객체별 로컬 학습 정확도를 비교하여 그 차이를 표준편차(Standard Deviation)로 산출한다. RSU별로 산출된 표준편차의 평균값을 구하고 평균값 이하의 객체 학습 정확도를 가진 RSU는 해당 학습 라운드에서 제외한다. 이 과정에서 객체별 학습 정확도의 표준편차는 RSU의 객체 분포가 학습 성능에 반영된 정도를 판단하기 위한 지표로 활용될 수 있다. 특정 객체에 대한 정확도 표준편차가 크게 나타나는 경우, 해당 RSU는 글로벌 모델에 충분히 반영되지 않은 다양한 객체를 포함할 가능성이 높다. 이러한 RSU를 선택함으로써 객체별 학습 정확도 편차를 감소시키고 불균형한 객체 분포 학습으로 인해 발생하는 데이터 편향을 효과적으로 완화하고자 한다.

객체별 학습 정확도를 고려한 1차 선정 이후 학습 안정성과 학습 성능 저하를 방지하기 위한 2차 선정 절차를 진행한다. 전체 객체에 대한 글로벌 모델의 학습 정확도와 각 RSU의 학습 정확도를 비교한다. 글로벌 모델보다 전체 학습 정확도가 낮은 RSU는 모델 수렴에 부정적인 영향을 미칠 가능성이 있으므로 해당 학습 라운드에서 제외한다. 이를 통해 객체 분포의 다양성을 고려한 RSU 선택 과정에서 학습 안정성을 동시에 확보하고 글로벌 모델의 전반적인 성능 저하를 방지하고자 한다.

2.2 제안 기법 동작 방식

연합학습 환경에서 RSU 집합을 $R = \{1, 2, \dots, N\}$, 객체 집합을 $O = \{o_1, o_2, \dots, o_k\}$ 로 정의한다. 중앙 서버는 글로벌 모델의 객체별 학습 정확도를 측정하고 각 RSU는 객체별 로컬 학습 정확도를 측정한다.

$$D_i(o_k) = |A_{global}(o_k) - A_i(o_k)| \quad (1)$$

식(1)은 객체 분포 불균형이 객체별 학습 정확도에 미치는 영향을 분석하기 위해 중앙 서버의 글로벌 모델과 RSU 간 객체별 정확도 편차 $D_i(o_k)$ 를 산출하는 과정을 나타낸다. RSU i 의 객체 o_k 에 대한 로컬 학습 정확도를 $A_i(o_k)$, 동일한 객체에 대한 글로벌 모델의 학습 정확도를 $A_{global}(o_k)$ 로 정의한다. 중앙 서버는 RSU에게 객체별 정확도 편차인 $A_i(o_k)$ 를 제어 패킷으로 패킷화하여 전송하도록 요청한다. RSU로부터 수집된 $A_i(o_k)$ 를 기반으로 중앙 서버는 객체별 학습 정확도 편차 $D_i(o_k)$ 를 산출한다.

$$\mu_k = \frac{1}{N} \sum_{i=1}^N D_i(o_k) \quad (2)$$

식(2)에서는 중앙 서버가 객체별 정확도 편차 $D_i(o_k)$ 를 산출한 후 이를 기반으로 객체별 평균 편차 μ_k 를 계산한다.

$$R = \bigcap_{k=1}^K R_k = \{i \in R \mid D_i(o_k) \geq \mu_k, \forall k\} \quad (3)$$

식(3)은 객체 o_k 에 대한 RSU i 의 객체별 정확도 편차 $D_i(o_k)$ 가 전체 RSU의 객체별 정확도 평균 편차인 μ_k 보다 작은 경우, 해당 RSU를 학습 참여 후보에서 1차 제외한다.

$$R = \{i \in R \mid A_i^{all} > A_{global}^{all}\} \quad (4)$$

1차 선정 이후, 중앙 서버는 RSU의 학습 정확도 A_i^{all} 를 글로벌 모델의 학습 정확도 A_{global}^{all} 와 비교하여 A_{global}^{all} 보다 학습 정확도 A_i^{all} 가 낮은 RSU를 2차 제외한다. 이는 객체별 정확도 편차 $D_i(o_k)$ 가 객체별 정확도 평균 편차 μ_k 보다 상회하여 1차 선정 기준을 만족하더라도 학습 기여도가 낮은 RSU의 참여로 인해 발생할 수 있는 글로벌 모델 학습 성능 저하를 방지한다.

III. 결론

본 논문에서는 선택적 참여 기반 연합학습 환경에서 발생할 수 있는 데이터 편향 문제를 완화하기 위해 중앙 서버와 RSU의 객체별 및 전체 학습 정확도를 고려한 RSU 선택 기법을 제안하였다. 제안 기법은 특정 객체 분포에 편향된 반복적인 학습을 제한하여 데이터 편향을 완화하고 다양한 객체 분포를 반영한 글로벌 모델의 학습을 지원한다. 또한 객체별 학습 정확도 기반 선택적 연합학습 기법을 통해 기존 선택적 연합학습 대비 학습 안정성을 기대할 수 있다. 향후 연구에서는 제안한 RSU 선택 기법의 성능을 검증하기 위해 다양한 시뮬레이션 환경에서의 성능 평가를 수행하여 기법의 성능을 정량적으로 분석하고자 한다.

ACKNOWLEDGMENT

This work was supported by the Commercialization Promotion Agency for R&D Outcomes(COMPA) grant funded by the Korean Government(Ministry of Science and ICT). (RS-2023-00304695), and supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant NRF-RS-2018-NR031059.

참고 문헌

- [1] T. Gong, L. Zhu, F. R. Yu and T. Tang, "Edge Intelligence in Intelligent Transportation Systems: A Survey," in IEEE Transactions on Intelligent Transportation Systems, vol. 24, no. 9, pp. 8919-8944, Sept. 2023.
- [2] F. Basma, Y. Tachwali and H. H. Refai, "Intersection collision avoidance system using infrastructure communication," 2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC), Washington, DC, USA, 2011, pp. 422-427
- [3] MCMAHAN, Brendan, et al. Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. PMLR, 2017. p. 1273-1282.
- [4] F. Maciel, A. M. De Souza, L. F. Bittencourt and L. A. Villas, "Resource Aware Client Selection for Federated Learning in IoT Scenarios," 2023 19th International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT), Pafos, Cyprus, 2023, pp. 1-8.
- [5] BENARBA, Nawel; BOUCHENAK, Sara. Bias in Federated Learning: A Comprehensive Survey. ACM Computing Surveys, 2025.