

무선 D2D 네트워크를 위한 자원 인지 및 위치 기반 클러스터링 연합학습

정준표, 고영배
아주대학교 AI 융합네트워크학과

{forjune, youngko}@ajou.ac.kr,

*임성화
*남서울대학교 멀티미디어학과

*sunghwa@nsu.ac.kr

Resource-aware, Location-based Clustered Federated Learning for Wireless D2D Networks

June-Pyo Jung, Young-Bae Ko

*Sung-Hwa Lim

Dept. of AI Convergence Network Ajou University *Dept. of Multimedia Namseoul University

요 약

클라우드 서버와 클라이언트들의 직접적인 통신을 통한 기존 연합학습(Federated Learning)방식의 문제점으로 꼽히는 많은 통신 자원의 낭비, 중앙처리에 따른 부하, 실제환경에서 적용의 어려움 등의 문제를 개선하기 위해 거리 기반 클러스터링 연합학습의 효율적 방식을 제안한다. 클라우드 서버가 각각의 모든 클라이언트들과 직접적인 일대일 통신을 통해 모델 파라미터를 교류함으로써 발생하는 통신 자원의 소모를 줄이는 방식으로, D2D 통신에 기반하여 클라이언트들의 위치에 따른 제한된 거리(distortion)기반 K-mean 클러스터링을 통해 그룹과 그룹 리더를 선정한다. 선정된 그룹 리더는 그룹 내의 클라이언트들의 모델을 1 차적으로 취합하여 서버로 전송함으로써 클라우드 서버의 부하를 줄이는 동시에 클라우드 서버와 클라이언트들의 통신 자원 소모를 줄이는 연합학습 방식을 제안한다.

I. 서 론

현재 네트워크 환경에서 기존의 머신 러닝은 디바이스에서 생성되는 데이터들을 클라우드 서버로 전송하여 데이터를 학습하는 방식을 사용한다. 하지만 클라우드 학습 처리 방식은 네트워크 대역폭, 개인정보 보호, 통신 자원 낭비 등의 문제점을 발생시키므로 모바일 환경에서는 적합 하지 않다[1].

이러한 문제점들을 해결하고자 최근 분산학습 방식의 하나인 연합학습(Federated Learning: FL)[1]이 주목받고 있다. 연합학습은 로컬 디바이스에서 생성되는 데이터를 클라우드로 전송하지 않고 로컬에서 학습함으로써 모델 파라미터 만을 클라우드 서버로 전송하여 업데이트 하는 방식이다.

하지만 기존 연합학습 방식을 모바일 환경에 효과적으로 적용하기 위해서는 학습 모델의 정확도의 수렴을 보장하는 동시에 제한된 통신과 컴퓨팅 자원에 대한 고려도 같이 이루어져야 한다. 따라서 연합학습의 다양한 프레임워크 중 하나로 클러스터링 기반 연합학습(Clustered Federated Learning: CFL)방식이 제안되었다[2]. 클라우드 서버와 모든 클라이언트들이 일대일 통신을 하는 기존의 방식을 개선하여, 클러스터링을 통해 그룹을 형성시켜 학습 모델을 취합한 후 그룹의 LN(Leader Node)만이 서버와 학습 모델을 교류하는 방식이다. 이를 통해 원거리에 위치한 클라우드 서버 와 클라이언트들 간의 통신 비용을 줄이면서 학습 모델의 수렴성을 보장한다[3].

기존의 CFL 관련 연구들은 학습 모델의 파라미터를 기준으로 클러스터링을 실시하였다. 그러나 모델 파라미터 기준의 클러스터링은 원거리에 위치한 클라이언트끼리 통신 가능성이 존재하므로 제한된 통신 자원의 모바일 환경에서 효율적이지 못하다. 따라서 본

논문에서는 D2D(Device to Device) 네트워크 환경을 가정하여 제한된 거리 내에서 클라이언트들의 위치를 기반으로 k-mean 방식의 클러스터링을 통해 그룹내의 모델들을 취합하여 LN 을 통해 원거리에 위치한 클라우드 서버와 클라이언트들의 직접적인 통신을 줄이는 자원 효율적 클러스터링 기반 연합학습 방식을 제안한다.

II. 제안기법

그림 1.의 알고리즘은 본 논문에서 제안하는 L-CFL (Location-based Clustered FL) 이다. 클라이언트의 수 n 명, 설정한 전체 학습 라운드 T , $range$ 범위에 존재하는 클라이언트들의 무작위 위치 배열 $location$, 모델 파라미터 w , 클라이언트간 최대 거리 임계 값 $threshold$ 를 입력 값으로 설정하였다. $location$ 값들을 입력으로 k-mean 클러스터링 함수를 통해 클라이언트들끼리 그룹화 하고 그룹의 개수에 따른 클러스터 내부 평균 거리 $distortion$ 값을 구한다. 이때 설정한 제한된 거리 $threshold$ 값을 넘지 않는 그룹의 최대 개수 m 을 구한다. 그림 2.는 $range$ 가 2000*2000 인 환경에서 k-mean 클러스터링 방식을 활용하여 $threshold$ 를 넘지 않는 그룹 m 개에 따른 그룹화된 산포도의 결과의 예를 나타낸다.

각 클러스터 내부에 위치한 클라이언트들 중 centroid 에 가장 가까운 클라이언트를 LN 으로 선정한다. 이후 기존 연합학습과 유사한 방식으로, 선정된 LN 은 클라우드 서버로부터 초기화된 학습 모델을 전송받은 후 클러스터 내부의 클라이언트들에게 학습 모델을 전송한다. 클러스터 내부의 클라이언트들은 전송받은 학습 모델을 각자의 로컬 데이터를 활용하여 훈련한다. 훈련이 끝나면 LN 으로 학습 모델을 전송하고 LN 은 학습 모델들을

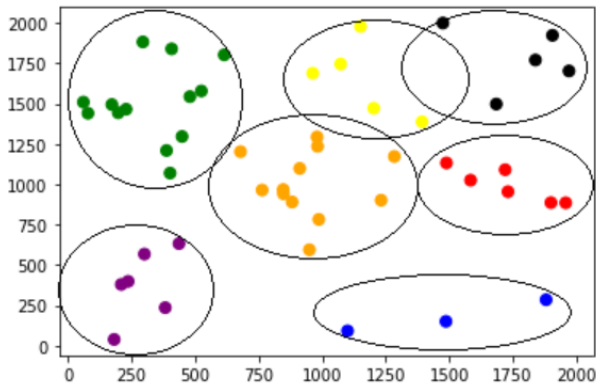
취합한 후 취합된 학습 모델을 클라우드 서버로 전송한다. 클라우드 서버는 LN 들로부터 받은 학습 모델을 취합한 후 학습 모델이 수렴할 때까지 T 라운드 동안 반복한다.

Algorithm 1: Location based Clustered Federated Learning

Input: Client set $C \leftarrow \{c_1, c_2, \dots, c_n\}$, Leader node set $LN \leftarrow \{LN_1, LN_2, \dots, LN_m\}$,
 Initialized group model set $W_0^G \leftarrow \{w_0^{g1}, w_0^{g2}, \dots, w_0^{gm}\}$,
 Total number of Communication T , Random location set $location$,
 Threshold for distance between nodes $threshold$

$kmmodels = Kmeans(location)$;
 $distortion = Distortions(kmmodels)$;
if $Max(distortion) \leq threshold$ **then**
 | $m = Max(distortion)$;
end
 $LN_m = findLN(centroid_m, client.location)$;
 Server sends Initialized model W_0^G to LN_k
for Communication Round $t=1, 2, \dots, T$ **do**
 | LN_m sends the model W_t^{gm} to all clients C belonging to the group;
 | Clients belonging to each group train the model W_t^{gm} using local data;
 | Clients belonging to each group update the model W_t^{gm} weights;
 | LN_m aggregate the models W_t^{gm} updated in the group and sends W_t^{gm} to server;
 | Server aggregates all LN_m models W_t^{gm}
end

<그림 1. Location -based Clustered Federated Learning Algorithm >



<그림 2. 클라이언트들의 위치에 따른 k-mean 클러스터링을 통한 그룹화된 산포도>

III. 성능 평가

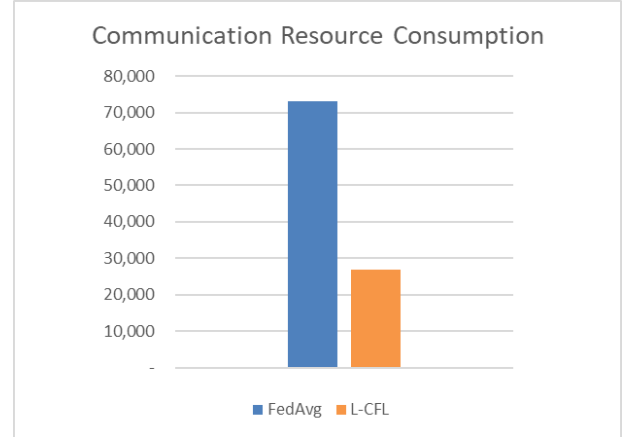
성능평가를 위해 D2D 네트워크 환경을 가정하였으며, 클라우드 서버의 위치는 2000m*2000m 의 중심점 (1000,100)으로 설정하였다. 연합학습시 소모되는 통신 자원의 단위를 거리(미터)에 따라 1 unit 으로 설정하였다. 클러스터링을 통해 조직된 그룹 내부에 존재하는 클라이언트들 간의 D2D 통신 비용은 다음과 같은 이유로 서버와 클라이언트들의 직접적인 통신 비용의 1/10 수준인 0.1 unit 으로 설정하였다.

- 1) 클라우드 서버와 LN 간의 직접적인 통신 비용은 클러스터 내부 통신 비용의 10 배이다[4]. [5].
- 2) 장거리 통신 거리와 잦은 backhaul 은 네트워크의 혼잡을 초래한다[4].

연합학습에 따른 통신 자원의 소모량을 비교하기 위한 벤치마크로 가장 보편적인 연합학습 방식인 FedAvg(Federated Averaging)[1]를 사용하였다.

그림 3. 은 본 논문에서 제안하는 L-CFL 기법의 통신 거리에 따른 자원 소모량을 FedAvg 와의 비교 그래프로

나타낸 것이다. FedAvg 방식의 거리에 따른 통신 비용은, 한 라운드동안 학습모델의 송신과 수신을 기준으로 $36502 * 2$ unit 이 소모되며, L-CFL 은 각 클러스터들의 내부 통신 비용 $1106 * 2$ unit, LN 과 서버와의 통신 $5621 * 2$ unit 을 합한 $6727 * 2$ unit 이 소모되었다. 이는 L-CFL 방식의 통신 비용이 FedAvg 방식의 통신 비용의 약 18%정도의 수준인 것을 확인할 수 있다.



<그림 3. FedAvg vs L-CFL 의 통신 거리에 따른 자원 소모량>

IV. 결론

본 논문은 기존 연합학습의 통신 자원 소모를 줄이기 위해 위치기반 k-mean 클러스터링을 통한 연합학습 기법을 제안하였다. 추후 본 연구를 확장하여 클라이언트들의 선택적 학습 참여를 통해 통신 자원 소모를 더 줄이면서 학습 모델의 유사도 기반의 클러스터링을 통한 연합학습 방식 연구를 통해 모바일 환경에서 자원 효율적이면서 높은 정확도의 학습 모델을 생성할 수 있는 연합학습에 대한 연구를 진행할 예정이다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 정보통신기획평가원의 대학 ICT 연구센터육성지원사업(IITP-2021-2018-0-01431)과 한국연구재단(No. 2021R1A2C1012776)의 지원을 받아 수행되었음.

참 고 문 헌

- [1] B. McMahan, et. al., "Communication-Efficient Learning of Deep Networks from Decentralized Data," in Artificial Intelligence and Statistics, pp. 1273- 1282, Apr. 2017.
- [2] Sattler, F, et. al., "Clustered federated learning: Model-agnostic distributed multitask optimization under privacy constraints". IEEE transactions on neural networks and learning systems. 2020.
- [3] Wang, Z., et. al. "Resource-Efficient Federated Learning with Hierarchical Aggregation in Edge Computing." In IEEE INFOCOM 2021-IEEE Conference on Computer Communications pp. 1-10. May. 2021
- [4] Y. C. Hu, M. Patel, D. Sabella, N. Sprecher, and V. Young, "Mobile edge computing—a key technology towards 5g," ETSI white paper, vol. 11, no. 11, pp. 1- 16, 2015.
- [5] E. Cuervo, A. Balasubramanian, D. Cho, A. Wolman, S. Saroiu, R. Chandra, and P. Bahl, "Maui: making smart-phones last longer with code offload," pp. 49- 62, 2010.