Re-Identification 문제에서의 향상된 표상을 얻기 위한 연합학습 시스템[1]

김성윤* 정우진* 윤세영

한국과학기술원

{curisam, gartland, yunseyoung}@kaist.ac.kr

Federated Learning Systems to Achieve Enhanced Feature Representations in Re-Identification

Seongyoon Kim* Woojin Chung* Seyoung Yun

KAIST

요 약

최근 빅 데이터 및 컴퓨팅 리소스의 발전으로 다양한 서비스에 인공지능(artificial intelligence, AI) 기술이 적용되고 있다. 일반적으로 클라우드로의 중앙 집중화를 통해 방대한 데이터를 수집하는데 이러한 과정에서 사용자 개인(클라이언트) 데이터의 유출 위험 이슈가 대두되고 있다. 최근 이러한 문제를 극복하기 위해 연합학습이 활발히 연구되고 있다. 연합학습은 각 클라이언트가 자신이 보유하고 있는 개별 데이터들로 직접 학습한 모델을 중앙으로 보내고 중앙에서는 클라이언트들로부터 받은 모델들을 취합하는 형식의 학습 방법이다. 본 논문에서는 이미지 분류(Image classification)에만 적용될 수 있는 연합학습보다 더 일반적인 Re-Identification(Re-ID)에 적용될 수 있는 연합학습 방법론 FedDKD를 제안하였고 Person Re-ID 벤치마크에서 비교 알고리즘보다 향상된 성능이 나옴을 확인했다.

1. 서 론

최근 클라우드 및 빅데이터 기술의 발전으로 다양한 서비스에 인공 지능(artificial intelligence, AI) 기술이 적용되고 있다. 이러한 인공 지능 모델을 잘 학습될 수 있던 배경에는 방대한 양의 데이터 그리고 방대한 파라미터 수의 모델에서의 연산을 수행할 수 있는 컴퓨터 리소스의 발전을 이유로 들 수 있다. 일례로, 클라우드 컴퓨팅은 리소스의 중앙 집중화를 기반으로 하기 때문에 모델 학습에 사용되는 모든 데이터를 클라우드 메모리에 저장해야 하는 제약이 있다. 데이터 중앙 집중화는 효율성 극대화라는 관점에서 많은 이점이 있으나, 사용자 개인(클라이언트) 데이터의 유출 위험이 있으며 클라우드 스토리지에 대한 제약도 존재한다.

최근 이러한 문제를 극복하기 위해 연합학습이 활발히 연구되고 있다. 연합학습은 클라우드 컴퓨팅과 같이 모든 클라이언트의 데이터들을 중앙에 모은 뒤 중앙 모델을 학습하는 것이 아닌, 각 클라이언트들이 클라이언트 모델 그리고 자신의 수요에 따른 차별화된 데이터(heterogeneous data)를 보유하고 있는 상황이다. 연합학습은 다음의 3가지 과정을 반복하여 모든 클라이언트들의 데이터들을 잘 예측할 수 있는 중앙모델을 얻어낸다. 1) 중앙에서 중앙 모델을 클라이언트들에게 전송. 2) 중앙 모델을 받은 각 클라이언트는 중앙 모델을 시작점으로 보유한 데이터들에 기반하여 직접 학습한 클라이언트 모델을 생성한 뒤 중앙으로 전송. 3) 중앙에서는 학습된 클라이언트 모델들을 취합. 이러한 연합학습은 클라이언트 개인의 데이터들을 중앙으로 수집하는 것이 아니기에 사생활 침해 소지가 적다. 또한 데이터를 직접적으로 공유하지는 않지만 각 클라이언트 개별 데이터로 학습된 모델들을 취합하는 과정을 통해 간접적으로 데이터를 공유하게 되어 더 나은 모델의 표상(feature)을 얻을 수 있다고 알려져 왔다.

본 논문에서는 연합학습을 Re-Identification(Re-ID)에 적용 가능한 **FedDKD**(**Fed**erated Learning via **D**ecoupling and **K**nowledge **D**istillation)을 제안한다. 일반적인 이미지 분류(Image Classification)를 위한 연합학습은 중앙에서 학습된 각 클라이언트들의 모델결과를 단순 취합하는데, 이러한 과정에서 중앙 모델과 클라이언트의 모델들이 동질적(Homogeneous)이어야 한다는 제약조건이 존재한다. 일반적인 AI 모델 θ는 [그림 1]과 같이 표상 추출기(feature extractor) θ_f 그리고 분류기(classifier) θ_c 로 구성이 된다.

[1] 이 논문은 2022 년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 [No. 2021-0-00907, 능동적 즉시 대응 및 빠른 학습이 가능한 적응형 경량 엣지 연동분석 기술개발, 90%] 및 [No. 2019-0-00075, 인공지능대학원지원(한국과학기술원), 10%].

Re-Identification 문제에서는 client가 주로 CCTV가 되는데, 각 CCTV가 보유하고 있는 객체(ID)의 개수(분류해야 하는 라벨의 개수)는 일반적으로 다르다. 따라서 Re-ID Task에서는 클라이언트(CCTV)에 배치하는 클라이언트 모델의 분류기 θ_c 부분은 이질적인 상황(Heterogeneous)에서 수행이 될 수 있는 보다 일반적인 연합학습 방법론이 필요하다. 이를 해결하기 위하여 본 논문에서는 추출기 θ_f, 그리고 분류기 θ_c 를 분리(decoupling)을 사용한다. 그리고 기존 baseline 대비 향상된 성능을 얻기 위해 중앙 추출기 모델 θ_f 를 선생 모델(Teacher Model)로 지정하는 지식 증류법(Knowledge Distillation)을 적용한다. 그리고 제안하는 FedDKD 알고리즘은 Person Re-ID 연합학습 벤치마크 세팅에서 비교대상 알고리즘에 비해 모든 CCTV의 데이터 셋에서 향상된 성능을 보여준다.

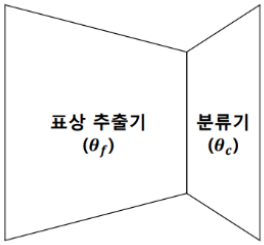

그림 1 모델 $\boldsymbol{\theta}$의 구조 ($\boldsymbol{\theta_f}.\boldsymbol{\theta_c}$)

2. 관련 연구

2.1 지식 증류법(Knowledge Distillation), 그리고 연합학습(Federated Learning) [1,2,3]

지식 증류법[1]은 학습된 거대 모델(Teacher Model)로부터 얻어진 입력 데이터에 대한 예측 값을 활용하여 가벼운 모델(Student Model)의 성능을 올리는 방법이다.

연합 학습[2]은 각 클라이언트들의 데이터들이 서로 공유가 되지 못하는 상황에서, 개별적으로 클라이언트들마다 학습된 모델들을 중앙에서 취합하는 방식으로 모든 클라이언트들의 수요를 만족시킬 수 있는 중앙 모델을 얻어내는 방법이다. [3]은 연합학습에서 지식 증류법을 활용해 변형된 손실함수(*LSD, NTD*)를 줄이는 것으로 더 나은 중앙 서버 모델의 성능을 얻을 수 있음을 확인한 논문이다. LSD, 그리고 NTD 모두 중앙 모델을 Teacher, 클라이언트 모델을 Student로 두어서 지식 증류를 하는 방법이다. NTD는 Not Ture Distillation의 약자로, LSD와는 다르게 증류 과정의 손실함수를 얻는 과정에서 입력 데이터에 해당하는 실제 예측 값 부분에 해당하는 한 개의 차원만을 없앤 벡터로부터 증류법을 적용한 방법이다.

2.2 연합 학습 통한 Re-Identification [4]

FedPAV[4]는 Re-ID에서 각 CCTV(클라이언트)가 보유한 데이터의 라벨수가 다른 것을 고려해 각 CCTV의 분류기의 모델 사이즈가 다를 수 있음을 허용한다. 또한 분류기와 추출기를 분리(decoupling)하여 추출기 부분만을 FedAVG [2]를 적용해 연합학습을 적용한 연구이다.

3. 제안 방법론

본 논문에서는 Re-ID 문제에서 연합학습을 적용할 수 있는 FedDKD[알고리즘 1]을 제안한다. 알고리즘에 사용되는 수식적 표기를 먼저 정의한 뒤 알고리즘을 구성하는 8가지 각 과정을 설명하겠다.

전체 클라이언트 수: N, 전체 라운드 수: R;
$n \in \{1, \cdots, N\}$번째 클라이언트가 보유하고 있는 데이터: D_n;
매 round에 선택되는 클라이언트 수: K;
초기 추출기 모델 (중앙 모델): θ_f^0;
초기 클라이언트들의 분류기 모델:$\{\theta_{c,n}^0\}_{n=1}^N$;
for *round* r=1,2, ... R **do**
  중앙 모델:
  (step1) N명의 클라이언트 중 K명의 클라이언트들의 인덱스 집합 S_r를 uniform하게 샘플링. $\rightarrow S_r \subseteq \{1, \cdots, N\}$
  (step 2) 중앙에서 중앙 추출기 모델 θ_f^{r-1}을 선택된 S_r 에 해당하는 클라이언트들에게 전송.
  for $k \in S_r$ **do**
    k번째 클라이언트 모델:
    (step 1) 중앙으로 받은 θ_f^{r-1}과 가지고 있는 분류기 $\theta_{c,k}^{\star}$을 결합하여 teacher model 구성. $\rightarrow \theta_k^{t,r} = (\theta_f^{r-1}, \theta_{c,k}^{\star})$
    (step 2) Student model $\theta_k^{s,r}$를 $\theta_k^{t,r}$로 초기화.
    (step 3) 지정된 student model, teacher model에서 D_k 바탕으로 ***LSD 손실함수*** 혹은 ***NTD 손실함수***을 줄이는 것을 목표로 학습된 student model $\theta_k^{s,r^\star}$을 얻어냄.
    (step 4) 학습된 student model $\theta_k^{s,r^\star}$의 추출기 부분 $\theta_{f,k}^{\star}$를 중앙 서버로 전송. $\rightarrow \theta_k^{s,r^\star} = (\theta_{f,k}^{r^\star}, \theta_{c,k}^{\star})$
  end
  중앙 모델:
  (step 1) 각 $k \in S_r$번째 클라이언트로부터 추출기 $\theta_{f,k}^{r^\star}$을 받음.
  (step 2) $\theta_f^r = \frac{1}{\sum_{k \in S_r} |D_k|} \sum_{k \in S_r} |D_k| \cdot \theta_{f,k}^{r^\star}$ 로 업데이트.
end

알고리즘 1 FedDKD

3.1 수식적 표기 정의

알고리즘 설명 이전에 수식적 표기를 먼저 설명하겠다. 먼저 N, R, 그리고 K는 각각 전체 연합학습에 참여하는 전체 클라이언트의 수, 연합학습동안 수행되는 전체 라운드 수, 그리고 매 라운드 마다 참여하는 클라이언트의 수이다. D_n, 그리고 $|D_n|$은 각각 n번째 클라이언트가 보유한 데이터 셋, 그리고 D_n 내에 있는 데이터의 개수이다. 마지막으로 θ_f^0 와 $\theta_{c,n}^0$ 는 학습 이전의 모델들로 각각 중앙 서버가 보유하고 있는 추출기, 그리고 n 번째 클라이언트가 보유하고 있는 분류기이다.

3.2 알고리즘 설명

FedDKD는 크게 8가지로 구성이 된다. 중앙서버에서 2개의 과정(클라이언트 샘플링 및 추출기 전송), 샘플링 된 각 클라이언트에서 4개의 과정(초기 선생 그리고 학생 모델 세팅, 학생 모델 학습, 그리고 중앙서버에게 추출기 전송), 그리고 중앙서버에서 2개의 과정(추출기 수용 및 취합)인데 다가오는 각 하위 섹션(3.2.1-3.2.3)에서 임의의 r 번째 라운드의 상황에서의 상황을 자세히 설명을 하겠다.

3.2.1 중앙서버: 클라이언트 샘플링 및 추출기 θ_f 전송

먼저, r 번째 라운드가 진행되기 이전에 중앙 서버는 r−1번째 라운드에서 취합된 추출기의 결과물인 θ_f^{r-1}을 보유하고

있는 상황이다. 첫번째 과정으로, 중앙서버에서 전체 N 명의 클라이언트들 중 r 번째 라운드의 연합학습에 참여할 K 명의 클라이언트들의 인덱스들을 $\{1,\cdots,N\}$ 집합으로부터 uniform 샘플링한 S^r 로 결정한다. 그 후 S^r 의 인덱스에 해당하는 클라이언트들에게 중앙서버는 추출기 θ_f^{r-1}를 전송한다.

3.2.2 클라이언트: 초기 선생 그리고 학생 모델 세팅, 학생 모델 학습, 그리고 중앙서버에게 추출기 전송

여기서는 S^r의 인덱스에 해당하는 K명의 각 클라이언트에서 일어나는 과정을 설명한다. 일례로, $k \in S^r$ 번째 클라이언트 모델에서 일어나는 것으로 대표해서 설명하겠다.
k 번째 클라이언트는 r 번째 라운드의 연합학습 참여하기 이전에 분류기 $\theta_{c,k}^*$를 보유하고 있다. $k \in S^r$번째 클라이언트는 중앙서버로부터 θ_f^{r-1}를 받는다. 추출기 θ_f^{r-1}로부터 선생($\theta_k^{t,r}$) 및 초기 학생($\theta_k^{s,r}$) 모델을 추출기와 분류기를 합친 형태인 ($\theta_f^{r-1}, \theta_{c,k}^*$)로 세팅한다. 다음, 클라이언트 k는 보유한 데이터 D_k 바탕으로, 그리고 선생 모델 $\theta_k^{t,r}$ 은 고정하고 LSD 혹은 NTD 손실함수를 줄이는 것을 목표로 학생 모델만 업데이트 하여 최종적으로 업데이트 된 학생 모델 θ_k^{s,r^*} 를 얻는다. 마지막으로, 클라이언트 모델은 θ_k^{s,r^*}의 추출기 $\theta_{f,k}^{r^*}$만을 중앙서버로 전송한다.

3.2.3 중앙서버: 추출기 수용 및 취합

앞선 과정으로 중앙서버는 S^r 의 인덱스에 해당하는 클라이언트들로부터 추출기를 받는다. 중앙서버는 받은 추출기들을 각 클라이언트들이 보유한 데이터 수 $|D_k|$ 에 비례하게 convex combination을 취하여 추출기 θ_f^r 를 업데이트 한다.

4. 실험 및 실험 결과

본 알고리즘은 유일하게 존재하는 연합학습 Re-id 벤치마크 FedPAV[4]의 모델: ResNet50, 그리고 하이퍼파라미터들을 동일하게 적용했다. 해당 벤치마크는 Person Re-ID로 총 9개의 CCTV들이 존재하며, 각 CCTV들은 순서대로 Market, DukeMTMC-reID, cuhk03-np-detected, cuhk01, MSMT17, viper, prid,, 3dpes, 그리고 ilids 데이터를 가진다. FedDTD는 FedPAV에 비해 지식 증류법을 위한 하이퍼파라미터 (β,τ)가 추가적으로 존재하는데 [3]과 동일하게 (1,3) 값을 적용했다.

Client	PAV	DKD *LSD*	DKD *NTD*	PAV*	DKD* *LSD*	DKD* *NTD*
1	78.24	78.83	**81.41**	69.18	74.76	**77.73**
2	72.71	71.54	**74.19**	65.66	68.54	**71.68**
3	23.14	23.50	**26.00**	17.29	23.43	**25.79**
4	**74.59**	71.81	73.46	66.67	70.06	**72.53**
5	41.79	41.50	**43.59**	33.55	33.10	**38.60**
6	44.94	43.99	**45.57**	33.86	37.66	**42.09**
7	22.00	**27.00**	**27.00**	11.00	14.00	**17.00**
8	67.89	67.89	**70.73**	62.60	69.10	**69.51**
9	79.59	82.65	**85.71**	79.59	78.57	**82.65**

표 1 Person Re-ID에서의 성능 결과

[표 1]은 베이스 라인 알고리즘 FedPAV[4], 그리고 제안한 알고리즘 FedDKD에서의 Rank 1 Accuracy를 기재하였다. PAV와 DKD는 각각 FedPAV와 FedDKD의 약어이다.
FedDKD는 LSD 혹은 NTD를 사용할 수 있는 두가지 옵션이 있어 사용한 손실함수는 별도로 표기하였다. 그리고 "*"가 표기된 것은 N=9, K=3으로 매 라운드마다 클라이언트를 샘플링 하는 Cross-Device 세팅, "*"가 표기 안 된 것은 N=9, K=9로 매 라운드마다 모든 클라이언트가 참여하는 Cross-Silo 세팅에서 진행됨을 나타낸다. 본 연구의 목표는 모든 클라이언트들 골고루 성능이 잘 나오는 것인데 Cross-Silo, 그리고 Cross-Device 세팅 모두에서 FedDKD-NTD 방법론이 대체적으로 우세한 것을 [표 1]에서 확인할 수 있다.

5. 결론 및 향후 연구

본 논문에서는 Re-ID에도 적용될 수 있는 더 일반적인 연합학습 방법론을 제안하였고 Person Re-ID 벤치마크에서 비교 알고리즘에 비해 성능이 향상됨을 확인했다. 허나 연합학습 관련 Re-ID 벤치마크가 Person만 존재하는 이유로 Vehicle Re-ID에서는 실험을 진행하지 못했다. Vehicle Re-ID에도 해당연구가 진행된다면 스마트 교차로 서비스 등에도 이용할 수 있을 것이다. 또한, 본 연구는 클라이언트들이 분류기의 사이즈를 이질적으로 가질 수 있음을 허용하지만 추출기 부분은 클라이언트들이 동일한 사이즈를 가져야 한다는 제약이 있다. 추출기 부분에 대한 가정도 완화할 수 있는 연구가 진행된다면 자연어처리 분야와 같은 분야에서도 해당 프레임 워크를 적용할 수 있을 것이다.

참 고 문 헌

[1] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531, 2015.

[2] Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Aguera y Arcas. Communication-efficient learning of deep networks from decentralized data. InArtificial Intelligence and Statistics, pages 1273 1282, 2017.

[3] Gihun Lee, Yongjin Shin, Minchan Jeong, and Se-Young Yun. Preservation of the global knowledge by not-true self knowledge distillation in federated learning. arXiv preprint arXiv:2106.03097, 2021.

[4] W. Zhuang, Y. Wen, X. Zhang, X. Gan, D. Yin, D. Zhou, S. Zhang, and S. Yi, "Performance optimization for federated person re-identification via benchmark analysis," arXiv preprint arXiv:2008.11560, 20.