미래사회
Journal of Future Society
2023. Vol. 14, No. 3, 01~11.

프라이버시를 보호하는 온라인 교육과정 추천 시스템

천 지 영* **노 건 태****

서울사이버대학교

본 논문에서는 사용자의 프라이버시를 보호하기 위한 연합 학습 기반의 추천 시스템을 제안한다. 기존 추천 시스템의 경우 데이터가 비독립 동일 분포(Non-IID)일 경우 추천의 성능이 저하되며, 데이터가 기관별로 나누어져 있는 경우 한곳에 모아서 모델링을 진행해야 하는 등의 데이터 프라이버시 이슈가 존재한다. 이러한 문제들을 해결하기 위해 제안하는 추천 시스템에서는 계층적 클러스터링을 통한 서버의 클라이언트 선택 전략과 연합 학습 기법을 사용하여, 비독립 동일 분포에서 나타나는 추천 성능 저하 문제와 사용자 프라이버시 문제를 동시에 해결하였다. 본 논문에서 제안하는 프라이버시를 보호하는 추천 시스템은 온라인 교육과정을 추천하는 환경에서 유용하게 활용될 수 있으며, 대표적으로 한국교육학술정보원(KERIS)에서 제공하는 맞춤 배움길 서비스 등에 활용될 수 있다.

주요어 : 연합 학습, Non-IID, 추천 시스템, 교육, 프라이버시

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1063992).
* 주저자: 천지영/서울사이버대학교 빅데이터·정보보호학과 조교수/서울시 강북구 솔매로 49길
/Tel: 02-944-5543/E-mail: jychun@iscu.ac.kr
** 교신저자: 노건태/서울사이버대학교 빅데이터·정보보호학과 조교수/서울시 강북구 솔매로 49길
/Tel: 02-944-5542/E-mail: gnoh@iscu.ac.kr

Ⅰ. 서론

코로나-19 팬데믹 이후, 전 세계의 온라인 교육 트렌드는 급속도로 변화되기 시작하였다. 이 변화의 중심에는 온라인 교육과정에 대한 높아진 관심과 그에 따른 폭발적인 수요 증가가 기반이 되었다. 이러한 상황 속에서 학습자들은 무수히 많은 교육 콘텐츠 중에서 자신에게 가장 적합한 것을 효과적으로 선택할 추천 시스템의 필요성이 대두되었는데, 이에 대한 대표적인 예로는 유튜브, 넷플릭스, 아마존과 같은 글로벌 기업들이 제공하는 추천 시스템이 있다. 이들 기업은 다양한 상품과 콘텐츠를 효율적으로 사용자에게 추천함으로써 그들의 매출과 기업 가치를 지속적으로 향상시키고 있다.

다행스럽게도 국내에서 위의 해결 방안으로 시작된 국가 차원의 교육과정 추천 서비스가 존재한다. 한국교육학술정보원(이하 KERIS)에서 제공하는 맞춤 배움길 서비스는 성인학습자가 자기주도적 평생교육을 설계할 수 있도록 배움 정보를 제공하는 서비스로, 20개의 원격대학으로부터 학습 데이터를 수집하여 AI 추천엔진을 사용해 2021년 5월부터 사용자에게 학습 추천 정보를 제공하고 있다. 이 서비스는 학습자에게 딱 맞는 교육과정을 인공지능 분석을 기반으로 하여 맞춤형으로 제공하는 데 의의가 있으며, 이러한 시도는 본인의 현재 상태는 알고 있지만, 현재 본인에게 필요한 교육과정이 무엇인지를 설계하기 어려운 학습자들에게 매우 적합한 학습 추천 서비스이다.

그러나 맞춤 배움길 서비스는 학습에 필요한 20개 원격대학의 학습 정보가 한곳에 모여서 인공지능 분석이 진행되는 문제가 존재한다. 이것은 기존 학습 데이터를 제공하는 개인들의 정보가 익명화 등의 처리 여부와 무관하게 한곳에 집중되어 발생되는 프라이버시 문제를 내포하고 있다. 또한, 대학알리미에서 제공하는 데이터를 분석한 결과, 학습정보를 제공하는 20개 원격대학은 모두 다른 특성을 가지고 있는 Non-IID(Non-Independent and Identically Distributed), 즉 비독립 동일 분포를 띄고 있어 추천의 성능이 저하될 우려가 있다.

따라서 본 논문에서는 사용자 데이터 프라이버시 문제와 비독립 동일 분포에서 나타나는 성능 감소 효과를 효율적으로 해결할 수 있는 프라이버시를 보호하는 연합 학습 기반의 온라인 교육과정 추천 시스템을 제안한다. 연합 학습(Federated Learning)은 데이터를 한 곳에 모으지 않고도 분산된 장치들로부터 인공지능 모델 학습이 가능한 기법이다. 이러한 연합 학습에 비독립 동일 분포 문제 해결을 위한 계층적 클러스터링(Hierarchical Clustering) 단계를 추가하여 추천 시스템에서의 프라이버시 문제와 성능 저하 문제를 동시에 해결하였다. 이렇게 제안된 추천 시스템은 기존의 맞춤 배움길 서비스의 개선에도 크게 기여할 것으로 기대된다.

Ⅱ. 배경 지식

본 장에서는 제안하는 기법을 이해하기 위해 필요한 연합 학습, 비독립 동일 분포(Non-IID), 계층적 클러스터링에 대한 개념과 맞춤 배움길 서비스에 대해서 살펴본다.

1. 연합 학습

연합 학습(Federated Learning)은 여러 분산 장치에서 생성된 로컬(Local) 데이터를 중앙에 집중시키지 않고, 각 장치가 자체 데이터로부터 계산한 로컬 모델을 중앙에 모아 글로벌(Global) 모델 학습에 활용하는 새로운 모델 학습 기법이다[1, 2]. 연합 학습은 데이터를 중앙에 집중시키지 않기 때문에 일정 수준의 프라이버시 보호를 가능하게 하며, 데이터를 중앙에서 처리하기 위한 저장 공간, 전력, 연산량 등의 비용을 절감할 수 있다.

연합 학습의 실제 적용 사례로는 대표적으로 구

글의 Gboard가 있으며, 사용자의 스마트폰에서 생성되는 데이터를 이용해 키보드 예측 기능을 개선하였다. 구글로부터 시작된 연합 학습은 다양한 분야에서 활용될 수 있는데, 대표적으로 의료 데이터, 교통 데이터, 금융 데이터 등 제도적으로 데이터의 공유가 어려운 환경에서 활용 가능성이 높다.

연합 학습의 대표 알고리즘으로는 FedAVG (Federated Averaging)가 있다[2]. 이 기법에서는 각각의 로컬 모델에서의 가중치의 평균으로 글로벌 모델을 계산한다. 전체 K명의 클라이언트가 존재할 때, 각각의 클라이언트 k는 글로벌 모델 w_t를 자신의 데이터로 학습한 후 다음과 같이 로컬 모델 w_{t+1}^k를 계산한다[3].

$$w_{t+1}^k \leftarrow w_t - \alpha \nabla F_k(w_t)$$

각각의 클라이언트는 각자 계산한 로컬 모델을 중앙 서버에게 전송하고, 중앙 서버는 로컬 모델들의 평균값을 계산하여 다음과 같이 글로벌 모델 w_{t+1}을 계산한다. 여기서 n_k는 클라이언트 k의 데이터의 개수이고, n은 모든 클라이언트 데이터의 개수를 합한 전체 데이터의 개수이다.

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^k$$

위와 같이 연합 학습은 각각의 클라이언트가 각각의 장치에서 자신의 로컬 데이터로 모델을 학습한 후 서버에게 전송하면 중앙 서버는 이 값들을 취합하여 전체 모델을 학습하는 방식으로, 중앙 서버가 각 장치들에 대한 데이터에는 접근하지 않는 방식이다.

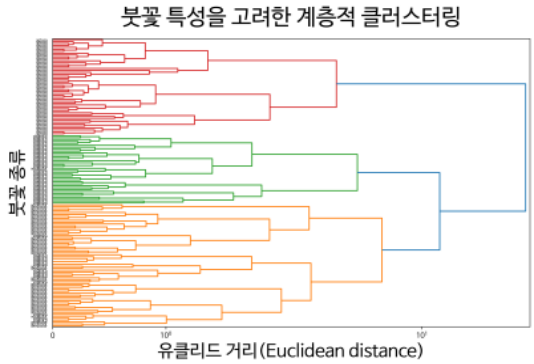
2. 비독립 동일 분포(Non-IID)

비독립 동일 분포(Non-Independent and Identically Distributed, Non-IID)는 데이터의 분포가 독립적이지 않거나 동일하지 않다는 특성을 말하며, 이는 데이터 간에 상관성이 존재하거나, 다른 확률 분포에서 추출된 데이터들이 포함되어 있는 경우를 의미한다. 예를 들어, 학습자들의 학습 데이터를 생각해보면, 각 학습자의 학습 패턴, 선호도, 능력 등은 독립적이지 않고 동일하지 않을 수 있다. 이는 각 학습자가 개별적인 배경, 경험, 지식을 가지고 있기 때문에 이런 경우 데이터는 Non-IID 특성을 가진다.

데이터가 Non-IID 특성을 갖는 경우, 데이터 분포 문제를 해결하기 위한 방안이 필요한데, 데이터를 선택적으로 사용하거나 합성 데이터 증강(Synthetic Data Augmentation) 등을 통해 IID 분포의 특성을 갖도록 한다.

3. 계층적 클러스터링

계층적 클러스터링(Hirarchical Clustering)은 데이터를 계층적 구조로 분할하는 방식으로, 전통적인 클러스터링 방법과는 달리 데이터 포인트들을 중첩되지 않는 부분집합으로 나누지 않는다. 이 방식은 데이터 포인트 간의 유사도나 거리를 기반으로 클러스터를 형성하며, 결과는 트리 구조로 시각화될 수 있다.



<그림 1> 계층적 클러스터링 예시

그림 1은 계층적 클러스터링의 대표적인 예시이다.

데이터는 사이킷런(Scikit-learn)에서 제공하는 붓꽃 데이터셋(Iris data set)을 사용하였으며, SciPy에서 제공하는 scipy.cluster.hierarchy 모듈을 사용하여 계층적으로 클러스터링하고 덴드로그램(Dendrogram)으로 시각화하였다[1].

계층적 클러스터링에는 병합적(Agglomerative) 방법과 분할적(Divisive) 방법이 있는데, 병합적 클러스터링은 각 데이터 포인트를 독립적인 클러스터로 간주하고 가장 유사한 클러스터들을 점진적으로 병합한다. 반면 분할적 클러스터링은 전체 데이터를 하나의 클러스터로 보고, 이를 점차적으로 분할한다.

계층적 클러스터링은 학습자의 특성과 선호도를 기반으로 유사한 학습자 그룹을 형성하는 데 효과적으로 사용될 수 있다. 온라인 교육 환경에서 학습자 데이터는 다양한 출처와 형태로 존재한다. 각 기관이 보유한 학습자 특성을 종합적으로 고려하여, 계층적 클러스터링을 통해 유사한 학습자들을 그룹화할 수 있다.

4. 맞춤 배움길

맞춤 배움길은 원격대학 졸업생 정보를 기반으로 하여 인공지능 분석을 통해 학습자 맞춤형 교육과정 정보를 제공하는 서비스로, KERIS에서 개발하여 운영하고 있다(그림 2 참고).

<그림 2> 맞춤 배움길 화면

이 서비스는 성인학습자가 자기주도적 평생교육을 설계할 수 있는 배움 정보를 제공하는 것을 목적으로 하며, 이를 위해 학습자와 유사도가 높은 선행학습자의 데이터를 기반으로 AI 추천엔진을 사용하여 교육과정 자가설계를 위한 추천이 이루어지도록 설계되었다. 2023년 7월 기준, 원격대학 20개교(고등교육법을 따르는 사이버대학 19개교 전체와 한국방송통신대학교)의 추천정보를 기반으로 하고 있다. 해당 서비스를 위해 KERIS는 성인학습자에 필요한 학습 데이터를 20개 원격대학으로부터 수집하여 한 곳에서 일원화된 서비스를 제공하고자 하였다.

Ⅲ. 기존 추천 시스템의 문제점

본 장에서는 기존 추천 시스템의 문제점을 분석하고, 기존 교육과정 추천 시스템의 문제점을 제시한다.

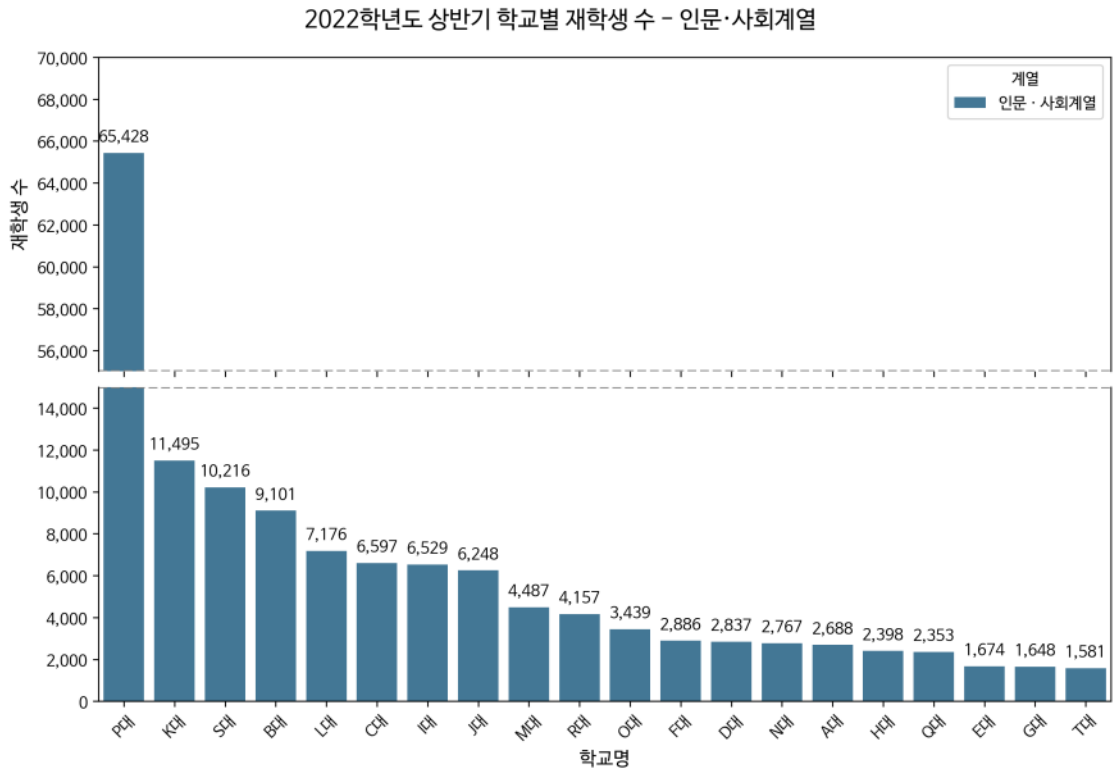
1. 데이터의 분포 문제

기존의 추천 시스템은 사용자(users)나 아이템(items)이 독립항등분포(Independent and Identically Distributed, IID)되었다고 가정하는 기법들이 대부분이었다. 이러한 가정하에 생성된 모델은 데이터가 IID 가정을 만족하지 못하는 경우 추천의 성능이 저하되는데, 예를 들면 뉴질랜드 새인 키위(kiwis)를 검색한 사용자에게 과일 키위(kiwi) 검색에 대한 연관정보를 추천하거나, 온라인 서점에서 전혀 관련 없는 책을 추천하기도 한다[4].

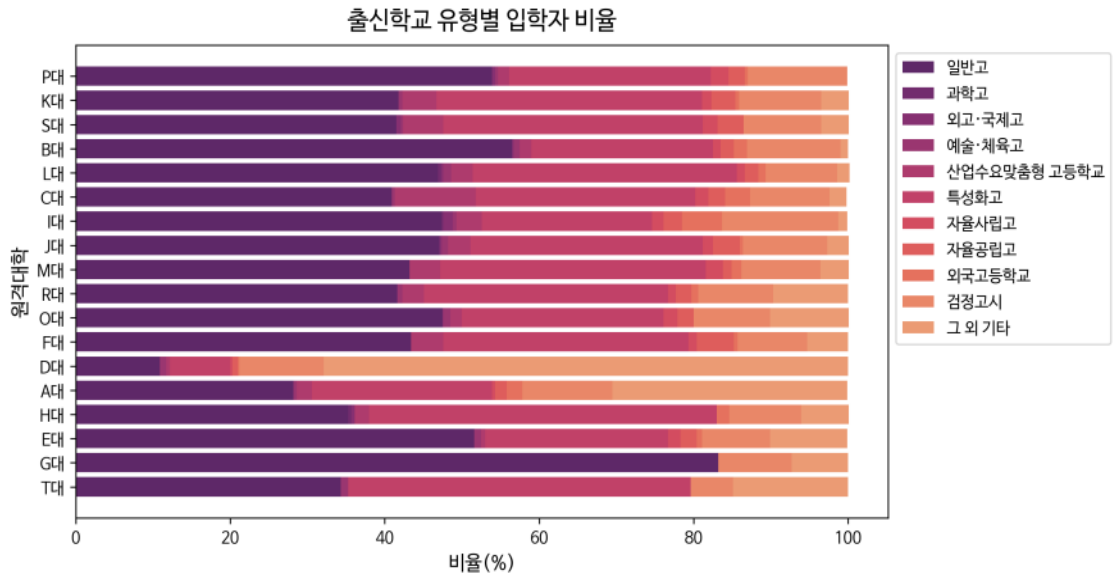
추천 시스템에서 사용하는 데이터의 실제 분포가 Non-IID한 경우가 많아 Non-IID 데이터를 가정하지 않는 모델의 경우, 학습이 제대로 이루어지지 않아 추천 시스템의 성능이 저하된다.

2. 데이터의 프라이버시 문제

추천 시스템의 성능은 데이터의 양에 의존하여, 사용하는 데이터가 많으면 많을수록 좋은 성능을 낸다[5].



<그림 3> 원격대학 재학생 수 – 인문·사회계열



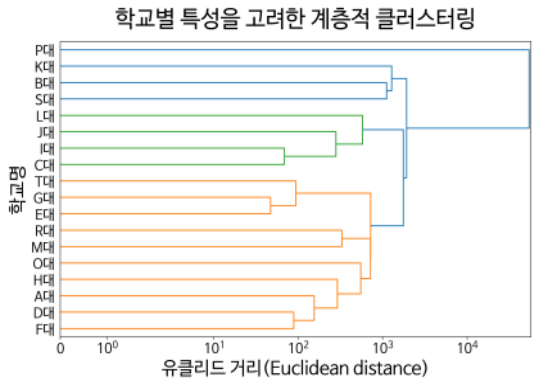
<그림 4> 원격대학 출신학교 유형별 입학자 비율

따라서 많은 데이터의 확보가 필요한데, 더 많은 데이터의 확보를 위해 여러 기관으로부터 다수의 데이터를 결합해서 사용할 수 있다면 더 나은 추천 시스템의 사용이 가능할 것이다. 하지만 프라이버시 및 보안 이슈로 인해 여러 기관 간의 데이터의 공유가 어렵다. 따라서 데이터는 각 기관의 로컬 서버에 보관한 상태에서 데이터의 공유 없이 모든 기관의 데이터로부터 학습된 글로벌 모델을 만들 수 있다면 프라이버시 및 보안 문제없이 추천 시스템 개발이 가능하다.

3. 기존 온라인 교육과정 추천 시스템의 문제점

2장 1절에서 기술한 바와 같이 기존의 교육과정 추천 시스템인 맞춤 배움길 서비스에서 KERIS는 20개 원격대학으로부터 학생들의 학습정보를 제공받는다. 대학알리미에서 제공하는 데이터를 분석한 결과, 학습정보를 제공하는 20개 원격대학의 데이터는 모두 다른 특성을 가지고 있는 Non-IID이며 데이터이다.

간단히 인문·사회계열 재학생 수만 보더라도 최대 41.38배나 차이가 나며(그림 3 참고), 출신학교 유형별 입학자 비율을 보더라도 입학자원의 구성 자체가 현격히 차이나는 것을 볼 수 있다(그림 4 참고).

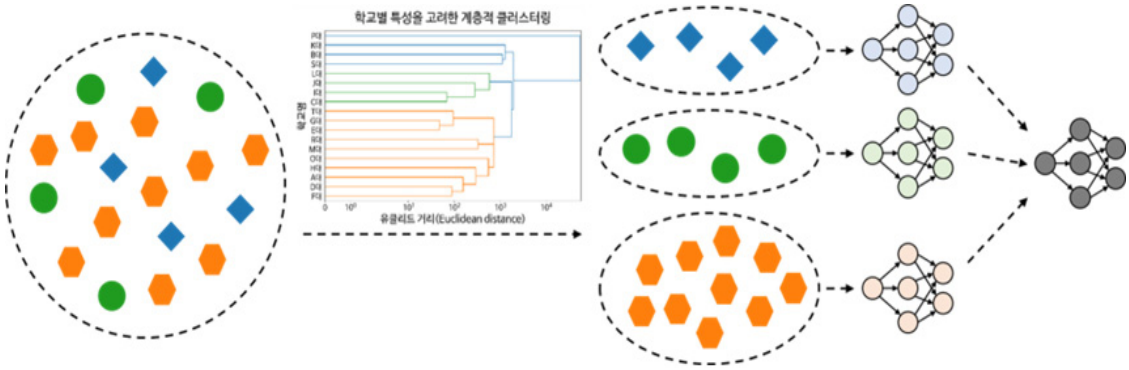


<그림 5> 학교별 특성을 고려한 계층적 클러스터링

이러한 데이터를 기반으로 계층적 클러스터링 결과를 시각화해보면 그림 5와 같다. 계층적 클러스터링을 위해서 SciPy에서 제공하는 scipy.cluster. hierarchy 모듈을 사용하였으며, 덴드로그램을 사용하여 시각화하였다.

Ⅳ. 제안하는 기법

본 연구에서는 프라이버시를 보호하는 교육과정 추천 시스템을 제안한다. 먼저 연합 학습 기반의 Non-IID 추천시스템을 제안하고, 교육 환경에 적용한 프라이버시를 보호하는 온라인 교육과정 추천 시스템을 제안한다(그림 6 참고).



<그림 6> 제안하는 연합 학습 기반의 Non-IID 추천 시스템 기본 개념도[6]

1. 제안하는 연합 학습 기반의 Non-IID 추천 시스템(Federated Non-IID Recommender System)

3장에서 제시한 데이터의 분포 문제와 데이터의 프라이버시 문제를 해결하기 위해 연합 학습 기반의 Non-IID 추천 시스템을 제안한다.

추천 시스템 모델의 개발에 연합 학습을 도입한다면 프라이버시를 보호하는 추천 시스템이 가능하다. 추천 시스템에서 사용하는 데이터의 경우 개인의 취향이라던지 사용자의 행동 패턴 등이 노출될 우려가 있기 때문에, 일반적인 머신러닝 모델 개발보다 더 강력한 프라이버시에 대한 고려가 필요하다.

따라서 연합 학습의 도입으로 이러한 문제를 줄일 수 있으며, 많은 양의 빅데이터를 한곳에 모으지 않고 각 기관에서 관리하게 함으로써 저장 용량이나 관리에 대한 비용 또한 줄일 수 있게 된다.

Non-IID 데이터 문제를 해결하기 위해 최근 계층적 군집화(Hierarchical Clustering)[3], 모델 개인화(Personalized Learning)[7], 지식 증류(Knowledge Distillation)[8] 등이 대안 기술로 연구되고 있다. 이 중 클러스터링 기반 연합 학습 기법이 추천 시스템에서 대중적인 기법[9]이며, 클러스터가 잘 형성되었을 시 용이하게 해결이 가능하다는 장점이 있다.

본 기법에서는 연합 학습시, Non-IID 문제를 해결하기 위하여 계층적 군집화를 활용한 서버의 클라이언트 선택 전략을 사용한다[3]. 이는 모든 클라이언트가 동시에 모델을 업데이트하는 것이 아닌, 특정 클라이언트를 선택하여 모델을 업데이트한다. 이러한 방법은 서버가 클라이언트의 데이터 분포를 고려하여 클라이언트를 선택하기 때문에 Non-IID 문제를 해결할 수 있다.

제안하는 기법은 다음과 같다. 본 기법에서는 중앙 서버와 K개의 클라이언트를 가정한다.

- **단계 1(초기화)**. 각각의 클라이언트는 다양한 데이터를 가지고 있고, 이 데이터는 Non-IID특성을 가질 수 있다. 제안하는 시스템에서는 연합 학습 기반의 추천 시스템을 사용한다[10]. 여기서 연합 학습 기반의 추천 시스템을 FedRecSys라고 한다.
- **단계 2(연합 학습)**. 초기에 모든 클라이언트는 각자의 데이터를 가지고 서버의 글로벌 모델을 기반으로 로컬 업데이트를 수행한다. 각각의 클라이언트는 자신의 데이터를 사용하여 모델을 로컬에서 훈련시키고(LocalUpdate 참조), 이 훈련된 모델의 가중치를 서버에 전송한다. 서버는 연합 학습의 대표 알고리즘인 FedAVG 알고리즘을 이용하여 글로벌 모델을 업데이트 한다(FedRecSys 참조).
- **단계 3(계층적 클러스터링)**. 충분한 학습을 진행한 후, 서버는 모든 클라이언트의 로컬 업데이트를 기반으로 클라이언트 간의 유사성을 평가한다. 유사성은 유클리드 거리를 사용하며, 모델의 매개변수를 벡터로 재구성하여 계층적 클러스터링 알고리즘의 입력으로 사용한다. 클러스터링 알고리즘은 유사성에 기반하여 클라이언트들로 구성된 계층적 클러스터를 형성한다(그림 6 참조).
- **단계 4(클러스터 기반 훈련)**. 유사성이 높은 클라이언트들로 구성된 각각의 클러스터는 독립적으로 훈련을 진행한다. 모든 클러스터의 훈련이 완료되면, 각 클러스터에서 훈련된 모델들은 글로벌 모델로 병합된다(FedRecSys & Hierarchical Clustering 참조).

다음은 단계 2에서 클라이언트 k가 자신의 데이터 D_k로부터 로컬 모델 w를 학습한 후 서버에 전송하는 절차를 나타낸다[3]. 여기서 E는 epoch의 수, α는 학습률(Learning Rate)을 나타낸다.

LocalUpdate(k,w)

$B \leftarrow$ Split D_k into batches of size B
for each local epoch i from *1* to E ***do***
  for batch $b \in B$ ***do***
    $w \leftarrow w - \alpha \nabla F_k(w)$
  end for
end for
return w to server

다음은 서버가 FedAVG 알고리즘을 이용하여 글로벌 모델을 업데이트하는 절차[3]로 단계 2에서 사용한 방법이다.

FedRecSys(w_t,K)

for each client $k \in K$ ***do***
  $w_{t+1}^{k} \leftarrow$ **LocalUpdate**(k,w_t)
end for

$$w_{t+1} \leftarrow \sum_{k=1}^{K} \frac{n_k}{n} w_{t+1}^{k}$$

다음은 연합 학습 기반의 추천시스템인 FedRecSys과 계층적 클러스터링 HC를 이용한 전체 과정을 나타낸다[3].

FedRecSys & Hierarchical Clustring

Initialize w_0
for each round $t \in [1, R]$ ***do***
  $w_{t+1} \leftarrow$ **FedRecSys**(w_t,K)
end for
$w \leftarrow w_{t+1}$
for each round $k \in K$ ***do***
  $\triangle w^{k} \leftarrow$ **LocalUpdate**(k,w)
end for
$C \leftarrow$ **HC**($\triangle w$, Hyperarameters)
for $c \in C$ ***do***
  $w_{c,0} \leftarrow w$
  for each round $t = 1, 2, \ldots$ ***do***
    $w_{c,t+1} \leftarrow$ **FedRecSys**($w_{c,t}$,K_c)
  end for
end for

이러한 절차를 통해 Non-IID 데이터 분포를 가진 클라이언트 간의 유사성을 기반으로 특화된 글로벌 모델을 생성한다.

2. 프라이버시를 보호하는 온라인 교육과정 추천 시스템

앞에서 제안한 연합 학습 기반의 Non-IID 추천 시스템을 맞춤 배움길에 적용한다면 20개 원격대학의 데이터가 한곳에 모이지 않고도 인공지능 분석을 기반으로 한 온라인 교육과정 추천 시스템을 제공할 수 있게 된다.

제안하는 연합 학습 기반의 Non-IID 추천 시스템을 맞춤 배움길에 적용한 기법을 설명하면 다음과 같다.

- **단계 1(초기화)**. KERIS는 각 원격대학을 하나의 클라이언트로 간주한다. 따라서 총 20개의 클라이언트가 존재한다.
- **단계 2(연합 학습)**. 초기에는 모든 클라이언트가 글로벌 모델을 기반으로 로컬 업데이트를 수행한다. 각 원격대학은 자신의 데이터를 사용하여 모델을 로컬에서 훈련시키고, 이 훈련된 모델의 가중치를 KERIS 서버에 전송한다.
- **단계 3(계층적 클러스터링)**. KERIS 서버는 모든 클라이언트의 로컬 업데이트를 기반으로 클라이언트 간의 유사성을 평가한다. 유사성에 기반하여 유사한 원격대학들로 구성된 계층적 클러스터를 형성한다.
- **단계 4(클러스터 기반 훈련)**. 유사성이 높은 원격대학들로 구성된 각각의 클러스터는 독립적으로 훈련을 진행한다. 이러한 절차를 통해 Non-IID 데이터 분포를 가진 클라이언트 간의 유사성을 기반으로 특화된 글로벌 모델을 훈련한다.

이러한 일련의 과정을 통해, KERIS는 각 원격대학의 학습 데이터 분포가 유사한 대학들끼리 클러스터를 형성할 수 있다. 이렇게 형성된 클러스터는 연합 학습에서 독립적으로 학습을 진행하게 되며, Non-IID 문제를 완화하고 연합 학습의 성능을 향상시킬 수 있다.

Ⅴ. 분석

본 논문에서 제안하는 연합 학습 기반 추천 기법은 각 클라이언트의 데이터 프라이버시 보호와 추천 시스템의 성능 향상을 동시에 추구한다. 연합 학습은 클라이언트의 원시 데이터를 로컬에서 처리하고 결과만을 중앙 서버와 공유함으로써 데이터의 중앙 집중화를 방지하고 프라이버시를 보호한다. 또한, 제안하는 계층적 클러스터링 방법은 Non-IID 문제를 완화하고, 비슷한 분포를 가지는 클라이언트들을 클러스터로 묶어 연합 학습의 성능을 향상시킨다. 이러한 접근 방식은 교육과정 추천 시스템의 효율성과 효과성을 크게 향상시킬 수 있다.

Ⅵ. 결론

본 논문은 온라인 교육과정 추천 서비스를 분산 환경에서도 인공지능 분석이 가능하도록 연합 학습을 적용할 것을 제안하였으며, Non-IID에서도 적용 가능하도록 구성하였다. 이를 통해 다양한 형태의 온라인 교육과정에서 추천 시스템으로 활용될 수 있을 것으로 기대된다.

특히, 본 논문에서 제안한 추천 시스템을 활용하여 KERIS의 맞춤 배움길 서비스가 지금보다 더 안전하고 효율적으로 제공될 가능성이 존재한다. 이를 위해 현재 20개 원격대학의 학습 데이터가 Non-IID 특성을 띄는 것을 분석하고, 이를 개선하기 위해 Non-IID에서 적용 가능한 연합 학습 아이디어를 제안하였다. 이러한 개선 방향을 통해 20개 원격대학의 학생들의 학습 데이터가 본인의 학교에서만 관리되며, 프라이버시 노출 위험성을 줄일 수 있게 되고, KERIS는 더욱 안전하고 정확한 서비스의 제공이 가능해질 것으로 기대된다.

참고문헌

https://en.wikipedia.org/wiki/Hierarchical_clustering

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. (2017) Communication-efficient learning of deep networks from decentralized data, *Artificial Intelligence and Statistics, 54,* 1273-1282.

Briggs, C., Fan, Z., & Andras, P. (2020). Federated learning with hierarchical clustering of local updates to improve training on non-IID data, *Proceedings of International Joint Conference on Neural Networks (IJCNN).*

Cao, L. (2016). Non-IID Recommender Systems: A Review and Framework of Recommendation Paradigm Shifting, *Engineering, 2*(2), 212-244.

Yang, L., Tan, B., Zheng, V. W., Chen, K., & Yang, Q. (2020). Federated Recommendation Systems, *Federated Learning,* 225-239.

Luo, S., Fu, S., Luo,Y., Liu,L., Deng,Y., & Wang, S. (2023) Privacy-Preserving Federated Learning with Hierarchical Clustering to Improve Training on Non-IID Data, *Proceedings of International Conference on Network and System Security,* 13983.

Hanzely, F., & Richtárik, P. (2020) Federated learning of a mixture of global and local models, arXIV, 2002.05516.

Yang, Q., Liu, Y., Chen, T., & Tong, Y. (2019)

Federated machine learning: Concept and applications, *ACM Transactions on Intelligent Systems and Technology, 10*(2), 1–19.

Ghosh, A., Chung, J., Yin, D., & Ramchandran, K. (2020) An efficient framework for clustered federated learning, Proceedings of the *34th International Conference on Neural Information Processing Systems (NIPS'20),* 19586-19597.

Perifanis, V., & Efraimidis, P. S. (2022) Federated Neural Collaborative Filtering, *Knowledge-Based Systems,* 242(C)

투고일자: 2023. 9. 4.
심사일자: 2023. 9. 27.
게재확정일자: 2023. 10. 6.

미래사회
Journal of Future Society
2023. Vol. 14, No. 3, 01~11.

Privacy-Preserving Recommender System for Online Course Enrollment

Ji Young Chun Geontae Noh

Seoul Cyber University

This paper proposes a recommender system based on federated learning to preserve user privacy. Traditio nal recommender systems often experience performance degradation when dealing with Non-IID (Non-Indep endent and Identically Distributed) data. Furthermore, when data is distributed across various institutions, it necessitates centralized modelling, leading to privacy concerns. To address these challenges, our proposed system employs hierarchical clustering for server-client selection strategies, coupled with federated learning techniques. This approach effectively mitigates recommendation performance issues arising from Non-IID distributions and simultaneously addresses user privacy concerns. The privacy-preserving recommender system presented in this paper is particularly beneficial for online educational course recommendation scenarios and can be applied to services such as the tailored learning paths provided by KERIS.

Keywords: Federated Learning, Non-IID, Recommender System, Education, Privacy