연합 학습을 이용한 개인 맞춤형 방송 정보 제공 플랫폼 연구

김현수, 문남미
호서대학교 벤처대학원 융합공학과
bluejfet@gmail.com, nammee.moon@gmail.com

A Study of Federated Learning base Broadcast Information recommendation platform

Hyunsoo Kim, Nammee Moon
Dept. of Convergence Engineering, Hoseo Graduate School of Venture, Hoseo Univ.

요 약

본 논문은 개인의 정보를 외부로 유출하지 않고, 소비자 방송 수신 단말 장치에 저장된 데이터를 이용하여 머신 러닝 모델을 학습하고, 소비자가 원하는 맞춤 방송 정보를 제공하는 시스템을 구글의 연합 학습[1] 을 기반한 설계에 관한 것이다. 이를 위하여, 소비자 사용 패턴 및 행동 데이터를 수집하고 저장하며 머신 러닝 학습을 진행 하는 단말 구조와 단말에서 생성된 학습 모델 파라미터 정보를 수집하고 평균화 하는 중앙 서버의 구조를 연구하고, 연합 학습을 이용한 학습 정보를 이용하여 개인 맞춤형 방송 정보를 제공하는 시스템을 연구한다.

1. 서론

주변의 수많은 디바이스 들이 인터넷으로 연결되면서 헤아릴 수 없는 양의 빅데이터를 생산하고 있는 시대에 우리는 살고 있다. 빅 데이터와 인공지능의 결합은 데이터를 기반으로 한 보다 정확한 분석 과 예측을 가능 하게 함으로써 삶의 질을 보다 윤택하게 하고 있다. 하지만, 개인 정보 보호에 대한 높은 관심과 법제화를 통한 규제는 보안에 민감한 데이터를 이용한 기계 학습 및 활용에 제약이 된다. 본 논문은 일상에서 사용하고 있는 방송 수신 단말 장치에 저장된 개인 정보 데이터를 외부로 유출하지 않고 기계학습 모델을 학습하고, 학습된 데이터를 이용하여 개인 맞춤형 방송 정보를 제공하는 시스템을 설계하고 제안한다.

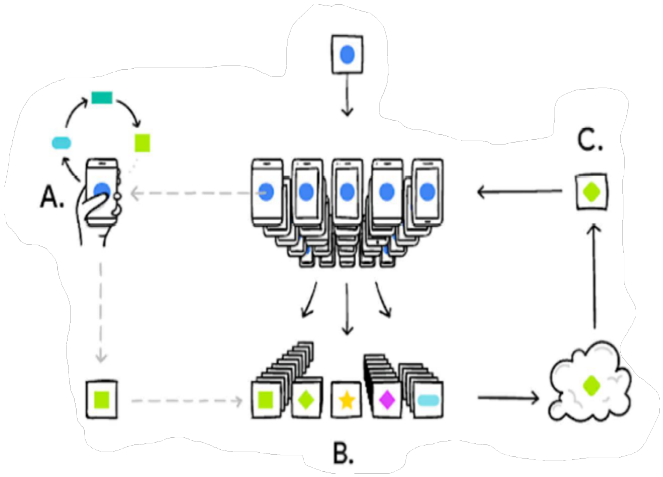
2. 시스템 설계를 위한 관련 연구

본 논문에서 제안하는 개인 맞춤형 방송 정보 제공 시스템은 구글에서 2016 년 발표된 연합 학습[1] 기술에 기반을 두고 있으며, 기계 학습 모델 학습을 위한 컴퓨팅 자원은 단말 장치의 자원을 활용하여 진행 한다. 다음과 같은 관련 분야의 연구가 선행되었다.

1) 연합 학습

구글에서 제안된 머신 러닝 모델 학습 기법으로서, 모바일 디바이스에 저장된 개인 정보를 중앙 서버로 전송하지 않고, 인공지능 모델 학습을 진행 할 수 있는 방법이다.

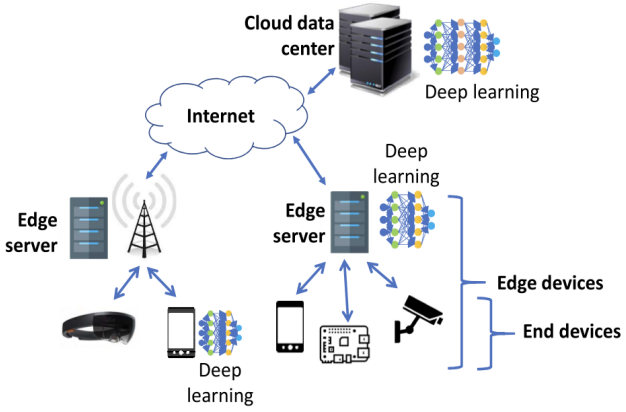
기존 인공지능 모델 학습은 모든 데이터를 중앙 서버에 모으고, 이를 라벨링 한 후 학습을 진행하는 방법으로 높은 학습 정확도와 빠른 학습 결과 도출이 가능 하였다. 하지만 민감한 개인 정보가 포함된 데이터의 경우 관련 정보 보안법의 규정으로 인한 정보 수집 및 활용에 제약이 발생함으로, 중앙 서버 방식은 적절하지 않다. 또한 분산된 데이터의 특성상 Non-IID 특성의 데이터를 이용하여 좋은 성과를 얻기 위한 연구들도 활발하게 진행되고 있다[2]. 분산된 데이터의 특성을 고려하고, 개인 정보를 보호하면서 인공지능 모델을 학습하기 위하여 고안된 방법이 연합학습[1] 이다.



(그림 1) 구글의 연합 학습 방법

2) 엣지 디바이스 머신러닝 처리 기법

산발적으로 발생되는 데이터를 중앙 서버에 모으고 이를 활용하여 인공 지능 모델을 학습함으로써 양질의 서비스를 제공하는 방법은 일반적인 인공지능 모델을 위한 시스템 설계의 기본 방법이다. 하지는 이는 개인 정보 보안 이슈 해결 방안 고려 및 데이터를 한군데 저장하기 위한 많은 노력과 비용이 요구되어 지며, 빅데이터를 한번에 처리하기 위한 높은 사양의 컴퓨팅 자원이 필요하다. 이와 같은 문제점들을 해결하기 위하여 데이터가 발생한 기기 혹은 근거리의 엣지 컴퓨팅을 이용하여 데이터를 처리하고 통합하는 기술이 점차 확산되고 있으며, 관련 연구들이 활발히 진행 되고 있다. [3]



(그림 2) 엣지 디바이스 기반 딥러닝 실행

3) 방송 시청 정보 추천 알고리즘

개인화 추천 시스템은 주로 협업 필터링과 내용 기반 필터링 방법이 사용된다. 내용 기반 필터링은 사용자의 입력 프로파일을 기반으로 사용자의 특성을 파악하고, 사용자의 선호 정보에 적합한 콘텐츠의 내용 분석을 통하여 추천하는 방식이며, 입력된 정보에 대한 정확성에 기반을 두고 추천함으로 정보가 부정확 하거나 변하는 경우 정확도가 떨어진다. 협업 필터링의 경우는 사용자들의 평가 정보를 바탕으로 유사도가 높은 콘텐츠의 추천이 가능한 필터링 기법으로서 기존 평가가 없는 경우나 새로운 사용자가 가입하는 경우 충분한 예측을 하기 어렵다. [4]

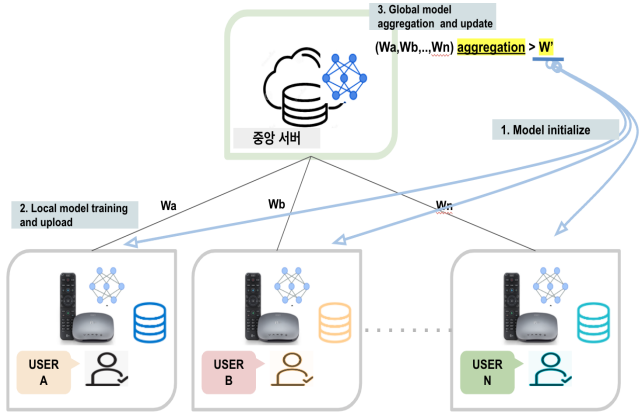
본 논문에서는 협업 필터링 기법을 적용하기 위하여 장르별 유사도가 높은 단말들을 그룹화 하고, 그룹간의 학습 결과를 취합, 평균화 함으로써, 적은 데이터를 이용하여 높은 예측률이 가능하도록 하였다.

(그림 3) 사용자 기반 협업 필터링

4) 연합 학습용 Framework

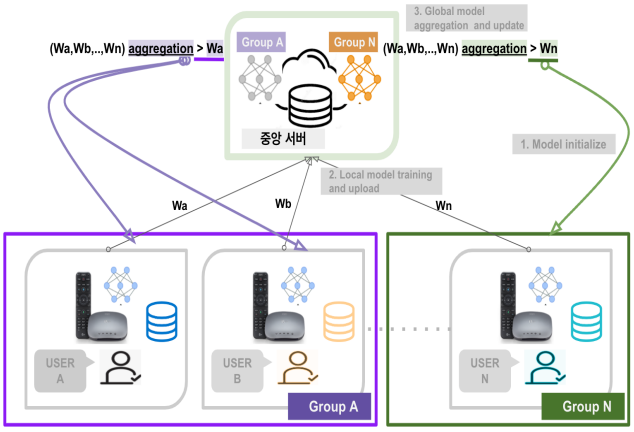
연합학습은 중앙 서버가 아닌 사용자 단말 단에 저장된 데이터 와 컴퓨팅 리소스를 이용하여 인공지능 학습을 진행 한다. 높은 컴퓨팅 자원을 이용하지 못함으로 단말의 자원을 활용하여 학습을 가능하게 하는 다양한 플랫폼들이 지원되고 있다. 본 논문에서는 구글에서 제공하는 TFF (TensorFlow Federated) framework 를 이용하여 시스템을 구축하는 것을 제안한다. TFF 는 아직 상용 수준의 기능은 제공하고 있지 않지만, 실험적인 시뮬레이션이 가능한 수준의 기능을 제공[5] [6] 한다. TFF 를 이용한 시뮬레이션은 구글 클라우드 플랫폼과 도커를 이용하여 실행 가능하다.[6]

3. 제안 시스템 구성



(그림 4) 연합 학습 기반의 개인 맞춤형 방송 정보 시스템 구성

제안 시스템 구성은 방송단말, 중앙서버 2 가지로 구성한다.
- 방송 단말 장치는 개인의 성향에 맞는 방송 콘텐츠를 제공하고 사용자의 방송 시청 정보를 장치 내의 저장소에 정해진 형태로 저장한다. 방송 단말장치내에 설치된 TFF 기반의 머신러닝 학습 모델은 저장된 유저 개인 정보를 학습 데이터로 활용하여 최적의 W 값을 도출 한다.
- 각 방송 단말에서는 시청 및 구매 이력 정보를 활용하여 선호 콘텐츠 정보를 확인하고, 협업 필터링을 위한 그룹화를 한다.
- 그룹화된 방송 단말은 해당 그룹에 맞는 협업 필터링을 이용하여 모델을 학습하고, 최적의 파라미터 값을 생성하여 서버로 전송한다.
- 중앙서버는 각 그룹별 파라미터 값을 취합, 평균을 취해 해당 그룹에 속한 단말의 모델을 업데이트 한다.
- 상기 과정의 반복적인 진행을 통하여 최적의 개인 맞춤 방송 정보를 제공한다.



(그림 5) 유사도 기반 그룹별 협업 필터링 구성도

4. 결론

본 논문에서는 연합학습을 이용하여 개인 정보의 유출 없이 머신러닝 모델을 학습 하고, 데이터의 품질과 양이 다른 Non-IID 데이터를 이용하여 최적을 학습 결과를 생성하며, 사용자 단말에 저장된 방송 데이터의 유사도에 따른 그룹화를 통하여, 높은 추천 정보의 예측이 가능한 시스템을 연구 하였다.

참고문헌

[1] H. Brendan McMahan, et al. “Communication-Efficient Learning of Deep Networks from Decentralized Data” arXiv:1602.05629v3 (2017)
[2] Yue Zhao, et al. “Federated Learning with Non-IID Data” arXiv:1806.00582v1 (2018)
[3] JIASI CHEN, et al. “Deep Learning With Edge Computing: A Review” Proceedings of the IEEE, Vol. 107, No.8. August 2019
[4] Jieun Son, et al. "Review and Analysis of Recommender Systems." Journal of the Korean Institute of Industrial Engineers41.2, 185-208, 2015
[5] Ghosh Bratin 3035437692 “Federated Learning Platform for Covid-19 Detection” FYP 20060 Final Report (2021)
[6] Chaoyang He, el al. “A Research Library and Benchmark for Federated Machine Learning” arXiv:2007.13518v4 (2020)