인공신경망에서의 블록체인 기반 개인정보보호 기술 동향

강예준[1], 김현지[1], 임세진[1], 김원웅[1], 서화정[1]
[1]한성대학교 IT융합공학과
etus1211@gmail.com, khj1594012@gmail.com, dlatpwls834@gmail.com,
dnjsdndeee@gmail.com, hwajeong84@gmail.com

Trends in Blockchain-based Privacy Preserving Technology for Artificial Neural Networks

Yea-Jun Kang[1], Hyun-Ji Kim[1], Se-Jin Lim[1], Won-Woong Kim[1], Hwa-Jeong Seo[1]
[1]Dept. of IT Convergence Engineering, Hansung University

요 약

최근 딥러닝이 다양한 분야에서 활용됨에 따라 중앙 집중식 서버, 적대적 공격 그리고 데이터 부족 및 독점화와 같은 다양한 문제점이 발생하고 있다. 또한 연합학습을 수행할 경우, 클라이언트가 잘못된 기울기를 서버에 제공하거나 서버가 악의적인 행동을 할 경우 심각한 문제로 이어질 수 있다. 이와 같은 보안 취약점을 해결하기 위해 딥러닝에 블록체인을 결합하여 중앙 집중식 서버를 분산화하고 각 참여자 노드에게 인센티브를 줌으로써 신뢰할 수 있는 데이터를 수집하는 기법이 연구되고 있다. 본 논문에서는 위와 같이 딥러닝의 문제점을 해결하기 위해 블록체인이 어떻게 적용되었는지 살펴본다.

1. 서론

최근 딥러닝은 다양한 분야에서 활용되고 있다. 하지만 오늘날의 딥러닝은 대부분 중앙 집중식 서버를 통해 학습하거나 연합학습을 수행한다. 하지만 이는 적대적 공격, 데이터 위·변조, 데이터셋 부족 및 독점화 그리고 클라이언트의 잘못된 기울기 제공과 같은 문제점이 존재하며, 서버가 악의적으로 행동할 경우에는 심각한 문제로 이어진다. 이러한 위협에 대응하기 위해 서버를 분산화 시키고 데이터를 각 클라이언트로부터 수집함으로써 데이터셋 부족 및 독점화를 방지하는 기법이 요구된다.

2. 관련 연구

2.1 블록체인

블록체인은 중앙 서버에서 장부를 관리하는 방식이 아닌, 암호화된 전자 장부를 거래에 참여한 모든 노드에 분산시켜 공유하는 기술이다[1]. 즉, 동일한 장부를 블록체인 네트워크 참여자들이 공동으로 기록 및 관리하는 탈중앙화한 방식이다. 블록체인은 제 3자가 거래를 중개함으로써 생기는 비용을 절감할 수 있으며 중앙 시스템에 의존하지 않아도 된다는 장점이 있다. 또한 보안이 강력하고 위·변조가 어렵다는 특징을 가지고 있어, 데이터 원본의 무결성을 보장할 수 있다. 이에 따라 최근 데이터의 무결성 증명이 요구되는 다양한 공공·민간 영역에서 활용되고 있으며, 이 외에도 다양한 분야에서 활용되고 있다.

2.2 인공신경망

딥러닝은 여러 층을 가진 인공신경망을 통해 학습이 수행된다. 인공신경망이란 인간의 뇌 속에 있는 뉴런의 연결구조를 본떠 만든 네트워크 구조이다[2]. 인간의 뇌에는 수많은 뉴런이 존재하며, 하나의 뉴런은 다른 뉴런으로부터 신호를 받고 또 다른 뉴런에게 신호를 전달한다. 이를 컴퓨터로 구현한 것이 인공신경망이다. 인공신경망에는 입력층, 은닉층, 출력층이 있다. 입력층은 학습하고자하는 데이터를 입력 받는 층이다. 그리고 신경망 외부에서는 접근할 수 없는 은닉층을 거쳐 출력층을 통해 최종 결과가 출력된다. 머신러닝과의 차이점은 데이터로부터 특징을 스스로 추출하여 학습한다는 점이다. 딥러닝은 컴퓨터 비전, 음성 인식, 자연어 처리, 신호 처리 등 다양한 분야에서 적용되고 있다.

2.3 순환신경망

순환신경망은 인공신경망의 한 종류로, 시계열 데이터를 학습하기 위한 인공신경망이다[3]. 순환신경망은 과거의 출력 데이터를 다시 입력으로 받는 재귀적인 구조를 가지고 있다. 즉, 신경망을 거친 벡터를 출력하기도 하고, 다음 은닉층 노드에 입력하기도 한다. 따라서 현재 결과가 이전 결과의 영향을 받으며 이전 단계에서 얻은 정보가 지속된다. 이러한 특성으로 순환신경망은 시계열 데이터를 다루기에 최적화된 신경망이다. 하지만 단계를 반복할 수록 과거 단계의 정보들이 지속되지 못하는 장기의존성 문제가 존재한다. 이를 해결하기 위해 새로운 모델인 장단기메모리(Long Short-Term Memory, LSTM)가 생겼으며 이는 Forget Gate, Input Gate, Output Gate 총 3개의 gate를 추가함으로써 장기의존성 문제를 해결하였다[4]. Forget Gate는 과거 정보를 얼마나 보존할 것인가를 결정하는 gate이고, Input Gate는 새로 입력된 정보를 얼마나 장기 상태에 활용할 것인가를 결정하는 gate이다. 그리고 마지막으로 Output Gate는 출력 정보를 얼마나 다음 단계로 활용할 것인가를 결정하는 gate이다. 최근에는 순환신경망보다 장단기메모리을 더 많이 사용하는 추세이며, 이는 자연어 처리, 음성 인식, 텍스트 생성 등 다양한 분야에서 활용되고 있다.

3. 인공신경망에서의 블록체인 기반 개인정보보호 기술 동형

딥러닝에서 발생하는 중앙 집중화, 데이터셋 부족 및 독점화 문제와 같은 문제점을 해결하기 위해 블록체인을 활용하는 연구가 진행되었다. 또한 연합학습에서 발생할 수 있는 보안 취약점을 보완하기 위해 블록체인을 활용하여 클라이언트의 악의적인 행동을 방지하는 연구도 있다. 이외에도 IoT 상에서 발생할 수 있는 문제점을 해결하기 위해 딥러닝을 블록체인과 결합하는 연구가 진행되고 있으며, 아래에서 소개하는 논문 외에도 [5], [6], [7] 등이 있다.

3.1 드론 식별 및 비행모드 감지

[8]에서는 드론을 식별하고 드론의 비행모드를 감지하기 위해 무선주파수 신호에 대해 인공신경망을 사용하는 프레임워크를 개발하였다. 해당 프레임워크에서는 스마트 계약을 기반으로 5G 네트워크와 무선주파수 신호를 통해 안전하게 데이터를 공유할 수 있다. 비행 모드에 따라 달라지는 무선주파수 신호는 해시함수를 통해 암호화되며, 블록체인의 각 블록은 이전 블록의 해시값을 포함한다. 체인의 모든 블록은 연결되어 있으며 블록에 담겨있는 데이터는 스마트 계약과 다른 노드 간의 암호화된 거래 목록을 관리하여 데이터의 위·변조를 방지한다. 따라서 시스템이 분산되어 있고 안정적인 5G 네트워크를 통해 데이터를 주고 받기 때문에 속도가 빠르다. 또한 네트워크에 참여하는 모든 노드가 실시간으로 데이터를 확인할 수 있어 데이터의 신뢰성이 높아진다. 인공신경망의 모델은 블록체인 네트워크로부터 받은 무선주파수 신호의 데이터 블록에서 드론을 식별하고 드론의 비행 모드를 감지한다. 탐지 및 자동화를 위해 엣지 디바이스 내부에 배포되어 있으며, 시계열 데이터를 다룰 수 있는 순환신경망이 사용되었다. 모델의 정확도는 2-class, 4-class, 10-class 탐지에 대해 각각 99.8, 84.5, 46.8%를 달성하였다.

3.2 DeepChain 프레임워크

[9]에서는 연합 학습에서 발생하는 여러가지 보안 문제를 블록체인을 통해 해결하였다. 연합 학습이란 다수의 로컬 클라이언트가 기울기를 서버에 업로드하여 서버가 수집된 기울기로 모델의 매개변수를 업데이트하는 학습 방법이다. 하지만 연합 학습은 로컬 클라이언트가 잘못된 기울기를 서버에 업로드할 수도 있고, 서버가 악의적으로 행동할 수도 있다. 이를 해결하기 위해 [9]에서는 DeepChain이라는 분산되고 안전하며 공정한 딥러닝 프레임워크를 제안하였다. 해당 프레임워크는 블록체인 기반으로 개발되었다. 이를 통해 각 클라이언트가 악의적인 행동을 하지 못하고 클라이언트의 데이터 프라이버시를 보장하여 전체 학습 과정에 대해 감사 가능성을 제공한다.

3.3 블록체인을 활용한 분산화 및 협력적인 딥러닝

머신러닝은 일반적으로 고도로 중앙집중화되어 있으며 데이터셋은 독점적이다. 또한 모델의 예측을 종종 쿼리별로 판매하는 경우도 있다. 모델을 새로운 데이터셋을 통해 계속해서 다시 학습시키지 않을 경우 해당 모델은 빠르게 구형 모델이된다. 이와 같은 문제점을 해결하기 위해 [10]에서는 참가자들이 협력적으로 데이터셋을 구축하고 스마트 계약을 사용하여 지속적으로 업데이트 되는 모델을 구축할 수 있는 프레임워크를 제안하였다. 해당 모델은 블록체

인에 공개적으로 공유되며 무료로 사용할 수 있다. 또한 모델에 신뢰할 수 있는 데이터를 제공하기 위해 인센티브 구조를 제안하였다. 즉, 블록체인을 통해 모델을 훈련시키고 신뢰할 수 있는 데이터를 수집하기 위해 인센티브 구조를 채택함으로써 모델의 정확도를 높였다.

3.4 DeepBlockIoTNet 프레임워크

최근 자율주행과 같은 차세대 사이버물리시스템에서 사물인터넷이 발달함에 따라 높은 정확도와 낮은 지연율로 빅데이터 분석이 요구되고 있다. 효율적인 빅데이터 분석을 위해 딥러닝은 강력한 분류 기능을 지원한다. 하지만 기존 연구는 중앙 집중화, 적대적 공격, 보안 및 개인정보보호와 같은 문제점을 가지고 있다. 해당 문제를 해결하기 위해 [11]에서는 IoT 네트워크를 위해 블록체인을 사용하여 안전한 딥러닝 접근을 할 수 있는 프레임워크 DeepBlockIoTNet을 제안하였다. 해당 프레임워크에서는 딥러닝 연산이 엣지 노드에 분산되고 안전한 방식으로 수행된다. 즉, 블록체인을 통해 안전한 딥러닝 연산을 제공하여 중앙 집중화 문제를 해결하였다. 하지만 센서가 다양하고 서로 다른 센서가 다른 종류의 데이터를 수집하여 다른 딥러닝 모델을 생성하는 문제점이 발생하였다. 이는 전이 학습 및 센서 융합 기술을 사용하여 해당 문제를 완화시킬 수 있다.

4. 결론

본 논문에서는 블록체인을 딥러닝과 결합하여 딥러닝에서 발생하는 문제점을 해결하는 기법에 대해 살펴보았다. 대부분 딥러닝의 중앙 집중식 서버 문제점을 해결하기 위해 블록체인을 활용하였으며, 데이터의 위·변조를 방지하기 위해 블록체인을 활용하는 경우도 있었다. 또한 일반적으로 딥러닝 학습에 필요한 데이터셋은 독점적이다. 이를 방지하기 위해 블록체인 네트워크 참여자들로부터 신뢰할 수 있는 데이터를 수집하여 참여자에게 인센티브를 줌으로써 신뢰할 수 있는 모델을 구축하는 연구들도 있다. 추가적으로 연합학습에서 발생하는 문제점을 방지하기 위해 블록체인을 활용하는 연구도 진행되고 있다. 이와 같이 딥러닝을 블록체인과 결합하여 중앙 집중식 서버, 데이터셋 독점, 연합학습에서의 보안 문제 등을 해결한다면, 블록체인과 딥러닝은 더 많은 분야에 적절히 적용될 수 있을 것으로 생각된다.

5. Acknowledgement

이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임(No.2018-0-00264, IoT 융합형 블록체인 플랫폼 보안 원천 기술 연구, 100%).

참고문헌

[1] Nakamoto, Satoshi. "Bitcoin: A peer-to-peer electronic cash system." Decentralized Business Review (2008): 21260.

[2] Wang, Sun-Chong. "Artificial neural network." Interdisciplinary computing in java programming. Springer, Boston, MA, 2003. 81-100.

[3] Mikolov, Tomas, et al. "Recurrent neural network based language model." Interspeech. Vol. 2. No. 3. 2010.

[4] Greff, Klaus, et al. "LSTM: A search space odyssey." IEEE transactions on neural networks and learning systems 28.10 (2016): 2222-2232.

[5] Zhu, Xudong, Hui Li, and Yang Yu. "Blockchain-based privacy preserving deep learning." International Conference on Information Security and Cryptology. Springer, Cham, 2018.

[6] Rathore, Shailendra, Yi Pan, and Jong Hyuk Park. "BlockDeepNet: a Blockchain-based secure deep learning for IoT network." Sustainability 11.14 (2019): 3974.

[7] Kumar, Rajesh, et al. "Blockchain-federated-learning and deep learning models for covid-19 detection using ct imaging." IEEE Sensors Journal 21.14 (2021): 16301-16314.

[8] Gumaei, Abdu, et al. "Deep learning and blockchain with edge computing for 5G-enabled drone identification and flight mode detection." Ieee Network 35.1 (2021): 94-100.

[9] Weng, Jiasi, et al. "Deepchain: Auditable and privacy-preserving deep learning with blockchain-based incentive." IEEE Transactions on Dependable and Secure Computing 18.5 (2019): 2438-2455.

[10] Justin, D., and B. W. Harris. "Decentralized & collaborative ai on blockchain." Proceedings of the 2019 IEEE International Conference on Blockchain (Blockchain), Atlanta, GA, USA. 2019.

[11] Rathore, Shailendra, and Jong Hyuk Park. "A blockchain-based deep learning approach for cyber security in next generation industrial cyber-physical systems." IEEE Transactions on Industrial Informatics 17.8 (2020): 5522-5532.