

시맨틱통신을 위한 GAN 기반 데이터 복원 기술

한대진, 나웅수*

공주대학교

djhan8733@gmail.com, *wsna@kongju.ac.kr

GAN-based data restoration technique for semantic communication

Deajin Han, Woongsoo Na*

Kongju National University

요약

최근 센서를 이용한 수집 데이터 전송 및 기기 간의 통신에서 무선통신의 사용량이 늘어 이의 정확한 데이터 및 빠른 전송이 요구되고 있다. 기존의 무선통신에서 소요 시간 증가 및 데이터의 손실이 가장 크게 발생할 수 있는 부분인 재전송으로 인한 HOL Blocking 발생 및 패킷 누락이 발생할 수 있다. 본 논문에서는 재전송으로 인한 시간 지연 및 데이터의 손실에 대하여 GAN을 이용한 복원을 통해 송신자의 재전송률을 근본적으로 낮추는 기법을 제안한다. 해당 기법의 우수성을 입증하기 위해 본 연구의 성능 평가를 위해 SolarCube의 태양열데이터의 전송 과정 중 데이터 Missing, Corrupted, Inconsistent의 손실이 일어났음을 가정 및 수행하였고, 그 결과 각각 90.23%, 93.12%, 93.73%가 나왔으며 이를 통해 데이터 전송 과정 중의 손실 검출 및 복원을 할 수 있다는 것을 확인할 수 있었다.

I. 서론

최근 컴퓨팅 기술의 발전으로 무선통신의 속도 향상 및 데이터를 수집하는 센서들의 성능 향상으로 무선 네트워크를 이용한 다양한 서비스 시도가 늘고 있다. 하지만 무선통신은 유선통신에 비해 장애물, 기상 조건 등 다양한 외부 요인이나 물리적 요인에 의해 데이터 전송에서의 속도가 일정하지 않으며 이는 데이터 손실 가능성으로 연결된다. 이에 따라 데이터 전체의 시간 지연으로 인해 발생하는 문제를 해결하기 위해 다양한 기법이 제안된다. [1]

본 논문에서 송신 과정에서의 시간 지연 감소를 최소화하기 위해 수신자 측에서 수신받은 데이터의 손실을 감지하고 이를 GAN을 이용해 복원하여 재전송률을 낮춰 전체적인 속도 향상 기법을 기술한다. [2]

II. 이론고찰 및 선행연구

Generative Adversarial Nets(GAN)는 Generator Model(G Model)과 Discriminator Model(D Model)을 이용한 학습 방법이다. 이는 GAN의 프레임워크의 컨셉인 '경쟁'을 이용한 방식으로 D Model은 입력받은 sample data가 G Model이 만들어 낸 sample data인지, 실제 training data distribution인지 판별하는 것을 학습하게 된다. 이때 G Model은 학습 데이터를 기반으로 D Model을 속이기 위한 진짜 같은 모델을 만들게 되며 반대로 D Model은 입력받은 데이터가 진짜인지거짓인지를 판단해 서로 적대적 학습을 통해 능력을 키워주는 것이다.

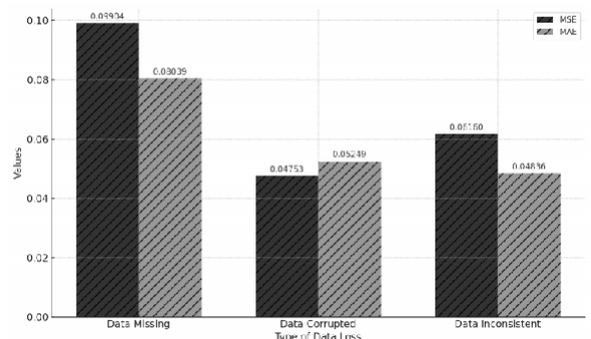
GAN은 학습 데이터를 바탕으로 가짜 데이터 및 학습 데이터에 근거해 입력 데이터가 올바른지 판단이 가능하며 이를 이용하여 누락 및 손상된 데이터를

복구하는 복원에 활용할 수 있다. 이러한 복원 과정에서도 모델링을 통해 G Model과 D Model의 적대적 학습이 가능하며 이는 G Model의 실제 데이터와 유사하도록 손상된 데이터를 복원하고 D Model은 G Model이 원본과 유사할 정도로 복원이 되었는지를 판별하게 할 수 있다.

III. 실험 및 결과 분석

본 연구에서는 태양열 데이터인 SolarCute Train set으로 학습 진행 후 test set을 이용하여 평가를 진행하였다. 해당 모델의 평가 분석을 위해 적용 상황인 무선 통신에서의 데이터 손실에 따른 복원을 위해 전송누락, 파일훼손, 데이터 불일치와 같은 손실이 일어났음을 가정했다.

그림 1은 진행한 실험 평가에 대해 선택한 수치적 평가 지표인 Mean Squared Error(MSE)와 Mean Absolute Error(MAE)에 대한 각 손실 유형에 따른 분석 결과이다.



<그림 1. 데이터 손실별 복원 결과>

먼저 MSE 는 예측값과 실제 값 간의 차이를 제공하여 평균화한 지표로, 본 연구 결과에서 전체적인 복원된 데이터가 원본 데이터는 얼마나 근접하는지를 측정하고 복원 과정에서 발생할 수 있는 큰 오차에 대한 민감도를 평가하기 위해 사용되었다. MSE 측정 결과로 각각 0.09903, 0.04752, 0.061595 가 나왔으며 이를 분석한 결과는 다음과 같다. 데이터 손실의 3 가지 유형 중 MSE 중심으로 분석하였을 때 Data Missing 이 가장 높게 나왔으며 이는 NaN 처리된 데이터의 복원이 다른 유형에 비해 더 어려웠음을 확인할 수 있다. 이는 복원에 있어서 오차가 상대적으로 크게 발생했음을 알 수 있다. 반대로 Data Corrupted 의 경우 가장 낮게 나타났으며 데이터의 변조 중 값의 부호 변경은 복원에서의 오차가 상대적으로 낮은 것을 확인할 수 있었으며 이는 다른 손실 유형에 비해 유의미한 결과를 나타내었다.

MAE 는 복원한 값과 실제 값 간의 절대 차이를 평균화한 지표로 실질적인 복원을 계산 및 본 연구에서 제안한 기법에 대해 평가를 위해 사용되었다. MAE 측정 결과로 각각 0.08038, 0.052488, 0.048361 로 이에 대한 분석 결과는 다음과 같다. MAE 측정 결과로 보았을 때 전체적인 결과는 MSE 와 같이 Data Missing 부분의 결과가 상대적으로 높게 나왔지만, Data Inconsistent 의 MAE 가 Data Corrupted 와 비슷해진 것을 보아 상대적으로 오차의 영향을 덜 받는 MAE 의 특성으로 인해 실제 복원을 진행하였을 때 전반적으로 복원의 평가가 좋은 것을 확인할 수 있으며 일부 데이터의 오차율이 큰 것을 확인할 수 있다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학 ICT 연구센터육성지원사업 (IITP-2023-RS-2022-00156353) 및 2023 년도 교육부의 재원으로 점연구소 지원사업 (2019R1A6A1A03032988)의 지원을 받아 수행된 연구임.

참 고 문 헌

- [1] E Ryu, M Ahn, J Park, J Lim, D Seong, and J Yoo. A congestion control scheme for minimizing data-loss in wireless multimedia sensor networks. *Journal of the KIISE: Computing Practices and Letters*, 19(5):278-282, 2013.
- [2] Wenzhong Li, Han Zhang, Shaohua Gao, Chaojing Xue, Xiaoliang Wang, and Sanglu Lu. Smartcc: A reinforcement learning approach for multipath tcp congestion control in heterogeneous networks. *IEEE Journal on Selected Areas in Communications*, 37(11):2621-2633, 2019.