

논문 2023-18-01

CNN 기반 대용량 시계열 데이터 압축 기법연구

(A Study of Big Time Series Data Compression based on CNN Algorithm)

황 상 호, 김 성 호, 김 성 재, 김 태 근*

(Sang-Ho Hwang, Sungho Kim, Sung Jae Kim, Tae Geun Kim)

Abstract : In this paper, we implement a lossless compression technique for time-series data generated by IoT (Internet of Things) devices to reduce the disk spaces. The proposed compression technique reduces the size of the encoded data by selectively applying CNN (Convolutional Neural Networks) or Delta encoding depending on the situation in the Forecasting algorithm that performs prediction on time series data. In addition, the proposed technique sequentially performs zigzag encoding, splitting, and bit packing to increase the compression ratio. We showed that the proposed compression method has a compression ratio of up to 1.60 for the original data.

Keywords : Data Compression, Time-series data, CNN, Machine Learning, Delta encoding

I. 서 론

농장과 공장에서는 재배와 생육환경관리를 위한 스마트 농장 기술과 제조, 유통, 검수 등의 공정에서 효율성 향상 및 품질 개선을 위한 스마트공장 및 디지털 트윈 기술이 점차 보급되고 있다.

디지털 트윈 기술의 보급으로 농장 또는 공장 내 설비 및 주변 환경에 설치되는 센서의 수는 급격히 늘어나고 있다 [1]. 최근에는 재배와 제조 공정에서의 효율성 향상, 또는 사용하고 있는 설비에 대한 고장 예지 등의 목적으로 인공지능 기술이 적용되고 있으며 인공지능 모델의 신뢰성 향상 및 분석 정확성 향상을 위해 방대한 양의 센서 데이터를 필요로 하고 있다 [2-4].

일반적으로 단말노드에서 수집된 센서 데이터는 분석, 모니터링, 저장을 위해 엣지 장비, 클라우드 데이터센터로 전송된다. 데이터 전송과 저장에는 많은 에너지가 소비되고 있다 [5, 6]. 이러한 데이터 전송 에너지 소비량 감소 또는 필요 저장 공간의 크기를 줄이기 위해 데이터 압축 기술이 필요하며, 센서에서 추출되는 데이터는 주로 시간의 순서에 따라 상관관계를 가지는 정형적인 시계열 데이터임을 이용하여 기존 압축 알고리즘들에 비해 압축율을 높이고 있다 [7-11].

본 논문은 IoT (Internet of Things) 장비에서 생성되는 시계열 데이터 저장에 필요한 용량을 줄이기 위해 CNN (Convolutional Neural Networks) 및 Delta 인코딩 (Encoding)을 선택적으로 적용하는 무손실 압축 기법을 제

안한다. 본 논문에서 제안하는 기법은 시계열 데이터에 대한 예측을 수행하는 Forecasting 알고리즘에서 CNN 및 Delta 인코딩을 상황에 따라 선택적으로 적용하여 인코딩 데이터의 크기를 줄였으며, Zigzag 인코딩, Splitting, Bit Packing을 순차적으로 적용하여 압축율을 높이고 있다.

이하 논문의 구성은 다음과 같다. 2장에서는 시계열 데이터 압축과 관련한 기존 연구에 대하여 기술한다. 3장에서는 본 논문에서 제안하는 하이브리드 시계열 데이터 압축 기법을 소개하고, 4장에서는 실험을 통해 기존의 기법들과 비교하고, 5장에서는 결론 및 향후 연구과제에 대하여 기술한다.

II. 배경지식 및 관련연구

압축 기법에는 손실 압축 기법 (Lossy Compression)과 무손실 압축 기법 (Lossless Compression)이 있다. 손실 압축 기법은 압축된 데이터의 복원 시 원본데이터와 동일하게 복원되지 않고 일부의 데이터정보가 손실되고, 무손실 압축 기법은 복원 시 원본 데이터와 동일한 데이터를 얻을 수 있다.

일반적으로 손실 압축 기법은 무손실 압축 기법에 비해 압축율이 높은 반면, 인코딩/디코딩 과정에서 소실되는 정보들이 발생한다. 인공지능, 빅데이터 분석 등의 기술이 발달함에 따라 많은 분야에서 무손실 압축의 활용도가 높아지고 있다 [7, 8].

무손실 데이터 압축 기법에는 ZIP, bzip2, 허프만 부호화, RLE (Run-Length Encoding) 등이 있으며, 이들 알고리즘들은 중복되는 문자열 또는 패턴을 같은 문자열로 치환하는 방법으로 압축을 수행하는 기법들이다. 이 알고리즘들은 과일로 저장될 수 있는 모든 데이터 형태에 대하여 압축을 수행할 수 있지만 압축율이 낮다는 단점이 있다.

일반적으로 시계열 데이터는 시간의 흐름에 따라 연속적으로 저장된 데이터로 현재의 값과 이전의 값이 상관관계를

*Corresponding Author (tgkim@knps.or.kr)

Received: Sep. 21, 2022, Revised: Oct. 27, 2022, Accepted: Jan. 2, 2023,
Sang-Ho Hwang: Gyeongbuk Institute of IT Convergence Industry
Technology (Senior Researcher)

Sungho Kim: Gyeongbuk Institute of IT Convergence Industry
Technology (Senior Researcher)

Kim Sung Jae: SMARTGIO (Research Director)

Kim Tae Geun: Korea National Park research institute (Researcher)

※ 본 연구는 중소벤처기업부의 규제자유특구혁신사업육성 지원에 의한
연구임 [P0020333].

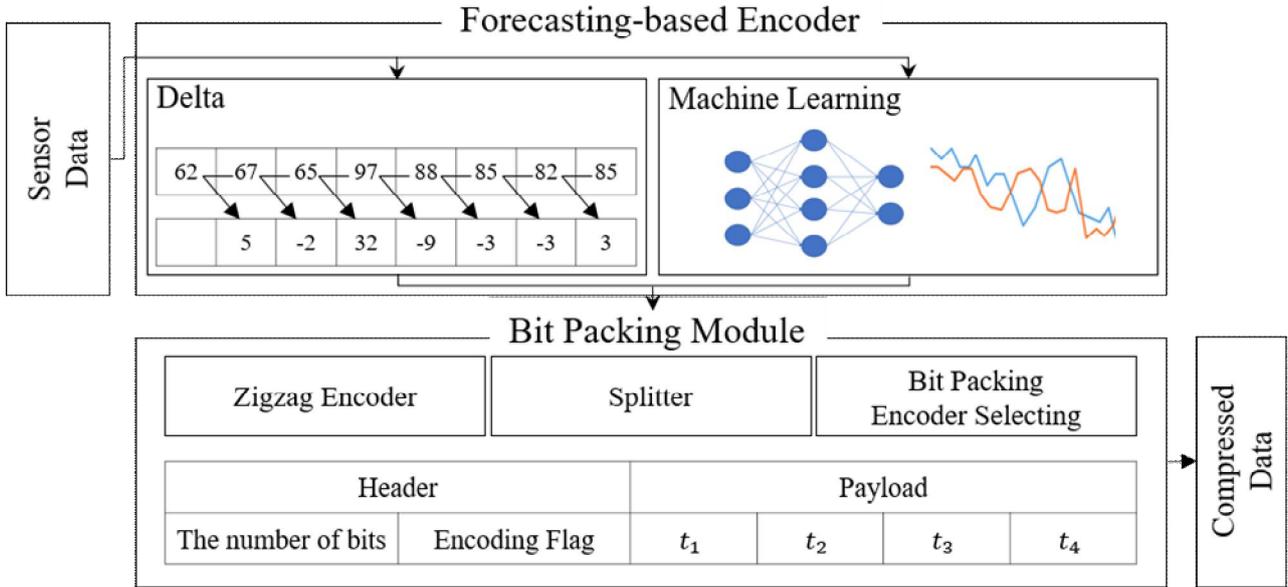


그림 1. 제안하는 압축기법의 구조
Fig. 1. Structure of the proposed compression method

가지는 특징이 있다. 이러한 시계열 데이터의 특징을 활용하여, 최근에는 점차 용량이 늘어나고 있는 시계열 데이터를 압축하기 위한 기법에 대한 연구와 활용이 늘어나고 있다 [9].

시계열 데이터에 대한 무손실 압축 기법들 중, GORILLA 알고리즘은 Facebook의 메모리 시계열 데이터베이스에서 사용하기 위해 설계되었다 [10]. GORILLA 알고리즘은 시계열 데이터에 포함되어 있는 타임스탬프 값을 2차 미분 (Delta of Deltas)으로 표현하여 사용하는 비트수를 줄이고, 시계열 데이터 값을 이전 값과의 XOR 코딩을 사용하여 표현에 필요한 비트 수를 줄이는 방법으로 압축을 수행한다. 이 기법은 값이 느리게 변하거나 자주 센싱하여 거의 비슷한 값이 연속되는 시계열 데이터에는 압축 효율이 높지만, 급격하게 변하거나 패턴이 없는 형태의 시계열 데이터에서는 압축 효율이 급격하게 떨어지는 단점이 있다.

Sprintz 알고리즘은 정수형 시계열 데이터 압축 기법으로 Davis Blalock 등에 의해 제안되었다. Sprintz 알고리즘은 시계열 데이터에 대한 예측을 수행하는 Forecasting, 예측된 값과 실제 값의 차이 (error)의 배열에 대한 비트를 줄이는 Bit packing, 0으로 이루어진 값들에 대한 비트를 줄이는 RLE (Run Length Encoding), 엔트로피 부호화로 허프만 부호화를 순차적으로 적용하고 있다 [11]. Forecasting을 통해, 시계열 데이터의 특징을 충분히 활용하여 데이터의 양을 줄이고, 그 후 전통적인 무손실 압축 알고리즘들을 순차적으로 적용함으로써 압축율을 높이고 있다. Sprintz 알고리즘은 Forecasting 모듈에 Delta, AR (Aggressive Regression) 등을 적용할 수 있으나, 이들 알고리즘을 고정적으로 적용하여 데이터의 종류와 패턴 변화에 따른 압축율의 편차가 큰 단점이 있다.

III. 하이브리드 시계열 데이터 압축 기법

그림 1은 제안하는 압축 기법의 전체적인 구조를 보여주고 있다. 제안하는 기법은 크게 시계열 데이터에 대한 예측을 기반으로 인코딩을 수행하는 Forecasting기반 인코더 모듈과 헤더 및 페이로드로 구성된 구조에 압축된 데이터를 나누어 저장하는 비트패킹 모듈 2가지로 구성되어 있다. 그림 1에서 첫 번째 단계인 Forecasting기반 인코더 모듈은 CNN 또는 Delta 인코딩에 기반하여 시계열 데이터 예측 및 예측 값과 실제 값의 차이를 추출한다. 두 번째 단계인 Zigzag 인코더는 최상위 비트에 위치하고 있는 부호비트 위치를 최하위 비트로 변경하고 세 번째 단계인 Splitter는 시계열 데이터를 고정된 수의 데이터를 한 개의 그룹으로 묶어주며 마지막 단계인 Bit Packing & 인코딩 선택 (Encoding Selecting)에서는 Forecasting기반 인코더 모듈에서 사용한 CNN과 Delta 인코더 중 가장 압축효율이 높은 인코더를 선택한다.

1. 인코딩 절차

1.1. Forecasting기반 인코딩

Forecasting기반 인코딩은 인코더로 Delta기반 인코딩과 CNN기반 인코딩 2가지를 사용하고 있다. Delta 인코딩은 이전 값과 차이를 이용하는 것으로 수식 1에 기반하고 있으며, 기본적으로 현재 값과 이전 값의 차이가 없을 때 0으로 시간의 흐름에 따라 값의 변화가 크지 않는 시계열데이터에서 높은 압축효과가 난다.

$$E_i = x_i - x_{i-1}. \quad (1)$$

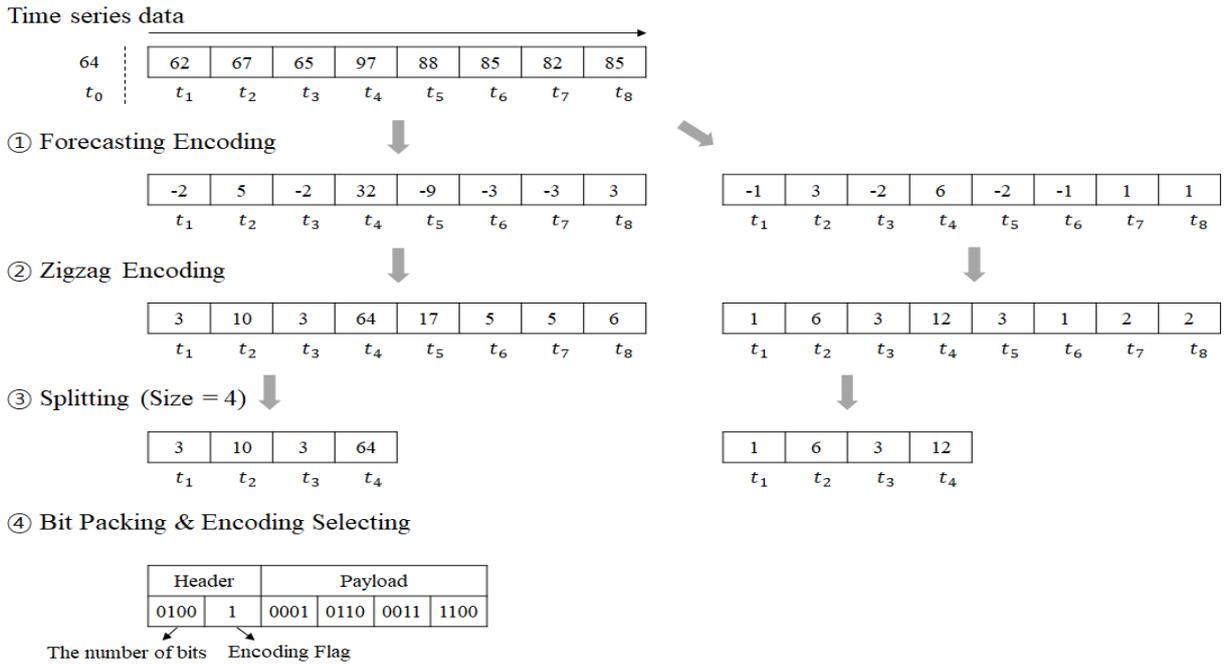


그림 2. 제안하는 압축 기법 동작의 예
Fig. 2. Example of the operation of the proposed compression technique

수식 1에서 E_i 는 i 번째 delta인코딩에 따른 차이 값을 의미하고 x_i 는 i 번째 센서 데이터를 의미한다.

CNN기반 인코딩은 시계열데이터 예측에 CNN 모델을 적용하고 있다. CNN은 연속된 데이터를 1차원 이미지로 취급하여 연속된 데이터에서 두드러진 요소를 추출한다. 추출된 특징에 기반하여 다음 데이터에 대한 예측을 수행하고 실제 데이터와의 차이 값을 압축에 활용한다. CNN기반 인코딩은 복잡하지만 일정한 패턴이 있는 시계열데이터에 높은 압축 효율을 보인다.

제안하는 기법은 압축 성질이 다른 2가지의 인코딩 기법을 동시에 활용하여 인코딩 선택기 (Encoding Selector) 모듈에서 가장 압축효율이 높은 인코딩을 선택하도록 한다. 그림 2의 ①은 Forecasting 인코딩의 예를 보여주고 있다. 그림 2에서 t_i 는 i 번째 시계열 데이터를 나타낸다.

1.2. Zigzag 인코딩

Forecasting 인코딩을 통해 추출된 예측된 값과 실제 값의 차이 값은 음수도 포함될 수 있다. 이러한 음수 값은 일반적으로 최상위 비트가 부호비트이기 때문에 필요한 최소 비트를 기준으로 압축을 수행하는 Bit Packing에서 압축 효율을 떨어뜨린다. 이를 방지하기 위해, Zigzag 인코딩은 부호비트를 최상위 비트에서 최하위비트로 변경하고 수식 2는 Zigzag 인코딩을 의미한다.

$$(n \ll 1) \wedge (n \gg k - 1). \tag{2}$$

그림 2의 ②는 ①의 결과를 바탕으로 Zigzag 인코딩을 수행한 예를 보여주고 있다.

1.3. Splitting

Zigzag 인코딩을 통해 추출된 값은 일정한 간격으로 그룹을 형성하여 Bit Packing이 이루어진다. 이렇게 일정한 그룹으로 묶는 과정을 Splitting이라 하며 그룹의 크기는 사용자 정의에 따른다. 그룹의 크기가 클수록 헤더 생성에서 발생하는 추가정보공간에 대한 비용이 줄어들지만 급격한 값의 변화율을 가진 시계열 데이터에서는 압축효율이 떨어질 수 있다. Splitting의 크기는 고정적으로 결정되고 알고리즘 동작에서 문맥적으로 적용한다. 그림 2의 ③은 ②의 결과를 바탕으로 Splitting 과정을 보여주고 있다. 그림 2의 ③에서는 분할크기가 4로 설정되어있어 $t_1 \sim t_4$ 의 데이터가 페이로드에 저장되게 된다.

1.4. Bit Packing & 인코딩 선택 (Encoding Selecting)

이 절차에서는 Splitting이 완료된 값들은 그룹별로 Bit Packing을 수행하고 대상이 될 인코딩 값을 선택한다. 앞선 절차로 인코딩된 값은 2가지 종류가 있으며, 이 중에서 Bit Packing시 데이터 크기를 더 줄일 수 있는 인코딩을 선택하고 인코딩 플래그 (Encoding Flag) 비트에 Delta 인코딩인 경우 0, CNN기반 인코딩인 경우 1을 기록한다. 그림 2의 ④는 ③의 결과를 바탕으로 수행한 결과를 보여주고 있다. 그림 2의 ④번에서 Zigzag 인코딩 후 분할된 데이터 중에서 CNN기반 인코딩으로 추출된 데이터의 사용될 수 있는 최대비트가 4로 Delta 인코딩의 6에 비해 작으므로 4번에서 선택된다. 헤더의 비트의 수는 4로, 인코딩 플래그는 1로 기록한다.

2. 디코딩 (Decoding) 절차

압축된 시계열 데이터에 대한 디코딩 절차는 제안한 기법이 수행한 인코딩 절차의 역순으로 진행된다. 첫 번째로 패킹된 구조의 헤더 안에 페이로드 데이터의 비트 수와 Forecasting 인코딩 플래그를 참조하여 언패킹 (Unpacking) 을 수행한다. 그 다음 Zigzag 디코딩을 수행하고 Forecasting 플래그에 해당하는 Forecasting 디코딩을 수행한다.

IV. 시험평가

1. 데이터 셋 및 실험환경

제안하는 기법의 성능을 평가하기 위해서, 본 실험에서는 UCI Machine Learning Repository내 데이터 셋을 활용하였다. 제공하고 있는 데이터 셋들 중 센서로부터 수집된 시계열 데이터를 포함하고 있는 데이터 셋 2개를 활용하였고, 해당 데이터 셋은 Gas sensor array temperature modulation과 Beijing Multi-Site Air-Quality이다 [12, 13]. Gas sensor array temperature modulation 데이터 셋은 가스 챔버 내 CO₂, 습도, 온도, MOX (Metal Oxide Semiconductor) 등을 측정된 데이터 값을 시간의 순으로 측정하여 제공하고 있으며, 본 논문에서는 2016년 10월 1일에 측정된 데이터를 활용하고 있다. Beijing Multi-Site Air-Quality 데이터 셋은 중국의 환경 모니터링 센터에서 제공하는 것으로 SO₂, NO₂, CO, O₃, 온도, 기압, 풍속 등의 시계열 데이터가 포함되어 있다. 본 논문에서는 Aotizhongxin 공기질 모니터링 데이터를 활용하였다. Gas sensor array temperature modulation 데이터 셋의 전체 용량은 576MB이며 Beijing Multi-Site Air-Quality 데이터 셋의 전체 용량은 31MB이며 그 중 일부의 데이터를 활용하였다. 제안하는 기법은 정수형 데이터 압축 기법으로 포맷의 형태가 1차원의 시계열 데이터인 경우 대용량까지 적용가능하다. 본 논문의 실험에서 사용한 시계열 데이터 셋은 최솟값, 최댓값을 기준으로 16bit 정수형 데이터로 양자화하여 사용하였다.

그림 3과 4는 Gas sensor array temperature modulation과 Beijing Multi-Site Air-Quality 데이터 셋의 특징을 보여주고 있다. 그림에서 사각형의 하단은 1사분위수이며 상단은 3사분위수를 나타내며 가운데 선은 중앙값이다. 사각형과 연결된 선의 상단은 최댓값이며 하단은 최솟값을 나타내며 점은 특이 값으로 경계를 벗어나는 값에 해당한다.

그림 3에서 R1~R3에 해당하는 센서 데이터는 낮은 평균값을 가지고 있으며 R4~R8의 센서 데이터는 상대적으로 평균값이 높다. R1과 R2의 경우에는 다른 데이터에 비해 특이 값이 많은 특징이 있다. 그림 4에서 PM2.5, PM10, NO₂, O₃ 센서 데이터는 평균이 낮은 반면 특이 값이 많이 분포하는 특징이 있으며 TEMP, PRES, DEWP 센서 데이터는 상대적으로 높은 평균값과 특이 값이 없는 특징이 있다.

성능 평가를 위해 실험에서는 압축되지 않은 원본 데이터와 Sprntz 알고리즘과 비교하였다. 시험평가에서 Sprntz

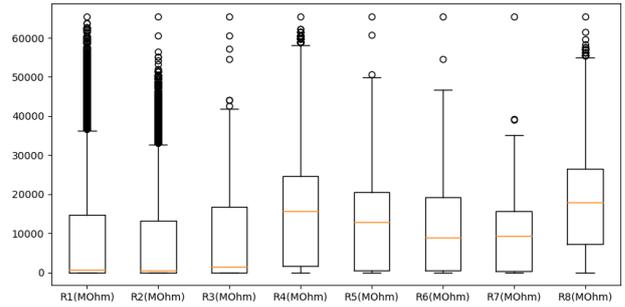


그림 3. Gas sensor array temperature modulation 데이터 셋의 특성
Fig. 3. Characteristics of Gas sensor array temperature modulation data set

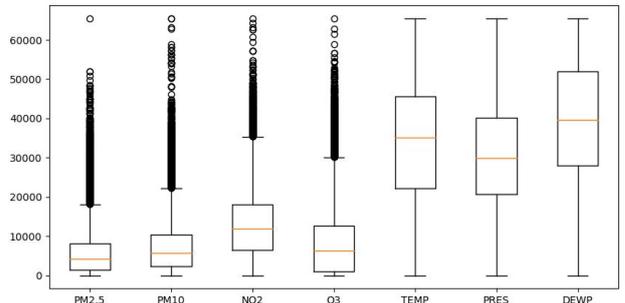


그림 4. Beijing Multi-Site Air-Quality 데이터 셋의 특성
Fig. 4. Characteristics of Beijing Multi-Site Air-Quality dataset

표 1. 제안하는 압축 기법의 CNN의 네트워크 구조
Table 1. Network structure of CNN of the proposed compression technique

Layer type (options)	Output shape
Input	(None, 4, 1)
Conv1D(filters=128, kernel size=2, activation=selu)	(None, 3, 128)
Conv1D(filters=128, kernel size=2, activation=selu)	(None, 2, 128)
MaxPooling1D(pool size=2)	(None, 1, 128)
Flatten	(None, 128)
Dense(units=64, activation=selu)	(None, 64)
Dense(units=64, activation=selu)	(None, 64)
Dense(units=1)	(None, 1)

알고리즘의 Forecasting은 Delta 인코딩을 사용하였다. 제안하는 기법의 Splitting 크기는 Sprntz 알고리즘과 동일하게 8로 설정하였고, Forecasting 인코딩에 사용한 CNN의 네트워크 구조는 표 1과 같다.

CNN기반 인코딩은 이전 4개의 데이터를 활용하여 다음 데이터에 대한 예측을 수행하며, 실험 내 모든 시계열 데이터에 대하여 동일한 네트워크를 적용하고 데이터 셋내의 절반의 데이터를 학습데이터로 사용하였고, 나머지 절반의 데이터에 대하여 압축 성능 평가 데이터로 활용하였다.

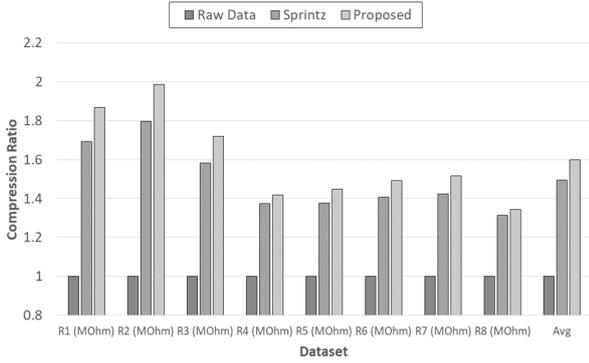


그림 5. Gas sensor array temperature modulation 데이터 셋에서의 압축율

Fig. 5. The compression ratio in gas sensor array temperature modulation dataset

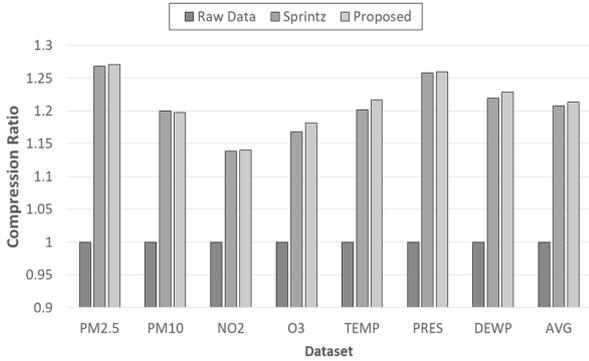


그림 6. Beijing Multi-Site Air-Quality 데이터 셋에서의 압축율
Fig. 6. The compression ratio in Beijing Multi-Site Air-Quality dataset

2. 실험결과

본 논문에서는 제안하는 기법의 성능을 평가하기 위해서 Sprintz 기법과 압축 성능비교를 하였다. 압축 성능 비교에는 수식 3의 압축율을 사용하였다.

$$Compression\ Ratio = \frac{Uncompressed\ size}{Compressed\ size} \quad (3)$$

그림 5는 Gas sensor array temperature modulation 데이터 셋에서의 압축율을 보여주고 있다. 그림 3에서 Raw Data는 압축 전의 원본 데이터를 의미하고 Sprintz는 Delta 인코딩을 적용하고 있다. Sprintz는 평균 1.49의 압축율을 보이고 있으며 제안하는 기법은 평균 1.60의 압축율을 보이고 있다. 그림 3의 센서 데이터간의 차이로 인하여 데이터간 압축율의 차이가 존재한다. 시계열 데이터내 값의 편차가 적을수록 압축율이 높은 경향을 보이고 있다.

그림 6은 Beijing Multi-Site Air-Quality 데이터 셋에서의 압축율을 보여주고 있다. 그림 6에서 Sprintz 기법은 1.20의 압축율을 보이고 있으며 제안하는 기법은 1.21의 압축율을 보이고 있다. 이에 대한 분석은 그림 9와 10에서 추가한다.

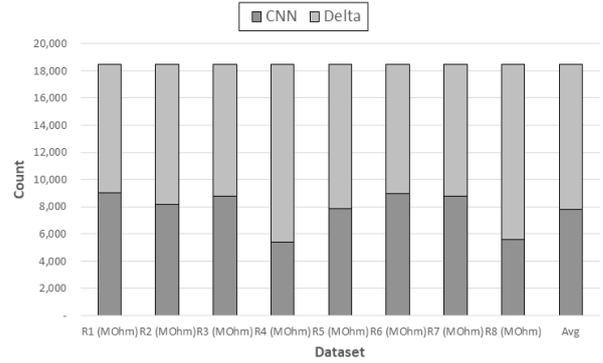


그림 7. Gas sensor array temperature modulation 데이터 셋에서 제안하는 기법이 선택한 Forecasting 인코딩의 비율
Fig. 7. Proportion of forecasting encoding selected by the proposed method in gas sensor array temperature modulation dataset

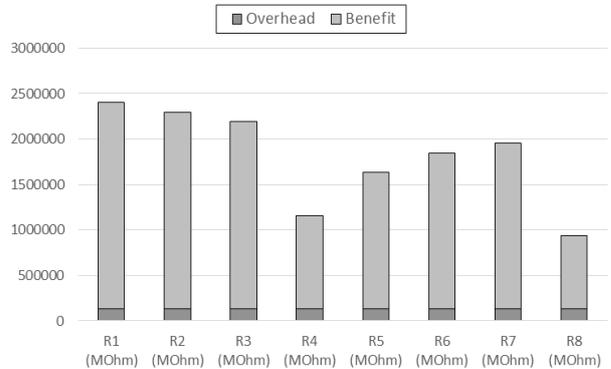


그림 8. Gas sensor array temperature modulation 데이터 셋에서 CNN인코딩을 선택함으로써 얻는 저장공간 이득
Fig. 8. Storage space gains of choosing CNN encoding in gas sensor array temperature modulation dataset

그림 7은 Gas sensor array temperature modulation 데이터 셋에서 제안하는 기법이 선택한 Forecasting 인코딩의 비율을 보여주고 있다. 그림 7에서 제안하는 기법은 평균 42.37%의 비율로 CNN기반의 인코딩을 선택하였으며 이는 그림 5에서 보는 것과 같이 제안하는 기법이 Sprintz 기법에 비해 높은 압축율을 보이는 원인이 된다.

그림 8은 Gas sensor array temperature modulation 데이터 셋에서 제안하는 기법이 선택한 CNN 인코딩으로 인해 얻는 저장 공간의 이득 (Benefit) 및 오버헤드 (Overhead)를 보여주고 있다. 그림 8에서 이득은 CNN기반 인코딩으로 인하여 순수하게 얻을 수 있는 저장 공간의 이득이며, 오버헤드는 제안하는 기법을 적용함으로써 추가되는 헤드에 대한 비용을 의미한다. 이득과 오버헤드의 합은 Delta 인코딩만 적용할 경우 필요로 하는 공간을 의미한다. 그림 7에서 제안하는 기법은 높은 확률로 CNN 인코딩을 Forecasting으로 선택하고 있으며 이는 높은 압축율의 결과로 이어진다. 반면 2개의 인코딩을 사용함으로써 발생하는 오버헤드의 비용은 크지 않음을 확인할 수 있다.

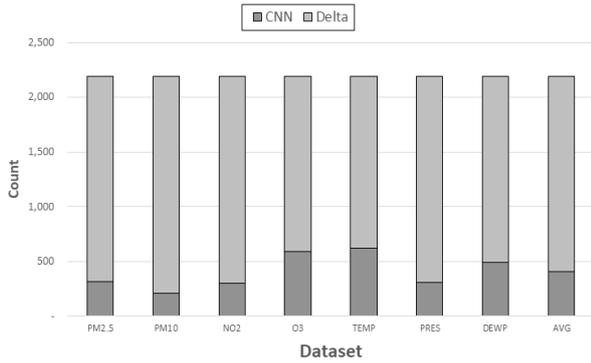


그림 9. Beijing Multi-Site Air-Quality 데이터 셋에서 제안하는 기법이 선택한 Forecasting 인코딩의 비율

Fig. 9. Proportion of forecasting encoding selected by the proposed method in Beijing Multi-Site Air-Quality dataset

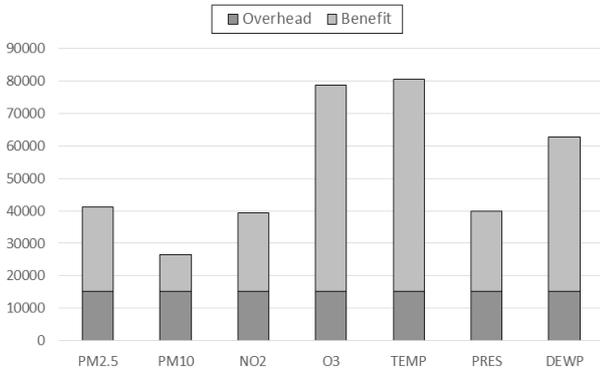


그림 10. Beijing Multi-Site Air-Quality 데이터 셋에서 CNN인코딩을 선택함으로써 얻는 저장공간 이득

Fig. 10. Storage space gains of choosing CNN encoding in Beijing Multi-Site Air-Quality dataset

그림 9는 Beijing Multi-Site Air-Quality 데이터 셋에서 제안하는 기법이 선택한 Forecasting 인코딩의 비율을 보여주고 있으며, 그림 7과 달리 CNN 인코딩을 선택한 비율이 평균 18%로 낮다. 그림 4에서 보듯이 Beijing Multi-Site Air-Quality 데이터 셋의 편차가 Gas sensor array temperature modulation 데이터 셋이 비해 높은 것과 데이터 패턴의 불규칙성에 기인한다.

그림 10은 Beijing Multi-Site Air-Quality 데이터 셋에서 제안하는 기법이 선택한 CNN 인코딩으로 인해 얻는 저장공간의 이득 및 헤더 추가로 인해 발생하는 오버헤드를 보여주고 있다. 그림 9에서 보듯이 CNN 인코딩을 선택하는 경우가 낮아서 그림 8에 비해 상대적으로 이득이 줄어들음을 확인할 수 있다.

제안하는 기법은 표 1에서 사용한 CNN 모델의 45,761개의 네트워크내 뉴런을 위해 메모리를 활용하고 있다. 이러한 메모리 및 계산의 활용으로 인해 메모리 사용량 및 계산에 오버헤드가 발생하지만 데이터 저장 공간을 줄일 수 있어 장기간 보관해야 할 시계열 데이터가 많은 분야에서 활용할 수 있다.

V. 결론

본 논문에서는 단말 장치에 부착된 센서에서 생성되는 시계열 데이터 저장에 필요한 용량을 감소시키기 위해 CNN 및 Delta 인코딩을 선택적으로 적용하는 무손실 압축 기법을 제안하였다. 본 논문에서 제안하는 기법은 시계열 데이터에 대한 예측을 수행하는 Forecasting 알고리즘에서 CNN 및 Delta 인코딩을 상황에 따라 선택적으로 적용하여 압축하는 방법을 적용하여 압축율을 높였으며, Zigzag 인코딩, Splitting, Bit Packing 및 인코딩 선택을 순차적으로 적용하여 압축율을 높이고 있다. 실험에서는 제안하는 기법이 원본데이터에 대해 최대 1.60의 압축율을 보였다. 제안하는 기법은 모든 정수형의 시계열 데이터에 대한 압축에 적용할 수 있으며 무손실 압축이 필요한 오디오 스트리밍 데이터에도 적용할 수 있다.

CNN 모델의 활용에 따른 메모리사용에 대한 오버헤드가 존재하며 이를 개선하기 위해 향후 연구에서는 시스템의 사용 에너지를 줄일 수 있도록 CNN 모델 및 적용 알고리즘에 대한 개선을 진행할 예정이다.

References

- [1] R. Krishnamurthi, A. Kumar, D. Gopinathan, A. Nayyar, B. Qureshi, "An Overview of IoT Sensor Data Processing, Fusion, and Analysis Techniques." *Sensors*, Vol. 20, No. 21, pp. 6076, 2020.
- [2] A. L. Diedrichs, F. Bromberg, D. Dujovne, K. Brun-Laguna, T. Watteyne, "Prediction of Frost Events Using Machine Learning and IoT Sensing Devices." *IEEE Internet of Things Journal*, Vol. 5, No. 6, pp. 4589-4597, 2018.
- [3] R. Akhter, S. A. Sofi, "Precision Agriculture Using IoT Data Analytics and Machine Learning." *Journal of King Saud University-Computer and Information Sciences*, Vol. 34, No. 8, pp. 5602-5618, 2021.
- [4] S. Ayvaz, K. Alpay, "Predictive Maintenance System for Production Lines in Manufacturing: A Machine Learning Approach Using IoT Data in Real-time." *Expert Systems with Applications*, Vol. 173, pp. 114598, 2021.
- [5] J. Azar, A. Makhoul, M. Barhamgi, R. Couturier, "An Energy Efficient IoT Data Compression Approach for Edge Machine Learning." *Future Generation Computer Systems*, Vol. 96, pp. 168-175, 2019.
- [6] H. M. Al-Kadhimi, H. S. Al-Raweshidy, "Energy Efficient Data Compression in Cloud Based IoT." *IEEE Sensors Journal*, Vol. 21, No. 10, pp. 12212-12219, 2021.
- [7] G. Chiarot, C. Silvestri, "Time Series Compression Survey." *ACM Computing Surveys (CSUR)*, Vol. 55, No. 10, pp. 1-32, 2021.
- [8] J. Azar, A. Makhoul, R. Couturier, J. Demerjian, "Robust IoT Time Series Classification with Data Compression and Deep Learning." *Neurocomputing*, Vol. 398, pp. 222-234, 2020.

[9] J. Azar, G. B. Tayeh, A. Makhoul, R. Couturier, "Efficient Lossy Compression for IoT Using SZ and Reconstruction with 1D U-Net." *Mobile Networks and Applications*, Vol. 27, pp. 984-996, 2022.

[10] T. Pelkonen, S. Franklin, J. Teller, P. Cavallaro, Q. Huang, J. Meza, K. Veeraraghavan, "Gorilla: A Fast, Scalable, in-memory Time Series Database." *Proceedings of the VLDB Endowment*, Vol. 8, No. 12, pp. 1816 - 1827, 2015.

[11] D. Blalock, S. Madden, J. Gutttag, "Sprintz: Time Series Compression for the Internet of Things." *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 2, No. 3, pp 1-23, 2018.

[12] J. Burgues, J. M.I Jimenez-Soto, S. Marco, "Estimation of the Limit of Detection in Semiconductor Gas Sensors Through Linearized Calibration Models." *Analytica Chimica Acta*, Vol. 1013, pp. 13-25, 2018.

[13] S. Zhang, B. Guo, A. Dong, J. He, Z. Xu, S. X. Chen, "Cautionary Tales on Air-quality Improvement in Beijing." *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 473:20170457, 2017.

Sang-Ho Hwang (황상호)



2009 Computer Engineering from Yeungnam University (B.S.)
 2013 Computer Engineering from Yeungnam University (M.S.)
 2017 Computer Engineering from Yeungnam University (Ph.D.)

2017~2019 Daegu Gyeongbuk Institute of Science & Technology (Researcher)

2019~Gyeongbuk Institute of IT Convergence Industry Technology (Senior Researcher)

Field of Interests: Big data, Deep learning, Embedded systems
 Email: shhwang@gitc.or.kr

Sungho Kim (김성호)



2012 Computer Engineering from Yeungnam University College (B.S.)
 2019 Computer Engineering from Yeungnam University (Ph.D.)
 2019~Gyeongbuk Institute of IT Convergence Industry Technology (Senior Researcher)

Career:

2021~Electrical/Electronic/SW Division, G-ARD in GITC

2021~Evaluation Committee, TIPA

Field of Interests: Big data, Deep learning, Embedded systems
 Email: shk@gitc.or.kr

Sung Jae Kim (김성재)



2003 Remote Sensing and Geographic Information System from Kyungil University (B.S.)
 2005 Remote Sensing and Geographic Information System from Kyungil University (M.S.)

2009 Remote Sensing and Geographic Information System from Kyungil University (Ph.D.)

Field of Interests: Remote Sensing, Big data, Deep learning
 Email: kimsungjae97@hanmail.net

Tae Geun Kim (김태근)



1990 Landscape Architecture from Mokpo University (B.S.)
 1997 Landscape Architecture from Mokpo University (M.S.)
 2007~Korea National Park research institute (Research Fellow)

2008 Geographic Information System from Inha University (Ph.D.)

Field of Interests: GIS, Remote Sensing, Big data
 Email: tgkim@knps.or.kr