논문 2024-19-24

# 에지노드에 적용 가능한 무손실 시계열데이터 압축 기법

## (Lossless Time Series Data Compression Technique Applicable to Edge Nodes)

이 수 성, 황 상 호, 김 성 호, 윤 장 규, 박 용 완*
(Soosung Lee, Sang-Ho Hwang, Sungho Kim, Jang-Kyu Yun, Yong-Wan Park)

Abstract : In this paper, we propose an improved technique called HAB (Huffman encoding Aware Bitpacking) that enhances the Sprintz method, which is a representative time series data compression technique. The proposed technique boosts compression rates by incorporating Huffman encoding-aware bitpacking in the secondary compression stage. Additionally, it can be applied to terminal nodes or edge nodes with limited available resources, as it does not require separate parameters or storage space for compression. The proposed technique is a lossless method and is suitable for fields that require the generation of artificial intelligence models and accurate data analysis. In the experimental evaluation, the proposed HAB showed an average improvement of 14.7% compared to the existing technique in terms of compression rate.

Keywords : Time series data, Senor data, Compression, Lossless, Huffman encoding

## Ⅰ. 서 론

스마트 공장, 스마트 팜, 자율주행 및 디지털 트윈과 같은 기술이 발전함에 따라 수집되는 센서 데이터가 폭발적으로 증가하고 있다 [1, 2].

많은 분야에서 수집되는 센서 데이터는 대부분 시계열 데이터로 시간 간격에 따라 대용량으로 수집되고 있다. 이러한 대용량의 시계열 데이터는 저장, 전송 및 처리에 있어 많은 비용과 리소스가 필요하고 그 수집되는 데이터의 양과 범주가 점점 늘어나고 있다. 이러한 문제를 해결하기 위해 센서 데이터에 대한 압축 기법에 대한 연구가 현재 활발하게 이루어지고 있다 [3-5].

데이터 압축 기법은 크게 손실 압축과 무손실 압축 2가지 방법으로 이루어지고 있다. 이중 손실 압축 기법은 비슷하게 재현이 가능한 모델을 활용하는 방법을 주로 사용하고 있어 데이터 압축비가 높은 장점이 있으나 데이터 손실이 발생하여 데이터분석 및 인공지능 모델 생성 등의 용도로 수집된 데이터 압축에는 적절하지 않다 [6].

인공지능 모델 생성이나 데이터 분석을 위해서는 데이터의 완전성과 정확성이 중요하며, 인공지능 모델을 훈련시킬 때 활용되는 학습데이터는 손실 없이 정확한 정보를 보존해야 생성되는 모델이 신뢰할 수 있는 예측이 가능하다. 또한, 데이터 분석을 통해 패턴을 파악하거나 통계적인 추론을 수행할 때도 원본 데이터의 완전한 정보가 필요하기 때문에 무손실 압축 기법을 적용하는 것이 적절하다. 다만, 무손실 압축 기법은 압축비가 낮은 문제가 있어 최근에는 시계열 데이터의 특징을 활용하여 압축비를 높이는 연구가 활발하게 이루어지고 있다 [7-9].

본 논문에서는 대표적인 시계열 데이터 압축 기법인 Sprintz기법을 개선한 HAB (Huffman encoding Aware Bitpacking) 기법을 제안한다. 제안하는 기법은 2차로 압축되는 허프만 인코딩을 고려한 비트 패킹을 수행하는 방법으로 압축율을 높이고 있다. 또한 압축을 위해 필요한 헤더 정보의 크기를 줄이고 별도의 파라메타 및 저장 공간을 사용하지 않아 가용 리소스에 제한이 있는 단말 노드 또는 엣지 노드에도 적용할 수 있다. 제안하는 기법은 무손실 기법으로 인공지능 모델 생성 및 정확한 데이터 분석이 필요한 분야에 적합하다.

이하 논문의 구성은 다음과 같다. 2장에서는 시계열 데이터 압축 기법 및 허프만 인코딩 등 기존 연구에 대하여 기술한다. 3장에서는 제안하는 HAB 기법에 대한 서술하고, 4장에서는 실험을 통해 기존 시계열 압축 기법과 비교 서술한다. 마지막 5장에서는 결론 및 향후 연구과제에 대하여 기술한다.

## Ⅱ. 관련 연구

### 1. 시계열 데이터 압축 기법

손실 압축은 데이터의 일부 정보를 제거하거나 근사화하여 데이터 크기를 줄이는 방식으로 데이터의 일부 손실을 허용하기 때문에 무손실 압축에 비해 더 높은 압축률을 달성할 수 있는 장점이 있다. 그러나 손실 압축은 원본 데이

*Corresponding Author (ywpark@yu.ac.kr)
Received: May. 29, 2024, Revised: Jul. 5, 2024, Accepted: Aug. 12, 2024.
S. Lee, J. K. Yun: GITC. (Principal Researcher)
S. H. Hwang, S. Kim: GITC. (Senior Researcher)
Y. W. Park: Yeungnam University. (Professor)
※ 본 연구는 국토교통부/국토교통과학기술진흥원의 지원으로 수행되었음. (22AMDP-C162334-02 : 자동차 통합보안 안전성 평가기술 개발)
(This work is supported by the Korea Agency for Infrastructure Technology Advancement(KAIA) grant funded by the Ministry of Land, Infrastructure and Transport (Grant 22AMDP-C162334-02).)

터의 일부 정보가 손실되므로 압축된 데이터를 완벽하게 복원할 수는 없어 음원, 동영상 등 데이터의 완전한 복원이 필요하지 않은 경우에 주로 사용이 된다.

반면 무손실 압축은 주어진 데이터의 모든 정보를 보존하므로 데이터의 완전성과 정확성을 유지할 수 있다. 무손실 압축은 원본 데이터를 왜곡 없이 압축하여 저장할 수 있어 정교한 인공지능 모델 생성이나 데이터 분석과 같은 작업에 적절하다.

대부분의 센서 데이터는 시계열 데이터이기 때문에 연속된 관측치가 서로 상관관계를 가지는 특징을 활용한 시계열 데이터 전용 압축 기법에 대한 연구가 많이 이루어지고 있다 [10-14].

Tuomas Pelkonen 등은 페이스북 (Facebook)의 메모리 시계열 데이터베이스에서 데이터 압축을 위해 GORILLA 알고리즘을 소개하였다 [15]. 이 알고리즘은 타임스탬프를 Delta of Deltas로 압축을 하고 각 타임스탬프에 대응하는 시계열 데이터 값은 이전 값과의 XOR 인코딩을 하는 형태로 압축을 한다. 64bit 부동소수점 형태의 시계열 데이터에 대한 압축으로 변화가 많지 않은 시계열 데이터에 대한 압축효율이 높지만 패턴이 일정하지 않은 데이터에서는 효율이 떨어진다.

Davis Blalock 등은 Forecasting 및 Bitpacking 등을 활용하여 무손실 시계열 데이터 압축을 수행할 수 있는 Sprintz 알고리즘을 제안하였다 [16]. 이 기법은 정수형 시계열 데이터에 적용할 수 있으며, Forecasting, Bitpacking 그리고 허프만 인코딩을 활용하여 시계열 데이터 압축을 수행하고 있다. Forecasting 인코딩은 Delta, Aggressive Regression을 적용할 수 있으며 효율에 따라 동시에 적용하여 인코딩을 수행할 수도 있다. ZigZag 인코딩 및 Bitpacking을 적용한 데이터에 허프만 인코딩을 적용하여 최종적으로 압축을 수행하여 파일로 저장한다. 무손실 압축으로 높은 압축효율을 가지고 있지만 Bitpacking 시 허프만 인코딩을 고려하지 않은 단점이 있다. 본 논문에서는 Sprintz 기법을 개선하여 Bitpacking에서 허프만 인코딩을 고려하여 압축효율을 높이고 있다.

### 2. 허프만 인코딩

허프만 인코딩은 데이터의 빈도를 기반으로 압축을 수행하여 빈도가 높은 데이터를 짧은 비트로 표현하고, 빈도가 낮은 데이터를 긴 비트로 표현하는 방식으로 압축을 수행하는 기법이다. 빈도가 높은 데이터들이 많이 포함이 될수록 허프만 인코딩의 압축효율은 높아진다.

허프만 인코딩은 다음의 절차로 이루어진다. 첫 번째로 입력 데이터의 빈도를 분석하여 빈도가 높은 데이터에 짧은 코드를 할당한다. 이를 위해 허프만 트리라는 자료구조를 사용하며, 허프만 트리는 데이터의 빈도를 바탕으로 구성되며, 빈도가 높은 데이터일수록 트리의 상위에 위치한다.

압축할 데이터를 허프만 트리를 이용하여 인코딩하는 과정에서는 각 데이터에 해당하는 비트열을 할당한다. 이때, 빈도가 높은 데이터는 짧은 비트열로 표현되고, 빈도가 낮은 데이터는 긴 비트열로 표현된다. 이렇게 허프만 인코딩은 인코딩된 데이터는 빈도가 높은 데이터에 대해서는 적은 비트를 사용하므로 전체적인 데이터 크기를 줄일 수 있다.

허프만 인코딩은 압축된 데이터를 다시 원래의 데이터로 복원하기 위해 허프만 트리를 사용하며, 이를 통해 데이터의 손실 없이 압축된 데이터를 복원할 수 있다.

이러한 방식으로 허프만 인코딩은 데이터 압축을 효율적으로 수행할 수 있으며, 특히 빈도가 높은 데이터가 자주 등장하는 경우에 효율이 높아, 허프만 인코딩은 다양한 응용 분야에서 데이터 압축에 활용되고 있다.

본 논문에서는 시계열 데이터 압축에서 많이 사용되고 있는 허프만 인코딩 기법의 특징을 활용하여, 압축이 되는 시계열 데이터를 빈도를 높이는 HAB 압축 기법을 제안한다.

## Ⅲ. 허프만인코딩을 고려한 시계열데이터 압축 기법

Sprintz 기법에서 적용하고 있는 허프만 인코딩은 중복된 데이터가 많을수록 압축 효율이 높아지기 때문에 제안하는 HAB기법은 저장되는 데이터의 통계적 중복성을 높이는 방법을 통해 기존 기법에 비해 압축효율을 높인다. 그림 1은 제안하는 기법의 구조 및 동작의 예를 보여주고 있다. 제안하는 기법은 시계열 데이터를 예측하여 예측된 값과 실제 값의 차이를 계산하는 Forecasting 인코딩, 최상위 비트 (Most Significant Bit, MSB)에 존재하고 있는 부호비트를 최하위 비트 (Least Significant Bit, LSB)로 변경하는 ZigZag 인코딩, 데이터수를 일정하게 자르는 Splitting, 데이터를 저장할 포맷을 결정하여 저장하는 Bitpacking으로 구성되어있다.

Forecasting 인코딩은 연속된 관측지가 서로 상관관계를 가지는 시계열 데이터의 특징을 활용하여 앞서 관측된 데이터 값에 기반하여 다음 측정되는 데이터 값을 예측하고 그 예측 값과 실제 값의 차이를 추출한다. 복원할 때도 앞서 복원한 값을 기반으로 예측 값을 구할 수 있으므로 쉽게 복원이 가능하다. Forecasting 인코딩을 활용하여 실제 데이터 표현에 필요한 정보가 아닌 차이 값만 반영하여 데이터를 압축한다. 시계열 데이터 예측에는 Delta, Double Delta과 같은 인코딩에서부터 CNN, LSTM 등의 인공지능 모델을 적용할 수 있으나 본 논문에서는 리소스가 한정적인 단말 노드에 적용하기 위해 Delta 인코딩을 사용한다. Delta 인코딩은 바로 이전 값과의 차이를 값이다. 그림 1의 Raw Data는 Forecasting 인코딩을 거치면서 Delta 인코딩이 된 배열로 변환된다. 동작 시 최초 수집된 센서 데이터는 서버로 별도로 보내어 초기 값으로 설정한다.

Delta 인코딩을 통해 추출되는 데이터는 음수도 포함되어 있기 때문에 데이터 표현에 필요한 최대 비트를 줄이기 위해 최상위 비트에 존재하는 부호비트를 최하위 비트로 변경하는 것이 필요하며 이를 위해 ZigZag 인코딩을 적용한다. ZigZag 인코딩은 최상위비트에 위치하고 있는 부호비트를 최하위비트로 변경하는 대표적인 알고리즘으로 k비트로 표

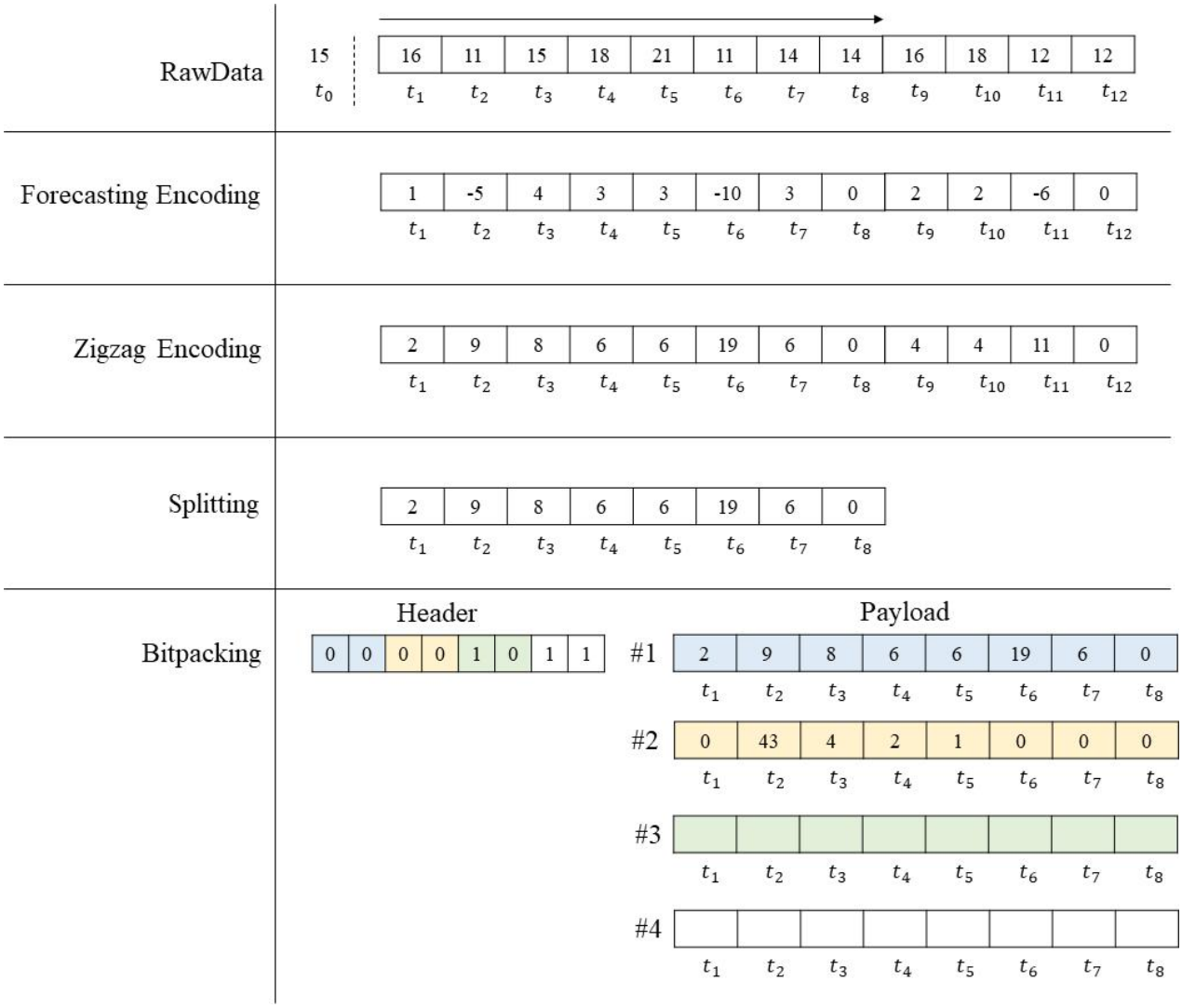


그림 1. HAB의 구조 및 동작의 예
Fig. 1. Structure and operation of HAB

현되는 데이터 n의 ZigZag 인코딩은 $(n \ll 1)\hat{}(n \gg k-1)$로 이루어진다. 그림 1의 ZigZag 인코딩에서는 Forecasting 인코딩을 통해 추출된 값의 부호비트 위치를 변경하여 배열을 재 생성한다.

시계열 데이터는 일반적으로 인근 데이터들은 비슷한 값을 가지는 특징이 있지만 값이 다른 특이 값들이 있을 수 있으며 이전 데이터로 예측되지 않는 데이터들이 포함될 수 있으므로 Forecasting 인코딩에서 큰 값들이 추출될 수 있다. 이러한 데이터들은 데이터표현을 위해 많은 비트를 차지하게 되는데 이는 데이터 압축을 어렵게 만든다. 이러한 데이터를 처리하기 위해 Splitting 및 Bitpacking을 적용해야한다.

이중 첫 번째로 적용되는 Splitting은 데이터를 일정한 길이의 그룹으로 나누는 역할을 한다. 데이터의 일부에만 특이 값이 포함되어 있다는 점에서 이러한 데이터들은 그룹을 통해 일부 데이터들과 함께 그룹을 지어 다른 데이터들과 격리함으로써 다른 데이터들이 데이터표현에 적은 비트만 소비될 수 있도록 만드는 것이다. 이러한 그룹의 크기를 적게 만들면 급격한 변화율을 가지는 시계열 데이터에 높은 압축효율을 제공할 수 있지만 데이터 그룹에 포함되는 페이로드로 인해 추가공간이 많이 소모되는 단점이 발생할 수 있다. 본 논문에서는 실험을 통하여 Splitting 길이를 8로 설정하였다. 그림 1의 Splitting에서는 센서 데이터 8개씩 묶어서 한 개의 그룹으로 생성한다.

Bitpacking은 실제 데이터 저장에 필요한 포맷으로 변경하는 과정을 수행한다. 본 논문에서는 2차로 압축을 수행하는 허프만 인코딩을 고려한 HAB을 수행한다. HAB는 우선 Splitting을 통해 들어온 그룹의 데이터 표현에 필요한 최대 비트 depth를 찾는다. 이때 HAB는 데이터 표현에 필요한 비트 depth의 크기를 허프만 인코딩에 적용하는 데이터 크기의 배수로 결정한다. 예를 들어 본 논문에서는 허프만 인코딩을 1byte로 수행하기 때문에 비트 depth의 크기를 0, 8, 16 등으로 8배수로 결정하였다. Bitpacking 헤더에는 해당

그룹의 비트 depth 크기가 있어야 디코딩 시 데이터들을 정확하게 복원할 수 있다. HAB는 이를 허프만 인코딩의 크기에 따른 배수로 결정함으로써 데이터들의 중복성을 유지하고 비트 depth의 크기를 기입하는 헤더의 사이즈를 줄임으로써 데이터 압축에 효율을 더하고 있다.

그룹 내 최대 비트 depth를 결정된 다음에, HAB는 Splitting을 통해 분리된 각각의 그룹들을 4개씩 묶어 하나의 헤더를 생성한다. 헤더는 8비트로 되어있고, 각 그룹의 2비트 정보를 연결하여 생성한다. 2비트 정보는 복원에 필요한 데이터 표현에 사용한 비트 depth의 크기를 의미하고 8비트 depth일 때 '00', 16 비트 depth일 때 '01', 0 비트 depth일 때 '10' 그리고 4개씩 묶을 때 그룹의 수가 맞지 않을 때 '11'을 표기하여 해당 payload가 비었음을 표시한다. payload에는 그룹으로 묶여진 데이터들을 저장하게 되는데 '00'과 '01'은 각각의 비트 depth 크기만큼 payload의 공간을 할당하여 저장하게 되고 나머지 '10'과 '11'은 그룹으로 묶여진 데이터가 모두 0으로 명확하거나 실제 데이터가 없는 그룹이기 때문에 payload공간을 채우지 않는다. 디코딩 시에는 헤더파일에 있는 flag를 참조하여 payload의 데이터들을 복원한다. Bitpacking이 완료된 후 파일의 헤더에는 시계열 데이터의 초기값 및 시계열데이터의 총 길이 값을 포함한 정보를 저장하여 디코딩 과정에서 활용한다.

그림 1의 Bitpacking에서는 제안하는 HAB의 예를 보여주고 있다. 그림 1에서는 Splitting이 완료된 4개의 그룹이 한 개로 묶어져서 헤더의 flag를 생성하고 있다. 첫 번째 및 두 번째 그룹은 8 비트 이내의 비트 depth를 가지는 배열로 flag에 '00'로 기입이 되며 세 번째 그룹은 해당 배열 내 값이 모두 0인 상태로 '10'을 표기한다. 그림 1의 동작의 예에서는 3개의 그룹에 해당하는 시계열 데이터를 가정하고 있으며 네 번째 그룹은 실제 배열이 존재하지 않는 상태를 의미한다. 따라서 flag 표기 규칙에 따라 네 번째의 마지막 그룹의 flag는 '11'로 표기한다. 세 번째 그룹과 네 번째 그룹은 헤더만 표기를 하고 페이로드에는 채우지 않는다. 첫 번째 와 두 번째 그룹의 데이터는 디코딩 시 페이로드의 데이터를 8bit 씩 묶어 복원한다. 세 번째 그룹의 값은 복원 시 모두 0인 8개의 배열로 복원하고 마지막 네 번째 그룹은 '11'로 데이터 복원 시 무시한다.

HAB를 적용된 값들은 최종적으로 허프만 인코딩을 통해 압축되어 파일로 저장된다. 압축된 파일에 대한 디코딩은 인코딩의 역순으로 진행된다.

## Ⅳ. 실험환경 및 실험결과

제안하는 기법의 성능을 평가하기 위해 기존 기법인 Sprintz와 비교하였다. Sprintz의 Forecasting 인코딩은 제안하는 기법과 동일하게 delta 인코딩을 적용하였으며, Spritting크기도 동일하게 8개로 설정하였다. 실험에서 사용한 데이터 셋은 UCI Machine Learning Repository에 공개되어있는 시계열 데이터 셋 중, Gas sensor array temperature modulation 데이터 셋과 Appliances energy prediction 데이터 셋을 평가에 사용하였다. Gas sensor array temperature modulation 데이터 셋은 화학적 센서를 통해 가스 챔버 내 여러 기체 정보를 수집한 데이터로 CO2, 습도, 온도, 벤젠 등을 측정한 시계열 데이터이다 [17].

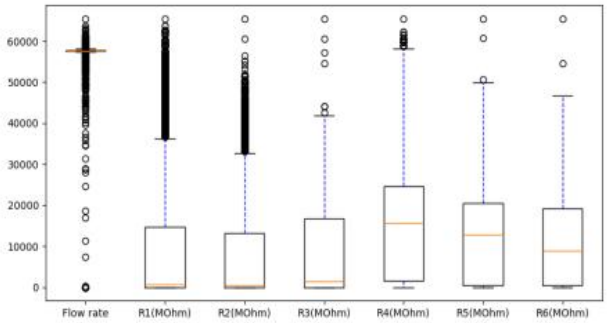


그림 2. Gas sensor array temperature modulation 데이터 셋의 특성
Fig. 2. Characteristics of gas sensor array temperature modulation data set

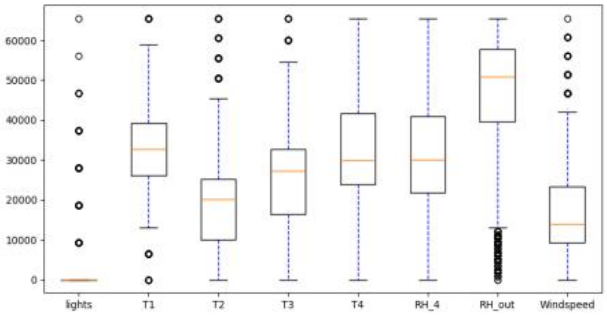


그림 3. Appliances energy prediction 데이터 셋의 특성
Fig. 3. Characteristics of Appliances energy prediction data set

Appliances energy prediction 데이터 셋은 가정의 온도 및 습도 등을 모니터링한 데이터로 4.5개월 동안 수집된 시계열 데이터로 이루어져있다 [18]. 그림 2와 3은 실험에서 활용한 데이터 셋의 특징을 보여주고 있다. 그림 2와 3에서 검은색 사각형의 하단 부분은 1사분위수 값이고 상단 부분은 3사분위수를 나타내며 주황색 선은 데이터 셋의 중간값이다. 파란색 선의 상단은 최대값, 하단은 최소값을 나타내며 최대값과 최소값 너머의 동그란 점은 특이값을 나타낸다. 그림 2의 Gas sensor array temperature modulation 데이터 셋에서 flow rate, R1, R2는 다른 데이터와 다르게 특이값이 많은 특징이 있고 R1, R2, R3은 낮은 중간값을 가지고 있다. 그림 3의 Appliances energy prediction 데이터 셋에서 대부분의 데이터는 특이값이 없는 특징이 있으나 RH_out 데이터는 많은 특이값을 가지고 있다. 실험에서 사용한 시계열 데이터는 모두 16bit 정수형 데이터로 양자화한 데이터를 활용하였다.

그림 4와 5는 Gas sensor array temperature modulation 데이터 셋과 Appliances energy prediction 데이터 셋에서 기존 기법인 Sprintz와 비교를 보여주고 있다. 성능평가에

사용한 압축율은 수식 1과 같다.

$$Compression\ Ratio = \frac{Uncompressed\ size}{Compressed\ size}. \quad (1)$$

그림 4의 그래프에서 x축은 센서 종류 및 평균이고 y축은 압축율을 나타내고 Raw Data는 압축이 이루어지기전 데이터를 의미한다. 그림 4에서 Sprintz는 최대 8.05배, 평균 3.1배의 압축율을 보이며 제안하는 HAB는 최대 9.58배 평균 3.56배의 압축율을 보여준다. 데이터 중 Flow rate의 경우 데이터의 변화율이 낮아 높은 압축율을 보이는 반면 R1~R6 챔버에서의 MOX 화합물의 경우 데이터의 변화율이 높아 상대적으로 낮은 압축율을 보이고 있다. 제안하는 HAB는 Gas sensor array temperature modulation 데이터 셋에서 Sprintz와 비교하여 평균적으로 14.6% 압축율을 개선하였다.

그림 5는 Appliances energy prediction 데이터 셋에서의 압축율 비교를 보여주고 있다. 그림 4의 Gas sensor array temperature modulation 데이터 셋과 비교하여 Appliances energy prediction 데이터 셋은 상대적으로 변화율이 낮은

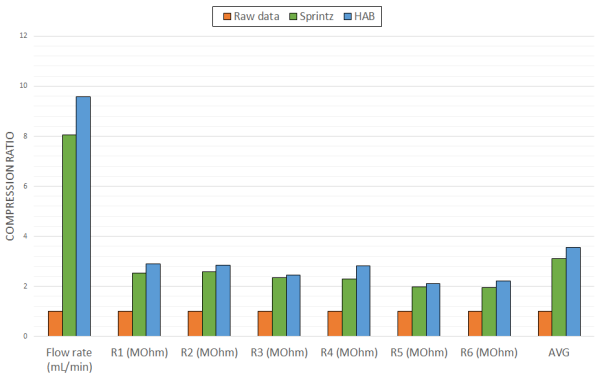


그림 4. Gas sensor array temperature modulation 데이터 셋의 압축율 비교
Fig. 4. The comparison of compression ratios of gas sensor array temperature modulation data sets

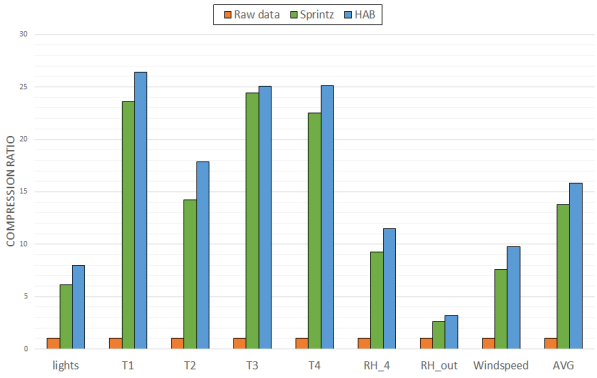


그림 5. Appliances energy prediction 데이터 셋의 압축율 비교
Fig. 5. The comparison of compression ratios of appliances energy prediction data sets

센서 데이터로 구성되어 있어 Sprintz 및 HAB 모두 높은 압축율을 보이고 있다. 그림 5에서 Sprintz는 최대 24.4배, 평균 12.66배의 압축율을 보이며 제안하는 HAB는 최대 26.41배 평균 15.85배의 압축율을 보여준다. 제안하는 HAB는 Appliances energy prediction 데이터 셋에서 Sprintz와 비교하여 평균적으로 14.8% 압축율을 개선하였다.

그림 6과 7은 Gas sensor array temperature modulation 데이터 셋내 R1과 R2 센서 데이터에서의 Sprintz 압축 기법과 제안하는 HAB 압축기법이 허프만 인코딩에 적용되기 전 통계적 중복성을 그래프로 나타낸 것이다. 그림 2에서 확인할 수 있는 것처럼, R1과 R2의 데이터는 특이값이 상대적으로 많이 분포하고 있는 데이터로 Forecasting Encoding을 통한 시계열 데이터의 예측이 어려워 인코딩된 값의 분포가 다른 데이터 셋에 비해 넓은 특징이 있다. 이에 본 논

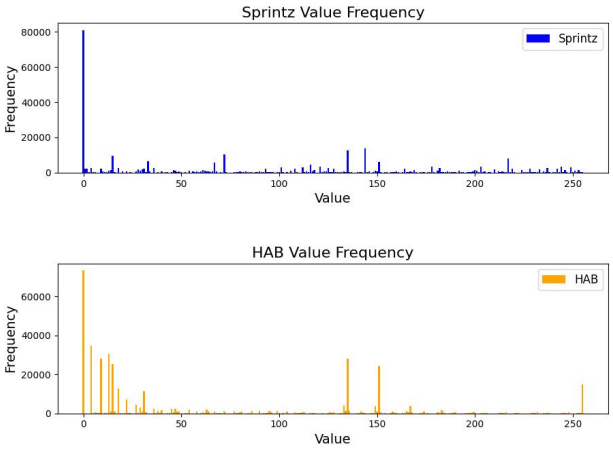


그림 6. Gas sensor array temperature modulation 데이터 셋내 R1에서의 데이터 중복성 비교
Fig. 6. The comparison of data redundancy in R1 within gas sensor array temperature modulation data set

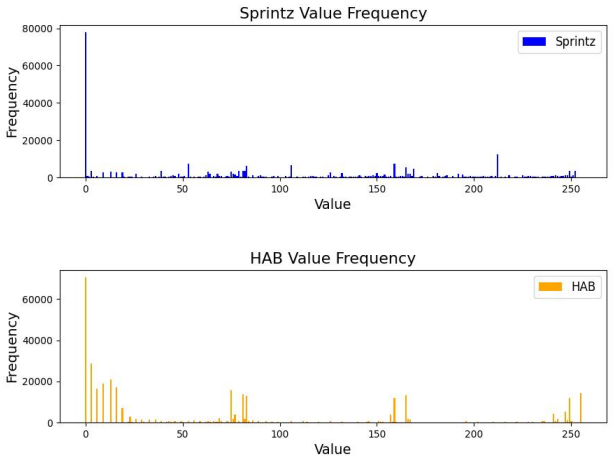


그림 7. Gas sensor array temperature modulation 데이터 셋 내 R2에서의 데이터 중복성 비교
Fig. 7. The comparison of data redundancy in R2 within gas sensor array temperature modulation data set

문에서는 대표적으로 R1과 R2의 데이터를 선택하여 데이터 중복성을 비교한다. 그래프에서 x축은 허프만 인코딩의 기본이 되는 8비트 값을 나타내며 y축은 빈도를 의미하며 위는 Sprintz 알고리즘 아래는 HAB알고리즘이다. 그림 4와 5에서 확인할 수 있듯이 제안하는 HAB가 Sprintz에 비해 데이터 중복성을 높이는 형태로 Bitpacking을 수행하기 때문에 허프만 인코딩을 적용한 후 시계열 데이터 압축율이 향상된다.

## V. 결 론

본 논문에서는 시계열 데이터 압축에서 활용할 수 있는 HAB 기법을 제안한다. 제안하는 기법은 2차로 압축되는 허프만 인코딩에 맞춰 비트 패킹을 수행하는 방법으로 압축율을 높이고 있다. 제안하는 HAB 기법은 시계열 압축을 위해 적은 파라메타를 사용하기 때문에 가용 리소스에 제한이 있는 단말 노드 또는 엣지 노드에도 적용할 수 있다. 제안하는 기법은 무손실 기법으로 인공지능 모델 생성 및 정확한 데이터 분석이 필요한 분야에 적용할 수 있다. 시험평가에 기존기법과 비교하여 평균적으로 14.7% 압축율이 개선됨을 보였다.

추후 시계열 데이터의 압축율을 높이기 위해 Forecasting 인코딩에 대한 연구를 진행할 예정이다.

## References


[1] M. R. Chowdhury, S. Tripathi, S. De, "Adaptive Multivariate Data Compression in Smart Metering Internet of Things," IEEE Transactions on Industrial Informatics, Vol. 17, No. 2, pp. 1287-1297, 2020.

[2] 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, 6076, 2020.

[3] G. Campobello, A. Segreto, S. Zanafi, S. Serrano, "RAKE: A Simple and Efficient Lossless Compression Algorithm for the Internet of Things," 2017 25th European Signal Processing Conference (EUSIPCO), pp. 2581-2585, 2017.

[4] S. K. Jensen, T. B. Pedersen, C. Thomsen, "Time Series Management Systems: a 2022 Survey," Data Series Management and Analytics. Association for Computing Machinery, 2022.

[5] K. Chirikhin, B. Ryabko, "Compression-based Methods of Time Series Forecasting," Mathematics, Vol. 9, No. 3, pp. 384, 2021.

[6] S. H. Hwang, K. M. Kim, S. Kim, J. W. Kwak, "Lossless Data Compression for Time-Series Sensor Data Based on Dynamic Bit Packing," IEMEK Journal of Embedded Systems and Applications, Vol. 18, No. 1, pp. 1-7, 2023.

[7] G. Chiarot, C. Silvestri, "Time Series Compression Survey," ACM Computing Surveys (CSUR), Vol. 55, No. 10, pp. 1-32, 2022.

[8] M. A. de Oliveira, A. M. da Rocha, F. E. Puntel, G. G. H. Cavalheiro, "Time Series Compression for IoT: A Systematic Literature Review," Wireless Communications and Mobile Computing 2023, 5025255, 2023.

[9] 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.

[10] C. J. Deepu, C. H. Heng, Y. Lian, "A Hybrid Data Compression Scheme for Power Reduction in Wireless Sensors for IoT," IEEE Transactions on Biomedical Circuits and Systems, Vol. 1, No. 2, pp. 245-254, 2016.

[11] L. Yan, J. Han, R. Xu, Z. Li, "Model-free Lossless Data Compression for Real-time Low-latency Transmission in Smart Grids," IEEE Transactions on Smart Grid, Vol. 12, No. 3, pp. 2601-2610, 2020.

[12] A. K. Idrees, S. K. Idrees, R. Couturier, T. Ali-Yahiya, "An Edge-fog Computing-enabled Lossless EEG Data Compression with Epileptic Seizure Detection in IoMT Networks," IEEE Internet of Things Journal, Vol. 9, No. 15, pp. 13327-13337, 2022.

[13] B. Barbarioli, G. Mersy, S. Sintos, S. Krishnan, "Hierarchical Residual Encoding for Multiresolution Time Series Compression," Proceedings of the ACM on Management of Data Vol. 1, No. 1, pp. 1-26, 2023.

[14] A. Bruno, F. M. Nardini, G. E. Pibiri, R. Trani, R. Venturini, "Tsxor: A Simple Time Series Compression Algorithm," SPIRE 2021: String Processing and Information Retrieval, pp 217-223, 2021.

[15] 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.

[16] D. Blalock, S. Madden, J. Guttag, "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.

[17] J. Burgués, J. M. Jiménez-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.

[18] L. M. Candanedo, V. Feldheim, D. Deramaix, "Data Driven Prediction Models of Energy Use of Appliances in a Low-energy House," Energy and Buildings, Vol. 140, pp. 81-97, 2017.

**Soosung Lee (이 수 성)**

1999 Control and Instrumentation Engineering from Daegu University (B.S.)
2024 Information and Communication Engineering from Yeungnam University (M.S.)
2018~Gyeongbuk Institute of IT Convergence Industry Technology (Principal Researcher)
Field of Interests: Embedded Systems, Hardware Design, Embedded AI, Automotive Electronics Components
Email: sslee@gitc.or.kr

**Sang-Ho Hwang (황 상 호)**

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 (김 성 호)**

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:
2021~2023 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

**Jang-Kyu Yun (윤 장 규)**

2007 Computer Engineering from Kyungpook National University (M.S.)
2012 Computer Engineering from Kyungpook National University (Ph. D.)
2012~2013 Post-Doc, POSTECH
2013~2016 Senior Researcher, Dentis Corp.
2016~Leader of Electrical Parts Reserch Center, Gyeongbuk Institute of IT Convergence Industry Technology
Field of Interests: Future-mobility, Automotive Cybersecurity, OTA Software Update
Email: jkyun@gitc.or.kr

**YongWan Park (박 용 완)**

1982~1984 Electrical Engineering from Kyungpook University (B.E./M.E.)
1989~1992 Electrical Engineering from State Univ. of New York(Buffalo). USA (M.S./Ph.D.)
1992~1993 California Institute of Technology as a research fellow.
1994~1996 Chief researcher for developing IMT-2000 system at SKT
1996~ Professor of Information and Communication Engineering at Yeungnam Univ.
Career:
2017~2021 Director of the Industry-Academia Cooperation Foundation at Yeungnam University
2021~2024 Dean of the College of Mechanical and IT Engineering at Yeungnam University
2023~2024 Dean of the Graduate School at Yeungnam University
2008~ Director of the Institute of Information and Communication at Yeungnam University
Field of Interests: Mobile Communications, Mobility, Vehicle Reliability, Sensor Systems, Vehicle Software
Email: ywpark@yu.ac.kr