

IoT 장비를 위한 시계열 데이터 압축 기법 구현

(Implementation of Time Series Data Compression Scheme for IoT Devices)

황상호[†], 석수영^{*}
(재)경북IT융합산업기술원

(Sang-Ho Hwang, Soo-Young Suk)

([†]Gyeongbuk Institute of IT Convergence Industry Technology (GITC).)

Abstract : Internet of Things(IoT) devices that collect sensor data typically have limited storage sizes and require battery life management. Large amounts of data not only require a lot of storage space, but also consume a lot of energy for transmission. In order to alleviate the constraints of these IoT devices, data volume management is essential. In this paper, we implement a data compression scheme to reduce the energy consumption of IoT devices such as sink detectors.

Keywords : IoT, Data Compression, Time Series Data, Lossless

I. 서론

센서는 현재 스마트팩토리, 스마트팜, 스마트시티, 웨어러블 디바이스 등의 ICT(Information & Communication Technology) 융합 서비스의 발전으로 그 종류와 수가 기하급수적으로 늘어가고 있다. 센서들은 실시간으로 많은 양의 데이터를 생성하고 있으며, 그 수도 점차 늘어날 것으로 예상된다 [1-3]. 센서 데이터를 수집하는 IoT(Internet of Things)장비들은 대체적으로 저장장치의 크기가 한정적이며 배터리 수명관리가 필요하다 [4]. 대용량의 데이터는 저장 공간을 많이 필요할 뿐만 아니라 전송에 많은 에너지를 소비하게 된다. 이러한 IoT 장비들이 가진 제약사항들은 완화하기 위해서 수집되는 데이터의 볼륨관리는 필수적이다.

IoT장비에서 센서로부터 생성되는 데이터의 볼륨을 줄이는 방법에는 압축이 널리 활용되고 있으며, 압축 기법은 압축된 데이터를 복원했을 때 일부

데이터가 손실되는 손실 압축 기법과 데이터 복원시 원본 그대로 값을 얻을 수 있는 무손실 압축 기법이 있다 [5]. 지금까지 많은 연구들은 비손실 기법들과 비교하여 압축 비율이 높은 손실 압축 방식을 선택하였다 [6-7]. 하지만 손실 압축 기법을 통해 수집된 센싱 데이터들은 품질이 저하되어 머신러닝 등의 알고리즘에서 부정적인 영향을 끼칠 수 있다.

본 논문에서는 침하감지기와 같은 IoT 장비의 에너지 소비량을 줄이기 위한 데이터 압축 기법을 구현한다. 구현한 데이터 압축 기법은 시계열 데이터의 시간흐름에 따른 센서 값의 변화량에 기반을 둔 비손실기법이다.

II. 시계열 데이터 압축 기법

센서로부터 수집된 각각의 시계열 데이터는

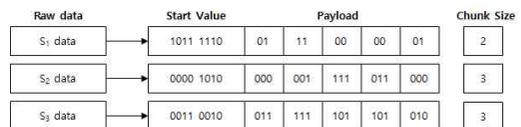


그림 1. 델타 값 기반 압축 수행의 예

Fig. 1. Example of performing delta-based compression

* 교신저자(Corresponding Author)

황상호, 석수영 : (재)경북IT융합산업기술원

※ 본 논문(또는 저서)은 과학기술정보통신부의 재원으로 2019년 과학기술기반 지역수요맞춤형 R&D지원사업의 지원을 받아 수행된 연구임 (CN19100GB001).

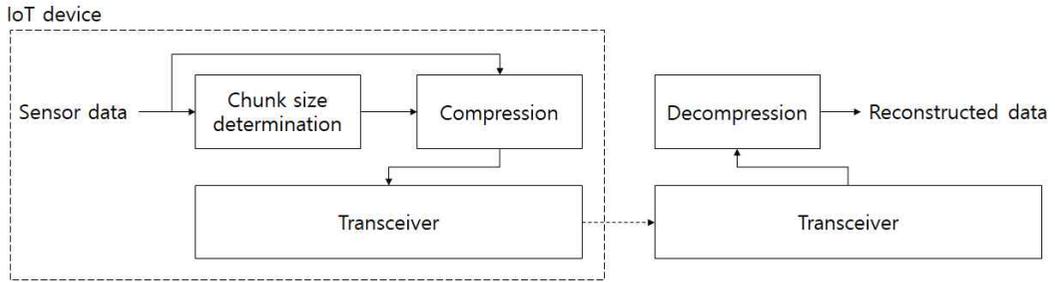


그림 2. 데이터 압축 기법을 포함하는 IoT 장비의 예
Fig. 2. Example of Iot device including data compression scheme

CSD(Chunk Size Determination)을 통해 결정된 압축 청크(chunk)의 크기에 기반하여 압축된다. 그림 1은 압축 수행의 예를 보여주고 있다. 수집된 센서 데이터는 센서 별로 압축되며, 압축은 이전 값과의 변화량(delta, Δ)에 기반한다. 예를 들어, 그림 1에서 센서 데이터들은 8bit단위로 수집이 되며, 각 센서 별로 6개의 데이터가 수집되었다. 첫 번째 데이터는 Start Value에 저장되고 이후 5개의 데이터는 변화량만 각 청크에 저장된다. 청크내 첫 번째 비트는 양과 음(0,1)의 부호를 나타내며 나머지 비트는 변화량을 의미한다. 그림 1의 예제에서 각 센서들의 청크 크기는 CSD에 의해 2, 3, 3으로 결정되었다.

그림 2는 본 논문에서 구현한 압축 기법을 포함하는 IoT 장비의 예를 보여주고 있다. 데이터는 CSD에 의해 결정된 청크 크기에 기반하여 압축 모듈을 통해 압축된다. 압축된 데이터는 송수신장치로 전송된 후 압축 해제된다.

그림 3은 실제 센서 데이터 및 그 변화량을 보여주고 있다. 그림 3에서 보는 것과 같이 변화량은 실제 데이터에 비해 적은 비트만으로 표현이 가능

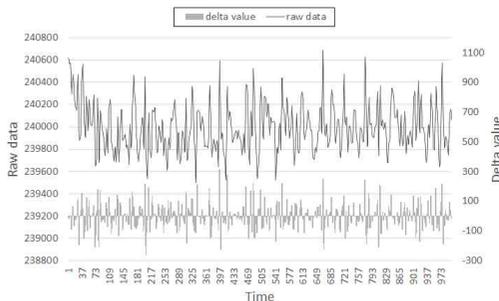


그림 3. 센서 데이터 및 변화량의 예
Fig. 3. Example of sensor data and delta value

하며, 이 변화량을 활용하면 실제 센서 값보다 적은 비트만 활용하여 데이터를 표현할 수 있다. 결국 변화량을 저장하는 청크의 크기가 압축률을 결정하며, 본 논문에서는 CSD에 의해 결정된다. CSD는 수식 1에 기반하여 고정된 길이의 시계열 데이터를 모니터링 하여 청크의 크기를 최소화할 수 있는 길이를 결정한다.

$$CS_i = \max(\log_2 \Delta_{ij}) \quad (1)$$

수식 1에서 i 는 센서 번호를 의미하며 Δ 는 변화량, j 는 데이터 순차번호를 의미한다. j 의 최대키는 사용자에게 의해 결정된다.

각 센서 데이터들은 CSD 및 Compression 모듈을 거치면 그림 4와 같이 헤더(Header)와 페이로드(Payload)로 구성된 데이터로 변환된다.

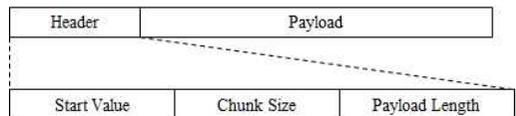


그림 4. 압축 데이터 포맷
Fig. 4. Compression data format

그림 4에서 헤더는 첫 번째 시작 값과 청크의 크기 그리고 페이로드의 전체 길이 정보를 가지고 있으며 페이로드에는 압축된 값이 저장된 청크들을 가지고 있다. 압축 데이터 검증을 위해 CRC(Cyclic Redundancy Check)가 포함될 수 있지만, 본 논문에서는 고려하지 않는다.

III. 실험 및 분석

1. 데이터 및 실험환경

본 논문에서는 구현한 압축 기법의 성능을 비교하기 위해서 UCI에 공개된 Gas sensor array temperature modulation Data Set[8]을 활용하였다. 이 데이터 셋은 각종 센서로부터 수집된 시계열 데이터로 가스 챔버 내부의 일산화탄소(CO), 습도 및 온도와 같은 측정값을 포함하고 있다. 실험에서는 데이터 셋 중에서 2016년 9월 30일에 수집된 데이터를 활용하였다.

실험에서 헤더의 Start Value의 크기는 32bit, Block Size의 크기는 5bit, Payload Length의 크기는 32bit로 설정하였다.

2. 실험결과

그림 5는 각 센서 데이터에 대한 압축률을 보여주고 있다. 실험에서 사용한 압축률은 아래 수식 2에 의해 계산되었다.

$$\text{Compression ratio} = \frac{\text{Size of original data}}{\text{Size of compressed data}} \quad (2)$$

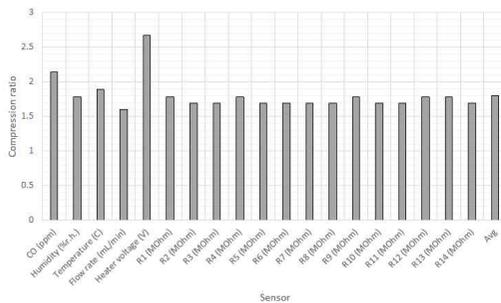


그림 5. Sensor data별 압축률

Fig. 5. Compression Ratio by sensor data

그림 5에서보는 것과 같이, 센서 데이터의 평균 압축률은 1.79를 보이고 있다. 데이터의 변화량이 낮은 Heater voltage 데이터 셋의 압축률이 가장 높은 2.66을 보이고 있으며, 모든 센서 데이터의 압축률은 1.59에서 2.66의 범위 내에 있다.

IV. 결론

본 논문에서는 IoT 장비들의 에너지 소비량을

줄이기 위한 데이터 압축 기법을 구현하였다. 구현한 데이터 압축 기법은 센서 값의 변화량에 기반을 둔 비손실기법이다. 실험에서 구현한 기법의 압축률은 평균 1.79이였으며, 센서 데이터의 볼륨을 줄이는데 도움이 된다. 추후 데이터의 압축률을 높일 수 있는 비손실 압축 기법에 대한 연구를 추가적으로 진행할 예정이다.

참고 문헌

- [1] R.A.A. Habeeb, F. Nasaruddin, A. Gani, I.A.T, Hashem, E. Ahmed, and M. Imran, "Real-time big data processing for anomaly detection: A Survey", International Journal of Information Management, Vol. 45, pp. 289-307, 2019.
- [2] F. Shi, Q. Li, T. Zhu, "A survey of data semantization in internet of things", Sensors, Vol. 18, No. 1, pp. 313, 2018.
- [3] 권지수, 조정훈, 박대진, "IoT 기반 간헐적 이벤트 로깅 응용에 최적화된 효율적 플래시 메모리 전력 소모 감소기법", 대한임베디드공학회 논문지, 제 14권, 제 2호, 87-96쪽, 2019.
- [4] C.J. Deepu, C.H. Heng, and Y. Lian, "A hybrid data compression scheme for power reduction in wireless sensors for IoT", IEEE transactions on biomedical circuits and systems, Vol. 11, No. 2, pp. 245-254, 2016.
- [5] K. Sayood, "Introduction to data compression", Morgan Kaufmann, 2017.
- [6] A. Ukil, S. Bandyopadhyay, A. Sinha, and A. Pal, "Adaptive Sensor Data Compression in IoT systems: Sensor data analytics based approach", In 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 5515-5519, 2015.
- [7] A. Moon, J. Kim, J. Zhang and S.W. Son, "Lossy compression on IoT big data by exploiting spatiotemporal correlation", In 2017 IEEE High Performance Extreme Computing Conference (HPEC), pp. 1-7, 2017.
- [8] J. Burgués, J.M. Jiménez-Soto, and 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.