

IoT 기상 데이터의 손실압축 기술 연구

문애경, 김기덕, 송윤정

한국전자통신연구원 지역산업IT융합연구실

akmoon@etri.re.kr, paul.kdkim@etri.re.kr, yjsong@etri.re.kr

Study of lossy compression technology for IoT weather data

Aekyung Moon, Kiduck Kim, Yunjeong Song

ETRI, Regional Industry IT Convergence Research Section

요약

본 논문은 다양한 신호 처리 알고리즘을 이용하여 데이터를 압축하고 복원하여 재구성된 데이터의 오류와 압축률을 비교함으로써 기상 데이터에 대한 손실압축 분석결과를 제시한다. 특히 IoT 기상 관측소에서 수집한 5종의 실제 센서 데이터에 대하여 평가한 실험 결과, 압축 계수가증가함에 따라 모든 압축 알고리즘의 데이터 복원오류도 증가하였고 DCT와 FWHT가 높은 압축결과를 보였다.

I. 서론

IoT 기상 스테이션에서 수집 한 데이터의 양이 증가함에 따라 애플리케이션 개발을 위한 효율적 데이터 관리 및 분석에 대한 요구가 증가하고 있다[1]. 수집된 IoT 데이터는 기하 급수적으로 증가하고 있고 사후 처리를 위해 측정된 데이터를 모두 클라우드로 전송하는 것은 많은 비용이 발생한다. 데이터를 원격 클라우드 또는 데이터 센터로 전송하는 대신, 에지(센서) 노드에서 필터링하고 장기 저장 및 데이터 분석을 위해 관련성 높은 데이터를 원격 노드로 주기적으로 전송하기 위한 연구가 진행 중이다. 이러한 관점에서 엣지 노드에서 IoT 데이터 세트를 효과적이고 안정적으로 활용하기 위한 빅데이터 압축 기술은 데이터 저장 및 통신오버헤드 등을 감소시키는 이점을 얻을 수 있다. 압축률을 높이기 위해서 손실압축(lossy compression)[2]이 고려되고 있고, 손실압축은 데이터 크기를 줄이는 데 도움이 되지만 오류가 발생할 수 있다.

최근 손실 압축 방법을 적용하여 발생하는 오류에 관한 연구들이 제안되었다[3]. 본 논문은 손실압축이 IoT 데이터 관리 및 분석에 미치는 영향을 이해하기 위해 데이터 변환 후 재구성 된 기상 센서 데이터에 대한 여러 압축 알고리즘을 평가한다. 무손실압축(lossless compression)에 비해 손실압축은 시공간 패턴의 이점을 활용하여 데이터 볼륨을 크게 줄일 수 있지만 압축 후 복원된 데이터가 원본데이터와 다르기 때문에 데이터 분석 결과, 품질에 영향을 줄 수 있다. 본 논문에서는 3개의 신호 변환 알고리즘(DCT (Discrete Cosine Transform)[4], FWHT (Fast Walsh-Hadamard Transform)[5], DWT (Discrete Wavelet Transform)[3])을 기반으로 기상 센서 데이터를 효율적으로 저장하기 위한 손실압축 알고리즘을 평가한다. IoT 기상 데이터의 데이터 복원시 오류와 압축률간의 관계 및 데이터 특성 정보를 제공하기 위하여 5개 범주의 기상 센서 데이터를 3가지 변환(DCT, FWHT, DWT) 기반 손실압축 알고리즘으로 실험한다. 실험 결과에 따르면 변환 계수, 압축 비율과 복원 오류 사이에 차이를 나타내고 변환계수가 높일수록 낮은 압축률과 오류가 발생한다.

II. 손실압축알고리즘

기상 데이터의 경우 개별 데이터 값은 임의성을 보여 주지만 전체 패턴은 시간적으로나 공간적으로 일관성을 제공한다. 이로 인해 변환 된 데이터가 명시적으로 데이터의 상관 관계를 나타내므로 무작위 값을 가지는 일반 센서 데이터에 비해 높은 압축률을 갖는다. 그림 1은 본 논문에서 평가 한 온도 데이터에 DCT를 적용한 결과를 보여준다.

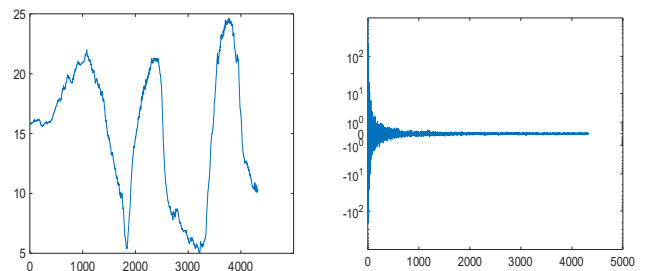


그림 1 온도데이터 DCT 변환

[6]에서 DCT, DWT 및 WHT의 손실압축알고리즘을 사용하여 기상 데이터를 실험하였다. 시뮬레이션 결과는 DCT로 변환 된 데이터를 사용하면 WHT(Walsh-Hadamard Transform) 및 DWT와 같은 다른 변환 방법에 비해 기상 데이터 복원 성능이 더 우수하였다. 그러나 압축 된 신호가 약간의 오류와 함께 재구성됨을 보여준다[7]. 본 논문에서는 미기상 데이터를 수집하고 3 가지 변환(DCT, FWHT, DWT)을 기반 손실 압축 알고리즘을 사용하여 압축한다. 그리고 서버에서 재구성 된 기상 센서 데이터의 오류를 평가한다. 본 논문의 목표는 손실압축 및 복원이 데이터 안정성에 미치는 영향을 평가하는 것이고 5개 타입의 기상 센서 데이터 세트를 사용하여 실험을 진행한다.

III. IoT 기상데이터 특성

평가를 위해 영천지역 과수원에 위치한 무선 기후 관측소에서 수집한

실제 데이터 세트를 사용한다. 설치된 기상 관측소에서 54일 동안 수집된 기상 데이터에서 온도, 습도, 광량, 풍향 및 풍속 등 중요한 5 가지 변수를 선택한다. 데이터는 1 분마다 지속적으로 모니터링 및 저장된다. 이 평가에서는 약 54 일의 샘플링에 해당하는 77,590 개의 데이터셋을 사용한다. 표 1은 표준 편차 (STD), 정규화된 표준 편차(NSTD), 왜도(Skewness) 및 첨도(Kurtosis)의 지표를 활용하여 원본 데이터의 통계적 속성을 보여준다. 정규화 표준 편차는 $STD(x)/Mean(x)$ 로 계산한다. 왜도는 평균값 주변의 데이터 비대칭 측정결과를 나타낸다. 음의 왜도값은 더 많은 데이터가 평균의 왼쪽에 흩어져 있음을 의미하고 양의 왜도값은 더 많은 데이터가 오른쪽에 흩어져 있음을 의미한다. 평균에 대해 대칭인 정규 분포는 왜도값이 0이다. 표 1의 첨도 측정은 분포가 얼마나 특이한 지를 나타낸다. 정규 분포의 첨도가 3이므로 첨도가 3보다 높은 분포는 일반적이지 않다. 광량 및 풍속은 다른 데이터 세트보다 STD가 더 높다. 왜도의 경우 태양 복사 및 풍속은 양수 값만 가진다. 표 1에 나타난 바와 같이, 실험 데이터에 대하여 패턴을 갖지 않는다.

표 2 테스트 IoT 기상데이터 특성

	표준편차	정규화된표준편차	왜곡	첨도
온도	5.6807	0.4659	-0.0907	3.0002
습도	20.3439	0.2548	-0.9132	2.5916
광량	5.9199	1.1748	0.8546	2.2679
풍향	112.6286	1.0666	0.5782	1.8434
풍속	0.7574	1.2284	1.3920	4.5762

IV 평가

각 압축 알고리즘의 성능을 평가하기 위해 다음 두 가지 성능평가지수를 사용한다.

- 압축 비율 : 달성 가능한 압축 비율. M : 압축 방법

$$R_M = \frac{|D| - |D'|}{|D|} \times 100\%$$

- NRMSE : 복원 오류, x , 원본 데이터, \hat{x} , 재구성된 복원데이터

$$NRMSE_M = \frac{SE_M}{Mean(x)} = \frac{1}{x} \sqrt{\frac{\sum_{n=1}^N (x(n) - \hat{x}(n))^2}{N}}$$

DCT, DWT 및 FWHT 변환을 기반으로 평가된 데이터 세트를 압축하는 방법은 다음과 같다[8]. DWT 웨이블릿 변환에는 Daubechies d4 웨이블릿을 사용한다. 1) 원본 데이터를 DCT, DWT, FTWH 기반 벡터로 분해한다. 2) 계수 벡터 S를 계수 값의 내림차순으로 정렬한다. 분류된 계수 벡터는 $SS = SS_1, SS_2, \dots, SS_n$ 으로 표시됩니다. 3) 신호에서 에너지의 δ 양을 나타내는 계수가 $0.0 \leq \delta \leq 1.0$ 또는 $0\% \leq \delta \leq 100\%$ 인 k 를 계산한다. 전체 DCT 계수에 저장된 에너지의 합은 1.0 (또는 100%)입니다. 4) 임계 값 δ 보다 작은 계수는 0으로 설정한다.

V. 결론

본 논문에서는 실제 기상 데이터의 5 가지 (온도, 습도, 광량, 풍향, 풍속)를 DCT, FWHT, 그리고 DWT를 사용하여 압축하고 복원된 데이터의 오류와 압축률을 비교하여 IoT 기상데이터에 손실 압축의 성능을 평가했다. 실험 결과는 DCT와 FWHT의 압축 비율이 높다는 것을 보여주고 DCT 및 FWHT의 경우 개별 데이터 집합에 대한 오류의 영향이 큰 차이

를 나타낸다. 특히 DCT는 온도 데이터에 비해 풍속 데이터의 복원 오류율이 67 % 정도 높다. 따라서 어플리케이션에서 허용 가능한 범위 내에서 압축 계수를 선택하는 것이 중요하다.

표 3 압축률 비교

	δ	온도	습도	광량	풍향	풍속
DCT	0.9	99.99	99.99	99.99	99.35	99.82
	0.999	99.68	99.73	97.81	30.25	35.17
DWT	0.9	74.21	68.8	88.57	84.88	89.02
	0.999	53.47	50.81	73.84	56.78	55.04
FWHT	0.9	99.99	99.99	99.93	99.09	99.66
	0.999	99.05	90.28	95.20	29.54	46.83

표 4 원본데이터와 복원시 오류 비교

	δ	온도	습도	광량	풍향	풍속
DCT	0.9	0.466	0.255	0.667	0.64	0.69
	0.999	0.049	0.046	0.07	0.065	0.07
DWT	0.9	0.48	0.45	0.67	0.63	0.69
	0.999	0.049	0.046	0.07	0.065	0.07
FWHT	0.9	0.45	0.395	0.65	0.63	0.68
	0.999	0.046	0.043	0.07	0.05	0.056

ACKNOWLEDGMENT

본 연구는 한국전자통신연구원 연구운영지원사업의 일환으로 수행되었음.[20ZD1110, 대경권 지역산업 기반 ICT융합기술 고도화 지원사업

참 고 문 헌

- [1] A. Ukil, S. Bandyopadhyay, and A. Pal, "IoT Data Compression: Sensor-Agnostic Approach," in 2015 Data Compression Conference, April 2015, pp. 303 - 312
- [2] J. J. Chou and L. A. Piegl, "Data reduction using cubic rational B-splines," IEEE Computer Graphics and Applications, vol. 12, no. 3, pp. 60 - 68, May 1992
- [3] M. M. Abo-Zahhad, A. I. Hussein, and A. M. Mohamed, "Compressive Sensing Algorithms for Signal Processing Applications: A Survey," International Journal of Communications, Network and System Sciences, vol. 8, no. 6, pp. 197 - 216, 2015
- [4] M. A. Razzaque, C. J. Bleakley, and S. Dobson, "Compression in wireless sensor networks: A survey and comparative evaluation," ACM Transactions on Sensor Networks, vol. 10, no. 1, p. 5, 2013.
- [5] B. J. Fino and V. R. Algazi, "Unified Matrix Treatment of the Fast Walsh-Hadamard Transform," IEEE Trans. Comput., vol. 25, no. 11, pp. 1142 - 1146, Nov. 1976
- [6] R. Sustika and B. Sugiarto, "Compressive Sensing Algorithm for Data Compression on Weather Monitoring System," TELKOMNIKA Telecommunication, Computing, Electronics and Control, pp. 974 - 980, 2016
- [7] R. Chaturvedi and Y. Yadav, "A Survey on Compression Techniques for ECG Signals," International Journal of Advanced Research in Computer and Communication Engineering, vol. 2, no. 9, 2013
- [8] A. Moon, X. Zhuo, J. Zhang, and S. W. Son, "AD2: Improving Quality of IoT Data through Compressive Anomaly Detection," IEEE Big Data, 2019