

균일 양자화를 이용한 센서 데이터 압축 기법

이석재, 여명호, 조용준, 박현호, 유재수

충북대학교 정보통신공학과

{sjlee, mhyeo, yjcho, phh}@netdb.cbnu.ac.kr, yjs@cbnu.ac.kr

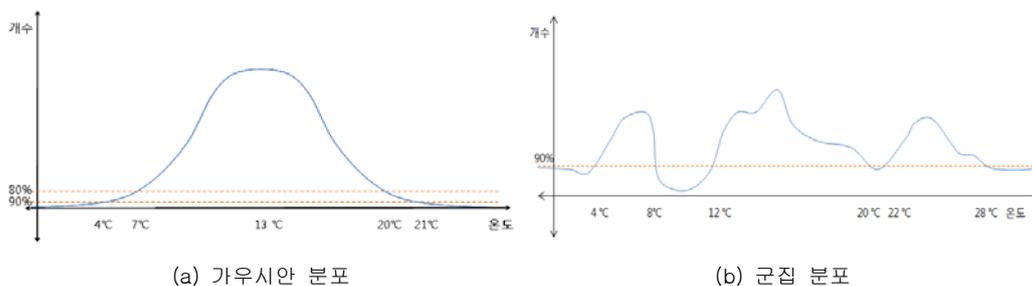
Efficient Sensor Data Compression Algorithms using Uniform Quantization

Seok-Jae Lee, Myoung-Ho Yeo, Yong-Jun Cho, Hyun-Ho Park, Jae-Soo Yoo
Dept. of Computer and Communication, Chungbuk National University

센서 데이터를 수집하는 과정에서 통신에 사용되는 전력소모를 줄이기 위해 데이터 병합, 압축 등의 기법들이 연구되고 있다. 데이터 병합 기법은 센서 네트워크 내에서 중복되는 데이터의 전송 횟수를 줄임으로써 전력소모를 줄인다. 데이터 압축 기법은 큰 데이터 전송에 사용되는 전력소모가 압축을 위한 계산에 사용되는 전력량이 더 적다는 점에 초점을 맞추어 연구되고 있다[1].

지금까지 연구된 대부분의 센서 데이터 압축 기법들은 센싱된 데이터에 웨이블릿 변환, 양자화, 가변길이 부호화 등과 같은 신호 압축 기법 또는 BWT, LZE, LZ77과 같은 스트림 압축 기법들을 적용하고 있다 [2]. 이와 같은 연구들은 각 센서별로 측정된 데이터만으로 압축을 처리하기 때문에 전체 센서 네트워크에서의 데이터 발생 분포, 빈도 등은 고려하지 않는다. 그러나 기상측정과 같은 응용에서는 발생하는 센서 데이터가 특정 값 또는 좁은 범위의 값으로 한정되는 경우가 대부분이다. 이 경우 센서 데이터의 특성을 각 센서가 알 수 있다면 보다 효과적인 데이터 압축 기법을 만들어 적용할 수 있다.

본 논문에서는 센서 데이터에 균일양자화 기법을 이용한 압축 기법을 적용해 센서의 전력소모를 감소시키는 데이터 압축 기법을 제안한다. 일반적으로 센서네트워크에서 센싱되는 값들은 가우시안 분포 또는 몇몇 값을 중심으로 유사한 값들이 군집을 형성하는 형태의 분포를 띄는 경우가 대부분이다. 제안하는 데이터 압축 기법은 센서 데이터의 이러한 특징을 고려해, 일정 주기마다 싱크 노드가 센서네트워크의 전체 센서 데이터를 수집해 값의 분포를 분석한다. 싱크 노드는 분석 결과를 기반으로 균일양자화를 위한 요약 정보를 생성해 전체 센서들에게 배포한다. 각 센서들은 이 요약정보를 이용해 센싱한 값을 압축하여 싱크 노드로 전송함으로써 통신비용을 효과적으로 줄일 수 있다. 그림 1은 센서네트워크에서 발생할 수 있는 데이터의 분포 형태 예를 나타낸다.



(a) 가우시안 분포

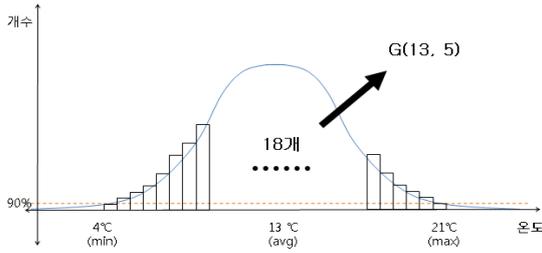
(b) 군집 분포

그림 1. 센서네트워크에서 발생하는 데이터 분포 예

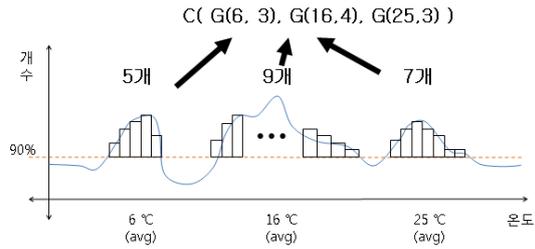
그림 1(a)에서 양자화 대상 센서 데이터 범위를 전체 발생 데이터의 90% 이상을 포함하는 범위로 지정하면, 최대값 21°C, 최소값 4°C, 중앙값은 13°C가 된다. 최소 센싱 단위가 1°C라면 양자화 비트로 표현할 데이터의 개수가 18개로 5비트로 모든 데이터를 표현할 수 있다. 이 경우 양자화 정보를 $G(13,5)$ 와 같이 중앙값과 데이터 표현 비트수를 이용해 생성한다. 그림 2(a)는 그림 1(a)의 분포에서 양자화 범위를 90%로 지정한 경우 양자화 결과를 나타낸다.

군집분포의 경우 비연속적인 각 군집영역을 나누어 가우시안 분포 형태로 모델링하고, 각 영역의 양자화 정보를 모아 배포한다. 군집분포에서 양자화 정보는 $C(G_1(\text{avg}, \text{bit}), G_2(\text{avg}, \text{bit}), \dots, G_n(\text{avg}, \text{bit}))$ 와 같이 여러 개의 양자화 정보가 연속적으로 포함된 형식으로 생성한다. 그림 2(b)는 그림 1(b)를 양자화 범위를 90%로 지정해 양자화 정보를 생성한 결과를 나타낸다.

* 이 논문은 2007년도 정보(과학기술부)의 재원으로 한국과학재단의 지원을 받아 수행된 연구임(No.R01-2006-000-10809-0)



(a) 그림 1(a)에 대한 양자화 정보



(b) 그림 1(b)에 대한 양자화 정보

그림 2. 데이터 분포에 따른 양자화 정보

또한 센서 데이터와 배포한 양자화 정보의 중심 값의 차가 클 경우, 양자화 범위를 벗어나는 값들이 발생할 수 있다. 제안하는 기법에서는 범위를 벗어나는 값들은 원본 데이터를 그대로 전송해 센서 데이터의 정확성이 떨어지는 것을 방지한다. 제안하는 기법의 우수성을 보이기 위해 센서 데이터를 제안한 압축 기법을 적용해 수집한 경우와 압축하지 않고 수집한 경우에 대해 시뮬레이션을 통해 성능평가를 실시하였다. 표 1은 성능평가에 사용된 요소들이다.

표 1. 성능평가 요소들

항 목	값	항 목	값
센서 노드 수	1,000개	양자화 데이터 양(범위)	80 ~ 100 (%)
센서 데이터 분포	가우시안, 균집, 균등 분포	센싱 주기	1 (sec)
센서 데이터 최소단위	1 (°C)	센싱 기간	24 (hours)
센서 데이터 발생범위	1 ~ 40 (°C)	양자화 정보 갱신주기	1 ~ 24 (hours)

표 2는 가우시안 분포 형태에서 양자화 범위를 전체 데이터의 80~100%를 포함하는 경우로 변경하면서 실험한 결과이다. 제안하는 압축 기법을 적용한 경우, 양자화 범위 95%에서 비압축 데이터에 비해 17.5%, 100%인 경우 15.6%로 전송 데이터 크기가 줄어들었다.

표 2. 가우시안 분포에서 양자화 적용 데이터 수에 따른 데이터전송량

양자화 대상 데이터 양(%)	양자화 범위(°C)	전송량(bit)	압축비(%)
비압축	-	2,764,800,000	100
80	△3	725,760,000	26.3
90	△6	535,680,000	19.4
95	△10	483,840,000	17.5
98	△17	487,296,000	17.6
100	△26	432,000,000	15.6

그림 3은 균집 분포에서 균집의 수에 따라 양자화 정보를 배포하는 비용의 차를 비교한 것이다. 실험은 3개의 균집이 형성된 경우(pattern 1)와 6개의 균집이 형성된 경우(pattern 2)를 갱신 주기에 따라 비교하였다. 갱신 주기가 짧고, 형성되는 균집 수가 늘어날수록 갱신에 필요한 전송 데이터의 크기가 비례해서 커지는 것을 알 수 있다. 표 3은 갱신주기에 따른 데이터 전송량이다.

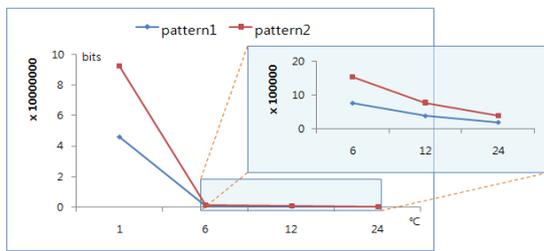


그림 3. 균집 수에 따른 양자화 정보 배포 비용

표 3. 갱신주기에 따른 데이터 전송량

주기	전송량(bit)
1h	15,360,000
6h	256,000
12h	128,000
24h	64,000

[참고문헌]

- [1] Naoto Kimura, Shahram Latifi, "A Survey on Data Compression in Wireless Sensor Networks," In Proceedings of the International Conference on Information Technology: Coding and Computing, April 2005.
- [2] 권영환, 김동국, 이좌형, 김윤, 정인범, "무선 센서 네트워크에서 압축을 이용한 네트워크 트래픽 감소 기법 구현," 한국정보과학회 가을 학술발표논문집, Vol.33, No.2(D), pp.699-704, 2006.