

센서 네트워크에서의 지연 갱신을 이용한 효율적인

예측 모델의 갱신

노현호, 김홍연, 민준기

한국기술교육대학교 컴퓨터공학부

e-mail : {oksknoh, zenweird, jkmin}@kut.ac.kr

An Efficient Update of Estimation Model Using Delayed Update in Sensor Networks

Hyun-ho Noh, Hongyeon Kim, Jun-Ki Min*

*School of CSE

Korea University of Technology and Education

요 약

센서 네트워크는 많은 수의 센서들로 구성되며 물질세계와 상호작용하여 실시간 감시를 할 수 있는 큰 규모의 네트워크이다. 현실 세계의 물리적 변화를 지속적으로 감시하기 위하여 센서 네트워크에 속해 있는 각 센서들은 주기적으로 측정값을 읽어서 기지국(base station)에 전달하는 기능을 수행한다. 이러한 경우, 각 센서들의 지속적인 통신에 따른 에너지 낭비가 발생하여 센서 네트워크 전체의 수명을 단축시키게 된다. 따라서 데이터 모니터링에 있어서 예측을 통한 근사 데이터 수집을 통해 센서 네트워크의 수명을 연장하고자 하는 연구들이 진행되고 있다. 그러나 현재의 예측을 통한 근사 데이터 수집은 측정값의 변화가 클수록 예측 모델의 갱신이 잦다. 따라서 본 논문에서는 지연 갱신을 이용한 효율적인 예측 모델 갱신 기법을 제안한다. 제안된 기법에서는 예측이 실패하였을 경우 즉시 예측 모델을 갱신하는 대신 문제가 되는 구간의 측정값들을 지연, 근사화하여 전송함으로써 예측 모델의 잦은 갱신을 방지할 수 있다.

1. 서론

최근 전자기술의 발전에 따라 센서 노드의 배터리 수명이 길어지고 전력 소모량이 줄어들면서 이전보다는 센서 노드의 수명이 길어지고 또한 센서 네트워크가 허용할 수 있는 대역폭이 늘어나고 있는 추세이다. 이와 같은 장점 때문에 센서 네트워크의 구성 범위가 광범위해지고 구성되는 노드의 수도 증가되면서 대규모 네트워크로 발전되어 센서 네트워크는 현실세계를 분석하기 위한 유용한 기술이 되고 있다.

센서 네트워크는 많은 수의 센서들로 구성되어 현실세계의 물리적 변화를 실시간 감시를 할 수 있는 큰 규모의 무선 네트워크로 분산 데이터베이스와 유사한 특징을 지닌다. 센서 노드는 배터리로 동작하는 저 전력, 저가의 장치로 배터리가 소모되어 센서 노

드의 동작이 멈추면 배터리를 다시 재충전하거나 교체하기가 용이하지 않다. 즉 무선 센서 네트워크의 노드들은 일회용 소모품의 성격을 지닌다. 따라서 무선 센서 네트워크의 전력을 효율적으로 사용하여 네트워크의 생존시간(lifetime)을 최대화하여야 한다.

센서의 중요 기능은 크게 관찰 대상의 정보를 수집하는 샘플링 기능, 데이터를 다른 노드에게 전달하는 전송 기능, 그리고 수집된 정보를 가공 및 저장하는 처리 기능으로 나뉘는데 일반적으로 전송 시 사용되는 센서의 전력 소모는 샘플링이나 데이터 처리에 사용되는 전력 소모보다 크기 때문에 데이터 전송 횟수 및 전송량을 최소화 하는 것은 노드의 수명에 상당 부분 영향을 끼친다 [2,3,4].

센서 네트워크의 집단함수 값을 효과적으로 추출하

기 위하여 in-network aggregation 기법[7]이 제안되었다. 그러나 MIN/MAX, SUM, AGG 와 같은 집단함수 값은 특정 응용 분야에서 전체 센서 네트워크의 변화를 분석하는데 적합하지 못하다.

현실 세계의 물리적 변화를 지속적으로 감시하기 위하여 센서 네트워크에 속해 있는 각 센서들은 주기적으로 측정값을 읽어서 기지국(base station)에 전달하는 기능을 수행한다. 이러한 경우, 각 센서들의 지속적인 통신에 따른 에너지 낭비가 발생하여 센서 네트워크 전체의 수명을 단축시키게 된다. 따라서 데이터 모니터링에 있어서 근사 데이터 수집(approximate data gathering) [1,2,3]을 통해 센서 네트워크의 수명을 연장하고자 하는 연구들이 진행되었다.

기존의 근사적 데이터 수집 기법에서는 센서의 예측을 통해 얻은 값과 실제로 측정된 값의 차이가 사용자가 정의한 오차 범위 이내일 경우 해당 센서가 측정값을 전송하지 않도록 하여 센서의 에너지 소비를 감소시킨다. 그러나, 측정값의 변화가 심할 경우 예측 모델이 부정확해질 확률이 크며, 이에 따라서 각 센서 노드들을 많은 에너지를 소비하게 된다. 따라서 본 연구에서는 예측 모델의 잦은 갱신을 방지하기 위해 선형회귀분석 기법을 활용한 지연 갱신을 이용함으로써 센서 노드의 에너지 소비를 줄이고자 한다.

2. 관련 연구

일반적으로 센서의 측정값 전송 주기는 샘플링 주기와 일치한다. 이때, 전송 횟수를 줄이기 위하여 축약(suppression) 기법을 사용할 수 있다. 축약 기법 중 하나인 시간적 축약 (temporal suppression)에서는 각 센서가 가장 최근에 보고한 측정값이 현재의 측정값과 동일하면 기지국(base station)에 현재 측정값을 전송하지 않도록 한다. 따라서 기지국에서는 특정 센서로부터 측정값이 전송되지 않으면 측정값에 변화가 없음을 알 수 있다.

Silberstein [3] 등은 시간적 축약과 지역적 축약(temporal suppression)을 결합한 Conch 기법을 제안하였다. Conch 기법은 이웃한 노드들 끼리는 유사한 측정값을 지닌다는 특징을 이용하여 센서노드들 중 보고자 노드들을 선정하고 이 보고자 노드들이 자신의 측정값과 이웃노드들의 측정값을 차이를 보내도록 하였

다.

PAQ [2]는 통계에서 많이 쓰이는 AR(Autoregression) 기법에 기반한 예측 모델을 사용하여 센서의 값을 예측한다. 특별히, [2]에서는 동적 AR(3) 모델을 사용하였는데, 최근의 세계의 측정값들을 이용하여 미래의 측정값을 예측하는 모델을 생성한다. 이 모델은 다음과 같은 수식으로 표현된다. $X(t) = aX(t-1)+b(t-2)+c(t-2)+b(w)N(0,1)$. 여기서 $b(w)N(0,1)$ 은 평균 0, 표준편차 $b(w)$ 를 가지는 가우시안 백색 잡음 (Gaussian white noise)를 나타낸다. 미래를 정확히 예측하기 위하여 계수 a, b, c 를 구하는 것이 PAQ 에서는 필요하다. 센서에 예측 범위를 설정한 후 범위를 만족하지 못할 경우에, 기지국으로 데이터를 전송하는 방식이다. 예측이 잘 이루어지고 있는 상황에서는 데이터 전송비용을 크게 감소시킬 수 있다.

Snapshot [1]은 노드가 자신과 이웃을 이용하여 그들을 대표할 수 있는 대표 노드들을 선정할 수 있다. 대표 노드는 선형회귀함수를 이용하여 이웃들의 측정값을 예측할 수 있다. 이 때, 각 노드들은 주변 노드들의 값을 알아야 하기 때문에 이웃 노드들은 주기적으로 자신의 측정값을 방송(broadcast)해야 한다. 이 때문에 각 노드들은 많은 에너지를 소비한다.

Deligiannakis 등 [8] 은 데이터 압축 기법을 활용하여 센서 노드들의 데이터 전송량을 감소시키는 SBR 기법을 제안하였다. 그러나 이 기법에서는 선형회귀분석기법을 확장하여 자신기반회귀분석(Self-Based Regression)을 제안하였다.

Jain 등은 센서 네트워크에서 예측모델로 칼만필터(Kalman Filter)를 활용하였다. 칼만 필터는 재귀적 데이터 처리 알고리즘으로 현재의 측정값과 이전의 예측값 만으로 미래의 측정값을 예측할 수 있도록 한다.

3. 배경지식

본 논문에서는 지연 갱신을 통한 빈번한 예측 모델의 갱신을 방지함으로써, 센서 노드의 에너지 소모를 감소시킨다. 예측 모델을 이용한 데이터 모니터링에서는 정확한 값보다는 근사값을 사용하기 때문에 지연 기간 동안 발생된 측정값 또한, 근사값을 생성하여 이를 전송시키는 것이 센서의 에너지 소비를 최소화 할 수 있다. 그러나 측정값의 변동이 클 경우, 예측 모델이 부정확해져서, 센서 노드들은 계속 측정값

을 보내야 한다.

따라서 본 연구에서는 예측모델이 맞지 않을 경우, 각 센서 노드들이 측정값을 바로 보내지 않지 지연시켜 빈번한 예측모델의 갱신을 방지한다. 이 지연 기간 동안의 측정값들은 선형회귀분석[5,6]을 이용하여 근사화한다.

선형회귀에서 중요한 일은 회귀계수인 모수 α 와 β 를 추정하는 것이다. α 의 추정치를 $\hat{\alpha}$, β 의 추정치를 $\hat{\beta}$ 라고 할 때 적합회귀선은 $\hat{y} = \hat{\alpha} + \hat{\beta}x$ 로 주어지고, 여기에서 \hat{y} 는 예측값 또는 적합값을 나타낸다. 표본 $\{(t_i, v_i); i = 1, 2, \dots, n\}$ 이 주어졌을 때 모수 α 와 β 의 최소 추정값(least squares estimates)을 가지는 계수 $\hat{\alpha}$ 와 $\hat{\beta}$ 는 수식 1 과 같이 계산된다.

$$\begin{aligned} \text{mean}_v &= \frac{\sum_{1 \leq i \leq n} v_i}{n} \\ \text{mean}_t &= \frac{\sum_{1 \leq i \leq n} t_i}{n} \\ \hat{\beta} &= \frac{\sum_{1 \leq i \leq n} (t_i - \text{mean}_t)(v_i - \text{mean}_v)}{\sum_{1 \leq i \leq n} t_i - \text{mean}_t} \\ \hat{\alpha} &= \text{mean}_v - \hat{\beta} * \text{mean}_t \end{aligned} \quad (1)$$

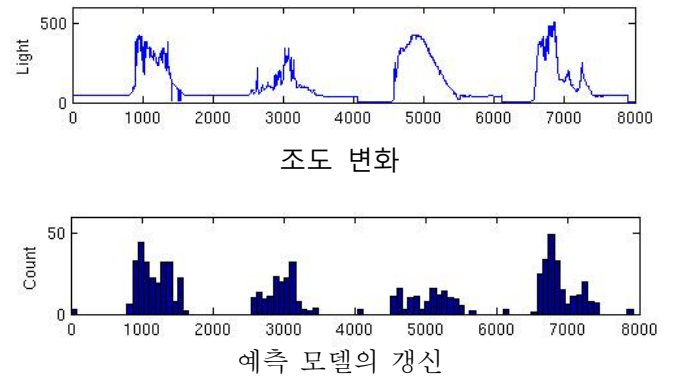
4. 지연 갱신을 이용한 예측 모델의 갱신

예측 모델을 이용한 근사적 데이터 수집 기법에서는 기지국에서 각 센서에 대한 예측 모델을 유지하여 데이터를 직접 전송 받지 않고 예측모델을 사용하여 각 센서들의 측정값을 예측한다. 따라서 각 센서를 예측값과 실제 측정값의 차이가 최대 허용 오차 (max error) 범위 이상일 경우에만 측정값을 전송하고 예측 모델을 보정한다. 기지국에서는 특정 센서로부터 측정값이 전송되지 않으면 예측값을 질의에 사용하여 사용자 질의를 수행하며, 특정 센서로부터 측정값이 전송되면 이 측정값을 사용하여 사용자 질의를 수행하고 예측 모델을 보정한다. 이와 같은 방법으로 기지국에 각 센서의 측정값을 오차 범위 이내로 유지할 수 있다.

기존에 사용되는 예측 모델들은 현재 데이터가 과거의 데이터와 유사할 것이라는 가정에 기초를 두고 있다. 따라서 시간에 따른 값의 변화가 크지 않을 때, 최적의 성능을 가진다. 그러나 값의 변화가 심한 경우에는 예측 모델이 자주 갱신되는 문제점이 있다. 그림 1 은 시간에 따른 조도 변화와 그에 따른 예측 모델 (선형회귀모델) 의 갱신을 히스토그램으로 표현

한 것인데 값의 변화가 심할수록 예측 모델의 갱신도 자주 일어남을 확인할 수 있다.

짚은 예측 모델 갱신을 방지하기 위하여 본 논문에서는 예측이 실패하면 바로 예측 모델을 갱신하는 기법 대신 지연 갱신을 통하여 전송 비용을 감소시키는 기법을 제시한다. 본 논문에서는 지연 갱신을 위하여 특정 구간의 측정값들을 선형회귀분석기법을 이용한 상관계수들만으로 표현하여 이를 전송하도록 한다.



(그림 1) 조도 변화에 따른 예측 모델의 갱신 빈도

본 논문에서는 예측 모델의 짚은 갱신을 방지하기 위해 측정값들을 지연, 근사하여 보내므로 기지국에서는 측정값들의 획득을 허용시간만큼 지연할 필요가 있다. 따라서 본 논문에서는 이를 위하여 TinyDB [4]에서 제안된 SQL 문을 다음과 같이 확장한다.

```
SELECT nodeid, attr
FROM sensors
SAMPLE INTERVAL t1 FOR t2
ALLOW MAX DELAY td
```

위 질의는 센서 네트워크에서 사용되는 기본적인 질의인 SELECT-FROM-WHERE 형태를 확장한다. 노드 식별자(node id)와 속성(attr)를 t₂ 동안 t₁ 간격으로 수집한다. 기존의 질의에서 확장된 부분은 ALLOW DELAY 절로써, 사용자는 측정값을 획득 시점을 최대 t_d 까지 지연하도록 한다. 이를 통하여 센서 네트워크의 수명을 늘일 수 있다.

한 노드에서 특정 시점 t 에 예측 모델이 갱신되고 이후 예측이 실패할 때까지의 구간을 예측 구간 (estimation interval)이라고 하고, 예측이 실패한 후 지연 갱신을 수행하는 구간을 지연 구간 (delay interval)

이라고 하자.

이때 각 노드는 지연 구간내의 측정값들 v 를 시간 t 에 대한 선형회귀분석식을 이용하여 $\hat{v} = \hat{a} * t + \hat{b}$ 으로 나타내고 \hat{a} 과 \hat{b} 을 전송하도록 한다. 이때, v 와 \hat{v} 의 차이는 최대허용오차(max_error)를 넘어서는 안된다. 따라서, 시간 t_i 에 이 차이가 허용 오차를 넘길 경우, \hat{a} 과 \hat{b} 그리고 t_i 시간의 측정값 v_i 를 전송한다. 또한, 지연구간의 길이는 최대 t_d 임으로 지연 구간의 길이가 t_d 를 넘길 경우에서 상관계수 \hat{a} 과 \hat{b} 및 현재 측정값을 전송하도록 한다.

기지국에서는 상관계수를 받게 되면 이를 이용하여 지연구간내의 측정값들의 근사값을 생성할 수 있다. 또한, 기지국과 센서 노드들은 이 근사값을 이용하여 예측 모델을 갱신한다.

앞에서 설명한 바를 정리하면 그림 2 의 알고리즘 1 과 같다. 알고리즘 1 은 센서가 외부환경을 측정하여 측정값을 생성할 때 마다 수행된다.

알고리즘 1 에서는 몇 개의 변수들을 사용한다. (\hat{a}, \hat{b}) 는 값을 예측하기 위한 계수들이며, max_error 는 최대 허용 오차, t_d 는 최대 허용지연시간이다.

또한 DT 는 지연구간에 속하는 측정값들의 집합이다. 따라서 시간 t_i 에서 DT 의 크기가 0 이면 t_{i-1} 시간까지의 측정값은 예측구간에 속하고 DT 의 크기가 0 보다 크다면 이면 t_{i-1} 시간까지의 측정값은 지연구간에 속한다는 것을 파악할 수 있다.

센서가 t_i 시간에 측정값 v_i 를 생성한 후 (line 1), t_{i-1} 시간의 값이 예측구간에 속했는지 지연 구간에 속했는지를 확인한다. (line 2).

만약 예측 구간이면 (line 3-7), 예측 모델(estimation mode)이 현재 측정값 v_i 를 적절히 반영하고 있으면 (line 4) 기지국의 현재 측정값을 보내지 않는다. 만약 예측모델이 정확하지 않다면 v_i 를 DT 에 삽입한다. (line 6-7)

지연 구간에서는 (line 8-15), 최대 지연 구간의 길이는 t_d 임으로 DT 크기가 t_d 보다 작다면 (line 9), 현재 측정값을 DT 에 추가해도 이에 대한 선형회귀분석식의 오차가 최대 허용 오차보다 작을 경우 이를 DT 에 추가 한다 (line 10-11).

만약 DT 의 크기가 t_d 이거나 현재 측정값을 DT 에 추가하면 최대허용오차를 벗어날 경우 (line 12), DT 를 이용한 선형회귀분석식의 상관계수 \hat{a}, \hat{b} 와 길이 |DT|

를 전송하고 DT 를 초기화 한다.

알고리즘1

```

1 read  $v_i$  at  $t_i$  //data sampling at time  $t_i$ 
2 If |DT| = 0 // estimation interval
3  $\hat{v} =$  estimation model( $t_i$ )
4 If  $|v_i - \hat{v}_i| \leq max\_error$ 
5 do nothing
6 else
7 DT =  $\{v_i\}$ 
8 else //delay interval
9 If |DT| <  $t_d$ 
10 if error(DT U  $\{v_i\}$ )  $\leq max\_error$ 
11 DT = DT U  $v_i$ 
12 If |DT| =  $t_d$  OR
error(DT U  $\{v_i\}$ )  $\leq max\_error$ 
13  $[\hat{a}, \hat{b}] =$  regression(DT)
14 send  $[\hat{a}, \hat{b}]$  and |DT|
15 DT = null

```

(그림 2) 지연 갱신을 이용한 예측 알고리즘

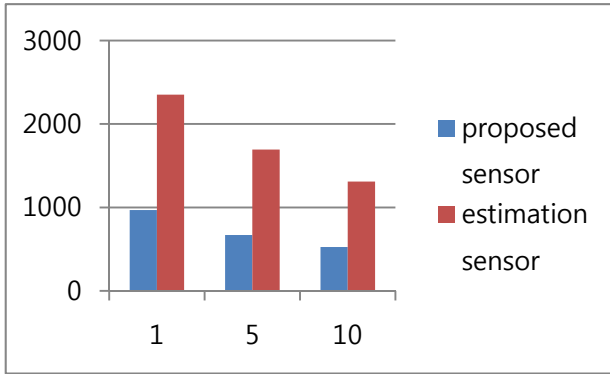
5. 성능평가

우리는 본 논문에서 제시한 알고리즘을 평가하기 위해 시뮬레이션을 진행하였다. 기존의 예측기법과 본 논문의 예측기법을 비교하며 예측 기법으로는 선형회귀를 사용하였으며 앞에서 언급한 바와 같이, 센서 에너지 소비 형태 중 가장 많은 부분을 차지하는 부분은 통신 에너지이므로 평가 항목으로는 전송 횟수만을 고려한다.

센서 네트워크에는 10 개의 센서가 존재하며 각 센서의 데이터는 Intel, Berkely reseach lab (<http://db.csail.mit.edu/labdata/labdata.html>)에서 측정한 조도를 이용하였다. 각 센서의 초당 샘플링 횟수는 1 회이고 센서 네트워크는 8000 초동안 유지된다.

허용 오차의 변화에 따른 전송 횟수의 변화

그림 4 는 센서 네트워크에서 허용 오차의 따른 센서들의 평균 전송 횟수를 비교한 결과이다. 제안한 기법의 경우 사용자는 10 초동안의 지연을 허용한다. 결과를 살펴보면 허용 오차에 관계 없이 제안한 기법이 기존의 예측모델의 결과보다 약 50%이상 전송 횟수가 감소하는 것을 알 수 있다.



(그림 3) 허용 오차의 변화에 따른 전송 횟수

허용지연시간의 변화에 따른 전송 횟수

그림 4 은 허용지연시간의 변화에 따른 전송횟수의 변화이다. 최대 허용 오차는 5 로 하였다. 그림 4 에서 보듯이 허용지연시간이 늘어날수록 데이터 전송 횟수가 줄어든 것을 확인할 수 있는데, 이는 지연 구간의 크기가 늘어남으로써 더 많은 측정값들을 근사화할 수 있기 때문이다.

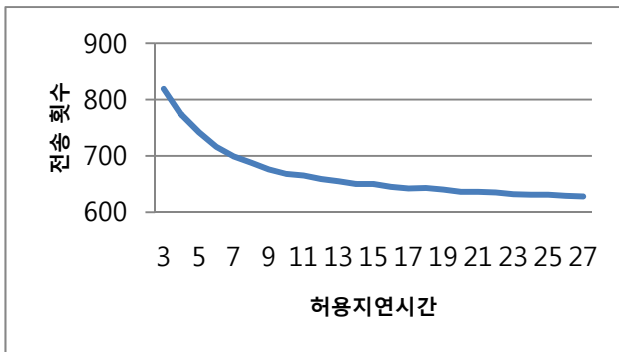


그림 4. 허용지연시간의 변화에 대한 데이터 전송 횟수

6. 결론

앞에서 우리는 기존의 예측 모델을 보완하기 위해서 센서 네트워크에서의 지연 갱신을 예측 모델의 갱신을 제안하였다. 기존의 예측 모델들은 측정값들을 실시간에 가깝게 획득할 수 있으나, 예측 모델의 잦은 갱신을 요구한다. 따라서 예측 모델의 잦은 갱신을 방지하기 위해 예측이 실패하면 허용지연시간만큼 값을 근사화하여 전송함으로써 데이터 전송을 최소화 할 수 있다.

참고 문헌

[1] Y. Kotidis. "Snapshot Queries: Towards Data-Centric Sensor Networks", In proceedings of ICDE. 2005.
 [2] D. Tulone and S. Madden, "PAQ: Time series forecasting for approximate query answering in sensor

networks", In Proceedings of EWSN, 2006.
 [3] A. Silberstein, R. Braynard and J. Yang. "Constraint Chaining: On Energy-Efficient Continuous Monitoring in Sensor network", In Proceedings of ACM SIGMOD, 2007.
 [4] Samuel Madden, Michael J. Franklin, Joseph M. Hellerstein and Wei Hong, "The Design of an Acquisitional Query Processor For Sensor Network", In Proceedings of SIGMOD, 2003.
 [5] E. Keogh, S. chu, D.Hard and M. Pazzani, Segmenting Time Series: A Survey and Novel Approach. Data mining in Time series databases, World Scientific.
 [6] Roland E. Walpole, Raymond H. Myers, Sharon L. Myers and Keying Ye. Probability & Statistic for Engineers & Scientists, Pearson Education. 2007.
 [7] S. Madden, M. J. Franklin, J. M. Hellerstein, and W. Hong. "Tag: A Tiny Aggregation Service for ad hoc Sensor Networks", In proceedings of OSDI Conference, 2002
 [8] A. Deligiannakis, Y. Kotids and N. Roussopoulos, "Compresson Historical Information in Sensor Networks." In Proceedings of ACM SIGMOD, 2004
 [9] A. Jain, E. Y. Chang and Y. F. Wang, "Adaptive Stream Resource management using Kalman Filters," In Proceedings of ACM SIGMOD, 2004.