

결함이 발생하는 데이터 중심 저장 환경에서 높은 정확도를 갖는 소실 데이터 보정 기법

박준호[○] 이효준 성동욱 유재수

충북대학교 정보통신공학과

arionfit@naver.com, {reverse999, seong.do}@gmail.com, yjs@chungbuk.ac.kr

A Data Correction Method for High Accuracy in Data Centric Storage Schemes with Faults

Junho Park[○] HyoJoon Lee DongOok Seong JaeSoo Yoo

Dept. of Information and Communication Engineering, ChungBuk Nat'l Univ.

최근 무선 센서 네트워크 환경에서 감지되는 데이터를 네트워크상에 효과적으로 저장하고 처리하기 위한 다양한 기법들이 제안되었다. 대표적인 연구로 데이터 중심 저장 기법이 있다. 데이터 중심 저장 기법은 수집한 데이터를 데이터의 값을 기반으로 하는 해시 함수나 지역적인 위치에 의해 분류 되어 특정 센서 노드에 저장한다. 그러므로 질의 생성 시, 네트워크 전체에 데이터를 요구하는 것이 아니라, 질의에 해당하는 데이터를 저장하는 노드만 데이터를 수집하기 때문에 질의 처리의 관점에서 효과적이다. 하지만, 센서 네트워크는 다양한 환경에 배포가 되기 때문에 통신 채널, 하드웨어 잡음과 같이 데이터 송수신에서 발생할 수 있는 오류에 노출되어 있고, 외부 요인에 의한 노드 과손도 쉽게 발생한다. 뿐만 아니라, 데이터 중심 저장 기법의 특성에 따라 저장소 집중과 질의 집중에 의한 에너지 소진에 따른 결함도 발생한다. 이러한 센서 네트워크에서의 데이터 중심 저장 기법은 노드 결함이 발생이 발생하면 특정 범위의 해당하는 전체 데이터가 소실되고, 소실된 데이터의 누락에 의한 질의 정확도 저하 문제가 발생한다. 이를 해결하기 위해 노드 결함에 따른 소실 데이터 보정 기법에 대한 연구가 제안 되었다. 기존 연구에서는 데이터 소실 영역에 대해 선형 회귀 분석 기법을 이용하여, 보정 모델을 생성하고, 모델에 따른 가상의 데이터를 질의 결과에 반영함으로써 질의 결과를 보정한다. 하지만, 기존 연구는 단지 선형 회귀 분석 기법을 이용하여 데이터 분포 패턴 및 변화율을 고려하지 않은 일직선 형태의 단순한 보정 모델을 생성함으로써, 낮은 정확도의 가상 질의 결과가 생성 되는 문제를 가지고 있다. 따라서 본 논문에서는 센서 노드나 네트워크 결함이 발생하는 데이터 중심 저장 기법에서의 수집 데이터 분포 패턴 및 변화율을 고려한 소실 데이터 보정 기법을 제안한다.

질의가 발생하면 질의 노드에서 질의 범위에 해당하는 데이터를 저장하는 노드로 데이터 요청 메시지가 전송된다. 일반적으로 센서 네트워크에서는 노드 간에 주기적인 통신을 통해 이웃 노드의 결함을 인지 할 수 있다. 따라서 질의 배포 시에 인접한 이웃 노드의 결함을 감지했을 경우, 결함 노드에 대한 정보를 질의 노드로 재전송한다. 질의 범위에 해당하는 모든 노드의 데이터 및 결함 노드 정보를 수집 및 취합을 담당하는 질의 노드에서는 소실 데이터 영역에 대한 가상 데이터를 생성하기 위한 준비 과정으로서, 질의 범위에 해당하는 모든 노드에서 전송 받은 데이터 및 결함 발생 노드에 대한 정보를 바탕으로 데이터 분포도의 작성 및 분석을 수행한다. 분석을 통해 생존 데이터 영역 및 소실 데이터 영역에 대한 정보를 그림 1과 같이 생성하여, 데이터 존재 영역 및 소실 영역을 인지하는 것이 가능하며, 이를 이용하여 센서 데이터의 특성에 기반을 둔 소실 데이터 보정 모델 생성 및 가상 데이터 생성 과정을 수행한다.

센서 네트워크를 통해 수집 되는 센서 데이터는 환경적인 특성상 물리적인 연관성을 지닌다. 다시 말해, 센서를 통해 측정하는 대부분의 물리량은 급격하게 변화하기 보다는 연속적인 값을 가지게 됨을 의미한다. 그러므로 수집 된 센서 데이터는 급격한 변화율을 가지기 보다는 인접한 데이터가 가지는 변화율의 연속적인 증가 및 감소 특성에 근거하여 변화하게 된다. 본 기법은 이러한 센서 데이터의 특성을 활용하여, 보정 모델

본 연구는 교육과학기술부와 한국산업기술재단의 지역혁신인력양성사업과 2009년도 정부(교육과학기술부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업의 결과임.(No. 2009-0080279)

을 생성한다. 앞서 수집한 전체 데이터를 분석하여 생존 데이터 영역 및 소실 데이터에 대한 취합을 마친 후, 생존 데이터 영역에서 각 변수 사이에 변화량을 나타내는 변화율 연산을 수행한다. 그림 3에서 생존 데이터 영역에 대한 미분 연산을 수행할 경우, 그림 4과 같이 변화율 그래프를 얻는 것이 가능하다.

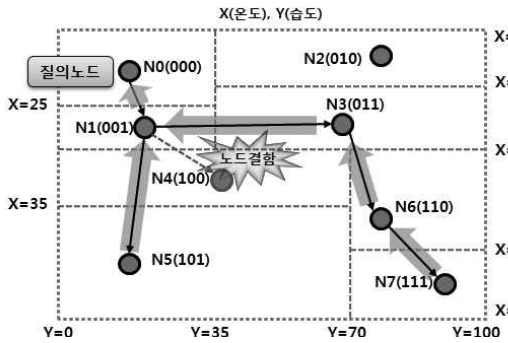


그림 1. 질의 배포와 노드의 결함

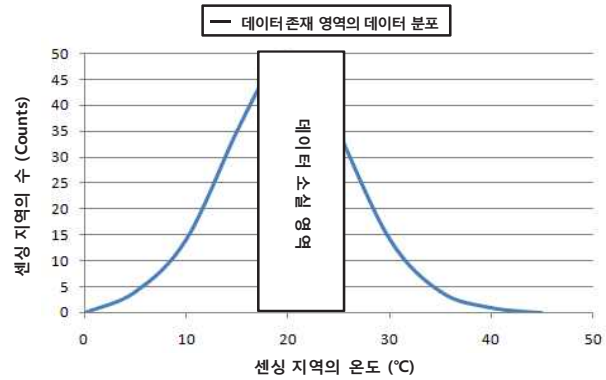


그림 2. 데이터 분포 및 데이터 소실 영역

수집된 센서 데이터의 분포 그래프를 $y = f(x)$ 로 표현한다고 할 때, 이는 온도 x 를 수집한 센서의 수를 나타낸다. 온도 $x = x_1$ 과 $x = x_2$ 사이의 수집한 센서의 변화는 $\Delta y = f(x_2) - f(x_1)$ 이다. 따라서 온도 $x_1 \leq x \leq x_2$ 동안의 센싱 데이터의 평균 증가율 혹은 감소율은 $\Delta y / \Delta x = f(x_2) - f(x_1) / x_2 - x_1$ 이다. 그러므로 그래프에서 소실 영역의 증가율 혹은 감소율은 정확하게 추정하는 것은 불가능하지만, 가상 평균 변화율은 인접한 생존 데이터의 변화율을 바탕으로 유추하는 것이 가능하다. 이와 같은 사실에 기반을 두어 소실 영역에 대한 가상 평균 변화율을 생성하고 평균 변화율에 대한 적분 연산을 수행할 경우, 기존의 소실된 영역의 데이터와 유사한 형태를 가지는 복원 모델을 생성할 수 있다. 이를 이용하여 소실 영역에 대한 가상의 데이터를 포함한 질의 결과를 반환할 경우, 기존의 기법에 비해 높은 수준의 정확도를 가지는 질의 결과를 반환하는 것이 가능하다.

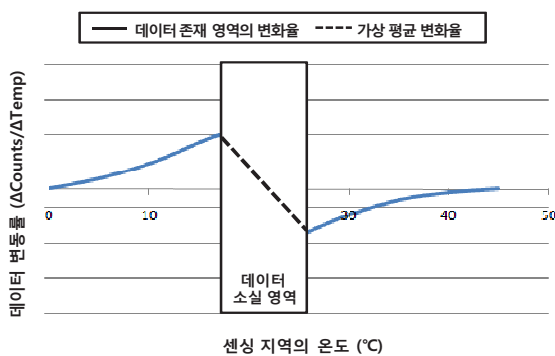


그림 3. 데이터 존재 영역의 데이터 변화율 및 소실 영역의 가상 평균 데이터 변화율

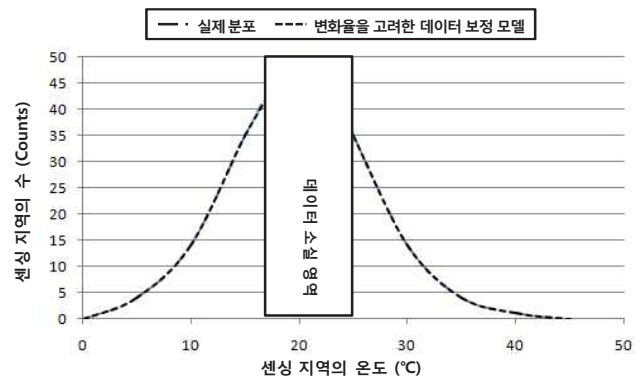


그림 4. 소실 영역에 대한 복원 모델

제안하는 기법의 우수성을 보이기 위해 시뮬레이션을 통해 기존의 소실 데이터 보정(E-KDDCS) 기법과 제안하는 기법과의 성능을 비교하였다. 그 결과 19% 이상의 질의 결과 정확도 향상을 보였고, 노드 결함율을 증가시키면서 수행한 평가에서도 기존 기법에 비해 평균 16% 이상의 질의 결과 정확도 향상을 보였다.

참고문헌

[1] 이효준, 박준호, 성동욱, 유재수, “데이터 중심 저장 환경에서 소실 데이터 보정 기법을 이용한 인-네트워크 질의 처리,” 한국정보과학회 2010 한국컴퓨터종합학술대회 논문집, pp. 337-342, 2010.