

무선 센서 네트워크 환경에서 기계 학습을 이용한 데이터 추론에 관한 연구

정용진 · 조경우 · 오창현

한국기술교육대학교

A Study on Data Inference using Machine Learning in WSN Environment

Yong-Jin Jung · Kyoung-Woo Cho · Chang-Heon Oh

Korea University of Technology and Education

E-mail : jungyj0211@koreatech.ac.kr

요 약

무선 센서 네트워크 환경에서 센서 노드로부터 수집되는 데이터의 소실은 센서 노드의 히든노드 및 전력 부족 등의 문제로 발생된다. 이러한 문제를 해결하기 위해 효율적으로 망을 유지하기 위한 연구가 활발히 진행되어 왔으나, 망 유지가 불가능한 상황의 연구는 전무하다. 따라서 망 유지가 불가능한 상황에서의 소실된 데이터를 추론하기 위한 연구가 필요하다. 본 논문에서는 소실된 데이터를 추론하기 위해 특정 도시의 미세 먼지 데이터를 이용한다. 기계 학습을 통해 축적된 데이터의 분석 및 소실된 데이터의 추론 가능성을 확인한다.

ABSTRACT

The loss of data collected from the sensor node in the wireless sensor network environment is caused by the hidden node of the sensor node and power shortage. In order to solve these problems, researches have been actively carried out to maintain the network effectively, but there is no study on the situation where the maintenance of the network is impossible. Therefore, research is needed to infer lost data in situations where network maintenance is impossible. In this paper, use particulate matter data of specific cities to deduce lost data. Analyze the accumulated data through machine learning and identify the possibility of inferring lost data.

키워드

Wireless Sensor Networks, Machine Learning, Data Inference

I. 서 론

무선 센서 네트워크(Wireless Sensor Networks)는 다양한 환경과 관점에서 여러 사용자에게 정보 및 편의를 제공하기 위해 사용되고 있다. 정보 및 편의를 제공하기 위해 센서 노드로부터 수집되는 데이터의 소실은 지속적인 서비스 제공에 문제를 야기한다. 데이터의 소실은 무선 센서 네트워크 환경에서의 에너지 고갈, 물리적 손상, 네트워크 문제 등으로 발생할 수 있다. 이러한 문제를 해결하기 위해 여러 연구가 진행되고 있으나 에너지 효율, 최적 경로 탐색 등 효율적인 망 유지를 위한 연구가 주를 이루고 있으며[1],[2], 히든 노드 발생 및 노드 고장 등의 물리적 문제로 인한 데이터 소실 상황에 대해 선행 연구 사례는 전무하다.

본 논문에서는 센서 노드들의 축적된 과거 데이터

의 기계 학습(Machine Learning)을 통해 분석 및 추론의 가능성을 확인한다.

II. 학습 모델 구축

축적 데이터는 2018년 4월 1일부터 동년 동월 30일까지 천안시 백석동, 성황동, 성성동의 측정소에서 수집된 미세 먼지(PM-10)를 이용하였다. 기계 학습의 학습 방법은 입력과 출력으로 구성된 샘플 데이터로부터 입력을 출력으로 사상하는 함수를 유추하는 지도 학습(Supervised Learning)을 이용하였다. 지도 학습은 분류와 회귀로 구성되며 연속적인 숫자, 또는 부동소수점수를 예측하기 위한 선형 회귀 모델과 결정 트리 회귀 모델을 이용하였다[3][4].

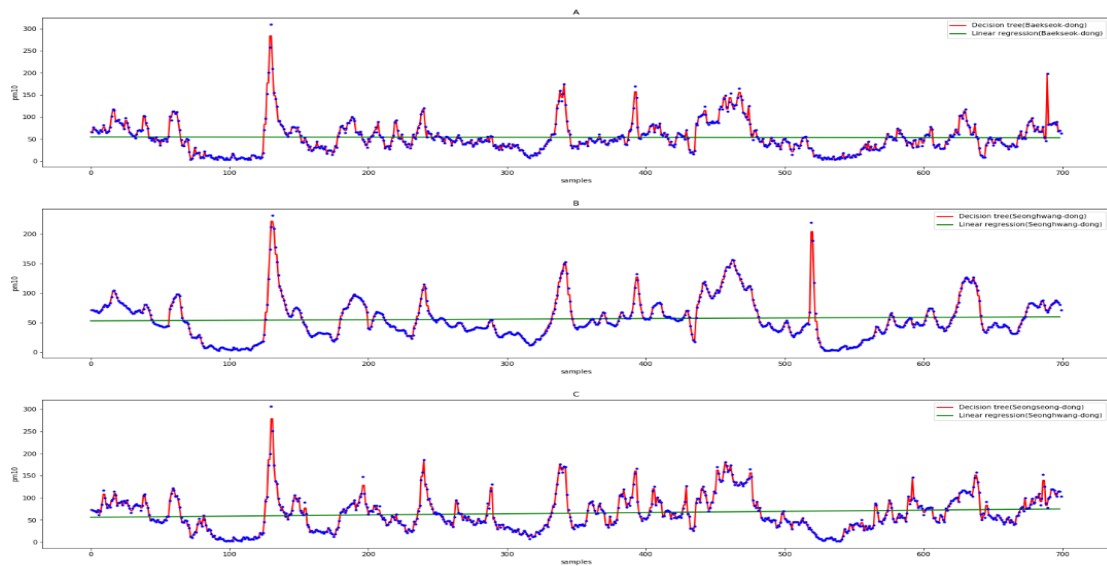


그림 1. 선형 회귀 모델과 결정 트리 회귀 모델 비교(백석동, 성황동, 성성동)

그림 1은 선형 회귀 모델과 결정 트리 회귀 모델을 비교한 그림이다. 위에서부터 백석동, 성황동, 성성동에 해당하며, x축은 720개의 시계열 샘플을 의미하며, y축은 미세 먼지 농도를 의미한다. 선형 모델은 선형 관계로만 모델링하므로 특성이 하나일 때 직선으로 나타난다. 결정 트리 회귀 모델은 데이터의 표현 형태에 따라 정확해 보이지만 복잡한 특성을 가진다. 미세 먼지와 같이 연속 데이터의 경우 여러 특성을 주어 강력한 회귀 모델로 만들어 사용할 수 있다[5].

회귀 모델에 여러 특성을 적용하기 위해 이산화(binng) 방법을 적용한다. 결정 트리 회귀 모델의 경우 데이터를 자유롭게 나눠 학습할 수 있으므로 특성의 값을 구간으로 나누는 이산화 방법이 무의미하다. 하지만 선형 회귀 모델의 경우 특성에 따른 각 구간마다 다른 값을 가지고 있는 것을 확인 하였으며, 이는 선형 회귀 모델은 이산화 방법이 적용되어 보다 유연한 모델로 사용 가능한 것을 확인하였다.

선형 회귀 모델에 더 많은 특성을 부여하기 위해 원본 데이터에 상호작용(interaction)과 다항식(polynomial)을 추가하는 방법이 있다. 상호작용은 이산화된 선형 회귀 모델에 기울기를 추가하는 방식으로 구간으로 분할된 데이터에 기존의 특성을 다시 추가하는 방법이다. 이 방법은 x축 특성이 하나이므로 기울기도 하나인 특징이 있으며, 이러한 특성으로 모든 구간에서 기울기가 같은 값을 가지게 된다. 다항식은 구간으로 분할된 데이터 특성과 x축 사이의 상호작용 특성을 추가하는 방식으로 구간 특성과 원본 특성을 곱하는 방법이다. 이 방법은 x축 사이에 특성이 추가되므로 각 구간에서 다른 기울기를 가지게 된다. 다항식 방법의 경우 상호작용 방법과 비교하여 각 구간에서 다른 기울기를 보여주고 있음을 확인하

였으며, 이는 각 구간에서 다른 기울기를 필요로 하는 미세 먼지 데이터의 경우 상호작용 방법보다 다항식 방법을 적용하는 것이 보다 유연한 모델로 사용이 가능하다는 것을 확인 할 수 있다.

그림 2는 이산화와 다항식 방법이 적용된 선형 회귀 모델이다. 그림 1과 그림 2를 비교하였을 경우 직선으로 표현되던 선형 회귀 모델이 이산화와 다항식 방법을 통해 추가된 특성으로 인해 원본 데이터와 유사한 모델이 되었음을 확인할 수 있다.

III. 분석 및 소실 데이터 추론

그림 2와 같이 천안시 3지역의 시간에 따른 미세 먼지 농도 변화 주기는 미세 먼지의 농도가 증가함과 감소하는 시기가 거의 일치함을 확인할 수 있다. 이는 해당 지역의 미세 먼지 농도변화 시 다른 지역의 미세 먼지 농도 비슷하게 변화한다는 것을 알 수 있으며, 특정 지역의 미세 먼지 측정 센서가 데이터 손실을 발생하여도 다른 지역과 비교하여 데이터를 추론할 수 있음을 알 수 있다.

추론의 과정은 다음과 같다.

1. A지역 데이터 손실 날짜 확인
2. B, C 지역 해당 시간의 값 추출
3. 각 지역에서 2번의 값과 동일한 시간 추출
4. 타 지역 추출 시간을 이용하여 A지역 해당 시간의 미세먼지 농도 추출

2018년 4월 11일 20시 성성동 미세 먼지 농도 데이터가 손실되었다고 가정하였을 경우, 해당하는 시간의 선형 회귀 모델 x좌표의 값은 251이며,

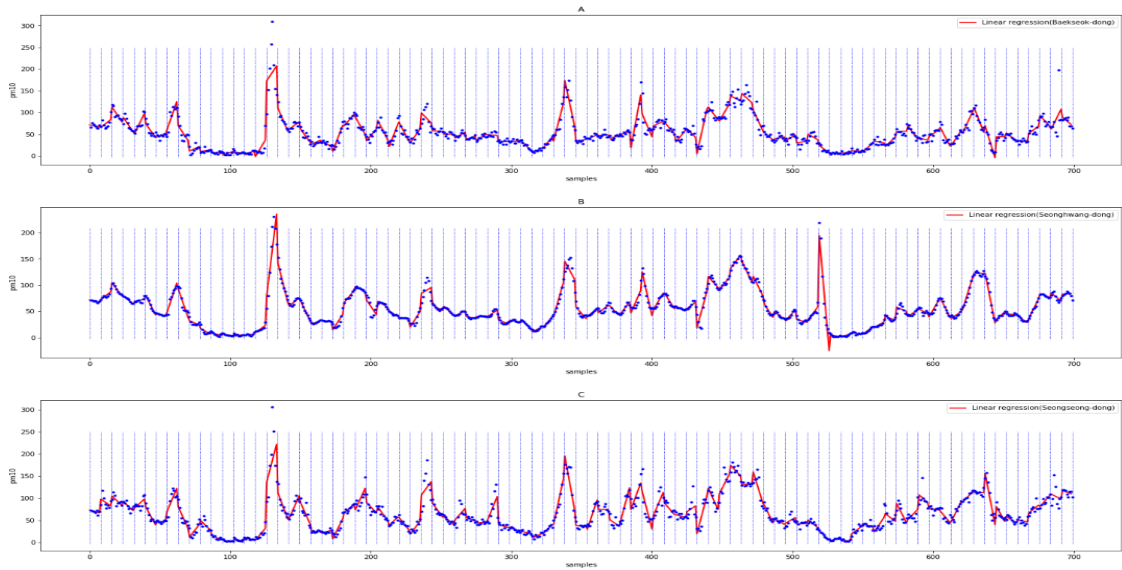


그림 2. 이산화, 방정식 방법이 적용된 선형 회귀 모델(백석동, 성황동, 성성동)

백석동과 성황동의 모델에 대입한 결과 54.0, 58.0의 값을 추출하였다. 추출된 값을 가지는 타 지역의 시간을 추출한 결과 20가지의 시간이 추출되었으며 해당 시간에 대한 성성동의 미세 먼지 농도 평균은 73.75로 추론할 수 있다. 추론된 미세 먼지 농도는 실제 미세 먼지 농도인 65와 8.75차이를 보이는 것을 확인하였다.

IV. 결 론

본 논문에서는 히든 노드 발생 및 노드 고장 발생 등에 따른 데이터 소실 상황에 대해 각 센서 노드들의 축적된 과거 데이터 기계 학습을 통해 소실 데이터 추론에 대해 연구하였다.

소실된 데이터 추론을 위해 기계 학습의 지도 학습 중 이산화, 다항식 방법이 적용된 선형 회귀 모델을 적용하여 천안시 지역 미세 먼지 데이터에 대한 모델을 구축하였다.

구축한 모델을 기반으로 인근 지역의 유사한 미세 먼지 농도 변화 추이를 확인하였으며, 데이터가 소실된 지역의 기계 학습 모델과 타 지역의 기계 학습 모델을 이용하여 가상의 소실 데이터를 추론하였다. 추론 결과 실제 미세 먼지 농도인 65와 8.75 차이를 보이는 73.75의 값을 추론하였다.

향후 다양한 모델의 적용과 데이터 추론의 오차를 줄이기 위한 연구가 필요하며, 모델에 대한 평가와 성능 향상에 대한 연구를 진행한다.

참고문헌

[1] 김지원, 조태환, 최상방, “데이터의 최단 거리 전송을 위한 가상 주소 라우팅 프로토콜.” *한국통신학회논문지*, vol. 42, no. 1, pp. 50-63, 2017.

[2] Aranda, J., Carrillo, H., & Mendez, D. “Enhanced multimodal switching mechanisms for node scheduling and data gathering in wireless sensor networks,” *In Communications and Computing (COLCOM), 2017 IEEE Colombian Conference on*, pp. 1-6, Aug. 2017.

[3] 최성찬, “사물인터넷 환경 머신러닝 기술 동향,” *IoT 사물인터넷 포럼*, vol.2, pp. 1-3, 2017.

[4] 문성은, 장수범, 이정혁, 이종성, “기계학습 및 딥러닝 기술 동향,” *한국통신학회지*, vol. 33, no. 10, pp. 49-56, 9. 2016.

[5] G. Sarah, M. C. Andreas, *Introduction to Machine Learning with Python*, O’Reilly Media, Inc, 2016.08