

분할 선형 회귀를 이용한 Emergency node 감지 모델 연구

김세준[○], 임환희^{*}, 이병준^{*}, 김경태^{*}, 윤희용^{**}

[○]성균관대학교 정보통신대학 전자전기컴퓨터공학과

^{**}성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {ksj105,lhh423,byungjun}@skku.edu[○], kyungtaekim76@gmail.com^{*}, youn7147@skku.edu^{**}

A Study on Emergency Node Detection Method based on Segmented Linear Regression

Se-Jun Kim[○], Hwan-Hee Lim^{*}, Byung-Jun Lee^{*}, Kyung-Tae Kim^{*}, Hee-Yong Youn^{**}

[○]Dept. of Electrical and Computer Engineering, Sungkyunkwan University

^{**}Dept. of Software, Sungkyunkwan University

● 요약 ●

본 논문에서는 산업 IoT (IIoT) 환경에서 생산 설비 내 각 센서 노드의 데이터 이상 여부를 게이트웨이에서 판단하는 Emergency node 선정 모델을 제안하였다. 이 모델은 IIoT 환경이 적용된 생산 설비의 Emergency 상태 즉, 이상 동작으로 인한 온도, 진동 데이터 등의 비정상적인 수집을 구분하여 즉각적으로 대응할 수 있도록 하는 것을 목표로 한다. 본 논문에서는 분할 선형 회귀를 통하여 주기 내 데이터의 허용 범위를 계산하여 기존의 Threshold 방식보다 정확하고 범용적으로 Emergency node를 분류한다.

키워드: 기계학습(machine learning), IoT(Internet of Things), 데이터 분류(Data Classification)

I. Introduction

최근 IoT 기술에 대한 관심이 증가하고 관련 연구가 활발히 이루어짐에 따라, 실제 환경에 대한 적용 방안 또한 다양하게 제시되고 있다. 이 중 특히 스마트 팩토리, 스마트 팜 등 산업 환경에 IoT를 적용하여 생산설비의 효율적인 모니터링 및 제어하는 Industrial IoT (IIoT) 의 연구가 집중적으로 이루어지고 있으며, 실제 적용 사례도 나타나고 있다. IIoT는 생산 설비에 주로 센서로 이루어진 IoT 노드들이 배치됨으로써 설비의 동작을 모니터링하고 지역적, 전역적으로 제어하는 것을 목표로 한다. 특히 이러한 모니터링은 생산 설비의 이상 유무를 감지하는데 주목적이 있기 때문에, 센서 노드로부터 전송되는 센싱 데이터가 정상적인지 분류하는 알고리즘이 요구된다 [1]. 기존에는 이를 위하여 센싱 데이터 값이 Threshold 를 넘으면 경고하는 기법이 적용되어 왔으나, 최고 혹은 최저점 사이에서는 이상 감지가 어렵다는 문제점이 있었다. 본 논문에서는 이를 해결하기 위하여 분할 선형 회귀의 Predicted interval을 이용하여 데이터의 경향에 근사한 이상 유무 판단 경계값을 찾는 기법을 제안한다.

II. Preliminaries

1. Related works

1.1 분할 선형 회귀

분할 선형 회귀는 일반적인 선형 회귀방식을 특정 구간별로 적용함으로써, 각 구간의 선형 회귀를 구하는 방법이다. 각 구간은 데이터의 시그마 함수의 이차 도함수가 0이 되는 지점을 통해 구해지는데, 먼저 주어진 데이터 셋 간의 오차 거리를 최소화하는 최적의 선을 찾기 위해 General least-squares method를 사용한다.

$$s = \sum_{i=1}^n (f_i - y_i)^2$$

이에 대한 1차 미분 값은 아래의 방정식과 같이 극점이 0에 도달하는 지점을 나타낸다.

$$\frac{ds}{db_i} = 2 \sum_{i=1}^n (f_i - y_i) \frac{df_i}{db_i}$$

2차 미분은 점이 유일하고 최소값에 도달했는지 여부를 확인한다. 아래의 식은 2차 미분을 보여준다. 여기서 s 는 Unknown parameters의 index를 나타낸다.

$$\frac{d^2s}{db_i^2} = 2 \sum_{i=1}^n [(f_i - y_i) \frac{d^2f_i}{db_i^2} + (\frac{df_i}{db_i})^2]$$

최종적으로 아래의 식을 만족할 때 극점이 결정된다.

$$\sum_{i=1}^n (f_i - y_i) \frac{df_i}{db_i} = 0$$

1.2 Predicted interval

Predicted interval은 특정 회귀 함수의 결과에 영향을 미치는 데이터의 범위를 나타낸 것으로, 본 논문에서는 데이터의 정상 범위를 나타낸다. 아래의 식은 선형 회귀 데이터의 예측 값, 즉 Predicted interval의 최대, 최소값을 나타낸다.

$$y_0 = \pm t_{\frac{\alpha}{2}, n-2} \sqrt{\frac{\sum (y - y^{\wedge})^2}{n-2}} \times (1 + \frac{1}{n} + \frac{(x_0 - x^{\wedge})^2}{\sum x - x^{\wedge}})$$

III. The Proposed Scheme

본 논문에서 제안하는 모델은 먼저 기존 데이터의 주기를 파악하여 분할 선형 회귀를 구하는 것으로부터 시작된다. 분할 선형 회귀를 구한 후에는 이에 대한 Predicted interval을 구한다. Predicted interval이 설정되고 난 후에는 게이트웨이는 실제로 센서에서 전송되는 데이터와 비교하여 데이터가 Predicted interval의 Upper bound를 수 회 이상 초과하면 해당 데이터를 전송하는 노드를 Emergency node로 설정한다. 이를 통하여 관리자가 설비 내에서 전송되는 모든 데이터를 분석하지 않아도 미리 게이트웨이에서부터 생산 설비의 이상 유무를 판별할 수 있게 된다.

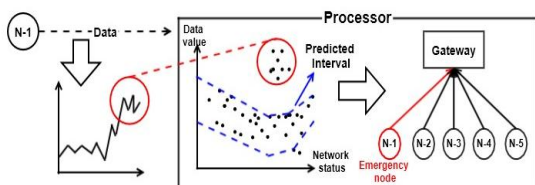


Fig. 1. 제안하는 Emergency node 판별 시스템 구조

IV. Conclusions

본 논문에서는 산업 IoT (IIoT) 환경에서 생산 설비 내 각 센서 노드의 데이터 이상 여부를 게이트웨이에서 판단하는 Emergency node 선정 모델을 제안하였다. 해당 모델은 기존의 Threshold를 이용한 데이터의 이상 유무 판별 기법보다 높은 정확도의 판별 성능을 나타낼 것으로 기대되며, 향후 연구로는 이를 실제로 구현하여 네트워크 환경에서 시뮬레이션 함으로써 성능을 평가한다.

ACKNOWLEDGEMENT

본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 정보통신-방송연구 개발 사업(No. 2016-0-00133, 초연결 IoT 노드의 군집 지능화를 통한 Edge Computing 핵심 기술 연구), SW중심대학지원사업(2015-0-00914), 한국연구재단 기초연구사업(No.2016R1A6A3A11931385, 실시간 공공안전 서비스를 위한 소프트웨어 정의 무선 센서 네트워크 핵심기술 연구, 2017R1A2B2009095, 실시간 스트림 데이터 처리 및 Multi-connectivity를 지원하는 SDN 기반 WSN 핵심 기술 연구), 삼성전자, BK21PLUS 사업의 일환으로 수행되었음.

REFERENCES

- [1] J. Xu, G. Guo, W. Li, H. Jiang, "A Packet scheduling Strategy for Heterogeneous Traffic of Internet of Things", Communications and Networks 2012 2nd International Conference, pp.1557-1560, 2012