

# 분산 게이트웨이 환경에서의 *K*-means Clustering을 이용한 센서 데이터 평준화 기법

이태호<sup>○</sup>, 임환희<sup>\*</sup>, 이병준<sup>\*</sup>, 김경태<sup>\*</sup>, 윤희용<sup>\*\*</sup>

<sup>○</sup>성균관대학교 정보통신대학 전자전기컴퓨터공학과

<sup>\*\*</sup>성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {leetachoO, lhh423<sup>\*</sup>, byungjun<sup>\*</sup>}@skku.edu, kyungtaekim76@gmail.com<sup>\*</sup>, youn7147@skku.edu<sup>\*\*</sup>

## Sensor Data Standardization using *K*-means Clustering in Distributed-Gateway System

Tae-Ho Lee<sup>○</sup>, Hwan-Hee Lim<sup>\*</sup>, Byung-Jun Lee<sup>\*</sup>, Kyung-Tae Kim<sup>\*</sup>, Hee-Yong Youn<sup>\*\*</sup>

<sup>○</sup>Dept. of Electrical and Computer Engineering, Sungkyunkwan University

<sup>\*\*</sup>Dept. of Software, Sungkyunkwan University

### ● 요약 ●

본 논문에서는 IIoT(Industrial IoT) 환경에서 사용되는 각 종 센서의 특성을 고려하여 *K*-means clustering을 이용해 측정 주기에 따른 군집화를 통해 평준화함으로써 센서에서 게이트웨이로의 데이터 전송 시 일어날 수 있는 1:1 독점 통신 현상 및 작업부하를 해결 할 수 있는 기법을 제안한다. 본 논문에서는 해당 기법의 효율을 보다 극대화할 수 있는 분산 게이트웨이 환경에서 실험을 진행하였으며, 해당 실험의 결과에 따르면 분산 게이트웨이 시스템에서 사용되는 게이트웨이들의 작업부하가 현저히 낮아졌고 각 종 센서들이 할당되는 빈도수가 일정하게 나타남으로써 신뢰성과 정확성을 확보에 보다 우수함을 보인다.

**키워드:** 평준화(Standardization), 군집화(Clustering), IIoT(Industrial IoT)

## I. Introduction

미국의 시장조사기관이자 IT분야 연구 및 자문을 담당하는 가트너(Gartner)에 따르면, 2017년 사물인터넷(Internet of Things, IoT) 기기가 84억대에 이르렀다고 한다. 또한 2020년까지 사물인터넷 기기가 약 204억대에 이를 것 이라고 발표했다[1]. 이와 같이 사물인터넷을 기반으로 우리 일상생활에 많은 변화가 일어나고 있다.

특히 생산 공장을 운영하는 기업과 같이 공장 자동화를 위해 사물인터넷을 이용한 IIoT(Industrial IoT)환경을 구축해 Smart factory, Smart plant 등이 점차 늘어나고 있다. IIoT 환경에서는 공장 자동화를 위해 게이트웨이를 중심으로 하여 하위의 각 종 센서를 연결하여 이용하며, 센서는 측정된 데이터를 상위의 게이트웨이로 전송하게 된다. 데이터를 전송받은 게이트웨이는 관리자들이 자동화 공정을 관리할 수 있도록 의미 있는 데이터를 선별하여 제공한다. 여기서 관리자에게 제공되는 데이터는 신뢰성과 정확성, 신속성 등의 조건이 요구된다.

본 논문에서는 IIoT 환경에서 측정 주기와 전송 속도, 측정 데이터의 중요도 등이 상이한 각 종 센서들의 데이터를 Machine learning 기법중 하나인 *K*-means clustering을 이용하여 게이트웨이로의 전송 과정에서 평준화(Standardization)시킴으로써 전송 지연 및 데이터 처리 비효율성을 줄이는 시스템을 제안한다.

## II. Preliminaries

현재 *K*-means clustering은 데이터 수집 과정에서 발생할 수 있는 노이즈 및 유사 데이터와의 군집화를 통해 신뢰성과 정확성을 증대시키는 방법으로 많이 사용되고 있다[2]. 하지만 특성이 다른 센서 데이터를 기반으로 각 데이터마다의 특징을 반영하여 개별적으로 적용한 사례는 전무하다.

## III. The Proposed Scheme

본 논문에서는 각 종 센서 데이터에 대해 *K*-means clustering을 적용하였을 때 보다 높은 효율을 나타낼 수 있는 분산 게이트웨이 시스템을 바탕으로 실험을 진행하여 해당 시스템에 대한 우수성을 표현한다.

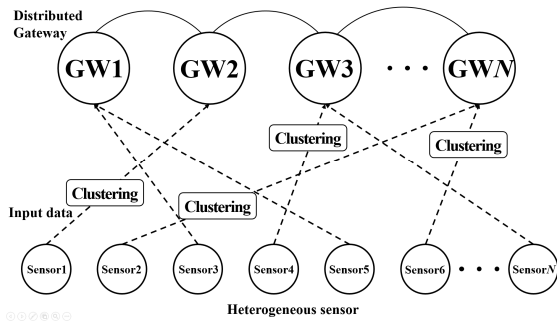


Fig. 1. Using K-means clustering in distributed-gateway

각 종 센서들 중 고속도 센서와 같이 ms단위의 높은 속도로 측정되는 센서의 경우 해당 센서에서 측정되는 모든 데이터를 지속적으로 게이트웨이로 전송하게 된다면 분산 게이트웨이 시스템에서 해당 센서와 전송받는 게이트웨이 간의 연결이 변경되지 않고 1:1 통신을 유지하게 되며, 게이트웨이는 방대한 데이터를 지속적으로 전송받음으로서 작업부하가 늘어나게 된다. 때문에 해당 센서의 측정 주기를 낮춤으로써 환산하여 군집화해 줌으로써 다른 종류의 센서 측정 주기와 평균화시켜 준다. 이렇게 된다면 측정 주기가 매우 짧은 센서가 하나의 게이트웨이를 독점하는 현상을 방지할 수 있고, 분산게이트웨이 시스템에서 효율적으로 각 종 센서들을 게이트웨이에 할당함으로써 데이터 처리의 지연시간 감소 및 신뢰성, 정확성을 증대시킬 수 있다. 또한 측정 주기가 비교적 긴 센서와 같은 경우에도 같은 특성을 가지는 센서별 군집화를 통해 효율적으로 게이트웨이에 할당 할 수 있다.

#### IV. Conclusions

본 논문에서는 사물인터넷(Internet of Things, IoT)분야 중 공장 자동화 시스템을 위하여 실시간 처리가 가능하며 신뢰성 및 정확성 높은 데이터를 요하는 IIoT(Industrial IoT) 환경에서 각 종 센서들이 상이한 특성을 가지고 있다는 것을 인지하고 이를 K-means clustering 을 이용한 방법으로 평균화해 줌으로써 데이터 처리를 진행하는 게이트웨이의 작업부하를 낮춰준다. 때문에 관리자는 보다 신뢰성과 정확성이 증대된 데이터를 제공 받을 수 있다.

향후 연구에서는 K-means clustering을 비롯한 다른 군집화 알고리즘과의 비교를 통해 조금 더 우수한 방법을 모색하는 연구를 진행하고 자 한다.

#### 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] Gartner [Online], Available: <https://www.gartner.com/newroom/id/3598917>, Feb. 2017.
- [2] S. Bandyopadhyay, C. Giannella, U. Maulik, H. Kargupta, K. Liu and S. Datta, "Clustering distributed data streams in peer-to-peer environments", Proceedings of Information Sciences, vol.176, issue.14, pp.1952-1985, 2006.