

## WSN 환경에서의 데이터 마이닝 기법 연구

김동현<sup>○</sup>, 김민우<sup>\*</sup>, 이병준<sup>\*</sup>, 김경태<sup>\*</sup>, 윤희용<sup>\*\*</sup>

<sup>○</sup>성균관대학교 정보통신대학 전자전기컴퓨터공학과

<sup>\*\*</sup>성균관대학교 소프트웨어대학 소프트웨어학과

e-mail: {kdh7263<sup>○</sup>, kimmw95<sup>\*</sup>, byungjun<sup>\*</sup>}@skku.edu, kyungtaekim76@gmail.com<sup>\*</sup>, youn7147@skku.edu<sup>\*\*</sup>

## A Study on Data Mining Techniques in WSN Environment

Dong-Hyun Kim<sup>○</sup>, Min-Woo Kim<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 ●

최근 인터넷 기술의 발달로 다수의 센서 노드로 구성된 Wireless Sensor Network (WSN) 환경이 증가하고 있으며, 이에 따라 무선으로 연결된 수많은 노드에 의해 생성되는 데이터의 양이 방대해지고 있지만, 데이터의 특성 및 패턴이 불규칙하여 기존 정적 분류 기법으로는 한계가 있다. 따라서 본 논문에서는 이러한 WSN 환경에서 생성되는 방대한 양의 데이터를 효율적으로 처리하기 위해 기계학습을 이용한 데이터 마이닝(Data mining) 기법에 대해 서술한다. 데이터 마이닝이란 데이터의 패턴 및 데이터 간의 관계를 이용하여 의사결정에 필요한 정보를 추출하는 것으로 다양한 기계 학습 알고리즘이 존재한다.

**키워드:** 데이터 마이닝(data mining), 무선 센서 네트워크(wireless sensor network), 분류(classification)

### I. Introduction

최근 인터넷 기술의 발달로 다수의 센서 노드 및 게이트웨이로 구성된 WSN 환경이 증가함에 따라 생성되는 데이터의 양이 방대하여 데이터를 효율적으로 처리하기 위한 기법이 요구된다[2]. 하지만 WSN 환경에서는 데이터의 특성 및 패턴이 불규칙하여 기존 비기계학습 알고리즘 만으로는 데이터의 특성을 파악하기에 한계가 있다. 데이터 마이닝은 데이터의 패턴 및 데이터 간의 관계를 이용하여 의사결정에 필요한 정보를 효율적으로 처리하는 것으로[1], 본 논문에서는 기계학습 중에서도 지도학습 (Supervised Learning)을 이용한 데이터 마이닝 기법에 대해 서술한다.

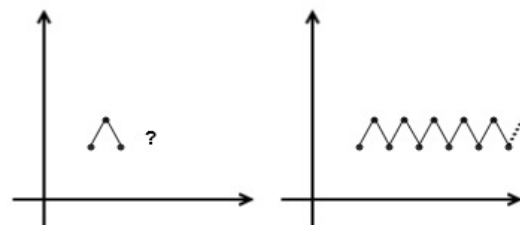


Fig. 1. Example of Data Pattern Prediction

### II. Preliminaries

#### 1. Related works

데이터 마이닝 기법에서는 데이터의 양이 증가 할수록 숨겨진 데이터의 패턴을 파악할 가능성도 증가한다.

#### 1.1 k-최근접 이웃 알고리즘(k-NN Algorithm)

가장 간단한 기계학습 알고리즘으로 데이터의 분류에  $k$ 개의 인접한 데이터를 사용한다. 분류에 사용되는 데이터와 이웃 데이터 사이의 거리를  $d$ 라고 할 때 각각의 이웃 데이터에게  $1/d$ 의 가중치를 주고, 최종 데이터의 분류 값은 과반수에 의해 결정된다.

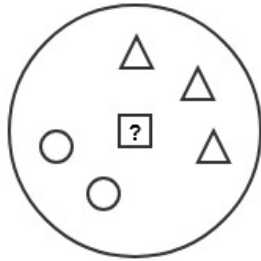


Fig. 2. Example of  $k$ -NN Algorithm ( $k = 5$ )

위 그림 1의 네모(?)는 과반수(동그래미=2/5, 세모=3/5)에 의해 세모 클래스로 분류된다.

### 1.2 결정 트리 학습법 (Decision Tree Learning)

결정 트리 학습법은 학습에 필요한 데이터에 대한 예측 모델로써 결정 트리를 사용하며, 데이터를 분석하여 패턴을 예측 가능한 규칙의 조합으로 나타낸다. 이러한 의사 결정트리에 사용되는 데이터에 대한 구분은 정보 엔트로피 (Information Entropy) 값을 기반으로 한다.

$$I(E) = - \sum_{e \in E} P(e) \log P(e)$$

위 수식에서  $E$ 는 모든 상태 집합,  $P(e)$ 는 해당 상태가 일어날 확률이다.

### 1.3 나이브 베이즈 분류 (Naive Bayes Classifier)

나이브 베이즈는 조건부 확률 이론을 기반으로 하는 모델로, 분류에 사용되는 데이터들은  $n$ 개의 특성을 나타내는 벡터  $x = \{x_1, x_2, \dots, x_n\}$ 로 정의하고, 각각을 독립된 요인으로 두고 조건부 확률을 이용해 해당 결과가 나올 확률을 계산하는 알고리즘이다.

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

위 수식에서  $P(Y|X)$ 는  $X$ 가 있어났을 때  $Y$ 가 일어날 확률이다.

## IV. Conclusions

본 연구에서는 다수의 센서노드로 구성된 WSN 환경에서 생성되는 방대한 양의 데이터를 효율적으로 처리하기 위한 방법으로, 데이터 마이닝 기법에 사용되는 기본적인 알고리즘에 대해 서술하였다. 향후 연구 계획으로는 주성분 분석 (Principal Component Analysis)과, 분류에 사용되는 기계 학습 알고리즘을 결합한 데이터 마이닝 기법에 대해 연구할 계획이다.

## ACKNOWLEDGMENT

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] N. Jain, V. Srivastava, "Data Mining Techniques: A Survey Paper," International Journal of Research in Engineering and Technology, Vol. 2, pp.116-119, 2013
- [2] X. Wu et al. "Data Mining with Big Data", IEEE Transactions on Knowledge and Data Engineering, Vol. 26, pp.97-107, 2014