

## 빅데이터 기반의 IoT 이상 장애 탐지 시스템 설계

나성일 · 김형중\*

고려대학교 정보보호대학원 빅데이터 응용 및 보안학과

## Design of Anomaly Detection System Based on Big Data in Internet of Things

Sung Il Na · Hyoung Joong Kim\*

Graduate School of Information Security, Korea University, Seoul 02841, Korea

### [요 약]

사물인터넷(IoT) 서비스는 스마트 환경이 발전하면서 다양한 데이터를 생산하고 있다. 이 데이터는 사용자의 상황을 판단하는 중요한 데이터로 사용된다. 그렇기 때문에 센서의 이상 상태를 실시간으로 모니터링하고 이상 데이터를 탐지하는 것이 중요하다. 하지만 데이터 구조와 프로토콜이 다양하기 때문에 표준화된 데이터 구조로 변환하는 과정이 필요하다. 그럼으로써 데이터의 품질을 보장하고 정확한 분석을 통해 서비스의 품질까지 좋아지는 효과를 기대할 수 있다.

본 논문은 수집된 센서의 이상탐지를 위해 빅데이터 기반의 이상탐지 시스템을 제안한다. 제안한 시스템은 이상탐지를 위해 데이터 표준화 전처리와 시계열 기반의 이상탐지가 우수한 SVM(Support Vector Machine) 모델을 적용하였다. 실험에서는 전처리와 전처리되지 않은 데이터를 각각 학습시키고 비교하였다. 그 결과, 전처리된 데이터는 이상 장애를 정확히 탐지하고 예측하였다.

### [Abstract]

Internet of Things (IoT) is producing various data as the smart environment comes. The IoT data collection is used as important data to judge systems's status. Therefore, it is important to monitor the anomaly state of the sensor in real-time and to detect anomaly data. However, it is necessary to convert the IoT data into a normalized data structure for anomaly detection because of the variety of data structures and protocols. Thus, we can expect a good quality effect such as accurate analysis data quality and service quality.

In this paper, we propose an anomaly detection system based on big data from collected sensor data. The proposed system is applied to ensure anomaly detection and keep data quality. In addition, we applied the machine learning model of support vector machine using anomaly detection based on time-series data. As a result, machine learning using preprocessed data was able to accurately detect and predict anomaly.

**색인어** : 이상탐지, 장애예측, 사물인터넷, 스마트서비스, 빅데이터분석

**Key word** : Anomaly Detection, Outlier Prediction, Internet of Things, Smart Service, Big Data Analytics

<http://dx.doi.org/10.9728/dcs.2018.19.2.377>



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

**Received** 22 December 2017; **Revised** 12 February 2018

**Accepted** 26 February 2018

**\*Corresponding Author; Hyoung Joong Kim**

**Tel:** +82-2-3290-4895

**E-mail:** khj-@korea.ac.kr

## 1. 서론

최근 사물인터넷(IoT : Internet of Things) 서비스의 발전은 다양한 상황을 판단할 수 있는 데이터를 생산하고 스마트 장치를 이용하여 하나의 연결된 사회로 구성하고 있다. 사물인터넷은 다양한 산업분야에서 센서와 인식 장치를 이용하여 자동화된 서비스로 편리함을 제공하는 핵심기술로 발전하고 있다. 이 과정에서 다양하게 변화되는 개인의 상황 정보는 실시간으로 수집되어 분석한다. 주변의 장치와 연결된 상태에서 사용자를 위해 인식 장치간의 자동설정, 자율제어, 최적의 운영 상태를 유지하는 시스템 모니터링까지 복잡한 서비스 구조로 발전하고 있다. 이러한 변화는 빅데이터 분석 환경을 기본으로 하고 있으며, 사용자별로 수집되는 센서 데이터를 구분하고 관리하는 역할을 지원하고 있다. 그리고 빅데이터 분석 환경은 인공지능 분석 기술과 결합되어 사물인터넷의 변화에 따라 양적으로 증가하는 데이터에서 의미가 있는 데이터를 생산하고 서비스를 구성하는 체계적인 환경으로 발전하고 있다.

특히 사물인터넷에서 생산되는 데이터는 센서마다 다양한 포맷과 프로토콜을 사용하고 있기 때문에 표준화된 수집 환경에서 데이터를 확보하고 있다. 그리고 인공지능 분석 기술과 기계학습 모델을 이용하여 맞춤형 개인화 서비스로 확장되고 있다. 이 과정에서 센서 데이터는 실시간으로 수집하고 데이터 표준화 특징을 이해하고 정확한 정규화 모델을 이용하여 데이터를 변환하면서 모든 데이터를 하나로 통합하여 분석하는 것이 중요하다. 만약 수집된 센서 데이터가 이상이 발생할 경우에는 정규화하고 통합하는 과정에서 데이터 오류로 인해 실패하거나 잘못된 데이터를 구성하게 된다. 이럴 경우에 사용자의 상황이나 스마트 장치의 이상 증상은 정확하게 판단되지 않기 때문에 사용자 서비스 장애가 발생하거나 실패하는 문제가 발생된다. 센서의 고장은 이동 중인 사용자의 상황을 판단하는 데이터의 유실로 인해 지역간의 서비스 연결성을 보장하지 못하는 문제와 그로 인한 맞춤형 개인화 서비스의 실패로 확장될 수 있다. 그리고 무인자동차나 자율주행과 같은 스마트 서비스 시스템의 경우에는 이상 장애로 인해 장치의 제어가 되지 않아 인명 사고로 이어질 수 있다. 그렇기 때문에 빅데이터 기반의 사물인터넷 서비스는 수집되는 센서의 데이터에 대한 이상 유무를 실시간으로 판단하고 모니터링하는 것이 중요하다. 이러한 상태에서 이상 상태를 예측함으로써 장애나 사고의 예방의 중요성도 높아지고 있다.

본 논문은 빅데이터 분석 환경에서 수집되는 센서의 이상 장애를 탐지하기 위한 빅데이터 기반의 이상탐지 시스템을 제안한다. 제안한 시스템은 센서 데이터의 이상 장애를 정확하게 탐지하기 위해 센서의 특징을 고려한 데이터 전처리 모델을 적용하였다. 그리고 수집된 센서 데이터가 시계열 구조를 가지고 있기 때문에 시계열 이상 탐지에 우수한 SVM(Support Vector Machine) 알고리즘을 개선하여 적용하였다. 제안한 모델을 평가하기 위한 실험에서는 도로의 경사면 상태를 감시하는 태양

광 센서 데이터를 이용하여 평가하였다.

## II. 빅데이터 기반의 IoT 이상 장애 탐지 시스템

### 2-1 시스템 개요

제안하는 시스템 모델은 빅데이터 기반의 센서 데이터 수집과 관리를 수행하는 사물인터넷 환경을 기반으로 한다. 최근의 사물인터넷은 다양한 센서에서 발생하는 데이터와 센서의 상태를 기록한 로그 데이터를 함께 전송하여 센서의 이상 상태를 모니터링하고 있다. 하지만 사물인터넷 환경에서 사용되는 센서는 다양한 데이터 포맷과 프로토콜로 인한 데이터의 이질성 때문에 데이터 통합이 어려운 상태로 개별 서비스를 개발하고 있다. 중복된 센서의 배치는 시설의 중복 투자와 함께 보안의 취약성을 가지고 있다. 이를 위해 본 논문은 사용자와 직접적으로 연관성이 높은 사물인터넷 환경에서 센서의 특징에 대한 사전 분석과 이상탐지를 위해 기계학습 모델을 적용하였다. 센서 데이터는 시계열 특징을 가지며, 시간차에 따라 수집되고 관리된다. 그림 1은 제안한 시스템의 이상 장애 탐지 시스템의 모델을 표현한 것이다.

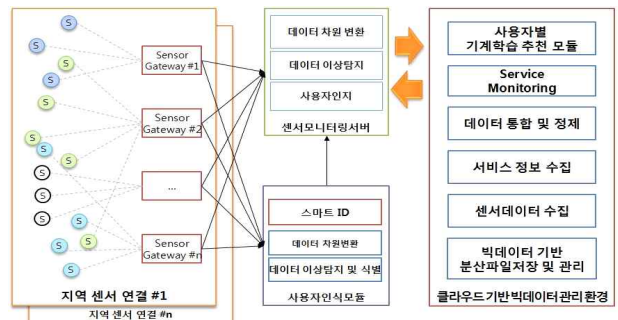


그림 1. IoT 이상 장애 탐지 시스템 모델  
Fig. 1. A model of IoT anomaly detection system

제안하는 시스템은 지역별 설치되어 있는 이기종 센서들을 하나로 통합하고 관리하면서 사용자에게 제공되는 센서 데이터와 상태를 모니터링하면서 실시간으로 이상 장애를 탐지하는 모델이다. 센서는 지역별 연결성을 보장하기 위해 프로토콜에 대한 범용성을 확보하고, 스마트 디바이스의 통신 모델을 기반으로 데이터 송수신한다. 통신보안은 암호화 키를 적용하고, 해당 키의 복호화는 클라우드 인프라 위에 개인용 실시간 빅데이터 분석 환경을 기준으로 설계하였다. 그래서 스마트 디바이스가 대량의 데이터를 분석하고 암호호화를 처리하기 위한 기능을 설계에서 보완하였다. 각각의 지역에는 이기종 센서를 고려하여 센서 게이트웨이를 이용하여 사용자가 수신하는 정책을 반영한다. 그림으로써 데이터 품질이 보장되는 서비스 환경 제공을 목표로 하였다. 사용자 인식 모듈은 센서 게이트웨이로부터 수신된 데이터를 개인데이터로 명시하고, 데이터의 차

원 변환과 정규화 처리를 담당한다. 센서 모니터링 서버는 지역의 모든 서버를 통합 모니터링하면서 수신된 데이터로부터 차원변환과 정규화를 사용자인식모듈과 동일하게 처리한다. 이는 지역 내 특정센서가 고장이나 이상으로 데이터 생산이 중단되는 경우 서비스의 연속성을 보장하기 위한 데이터를 보완하여 사용자에게 전달하기 위함이다. 그리고 이상탐지와 식별을 위해 기계학습 모델을 적용하여 실시간으로 이상상태를 탐지하도록 설계하였다.

제한하는 시스템에서 데이터 통합은 동일 시간대에 수신된 센서 데이터를 통합하기 위한 센서 게이트를 기준으로 통합하도록 설계하였다. 각각의 센서는 전원이 0이 될 때까지 센서 게이트웨이로 자신의 상태정보와 생산하는 데이터를 전송한다. 그리고 센서 게이트웨이는 수신된 데이터를 사용자와 모니터링 서버로 전달하기 위한 통합 처리 과정을 모델링하였다. 그림 2는 제한한 모델에서 사물인터넷에서 발생된 데이터를 통합하고 정제하는 모델이다.

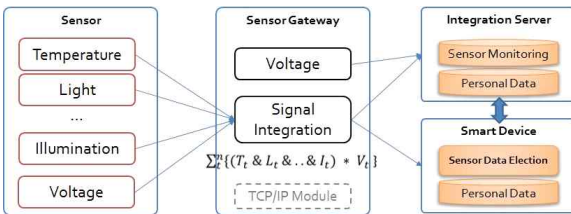


그림 2. 사물인터넷 데이터 통합 및 정제 모델  
Fig. 2. A model of data cleansing and integration in IoT

제한한 시스템의 데이터 통합 및 정제 모델에서 센서는 서버를 위해 생산되는 데이터와 센서 자신의 전압을 함께 센서 게이트웨이로 전달한다. 그림 2에서와 같이 센서 게이트웨이는 수신된 데이터를 구분자(delimiter)로 구분하여 센서 별 데이터를 통합하여 사용자와 모니터링 서버로 전달한다. 구분자는 CSV 포맷 기준으로 콤마와 탭으로 센서와 데이터를 구분하도록 설계하였다. 예를 들면 하나의 단위 지역에 온도 센서가 2개와 빛 센서 1개가 존재한다면, “{센서 ID}, {생산데이터} 탭 {센서전원데이터}, {센서 ID}, {생산 데이터} 탭 {센서전원데이터}, ...”로 표현할 수 있다. 센서의 생산 데이터가 다양할 경우에는 동일한 구조로 센서전원 데이터를 측정된 동일 값으로 표현한다. 이렇게 통합된 데이터를 기준으로 통합 서버와 개인의 스마트 장치에서 이상데이터를 찾아내고 식별하기 위해 센서의 특징에 맞는 데이터 정제 모델을 적용하였다. 예를 들면 태양광 충전방식의 센서는 센서의 충전 상태를 계속적으로 확인하기 위해 생산된 데이터와 연관성을 가지거나 연관정보를 별도로 생산한다. 그렇기 때문에 센서의 특징을 이해하지 못한 상태에서 전처리되지 않는 데이터로 이상 장애를 탐지하는 것은 잘못된 데이터 학습이나 학습의 정확도를 떨어뜨리는 단점을 가질 수 있다. 그렇기 때문에 제한한 시스템은 센서의 특징을 판별하기 위해 센서 게이트웨이에 센서의 메타정보와 센서 고유번호를 통합하여 데이터를 구성하도록 설계하였다.

### 2-2 이상탐지 기계학습 모델

사물인터넷의 센서 데이터 고장 유형은 매우 단순하다. 전원에 공급 상태에 따라 센서 데이터 생산을 결정할 수 있기 때문이다. 만약 특정 전압의 범위를 유지하지 못한다면 센서는 데이터를 생산하지 못하고 대기신호만을 센서 게이트웨이에 전달할 것이다. 그리고, 그 신호의 의미를 정확하게 찾아내는 학습이 필요하다. 각각의 센서의 데이터는 시간에 변화되고 있는 상태정보가 발생하지 않을 경우는 단순 고장이나 전원이상으로 판단할 수 있다. 그렇기 때문에 계산 복잡도가 높은 딥러닝 기계학습 모델보다는 단순신경망을 이용한 기계학습 모델이 적합하다. 사용자의 스마트 디바이스는 딥러닝을 수행하기에는 자원이 매우 부족하기 때문에 단순한 기계학습 모델을 적용해야 한다.

본 논문은 이상 데이터 식별 및 장애 탐지를 위해 단순신경망 구조를 가진 SVM을 개선하여 적용하였다. SVM은 선형과 비선형 데이터를 학습시킬 수 있는 기계학습 모델이며, 시계열 데이터에서 이상데이터 식별과 장애를 탐지하는 성능이 우수하다. 센서 데이터의 이상이나 장애가 시간에 따라 규칙적으로 발생되지 않기 때문에 SVM의 선형 모델보다는 비선형 모델이 적합하다. 본 논문은 센서 데이터에서 발생하는 이상이나 장애에 대한 예측을 추정하기 위해 비선형 구조의 데이터를 학습하는 SVM 모델을 적용하였다. SVM 알고리즘은 원 훈련 데이터를 비선형 매핑으로 고차원 공간 데이터로 변환한다. 차원변환을 2차원에서 3차원으로 높게 변환하기 위해서는 식 (1)을 이용하였으며, 차원이 높게 변형되었을 때에는 선형 분리 초평면을 가지게 하였다. 하지만 고차원으로 비선형 매핑을 지속적으로 하면 MMH(maximum margin hyperplane)를 구하기 위해 연산 속도와 성능이 떨어지는 문제점을 커널 트릭으로 개선하였다.

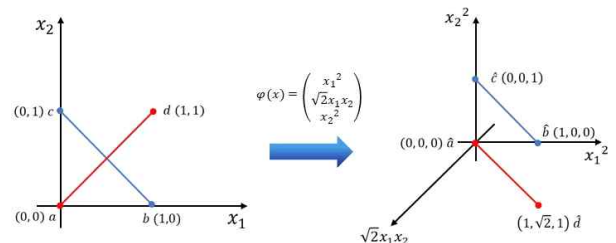


그림 3. 비선형 SVM의 원데이터 차원변형 모델  
Fig. 3. A data dimension transform model of nonlinear SVM

$$\varphi(x) = \begin{pmatrix} x_1^2 \\ \sqrt{2}x_1x_2 \\ x_2^2 \end{pmatrix} \quad (1)$$

비선형 SVM의 성능을 개선하기 위해 일반적으로 커널 트릭(kernel trick)을 사용하여 고차원으로 변환한 후 벡터 내적을 계

산하도록 하였다. 커널 함수(kernel function)와 벡터 내장은 식 (2)와 같다. 커널 기법은 선형 SVM으로 분리가 불가능한 훈련에 적용가능하며, 입력 데이터를 어떤 특성에 조직화하고 이를 고차원의 특성 공간(feature space)에 매핑하는 기법이다. 커널 함수는 입력 데이터를 특성 공간에 매핑하기 때문에 두 개의 데이터를 하나의 특성으로 조직화할 수 있다.

$$k(x_i, x_j) = \varphi^T(x_i)\varphi(x_j) = \sum_{j=1}^{\infty} \varphi_j(x_i)\varphi_j(x_j), i=1,2,\dots,N_s \quad (2)$$

일반적으로 기계학습 알고리즘에서  $\varphi^T(x_i)\varphi(x_i)$  가 존재하는 모든 곳에는  $k(x_i, x_i)$ 로 대체할 수 있다. 본 연구에서는 Gaussian radial basis function과 sigmoid kernel function을 결합한 커널 트릭을 사용하였다. 각각의 커널 트릭 함수는 힐버트 공간 커널 함수 평가 시 가우스 커널 함수를 적용하였으며, 유클리드 공간 커널 함수 평가 시 로직스틱 커널 함수를 적용하였다. 그럼으로써 비선형 SVM 기계학습에서 발생할 수 있는 차원 연산속도와 성능 문제를 해결하였다. 본 논문은 식 (3)(4)와 같이 전치 행렬을 이용하여 차원을 매핑하는 모델로 개선하였다.

힐버트 공간 커널 함수 (3)  
 $\rightarrow k(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2), \gamma > 0$

유클리드 공간 커널 함수 (4)  
 $\rightarrow k(x_i, x_j) = \tanh(\gamma x_i^T x_j + c_0)$

### III. 실험 및 평가

#### 3-1 데이터 구조

실험을 위해 사용된 데이터는 도로의 경사면에 설치되어 도로의 상태를 감시하는 태양광 충전방식의 센서 데이터를 이용하였다. 실험에 사용된 데이터는 도로경사면을 감시하는 기업에서 수집한 데이터이다. 2017년에 수집된 데이터 중에서 데이터 품질이 양호한 2017년 7월과 8월 사이의 데이터를 사용하였다. 데이터 수집 단위는 시간(H) 단위이며, 총 건수는 41,970건이다.

gatewayIndex	gatewayVoltage	sensorIndex	sensorVoltage	sensorIllumination	sensorTemperature	createdDate
100	0		3485	52	132	2017.7.16 17:13
100	0		3518	505	108	2017.7.19 10:24
100	0		3522	289	108	2017.7.19 10:34
100	3784		3549	261	108	2017.7.19 11:34
100	3786		3547	279	108	2017.7.19 12:35
100	3864		3568	357	108	2017.7.19 13:35
100	3860		3576	272	108	2017.7.19 14:36
100	3786		3575	178	108	2017.7.19 15:36
100	3788		3600	182	108	2017.7.19 16:37
100	3779		3609	149	108	2017.7.19 17:37
100	3775		3608	117	108	2017.7.19 18:38
100	3770		3604	67	108	2017.7.19 19:38

그림 4. 평가를 위해 사용된 데이터 구조  
 Fig. 4. Data structure used for evaluation

해당 센서는 하나의 센서 게이트웨이에 연결되어 측정된 데이터를 수집하고, 센서 게이트웨이는 5개의 센서가 연결되어 센서 데이터를 전송하고 있다. 데이터의 구조는 그림 4와 같은 구조로 센서를 연결하는 게이트웨이 인덱스, 게이트웨이 전압 상태 값, 센서 인덱스, 온도, 조도, 수집 시간 등으로 구성된다. 센서 게이트웨이는 연결성을 관리하고 있기 때문에 전압이 없는 경우에는 이상 장애로 판단할 수 있다. 하지만 도로의 경사면 데이터를 전송하는 센서는 센서의 전압이 특정값 이상의 범위를 유지해야 데이터를 전송할 수 있다. 만약, 전압이 해당값 이하되면 도로의 상태값을 전송하지 못하기 때문에 도로 경사면의 붕괴나 낙석과 같은 사고로 이어질 수 있다. 그래서 센서의 고장이나 이상 증상을 전압으로 감시하고 있는 실정이다.

센서의 이상 탐지를 위한 실험은 제공된 태양광 센서의 이상 증상을 탐지하기 위해 데이터간의 연관성을 확인하였다. 충전에 영향 변수의 특징을 찾기 위해 에너지 변환 효율을 이용하여 전처리 정규화를 진행하였다. 제공된 데이터의 수집 시간이 약한 시간 간격이기 때문에 시간에 대한 연속성을 부여하기 위해 수집 시간을 순차적 인덱스로 변환하여 처리하였다. 태양광 충전 방식의 에너지 변환 효율은 식 (5)와 같이 적용되고 있다.

$$\text{변환효율}(\%) = (\text{생산된전력량}) / (\text{태양빛에너지}) \times 100 \quad (5)$$

태양광 충전방식은 식(5)에 따라 태양광 충전 효율이 달라지기 때문에 태양빛 에너지를 흡수하는 온도값과 조도 값이 중요하다. 특히 실험에서 사용된 데이터는 도로의 경사면에서 설치되어 자연적 요인(낙하물, 풀, 나무, 먼지, 자동차 진동등, 가로등 등)에 의하여 설치 각이 변경되거나 태양빛을 정확하게 흡수하지 못하는 상태가 많이 발생한다. 그렇기 때문에 설치된 태양광 반사판에 대한 상태값을 고려한 정규화 모델이 수립하고 적용하였다.

#### 3-2 실험 결과 및 평가

수정된 이상 장애 탐지 기계학습 모델은 태양광 전지의 에너지 효율을 기준으로 정규화한 데이터를 이용하여 전압(x축)과 조도(illumination, y축), 온도(Temperature, y축)의 연관성을 각각 평가하였으며, 그 결과는 그림 5와 같이 도출되었다.

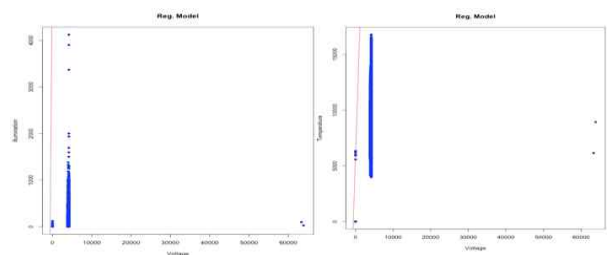


그림 5. 충전 전압과 온도, 조도의 상관성 분석 결과  
 Fig. 5. The result of correlation analysis between voltage

and temperature, voltage and illuminance

그림 5는 태양광 전지의 이상 상태를 예측하기 위해 충전율에 영향을 주는 조도와 온도의 연관성 분석을 하였다. 이는 전압의 이상상태가 전압 충전에 있기 때문에 전압의 연관성을 기준으로 해당 변수의 변화를 모니터링하고 이상상태를 탐지하고 예측하는 것이 중요하다. 실험에서는 이를 위해서 충전 전압과 온도, 충전 전압과 조도, 충전 전압과 온도 및 조도를 결합한 값을 각각의 연관 관계를 측정하였다. 그 결과, 충전 전압은 조도와 온도에 대해 개별적인 연관성이 떨어지는 것을 확인하였다. 전압 추세선은 온도와 조도를 기준으로 수평하게 구성되었으며, 일부의 태양광 충전 성능은 계절적 특성과 설치된 장소에 따라 온도와 조도를 결합한 형태에서 좋은 성능을 가지고 있음을 확인하였다.

실험을 위해서 태양광 변환 효율을 적용하여 정제된 데이터를 전체적으로 표현하면 그림 6과 같다. x축은 수집된 시간을 인덱스로 표현한 값이고, y축은 온도와 전압의 관계성을 표현한 것이다.

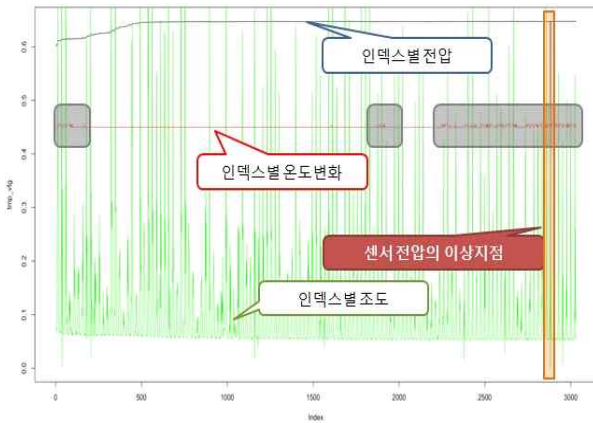


그림 6. 원시 데이터에 대한 단일 정제 처리 결과 시각화  
 Fig. 6. Visualization of simple cleansing with raw data analytics

시간당 태양광 전지의 전압 상태 값은 충전이 지속되는 동안 일정한 수준의 값을 유지하지만 충전이 미흡할 경우에는 급격히 떨어지는 구간 특징을 가지고 있다. 수집 시간의 차이는 있지만 이를 고려하더라도 태양광 에너지의 변환효율은 조도와 온도에 따라 충전 상태가 결정됨을 확인하였다. 온도는 전압의 변화보다 적은 크기로 변화를 가지며 정규화를 하더라도 3곳의 패턴 연관성이 적지만 후반부의 길어진 상태 패턴의 경우에 조도와 함께 전압의 충전 효율을 떨어뜨리는 연관성이 발견되었다. 조도는 일반적으로 태양빛이 좋은 상태와 그렇지 않은 상태에 따라 전압 충전 상태가 많은 차이를 가진다. 하지만 도로 상에 설치된 태양광 센서는 조도에 영향을 주는 차량의 불빛과 가로등, 풀잎, 나무 줄기, 먼지 등에 따라 충전 상태가 달라짐을 알 수 있다. 그렇기 때문에 태양광 에너지 변환 효율을 고려한 태양광 센서의 충전 모델 정규화는 전압의 이상

지점과 동일한 구간에서 온도와 조도가 결합된 상태 값의 변화가 가장 큰 영역을 이상 장애로 탐지할 수 있었다. 이를 통해 태양광 센서의 이상 장애를 탐지하기 위해 수집한 초기 가설은 온도와 조도를 결합한 에너지 변환 효율을 함께 적용하여 이상 장애 탐지 기계학습 모델의 학습성능을 보완하였다.

해당 가설에 따라 데이터를 정제하고 정규화한 데이터를 이상탐지 기계학습 모델에 적용하였다. 사용된 데이터는 정제된 수집 데이터의 70%를 훈련데이터로 사용하였으며, 나머지 30% 데이터를 검증용으로 사용하였다. 기계학습에 사용된 하이퍼파라미터는 R의 SVM 패키지에서 제공하는 기본 설정 값을 적용하였으며, R 모듈에서 초기 학습된 결과를 바탕으로 추천된 하이퍼파라미터 추천 값(Cost = 0.5, gamma = 1, epsilon = 0.6)을 반영하여 재학습을 진행하였다. 그 결과, Support Vector의 개수가 1129개를 형성하였으며, 이 중에 이상장애에 대한 수렴이 잘 훈련되고 있음을 확인하였다. 그리고 검증용 데이터를 이용하여 학습 모델에 적용하고 이상 데이터가 이상 상태를 감지하는 범위에 포함되는지를 검증하였다.

그림 7은 태양광 센서의 이상 장애 탐지 기계학습을 검증한 결과이다. x축은 시간단위당 충전 전압 검증 데이터이고 y축은 온도와 조도를 결합한 데이터 검증 데이터이다.

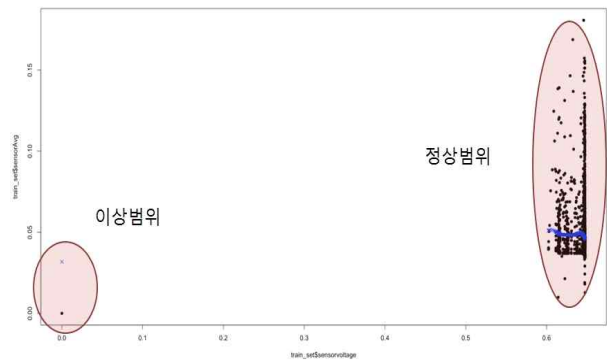


그림 7. 통합 정제 데이터 기준 이상탐지 학습 결과  
 Fig. 7. Anomaly detection result in integrated cleansing data

그림에서와 같이, 검정색 점으로 표현된 데이터는 검증용 데이터의 전압값을 표현한 것이다. 그리고 파랑색 점은 정제된 온도와 조도의 결합 데이터로 훈련된 기계학습 모델에서 추출된 데이터 값이다. 태양광 전압의 정상 범위에 들어가는 값과 전압 이상 범위에 들어가는 값이 거의 일치하고 있음을 확인하였다.

표 3. Support Vector의 개수  
 Table 3. Number of support vector

구 분	Support Vector
조도 기준의 Support Vector 수	876
온도 기준의 Support Vector 수	813
조도, 온도의 변환 효율 기준의 Support Vector 수	1129

정제된 데이터로 학습된 결과는 정상의 범주에 포함되는 Support Vector의 수가 정확하게 탐지하고 있기 때문에 이 범주에 포함되지 않은 데이터를 이상데이터로 탐지가 가능하다.

#### IV. 결 론

본 논문은 사물인터넷 환경에서 생산되는 반정형 구조의 센서 데이터를 기준으로 이상 장애를 탐지하기 위한 빅데이터 기반의 이상 장애 탐지 시스템을 제안하였다. 제안된 시스템은 수집 데이터의 특징을 고려한 정제 모델과 이상탐지 기계학습 모델을 적용하였을 때와 정제되지 않은 원시 데이터와 이상탐지 기계학습 모델을 적용하여 각각의 학습 결과를 평가하였다. 정제되지 않은 원시 데이터를 사용한 이상탐지 기계학습은 노이즈가 많은 상태로 학습 성능이 좋지 않았으며, 이상 탐지에서도 잘못된 탐지가 많이 발생하였다. 하지만 센서의 특징이 반영된 정제 모델과 이상탐지 기계학습 모델은 정확한 이상탐지가 가능하였다. 이는 기계학습으로 사용되는 데이터의 정제가 중요하며, 센서 데이터의 특징이 반영된 정제 모델과 통합 과정이 반영되어야 함을 확인하였다.

이러한 실험결과를 통해 사물인터넷에서 사용되는 태양광 센서 데이터 이상탐지는 빅데이터 분석 환경에서 데이터를 수집할 때 센서의 특징을 고려한 정제가 반드시 필요하며, 그럼으로써 구성된 기계학습 모델에서도 좋은 성능을 발휘할 수 있을 것이다. 향후 연구는 제안한 시스템 모델을 구현하고, 구현된 시스템에서 태양광 센서의 특징을 고려한 자동화된 센서 데이터 정제 및 통합 모델을 설계하고 반영하는 것이다.

#### 참고문헌

[1] J. C. Choi, I. K. Jang, D. H. Lee, M. E. Kim, and Y. S. Soen, "Technology status and prospects for develop intelligence internet of Things," IITP Series, No.1800, 2017. [Internet]. Available : <http://www.itfind.or.kr/WZIN/jugidong/1800/file/6216979542456700532-1800.pdf>

[2] B. I. Jang and C. S. Kim, "A study on the security sechnology for the internet of things," *Journal of the Korea Security Engineering*, vol. 11, no.5 , pp. 429-438, 2014.

[3] D. H. Kim, S. Y. Yoon, and Y. P. Lee, "Security for IoT services," *Journal of the Korea Institute of Communications and Information Sciences*, vol. 30, no. 8, pp. 53-59, 2013.

[4] D. H. Shin, J. Y. Jeong, and S. H. Kang, "Trend and prospects internet of things," *Journal of the Korea Society for Internet Information*, vol. 14, no. 2, pp. 32-46, 2013.

[5] M. S. Kong, H. J. Chae, and B. H. Ryu, "Trends and prospects of internet (IoT) technology," *Journal of the Korea Society of Mechanical Engineers*, vol. 56, no. 2, pp. 32-36, 2016.

[6] N. H. Kim, D. An, and J. H. Choi, *Prognostics and Health*

*Management of Engineering Systems: An Introduction*, Springer International Publishing, 2017.

[7] B. I. Kwak, M. R. Han, A. R. Kang, and H. K. Kim, "A study on detection methodology of threat on cars from the viewpoint of IoT," *Journal of the Korea Institute of Information Security & Cryptology*, vol. 25, no. 2, pp. 411-421, 2015.

[8] B. Y. Lee, W. S. Choi, and D. H. Lee, "Privacy-preserving outlier detection in healthcare services," *Journal of the Korea Institute of Information Security & Cryptology*, vol. 25, no. 5, pp. 1187-1199, 2015.

[9] S. J. Lee and D. H. Lee, "Real time predictive analytic system design and implementation using bigdata-log," *Journal of the Korea Institute of Information Security & Cryptology*, vol. 25, no. 6, pp. 1399-1410, 2015.

[10] W. D. Cho, J. H. Seong, S. T. Choi, J. H. Kim, J. H. Ha, J. S. Baek, and J. Y. Lee, "Adaptive sensing and monitoring technologies for detecting big-data based on IoT multi-sensors," *Journal of The Korea Institute of Information Scientists and Engineers*, vol. 35, no. 7, pp. 26-34, 2017.

[11] J. H. Lim, Y. K. Park, J. M. Kwon, and J. U. Seo, "Machine learning technology in the internet of things environment," *Journal of the Korea Institute of Communications and Information Sciences*, vol. 33, no. 5, pp. 48-54, 2016.

[12] B. T. Jang, "Next-generation machine learning technologies," *Journal of The Korea Institute of Information Scientists and Engineers*, vol. 25, no. 3, pp. 96-107, 2007.

[13] T. Armerding, S. Ambati, and A. Gray "Machine Learning Guide," International Data Group(IDG), InfoWorld, 2016. [Internet]. Available : <http://www.itworld.co.kr/techlibrary/97428>

[14] C. W. Tsai, C. F. Lai, M. C. Chiang, and L. T. Yang, "Data mining for internet of things: A survey," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 1, pp. 77-97, 2014.

[15] K.T. Choi, "Neural networks optimization for multi-dimensional digital signal processing in IoT devices," *Journal of Digital Contents Society*, vol. 18, no. 6, pp. 1165-1173, 2017.



**나성일(Sung-II Na)**

2003년 : 호원대학교 환경화학공학과 학사

2016년~현재 : 고려대학교 빅데이터응용 및 보안학과 (석사과정)

2016년~현재 : 에이바이트 대표이사

※ 관심분야 : 개인정보보호(Personal Information Security), 머신러닝, AI, 빅데이터분석 등



**김형중(Hyoung Joong Kim)**

1978년 : 서울대학교 전기공학과 학사

1986년 : 서울대학교 제어계측공학과 (공학석사)

1989년 : 서울대학교 제어계측공학과 (공학박사)

1989년~2006년: 강원대학교 교수

2006년~현재: 고려대학교 정보보호대학원 교수

※ 관심분야 : 컴퓨터보안, 패턴인식, 가역정보은닉, 머신러닝, 빅데이터분석 등