

AIoT 환경에 최적화된 머신러닝 기반의 IoT 데이터 처리 기법

정윤수¹, 김용태^{2*}

¹목원대학교 정보통신융합공학부 교수, ²한남대학교 멀티미디어공학과 교수

IoT data processing techniques based on machine learning optimized for AIoT environments

Yoon-Su Jeong¹, Yong-Tae Kim^{2*}

¹Professor, Dept. of information Communication Convergence Engineering, Mokwon University

²Professor, Dept. of Multimedia Engineering, Hannam University

요약 최근 IoT와 연계된 서비스들이 다양한 환경에서 활용되면서 IoT와 인공지능 기술이 융합되고 있다. 그러나, IoT 데이터를 안정적으로 처리하는 기술들이 완벽하게 지원되고 있지 않아 이를 위한 연구가 필요한 상황이다. 본 논문에서는 IoT 데이터를 머신러닝 기반으로 임베디드 벡터를 생성한 후 IoT 데이터를 최적화 할 수 있는 처리 기법을 제안한다. 제안 기법에서는 처리 효율을 위해서 IoT 데이터의 인덱스, 수집 위치(X와 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등을 QR 기반으로 임베디드 벡터화를 수행한다. 또한, IoT 데이터를 비대칭적으로 연계하도록 IoT 데이터 수집 과정에서 로드밸런싱을 수행할 수 있도록 다양한 IoT 장치에서 생성한 데이터를 통합 관리한다. 제안 기법은 비대칭적으로 IoT 데이터를 그룹화할 수 있도록 IoT 데이터를 해쉬기반으로 서로 직교화하도록 처리한다. 또한, IoT 데이터 종류 및 특성에 따라 주기적으로 생성 및 그룹화하기 때문에 IoT 데이터 간 간섭은 최소화할 수 있다. 향후 연구에서는 IoT 서비스를 제공하는 여러 환경에서 제안 기법을 비교 평가할 계획이다.

키워드 : 사물인터넷, 머신러닝, 해쉬체인, 최적화, 데이터 처리

Abstract Recently, IoT-linked services have been used in various environments, and IoT and artificial intelligence technologies are being fused. However, since technologies that process IoT data stably are not fully supported, research is needed for this. In this paper, we propose a processing technique that can optimize IoT data after generating embedded vectors based on machine learning for IoT data. In the proposed technique, for processing efficiency, embedded vectorization is performed based on QR such as index of IoT data, collection location (binary values of X and Y axis coordinates), group index, type, and type. In addition, data generated by various IoT devices are integrated and managed so that load balancing can be performed in the IoT data collection process to asymmetrically link IoT data. The proposed technique processes IoT data to be orthogonalized based on hash so that IoT data can be asymmetrically grouped. In addition, interference between IoT data may be minimized because it is periodically generated and grouped according to IoT data types and characteristics. Future research plans to compare and evaluate proposed techniques in various environments that provide IoT services.

Key Words : IoT, Machine learning, Hashchain, Optimization, Data processing

*Corresponding Author : Yong-Tae Kim(ky7762@hnu.kr)

Received January 26, 2022

Accepted March 20, 2022

Revised March 3, 2022

Published March 28, 2022

1. 서론

최근 인공지능 기술이 발전되면서 IoT와 연계된 서비스들이 다양한 환경(제조, 금융, 법률, 물류·유통 등)에서 활용되고 있다. 특히, IoT와 인공지능을 융합한 기술들은 의료 및 산업 분야에서 활발하게 사용되고 있다[1].

IoT 데이터를 수집한 후 축적된 데이터 셋들은 인공지능 기술을 이용하여 클러스터링 분석을 통해 플랫폼에서 활용할 수 있도록 최적함으로써 데이터 유형에 따라 R, 엑셀, Tableau, Gephi 등을 사용하여 다양한 그래프 시각화(바차트, 3D 차트, 히트맵, parallel 그래프, 네트워크 그래프 등)가 가능하다[2-4].

IoT 데이터는 환경 및 목적에 따라 다양한 결과들이 만들어지기 때문에 IoT 데이터 결측값을 최소화하는 연구들이 다양하게 이루어지고 있다[5]. 특히, 최근 머신러닝 딥러닝이 대두되면서 딥러닝 관련 기술들을 중심으로 실시간으로 IoT 데이터를 검증할 수 있는 기술들이 나타나고 있다[6]. 머신러닝 딥러닝 기술들은 IoT 데이터를 사전 처리 단계에서 추출하기 때문에 IoT 데이터의 차원을 줄여 정확도와 처리 효율을 향상시킨다. 또한, IoT 선별 작업 과정에서 IoT 데이터 간 비선형 상관관계의 특징 정보를 포착하도록 주성분 분석(PCA)을 사용한다.

본 논문에서는 AIoT 환경에서 다양한 유형의 IoT 데이터를 머신러닝 기반으로 임베디드 벡터를 생성하여 IoT 데이터를 최적화할 수 있는 처리 기법을 제한한다. 제안 기법은 수집된 IoT 데이터의 인덱스, 수집 위치(X축과 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등을 통합 관리하기 위해서 QR 기반으로 임베디드 벡터를 생성하도록 하였다. 또한, 제안 기법의 목적은 산업환경에서 다양한 IoT 장치에서 생성되는 데이터를 통합 활용할 수 있도록 하는 것이다. 특히, 제안 기법은 IoT 데이터를 비대칭적으로 연계하도록 IoT 데이터를 수집 과정에서 로드밸런싱을 수행한다. 제안 기법은 비대칭적으로 IoT 데이터를 그룹화할 수 있도록 IoT 데이터를 해시 기반으로 서로 직교화하도록 처리한다. 또한, IoT 데이터 종류 및 특성에 따라 그룹화하기 때문에 IoT 데이터는 주기적으로 생성하고 IoT 데이터 간에는 서로 간섭을 최소화할 수 있도록 하였다.

이 논문의 구성은 다음과 같다. 2장은 IoT 딥러닝 개요 및 기존 연구 결과들에 관해서 설명한다. 3장에서

는 AIoT 환경에 최적화된 머신러닝 기반의 IoT 데이터 처리 기법을 제안하고, 4장에서는 시뮬레이션을 통한 성능평가를 수행한다. 마지막으로 5장은 결론을 맺는다.

2. 선행연구

2.1 IoT 딥러닝

IoT는 모든 사물을 연결하여 정보를 공유하고 지능적으로 동작할 수 있도록 자동화, 정보교류 및 가공을 목적으로 사용된다. IoT의 기반 기술은 크게 센싱 기술, 유·무선 통신 및 네트워크 인프라 기술, 서비스 인터페이스 등으로 구분된다. 통신기술과 융합된 IoT 장치는 생성 및 수집되는 데이터가 기존 데이터 분석역량을 점점 넘어서고 있다[7]. IoT는 인공지능(Artificial intelligence) 중 딥러닝과 서로 밀접하게 관련되어 있다. IoT 장치에서 생성되는 데이터는 초기에는 단순한 데이터들뿐이었지만 인공지능과 관련된 기술이 접목되면서 IoT 데이터의 양과 질(정확도 등) 측면에서 학습·연산이 가능하다. IoT 딥러닝은 다양한 분야에서 사용되지만, 그중에서 의료 분야에서 주로 많이 사용된다. 그 이유는 이미지 인식과 자연어 처리를 모두 활용할 수 있기 때문이다. 최근 IoT 장치들은 스마트폰과 통신 기술을 연동하는 제품들이 많이 출시되고 있다. 이러한 장치들은 임베디드 기술을 이용하였기 때문에 Wi-Fi 망을 이용해 실시간으로 확인할 수 있다. 그러나, IoT 딥러닝에 사용되는 알고리즘들은 개발 단계로써 대용량의 데이터를 학습하여 특징을 추출하는데 걸리는 시간을 단축해야 한다.

2.2 AIoT

AIoT(Artificial Intelligence of Things)는 IoT 기술들이 다양한 환경(제조, 금융, 법률, 물류·유통 등)에서 활용되면서 수집되는 데이터를 효율적으로 처리/분석하기 위해서 인공지능과 융합된 기술을 의미한다[8].

AIoT는 수집된 데이터 셋을 인공지능 기술을 이용하여 플랫폼에서 활용할 수 있도록 데이터 유형에 따라 R, 엑셀, Tableau, Gephi 등을 사용하여 다양한 그래프 시각화(바차트, 3D 차트, 히트맵, parallel 그래프, 네트워크 그래프 등)를 통해 처리된다.

AIoT 핵심 기술의 특징은 초연결성(Hyperconnectivity), 초지능성(Superintelligence),

그리고 초융합성(Hyperconvergence) 이다. AIoT 가 활용되는 분야는 공장 자동화, Nest의 학습형 온도 조절기, 스마트홈 분야 등이 있다

2.3 기존 연구 결과

IoT 데이터가 다양한 환경 및 장치로부터 수집되어 인공지능 관련 기술들을 이용하여 다양한 환경(제조, 금융, 법률, 물류·유통 등)에서 사용되면서 IoT 데이터의 결측값을 탐지하는 연구들이 다양하게 진행되고 있다[9-11]. IoT 데이터 결측값의 발생 원인은 IoT 데이터 유형(침입 경로, 공인 데이터 셋 부족 등)에 따라 다르게 나타나고 있다.

인공지능 기술 중에서도 딥러닝 관련 기술들은 결측값을 실시간으로 점검할 수 있어서 IoT 데이터에 적합하다. 특히, IoT 데이터의 특징 정보들은 IoT 데이터를 사전 처리 단계에서 추출하기 때문에 IoT 데이터 차원을 줄여 정확도와 처리 효율을 향상시키고 있다[12].

IoT 데이터는 환경 및 목적에 따라 수집되는 데이터가 다양해질 수 있어서 IoT 수집 단계에서 IoT 데이터

를 정확하게 선별할 수 있어야 한다. 딥러닝 관련 기술에서는 IoT 선별 작업을 위해서 선형 판별 분석 대신 주성분 분석(PCA)을 사용한다. 그 이유는 IoT 데이터의 정확성을 높일 뿐만 아니라 IoT 데이터 간 비선형 상관관계의 특징 정보를 포착하지 않기 때문이다[13].

3. 머신러닝 기반의 IoT 데이터 처리 기법

3.1 IoT 데이터 처리 흐름도

제안 기법에서는 IoT 데이터의 정확한 처리를 위해서 머신러닝 기반으로 IoT 데이터의 수집 정보들을 IoT 데이터의 인덱스, 수집 위치(X축과 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등을 통합 관리할 수 있도록 QR 기반의 임베디드 벡터를 생성하여 관리할 수 있도록 IoT 수집 정보를 Fig. 1처럼 처리한다. Fig. 1과 같이 제안 기법은 수집된 IoT 데이터의 사용 목적 및 특성에 맞게 임베디드 벡터를 생성하기 때문에 머신러닝 기반의 알고리즘들을 이용하여 모델을 생성한 후 학습시킴으로써 성능평가를 수행한다.

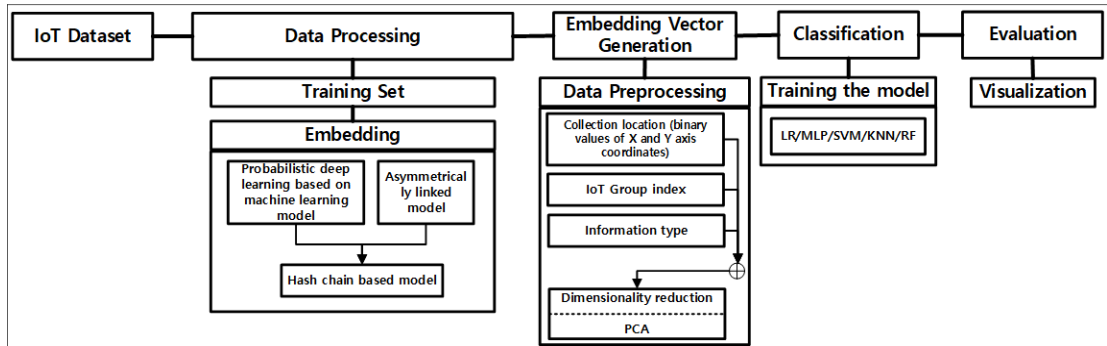


Fig. 1. Data Processing Flow for Proposed Scheme

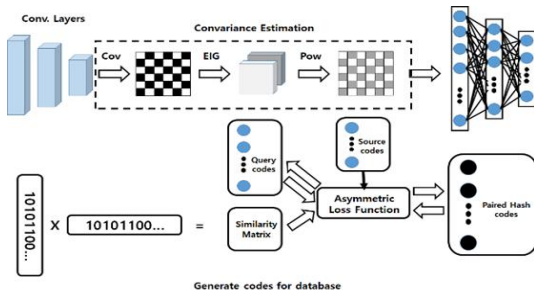


Fig. 2. Binary code vectorization of IoT information

제안 기법은 임베디드 벡터를 Fig. 2처럼 생성하며, IoT 데이터를 비대칭적으로 연계할 수 있도록 IoT 수집 과정에서 로드밸런싱을 수행한다. Fig. 2에서는 IoT 데이터의 가중치 확률을 이용하여 IoT 데이터를 비대칭적으로 그룹화하도록 IoT 데이터를 서로 직교화하도록 처리한다. 그 이유는 IoT 데이터를 외부적으로 안전하게 보호할 수 있도록 해시 처리가 가능하기 때문이다. 또한, IoT 데이터를 IoT 데이터 종류 및 특성에 따라 그룹화한 후 주기적으로 생성하도록 함으로써 IoT 데이터 간에 서로 간섭을 최소화하기 위해서이다.

3.2 임베디드 벡터 생성

제안 기법에서 IoT 데이터를 이용하여 임베디드 벡터를 생성할 때 사용되는 모델은 빈도 기반의 통계 모델을 사용한다. 그 이유는 IoT 데이터의 특성상 데이터의 분류가 복잡하지 않고 일정한 패턴으로 IoT 데이터가 생성하여 처리시키기 때문이다. IoT 데이터의 처리 목적 및 환경에 따라 모델은 변경할 수 있다. 제안 기법에서 수집되는 IoT 데이터들은 특정 지역에서 동시다발적으로 IoT 데이터들이 생성되기 때문에 식 (1) 처럼 동시다발적으로 IoT 데이터 분포를 확률적으로 변환하기 위해서 변환된 값에 음수(-)를 취하여 로그를 통해 정규화시킨다.

$$W_{ij} = -\log_2 \left(\frac{w_{ij}}{\sum w_{ij}} \right) \quad (1)$$

여기서, w_{ij} 는 동시다발적으로 수집되는 IoT 데이터를 의미하며, i, j 는 IoT 데이터를 행렬 형태로 구성하기 위한 행과 열을 의미한다.

또한, IoT 데이터의 수집 형태를 비대칭적으로 처리하도록 식 (2)와 같은 확률 함수를 사용한다.

$$E(D_i^x) = -\sum_{i=1}^n \frac{1}{n} \log \frac{1}{n} = \log n \quad (2)$$

여기서, x 는 IoT 데이터의 비대칭 연계 그룹 인덱스를 의미하고 i 는 IoT 데이터의 연계 수를 의미한다.

3.3 IoT 데이터의 결측값 탐지

제안 기법에서 IoT 데이터의 결측값은 Fig. 3의 과정을 통해 탐지하며, IoT 데이터를 임베디드 벡터화한 후 임베디드 벡터의 차원을 다르게 함으로써 IoT 데이터 수집 모델을 다양화하도록 하였다.

Fig. 3에서 사용한 드롭아웃(dropout) 비율은 IoT 데이터의 결측값을 탐지하는 과정에서 IoT 데이터의 과대 적합을 방지하기 위한 규제로서 사용하고 있다. 과대 적합을 최소화하기 위해서 IoT 데이터의 수에 따라 Dense를 2ⁿ번 크기로 증가하면서 IoT 데이터 수집 모델의 표현력을 향상시킬 수 있다.

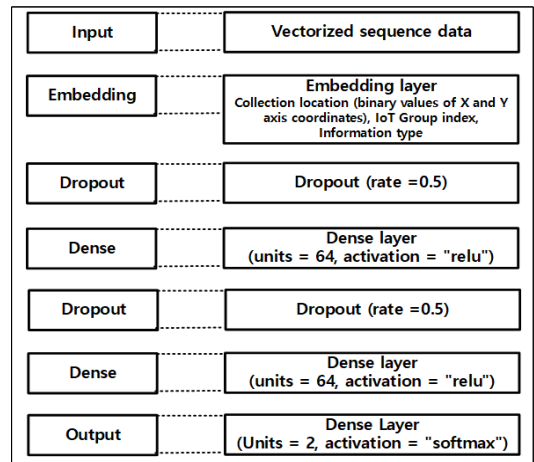


Fig. 3. Detection of missing values of IoT data

3.4 IoT 데이터 처리

3.4.1 IoT 데이터 추출

IoT 데이터 추출은 Table 1과 같은 머신러닝에서 사용되는 확률적 통계 기법들을 사용한다.

Table 1. IoT Data Features Extract and Selection

Feature	Equation
Mean	$\bar{X} = \frac{1}{n} \sum_{i=1}^N X_i$
Variance	$\sigma^2 = \frac{1}{n} \sum_{i=1}^N (X_i - \bar{X})^2$
Maximum	$X_{\max} = \max_{i=1,2,\dots,N} \{X_i\}$
Minimum	$X_{\min} = \min_{i=1,2,\dots,N} \{X_i\}$
Root mean square	$S = \sqrt{\frac{1}{n} \sum_{i=1}^N X_i^2}$
Energy	$E = \frac{1}{n} \sum_{i=1}^N X_i ^2$

Table 1의 데이터 추출 방법들은 IoT 데이터를 머신러닝 딥러닝 모델을 이용한 계층(Layer)을 구성하여 입력과 출력 데이터를 설정하여 학습을 진행하여 추출값을 획득한다. 제안 기법은 d 차원을 k 차원의 특징 부분공간($k < d$)으로 줄이도록 순차적 특징 선택 알고리즘 ()을 사용하여 계산 효율을 향상시켰다.

3.4.2 가중치 계산

가중치 계산은 IoT 장치로부터 동시다발적으로 수집되는 IoT 데이터를 식 (3) 처럼 행렬로 정의한 후 백

터 처리한다. 식 (3) 에 사용된 d_{ij} 는 쌍대비교 행렬로 고유벡터를 딥러닝 처리를 위해서 상관관계의 가중치를 생성한다.

$$V = \begin{matrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{matrix} \begin{bmatrix} 1 & d_{12} & \cdots & d_{1n} \\ d_{11} & 1 & \cdots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \cdots & 1 \end{bmatrix} \approx \begin{matrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{matrix} \begin{bmatrix} w_1/w_1 & w_1/w_2 & \cdots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \cdots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \cdots & w_n/w_n \end{bmatrix} \quad (3)$$

IoT 장치로부터 동시다발적으로 수집되는 데이터들은 결측값 없이 IoT 데이터를 처리하기 위해서 IoT 데이터들은 주기적인 시간 간격으로 데이터를 수집한다. 그 이유는 IoT 데이터를 임의의 크기로 그룹화 처리하기 위해서이다. IoT 데이터 수집 간격은 IoT 데이터 수집 크기 및 양에 따라 조절한다. IoT 데이터 중 결측값이 존재 유·무를 판단하여 결측값이 존재하는 데이터들은 모두 non-블록체인으로 처리하며, 정상적인 데이터인 경우에는 블록체인으로 처리한다. 이같이 결측값의 존재 유·무에 따라 블록체인과 non-블록체인으로 구분하는 이유는 IoT 데이터 간 간섭을 최소화하기 위해서이다.

3.4.3 데이터 연계

제안 기법은 IoT 데이터를 연계하기 위해서 IoT 데이터 간 부하를 낮추면서 비트 형태의 보안 정보 (Security Information) 값을 사용한다. IoT 데이터의 서명 값과 IoT 데이터를 해시 체인으로 연계 처리하도록 랜덤하게 할당한다. 이때, IoT 데이터들은 비대칭적으로 처리하기 때문에 IoT 데이터를 식 (4) 처럼 쌍대 비교 행렬 d_{ij} 을 이용하여 i 번째 벡터와 j 번째 벡터 간 연계 정보 R_{ij} 을 생성한다.

$$R_{ij} = \sum_{i=1, k=1}^n D_{ik}^x \times D_{jk}^y \quad (4)$$

IoT 데이터 연계에 사용되는 가중치는 식 (5) 처럼 계산한다.

$$y = wx + \beta \quad (5)$$

여기서, w 는 가중치, β 는 편향 값을 의미한다.

4. 연구 결과

4.1 실험 환경

시뮬레이션 환경은 Table 2와 같다. 시뮬레이션은 VMware 가상환경을 사용하여 실험하였고 시뮬레이션에 사용된 데이터 셋은 머신러닝을 위해서 Google Colab에서 파이썬, 사이킷런, 판다스, 넘파이 그리고 메타스플로잇 등의 개발 언어 및 라이브러리를 사용하였다.

Table 2. Simulation environment

Specification		
OS	Windows 10 Professional	
CPU	AMD Ryzen 5500X	
VMware Workstation	VMware Workstation 16	
	OS	Windows 10 Home
	CPU	4 core
	RAM/HDD	8GB / 100GB
Language	Python 3.9.8	
Tool	Google Colab	
Library	Scikit-learn, pandas, numpy, matplotlib, etc	

4.2 시뮬레이션 모델 구성

시뮬레이션에서 사용된 데이터 셋은 [14, 15]에서 25,000개를 수집하였으며, 수집된 데이터는 실시간 트래픽 데이터(raw data)로써 훈련 데이터(train set)와 테스트 데이터(test set)의 비율은 8:2로 유지하였다. 분류에 사용된 알고리즘은 LR(Logistic Regression), MLP(Multi-layer Perceptron), SVM(Support Vector Machine)이며, 학습에 사용된 알고리즘은 KNN(K-nearest neighbors), RF(Random Forest)이다. 기계 학습을 위해 5가지 선별된 특성(인덱스, 수집 위치(X축과 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등)과 조합하여 모델을 구성하였다. 머신러닝 학습 모델에 사용되는 알고리즘별 파라미터 값은 Table 3과 같다.

Table 3에서 SVM 알고리즘의 경우 정규화 파라미터 $\text{t}(Regularization parameter(C))$ 는 [0.001, 0.01,

0.1, 1, 5, 10, 25, 50, 100]으로 각각 평가하였으며, 그 결과 평균적으로 정규화 파라미터(Regularization parameter(C))가 10일 경우가 가장 좋은 성능을 가졌다. KNN 알고리즘의 경우 이웃의 수는 [1, 3, 5, 7, 10]으로 평가하였을 경우 평균적으로 5가 가장 좋은 성능을 가졌다.

Table 3. Parameter values of machine learning

Algorithm	Parameter Value
LR	Regularization strength(C) : 10 Solver : lbfgs penalty : l2
MLP	Hidden_layer : 100 Activation : relu Weight optimization : adam Learning rate : 0.001
SVM	Regularization parameter(C) : 0.001, 0.01, 0.1, 1, 5, 10, 25, 50, 100 Kernel : rbf Probability : true
KNN	Number of neighbors : 5 Weights : uniform
RF	number of trees : 100

4.3 성능 평가 척도

성능 평가에서 사용되는 시뮬레이션 성능 평가 척도는 식 (6)~식 (9) 와 같다.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (6)$$

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (9)$$

4.4 결과 및 분석

제안 기법의 성능평가 결과는 Table 4와 같다. Table 4의 성능평가는 정확도와 F1 스코어의 평균값을 라운드 수에 따라 비교하였다. Table 4의 결과처럼, 제안 기법은 머신러닝 학습 모델에 사용되는 알고리즘에 수집된 IoT 데이터를 분석한 결과 여러 특성을 선택한 IoT 데이터가 단일 특성을 선택한 IoT 데이터보다 좋은 결과를 보였고, IoT 데이터의 인덱스, 수집 위치

(X축과 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등을 조합한 경우가 탐지 성능이 향상된 것을 확인할 수 있었다. 그리고, 차원 축소 전/후를 비교한 결과 차원 축소 전에는 MLP(Multi-layer Perceptron)가 가장 좋은 성능을 보였고, 차원 축소 후에는 RF(Random Forest)가 좋은 성능을 보였다. IoT 수집 데이터의 수에 따른 학습 평가에서는 IoT 데이터의 수집량이 적을수록 LR(Logistic Regression) 알고리즘과 SVM(Support Vector Machine) 알고리즘이 성능이 좋았으며, IoT 데이터 수집량이 증가할수록 KNN(K-nearest neighbors) 알고리즘의 성능이 좋게 평가되었다. IoT 수집 데이터의 특성 수를 조합하였을 때의 성능은 MLP(Multi-layer Perceptron)와 RF(Random Forest)가 좋은 성능을 보였다.

Table 4. Performance result unit: %

		unit: %					
		CN	1	2	3	4	5
LR	A	84.52	88.75	86.75	89.79	90.01	91.75
	LT	0.008	0.007	0.011	0.009	0.012	0.01
	F1	80.85	84.26	82.37	86.63	83.81	88.05
	CN	1	2	3	4	5	6
MLP	A	94.49	91.87	93.47	92.89	93.63	94.47
	LT	0.007	0.005	0.003	0.001	0.003	0.004
	F1	93.17	91.24	89.82	90.71	88.58	92.41
	CN	1	2	3	4	5	6
SVM	A	87.28	88.35	87.37	89.24	90.21	88.01
	LT	0.008	0.009	0.008	0.007	0.006	0.005
	F1	86.49	90.17	88.34	89.02	87.07	91.57
	CN	1	2	3	4	5	6
KNN	A	87.21	88.74	87.98	91.14	90.11	89.72
	LT	0.002	0.005	0.004	0.005	0.006	0.005
	F1	82.28	86.07	83.28	88.73	86.47	89.17
	CN	1	2	3	4	5	6
RF	A	90.92	91.75	89.41	92.52	93.71	91.87
	LT	0.003	0.005	0.004	0.003	0.004	0.005
	F1	89.39	88.27	90.59	93.06	90.77	92.14
	CN	1	2	3	4	5	6

CN : The Count number of simulations

LT : Learning Time A : Accuracy F1 : F1-Score

5. 결론

인공지능 기술과 IoT 기술이 융합되면서 다양한 환경(제조, 금융, 법률, 물류·유통 등)에서 IoT 서비스가 활용되고 있다. 본 논문에서는 IoT 데이터의 안정적인 서비스를 지원하기 위한 머신러닝 기반의 IoT 데이터

최적화 기법을 제안하였다. 제안 기법은 임베디드 벡터를 사용하여 IoT 데이터의 인덱스, 수집 위치(X축과 Y축 좌표의 이진값), 그룹 인덱스, 타입, 종류 등을 QR 기반으로 비대칭적으로 연계 처리하였다. 또한, IoT 데이터의 수집 과정에서 로드밸런싱을 유지하도록 IoT 데이터를 그룹 관리하도록 IoT 데이터를 해시 기반으로 서로 직교화하도록 처리하였다. 성능평가, IoT 데이터의 차원 축소 전/후에 따라 차원 축소 전에는 MLP(Multi-layer Perceptron)가 차원 축소 후에는 RF(Random Forest)가 좋은 성능을 보였다. IoT 데이터의 학습 평가에서는 IoT 데이터량이 적으면 LR(Logistic Regression) 알고리즘과 SVM(Support Vector Machine) 알고리즘이 좋았으며, IoT 데이터 수집량이 증가할수록 KNN(K-nearest neighbors) 알고리즘 성능이 좋게 평가되었다. 향후 연구에서는 연구 결과를 기반으로 IoT 서비스를 제공하는 여러 환경에서 비교 평가를 수행할 계획이다.

REFERENCES

[1] K. H. Kim. (2021). Current Status and Implications of Artificial Intelligence (AI) Introduction by Major Industries. *Jincheon : KISDI*.

[2] J. S. Lee. (2013). A Study on Visualizing Method and Expression of Information Design for Big Data. *Journal of Basic Design & Art*, 14(3), 259-269. UCI : G704-001069.2013.14.3.026

[3] G. S. Choe, Y. G. Ham & S. H. Kim. (2013) Bigdata Visualization. *KSCI Review*, 21(1), 33-43.

[4] Friendly, M. (2008). A brief history of data visualization. *In Handbook of data visualization* (pp. 15-56). Springer, Berlin, Heidelberg. DOI : 10.1007/978-3-540-33037-0_2

[5] X. Sun & N. Ansari. (2016). EdgeIoT: Mobile edge computing for the internet of things. *IEEE Communications Magazine*, 54(12), 22-29.

[6] Y. J. Song & S. H. Hong. (2021). Build a Secure Smart City by using Blockchain and Digital Twin. *INTERNATIONAL JOURNAL OF ADVANCED SCIENCE AND CONVERGENCE*, 3(3), 9-13. DOI : 10.22662/IJASC.2021.3.3.009

[7] S. H. Sim. (2020). A Study on the Utilization and Prospect of ICT-Based Movement Support for the Mobility Handicapped. *INTERNATIONAL JOURNAL OF ADVANCED SCIENCE AND*

CONVERGENCE, 2(4), 1-6. DOI : 10.22662/IJASC.2020.2.4.001

[8] S. H. Sim. (2020). A Study on the Prospect of Using IoMT for Disease Prevention and Management. *INTERNATIONAL JOURNAL OF ADVANCED SCIENCE AND CONVERGENCE*, 2(3), 9-14. DOI : 10.22662/IJASC.2020.2.3.009

[9] U. Imtiaz & H. M. Qusay. (2019). A Two-Level Hybrid Model for Anomalous Activity Detection in IoT Networks. *Proceedings of the 2019 16th IEEE Annual Consumer Communications & Networking Conference(CCNC)*, 1-6.

[10] Machine Learning Repository. (2021). <https://archive.ics.uci.edu/ml/index.php>

[11] NSL, <http://www.nsl.cs.unb.ca/NSL-KDD>.

[12] Machine Learning Repository. (2021). <https://archive.ics.uci.edu/ml/datasets.php?form at=&task=&att=&area=&numAtt=greater100&num Ins=&type=seq&sort=nameDown&view=list>.

[13] B. Ravishankar, P. Kulkarni & M. V. Vishnudas, (2020). Blockchain-based Database to Ensure Data Integrity in Cloud Computing Environments. *Proceedings of the 2020 International conference on Mainstreaming Block chain Implementation (ICOMBI)*, 1-4.

[14] Machine Learning Repository. (2021). <https://archive.ics.uci.edu/ml/datasets.php?form at=&task=&att=&area=&numAtt=greater100&num Ins=&type=seq&sort=nameDown&view=list>.

[15] Y. Meidan, M. Bohadana, A. Shabtai, M. Ochoa, N. O. Tippenhauer, J. D. Guarnizo & Y. Elovici. (2017). *Detection of unauthorized iot devices using machine learning techniques*. preprint arXiv:1709.04647.

정 윤 수(Yoon-Su Jeong)

[정회원]



- 1998년 2월 : 청주대학교 전자계산학과(공학학사)
- 2000년 2월 : 충북대학교 전자계산학과(이학석사)
- 2008년 2월 : 충북대학교 전자계산학과(이학박사)

- 2012년 3월~현재 : 목원대학교 정보통신융합공학부 교수
- 관심분야 : IoT/AIoT/IIoT, 유·무선 통신 보안, 정보보호, 빅데이터, 클라우드 컴퓨팅
- E-Mail : bukmunro@mokwon.ac.kr

김 용 태(Yong-Take Kim)

[정회원]



- 1984년 2월 : 한남대학교 계산통계학과(공학학사)
- 1988년 2월 : 송실대학교 전자계산학과(공학석사)
- 2002년 2월 : 충북대학교 전자계산학과(이학박사)

- 2010년 10월~현재 : 한남대학교 공과대학 멀티미디어학과 교수
- 관심분야 : 모바일 웹서비스, 정보 보호, 센서 웹, 모바일 통신보안
- E-Mail : ky7762@hnu.kr