



스마트 팩토리 환경에서의 딥러닝 기반 제품 데이터 시각화 및 지능형 모니터링 기술 연구

ASHIQUZZAMAN AKM · 이동수 · 오승민 · 이지훈 · 김진술*
전남대학교 전자컴퓨터공학부

A Study on Deep Learning-based Product Data Visualization and Intelligent Monitoring Technology in Smart Factory Environment

ASHIQUZZAMAN AKM · Dongsu Lee · Seungmin Oh · Jihoon Lee · Jinsul Kim*

School of Electronics and Engineering, Chonnam National University, Gwangju, Korea

[요약]

오늘날의 스마트 공장에서 실시간 제품 데이터 모니터링 및 예측을 위한 클라우드 기반의 가상화 플랫폼은 산업 프로세스 자동화를 위해 매우 중요하다. MEC(Multi-Access Edge Computing) 기반의 서버와 같은 클라우드 플랫폼은 실시간 모니터링 환경에서 저지연과 고성능을 지원하기 때문이다. 공장 데이터 시각화 및 모니터링에 관한 많은 연구는 기존에도 많이 이루어졌으나 본 논문에서는 데이터 시각화 및 클러스터링을 위한 DNN(Deep Neural Network : 이하 DNN)을 제안했다. 본 논문을 통해 이전에는 없었던 스마트 팩토리 환경에서의 클러스터링 모듈과 데이터 시각화 및 DNN을 통한 최신 스마트 팩토리 데이터 모니터링 및 디스플레이 모듈에 대한 새로운 시스템을 제안하였다.

[Abstract]

Cloud based virtual platform for real-time product monitoring and prediction is crucial for automation of Industrial processes. The MEC based server and cloud platform gives a low latency and high performance in real-time environment monitoring. Although many research has been done in the data visualization, neural network and factory data monitoring, combining these methods to make a MEC based Smart factory data monitoring system has never been proposed. In this research article, we have proposed a data visualization and DNN (Deep Neural Network) with Clustering module. This proposed method for training the neural network achieves a 98% accuracy in the prediction and the cluster can give a 87% accuracy for 2 clusters. This accuracy is state of the art for factory production data classification. This state of art smart factory data monitoring and display module gives the smart factory system a MEC based and high accuracy based monitoring capability.

색인어 : 딥러닝, 클라우드 컴퓨팅, 모바일 엣지 컴퓨팅, 플랫폼 가상화, IoT, 지능형 예측

Key word : Deep learning ,Cloud Computing, MEC, Platform Virtualization, IoT, Intelligent Prediction

<http://dx.doi.org/10.9728/dcs.2019.20.10.1933>



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 04 September 2019; Revised 01 October 2019

Accepted 20 October 2019

*Corresponding Author; Jinsul Kim

Tel:

E-mail: zamanashiq3@gmail.com

I. 서 론

최근 제조 공장에서 제품 생산 데이터를 모니터링하고 예측하기 위해 구축한 지능형 프로그램은 전통적인 제품 생산 관리 방법론으로부터 급격한 변화를 주었다. “스마트 팩토리”로 알려진 대부분의 신세대 공장이나 생산 기관에서 생산 자동화 시스템을 구축함으로써 클라우드와 같은 방대한 공장 데이터 수집 솔루션을 통해 이전보다 사람이 직접 처리해야하는 업무들이 감소되었다. 또한 데이터 시각화는 차트, 그래프와 같은 시각적 요소를 사용하여 데이터 시각화 도구로 데이터의 추세, 특정치 및 패턴을 사람이 보고 이해할 수 있는 접근성을 제공한다. 데이터 시각화는 사람의 눈에 데이터를 표시하는 가장 좋은 방법이며 해당 데이터를 가장 잘 이해하기 위해 가장 적절한 방법이 될 수 있다. 또한 뉴럴 네트워크는 학습을 통한 정확도 향상과 실제 테스트를 통한 일반화로 최근 관련 연구 분야에서 주목을 받아왔다. 이를 통해 공장 빅 데이터를 분류할 경우 생산 결합과 빅 데이터베이스에 대한 분류 및 모니터링을 좀 더 쉽게 할 수 있게 된다. 이러한 프로세스는 지도학습의 일종으로서 경우에 따라 데이터 레이블이 요구되며 빅 데이터일 경우 엄청난 시간 관리 문제로 이어지게 된다. 또한 비지도학습의 가장 잘 학습된 예로서 클러스터링은 많은 알고리즘이 존재하지만 그 중 가장 많이 사용되는 알고리즘으로 K-Means Clustering이 있다. K-Means 알고리즘은 k개의 centroid를 식별한 다음 모든 데이터 포인트를 가장 가까운 클러스터에 할당하는 동시에 centroid는 최대한 작게 유지하는 과정을 의미한다. 주요 알고리즘은 데이터를 원하는 클러스터로 클러스터링하기 위해 다양한 N차원 포인트 벡터 또는 데이터 포인트를 처리할 수 있다. K-Means라는 용어는 군집 중심부까지의 평균 거리를 수를 가리킨다.

1950년대에는 IBM 연구소의 Nathaniel Rochester는 신경망 시뮬레이션에 대한 실험을 처음으로 시도했다. 그러나 당시 알고리즘으로서는 현대 신경망으로서의 다중 퍼셉트론 시뮬레이션에 필수적으로 요구되는 학습과 정확도 달성을 이루기에는 한계가 있었다.

1989년이 되어서야 역전파 알고리즘은 신경망을 훈련시키고 활용하는데 사용되기 시작함으로써 그 후 다시 새로운 세대의 신경망과 머신러닝 기반 모델 개발을 위한 다양한 응용으로 발전하였다.

최근에는 다양한 분야에서 딥러닝 기반의 분류와 모델링을 위한 다양한 모델들이 빅 데이터 중심 학습을 통해 엄청난 관심과 높은 정확도를 얻고 있는 상황이다.

신경망 개발 및 K-Means 데이터 분류 및 처리에 대해 섹션3에 상세히 기술되어 있지만 공장 데이터를 DNN에 통합하는 연구는 진행된 적이 없기에 본 논문에서는 데이터 처리에 관한 주요 중점 연구가 이루어졌다. 공장 생산 데이-

터베이스에서 데이터를 시각적으로 표현하기 위하여 먼저 수집하고 데이터를 사전에 처리한 후 DNN과 K-Means 모두 결합 생산 데이터를 모니터링 및 예측하도록 학습시킨다. 이 모델의 주요 문제는 DNN과 K-Means가 학습시킬 수 있는 불균형 데이터 셋이다. 이 모델은 지능적 모니터링에서 편향 분류를 최소화하기 위한 적합한 예측과 클러스터를 얻기 위해 새로운 학습 및 리샘플링 방법을 도입하였다.

본 논문은 다음과 같이 정리되어 있다. 본 논문의 2장에서는 기준 관련 연구에 대해 기술하고 3장에서는 심층 학습 및 클러스터링 모델을 학습시키기 위한 데이터 셋과 데이터 사전 처리를 설명한다. 그런 다음 4장에서는 실험 및 결과를 설명하고 마지막으로 5장에서는 결론으로 본 논문을 마무리한다.

II. 관련 연구 및 배경

데이터 시각화는 공장 환경에서의 생산 데이터 모니터링과 관련하여 광범위하게 쓰이는 용어 중 하나이다. 그러나 현재까지 데이터 시각화의 표준에 대한 지침과 상세한 연구가 이루어지지 않고 있었다. 본 논문에서는 데이터 시각화를 위해 실험 및 결과로서 처리된 데이터는 몇 개의 표과 그래프의 형태로 표현하였다. K-Means는 클러스터링 문제를 해결하기 위해 활용되고 있는 가장 간단한 비지도 학습 알고리즘 중 하나로 잘 알려져 있다. K-Means의 절차로는 원점에서 고정된 특정 수의 클러스터(k 클러스터로 가정)를 통해 주어진 데이터 셋에 대한 분류를 포함한다. 메인 아이디어는 각 cluster에 하나씩 k centroids를 정의하는 것인데 여기서 centroids는 각각의 위치가 다르기 때문에 정교하게 가능한 한 멀리 떨어져 배치되어야 한다. 다음으로 주어진 데이터 셋에 속하는 각 점들을 취하여 가장 가까운 중심에 연결한다. 이 때 결점이 존재하면 첫 번째 단계가 완료되고 초기 그룹핑은 여기서 완료된다. 이는 이전 단계에서 발생한 클러스터의 적절한 중심으로서 k개의 새로운 중심부를 재계산하는 과정이 된다. k개의 새로운 중심점을 얻은 다음 동일한 데이터 셋 포인트와 가장 가까운 새로운 중심점 사이에서 새로운 결합이 수행하기 위해 루프가 생성되는데 이 루프의 결과로 k centroids가 더 이상 변화가 없을 때까지 그들의 위치가 조금씩 변화해 간다는 것을 알 수 있었다. 즉 중심부는 더 이상 이동하지 않으면서 해당 알고리즘은 이러한 경우 오차 제곱 합수인 목표 합수를 최소화하는 것을 목표로 둔다. 목표 합수에 대한 계산식은 다음과 같다.

$$J = \sum_{j=1}^k \sum_{i=1}^n \| x_i^{(j)} - c_j \| \quad (1)$$

여기서 $\|x_i^{(j)} - c_j\|$ 는 data point $x_i^{(j)}$ 와 cluster center c_j (각 클러스터 센터에서 n개의 데이터 지점의 거리를 나타내는 지표) 사이에 선택된 거리 측정을 나타낸다.

이는 히든 레이어를 포함하는 MLP 또는 DNN을 적용한 주요 3개의 계층을 가지는 모델로서 도표화하여 표현될 수 있는데 그림 1에서 DNN의 단순화된 모델 아키텍처를 보여주고 있으며 중간에 표시된 히든 레이어는 여러 층으로 확장될 수 있다. 수학적으로, one-hidden-layer MLP는 공식 $f: R^d \rightarrow R^L$ 로 표현하며 여기서, D 는 입력 벡터 x 의 크기, L 은 출력 벡터 $f(x)$ 의 크기, 즉 매트릭스 표기 기법

$$f(x) = G(b^{(2)} + W^{(2)} + (s(b^{(1)} + w^{(1)}x))) \quad (2)$$

편향 벡터 $b^{(1)}, b^{(2)}$ 가중치 매트릭스 $W^{(1)}, W^{(2)}$ 와 활성화 함수 G 와 s .

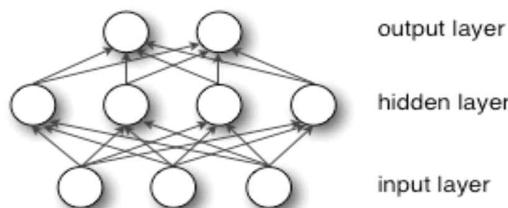


그림 1. 간소화된 딥러닝 뉴럴 네트워크

Fig. 1. Deep learning Neural Network simplified Architecture

벡터 $h(x) = \phi(x) = s(b^{(1)} + W^{(1)}x)$ 는 히든 레이어가 된다. $W^{(1)} \in R^{D \times D_h}$ 는 입력 벡터와 히든 레이어를 연결하는 가중치 행렬이다. 각각의 Column $W^{(1)}$ 은 입력 단위에서 i번째 단위까지의 가중치를 나타낸다. tanh 를 포함한 s 에 대한 일반적인 선택 사항으로 $\tanh(x) = (e^x - e^{-x}) / (e^x + e^{-x})$, 또는 로지스틱 시그모이드 함수인 $\sigma(x) = 1 / (1 + e^{-x})$ 는 DNN 활성화를 위한 연구는 최근 관련 분야에서 Relu(Rectified Linear Unit)가 비선형성을 위해 더욱 유용하게 사용되어 왔지만 특히 [4]에서는 Relu 함수를 주로 사용하여 구현한 내용임을 알 수 있었다.

반면에 신경망 기반 생산 모니터링과 지능적 예측에 대한 최근 연구로 Munir 등이 연구한 내용에 따르면 [5], 스마트 그리드를 통한 스마트 홈의 에너지 수요 대응을 관리하기 위해 RNN 기반 에너지 관리 수요 생산을 제안하였다. 또한 스마트 홈 사용자를 위해 RNN을 사용하여 지능형 에너지 예측 기기를 도입함으로써 기존의 문제를 해결하려는 시도가 있었다.

[6]에서 수행된 연구에서는 적시에 고장을 감지하여 전력

시스템을 보호하기 위해 전력 SCADA(감시 제어 및 데이터 수집) 시스템에서 디지털 보호 릴레이를 이용한 메서드를 개발하였고 이 논문에서 등장한 지능형 의사결정이 지원하는 디지털 제어는 해당 연구를 수행함에 있어 필수적인 요소임을 알 수 있었다. 응우옌 외 [7]에서 수행된 연구에서는 실제 공장에서 LSTM 기반 이상 징후 감지에 관하여 실험하였고 LSTM 기반 스트림의 데이터 이상 징후 감지에 초점을 둔 모델은 스마트 팩토리 모니터링 데이터를 기반으로 해당 제품이 불량 여부를 식별한다. 그러나 본 논문과의 차이점으로 현재 스마트 공장 상황에서 활용할 수 있는 실시간 예측과 시각화 기술에 관한 내용은 언급되어 있지 않았다.

IEC61400-25에 기초한 풍력발전소 SCADA 시스템의 데이터베이스 방법은 [8]에 설계되어 있었다. 그러나 본 논문과의 차이점으로 공장 데이터베이스와 전력 SCADA 시스템에는 데이터 시각화 및 지능형 설계 또는 딥러닝 기반 의사결정 지원 시스템에 대한 연구에 초점을 두지 않았다. 딥러닝 방식으로 IoT 기반의 홈 시스템 데이터 처리 및 모니터링은 최근 몇 년 사이 많은 발전이 이루어졌고 딥러닝을 통한 가정용 IoT 센서 데이터 처리에 관해서는 [9]에서 연구되었다. 그러나 본 논문과의 차이점으로 딥러닝을 이용한 스마트 IoT 홈 데이터 분석과 디바이스 제어 알고리즘에 초점을 두었다. [10]에서는 딥러닝 및 증강현실을 활용한 재해 대응능력 향상을 위한 네트워크 자원 확보, 딥러닝 및 데이터 통찰 기법을 통해 네트워크 자원 할당에 중점을 두었고 이와 동일한 종류의 대한 연구가 다른 곳에서도 계속해서 진행되어 왔음을 알 수 있었다[11-12]. 그러나 앞서 인용한 논문들에 나타난 연구들의 범위는 공장 데이터 모니터링에 집중되지 않았음을 알 수 있었다.

산업이나 농업 환경에 전반적으로 모니터링 기술이 점점 더 보편화되고 있는 가운데 [13]의 연구는 실시간 제품 데이터 모니터링을 위한 프레임워크 개발에 관한 연구에 초점을 두었다. 그러나 이 논문에서는 딥러닝 기반의 공장 데이터 모니터링에 관한 내용은 따로 언급되지 않았다. 클러스터 분석 및 자주 발생하는 패턴 마이닝에 기초한 풍력 터빈 고장 감지 접근법을 제안한 [12]에서는 공장 설정 및 심층 학습의 범위에서 빅 데이터 셋에 자주 사용되는 패턴 마이닝을 구현했다는 면에서는 [9]에서도 관련된 내용들이 있었다. 또한 스마트 팩토리와 스마트 시티에서 데이터 최적화와 관련된 연구는 [14], [15]에서 수행되었다. 그러나 본 논문과의 차이점으로 스마트 공장 빅 데이터 처리와 불균형 데이터 셋 학습에 대해서는 앞서 인용한 논문들에서 따로 언급되지 않았음을 알 수 있었다.

III. 제안된 방안

본 논문에서 제안된 스마트 팩토리 제품 데이터 시각화 및

지능형 모니터링 시스템은 몇 가지 사항들이 필수적으로 요구된다. 특히 공장 데이터 수집 및 처리 과정에서 이전에 시도되지 않았던 새로운 시각화 및 기계 학습 기반 데이터 사전 처리는 딥러닝 K-Means 클러스터링을 통해 구현하였다. 제안된 시스템은 모바일 엣지 컴퓨팅(MEC) 기술을 적용함으로써 클라우드 기반에서 연산량을 줄이거나 균형을 맞추는 효과가 있었다. 본 논문에서 제안된 모델은 프레임워크 아키텍처, 공장 데이터 수집 및 처리 단계와 시각화 및 기계 학습 기반 지능형 예측 모델, DNN K-means Clustering의 구성에 대하여 전체적으로 상세히 설명하였다.

3-1 프레임워크 아키텍처 개요

공장 데이터 시각화 솔루션을 위해 개발 중인 아키텍처의 전체 프레임워크는 그림 2에서 표시된 블록 다이어그램 순서대로 나타낼 수 있다. 본 공정은 먼저 생산된 데이터 및 품질 확인 값의 모든 데이터를 해당 공장 환경에 맞게 수집하여 누락된 값과 타임스탬프 생성 모듈을 확인하기 위해 데이터를 전송한다. 더 자세히 말하면 프로세스를 확인하고, 계속해서 입력되는 데이터 학습 프로세스에서 누락된 값을 모니터링하고 채운 다음, 데이터를 데이터 시각화 알고리즘과 정규화 데이터 모듈로 전송하여 수집된 데이터를 안전하게 보관하기 위해 데이터 저장과 기계 학습 모듈로 학습시켰다. 나중에 데이터는 DNN 학습 모델, K-means 모듈 모두에 전송된다. 그런 다음 기계 학습 기반 데이터 시각 모듈에서 처리된 데이터는 별도로 분리된 모듈과 같은 구성 요소 세션과 예측 분석 모듈에 표시된다.

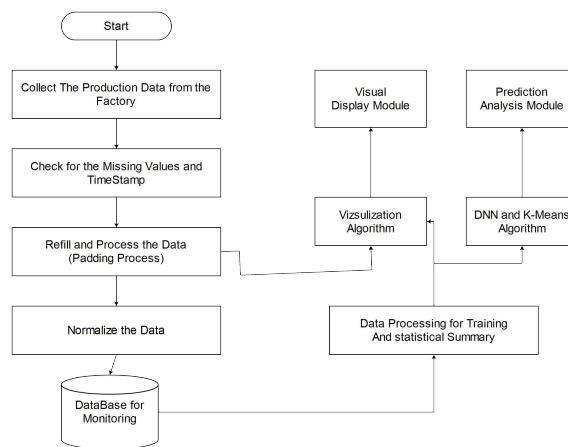


그림 2. 제안된 아키텍처 개요

Fig. 2. Proposed architecture Overview

3-2 데이터 수집 및 처리

이 과정에서 수집된 데이터는 메인 모듈에서 두 가지의 주요 처리를 한다. 처리 데이터는 어떤 형태로든 수집되어야 하

고 수집 과정에서 서버 컴퓨터와 모듈이 누락된 null 값과 누락된 데이터를 확인하여 결합 데이터를 제거하고 결합 데이터의 회수를 다시 요청함으로써 데이터에 맞는 구성을 보장한다. 그림 3과 같이 공장의 주요 데이터는 데이터베이스와 데이터 삭제 및 점검 모듈로 재수집된다.

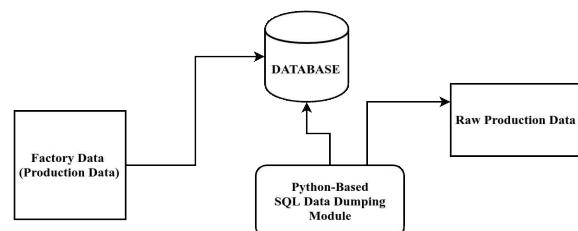


그림 3. 공장 데이터 수집

Fig. 3. Factory Data Collection

Python 기반 SQL(Structured Query Language) 모듈은 기계 학습 및 시각 모듈에 필요한 SQL 데이터를 사용한다. 그 후 모든 데이터는 일시적으로 수집 데이터베이스에 저장되고 기계 학습 모듈과 데이터 시각화 모듈에 필요한 모든 데이터는 리필과 패딩 프로세스 알고리즘으로 처리된다.

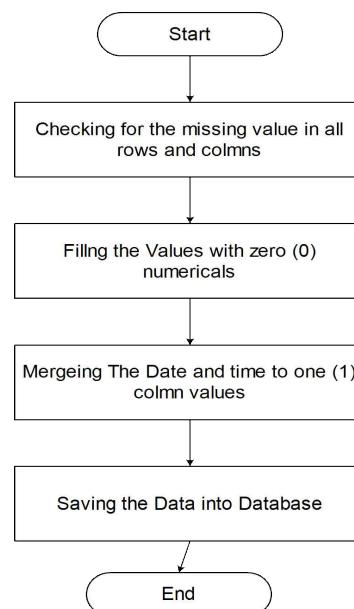


그림 4. 데이터 처리 및 패딩 과정

Fig. 4. Data cleaning and padding process

그림 4와 같이 행과 열별로 결측값의 모든 데이터를 검사하고 빈 셀에 유의하지 않은 값을 설정하기 위한 모듈의 처리 과정으로 기계학습 모듈과 데이터 시각화 모듈은 정확한 값을 학습시키기 위해 필수적으로 요구된다.

IV. 실험 및 결과

4-1 데이터 세트

제품 데이터 셋 모니터링을 위한 주요 데이터 셋은 빅 데이터베이스로, 42개의 개별 열과 27,430개의 데이터 행으로 구성된다. 데이터의 메인 열과 데이터 유형은 다음과 같다. 표 1과 같이 원시 데이터의 전체 모델은 42개의 서로 다른 색상의 데이터를 포함하여 구성되어 있으며 이 때 일부 데이터는 상황에 따라 누락된 값을 발생할 수 있으며 새로운 처리를 위한 모델로 전환된다. 데이터베이스가 모델을 학습시키는 과정에서 항상 동일한 중요성을 가지는 것은 아니기 때문에 중요하지 않은 데이터를 줄이기 위해서 어떤 경우에는 몇 개의 데이터는 관리자가 직접 입력한다. 관련 없는 데이터를 선택한 후, 23개의 열은 보여주고 표시하는 데 중요한 값과 데이터를 가지고 있기 때문이다. 메인 디스플레이 및 학습 값의 23개의 열이 결정된 후 23개의 열을 호출하여 추출시키고 SQL 모듈을 통해 새 파일에 저장한다. 또한 여기서 중요한 부분으로 데이터 처리 및 보고과정인데 시각화와 데이터 학습 모듈이 23개의 열에 초점을 맞춰 수행한다는 점이 이에 해당된다.

이 과정에서 데이터 셋을 삭제하게 되면 발생할 수 있는 문제는 일부 데이터를 문자열 값이 아닌 숫자 값으로 전송한다는 점에 있다. 현재 대부분의 데이터는 문자열로 변환된 수치 값을 가지고 있다면 값을 변경하지 않고 변환됨으로써 문자열의 트랜스코딩된 값은 범주형 값으로 전이된 다음 범주형 데이터를 숫자 값으로 전송된다. 그런 다음 23개의 열에 있는 모든 데이터를 검사하고 누락된 값을 확인한 다음 중요하지 않은 데이터는 폐드에 추가된다.

표 2에서는 시각적 프로그램 및 DNN 및 K-Means 프로그램에서 학습에 사용되는 최종 처리 데이터가 표시되어 있다. 이는 전체 데이터 셋을 전송하고 숫자 값만 사용하여 데이터 프레임으로 변환하는 과정이다. 이전 표는 별도의 날짜와 시계열로 다소 중복된 부분이 있었지만 새로운 데이터에서 결합되어 칼럼으로 재생성된다. 모든 문자열 범주형 값은 이해하기 쉽도록 최상의 고유 숫자 표현을 제공한다. 최종 데이터 셋은 계산 공간을 줄이기 위해 Python 기반 Pandas 라이브러리 데이터 프레임 개체로 변환되었다. 데이터 프레임 27,430개의 행이 있는 마지막 23개의 열에는 출력을 위해 필요한 6.1MB의 계산 공간만 있는데 이는 모바일 엣지 컴퓨팅 (MEC)을 통해 데이터 전송 및 연산 성능이 기준보다 높게 향상된 결과임을 나타낸다.

표 1. 주요 데이터 시트 컬럼명 및 데이터 타입

Table 1. MAIN DATA SHEET COLUMN NAMES AND DATA TYPES

Column Name	Data Type
QMS INDEX	int64
DAY INSP	string
TIME INSP	string
INSP TYPE	string
INSP TYPE NAME	string
OPER INSP TYPE INDEX	int64
OPER INSP TYPE	string
ITEM INDEX	int64
ITEM CODE	string
ITEM NAME	string
SAMPLE COUNT	int64
JUDGE CODE	string
JUDGE NAME	string
USER INDEX	int64
USER NAME	string
REMARKS	string
INSP NAME	string
INSP CODE DESC	string
INSP STANDARD	string
INSP METHOD	string
INSP LEVEL	string
INSP CYCLE	string
MIN	float64
MAX	float64
FIRST VALUE01	string
FIRST JUDGE01	string
FIRST VALUE02	string
FIRST JUDGE02	string
FIRST VALUE03	string
FIRST JUDGE03	string
MIDDLE VALUE01	string
MIDDLE JUDGE01	string
MIDDLE VALUE02	string
MIDDLE JUDGE02	string
MIDDLE VALUE03	string
MIDDLE JUDGE03	string
LAST VALUE01	string
LAST JUDGE01	string
LAST VALUE02	string
LAST JUDGE02	string
LAST VALUE03	string
LAST JUDGE03	string

표 2. 처리된 최종 데이터**Table 2. FINAL PROCESSED DATA**

Data Column Name	Rows	Cleaning Process	Datatype
DAY TIME INSP	27430	date-day	datetime 64[ns]
QMS INDEX	27430	int-val	float64
OPER INSP TYPE INDEX	27430	non-null	float64
ITEM INDEX	27430	non-null	float64
ITEM CODE	27430	int-val	float64
SAMPLE COUNT	27430	int-val	float64
JUDGE CODE	27430	non-null	float64
INSP NAME	27430	non-null	float64
INSP METHOD	27430	non-null	float64
MIN	27430	float-values	float64
MAX	27430	float-values	float64
FIRST VALUE01	27430	non-null	float64
FIRST JUDGE01	27430	non-null	float64
FIRST VALUE02	27430	non-null	float64
FIRST JUDGE02	27430	non-null	float64
FIRST VALUE03	27430	non-null	float64
FIRST JUDGE03	27430	non-null	float64
MIDDLE VALUE01	27430	non-null	float64
MIDDLE JUDGE01	27430	non-null	float64
MIDDLE VALUE02	27430	non-null	float64
MIDDLE JUDGE02	27430	non-null	float64
MIDDLE VALUE03	27430	non-null	float64
MIDDLE JUDGE03	27430	non-null	float64
LAST VALUE01	27430	int-val	float64
LAST JUDGE01	27430	non-null	float64
LAST VALUE02	27430	int-val	float64
LAST JUDGE02	27430	non-null	float64
LAST VALUE03	27430	int-val	float64
LAST JUDGE03	27430	non-null	float64

4-2 불균형 데이터 셋에 대한 리샘플링의 중요성

불균형 클래스 학습은 기계학습 분류에서 흔히 발생하는 문제로서 각 학습에서는 판측치의 불균형 비율이 존재할 수 있는데 클래스 불균형은 의료 진단, 스팸 필터링, 부정행위 탐지 등 다양한 분야에서 발생하고 있음을 알 수 있다. 그러나 본 논문에서 테스트한 공장 데이터에서의 생산 데이터 중 결함 데이터에 대한 분류가 매우 낮기 때문에 딥러닝 K-Means 클러스터링을 하기에는 한계가 있었다. 대부분의 기계 학습 알고리즘은 각 클래스의 샘플 수가 거의 같거나 완전히 같을 때 가장 학습에 대한 결과가 잘 나타나는데 이는 대부분의 알고리즘이 정확도를 극대화하고 오류를 감소시키는 데 그 목적을 두고 설계되었기 때문이다. 리샘플링 방안은 많은 종류

의 알고리즘이 존재하지만 본 논문에서의 실험을 위해 소수 계층 클래스는 전체 데이터 집합에 비해 매우 적은 예시를 가지고 있었기 때문에 소수계층 클래스를 오버샘플링하고 전체 데이터 셋을 새로운 데이터 셋으로 리샘플링하는 아이디어를 검증하고자 하였다. 여기서 오버샘플링은 소수 계층의 복사본 추가하는 것으로서 오버샘플링은 처리할 데이터가 부족하거나 없을 때 좋은 선택이 될 수 있다는 점을 검증하기 위해 오버샘플링 후, 데이터 셋의 부피를 2배로 늘리기 위해 소수계층 클래스의 사본을 여러 개 추가한 다음, 신경망과 클러스터 포인트에서 수정된 데이터 셋을 학습하고 예측하는 방안에 대해 제안하였다.

4-3 데이터 시각화

실시간 시각화를 위한 실험은 python 기반의 Matplotlib 라이브러리와 serial data reading 라이브러리로 수행하였다. 그런 다음 실시간 데이터를 적절한 값으로 변환하였고 메인 시각화 프로그램을 전송하면 그림. 5에 표시된 그래프와 같이 나타난다. 이는 해당 날짜에 대한 모든 데이터를 표시한 다음 3가지 종류의 판단 값과 최종 출력값을 예측한다.

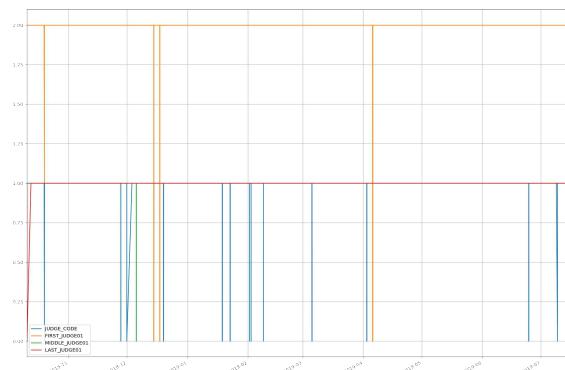


그림 5. 파이썬 포트 리스너 프로그램으로 구현한 실시간 제품 모니터링 데이터 시각화(2번째 패널)

Fig. 5. Data Visualization (1st panel) for product monitoring in real-time with python port listener program

또한 그림.6은 프로그램의 최소값과 최종 판단 코드를 보여주는 시각 프로그램을 위한 또 다른 패널을 나타낸다. 본 프로그램은 공장 생산 환경을 모니터링을 위해 원본 데이터를 실시간으로 보여주는 역할을 담당한다.

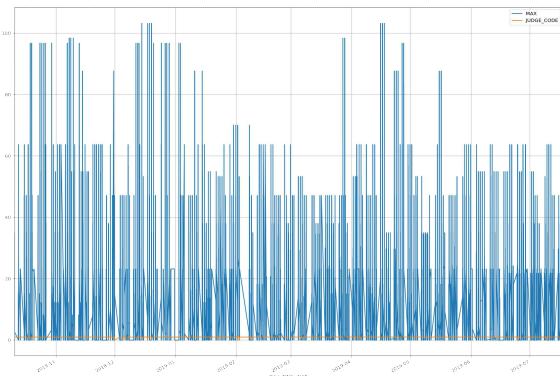


그림 6. 파이썬 포트 리스너 프로그램으로 구현한 실시간 제품 모니터링 데이터 시각화(2번째 패널)

Fig. 6. Data Visualization (2nd panel) for product monitoring in real-time with python port listener program

4-4 신경망과 K-Means 예측 정확도

신경망을 위한 DNN 모듈은 수정된 데이터 셋으로 학습시켰다. 더 자세히 말하면 DNN의 경우 27개, 13개, 2개의 히든 레이어가 있는 LSTM의 2개 레이어. 전체 신경 네트워크는 Adadelta 최적화 및 손실 함수에 대한 평균 오차 제곱(MSE)과 함께 Relu 활성화 함수[16]로 구현하였다. 신경망은 1000 batch size, 100 epoch에 대해 학습시켰으며 학습된 DNN의 손실 그래프는 그림 7과 같이 나타났다.

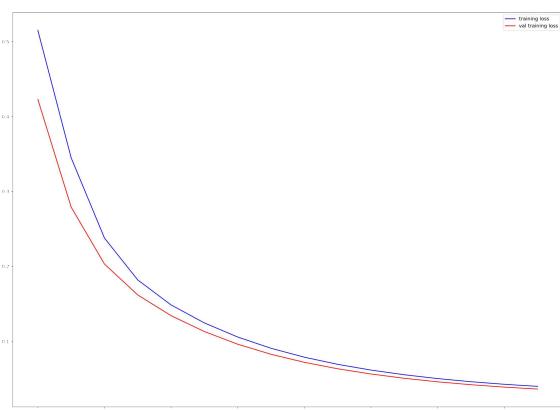


그림 7. 학습 중 표현된 DNN 손실

Fig. 7. DNN loss during training

또한 학습된 DNN의 대한 정확도 그래프는 그림 8에 나와 있다.



그림 8. 학습중 표현된 DNN 정확도

Fig. 8. DNN gaining accuracy during training

신경망의 높은 정확도는 클래스 가중치의 균형을 맞추고 리샘플링된 클래스 데이터와 함께 재학습시킴으로써 그 결과가 도출되었다. 학습을 위해 가중치 균형 기술은 공장 데이터 DNN 분류에 대한 새로운 접근법으로 학습시킨 후 모델은 데이터를 시험하고 학습 중 99%의 정확도가 도출되었으며 시험 데이터로는 98.87%의 정확도가 도출되었다. 한편, 생산과정에 대한 클러스터링 데이터는 높은 정확도를 위해 랜덤 초기화 및 k = 2 centerfolds로 학습시켰고 이는 이상 징후 감소 효과가 있었다. 리샘플링된 데이터 클러스터링은 그림 9에 표시된 것처럼 전체적으로 오버샘플링된 데이터 셋과 비교하면 약 87%의 정확도를 보였다.

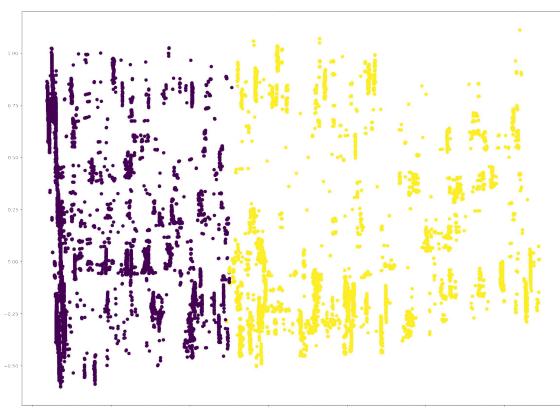


그림 9. cluster k = 2 데이터 시각화

Fig. 9. cluster k = 2 data visualization

데이터 셋은 23차원 데이터를 구성되었기 때문에 2D 평면에 표시할 수 없었고 시각화를 위한 차원감소가 요구되었다. PCA(Principal Component Analysis) 원리를 이용하여 구성요소를 분석함으로써 2개의 주요 구성요소를 얻기 위해 수행된 다음, 데이터가 클러스터 K-means 레이블과 함께 2D 평면에 표시될 수 있게 하였다.

4-5 MEC 기반 애플리케이션 배포

실시간 공장 제품 생산 데이터 모니터링 및 불량품 생산 예측을 위한 클라우드 기반 가상화 플랫폼은 최근 IoT 기반 엔티티 자동화에 매우 중요한 요소로서 MEC 기반 서버 및 클라우드 플랫폼은 실시간 환경 모니터링을 위해 짧은 대기 시간 및 빠른 계산 속도에 대한 높은 성능을 보였다. 공장 데이터는 실시간 포트 수신기 프로그램에서 수집되며 프로그램은 서버와 데이터 사이에 MEC 환경 기반으로 작동되며 직접적으로는 시각화기에서 데이터가 공급된다고 볼 수 있다. 시각화기는 데이터를 실시간으로 처리하여 사람이 데이터를 읽을 수 있도록 가공할 클라우드의 딥러닝 섹션에 있는 공장 실시간 데이터를 공급한다. 그런 다음 데이터는 MongoDB와 Python에서 개발한 데이터베이스로 전송되어 Amazon E2C 클라우드 컴퓨팅 인스턴스에 위치한 클라우드 플랫폼에서 향후 검토를 위해 안전하게 기록되고 보관된다. 이를 통해 해당 모델에 가상 플랫폼 기반 모듈을 구획하여 전체적으로 배치한 후 해당 공장 환경에 맞도록 적절한 MEC 기반 애플리케이션을 제공할 수 있었다.

V. 결론

본 논문에서는 클라우드 기반 MEC 기술을 적용한 실시간 공장 생산 데이터 시각화 서비스 및 딥러닝 기반 자원 모니터링 시스템을 제안하였다. 제안된 클라우드 기반 실시간 프로세스 관리는 이러한 공장 환경에서 매우 유용한 시스템임을 알 수 있었다. 클라우드 기반 플랫폼으로 이루어진 데이터 처리 및 시각화는 기존의 데이터 모니터링 방식과 비교했을 때 그 방식이 새롭고 향상된 성능을 가진 방안임을 실험 및 결과로 증명하였다. 제안된 새로운 모델은 시각화 프로세스와 데이터 수집, 삭제 및 처리를 과정을 위해 신경망을 학습시켰고 불량품 생산 예측에 대한 98%의 정확도를 도출하였다. 또한 두 군집에 대해서는 87%의 정확도를 나타내었다. 시각화 알고리즘은 실시간 데이터 생산 및 모니터링 관점에서 해당 프로세스를 위해 설계된 머신러닝 알고리즘으로 데이터 분류 또는 지능형 제품 데이터 모니터링을 정확하게 예측했다. 지능형 모니터링을 위하여 DNN 모델, K-means 클러스터링 모듈을 적용함으로써 현재 데이터에 대한 기술적 정확도의 상태를 알 수 있었으며 클라우드 기반의 시각화된 플랫폼으로 구현한 본 프레임워크는 MEC의 최대 유틸리티를 얻을 수 있도록 설계되었다. 향후에는 IoT 기반 스마트 시장과 공장의 증가로 인해 리소스 예측 및 모니터링을 위한 클라우드 기반의 MEC 서버에 관한 연구가 활발해지고 그 수요 또한 높아질 것으로 전망한다.

감사의 글

본 논문은 중소벤처기업부에서 지원하는 2018년도 산학연 협력 기술개발사업(No.S2655639)의 연구수행으로 인한 결과물이며 또한, 과학기술정보통신부(MEST)가 후원하는 한국연구재단(NRF)을 통해 기초과학연구프로그램의 지원을 받아 수행되었음(Grant No. NRF-2017R1D1A1B03034429).

참고문헌

- [1] J. A. Hartigan and M. A. Wong, "Algorithm as 136: A k-means clustering algorithm," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 28, no. 1, pp. 100–108, 1979.
- [2] G. Zini and G. d'Onofrio, "Neural network in hematopoietic malignancies," *Clinica chimica acta*, vol. 333, no. 2, pp. 195–201, 2003.
- [3] Y. Hirose, K. Yamashita, and S. Hijiya, "Back-propagation algorithm which varies the number of hidden units," *Neural Networks*, vol. 4, no. 1, pp. 61–66, 1991.
- [4] H. Hu, R. Peng, Y. Tai, C. Tang, and N. Trimming, "A data-driven neuron pruning approach towards efficient deep architectures. arxiv preprint," arXiv preprint arXiv:1607.03250, 2016.
- [5] M. S. Munir, S. F. Abedin, M. G. R. Alam, D. H. Kim, and C. S. Hong, "Rnn based energy demand prediction for smart-home in smart-grid framework," in Korea Software Congress, 2017, pp. 437–439.
- [6] Sung-Hwan Lee, Sang-Bae Park and Ihn-Seok Ahn, "Power Monitoring and Distributed Control by Digital Relay," *The Korean Society of Industrial Application*, vol. 4, no. 3, pp. 263–266, 2001.
- [7] V. Q. Nguyen, L. Van Ma, and J. Kim, "Lstm-based anomaly detection on big data for smart factory monitoring," *Journal of Digital Contents Society*, vol. 19, no. 4, pp. 789–799, 2018.
- [8] Chang Hun Chae, Hyo Yul Choi, Jun Suk Choi, "A Database Design Method for Wind Power Plant SCADA System based on IEC61400-25", *KIPS Transactions on Computer and Communication Systems*, vol. 1, no. 3, pp. 151–160, 2012.
- [9] I.-G. Lee and B.-Y. Lee, "Proceedings of the Korea Environmental Sciences Society Conference" in Proceedings of the Korean Environmental Sciences Society Conference. The Korean Environmental Sciences Society, 2018, pp. 67–67.
- [10] Younghwan Shin Jusik Yun Sunho Seo and Jong-Moon Chung, "Deployment of Network Resources for

- Enhancement of Disaster Response
Capabilities with Deep Learning and Augmented Reality
, "Journal of Internet Computing and Services, vol. 18, no. 5,
pp. 69–77, 2017.
- [11] Kim, Hwa-Jong , "Deep Learning City: A Big Data Analytics Framework for Smart Cities," *informationization policy*, vol. 24, no. 4, pp. 79–92, 2017.
- [12] F. Elijorde, S. Kim, and J. Lee, "A wind turbine fault detection approach based on cluster analysis and frequent pattern mining." *KSII Transactions on Internet & Information Systems*, vol. 8, no. 2, 2014.
- [13] D. L. White, J. L. Sharp, G. Eidson, S. Parab, F. Ali, and S. Esswein, "Real-time quality control (qc) processing, notification, and visualization services, supporting data management of the intelligent river," 2010.
- [14] A. Ashiquzzaman, L. Van Ma, S. Kim, D. Lee, T.-W. Um, and J. Kim, "Compacting deep neural networks for light weight iot & scada based applications with node pruning," in 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC). IEEE, 2019, pp. 082–085.
- [15] M. Mohammadi, A. Al-Fuqaha, S. Sorour, and M. Guizani, "Deep learning for iot big data and streaming analytics: A survey," *IEEE Communications Surveys & Tutorials*, 2018.
- [16] D.-A. Clevert, T. Unterthiner, and S. Hochreiter, "Fast and accurate deep network learning by exponential linear units (elus)," arXiv preprint arXiv:1511.07289, 2015.



ASHIQUZZAMAN AKM

2017년 : B.S. in Computer Science and Engineering from University of Asia Pacific, Dhaka Bangladesh.
2018년~현재 : 전남대학교 전자컴퓨터공학과 석사과정

※ 관심분야 : 지능형 네트워킹 시스템, 인간-컴퓨터 상호작용, 컴퓨터 비전, 딥러닝 및 머신러닝



이동수(Dongsu Lee)

2018년 : 광주대학교 정보보안학과 학사
2018년~현재 : 전남대학교 전자컴퓨터공학과 석사과정

※ 관심분야 : 모바일 클라우드 컴퓨팅, 스마트팩토리, 사물인터넷

오승민(Seungmin Oh)



2019년 : 나사렛대학교 디지털콘텐츠학과 학사
2019년~현재 : 전남대학교 전자컴퓨터공학과 석사과정

※ 관심분야 : 모바일 클라우드 컴퓨팅, 에너지 절약 시스템, 사물인터넷, 인공지능 및 강화학습

이지훈(Jihoon Lee)



2019년 : 전남대학교 전자컴퓨터공학부 컴퓨터정보통신학과 학사
2019년~현재 : 전남대학교 전자컴퓨터공학과 석사과정

※ 관심분야 : 모바일 클라우드 컴퓨팅, 스마트팩토리, 사물인터넷

김진술(Jinsul Kim)



2001년 : University of Utah, Salt Lake City, Utah, USA (B.S. Degree).
2005년 : 한국과학기술원 (KAIST) 석사
2008년 : 한국과학기술원 (KAIST) 박사

2005년 ~ 2008년: Researcher in IPTV Infrastructure Technology
Research Laboratory, Broadcasting/Telecommunications
Convergence Research Division, 한국전자통신연구원 (ETRI).

2009년 ~ 2012년: 나사렛대학교 조교수

2012년 ~ 현재: 전남대학교 부교수

※ 관심분야 : QoS/QoE, 측정/관리, 모바일 IPTV, 스마트 TV, 멀티미디어 통신, 디지털 미디어 아트 및 네트워크 지능기술