

전력 사용량 데이터 패턴 분석을 위한 시계열 이미지 생성 방안 연구

윤 준 철¹ · 오 상 원^{2*} · 김 진 술³¹전남대학교 ICT융합시스템공학과 박사과정 ^{2*}전남대학교 ICT융합시스템공학과 석사과정³전남대학교 ICT융합시스템공학과 교수

A Study of Time Series Image Generation Method for Analyzing Electric Power Usage Patterns

Junchul Yoon¹ · Sangwon Oh^{2*} · Jinsul Kim³¹Korea Electric Power Corporation, 55 Jeollyeok-ro, Naju-si, Jeollanam-do, Korea^{2*}Master's Course, Department of ICT Convergence System Engineering, Chonnam National University, Gwangju, Korea³Professor, Department of ICT Convergence System Engineering, Chonnam National University, Gwangju, Korea

[요 약]

시계열 데이터의 패턴을 분석하기 위해서는 주성분 분석이나 오토인코더와 같은 통계적 기법을 사용하거나 순환신경망 같은 딥러닝 모델을 기반으로 시계열 데이터의 특징을 활용한다. 그러나 데이터가 기록되는 환경 및 원인이 단순하지 않고 다양한 변인이 영향을 미치기 때문에 간단한 통계적 기법 또는 순환신경망 기반 딥러닝 모델만으로 좋은 성능을 기대하기 어렵다. 본 논문에서는 시계열 데이터를 이미지화시켜서 대표적인 이미지의 이진(Binary) 분류 모델인 컨볼루션 신경망 기반의 딥러닝 모델을 사용하여 전력 데이터를 분류하는 방법을 제안했다. 모델을 학습하기 위해 빌딩별 전력 소모 데이터를 하루마다 이미지로 만들어서 총 85개의 이미지를 사용했고, 주중 혹은 주말 데이터로 이진 분류했다. 이미지화를 위한 방법으로 RP (Recurrence Plot), GAF (Gramian Angular Field), MTF (Markov Transition Field) 알고리즘을 사용했다. 모든 시계열 이미지 모델이 기존 LSTM (Long Short-Term Memory) 기반 모델보다 동등하거나 더 높은 정확도를 보여줬고, 이미지화 기반 컨볼루션 신경망 모델 중에서 MTF 알고리즘으로 이미지화한 방법이 0.96의 F1-Score를 도출하여 가장 정확한 결과를 보여줬다.

[Abstract]

To analyze the pattern of time series data, statistical techniques such as Principal Component Analysis (PCA) or autoencoder are used, or features of time series data are utilized based on deep learning models such as Recurrent Neural Network (RNN). However, it is difficult to expect good performance only with simple statistical techniques or RNN-based deep learning models because the environment and causes in which the data is recorded are not simple and various variables affect it. In this paper, we propose a method to image time series data to classify power data using Convolutional Neural Network (CNN)-based deep learning models, which are binary classification models of representative images. To train the model of proposed method, a total 85 images were used by generating images of power usage data for each building every day, and they were binary classified as weekday or weekend data. Recurrence Plot (RP), Gramian Angular Field (GAF), and Markov Transition Field (MTF) algorithms were used as methods for imaging. All time series image-based models showed equal or higher accuracy than conventional LSTM-based models, and among imaging-based CNN models, imaging methods with MTF algorithms derived the highest F1-Score (0.96).

색인어 : 시계열 데이터, 시계열 이미지, 패턴 분석, 데이터 분류, 컨볼루션 신경망**Keyword** : Time Series Data, Time Series Image, Analyzing Patterns, Classification, Convolutional Neural Networks<http://dx.doi.org/10.9728/dcs.2022.23.8.1469>

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 21 June 2022; Revised 25 July 2022

Accepted 16 August 2022

*Corresponding Author; Sangwon Oh

Tel: 

E-mail: osw0788@gmail.com

I. 서 론

1-1 연구 배경 및 목적

수많은 데이터와 기술들이 연구되고 있는 현대 시대는 4차 산업혁명을 맞이하여 디지털 혁명이 이루어지고 있다[1]. 특히 주변에서 일어나는 모든 상황, 행동, 현상들이 데이터로 변환되어 저장되는 빅데이터 시대에는 여러 현상에 대한 분류와 분석에 관한 연구가 필요하다. 정보 통신 기술의 발전은 빅데이터 시대와 접목되어 사회의 다양한 분야에서 새로운 가치 창출을 일으키고 있다. 특히 에너지 관련 산업에서는 전력망과 정보 통신 기술을 결합한 스마트 그리드에 관한 연구가 진행되고 있다[2]. 그리고 전 세계에서 산업이 발전될수록 전력 에너지의 사용량도 증가하고 있어서 안정적인 전력 공급과 더불어 효율적인 전력 공급 전략을 수립하는 것이 절대적으로 필수적이다[3].

전력 공급 전략은 전력을 사용하는 소비자의 전력 사용 패턴을 파악하는 것이 중요하다. 일반적으로 전력 검출기가 시간 단위에 따라 전력 사용량을 기록하기 때문에 소비자가 사용하는 전력 데이터는 시계열 형태로 저장 및 기록된다[4]. 시계열 데이터는 일정한 주기를 가지거나 시간에 따른 추세를 보여주는 등 다양한 성질을 지니고 있다. 일반적으로 전력 소비 패턴은 데이터의 통계적인 속성 각각의 소비자마다 지니고 있지만, 전력 소비자들의 위치 혹은 종사 직업에 따라 전력 소비 패턴이 유사하거나 크게 다르지 않을 수 있다. 따라서 효과적으로 전력 소비 패턴을 분석하기 위해서는 유사한 전력 패턴들을 클러스터링할 필요가 있다[5].

일반적으로 통계 기법 또는 LSTM (Long Short-Term Memory)과 같은 RNN (Recurrent Neural Network) 기반 딥러닝 모델은 시계열 데이터를 예측하거나 패턴을 분석하기 위해 사용된다. 통계적 기법에는 주성분 분석[6] 및 오토인코더[7]와 같은 차원 감소를 통해 특징을 추출하여 시계열 데이터의 특징을 기반으로 패턴을 분석하는 것이 포함된다. 딥러닝 모델 기반 기술에는 원본 데이터를 예측하고 각각 다른 특징을 지닌 시계열 데이터를 분류하기 위한 LSTM 및 GRU 기반 모델 학습, 완전히 연결된 계층 기반 이진 분류를 분류하는 모델 활용 등이 포함된다[8].

그러나 각 전력 데이터마다 기후, 지리적 특성, 사회적 요인과 같은 다양한 외부요인들이 일정하지 않은 시간 주기를 가지고 전력 데이터가 기록될 수 있다. 또한, 주 외부요인에 따라서 패턴들을 구분하는 방식이 달라질 수 있다. 기존에는 다양한 외부요인을 고려하기 위해 사전에 차원 감소 기법을 사용한 복잡한 딥러닝 모델들을 사용한다. 전처리 단계에서 시계열 데이터의 윈도우 크기를 결정하고, 적절한 차원 감소 기법을 통해 특징을 추출하는 과정이 필요하므로 자동화하기 힘들고 데이터 정제과정이 복잡하다.

따라서 본 논문에서는 전력 소비 데이터의 패턴 분류를 위해 LSTM 기반 및 차원 감소 기반의 모델이 아니라 CNN

(Convolutional Neural Netowkr) 기반의 이진 분류 모델을 사용하기 위해 RP (Recurrence Plot), GASF (Gramian Angular Summation Field), GADF (Gramian Angular Difference Field) 및 MTF (Markov Transition Field) 알고리즘을 활용하는 방법과 기존 딥러닝 모델 기반 방법과 비슷하거나 더 높은 성능을 도출하는 방법을 제안했다.

II. 이론적 배경

본 논문에서는 시계열 데이터의 패턴을 분류하기 위해 데이터를 이미지로 변환한다. 기존 시계열 데이터를 사용하여 데이터 분류 모델을 학습하면, 이전 시간 단위에 기록된 데이터에 의존되어서 데이터 패턴의 특징을 포착하기 어렵다. 우리는 시계열 데이터를 이미지화하여, 시간 종속성에 영향을 덜 받고 데이터 패턴의 특징에 영향을 더 받도록 했다. 다만, 시간에 따라 변화하는 시계열 데이터의 추세와 시간 종속성을 한 이미지에 표현하기 위해서 거리행렬, 극좌표 및 전이확률을 활용한 이미지화 기법들을 아래와 같이 사용했다.

2-1 Recurrence Plot

RP 알고리즘은 2차원에 데이터값의 회귀를 표현함으로써 m-차원 위상 공간 궤적을 탐색하는 것을 목표로 하는 시각화 알고리즘이다[9-10]. 시계열 데이터를 RP를 이용해 이미지로 변환하기 위해서는 먼저 시계열 데이터의 m-차원 공간 궤적을 구성해야 한다. 시계열 데이터가 $X = \{x_1, x_2, \dots, x_i\}$ 와 같이 주어졌을 때, m-차원 공간 궤적은 식 (1)과 같이 정의할 수 있다. 러닝은 그 자체로 핵심 기술이 될 만큼 본 연구에서 특별하지 않다. 하지만 본 연구에서 사용할 딥러닝 범주에 속하는 주요 알고리즘은 CNN이라고 불리는 딥러닝 신경망의 파생된 알고리즘이다. CNN은 학습 데이터에서 중요한 특성을 추출하기 위해 합성곱의 개념을 사용한다. 합성곱을 통해 데이터를 처리하면 해당 신경망에서 특정 위치에 수렴하는 특징(Convergence Feature)이 발생한다. 이에 해당하면 큰 값을 할당하고, 그렇지 않으면 낮은 값을 할당한다. 즉, 각 커널 셀 값의 요소별 곱셈과 그에 상응하는 영상 픽셀(Image Pixel)값이 합성곱 행렬의 주어진 위치에서 커널 셀과 겹쳐지는 부분을 합산한다. 정밀한 공식은 식 (1)에서 설명한다.

$$S = \{s_1 = (x_1, x_2), s_2 = (x_2, x_3), \dots, s_n = (x_n, x_{n+1})\} \quad (1)$$

s_n 은 x_n 에서 x_{n+1} 까지 시계열 데이터의 궤적이라고 볼 수 있고 S는 궤적들의 집합이다. 행렬 형태로 궤적 정보를 표현하기 위해 공간 궤적 데이터를 기반으로 거리 행렬을 정의할 수 있다. 이를 수학적으로 표현하기 위해 식 (2)처럼 행렬의 원소들을 정의할 수 있다.

$$R_{i,j} = \theta(\epsilon - \| \vec{s}_i - \vec{s}_j \|) \quad (2)$$

ϵ 는 거리 역치값이며, $\theta(x)$ 는 단위 계단 함수이다. $R_{i,j}$ 는 거리 행렬의 원소값이며 수식 상으로 s_i 와 s_j 의 거리를 의미한다. 두 궤적 간의 거리를 계산하기 위해서 유클리드 공간에서 두 점의 거리를 계산하는 L2 노름을 사용한다. 따라서 거리 행렬 R 은 두 s 간의 거리를 행렬로 표시한 꼴이다. R 의 대각원소들은 자기 자신과의 거리를 정의하기 때문에 0이 된다. 그리고 $R_{i,j}$ 와 $R_{j,i}$ 는 식 (2)의 정의에 따르면 같으므로 R 은 대칭행렬이 된다.

RP 알고리즘에 의해 변환된 시계열 이미지는 궤도가 같은 위치에 있는 시간 쌍의 모음을 묘사한다. 궤도의 주기가 T 라고 하면 존재하는 모든 시간 쌍은 T 의 배수로 분리되어 시계열 이미지에서 대각선으로 표시된다. 즉, RP의 시각적인 표현은 시계열 데이터의 주기 및 변화폭에 대한 정보를 제공한다.

2-2 Gramian Angular Field

GAF (Gramian Angular Field)는 각 시점 간의 시간적 상관관계를 극좌표를 기반으로 표현하는 알고리즘이며, 시계열 데이터를 극좌표 기반 행렬 형태의 이미지로 변경하는 것은 시간 상관관계를 보존할 수 있다는 장점이 있다[11]. 일반적으로 GAF 알고리즘을 적용하기 전에 시계열 데이터의 값을 극좌표 값으로 변환할 때 치역을 통일하기 위해서 $[-1, 1]$ 혹은 $[0, 1]$ 구간으로 정규화시킨다. 이때, 정규화된 신호를 \tilde{x}_i 라고 하고 시간 인덱스를 t_i , 반지름을 r_i 로 표현한다면 식 (3)처럼 극좌표로 변환할 수 있다.

$$\begin{cases} \phi_i = \arccos(\tilde{x}_i), & -1 \leq \tilde{x}_i \leq 1, \text{ and } \tilde{x}_i \in \tilde{X} \\ r_i = \frac{t_i}{N}, & t_i \in N \end{cases} \quad (3)$$

여기서 N 은 극좌표계의 범위에 따른 정규화 상수이다. GAF는 각도의 합과 차에 따라 두 가지 방식으로 나뉜다. GASF는 i 와 j 의 시간쌍으로 이루어진 극좌표계 시계열 데이터의 각도의 합으로 나타내며, 식 (4)와 같이 정의한다.

$$GASF = [\cos(\phi_i + \phi_j)] = \tilde{x}' \cdot \tilde{x} - \sqrt{1 - \tilde{x}'^2} \cdot \sqrt{1 - \tilde{x}^2} \quad (4)$$

GADF는 GASF와 반대로 극좌표계의 각도의 차로 식 (5)와 같이 정의한다.

$$GADF = [\sin(\phi_i - \phi_j)] = -\sqrt{1 - \tilde{x}'^2} \cdot \tilde{x} - \tilde{x}' \cdot \sqrt{1 - \tilde{x}^2} \quad (5)$$

GAF는 왼쪽 상단에서 오른쪽 하단으로 이동함에 따라 시간이 증가하기 때문에 시간 종속성을 보존한다. 주대각선은 원시데이터의 값과 각도 정보를 포함하기 때문에 이를 이용하여 원시 시계열 데이터를 복구할 수 있는 장점이 있다.

2-3 Markov Transition Field

MTF 알고리즘은 이산화한 시계열 데이터의 전이 확률을 나타내는 알고리즘이다[11]. MTF를 구성하기 위해 주어진 시계열 데이터 집합 X 를 값에 따라 Q 개의 구간으로 분할한 뒤, 시간 인덱스 t_i 의 시계열 데이터 값 x_i 에 맞는 구간 q_j ($j \in [1, Q]$)에 배정한다. 시간 축을 따라 1차 마르코프 체인 방식으로 $Q \times Q$ 크기의 가중치 인접 행렬 W 을 아래 식 (6)과 같이 구성한다.

$$W = \begin{pmatrix} w_{11} P(x_i \in q_1 | x_{i-1} \in q_1) & \dots & w_{1Q} P(x_i \in q_1 | x_{i-1} \in q_Q) \\ \vdots & \ddots & \vdots \\ w_{Q1} P(x_i \in q_Q | x_{i-1} \in q_1) & \dots & w_{QQ} P(x_i \in q_Q | x_{i-1} \in q_Q) \end{pmatrix} \quad (6)$$

$w_{i,j}$ 는 q_i 구간에서 q_j 구간으로 전이하는 빈도를 나타낸다. W 의 각 열의 합을 1로 정규화함으로써 마르코프 전이 매트릭스를 구성할 수 있다. 이 과정에서 W 는 X 의 분포와 시간 인덱스 t_i 에 대한 시간 종속성이 제거된다. 이러한 W 의 정보 손실을 극복하기 위해, 시간 순서를 따라 각 확률을 정렬하여 MTF를 아래 식 (7)과 같이 정의한다.

$$M = \begin{bmatrix} M_{11} & \dots & M_{1n} \\ \vdots & \ddots & \vdots \\ M_{n1} & \dots & M_{nn} \end{bmatrix} = \begin{bmatrix} w_{i,j|x_1 \in q_i, x_1 \in q_j} & \dots & w_{i,j|x_1 \in q_i, x_n \in q_j} \\ \vdots & \ddots & \vdots \\ w_{i,j|x_n \in q_i, x_1 \in q_j} & \dots & w_{i,j|x_n \in q_i, x_n \in q_j} \end{bmatrix} \quad (7)$$

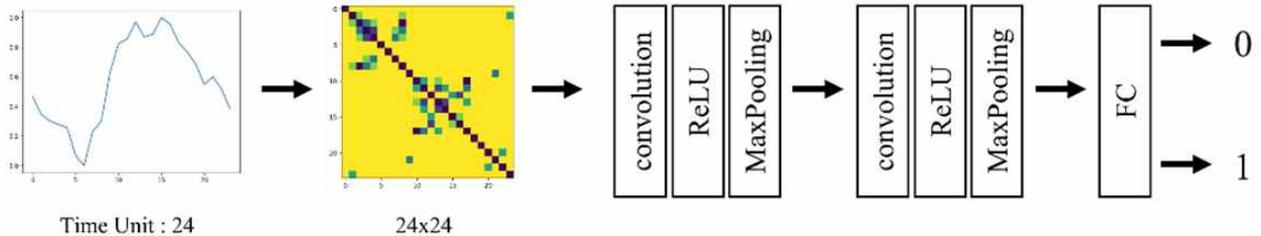


그림 1. 시계열 이미지를 활용한 제안된 데이터 분류 방법
 Fig. 1. Proposed data classification method using time series images

MTF의 i 행 j 열의 값인 $M_{ij} | | i-j | = k$ 는 시간 인덱스 t_i 일 때의 데이터 값이 속한 구간인 q_i 에서 시간 인덱스 t_j 의 데이터 값이 속한 구간인 q_j 로 전이할 확률, 즉 두 시간의 차이가 k 인 지점 간의 전이 확률을 나타낸다. 구간의 너비가 크면 대부분의 값을 평균에 가장 가까운 구간으로 집계하고, 구간의 너비가 작으면 극한 구간에서 집계되는 값이 적어진다. 따라서 시계열 이미지 안에서 구간 크기에 따라 시계열 데이터의 표준편차 및 분포를 시각적으로 확인할 수 있다[11].

III. 제안하는 방법

본 실험에서는 시계열 이미지를 생성하기 위해서 RP 알고리즘을 활용해서 시계열 데이터 셋의 개수만큼 거리 행렬을 생성한다. 그림 1.과 같이 24개의 시간 단위에서 기록된 시계열 데이터를 RP, GASF, GADF, MTF 알고리즘으로 시계열 이미지를 생성한다.

RP 알고리즘을 사용하여 이미지 데이터를 생성하기 위해 역치값을 사용하는 CRP (Cross Recurrence Plot)과 일반 RP 알고리즘을 고려했다. CPR은 거리 행렬의 값이 일정 수준 이상이면 1, 그 미만이면 0으로 설정하여 이진값을 구성된 RP 행렬을 계산하는 알고리즘이다[12]. 실험 과정에서 두 가지의 RP 알고리즘을 사용하여 이미지 데이터셋을 만들고 그 중 하나를 택하여 RP 알고리즘 기반 이미지 셋으로 정한다.

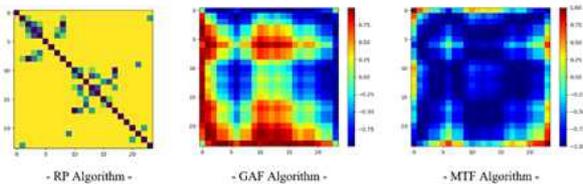


그림 2. 변환된 시계열 이미지
Fig. 2. Transformed time series image

표 1. 제안된 이진 분류 모델의 합성곱 신경망 구조
Table 1. CNN structure of the proposed binary classification model

Layer	Channel	Kernel
Input	1*24*24	1*1*24
Conv1	24*24	3
Relu	24*24	3
MaxPooling	24*24	3
Conv2	16*16	3
Relu	16*16	3
MaxPooling	16*16	3
Fully Connected	1	N/A

GAF 알고리즘 기반으로 이미지 데이터를 생성하기 위해 우선 일일 시계열 데이터들을 모두 [0, 1]의 범위를 가지도록 정규화한 후에 GASF, GADF 알고리즘을 적용하여 이미지 데이터 셋을 구축했다. MTF 알고리즘 기반으로 이미지 데이터를 생성하기 위해 상위 차원의 시계열 데이터를 합쳐서 RP, GAF 알고리즘을 사용하는 시계열 데이터와는 달리 범위와 시간 단위가 많은 시계열 데이터를 사용했다. 그리고 실험 단계에서 다른 알고리즘과 같은 차원의 이미지 데이터 셋을 생성하기 위해서 24개의 구간을 분할하여 마르코프 전이 매트릭스를 구성했다.

4가지 알고리즘을 사용해서 만든 이미지 데이터 셋들을 이진분류 하기 위해 합성곱 신경망을 사용한다. 표 1.에서 이진분류 하기 위한 신경망 구조가 명시되어 있다. 본 실험에서 사용하는 이미지 데이터의 크기가 24*24이므로 입력 데이터의 차원을 1*24*24로 설정하였다. 신경망을 통하여 입력 이미지 데이터의 패턴을 판단하여 분류하게끔 신경망을 구축했다.

IV. 실험 및 결과

본 실험에서 사용한 시계열 데이터는 한국에너지공단에서 제공한 빌딩별 전력 사용 데이터이다. 60개의 빌딩에서 검출된 전력 사용 데이터가 2020년 6월 1일부터 2020년 8월 24일까지 1시간 간격마다 기록됐다. 위 데이터는 총 122,400개의 행이 존재하고, 이상치 및 결측치는 존재하지 않는다. 전력 사용량 외의 온도, 습도 등 다양한 열이 존재하지만, 온전한 전력 시계열 데이터를 사용하기 위해 전처리 과정에서 삭제했다.

전력 데이터의 패턴을 구분 짓기 위해서 주중 데이터와 주말 데이터로 나누어서 이진 분류를 시행한다. 시간 기준이 하루이므로 시간마다 기록된 전력 소모 데이터들을 24개의 시간 단위로 묶어 전력 데이터를 준비한다. 그림 3-6 같이 85일 동안 기록된 시계열 데이터를 이미지화했다. 빌딩마다 85일 동안 기록된 시계열 데이터들을 85개의 RP, GASF, GADF, MTF 알고리즘을 사용하여 이미지화했다. 그림 3은 특정 빌딩의 85일간 기록된 일일 시계열 데이터를 RP 알고리즘을 사용하여 Colorbar 형식으로 나타낸 이미지이고, 그림 4는 GASF, 그림 5는 GADF, 그림 6은 MTF 알고리즘을 사용하여 변환된 이미지들이다.

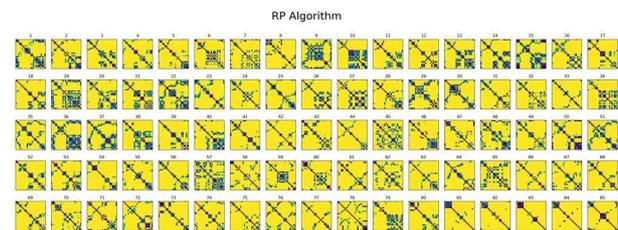


그림 3. 85일 동안 검출된 전력 소모량 시계열 이미지 (RP)
Fig. 3. Power usage time series image detected over 85 days (RP)

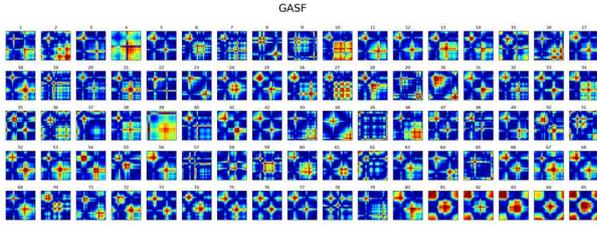


그림 4. 85일 동안 검출된 전력 소모량 시계열 이미지 (GASF)
 Fig. 4. Power usage time series image detected over 85 days (GASF)

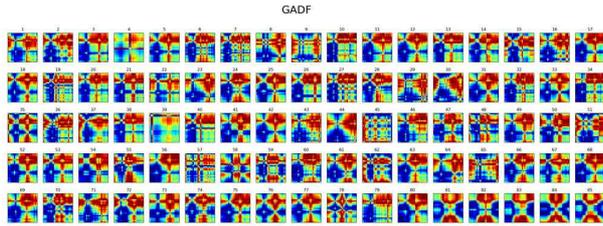


그림 5. 85일 동안 검출된 전력 소모량 시계열 이미지 (GADF)
 Fig. 5. Power usage time series image detected over 85 days (GADF)

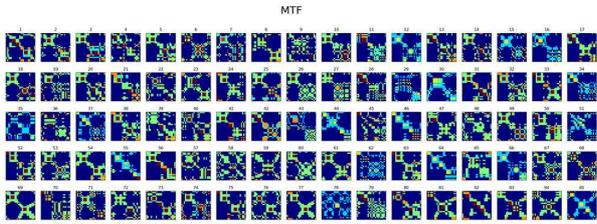


그림 6. 85일 동안 검출된 전력 소모량 시계열 이미지 (MTF)
 Fig. 6. Power usage time series image detected over 85 days (MTF)

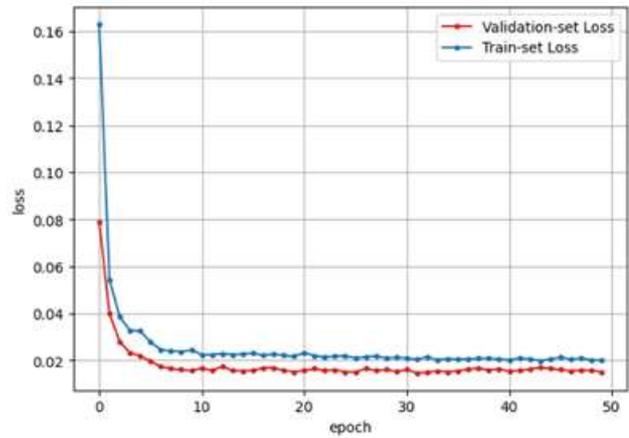
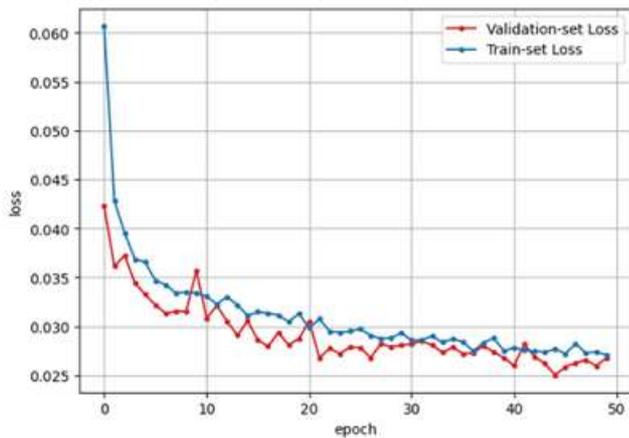
MTF 알고리즘을 사용한 방법은 별도의 초모수 설정이 없어 바로 알고리즘을 적용하여 이미지화한 데이터를 사용했다.

기존의 (60, 85, 24)의 shape를 가진 데이터를 (60, 85, 24, 24) shape로 차원을 증가했다. 원래 데이터와 차원이 증가한 데이터를 패턴 분류 모델의 훈련 및 시험 데이터로 사용하여 두 결과를 비교했다.

학습을 위하여 (60, 85, 24) 차원을 지닌 시계열 데이터를 빌딩 순서와 관계없이 5100개의 일일 시계열 데이터로 변환한 후, 6:2:2 비율로 훈련용 데이터 셋, 검증용 데이터 셋, 그리고 테스트용 데이터 셋으로 분할했다. 그리고 각 데이터가 주중인지, 주말인지에 따라 라벨링을 하고, RP, GASF, GADF, MTF 알고리즘을 사용해 이미지화해서 이진 분류 모델의 데이터 셋으로 사용했다. 시계열 데이터 셋은 LSTM 기반 이진 분류 모델의 학습 및 테스트 데이터로 사용되고, 이미지화한 데이터 셋은 CNN 기반 이진 분류 모델의 학습 및 테스트 데이터로 사용된다. LSTM 기반 이진 분류의 모델과 CNN 기반 이진 분류의 모델은 표 1.과 같이 구성했다. Epoch는 50회, Batch size는 85로 설정한 후 이진 분류 모델을 학습했다.

패턴 분류 학습은 그림 7와 같이 CNN 기반 모델의 손실 값이 LSTM 기반 모델의 손실 값과 유사한 것을 확인할 수 있다. 시계열 데이터를 사용해서 해당 일일 전력 소모 데이터가 주중 혹은 주말에 기록된 데이터인지를 판단할 때와 이미지화시킨 데이터를 사용해서 해당 이미지가 주중 혹은 주말에 기록된 전력 소모 시계열 이미지인지를 판단할 때 모두 Epoch가 증가할수록 손실 값이 감소하는 것을 확인했다. 그리고 CNN 기반의 모델 손실 감소 추세가 더 안정적임을 확인할 수 있다. 또한, 표 2과 같이 각 경우에 따른 시계열 데이터 분류 결과를 보였다. 정확도 지표는 아래 식 8과 같이 F1-Score를 사용했다.

$$F1\ Score = \frac{2}{\frac{TP+FP}{TP} + \frac{TP+FN}{TP}} \quad (8)$$



LSTM based Binary Classification

CNN based Binary Classification

그림 7. 시계열 데이터 기반 LSTM 분류 모델 손실 경과(좌), 시계열 이미지 기반 CNN 분류 모델 손실 경과(우)
 Fig. 7. LSTM classification model loss progression (left), CNN classification model loss progression (right)

여기서 TP(True Positive)는 참을 참이라고 올바르게 탐지한 것, FP(False Positive)는 거짓을 참이라고 잘못 탐지한 것, FN(False Negative)은 참을 거짓이라고 잘못 탐지한 것을 의미한다. 본 실험에서 주중 데이터와 주말 데이터 간의 불균형이 존재하므로 F1 Score를 사용했다.

실험 결과, 시계열 데이터의 패턴 분류 분야에서 RP 알고리즘으로 이미지화한 시계열 데이터를 활용한 CNN 기반 분류 모델의 성능이 LSTM 기반 분류 모델의 성능만큼을 보였고, GAF 기반 알고리즘으로 이미지화한 시계열 데이터 기반 모델은 RP 알고리즘 기반 모델과 비슷한 성능을 보였다. MTF 알고리즘으로 이미지화한 시계열 데이터 기반 모델이 가장 높은 정확도를 보였다.

V. 결론

본 논문에서는 시간에 따라 기록된 전력 소비 데이터의 패턴을 분석하기 위해 시계열 이미지를 사용하는 CNN 기반 분류 모델을 제안했다. 기존 연구에서는 시계열 데이터의 통계적 특성을 고려하기 위해 통계적 모델 혹은 RNN 기반의 지도 학습 모델을 사용하여 전력 소모 데이터를 분류했다.

표 2. 이미지화 알고리즘 별 제안된 모델의 정확도
Table 2. F1-score of the proposed models

Epoch	F1-Score				
	LSTM	CNN-RP	CNN-GASF	CNN-GADF	CNN-MTF
10	0.75	0.64	0.62	0.64	0.66
20	0.91	0.84	0.72	0.86	0.84
30	0.92	0.92	0.92	0.92	0.94
40	0.93	0.93	0.93	0.92	0.95
50	0.93	0.94	0.93	0.94	0.96

제안된 방법은 이진 분류에 특화된 CNN 기반의 지도 학습 모델을 사용하기 위해 시계열 데이터를 시계열 이미지 데이터로 변환하여 모델을 학습시킨다. 제안된 방법을 사용하기 위해 시계열 데이터를 이미지로 변환하는 방법을 4가지를 사용하여 성능을 비교했다.

RP, GASF, GADF, MTF 알고리즘을 사용하여 본 연구에서 쓴 전력 시계열 데이터를 이미지 형태로 변환하여 CNN 기반 이진 분류 모델을 사용하여 전력 소모 데이터를 주중 데이터와 주말 데이터로 분류했다. 기존 전력 소모 시계열 데이터만을 사용한 LSTM 기반 분류 모델의 성능보다 이미지화한 전력 소모 시계열 이미지 데이터를 사용한 CNN 기반의 이진 분류 모델의 성능이 더 우수했다. 특히 MTF 알고리즘

을 사용해서 구축한 시계열 이미지 데이터를 사용한 이진 분류 모델의 정확도가 0.96으로 가장 높게 도출됐다.

시계열 데이터를 분류하기 위해 데이터를 이미지화하여 CNN 계열의 지도 학습 모델을 사용하는 것이 원데이터를 사용한 RNN 계열의 지도 학습 모델을 사용하는 것과 동등하거나 그 이상의 성능을 보여줬다. 따라서 시계열 데이터의 이상 탐지를 위해 시계열 데이터의 차원을 증강하거나 이미지화하는 방안도 고려해볼 만하다. 향후 연구로는 RNN 계열의 지도 학습 모델로 LSTM뿐만 아니라 GRU (Gated Recurrent Unit), MGU (Minimal Gated Unit) 등 다양한 모델들을 사용하여 비교해보고, 본 논문에서 사용한 이미지화 방법뿐만 아니라 다른 차원 증강기법을 사용하여 데이터를 생성할 것이다.

감사의 글

본 연구는 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원(No. 2021-0-02068, 인공지능 혁신 허브 연구 개발)과 2022년도 한국전력공사의 재원으로 전력연구원(KEPRI) 연구개발과제의 지원을 받아 수행된 연구(과제번호 :R20IA02)로서, 관계부처에 감사드립니다.

참고문헌

- [1] H. Dirk, "Societal, economic, ethical and legal challenges of the digital revolution: from big data to deep learning, artificial intelligence, and manipulative technologies," *Towards digital enlightenment*. Springer, Cham, pp. 47-72, August 2018. https://doi.org/10.1007/978-3-319-90869-4_6
- [2] X. Fang, S. Misra, G. Xue, and D. Yang, "Smart grid—The new and improved power grid: A survey," *IEEE communications surveys & tutorials*, Vol. 14, No. 4, pp. 944-980, December 2011. <https://doi.org/10.1109/SURV.2011.101911.00087>
- [3] K. Frauendorfe, H. Glavitsch, and R. Bacher, *Optimization in Planning and Operation of Electric Power Systems: Lecture Notes of the SVOR/ASRO Tutorial Thun, Switzerland, October 14-16, 1992*, Springer Science & Business Media, 2013.
- [4] P. Chujai, N. Kerdprasop, and K. Kerdprasop, "Time series analysis of household electric consumption with ARIMA and ARMA models," *Proceedings of the international multiconference of engineers and computer scientists*, Hong Kong: IAENG, Vol. 1, pp-295-300, March 2013.
- [5] C. Zhong, J. Shao, F. Zheng, K. Zhang, H. Lv, and K. Li, "Research on electricity consumption behavior of electric power users based on tag technology and clustering algorithm," *2018 5th International Conference on*

- Information Science and Control Engineering (ICISCE)*, Zhengzhou, pp. 459-462, July 2018.
- [6] L. Huang, XL. Nguyen, M. Garofalakis, M. Jordan, A. Joseph, and N. Taft, "In-network PCA and anomaly detection," *Advances in neural information processing systems*, Vol. 19, 2006.
- [7] Y. Guo, W. Liao, Q. Wang, L. Yu, T. Ji, and P. Li, "Multidimensional time series anomaly detection: A gru-based gaussian mixture variational autoencoder approach," *Proceedings of The 10th Asian Conference on Machine Learning*, Vol. 95, pp. 97-112, November 2018.
- [8] P. Malhotra, A. Ramakrishnan, G. Anand, L. Vig, P. Agarwal, and G. Shroff, "LSTM-based encoder- decoder for multi-sensor anomaly detection," *arXiv preprint arXiv:1607.00148*, July 2016.
<https://doi.org/10.48550/arXiv.1607.00148>
- [9] J. P. Eckmann, S. O. Kamphorst, and D. Ruelle, "Recurrence plots of dynamical systems," *World Scientific Series on Nonlinear Science Series A*, Vol. 16, pp. 441-446, 1995.
<http://doi.org/10.1209/0295-5075/4/9/004>
- [10] H. Lee, J. Lee, and M. Shin, "Using wearable ECG/PPG sensors for driver drowsiness detection based on distinguishable pattern of recurrence plots," *Electronics*, Vol. 8, No. 2, February 2019.
<https://doi.org/10.3390/electronics8020192>
- [11] W. Zhiguang, and T. Oates, "Encoding time series as images for visual inspection and classification using tiled convolutional neural networks," *Workshops at the twenty-ninth AAAI conference on artificial intelligence*, Austin Texas, January 2015.
- [12] N. Marwan, and J. Kurths, "Cross recurrence plots and their applications," *Mathematical physics research at the cutting edge*, pp. 101-139, 2004.



윤준철(Junchurl Yoon)

1996년 : 숭실대학교 전자공학과 (공학사)
 2011년 : Alto대학교(핀란드) IB-MBA경영학 (석사)
 2021년~ 현재 : 전남대학교 ICT융합시스템공학과 박사과정

※ 관심분야 : 미디어처리, 상황인지 컴퓨팅, 스마트 워크



오상원(Sangwon Oh)

2020년 : 전남대학교 컴퓨터정보통신공학전공 (학사)
 2021년~ 현재 : 전남대학교 ICT융합시스템공학과 석사과정

※ 관심분야 : 딥러닝 및 머신러닝, 데이터 분석, 데이터 전처리, 인간-컴퓨터 상호작용



김진술(Jinsul Kim)

2001년 : Computer Science from University of Utah, Salt Lake City, Utah, USA (공학사)
 2005년 : 한국과학기술원 정보통신공학 (공학석사)
 2008년 : 한국과학기술원 정보통신공학 (공학박사)

2005년~2008년: 한국전자통신연구원 IPTV 인프라 기술, 융·복합 방송/통신 분야 연구원

2009년~2012년: 나사렛대학교 멀티미디어학과 교수

2012년~ 현재 : 전남대학교 전자컴퓨터공학과 교수

※ 관심분야 : QoS/QoE 예측/분석/관리, 모바일 미디어 처리/통신, 클라우드 컴퓨팅 디지털 미디어 및 네트워크 지능기술