

## 정확도 향상을 위한 이미지 변환 기반의 시계열 데이터 증강 기법

최문규, 신수용

금오공과대학교 IT융복합공학과

eo97@kumoh.ac.kr, wdragon@kumoh.ac.kr

## A Time Series Data Augmentation Technique Based on Image Conversion to Improve Accuracy

Mun Kyu Choi, Soo Young Shin

Dept. of IT Convergence Engineering

Kumoh National Institute of Technology

## 요 약

대부분의 딥러닝 모델의 성능은 학습 데이터의 양과 질에 의해 좌우된다. 그러나, 일반적으로 질 좋고 많은 데이터 셋을 모으는 것은 많은 시간과 비용이 소비된다. 이미지 데이터를 사용한 딥러닝의 경우 데이터 증강 기법을 사용하여 이 문제점을 극복한다. 하지만 시계열 데이터를 사용한 딥러닝의 경우에는 시간에 종속적이라는 특성 때문에 일반적인 데이터 증강 기법이 제한된다. 본 논문에서는 이러한 문제를 해결하기 위해서 시계열 데이터를 이미지화 하여 데이터를 증강하고 이에 따른 정확도 향상을 검증한다.

## I. 서 론

일정한 시간동안 수집된 데이터들로 구성된 수열인 시계열 데이터는 의학, 기상학, 경제학, 천문학 등 다양한 분야에서 사용된다 [1]. 시계열 데이터는 연속된 관측 내에서 발견한 특정한 법칙성을 바탕으로 과거 데이터로부터 미래의 정보를 예측할 수 있을 뿐만 아니라, 클래스 분류, 비정상 데이터 감지 등도 가능하다. 시계열 데이터 또한 이미지로 변환해서 대표적인 딥러닝 네트워크인 CNN을 통해 분석하는 다양한 연구가 이루어졌고 좋은 성능을 보여주었다 [2, 3].

컴퓨터 비전에서 좋은 성능을 내기 위해서는 충분한 양의 데이터 셋이 바탕이 되어야 한다. 그래서 주로 데이터 증강이라는 기법이 사용된다. 하지만 시계열 데이터는 다르다. 충분한 데이터 확보에 많은 시간과 비용이 소요되고 시간에 종속적이라는 시계열의 특성 때문에 데이터 증강 기법도 제한된다.

본 논문에서는 충분한 양의 시계열 데이터를 확보하기 위해 3가지 기법의 시계열 데이터 이미지화를 사용하고 데이터를 증강하여 분류기의 정확도를 향상하는 방법을 제안한다. 그 후 증강된 데이터에서 추가로 이미지에 대한 데이터 증강 기법을 사용한 실험을 통해 데이터 양에 따른 정확도 향상에 대한 결과도 검증한다.

## II. 시계열 데이터 이미지화

시계열 데이터 이미지화(Imaging Time Series)는 시계열 데이터를 이미지로 변환하는 기술이다. Imaging Time Series 기술은 시계열 데이터 특징을 재구성하여 시각화가 가능하게 한다. 본문에서는 3가지 Imaging Time Series 기술(Recurrence plot, Gramian Angular Field, Markov Transition Field)을 사용하여 분류기 성능을 개선 및 비교하였다.

## 1. Recurrence plot

Recurrence plot(RP)은 비선형 데이터를 분석하는데 제안되는 기술이다 [3]. 데이터 값의 회귀에 대하여 2차원에 표시하고 m차원 위상 공간 궤도를 탐색하는 것을 목표로 한다. RP는 식 (1)과 같이 나타낼 수 있다.

$$RP_{i,j} = \theta(\epsilon_i - \|\vec{x}_i - \vec{x}_j\|)$$

$$\vec{x}_i \in \mathbb{R}^m, i, j = 1, \dots, N$$

여기서  $N$ 은 다른 시간  $j$ 에서 시간  $i$ 에서의 상태  $x_i$ 의 개수이다.  $\epsilon_i$ 는 임계 값 거리이고  $\|\cdot\|$ 는 norm 그리고  $\theta(\cdot)$ 는 단위계단함수를 의미한다.

## 2. Gramian Angular Field

Gramian angular field(GAF)는 시계열 데이터에 대한 시간적 상관관계를 생성하는 gramian 행렬을 사용하는 극 좌표 시스템이다. 극 좌표에 기반한 행렬은 시계열 데이터를 이미지화 할 경우 상관관계를 보존할 수 있다. GAF는 식(2)와 같이 나타낼 수 있다.

$$\tilde{x}_i = a + (b - a) \times \frac{x_i - \min(X)}{\max(X) - \min(X)}$$

$$GAF_{i,j} = \cos(\phi_i + \phi_j), i, j = 1, \dots, N$$

여기서 시계열 데이터는  $-1 \leq a < b \leq 1$  범위 안에서  $[a, b]$ 로 정규화 한다. 그리고 각  $\phi_i = \arccos(\tilde{x}_i)$ 는  $-1 \leq \tilde{x}_i \leq 1$ 의 범

위를 가진다.  $X = \{x_1, x_2, \dots, x_N\}$ 는 주어진  $N$ 개 관측치의 시계열을 나타낸다.

### 3. Markov Transition Field

Markov transition field (MTF)는 시간 영역에서 markov 전이 확률을 순차적으로 나타내는 동적 전이 통계이다. MTF는 식 (3)과 같이 나타낼 수 있다.

$$MTF_{i,j} = w_{ij|vertx_i \in q_i, x_j \in q_j}, i, j = 1, \dots, N$$

여기서  $w_{i,j}$ 는  $q_j$ 의 점이  $\sum_j w_{u,j} = 1$ 에서  $q_i$ 의 점이 이어지는 빈도로 주어진다.

### III. 실험

정확도 향상 검증에 위해서 UCR 시계열 데이터 셋을 활용했다. 그 중 2가지 클래스로 구성된 Wafer 시계열 데이터를 선택했고 훈련데이터의 양은 1000개이다. 그래프 스냅샷에서 3가지 Imaging Time Series 기법을 사용해 이미지로 변환한 뒤 YOLOv5 classification에서 학습을 진행했다. Classification을 위해 사용한 사전학습 모델은 yolov5s-cls.pt를 사용했다. 훈련은 이미지 크기 224, epoch 30으로 진행했다.

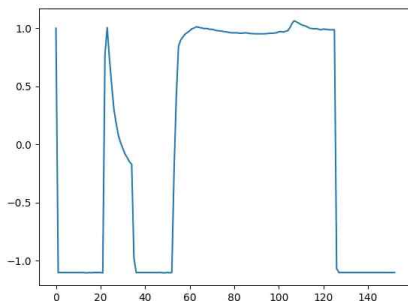


그림 1. 그래프 스냅샷



그림 2. GAF, MTK, RP를 사용하여 변환된 이미지

	스냅샷	RP	GAF	MTF
이미지 개수	550개	1,000개	1,000개	1,000개
정확도	0.546	0.994		

표 1 데이터 셋 개수 및 정확도

위의 그림1, 2와 같이 시계열 데이터를 이미지로 변환하여 훈련을 진행했고 테스트 데이터 6000장에 대해 정확도를 검증했다. 그래프 스냅샷으

로 훈련 데이터는 테스트 데이터에 대해 0.546의 정확도를 보였다. 3가지 Imaging Time Series기법을 통해 3000장으로 데이터가 증가된 데이터는 테스트 데이터에 대해 0.994의 정확도를 보였다. 또 추가로 데이터 양에 따른 정확도 향상을 검증하기 위해서 변환된 이미지에 데이터 증강 기법을 적용해서 데이터의 양을 늘렸다. 데이터에 노이즈를 추가해서 7000장으로 늘렸고 데이터는 0.997의 정확도를 보였다.

### IV. 결론

본 논문에서는 시계열 데이터의 정확도 향상을 위한 데이터 증강 기법을 제안했고 데이터 증강에 따른 정확도 향상을 검증했다. 데이터 양에 따라 정확도가 향상됨을 확인했다. 향후에는 변환된 이미지에 효과적인 데이터 증강기법을 적용하여 적은 수의 시계열 데이터에서 정확도 향상을 검증한다.

### ACKNOWLEDGMENT

이 논문은 2022년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2018R1A6A1A03024003)

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신 인재양성(Grand ICT연구센터) 사업의 연구 (IITP-2023-2020-0-01612) 및 2021년도 정부(교육과학기술부)의 재원으로 한국연구재단의 대학중점연구소 지원사업으로 수행된 연구임(2018R1A6A1A03024003)

### 참 고 문 헌

- [1] <http://www.bosa.co.kr/news/articleView.html?idxno=2173856>
- [2] Ismail Fawaz, H., Forestier, G., Weber, J. et al. Deep learning for time series classification: a review. Data Min Knowl Disc 33, 917 - 963 (2019).
- [3] 박지연, 서동호 and 남해운. (2021). 이미지화 알고리즘 및 딥러닝을 이용한 자동 변조 분류. 한국전자과학회 논문지, 32(4), 328-333.