

고차원 매핑기법과 딥러닝 네트워크를 통한 정형데이터의 분류

김경택¹, 장원두^{2*}

¹부경대학교 인공지능융합학과 학생, ²부경대학교 컴퓨터·인공지능공학부 교수

Classification of Tabular Data using High-Dimensional Mapping and Deep Learning Network

Kyeong-Taek Kim¹, Won-Du Chang^{2*}

¹Student, Artificial Intelligence Convergence, Pukyong National University

²Professor, Division of Computer Engineering and Artificial Intelligence, Pukyong National University

요약 최근 딥러닝은 다양한 분야에서 전통적인 기계학습에 비해 월등히 높은 성능을 보이고 있으며, 패턴인식을 위한 보편적인 방법으로 자리 잡아 가고 있다. 하지만, 이에 비해 정형데이터를 사용하는 분류 문제에서는 여전히 머신러닝 기법이 주류를 이루고 있다. 본 논문에서는 정형데이터를 고차원 텐서로 변환하는 네트워크 모듈을 제안하며, 이 모듈을 보편적인 딥러닝 네트워크와 함께 구성하여 정형데이터의 분류 문제에 적용하였다. 제안된 방법은 4종의 데이터셋을 활용하여 학습 및 검증되었으며, 제안된 방법은 90.22%의 평균 정확도를 달성하여, 최신 딥러닝 모델인 TabNet에 비해 2.55%p 높은 정확도를 보였다. 제안된 방법은 컴퓨터 비전 분야에서 높은 성능을 보이는 다양한 네트워크 구조를 정형데이터에 활용할 수 있다는 점에서 의미가 있다.

주제어 : 딥러닝, 정형데이터, 머신러닝, 합성곱 신경망, 패턴 분류

Abstract Deep learning has recently demonstrated conspicuous efficacy across diverse domains than traditional machine learning techniques, as the most popular approach for pattern recognition. The classification problems for tabular data, however, are remain for the area of traditional machine learning. This paper introduces a novel network module designed to tabular data into high-dimensional tensors. The module is integrated into conventional deep learning networks and subsequently applied to the classification of structured data. The proposed method undergoes training and validation on four datasets, culminating in an average accuracy of 90.22%. Notably, this performance surpasses that of the contemporary deep learning model, TabNet, by 2.55%p. The proposed approach acquires significance by virtue of its capacity to harness diverse network architectures, renowned for their superior performance in the domain of computer vision, for the analysis of tabular data.

Key Words : Deep Learning; Tabular data, Machine Learning, CNN, Pattern Classification

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 지역지능화혁신인재양성(Grand ICT연구센터) 사업 (IITP-2023-2016-0-00318)과 과학기술정보통신부의 재원으로 수행된 연구개발특구진흥재단-기술사업화 역량강화 사업의 지원을 받아 수행된 연구임.(No. 2023-BS-RD-0061 / 지능형 보안 감시 시스템 고도화 및 상용화 기술 개발)

*교신저자 : 장원두(chang@pknu.ac.kr)

접수일 2023년 10월 26일 수정일 2023년 11월 22일 심사완료일 2023년 11월 24일

1. 서론

여러 개의 네트워크층으로 구성된 딥러닝 네트워크는 인공지능망의 대표적인 모델로써 다양한 분야에서 활용되고 있다. 딥러닝은 이미지넷[1] 등과 같은 거대 공개 데이터셋을 통해 그 성능이 빠르게 향상되었으며[2], 영상인식[3,4], 시계열 데이터 분석[5], 자연어 처리[6,7] 등과 같이 비정형 데이터를 활용하는 최신 인공지능 모델들은 딥러닝을 활용하는 경우가 대다수다.

이와 대조적으로, 정형데이터의 경우에는 아직 전통적인 머신러닝 기법들이 딥러닝 네트워크에 비해 더 많이 사용되고 있다[8]. 정형데이터는 행과 열의 표 형태로 구성되는 데이터로써[9], 서로 다른 종류의 데이터가 함께 섞여 있는 경우가 많다[10]. 이로 인해 비정형 데이터에 비해 특징 간의 연관성이 상대적으로 낮으며[10], 부스팅 알고리즘과 같은 전통적인 머신러닝 기법들이 딥러닝에 비하여 좋은 성능을 나타내는 경우가 많다[9,11].

하지만 최근, TabNet[8], SAINT[12] 등과 같이 정형데이터의 분석과 분류 등을 위한 딥러닝 네트워크가 제안되었으며, 데이터에 따라 전통적인 방법들에 비해 더 높은 성능을 보이는 경우가 있음이 입증되었다. TabNet은 순차적 어텐션을 사용하여 각 결정 단계에서 추론할 특징을 선택하는 방식으로 특징별 중요도를 산출하며 학습을 진행한다. Forest Cover Type, Poker Hand 등의 데이터셋에서 머신러닝 기법들과 비교하여 96.99%와 99.2% 등 높은 정확도를 보였다[8]. SAINT는 정형데이터의 행과 열에 대하여 어텐션을 수행하는 방법으로 행과 열에 대하여 통합 어텐션 점수를 계산하며, 계산된 점수를 통하여 고려할 특징을 최적화한다. 고객 정보를 통해 잔존 여부를 예측하는 Blastchar, 심부정맥 유무를 구별하여 아웃라이어(outlier)와 인라이어(inlier)로 분류하는 Arrhythmia 등 총 16개의 데이터셋에서 SAINT가 가장 높은 평균AUROC (93.13%)을 기록하였다[12].

정형데이터의 분석에 딥러닝 네트워크를 활용하기 위해 정형데이터를 이미지 형식으로 변환하여 활용하는 연구들도 있다. 데이터를 2/3 차원 행렬 형태로 변환한다면 이미지 분석 등을 위해 개발 및 사전 학습되어 공개된 다양한 딥러닝 네트워크 모델을 활용할 수 있으므로, 비교적 쉽게 다양한 모델에 적용 가능하다는 장점이 있다. Buturović 과 D. Miljković은 정형데이터를 임의의 이미지와 합성하여 데이터를 이미지로 변형시키는 방법을 제안하였으며, 박테리아 또는 바이러스 감염 환자 혈액 분류 문제에서 90%의 정확도를 기록하였다[13]. Iqbal

등은 정형데이터를 2차원 공간상에 차례대로 배치하는 방법을 제안하였다[14]. 이때, 통계적 기법을 활용하여 각 특징의 가중치를 계산하고, 해당 가중치를 기반으로 이미지 공간에 특징을 할당하는 방법이 사용되었는데, 총 6개의 데이터셋에서 98%의 정확도를 보였다. Zhu 등은 특징간의 유사도를 기반으로 유사도가 높은 특징들이 가까이 배치될 수 있도록 이미지를 생성하는 방법을 제안하였다[15]. 이 방식을 통해 약물 스트리밍 데이터셋에서 $R^2=0.856$ 으로 가장 높은 성능을 보였다.

한편, 정형데이터에 대한 딥러닝 기법들의 성능에 대해서는 여전히 다양한 논의들이 존재한다. Shwartz-Ziv 등은 TabNet이 더 높은 성능을 보이는 데이터는 매우 제한적이며, XGBoost[16]와 같은 전통적인 기법들이 더 높은 성능을 보이는 경우가 많다고 주장하였다[9].

본 연구는 정형데이터를 고차원 배열로 변환하는 네트워크 모듈을 제안하고, 제안된 네트워크를 사용하여 널리 알려진 데이터셋들의 분류 정확도를 비교 및 분석하였다. 제안된 방법은 정형데이터를 이미지 형식으로 변환한다는 점에서 Zhu 등의 연구[15]와 유사하지만, Zhu 등은 유사도를 직접적으로 계산하였던 반면, 본 연구는 네트워크의 학습 과정을 통해 최적의 고차원 텐서를 만들어 낸다는 점에서 기존 방법과 다르다.

이 논문의 구성은 다음과 같다. 2장에서는 실험에 사용된 데이터와 학습 시스템의 구성, 딥러닝 네트워크의 구조에 대하여 기술하였다. 3장에서는 제안된 방법의 성능을 기존의 네트워크들과 비교하여 성능 검증을 진행하였다. 마지막으로 4장에서는 결론 및 향후 연구에 관하여 기술하였다.

2. 분석 방법

2.1 데이터

본 연구에서는 정형데이터 분석에서 널리 사용되는 데이터셋인 Vehicle(UCI Vehicle silhouettes), Landsat(UCI Landsat Satellite), Segment(UCI Image Segmentation), CovType(Forest Cover Type)의 4가

〈Table 1〉 Specification of datasets

	Vehicle	Landsat	Segment	CovType
No. Features	18	36	19	54
No. Categories	4	6	7	7
No. Instances	946	6435	2310	581012

지 데이터셋을 사용하여 모델을 평가하였다. 각 데이터셋의 특징, 카테고리, 인스턴스 수 등은 Table 1과 같다. Vehicle은 차량의 실루엣에서 추출된 특징들을 순차적으로 나열한 데이터다. 오펔 만타 400, 사브 9000, 이층 버스, 쉐보레 밴 등 4개 차량의 실루엣 특징 데이터가 각각 240, 240, 240, 226개씩 총 946개의 인스턴스로 구성되어 있다. Landsat은 Landsat 다중 스캐너 이미지에서 추출된 데이터셋으로 위성 이미지의 다중 스펙트럼 값을 기록한 데이터다. 적색도, 목화, 회색도, 습한 회색도, 짙이 있는 토양, 매우 습한 회색도 인스턴스가 각각 1533, 703, 1358, 626, 707, 1508개로 총 6435개의 인스턴스로 구성되어 있다.

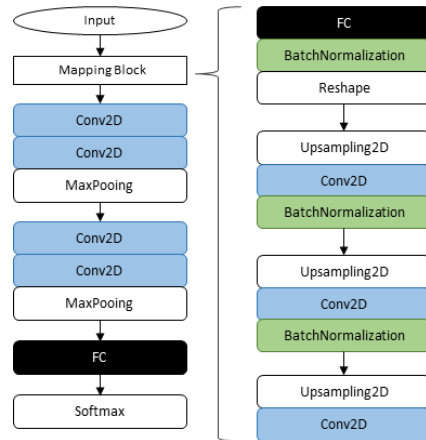
Segment는 야외에서 촬영된 7장의 사진에서 3×3 의 이미지 그리드를 랜덤하게 추출한 후, 추출된 이미지 그리드에서 계산된 19개의 특징 데이터다. 이미지 그리드의 특징으로부터 원본 이미지가 무엇인지 예측하는 것을 목표로 한다. Brickface, Sky, Folige, Cement, Window, Path와 Grass로 명명된 7개 이미지로부터 추출된 인스턴스가 각 330개씩 있으며, 총 2310개의 인스턴스로 구성되어 있다.

CovType은 토양 유형, 기후 등에 대한 54개의 특징으로 구성된 데이터다. 이들 특징을 사용하여 숲을 구성하는 수목의 종류, 즉 산림 피복유형(Forest Cover Type)을 예측하기 위한 목적으로 개발되었다. 데이터는 7개 피복 유형(가문비나무, 로지폴 소나무, 폰테로사 소나무, 미루/버드나무, 사시나무, 더글라스퍼, 왜성편형수림 등)으로 나누어져 있으며, 총 581012개의 데이터로 구성되어 있다.

2.2 네트워크 모델

본 연구에서 사용된 딥러닝 모델의 전체적인 구조는 Fig 1과 같다. 제안된 모델은 Mapping Block과 합성곱 층, 완전 연결 층이 직렬로 연결된 구조이며, Mapping Block은 1차원 벡터를 2차원 공간으로 사상(mapping)하는 역할을, 합성곱 층은 2차원 공간의 벡터를 다시 1차원 벡터로 변환하는 역할을 한다.

본 연구에서 제안된 Mapping Block은 3개의 합성곱 층과 3개의 업샘플링(Upsampling) 층 그리고 1개씩의 Reshape 층과 완전 연결 층으로 구성된다. Mapping Block에 입력된 데이터는 완전 연결 층을 통해 동일 차원에서 변형되며, Reshape 층에서 2차원 형식으로 변환된 후, Upsampling 층과 합성곱 층을 통해 3차원으로 매핑(mapping)된다. 이 과정에서 배치 정규화 층이 학습의 안정성을 높이기 위해 활용되었으며, 마지막 합성



[Fig. 1] Proposed network structure with Mapping Block.

곱 층의 필터 개수는 1로 조정하여 Mapping Block의 출력이 2차원의 이미지 형태를 띠도록 하였다. 이를 통해 Mapping Block의 동작 형태를 이미지의 형태로 시각화할 수 있다.

Mapping Block의 출력은 2차원 배열 형태이므로, Mapping Block은 이미지 분류 등에 사용되는 각종 네트워크 모델과 쉽게 결합할 수 있다. 본 연구에서는 이미지 분류에 널리 사용되는 직렬(sequential) 구조의 모델을 결합하여 실험을 수행하였다. 2차원 합성곱 층 2개와 MaxPooling 층 1개로 구성된 Block을 2개 사용하였고, 완전 연결 층을 사용하여 데이터를 분류하도록 하였다.

2.3 데이터 전처리와 검증방법

실험에 앞서 각 특징 데이터는 min-max 정규화 기법을 사용하여 정규화되었다. X 가 특징의 값, X_{min} 과 X_{max} 가 각각 해당 특징 데이터의 최솟값과 최댓값일 때, 데이터 정규화를 위한 수식은 다음과 같다.

$$X_{normalization} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (1)$$

제안된 모델의 검증에는 10 폴드(fold) 검증 기법이 사용되었다. 즉, 데이터셋을 10개의 폴드로 분할한 후, 9개 폴드로 학습된 모델을 나머지 1개의 폴드를 사용하여 테스트한다. 이 과정을 10회 반복하여 모든 데이터가 테스트에 사용되도록 하였으며, 학습 데이터의 선정은 각 정답 값에 대하여 독립적으로 이루어졌다. 예를 들어 Vehicle 데이터셋의 경우, 4개 차량 데이터에 대해 각각 9:1로 데이터를 분리하여 사용하였다.

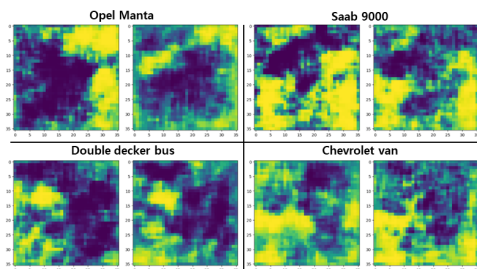
<Table 2> Classification accuracy for each dataset

Model	Accuracy(%) according to Datasets				
	Vehicle	Landsat	Segment	Cov Type	Avg.±stdev
MLP	79.3	84.9	93.8	81.08	84.77±6.46
XGBoost	77.28	92.06	98.26	89.34	89.24±8.8
TabNet[8]	81.44	89.44	82.8	96.99	87.67±7.13
Proposed	83.5	88.4	95.4	93.58	90.22±5.37

3. 결과

제안된 모델의 정확도는 Table 2와 같다. 제안된 모델의 평균 정확도는 90.22%로 가장 높은 성능을 보였으며, 이것은 두 번째로 높은 평균 정확도를 보인 XGBoost보다 0.98%p 높은 것이다. 제안된 모델은 Vehicle, Landsat, Segment, CovType 데이터셋에 대해 각각 83.5%, 88.4%, 95.4%, 93.58%, 90.22%의 정확도를 달성하였다.

본 연구에서, MLP를 제외한 다른 3개 모델은 서로 다른 측면에서 그 강점이 드러났다. MLP는 평균 정확도 84.77%로 가장 낮은 성능을 보였으며, Vehicle, Landsat, Segment, CovType 데이터셋에 대해 각각 3위, 4위, 3위, 4위에 위치하였다. XGBoost는 두 번째로 높은 평균 정확도를 보였으나, 정확도의 편차가 다른 모델에 비해 높게 나타났다. Segment와 Landsat에서는 각각 98.26%, 92.06%로 1위에 위치하였으나, CovType에서는 3위, Vehicle에서는 4위에 위치하였다. 딥러닝 모델을 사용하는 TabNet[8]은 CovType에서 96.99%로 가장 높은 정확도를 보였으나, 다른 데이터셋에서 상대적으로 낮은 정확도를 보이며, 평균 정확도 3위에 위치하였다. 이 결과는 딥러닝 모델이 데이터셋에 따라 XGBoost 등의 전통적인 방법들에 비해 성능이 낮게 나타나는 경우가 많다는 Shwartz-Ziv와 Armon의 연구[9]와도 동일한 결론을 도출한다.



[Fig. 2] Visualization of Mapping Block layers

<Table 3> Training time according to datasets

	Vehicle	Landsat	Segment	CovType
TabNet	383.66	181.35	149.64	25260.88
Proposed	26.67	140.12	6.39	1,524.45

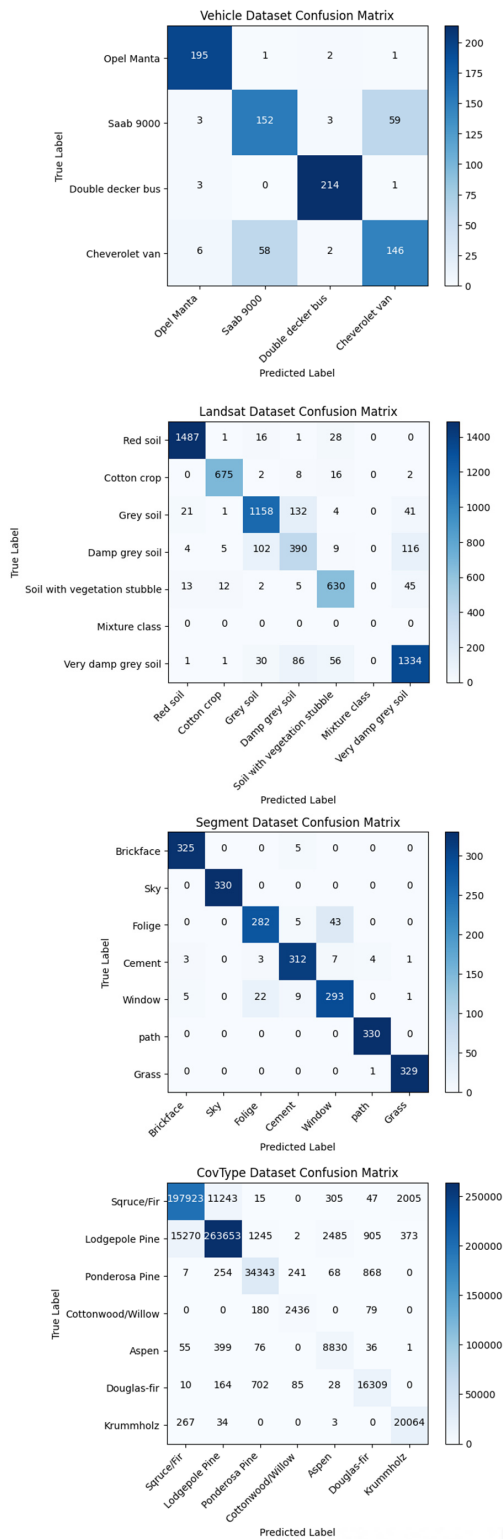
반면, 제안된 방법은 Vehicle 데이터에서 1위, Segment와 CovType 데이터셋에서 2위, Landsat 데이터셋에서 3위에 위치하였으며, 1위를 차지하지 못한 경우에도 1위와의 편차가 각각 2.86%p(Segment), 3.41%p(CovType), 3.66%p(Landsat)로 비교적 낮은 편이었다.

각 모델의 성능은 모델의 레이어 수, 필터 개수 등과 밀접한 관련이 있다. 따라서, 본 연구에서는 각 모델(MLP, XGBoost, TabNet, 제안된 방법 등)의 네트워크 크기를 그리드 탐색을 통해 다양하게 변화시키면서 10폴드 검증을 수행한 후, 가장 높은 성능을 달성한 모델의 결과를 Table 2에 표시하였다. 딥러닝 모델의 경우 학습률, 레이어 수, 배치 크기가 하이퍼파라미터 최적화에 포함되었으며, 비교 모델들의 초기 하이퍼파라미터는 기존 논문을 바탕으로 지정되었다.

Mapping Block을 활용한 제안하는 모델과 TabNet의 학습시간은 Table 3과 같다. 두 모델 모두 Covtype에서 학습시간이 길었으며, 제안된 모델은 수행속도 면에서 TabNet보다 월등히 높은 성능을 보였다.

한편, 실험에서 사용된 Mapping Block의 output은 2차원 형태를 가지므로 레이어의 상태를 이미지로 시각화하여 분석할 수 있다. Fig. 2는 Vehicle 데이터셋의 Mapping Block 출력을 카테고리별로 시각화한 것이다. 오펀 만타(Opel Manta)와 2층버스(Double decker bus)는 다른 그룹의 데이터와 인지적으로 다른 형태를 가지고 있음을 쉽게 알 수 있으며, 사브(Saab) 9000과 쉐보레 밴(Chevrolet van)은 상대적으로 유사한 형태의 이미지 구조를 가진다.

Fig. 3은 Vehicle, Landsat, Segment, Covtype 데이터셋의 테스트 데이터를 사용하여 예측 및 계산된 혼동행렬이다. Vehicle 데이터셋에서는 사브 9000과 쉐보레 밴 사이에서 대부분의 오인식이 발생하였는데, 이는 Mapping Block의 출력상에서 사브 9000과 쉐보레 밴이 유사한 형태로 변환된 것과 연관성이 높은 것으로 보인다(Fig. 2). Landsat 데이터셋에서는 “매우 습한 회색토(Very damp grey soil)”에서 가장 많은 양의 오인식이 발생하였으나, 이것은 해당 카테고리의 인스턴스 수가 다른 카테고리의 인스턴스 수에 비해 많기 때문이다. 통상적으로 혼동행렬의 오인식 수는 전체 인스턴스



수에 비례하며, 실제 오인식률은 약 1.40% 수준으로 매우 낮다. Covtype 데이터셋의 혼동행렬에서는 가문비나무(Spruce/Fir)와 로지폴 소나무(Lodgepole Pine) 사이에서 가장 많은 오인식(10.69%)이 발생하였으며, 미루/버드나무(Cottonwood/Willow)를 폰테로사 소나무(Ponderosa Pine)로 오인식한 경우가 6.68%, 사시나무(Aspen)를 로지폴 소나무로 오인식한 경우가 4.25%로 나타났다.

4. 결론

본 논문에서는 정형데이터의 인식을 위해 1차원의 입력 특징 벡터를 고차원으로 매핑하는 Mapping Block 모델을 제안하였으며, 4가지 데이터셋을 사용하여 전통적인 방법 및 최신 딥러닝 네트워크의 성능과 비교 및 분석하였다. 제안된 방법은 평균 90.22%의 정확도를 보였으며, 이는 최신 딥러닝 네트워크인 TabNet의 결과에 비해 2.55%p, XGBoost에 비해서도 0.98%p 높은 것이다. 특히, 제안된 네트워크는 5.37%의 가장 낮은 표준편차를 보이며 데이터셋 전반에 대해 높은 성능을 보였다.

다만, 제안된 방법은 Vehicle 데이터셋을 제외한 다른 개별 데이터셋에서 가장 높은 성능을 보이지는 못하였으며, 4개의 데이터셋을 통해서 제한적으로 검증되었다. 제안된 방법은 분류하는 클래스의 수가 적은 경우 (Vehicle)에만 높은 성능을 보인 것으로 이해해 볼 수 있는데, 이것은 전반적인 분류 성능을 향상시키므로 접근하여야 할 것으로 여겨진다. 향후 연구에서는 매핑 블록의 출력 차원을 3차원 이상의 고차원으로 변경하거나 손실함수를 새롭게 디자인하여, 많은 클래스가 제한된 공간 안에서 충분한 차이를 나타낼 수 있도록 하는 방법을 고려해 볼 수 있다. 또한, 추가적인 데이터셋을 통한 실험을 수행하고 ResNet, 비전 트랜스포머 등과 같이 널리 사용되는 백본(back-bone) 네트워크 등과 결합하는 등의 실험을 통해, 제안된 모델의 성능을 추가로 향상하고자 한다.

한편, 제안된 방법은 학습의 중간 과정을 시각적으로 확인해 볼 수 있는 특성이 있으며, 고차원 매핑 결과와 분류 성능의 연관성을 고려해 볼 때, 향후 데이터가 학습되는 과정을 이미지로 가시적으로 표현하는 설명 가능한 AI의 일환으로도 사용될 수 있을 것으로 여겨진다.

[Fig. 3] Confusion matrix for each dataset

REFERENCES

- [1] A.Krizhevsky, I.Sutskever and G.E.Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," Advances in Neural Information Processing Systems 25, 2012.
- [2] C.Szegedy, W.Liu, Y.Jia, P.Sermanet, S.Reed, D.Anguelov, D.Erhan, V.Vanhoeck and A.Rabinovich, "Going Deeper with Convolutions," Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp.1-9, 2015.
- [3] A.Dosovitskiy, L.Beyer, A.Kolesnikov, D.Weissenborn, X.Zhai, T.Unterthiner, M.Dezhghani, M.Minderer, G.Heigold, S.Gelly, J.Uzkoreit and N.Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," International Conference on Learning Representations, 2021.
- [4] K.He, X.Zhang, S.Ren and J.Sun, "Deep Residual Learning for Image Recognition," Computer Vision and Pattern Recognition, pp.770-778, 2016.
- [5] B.Lim, S.Ö.Arik, N.Loeff and T.Pfister, "Temporal Fusion Transformers for Interpretable Multi-Horizon Time Series Forecasting," International Journal of Forecasting, Vol.37, No.4, pp.1748-1764, 2021.
- [6] J.Devlin, M.W.Chang, K.Lee and K.Toutanova. "BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding," Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Vol.1, 2019.
- [7] A.Vaswani, N.Shazeer, N.Parmar, J.Uzkoreit, L.Jones, A.N.Gomez, L.Kaiser and L.Polosukhin, "Attention Is All You Need," Advances in Neural Information Processing systems 30, pp.5998-6008, 2017.
- [8] S.Ö.Arik and T.Pfister, "Tabnet: Attentive Interpretable Tabular Learning," Proceedings of the AAAI Conference on Artificial Intelligence, Vol.35, No.8, pp.6679-6687, 2021.
- [9] R.Shwartz-Ziv and A.Armon, "Tabular Data: Deep Learning is Not All You Need," Information Fusion, Vol.81, pp.84-90, 2022.
- [10] V.Borisov, T.Leemann, K.Seßler, J.Haug, M.Pawelczyk and G.Kansneci, "Deep Neural Networks and Tabular Data: A Survey," IEEE Transactions on Neural Networks and Learning Systems, 2022.
- [11] I.Shavitt and E.Segal, "Regularization Learning Networks: Deep Learning for Tabular Datasets," International Conference on Neural Information Processing Systems 31, pp.1379-1389, 2018.
- [12] G.Somepalli, M.Goldblum, A.Schwarzschild, C.B.Bruss and T.Goldstein, "SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training," arXiv:2106.01342, 2021.
- [13] L.Buturović and D.Miljković, "A Novel Method for Classification of Tabular Data Using Convolutional Neural Network," BioRxiv, 2020.
- [14] M.I.Iqbal, S.H.Mukta and A.R.Hasan, "A Dynamic Weighted Tabular Method for Convolutional Neural Networks," IEEE Access, 10, pp.134183-134198, 2022.
- [15] Y.Zhu, T.Brettin, F.Xia, A.Partin, M.Shukla, H.Yoo, Y.A.Evard, J.H.Doroshov and R.L.Stevens, "Converting Tabular Data into Images for Deep Learning With Convolutional Neural Networks," Scientific Reports 11.1, 2021.
- [16] T.Chen and C.Guestrin, "XGBoost: A Scalable Tree Boosting System," The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785-794, 2016.

김 경 택(Kyeong-Taek Kim)

[준회원]



- 2022년 8월 : 부경대학교 컴퓨터공학과 (공학사)
- 2022년 9월 ~ 현재 : 부경대학교 인공지능융합학과 석사과정

〈관심분야〉

딥러닝, 패턴인식, 머신러닝, 정형데이터, 컴퓨터비전

장 원 두(Won-Du Chang)

[정회원]



- 2003년 2월 : 부산대학교 정보컴퓨터공학부 (공학사)
- 2005년 2월 : 부산대학교 컴퓨터공학과 (공학석사)
- 2011년 3월 : Aizu Univ. (일) Info. Sys. Dept. (컴퓨터이공학박사)
- 2011년 8월 ~ 2013년 6월 : 몽골국제대학 IT학부 조교수
- 2017년 3월 ~ 2020년 2월 : 동명대학교 전자및의용공학부 조교수
- 2020년 3월 ~ 현재 : 부경대학교 인공지능융합학과 조교수/부교수

〈관심분야〉

패턴인식, 신호처리, 딥러닝, 영상처리, 헬스케어