

설명가능한 의사결정을 위한 마이닝 기술

정경용

경기대학교 AI컴퓨터공학부

Research on Mining Technology for Explainable Decision Making

Kyungyong Chung

Division of AI Computer Science and Engineering, Kyonggi University, South Korea

요약 데이터 처리 기술은 의사결정을 위해 중요한 역할을 하며, 데이터 결측값 및 이상값 처리, 예측, 추천 모델 등이 포함된다. 이는 모든 과정과 결과의 타당성, 신뢰성, 정확성에 대한 명확한 설명이 필요하다. 또한 의사결정트리, 추론 등을 이용한 설명가능한 모델을 통해 데이터의 문제를 해결하고, 다양한 유형의 학습을 고려하여 모델 경량화를 진행할 필요가 있다. 육하원칙을 적용한 다중 계층 마이닝 분류 방법은 데이터 전처리 후 트랜잭션에서 빈번하게 발생하는 변수와 속성 간의 다차원 관계를 발견하는 방법이다. 이는 트랜잭션에서 마이닝을 이용하여 유의미한 관계를 발견하고, 회귀분석을 통해 데이터를 모델링하는 방법을 설명한다. 이에따라 확장 가능한 모델과 로지스틱 회귀모델을 개발하고, 데이터 정제, 관련성 분석, 데이터 변환, 데이터 증강을 통해 클래스 레이블을 생성하여 설명가능한 의사결정을 위한 마이닝 기술을 제안한다.

• 주제어 : 설명가능한 인공지능, 데이터마이닝, 딥러닝, 머신러닝, 인공지능

Abstract Data processing techniques play a critical role in decision-making, including handling missing and outlier data, prediction, and recommendation models. This requires a clear explanation of the validity, reliability, and accuracy of all processes and results. In addition, it is necessary to solve data problems through explainable models using decision trees, inference, etc., and proceed with model lightweight by considering various types of learning. The multi-layer mining classification method that applies the sixth principle is a method that discovers multidimensional relationships between variables and attributes that occur frequently in transactions after data preprocessing. This explains how to discover significant relationships using mining on transactions and model the data through regression analysis. It develops scalable models and logistic regression models and proposes mining techniques to generate class labels through data cleansing, relevance analysis, data transformation, and data augmentation to make explanatory decisions.

• Key Words : Explainable AI, Data Mining, Deep Learning, Machine Learning, Artificial Intelligence

Received 23 November 2023, Revised 25 December 2023, Accepted 28 December 2023

* **Corresponding Author** Kyungyong Chung, Division of AI Computer Science and Engineering, Kyonggi University, 154-42, Gwanggyosan-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do, South Korea E-mail: dragonhci@gmail.com

I. 서론

인공지능, 빅데이터 기술의 발전에 따라 환경, 식습관, 센싱, 생활활동에서 관측된 데이터들이 수집이 가능하다. 이와 같은 데이터는 기술의 발전에 따라 빅데이터 규모로 급증하고 있다. 다양한 분야에서 생성되는 활동, 환경, 의료, 건강 등 정보는 다양한 비정형, 반정형, 정형 데이터의 형태를 가지고 있다. 최근 머신러닝 기반 잠재적 질병 예측, 건축 구조물 강도 예측, 신약 개발, 약물 부작용, 이상 탐지 등 연구가 다양하게 진행되고 있다[1]. 데이터 마이닝 연구는 통계 분석과는 다른 특성이 있다. 연구의 대상이 되는 기존 의사결정의 문제점과 인공지능 데이터의 한계점을 해결하기 위해서 육하원칙을 반영한 경량화된 모델과 인간의 학습원리, 고차원적 인지기능을 통해 의미적 유용성을 확인하기 위한 프로세스가 요구된다[2]. 비정형 데이터는 텍스트로 기술되어 있고, 약어와 기호가 사용되고 있는 특성이 있다. 자연어 처리 기법과 같은 비정형 형태의 마이닝 기술의 효과적인 설명가능한 의사결정에 관한 연구는 매우 중요하며, 다양한 형태의 데이터를 처리하기 위한 마이닝 고도화 기법이 필요하다. 의사결정 지원 시스템의 요소 기술로 대표되는 기계학습, 강화학습, 데이터마이닝은 연관규칙, 상관관계, 회귀분석, 군집 및 분류, 예측의 기법을 사용하여 데이터를 처리한다. 이를 기반으로 작동하는 앙상블 모델링 기법과 같은 설명가능한 인공지능 기술은 빅데이터 처리에 중요한 요소가 될 것으로 보인다.

빅데이터 분야에서 설명가능한 의사결정의 정확성, 투명성, 신뢰성 요구가 증가하고 있다. 해석가능한 기계학습을 통한 의사결정 및 의미론적 관계 발견을 위한 데이터마이닝 고도화 연구는 중요하다[3]. 전문가의 경험에 의한 의사결정은 데이터 기반 수학적 통계 모델에 근거하여 지식을 생성하기 때문에 결과가 어떠한 근거나 과정으로 도출되었는지 알 수 없다. 결과에 대한 설명성이 부족하고, 신뢰성 문제가 지속적으로 발생하는 한계점을 해결하고자 한다. 빅데이터 처리의 핵심은 다양한 데이터의 확보와 분석의 정확성이며, 최근의 빅데이터 분석 알고리즘은 정확도와 속도면에서 긍정적 효과를 보여주고 있다[3]. 전문가에게 단시간 내에 분석, 연구 결과를 제공하고 설명가능한 예측 및 시뮬레이션 제공이 가능해졌다. 클라우드 시스템은 프라이버시를 보호하고, 데이터 관리 비용을 효율적으로

로 개선한다. 또한 빅데이터 처리 기술은 사용자에게 설명이 가능한 예측 결과를 제공한다. 이를 통하여 전문가들은 분석 내용을 이해하고 의사결정에 합리적인 근거를 마련할 수 있게 된다. 본 논문에서는 설명가능한 의사결정을 위한 마이닝 기술을 제안한다. 제안하는 방법은 설명가능한 의사결정을 위한 데이터 처리, 육하원칙을 적용한 다중 계층 마이닝의 분류, 의사결정을 위한 앙상블 모델을 제시하여 성능 고도화를 진행한다.

본 논문의 구성은 다음과 같다. 2장에서는 의사결정을 위한 분야별 기술 동향 및 한계점에 대해 제시하고 3장에서 제안하는 설명가능한 의사결정을 위한 마이닝 기술에 대해 기술한다. 4장에서는 결과에 대해 기술하고, 5장에서는 결론을 맺는다.

II. 관련연구

2.1 의사결정을 위한 헬스케어 기술 동향

설명가능한 인공지능은 딥러닝, 머신러닝 모델의 결괏값에 대한 설명가능한 이유를 인간이 이해할 수 있도록 모듈을 분해 및 파악하여 제공한다. 이를 통해 인간과 컴퓨터의 상호작용에 합리성을 확보하고 인공지능의 정확한 판단을 사람이 이해할 수 있도록 설명함으로써 의사결정 모델이 근거, 추론 과정 등을 논리적으로 나타낼 수 있게 한다.

설명가능한 인공지능 기술 종류로 의료 분야에서는 CAM, GRAD-CAM[2], LRP[3], Deep SHAP[4], trainable Attention을 사용한다. H. M. Bas et al.[1]은 딥러닝 기반 의료영상 분석에서 설명가능한 인공지능을 제안하였다. 이는 CT 영상에서 이미지 조각당 칼슘을 추출하였고 딥콘볼루션을 사용하여 시각화를 진행한다. CAM(Class Activation Mapping) 방법을 사용하여 다양한 공간 위치에서 시각적 패턴의 존재에 대한 가중치형합계를 기반으로 Grad-CAM을 사용한다. 또한 유도역전파를 고도화하여 레이어별 전파 방법을 사용하여 분류한다. 훈련가능한 어텐션을 적용하여 네트워크에서 분류를 위해 입력 이미지에 주의를 기울인 위치와 비율을 강조한다. 훈련가능한 어텐션을 사용하여 관련 영역을 증폭시키거나 삭제하는 방법을 통해 의사결정 과정에서 이미지에서 알려지지 않은 정보를 식별한다. 하지만 의사결정 모델이 계산한 내용에 완전히 충실하지 않은 설명을 제시하는 한계가 존재한다. 그림 1은

Grad-CAM 구조를 나타낸다.

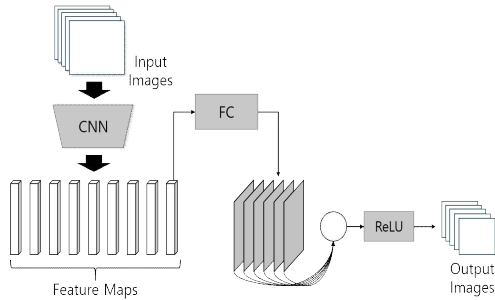


Fig. 1. Grad-CAM Structure

2.2 분야별 의사결정을 위한 기술 한계점

금융 분야에서 의사결정 모델은 SHAP, LIME[5], LRP를 사용한다. T. Hong et al.[6]은 딥러닝 기반의 설명가능한 인공지능을 활용한 암호화폐 가격 예측 방법을 제안하였다. 이는 비트코인 가격 데이터와 구글 검색량 데이터를 전처리하여 CNN을 적용하여 등락 예측 모형을 구축한다. SHAP(SHapley Additive exPlanations)를 활용해 이전 단계에서 구축한 딥러닝 예측모형이 산출해내는 개별 등락 예측값에 대한 분석을 진행한다. LIME(Local Interpretable Model-Ahnostic Explanation)를 적용하여 어떤 단어가 자주 등장할 때 이상 행동을 탐지한다. 또한 코퍼스에서 인공지능 의사결정 모델이 어떠한 영향을 미치는지 파악한다. 기존의 딥러닝 기법을 이용하여 암호화폐 및 주식 예측과 관련된 연구에서는 높은 성능을 보임에도 불구하고 결과를 산출한 근거에 대한 제시가 부족하다. 규제 산업에 포함되는 금융권에서는 딥러닝을 적용하기 힘들다는 단점이 있다. 이는 결과에 대해 어떤 변수가 얼마만큼의 영향을 주는지 파악할 수 있는 설명가능한 인공지능 기술 중에 하나인 SHAP을 활용해 결과의 이유를 파악할 수 있게 한다. 하지만 암호화폐의 특성인 대용량 데이터를 사용하면 연산량이 커진다는 점과 기존에 학습된 데이터가 없다면 학습 데이터를 기반으로 이와 유사한 가상의 데이터를 증강하는 한계점이 존재한다.

법률 분야에서 의사결정 마이닝 모델은 LEXAI[7]를 사용한다. J. Bae. et al [8]은 설명가능한 AI를 이용한 법문서 유사도 분석 서비스를 제안하였다. 이는 대법원 판결문을 입력받아 문서들의 단어에 가중치를 부여한다. 문서의 특징 벡터를 추출하는 과정을 거쳐 표현

값을 도출하는 LEXAI(Legal document analysis service using explainable AI) 기법을 사용한다. 문서의 단어에 가중치를 부여하기 위해서 벡터 차원에 임베딩을 진행한다. 문서의 단어 출현 빈도수를 가지고 접근하는 문서 유사도를 TF-IDF 방법을 사용한다. 비슷한 단어가 사용된 문서들에 대하여 동일한 단어의 집합에서 배치와 순서에 따른 문맥 흐름이 변경되는 문제점이 있다. 언어적인 논리적 흐름이 달라질 수 있다. 그림 2는 LEXAI 프로세스를 나타낸다.

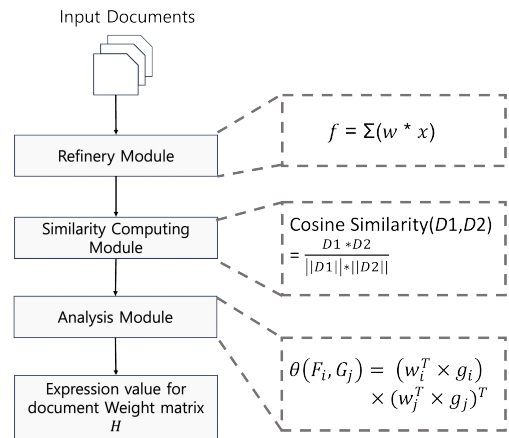


Fig. 2. Process of LEXAI

III. 설명가능한 의사결정을 위한 마이닝

3.1 의사결정을 위한 데이터 처리 기술

빅데이터 처리는 데이터 결측값 및 이상값 처리, 데이터 모델링, 예측을 통해 의사결정을 지원한다. 알고리즘의 복잡성으로 인해 모델 내부는 도출한 결과에 대한 근거, 과정을 명확하게 설명하는 것은 어려운 문제이다. 의사결정 지원 시스템으로부터 생성되는 모든 과정과 결과의 타당성, 신뢰성, 정확성에 대한 명확한 설명이 필요하다. 또한 알고리즘의 특성과 장단점에 대한 단계적인 설명을 제공해야 한다. 설명가능한 모델은 심층 설명학습, 의사결정트리, 추론, 모델 유추 등으로 구성한다. 심층 설명학습은 신경회로망의 은닉 계층의 노드에 설명 라벨을 붙이는 방식으로 개선하고 기존 신경회로망을 변형하거나 하이브리드 형태로 보완하는 딥러닝 기술이다. 네트워크에서 주목하는 노드를 역추적하여 최종 결론에 도달한 의미 해석이다. 의사결정트리는 머신러닝을 통한 의사결정 로직을 학습하여 학습 네트워크 구조에서 의사결정 과정과 연계하

여 설명하고, 의사결정 트리와 같은 해석력이 높은 학습 방법과 결합하여 결과의 일치성을 확인하고 모델을 도출한다. 모델 유추는 통계적 모델로 실험과 관찰을 통하여 블랙박스 모델의 결과를 추론하여 설명한다. 모달 통합에서 지도학습, 비지도학습, 준지도학습, 강화학습 범주를 고려하여 구성하고, 클리닝, 통합, 변환, 축소, 이산화 등을 이용하여 충분하지 않은 양의 훈련 데이터 문제, 샘플링 편향 문제, 에러, 이상치, 잡음 등 데이터 문제, 과대적합 및 과소적합 문제를 해결한다. 정형, 반정형, 비정형으로 구성된 불안전 및 불균형 데이터로부터 존재하는 관계를 학습하고, 공통정보를 추출하는 데이터 처리, 레이블되지 않은 데이터를 심층 오토인코더로 훈련한다. 잡음제거 오토인코더의 입력층에 가우시안 노이즈를 사용해 잡음제거, 비지도 학습이나 차원 축소를 위해 시계열 데이터에 대해 순환 오토인코더를 사용하여 입력 시퀀스를 벡터로 압축하는 시퀀스-to-벡터 RNN을 사용하여 인코더를 구성한다. 벡터-to-시퀀스 RNN을 사용하여 디코더하여 시공간, 다변량, 비정형 데이터 처리 기법 및 구조변경을 통한 모델 경량화를 진행한다.

3.2 육하원칙을 적용한 다중 계층 마이닝 분류

다중 계층 마이닝은 데이터 정제와 분석 변수 처리 등 데이터 전처리를 거친 트랜잭션에서 빈번하게 발생하는 변수와 속성 간의 다차원 관계를 발견하는 방법이다. 빈발항목집합으로부터 다차원 규칙을 육하원칙을 적용하여 발견하는 방법으로 정량적 속성의 정적 이산화를 통한 연관성을 발견한다. 트랜잭션에서 빈번한 항목을 추출하며, 항목들간의 연관성이 최소지도를 만족하는 규칙을 찾는 방법이다. 또한 효율성과 확장성을 향상시키기 위해 증강, 감소, 분할, 샘플링 등을 적용한다. 트랜잭션에서 마이닝을 이용하여 유의미한 관계를 발견하기 위해 다중 계층 마이닝을 통해 항목 사이의 연관성을 발견한다. 회귀분석은 선형적인 상관성을 수학적으로 추정하고 독립변수와 종속변수를 이용하여 데이터를 모델링한다. 독립변수는 설명변수라고도 하며, 원인이 되는 변수, 예측에 필요한 변수이다. 종속변수는 목표변수, 반응변수라고도 하며 예측의 결과가 되는 변수이다. 위험도 예측을 위한 선형회귀, 다중회귀, 비선형회귀는 인과관계를 통하여 독립변수가 종속변수에 영향을 미치는 정도를 발견한다. 선형회귀는 종속변수와 독립변수의 선형 상관관계를 모델

링하며 독립변수의 수에 따라 단순 선형과 다중 선형으로 구분한다. 육하원칙을 적용한 계층 마이닝의 고도화로 확률적 경사 하강을 사용하여 확장 가능한 모델과 로지스틱 회귀모델을 개발한다. 컴퓨팅에서 확장 가능한 데이터 구조와 예측 모델을 개발하여 위험도를 판단한다. 데이터 정제, 관련성 분석, 데이터 변환, 데이터 증강을 통해 사전에 지정한 속성의 범위에 따른 클래스 레이블에 속하게 하는 방법이다. K-NN, SVM, 의사결정트리, 랜덤포레스트, 신경망을 사용하여 각각 클래스로 분류하며 설명가능한 특징을 추출한다. 회귀 분석, 다변량분석, 시계열분석, 베이스 기법, 비정형 데이터 분석, 회귀 트리, 분류 트리에 대해 검증된 데이터를 처리할 수 있도록, 육하원칙을 적용한 의사결정 트리에 의한 분류를 진행한다. 그림 3은 육하원칙을 적용한 의사결정트리에 의한 분류 결과를 나타낸다.

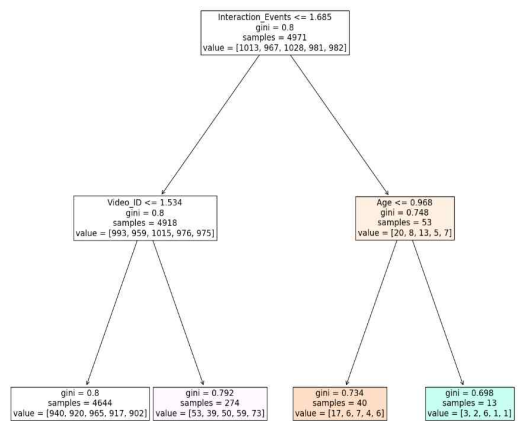


Fig. 3. Classification results from a decision tree using the 5w1h principle

IV. 설명가능한 의사결정을 위한 마이닝 결과

제안하는 설명가능한 의사결정을 위한 마이닝 기술의 결과는 k-means 클러스터링을 활용한다. 이는 군집의 개수를 k개로 설정하여 각 군집에 포함된 데이터의 샘플을 평균 중심으로 이동하여 군집을 진행한다. 그러나 가장 적합한 클러스터의 개수를 설정하기 위해서는 Elbow 방법을 활용한다[9]. Elbow는 Inertia가 빠르게 변하는 지점을 가장 적합한 k로 설정한다. Inertia는 각 군집에 따른 오차의 제곱합으로 군집 내 분산으로

정의하여, k의 수가 증가할 경우 샘플이 할당된 중심에 가까워지기 때문에 inertia가 감소한다. 그림 4는 Elbow 방법을 이용한 군집 개수 평가 결과를 나타낸다. 이는 가로축은 k의 개수, 세로축은 분포의 점수를 나타낸다.

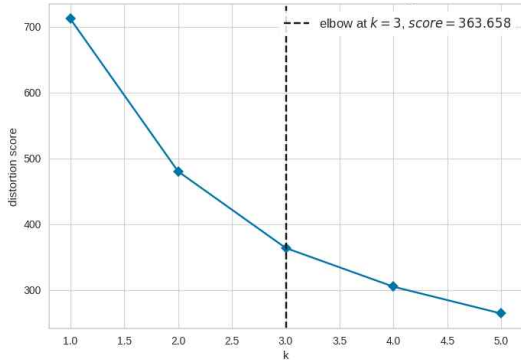


Fig. 4. Results of cluster count assessment using the Elbow method

그림 4의 결과에서 k=3일 경우 가장 적합한 군집의 개수를 나타낸다. 이에 따라 그림 5는 k가 3일 경우의 군집 결과를 나타낸다.

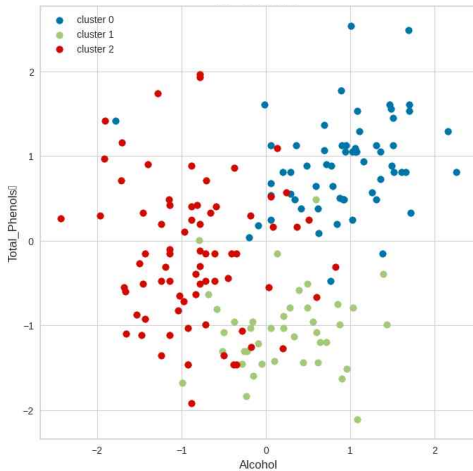


Fig. 5. Clustering results for k=3

의사결정을 위한 앙상블 모델은 다양한 모델의 결과를 이용해 적합한 예측 결과를 도출한다. 다수결 방식을 사용하는 부트스트랩 베깅 예측, 랜덤 포레스트, 부스팅 등이 있다. 다수결 방식은 비상관화를 가진 의

사결정트리를 구성하고, 적절한 결과를 결정하는 방법이다. 이는 랜덤 포레스트가 노이즈가 포함된 데이터 집합에서 유용하게 동작한다. 트리 랜덤화는 베깅 과정을 통해서 구성되며, 훈련 데이터 집합을 통하여 트리를 훈련하고 투표 방식을 통하여 결합하여 구성한다. 이는 일반화 성능이 좋지 않게 과적합 되기 쉽다는 단점을 해결한다. 예측도가 떨어지는 오류 데이터에 가중치를 부여하여 부정적인 결과를 보여주는 모델을 수정한다. 모델의 성능이 떨어지더라도, 최종 모델은 개선된 결과를 보여준다. 우수한 성능을 활용하기 위한 다양한 파생 알고리즘, 패키지, 라이브러리가 개발되고 있다. XGBoost[10], Light GBM[11], CatBoost를 이용하여 성능을 개선하고, 상당한 계산량이 요구되는 빅데이터 처리 기술에 응용된다. 하드웨어적으로도 분산 및 병렬 처리 결합, 자원소모 최소화, 효율적 분할을 통해 다양한 앙상블 기법을 개발한다[12]. 학습별로 발생하는 차원 수와 분포의 문제가 있으며, 관계, 시간에 따라 효과적인 성능을 보이는 모델별 시각화가 있다. 의사결정을 위한 앙상블 기법의 성능을 위해서 형태를 조율해야 한다[13]. 파라미터로 효과적인 성능을 나타낼 수 있는 기계학습 기법을 사용하여 결정트리 차원 수를 줄이고, 수행 시간을 개선한다.

V. 결론

불완전한 데이터 확장 기반 이상 탐지와 육하원칙을 적용한 모델링과 의미론적 관계를 확장한 공유속성 추출 기법과 데이터 마이닝 기술은 고성능 병렬 컴퓨팅 환경에서 설명가능한 비즈니스에 활용할 수 있다. 설명가능한 예측을 위한 의사결정 마이닝을 통해 가치를 창출하는 점에서 4차산업의 패러다임을 주도할 수 있고, 정보의 설명성과 가용성 측면에서 실현 가능성을 입증하여 기술적으로 구현하기 위한 융합 연구로 활용이 가능하다. 기계학습을 마이닝 프로세스에 적용하여 빅데이터 분산 공유 및 신뢰성, 투명성, 설명성 이슈를 해결하는 연구 결과를 도출한다. 설명가능한 의사결정을 위한 마이닝 모델을 통해, 다양한 상태를 다각적으로 평가할 수 있고 기계학습으로 데이터 활용한 지능정보화 인프라 조성을 통해 지능형 산업에서 새로운 부가가치 창출과 미래 지향적 사회요구 대응 기술이 가능하다.

ACKNOWLEDGMENTS

이 논문은 2023학년도 경기대학교 연구년 수혜로 연구되었음.

REFERENCES

- [1] B. H. Van der Velden, H. J. Kuijff, K. G. Gilhuijs, M. A. Viergever, "Explainable artificial intelligence (XAI) in deep learning-based medical image analysis," *Med. Image Anal.*, vol. 79, pp. 102470-102490, 2022.
- [2] D. Song, J. Yao, Y. Jiang, S. Shi, C. Cui, L. Wang, F. Dong, "A new xAI framework with feature explainability for tumors decision-making in Ultrasound data: comparing with Grad-CAM," *Comput. Methods Programs Biomed.*, vol. 235, pp. 107527-107528, 2023.
- [3] L. Weber, S. Lapuschkin, A. Binder, W. Samek, "Beyond explaining: Opportunities and challenges of XAI-based model improvement," *Inf. Fusion*, vol. 92, pp. 154-176, 2022.
- [4] A. Singh, S. Sengupta, V. Lakshminarayanan, "Explainable deep learning models in medical image analysis", *J. Imaging*, vol. 6, no.6, pp. 52-70, 2020.
- [5] M. R. Zafar, N. Khan, "Deterministic local interpretable model-agnostic explanations for stable explainability", *Mach. learn. knowl. extr.* vol. 3, no. 3, pp. 525-541, 2021.
- [6] T. Hong, J. Won, E. Kim, M. Kim, "The Prediction of Cryptocurrency Prices Using eXplainable Artificial Intelligence based on Deep Learning", *J. Intell. Inform. Syst.*, vol. 29, no. 2, pp. 129-148, 2023.
- [7] L. K. Branting, C. Pfeifer, B. Brown, L. Ferro, J. Aberdeen, B. Weiss, B. Liao, "Scalable and explainable legal prediction", *Artif. Intell. Law.* vol. 29, pp., 213-238, 2021.
- [8] J. Bai, S. Park "LEXAI : Legal Document Similarity Analysis Service using Explainable AI", *J. of KIISE*, vol. 47, no. 11, pp. 1061-1070, 2020.
- [9] M. Cui, "Introduction to the k-means clustering algorithm based on the elbow method. *Accounting*", *J. Account. Audit. Finance*, vol. 1, no. 1, pp. 5-8, 2020.
- [10] Z. Li, "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost," *Comput. Environ. Urban Syst.*, vol. 96, pp. 101845-101872, 2022.
- [11] J. Ren, Z. Yu, G. Gao, G. Yu, J. Yu, "A CNN-LSTM-LightGBM based short-term wind power prediction method based on attention mechanism," *Energy Rep.*, vol. 8, pp. 437-443, 2022.
- [12] B. U. Jeon, K. Chung, "CutPaste-based Anomaly Detection Model using Multi-scale Feature Extraction in Time Series Streaming Data," *KSII Transactions on Internet and Information Systems*, Vol. 16, No. 8, pp. 2787-2800, 2022.
- [13] H. Yoo, R. C. Park, K. Chung, "IoT-Based Health Big-Data Process Technologies: A Survey," *KSII Transactions on Internet and Information Systems*, Vol. 15, No. 3, pp. 974-992, 2021.

저자소개

정 경 용 (Kyungyong Chung)



2000년 2월 : 인하대학교

전자계산공학과 (공학사)

2002년 2월 : 인하대학교

전자계산공학과 (공학석사)

2005년 8월 : 인하대학교

컴퓨터정보공학부 (공학박사)

2006년 3월 ~ 2017년 2월 :

상지대학교 컴퓨터정보공학부 교수

2017년 3월 ~ 현재 : 경기대학교

AI컴퓨터공학부 교수

관심분야 : 데이터마이닝, 헬스케어, 빅데이터, 헬스케어, 인공지능, 데이터 분석, 정보검색, 추천