



A Design Case Study of Artificial Intelligence Pipeline Visualization

Yoonha Park¹, Jae Young Yun^{2*}

¹Department of Visual Communication Design, School of Design, Student, Hongik University, Seoul, Korea

²Department of Visual Communication Design, School of Design, Professor, Hongik University, Seoul, Korea

Abstract

Background Artificial intelligence has been variously used and played a key role as we move toward intelligent information society. However, although it was designed by human, the current artificial intelligence solution can be said to be very closed to 'black box' as it is difficult to look into its internal decision making process. Although the necessity and importance of modeling analysis is increasing, there has been lack of academic discussions on how to approach the visualization of its process. Therefore, in this study, we examine the visualization approach for each process through the analysis of preceding studies that visualize the execution process of artificial intelligence.

Methods In this study, machine learning and deep learning pipelines in general predictive modeling projects were examined and the papers applied to them were examined, and the process of performing artificial intelligence was divided into three stages (Collection, Learning, Inference). In addition, we analyzed how visualization is performed at each stage based on the research framework.

Results The processes of artificial intelligence are categorized as ① Collection process (visualization of data collection status) ② Learning process(preprocessing, characteristic extraction and selection) ③ Inference(valuation of accuracy and significance of algorithm). As a result of examining visualization method of each procedural step, we found that there is difference in providing data visualization method depending on the research purpose (understanding model operation/identifying the cause of misclassification, debugging support, etc.) and subject(expert/non-expert).

Conclusions The items derived from this study will be used as basic data for analyzing data visualization design cases according to the execution process of artificial intelligence in the future. Based on this study, if case analysis from more diverse perspectives is conducted in the future, it is expected that strategy establishment and theoretical foundation to visualize the execution process of artificial intelligence will be possible.

Keywords Artificial Intelligence, Data Visualization, Modeling Analysis Suitability

This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2018S1A5A8029352)

*Corresponding author: Jae Young Yun (ryun@hongik.ac.kr)

Citation: Park, Y., & Yun, J. Y. (2021). A Design Case Study of Artificial Intelligence Pipeline Visualization. *Archives of Design Research*, 34(1), 133-155.

<http://dx.doi.org/10.15187/adr.2021.02.34.1.133>

Received : Aug. 20. 2020 ; **Reviewed :** Jan. 22. 2021 ; **Accepted :** Feb. 10. 2021

pISSN 1226-8046 **eISSN** 2288-2987

Copyright : 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 educational and non-commercial use, provided the original work is properly cited.

1. 연구의 배경 및 목적

최근 사용자의 의사결정을 돕는 인공지능 혁신 사례들이 지속적으로 발표되고 있지만, 지금까지의 사례들은 인공지능이 예측하는 ‘결과’ 위주로 보여줄 뿐, 어떻게 이 판단에 이르렀는지에 대한 설명이 부족하거나, 시각적으로 보여주지 않는다는 한계가 있다(Došilović & Hlupić, 2018). 사용자의 입장에서 현재 인공지능 시스템에서 주로 보이는 ‘인공지능이 도출한 결과’와 ‘해당 결과에 대한 확률적 정확도’만 가지고는, 해당 시스템에서 입력 데이터로부터 어떻게 학습이 이루어졌고 어떻게 수행되었는지에 대한 과정을 자세히 이해하기 어려워서, 사용자가 최종 결정을 내릴 때 부담이 있을 수 있다. 이에 인공지능 내부에서 일어나는 과정을 투명하게 시각화하기 위한 연구가 여러 학계와 산업계에서 다방면으로 진행되고 있지만, 아직까지는 정확성과 같은 결과적인 부분에 더 초점이 맞춰져 연구가 진행되고 있다(Gulshan et al., 2016; Esteva et al., 2017). 따라서 본 연구에서는 일반적인 예측 모델링 프로젝트에서의 머신러닝과 딥러닝 파이프라인(Pipeline)을 살펴보고, 이를 적용한 논문들을 조사하여, 인공지능의 수행과정을 세 단계(탐색, 학습, 추론)로 구분하였다. 그리고 각 단계별로 어떻게 시각화가 이루어지고 있는지 리서치 프레임워크(Research Framework)를 토대로 분석하고 이해하는 데 연구의 목적을 둔다.

1. 1. 인공지능 시각화 연구의 한계점

최근 인공지능 분야에서 데이터 시각화를 활용하여 시각적 분석을 시도하고, 인공지능이 예측하는 결과를 사용자에게 보여주는 방법에 관한 연구가 활발히 진행되고 있다. 예를 들어 Jang & Yun(2018)은 주식투자자들의 의사결정에 도움을 주기 위해 인공지능 기반의 디자인 전략을 제안하였고, Lee & Pan(2019)은 인공지능 기술을 활용한 디자인 시스템 구조의 현황과 특징을 4단계(스타일 가이드, 템플릿, 페이지, 컴포넌트)로 구분하여 조사하였으며, Min(2018)은 추천 시스템의 투명성과 직관성을 높이기 위해 인공지능이 제안하는 결과값을 시각화하여 제공하였다. Mun & Lee(2016)는 통계적인 데이터 분석과 시각적인 데이터 분석을 결합한 예측모형을 검증하였고, Ryu et al.(2017)은 머신러닝 기술 기반의 주가 패턴 유형에 따른 클러스터 결과를 시각화하여 향후 주가 방향을 분석하였다. 특히 인공지능이 적극적으로 활용되고 있는 의료 분야에서의 AI 혁신 사례로, Gulshan et al.(2016)은 당뇨병성 망막병증을 정확하게 발견할 수 있는 딥러닝 기반의 인공지능을 발표하였고, 미국 스탠퍼드대 연구진 Esteva et al.(2017)은 피부암을 진단하는 딥러닝 기반의 인공지능을 개발하였다. 이처럼 사용자들의 의사결정을 돕기 위한 목적으로 인공지능의 결과값을 시각화하려는 사례들이 지속적으로 발표되고 있지만, 아직까지는 인공지능이 예측하는 ‘결과’ 위주로 보여줄 뿐, 어떻게 이 판단에 이르렀는지에 대한 설명이 부족하거나, 시각적으로 보여주지 않는다는 한계가 있다(Došilović & Hlupić, 2018).

1. 2. 블랙박스 이슈에 따른 인공지능 수행과정 시각화의 필요성

인공지능에 대한 기대에도 불구하고, 지금까지의 인공지능은 의사결정을 지원할 때 결과만 알려주고 어떤 근거로 이 판단에 이르렀는지 논리적으로 설명할 수 없는 점이 한계로 지적되므로 “블랙박스”로 불리기도 한다. 인공지능 모델이 만족스럽지 않은 수행을 했을 경우, 사용자는 입력 데이터로부터 어떻게 학습이 이루어졌고 어떻게 수행되었는지에 대한 과정을 이해하지 못함에 따라 문제 해결 방법을 알지 못한다는 한계를 지닌다. 또한 생성된 결과나 예측에 대한 해석이 어렵다면, 사용자들은 인공지능이 제시한 결과를 완전히 신뢰할 수 없을 것이다(Jung et al., 2019).

인공지능이 제안하는 결과(예측)값을 활용하려면 1) 인공지능 제안하는 ‘예측값에 대한 신뢰’와 2) 예측값을 도출하기 위해 사용된 인공지능 ‘모델에 대한 신뢰’를 갖는 것이 중요한데, 이 두 가지 모두 사용자가 모델의 작동을 얼마나 이해할 수 있는가에 영향을 받는다. 이에 인공지능 컴퓨팅 분야 선도기업인 NVIDIA는 딥러닝 네트워크를 “시각화”하는 방법에 대해 검토하고 있다고 밝혔고(Tech M, 2017), 여러 연구자들도 입력 데이터로부터 결과값이 도출되기까지의 과정을 시각화하여 인공지능 모델에 대한 신뢰도와 해석 가능성을 높여야 함을 주장하였다. 그 예로, Choo et al.(2010)은 머신러닝 알고리즘이 복잡해질수록 사용자가 이를 해석하는 데 어려움을 겪음에 따라 상호작용 가능한 툴 개발의 필요성에 대해 주장하였고, Tzeng & Ma(2005)는 과정에 대한 시

각화의 필요성을 강조하며, 입력 데이터와 뉴럴 네트워크에 대한 시각화 제공 시 사용자가 복잡한 데이터 관계를 파악하고 이해하는 데 도움이 될 것이라고 주장하였다.

사용자의 이해를 돕기 위해, 모델링 해석의 필요성과 중요성이 나날이 높아져 가고 있지만, 그 프로세스를 시각화할 때 어떻게 접근해야 할지에 대한 학술적인 논의는 부족한 실정이다. 따라서 본 연구에서는 인공지능의 프로세스를 시각화한 선행연구 분석을 통해, 각 과정별 시각화 접근방법을 살펴보고자 한다.

2. 리서치 프레임워크(Research Framework) 도출

본 연구에서는 머신러닝과 딥러닝 모델링 프로세스를 다룬 논문들을 조사하여, 인공지능의 수행과정을 세 단계로 구분하였다. 그리고 각 단계별로 어떻게 시각화가 이루어지고 있는지 이해하기 위해 인공지능 수행과정에 대한 시각화가 분명히 제시된 선행연구 9편을 선정하고, 리서치 프레임워크(Research Framework)를 토대로 분석하였다.

2. 1. 인공지능 수행과정 구분

(1) 머신러닝의 모델 구축 프로세스

머신러닝 소프트웨어를 제작할 때 구상부터 시작해서 데이터 탐색, 분석, 검증, 배포와 모니터링에 이르기까지의 과정을 “머신러닝 라이프사이클 (Machine Learning Life Cycle)”이라고 표현한다. 머신러닝도 기본적으로 하나의 소프트웨어라고 할 수 있기 때문에 개발 라이프사이클을 가지게 되며, 데이터가 처음 돌아가게 되는 흐름부터 산출되는 최종적인 아웃풋까지의 일련의 흐름을 파이프라인이라고 한다. 이는 프로젝트에 따라 진행과정이 조금씩 다르지만, 일반적인 예측 모델링 프로젝트(predictive modeling)는 다음과 같은 프로세스를 따른다.

- 데이터 수집 : 프로젝트를 진행하기 위해서는 데이터가 필요하며, 데이터 수집 방법에는 여러 가지가 있다. 데이터베이스에 데이터가 쌓여 있는 경우 데이터 수집 과정은 간단하게 데이터베이스나 데이터 파일에서 데이터를 불러오는 것으로 충분할 수 있다.
- 데이터 전처리 : 수집된 그대로의 데이터는 결측값이 있거나, 이상한 값이 들어간 경우가 많다. 따라서 이러한 데이터를 정제해서 머신러닝 모델의 입력에 적합한 형태로 바꿔주는 단계를 ‘데이터 정제(data cleaning)’라고 부르며, 데이터에서 기존 속성(feature)을 조합해서 새로운 속성을 만들어내기 위해 데이터 전처리가 필요한 경우도 있다. 즉 데이터 전처리란, 데이터를 정제하고 가공하는 일 외에 데이터의 스케일을 맞추고(feature scaling), 더미화하고(dummification), 차원을 줄이는 일(dimensionality reduction) 등을 모두 포함하는 개념이다.
- 탐색적 데이터 분석(EDA: Exploratory Data Analysis) : 전처리된 데이터를 이해하기 위해 데이터의 특징을 찾고, 숨겨진 패턴을 발견하는 과정을 ‘탐색적 데이터 분석’이라고 부른다. 탐색적 데이터 분석 단계에서는 데이터의 히스토그램이나 두 변수 사이의 산포도를 그려보고, 변수들의 상관관계를 보는 일 등을 포함한다. 탐색적 데이터 분석과 데이터 전처리, 모델 선택 과정은 순차적이라기보다 반복적인 관계로, 탐색적 데이터 분석을 통해 어떤 전처리가 필요한지 알 수 있고, 전처리 후에 데이터를 더 잘 이해할 수도 있다. 모델 선택 과정 역시 데이터 탐색과 밀접한 관계에 있다. 데이터를 이해하고 나서 데이터에 적합한 모델을 선택할 수 있고, 원하는 만큼 모델의 정확도가 나오지 않을 경우 그 이유를 찾기 위해 데이터 탐색 과정으로 돌아올 수 있다.
- 모델 선택 : 모델을 선택한다는 것은 말 그대로 예측값을 계산하는 알고리즘을 선택하는 것이다. 모델이 사용할 속성들(features)을 선택하는 것은 때때로 의미 없는 속성이 들어갈 때 모델 성능이 더 떨어지는 경우가 있기 때문에, 모델에 중요한 속성들을 골라내는 일도 모델 선택과정에서 필요한 일이다.
- 평가 및 적용 : 만들어진 머신러닝 모델의 성능을 평가하고, 모델을 활용하여 새로운 데이터에 대한 예측을 하는 단계이다.

파이프라인은 데이터의 흐름의 관점에서 말하는 것으로, Microsoft Research팀은 머신러닝 파이프라인을 9가지 단계(e.g., 모델 요구사항, 데이터 수집, 데이터 정제, 데이터 라벨링, 피처 엔지니어링, 모델 훈련, 모델 평가, 모델 배포, 모델 모니터링)로 구분하였고(Amershi et al., 2019), Wikibon Research의 Gilbert et al.(2016)은 '데이터 입수 - 데이터 준비 - 모델 교육 - 예측 제공' 4단계로 구성하였다(George, 2016). 이외에도 Wang et al.(2019)은 머신러닝의 기본적인 Workflow를 크게 세 단계(준비-모델링-배포)로 구분하였고, 이식 외(2019)는 5단계(데이터 수집, 데이터 가공, 피처 추출, 학습모델 구현 및 검증, 테스트)로 구분하였다.

(2) 딥러닝의 모델 구축 프로세스

송경두 외(2019)는 의학 분야에서 딥러닝 모델링 프로세스를 주제 선정, 데이터 수집, 데이터 탐색 및 정제, 알고리즘 개발, 알고리즘 평가, 임상적용의 단계로 구분하였다.

- 주제 선정 : 주제 선정이란 해결하고자 하는 문제를 찾는 것으로, 여러 해결과제 중에서 딥러닝에 적합한 과제를 찾는 단계를 말한다.
- 데이터 수집 : 데이터 프로젝트 진행을 위해 데이터를 수집하는 단계이다. 머신러닝에서는 검출하고자 하는 피처(feature)를 전문가가 미리 정의하였지만, 딥러닝은 피처의 정의 없이 데이터로부터 자동적으로 피처를 학습한다.
- 데이터 탐색 및 정제 : 데이터가 수집된 이후에는 데이터를 탐색하는 과정이 필요하다. 예를 들어, 수집된 데이터의 질이 알고리즘 개발에 적합하지, 데이터의 분포가 왜곡되어 있지는 않은지, 데이터의 특성은 어떠한지 등에 대한 탐색이 필요하다. 즉 탐색적 자료 분석(exploratory data analysis)을 통해 데이터 수집 과정에 문제가 없었는지 파악하고, 데이터 정제 방법에 대한 아이디어를 얻을 수 있다. 데이터 정제 과정을 세부적으로 살펴보면 데이터 알고리즘 개발에 적합하게 변환하는 과정과 데이터 레이블링 과정이 있다. 처음 수집된 데이터는 각각의 데이터마다 특성이 다르며 노이즈로 작용할 수 있는 부분들을 포함하고 있기 때문에, 각각의 영상데이터의 크기를 동일하게 맞추고, 실제 알고리즘 개발에 입력할 영상 이외의 부분을 잘라내는 등의 과정이 필요하다.
- 알고리즘 개발 : 알고리즘 개발 단계는 전처리, 모델학습, 후처리 세 가지 단계로 나눌 수 있다. 전처리는 각 영상데이터가 원래 저장되어 있던 형태에서, 딥러닝 모델이 가장 효율적으로 문제를 해결할 수 있는 형태로 변환하는 과정이며, 후처리는 모델이 예측한 이후에 의사나 사용자가 인식할 수 있는 형태로 변환하는 과정을 말한다.
- 알고리즘 평가 : 알고리즘이 개발되면 이를 평가하는 단계가 필요하다. 가장 일반적인 방법은 개발된 알고리즘을 평가 데이터에 적용하여 그 정확도를 평가하는 방법이며, 과제에 따라서 해당 분야의 전문가와 정확도 비교를 통해 개발된 알고리즘을 평가하기도 한다.

관련된 연구로, Gibson et al.(2017)은 딥러닝의 경우 3단계(모델 선택: 학습 데이터에 대한 모델 선택 및 설정, 모델 평가: 검증 데이터를 통해 모델 성능 측정, 모델 배포)로 구성된 일반적인 머신러닝 파이프라인을 채택한다고 하며, 이 과정은 매우 단순해보이지만 실제로는 좀 더 복잡한 과정이 들어가게 된다고 주장하였다.

이외에도 인공지능과 관련된 연구를 진행하며 모델링 구축 프로세스를 언급한 사례로, Park et al.(2018)은 인공지능 개발 사례분석을 위해 스캐터랩의 서비스 사례를 데이터 수집 - 레이블링 - 분석(패턴추출, 패턴인식 등) - 활용 4단계로 구분하여 연구를 진행하였다. Liu et al. (2018)은 스마트 헬스케어 기반으로 Digitize(데이터 수집, 저장) - Analysis(인공지능을 통한 분석, 결과 도출) - Service(분석된 결과 기반으로 맞춤형 서비스 제공)로 구분하여 사례 조사를 실시하였다. Yun & Ahn(2017)은 인공지능 기술을 활용한 가짜 뉴스 예측 모형을 위해 데이터 수집과 클래스 구분(진실, 판단유보, 거짓) - 특징값들로 변환 - 학습용과 검증용 데이터 세트 구분, 특징값들을 다양한 기계학습 기법에 적용, 분류 예측 수행 - 성과 비교로 구성하여 제안하였다. 마지막으로 Lee et al.(2016)은 기계학습의 적용단계를 데이터 수집 - 알고리즘을 통한 데이터 학습 - 알고리즘에 대한 유의미성 단계로 구분하였다.

머신러닝과 딥러닝의 파이프라인은 단순해보이지만, 실제로는 좀 더 복잡한 과정이 들어가게 된다. 각 단계별 시각화 방법을 효과적으로 보여주기 위해, 본 연구에서는 머신러닝과 딥러닝의 파이프라인을 다른 논문들을 토

대로, 인공지능의 수행과정을 ① 탐색단계 (데이터 수집, 전처리, 다양한 기계학습 기법에 적용 등 모형개발 전 단계에 관한 시각화) ② 학습단계 (특징 추출, 특징 선택 등 모형 선택 후 실제 데이터 학습 과정에서의 시각화) ③ 추론단계 (알고리즘의 정확성과 유의미성 평가 등 모형의 성능에 관한 시각화)로 구분하여 연구를 진행해나간다.

2. 2. 선행연구 선정

본 연구의 신뢰도 확보를 위해 선행연구는 다음의 기준에 의거하여 선정하였다. 우선 해외는 구글 스칼라 (Google Scholar), ACM, Science Direct 등을 통해, 국내는 DBpia 등의 데이터베이스에서 주제어 검색을 통해 논문을 조사하였다. 또한 다양한 자료를 얻기 위해 수집한 선행연구들의 참고문헌을 통해 분석대상 연구들을 추가로 조사하였다. 국내의 선행연구 검색 시 주제어는 “Artificial Intelligence”, “Data Visualization”, “Modeling Analysis”, “Visual Exploration”, “Explainable AI”, “기계학습 시각화”등을 포함하여 검색하였다. 해당 검색어 이외의 제한을 두지 않았을 때는 중복된 논문을 포함하여 총 16,700편의 선행연구가 검색되었다. 이 중 인공지능이 주목받기 시작한 2015년부터 2020년까지의 연구를 대상으로 한 결과 총 2,272편이 검색되었다. 이 중에는 알고리즘 개발 혹은 알고리즘 제작 후 결과값의 정확도 향상 등을 중점적으로 다룬 공학 분야의 연구가 다수 포함되어 있었다. 따라서 인공지능의 시각화를 시도한 연구더라도 프로세스와 관련이 없는 논문, 알고리즘 개발이 목적인 논문을 제외한 결과 32편이 남았다. 마지막으로 본 연구의 방향성을 고려하여 이미지나 메시지 분류 작업과 같이 특수한 성격의 시각화 논문도 제외하고, 수행과정에 대한 시각화가 분명히 제시된 최종 9편의 연구를 분석 대상으로 선정하였다.

3. 리서치 프레임워크(Research Framework)에 따른 분석결과

본 장에서는 선정된 선행연구를 연구목적(모델 작동 이해, 모델 오분류 원인 파악, 인공지능 기술 이해, 예측모델 판단과 해석에 도움)에 따라 구분하여 살펴본 후 [Figure 1], 각 단계별(탐색, 학습, 추론)로 어떻게 시각화하였는지 정리하였다.

머신러닝 실무자들의 모델 작동 이해를 목적으로 한 연구로, Kahng et al.(2017)은 사용자들의 딥러닝 모델 해석을 돕기 위한 시각화 툴이 개발되고 있지만, 다양한 산업에서 배포되는 인공지능 모델의 복잡성과 데이터 세트(data set)의 다양성 등으로 인해 기존 시각화 툴로는 해결이 어려운 문제가 있음을 지적하였다. 또한 코드기반 환경(command line interface)에서의 딥러닝 작업 수행 시 직관적인 모델 구현과 이해에 어려움이 있음을 인식하고, 이를 해결하기 위해 페이스북에서 근무하는 15명 이상의 연구원, 데이터 과학자, 엔지니어들을 대상으로 참여디자인 세션을 진행하여 딥러닝 모델을 개발 및 해석해주는 인터랙티브 시각화 시스템 ‘ACTIVIS’를 개발하였다(Figure 1-1).

머신러닝 실무자들의 모델 오분류 원인 파악에 도움을 주고자 한 연구로, Amershi et al.(2015)은 Microsoft 실무자들이 머신러닝 모델 구축 시 오류 분석을 위해 상황을 중단하는 경우가 빈번함을 발견하고, 이를 해결하기 위해 실사용자로부터 모델 구축 시 생기는 오류 등에 대해 반복적 피드백을 얻어, 모델 구축 단계에서 오류 분석으로의 컨텍스트 전환으로 인한 중단을 줄이는데 도움을 주는 인터랙티브 시각화 시스템 ‘ModelTracker’를 개발하였다(Figure 1-2). Siang et al.(2018)은 사용자의 움직임에 따라 변화되는 패턴을 인지하는 ‘인간 동작 인지 기술’의 활용에 있어서, 실제 행위를 다른 행위로 오분류할 경우 환자에게 치명적인 영향을 미칠 가능성이 있다고 주장하였다. 따라서 학습모델이 제대로 분류하지 못하는 특정 동작들에 대한 원인을 파악하고자, UCI Machine Learning Repository에 공개된 데이터를 활용하여 DNN(Deep Neural Network) 기반의 인간 동작 인지모델을 만들고, 분류 성능이 좋지 않은 동작의 예측 결과를 구체적으로 확인하기 위해 t-SNE 알고리즘을 이용하여 데이터를 시각화하였다(Figure 1-3).

데이터 과학자들의 예측모델 판단에 도움을 주고자 한 연구로, Krause et al.(2016)은 데이터 과학자들을 대상으로 인터랙티브 부분 의존성 진단(partial dependence diagnostic)을 제공하여 피처가 전체적으로 어떤 영향을 미쳤는지 파악할 수 있고, 특정 데이터 포인트가 예측되는 방법과 이유를 이해할 수 있으며, 특징값을 조정했을 때 예측이 어떻게 되는지 확인할 수 있는 인터랙티브 시각적 분석 시스템 ‘Prospector’를 개발하였다(Figure 1-4). 비슷한 사례로, Hohman et al.(2019)은 예측 모델을 해석할 수 있는 적절한 툴이 없다면 데이터 과학자들이 잘못된 선택을 내릴 수 있다고 주장하며, 머신러닝 전문가 및 실무자들과의 반복적인 디자인 프로세스를 통해 예측모델 해석에 도움을 주는 시각적 분석 툴 ‘Gamut’을 개발하였다(Figure 1-5).

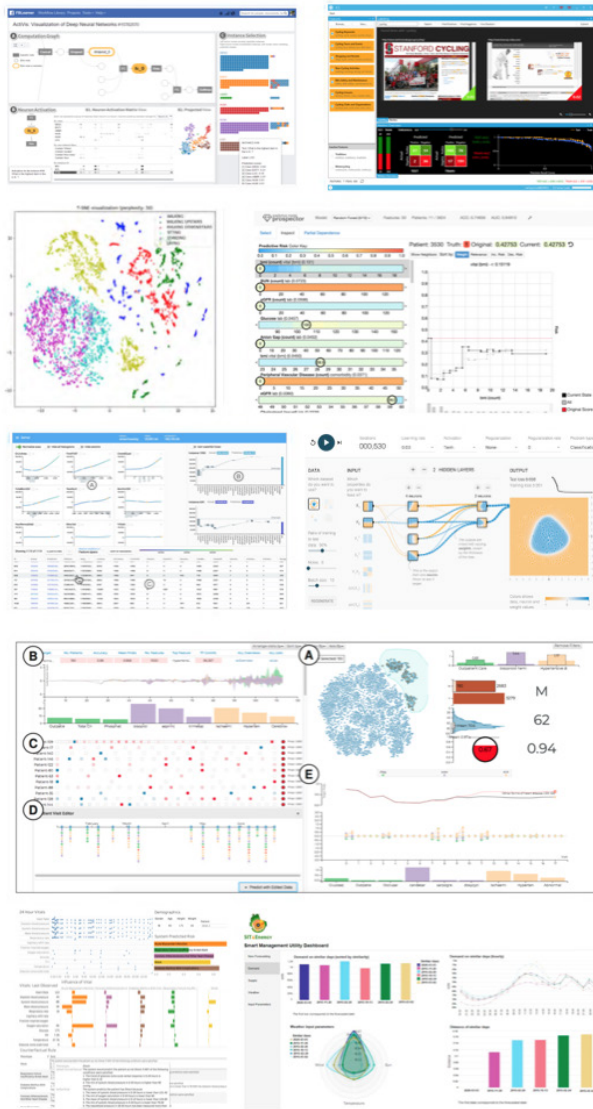


Figure 1 Visualization representative screen of the selected study: From the top left, ACTIVIS (Kahng et al., 2017), ModelTracker (Amershi et al., 2015), T-SNE View (Siang et al., 2018), Prospector (Krause et al., 2016), Gamut (Hohman et al., 2019), Playground (Smilkov et al., 2016), RetainVis (Kwon et al., 2018), AI-driven Medical Diagnosis Tool (Wang et al., 2019), Smart Energy Management Dashboard (Grimaldo & Novak, 2020)

비전문가들의 인공지능 기술 이해에 도움을 주고자 한 연구로, Smilkov et al.(2016)은 비전문가들이 시각화된 신경망을 직접 사용해보며 쉽게 배울 수 있는 교육 목적의 오픈소스 시각화 툴 ‘Playground’를 제작하였다(Figure 1-6). 사용자는 데이터의 비율, 모델 학습 속도 등을 변화시켜가며 시스템의 작동 원리를 빠르게 이해하고 파악할 수 있다.

건강 전문가(의사, 의료 연구원)들의 모델 해석과 상호작용을 향상시키기 위한 연구로, Kwon et al.(2018)은 최근 전자의무기록(EMR)에서 RNN 모델을 기반으로 하여 환자의 현재와 미래 상태를 예측해주는 수많은 애플리케이션들을 볼 수 있지만, 사용자는 모델이 왜 그렇게 예측하였는지 이해하기 어렵다는 점을 지적하였다. 따라서 RNN 모델의 해석력과 상호작용을 증진시키기 위한 목적으로, 연구진들은 HIRA(Health Insurance Review and Assessment Service)에서 제공하는 환자의 데이터를 사용하여 전문가(인공지능 과학자, 시각분석 연구원, 의료 전문가)들과의 반복적인 디자인 평가를 통해 RNN 모델 기반의 해석 가능한 시각 분석 툴 'RetainVis'를 개발하였다(Figure 1-7). 비슷한 사례로, Wang et al.(2019)은 의료진들이 인공지능 시스템을 이해하고 신뢰하며 효과적으로 사용하는 데 도움을 주기 위해 사용자 중심의 XAI(eXplainable AI) 프레임워크를 제작하였다(Figure 1-8).

Grimaldo & Novak(2020)은 재생 에너지원의 증가로 에너지 공급과 수요를 효과적으로 관리하기 위한 새로운 방법들이 개발되고 있지만, 이러한 애플리케이션의 대부분은 복잡한 머신러닝 프로세스를 포함하고 있기 때문에 기술적 지식이 없는 사용자들이 이해하기에는 어려움이 있다고 주장하였다. 결과적으로 머신러닝 분야에 대한 전문 지식이 없어도 에너지 수요와 공급에 대한 예측 모델을 이해하는데 도움을 주기 위한 목적으로, 설명 가능한 머신러닝 모델과 시각적 분석을 결합하여 지역 에너지 시스템의 단기 에너지 수요를 예측하는 툴을 개발하였다(Figure 1-9).

이와 같이 선정한 선행연구에서 시도된 시각화 방법을 인공지능 수행과정(탐색단계 - 학습단계 - 추론단계)에 따라 다음과 같이 살펴보도록 한다(4.1~4.3).

3. 1. 탐색단계

현 탐색단계에서는 크게 세 가지 유형(데이터 특성 시각화, 원시 데이터 시각화, 차원 축소)에 따라 시각적으로 다르게 제공하고 있다. 먼저 데이터 특성에 따라 여러 형태의 그래프를 활용하여 데이터 분포나 크기 등의 파악에 도움을 주고자 하였으며, 원시 데이터를 시각화하기 위해서는 'Interactive Table' 등을 활용하여 사용자들이 직접 비교, 필터링 등이 가능하도록 하였다. 마지막으로 분별력을 유지하며 차원을 줄이기 위해서는, 산점도 차트를 사용하여 분류 분석 결과를 시각적으로 이해하고 빠른 데이터 탐색이 가능하도록 하였다. 탐색단계를 구현한 시각화 사례들은 다음과 같다.

1) 데이터 특성 시각화

Kwon et al.(2018)은 모든 환자의 개요를 스캐터 플롯(A)으로, 환자의 속성(Feature)은 바 차트(B, C)로 제공한다. 오른쪽 상단 바 차트(B)에서는 상위 3가지 공헌인자와 평균 점수를 보여주고, 차트 C에서는 성별, 연령, 예측된 진단 위험도의 평균점수를 각각 바 차트(Bar Chart), 면적 차트(Area Chart), 원형 차트(Circle Chart)로 보여준다(Figure 2). 사용자들은 차트의 컬러와 xy 각 축의 정보를 직접 선택할 수 있다. 여러 형태의 그래프를 활용한 시각화 제공으로 사용자(건강 전문가)들이 데이터 분포나 특성, 크기를 한 눈에 쉽고 빠르게 파악하는데 도움을 줄 수 있을 것이다.

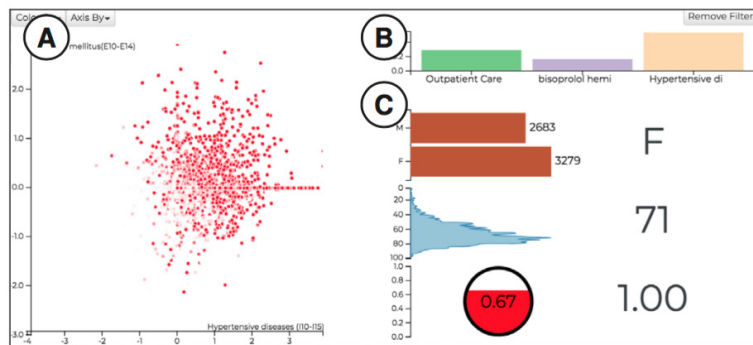


Figure 2 Visualization of all collected patients(A) and features(B,C) (Kwon et al., 2018)

Krause et al.(2016)은 전자 의료 기록을 통해 정의한 진단 코드와 절차, 환자 상태 등을 기반으로 환자의 기록 특성을 두 개의 열(당뇨병 확진을 받은 집단, 당뇨병에 걸리지 않은 집단)로 보여주고, 각 칼럼은 0.1 단계씩 예상 확률로 분류되어 사용자들은 해당 기준에 맞는 환자 선택이 가능하다. 영역이 너무 작아 마우스 클릭이 어려울 경우를 고려하여 각 영역 옆에 박스가 표시되며, 모집단 선택 시엔 환자 개개의 예측 결과를 숫자로 확인할 수 있다(Figure 3). 색과 면적을 기준으로 표시되는 트리맵 차트의 제공으로 집단별 상대적인 크기를 비교하는 것이 가능하지만, 기본적인 막대 차트나 원 차트가 아닌 트리맵 차트를 사용함으로써 대부분의 사람들은 데이터 파악에 어려움을 겪게 될 가능성이 있다.

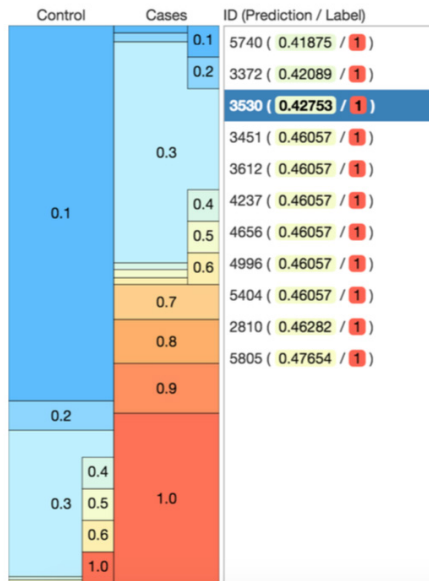


Figure 3 The experimental group and the control group are organized into two columns based on the recording characteristics (Krause et al., 2016)

2) 원시 데이터 시각화

Wang et al.(2019)은 각 환자별 24시간동안 측정된 심박수, 확장기 혈압, 수축기 혈압, 혈당, 체온 등을 원형의 데이터 포인트로 보여주지만(Figure 4), 각각의 데이터 포인트를 추세선으로 보여주지 않기 때문에 특징적인 패턴을 한 눈에 파악하기 어려울 수 있다.

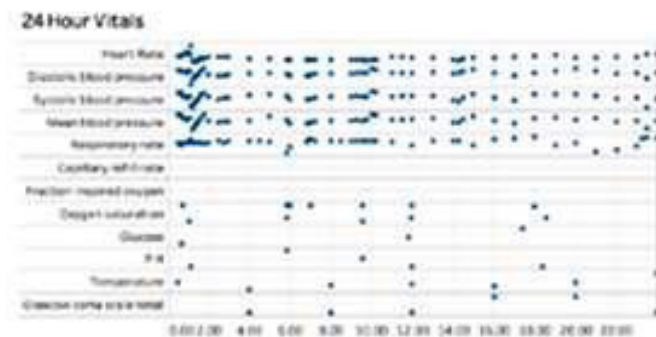


Figure 4 Vital visualization of patients measured for 24 hours (Wang et al., 2019)

Hohman et al.(2019)은 로우 데이터를 Interactive Table로 제공하며, 사용자가 직접 데이터 정렬과 필터링 등을 할 수 있다(Figure 5). 전체 데이터를 한 눈에 볼 수 있는 Interactive Table을 통해 데이터 과학자들이 원시 데이터를 측정항목과 기준에 따라 한 눈에 파악하는데 도움을 줄 수 있고, 바로 비교, 필터링 및 구성을 할 수 있다.

ID	Actual	Predicted	Differences	Neig. ...	LotArea	GrLivArea	FirstFlrSF	Overall...	BsmFlr...	Second...	TotallBs...	YearBuilt	Overall...	Enclose...	Kitchen...	Garage...
833	192000	188383.86...	3616.13818...	1.18187022...	8749	1459	1459	7	0	0	1459	2002	5	0	1	527
614	225000	233550.63...	-8559.6377...	1.18238997...	9158	1496	1496	8	0	0	1496	2007	5	0	1	474
639	139000	142990.90...	-3490.9027...	1.18266194...	9500	1489	944	6	0	545	778	1940	7	0	1	240
721	201000	231785.510...	-30785.510...	1.18421592...	8925	1466	1466	16	0	0	1466	2007	5	0	1	610
616	143500	122110.518...	21389.481...	1.18463303...	8197	1285	1285	0	0	0	660	1977	5	0	1	528
201	175500	176693.40...	-1193.40473...	1.1849362...	12090	1422	1422	6	588	0	1422	1981	6	0	1	576
1111	287090	252843.03...	34246.963...	1.18501411...	9262	1578	1578	8	0	0	1573	2008	5	0	1	840
961	250000	256824.06...	-6824.0685...	1.1850592...	11428	1634	1634	8	0	0	1626	2006	5	0	1	666
1039	189000	207882.38...	-18882.381...	1.18809881...	9842	2601	990	5	0	1611	612	1921	6	301	1	621
601	164000	163636.135...	363.8686...	1.18874491...	7700	1870	1061	6	0	798	766	1968	6	316	1	447

Figure 5 Provides raw data as an interactive table

3) 차원 축소

분별력을 유지하며 차원을 축소하는 사례로, Siang et al.(2018)은 수집된 센서 데이터에서 추출한 561차원의 요인들을 T-SNE 알고리즘을 활용하여 2차원으로 축소해 동작별로 시각화하여 제공한다. 이를 통해 어떤 동작이 모델의 정확성을 떨어뜨리는지 확인할 수 있고, 유사패턴을 갖는 동작에 대한 정보파악도 가능하다(Figure 6, 좌측). Kahng et al.(2017)의 연구에서도 사용자가 고차원적인 패턴을 더 잘 탐색 할 수 있도록 ‘t-SNE 뷰’를 제공한다(Figure 6, 우측). 즉 변수들 사이의 데이터 분포를 확인하기 쉬운 산점도 차트를 사용하여 각 요소들의 위치, 크기에 따른 값들의 관계 등 분류 분석 결과를 시각적으로 이해하고 빠른 데이터 탐색에 유용하다(Byun & Park, 2016).

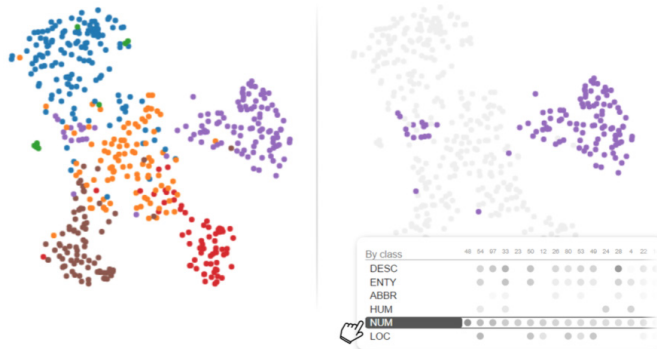


Figure 6 Left: Visualization of extracted factors for each motion to grasp the information of motions with similar patterns (Siang et al., 2018), Right: t-SNE View (Kahng et al., 2017)

3. 2. 학습단계

현 학습단계는 크게 여섯 가지 유형(모델 학습 관련 데이터 시각화, 은닉계층 시각화, 가중치 시각화, 단일 입력 변수와 예측 사이 관계 시각화, 시퀀스 기반 데이터 시각화, 피쳐 탐색 시각화)에 따라 시각적으로 다르게 구성하여 제공하고 있다. 먼저 모델 학습 관련 데이터를 시각화하기 위해서는 데이터의 색상과 배치를 다르게 구성하여 제공하고 있었고, 이 과정에서 수집된 데이터의 비율, 배치 사이즈(한 번에 넣을 자료의 양) 등은 사용자가 직접 설정해서 실험해볼 수 있도록 하였다. 은닉계층의 정보를 보여주기 위해서는 뉴런의 활성화 정보를 계층적으로 시각화하거나 피쳐의 영향력에 따라 밝기의 활성화에 차이를 주었고, 가중치의 경우 선의 굵기를 다르게 표현하여 제공하고 있었다. 모델의 예측이 단일 입력 변수에 어떻게 의존하는지를 보여주기 위해서는 부분 의존성(Partial Dependence) 플롯을 시각화하여 제공하고 있는 것으로 나타났으며, 시퀀스 기반 데이터를 시각화하기 위해서는 시간의 흐름에 따른 양적 진전을 면적 차트로 제공하고 있었다. 마지막으로 피쳐 탐색을 시각화하기 위해서는 Feature Sidebar 등을 통해 모델에서의 중요한 피쳐를 확인하고 비교할 수 있도록 구성하였다. 학습단계를 구현한 시각화의 사례들은 다음과 같다.

1) 모델 학습 관련 데이터 시각화

Smilkov et al.(2016)도 데이터 종류(훈련데이터, 검증데이터)를 두 가지 색상으로 구분해서 보여준다. 사용자는 테스트 데이터셋을 직접 선택해서 사용해 볼 수 있고, 'Show Test Data' 버튼 클릭 시 테스트 데이터의 외곽선은 검은 원형으로 구분되어 표시된다. 또한 슬라이드 바(slider)를 통해 학습과 검증에 사용할 데이터의 비율, 노이즈를 얼마로 둘 것인지, 배치 사이즈를 몇 개로 할지 정할 수 있다(Figure 7). 이처럼 데이터 종류나 데이터 비율 등을 사용자가 직접 설정해서 실험해 볼 수 있도록 설계한다면, 비전문가들의 인공지능 기술과 모델 작동 이해에 도움을 줄 수 있을 것이다.

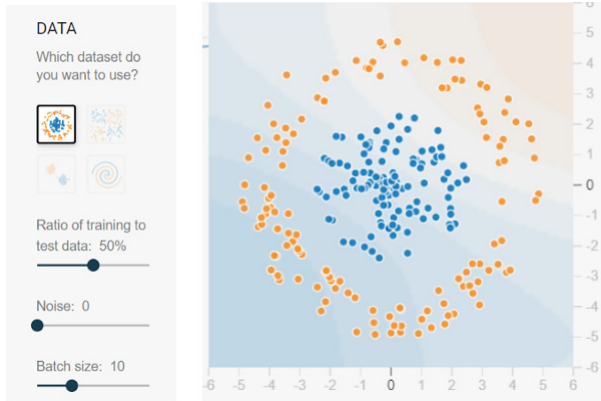


Figure 7 Visualization of the test dataset (playground.tensorflow.org)

Amershi et al.(2015)은 레이블이 있는 박스의 색상을 다르게 제공하며 (녹색-증가, 적색-감소), 학습 데이터는 디스플레이 상단에, 검증 데이터는 하단에 표시된다. 모델의 예측 점수(0~1)에 따라 박스가 배치되기 때문에, 성능이 좋은 모델의 경우 녹색 박스가 오른쪽으로, 빨간 박스가 왼쪽으로 배치된다. 만일 사용자가 모델 구축을 반복하며 더 많은 데이터를 추가할 시, 박스는 재배열되고 자동으로 업데이트 된다(Figure 8). 이를 통해 모델의 성능을 시각적으로 한 눈에 파악, 전달하는 것과 동시에 사용자가 직접적으로 데이터의 오류 등을 검사하는 것이 가능할 수 있다.



Figure 8 Visualization of a labeled box (Amershi et al., 2015)

2) 은닉계층 시각화

은닉계층의 정보를 시각화한 사례로, Amershi et al.(2015)의 'ModelTracker'에서는 실무자들이 모델 구축을 반복하며 데이터를 추가로 입력할 시 박스도 추가된다. 예측 점수가 바뀔 경우 박스는 재배열되며 자동 업데이트되며, 속성(feature)이 영향을 미치는 정도에 따라 박스 밝기의 활성화로 구분된다(Figure 9). 활성화를 시각화하여 제공함으로써 어느 부분이 결정에 중요하게 기여하는지 이해할 수 있고, 분류에 실수가 있는 경우 결정 과정을 디버깅하는데 도움을 줄 수 있을 것이다.

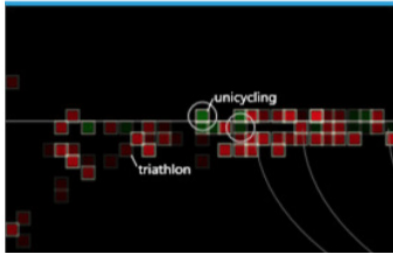


Figure 9 According to the degree of influence of the feature, it is divided into box brightness activation (Amershi et al., 2015)

Kahng et al.(2017)은 활성화된 인스턴스(같은 클래스에 속하는 개개의 개체)의 하위 세트 및 클래스의 활성화 정보를 ‘뉴런 활성화 패널 행렬 뷰’를 통해 계층적으로 시각화하여 제공한다. 녹색 화살표가 가리키는 칼럼(column)은 뉴런을 나타내며, 활성화가 강할수록 어두운 회색으로 보인다(Figure 10). ‘ModelTracker’와 마찬가지로 활성화를 시각화하여 보여줌으로써 어느 부분이 모델의 결정에 중요하게 기여하는지 파악하는데 도움을 줄 수 있을 것이다.

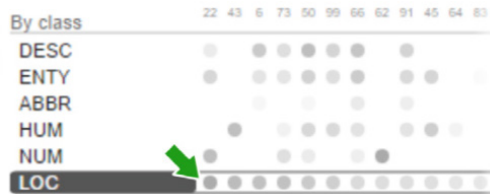


Figure 10 Neuron activation panel matrix view (Kahng et al., 2017)

Kwon et al.(2018)의 연구에서는 환자별 방문 기록은 가로로 배열된 박스를 통해 제공한다. 각 상자는 방문 시 산정된 모든 의료 코드의 공헌도 점수의 합을 나타내며, 그 점수에 따라 청색-백색-적색(음-0-양)의 척도로 변화하도록 구성하였다(Figure 11).



Figure 11 Sequence-based data visualization 2 (Kwon et al., 2016)

3) 단일 입력 변수와 예측 사이 관계 시각화

모델의 예측이 단일 입력 변수에 어떻게 의존하는지 보여준 사례로, Krause et al.(2016)은 부분 의존성(Partial Dependence: 소수의 입력 변수와 예측 사이의 함수관계 나타냄) 플롯을 시각화하여 제공한다 (Figure 12). 이를 통해 예측이 관심 입력 변수의 값에 따라 부분적으로 영향을 받는지 확인할 수 있다. 즉 특성과 예측 사이의 인과관계를 분석하여 명시적으로 보여줌으로써 데이터 과학자들의 빠른 원인 해석에 도움을 줄 수 있지만, Figure 12의 경우 특성 분포가 표시되지 않아 사용자들의 오해를 유발할 가능성이 있다.

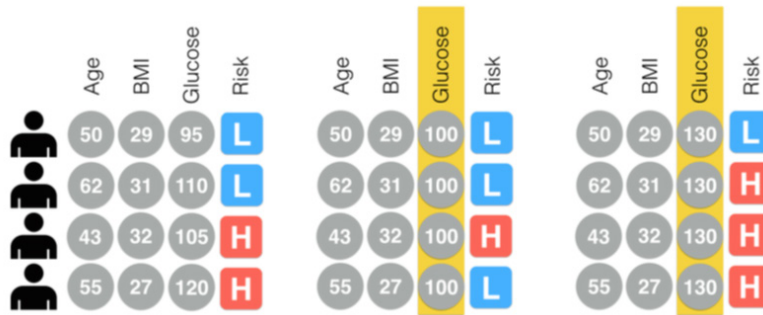


Figure 12 For understanding the relationship between input variables and predictions Partial dependence plot (Krause et al., 2016)

4) 피쳐 탐색 시각화

Hohman et al.(2019)은 각 속성(Feature)에 대한 데이터 밀도를 라인차트로 제공하며, 사용자는 오른쪽의 Feature Sidebar에서 어떤 속성을 보여줄지 선택할 수 있다. Feature Sidebar에 나열된 키워드 순서는 모델에서의 중요도에 따라 정렬된 것이다(Figure 13). 이를 통해 데이터 과학자들이 모델에서의 중요도에 따른 키워드를 한 눈에 파악하는 데 도움을 줄 수 있고 바로 비교, 필터링 및 구성이 가능하다.

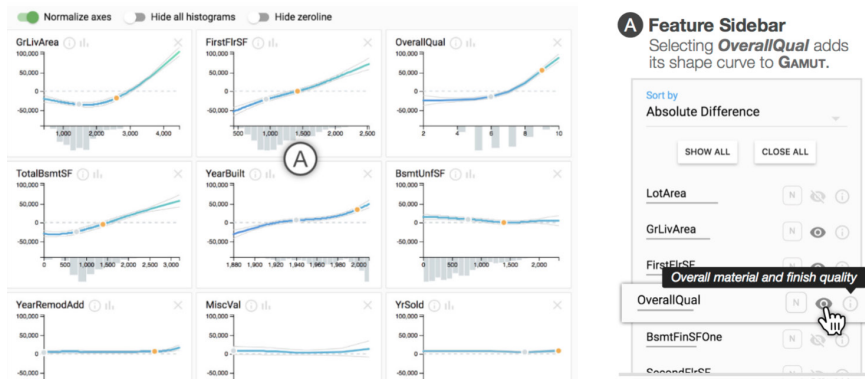


Figure 13 Feature Sidebar (Hohman et al., 2019)

Wang et al.(2019)은 환자의 바이탈 수치가 시스템이 예측하는 5가지 위험상황(e.g., 쇼크, 급성 심근 경색, 관상동맥심장질환 등)에 어느 정도의 영향을 미치는지 토네이도 다이어그램(Tornado Diagram)을 통해 보여줌으로써(Figure 14), 의료진들이 바이탈 수치와 예측되는 위험 상황간의 인과관계를 빠르게 파악하는데 도움을 줄 수 있다.

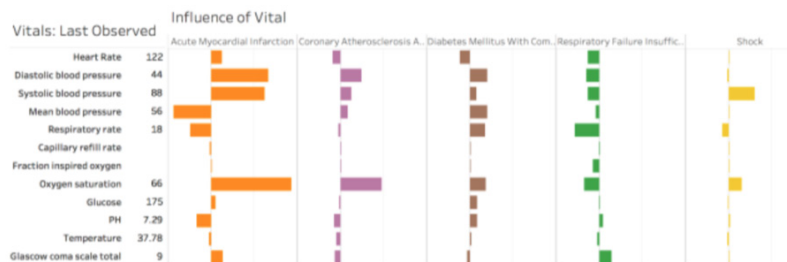


Figure 14 Tornado Diagram (Wang et al., 2019)

Grimaldo & Novak(2020)의 연구에서는 예측 날짜(결과값)의 시간별 에너지 수요를 라인차트(Line Chart)를 통해

보여주며, 시간별 데이터는 과거 유사한 날의 평균값을 기반으로 제공한다. 또한 레이더 차트(Radar Chart)를 통해 예측 날짜와 가장 유사한 과거 5일의 바람, 온도, 일조량 등을 비교해서 보여주며, 사용자는 대시보드에서 필터 기능을 적용하여 특정 요소만을 선택해서 보는 것도 가능하다(Figure 15). 예를 들어, 일조량과 온도가 비슷한 두 개의 날을 선택해서 생산량과 소비량을 분석한 후 최대 소비 시간을 비교할 수 있다. 레이더 차트를 통해 다양한 요소를 한 눈에 비교하여 볼 수 있지만, Figure 15의 레이더 차트에서는 뚜렷하게 경계가 구분되는 색상을 활용하지 않아 해석에 오류가 생길 가능성이 있다.

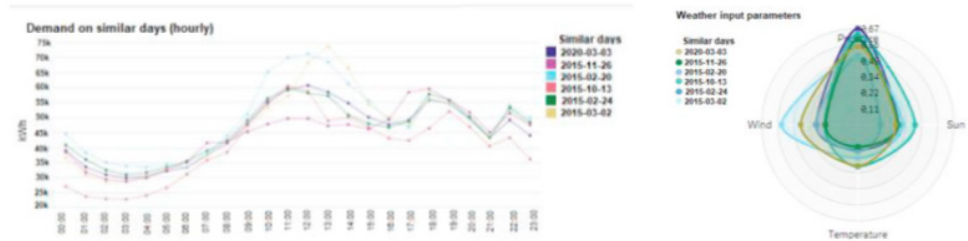


Figure 15 Demand on similar days (Left), Weather input parameters (Right) (Grimaldo & Novak, 2020)

5) 가중치 시각화

Smilkov et al.(2016)이 개발한 'Playground'에서는 뉴런의 활성화 함수 노드와 히든 레이어(hidden layer)를 추가시켜가며 학습 진행과정을 파악할 수 있다. 사용자는 학습률(Learning rate)과 문제유형(Problem type) 등을 선택할 수 있고, 플레이 버튼을 누르면 각 뉴런에서 어떤 종류의 데이터에 대한 가중치가 높아지는지 선의 굵기 변화를 통해 시각적으로 확인할 수 있다(Figure 16). 사용자는 클릭 몇 번으로 쉽게 자신만의 인공신경망을 설계하고 학습, 실행시킬 수 있으며, 인공지능 신경망이 자료를 분류해주는 것을 그래픽으로 보여줌으로써 인공지능 신경망의 원리를 파악하는데 도움을 줄 수 있다.

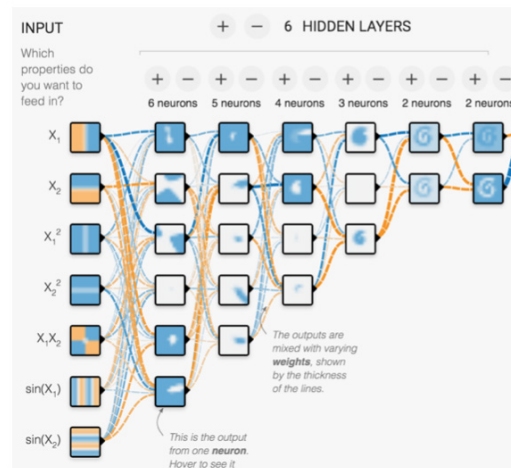


Figure 16 Visualize the learning process as it passes through nodes and layers (playground.tensorflow.org)

6) 시퀀스 기반 데이터 시각화

시퀀스 기반 데이터의 학습 과정을 시각화한 사례로, Kwon et al.(2018)은 선택한 환자 정보를 면적 차트와 코드 바 차트(Code Bar Chart)를 통해 제공한다. 면적 차트에서는 시간이 지남에 따라 산정된 평균과 표준편차를 수평축을 따라 제공하고, 코드 바 차트(Code Bar Chart)에서는 환자들의 공헌인자 9가지를 보여준다(Figure 17). 시간의 흐름에 따른 양적 진전을 면적 차트로 보여주어 사용자가 그 추이를 쉽게 확인하는데 도움을 줄 수 있지만, 한 눈에 직관적으로 파악하기 어려울 수 있다.

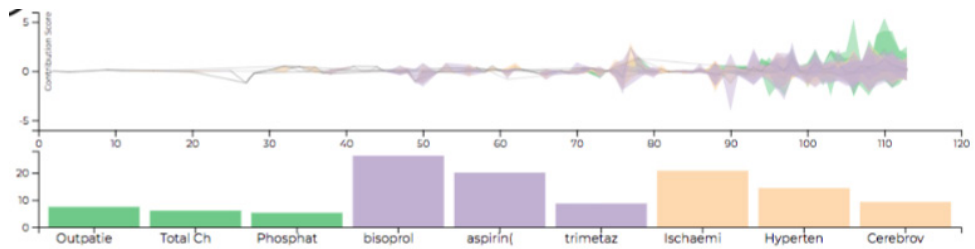


Figure 17 Sequence-based data visualization 1 (Kwon et al., 2018)

3. 3. 추론단계

현 추론 단계에서는 네 가지 목적(분류 결과 시각화, 예측 결과 시각화, 모델 비교 및 선정, 모델 성능 평가)에 따라 시각적으로 다르게 구성하여 제공하고 있다. 먼저 분류 결과를 시각화하기 위해, ‘인스턴스 선택 패널’을 사용하여 오분류된 인스턴스와 올바르게 분류된 인스턴스를 구분 가능하도록 하였고, 결과값이 도출되는 과정에서의 오류가 줄어드는 상황 등을 수치적으로 보여주었다. 예측 결과를 시각화할 때는, 인공지능이 도출한 결과에 대한 근거를 제공함으로써 사용자들이 모델의 예측 결과를 올바르게 해석하는데 도움을 주고자 하였다. 여러 모델을 비교한 후 가장 적합한 모델 선정을 위해서는, 하나의 플롯에서 세 가지 모델을 동시에 볼 수 있도록 구성하여, 데이터 과적합 문제 등을 빠르게 확인하고 해결할 수 있도록 하였다. 마지막으로 모델의 성능을 평가하고 예측 결과의 구체적 확인을 위해서는 Confusion Matrix를 사용하거나 AUC 값 등을 수치적으로 보여주었고 있었다. 추론 단계를 구현한 시각화의 사례들은 다음과 같다.

1) 분류 결과 시각화

Kahng et al.(2017)은 ‘인스턴스 선택 패널’을 통해 분류 결과를 제공하며(오-오분류된 인스턴스, 원-올바르게 분류된 인스턴스), 오분류 된 인스턴스에 마우스 오버 시 그 이유를 확인할 수 있다(Figure 18). 만일 분류에 실수가 있을 경우, 사용자들이 ‘인스턴스 선택 패널’을 통해 인공지능의 결정 과정을 확인하고 이해하는데 도움을 줄 수 있을 것이다.

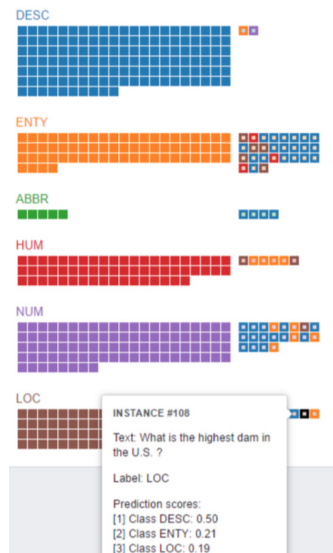


Figure 18 Instance selection panel for visualization of classification results (Kahng et al., 2017)

Smilkov et al.(2016)은 출력 결과를 그래프로 제공하며(Figure 19), 오류가 줄어드는 상황은 ‘OUTPUT’의 ‘Test loss’를 통해 확인할 수 있고, 계산 횟수는 플레이 버튼 옆 ‘Epoch’ 횟수로 알 수 있다. 학습 단계에서 자신이 설계한 인공지능망의 분류 결과를 시각적으로 확인 가능함으로써, 사용자는 어떤 feature가 회오리 무늬를

만들어낼 수 있을지, 몇 겹의 hidden layer가 적당할지 고민해 볼 수 있고, 계산횟수가 증가함에 따라 모델이 어떻게 발전해나가는지 확인해 볼 수 있다.

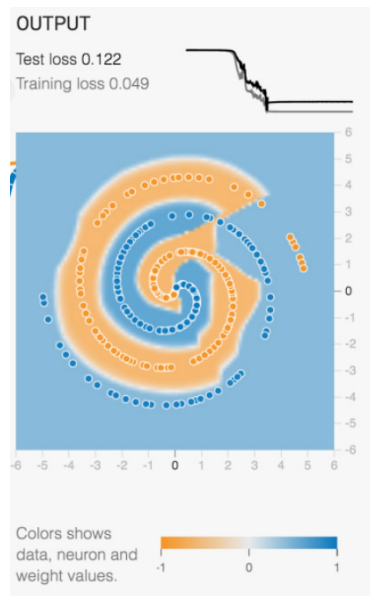


Figure 19 Visualize output results (playground.tensorflow.org)

2) 예측 결과 시각화

Kwon et al.(2018)은 시간의 진행에 따라 예측된 진단 위험을 라인차트(Line Chart)로 제공한다. 라인차트 바로 아래 코드 바 차트(Code Bar Chart)를 통해 9가지 상위 공헌도 의료 코드를 보여주어, 예측된 진단 위험의 진행과 왜 이러한 예측이 이루어졌는지 확인할 수 있다. 만일 사용자가 x축을 따라 이동하면, 시간점까지 이전의 방문에 대한 의료 코드의 업데이트된 공헌도 점수를 확인할 수 있다(Figure 20). 단순 결과값이 아닌 라인 차트와 코드 바 차트를 통해 예측 결과에 대한 근거를 제공하여, 사용자들(의료진)의 효율적이고 효과적인 처방에 도움을 줄 수 있을 것이다.

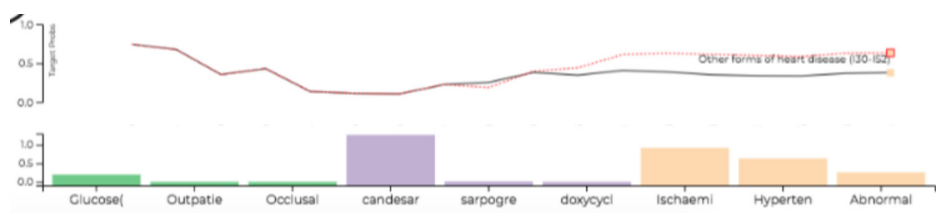


Figure 20 Visualize prediction scores (Kwon et al., 2019)

Wang et al.(2019)은 각 환자별 예측되는 질병 위험도를 색상이 다른 바 차트(Bar Chart)를 통해 제공하며, 어떤 조건을 충족시켜야 환자가 위험 상황(쇼크, 급성 심근 경색 등)을 예방할 수 있을지 텍스트를 통해 제시한다(Figure 21). 인공지능이 제안하는 결과값(예측되는 질병 위험도)만을 제공하는 것이 아닌 마우스 오버 시 환자별 예방법을 의료진에게 제공으로써, 사용자(의료진)들의 잘못된 판단을 줄이는 데 도움을 줄 수 있을 것이다.



Figure 21 System Predicted Risk(Left), Counterfactual Rule(Right) (Wang et al., 2019)

Grimaldo & Novak(2020)은 사용자(기술적 지식이 없는 비전문가)가 예측 결과를 이해하고 분석하는데 도움이 되도록 예측 날짜와 과거 날짜 사이의 평균 에너지 수요를 유사성에 따라 바 차트(Bar Chart)로 정렬하여 보여준다(Figure 22). 예측하는 날이 가장 왼쪽에 표시되며, 소비량이 유사한 과거 날짜 5일이 순서대로 표시된다. 기본적으로 많이 활용되는 바 차트를 제공하여 사용자들이 한 눈에 에너지 수요를 쉽게 비교, 파악하는 것이 가능할 수 있다.

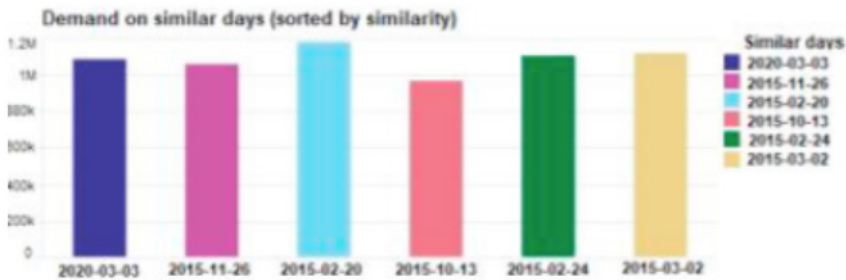


Figure 22 Demand on similar days (Grimaldo & Novak, 2020)

Hohman et al.(2019)의 예측 결과 그래프에서 가로축은 모든 속성(feature)의 카테고리, y축은 최종 예측 값을 나타낸다. Figure 23에서 Instance(인스턴스: 같은 클래스에 속하는 개개의 객체) 550과 798은 속성(feature)이 영향을 미치는 정도에 따라 박스 밝기의 활성화로 구분된다. 결과적으로 waterfall chart가 비슷하게 그려지고, 예측값(190,606과 188,620)도 큰 차이가 없는 것으로 보이지만, 'OverallQual' 피쳐에 마우스 커서를 갖다 댈 시 Instance 550에서는 +22,295로 기여하는 반면, Instance 798에서는 -14,340으로 기여함을 알 수 있다. 즉 인공지능이 도출한 결과에 대한 근거 확인이 가능하기 때문에 사용자들이 모델의 예측 결과를 올바르게 해석하는 데 도움이 될 것이며, 각 속성(feature)이 영향을 미치는 정도에 따라 박스 밝기를 다르게 시각화함으로써 어느 부분이 결정에 중요하게 기여하는지 파악할 수 있다.

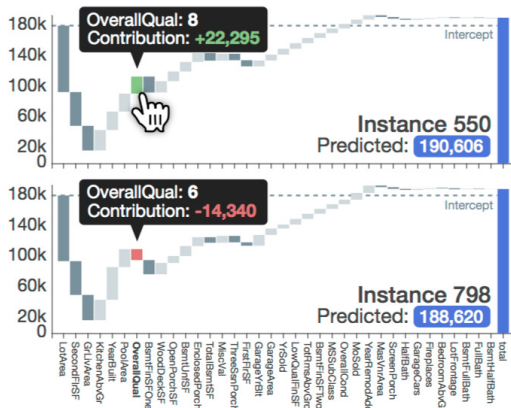


Figure 23 Possible to confirm the basis for the result value derived by artificial intelligence (Hohman et al., 2019)

Krause et al.(2016)의 Prospector에서는 환자별 예측 위험도를 바 차트(Bar Chart)로 제공하며(Figure 24,

좌), 예측된 위험을 가장 크게 감소시킬 수 있는 값에 대해 외곽선이 흰색인 원으로 표시하여 제안한다(Figure 24, 가운데). 또한 사용자가 직접 값을 변경했을 경우 (bmi 값을 0에서 1로 변경), 예측 위험도가 어떻게 변화하는지 오른쪽의 바 차트를 통해 확인할 수 있다(Figure 24, 우).

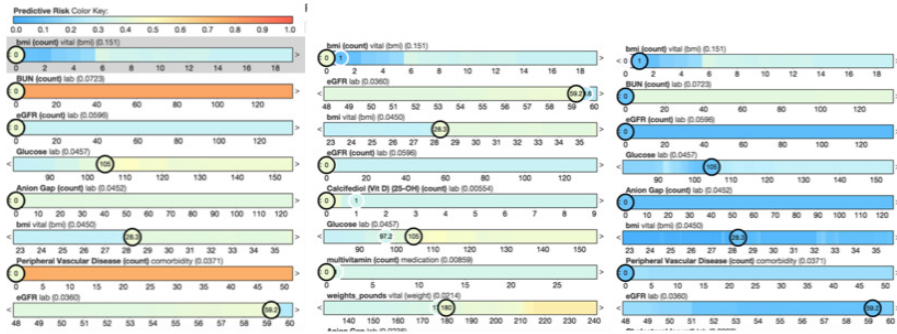


Figure 24 System Predicted Risk(Left), Suggest a value that can reduce the predicted risk (Middle), Predicted risk change as the value changes (Right) (Krause et al., 2016)

3) 모델 성능 평가

구현한 모델의 성능 평가 및 결과값의 구체적 확인을 목적으로 한 사례로, Amershi et al.(2015)은 Confusion Matrix(분류 모델의 정확성을 평가할 때 활용하는 오차행렬)를 제공한다. Figure 25의 AUC값은 모델 성능을 요약하여 보여주는 것으로 점수가 높을수록 수행능력이 좋음을 의미하며, 오른쪽의 'Precision Recall Curves'는 알고리즘의 정밀도(Precision)와 재현율(Recall)의 관계를 나타낸 것이다. Confusion Matrix를 단순 텍스트로 제공하는 것이 아닌 집계한 값에 비례해서 색상이 다른 2차원 자료로 시각화하여 모델이 얼마나 정밀하고, 얼마나 실용적인지 한 눈에 파악하는데 도움을 줄 수 있다. Precision Recall Curves의 경우, 모델 비교 후 선택하고자 할 때 유용하게 사용될 수 있다.

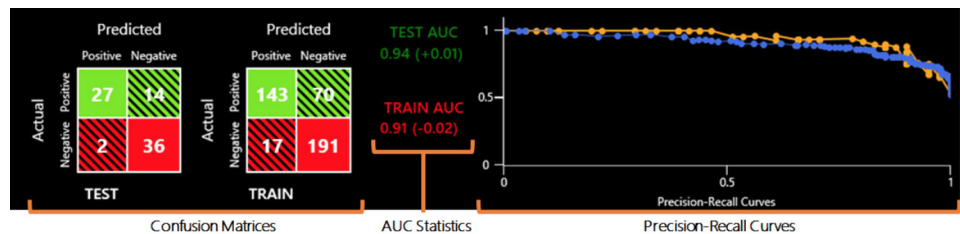


Figure 25 Confusion Matrix for model performance evaluation (Amershi et al., 2015)

Siang et al.(2018)은 분류 성능이 좋지 않은 동작에 대한 예측 결과의 구체적 확인을 위해 Confusion matrix를 제공한다. 세로축은 실제 동작을, 가로축은 모델이 분류한 동작을 의미한다(Figure 26, 좌). (Figure 26, 우)를 보면 sitting이 standing으로 잘못 분류되어 Recall(관련 있는 것으로 분류된 항목들 중 실제 검색된 항목들의 비율)값이 가장 낮게 나온다. 하지만 standing은 sitting으로 오분류되지 않는 것을 보아, 학습한 DNN 모델이 두 가지 동작(standing, sitting) 중에서 standing으로 분류하는 성향이 높음을 알 수 있다. 이를 통해 모델이 얼마나 정밀하고, 얼마나 실용적인지 한 눈에 파악하는데 도움을 줄 수 있다.

동작	Precision	Recall
walking	0.899	0.972
Walking upstairs	0.946	0.896
walking downstairs	0.959	0.950
sitting	0.984	0.741
standing	0.807	0.977
laying	0.998	0.994
평균	0.932	0.922

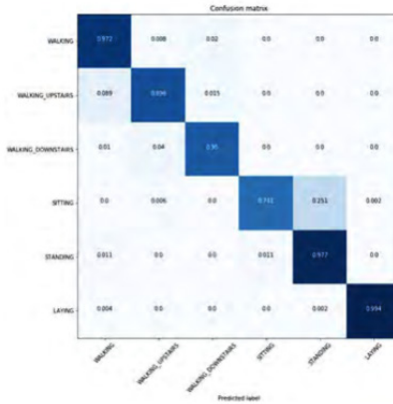


Figure 26 Confusion Matrix for concrete confirmation of prediction results (Siang et al., 2018)

4) 모델 비교 및 선정

모델의 성능 비교를 가능하게 한 사례로, Krause et al.(2016)은 하나의 플롯(plot, 여러 값 사이의 수치적인 관계를 좌표계에 점으로 찍어서 선 또는 면으로 나타낸 것)에서 세 가지 모델을 동시에 볼 수 있도록 구성하였다. 이를 통해 사용자는 본인이 진행하는 프로젝트에서의 적합한 모델을 선택하는데 도움을 얻을 수 있다(Figure 27).

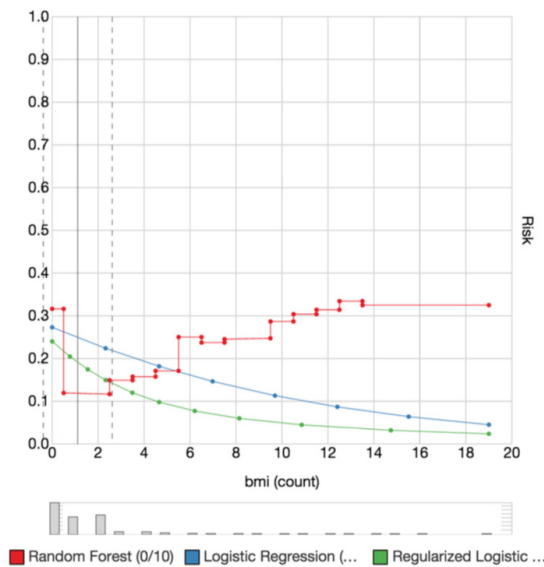


Figure 27 Multiple models can be compared (Krause et al., 2016)

4. 논의 및 결론

인공지능에 대한 기대와 여러 혁신 사례에도 불구하고, 지금까지의 인공지능은 의사결정을 지원할 때 결과만 알려주고 어떤 근거로 이 판단에 이르렀는지 논리적으로 설명할 수 없는 점이 한계로 지적되며 “블랙박스”로 취급되었다. 이에 인공지능의 신뢰와 직결되는 ‘설명 가능성’이 인공지능 분야의 화두로 떠오르고 있지만, 인공지능 프로세스를 시각화할 때 어떻게 접근해야 할지에 대한 학술적인 논의는 매우 부족한 실정이다. 따라서 본 연구에서는 인공지능 프로세스를 시각화한 선행연구 분석을 통해 각 과정별 시각화 접근방법을 살펴보고자 하

였다.

연구목적 달성을 위해 머신러닝과 딥러닝 모델링 프로세스를 다룬 논문들을 조사하여, 인공지능의 수행과정을 ② 탐색단계 (데이터 수집, 전처리, 다양한 기계학습 기법에 적용 등 모형개발 전 단계에 관한 시각화) ① 학습단계 (특징 추출, 특징 선택 등 모형 선택 후 실제 데이터 학습 과정에서의 시각화) ③ 추론단계 (알고리즘의 정확성과 유의미성 평가 등 모형의 성능에 관한 시각화)로 구분 후, 선정된 선행연구에서의 시각화 방법을 각 단계별로 살펴보았으며, 각 사례별 장단점을 기술함으로써 머신러닝 시각화 도구의 디자이너가 개별 디자인 사례로부터 무엇을 배울 수 있을지 파악하는데 도움을 주고자 하였다.

결과적으로, 각 단계별 목적(모델 작동 이해/오분류 원인 파악, 디버깅 지원 등) 및 대상(전문가/비전문가)에 따라 데이터 시각화 제공 방법에 차이가 있는 것으로 나타났다. 구체적으로, 머신러닝 실무자를 대상으로 한 연구의 경우 모델작동 이해 및 모델 오분류 원인 파악에 도움을 주기 위한 목적으로 시각화를 진행하였으며, 데이터 과학자를 대상으로 한 연구의 경우 예측모델 판단과 해석에 도움을 주는데 목적을 두었다. 비전문가를 대상으로 한 연구의 경우, 인공지능 기술 이해에 도움을 주거나 결과값에 대한 근거를 제시함으로써 예측모델에 대한 신뢰도를 높이기 위한 목적으로 시각화를 진행하였다. 조사한 연구의 대부분은 최종 사용자가 아닌 데이터 과학자 전문가에 초점을 맞춰 진행되었기 때문에, 시각화 툴의 이해를 위해서는 전문적인 지식이 필요함을 알 수 있었다.

인공지능의 수행과정을 시각화한 선행 연구들은 비교적 최근의 연구들이다. 따라서 향후 인공지능의 수행과정에 따른 시각화 전략 수립과 이론적 토대 마련을 위해서는 다음의 사항에 대한 논의가 필요해 보인다.

첫 번째, 최종 사용자 중심의 시스템 디자인 제안이 필요하다. 앞서 말했듯 인공지능 수행과정 시각화의 중요성을 인지한 다수의 연구자들은 다양한 프로젝트와 툴을 개발하고 있지만, 시도 사례의 공통점은 매우 복잡해 보이고 데이터 과학자 전문가를 위한 시각화에 가깝다는 점이다. 인공지능의 역할이 점차 확대됨에 따라 투명성과 신뢰가 요구되는 분야(e.g., 의료 분야, 금융 분야 등)에서 사용자-중심 디자인의 중요성은 높아지고 있다. Waller et al. (2006)에 의하면, 사용자-중심 디자인적 접근은 디자인 과정에서 반복적이고 지속적인 사용자 평가에 의해 소프트웨어 디자인의 활용도를 높일 수 있다고 하였다. 따라서 인공지능 기반의 시각화 툴을 직접 사용하게 될 최종 사용자(e.g., 의료 분야에서는 의료진, 금융 분야에서는 펀드매니저 등)들의 행태와 환경을 이해한 후, 그들의 입장에서 가장 필요한 정보를 파악하여, 그들이 이해할 수 있는 방식의 설명 가능한 모델을 제작하고 시스템을 디자인하는 것이 필요한 시점임을 알 수 있다.

두 번째, 모델의 훈련 과정에서 사용자가 직접 버그를 찾아서 수정하거나 잘못된 부분을 찾아 고치는 등의 상호작용 가능한 시스템은 많지만, 인공지능이 제시한 결과값을 얻고 난 후 추가적으로 특징값을 조정했을 때 나타나는 예측값의 변화를 확인할 수 있는 시스템은 비교적 적음을 알 수 있었다. 따라서 Krause et al.(2016)의 연구에서처럼 모델의 훈련 과정에서뿐 아니라 모델이 제시한 결과값에 대해서도 사용자가 직접 피쳐값을 조정해 보며 변화를 주어 최악의 상황까지 미리 진단해볼 수 있도록 제작한다면, 모델에 대한 사용자의 신뢰도를 높일 수 있을 것으로 기대한다.

세 번째, 모형 개발 전 단계에 관한 시각화가 필요하다. 모델 제작 시 데이터 세트를 구분하고 다양한 기계학습 기법에 적용해보는 것 등은 필수적으로 거치는 단계이다. 인공지능 수행과정별 데이터시각화 전략 수립을 위해 인스턴스의 배치 등을 더욱 다양한 관점에서 시각적으로 보여주는 시도가 필요하지만, 아직까지는 알고리즘의 정확성과 같은 결과적인 부분에 더 초점을 맞춘 연구가 진행되고 있다. 앞서 말했듯 프로세스의 시각화는 사용자가 모델의 작동을 얼마나 이해할 수 있는가에 중요한 영향을 미치기 때문에, 탐색단계의 시각화부터 더욱 다양하게 시도될 필요가 있다.

본 연구에서 도출된 항목들은 앞으로 인공지능의 수행과정에 따른 데이터시각화 디자인 사례를 분석하기 위한 기초 자료로 사용될 것이다. 본 연구를 토대로 향후 더욱 다양한 관점에서의 사례 분석이 진행된다면 인공지능의 수행과정을 시각화하기 위한 전략 수립과 이론적 토대 마련이 가능할 것으로 기대한다. 향후 연구에서는 머신러닝 관련 시각화 전반을 분석하는 것이 아닌 Explainable AI 분야로 한정하여 사례 분석을 실시할 예정이다.

References

1. Amershi, S., Chickering, M., Drucker, S. M., Lee, B., Simard, P., & Suh, J. (2015). Modeltracker: Redesigning performance analysis tools for machine learning. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 337–346. ACM.
2. Byun, J. Y., & Park, Y. B. (2016). A guiding system of visualization for quantitative bigdata based on user intention. *KIPS Transactions on Software and Data Engineering*, 5(6), 261–266.
3. Choo, J., Lee, H., Kihm, J., & Park, H. (2010). iVisClassifier: An interactive visual analytics system for classification based on supervised dimension reduction. In *2010 IEEE Symposium on Visual Analytics Science and Technology*, 27–34. IEEE.
4. Došilović, F. K., Brčić, M., & Hlupić, N. (2018). Explainable artificial intelligence: A survey. In *2018 41st International convention on information and communication technology, electronics and microelectronics (MIPRO)*, 0210–0215. IEEE.
5. Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639), 115.
6. George Gilbert (2016). Machine Learning Pipeline: Chinese Menu of Building Blocks. Retrieved from <https://wikibon.com/machine-learning-pipeline-chinese-menu-of-building-blocks/>
7. Gibson, E., Li, W., Sudre, C., Fidon, L., Shakir, D., Wang, G., Rosen, Z., Gray, R., Doel, T., Hu, Y., Whyntie, T., Nachev, P., Modat, M., Barratt, D., Ourselin, S., Cardoso, M., & Vercauteren, T. (2018). NiftyNet: a deep-learning platform for medical imaging. *Computer methods and programs in biomedicine*, 158, 113–122.
8. Grimaldo, A. I., & Novak, J. (2020). Combining Machine Learning with Visual Analytics for Explainable Forecasting of Energy Demand in Prosumer Scenarios. *Procedia Computer Science*, 175, 525–532.
9. Gulshan, V., Peng, L., Coram, M., Stumpe, M. C., Wu, D., Narayanaswamy, A., Venugopalan, S., Widner, K., Madams, T., Cuadros, J., Kim, R., Raman, R., Nelson, P., Mega, J., & Webster, D. (2016). Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316(22), 2402–2410.
10. Hohman, F., Head, A., Caruana, R., DeLine, R., & Drucker, S. M. (2019). Gamut: A design probe to understand how data scientists understand machine learning models. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, 1–13.
11. Jang, S., & Yun, J. (2018). AI Design Strategy for Investment Judgement Support Aid. *Proceedings od HCI Korea 2018*, 62–65.
12. Jung, S., Byun, J., & Kim, C. (2019). Introduction of explainable Artificial Intelligence technology. *The Magazine of the IEK*, 46(2), 55–63.
13. Kahng, M., Andrews, P. Y., Kalro, A., & Chau, D. H. P. (2017). ActiVis: Visual Exploration of Industry-Scale Deep Neural Network Models. *IEEE transactions on visualization and computer graphics*, 24(1), 88–97.
14. Krause, J., Perer, A., & Ng, K. (2016). Interacting with predictions: Visual inspection of black-box machine learning models. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 5686–5697. ACM.
15. Kwon, B. C., Choi, M. J., Kim, J. T., Choi, E., Kim, Y. B., Kwon, S. W., Sun, J., & Choo, J. (2018). Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records. *IEEE transactions on visualization and computer graphics*, 25(1), 299–309.
16. Lee, H., Chung, S., & Choi, E. (2016). A Case Study on Machine Learning Applications and Performance Improvement in Learning Algorithm. *Journal of Digital Convergence*, 14(2), 245–258.
17. Lee, S., & Pan, Y. (2019). A Case Study of Screen-based Web & App Design System and Artificial Intelligence-based design software. *Journal of Korea Society of Design Trend*, 63, 139–150.
18. Lee, S., Kim, D., Cho, Y., Myung, J., Moon, D., Lee, J., & Yoon, M. (2019). Machine learning based security data analysis research. *Korea Institute of Information Security and Cryptology*, 29(3), 6–13.
19. Liu, Y., Jeon, H., & Pan, Y. (2018). A Smart Healthcare Service with O2O – Focus on Healthcare Company in Shenzhen. *Journal of the Korean Society Design Culture*, 24(2), 231–239.
20. Min, Z. (2018). Guide and Retain Users: Interactive Recommender System. In *Proceedings of the*

- 2018 10th International Conference on Computer and Automation Engineering, 44–48. ACM.
21. Mun, S., & Lee, K. (2016). Data analysis by Integrating statistics and visualization: Visual verification for the prediction model. *Design Convergence Study*, 15(6), 195–214.
 22. "Open the Black Box of AI", Tech M, Retrieved January, 2017, from <https://www.techm.kr/news/articleView.html?idxno=3574>
 23. Park, J., Park, D., Choi, J., Kim, S., & Kim, J. (2018). A case study on an effective method of collecting user's message data and a data utilization. *Proceeding of HCI Korea*, 977–981.
 24. Ryu, J., Shin, H., Kim, M., & Baek, J. (2017). Patter Analysis of Stock Prices Using Machine Learning and Data Visualization. *Journal of Information Technology and Architecture*, 14(2), 189–197.
 25. Siang, C., Shin, G., Kim, Y., & Yun, M. (2018). Human Activity Recognition Using Deep Neural Network. *Proceeding of HCI Korea*, 716–720.
 26. Smilkov, D., Carter, S., Sculley, D., Viégas, F. B., & Wattenberg, M. (2017). Direct-manipulation visualization of deep networks. *arXiv preprint arXiv:1708.03788*.
 27. Song, K., Kim, M., & Do, S. (2019). The Latest Trends in the Use of Deep Learning in Radiology Illustrated Through the Stages of Deep Learning Algorithm Development. *Journal of the Korean Society of Radiology*, 80(2), 202–212.
 28. Tzeng, F. Y., Lum, E. B., & Ma, K. L. (2005). An intelligent system approach to higher-dimensional classification of volume data. *IEEE Transactions on visualization and computer graphics*, 11(3), 273–284.
 29. Wang, D., Weisz, J. D., Muller, M., Ram, P., Geyer, W., Dugan, C., Tausczik, Y., Samulowitz, H., & Gray, A. (2019). Human-AI Collaboration in Data Science: Exploring Data Scientists' Perceptions of Automated AI. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–24.
 30. Yun, T., & Ahn, H. (2018). Fake New2112s Detection for Korean News Using Text Mining and Machine Learning Techniques. *Journal of Information Technology Applications & Management*, 25(1), 365–368.

인공지능 수행과정 시각화에 대한 디자인 사례 연구

박윤하¹, 윤재영^{2*}

¹홍익대학교 디자인학부 시각디자인 전공, 학생, 서울, 대한민국

²홍익대학교 디자인학부 시각디자인 전공, 교수, 서울, 대한민국

초록

연구배경 지능정보사회(intelligent information society)로 나아감에 따라 인공지능은 다양한 분야에서 미래의 정보사회를 이끌어 나갈 기술로 주목받고 있다. 이에 최근 몇 년간 AI 기반의 여러 혁신 사례가 지속적으로 발표되고 있지만, 지금까지의 인공지능은 의사결정을 지원할 때 결과만 알려주고 어떤 근거로 이 판단에 이르렀는지 논리적으로 설명할 수 없는 점이 한계로 지적되며 ‘블랙박스’로 취급되었다. 이에 모델링 해석의 필요성과 중요성이 높아지고 있지만, 그 프로세스를 시각화할 때 어떻게 접근해야 할지에 대한 학술적 논의는 매우 부족한 실정이다. 따라서 본 연구에서는 인공지능의 수행과정을 시각화한 선행연구 분석을 통해 각 과정별 시각화 접근 방법을 살펴보고자 한다.

연구방법 본 연구에서는 일반적인 예측 모델링 프로젝트에서의 머신러닝과 딥러닝 파이프라인(Pipeline)을 살펴보고, 이를 적용한 연구들을 조사하여, 인공지능의 수행과정을 세 단계(탐색, 학습, 추론)로 구분하였다. 그리고 각 단계별로 어떻게 시각화가 이루어지고 있는지 리서치 프레임워크(Research Framework)를 토대로 분석하였다.

연구결과 인공지능 파이프라인의 시각화를 다룬 논문들을 조사하여, 인공지능의 수행과정을 ① 탐색단계(데이터 수집, 전처리, 다양한 기계학습 기법에 적용 등 모형개발 전 단계에 관한 시각화) ② 학습단계(특징 추출, 특징 선택 등 모형 선택 후 실제 데이터 학습 과정에서의 시각화) ③ 추론단계(알고리즘의 정확성과 유의미성 평가 등 모형의 성능에 관한 시각화)로 구분하였고, 각 단계별 시각화 방법을 살펴본 결과 목적(모델 작동 이해/오분류 원인 파악, 디버깅 지원 등) 및 대상(전문가/비전문가)에 따라 데이터시각화 제공방법에 차이가 있는 것으로 나타났다.

결론 본 연구에서 도출된 항목들은 앞으로 인공지능의 수행과정에 따른 데이터시각화 디자인 사례를 분석하기 위한 기초 자료로 사용될 것이다. 본 연구를 토대로 향후 더욱 다양한 관점에서의 사례 분석이 진행된다면 인공지능의 수행과정을 시각화하기 위한 전략 수립과 이론적 토대 마련이 가능할 것으로 기대한다.

주제어 인공지능, 데이터 시각화, 모델링 해석

이 논문은 2018년 대한민국 교육부와 한국연구재단의 지원을 받아 수행된 연구임(NRF-2018S1A5A8029352)

*교신저자 : 윤재영 (ryun@hongik.ac.kr)