

인공지능 학습데이터 라벨링 정확도에 따른 인공지능 성능

이지훈¹, 신지은^{2*}

¹건양대학교 의과대학 정보의학교실 석사과정 학생

²건양대학교 의과대학 정보의학교실 교수, 건양대학교 병원 헬스케어데이터 검증센터 센터장

AI Performance Based On Learning-Data Labeling Accuracy

Ji-Hoon Lee¹, Jieun Shin^{2*}

¹Student, Dept. of Biomedical Informatics, College of Medicine, Konyang University, The Master's Course

²Professor, Dept. of Biomedical Informatics, College of Medicine, Konyang University, Ph.D and The Head of HealthCareData Verification Center

요약 본 연구는 데이터의 품질이 인공지능(AI) 성능에 미치는 영향을 검토한다. 이를 위해, 데이터 특성변수(Feature)의 유사도와 클래스(Class) 구성의 불균형을 고려한 모의실험(Simulation)을 통해 라벨링 오류 수준이 인공지능의 성능에 미치는 영향을 비교 분석하였다. 그 결과, 특성변수 간 유사성이 높은 데이터에서는 특성 변수 간 유사성이 낮은 데이터에 비해 라벨링 정확도에 더 민감하게 반응하였으며, 클래스 불균형이 증가함에 따라 인공지능 정확도가 급격히 감소되는 경향을 관찰하였다. 이는 인공지능 학습데이터의 품질평가 기준 및 관련 연구를 위한 기초자료가 될 것이다.

키워드 : 인공지능, 학습데이터, 데이터 품질, 라벨링 정확도, 데이터

Abstract The study investigates the impact of data quality on the performance of artificial intelligence (AI). To this end, the impact of labeling error levels on the performance of artificial intelligence was compared and analyzed through simulation, taking into account the similarity of data features and the imbalance of class composition. As a result, data with high similarity between characteristic variables were found to be more sensitive to labeling accuracy than data with low similarity between characteristic variables. It was observed that artificial intelligence accuracy tended to decrease rapidly as class imbalance increased. This will serve as the fundamental data for evaluating the quality criteria and conducting related research on artificial intelligence learning data.

Key Words : Artificial intelligence, Learning data, Data quality, Labeling accuracy, Data

*Corresponding Author : Ji-Eun Shin(jeshin@konyang.ac.kr)

Received December 11, 2023

Accepted January 20, 2024

Revised December 28, 2023

Published January 28, 2024

1. 서론

1.1 연구의 필요성

인공지능의 발전은 다양한 분야에서 수많은 이들에게 혜택을 제공하고 있다. Apple Siri와 Amazon Alexa와 같은 개인 비서의 형태를 가진 인공지능 서비스를 통해 사용자에게 특화되어 다양한 분야의 정보를 제공하며, 채팅을 통한 챗봇은 뉴스와 같은 정보나 헬프데스크 서비스를 지원하는 단계까지 운영되고 있다[1]. 그러나 인공지능의 혜택만큼이나 문제점도 존재한다. 스캐터랩 기업에서 개발한 일상대화형 챗봇 '이루다'는 서비스 이후 동성애, 장애인 등에 대한 차별과 혐오 발언 등 다양한 오류로 인한 문제들이 발생하여 서비스를 종료하였다[2]. 또한 마이크로소프트의 인공지능 챗봇 '테이'는 사용자에 의한 편향된 학습데이터로 다양한 차별과 혐오 발언을 쏟아내는 문제를 발생시켰다[3]. 이같이 인공지능의 오류는 학습하는 데이터의 오류로 인한 편향, 즉 데이터의 낮은 품질이 하나의 원인이다[4, 5].

인공지능 모델은 초기 학습을 위한 학습데이터, 학습을 검증하기 위한 검증데이터, 그리고 결과를 확인하기 위한 테스트데이터로 나눈다[6]. 학습, 검증, 테스트데이터는 데이터와 정답 값을 가지는 데이터 라벨(Label)로 구성되고, 테스트데이터 없이 하나의 데이터셋을 k개로 나누어 모델을 평가하기도 한다[6]. 특히 학습데이터는 인공지능의 핵심 부분으로, 데이터의 품질이 매우 중요하다. 이에 따라 해외의 경우 데이터셋의 구축을 위한 공정 과정을 정의하고 있으며, 정부와 민간의 협업과 투자를 통해 다양한 학습데이터 세트를 구축하고 있다[7]. 국내에서는 한국지능정보사회진흥원과 정부가 데이터 품질에 대한 기준을 제시하고 지원하고 있으나[8], 데이터 품질과 관련된 문제는 여전히 발생하고 있다.

라벨링은 작업목적의 명료함이나 작업자의 능력과 작업환경에 따라 다양한 라벨링 품질의 문제가 발생할 수 있다[9]. 또한 라벨링은 주로 수작업을 통해 이루어지며, 이로 인해 발생하는 인적 오류는 데이터 라벨링에 반영되어 데이터의 품질을 낮출 수 있다[10]. 낮은 품질의 데이터로 학습한 인공지능 모델은 데이터가 내포한 오류를 학습하게 되어 편향된 결과를 초래하거나 학습 자체에 문제를 일으킬 수 있다[11].

이에 본 연구에서는 라벨링 오류에서 기인한 낮은 데이터 품질이 인공지능 학습 결과에 미치는 영향을 검토하

고자 모의실험을 통해 분석하고, 라벨링 오류와 관련된 품질지표 수준을 마련하기 위한 기초자료를 제공하고자 한다.

1.2 연구 목적

본 연구의 목적은 데이터 라벨링 정확도가 인공지능 학습 결과에 미치는 영향을 검증하는 데에 있다. 이를 위해 다음 두 가지 목표를 설정하였다.

첫째로, 집단(class) 간 특성변수(feature)의 유사성 정도에 따라 데이터 라벨링의 정확도가 낮아지는 경우, 인공지능 학습 결과에 어떠한 차이가 나타나는지 확인한다. 현재 사용되는 품질 지표는 특성변수의 유사성 변화에 따른 라벨링 오류 비율을 고려하지 않기 때문에, 이 측면에서의 영향을 검토하고자 한다.

둘째로, 집단(class)을 구성하는 자료의 개수 불균형이 인공지능 학습 결과에 미치는 영향을 확인한다. 현재의 품질 지표는 자료 개수의 불균형을 고려하지 않고 사용되고 있어, 이러한 불균형 변화에 따른 라벨링 오류 비율이 인공지능 학습 결과에 어떻게 영향을 미치는지 살펴본다. 이를 통해 집단별 자료의 개수 불균형이 학습 결과에 미치는 영향을 검토한다.

2. 연구방법

본 연구에서는 실제 데이터에서 발생할 수 있는 데이터의 특성을 반영하여 생성한 데이터를 활용하여 모의실험(Simulation)을 통한 데이터 라벨링 정확도가 인공지능 학습 결과에 미치는 영향을 검증한다.

2.1 모의실험(Simulation) 데이터 생성

본 연구를 위해 생성하는 데이터는 '특성변수(feature)'와 3개의 '집단(class)'으로 구성되어 있다. 여기서 '특성변수'는 인공지능의 정답 예측에 영향을 미치는 변수들을 나타내며, '집단'은 인공지능 모델이 예측하려는 목표를 나타낸다.

본 모의실험 데이터는 세 개의 집단(A, B, C)을 사용하였으며, 각 집단은 2,500개의 개체(case)로 구성되었다. 각 개체는 다음과 같은 과정을 통해 6개의 특성변수를 갖도록 생성하였다(Table 1 참고).

Table 1. Example of simulation data class and feature

	Feature1	Feature2	Feature3	Feature4	Feature5	Feature6	Class
1	0.52	0.22	0.60	0.85	0.34	0.83	A
2	0.11	0.15	0.44	0.89	0.91	0.71	A
...
2500	0.77	0.28	0.49	0.48	0.36	0.30	A
1	0.17	0.75	0.40	0.50	0.85	0.01	B
2	0.48	0.10	0.14	0.78	0.41	0.59	B
...
2500	0.03	0.47	0.87	0.74	0.41	0.46	B
1	1.06	1.01	1.63	0.61	1.76	1.17	C
2	1.54	0.56	1.71	1.14	0.86	1.23	C
...
2500	1.38	0.90	1.36	1.64	0.81	0.23	C

Each row has six feature variables and represents an object (case) that belongs to a certain class, and each column represents a variable that represents a feature variable(Feature1~Feature6) and a class(figures are for illustrative purposes only).

- 1단계 : 각 개체는 6개의 '특성변수(feature)'를 생성하기 위해 0부터 1 사이의 범위에서 균등분포 [12]를 따르는 6개 난수를 생성하여 사용한다.
- 2단계 : 6개의 '특성변수'를 가진 총 5,000개의 개체를 생성한다.
- 3단계 : '특성변수' 분포가 유사한 집단을 생성하기 위해 생성된 5,000개의 개체를 임의로 2개의 '집단(A, B)'으로 배정한다.
- 4단계 : '특성변수' 분포가 상이한 집단을 생성하기 위해 '집단 B'의 6개 '특성변수' 값에 2배를 한 값을 사용하여, 2,500개의 개체를 갖는 '집단 C'를 생성한다.

여기서, 집단 A, B, C가 인공지능 모델이 예측하려는

목표로 라벨링 값이다.

생성된 모의실험 데이터 특성을 시각적으로 분석하기 위해, 높은 차원의 데이터 거리를 확률분포를 이용하여 2~3차원으로 축소시켜 시각화하는 매니폴드 학습 기법인 T-SNE를 활용하였다[13]. 모의실험에 사용되는 데이터의 특성은 집단 A와 B의 경우 특성변수 분포가 비슷한 데이터로 2D와 3D 그래프 모두 구별하기 어렵게 나타난다(Fig. 1, Fig. 2 참고). 반면, 집단 A와 C는 특성변수의 분포가 상이한 데이터로 2D와 3D 그래프 모두 구별하기 쉽게 나타났다(Fig. 1, Fig. 2 참고).

2.2 인공지능 학습

인공지능 모델 학습 결과에 미치는 영향을 확인하기 위해 서포트 벡터 머신(SVM) 모델을 사용하였다[14].

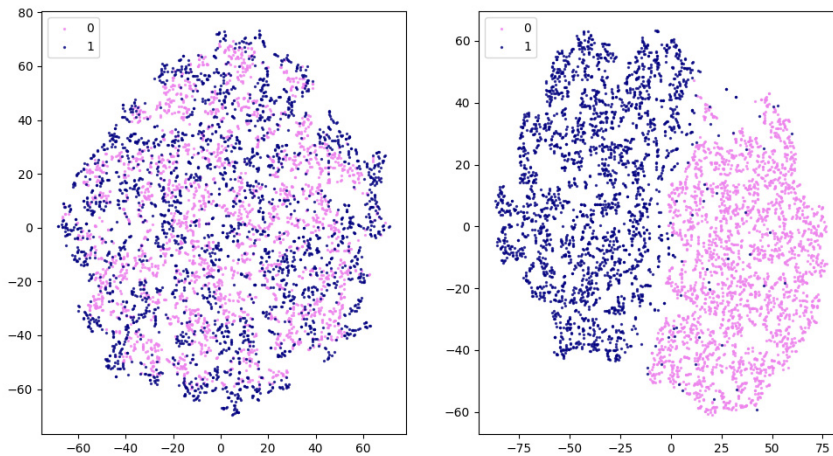


Fig. 1. 2D results based on data classification using T-SNE / Left. Similar groups(class A, B) / Right. Different groups (class A, C)

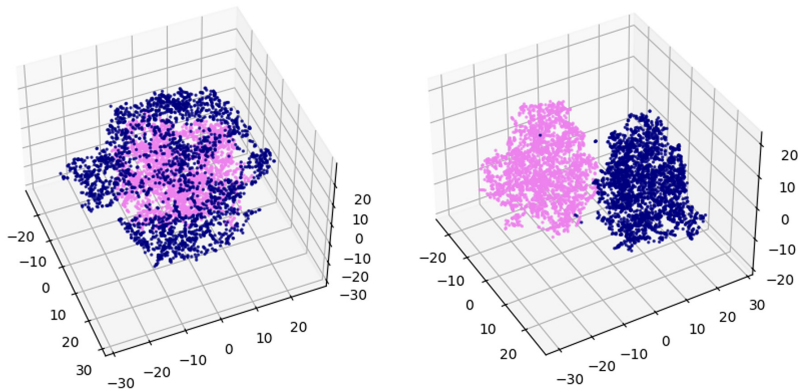


Fig. 2. 3D results based on data classification using T-SNE / Left. Similar groups(class A, B) / Right. Different groups (class A, C)

SVM은 분류와 회귀 등 다양한 조건에서 높은 성능을 가지는 지도학습 모델로 특정 조건을 만족하며 클래스를 분류하는 초평면(Hyperplane)을 찾는 기법이다[14].

2.3 모델 성능 평가

모델의 성능 평가를 위해 예측 결과와 테스트데이터 값의 일치 정도를 확인하는 정확도(accuracy)를 측정한다. 학습과 성능 평가를 위해 각 집단에 대해 학습데이터(Training data)와 테스트데이터(testing Data)로 구분하여 4:1 비율로 나누어 적용하였다.

2.4 라벨링 오류율 조정

본 연구에서 생성한 모의실험 데이터의 초기 라벨링 정확도는 100%이다. 특성변수 분포가 유사한 집단인 A와 B의 라벨을 교차적으로 변경함으로 A의 라벨을 B로, B의 라벨을 A로 바꾸어 라벨링 정확도를 감소시켜 오류율을 증가시켰다. 이러한 과정을 ‘교차변경’으로 지칭한다. 교차변경을 통해 라벨링 오류율은 0%부터 시작하여 10%p씩 증가시켜 최대 50%까지 높인다. 특성변수 분포가 상이한 두 집단인 A와 C에 대해서도 동일한 방식으로 라벨링 오류율을 조정한다.

2.5 클래스 불균형

집단(class)을 구성하는 데이터의 개수에 불균형을 만들기 위해, 각 집단의 구성 비율을 1:1, 1:2, 1:5, 1:10으로 조정한다.

3. 연구결과

3.1 특성변수 분포가 유사한 두 집단 비교

특성변수가 유사성을 가지는 집단의 경우, 라벨링을 교차변경하지 않은 모델(기준 모델)에서 인공지능 모델의 정확도는 약 92%로 나타났다. 교차변경 10%와 20%는 기준모델과 비교해 인공지능 모델의 정확도 수준이 1%p 이하의 차이로 비슷하게 나타났다. 교차변경 30%에서 인공지능 모델의 정확도는 약 89%로 기준모델보다 정확도가 약 3%p 낮아졌다. 교차변경 40%에서 인공지능 모델의 정확도는 약 81%로 기준모델보다 정확도가 약 12%p 낮아졌다. 또한, 교차변경 50%에서 인공지능 모델의 정확도는 약 53%로 기준모델보다 정확도가 약 39%p 낮아졌다(Fig. 3, Table 2 참고).

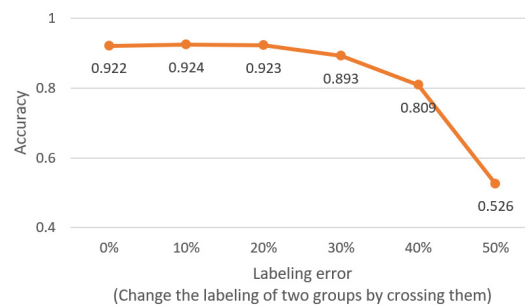


Fig. 3. Classification using random number data, ambiguous groups (class A, B) / a: data ratio 1:1

3.2 특성변수 분포가 상이한 두 집단 비교

특성변수가 상이함을 갖는 집단의 경우, 라벨링을 교

차변경하지 않은 모델(기준 모델)에서 인공지능 모델의 정확도는 99.7%로 나타났다. 교차변경 10~30%에서는 기준모델과 비교해 인공지능 모델의 정확도 수준이 1%p 이하의 차이로 비슷하게 나타났다. 교차변경 40%에서 인공지능 모델의 정확도는 약 94%로 기준모델보다 정확도가 약 6%p 낮아졌다. 교차변경 50%에서 인공지능 모델의 정확도는 약 54%로 기준모델보다 정확도가 약 45%p 낮아졌다(Fig. 4, Table 2 참고).

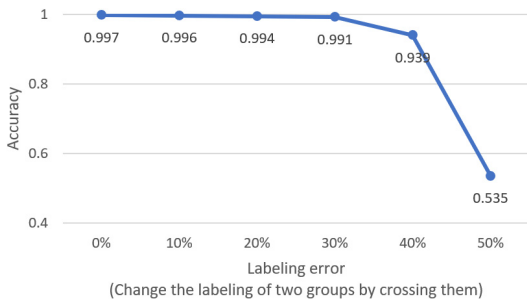


Fig. 4. Classification using random number data, distinguishable groups (class A, C) / a: data ratio 1:1

Table 2. Classification result of random number data / data ratio 1:1

Flips Size	Ambiguous Data / Ambiguous Data(A, B)					
	0%	10%	20%	30%	40%	50%
1:1	0.922	0.924	0.923	0.893	0.809	0.526
Size	Distinguishable Data / Distinguishable Data(A, C)					
	1:1	0.997	0.996	0.994	0.991	0.939

3.3 클래스 불균형

3.3.1 특성변수 분포가 유사한 두 집단 비교

특성변수가 유사성을 가지는 집단의 경우, 1:1 클래스 비율을 가지는 균형적 집단을 기준으로 비교하면 인공지능 모델의 정확도는 각 교차변경 수준(0%, 10%, 20%, 30%, 40%)에서 클래스 불균형 비율이 높아질수록 감소하는 것으로 나타났다. 반면, 교차변경 50%에서는 클래스가 균형적인 경우(1:1) 인공지능 모델 정확도가 약 52%로 낮았으며 클래스 불균형 비율이 높아져도 인공지능 모델 정확도는 약 5%p 이내의 차이로 감소하였다.

클래스 불균형 비율 1:2는 균형비율(1:1)과 모든 교차변경 수준에서 인공지능 모델 정확도 차이가 약 3%p 이내의 차이로 비슷한 수준으로 나타났다. 반면 클래스 불균형 비율 1:5의 인공지능 모델 정확도는 50%~80%로 낮

았다. 또한, 클래스 불균형 비율 1:10에서 교차변경을 하지 않은 경우 인공지능 모델 정확도는 약 65%로 매우 낮았으며, 교차변경 10% 이상에서 인공지능 모델 정확도는 50% 이하로 급격하게 낮아졌다(Fig. 5, Table 3 참고).

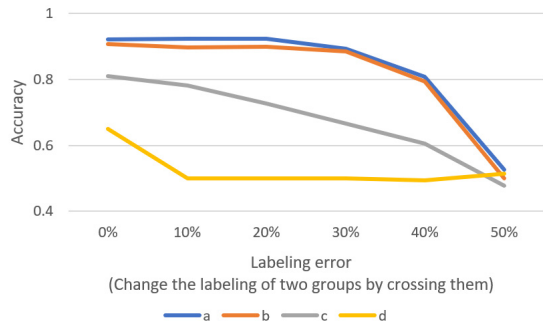


Fig. 5. Classification using random number data, ambiguous groups (class A, B) a: data ratio 1:1, b: data ratio 1:2, c: data ratio 1:5, d: data ratio 1:10

3.3.2 특성변수 분포가 상이한 두 집단 비교

특성변수가 상이함을 갖는 집단의 경우, 1:1 클래스 비율을 가지는 균형적 집단을 기준으로 비교하면, 인공지능 모델 정확도는 클래스 불균형 비율 1:10인 경우를 제외하고 비슷한 양상을 보였다. 인공지능 모델 정확도는 클래스 불균형 1:1, 1:2, 1:5 비율과 교차변경 0~40% 수준에서 약 5%p 이내의 차이로 비슷한 수준으로 나타났으며, 교차변경 50%에서는 약 10%p 이내의 차이를 보였다. 반면 클래스 불균형 비율 1:10은 교차변경 40%에서 정확도가 약 60% 이하로 급격하게 낮아졌다(Fig. 6, Table 3 참고).

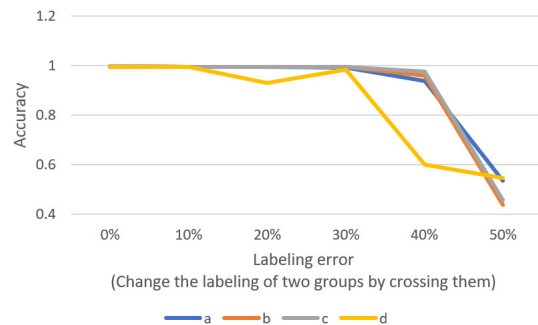


Fig. 6. Classification using random number data, distinguishable groups (class A, C) a: data ratio 1:1, b: data ratio 1:2, c: data ratio 1:5, d: data ratio 1:10

Table 3. Classification result of random number data

Flips Size	Ambiguous Data / Ambiguous Data(A, B)					
	0%	10%	20%	30%	40%	50%
1:1	0.922	0.924	0.923	0.893	0.809	0.526
1:2	0.907	0.898	0.899	0.885	0.793	0.501
1:5	0.810	0.781	0.727	0.667	0.605	0.477
1:10	0.650	0.500	0.500	0.500	0.495	0.514
Size	Distinguishable Data / Distinguishable Data(A, C)					
	0%	10%	20%	30%	40%	50%
1:1	0.997	0.996	0.994	0.991	0.939	0.535
1:2	0.997	0.996	0.994	0.995	0.960	0.438
1:5	0.998	0.995	0.994	0.994	0.975	0.457
1:10	0.995	0.996	0.930	0.984	0.599	0.546

4. 논의 및 결론

본 연구에서는 라벨링 오류가 인공지능 학습 결과에 미치는 영향을 모의실험을 통해 분석하고, 라벨링 오류와 관련된 품질지표 수준을 마련하기 위한 기초자료를 제공하기 위해 시도하였다.

데이터의 품질 중 라벨링 정확도는 정량적 목표치를 제시하도록 하고 있으나, 이에 대한 명확한 기준이 제시되고 있지 않다[15]. 본 연구의 모의실험 결과, 데이터의 특성이 상이한 데이터에서는 라벨링 오류율이 40% 이상일 때 인공지능의 정확도가 낮아지는 경향을 관찰하였다. 이는 데이터의 특성이 유사한 데이터에서 라벨링 오류율이 30% 이상일 때 정확도가 낮아지는 것과 차이를 보인다. 데이터 특성의 차이가 유사한 경우 라벨링 오류에 민감하게 반응한다고 할 수 있다. 따라서 라벨링 오류율은 데이터의 특성을 고려하여 정량적 목표치 및 기준이 마련되어야 할 것이다.

기존의 연구에서는 학습데이터 집단의 불균형 문제를 해결하기 위해 소수의 데이터를 증폭하는 오버 샘플링(Over Sampling) 기법과 다수의 데이터를 소수의 데이터 만큼 축소하여 활용하는 언더 샘플링(Under Sampling) 등 현재까지 다양한 방법이 제시되었다[16]. 그러나, 집단의 불균형에 따른 라벨링 오류율이 인공지능 결과에 영향을 미치는지 확인한 연구는 없었다. 본 연구의 모의실험 결과, 두 집단을 구성하는 클래스가 불균형하더라도 데이터 특성의 분포가 상이한 경우는 라벨링 오류율이 40% 이상일 경우에 인공지능 모델 성능 중 하나인 정확도 수준이 낮아졌다. 이는 유사한 데이터에서 클래스가 불균형한 경우 보다 낮은 라벨링 오류율에서 정확도 수준이 낮아지는 것과 차이를 보인다. 따라서 데이터의 비율이 증가하여 불균형이 커질수록 특성변수의 특성이 모호한 경

우 데이터의 라벨링 정확도를 높여야 한다. 다시 말해, 실제 데이터의 특성변수의 분포 특성 및 클래스 불균형 등을 고려하여 라벨링 정확도 수준이 결정되어야 한다.

본 연구의 중요성은 모의실험을 통해 라벨링 오류가 인공지능 학습 결과에 미치는 영향을 검토하고 데이터 특성에 따라 라벨링 오류율이 인공지능 학습 결과 및 성능에 미치는 영향이 다르다는 것을 확인하는 것에 있다. 그러나 이 연구는 몇 가지 한계점을 지닌다. 본 연구는 모의 실험 데이터를 활용하였고 라벨링 오류율을 인위적으로 조정하여 결과를 비교하였다. 이는 실제 데이터의 다양성을 반영하지 못하며, 이미지, 텍스트, 사운드, 비디오와 같은 다양한 데이터의 종류와 변수를 고려하기 어려운 한계점이 있다[7]. 또한 현재까지 개발된 모든 인공지능 모델에 일반화하기 어려우며, 인공지능 모델 성능지표를 정확도만을 제시했다는 한계를 갖는다. 따라서 실제 데이터와 다양한 인공지능 모델을 기반으로 한 추가 실험을 통해 더 정확한 결과를 도출할 필요성이 있다.

실제 데이터 특성은 사용자의 목적에 따라 학습데이터의 구성과 형태가 다양해질 수 있다. 이에 따라 향후 연구에서는 정량적으로 실제 연구된 모델과 실제 데이터를 대상으로 실험 데이터와 비교 분석하여 실제 데이터 라벨링 오류가 인공지능 학습 결과 및 성능과 관련성을 분석할 필요가 있다. 또한 라벨링 오류와 관련된 품질 지표를 검토하여 데이터 특성에 적합한 라벨링 오류 수준의 기준을 마련할 수 있을 것이다.

REFERENCES

- [1] Dong-ah Park (2017). A Study on Conversational Public Administration Service of the Chatbot Based on Artificial Intelligence. *Journal of Korea Multimedia Society*, 20(8), 1347-1356. DOI : 10.9717/kmms.2017.20.8.1347
- [2] Choi S.S. & Hong A.R (2021). Identifying Issue Changes of AI Chatbot 'Iruda' Case and Its Implications. *Electronics and Telecommunications Trends*, 36(2), 93-101. DOI : 10.22648/ETRI.2021J.360210
- [3] Yun Sangoh. (2018). Issues of Public Service Using Artificial Intelligence -Focused on ChatBot Service-. *Korean Public Management Review*, 32(2), 83-104.

- DOI : 10.24210/kapm.2018.32.2.004
- [4] Jung Won-Sup. (2020). Discrimination and Bias of Artificial Intelligence. *HUMAN BEINGS, ENVIRONMENT AND THEIR FUTURE*, (25), 55-73. DOI: 10.34162/hefins.2020..25.003
- [5] Jang Jun Hee & Seok-Joo Koh. (2020, November). A Study on the Policy Direction for Building AI Learning Data, *Proceedings of Symposium of the Korean Institute of Communications and Information Sciences*. (pp. 305-306). Online : KICS
- [6] Berrar, D. (2019). Cross-Validation. *Encyclopedia of Bioinformatics and Computational Biology*, 1, 542-545. DOI : 10.1016/B978-0-12-809633-8.20349-X
- [7] Jongwook Yoon (2019). A Pilot Study on the Standardization of Machine Learning Dataset Construction Process. *The Journal of Internet Electronic Commerce Resarch*, 19(5), 199-217. DOI : 10.37272/JIECR.2019.10.19.5.199
- [8] Kim, S. H. & Ryu, D. (2023, June). Method for improving video/image data quality for AI learning of unstructured data. *Jouranl of Information and Security*, 23(2), 55-66. DOI : 10.33778/kcsa.2023.23.2.055
- [9] Soon-Jae Kim, Woo-Hyeok Son, Ji-Hye Lee, Hoa-Hung Nguyen & Han-You Jeong. (2023). Sampling-based Analysis of Labeling Errors in AI-Hub Traffic-Light Datasets. *Journal of the Institute of Electronics and Information Engineers*, 60(3), 109-112. DOI : 10.5573/ieie.2023.60.3.109
- [10] Northcutt, C. G., Athalye, A., & Mueller, J. (2021, December). Pervasive label errors in test sets destabilize machine learning benchmarks. *Conference on Neural Information Processing Systems*. Online : NeurIPS
- [11] Oh Yoehan & Hong Sungook. (2018). Does Artificial Intelligence Algorithm Discriminate Certain Groups of Humans?. *Journal of Science & Technology Studies*, 18(3), 153-215. DOI : 10.22989/jsts.2018.18.3.004
- [12] L'Ecuyer, P. (1994). Uniform random number generation. *Annals of Operations Research*, 53, 77-120. DOI : 10.1007/BF02136827
- [13] Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of machine learning research*, 9(11).
- [14] Joong-jo Park, Tae-Woong Kim & Kyoung-min Kim. (2010). Handwritten Numeral Recognition using Composite Features and SVM classifier. *Journal of the Korea Institute of Information and Communication Engineering*, 14(12), 2761-2768. DOI : 10.6109/jkiice.2010.14.12.2761
- [15] NIA. (2023). Guidelines for Data Quality Management for Artificial Intelligence Learning v3.0. Daegu : NIA.
- [16] Kyunam Lee, Jongtae Lim, Kyoungsoo Bok, & Jaesoo Yoo (2019). Handling Method of Imbalance Data for Machine Learning : Focused on Sampling. *JOURNAL OF THE KOREA CONTENTS ASSOCIATION*, 19(11), 567-577. DOI : 10.5392/JKCA.2019.19.11.567

이 지 훈(Ji-Hoon Lee)

[정회원]

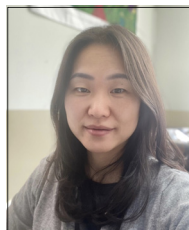


- 2021년 2월 : 건양대학교 의공학부 (의공학학사)
- 2022년 3월~현재 : 건양대학교 의과대학 정보의학교실 석사과정

- 관심분야 : 인공지능, 데이터
- E-Mail : watcherz9012@gmail.com

신 지 은(Jieun Shin)

[정회원]



- 2021년 3월~현재 : 건양대학교 의과대학 정보의학교실 교수

- 관심분야 : 빅데이터, 데이터 품질
- E-Mail : jeshin@konyang.ac.kr