Korean Chem. Eng. Res., 61(3), 388-393 (2023)
https://doi.org/10.9713/kcer.2023.61.3.388
PISSN 0304-128X, EISSN 2233-9558

신경 망의 지도 학습을 위한 로그 간격의 학습 자료 구성 방식과 손실 함수의 성능 평가

송동규‡ · 고세헌‡ · 이효민†

제주대학교 화학공학과
63243 제주시 제주대학로 102
(2023년 7월 11일 접수, 2023년 7월 19일 채택)

Performance Evaluation of Loss Functions and Composition Methods of Log-scale Train Data for Supervised Learning of Neural Network

Donggyu Song‡, Seheon Ko‡ and Hyomin Lee†

Department of Chemical Engineering, Jeju National University, 102 Jejudaehak-ro, Jeju-si, 63243, Korea
(Received 11 July 2023; Accepted 19 July 2023)

요 약

지도 학습 기반의 신경 망을 활용한 공학적 자료의 분석은 화학공학 공정 최적화, 미세 먼지 농도 추정, 열역학적 상평형 예측, 이동 현상 계의 물성 예측 등 다양한 분야에서 활용되고 있다. 신경 망의 지도 학습은 학습 자료를 요구하며, 주어진 학습 자료의 구성에 따라 학습 성능이 영향을 받는다. 빈번히 관찰되는 공학적 자료 중에는 DNA의 길이, 분석 물질의 농도 등과 같이 로그 간격으로 주어지는 자료들이 존재한다. 본 연구에서는 넓은 범위에 분포된 로그 간격의 학습 자료를 기계 학습으로 처리하는 경우, 사용 가능한 손실 함수들의 학습 성능을 정량적으로 평가하였으며, 적합한 학습 자료 구성 방식을 연구하였다. 이를 수행하고자, 100×100의 가상 이미지를 활용하여 기계 학습의 회귀 과업을 구성하였다. 4개의 손실 함수들에 대하여 (i) 오차 행렬, (ii) 최대 상대 오차, (iii) 평균 상대 오차로 정량적 평가하여, mape 혹은 msle가 본 연구에서 다룬 과업에 대해 최적의 손실 함수가 됨을 알아내었다. 또한, 학습 자료의 값이 넓은 범위에 걸쳐 분포하는 경우, 학습 자료의 구성을 로그 간격 등을 고려하여 균등 선별하는 방식이 높은 학습 성능을 보임을 밝혀내었다. 본 연구에서 다룬 회귀 과업은 DNA의 길이 예측, 생체 유래 분자 분석, 콜로이드 용액의 농도 추정 등의 공학적 과업에 적용 가능하며, 본 결과를 활용하여 기계 학습의 성능과 학습 효율의 증대를 기대할 수 있을 것이다.

Abstract – The analysis of engineering data using neural network based on supervised learning has been utilized in various engineering fields such as optimization of chemical engineering process, concentration prediction of particulate matter pollution, prediction of thermodynamic phase equilibria, and prediction of physical properties for transport phenomena system. The supervised learning requires training data, and the performance of the supervised learning is affected by the composition and the configurations of the given training data. Among the frequently observed engineering data, the data is given in log-scale such as length of DNA, concentration of analytes, etc. In this study, for widely distributed log-scaled training data of virtual 100×100 images, available loss functions were quantitatively evaluated in terms of (i) confusion matrix, (ii) maximum relative error and (iii) mean relative error. As a result, the loss functions of mean-absolute-percentage-error and mean-squared-logarithmic-error were the optimal functions for the log-scaled training data. Furthermore, we figured out that uniformly selected training data lead to the best prediction performance. The optimal loss functions and method for how to compose training data studied in this work would be applied to engineering problems such as evaluating DNA length, analyzing biomolecules, predicting concentration of colloidal suspension.

Key words: Machine learning, Supervised learning, Neural network, Loss function, Training data

1. 서 론

최근 신경망을 활용한 예측 모델링은 다양한 분야에서 중요한 도구로 사용되고 있다. 사진 분석을 통한 대기질 예측[1], 광학 분야의 적응형 복조기(adaptive demodulator) 개발[2], 열역학적 상 평형 예

†To whom correspondence should be addressed.
E-mail: fluid@jejunu.ac.kr
‡These authors contributed equally.
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 non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

측[3], 이동 현상 계의 물성 예측[4] 등이 그 예시이다. 신경 망의 학습 방법으로 지도 학습, 비지도 학습과 강화 학습이 있으며, 공학적인 자료의 회귀 및 분류 과업은 지도 학습을 통해 주로 이루어진다.

신경 망의 지도 학습에는 학습 자료가 필요하다. 특정한 경우에는, 사용하고자 하는 학습 자료가 복수 이상의 로그 간격(log scale)으로 발견되는 경우가 있다. 일례로, 만들어지는 방법과 세포의 위치에 따라 세포-유리 DNA(cell-free DNA)를 분류할 수 있고 이들의 길이는 서로 다른 로그 간격으로 발견된다[11]. 세포-유리의 발생 기작에 따라, 세포자살(apoptosis)로 인한 세포-유리 DNA는 166 bp 이하, 세포괴사(necrosis)에 의한 경우에는 10,000 bp 이하, 활성 압출(active extrusion)에 의한 경우에는 대략 1,000~3,000 bp의 길이를 갖는다. 또한, 세포-유리의 발생 위치에 따라, 세포외 소포체(extracellular vesicles)에서 발생한 경우 대략 150~6,000 bp, 미트콘드리아(mitochondria)에서 발생한 경우 대략 40~300 bp의 길이를 갖는다. 따라서, $O(10^1)$, $O(10^1)$, $O(10^2)$, $O(10^3)$, $O(10^4)$의 로그 간격을 갖는 자료라고 할 수 있다. 또한, 상기의 상세한 세포-유리 DNA 분류를 제외하고, 일반적으로 건강한 사람의 혈액 속 세포-유리 DNA는 70~200 bp에서 많이 발견되고 순환 종양 DNA(circulating tumor DNA)는 대략 200 bp~1,000 bp의 길이로 관측된다[5]. 건강한 사람의 세포-유리 DNA와 순환 종양 DNA의 혼합 시료는 $O(10^1)$, $O(10^2)$, $O(10^3)$의 로그 간격을 갖는 자료라고 할 수 있다. 다른 예로, 프로타솜, 온코솜 등의 이중 지질막 소포체(lipid bilayer vesicle)는 유래된 세포의 위치에 따라 각각 여러 범위의 크기를 갖는다[6]. 전립선 세포에서 유래한 프로타솜(prostasome)의 크기(지름)는 30~500 nm이고 암세포에서 유래한 온코솜(oncosome)의 크기는 1,000~10,000 nm이다[7,8]. 따라서, 이중 지질막 소포체의 크기는 $O(10^1)$, $O(10^2)$, $O(10^3)$, $O(10^4)$의 로그 간격을 갖는 자료라고 할 수 있다.

로그 간격으로 주어진 학습 자료의 문제점은 신경망의 목표 값(target value)의 대소 차이가 클 수 있다는 것이다. 학습 과정에서 경사 하강법(gradient descent)에 의해 조정되는 가중치(weight)는 목표 값의 크기(norm)에 비례한다[9]. 그러므로, 작은 목표 값은 신경 망의 학습에 유의미한 영향을 미치지 않을 수 있다. 다시 말해, 작은 목표 값을 가지는 학습 자료는, 큰 목표 값을 가지는 학습 자료에 비해, 학습 과정 중 중요도가 떨어지며 상당한 예측 오차를 유발할 수 있다.

본 연구에서는 로그 간격으로 주어지는 학습 자료의 기계 학습에서 발생할 수 있는 문제를 해결하고자, 회귀 과업(regression task)에서 주로 사용되는 손실 함수들(평균-제곱-오차, 평균-절대-백분율-오차, 평균-절대-오차, 평균-제곱-로그-오차)의 정량적 평가와 적합한 학습 자료 구성 방식을 연구하였다. 모델의 예측 값(predicted value)과 목표 값 간의 차이를 측정하는 지표인 손실 함수와 한정된 개수의 학습 자료 구성 방식은 모델의 학습 성능에 영향을 준다. 손실 함수는 기계 학습의 학습 알고리즘인 경사 하강법에 직접적으로 작용하며, 그 원리는 다음과 같다[9].

$$w_{ij} \leftarrow w_{ij} + \alpha \frac{\partial \mathrm{J}}{\partial w_{ij}} \tag{1}$$

식 (1)에서 손실 함수의 미분 항($\frac{\partial \mathrm{J}}{\partial \mathrm{w}_{ij}}$)은 손실 함수에 따라 달라지므로, 선택하는 손실 함수에 따라 신경망의 학습 성능이 영향 받을 수 있다. 또한, 지도 학습의 특성상, 학습 자료가 임의의 목표 값에 편향된 경우 신경 망의 예측 성능은 편향될 수 있다. 따라서, 손실 함수와 학습 자료의 구성 방식이 기계 학습에 미치는 영향을 정량적으로 분석하여, 로그 간격의 학습 자료에서 발생할 수 있는 학습 성능의 저하를 해결하고자 하였다.

본 연구에서 다룬 회귀 과업, 손실 함수 선택, 학습 자료 구성 방식들은 DNA 길이 예측, 생체 유래 분자 분석, 콜로이드 용액의 농도 추정 등의 공학적 과업에 적용 가능하며, 본 결과를 활용하여 기계 학습의 성능과 학습 효율의 증대를 기대할 수 있을 것이다.

2. 기계 학습 방법론

2-1. 가상 이미지를 이용한 회귀 과업 구성

넓은 범위의 로그 간격으로 주어지는 학습 자료의 효율적인 기계 학습 처리 방법을 연구하기 위하여, 본 연구에서는 Fig. 1의 가상 이미지를 이용한 회귀 과업을 구성하였다. 로그 간격으로 주어지는 학습 자료는 실제 공학적인 상황에서 다양하게 발견된다. 일례로, 세포-유리 DNA와 순환 종양 DNA는 DNA의 길이가 로그 간격으로 차이가 난다[5]. 건강한 사람의 혈액 속 세포-유리 DNA의 평균 길이는 70~200 bp이고 순환 종양 DNA는 200 bp~1000 bp 이상으로 $O(10^1)$, $O(10^2)$, $O(10^3)$의 로그 간격의 자료가 된다. 또 다른 예시로, 수질 분석 모델 중 전기 전도도(electrical conductivity)를 기반해 만든 모델의 전기전도도 수치가 100 μS cm^{-1} 이하부터 800 μS cm^{-1} 사이에서 발견되며, $O(10^1)$, $O(10^2)$의 로그 간격의 자료가 발견된다[10]. 실제 발견되는 공학적 자료의 경우, 측정 오차 및 수집 오차 등의 의도치 않은 변수가 학습 자료가 개입될 수 있으므로, 본 연구에서는 의도치 않은 변수가 배제된 가상 이미지로 학습 자료를 구성하였다. Fig. 1은 100×100의 이미지들이며, 무작위로 설정된 N_{pixel}의 흰 픽셀을 무작위 위치에 배치하여 가상 이미지를 만들었다. N_{pixel}이 동일하더라도, 흰 픽셀의 배치가 달라질 수 있으므로, Fig. 1의 윗 줄과 아랫 줄의 이미지들과 같이 서로 다른 이미지를 얻을 수 있었다. 이러한 가상 이지미지의 N_{pixel}의 범위를 1부터 10,000 사이의 자연수로 설정하였으며, 이 범위는 넓은 범위의 로그 간격을 형성하게 된다. 예를 들어, N_{pixel} = 5인 경우 $O(10^0)$의 로그 간격 학습 자료가 된다. 혹은, N_{pixel} = 500 이면, $O(10^2)$의 로그 간격 학습 자료이다. 따라서 가상 이미지를 통해 로그 간격으로 주어지는 학습 자료를 준비할 수 있다. 학습 자료는 심층 신경 망으로 입력되며, 심층 신경 망은 이미지를 분석하여 N_{pixel}의 예측 값을 출력하는 회귀 과업이 이루어진다.

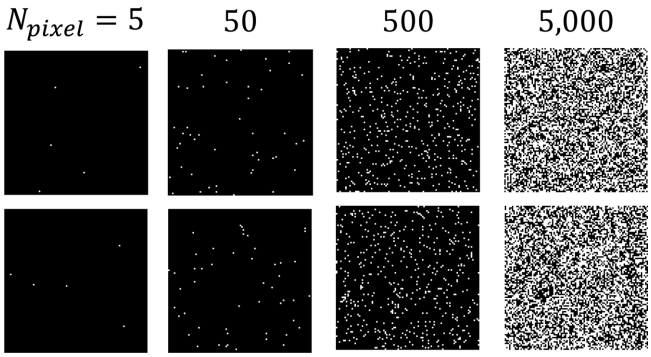

Fig. 1. Examples of virtual images (RGB color) as train and validation data for neural network. Each image has the resolution of 100×100.

2-2. 심층 신경 망

Fig. 2(a)는 가상 이미지를 이용한 회귀 과업을 진행하기 위해 만든 기계 학습 모델을 도식화한 것이다. 만들어진 가상 이미지는 입력층(input layer)을 통해 신경 망으로 입력된다. 입력 층 다음에는 합성곱 신경 망(convolution neural netwrok)을 두었다. 합성곱 신경 망은 합성곱 층(convolution layer)과 최대-풀링 층(max-pooling layer)을 교차로 두어 구성하였다. 합성곱 층은 3×3 크기의 커널(kernel)을 이용하였으며, 두 합성곱 층은 각각 32개의 필터를 설정하였다. 최대-풀링 층은 3×3의 풀링 크기로 설정하였다. 완전-연결 신경 망(fully-connected neural network)은 두 개의 은닉 층(hidden layer)을 두었으며, 각 층의 뉴런은 100개씩 설정하였다. 합성곱 층과 은닉 층의 활성화 함수(activation function)는 ReLU(Rectified Linear Unit) 함수를 사용하였다. 마지막 출력 층(ouput layer)은 1개의 뉴런을 두었으며, 활성화 함수는 선형 함수를 사용하였다. 기계 학습의 알고리즘으로 adam(adaptive moment estimation)을 이용하여 신경 망을 학습시켰다.

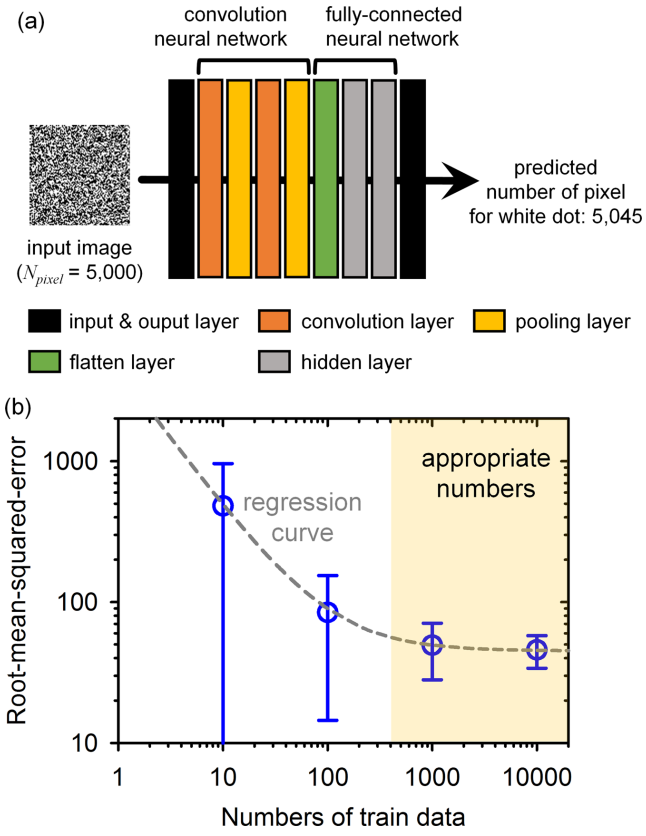

Fig. 2. (a) Configuration of neural network. (b) Root-mean-squared-error of train data as a function of the numbers of train data. Symbols and error bars are averaged errors and standard deviations, respectively. The shaded region shows appropriate numbers of train data.

기계 학습의 경우, 학습 자료가 많아질수록 학습의 성능이 향상된다. 그러나, 무한개의 학습 자료를 마련하는 것은 불가능하므로, Fig. 2(b)와 같이 적정의 학습 자료 개수를 찾고자 하였다. Fig. 2(b)는 학습 자료 개수에 따른 신경 망의 예측 제곱근-평균-제곱-오차(root-mean-squared-error)를 나타낸 것이다. 학습 자료가 적은 경우, 오차와 오차의 변화 폭이 모두 크다. 학습 자료를 늘려갈수록, 오차와 오차의 변화 폭이 줄어들며 400 개 이상의 학습 자료의 경우 오차가 크게 변하지 않은 수렴구간에 들어가게 된다. 따라서 본 연구에서는 1,000개의 학습 자료를 이용하여 로그 간격으로 주어지는 학습 자료의 기계 학습 처리 방법을 연구하였다.

신경 망의 학습에는 손실 함수(loss function)가 필요하며, 본 연구에서는 기계 학습에서 주로 사용되는 손실 함수를 이용하였다. Table 1에 이용한 손실 함수 4 종류 및 각 함수 들의 용도를 정리하였다.

3. 결과 및 고찰

3-1. 손실 함수별 오차 행렬 가시화

손실 함수별(mse, mape, mae, msle) 신경망의 예측 정확도를 나타내기 위하여, Fig. 3과 같은 오차 행렬(confusion matrix)을 가시화하는 방법을 사용하였다. 가시화된 오차 행렬은 신경망의 예측 값(predicted value)이 목표 값(target value)에 가까워질수록 그래프의 검은 실선 근처에 점이 놓이는 것으로 예측 정확도를 정량적으로 표현할 수 있다. Fig. 3(a)~3(c)까지 mse 손실 함수를 회색 원 기호로 표기하였으며, 다른 손실 함수와 비교할 수 있다. mse는 $O(10^1)$ 이하의 작은 목표 값에 대하여 오차의 정도가 크다. 그러나, mape, mae, msle는 $O(10^1)$ 이하의 작은 목표 값에 대해서도 작은 오차를 보인다. $O(10^3)$ 이상의 큰 목표 값의 경우, 손실 함수별 차이는 없으며, 모든 손실 함수가 적합한 예측 성능을 보인다.

상기의 결과를 토대로, 넓은 범위의 목표 값이 존재하는 경우의 적합한 손실 함수를 선택할 수 있을 것이다. 예를 들면, 세포-유리 기작과 세포-유리의 위치에 따라 세포-유리 DNA가 서로 다른 길이를 가지며, 이 DNA들의 길이가 넓은 범위의 목표 값에 대응될 수 있다[15]. 생성-기작에 따라, 세포자살로 발생한 경우 166 bp 이하, 세포괴사의 경우 10,000 bp 이하, 활성 압출의 경우 대략 1,000~3,000 bp의 길이를 갖는다. 또한, 세포-유리의 위치에 따라, 세포외 소포체에서 발생한 경우 대략 150~6,000 bp, 미트콘드리아에서 유리된 경우 대략 40~300 bp의 길이를 갖는다. 따라서, 목표 값이 $O(10^0)$~

Table 1. Loss functions used in this work

Name	Mathematical form	Abbreviation	Usages
Mean squared error	$\frac{1}{N}\sum_i^N (y_i-\hat{y}_i)^2$	mse	Prediction of pollutant concentrations [11]
Mean absolute percentage error	$\frac{100\%}{N}\sum_i^N \left\|\frac{(y_i-\hat{y}_i)}{y_i}\right\|$	mape	Prediction of surface chloride concentration [12]
Mean absolute error	$\frac{1}{N}\sum_i^N \|y_i-\hat{y}_i\|$	mae	Prediction of PM_{10} concentration [13]
Mean squared logarithmic error	$\frac{1}{N}\sum_i^N [\log(y_i+1)-\log(\hat{y}_i+1)]^2$	msle	Prediction of impending earthquake [14]

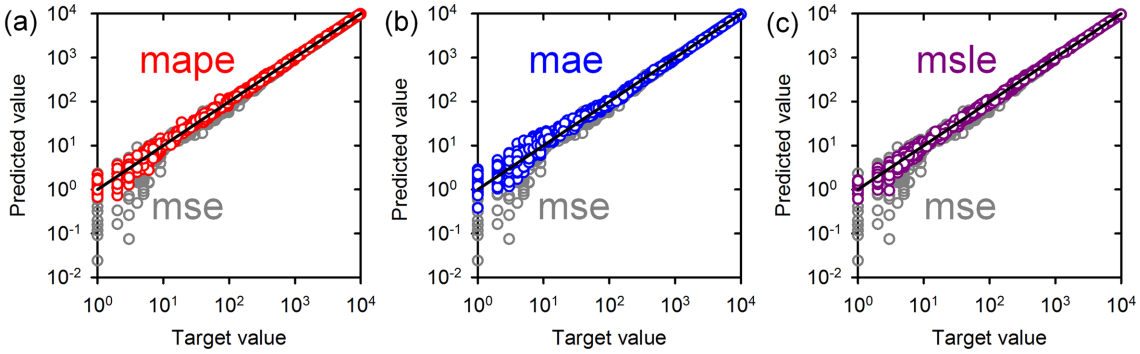

Fig. 3. Comparison of confusion matrix for (a) mse and mape, (b) mse and mae, and (c) mse and msle.

$O(10^4)$까지 형성된다. 다른 예로, 이중 지질막 소포체[16]와 순환종양세포[17]의 미세유체역학적 분류 과업은 각 생체 유래 분자의 크기 정보에 의존한다. 순환종양세포의 길이 범위는 200~1000 bp, 즉 68~340 nm이다[5]. 반면, 이중 지질막 소포체는 소포체를 생성한 세포의 위치에 따라 각각 여러 범위의 길이를 갖는다[6]. 이중 지질막 소포체 중, 전립선 세포에서 유래한 프로타솜의 크기(지름)는 30~500 nm이고 암세포에서 유래한 온코솜의 크기는 1,000~10,000 nm이다. 따라서, 넓은 목표 값이 존재하는 공학적 과업은 다수 발견될 수 있다. 상기의 예시와 같은 과업들에 대한 손실 함수는 mse가 아닌 mape 등의 특정 손실 함수가 사용된다면, 기계 학습의 성능을 향상시킬 수 있다. 각 손실 함수별 목표 값 구간에 대한 자세한 오차 비교는 다음 절에 하고자 한다.

3-2. 신경망 성능에 대한 학습 자료 구성 방식의 영향

Fig. 4는 학습 자료의 구성 방식에 따른 각 목표 값들의 예측 상대 오차를 차용한 손실 함수별(mse, mape, mae, msle)로 보여준다. 목표 값들의 분포가 로그 간격이므로, 학습 자료의 구성 방식을 (i) 무작위 선별(randomly selected)와 (ii) 각 로그 간격별 균등 선별(uniformly selected)로 설정하였다. 예를 들면, 무작위 선별의 경우, 10^0~10^4 까지의 목표 값 1,000개를 무작위 선별한다. 반면, 균등 선별의 경우, 10^0~10^1까지의 목표 값 250개, 10^1~10^2까지의 목표 값 250개의 방식으로 각 로그 간격당 학습 자료의 개수를 균등하게 선별하였다.

Fig. 4(a)는 무작위로 선별된 학습 자료들의 예측 상대 오차이다. msle를 제외한 손실 함수들(mse, mape, mae)은 $O(10^2)$ 이하의 목표 값에 대해 200% 이상의 최대 오차를 보인다(mse: 947%, mae: 457%, mape: 245%, msle: 131%). msle 손실 함수의 경우 상대 오차가 비교적 작지만, $O(10^0)$의 목표 값에는 여전히 상대 오차가 증가하는 경향을 보인다. $O(10^2)$의 중간 값의 상대 오차 또한 무시할 수 없는 크기의 결과를 얻었다.

Fig. 4(b)는 균등 선별된 학습 자료들의 예측 상대 오차이다. Fig. 4(a)와 비교하여, $O(10^2)$ 이하의 목표 값들의 상대 오차가 상당히 낮아진 것을 알 수 있다. $O(10^2)$의 중간 목표 값에 대해서도, 상대 오

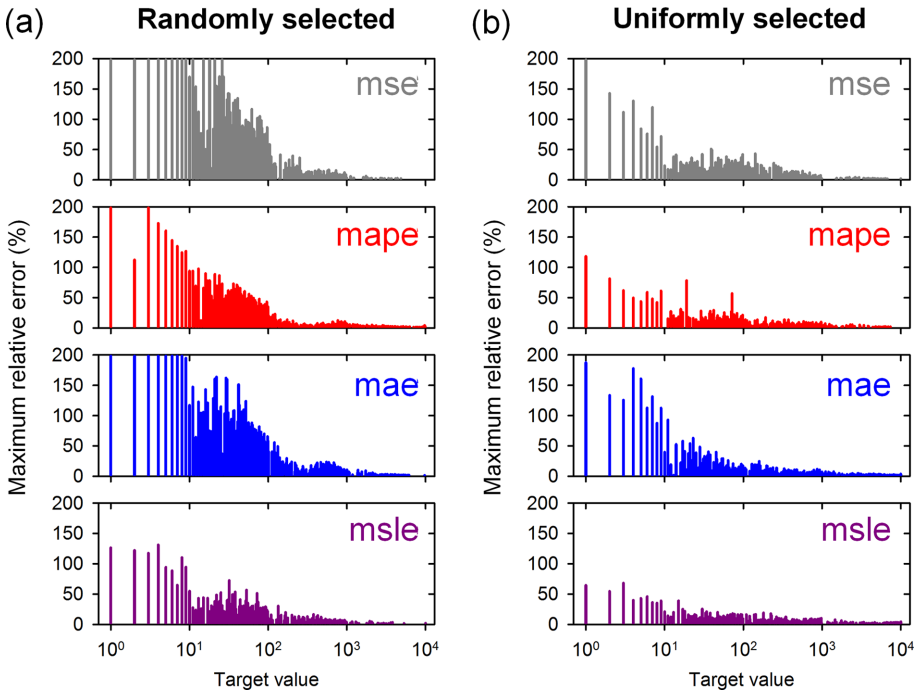

Fig. 4. Maximum relative error as a function of target value in the case of (a) randomly selected train data and (b) uniformly selected train data.

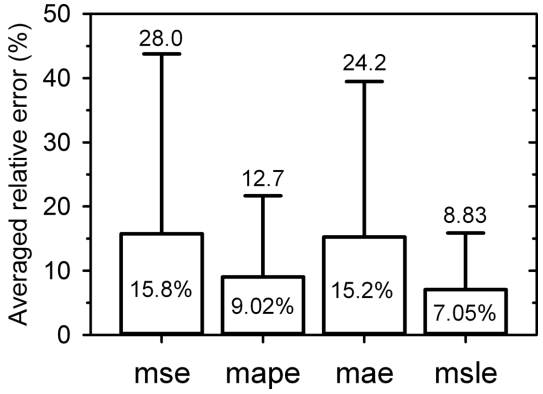

Fig. 5. Averaged relative error for the specific loss functions such as mse, mape, mae and msle.

차가 현저히 낮아졌기 때문에 학습 자료의 구성 방식이 신경망의 학습 결과에 영향을 미침을 알 수 있다. 더불어, 네 종류의 손실 함수 중 msle의 상대 오차가 가장 작게 분포한 것을 통해 msle의 학습 성능이 우수한 것을 알 수 있다. msle의 특성 상 로그 값을 정의할 수 없는 경우(e.g. 음수의 목표 값 등)에는 mape가 차선책이 될 수 있다.

3-3. 손실 함수의 적합성

본 연구에서 사용한 4개의 손실 함수(mse, mape, mae, msle)의 예측 성능을 정량적으로 평가하고자, 평균 상대 오차와 표준 편차를 Fig. 5에 나타내었다. 최대의 평균 상대 오차를 보이는 손실 함수는 mse였고(15.8%), 최저의 평균 상대 오차는 msle에서 관찰되었다(7.05%). mape 손실 함수는 msle와 상대 오차 분포가 유사하며, mae 손실 함수는 mse와 상대 오차 분포가 유사했다. 그러므로, 로그 간격으로 주어지는 학습 자료에 대해, 최적의 손실 함수로써 mape 혹은 msle를 선택하는 것이 합리적이다.

Mape와 msle 중 하나의 손실 함수를 선택해야할 때, 고려해야할 점이 있다. 신경망의 출력이 음수가 나오는 경우에는 msle의 로그항($\log(\hat{y}_i+1)$, 여기에서 $\hat{y}_i$는 i번째 자료에 대한 신경망의 출력 값)이 정의되지 않을 수 있기 때문에, 이를 방지할 추가적인 제약이 필요하다. 음수가 배제된 목표 값들로만 이루어진 과업은 출력 층의 활성화 함수를 ReLU로 설정하면, msle를 문제 없이 사용할 수 있다. 본 연구에서 다룬 이미지 분석 혹은 농도 추정, DNA 길이 예측 등의 과업들이 음수가 배제된 목표 값들로만 이루어진 과업의 한 예이다. 그러나, 음수가 포함된 목표 값들로 이루어진 과업은 msle보다 mape를 선택하는 것이 안정적인 학습이 될 것이다.

4. 결 론

본 연구에서는 넓은 범위에 분포된 학습 자료를 이용한 기계 학습에 대해, 손실 함수들(mse, mape, mae, msle)의 정량적 평가와 적합한 학습 자료 구성 방법을 수행하였다. 이를 수행하고자, 100×100의 가상 이미지를 활용하여 기계 학습의 회귀 과업을 구성하였다. 4개의 손실 함수들에 대하여 (i) 오차 행렬, (ii) 최대 상대 오차, (iii) 평균 상대 오차로 정량적 평가하여, mape 혹은 msle가 본 연구에서 다룬 과업에 대해 최적의 손실 함수가 됨을 알아내었다. 또한, 학습 자료의 값이 넓은 범위에 걸쳐 분포하는 경우, 학습 자료의 구성을 로그 간격 등을 고려하여 균등 선별하는 방식이 높은 학습 성능을 보임을 밝혀내었다. 본 연구에서 다룬 회귀 과업은 DNA의 길이 예측, 생체 유래 분자 분석, 콜로이드 용액의 농도 추정 등의 공학적 과업에 적용 가능하며, 본 결과를 활용하여 기계 학습의 성능과 학습 효율의 증대를 기대할 수 있을 것이다.

감 사

이 논문은 2023학년도 제주대학교 교원성과지원사업에 의하여 연구되었습니다.

사용기호

α	: Learning rate
J	: Loss function
N	: Total number of data samples
N_{pixel}	: Number of pixels
w_{ij}	: Weight of i-th layer and j-th node for neural network
y_i	: Target value of i-th data sample
$\hat{y}_i$	: Predicted value for i-th data sample

References

1. Wang, X. *et al.* in *2020 IEEE International Conference on Image Processing (ICIP)*. 768-772.
2. Li, J., Zhang, M. and Wang, D., "Adaptive Demodulator Using Machine Learning for Orbital Angular Momentum Shift Keying," *IEEE Photonics Technology Letters*, **29**, 1455-1458(2017).
3. Poort, J. P., Ramdin, M., van Kranendonk, J. and Vlugt, T. J. H., "Solving Vapor-liquid Flash Problems Using Artificial Neural Networks," *Fluid Phase Equilibria*, **490**, 39-47(2019).
4. Valderrama, J. O., Reátegui, A. and Rojas, R. E., "Density of Ionic Liquids Using Group Contribution and Artificial Neural Networks," *Industrial & Engineering Chemistry Research*, **48**, 3254-3259(2009).
5. Han, X., Wang, J. and Sun, Y., "Circulating Tumor DNA as Biomarkers for Cancer Detection," *Genomics, Proteomics & Bioinformatics*, **15**, 59-72(2017).
6. Lee, Y., El Andaloussi, S. and Wood, M. J., "Exosomes and Microvesicles: Extracellular Vesicles for Genetic Information Transfer and Gene Therapy," *Human molecular genetics*, **21**, R125-134(2012).
7. Baskaran, S., Panner Selvam, M. K. and Agarwal, A. in *Advances in Clinical Chemistry* Vol. 95 (ed Gregory S. Makowski) 149-163 (Elsevier, 2020).
8. Spada, S. and Galluzzi, L. *Extracellular Vesicles*. (Elsevier Science, 2020).
9. Goodfellow, I., Bengio, Y. and Courville, A. *Deep Learning*. (MIT Press, 2016).
10. Aldrees, A. *et al.*, "Multi-Expression Programming (MEP): Water Quality Assessment Using Water Quality Indices," *Water*, **14**, 947 (2022).
11. Alpan, K. and Sekeroglu, B., "Prediction of Pollutant Concentrations by Meteorological Data Using Machine Learning Algorithms," *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*,

XLIV-4/W3-2020, 21-27(2020).
12. Cai, R. *et al.*, "Prediction of Surface Chloride Concentration of Marine Concrete Using Ensemble Machine Learning," *Cement and Concrete Research*, **136**, 106164(2020).
13. Bozdağ, A., Dokuz, Y. and Gökçek, Ö. B., "Spatial Prediction of PM10 Concentration Using Machine Learning Algorithms in Ankara, Turkey," *Environmental Pollution*, **263**, 114635(2020).
14. Sadhukhan, B., Chakraborty, S. and Mukherjee, S., "Predicting the Magnitude of An Impending Earthquake Using Deep Learning Techniques," *Earth Science Informatics*, **16**, 803-823(2023).
15. Bronkhorst, A. J., Ungerer, V. and Holdenrieder, S., "Comparison of Methods for the Isolation of Cell-free DNA from Cell Culture Supernatant," *Tumour Biology*, **42**(4), 1010428320916314(2020).
16. Theel, E. K. and Schwaminger, S. P., "Microfluidic Approaches for Affinity-Based Exosome Separation," *International Journal of Molecular Sciences*, **23**, 9004(2022).
17. Descamps, L., Le Roy, D. and Deman, A. L., "Microfluidic-Based Technologies for CTC Isolation: A Review of 10 Years of Intense Efforts towards Liquid Biopsy," *Int. J. Mol. Sci.*, **23**(4), 1981(2022).

Authors
Doggyu Song: Undergraudate Student, Department of Chemical Engineering, Jeju National University, Jeju 63243, Korea; 98donggyu@naver.com
Seheon Ko: Undergraudate Student, Department of Chemical Engineering, Jeju National University, Jeju 63243, Korea; koseheon0541@naver.com
Hyomin Lee: Associate Professor, Department of Chemical Engineering, Jeju National University, Jeju 63243, Korea; fluid@jejunu.ac.kr