JDCS
디지털콘텐츠학회논문지
Journal of Digital Contents Society
Vol. 20, No. 1, pp. 127-133, Jan. 2019

유전적 알고리즘이 데이터셋 생성에 미치는 영향에 대한 연구

박 상 현 · 조 영 복*
대전대학교 정보보안학과

An Effect of Genetic Algorithm for Creating a Dataset

Sang-hyun Park · Young-bok Cho*
Department of Information Security, Daejeon University, Daejeon, 34520 Korea

[요 약]

현재 다양한 분야에서 딥러닝 기술이 활용되고 있다. 딥러닝을 적용시키기 위해서는 모델의 구성도 중요하지만 학습, 검증을 위한 데이터셋도 중요하다. 딥러닝모델의 정확도를 위해서는 학습 데이터셋의 크기가 매우 중요한 데 이는 데이터셋이 정확도에 많은 영향을 주고 있기 때문이다. 또한 다양한 환경에서 수집된 데이터가 많이 존재하는 상황일수록 정확도가 높아진다. 이로 인해 다양하고 많은 데이터를 수집하기 위해 자본을 소비한다. 그러나 한정된 환경으로 인해 더 이상 데이터셋을 수집할 수 없는 상황에서는 기존 데이터를 회전을 하거나 확대 등 변형 작업을 통해 새로운 데이터셋을 생성한다. 본 연구에서는 데이터셋을 생성하고 학습시키는 과정에서 쓰이는 데이터 부풀리기에서 유전적 알고리즘의 세대, 교배, 돌연변이가 정확도에 미치는 영향에 대한 연구 방법을 제시하고자 한다.

[Abstract]

Currently, deep-learning technology is used in various fields. In order to apply deep learning, model configuration is important, but data sets for learning and testing are also important. For the accuracy of the deep learning model, the size of the training dataset is very important because the dataset has a significant impact on accuracy. Also, the more data collected in various environments, the higher the accuracy. This consumes capital to collect a large amount of data. However, when a data set can not be collected due to a limited environment, a new data set is created through a transformation operation such as rotating or enlarging existing data. In this study, we propose a study on the effect of generation, mating, and mutation of genetic algorithms on accuracy of data inflation in the process of generating and learning data sets.

색인어 : 유전 알고리즘, 데이터셋, 딥러닝, 하이퍼파라미터, 적합도

Key word : Genetic Algorithm, Dataset, Deep-Learning, Hyperparameter, Fitness

http://dx.doi.org/10.9728/dcs.2019.20.1.127

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-CommercialLicense(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.

Received 01 November 2018; **Revised** 02 December 2018
Accepted 20 January 2019

***Corresponding Author; Young-Bok Cho**

Tel: +82-42-280-2406
E-mail: pshboom@edu.dju.ac.kr

Ⅰ. 서 론

알파고로 인해 화제가 된 딥러닝은 세상을 바꿀 핵심기술이라고 평가받고 있다[1]. 현재 딥러닝 관련 논문과 서적이 쏟아져 나오고 있으며 딥러닝을 실제로 적용한 사례도 많이 등장하고 있다. 우선 딥러닝을 시작하기 위해서는 방대한 양의 데이터셋이 필요하다. 데이터셋이 존재해야만 모델을 학습시킬 수 있고 모델에 따라 필요한 데이터셋이 다르므로 데이터셋 수집에 많은 시간과 자본이 소요되기 마련이다. 따라서 딥러닝에서 데이터셋의 수집은 매우 중요한 것이다. 그러나 딥러닝을 위해 단순히 데이터셋의 양이 중요한 것이 아니라 노이즈가 없이 정제된 상태의 데이터셋이 많을수록 정확도가 높아지게 된다. 따라서 데이터셋은 딥러닝의 시작이자 전부를 담당하는 부분이라 해도 과언이 아닐 것이다. 그러나 현재 딥러닝을 위한 데이터셋 수집 과정에서 많은 한계에 부딪히는 상황에 도달하게 되는 경우가 있다. 이러한 상황에 놓이게 되면 기존의 수집된 데이터를 부풀리는 방법을 모색하게 된다. 이미지의 경우에는 기존의 데이터를 이용하여 확대, 축소, 이동, 회전 등을 이용하여 새로운 이미지를 생성하게 된다.

본 연구에서는 유전적 알고리즘이 이러한 데이터셋 부풀리기에 미치는 영향을 실험을 통해 분석한다. 실험을 위해 이미지 분류에 널리 사용되는 컨볼루션 신경망 모델을 사용하였다. 여러 신경망의 근간을 이루는 다층 신경망은 여러 층으로 구성되어져 복잡한 자료를 모델링하기에 적합하다[2]. 또한 컨볼루션 신경망은 고전적 인공신경망보다 깊은 계층의 신경망 학습이 잘 되는 장점을 가진다[3]. 이 신경망을 기반으로 데이터셋을 학습시키는 과정에서 유전적 알고리즘이 얼마나 많은 영향을 미치고 기존의 방법과 비교하여 정확도에 미치는 영향이 있는지를 실험을 통해 확인 하였다.

본 논문의 구성은 2장에서 유전 알고리즘과 데이터셋에 대해 기술하고 3장에서는 실험에 사용한 유전 알고리즘에 대해 기술한다. 4장에서는 논문의 결과와 5장에서는 향후 연구에 대해 기술한다.

Ⅱ. 연구의 내용

2-1 유전알고리즘의 특징

최근 공학의 여러 가지 응용 분야 중 복잡한 제약성을 가진 대규모의 최적화 문제들은 일반적인 수학적 프로그램이나 최적화 알고리즘의 조합을 이용하여 짧은 연산 시간 안에 최적의 해를 구하기가 매우 어렵다. 이러한 문제들을 적당한 연산 시간 안에 풀 수 있는 주요한 메커니즘으로써 진화 알고리즘(Evolutionary Algorithm)이 사용된다. 진화 알고리즘은 풀고자 하는 문제에 대한 가능한 솔루션들을 정해진 형태의 자료 구조로 표현한 다음 이들을 점차적으로 변형함으로써 점점 더 좋은 해들을 생성한다. 각각의 가능한 해를 하나의 유기체(organism) 또는 개체(individual)로 보며 이들의 집합을 개체군(population)이라 한다. 하나의 개체는 보통 한 개 또는 여러 개의 염색체(chromosomes)로 구성되며 염색체를 변형하는 연산자들을 유전 연산자(genetic operation)라 한다. 이러한 진화 알고리즘은 구조가 단순하고 방법이 일반적이어서 응용 범위가 상당히 넓은 것이 특징이다. 여러 진화 알고리즘에 의한 탐색은 탐색공간에 대한 제약, 예를 들어 공간의 연속성과 같은 제약 조건을 갖지 않기 때문에 여러 가지 문제에 적용이 가능하고 또 주어진 상황에 적응적으로 대처해서 탐색한다는 특징을 가지고 있다. 유전자 알고리즘은 복잡한 최적화 작업이나 학습을 이용한 구조에 자주 사용되는 자연도태의 유전적인 메커니즘에 기초한 탐색 알고리즘으로써 선택(selection)과 교차(crossover), 돌연변이(mutation), 반전(inversion) 연산자를 이용하여 문제에 대한 후보해 (candidate solution)또는 유기체의 집합인 개체군을 새로운 개체군으로 반복적으로 변형한다. 따라서 유전적 알고리즘이란 생명체가 번식을 통해 진화하는 것처럼 컴퓨터가 해당 상황에서 가장 좋은 방법만을 선택하여 컴퓨터 스스로 해결책을 모색하는 알고리즘이다[4]. 유전 알고리즘은 설계자가 원하는 방향에서 그 값이 증가하도록 적합도 함수를 결정한 후 이 함수 값만을 계산하여 그 결과로부터 최적화를 행하므로 기타 정보가 필요치 않아 실제 시스템에 비교적 쉽게 적용할 수 있다는 장점이 있다[5]. 따라서 어느 부분에 대해 유전자를 생성하고 이를 이용해 나아갈 것 인지에 따라 결과는 천차만별이 된다.

2-2 딥러닝에서의 훈련 데이터셋의 중요성

인간의 뇌는 수많은 뉴런으로 서로 연결되어 있다. 인간의 뇌와 유사한 인공적인 신경망은 다층구조라 할 수 있고, 이러한 다층구조를 딥러닝이라 부른다[6]. 딥러닝은 무(無)에서 유(有)를 창조하는 기술이 아닌 학습을 통해 업무의 자동화를 이끌어 나가는 기술이다. 데이터셋이 존재하지 않는다면 학습조차 이루어지지 않으며 진행조차 할 수 없다. 딥러닝에서 데이터셋의 역할은 훈련과 시험, 검증에 있다. 즉, 데이터셋은 딥러닝에서 정확도를 담당하는 역할인 것이다[7]. 딥러닝에서는 구현하고자 하는 모델에 따라서 데이터셋이 정해질 수 있으며 해결하고자 하는 문제에 따라 데이터셋이 정해질 수도 있다. 즉 사전에 데이터셋을 미리 준비한다고 하더라도 그 상황에 맞추어 수집해야하기 때문에 몇몇 데이터셋이 무용지물이 되어버리는 경우가 있다. 따라서 모델과 문제, 기타 요소들이 전부 정해진 순간부터 데이터셋 수집에 들어가기 때문에 충분한 데이터셋이 모이지 않은 경우가 잦다. 또한 충분한 데이터셋이 모여 있음에도 일부러 노이즈를 추가한 데이터셋을 집어넣어 학습률을 높이는 경우가 있다. 이러한 복합적인 이유로 데이터셋을 부풀리는 선택을 하게 된다. 데이터셋 부풀리기는 원본 이미지를 선택한 뒤 이 이미지에 대해 미리 지정한 속성 값 내에서 다른 이미지를 만들어내는 기술이며 현실적인 한계로 인해 더 이상 데이

터셋 수집이 불가능한 상황일 때 유용하게 쓰이는 기술이다.

```
keras.preprocessing.image.ImageDataGenerator(featurewise_center=False,
samplewise_center=False,
featurewise_std_normalization=False,
samplewise_std_normalization=False,
zca_whitening=False,
rotation_range=0.,
width_shift_range=0.,
height_shift_range=0.,
shear_range=0.,
zoom_range=0.,
channel_shift_range=0.,
fill_mode='nearest',
cval=0.,
horizontal_flip=False,
vertical_flip=False,
rescale=None,
preprocessing_function=None,
data_format=K.image_data_format())
```

그림 1. ImageGenerator 클래스를 설명하는 Keras.io 페이지
Fig. 1. Keras.io page to explain ImageGenerator class

그림 1은 케라스에서 ImageDataGenerator 함수를 이용해 제공되는 데이터 부풀리기 기능으로 keras.io 페이지를 나타낸 것이다. 데이터셋을 부풀리기 위해서 케라스의 ImageDataGenerator를 사용할 것이다. ImageDataGenerator는 회전, 이동, 확대, 축소, 밝기 등을 무작위로 정하여 이미지 데이터의 배치를 생성한다. 또한 원하는 만큼 이미지를 생성할 수 있어 손쉽게 데이터셋을 부풀릴 수 있다[8].

표 1. ImageDataGenerator 속성
Table 1. ImageDataGenerator property

Name	Contents
rotation_range	Rotation within the angular range
width_shift_range	Move in horizontal range
height_shift_range	Move in vertical range
shear_range	Rotation in Shear Strength
zoom_range	Zoom in / out of image in range
horizontal_flip	Flip image horizontally
vertical_flip	Flip image vertically

표 1은 ImageDataGenerator의 속성을 설명한 것이다. 이러한 속성값을 적절히 바꾸면서 훈련 데이터셋을 생성하게 된다. 또한 속성들을 혼합하여 데이터 부풀리기를 수행할 수도 있다. 이렇게 생성된 데이터셋을 훈련시켜 학습을 하게 된다면 기존의 학습에 비해 성능향상이 이루어진다. 또한 이렇게 생성한 데이터셋을 훈련셋에 추가하는 것이 아닌 시험셋에 추가하여 자신의 신경망이 제대로 학습되었는가도 확인할 수 있다. 단, 훈련셋과 시험셋 모두 ImageDataGenerator로 생성한 데이터셋을 사용하게 된다면 아무리 정확도가 높더라도 실전에서는 유용한 결과를 얻어내기 힘들다. 따라서 어느 부분에 대해 데이터셋 부풀리기를 적용할지를 결정하는 것이 중요하다.

Ⅲ. 유전적 알고리즘이 데이터셋 생성에 미치는 영향 분석

딥러닝을 위한 데이터셋이 중요한데 이런 데이터셋을 구하는 방법은 어려운 실정이다. 따라서 본 논문에서는 유전적 알고리즘이 데이터셋 부풀리기에 미치는 영향을 분석하기 위한 실험환경을 설정한다. 제안된 케라스 기반의 ImageDataGenerator의 속성을 사람이 직접 입력할 수도 있다. 그러나 딥러닝은 안정된 데이터만 사용한다면 실전에서 벌어지는 노이즈 및 예상치 못한 변수로 인해 원하는 데이터를 도출하기 힘들다. 유전적 알고리즘을 데이터셋 부풀리기에 적용해 실전에서 등장하는 노이즈를 재현하기 위해서는 속성들을 어떻게 연산할지가 중요하다. 또한 유전 알고리즘을 적용하여 실제 최적의 해를 구하지는 못하더라도 최적해에 가까운 해를 구하기 위한 방법으로써 접근할 수 있다[9]. 따라서 우선 ImageDataGenerator에서 언급한 표 1에서 정의된 속성들의 값을 무작위로 선정하고 이러한 속성 값들을 실수(real number) 코딩 유전 알고리즘을 이용하여 개체를 표현한다. 실수 코딩이란 유전자를 이진화(binarization)하여 유전 알고리즘을 적용하는 이진 코딩과는 달리 유전자의 값 자체를 실수로 지정하여 그대로 적용하는 기술을 말한다. 속성 값이 실수일 경우 속성 값과 유전자가 일대일 대응되므로 간단명료하게 표현할 수 있다. 그림2는 ImageDataGenerator를 실제로 적용시킨 예제이다.

```
data_aug_gen = ImageDataGenerator(rescale=1./255,
                                  rotation_range = 353,
                                  width_shift_range = 0.79,
                                  height_shift_range = 0.88,
                                  shear_range = 0.19,
                                  zoom_range = [1.62, 1.87],
                                  horizontal_flip = False,
                                  vertical_flip = False,
                                  fill_mode='nearest')
```

그림 2. 실제로 적용시킨 오브젝트의 예제
Fig. 2. Example of applying object

그림2의 코드 중 rotation range 단위는 각도(°)이며 0부터 지정 범위 내에서 이미지가 임의로 회전한다. width shift range의 단위는 비율이며 0부터 지정 범위 내에서 임의로 이미지의 폭을 이동시킨다. height shift range의 단위는 비율이며 0부터 지정 범위 내에서 임의로 이미지의 높이를 이동시킨다. shear range의 단위는 라디안(π)이며 0부터 지정 범위 내에서 임의로 이미지를 변형한다. zoom range의 단위는 배율이며 지정 범위 내에서 확대하거나 축소하여 크기를 변화시킨다. horizontal flip은 참, 거짓으로 나뉘며 해당 속성이 참이라면 수평 방향으로 이미지를 뒤집는다. vertical flip은 참, 거짓으로 나뉘며 해당 속성이 참이라면 수직 방향으로 이미지를 뒤집는다. 이렇게 생성한 개체들을 ImageDataGenrator에 적용해 정확도를 비교한 뒤 개체별 정확도에 따라 각 개체에게 적합도를 부여하게 되는데 이것은 부모 유전자 선택 방법 중 그림 3의 알고리즘인 토너먼트 선택(Tournament Selection)을 적용하기 위함이다. 토너먼트 선택이란 가장 간단한 선택 연산 메서드로 염색체를 무작위로 선정 한 뒤 그 중 가장 높은 적합도를 가진 염색체를 부모 염색체로 선정하는 방법이다.

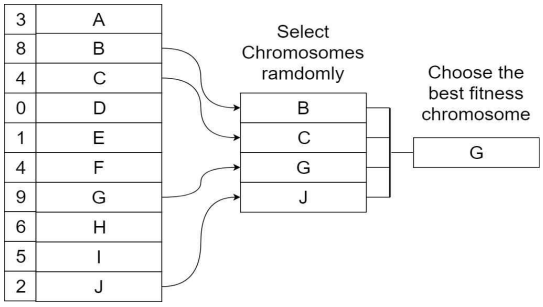

그림 3. 토너먼트 선택 알고리즘
Fig. 3. Tournament selection algorithm

즉 세대에서 두 개의 염색체를 임의로 선택하고, 난수 p를 생성한다. 이 난수 p와 미리 정의한 선택압 k의 대소를 따지고, p>k이면 적합도 함수에 의해 평가된 적합도가 다른 하나보다 낮은 염색체를 선택하고 반대의 경우 높은 염색체를 선택하게 된다. 이때 선택압 k는 우수한 염색체를 남기는 비율로 p와 동일한 정의역을 갖게 된다. 현재 세대 적용이 끝났다면 기존 세대의 유전자는 삭제하고 선정된 부모 유전자를 통해 단순 교배를 실행한다. 그림 4는 유전적 알고리즘에서 나타내는 단순 교배의 예시를 도식화한 것이다.

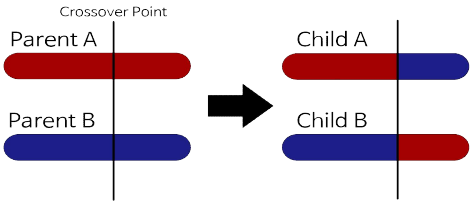

그림 4. 단순교배 알고리즘
Fig. 4. Simple crossover algorithm

그림 4의 단순 교배란 가장 간단한 교배법이며 교배하는 위치, 교차점을 임의로 결정하여 교차점을 경계로 부모의 염색체 사이에서 유전자를 교환하는 방법이다[10]. 이러한 방법으로 세대를 생성한 뒤 마지막 유전자에는 초깃값 생성처럼 모든 값이 무작위인 돌연변이를 추가해준다. 돌연변이를 추가하지 않는다면 유전자가 한 방향으로 수렴하는 경우가 생긴다. 이러한 세대교체 작업을 진행한 뒤 해당 세대를 다시 적용하는 것을 반복하여 세대가 진행될수록 정확도를 높여나가는 것이 데이터셋 생성에서의 유전 알고리즘 적용이다.

Ⅳ. 실험 및 결과 분석

4-1 실험 환경

본 논문에서 표2에서 정의된 환경에서 실험할 대상은 도형 판별을 위해 기본도형인 triangle, square, circle을 분류하는 학습모델을 이용해 실험하였다. 실험환경은 Intel(R) Core(TM) i5 7200U 2.7GHz CPU와 8GB RAM, NVIDIA GeForce 940MX GPU가 장착된 PC상에서 실험하였다. 또한 python으로 염색체 초기화, 적합도 판별, 토너먼트 선택, 단순 교배, 돌연변이등 유전알고리즘 전체를 구현한 코드를 이용하여 실험하였다. 또한 컨볼루션 신경망의 구현은 케라스를 이용하였다. 표 2에서는 실험환경을 정리하였다.

표 2. 실험 환경
Table 2. Experimental Environment

Type	Description
CPU	Intel(R) Core(TM) i5 7200U 2.7GHz
RAM	8GB DDR4 Memory
GPU	NVIDIA GeForce 940MX Graphics with 2GB Graphic Memory
Development Langage	Python3

이 모델을 통해 손으로 그린 각 도형을 정확히 판별한 학습의 정확도를 비교하였다. 그림 5는 컨볼루션 신경망 모델 학습에 사용된 데이터셋의 일부이다. 각 열의 마지막 도형은 데이터셋 부풀리기에 사용된 원본이미지이다.

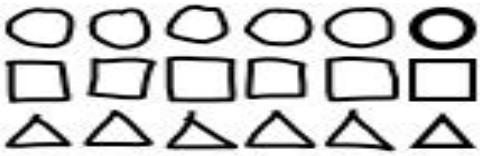

그림 5. 실험에 사용된 데이터셋 중 일부분
Fig. 5. Part of the datasets used in this experiment

컨볼루션 신경망 모델에 적용시킨 훈련셋은 각 도형마다 손수 만든 이미지 15개와 포토샵을 통해 생성한 원본 이미지로 데이터 부풀리기를 진행한 이미지 1,485개를 합하여 1,500개씩 총 4,500개의 데이터셋을 준비하였고 시험셋은 각 도형마다 14개의 이미지, 총 42개의 데이터셋을 준비하였다. 에포크(epoch)는 50회로 지정하였다. 그림6은 이 실험에서 사용된 유전 알고리즘의 흐름을 도식화한 것이다.

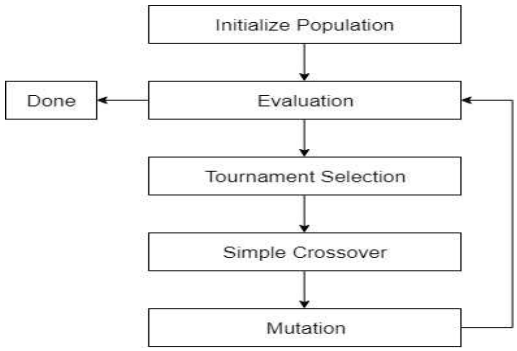

그림 6. 실험에 사용된 알고리즘의 흐름도
Fig. 6 A plot of the algorithm used in this experiment

초기 염색체 생성은 무작위로 선정하였고 이 염색체 값을 통해 데이터셋 부풀리기를 진행하였다. 이렇게 생성된 데이터셋

을 이용하여 정확도를 계산한 뒤, 정확도를 유전적 알고리즘의 적합도로 선택한다. 그 다음 선택연산을 적용하고 교배 연산을 적용하여 새로운 세대를 생성하였다. 선택연산은 토너먼트 선택을 적용하였고 교배연산은 단순 교배를 적용하였다.

4-2 실험 결과

실험 환경을 토대로 진행한 실험의 결과는 다음과 같다. 표 3은 세대 당 평균 정확도, 상위 5개의 염색체의 정확도, 최고 정확도를 적은 표이다.

표 3. 세대별 평균 정확도 결과
Table 3 Result of average accuracy per generation

	Average Accuracy	Top 5 Accuracy	Best Accuracy
Generation 1	49.33%	58.67%	66.67%
Generation 2	48.00%	56.00%	66.67%
Generation 3	56.67%	64.00%	66.67%
Generation 4	54.67%	61.33%	66.67%
Generation 5	58.00%	62.67%	73.33%
Generation 6	53.33%	65.33%	73.33%
Generation 7	52.67%	65.34%	66.67%
Generation 8	70.67%	74.67%	86.67%
Generation 9	70.67%	80.00%	86.67%
Generation 10	71.33%	74.66%	80.00%

무작위 값으로 시작한 첫 세대 이후로 세대가 진화하면서 유전알고리즘이 적용됨과 동시에 꾸준히 성장하는 모습을 보였다. 그러나 4세대부터 7세대까지 두 가지의 유전자가 반복하는 모습을 보였다. 이로 인해 성장률이 저조한 모습을 보였다. 이후 7번째 세대에서 뛰어난 성능을 보인 돌연변이를 만나 그림 7에 나타난 것과 같이 8번째 세대에서 20퍼센트가 넘는 급격한 성능 향상이 이루어졌다. 또한 이러한 성능 향상을 토대로 이후의 염색체에서는 비슷한 성질의 우수한 염색체가 부모로 선정되어 한 염색체의 유전자로 수렴하여 세대가 진행되어도 더 이상 진화하지 않고 비슷한 염색체가 반복되는 모습을 보여주었다.

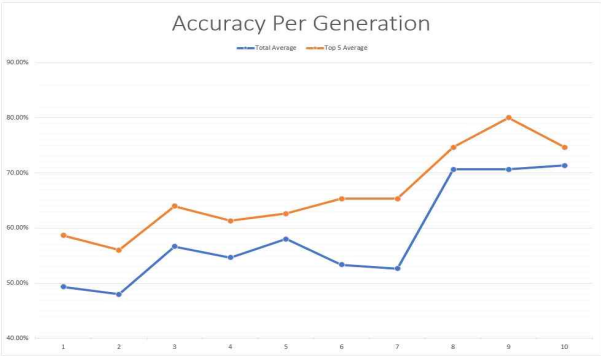

그림 7. 세대별 평균 정확도의 차트
Fig. 7 Chart of result of average accuracy per generation

표 4는 데이터셋 부풀리기를 진행하지 않고 오직 15개의 훈련 데이터셋으로 학습한 모델과 사람이 작성한 ImageDataGenerator 속성으로 데이터셋을 부풀린 모델과 유전 알고리즘을 통해 작성된 ImageDataGenerator 속성으로 학습한 모델 중 가장 뛰어난 결과를 보인 모델과의 비교이다. 데이터셋 부풀리기를 전혀 진행하지 않은 모델은 40%의 정확도를 보여주었으며 사람이 작성한 속성은 73%의 정확도를 보여주었지만 유전 알고리즘을 통해 작성된 속성은 86%의 정확도를 보여주었다.

표 4. 사람이 임의로 설정한 데이터와 유전 알고리즘이 적용된 데이터와의 비교
Table 4 Comparison of human generated data with genetic algorithm applied data

	basic data	human generated data	genetic algorithm applied data
rotation_range	–	90	19
width_shift_range	–	0.1	0.01
height_shift_range	–	0.1	0.56
shear_range	–	0.5	0.94
zoom_range_min	–	0.8	0.7
zoom_range_max	–	1.5	1.58
horizontal_flip	–	TRUE	TRUE
vertical_flip	–	TRUE	FALSE
accuracy	40%	73.33%	86.67%

즉 데이터셋 부풀리기 행위 자체는 모델의 정확도 향상에 많은 영향을 미쳤으며 사람이 생각한 속성보다 유전 알고리즘을 통해 생성된 속성이 정확도를 높이는 부분에 있어서 더 좋은 결과를 도출해냈다. 현재 이 실험은 간단한 컨볼루션 신경망 모델을 통해 실험을 하였으나 이러한 데이터 부풀리기에 대한 유전 알고리즘 적용은 다양한 신경망 모델에도 적용 가능한 것으로 보인다.

Ⅴ. 결 론

본 논문에서 제시된 유전적 알고리즘은 생명이 진화한 것처럼 무작위 개체에서 의미 있는 값을 찾아내는 과정이다. 따라서 인간이 생각할 수 없는 범위 내에서 결과를 도출해내는 과정은 의미 있는 과정이라고 볼 수 있다. 또한 이 논문을 통해 유전적 알고리즘이 데이터셋 생성에 미치는 영향은 많다는 결론을 도출해낼 수 있었다. 하지만 본 논문에서는 한 가지 신경망 모델에 대한 유전적 알고리즘 적용을 진행하였으므로 다른 신경망 모델에 적용을 위해서는 더 많은 연구가 진행되어야 할 것이다. 이러한 방법을 통해 데이터셋을 부풀리는 과정에서 조금 더 정확도를 높일 수 있는 방안이 마련될 것이라고 유추된다.

향후 딥러닝을 사용하는 다양한 분야 중 특히나 민감 정보를 다루는 의료정보 분야에서 데이터 부풀리기는 정확한 학습을 위해 절실히 요구되는 것으로 영상데이터의 특성을 고려한 데이터 부풀리기에 대한 실험이 지속되어야 할 것이다.

감사의 글

이 성과는 2018년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임. (No. 2018R1C1B5083789)

참고문헌

[1] Gartner, Gartner Top 10 Strategic Technology Trends for 2018 [Internet]. Available : https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/.

[2] S. M. Chi, "Architectures of Convolutional Neural Networks for the Prediction of Protein Secondary Structures", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 5, pp.728-733, May, 2018.

[3] J. H. Kim, D. S. Choi, H. S. Lee and J. W. Lee, "Target Classification of Active Sonar Returns based on Convolutional Neural Network", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 10, pp.1909-1916, Oct, 2017.

[4] Y. H. Hong, "A Study on the Automatic Appraisal of the Posted Land Price Using Genetic Algorithms and Artificial Neural Networks", M.S. dissertation, Ajou University, Suwon, pp.2, 2005.

[5] Y. K. Choi and J. H. Park, "Predictive Control based on Genetic Algorithm for Mobile Robots with Constraints", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no. 1, pp.9-16, Jan, 2018.

[6] C. W. Son and S. B. Lee, "Realization of home appliance classification system using deep learning", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 21, no. 9, pp.1718-1724, Sep, 2017.

[7] Wikipedia, Article of Keras [Internet]. Available : https://en.wikipedia.org/wiki/Keras.

[8] S.H Par, and Y.B Cho, "Convolution Neural Network based TW3 Maximum Height Prediction System", *Journal of the Korea Institute of Information and Communication Engineering*, Vol. 22, No. 10: 1314~1319, Oct. 2018

[9] Keras Documentation, "Image Preprocessing" [Internet]. Available: https://keras.io/preprocessing/image/#imagedatagenerator.

[10] K. B. Kim, "Combining A* and Genetic Algorithm for Efficient Path Search", *Journal of the Korea Institute of Information and Communication Engineering*, vol. 22, no.7, pp.943-948, Jul. 2018.

[11] C.H Hwang, H.S. Kim and H.K. Jung,"Detection and Correction Method of Erroneous Data Using Quantile Pattern and LSTM", *Journal of information and communication convergence engineering,* vol 16.No4. pp. 242-247, Dec. 2018

박상현(Sang-Hyun Park)

2013년～현 재: 대전대학교 정보보안학과 재학

※관심분야 : Deep-learning, 유전알고리즘 등

조영복(Young-Bok Cho)

2005: 충북대학교 전자계산학과 공학석사
2012: 충북대학교 전자계산학과 공학박사
2016: 충북대학교 의학과 박사과정수료
2012-2018: 충북대학교 소프트웨어학과 초빙교수

2018년～현 재: 대전대학교 정보보안학과 조교수
※관심분야: 의료영상처리, 정보보안, 의료정보보호, 모바일보안 등