Journal of Korea Multimedia Society Vol. 21, No. 10, October 2018(pp. 1195-1202)
https://doi.org/10.9717/kmms.2018.21.10.1195

전이학습을 이용한 효율적인 기타코드 분류 시스템

박선배[†], 이호경[††], 유도식[†††]

An Efficient Guitar Chords Classification System Using Transfer Learning

Sun Bae Park[†], Ho-Kyoung Lee[††], Do Sik Yoo[†††]

ABSTRACT

Artificial neural network is widely used for its excellent performance and implementability. However, traditional neural network needs to learn the system from scratch, with the addition of new input data, the variation of the observation environment, or the change in the form of input/output data. To resolve such a problem, the technique of transfer learning has been proposed. Transfer learning constructs a newly developed target system partially updating existing system and hence provides much more efficient learning process. Until now, transfer learning is mainly studied in the field of image processing and is not yet widely employed in acoustic data processing. In this paper, focusing on the scalability of transfer learning, we apply the concept of transfer learning to the problem of guitar chord classification and evaluate its performance. For this purpose, we build a target system of convolutional neutral network (CNN) based 48 guitar chords classification system by applying the concept of transfer learning to a source system of CNN based 24 guitar chords classification system. We show that the system with transfer learning has performance similar to that of conventional system, but it requires only half the learning time.

Key words: Artificial Neural Network, Audio Signal Processing, Convolutional Neural Network, Transfer Learning

1. 서 론

높이가 다른 둘 이상의 음을 동시에 울릴 때 발생하는 합성음을 코드(chord)라고 하는데, 코드를 통해 음악은 풍부한 감정과 분위기를 표현할 수 있다. 코드는 구성 음들의 배치에 따라 다양한 종류가 존재하는데, 구성 음들 중 가장 낮은 음인 근음(root)에 따라 전체적인 높이가 결정되고, 나머지 음들의 높낮이에 따라 코드의 분위기가 결정된다. 다양한 악기 혹은 악기들의 합주로 코드 음을 낼 수 있지만, 그중에서도 여섯 개의 현을 이용하여 소리를 내는 기타는 두 개 이상의 음을 동시에 내기 쉬워 '작은 오케스트라'

※ Corresponding Author : Do-Sik Yoo, Address: (04066) Dept. of Electronic and Electrical Engineering., P-705, Hongik University, Mapo-Gu Wausan-Ro 94, Seoul, Korea, TEL : +82-2-320-3048, FAX : +82-2-320-1193, E-mail : yoodosik@hongik.ac.kr
Receipt date : Jul. 25, 2018, Revision date : Sep. 13, 2018
Approval date : Sep. 27, 2018
[†] Dept. of Electronic and Electrical Engineering., Graduate School, Hongik University
(E-mail: sun882125@gmail.com)
[††] Dept. of Electronic and Electrical Engineering., Graduate School, Hongik University
(E-mail: hklee@hongik.ac.kr)
[†††] Dept. of Electronic and Electrical Engineering., Graduate School, Hongik University
(E-mail: yoodosik@hongik.ac.kr)
※ This research was supported by Basic Science Research Program through the National Research Foundation of Korea funded by the Ministry of Education under grant (NRF-2018R1D1A1B07045719, NRF-2016 R1D1A1 B03932502) and Korea Electric Power Corporation (Grant number:R18XA02).

라 불릴 정도로 코드 음을 만들기 적절한 악기이다.

기타 애호가들이라면 어떤 기타 음악이 연주되고 있을 때, 그 음악의 코드가 어떻게 진행되는지 궁금한 경우가 있게 마련이다. 따라서, 연주되는 음악소리로부터 코드 진행을 자동적으로 추출해 낼 수 있는 장치가 있다면 기타 애호가들에게 큰 도움이 될 것이다. 과거 음성신호처리기술이 그리 발달되지 않았던 때는 단순히 합성음의 주파수를 인지하여 어떤 코드인지 알아내는 방법을 사용하였다. 이 방법은 구현이 간단하지만 좋은 성능을 내지 못했는데, 그것은 지판을 운지하는 위치나 완력에 따라 음의 주파수가 미세하게 변하기 때문이다. 최근에 주목받고 있는 기계학습 기법은 기타 코드 분류 시스템의 구성에도 이용되어 좋은 성능을 보이고 있다. 기계 분류 시스템에서는 효과적인 분류를 위하여, 시시각각으로 입력되는 데이터로부터 제한된 갯수의 데이터들을 만들어 낸 다음 특징이라 불리는 이 데이터들을 사용하여 분류 작업을 실행하는 것이 보통이다. 특징 추출 기법으로는 은닉 마르코프 모델(hidden markov model)을 이용한 확률적 분석을 통한 기법이나[1,2], 필터링 등을 통해 구성 음들의 배음 성분들을 한데모아 소수의 효과적인 특징들로 압축하는 기법[3,4]들이 있다. 더 나아가, 간단한 전처리과정을 거친 코드 음으로 합성곱 신경망(Convolutional neural network)을 학습하여 특징 추출과 분류까지 한 번에 수행하는 기법도 제안된 바 있다[5].

한국어 화자는 일본어를 배울 때 다른 외국어에 비해 습득이 빠르다. 그 이유는 일본어의 경우 한국어와 문법구조가 유사하기 때문에 이를 새로 학습하지 않고 일본어의 어휘 정도만 익혀도 기초적인 구사가 가능하기 때문이다. 기계학습에도 이와 유사한 개념의 학습방법이 있는데, 데이터 분류 등의 문제를 해결하기 위해 이미 학습이 완료된 모델을 이용하여 빠른 학습을 수행하는 '전이학습(transfer learning)'이 그것이다[6-8]. 전이학습은 다양한 시스템에 적용할 수 있으나 합성곱 신경망에 특히 용이하게 적용할 수 있다[7-10].

현재까지 전이학습을 이용한 인공신경망 분류시스템 연구는 영상처리 분야에 국한되어 있으며, 기타 코드 분류를 비롯한 음성처리 분야에는 연구가 진행되지 않고 있다. 따라서 우리는 전이학습을 통한 기타 코드 분류시스템을 효율적으로 구성함으로써, 음성처리 분야에 전이학습을 활용할 수 있는 기반을 다지고자 한다. 이를 위하여 본 논문에서는 24종류의 기타 코드 분류 시스템을 학습해놓은 후, 이 시스템을 이용하여 48종류의 기타 코드 분류 시스템을 빠르게 학습하는 방법을 제시한다. 우리는 제안하는 전이학습을 이용할 경우, 전이학습을 수행하지 않고 처음부터 학습을 진행한 시스템과 비교하여 학습해야 할 변수의 개수와 분류 정확도는 비슷한 수준이 유지됨에도 학습 시간은 절반으로 감소하는 것을 모의실험을 통해 보인다.

2. 관련연구

2.1 합성곱 신경망

영상처리 분야에서는 커널(kernel)이라 불리는 행렬과 영상을 합성곱하여 영상 내 사물의 곡선과 선의 모양을 탐지하여 이 특징들로 사물을 분류하는 방법을 널리 사용하여 왔다. 인공신경망 연구가 진행됨에 따라 커널 행렬의 모양마저 학습을 통해 최적화하는 방법이 도입되어 영상 내 사물의 어떠한 모양을 탐지할지까지 시스템이 스스로 결정하는 합성곱 신경망이 제안되었다[9,10]. 합성곱 신경망은 영상처리 분야에서 좋은 성능을 보일 뿐만 아니라 층의 개수를 증가시켜도 학습해야 할 변수의 수가 크게 증가하지 않아 수십 개의 층도 쉽게 구현이 가능하기 때문에 더 복잡한 문제도 해결할 수 있게 되었다. 본 논문에서는, 사용하는 데이터가 1차원 벡터인 음성 데이터이기 때문에 커널이 열벡터인 1차원 합성곱 신경망을 다룬다.

합성곱 신경망은 합성곱 층, 풀링 층, 완전연결 층(fully-connected layer)의 세 가지 종류의 층으로 구성되어 있다. 합성곱 층에서는 커널 벡터들과 입력 데이터를 합성곱 하는데, 이를 수식으로 나타내면 식(1)과 같다.

$$\underline{h}_i = f(\underline{x} * \underline{k}_i + b_i \underline{J}), \quad i = 0, 1, \cdots, N_{kernel} - 1 \tag{1}$$

여기서 $\underline{h}_i$는 i번째 커널벡터 $\underline{k}_i$를 이용하여 합성곱으로 특징을 추출한 결과이며, b_i는 바이어스 값이고 $\underline{J}$는 모든 성분이 1로 이루어진 벡터이다. $f(\cdot)$는 활성함수이며, 주로 ReLU(rectifier linear unit)함수를 이용한다. 합성곱 신경망에선 보통 다수의 커널 벡터를 이용하기 때문에, 합성곱 층 출력 벡터의 수

가 상당히 증가한다. 따라서 최대값 풀링(max-pooling)[9]을 통해 차원의 수를 줄인다. 최대값 풀링은 입력된 벡터의 인접한 두개의 성분 중 큰 값만을 취하는 것으로, 수행하고 나면 벡터의 차원 수가 절반으로 줄어들게 된다. 풀링을 마친 뒤에는 출력 벡터들을 하나의 긴 벡터로 합친 뒤 완전연결 층에 입력하게 된다. 완전연결 층은 일반 퍼셉트론 층과 같이 가중치 행렬을 곱하고 바이어스 벡터를 더한 뒤 활성함수를 적용하는 층이다. 이를 수식으로 표현하면 식 (2)와 같다.

$$\underline{y} = f(W\underline{h} + \underline{b}) \tag{2}$$

여기서 W, $\underline{b}$는 각각 가중치 행렬과 바이어스 벡터이며, 활성함수 $f(\cdot)$는 주로 소프트맥스(softmax) 함수를 이용하게 된다. 합성곱 신경망은 구축하고자 하는 구조에 따라 수십 층 이상의 다층 구조로도 구축할 수 있지만, 본 논문에서는 '합성곱 층-풀링 층-완전연결 층'의 구조를 가진 '단층 합성곱 신경망'과 '합성곱 층-풀링 층-합성곱 층-풀링 층-완전연결 층'의 구조를 가진 '복층 합성곱 신경망'을 다룬다.

2.2 전이학습

전이학습이란 어떤 기계학습 시스템을 만들 때 이미 학습이 완료된 모델을 이용하여 학습을 빠르게 진행하는 방법을 말한다[6]. 전이학습의 장점은 크게 두 가지가 있다. 첫째, 학습시간을 단축할 수 있다. 이미 특정 기계학습 시스템을 구성하여 사용하고 있는 상황에서 측정방법의 발달 혹은 사용자의 설정변경에 데이터 형태나 데이터마다 부여된 부류가 달라지는 경우가 생길 수 있다. 데이터의 형태나 부류가 달라지는 경우, 일반적인 기계학습 시스템에서는 학습과정을 처음부터 진행해야 한다. 이때, 복잡한 구조의 딥러닝 시스템일 경우 학습에 필요한 시간이 매우 길다. 전이학습을 이용하여 새로운 시스템의 일부만을 빠르게 학습하면 많은 노력을 절약할 수 있다. 둘째, 데이터가 충분히 확보되어 있지 않은 경우에도 인공신경망에 전이학습의 개념을 도입하면 상당한 수준까지의 학습이 가능하다. 우수한 성능의 인공신경망 시스템을 구성하기 위해선 일정규모 이상의 데이터셋이 필요한데, 의학영상과 같이 일반인이 접근하기 어려운 데이터는 충분히 양을 확보하는데 상당한 난점이 있다. 전이학습을 이용하면, 원하는 분야와 유사한 분야의 대규모 데이터를 이용해 시스템을 일차로 학습한 다음 주어진 소수의 데이터를 이용하여 미세조정(fine-tuning)함으로써 수준 높은 학습 효과를 얻을 수 있다.

전이학습을 이해하기 위해서는 소스시스템(source system)과 목표시스템(target system)이라는 개념에 대해 알아 둘 필요가 있다. 소스시스템은 소스데이터(source data)와 소스작업(source task)로 구성되는데, 여기서 소스작업이란 소스데이터와 소스출력(source output)을 이용하여 학습이 완료된 기계학습 시스템을 말한다. 목표시스템은 새로 추가된 목표데이터(target data)와 그에 대응되는 목표출력(target output)을 이용하여 학습을 수행하려는 기계학습 시스템인 목표작업(target task)으로 구성된다. Fig. 1은 전이학습의 처리과정을 나타내는데, 이미 학습이 완료된 소스시스템의 일부 요소들을 이용하여 목표시스템을 효율적으로 구성하는 것이 전이학습의 궁극적 목적이라 할 수 있다.

S.J. Pan[6]은 전이학습 방법을 몇 가지로 분류하였는데 그중에서도 소스작업의 변수들 중 일부를 그대로 목표작업에 적용하는 '변수전이(parameter-transfer)'를 눈여겨볼 필요가 있다. 그 이유는, 변수전이가 합성곱 신경망에서 학습한 커널을 다른 시스템에 재사용하는 방식으로 쉽게 구현할 수 있어, 합성곱 신경망을 주로 사용하는 영상처리 분야에서 연구가 활발히 진행되고 있기 때문이다[7-10]. H.C. Shin[7]은 의학영상 분류 시스템을 인공신경망으로 구성할 때 확보할 수 있는 데이터의 양이 적어 충분히 학습할 수 없다는 난점을 극복하기 위하여, Image Net 데이터셋에 있는 동식물 영상을 이용하여 합성곱 신경망의 커널들을 학습한 뒤, 인간의 의학영상 데이터를 이용해 이 커널들을 미세조정 하는 방식으로 비교적 적은 데이터를 가지고 시스템을 충분히 학습하는 방법을 제시하였다. M. Oquab[8]은 영상에

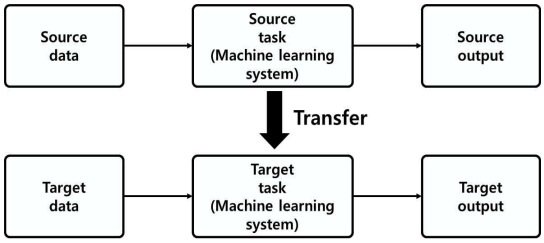

Fig. 1. Processing Diagram of Transfer Learning.

서 사물들의 위치를 특정하는 시스템을 인공신경망으로 구성하였는데, 공간 안에 각 사물들이 위치해있는 영상 데이터를 충분히 확보하지 못하는 문제를 전이학습 기법을 이용하여 해결하였다. 보다 구체적으로 각 사물들을 분류하는 시스템을 먼저 구성해두고, 전이학습을 이용하여 사물 분류 시스템에서 학습한 변수들을 위치 특정 시스템으로 가져와 구성하였다. D.C. Ciresan[9]는 라틴 문자를 인식하는 시스템으로부터 변수들을 전이받아 한자(漢字)를 인지하는 시스템을 구축하였다. 외견상 라틴 문자와 한자는 큰 차이가 있기 때문에 한자 인식 시스템을 구축하는데 큰 도움이 되지 않을 것 같지만, 전이학습을 이용한 한자 인식 시스템이 더 짧은 학습시간과 높은 분류 정확도를 가진다는 것을 보였다. S.A Prajapati[10]는 치과질환이 있는 엑스레이 영상을 분류하는 시스템을 구축하였다. 이때 251개의 엑스레이 영상으로만 학습할 때보다 1000여종의 동식물 데이터가 있는 대규모 데이터셋인 ILSVRC-2012을 이용하여 합성곱 커널들을 학습한 뒤, 이 커널들에 엑스레이 영상을 이용하여 미세조정 하는 방식으로 구축한 시스템이 더 좋은 성능을 나타내는 것을 보였다.

3. 제안하는 방법

본 논문에서는 전이학습을 이용하여 효율적인 기타 코드 분류시스템을 구축하는 방법을 제안하고 그 성능을 검증한다. 이를 위하여 우선 소스시스템으로 사용할 '24종 기타 코드 분류시스템'을 합성곱 신경망으로 만들어 학습해둔다. 그리고 24종류의 새로운 기타 코드 데이터를 추가하여 목표시스템인 '48종 기타코드 분류시스템'을 구축할 때, 전이학습을 이용하여 보다 빠르게 학습을 완료할 수 있는 것을 보인다. 제안하는 방법으로 구축한 시스템의 성능을 보다 다양한 상황에서 검증하기 위해 네 가지의 전이학습 시스템과 이와 동일한 구조의 비전이학습 시스템과 비교하는데, 그 기준은 학습해야 하는 변수의 개수, 분류 정확도, 그리고 학습 시간이다. 제안하는 구성 방법은 Fig. 2와 같다.

Fig. 2의 상단부분은 소스시스템으로 12개의 음을 근음으로 가지는 메이저코드와 마이너코드 24종류[13]를 분류하는 시스템이다. 소스데이터는 512차원의 데이터이며, 이를 이용하여 단층 합성곱 신경망인 소스작업 1과 복층 합성곱 신경망인 소스작업 2를

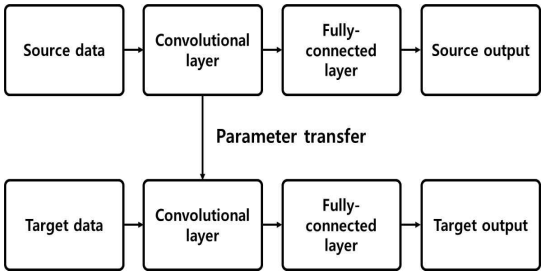

Fig. 2. Processing Diagram of Proposed scheme.

학습해둔다. 학습이 완료된 후 두 종류의 소스작업에서 커널 벡터들의 값을 저장해둔다.

Fig. 2의 하단부분은 목표시스템으로서 12개 음을 근음으로 가지는 메이저코드, 마이너코드, 메이저 세븐코드와 마이너 세븐코드의 총 48종의 코드를 분류하는 시스템이다. 목표데이터는 48종의 데이터를 목표작업에 따라 512차원 벡터와 그 절반의 저주파 성분인 256차원 벡터로 만들어 구성하였다. 또한 보다 다양한 모의실험을 위하여 네 종류의 목표작업을 구축하는데, 목표작업 1과 목표작업 2는 각각 소스작업 1과 소스작업 2의 커널벡터들을 그대로 재사용하는 단층과 복층 합성곱 신경망으로서, 512차원 벡터들을 이용하여 완전연결 층의 변수들을 학습한다. 목표작업 3과 목표작업 4도 마찬가지로 각각 소스작업 1과 소스작업 2를 이용한 단층과 복층 합성곱 신경망이지만, 완전연결 층의 변수들을 256차원 벡터 데이터로 학습한다. 목표작업 3과 목표작업 4는 512차원 벡터를 이용하여 학습한 커널을 256차원 벡터를 분류하는 시스템에 적용할 때, 즉, 관측환경이 바뀔 때의 성능을 확인하기 위해 구축하였다.

4. 시뮬레이션 및 결과 분석

4.1 데이터셋

시뮬레이션을 위한 데이터셋은 전기기타를 PC에 연결, 직접 코드 음을 연주하면서 녹음하는 방법으로 확보하였으며, 12음의 근음마다 메이저코드, 마이너코드, 메이저 세븐코드와 마이너 세븐코드 총 48종류로 구성되어 있다. 녹음 데이터의 샘플링은 44.1kHz, 16bit로 설정하였고, 각 샘플을 약 92ms초 단위로 잘라 4096차원 벡터로 만든 뒤 4096개의 0을 붙이는 제로패딩을 수행했다. 이후 이산푸리에 변환한 뒤 절반의 대칭성분을 제거하여 다시 4096차원 벡터로 만

들었다. 추가적으로 각 복소 데이터 벡터들의 성분별 절대값을 취해 실수 데이터 벡터로 변환한다. 그리고 실수화 된 데이터 벡터들의 스케일을 통일하기 위한 정규화 과정을 식 (3)에 따라 수행한다. 즉, 정규화된 벡터 $\underline{Y}$는

$$\underline{Y} = \frac{\underline{X}}{\sum_i X_i} \tag{3}$$

과 같이 구한다. 여기서 $\underline{X}$는 위에서 설명한 바와 같이 복소 데이터 벡터들의 성분별 절대값을 취해 얻은 실수 벡터를 의미하며, X_i는 $\underline{X}$의 i번째 성분을 뜻한다. 마지막으로 정규화 된 데이터 벡터에서 필요에 따라 저주파성분 512개와 256개를 취하였다. 데이터의 총 개수는 학습 데이터 62000개, 검증 데이터 6200개로 구성되어 있다. 이 방법은 최근의 기타 코드 음 분류 연구를 참고하여 진행하였다[3].

데이터 샘플에 코드별 라벨을 부여할 때는 다루는 부류의 수에 따라 24, 혹은 48차원 벡터를 만들어 각 코드에 해당하는 자리만 1이고 나머지 성분은 0이 되는 '원 핫 인코딩(one-hot encoding)'을 수행하였다[14]. 원 핫 인코딩은 기계학습에서 라벨을 부여할 때 많은 장점이 있으며, 특히 교차 엔트로피 비용함수 연산을 빠르게 수행할 수 있는 장점이 있다.

4.2 시스템 구성

시뮬레이션은 CPU로는 인텔 코어 i7-7700K, 그래픽카드로는 NVIDIA GTX 1060, 그리고 메모리로는 16GB DRAM을 장착한 시스템에 우분투 16.04 운영체제을 설치한 환경에서 실행하였다. 신경망 알고리듬은 구글이 공개한 오픈소스 기계학습 라이브러리인 텐서플로우 GPU 버전[15]을 이용하여 구현하였고, 모든 학습할 변수들은 평균이 0이고 표준편차가 0.1인 '잘린 정규분포(truncated normal distribute)'로 초기화하였다.

소스시스템을 구축할 때는 전체 48종류의 코드 데이터 중 12개의 근음을 가지는 메이저코드와 마이너코드만을 포함시킨 24종류의 데이터를 활용하였는데, 그 수는 학습데이터 30000개와 검증 데이터 3000개이다. 단층 합성곱 신경망인 소스작업 1은 Fig. 3과 같이, 먼저 128차원 커널 벡터 8개를 이용하여 합성곱을 수행하고 ReLU 활성함수를 적용하여 특징을 추출한 다음 특징 벡터들을 최대값 풀링하여 그 차원

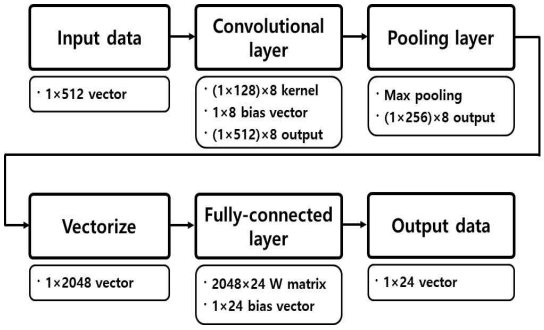

Fig. 3. System Structure of Single Layer Convolutional Neural Network for Source Task 1.

을 절반으로 낮춰준다. 그 후 절반의 차원이 된 특징 벡터들을 일렬로 나열하여 2048차원 벡터로 만들고, 여기에 2048×24차원 가중치 행렬을 곱한 뒤 24차원 바이어스 벡터를 더한다. 이후 소프트맥스 활성함수에 입력하여 얻은 결과물인 24차원 벡터로 최종 분류를 한다.

복층 합성곱 신경망인 소스작업 2는 Fig. 4와 같이, 먼저 소스작업 1과 동일한 합성곱 층과 풀링 층을 거친 뒤 다시 64차원 커널 벡터들을 이용하여 특징을 추출한다. 그 후, ReLU 활성화함수를 적용하고 최대값 풀링을 수행하여 다시 차원을 절반으로 낮춘다. 이 과정을 거치면 8개의 특징벡터들은 각각 128차원 벡터가 된다. 그 후 특징벡터들을 일렬로 나열하여 1024차원 벡터로 만든 후, 여기에 1024×24 가중치 행렬과 24차원 바이어스 벡터와 소프트맥스 활성함수를 이용하여 최종 분류를 한다.

목표시스템을 구축할 때는 전체 데이터셋인 48종류 코드 데이터 62000개를 이용하여 학습하고, 6200

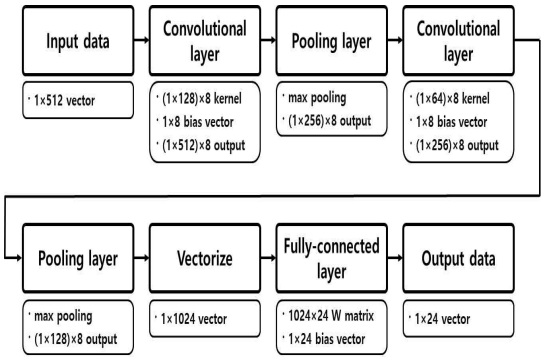

Fig. 4. System Structure of Double Layer Convolutional Neural Network for Source Task 2.

개의 데이터로 검증을 한다. 또한 앞서 언급한대로 합성곱 신경망의 모든 변수들을 학습하는 것이 아니라, 소스작업에서 학습을 마친 합성곱 커널과 바이어스를 전이받아 활용하고, 완전연결 층의 가중치 행렬과 바이어스 벡터만을 학습한다. 추가로, 대조군으로 사용하기 위하여 목표작업 1-4와 동일한 구조의 합성곱 신경망 시스템 네 종류도 구현하여 전이학습을 이용하지 않고 처음부터 학습시킨다.

학습은 소스작업과 목표작업 모두 각각 1000개의 샘플을 임의로 취하여 학습하는 미니배치학습(mini batch learning[15])을 진행하였다. 최적화기법은 Adam Optimizer[16]를 학습률 10^{-3}로 설정하여 각 배치마다 1000회 수행하였다. 결과값은 각 8종류 기계학습 시스템을 100회 구현한 결과를 평균하여 기록하였다.

4.3 결과 분석

본 절에서는 목표작업 4종류와 대조군으로 전이학습을 수행하지 않고 처음부터 학습을 진행하는 시스템 4종류, 총 8종류의 기계학습 시스템을 구현하여 모의실험을 진행한 결과에 대해 설명한다.

우선 각 시스템에서 학습해야 할 변수를 살펴보면 Table 1과 같다. 합성곱 신경망에서 합성곱 층의 학습해야 할 커널 수는 사용자가 설정한 크기와 개수에 따라 산술적으로 증가한다. 반면, 완전연결 층은 입력하는 데이터의 차원에 따라 가중치 행렬 성분 수가 기하적으로 증가하기 때문에 대부분의 학습해야 할 변수들은 완전연결 층에 존재한다. 전이학습을 이용한 시스템에서는 완전연결 층의 변수들만 학습을 수행하기 때문에 학습해야 할 변수의 개수에는 큰 차이가 없다.

분류 정확도는 Table 2와 같이 전이학습을 수행했을 때 약간 감소하는 것을 볼 수 있다. 그 이유는 소스작업에서 학습한 커널들이 24종류의 기타 코드들의 특징을 추출하는데 최적화되어있기 때문이다. 하지만 분류 정확도 감소가 최소 0.07%에서 최대 1.32%로 현저한 수준은 아니며, 목표작업 3과 목표작업 4와 같이 관측환경이 바뀐 상황에서도 성능이 유지된다. 분류 정확도 감소는 완전연결 층의 학습이 끝난 후, 커널 벡터를 포함한 전체 변수에 대하여 미세하게 학습을 더 수행하는 미세조정 과정으로 해결할 수 있다.

마지막으로 Table 3에 기록된 1000개의 데이터로 이루어진 데이터 배치를 1회 학습하는데 필요한 시간을 보면 그 차이가 확연히 드러난다. 전이학습과 비전이학습은 학습해야 할 변수의 수는 차이가 나지 않지만, 전이학습의 경우 합성곱 커널들을 따로 학습하지 않고 재사용하기 때문에 학습시간에서 차이난다. 이것은 일반적으로 합성곱 층의 학습이 완전연결 층보다 오래 걸리기 때문이다. 학습시간의 차이는 거의 2배 정도이며, 특히 복층 합성곱 신경망에서 더 큰 차이를 보였다. 데이터의 수가 더 많고 층의 개수가 더 많은 합성곱 신경망일수록 전이학습의 학습속도 차이는 더욱 커질 것으로 예상된다.

Table 1. Number of learning variables

Input dimension	System	Transfer learning	Non-transfer learning
512	1 layer CNN	98,304	99,384
	2 layer CNN	49,200	50,744
256	1 layer CNN	49,200	50,232
	2 layer CNN	24,624	26,176

Table 2. Simulation result on classify accuracy [%]

Input dimension	System	Transfer learning	Non-transfer learning
512	1 layer CNN	99.288	99.420
	2 layer CNN	98.479	99.799
256	1 layer CNN	99.338	99.467
	2 layer CNN	98.670	98.845

Table 3. Simulation result on mean learning time per single batch [ms]

Input dimension	System	Transfer learning	Non-transfer learning
512	1 layer CNN	10.82	19.90
	2 layer CNN	14.41	27.34
256	1 layer CNN	7.28	12.82
	2 layer CNN	9.25	19.40

5. 결 론

본 논문에서는 기타 코드 음을 분류하는 시스템을 이전에 학습한 모델의 변수들을 그대로 이용하여 학습시간을 줄이는 전이학습 기법을 활용하여 구현하

고 그 성능을 평가해 보았다. 결론적으로, 전이학습을 수행할 때 학습해야 할 변수의 수는 큰 차이가 없고, 분류 정확도가 0.07-1.32%정도 낮아지지만 학습시간을 거의 절반 가까이로 줄일 수 있다는 것을 확인하였다. 또한 관측 환경이 바뀌어도 분류 정확도가 유지된다는 점에서 다양한 유사 분야의 인공신경망을 이용한 분류시스템으로 확장이 가능할 것으로 보인다.

전이학습은 이미 학습이 완료된 모델이 있어야만 사용할 수 있는 방법이지만, 기계학습 분야에서 시스템을 구성하여 사용하던 도중 관측 환경의 변화, 그리고 목표로 하는 데이터의 부류의 수가 변화하는 등의 상황은 충분히 발생할 수 있다는 점에서 필요성이 높다고 판단된다. 특히, 빅데이터 시대의 도래로 다양한 데이터들이 쏟아지고 있는 현 상황에서 전이학습은 시스템 구축 시간을 확연히 줄일 수 있는 효율적인 방안이라 판단된다.

향후 우리는 보다 더 다양한 기타 코드 분류시스템을 구성해두고, 피아노 등 다른 악기들과 합주하는 데이터를 확보하여 전이학습을 이용해 다양한 악기로 구성된 코드 음을 분류하는 시스템으로 확장이 가능한지 살펴볼 계획이다. 또한 미세조정 학습을 얼마나 진행해야 비전이학습과 동일한 수준의 성능을 보일 수 있는지에 대한 추가연구를 진행할 계획이다.

REFERENCE

[1] A.M. Barbancho, A. Klapuri, L.J. Tardon, and I. Barbancho, "Automatic Transcription of Guitar Chords and Fingering from Audio," *IEEE Transactions on Audio, Speech, and Language Processing*, Vol. 20, No. 3, pp. 915-921. 2012.

[2] A. Sheh and D. Ellis, "Chord Segmentation and Recognition Using EM-trained Hidden Markov Models," *Proceedings of the 4th International Society for Music Information Retrieval Conference*, pp. 183-189, 2006.

[3] J. Guerrero-Turrubiates, S. Ledsema, S. Conzalez-Reyna, and G. Avina-Cervates. "Guitar Chords Classification Using Uncertainty Measurements of Frequency Bins," *Mathematical Problems in Engineering*, Vol. 2015, Article ID. 205369, pp. 1-9, 2015.

[4] S. Arun and Y. Wang. "Key, Chord, and Rhythm Tracking of Popular Music Recordings," *Computer Music Journal*, Vol. 29, No. 3, pp. 75-86, 2005.

[5] S.B Park and D.S Yoo, "Performance Comparison of Guitar Chords Classification Systems Based on Artificial Neural Network," *Journal of Korea Multimedia Society*, Vol. 21, No. 3, pp. 391-399, 2018.

[6] S.J. Pan and Q. Yang. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering*, Vol. 22, No. 10, pp. 1345-1359, 2010.

[7] H.C. Shin, H.R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, et al., "Deep Convolutional Neural Networks for Computer-aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Transactions on Medical Imaging*, Vol. 35, No. 5, pp. 1285-1298, 2016.

[8] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1717-1724, 2014.

[9] D.C. Cireşan, U. Meier, and J. Schmidhuber, "Transfer Learning for Latin and Chinese Characters with Deep Neural Networks," *Proceeding of the 2012 International Joint Conference on Neural Networks,* pp. 1-6, 2012.

[10] S.A Prajapati, R. Nagaraj, and S. Mitra, "Classification of Dental Diseases using CNN and Transfer Learning," *Proceeding of 2017 5th International Symposium on Computational and Business Intelligence,* pp. 70-74, 2017.

[11] Y. LeCun, B.D. Boser, J.S. Denker, D. Henderson, R.E. Howard, W. Hubbard, et al., "Backpropagation Applied to Handwritten Zip Code

Recognition," *Neural Computation*, Vol. 1, No. 4, pp. 541-551, 1989.

[12] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, et al., "Recent Advances in Convolutional Neural Networks," A*rXiv Preprint arXiv:1512.07108*, 2015.

[13] Music Theory/Chords. https://en.wikibooks.org/wiki/Music_Theory/Chords (accessed Aug., 12, 2018).

[14] A.C. Timothy, "Random Forests, Decision Trees, and Categorical Predictors: The" Absent Levels" Problem," A*rXiv Preprint ArXiv: 1706.03492*, 2017.

[15] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, et al., "Tensorflow: Large-scale Machine Learning on Heterogeneous Distributed Systems." A*rXiv Preprint ArXiv: 1603.04467*, 2016.

[16] P.K. Diederik and B. Jimmy, "Adam: A Method for Stochastic Optimization," A*rXiv Preprint ArXiv:1412.6980*, 2014.

박 선 배

2015년 2월 홍익대학교 전자전기공학부(공학사)
2017년 2월 홍익대학교 대학원 전자전기공학과(공학석사)
2017년 3월~현재 홍익대학교 대학원 전자전기공학과 박사과정
관심분야 : 신호처리, 패턴인식, 영상처리 및 필터링 이론, 머신러닝 등

이 호 경

1994년 8월 남가주대학교 전기공학과(공학박사)
1994년 9월~1999년 3월 홍익대학교 전자전기공학부 조교수
1999년 4월~2004년 8월 홍익대학교 전자전기공학부 부교수
2004년 9월~현재 홍익대학교 전자전기공학부 교수
관심분야 : 이동통신, 채널부호이론, 정보이론, 신호처리, IoT 등

유 도 식

2002년 2월 미시간대학교 전기컴퓨터공학과(공학박사)
2006년 9월~2011년 3월 홍익대학교 전자전기공학부 조교수
2011년 4월~2016년 8월 홍익대학교 전자전기공학부 부교수
2016년 9월~현재 홍익대학교 전자전기공학부 교수
관심분야 : 통신 및 신호처리, 정보이론, 어레이신호처리, 영상처리 및 필터링, 신호처리, 변환광학, 머신러닝 등