

지속적 학습 환경에서 지식전달에 기반한 LwF 개선모델

강석훈^{1*} · 박성현²

Advanced LwF Model based on Knowledge Transfer in Continual Learning

Seok-Hoon Kang^{1*} · Seong-Hyeon Park²

^{1*}Professor, Department of Embedded Systems Engineering, Incheon National University, Incheon, 22012 Korea

²Graduate Student, Department of Embedded Systems Engineering, Incheon National University, Incheon, 22012 Korea

요 약

지속적 학습에서의 망각현상을 완화시키기 위해, 본 논문에서는 지식전달 방법에 기반한 개선된 LwF 모델을 제안하고, 이의 효율성을 실험 결과로 보인다. LwF에 지속적 학습을 적용할 경우, 학습되는 데이터의 도메인이 달라지거나 데이터의 복잡도가 달라지면, 이전에 학습된 결과는 망각현상에 의해 정확도가 떨어지게 된다. 특히 복잡한 데이터에서 단순한 데이터로 학습이 이어질 경우 그 현상이 더 심해지는 경향이 있다. 본 논문에서는 이전 학습 결과가 충분히 LwF 모델에 전달되게 하기 위해 지식전달 방법을 적용하고, 효율적인 사용을 위한 알고리즘을 제안한다. 그 결과 기존 LwF의 결과보다 평균 8% 정도의 망각현상 완화를 보였으며, 학습 태스크가 길어지는 경우에도 효과가 있었다. 특히, 복잡한 데이터가 먼저 학습된 경우에는 LwF 대비 최대 30% 이상 효율이 향상되었다.

ABSTRACT

To reduce forgetfulness in continuous learning, in this paper, we propose an improved LwF model based on the knowledge transfer method, and we show its effectiveness by experiment. In LwF, if the domain of the learned data is different or the complexity of the data is different, the previously learned results are inaccurate due to forgetting. In particular, when learning continues from complex data to simple data, the phenomenon tends to get worse. In this paper, to ensure that the previous learning results are sufficiently transferred to the LwF model, we apply the knowledge transfer method to LwF, and propose an algorithm for efficient use. As a result, the forgetting phenomenon was reduced by an average of 8% compared to the existing LwF results, and it was effective even when the learning task became long. In particular, when complex data was first learned, the efficiency was improved more than 30% compared to LwF.

키워드 : LwF, 지속적학습, 지식전달, 신경망, 치명적 망각

Keywords : LwF, Continual Learning, Knowledge Transfer, Neural Network, Catastrophic Forgetting

Received 20 December 2021, Revised 24 December 2021, Accepted 3 February 2022

* Corresponding Author Seok-Hoon Kang(E-mail:hana@inu.ac.kr, Tel:+82-32-835-8422)

Professor, Department of Embedded Systems Engineering, Incheon National University, Incheon, 22012 Korea

Open Access <http://doi.org/10.6109/jkiice.2022.26.3.347>

print ISSN: 2234-4772 online ISSN: 2288-4165

© 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.
Copyright © The Korea Institute of Information and Communication Engineering.

I. 서론

지속적 학습환경에서 과거 학습 데이터를 재학습하지 않고 순차적으로 새로운 데이터로 모델을 계속 학습시키면 치명적 망각 현상이 발생한다[1, 2]. 지속적 학습 환경에서 모델이 이전에 학습한 태스크들의 정보를, 학습이 거듭될수록 급격히 잊는 경우가 생기는 현상이다. 이 현상은 지속적 학습 환경에서 발생하는 주된 문제이며 동시에 해결해야 할 문제다. 이러한 망각현상을 제어하거나 개선하기 위해, 몇가지 분류의 방법들이 제시되고 있다. 정규화방법, 네트워크 동적확장방법, 메모리 재생방법이 가장 대표적이다.

정규화 방법은 이전에 학습된 데이터와 연관된 가중치들의 변화를 억제하는 방법이며, LwF(Learning without Forgetting)와 EWC(Elastic Weight Consolidation) 그리고 SI(Synaptic Intelligence)가 대표적인 방법이다. LwF[2, 3]는 과거 학습 데이터로 학습했던 모델의 출력값을 유지하여 과거 학습 데이터의 지식을 우회적으로 보존하여 사용하는 방법이다. EWC[2, 4]는 과거 학습 데이터에 대한 가중치 별 중요도를 측정하고 이를 사용하여, 새로운 데이터를 학습시 과거 학습데이터에게 중요도가 높은 가중치의 변화를 억제시키는 방법으로, 직접적으로 과거 학습 데이터의 지식을 보존한다. SI[2, 5]는 학습 과정에서 매 배치마다 변화하는 손실값의 변화량을 측정하고 이를 최소화 하는 방법으로, 과거 학습 데이터의 지식을 보존하는 방법이다. 이 방법은 과거 학습 데이터를 저장하는 추가 메모리가 필요 없고 모델의 변동이 없거나 적은 특징을 갖고 있다. 또한 유사한 데이터들을 순차적으로 학습할 때 좋은 성능을 보인다[6].

네트워크 동적 확장 방법은 필요에 따라 모델의 구조를 변경하여 모델의 정보 저장 능력을 동적으로 설정하는 방법이다. 그 중 한 방법인 DEN (Dynamically Expandable Networks)[7]은 희소 정규화를 사용하여 모델을 훈련하고 각 레이어에 추가할 뉴런 수를 결정하는 선택적 재훈련을 수행하여 학습하는 방식이다. 이러한 방법은 학습하는 데이터의 특징에 따라 모델의 크기가 동적으로 커지는 특징을 가지고 있다.

메모리 재생 방법은 과거 학습 데이터를 생성하거나 저장하는 방법을 사용하여 새로운 학습 데이터를 학습할 때 생성되어 저장된 데이터를 추가해 학습하여 이전 학습 데이터의 정보를 유지하는 방법이다. Deep

Generative Replay[8] 방법은 과거 학습 데이터를 생성하는 생성 모델을 추가로 학습하여 사용한다. 이러한 방법은 과거 학습 데이터를 저장하는 공간이 필요하거나 생성 모델을 위한 공간이 필요하므로 추가 메모리가 필요하고, 생성 모델을 사용하는 방법인 경우 2가지 모델을 동시에 학습해야 하므로 학습 과정이 복잡한 특징을 가지고 있다.

지속적 학습 환경에서 모델 성능을 평가하는 기준은 아직 따로 정해지지 않았고, 관련 논문마다 다양한 방법을 사용한다[1, 2]. 대부분의 연구에서는 성능 평가 방법으로 Multi-Task(MT) 방식과 Single-Incremental-Task(SIT) 방식 2가지를 선택하여 사용하고 있다. MT 방식은 각 학습 태스크마다 독립적으로 정확도를 계산하고, SIT 방식은 모든 학습 태스크를 통합하여 정확도를 계산하여 사용한다. MT 방법은 각 태스크별로 출력 뉴런의 그룹이 따로 존재하는 반면, SIT 방법은 모든 태스크가 하나의 출력 뉴런 그룹을 사용하는 구조이다. 이 가운데 MT 방법이 가장 많이 사용되고 있다[3, 5]. 본 논문은 LwF를 기반으로 지식전달 방법의 개념을 변형 적용하여 망각현상을 개선하고자 하며, 성능 평가방법으로 MT방법을 이용하고자 한다.

II. 본론

본 논문은 LwF를 바탕으로 하고 있다. LwF는 지식 종류[9]를 사용하여 과거 네트워크 출력을 유지하고 과거 지식을 유지한다. 과거에 학습한 데이터와 새롭게 학습할 데이터의 도메인이 동일하거나 유사한 경우 좋은 성능을 보이지만, 도메인이 다르거나 유사도가 낮다면 상대적으로 성능이 떨어지는 모습을 보인다[3]. LwF는 도메인이 다른 데이터들의 입력이 지속될 때 성능이 떨어지게 된다. 특히 상대적으로 복잡한 정보를 가지는 데이터를 학습한 이후, 단순한 정보를 가지는 데이터를 학습하게 되었을 때, 이전 데이터에 대한 망각현상이 두드러지는 경우가 많다[1,10]. 이 경우 이전 데이터로 학습된 가중치들이 구성하는 내부 구조가 새로운 데이터를 설명하기에 부족한 정보를 가지고 있으므로 학습 과정에서 새로운 데이터를 잘 설명할 수 있는 방향으로 구조가 변경된다. 따라서 가중치에 급격한 변화가 생기게 되고 이 변화로 인해 망각 현상이 더 크게 발생하는 것이

다. LwF는 과거 네트워크 출력을 유지하여 과거 데이터와 관련된 가중치의 변화를 우회적으로 억제하지만 도메인의 차이가 커짐에 따라 가중치의 변화가 점점 심해지므로 성능의 하락이 발생하게 된다. 본 논문은 이 문제에 집중하여, 복잡한 과거 데이터의 망각 현상을 완화시키는데 초점을 맞춘다.

지식 전달(Knowledge Transfer) 방법은 큰 네트워크의 중간 레이어 출력과 동일한 출력을 작은 네트워크의 중간 레이어가 출력하도록 하여 큰 네트워크의 지식을 작은 네트워크로 전달하는 모델 압축(Model Compression) 방법에서 사용된다. AT(Attention Transfer)[11], AB(Activation Boundaries)[12] 등과 같은 다양한 방법들이 존재한다. AT는 컨볼루션 레이어의 출력을 사용하여 네트워크가 중요하게 집중하는 부분의 정보를 전달하는 방법을 사용한다. 이때 상위 레이어부터 하위 레이어까지 다양한 레이어들의 출력을 사용하여 높은 수준부터 낮은 수준까지의 표현 정보를 사용한다. 이러한 정보를 가진 크고 강력한 Teacher 네트워크에서 작은 규모의 Student 네트워크로 정보를 전달하여 Student 네트워크가 Teacher 네트워크의 지식을 습득할 수 있게 하는 것이다. AB방식은 AT 방법과 비슷한 방법이다. 이 방법은 AT에서 사용한 중간 레이어의 출력 정보 대신, 활성화 정보를 사용한다. 기존 방법들은 강한 응답에 집중하여 학습되기 때문에 Decision Boundary의 형성에 악영향을 미칠 수 있으므로 출력의 값이 아닌 활성화, 비활성화 여부를 학습하는 것이 더 효과적이라고 보는 방법이다.

LwF는 말단의 일부 정보를 유지하여 과거 학습정보를 유지하려고 하지만, 네트워크의 중간 가중치 값은 새로운 태스크에 의해 지속적으로 변경되게 되며, 이것이 결국 망각으로 이어지게 된다. 따라서 하나의 네트워크에 유사하지만 다양한 도메인의 데이터를 학습시키고 하는 환경에서는, LwF가 보존하는 데이터까지, 이전의 학습정보를 적절하게 잘 전달할 필요가 있는 것이다. 본 논문에서는 지식전달 방법을 이용하여 LwF의 일부분을 재설계하고, 이로 인해 이전의 학습정보가 새로운 태스크로 잘 전달되게 하여 시스템에 부하가 적으면서도 이전 학습정보의 전달이 개선될 수 있도록 한다. 기존 LwF가 단순한 데이터로 시작해서 복잡한 데이터로 점차 학습해 가는 과정은 지식유지를 잘 하는 반면, 복잡한 데이터로 시작해서 단순한 데이터로 학습해갈 때 망각현상이 두드러지게 되는데, 본 논문의 방법은 전자

의 경우에는 유사한 망각비율을 가지지만, 후자의 경우에는 개선된 결과를 보이게 개선한다.

III. 지식전달을 이용한 개선된 LwF 모델

LwF의 한계점인 도메인 차이에 따른 성능 하락 현상을 해결하기 위한 방법 중 가장 단순한 방법은 이전 학습 데이터와 관련된 모든 가중치들을 고정시키는 것일 것이다. 하지만 이러한 방법은 새로운 데이터들을 위한 매개변수 공간을 축소시켜 새로운 데이터의 저조한 성능을 일으킬 수 있다. 따라서 가중치를 직접 고정하는 방법을 사용하지 않고 우회적으로 조절하는 LwF를 기반으로 과거 데이터와 관련된 가중치의 변화를 더 효과적으로 억제할 방법이 필요하다. 따라서 네트워크의 마지막 레이어 출력 분포뿐만 아니라, 추가적으로 중간 레이어 출력을 최대한 유지하게 하는 방법을 적용한다면, 기존 LwF보다 더 강한 제약이 걸리게 되고 가중치 변화를 효과적으로 억제할 수 있다. 본 논문은 이 부분에 지식 전달 방법을 변형하여 사용한다. 지식을 전달할 네트워크가 과거에 학습한 지식을 가지고 있는 네트워크라고 가정하고, 지식을 전달받을 네트워크를 새로운 지식을 학습할 네트워크라고 가정한다면 LwF와 지식 전달 방법을 결합할 수 있다. 따라서 이와 같은 형태로 이전 태스크들의 학습 지식을 새 태스크로 전달하는 구조로 변경하여 LwF 모델에 적용한다. 그림 1에 본 논문에서 제안하는 지식전달을 이용한 변형된 LwF 모델의 구조를 나타내었다.

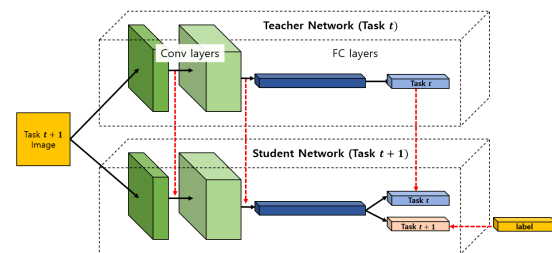


Fig. 1 Proposed Model using Knowledge Transfer

LwF는 과거 태스크의 출력을 유지하기 위해 현재 태스크를 사용해 출력분포를 사전에 생성하여 학습을 진행한다. 그러나 본 논문의 방법인 지식 전달을 적용한 변형된 LwF는, Teacher Network, Student Network 구조

를 통해, 이전 결과인 Teacher Network의 출력을 사용하므로, 다음 태스크를 위해 출력분포를 사전에 다시 생성할 필요가 없다. 그림 1의 본 논문의 구조에서 적용하는 LwF의 손실함수는 수식 (1), (2)와 같이 나타낼 수 있다.

$$L_{total} = L_{lwf} + L_{kt} \quad (1)$$

$$L_{kt} = \| F(T(I)) - F(S(I)) \|_2 \quad (2)$$

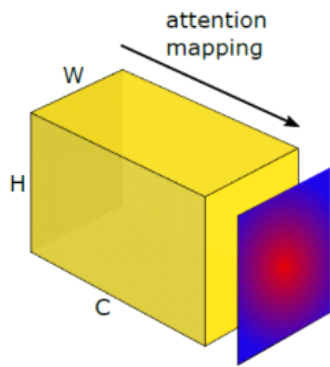


Fig. 2 Attention Calculation of the Attention Transfer

이 때, L_{total} 은 기존 LwF의 손실함수인 L_{lwf} 와 AT의 손실함수인 L_{kt} 의 합으로 정의하였다. L_{kt} 에서 $T(I)$, $S(I)$ 는 Teacher Network, Student Network가 학습 이미지 I 를 입력으로 받아 출력하는 중간 레이어의 값을 의미한다. 함수 $F()$ 는 그림 2와 같은 방법으로 N-Channel의 Feature를 Point-wise하게 평균을 구한다[11]. 이 수식을 사용하여 Teacher와 Student 출력의 차이를 줄이는 방향으로 기본적인 학습이 진행된다. 그러나 이 Loss 함수를 그대로 사용하면, AT 적용방법은 새로운 지식 습득에 유리한 반면, 상대적으로 과거 지식 유지 성능은 떨어지는 현상이 발생한다. 따라서 이 망각현상 문제의 완화를 위해서는 손실함수 수식을 수정해야 한다. 본 논문에서는 새로운 지식 습득 능력과 과거 지식 유지 성능의 적절한 조절을 위해 수식 (3)과 같이 손실함수를 변경하였다.

$$L_{kt} = \| T(I) - S(I) \|_2 \quad (3)$$

AT 방법은 중간 레이어의 출력을 채널 기준으로 평균을 내어 Attention Map으로 사용한다. 채널 기준으로 평균을 내는 방법은 출력 결과가 가진 많은 정보들을 제거하므로 지식 전달 사전 실험에서 AT 방법의 적용이

과거 지식 유지 성능이 떨어질 수 있게 되는 것이다. 따라서 본 논문에서는 채널 기준으로 평균을 내지 않은 출력값을 사용하여 최대한 과거 지식이 유지 될 수 있도록 했다.

Table. 1 LwF Learning Flow of This Paper

Algorithm 1. LwF with Attention Transfer
Require: Dataset: $D = \{D_1, \dots, D_T\}$, Teacher Network: Teacher , Student Network: Student
for each Task $t=1, \dots, T$ do
if $t=1$ then
Train the Student using normal cross entropy loss
else
Copy Student to Teacher
Train the Student using Eq. (3)

표 1은 본 논문에서 제안하는 AT를 적용한 LwF의 학습 과정을 나타낸 것이다. 첫 번째 태스크는 일반적인 교차 엔트로피 손실을 사용하여 학습을 진행한다. 이후 태스크는 이전 태스크까지 학습이 완료된 Teacher Network의 출력을 사용하여, 새로운 태스크를 학습하는 Student Network의 학습을 진행하게 한다. 학습이 완료되면 다음 태스크 학습을 위해 Student Network는 Teacher Network로 변경시키고 새로운 Student Network 학습에 사용되게 한다. 즉, 이전에 학습완료된 값들인 Student Network를, 다음 차례에는 Teacher Network가 되게 한뒤 다음 학습을 진행하여, LwF에서 수행하던 Y_0 를 얻기 위한 계산작업을 하지 않게 하였다.

AT 방법은 Teacher가 Student에게 학습 데이터에서 집중해야할 특징을 전달함으로써 Student가 학습 데이터를 더 잘 학습할 수 있게 돕는 것이다. LwF는 과거 출력 분포를 학습함으로써 과거 지식을 유지한다. 과거 출력 분포를 AT 방법에서 학습할 데이터셋으로 가정하여 LwF에 적용이 가능하게 된다. 따라서 본 논문의 방식은 Teacher가 Student에게 학습 데이터에서 기존에 집중했던 특징을 전달함으로써 과거 지식을 더 잘 유지할 수 있게 만드는 구조를 가지게 된다.

아울러, LwF는 학습되는 태스크가 늘어날수록 출력 레이어가 늘어나는 구조이므로 학습할 태스크가 몇 개의 클래스를 구분해야 하는지에 따라 계산량이 늘어나게 된다. 지식 전달 기반 방법들은 중간 출력 레이어의

출력을 비교해야 하므로 중간 레이어의 출력 크기만큼 계산량이 증가하게 된다. 일반적으로 중간 레이어 출력 크기는 출력 레이어의 출력 크기와 비교했을 때 상대적으로 큰 크기를 가지고 있다. 그러나 본 논문의 방법은 매 태스크가 늘어남에 따라 늘어나는 LwF의 출력 레이어와 달리 고정된 개수만을 Teacher Network로 변경하여 사용하도록 구성하였다. 따라서 기존 LwF에 비하여 추가적인 시스템 부하가 거의 없다.

IV. 실험 및 결과

지속적 학습 환경에서 발생하는 치명적 망각 현상은 태스크 데이터 사이의 유사도(도메인 유사도)가 낮아질수록 강도가 심해지는 경향을 보인다[3, 5]. 이를 참고하여 단순한 특징을 가진 데이터에서 점차 복잡한 특징을 갖는 데이터로 학습이 진행되는 시나리오와, 반대로 복잡한 특징을 갖는 데이터에서 점차 단순한 특징을 갖는 데이터로 진행되는 시나리오로 구성된 2가지 시나리오를 실험하였다. 본 실험에서 사용할 데이터는 Mnist, Emnist, Fashion mnist, Cifar10 데이터이다. 표 2에 이 데이터를 이용한 실험 시나리오를 나타내었다. 여기에서 데이터간의 복잡도는 LwF의 워업방법을 사용하여 진행하였다. 이 과정에서 데이터간의 정확도를 측정하면 상대적인 복잡도 결과를 얻을 수 있게 된다.

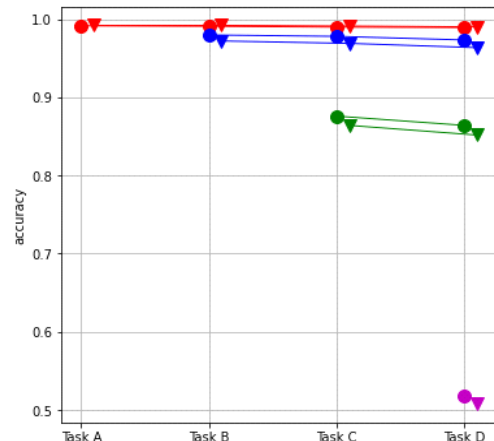
Table. 2 Experimental Scenario

Task Scenario	Task A	Task B	Task C	Task D
1	Mnist	Emnist	Fashion	Cifar10
2	Cifar10	Fashion	Emnist	Mnist

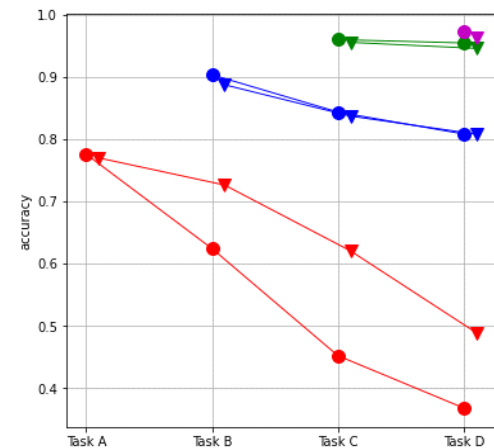
표 2에서, 시나리오 1은 단순한 특징을 가진 데이터에서 점차 복잡한 특징을 갖는 데이터로 학습이 진행된다. 시나리오 2는 복잡한 특징을 갖는 데이터에서 점차 단순한 특징을 갖는 데이터로 진행된다. 각 데이터를 단일로 학습했을 때 달성하는 최대 정확도에 따라 학습의 난이도를 판별하였다. 최대 정확도를 기준으로 오름차순으로 정렬하면 Cifar10, Fashion, Emnist, Mnist 순서가 된다. 이 결과를 이용하여 단순한 특징을 가진(학습하기 쉬운) 데이터로는 Mnist, Emnist 데이터를 할당하였고, 복잡한 특징을 가진 데이터(상대적으로 학습하기 어려

운)로는 Fashion mnist, Cifar10 데이터를 할당하였다.

학습 배치 크기는 100개, epoch는 각 태스크별로 20회로 설정하였고, 모델은 기본 LwF 방법을 사용하는 모델과 동일한 실험을 진행하여 비교하였다. 실험결과는 그림 3과 표 3에 나타내었다.



(a) Scenario 1



(b) Scenario 2

Fig. 3 Accuracy Comparison Graph of Experimental Results with LwF

그림 3은 두가지 시나리오에 대해, 태스크가 진행됨에 따른 각 태스크의 정확도를 비교하여 나타낸 것이다. 표 3은 그 중, 마지막 태스크단계에서의 정확도 값을 비교하여 나타내어, 최종단계에서 학습유지정도를 비교

할 수 있게 하였다.

Table. 3 Final Accuracy Test Results for Each Task

Scenario	Method	Task A	Task B	Task C	Task D
1	LwF	98.9	97.3	86.4	51.7
	This Paper	99.0	97.2	86.4	50.8
2	LwF	36.7	80.8	95.4	97.2
	This Paper	48.8	80.9	95.1	97.1

그림 3과 표 3의 시나리오 1의 결과에서, 단순한 특징을 갖는 데이터에서 점차 복잡한 특징을 갖는 데이터로 학습이 진행됨에 따라 본 논문의 결과가 LwF의 결과보다 미묘하게 저하되었지만 크게 변화가 없는 값이다. 이는 LwF 손실값에, 과거 학습데이터의 지식 유지를 위한 추가적인 손실이 추가됨에 따라 생기는 제약때문에 발생하는 성능의 변화이다. 이러한 현상은 기존 LwF 손실의 가중치를 크게 설정했을 때, 즉 강한 제약을 걸게 되었을 경우에도 발생하는 현상이다. 반면 시나리오 2의 경우, 학습하기 어려운 데이터에서 점차 학습하기 쉬운 데이터로 학습이 진행됨에 따라, LwF와 비교하여 본 논문의 방법이 과거 학습데이터 지식 유지 능력이 더 높은 것을 알 수 있다. 특히 본 논문의 목표인 이전 데이터의 망각비율 감소를 고려해서 관찰해 볼 경우, 태스크 A의 경우 본 논문의 방법은 LwF에 비해 정확도가 약 30% 이상 향상되었으며, 평균 약 8% 정도 향상된 것을 알 수 있다.

치명적 망각 현상은 모델의 규모가 클수록 크게 발생할 수 있으므로, 이를 위한 실험도 진행하였다. 작은 규모의 모델은 2개의 컨볼루션 레이어와 풀링 레이어 그리고 2개의 FC 레이어로 구성된 작은 모델을 사용하였고, 큰 규모의 모델은 VGG16 모델을 기반으로 한 모델을 사용하였다. 2가지 모델의 자세한 구조는 표 4와 같다.

Table. 4 Model Setup of two Structures with Different Scales

Model	Layer Name	Output Dim	Filter Size	Weight #
Small	Conv1	30x30x32	3x3	896
	Conv2	28x28x64	3x3	18496
	Pool1	14x14x64	2x2	-
	FC1	128	-	165760
	FC2	10	-	1290

Model	Layer Name	Output Dim	Filter Size	Weight #
Big	Conv1	32x32x64	3x3	1792
	Conv2	32x32x64	3x3	36928
	Pool1	16x16x64	2x2	-
	Conv3	16x16x128	3x3	73856
	Conv4	16x16x128	3x3	147584
	Pool2	8x8x128	2x2	-
	Conv5	8x8x256	3x3	295168
	Conv6	8x8x256	3x3	590080
	Conv7	8x8x256	3x3	590080
	Pool3	4x4x256	2x2	-
	FC1	128	-	524416
	FC2	10	-	1290

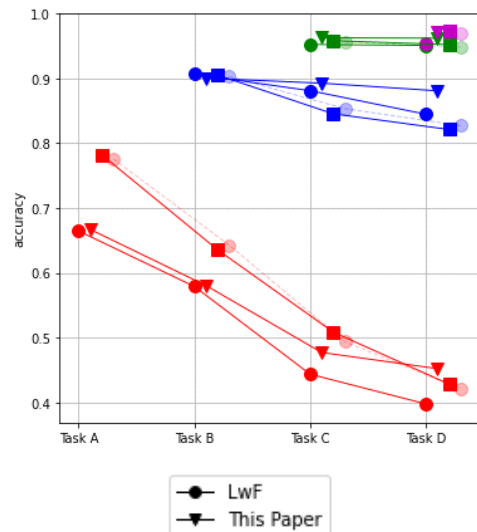


Fig. 4 Accuracy Comparison Graph of Two Structures with Different Scales

그림 4는 본 논문의 방식을 세모와 네모로 표현하여, 동그라미로 표현한 LwF 방식과 비교한 것이다. 과거의 학습정확도를 나타내는 망각비율에서 본 논문의 방법이 기존 LwF보다 약 10% 이상 개선되었음을 확인할 수 있다.

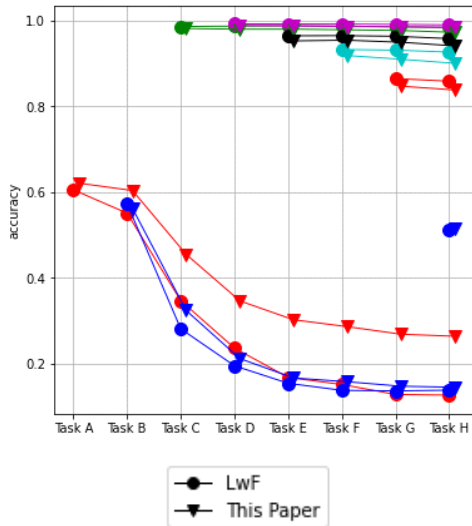


Fig. 5 Accuracy Comparison Graph with Long Task

그림 5에는 태스크가 길어진 경우에 대한 결과 그래프를 나타내었다. 각 태스크는 Cifar100-Cifar 10-Mnist-Emnist-Emnist-Emnist-Fashion-Cifar10 과 같이 이루어 지도록 하여, 실험 데이터셋으로 구성할 수 있는 범위에서 복잡-단순-복잡의 형태를 가지도록 시나리오를 구성하였다. 태스크 A와 B는 각각 모든 Cifar100의 클래스 중 25개의 클래스를 할당하였고 (A: 0~25 클래스 / B: 26~50 클래스) D, E, F 는 각각 Emnist의 아라비아 숫자 10개, 영문 대문자 26개, 영문 소문자 26개를 할당하여 실험한 것이다. 복잡한 데이터가 학습된 후 다시 단순한 데이터가 학습되고 이어서 다시 복잡한 데이터가 처리된 경우의 결과이다. 단순한 데이터의 경우 제약이 적은 LwF의 결과가 정확도에서 약간 우세하나, 본 논문의 결과도 거의 유사하게 나왔으며, Task A의 제일 복잡한 데이터의 경우, 7개의 태스크가 추가 학습되는 동안, LwF의 경우는 정확도가 처음 학습결과인 0.61에서 0.14로 약 22.9% 수준까지 떨어졌지만, 본 논문의 경우는 0.62에서 0.29로 약 46.8% 정도를 유지하고 있음을 보이고 있어서, 본 논문의 방법이 태스크가 길어져도 충분한 이전 학습결과를 유지하고 있으며, 복잡한 데이터의 학습결과에 대한 망각률이 LwF와 비교하여 개선되었음을 보이고 있다.

V. 결론

본 논문에서는 LwF에서 데이터 복잡도에 의한 도메인 차이에 따른 성능 하락 현상을 해결하기 위해, 지식 전달 방법을 LwF에 접목시켰고, 이를 위해 LwF에서 사용할 수 있는 Teacher Network, Student Network를 정의하였다. 또한 구체적으로 이를 적용할 수 있는 진행 흐름을 제안하고 구현하였다. 이를 통해, 이전에 학습된 Student Network 결과를 다음 태스크의 Teacher Network로 변경하여 사용할 수 있게 하여, 이전 단계의 지식이 계속해서 다음 태스크로 이어질 수 있도록 하였다. 그 결과 본 논문의 방식은 데이터의 내용이 상대적으로 복잡해지는 경우는 기존의 LwF와 유사한 능력을 유지하였으며, 데이터의 내용이 상대적으로 단순해지는 경우에 대해서는 약 8% 가량의 정확도 개선을 가져올 수 있었다. 특히, 본 논문이 목표로 하는 치명적 망각의 정도를 파악하기 위한 첫 번째 태스크 데이터의 경우, 그 정확도가 30%가량 개선되어, 본 논문의 방법이 LwF에서 과거 학습 태스크의 지식 유지 성능을 강화시키고 있음을 알 수 있었다. 향후 지식전달의 다양성을 포함할 수 있는 연구가 필요하다.

References

- [1] R. M. French, "Catastrophic forgetting in connectionist networks," *Trends in cognitive sciences*, vol. 3, no. 4, pp. 128-135, Apr. 1999.
- [2] G. I. Parisi, R. Kemker, J. L. Part, C. Kanan, and S. Wermter, "Continual lifelong learning with neural networks: A review," *Neural Networks*, vol. 113, pp. 54-71, May. 2019.
- [3] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 12, pp. 2935-2947, vol. 40, no. 12, pp. 2935-2947, Dec. 2018.
- [4] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," in *Proceedings of the national academy of sciences*, vol. 114, no. 13, pp. 3521-3526, Mar. 2017.
- [5] F. Zenke, B. Poole, and S. Ganguli, "Continual learning through synaptic intelligence," in *Proceedings of the 34th*

- International Conference on Machine Learning*, vol. 70, pp. 3987-3995, 2017.
- [6] Y. Hsu, Y. Liu, A. Ramasamy, and Z. Kira, "Re-evaluating continual learning scenarios: A categorization and case for strong baselines," *arXiv preprint*, arXiv:1810.12488, 2018.
- [7] J. Yoon, E. Yang, J. Lee, and S. J. Hwang, "Lifelong learning with dynamically expandable networks", *arXiv preprint*, arXiv:1708.01547, 2017.
- [8] H. Shin, J. K. Lee, J. Kim, and J. Kim, "Continual Learning with Deep Generative Replay," *arXiv preprint*, arXiv: 1705.08690, 2017.
- [9] G. Hinton, O. Vinyals, and J. Dean, "Distilling the Knowledge in a Neural Network," *NIPS Workshop*, 2014.
- [10] K. McRae, and PA. Hetherington, "Catastrophic Interference is Eliminated in Pretrained Networks," in *Proceedings of the 15th Annual Conference of the Cognitive Science Society*, pp. 723-728, 1993.
- [11] S. Zagoruyko and N. Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *arXiv preprint*, arXiv:1612.03928, 2016.
- [12] B. Heo, M. Lee, S Yun, and JY. Choi, "Knowledge Transfer via Distillation of Activation Boundaries Formed by Hidden Neurons," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019.



강석훈(Seok-Hoon Kang)

1995년 8월 한양대학교 전자통신공학과 공학박사
2004년 3월 ~ 현재 인천대학교 임베디드시스템공학과 교수
※ 관심분야 : 인공지능, 딥러닝, 인식시스템, 임베디드 시스템, 모바일 시스템



박성현(Seong-Hyeon Park)

2020 3월 ~ 현재, 임베디드 시스템 공학과 석사과정
※ 관심분야 : 인공지능, 딥러닝, 임베디드 시스템