

# Initial Research for Continual Learning Methods on Class-incremental Scenario

이름 김태희

지도교수 유종빈

## 연구 배경 및 목적

현재 딥러닝을 기반으로 한 여러가지 솔루션은 각각의 데이터셋과 주어진 상황에서 큰 성과를 보이며 빠르게 성장하고 있다. 하지만 기존의 딥러닝 모델의 예측 성능을 평가하는 방법은 하나의 데이터셋을 한 번에 걸쳐 학습하는 방식으로, 정해진 수의 클래스가 존재하여 그 클래스에 대한 예측 성능만을 평가하는 방법이다. 하지만, 우리의 실생활에 딥러닝 기술이 적용되는 자율 주행이나 안면 인식과 같은 응용 분야에서, 이러한 특성은 실용적이지 못하다. 모델이 기존에 학습한 데이터만으로 실제 상황에서 수많은 종류의 클래스를 예측해야 하기 때문이다. 따라서 딥러닝 모델을 실제 상황의 새로운 클래스에 지속적으로 학습(Class-incremental Scenario)하여 그 실용성을 제고할 필요가 있다.

한편, 현재의 모든 딥러닝 모델은 새로운 클래스를 추가적으로 학습할수록 이전에 학습한 클래스에 대한 예측 성능이 급격하게 감소하는 'Catastrophic Forgetting' 현상을 겪고 있다. 본 연구에서는 이러한 현상을 해결하기 위한 'Continual Learning' 방법을 본격적으로 개발하기에 앞서, 현재까지 진행된 선행 연구를 살펴보고자 한다. 즉, 본 연구는 기존 접근 방법의 장단점을 분석하고, 이를 기반으로 추후에 진행될 심화 연구에 대한 접근 방향을 설정하는 것에 목적이 있다.

## 연구 진행 과정

Continual Learning을 위한 기존 접근 방법은 크게 3가지로 나눌 수 있다. 학습되어 있는 딥러닝 모델에 대해 새로운 학습으로 인한 모델 파라미터의 급격한 변화를 막는 *Regularization* 기반의 방법, 이전에 학습한 클래스에 대한 대표성을 띄는 데이터를 저장하고, 새로운 학습에 이를 다시 활용하는 *Replay* 기반의 방법, 학습된 모델을 그대로 보존하고 새로운 학습을 위해 파라미터를 추가하는 방식인 *Architecture* 기반의 방법이 있다.

본 연구에서는 각 접근 방법의 근간이 되는 논문을 중심으로 파생되는 다양한 논문에 대해 조사하고 추후 연구에 적용될 수 있는 핵심 아이디어를 분석했다.

접근 방법	장점	단점	개선 방안
LwF (Learning without Forgetting)	• 이전에 학습한 클래스에 대한 데이터 불필요 • 학습 리소스 부담 적음	• 비교적 낮은 성능 • 이전에 학습한 딥러닝 모델 저장 필요	전체 파라미터에 대한 정규화가 아닌 중요도에 따른 선별적인 적용 필요
EWC (Elastic Weight Consolidation)	• 중요도에 따른 선별적인 파라미터 정규화 • LwF에 비해 높은 성능	• 중요도 계산을 위한 Overhead 존재 • Task가 증가할수록 성능이 크게 하락	여러 개의 Task에 대해서도 견고한 접근 방법 필요
SI (Synaptic Intelligence)	• 파라미터 별 중요도가 실시간으로 계산 • EWC에 비해 높은 성능	• 학습 리소스 부담 증가	중요도 계산을 보다 간단히 수행할 수 있는 방법 필요
BIC (Bias Correction)	• Task가 증가하더라도 견고한 일반화 성능	• 실용적 적용이 어려운 2-Stage 학습 방법	End-to-End 방식에서의 접근 필요

[표1] Regularization 기반의 접근 방법에 대한 분석 표

Regularization 기반의 접근 방법에서 가장 기본적인 방향을 설정한 논문은 LwF다. LwF는 이전 클래스에 대해 학습한 모델의 Output에 대해서 Distillation Loss를 이용해 지속적인 학습을 가능하게 하는 방법으로, 이를 개선해 가면서 접근 방법이 발전되었다. Regularization 대상이 되는 파라미터를 선별적으로 구성하거나 Classifier의 효과적인 학습을 유도하는 방식으로 발전해온 것을 확인할 수 있다. 하지만 Regularization 방식은 리소스 부담이 적다는 장점이 있지만 낮은 성능과 불안정한 학습으로 인해 타 접근 방법과 융합되어 사용되고 있다.

접근 방법	장점	단점	개선 방안
iCaRL (Incremental Classifier and Representation Learning)	• 정규화 기법에 비해 높은 일반화 성능	• 추가적인 메모리 공간 필요 • 휴리스틱한 Exemplar	각 클래스에 대한 Exemplar 선별 방식 개발
Mnemonics	• Bi-level Optimization을 통한 최적의 Exemplar 선별	• 고해상도에 이미지에 대한 학습 리소스 부담 • 학습 불안정	Exemplar 선별 과정에서 Herding 알고리즘 도입
LUCIR (Learning a Unified Classifier)	• 새로운 Loss 함수 구성만으로 iCaRL에 비해 높은 성능	• 데이터셋에 따른 적절한 하이퍼파라미터를 찾기 어려움	-
DER (Dark Experience for General Continual Learning)	• 비교적 단순한 방법으로 현재까지 가장 높은 성능	• Task 간의 Domain Gap이 큰 경우에 대해 불안정	데이터 분포가 크게 변화하는 경우를 대비할 수 있는 방법 필요

[표2] Replay 기반의 접근 방법에 대한 분석 표

Replay 기반의 접근 방법에서는 iCaRL이 기초적인 접근 방법으로, 이전 클래스의 대표성을 띄는 Exemplar를 Buffer Memory에 저장하여 새로운 학습에 다시 활용하는 방식이다. 후속 연구에서는 Exemplar를 보다 효과적으로 선별하는 방법, Classifier에서 발생하는 새로운 Task에 대한 Bias를 해결하는 방법 등을 통해 발전해왔다. Replay 기반의 방식은 학습 리소스에 대한 부담이 적을 때, 가장 큰 효과를 볼 수 있으며 그 성능 또한 우수하여 Continual Learning 분야에 큰 영향을 끼쳤다. 다만 Exemplar를 저장하기 위한 상황이 구성되지 않는다면 사용될 수 없다는 한계점이 존재한다.

접근 방법	장점	단점	개선 방안
DER (Dynamically Expandable Representation)	• 이전 학습의 파라미터를 지속적으로 학습에 활용 가능	• Task마다 저장해야 할 모델이 늘어나 리소스에 부담	Model Pruning을 통해 동적으로 모델 경량화 필요
FOSTER (Feature boosting and Compression)	• Task 간의 분포가 크게 변화할 때에도 Robustness 보장	• 2-Stage 학습 방식 • 다소 복잡한 학습 방법 & 모델 구조	Minor Class에 대한 Residual을 유지할 수 있는 방법 개발

[표3] Architecture 기반의 접근 방법에 대한 분석 표

Architecture 기반의 접근 방법의 경우, 기존에 학습된 모델에 대해 어떠한 학습도 진행하지 않고, 추가적인 모델을 구성하여 이들을 융합하는 방식이 주를 이룬다. 이로 인해 과거에 학습한 지식에 대한 보존이 잘되어 타 접근 방법에 비해 월등한 성능을 가지는 경우가 많다. 하지만 Task가 늘어날수록 학습된 모델을 계속 저장하거나 확장해야 하기 때문에 리소스 부담이 다른 학습 방법에 비해 크다는 단점이 있다.

## 결과 및 분석

본 연구에서는 딥러닝 모델을 여러 데이터 분포를 가진 데이터셋을 서로 다른 시기에 학습할 때 나타나는 'Catastrophic Forgetting' 현상을 해결하기 위한 Continual Learning 방법을 3가지 분류로 나누어 조사하고 분석했다.

연구 초기에는 Regularization 기반의 접근 방법이 초기에 많이 제시되었으며, 최근 연구로 올수록 Architecture 기반에 이를 융합하는 방식으로 발전되었다. Replay 기반의 접근 방법은 Memory에 데이터를 저장할 수 있는 환경에서의 지속적인 학습 방법을 제안했고, 여러 파생 분야가 개척되며 큰 영향력을 끼쳤다. 결국 본 연구를 통해, 학습 환경, 모델의 크기, 실용성, 학습해야 할 Task의 개수 등과 같은 다양한 환경 변수에 의해 적절한 접근 방법이 사용될 수 있다는 결론을 얻을 수 있었다.

한편, 최근에는 Vision Transformer 모델을 기반으로 Prefix, Adapter 등을 Parameter-efficient Fine-tuning(PEFT)하는 방식인 L2P, DualPrompt 방식이 제안되고 있다. 추후 연구에서는 이번 연구를 통해 학습한 통찰을 바탕으로 최신 연구 동향에 맞는 새로운 접근 방법을 고안해내고자 한다.