

# 강건하고 공정하며 확장 가능한 데이터 중심의 연속 학습

한국과학기술원 ■ 이재길<sup>\*</sup>·신기정<sup>\*\*</sup>·박찬영·황의중<sup>\*\*</sup>

(주)에스아이에이 ■ 최예지

## 1. 서 론

최근 들어 딥러닝(deep learning)은 비약적인 발전을 이루었으며, 이로 인해 이미지 분류/인식, 영상 인식, 음성 인식, 자연어 번역 등의 성능이 크게 향상되었다. 이러한 딥러닝 발전의 주요 원인 중 하나는 학습 데이터로 사용 가능한 빅데이터(big data)의 양이 급격하게 증가하고 있기 때문이다. 최근 인공지능 학계에서는 모델보다 데이터에 더 집중해야 한다는 데이터 중심 인공지능(data-centric AI)으로의 전환이 많은 관심을 받고 있다. 한편, 연속적으로 축적되는 지도·비지도 데이터를 실시간으로 평생 학습할 수 있는 딥러닝 기반의 인공지능 모델 기술이 필요해졌다. 예전의 데이터로 학습되어 얻은 지식을 유지하면서 새로운 데이터로 학습되어 얻은 지식을 추가로 획득해야 한다. 이를 위해 연속 학습(continual learning) 혹은 평생 학습(lifelong learning)이 인공지능 혹은 기계학습 분야의 주요 연구 분야로 떠오르기 시작하였다. 그러나, 현재까지 개발된 연속 학습 기술은 초기 단계에 머무르고 있으며, 실세계의 빅데이터를 충분히 고려한 환경에서의 연구가 진행되지 않았다. 따라서, 실세계의 데이터 품질, 데이터(태스크) 변화, 데이터 종류를 반영할 수 있는 강건하고 공정하며 확장 가능한 “데이터 중심”의 연속 학습에 대한 연구가 필요하다 (그림 1).

## 2. 데이터 중심 연속학습

### 2.1. 실세계 데이터 품질을 고려한 연속 학습

실세계 학습데이터에는 노이즈가 포함될 수 있고,

편향될 수 있고, 클래스 별로 데이터가 충분하지 않은 품질 문제가 매우 흔하게 발생한다[1]. 일례로 널리 사용되는 10가지 벤치마크 데이터에 평균 3.4%의 레이블 오류가 포함되어 있다. 또한 기계변역 모델 훈련에 사용된 학습데이터에 남성에 관한 글이 여성보다 더 많다. 일반적으로 대다수의 실세계 데이터는 클래스별로 Long-Tailed 분포를 따르기 때문에 꼬리에 해당하는 클래스는 데이터가 충분하지 않다. 이 때문에 일반적인 딥러닝 모델에서는 노이즈에 강건한(robust) 학습 방법론, 편향성에 공정한(fair)한 학습 방법론, Few-Shot 학습 방법론이 활발하게 연구되었다. 반면 연속 학습 연구는 아직 초창기이기 때문에, 학습데이터는 노이즈가 없이 깨끗하고, 편향성이 심하지 않고, 클래스별로 충분한 데이터가 있는 기초적이지만 실세계에서 비현실적인 환경을 가정하고 있다. 따라서 데이터 품질을 고려한 강건하고 공정한 연속 학습 기술을 개발하고 있다.

### 2.2. 실세계 데이터 변화를 고려한 연속 학습

연속 학습을 장기간 수행하면서 태스크와 이에 따른 데이터 분포가 변화하면 이러한 변화를 탐지하고 새로운 태스크에 잘 적응하는 것이 필요하다. 태스크 변화는 항상 순식간에 발생하지 않으며 점진적으로 변화할 수 있으며 이로 인해 태스크 간에 중첩이 발생할 수 있다. 따라서 태스크 경계가 분명하지 않고 불분명한 경우가 흔히 존재할 것으로 예상된다. 또한, 기존의 연속 학습 연구에서는 태스크 변화가 다소 인위적이다. 예를 들어 1과 2를 분류하는 태스크에서 3과 4를 분류하는 태스크 같은 설정을 가정한다. 현재 연속 학습 연구는 아직 초창기이기 때문에, 태스크가 순간적으로 명확하게 변하는 비현실적인 환경을 가정하고 있다. 따라서 데이터 변화를 고려한 강건하고 공정한 연속 학습 기술을 개발하고 있다.

\* 종신회원

\*\* 정회원

<sup>†</sup> 이 논문은 2022년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2022-0-00157, 강건하고 공정하며 확장 가능한 데이터 중심의 연속 학습)



그림 1 데이터 중심 연속 학습의 AS-IS 및 TO-BE.

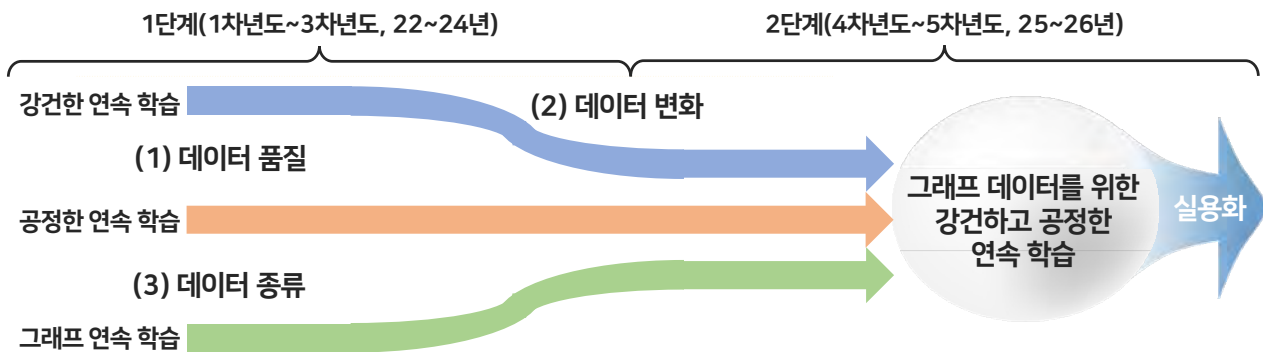


그림 2 본 연구의 수행 단계 및 목표.

### 2.3. 실세계 데이터 종류를 고려한 연속 학습

그래프 데이터는 실세계의 다양한 관계 및 현상을 모델링 할 수 있어서, 다양한 응용에서 널리 도입되었다. 예를 들어, SNS 혹은 e-Commerce에서 사용자 간 혹은 사용자와 상품 간의 상호 작용을 나타낼 수 있다. 또한, 바이오 인포매틱스에서 단백질 기능 예측을 위해 단백질-단백질 상호 작용을 나타낼 수 있다. 따라서, 인공지능 및 기계학습 학계 및 산업계에서는 그래프 데이터를 위한 심층망 연구가 매우 활발하게 진행되고 있다. 반면 그래프 데이터를 위한 연속 학습 기술은 거의 연구된 바가 없다. 하지만 SNS, WWW, e-Commerce, 단백질 상호 작용 등 상기 실세계 그래프는 시간에 따라 끊임없이 변화하기 때문에 연속 학습이 필수적이다. 연속 학습은 주로 이미지 데이터를 위주로 연구되었고, 이러한 연구를 그래프 데이터에 그대로 적용하기 어려우며, 그래프 데이터를 위한 연속 학습 기술은 매우 도전적인 주제이다. 인스턴스(정점, 간선, 부분 그래프 등)들이 독립적이지 않으며, 서

로 연결되어서 의미를 형성한다는 그래프 데이터의 고유한 특성과 그래프 신경망의 고유한 구조 때문에, 기존 합성곱 신경망 등에서 사용되는 기술을 단순 적용하는 것은 어렵다. 따라서 데이터 종류를 확장하는 그래프 데이터를 위한 연속 학습 기술을 개발하고 있다.

### 3. 연구 단계 및 목표

데이터 노이즈에 강건하고, 데이터 편향에 공정하며, 그래프 데이터로 확장가능한 데이터 중심의 연속 학습 기술을 두 단계에 걸쳐 개발하고 있다 (그림 2). 1단계로 데이터 품질과 데이터 변화를 동시에 고려하는 강건한 연속 학습과 공정한 연속 학습이 하나의 축이며, 데이터 종류를 고려하는 그래프 데이터 연속 학습이 나머지 하나의 축으로 구성된다. 강건한 연속 학습과 공정한 연속 학습은 ① 노이즈와 편향성이 변하는 환경, ② Long-Tailed 분포 및 Few-Shot 환경, ③ 태스크 경계가 불분명한 환경 등 실세계 데이터를 최대한 반영하는 매우 도전적인 문제를 해결하고자



그림 3 본 연구에서 지원하는 데이터·도메인 및 변화 요인.

한다. 그래프 데이터 연속 학습은 ① 태스크 변화 환경, ② 도메인 변화 환경, ③ 클래스 변화 환경을 모두 지원하여 기본에 충실히 하고자 한다. 2단계에서는 1단계에서 개발한 강건한 연속 학습, 공정한 연속 학습, 그래프 데이터 연속 학습을 통합하고자 하며, 참여기업을 통해 사업화를 추진하고자 한다. 그래프 데이터 연속 학습도 태스크 경계가 불분명한 환경을 고려하도록 확장하고자 한다. 강건한 연속 학습과 공정한 연속 학습을 먼저 통합하며, 그래프 데이터 연속 학습을 추가하여 최종 목표를 달성하고자 한다.

데이터·도메인은 이미지, 테이블 데이터뿐만 아니라 그래프 데이터를 지원하며, 변화 요인은 데이터 분포 변화와 새로운 지식의 발견을 지원한다 (그림 3).

#### 4. 기술개발의 중요성 및 원천성

본 연구는 네 가지 측면에서 새롭고 중요하다고 할 수 있다. 먼저 책임 있는 인공지능(responsible AI)은 데이터 노이즈 혹은 편향성에도 불구하고 믿음직하고 공정한 결과를 내어주어 인공지능이 책임감을 가져야 한다는 요구이다. 이러한 흐름에 발맞추어 연속 학습도 Responsible AI에 부합되게 발전해야 하는 것이 자명하며, 강건하고 공정한 연속 학습에서 발생하는 기술적인 난제를 해결할 예정이다. 둘째, 그래프(graph) 데이터를 위한 연속 학습이다. 그래프 데이터는 우리가 일상에서 접할 수 있는 거의 모든 서비스의 바탕에 사용되고 있다. 그래프 데이터의 고유한 특성과 그래프 신경망의 고유한 구조를 감안하여, 그래프 데이터를 위한 연속 학습 분야에서 선도적인 위치를 점하고자 한다. 셋째는 멀티 모달(multi-modal) 연속 학습이다. 이미지, 텍스트 등 기존의 레거시 데이터뿐만 아니라 그래프 데이터를 추가하여, 2단계 연구에서 연구성과가 통합되면 레거시 데이터와 그래프 데이터를 동시에 받아들여 학습할 수 있는 멀티 모달 연속 학습이 세계 최초로 가능해질 것이라 예상된다. 넷째는 적용 가능한 응용의 확장이다. 그래프 데이터와 멀티 모달 연속 학습으로 인해 기존의 이미지 데이터를 위한 응용(예: 위성 영상 분석)에서부터 아직 연속 학

습이 적용되지 못했던 e-Commerce(예: 상품 추천), 바이오 인포매틱스(예: 단백질 기능 예측)까지도 지원 가능해질 것이다.

### 5. 연구개발 대상의 기술 현황

#### 5.1. 국내외 기술 현황

먼저 새로운 환경, 새로운 유형의 지식을 습득하는 연속 학습 모델 개발 연구가 있다. 현재 국내 기관에서 딥러닝 기술을 활용하는 시도가 많이 이루어지지만, 딥러닝 시스템을 위해 수집되는 데이터는 정적인 상황에 국한되어 있으며, 새로운 환경, 새로운 유형의 데이터 등장과 같은 연속 학습을 위한 학습데이터가 거의 축적되지 않았다. 국내 기업과 연구소에서 연속 학습을 위한 데이터 준비와 시스템 개발을 시도하고는 있으나, 연속 학습 모델을 훈련할 수 있는 데이터 부족으로 인해 아직 소프트웨어를 상용화하기까지 시간이 소요될 것으로 파악된다. 국내 다수 기업은 새로운 환경의 데이터에 대해, 전체 데이터를 모델을 다시 학습하여, 컴퓨팅 및 저장 비용 측면에서 비효율적으로 딥러닝 시스템을 구축하고 있다. SK T-Brain에서는 사전 지식을 모방하는 생성모델 기법[2]을 제시했으나 생성모델이 가지는 훈련의 불안정성 및 Mode Collapse 문제 등으로 복잡한 문제에서 한계를 지닌다. LG, University of Toronto 공동연구진은 배치 레벨의 Replay를 활용해 메모리 효율화를 이룬 모델[3]로 CVPR 2020 국제학술대회의 연속 학습 경연 대회에서 우승하였으나, 앞서 언급한 연속 학습의 근원적인 문제를 해결하지는 않았다. NAVER Labs에서는 도메인 변화 환경에서 치명적 망각 문제를 해결하기 위해 데이터 증강(augmentation)과 메타 학습 기법[4]을 개발하였으나, 도메인 간 클래스가 동일한 문제를 가정하므로 모델의 적용 가능성이 떨어진다.

그래프 데이터 분야의 연속 학습의 국내 연구는 전무하다. 라벨링 노이즈 및 모델 강건성에 대한 연구의 경우, 기존 국내 기관의 연속 학습 모델 연구는 변화하는 환경, 새로운 유형에 대해 망각을 하지 않는 것에 초점을 두어, 새로운 업무의 지식을 습득해나갈 때

라벨링 노이즈 분포의 변화를 전혀 고려하지 않고 있다. 데이터, 태스크에 따른 라벨링 노이즈 분포의 변화는 실제 세계 데이터에서 빈번히 등장하는 사례로, 연속 학습 모델의 수행 능력에 영향을 줄 수 있다. 다양한 태스크에 대한 일반화된 연속 학습 모델을 구현하기 위해서 라벨링 노이즈 분포의 변화에도 강건한 연속 학습 모델 기법 연구가 필요하다.

본 연구팀의 사전 연구를 설명하자면 기계학습의 강건성 및 공정성, 그래프 데이터 분석에 관한 세계 최고 수준의 연구를 수행하였다. 이재길 교수 연구팀은 기계학습의 강건성에 관한 다수의 연구를 수행하였다[5-9]. 황의중 교수 연구팀은 기계학습의 공정성과 공정성 및 강건성의 결합에 관한 다수의 연구를 수행하였다[10-14]. 신기정 교수 연구팀은 그래프 데이터를 위한 온라인 알고리즘을 다수 개발하였으며, 특히 실시간으로 변화하는 그래프 스트림에서의 비지도학습(이상탐지, 압축, 구조 분석, 간선 예측)과 관련한 다수의 연구를 수행하였다[15-18]. 박찬영 교수 연구팀은 다양한 종류의 그래프 데이터(동종(homogeneous), 다중(multiplex), 이종(heterogeneous) 그래프)를 위한 표현 학습 기법을 다수 개발하였다[19-25]. 최예지 박사가 연구책임자로 참여하는 SIA 기업은 기상, 위성 분야 인공지능 분석기술 개발의 대표적인 국내 기업이다.

## 5.2. 국외 기술 동향 및 수준

새로운 환경, 새로운 유형의 지식을 습득하는 연속 학습 모델 개발이 이루어졌다. IST Austria 대학 연구진은 클래스 변화 환경에서 새로운 클래스에 유동적으로 적응 가능한 특징 추출을 위해 전이 학습을 대표할 수 있는 사례 데이터를 리허설하는 방식[26]을 도입하였으나, 클래스 변화에 따른 메모리 증가로 효율성이 크지 않았다. Facebook에서는 이전에 학습한 데이터 중 일부를 메모리에 저장하고, 이를 이용해 계산된 그라디언트(gradient)와 새로운 데이터에 대한 그라디언트를 비교하여 과거 학습 내용을 잃지 않는 방향으로 모델을 학습하는 기법[27]을 개발하였으나, 이는 규제(regularization) 기반의 방법에 비해 비효율적 메모리 사용 등의 한계를 가지고 있다. 기존의 리허설 방식의 연속 학습 모델은 태스크의 종류가 많아지게 될 경우, 적용하기 어려우며 태스크마다 데이터를 모두 리허설 하는 것은 효율적이지 못하다. Deepmind 연구진은 기존 학습된 모델이 새로운 환경에 대해 학습할 때, 지식을 보존해야 하는 부분을 찾

아내는 연속 학습 모델[28]을 제안하였다. 사전 지식에 대해서는 모델을 유보하면서, 새로운 지식에 대해서만 모델을 학습하는 효율적인 구조를 제안하였다. University of Electronic Science and Technology of China는 그래프 데이터 분야에서 전이 학습을 대표할 수 있는 대푯값(prototype)을 선택하는 다양한 기법[29]을 개발하였으나, 이외에는 그래프 데이터 분야에서의 연속 학습 연구는 미진하다.

라벨링 노이즈 및 모델 강건성에 대한 연구의 경우, 기존의 연속 학습 모델 연구는 대체로 어떻게 치명적 망각(catastrophic forgetting)을 방지하는지에 중점적으로 다루었으나, 사전 지식과 새로운 지식의 라벨링 노이즈, 모델 강건성에 대해서는 거의 다루어지지 않았다. University of Illinois at Chicago, ETH Zurich 공동 연구팀은 사전 지식과 상반되는 새로운 지식이 주어질 때도 모델이 강건하게 새로운 지식을 습득할 방법[30]을 제안하였으나, 비슷한 태스크를 탐지하는데 매우 긴 시간이 걸려 효율성과 확장성(scalability)에 문제가 있는 것으로 판명되었다. University of Modena and Reggio Emilia에서는 사전 지식과 새로운 지식 간의 도메인이 모호하거나, 급작스러운 라벨링 분포의 변화와 같은 실제 응용 사례에서 발생할 수 있는 문제를 제기하였다[31].

## 6. 연구 개발 계획

데이터 노이즈에 강건하고, 데이터 편향에 공정하며, 그래프 데이터로 확장가능한 데이터 중심의 연속 학습 기술을 총 5년에 걸쳐 체계적으로 개발하고자 한다. 먼저 그래프 데이터 상에서의 태스크, 도메인, 클래스 변화 환경을 모두 지원할 수 있도록 그래프 신경망을 위한 종합적인 연속 학습 방법 체계를 개발하고 더 나아가 강건성과 공정성도 추가적으로 지원하고자 한다. 또한, 연속 학습 측면에서의 공정성을 정의하고, 이를 위한 공정한 샘플링 기반의 연속 학습 방법을 개발하고자 한다. 노이즈/편향성 분포도 데이터 분포와 함께 변화하는 도전적인 환경에서의 강건하고 공정한 Replay 혹은 Regularization 기반 방법을 개발하고자 한다. Few-shot 환경에서 노이즈에 강건한 연속 학습 방법 등 다양한 데이터 품질 문제를 종합적으로 고려하는 연속 방법을 주력으로 개발할 예정이다. 기존 데이터 종류를 위한 연속 학습 방법과 그래프 데이터를 위한 연속 학습 방법을 결합하여 두 종류 이상의 데이터를 동시에 지원하는 체계도 개발

---

하고자 한다. 마지막으로 위성/기후 데이터 AI 기반 분석 SW에 본 연구 성과를 도입하여 실용화 및 사업화를 수행할 예정이다.

### 참고문헌

- [ 1 ] Steven Whang, Jae-Gil Lee: Data Collection and Quality Challenges for Deep Learning. Proc. VLDB Endow. 13(12): 3429-3432 (2020).
- [ 2 ] Hanul Shin, Jung Kwon Lee, Jaehong Kim, Jiwon Kim: Continual Learning with Deep Generative Replay. NIPS 2017: 2990-2999.
- [ 3 ] Zheda Mai, Hyunwoo Kim, Jihwan Jeong, Scott Sanner: Batch-level Experience Replay with Review for Continual Learning. CoRR abs/2007.05683 (2020).
- [ 4 ] Riccardo Volpi, Diane Larlus, Grégory Rogez: Continual Adaptation of Visual Representations via Domain Randomization and Meta-learning. CVPR 2021: 4443-4453.
- [ 5 ] Dongmin Park, Hwanjun Song, Minseok Kim, Jae-Gil Lee: Task-Agnostic Undesirable Feature Deactivation Using Out-of-Distribution Data. NeurIPS 2021: 4040-4052
- [ 6 ] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, Jae-Gil Lee: Robust Learning by Self-Transition for Handling Noisy Labels. KDD 2021: 1490-1500
- [ 7 ] Hwanjun Song, Minseok Kim, Jae-Gil Lee: SELFIE: Refurbishing Unclean Samples for Robust Deep Learning. ICML 2019: 5907-5915
- [ 8 ] Hwanjun Song, Minseok Kim, Sundong Kim, Jae-Gil Lee: Carpe Diem, Seize the Samples Uncertain at the Moment for Adaptive Batch Selection. CIKM 2020: 1385-1394
- [ 9 ] Hwanjun Song, Minseok Kim, Dongmin Park, Yooju Shin, Jae-Gil Lee: Learning from Noisy Labels with Deep Neural Networks: A survey. IEEE Transactions on Neural Networks and Learning Systems (2022)
- [10] Ki Hyun Tae, Steven Euijong Whang: Slice Tuner: A Selective Data Acquisition Framework for Accurate and Fair Machine Learning Models. SIGMOD 2021: 1771-1783
- [11] Yuji Roh, Kangwook Lee, Steven Euijong Whang, Changho Suh: FairBatch: Batch Selection for Model Fairness. ICLR 2021
- [12] Yuji Roh, Kangwook Lee, Steven Euijong Whang, Changho Suh: Sample Selection for Fair and Robust Training. NeurIPS 2021: 815-827
- [13] Jae-Gil Lee, Yuji Roh, Hwanjun Song, Steven Euijong Whang: Machine Learning Robustness, Fairness, and their Convergence. KDD 2021: 4046-4047
- [14] Yuji Roh, Kangwook Lee, Steven Whang, Changho Suh: FR-Train: A Mutual Information-Based Approach to Fair and Robust Training. ICML 2020: 8147-8157
- [15] Siddharth Bhatia, Bryan Hooi, Minji Yoon, Kijung Shin, Christos Faloutsos: Midas: Microcluster-Based Detector of Anomalies in Edge Streams. AAAI 2020: 3242-3249
- [16] Jihoon Ko, Yunbum Kook, Kijung Shin: Incremental Lossless Graph Summarization. KDD 2020: 317-327
- [17] Dongjin Lee, Kijung Shin, Christos Faloutsos: Temporal locality-aware sampling for accurate triangle counting in real graph streams. VLDB J. 29(6): 1501-1525 (2020)
- [18] Minseok Kim, Hwanjun Song, Yooju Shin, Dongmin Park, Kijung Shin, Jae-Gil Lee: Meta-Learning for Online Update of Recommender Systems. AAAI 2022: 4065-4074
- [19] Namkyeong Lee, Junseok Lee, Chanyoung Park: Augmentation-Free Self-Supervised Learning on Graphs. AAAI 2022: 7372-7380
- [20] Dongha Lee, Su Kim, Seonghyeon Lee, Chanyoung Park, Hwanjo Yu: Learnable Structural Semantic Readout for Graph Classification. ICDM 2021: 1180-1185
- [21] Chanyoung Park, Carl Yang, Qi Zhu, Donghyun Kim, Hwanjo Yu, Jiawei Han: Unsupervised Differentiable Multi-aspect Network Embedding. KDD 2020: 1435-1445
- [22] Baoyu Jing, Chanyoung Park, Hanghang Tong: HDML: High-order Deep Multiplex Infomax. WWW 2021: 2414-2424
- [23] Chanyoung Park, Donghyun Kim, Jiawei Han, Hwanjo Yu: Unsupervised Attributed Multiplex Network Embedding. AAAI 2020: 5371-5378
- [24] Chanyoung Park, Donghyun Kim, Qi Zhu, Jiawei Han, Hwanjo Yu: Task-Guided Pair Embedding in Heterogeneous Network. CIKM 2019: 489-498
- [25] Seonghyeon Lee, Chanyoung Park, Hwanjo Yu: BHIN2vec: Balancing the Type of Relation in Heterogeneous Information Network. CIKM 2019: 619-628
- [26] Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, Christoph H. Lampert: iCaRL: Incremental Classifier and Representation Learning. CVPR 2017: 2001-2010.

- [27] David Lopez-Paz, Marc'Aurelio Ranzato: Gradient Episodic Memory for Continual Learning. NeurIPS 2017.
- [28] Chrisantha Fernando, Dylan Banarse, Charles Blundell, Yori Zwols, David Ha, Andrei A. Rusu, Alexander Pritzel, Daan Wierstra: PathNet: Evolution Channels Gradient Descent in Super Neural Networks. CoRR abs/1701.08734 (2017).
- [29] Fan Zhou, Chengtai Cao: Overcoming Catastrophic Forgetting in Graph Neural Networks with Experience Replay. AAAI 2021: 4714-4722.
- [30] Zixuan Ke, Bing Liu, Xingchang Huang: Continual Learning of a Mixed Sequence of Similar and Dissimilar Tasks. NeurIPS 2020.
- [31] Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, Simone Calderara: Dark Experience for General Continual Learning: a Strong, Simple Baseline. NeurIPS 2020.

## 약 력



### 이재길

2022~현재 KAIST 전산학부 교수  
2014~2022 KAIST 전산학부, 산업및시스템공학과 부교수  
2010~2014 KAIST 지식서비스공학과 조교수  
2008~2010 IBM Almaden Research Center 박사후연구원

2006~2008 University of Illinois at Urbana-Champaign 박사후연구원  
2005 KAIST 전산학부 졸업(박사)  
1999 KAIST 전산학부 졸업(석사)  
1997 KAIST 전산학부 졸업(학사)  
관심분야: 빅데이터, 인공지능, 데이터 마이닝  
Email: jaegil@kaist.ac.kr



### 박찬영

2020~현재 KAIST 산업및시스템공학과 조교수  
2019~2020 University of Illinois at Urbana-Champaign 박사후연구원  
2014~2019 포항공과대학교 졸업(박사)  
2014 서강대학교 컴퓨터공학과 졸업(학사)  
관심분야: 그래프 기계학습, 추천 시스템, 인공지능,

데이터마이닝

Email: cy.park@kaist.ac.kr



### 신기정

2019~현재 KAIST 김재철AI대학원 조교수  
2019 Carnegie Mellon Univ. Computer Science 학과 졸업(박사)  
2015 서울대학교 컴퓨터공학부 졸업(학사)  
관심분야: 빅데이터, 인공지능, 그래프 알고리즘, 데이터마이닝

Email: kijungs@kaist.ac.kr



### 황의종

2020~현재 KAIST 전기및전자공학부 부교수  
2018~2020 KAIST 전기및전자공학부 조교수  
2012~2018 Google Research 연구원  
2012 Stanford Univ. Computer Science 학과 박사후연구원  
2012 Stanford Univ. Computer Science 학과 졸업(박사)

2007 Stanford Univ. Computer Science 학과 졸업(석사)

2003 KAIST 전산학부 졸업(학사)

관심분야: 빅데이터, 인공지능, 데이터 중심 인공지능, 신뢰 가능한 인공지능

Email: swhang@kaist.ac.kr



### 최예지

2020~현재 (주)에스아이에이 책임연구원  
2018~2020 한국과학기술정보연구원 박사후연구원  
2012~2018 연세대학교 지구환경연구소 연구원  
2018 연세대학교 대기과학과 졸업(박사)  
2007 연세대학교 대기과학과 졸업(학사)  
관심분야: 위성데이터, 기상, 인공지능, 딥러닝

Email: yejichoi@si-analytics.ai