ChatGPT

데이터 전처리 기법

- **데이터 수집:** 효율적 데이터 수집을 위해 **능동 학습(Active Learning)**이 활용된다. 모델이 가장 정보량이 많은 미라벨(unlabeled) 샘플을 선택해 레이블링을 요청함으로써, 최소한의 데이터로도 높은 성능을 달성할 수 있다 [1]. 또한 **약한 감독 학습(Weak Supervision)** 기법인 Snorkel 등이 도입되어 규칙 기반 레이블링으로 대량의 훈련데이터를 신속히 확보한다. 한편, **대규모 약지도 사전학습** 연구에서는 소셜 미디어의 해시태그 등을 레이블로 활용하여 **수십 억 이미지** 규모로 모델을 학습시켰고, ImageNet 분류 정확도 Top-1 85.4%의 최고 성능을 보고하였다 [2] [3]. 이러한 웹 크롤링/크라우드소싱 기반 데이터 수집은 최소한의 수작업으로 방대한 데이터를 확보해 모델 성능을 향상시켰다.

- **데이터 정제: 노이즈가 있는 레이블**을 다루기 위한 기법들이 제안되었다. 예를 들어 **Co-teaching**은 두 개의 신경망을 동시에 학습시키면서 서로 **신뢰할 수 있는 데이터**를 교환하여 훈련하도록 함으로써, 레이블 오류가 많아도 견고한 모델을 얻는 새로운 학습 패러다임이다 [4]. 실제로 Co-teaching은 MNIST, CIFAR-10/100의 **노이즈 레이블** 실험에서 기존 기법 대비 우수한 성능을 보였다. 또한 **Confident Learning** 기법은 데이터의 **레이블 품질**에 집중하여, 확률 모형을 통해 데이터셋 내 숨은 **레이블 오류를 식별 및 제거**하고 신뢰도 높은 샘플로 학습함으로써 모델 정확도를 높였다 [5]. 예를 들어, Confident Learning을 ImageNet에 적용하여 일부 잘못 레이블된 이미지를 찾아내 정정하고, 해당 정제된 데이터로 ResNet을 재훈련하여 성능 향상을 이끌어냈다 [6].

- **데이터 증강:** 제한된 데이터를 늘리기 위해 다양한 **데이터 증강** 기법이 활용된다. **mixup** 기법(Zhang 등, 2018)은 임의의 두 샘플과 라벨을 선형 혼합하여 학습에 사용함으로써 모델의 **일반화 성능을 향상**시키고, 잘못된 라벨이나 적대적 공격에 대한 **강건성을 높였다** [7]. 또한 구글에서 제안한 **AutoAugment**(Cubuk 등, 2019)은 강화학습으로 최적의 증강 정책을 자동 탐색하여 **분류 정확도를 향상**시켰다 [8]. AutoAugment로 찾은 증강 정책은 CIFAR-10 등에서 기존 대비 오류율을 크게 낮추고, ImageNet에서도 최고 성능을 달성하였다. 더 나아가, **생성 모델**을 활용한 증강도 시도되었다. 예를 들어 **DAGAN**(Antoniou 등, 2018)은 **GAN 기반으로 새로운 이미지 샘플을 생성**하여 소량의 얼굴이미지로 학습할 때 분류기의 성능을 향상시켰음을 보였다 [9]. 이처럼 데이터 증강 기법들은 데이터 다양성을 높여 과적합을 막고 모델의 **범용 성능**을 높이는 데 핵심적인 역할을 한다.

AI 모델 개발

- **아키텍처 설계:** 모델 구조 설계 분야에서는 **신경망 아키텍처 검색(NAS)** 기법이 주목받았다. 예를 들어 Zoph 등(2018)은 강화학습으로 합성곱 셀 구조를 검색하여 NASNet을 개발했는데, 이는 **ImageNet 분류 정확도 Top-1 82.7%**로 당시 최고 성능을 인간이 설계한 모델 대비 1.2%p 높이면서 **계산량을 28% 절감**시켰다 [10]. 또한 구글의 **EfficientNet**(2019)은 NAS로 얻은 기반 모델을 너비·깊이·해상도를 균형 있게 스케일업하는 **Compound Scaling**을 제안하여, **매우 적은 파라미터로 최고 정확도**를 달성했다 [11]. EfficientNet-B7 모델은 ImageNet Top-1 정확도 84.3%로 당시 SOTA를 기록하면서도 이전 최첨단 ConvNet보다 **모델 크기는 8.4배 작고 추론은 6.1배 빠른** 효율성을 보였다 [11]. 이처럼 NAS와 모델 스케일링 기법을 통한 아키텍처 설계는 모델의 성능과 효율을 동시에 크게 향상시켰다.

- **설명 가능한 AI(XAI):** 복잡한 AI 모델의 **투명성**을 높이기 위한 다양한 XAI 기법이 등장하였다. **LIME**(2016)은 모델 예측에 대한 국소적인 선형 설명자를 학습시켜 개별 예측의 근거를 사람 친화적으로 설명하는 기법으로, 예를 들어 환자 진단 모델의 "독감" 예측 시 주요 증상 특징을 하이라이트하여 **의사가 모델을 신뢰하는 데 도움**을 준다 [12]. **SHAP**(2017)은 게임이론의 **셰플리 값**을 활용하여 각 특징이 예측에 기여하는 중요도를 산출하는 방법으로, 일관성과 추가적 특성 보장 등의 이론적 우수를 갖는다 [13]. SHAP은 모델에 관계없이 적용 가능하며 특정 예측에 대한 **특성 기여도를 직교적으로 설명**해준다. **Grad-CAM**(2017)은 **CNN의 출력에 대한 그래디언트**를 이용해 입력 이미지의 중요한 영역을 시각화하는 기법으로, 분류나 VQA 등의 **딥러닝 결정 근거**를 열지도

형태로 나타낸다 [14]. Grad-CAM은 별도 구조 수정 없이 다양한 CNN 기반 모델에 적용되어, 예를 들어 자율주행 영상에서 모델이 주목한 물체나 영역을 하이라이트함으로써 **모델 판단에 대한 직관적 해석**을 제공한다 [15]. 이러한 XAI 기법들은 AI 모델의 **신뢰성과 이해도**를 높여주어 산업 현장에서 중요한 역할을 한다.

- **AI 모델 학습 및 평가: 학습 곡선** 분석은 모델의 학습 상태를 진단하는 핵심 도구이다. 에포크 진행에 따른 훈련/검증 셋의 오차나 정확도 추이를 그린 **러닝 커브**를 보면, **훈련 정확도는 높지만 검증 정확도가 현저히 낮아지는 경우 과적합**을 의심할 수 있고, 반대로 **훈련·검증 정확도 모두 낮고 개선되지 않으면 모델 용량 부족에 따른 미적합**으로 판단한다 [16]. 이러한 지표를 통해 데이터량 증강, 모델 복잡도 조절, 정규화 등 대응책을 세운다. 한편 **평가 지표** 선택도 중요하다. **정확도(Accuracy)**는 클래스 불균형 상황에서 오해를 낳을 수 있어 보완 지표가 필요하다 [17]. 예를 들어 전체 샘플 중 95%가 부정 클래스인 데이터셋에서 모든 예측을 부정으로 하면 정확도 95%가 나오지만 이는 무의미하다 [17]. 따라서 **정밀도-재현율(Precision-Recall)** 및 **F1-스코어** 등이 분류 성능을 더 잘 측정하며, **ROC-AUC**나 **평균 정밀도(mAP)** 등 과제별 지표도 활용된다. 특히 불균형 데이터셋의 경우 **Balanced Accuracy**(클래스별 정확도의 평균) 등이 권장되며, 모델의 다양한 성능 측면을 고려해 종합 평가한다 [17].

- **AI 모델 튜닝 기법: 하이퍼파라미터 최적화(HPO)** 분야에서는 Grid Search보다 **Random Search**가 효율적임이 밝혀졌다. Bergstra&Bengio(2012)는 이론 및 실험적으로, 제한된 예산 하에서 무작위 샘플링이 **격자 탐색 대비 더 우수한 하이퍼파라미터 조합**을 찾음을 보였다 [18]. 나아가 Bayesian Optimization 기법(Snoek 등, 2012)은 **가우시안 프로세스**로 하이퍼파라미터-성능 관계를 모델링하여 자동 튜닝을 수행, 다수 작업에서 **인간 전문가 수준 또는 그 이상의 최적화 성능**을 달성했다 [19]. 한편 **클래스 불균형** 문제를 완화하기 위한 기법으로 데이터 측면과 알고리즘 측면 접근이 모두 활용된다. **SMOTE**(Chawla 등, 2002)와 같은 오버샘플링 기법은 소수 클래스 샘플들을 주변 공간에 **새로운 합성 샘플**로 생성하여 학습 데이터를 증가시킴으로써 분류기의 민감도를 높인다 [20]. 알고리즘 측면에서는 **비용 민감 학습**이나 **가중치 조정**을 통해 소수 클래스 오류에 큰 패널티를 부여하거나, **Focal Loss**(Lin 등, 2017)처럼 **손실 함수 자체를 수정**하는 방법이 쓰인다 [21]. Focal Loss는 검출 등에서 등장하는 **극심한 클래스 불균형**을 다루기 위해 쉽거나 명확히 분류된 사례의 손실을 다운웨이트하여 어려운 사례 학습에 집중하도록 함으로써, one-stage 객체 검출기의 성능을 기존 대비 크게 향상시켰다 [21].

AI 시스템 구축

- **ML 파이프라인 설계 및 배포: 머신러닝 파이프라인**은 데이터 준비부터 모델 배포까지의 과정을 자동화하여 일관성 있게 재현하는 데 핵심적이다. 예를 들어 구글의 TFX는 **일련의 파이프라인 컴포넌트**로 ML 시스템을 구현하며, 대용량 데이터 처리와 모델 학습/검증/배포 단계들을 모듈화하여 **고확장성 고성능 ML 파이프라인** 구축을 지원한다 [22]. 이러한 파이프라인은 데이터 수집-전처리-특성추출-훈련-검증-배포로 이어지는 작업 흐름을 코드로 구현하고, CI/CD와 연계하여 **모델 업데이터의 지속적 통합·배포**(Continuous Deployment)를 가능케 한다. **컨테이너화** 기술(Docker 등)을 활용해 모델을 마이크로서비스로 배포하고, **쿠버네티스** 기반으로 스케일 아웃하여 대규모 트래픽에도 대응한다. 또한 MLflow, Kubeflow와 같은 MLOps 프레임워크를 쓰면 **실험 추적, 모델 저장소, 배포**까지 통합 관리가 가능하여 전체 ML 수명주기를 효율화한다 [23]. 예를 들어 MLflow는 실험별 파라미터와 성능을 관리하고 모델을 등록한 뒤, REST API로 서빙하는 등 **모델 실험부터 서비스까지의 반복 과정을 체계화**해준다 [23].

- **AI 시스템 모니터링 및 자동화:** 모델을 배포한 이후에는 **모델 성능 모니터링**과 자동화된 유지보수가 중요하다. **ML 모델 모니터링**이란 프로덕션 환경에서 **지속적으로 모델 품질 지표를 추적**하고 이상을 탐지하는 것으로, 소프트웨어 일반 모니터링과 달리 예측 **정확도나 오류율**, **데이터 분포의 변화(drift)** 등을 중점적으로 관찰한다 [24]. 예를 들어 일정 기간 동안 입력 데이터의 분포가 훈련 시와 크게 달라지면 **데이터 드리프트** 경고를 발생하고, 모델 예측 출력의 클래스 분포 변화나 정확도 저하가 감지되면 **개념 드리프트**를 의심하여 관리자에게 알린다. 이를 위해 **대시보드**를 통해 실시간 지표를 시각화하고, 임곗값 기반 **알림 시스템**을 구축하기도 한다. **자동화** 측면에서는, 모니터링 결과 이상 징후 시 **자동 재학습 파이프라인**을 트리거하여 최신 데이터로 모델을 주기적으로 업데이트하거나, 성능 저하 시 이전 모델로 **자동 롤백**하는 절차 등을 마련한다. 예컨대 데이터 드리프트가 임계치를 넘으면 새로운 데이터로 모델을 재훈련·배포하는 등 **MLOps 자동화**를 구현하여 인적 개입 없이 모델을

지속적으로 개선할 수 있다. 이처럼 모니터링 및 자동화 기법은 모델의 **신뢰성 유지**와 **운영 비용 절감**에 필수적이다.

- **AI 시스템 최적화:** 대규모 AI 모델의 실서비스 적용을 위해 **시스템 최적화**가 적극 연구되고 있다. 하나의 방향은 모델 경량화로, **모델 압축** 기법들이 대표적이다. **딥 컴프레션(Deep Compression)** 방법은 불필요한 가중치를 제거하는 **모델 가지치기(pruning)**와 가중치 값을 클러스터링하여 비트수를 줄이는 **양자화(quantization)**를 순차적으로 적용하고, 마지막으로 허프만 코딩으로 저장공간을 최적화한다 [25]. Han 등(2016)의 딥 컴프레션은 **알렉스넷 모델을 240MB에서 6.9MB로 35배 축소**하고도 정확도 저하가 없음을 보였고, VGG-16도 49배 압축(552MB→11.3MB)하여 **동일 정확도**를 유지했다 [26]. 이렇게 축소된 모델은 메모리 내 적재가 수월해져 **추론 속도가 3~4배 향상**되고 에너지 효율도 크게 높아졌다. 또 다른 최적화 방향은 **지식 증류(Knowledge Distillation)**이다. Hinton 등이 제안한 증류 방법은 큰 **교사 모델**의 출력 분포(soft logits)를 이용해 **작은 학생 모델**을 훈련시킴으로써, 작은 모델이 큰 모델의 지식을 배우도록 한다 [27] [28]. 이를 통해 파라미터 수가 훨씬 적은 경량 모델도 성능 저하를 최소화하며 원래 모델과 유사한 예측 능력을 갖출 수 있다. 지식 증류는 복잡한 **앙상블 모델을 단일 모델로 압축**하거나, 거대한 언어모델을 모바일 환경에서 동작 가능한 크기로 줄이는 등에 활용되고 있다. 이 밖에도 하드웨어 최적화 측면에서 **GPU/TPU 가속**, **연산 그래프 최적화**(XLA 등), **배치닝 및 파이프라이닝**으로 **고처리량 실시간 추론**을 구현하는 노력이 병행되고 있다. 이러한 시스템 최적화 기법들은 AI 모델의 **서비스 응답 지연을 줄이고** 자원 소모를 절감하여, 대규모 AI 서비스를 현실화하는 기반이 되고 있다.

주요 AI 기술 트렌드 (2018년 이후)

- **Transformer 혁신과 거대 언어모델:** 2017년 트랜스포머(Transformer) 구조 등장 이후, **자연어 처리(NLP)** 분야는 트랜스포머 기반 사전학습 모델로 급격히 발전했다. 2018년 제안된 **BERT**(Devlin 등)는 **양방향 트랜스포머 언어모델**을 미리 학습한 뒤 다운스트림 태스크에 파인튜닝하는 방법으로, 질문응답, 추론 등 **11개 NLP 태스크에서 기존 대비 대폭 향상된 SOTA 성능**을 달성했다 [29]. 이후 GPT 계열로 대표되는 **대규모 언어모델** 시대가 열려, OpenAI의 **GPT-3**(2020)은 **1750억 파라미터**의 초거대 모델로서 별도 파인튜닝 없이 **Few-shot 학습**만으로도 번역, 질의응답, 산술 등 다수 과제를 인간에 준하는 성능으로 수행해 주목받았다 [30]. 이는 대용량 모델이 **범용 인지 능력**을 획득할 가능성을 보여주어, **ChatGPT**와 같은 고도화된 대화형 AI의 등장을 촉진했다.

- **차세대 비전 모델: 컴퓨터 비전** 분야에서도 2018년 이후 새로운 모델들이 속속 등장했다. **ResNet**(2015)으로 대표되는 매우 깊한 CNN의 성공 이후, 2019년 **EfficientNet**이 NAS+스케일링으로 **파라미터 효율을 극대화한 모델**로 ImageNet 분류 SOTA를 기록했고 [11], 2020년에는 **Vision Transformer(ViT)**가 제안되어 **CNN 없이** 순수 트랜스포머로도 시각 인식을 성공할 수 있음을 보였다 [31]. ViT는 입력 이미지를 패치로 쪼개 순차열로 트랜스포머에 투입하는 방식으로, **대규모 데이터 사전학습 후 ImageNet 등에서 최고 수준 정확도**를 달성했다 [32]. 이는 **"ResNet 이후"** 비전 아키텍처의 새로운 패러다임으로 평가받으며, 이후 **Swin Transformer** 등 지역적 자기어텐션을 도입한 모델들이 검출/분할에서 우수한 성능을 보였다. 한편 CNN 계열에서도 **ResNeXt**(2017, 카드inality 개념 도입), **DenseNet**(2017, 밀집 연결) 등의 구조 개선이 잇따랐고, **RegNet**(2020)처럼 아키텍처 패밀리를 규칙적으로 설계하는 시도도 나왔다. 이처럼 비전 모델은 **트랜스포머 도입**과 **CNN 구조 개선** 두 방향으로 발전하며, 다양한 시각 과제에서 성능 향상을 이어가고 있다.

- **초거대 생성모델과 확산 모델: 생성적 적대네트워크(GAN)** 열풍 이후, 최근에는 **확산 모델(Diffusion model)**이 등장하여 이미지 생성 분야의 새로운 트렌드가 되었다. 2020년 Ho 등의 디퓨전 모델은 **노이즈를 단계적으로 제거**하는 과정을 통해 이미지를 생성하는 방식으로 제안되었고, 2021년 **Dhariwal & Nichol** 연구에서는 확산 모델이 ImageNet 기반 **이미지 생성 품질(FID)**에서 GAN 등 기존 기법을 뛰어넘음이 보고되었다 [33]. 특히 **classifier guidance**나 **업샘플링 확산모델** 등의 기법을 통해 확산 모델은 다양한 해상도에서 최고 성능을 얻어, 결과적으로 **Diffusion 모델이 GAN을 제치고 SOTA 이미지 생성**을 달성했다 [33]. 이러한 기술 기반으로 OpenAI의 **DALL·E 2**, Stability AI의 **Stable Diffusion**, 구글의 **Imagen** 등 **텍스트-투-이미지** 생성 모델들이 잇달아 공개되어, 단순 문장 입력만으로 고화질의 이미지를 생성하는 획기적인 데모를 선보였다. 요약하면, **트랜스포머 기반 모델의 범용화**, **초대규모 파라미터의 활용**, **멀티모달 및 생성 모델의 발전**이 2018년 이후 AI 분야의 핵심 트렌드로 자리잡고 있으며, 이는 곧 산업 전반의 **AI 혁신**을 가속화하고 있다.

1 A Survey on Deep Active Learning: Recent Advances and New Frontiers
https://arxiv.org/html/2405.00334v2

2 3 Exploring the Limits of Supervised Pretraining
https://openaccess.thecvf.com/content_ECCV_2018/papers/Dhruv_Mahajan_Exploring_the_Limits_ECCV_2018_paper.pdf

4 [1804.06872] Co-teaching: Robust Training of Deep Neural Networks with Extremely Noisy Labels
https://arxiv.org/abs/1804.06872

5 6 [1911.00068] Confident Learning: Estimating Uncertainty in Dataset Labels
https://arxiv.org/abs/1911.00068

7 [1710.09412] mixup: Beyond Empirical Risk Minimization
https://arxiv.org/abs/1710.09412

8 A Comprehensive Survey on Data Augmentation
https://arxiv.org/html/2405.09591v3

9 Data Augmentation Generative Adversarial Networks - OpenReview
https://openreview.net/forum?id=S1Auv-WRZ

10 Learning Transferable Architectures for Scalable Image Recognition
https://openaccess.thecvf.com/content_cvpr_2018/papers/Zoph_Learning_Transferable_Architectures_CVPR_2018_paper.pdf

11 [1905.11946] EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks
https://arxiv.org/abs/1905.11946

12 Local Interpretable Model-Agnostic Explanations (LIME): An Introduction – O'Reilly
https://www.oreilly.com/content/introduction-to-local-interpretable-model-agnostic-explanations-lime/

13 [1705.07874] A Unified Approach to Interpreting Model Predictions
https://arxiv.org/abs/1705.07874

14 15 Grad-CAM: Visual Explanations From Deep Networks via Gradient-Based Localization
https://openaccess.thecvf.com/content_ICCV_2017/papers/Selvaraju_Grad-CAM_Visual_Explanations_ICCV_2017_paper.pdf

16 Learning Curve To Identify Overfit & Underfit - GeeksforGeeks
https://www.geeksforgeeks.org/machine-learning/learning-curve-to-identify-overfit-underfit/

17 Precision and recall - Wikipedia
https://en.wikipedia.org/wiki/Precision_and_recall

18 Random Search for Hyper-Parameter Optimization
https://jmlr.org/papers/v13/bergstra12a.html

19 Practical Bayesian Optimization of Machine Learning Algorithms
https://proceedings.neurips.cc/paper/2012/file/05311655a15b75fab86956663e1819cd-Paper.pdf

20 [1106.1813] SMOTE: Synthetic Minority Over-sampling Technique
https://arxiv.org/abs/1106.1813

21 ICCV 2017 Open Access Repository
https://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

22 Architecture for MLOps using TensorFlow Extended, Vertex AI ...
https://cloud.google.com/architecture/architecture-for-mlops-using-tfx-kubeflow-pipelines-and-cloud-build

23 Open Source MLOps: Platforms, Frameworks and Tools - Neptune.ai
https://neptune.ai/blog/best-open-source-mlops-tools

24 Model monitoring for ML in production: a comprehensive guide
https://www.evidentlyai.com/ml-in-production/model-monitoring

25 26 [1510.00149] Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding
https://arxiv.org/abs/1510.00149

27 28 Knowledge Distillation: Principles, Algorithms, Applications
https://neptune.ai/blog/knowledge-distillation

29 [1810.04805] BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding
https://arxiv.org/abs/1810.04805

30 [2005.14165] Language Models are Few-Shot Learners
https://arxiv.org/abs/2005.14165

31 32 [2010.11929] An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale
https://arxiv.org/abs/2010.11929

33 [2105.05233] Diffusion Models Beat GANs on Image Synthesis
https://arxiv.org/abs/2105.05233