자율주행 데이터를 위한 연합학습 기반 불확실성 샘플링 벤치마킹 연구

정혁준, 조찬혁, 최지헌, 오상윤*
아주대학교
jpo7173@ajou.ac.kr, mojoid0913@ajou.ac.kr, unidev@ajou.ac.kr, *syoh@ajou.ac.kr

Benchmarking Uncertainty Sampling for Autonomous Driving Data in Federated Learning

Hyeokjun Jeong, Chanhyeok Jo, Jiheon Choi, Sangyoon Oh*
Ajou University

요 약

본 연구는 자율주행 시스템 적용을 위한 기초 연구로서, 연합학습 환경에서 다양한 불확실성 기반 능동 학습 방법들의 성능을 비교 분석하였다. 일반적인 이미지 분류 작업에서 불확실성 샘플링 (margin, mc-dropout, max-entropy) 방법의 효과를 검증함으로써 자율주행 도메인 적용 가능성을 확인하였다. MNIST, FashionMNIST, CIFAR-10 데이터셋에서 각 샘플링 방법에 따른 정확도 변화를 측정하였다. 실험 결과, 불확실성 기반의 샘플링이 무작위 샘플링 대비 높은 정확도를 보였다. 특히 마진 기반의 샘플링이 우수한 성능을 보였다.

Ⅰ. 서 론

인공지능 (Artificial Intelligence) 기술은 자율주행 시스템 (Autonomous Driving System)의 핵심 기술로, 주행 안정성을 강화하거나, 전기차(BEV: Battery Electric Vehicle)의 배터리 에너지 효율 최적화 등에 적극적으로 활용되고 있다. 인공지능 기반의 정확도 높은 예측 모델을 개발하기 위해서는 주행 조건과 환경을 포함하는 다양한 데이터가 반드시 필요하다. 그러나, 차량의 주행 데이터는 각 제조사의 핵심 자산이자 민감한 정보이기 때문에, 각 자동차 제조사 (vendor)는 자사 중심의 독립적인 데이터 사일로(silo) 형태로 보관 및 운영되고 있다. 따라서, 높은 다양성과 더 많은 데이터를 기반으로 높은 정확도의 모델을 개발하기 위한 제조사 간 데이터 공유 혹은 중앙 서버로의 통합하는 것은 현실적으로 많은 어려움이 존재한다.

이러한 데이터 통합의 문제에 대해 연합학습(federated learning) 기법이 해결책이 될 수 있다. 연합학습은 각 제조사가 자사의 데이터를 직접 공유하지 않고 각 제조사, 즉 제조사의 데이터 사일로 별로 독립적 모델 학습을 수행하고, 생성된 모델 가중치(weight) 혹은 그레디언트(gradient)만을 수집하여 전역 모델 (global model)을 생성하는 분산 학습 방법이다.

이 기법을 통해 데이터 공유가 어려운 상황에서도 모든 데이터를 한곳에 모아 학습한 것과 유사한 수준의 고성능 모델을 얻을 수 있다. 하지만, 이 기법을 제조사 사일로 기반으로 적용하기 위해서는 다음의 문제를 해결하는 것이 필요하다. 현재까지 제안된 연합학습 기법들 대부분은 참여자 (silo)가 각자 학습을 수행하는 데이터가 모두 라벨링(labeling)되어 있다는 전제를 가진다. 그러나, 자율주행을 위한 이미지 데이터에 객체를 표시하거나, 주행 데이터의 정확한 에너지 소비량을 태깅(tagging)하는 등의 라벨링 작업은 많은 시간과 비용을 요구하면서 동시에 각 제조사 기술에 대한 전문적 지식이 필수적으로 필요하며, 데이터 자체에 대한 공유가 어려운 것처럼 이에 대한 공유도 어렵다. 따라서 라벨이 없거나, 상이한 데이터가 대부분인 실제 자율주행 데이터 환경에서 연합학습의 실효성은 매우 떨어지게 된다.

이를 위해 다양한 라벨링 기법들이 제시되고 있으며, 이들은 각 상황에 따라 상이한 성능을 보여 조건에 따른 최적의 라벨링 기법을 선택하는 것이 매우 중요하다. 본 연구는 연합학습 데이터의 효과적인 라벨링을 위한 다양한 불확실성(uncertainty) 기반의 능동학습(active learning) 방법들의 기본 성능을 일반적인 이미지 분류 작업을 통해 벤치마크하고, 자율주행 도메인에서 요구되는 높은 안정성을 만족시키는지 여부를 판단하기 위해 필요한 각 기법들에 대한 기초적인 성능을 확인하는 것을 목표로 한다.

Ⅱ. 연구 배경 및 관련 연구

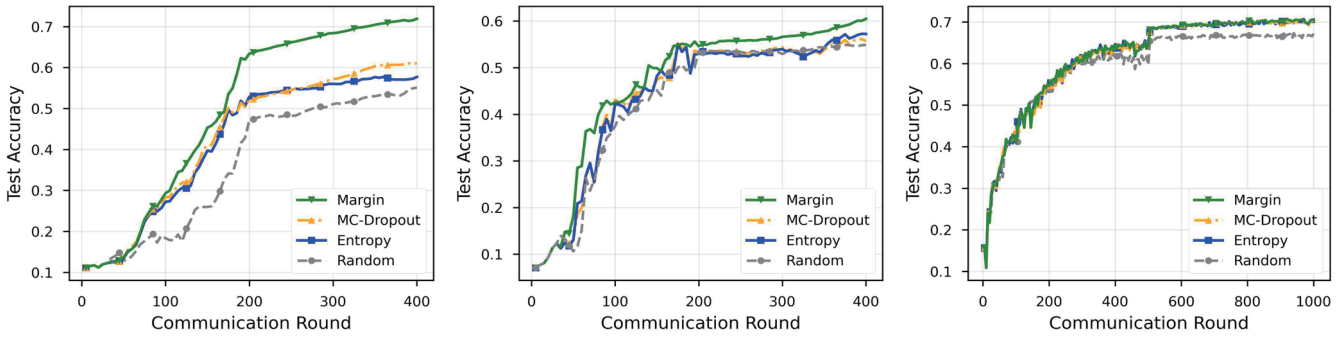
능동 학습의 핵심 목표는 훈련 과정에 가장 유용한 데이터를 선택적으로 샘플링하여, 최소한의 라벨링으로 높은 정확도를 달성하는 것이다. 능동 학습 방법은 크게 불확실성 기반(uncertainty-based)과 표현 기반(representation-based) 샘플링으로 구분된다 [1]. 이 중, 불확실성 기반의 샘플링은 모델이 가장 확신하지 못하는 샘플을 우선적으로 선택하는 방식이며, 불확실성을 정량화하기 위해 Monte-Carlo Dropout (MC-Dropout)[2], Max-Entropy[3], Margin-Sampling[4] 등의 기법이 주로 활용된다.

연합학습 환경에서 라벨링된 데이터가 부족한 경우, 어떤 데이터를 우선적으로 라벨링할 지 결정하는 연구가 활발히 진행되고 있다. 특히 의료 분야에서는 각 병원 간의 환자 데이터 공유 없이 피부 병변 이미지를 분류하기 위해 능동 학습을 적용한 사례[5]가 있다. 해당 연구는 중앙 서버의 전역 모델을 앙상블하여 엔트로피(ensemble-entropy) 값을 기준으로 라벨링 대상을 선택했다. 또한, 불필요한 라벨링 비용을 줄이기 위한 연구[6]도 제안되었다. 이 연구는 전역 모델과 로컬 모델 간의 예측 일치성을 기준으로 샘플을 선택한다. 즉, 예측이 일치하는 샘플은 신뢰도가 높다고 판단하여 의사 라벨(pseudo-label)을 부여하고, 예측이 불일치하는 샘플은 예측 신뢰도가 낮다고 판단하여 전문가에게 라벨링을 요청하게 된다.

이와 같이 연합학습 환경에 능동 학습을 도입하는 연구들의 핵심은, 전역 모델의 성능 향상에 가장 크게 기여할 샘플을 선별하는 전략을 수립하는 것이다. 따라서, 본 연구는 기존 대표적인 불확실성 기반 샘플링 방법들을 연합학습에 적용하여, 라벨링된 데이터가 부족한 로컬 클라이언트가 전역 모델의 정확도에 유의미한 기여를 할 수 있는지 검증하고자 한다.

Method	MNIST				FashionMNIST				CIFAR-10			
	20%	40%	60%	80%	20%	55%	70%	85%	20%	40%	60%	80%
Margin	11.10	16.87	**58.27**	**68.52**	7.07	**36.34**	**54.99**	**56.69**	44.63	**61.45**	63.03	69.57
MC-Dropout	11.10	**17.32**	48.95	57.32	7.07	26.22	54.86	53.78	44.53	60.75	64.64	69.25
Max-Entropy	11.10	16.73	48.87	55.98	7.07	26.82	54.36	53.53	45.54	61.04	**64.68**	**69.89**
Random	11.10	15.25	37.66	51.29	7.07	26.38	49.41	53.32	**46.59**	60.39	60.94	66.02

표 1 데이터셋별 질의(query) 비율에 따른 정확도 변화



(a) MNIST (b) FashionMNIST (c) CIFAR-10

그림 1 샘플링 횟수 및 연합학습 라운드에 따른 정확도 변화

Ⅲ. 실험 및 평가

본 연구의 실험은 NVIDIA GeForce RTX 2080 (VRAM 8GB), Intel(R) Core(TM) i7-8700와 64GB DDR4 메모리 기반의 환경에서 수행되었으며, CUDA 12.8, PyTorch 2.8.0 을 사용하였다. 연합학습 환경의 non-IID 특성을 반영하기 위해 Dirichlet alpha 값을 0.3으로 설정하였고, 이 값을 각 클라이언트별 데이터 분배에 활용하였다. 주요 하이퍼 파라미터 설정은 표 2에 정리하였다.

	MNIST	FashionMNIST	CIFAR-10
Learning Rate	0.0005	0.001	0.0012
Batch Size	32	32	16
FL Rounds	400	400	1000
Query Interval	30	30	10
Query Budget	5	20	5

표 2 주요 하이퍼파라미터 설정

그림 1은 데이터셋별로 라운드 변화에 따라 정확도 변화를 보인다. 그 결과, 모든 데이터셋에서 불확실성 샘플링 방법이 무작위 샘플링 대비 빠른 정확도 향상과 수렴을 보였다. 표 1은 능동학습 질의(query)를 통해 라벨링한 샘플 비율에 따른 정확도를 비교한 것이다. 학습 초반을 제외한 모든 경우에서 불확실성 기반의 샘플링 방법들이 라벨링 비율 대비 높은 정확도를 보였다. 특히 불확실성 기반 기법들 중에서도 MNIST, FashionMNIST 데이터셋에 대해서는 마진 기반 샘플링이, CIFAR-10에 대해서는 불확실성 기반의 샘플링 방법이 모두 유사하게 높은 성능을 보였다.

실험 결과에서 마진 기반의 샘플링은 MNIST와 FashtionMNIST에서 다른 방법들에 비해 일관되게 높은 성능을 보였으나, CIFAR-10에서는 방법들 간의 성능 차이가 상대적으로 줄어들었다. 이는 데이터셋의 복잡도가 MNIST와 FashionMNIST와 같이 클래스 간 경계가 비교적 명확하고 단순한 데이터셋에 대해서는 마진 기반의 방법이 결정 경계 근처의 유용한 샘플을 효과적으로 식별할 수 있다. 그러나 CIFAR-10과 같이 복잡한 데이터셋에서는 모든 불확실성 측정 방법이 비슷한 수준의 한계를 보이는 것으로 해석된다.

Ⅳ. 결론

본 연구에서는 연합학습 환경에서 불확실성 기반의 샘플링이 무작위 샘플링에 비해 더 적은 라벨 수로 높은 정확도를 달성함을 보였다. 이후 연구에서는 자율주행 도메인의 요구사항인 안정성을 고려하면서, 실제 환경의 통계적, 시스템 이질성을 고려한 샘플링 기법의 제안이 필요하다.

ACKNOWLEDGMENT

이 연구는 2025년도 산업자원통상부 및 산업기술평가관리원(KEIT)-자율주행기술개발혁신사업(20018248, 주변 상황 인식 센서 성능 및 판단 기능 부족으로 인한 사고 위험 대응 기술 개발)의 지원을 받아 수행하였음.

참 고 문 헌

[1] Ren, Pengzhen, et al. "A survey of deep active learning." ACM computing surveys (CSUR) 54.9 (2021): 1-40.

[2] Gal, Yarin, and Zoubin Ghahramani. "Dropout as a bayesian approximation: Representing model uncertainty in deep learning." international conference on machine learning. PMLR, 2016.

[3] Nguyen, Vu-Linh, Mohammad Hossein Shaker, and Eyke Hüllermeier. "How to measure uncertainty in uncertainty sampling for active learning." Machine Learning 111.1 (2022): 89-122.

[4] Balcan, Maria-Florina, Andrei Broder, and Tong Zhang. "Margin based active learning." International Conference on Computational Learning Theory. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[5] Sener, Ozan, and Silvio Savarese. "Active learning for convolutional neural networks: A core-set approach." arXiv preprint arXiv:1708.00489 (2017).

[6] Deng, Zhipeng, et al. "Fedal: An federated active learning framework for efficient labeling in skin lesion analysis." 2022 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, 2022.

[7] Zhang, Enzhi, and Liu Yang. "Unnecessary Budget Reduction in Federated Active Learning." 2024 IEEE 36th International Conference on Tools with Artificial Intelligence (ICTAI). IEEE, 2024.