

데이터셋 적응적 옵티마이저 시스템

정민영

안산대학교

myjung@ansan.ac.kr

Dataset-Aware Adaptive Optimizer

Minyoung Jung

Ansan Univ.

요약

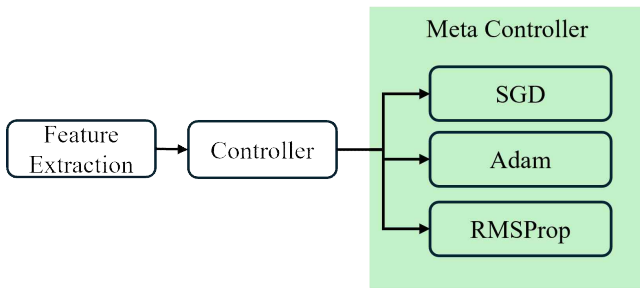
본 논문에서는 데이터셋의 통계적 특성을 인식하여 최적화 가중치를 동적으로 조정하는 데이터셋 인식형 적응적 옵티마이저 시스템을 제안한다. 기존의 단일 옵티마이저(SGD, Adam, RMSProp 등)는 서로 다른 데이터 분포와 손실 함수에 따라 성능이 크게 달라지는 한계를 가진다. 이를 극복하기 데이터 통계 기반 피쳐 벡터(feature vector)를 입력으로 받아, 학습 과정 중 각 옵티마이저의 기여도를 자동 조절하는 메타컨트롤러(meta-controller)를 설계하였다.

I. 서론

딥러닝 모델의 성능은 네트워크 구조, 학습 데이터의 특성, 그리고 옵티마이저의 선택에 의해 결정된다. 특히 학습 단계에서 사용되는 최적화 함수(optimizer)는 모델의 수렴 속도, 안정성, 그리고 일반화 능력에 직접적인 영향을 미친다. 이와 관련 이미지 분류 데이터셋에 따른 최적화 알고리즘의 대한 성능 비교에 대한 연구가 있으며[1], 본 연구는 입력 데이터의 통계적 특성에 기반하여 최적화 전략을 실시간으로 재구성하는 적응형 최적화 시스템을 제안한다.

II. 본론

제안하는 데이터셋을 특성 인식에 기반한 적응적인 옵티마이저 데이터의



Adaptive α Parameter update

특성을 실시간으로 반영하기 위해 상호 유기적으로 작동하기 위해 그림 1에서 보는 바와 같이 특징 추출기(Feature Extractor), 컨트롤러(Mcontroller), 그리고 파라미터를 업데이트 하는 옵티마이저 앙상블(Optimizer Ensemble)로 구성된다.

Fig. 1. 데이터셋 적응적 시스템 프레임워크

특징 추출기는 매 학습 단계(Step)마다 입력되는 미니 배치 데이터로부터 평균(Mean), 분산(Variance), 그라디언트 노름(Gradient Norm), 엔트로피(Entropy) 등 데이터의 분포와 학습 상태를 대변하는 통계적 지표를 계산한다. 이어서 메타 컨트롤러는 추출된 특징 벡터를 입력으로 받아

신경망 연산을 수행하고, 이를 통해 각 기저 옵티마이저에 부여할 중요도를 나타내는 가중치 벡터를 산출한다. 마지막으로 옵티마이저 앙상블 모듈은 SGD[2], Adam[3], RMSProp[4] 등의 옵티마이저가 제안하는 업데이트 방향을 가중치 벡터를 사용하여 가중 평균(Weighted Average)함으로써 최종 파라미터를 갱신한다. 이 모든 과정은 학습이 진행되는 동안 반복적으로 수행되며, 데이터의 특성 변화에 따라 값이 실시간으로 적응(Adaptation)하여 최적의 학습 경로를 탐색하도록 설계되었다.

$$\alpha = [\alpha_1, \alpha_2, \alpha_3, \alpha_4], \text{ where } \sum_{i=1}^4 \alpha_i = 1 \quad (1)$$

본 프레임워크의 특징 추출기(Feature Extractor)는 입력 데이터로부터 평균, 분산, 그라디언트 노름(Gradient Norm), 엔트로피(Entropy)등의 통계적 지표를 계산하여 데이터의 특성을 인식하는 역할을 수행한다. 이를 바탕으로 메타 컨트롤러는 식 (1)에 정의된 가중치 벡터를 계산하여 각 옵티마이저의 기여도를 결정한다. 옵티마이저 앙상블(Optimizer Ensemble) 모듈은 SGD, Adam, RMSProp의 개별 업데이트 결과를 해당 값으로 가중 평균하여 최종 파라미터를 갱신한다. 이러한 과정은 학습 전반에 걸쳐 매 스텝마다 반복 수행되므로, 데이터 특성의 변화에 따라 최적화 전략이 실시간으로 적응되는 효과를 갖는다. 메타 컨트롤러 M 은 특징 추출기로부터 입력받은 특징 벡터 $f \in \mathbb{R}^n$ 에 대하여, 다음의 Softmax기반 함수를 통해 각 옵티마이저에 적용할 가중치 벡터 α 를 산출한다.

$$\alpha_i = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}}, \text{ where } \mathbf{z} = W_m \mathbf{f} + b_m \quad (2)$$

위 식(2)은 각 기저 옵티마이저(Base Optimizer)의 상대적 신뢰도를 확률 분포 형태로 변환하는 역할을 수행한다. 여기서 K 는 앙상블에 사용된 옵티마이저의 개수(본 연구에서는 4)를

의미하며, W_m 과 b_m 은 메타 컨트롤러 내부 신경망의 학습 파라미터이다. 학습 초기 단계에서 α 는 편향성을 배제하기 위해 균등 분포(Uniform Distribution)인 $[0.25, 0.25, 0.25, 0.25]$ 로 초기화 된다. 이후 학습이 진행됨에 따라 손실 감소율(ΔL)과 그라디언트 분산(Gradient Variance)과 같은 피드백 정보를 기반으로 점진적으로 가중치를 조정하여 최적의 구성을 탐색한다.

초기에는 모든 옵티마이저의 가중치가 균등하지만, 학습이 진행되면서 데이터 특성에 따라 가중치가 재분배된다. 예를 들어 초기에는 Adam과 RMSProp이 높은 비중을 차지하나, 후반부에는 SGD의 비율이 증가하여 일반화 성능을 향상시킨다.

III. 결론

본 연구에서는 제안하는 DAO-E 프레임워크의 성능을 객관적으로 검증하기 위해, 이미지 분류 분야의 표준 벤치마크인 CIFAR-10 데이터셋을 사용하였다. 해당 데이터셋은 10개의 클래스로 구성된 50,000장의 학습 이미지와 10,000장의 테스트 이미지로 이루어져 있다.

실험 모델로는 심층 신경망의 표준 아키텍처인 ResNet-18을 채택하였으며, 모든 실험은 공정한 비교를 위해 동일한 하이퍼파라미터 설정 하에 수행되었다. 배치 크기(Batch size)는 128, 총 학습 에폭(Epoch)은 100으로 고정하였으며, 비교군으로는 SGD, Adam, RMSProp, Lion 옵티마이저를 사용하였다. 실험은 NVIDIA RTX 4090 GPU, CUDA 12.6, 그리고 PyTorch 2.6.0 프레임워크 환경에서 진행되었다.

알고리즘의 성능을 검증하기 위해 CIFAR-10 데이터셋과 ResNet-18 모델을 사용하였다. 비교대상은 SGD, Adam, RMSProp, Lion 네 가지 단일 옵티마이저이다. 정확도가 91.6% 달성하여, 단일 SGD, RMSProp에 비해 우수한 성능을 나타내었다. 옵티마이저의 선택을 모델 설계단계가 아닌 학습 과정 단계에서 가능하다. 이는 최적화 선택 문제를 메타 학습적 관점에서 해결 한 것으로, 다양한 데이터셋과 손실구조에 대한 범용성을 확보 가능하다. 또한 GPU 병렬처리에 적합하게 설계되어, 각 옵티마이저 업데이트를 동시에 계산한 후 가중합으로 결합할 수 있다. 이를 통해 추가 연산비용을 최소화하면서 실시간 적응성을 유지한다. 그리고, Lion[4]과 같은 옵티마이저를 추가하여, 향후에 연구가 필요하다.

참 고 문 헌

- [1]정민영, “이미지 분류에서 최적화 알고리즘의 성능 비교,” JCCI 2025
- [2]H. Robbins and S. Monro, “A Stochastic Approximation Method,” Annals of Mathematical Statistics, pp. 400-407
- [3] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” Proc. International Conference on Learning Representations(ICLR) 2015.
- [4]Ruder, S. “An overview of gradient descent optimization algorithms,”ArXiv, 2016.
- [5]Z. Liu et al., “Lion: Symbolic Discovery of Optimization Algorithms,”ArXiv, 2023.