

ReRAM 뉴로모픽 소자의 안정적 학습을 위한 확률적 학습 기법 연구

왕우진, 권세은, 박은솔, 박미소, 신용구*

고려대학교

{wwj604, senikk1118, espark_82, miso419, *ygshin92}@korea.ac.kr

A Study on Stochastic Learning Techniques for Stable Training of ReRAM Neuromorphic Devices

Woo Jin Wang, Se Eun Kwon, Eun-Sol Park, Mi so Park, and Yong Goo Shin*

Korea University

요약

본 논문에서는 다양한 모델에 ReRAM 기반 소자를 매핑하고, Oxford 102 Flower, Stanford Cars, Food-101과 같은 고해상도 데이터 세트에서 패턴 인식 성능을 분류했다. 또한, 학습 시 발생할 수 있는 불안정성을 해결하기 위해 두 가지 핵심 전략을 제안한다. 첫째, ReRAM 소자의 표현 범위를 확장하고 안정성을 높이기 위해 Z-score 표준화와 중복 조합 방식을 도입한다. 둘째, 기울기 크기에 따라 가중치 업데이트 여부를 확률적으로 결정하는 새로운 업데이트 메커니즘을 제시하여 학습 안정성을 확보하고자 한다. 제안하는 확률적 업데이트 방식은 CIFAR-10 데이터 세트에서 90.52%를 달성하며 가장 우수한 분류 정확도를 보여준다. 또한 다양한 대규모 모델 및 데이터 세트에서도 견고한 성능을 입증한다.

I. 서론

최근 ReRAM 기반 뉴로모픽 소자를 이용한 패턴인식 연구가 활발히 진행되고 있다[5, 6]. 그러나 기존 연구들은 작은 네트워크와 데이터 세트에 국한되어 왔다. 이러한 제한적인 환경에서의 성공에도 불구하고, 실질적인 응용을 위해서는 대규모 네트워크와 데이터 세트에서의 성능 검증이 필수적이다.

본 연구에서는 ReRAM 기반 뉴로모픽 소자의 실질적 응용 가능성을 확대하기 위해, 기존 연구의 한계를 보완하고자 한다. 구체적으로 ReRAM 기반 뉴로모픽 소자를 대규모 신경망 구조에 매핑하고, Oxford 102 Flower [2], Stanford Cars [8], Food-101 [1]과 같은 대규모 데이터 세트를 활용하여 패턴인식 성능을 분석한다. 이를 통해 ReRAM 기반 뉴로모픽 컴퓨팅이 대규모 연산 환경에서도 견고한 성능을 유지할 수 있는지 검증한다. 더 나아가, 대규모 네트워크 구현 및 학습 과정에서 발생할 수 있는 불안정성을 해결하기 위한 ReRAM 기반 소자값 표현 범위 확장과 새로운 확률 기반 업데이트 방식을 제시하고, 이러한 접근 방식이 학습 안정성 확보 및 성능에 미치는 영향을 분석하고 검증한다.

II. 본론

2.1. 소자 값의 안정성과 표현력 향상

ReRAM 소자의 저항값 변화를 통해 신경망의 가중치를 업데이트하는 방식은 효율적인 인공지능 하드웨어 구현에 대한 잠재력이 있지만, 값이 매우 작다는 소자 특성으로 인해 학습 과정에서 어려움이 존재한다. 일반적으로 ReRAM 소자의 값은 매우 작기 때문에 가중치에 바로 사용하면 값의 변동 폭이 작아 효과적인 학습이 어려울 수 있다. 이러한 문제점을 해결하고 ReRAM 기반 신경망의 학습 성능을 향상하기 위해 식 (1)에 나와 있는 Z-score 표준화 방법을 도입한다.

$$z_i = \frac{x_i - \mu}{\sigma} \quad (1)$$

여기서 x_i 는 i 번째 소자의 저항값, μ 는 전체 저항값의 평균, σ 는 전체 저항값의 표준편차를 나타낸다. Z-score 표준화를 통해 소자들은 평균이 0, 표준편차가 1인 분포를 가지게 된다.

이후, 소자에서 시냅스 가중치가 증가하는 특성과 감소하는 특성에서 표현될 수 있는 값들을 활용하여 새로운 값을 생성한다. 구체적으로 임의의 두 표준화된 소자 값 z_p 와 z_d 를 더하여 새로운 값으로 사용했다. 이러한 중복 조합 방식을 통해 원래 소자들이 가질 수 있는 다양한 값을 표현할 수 있도록 한다.

2.2. 확률적 가중치 업데이트 메커니즘

일반적으로 인공신경망은 기울기를 기반으로 가중치를 연속적으로 조정하며 학습한다. 반면, 소자를 이용한 학습에서는 가중치 조정이 연속적이지 않고 이산적으로 이루어진다. 이러한 이산적인 가중치 업데이트 방식은 최적값에 도달하기 어렵다. 따라서 최적값에 효과적으로 수렴하는 것이 중요하다. 위 문제를 완화하고 안정적인 학습을 하기 위해, 본 연구에서는 기울기의 크기에 따라 가중치 업데이트 여부를 확률적으로 결정하는 메커니즘을 제안한다.

$$p_i = \frac{g_i - \min(g)}{\max(g) - \min(g)} \quad (2)$$

식 (2)는 가중치에 대한 기울기를 계산한 후 min-max 정규화를 통해 0에서 1 사이의 범위 값으로 변환하고, 이를 확률로 사용하는 과정을 나타낸다. 여기서 g_i 는 i 번째 가중치에 대한 기울기 값, $\min(g)$ 와 $\max(g)$ 는 각각 전체 기울기 값의 최솟값과 최댓값을 나타낸다. 이렇게 얻어진 확률값 p_i 는 해당 가중치를 업데이트할 확률로 사용한다. 즉, 기울기의 크기가 클수록 업데이트될 확률이 높아지고, 기울기의 크기가 작을수록 업데이트될 확률이 낮아진다. 이는 학습에 큰 영향을 미치는 가중치들은 우선

업데이트 방법	정확도 (%)	
	M/ATA/S	Mxene
이산형	10.00	88.69
p=0.25	90.50	89.88
p=0.5	77.17	90.10
p=0.75	19.59	89.62
cdf	46.69	89.72
min-max 정규화	90.52	90.36

표 1. 업데이트 방법별 성능 비교

적으로 업데이트하고, 작은 변화를 보이는 가중치들의 불필요한 업데이트를 줄여 학습의 안정성을 확보하는 데 기여한다.

2.3. 실험 결과

제안하는 업데이트 방법에 대한 성능을 검증하기 위해, 이전 연구 [5, 6]에서 채택한 모델을 사용하여 CIFAR-10 데이터 세트에서 확률적 업데이트 방법에 따른 간단한 이미지 분류 실험을 수행했다. 모든 학습은 200 에폭, 배치 크기 256으로 동일하게 설정하고, 각 방법의 최고 분류 정확도를 비교했다. 표 1에서 확인할 수 있듯이, 제안된 min-max 정규화 기반 확률적 업데이트를 적용했을 때 M/ATA/S 소자와 Mxene 소자에서 각각 90.52%와 90.36%의 가장 높은 분류 정확도를 달성했다. 이는 이산적인 방법 대비 향상된 결과이며, 동시에 학습 안정성 또한 보여주었다. 이러한 결과는 동일한 딥러닝 네트워크 환경에서 제안된 방법이 향상된 성능을 제공할 수 있음을 보여준다.

그러나 이전 연구 [5, 6]에서 사용된 딥러닝 네트워크는 비교적 단순하며 실제 이미지 분류 작업에서 일반적으로 사용되는 모델의 복잡성을 반영하지 않는다. 따라서 표 1에 나와있는 실험 결과만으로 제안되는 업데이트 방식이 다양한 네트워크 구조에서 효과적으로 작동하는지 평가하기 어렵다. 따라서, 제안하는 방법의 일반적인 적용 가능성을 평가하기 위해, 고해상도 이미지 분류에 사용되는 GoogleNet [7], ResNet34 [3], ResNet50 [3], DenseNet121 [4] 딥러닝 네트워크로 추가 실험을 진행했다. 해당 실험은 대규모 고해상도 데이터 세트 (Oxford 102 Flower, Stanford Cars, Food-101)를 사용했다. 표 2의 결과에서 볼 수 있듯이, 제안된 업데이트 방법은 데이터 세트 관계없이 일관되게 높은 정확도를 보여주며 min-max 정규화 업데이트 방식의 견고성과 학습 안정성을 입증했다.

III. 결론

본 논문에서는 ReRAM 기반 뉴로모픽 시스템의 패턴 인식 성능을 다양한 딥러닝 네트워크와 고해상도 데이터 세트에서 검증하고, 학습 안정성을 위한 새로운 방법을 제안했다. 이를 통해 이산적인 가중치 조절 과정에서 발생할 수 있는 불안정성을 완화하고, 기울기가 큰 가중치 위주로 업데이트를 수행하여 학습의 안정성을 확보했다. 제안된 방법은 다양한 모델과 고해상도 데이터 세트에서도 안정적인 동작과 높은 분류 정확도를 달성했다. 이는 본 연구에서 제안한 기법들이 ReRAM 기반 뉴로모픽 컴퓨팅 시스템에 학습 안정성을 효과적으로 개선하여 뉴로모픽 컴퓨팅의 발전에 기여할 것으로 기대한다.

ACKNOWLEDGMENT

This work was supported by IITP grant funded by MSIT (Grant#.

네트워크	정확도 (%)					
	데이터 세트					
	Oxford 102 Flower		Stanford Cars		Food-101	
	M/AT A/S	Mxene	M/AT A/S	Mxene	M/AT A/S	Mxene
GoogleNet	79.71	80.93	70.68	63.97	77.74	77.68
ResNet-34	84.60	82.64	63.50	59.61	79.22	78.75
ResNet-50	83.74	81.05	77.04	74.16	79.54	78.32
DenseNet-121	83.99	84.84	70.65	73.02	78.19	78.93

표 2. 다양한 모델 및 데이터 세트별 성능 비교

RS-2025-02263277, 50%). This work was supported by ITRC support program supervised by the IITP and funded by MSIT (Grant#: IITP-2025-RS-2023-00258971, 50%)

참 고 문 헌

- [1] L. Bossard, M. Guillaumin and L. Van Gool, "Food-101 - mining discriminative components with random forests." *Computer vision - ECCV 2014: 13th European conference*, 2014
- [2] A. Gurnani, V. Mavani, V. Gajjar and Y. Khandhediya, "Flower categorization using deep convolutional neural networks." *arXiv preprint arXiv:1708.03763*, 2017
- [3] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016
- [4] G. Huang, Z. Liu, L. Van Der Maaten and Kilian Q. Weinberger, "Densely connected convolutional networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017
- [5] J.H. Ju, S. Seo, S. Baek, D. Lee, S. Lee, T. Lee, B. Kim, J.J. Lee, J. Koo, H. Choo, S. Lee and J.H. Park, "Two-Dimensional MXene Synapse for Brain-Inspired Neuromorphic Computing." *Small (Weinheim an der Bergstrasse, Germany)* vol. 17,34, 2021
- [6] J.P. Kim, S.K. Kim, S. Park, S. Kuk, T. Kim, B.H. Kim, S.H. Ahn, Y.H. Cho, Y.J. Jeong, S.Y. Choi and S. Kim, "Dielectric-engineered high-speed, low-power, highly reliable charge trap flash-based synaptic device for neuromorphic computing beyond inference." *Nano Letters* 23,2, 2023, pp:451-461
- [7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015
- [8] K. Valev, A. Schumann, L. Sommer and J. Beyerer, "A systematic evaluation of recent deep learning architectures for fine-grained vehicle classification." *Pattern Recognition and Tracking XXIX*, Vol. 10649, 2018