




졸업작품/논문 제안서

2018 년도 제 1 학기

작품/논문	○작품() ○논문(○) ※ 해당란에 체크
제목	Regularization을 이용한 Resource-constraint CNN 최적화 시스템
GitHub URL	https://github.com/seonghye/graduation
팀원명단	한성혜 (인) (학번: 2015313050) 

2018 년 3 월 23 일

지도교수 : 신 동 군 서명 _____

1. 과제의 필요성

1. 1. Abstract

딥 러닝에서는 시간과 성능에 최적화된 신경망 구조를 설계하는 것이 중요하다. 본 졸업논문에서는 제한된 자원에 최적화된 신경망 구조를 만들어주는 딥 러닝 시스템을 제안하고자 한다.

이 시스템은 신경망의 레이어와 채널 개수를 줄이고 늘리는 것을 반복함으로써, 신경망 구조를 자원 제한 조건(Resource Constraint)에 맞게 최적화하도록 학습하는 MorphNet 시스템을 바탕으로 개발한다. MorphNet에서는 신경망 크기를 줄이는 데는 sparsifying regularizer를, 신경망 크기를 늘리는 데는 uniform multiplicative factor를 사용한다[4].

선행 연구에서는 자원 제한 조건으로 FLOPS, model size를 사용하고 있지만, 본 졸업논문에서는 power usage 등의 요소를 추가할 것이다. 또한, 학습 과정에서 node의 개수를 줄이는 shrinking[4], 학습 없이 weight의 상태를 참조하여 node의 개수를 줄이는 network pruning[2] 등을 사용해서 학습 시간을 개선할 것이다.

1. 2. 서론

딥 러닝에서의 신경망 구조 설계는 과학보다는 창의에 가까운 분야였다. 오랜 기간 전문가들은 직관과 경험에 의존한 trial-and-error 방법으로 신경망 구조를 설계하여 신경망의 정확도와 속도를 높여왔다. 그러나, 최근에는 딥 러닝의 활용도를 높이기 위해 다양한 모델이 개발되고 다양한 사양의 하드웨어에서 동작함에 따라, trial-and-error 방법은 설계 시간이 너무 오래 걸려서 한계를 보이고 있다. 따라서, 자동 신경망 구조 설계의 중요성이 커지고 있다.

자동 신경망 구조 설계에서 가장 중요한 방법 중 하나는 sparsifying regularizer다. 최근에는 weight가 아닌 neuron에 중점을 둔 regularizer들이 많이 사용되고 있다. 이를 사용하는 경우 parameter의 수를 줄일 수는 있지만, 특정 자원의 사용량을 줄이는 데는 사용할 수 없다.

최근 신경망 구조 설계에서는 보조 신경망을 사용한 shrinking 기법이 제안되었다.[5] 이 기법의 경우 학습에 적게는 며칠에서 많게는 몇 달의 시간이 걸리기 때문에 현실적으로 이 기법을 사용하는 것은 어렵다고 할 수 있다.

이러한 최근 동향에서 resource-constraint 상황에서의 빠른 자동 신경망 구조 설계 방법은 큰 데이터 셋으로 확장이 가능하다는 점, 특정 자원을 목표로 최적화 할 수 있다는 점, 특정 자원 사용을 줄이며 성능 향상을 가져올 수 있다는 점의 크게 3가지의 이점을 가져다 줄 수 있다.[4]

2. 선행연구 및 기술현황

2. 1. P. M. Williams. Bayesian regularization and pruning using a laplace prior. Neural computation, 7(1):117–143, 1995.

Sparsity를 유도하여 loss에 대한 parameter의 영향을 줄이는 L1 regularization 방법을 제시하였다.

2. 2. S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In Advances in Neural Information Processing Systems, pages 1135–1143, 2015.

먼저 sparsity를 유도하는 방향으로 학습을 진행한 후, 이러한 새 구조를 이용해 학습을 다시 진행하는 방법을 여러 번 반복하는 multi-step training 방법을 제시하였다.

2. 3. B. Liu, M. Wang, H. Foroosh, M. Tappen, and M. Pensky. Sparse convolutional neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 806–814, 2015

Pruning 기법, Sparsifying DNN을 사용하여 메모리 사용량을 줄이는 방법을 제시하였다.

2. 4. Gordon, A., Eban, E., Nachum, O., Chen, B., Yang, T. J., & Choi, E, MorphNet: Fast & Simple Resource-Constrained Structure Learning of Deep Networks. 2017.

FLOPS, model size의 resource-constraint 상황에서 최적화된 신경망 학습 모델을 만드는 방법을 제시하였다.

2. 5. Yao, S., Zhao, Y., Zhang, A., Su, L., & Abdelzaher, T. Deepiot: Compressing deep neural network structures for sensing systems with a compressor-critic framework. In: Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems. ACM. 2017.

보조 신경망을 이용해 학습과정에서 신경망 node의 개수를 줄이는 Shrinking 기법을 제시하였다.

2. 6. 선행연구와의 차이점

2. 4.의 MorphNet을 참고하여 개선하고자 하는 점이 본 연구의 목적이다. MorphNet으로부터 크게 두 가지를 개선하고자 한다. 첫째, MorphNet의 경우 FLOP, model size만을 resource-constraint로 사용하였다. 본 연구의 경우 power usage등의 항목을 resource-constraint 항목에 추가할 것이다. 둘째, 신경망 학습 과정에서 node의 개수를 줄이는 shrinking, pruning등의 기법을 사용하여 신경망 성능은 유지하면서 학습 시간을 단축하는 것을 목표로 할 것이다.

3. 작품/논문 전체 진행계획 및 구성

표 1. 월별 진행 계획

	3	4	5	6	7	8	9	10	11
주제선정									
자료조사 및 주제 구체화									
실험 환경 설정									
실험									
결과분석									
논문작성									
GitHub 관리									
일정	제안서 제출						중간 보고서 제출		최종 보고서 제출

3. 1. 선행 연구 조사

- 주제 관련 기본 개념 습득과 최신 동향 분석을 통한 주제 구체화
- 관련 논문들을 찾아 읽고 정리

3. 2. 실험 환경 설정

- 실험에 이용할 컴퓨터, 서버 기본 설정
- 실험할 resource-constraint 선정
- 실험에 사용할 신경망 모델 선정

3. 3. 시스템 구현 및 실험

- 신경망 모델 구현 및 대조군과 실험군 구성
- resource-constraint 조건을 바꾸어가며 반복적으로 실험
- 신경망 학습 과정에 shrinking, pruning 등을 적용하여 실험 진행

3. 4. 실험 결과 분석

- 실험에 대한 결과 확인 및 분석
- 실험 결과 그래프 작성 후 동향 및 원인 분석

3. 5. 개선 방안 제시

- 결과 분석을 토대로 개선을 위한 방안 탐구
- 개선안 제시를 위한 추가 자료 조사

4. 기대효과 및 개선방향

2. 4.에서 소개한 MorphNet을 기본으로 하여 resource-constraint 가능한 자원에 power usage등의 몇몇 다른 요소들을 추가한다. 또한 자동 신경망 구조 설계의 학습과정에서 node의 개수를 줄이는 shrinking, pruning등의 기법을 사용하여 학습 시간을 단축한다. 이러한 두 가지의 사항을 개선하여 다음과 같은 세 가지의 효과를 기대한다. 첫째, 자동 신경망 구조 설계의 학습시간 단축을 통해 더 큰 네트워크로의 확장성을 지원한다. 둘째, 특정 자원사용량 절감을 목표로 최적화한다. 셋째, 특정 자원이 제한된 상황에서 네트워크의 성능을 높인다.

이 연구를 기반으로 하여 다른 자원 제한 상황에 대한 최적화, 자동 신경망 구조 설계 학습의 성능, 시간, 정확도등의 개선 등 앞으로의 자동 신경망 구조 설계에 관련된 더 많은 발전이 있을 것으로 생각된다.

5. 참고문헌

- [1] P. M. Williams. Bayesian regularization and pruning using a laplace prior. *Neural computation*, 7(1):117–143, 1995.
- [2] S. Han, J. Pool, J. Tran, and W. Dally. Learning both weights and connections for efficient neural network. In *Advances in Neural Information Processing Systems*, pages 1135–1143, 2015.
- [3] B. Liu, M. Wang, H. Foroosh, M. Tappen, and M. Pensky. Sparse convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 806–814, 2015.
- [4] Gordon, A., Eban, E., Nachum, O., Chen, B., Yang, T. J., & Choi, E, "MorphNet: Fast & Simple Resource-Constrained Structure Learning of Deep Networks." 2017.
- [5] Yao, S., Zhao, Y., Zhang, A., Su, L., & Abdelzaher, T. Deepiot: Compressing deep neural network structures for sensing systems with a compressor-critic framework. In: *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. ACM. 2017.

Appendix. 교수님 출장으로 인한 확인 메일



한성혜 <hansu0713@gmail.com>

Dongkun에게 ▾

안녕하세요 교수님
학부연구생 한성혜입니다.

학부 졸업 논문 제안서를 첨부해 놓았습니다. 검토 및 확인 부탁드립니다.

감사합니다.
한성혜드립니다.



Dongkun Shin

나에게 ▾

일단 제출하도록 해라.

FLOP을 줄이는 것과 Power를 줄이는 것은 같은 것 같은데, 구체적으로 어떤 기법을 개발하려고 하는지 잘 모르겠으니 다음 주에 구체적인 설명을 할 수 있도록 준비해둬라.

신동균.

2018년 3월 22일 오후 11:12, 한성혜 <hansu0713@gmail.com>님이 작성:



한성혜 <hansu0713@gmail.com>

Dongkun에게 ▾

네 알겠습니다.

감사합니다.
한성혜드립니다.

2018년 3월 23일 금요일, Dongkun Shin<dongkun@skku.edu>님이 작성한 메시지:
