

리뷰 - Decoupled Neural Interfaces using Synthetic Gradients

개요

나온거는 2016. 발표는 ICML 2017.

FP, BP가 반복적으로 진행됨. 그래서 하자가 진행 중엔 다른 것은 블러킹됨.

예측 계산 모델을 도입하여 이 제약을 제거.

비동기로 학습 가능

순차적 동기화된 업데이트

일반 NN은 다음의 locking이 있다.

- FF locking : 이전 층에서 입력이 전달될 때까지 기다린다.
- BP locking : 다음 층에서 gradient가 전달때 까지 기다린다.
- update locking : 다음 층이 업데이트 될 때 까지 기다린다.

아이디어 핵심

모듈 i의 웨이트 theta는 다음 값을 가지고 업데이트 한다.

$$\frac{\partial L}{\partial \theta_i} = f_{\text{Bprop}}((h_i, x_i, y_i, \theta_i), \dots) \frac{\partial h_i}{\partial \theta_i}$$

- L : 최소화 할 전체 loss
- x : 입력
- y : 타겟값
- h : 모듈의 출력
- theta : 웨이트
- 모듈 : 특정 층 1개

x는 모듈 i-1, y는 모듈 i+1에서 오기 때문에 모듈 i는 모듈 i-1, 모듈 i+1에 의존적이다.

이를 과감하게 다음으로 근사화 한다.

$$\simeq \hat{f}_{\text{Bprop}}(h_i) \frac{\partial h_i}{\partial \theta_i}$$

DNI

Decoupled Neural Interface

각 모듈(층) 간의 연결을 직접 하지 않고, DNI로 대체한다.

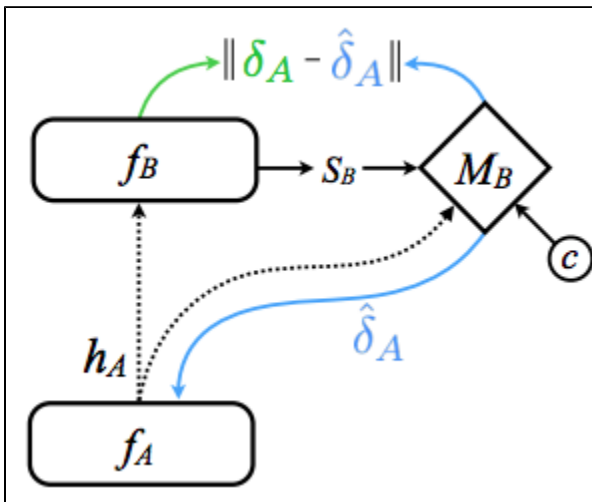
DNI의 입출력은 다음과 같다.

- 입력 : 이전 층의 출력
- 출력 : 해당 입력에 의한 다음 층의 경사

DNI의 출력을 SG(Synthetic Gradient)라 한다.

이전 층은 DN의 출력을 가지고 업데이트 한다.

이런 DNI를 사용하여 BP나 FF의 기다림 없이 비동기적으로 업데이트 할 수 있다.



A, B는 각 모듈. 일반적인 층

B의 유틸모델 M-B는 모듈 A의 출력(h_A), B의 상태(S_B), 부가적인 정보(C)를 입력으로 SG를 출력한다.

A는 이를 받아 즉각 업데이트 한다.

B가 C에서 받은 true gradient를 받아서 true gradient와 SG의 차이를 적게하도록 MB를 업데이트 한다.

논문에서는 부가적인 정보 C에 대한 별개의 언급이 없다. 단지 C가 있을 수도 있다는 것?

FF network에서의 SG

특정 층의 웨이트는 다음으로 업데이트 된다.

$$\theta_i \leftarrow \theta_i - \alpha \delta_i \frac{\partial h_i}{\partial \theta_i} ; \delta_i = \frac{\partial L}{\partial h_i}$$

그런데 DNI를 사용하여 δ_i 대신 $\hat{\delta}_i$ 을 구하여 기다림 없이 업데이트 한다.

$$\theta_n \leftarrow \theta_n - \alpha \hat{\delta}_i \frac{\partial h_i}{\partial \theta_n}$$

DNI는 delta-i와 delta-i hat의 차이

$$\|\hat{\delta}_i - \delta_i\|_2^2$$

를 최소화 하도록 학습한다.

RNN에서의 SG

다음과 같이 정리할 수 있다.

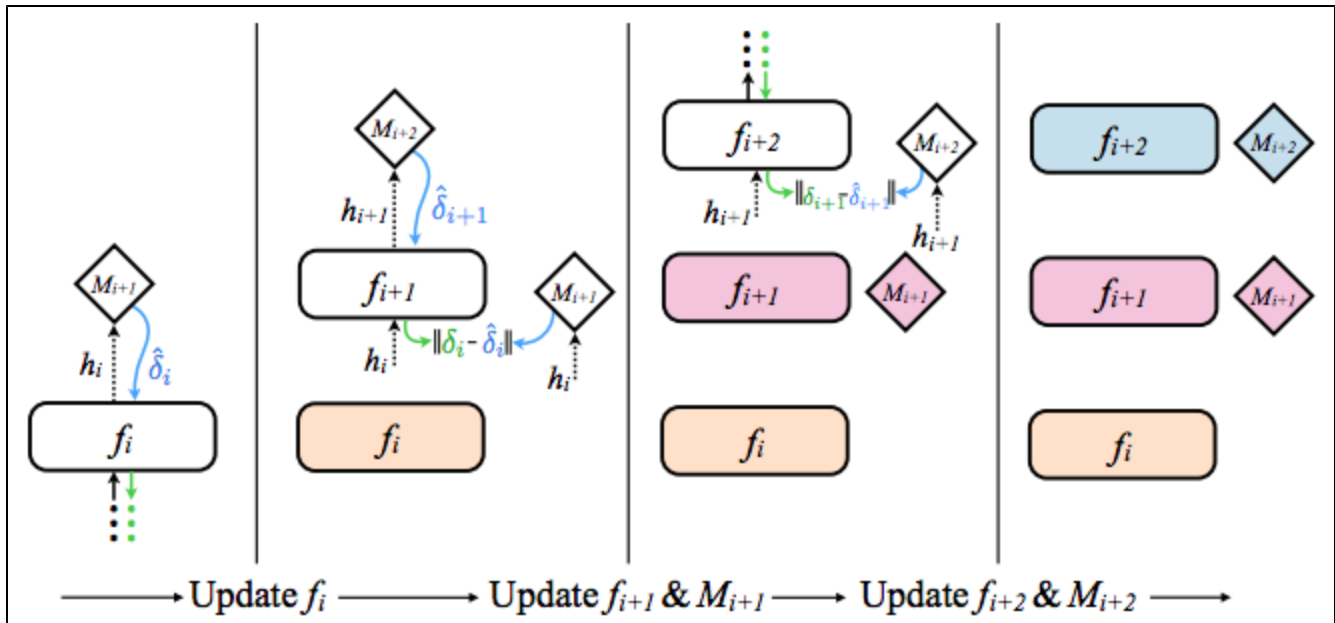
$$\begin{aligned} \theta - \alpha \sum_{\tau=t}^{\infty} \frac{\partial L_{\tau}}{\partial \theta} &= \theta - \alpha \left(\sum_{\tau=t}^{t+T} \frac{\partial L_{\tau}}{\partial \theta} + \left(\sum_{\tau=T+1}^{\infty} \frac{\partial L_{\tau}}{\partial h_T} \right) \frac{\partial h_T}{\partial \theta} \right) \\ &= \theta - \alpha \left(\sum_{\tau=t}^{t+T} \frac{\partial L_{\tau}}{\partial \theta} + \delta_T \frac{\partial h_T}{\partial \theta} \right) \end{aligned}$$

첫줄 3번째 sigma의 tau의 시작은 t+T+1이어야 한다. 오타인듯.

보통 실제적인 한계로 인해 delta-t를 0으로 처리해서 truncated BPTT가 발생.

그런데 요 값을 DNI로 값을 구해서 처리한다.

층 간의 업데이트



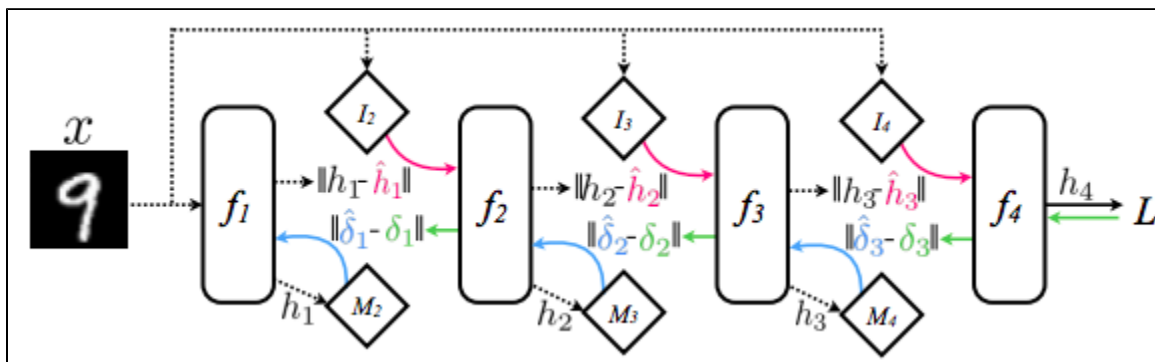
f_i 가 출력 h_i 를 내고, M_{i+1} 은 δ_i 를 출력. f_i 는 이를 가지고 즉각 웨이트를 업데이트.

f_{i+1} 는 역시 h_{i+1} 을 내고, δ_{i+1} 를 받음. 이를 가지고 M_{i+1} 을 업데이트.

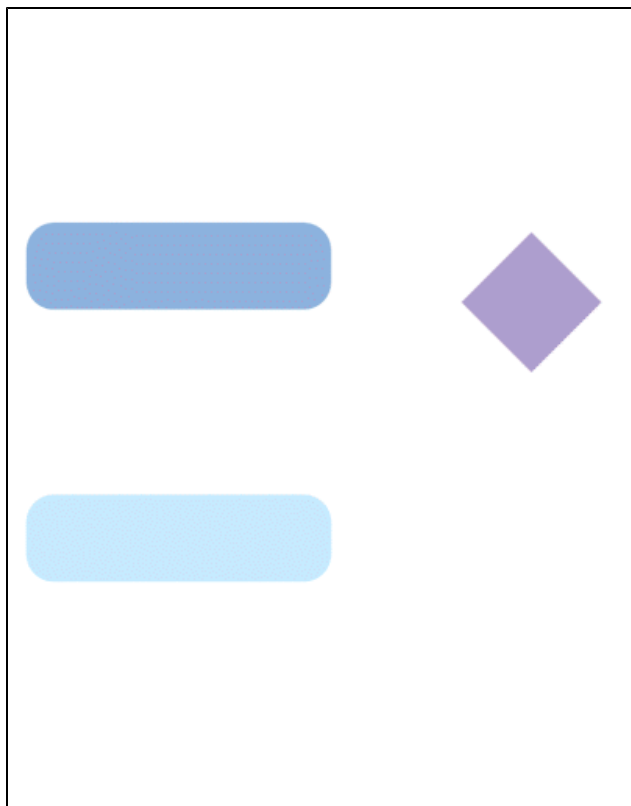
Synthetic Input 모델

SG와 반대 개념. 층의 출력 h 를 대신한다. 실제 출력 h 와 이 모델의 출력 \hat{h} 이 유사하도록 학습된다.

모델의 입력을 원 입력 데이터에서 받는다.



방법 도해

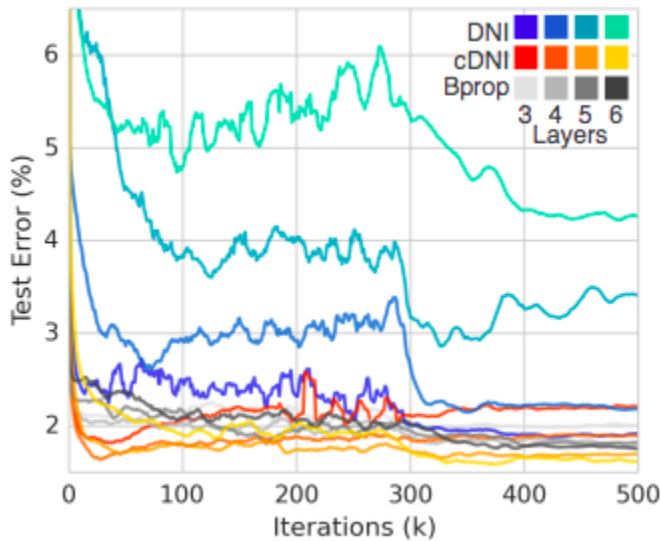


(from <https://deepmind.com/blog/decoupled-neural-networks-using-synthetic-gradients/>)

실험 결과

정확도는 떨어진다.

Layers		MNIST (% Error)				CIFAR-10 (% Error)			
		No Bprop	Bprop	DNI	cDNI	No Bprop	Bprop	DNI	cDNI
FCN	3	9.3	2.0	1.9	2.2	54.9	43.5	42.5	48.5
	4	12.6	1.8	2.2	1.9	57.2	43.0	45.0	45.1
	5	16.2	1.8	3.4	1.7	59.6	41.7	46.9	43.5
	6	21.4	1.8	4.3	1.6	61.9	42.0	49.7	46.8
CNN	3	0.9	0.8	0.9	1.0	28.7	17.9	19.5	19.0
	4	2.8	0.6	0.7	0.8	38.1	15.7	19.5	16.4



Reference

- paper : <https://arxiv.org/pdf/1608.05343>
- following paper : Understanding Synthetic Gradients and Decoupled Neural Interfaces, <https://arxiv.org/abs/1703.00522>
- 딥마인드 저자의
 - 내 짧은 소개 아티클 : <https://deepmind.com/research/publications/understanding-synthetic-gradients-and-decoupled-neural-interfaces/>
 - 상세 설명 포스트 : <https://deepmind.com/blog/decoupled-neural-networks-using-synthetic-gradients/>
- 구현 코드 : <https://github.com/vyraun/DNI-tensorflow>
- 텐서플로우 블로그의 논문 소개 : <https://tensorflow.blog/tag/synthetic-gradient/>
- 텐서플로우 블로그의 논문 요약 : <https://tensorflow.blog/2016/08/22/decoupled-neural-interfaces-using-synthetic-gradients1608-05343-summary/>
- 모두의 연구소 자료 : https://norman3.github.io/papers/docs/synthetic_gradients
- 설명 ppt
 - <https://www.slideshare.net/anujgupta5095/synthetic-gradients-decoupling-layers-of-a-neural-nets>
 - https://www.slideshare.net/aureliengeron/synthetic-gradients-tutorial?qid=5b6f8f07-6697-4fb4-a9e1-c5c0bea0283c&v=&b=&from_search=1
 - 설명 동영상 : https://www.youtube.com/watch?v=1z_Gv98-mkQ