

CIFAR-10 CNN

2017-11-16

홍만수

1. Layer

총 레이어: 5 (Convolution: 3, Fully-connected: 2)

> Convolution

+ conv2d(strides=[1, 1, 1, 1], padding='SAME')

+ max_pool(ksize=[1, 2, 2, 1], strides=[1, 2, 2, 1]): 매번 반으로 줄임

kernel(3x3 -32) → batch normalization → relu → max pool

→ kernel(3x3 -128) → batch normalization → relu → max pool

→ kernel(5x5 -256) → batch normalization → relu → max pool

> fully-connected layer

→ kernel(4 * 4 * 256 -512) → batch normalization → relu → drop out

→ kernel(512 -10)

2. 인자 탐색 과정 및 결과

솔직히 알고 한 것이 아니라 그때그때 결과를 보고 건드렸다. 탐색 과정이라고 하기에는 가능한 아무거나 손을 댄 느낌이다. 예를 들자면, AdamOptimizer 는 좋다고 해서 사용했을 뿐이다. 그래서 시도를 실패했을 때, 그 전으로 되돌릴 수가 없었다.

결과는 test accuracy 0.78 가량.

3. 모델 성능 향상 과정 서술

처음에는 맥북에어로 시작했다. Convolution 을 두 번 거치는 작은 것이었다. 시험삼아 Epoch 도 10 번만 돌렸다. 설마 이 정도도 돌리지 못하진 않겠지라고 생각했다. 정말 큰 실수였다. 맥북에어는 NVIDIA GPU 가 없다. Intel HD 6000 따위나 쓰는 커다란 스마트폰이다. 첫 epoch 을 수행하는데 무려 4 분 가량이 걸렸다. 기다리고 기다려 마

친 첫 세션의 테스트 정확성은 0.2 를 넘지 못했다. 10 개 중 1 개만 맞췄다는 말이다. 수능 문제를 전부 찍어도 1/10 보다는 높은 확률로 맞출 것이다.

레이어를 하나 더 추가했다. 그리고 또다시 40 여분을 기다렸다. 정확성은 여전히 0.2 언저리를 알짱거렸다. 구성대에 심각한 문제가 있는 것이 틀림없었다. 심각한, 그리고 많은 수정이 필요할 것으로 보였다. 하지만 CPU 만을 써서는 마음 놓고 몇 번이고 돌릴 수 없었다. 한줄 쓰고 40 분, 한 줄 빼고 40 분. 결국 모델의 향상을 위해서 하드웨어를 교체할 필요가 있음을 인정해야만 했다.

GeForce 1050ti 를 담은 DELL 노트북으로 변경했다. 이제 tensorflow-gpu 를 설치하고 빠른 출력을 기할 수 있으리라고 생각했으나 생각지도 못한 복병에 뒤통수를 후려맞게 되었다. 노트북의 운영체제가 Ubuntu 17.04 였는데 어떤 수를 써도 CUDA Toolkit 9(이하 CUDA-9.0)만 설치되었다. tensorflow-gpu 는 CUDA Toolkit 8(이하 CUDA-8.0)을 이용한다. 몇 시간 동안의 검색 끝에 간신히 돌아서 CUDA-8.0 을 설치할 수 있었다. 그제야 제대로 왔다고 진행이 가능해졌다.

Convolution layer 는 3, fully connected layer 는 2 였다. 하드웨어가 갖춰졌으니 epoch 를 100 으로 높이고 실행했다. 그럼에도 불구하고 0.2 ~0.3 언저리에서 결과가 나왔다. 학습이 전혀 되고 있지 않았다. 단순히 레이어를 늘리거나 epoch 를 늘리는 것만으로 결과를 얻기에는 지나치게 비효율적으로 보였다. 코드를 다시 훑어보던 중 어처구니 없는 것을 발견할 수 있었다. Optimizer 의 learning rate 이 0.01 밖에 되지 않았던 것이다. Learning rate 를 0.001 로 재설정하자 당장 0.6 의 결과를 얻을 수 있었다.

더 높은 결과를 얻기 위해 다음은 batch normalization 을 시도했다. relu 전 batch 를 매번 정규화해주자 곧 0.7~0.8 의 정확성을 얻어냈다. 하지만 epoch 50 회 이후에서나 간간히 0.8 을 넘고 주로 0.79 언저리에서 노니는 것을 보니 무언가 변화가 필요함을 느꼈다.

다시 레이어를 추가했다. Convolution layer 를 8 개 더 추가했다. 하지만 문제가 일어났다. logit 과 label 의 크기가 달라 softmax_cross_entropy_with_logits 가 돌아가지 않았다. 몇 번이고 고쳐보려고 했지만 같은 에러 메세지만 받을 뿐이었다. 그래서 레이어를 다시 줄이기 시작했다. 마침내 계속 줄이다보니 다시 convolution layer 가 3 밖엔 남지 않게 되었다. 더이상 레이어를 줄일 수 없다는 생각에 몇 차례 수정을 해보다보니 다시 logit 과 label 사이즈가 맞아 수행이 되었다. TF 를 완전히 이해하지 못한 탓이다.

이전 코드와 완전히 같지 않았는지 테스트 정확성이 0.8 을 넘지 못하게 되었다. (따로 저장해둘 걸 그랬다.) 새 레이어를 추가하는 것이 두려웠기 때문에 대신 drop-out probability 를 0.7 로 도입했다. 보다 generalized 된 결과를 얻을 수 있을 것을 기대했다. 결국 정확성이 미묘하게 떨어졌다.

명일 발표 준비에 타수업 보고서 준비도 필요했다. 초기에 맥북에어를 쓰고 노트북 교체 후 설정하는데 시간을 너무 많이 썼다. Training 이 overfitting 이 아닌지 validation, traning accuracy 와 test accuracy 비교도 필요한데 더 이상 시간을 쓰기 어려워 불만족스럽지만 아쉽게도 이 정도로 그만둔다.