

Assignment 2

Introduction to Deep Learning

201911013 곽현우

Figure 1. Accuracy comparison for test set

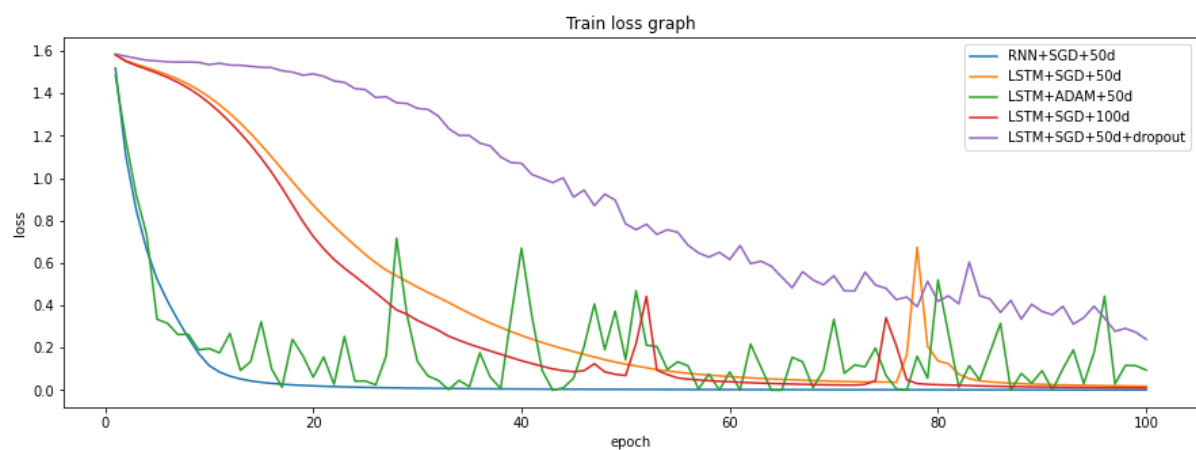
Test Accuracy	
RNN+SGD+50d	80.357143
LSTM+SGD+50d	75.000000
LSTM+ADAM+50d	66.071429
LSTM+SGD+100d	69.642857
LSTM+SGD+50d+dropout	55.357143

Figure 2. All emojis for test set

	RNN+SGD+50d	LSTM+SGD+50d	LSTM+ADAM+50d	LSTM+SGD+100d	LSTM+SGD+50d+dropout	Ground Truth
I want to eat	🍴	🍴	🍴	🍴	🍴	🍴
he did not answer	😞	😞	😞	😞	😞	😞
he got a very nice raise	😄	❤️	😄	❤️	😄	😄
she got me a nice present	😄	😄	😄	😄	😄	😄
ha ha ha it was so funny	😄	😄	😄	😄	😄	😄
he is a good friend	😄	❤️	😄	❤️	❤️	😄
I am upset	😞	😞	❤️	😞	😞	😞
We had such a lovely dinner tonight	😄	😄	😄	😄	❤️	😄
where is the food	🍴	🍴	🍴	🍴	🍴	🍴
Stop making this joke ha ha ha	😄	😄	😄	😄	😄	😄
where is the ball	🏀	🏀	🏀	🏀	🏀	🏀
work is hard	😞	😞	❤️	😞	😞	😞
This girl is messing with me	😞	😞	😞	😞	😞	😞
are you serious	😞	😞	😞	😞	🍴	😞
Let us go play baseball	🏀	🏀	🏀	🏀	🏀	🏀
This stupid grader is not working	😞	😞	😞	😞	😞	😞
work is horrible	😞	😞	😞	😞	😞	😞
Congratulation for having a baby	😄	😄	❤️	😄	😄	😄
stop pissing me off	😞	😞	😞	😞	😞	😞
any suggestions for dinner	🍴	🍴	❤️	😄	😄	🍴
I love taking breaks	❤️	❤️	😄	❤️	❤️	❤️
you brighten my day	😄	😄	😄	😄	😄	😄
I boiled rice	🍴	🍴	🍴	🍴	🍴	🍴
she is a bully	❤️	❤️	😞	❤️	❤️	😞
Why are you feeling bad	😞	😞	😞	😞	❤️	😞
I am upset	😞	😞	❤️	😞	😞	😞
give me the ball	🏀	🏀	🏀	🏀	🏀	🏀
My grandmother is the love of my life	😄	😄	😄	❤️	❤️	❤️
enjoy your game	🏀	🏀	🏀	🍴	🏀	🏀
valentine day is near	😄	😄	❤️	😄	😄	😄
I miss you so much	❤️	❤️	❤️	❤️	❤️	❤️
throw the ball	🏀	🏀	🏀	🏀	🏀	🏀

My life is so boring	😞	😞	😞	😞	😞	😞
she said yes	😞	😞	😞	😞	😞	😞
will you be my valentine	😞	😞	❤️	😞	🤖	😞
he can pitch really well	🤖	🤖	😞	😞	🤖	🤖
dance with me	😞	😞	😞	😞	😞	😞
I am hungry	🤖	🤖	🤖	🤖	🤖	🤖
See you at the restaurant	🤖	🤖	🤖	😞	🤖	🤖
I like to laugh	❤️	😞	😞	❤️	❤️	😞
I will run	🤖	🤖	❤️	🤖	🤖	🤖
I like your jacket	❤️	❤️	❤️	❤️	❤️	❤️
i miss her	❤️	❤️	❤️	❤️	❤️	❤️
what is your favorite baseball game	🤖	🤖	🤖	🤖	😞	🤖
Good job	😞	😞	😞	😞	😞	😞
I love you to the stars and back	😞	❤️	😞	❤️	❤️	❤️
What you did was awesome	😞	😞	😞	😞	😞	😞
ha ha ha lol	😞	😞	😞	😞	😞	😞
I do not want to joke	😞	😞	❤️	❤️	❤️	😞
go away	😞	🤖	😞	🤖	🤖	😞
yesterday we lost again	🤖	😞	😞	😞	😞	😞
family is all I have	❤️	🤖	😞	🤖	🤖	❤️
you are failing this exercise	😞	😞	😞	😞	😞	😞
Good joke	😞	😞	😞	😞	😞	😞
You deserve this nice prize	😞	😞	😞	😞	😞	😞
I did not have breakfast	🤖	😞	🤖	😞	😞	🤖

Figure 3. Training Loss Graph



Description

* experiments

A: RNN + SGD + 50d

B: LSTM + SGD + 50d

C: LSTM + ADAM + 50d

D: LSTM + SGD + 100d

E: LSTM + SGD + 50d + Dropout

1. Comparison of Optimizer

사용된 Optimizer 는 ADAM 과 Stochastic Gradient Descent(SGD)가 있다. SGD 의 경우 전체 Dataset 을 다 보고 Weight 를 Update 시키는 것이 아니라 일부 Data 들을 보고 Weight 를 Update 시키는 Optimizer 이다. ADAM 의 경우 Update 시 Momentum 인자를 도입하여 이전에 Update 시킨 방향을 참고하고 Learning Rate 또한 이전 Update 된 한 Step 의 크기를 고려하여 조절되도록 하였다. 즉 ADAM 이 SGD 보다 좀 더 Global Minimum 에 빠르게 도달할 수 있다.

이는 위 Los Graph 에도 A 와 B 실험을 비교해보았을 때 ADAM 을 사용한 B 실험이 Loss 가 빠르게 수렴한다는 것을 통해 알 수 있다.

2. RNN Structure

사용한 RNN structure 는 RNN 과 LSTM 이다. RNN 은 특정 Sequence Data 가 순차적을 들어왔을 때 Hidden State 값을 도입하여 이전의 Data 들의 정보를 참고해 재귀적으로 Inference 하는 가장 기본적인 Model 이다. 그러나 이러한 RNN 은 Sequence Data 의 길이가 길어지면 Vanishing Gradient 문제가 발생하여 Weight 가 제대로 Update 되지 않는 문제가 발생한다.

이러한 문제를 해결하기 위해 LSTM 이라는 모델을 도입하였다. LSTM 은 Cell State 라는 값을 도입하여 장기기억과 단기기억을 통해 좀 더 RNN 보다 맥락 추론을 잘 할 수 있도록 한다.

위 실험에서는 RNN Structure 만 다르고 다른 조건은 동일한 실험 결과를 비교해보았을 때 RNN 이 LSTM 보다 좀 더 좋은 성능을 내고 있다. 이는 문장을 이루는 단어의 개수가 크지 않아 Long-Term Dependency 가 없었을 수 있다. 또한 LSTM 은 Cell 내부에 많은 Parameter 가 존재하므로 Optimization 하는 과정이 어렵기 때문에 RNN 이 LSTM 보다 좋은 성능을 냈을 것이라고 추측할 수 있다.

3. The Length of glove vectors

위 실험에서는 glove vector 를 50 과 100 을 사용하였다. Test Accuracy 를 보면 차원이 50 이었을 때가 100 이었을 때 보다 더 높게 측정되었다. 이는 단어를 100 차원으로 표현했을 때 좀 더 분별성을 뿜 수 있고 단어를 더 잘 표현했다고 해석할 수 있다.

4. Dropout

Dropout 은 특정 비율로 model 내 Node 들을 비활성화 시켜서 Overfitting 을 막는다. Train Loss Graph 에서 B 와 E 실험의 graph 를 보면 Dropout 을 적용했을 때 Loss 값이 느리게 떨어지는 것을 볼 수 있는데 이는 Train Dataset 에 Overfitting 되지 않도록 한다는 것을 알 수 있다.

5. The Difference between Word2Vec and Glove

NLP 에서 특정 단어들을 하나의 벡터로 나타내는 것을 Embedding 이라고 하는데 Word2Vec 과 Glove 가 해당한다.

Word2Vec 은 SkipGram 과 CBOW 라는 방식이 있다.. SkipGram 이라는 방식은 특정 단어에 대한 주변 단어들과의 관계를 고려하여 Input 을 특정 단어의 Vector(One-Hot Encoding) Target 을 Input 의 이웃 단어의 Vector(One-Hot Encoding)으로 하여 Hidden Layer 를 추가한 뒤 학습시킨다. 학습이 완료된 후 특정 Word 를 Hidden Layer 들을 거쳐서 Vector 로 나타내는 방식이 Word2Vec 이다.

CBOW 는 SkipGram 과 반대로 주변 단어들을 Input 으로 중심 단어를 Target 으로 훈련시킨 후 Embedding 하는 방식이다.

Glove 는 Word2Vec 가 주변 단어들만 보고 Corpus 의 전체적인 통계정보를 반영하지 못한다는 한계를 해결한다. Glove 는 동시 등장행렬을 목표로 학습하도록 하며 최종적으로 두 단어의 Embedding Vector 의 곱이 동시 등장 확률의 Log 값이 되도록 하는 Embedding 방식이다.