

배치정규화를 적용한 CBOW 모델 성능 향상 개선에 관한 연구

윤국범, 이정수*

부산외국어대학교, *부산외국어대학교

rnqja486@naver.com , *rhee@bufs.ac.kr

A Study on the Improving Loss Rate of CBOW Model Using Batch Normalization systems

Yoon Guk Beom, Rhee Jung Soo*

Busan University of Foreign Studies., *Busan University of Foreign Studies.

요 약

CBOW 모델은 주변 단어로부터 표적(target) 단어를 예측하는 분산 표현 방식의 자연어 처리 모델이다. 본 논문은 CBOW 모델에 배치정규화 (BatchNormalization)를 적용하여 PTB 말뭉치를 학습시켜 CBOW 모델의 학습 손실률을 개선하는 것을 목표로 한다. 기존 모델과 제안 모델의 실험 결과 학습 손실률이 1.53에서 0.82로 줄어들었음을 확인 할 수 있다.

I. 서론

본 논문에서는 자연어 처리 분야에서 사용되고 있는 CBOW(Continuous Bag-of-Words) 모델의 성능 향상을 위한 방법을 제안한다. CBOW는 말뭉치(corpus)를 학습하는 기본 신경망 모델로써 전이 학습에 활용된다. 본 연구에서는 다양한 종류의 말뭉치 중에서도 PTB(Penn Treebank) 말뭉치를 사용한다. PTB 말뭉치란 자연어 처리와 컴퓨터 언어학에서 중요한 역할을 하는 중형 말뭉치이며, 주로 1989년 월스트리트 저널에서 발췌한 2,499개의 기사로 구성된 말뭉치이다. 약 100만개의 단어와 49,208의 문장이 포함되어있는 PTB 말뭉치는 CBOW 모델의 성능을 평가하는 데 있어 대표적인 벤치마크로 널리 활용되고 있다. 본 논문에서는, SoftmaxWith Loss 계층에서의 기존의 손실을 줄이기 위해 CBOW 모델에 배치 정규화 (Batch Normalization) 기법을 적용한다. 실제로 CBOW 모델에 배치 정규화 기법을 적용하는 실험을 통해 제안된 방법이 CBOW 모델의 손실률 개선에 효과적임을 검증한다.

II. 본론

CBOW 모델은 주변 단어로부터 표적 단어를 학습하는 신경망 모델이다. CBOW 모델은 주어진 문장에서 주변에 있는 단어들을 입력으로 받아 벡터로 표현한다. 이 벡터들은 각 단어의 의미를 압축해서 나타내는 숫자 배열이다. 이렇게 표현된 벡터들은 전이학습으로 활용된다.

초기가중치로부터 학습을 진행하고 실제 단어와 예측단어의 손실은 SoftmaxwithLoss 계층에서 이루어진다. 이 손실을 줄이기 위하여 역전파 방식으로 가중치의 매개변수를 조정한다. 그림1은 simple CBOW 모델의 순전파 과정을 시각적으로 표현하고, 아래는 simple CBOW의 모델 구성을 기술한 것이다.

입력 단어 'you'와 'goodbye'에 해당하는 단어 index(0과 2)와 'W_in' 가중치와의 행렬곱을 'Embed' 계층을 통과하여 임베딩 벡터로 변환한다. 임베딩 계층에서 사용된 클래스는 EmbeddingDot로서 아래 그림2와 같다.

두 임베딩 벡터는 더해진 후, 평균을 내기 위해 0.5를 곱한 후, 평균화된 임베딩 벡터가 은닉층의 활성화 벡터(h)로 사용되며 'W_out' 가중치 행렬과 곱해져 최종 출력 벡터가 생성되고 최종적으로 출력 벡터는 손실 함수와 비교되어 손실 값이 계산된다. 손실을 줄이기 위해 역전파 최적화 알고리즘 Adam을 적용하여 가중치 'W_in'을 수정하여 학습을 반복한다. 또한 네거티브 샘플링을 적용해 실험의 불필요한 계산량을 줄여 학습 속도를 향상시킨다. 본 실험에서는 10 epoch로 설정해 PTB를 학습한다.

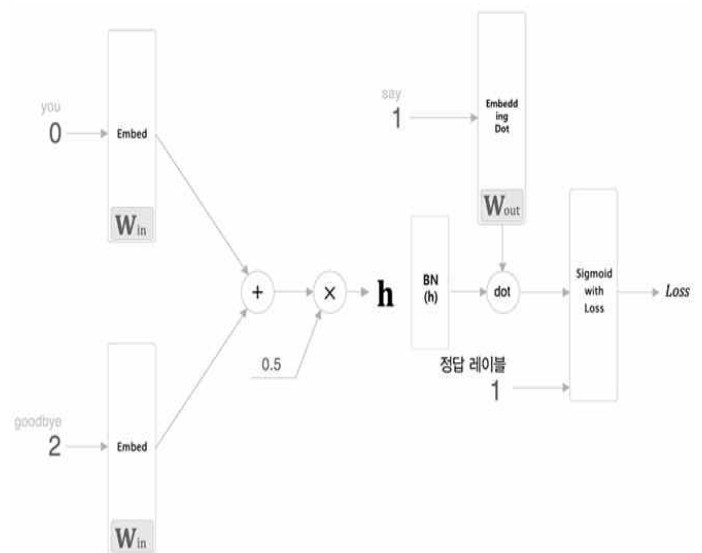


그림 3 배치정규화를 적용시킨 CBOW 모델 구성도

그림3과 같이 가중치가 적용된 결과는 중간 히든 레이어 인 h로 전달되고 히든 레이어 이후에 배치 정규화 레이어가

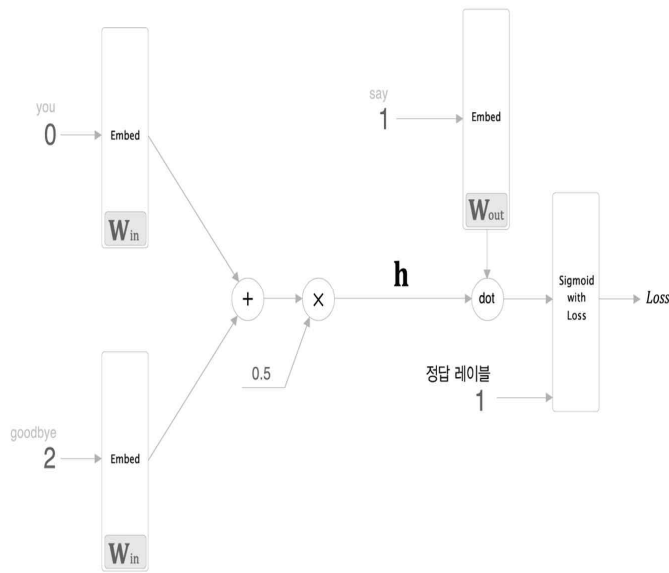


그림 1 simple CBOW 모델 구성도

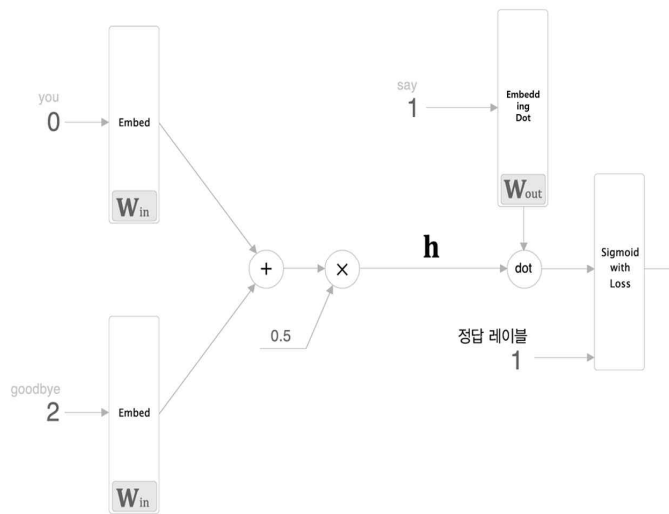


그림 2 Embedding Dot를 적용한 CBOW 모델 구성도

추가된다. 배치 정규화는 입력값을 정규화하여 학습의 안정성과 속도를 향상시키는 기법으로, 과적합 방지 및 모델의 일반화 능력을 향상시키는데 기여한다.

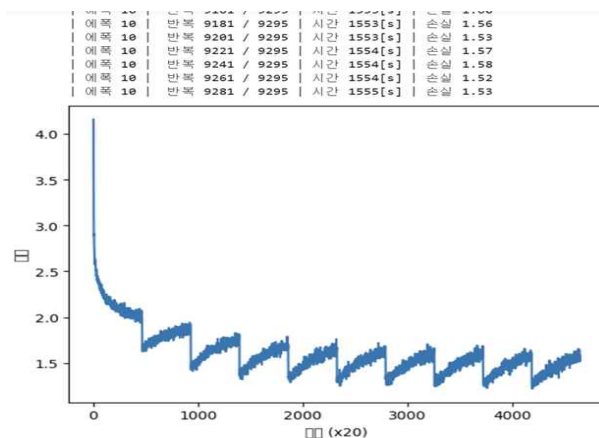


그림 4 일반 CBOW 모델 구현

실험 결과 Windos size가 5인 CBOW 모델에 배치 정규화를 적용했을 때 손실률이 1.53(그림 4)에서 0.82(그림 5)로 감소하는 것을 확인했다.

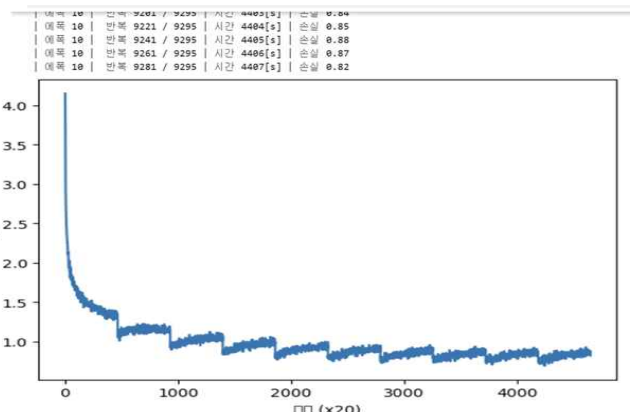


그림 5 Batch Normalization을 적용한 CBOW 모델 구현

이는 배치 정규화가 CBOW 모델 학습 과정에서 발생하는 무분별한 가중치 문제를 해결하고 학습 속도를 향상시켜 손실률을 개선하는 데 효과적임을 보여준다.

III. 결 론

본 논문에서는 CBOW 모델에 배치정규화를 적용하여 손실률을 개선했다. 실험을 통해 제안된 방법이 CBOW 모델의 손실률 개선에 효과적임을 검증한다. 즉, 손실률이 1.53(그림 4)에서 0.82(그림 5)로 감소하는 것을 확인한다.

ACKNOWLEDGMENT

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 ICT혁신인재4.0사업의 연구결과로 수행되었음 (IITP-2024-2020-0-01825)

참 고 문 헌

- [1] Saito Koki, Deep Learning from Scratch ②
- [2] Sergey Ioffe and Christian Szegedy(2015): Batch Normalization:Accelerating Deep Network Training by Reducing Internal Covariate Shift.arXiv:1502.03167[cs](February 2015).
- [3] Levy, Omer, and Yoav Goldberg, and Ido Dagan: "Improving distributional similarity with lessons learned from word embeddings." Transactions of the Association for Computational Linguistics 3 (2015):211-225.Standard(AE S),"FIPS PUB ZZZ, 2001, (http://www.nist.gov/aes).
- [5] Daemen, J., and Rijmen, V. "AES Proposal: Rijndael, Version2," Submission to NIST, March 1999.