JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 30 No. 1, pp. 65-72, January 2025
https://doi.org/10.9708/jksci.2025.30.01.065

Minimizing the Loss Values in the CBOW Models

Gukbeom Yoon*, Jungsoo Rhee**

*Graduate student, Department of Smart Convergence Security, Busan University of Foreign Studies, Busan, Korea
**Professor, Department of Smart Convergence Security, Busan University of Foreign Studies, Busan, Korea

[Abstract]

The Continuous Bag of Words (CBOW) model is a popular technique in natural language processing (NLP) used to generate word Embeddings. It predicts a target word given its surrounding context words. The model consists of an input layer, a hidden layer, and an output layer. The PTB, which is commonly used as a benchmark for evaluating the performance of the CBOW model, is a medium-sized corpus that plays an important role in natural language processing and computational linguistics. It consists of 2,499 articles extracted from the Wall Street Journal in 1989, containing approximately 1 million words and 49,208 sentences. This paper aims to improve the average loss value of the loss function by applying batch normalization to the CBOW model and training it on the PTB dataset. To achieve the objective of this paper, experiments were conducted in a notebook environment equipped with CuPy, comparing the original CBOW model with the batch normalized CBOW model. The results shows that average loss value decreased from 1.25 to 0.65. Therefore, this paper demonstrates the effectiveness of batch normalization in improving the performance of the CBOW model and is expected to contribute to refining distributional representation of words for transfer learning.

▸ **Key words**: Batch Normalization, CBOW, CNN, CuPy, Embedding, Negative Sampling, Word2Vec

[요 약]

CBOW 모델은 주변 단어로부터 표적 단어를 예측하는 분산 표현 방식의 자연어 처리 신경망 모델이다. CBOW 모델의 성능을 평가하는 데 있어 대표적인 벤치마크로 사용되는 PTB는 자연어 처리와 컴퓨터 언어학에서 중요한 역할을 하는 중형 말뭉치이며, 1989년 월스트리트 저널에서 발췌한 2,499개의 기사로 구성되어, 약 100만개의 단어와 49,208개의 문장이 포함되어 있다. 본 논문은 CBOW 모델에 배치정규화를 적용하여 PTB를 학습시켜 손실함수의 평균 손실값을 개선하는 것을 목표로 한다. 논문의 목표를 구현하기 위해, CuPy를 탑재한 노트북 환경에서 기존 CBOW 모델과 배치정규화 적용 모델을 비교 실험하였으며, 그 결과 평균 손실값이 1.25에서 0.65로 감소함을 확인하였다. 그러므로 본 논문은 CBOW 모델 성능 향상에 있어 배치정규화의 유효성을 입증하고, 전이학습을 위한 분산 표현의 정교화를 이루는 데 도움을 줄 것으로 기대된다.

▸ **주제어**: 배치정규화, CBOW, CNN, CuPy, 임베딩, 네거티브 샘플링, Word2Vec

• First Author: Gukbeom Yoon, Corresponding Author: Jungsoo Rhee
*Gukbeom Yoon (rnrqja486@naver.com), Department of Smart Convergence Security, Busan University of Foreign Studies
**Jungsoo Rhee (rhee@bufs.ac.kr), Department of Smart Convergence Security, Busan University of Foreign Studies
• Received: 2024. 12. 06, Revised: 2024. 12. 24, Accepted: 2024. 12. 30.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

CBOW(Continuous Bag Of Word)는 말뭉치(corpus)를 학습하는 기본 신경망 모델로서 RNN과 같은 타 신경망으로의 전이 학습에 활용된다[1]. 본 연구에서는 다양한 종류의 말뭉치 중에서도 PTB(Penn Treebank)를 사용한다. PTB 는 자연어 처리와 컴퓨터 언어학에서 중요한 역할을 하는 중형 말뭉치이며, 1989년 월스트리트 저널에서 발췌한 2,499개의 기사로 구성되어 있다. 약 100만개의 단어와 49,208개의 문장이 포함되어 있는 PTB는 CBOW 모델을 학습시키는데 있어 대표적인 벤치마크로 널리 활용되고 있다[2].

RNN을 활용한 언어 모델에서도 단어의 분산 표현을 얻을 수 있지만, 어휘 수의 증가에 따른 대응이나 단어 분산 표현의 질적 개선을 위해 언어의 의미를 학습하는 AI 신경망 모델로서 CBOW 모델과 Skip-Gram 모델로 구성된 Word2Vec을 Tomas Mikolov가 2013년에 처음 소개하였다[3, 4]. 특히, CBOW와 Skip-Gram 모델은 단일 은닉층을 사용하는 간단한 신경망 구조로 설계되어 효율적인 학습을 가능하게 한다. 이와 더불어 Tomas Mikolov는 2015년 5월 19일에 Word2Vec 기술과 관련된 미국 특허(US9037464)를 등록하였다[5]. 이 특허는 Word2Vec이 단순한 언어 모델 이상의 역할을 수행하며, 단어의 의미를 효과적으로 벡터 공간에 표현하고 이를 다양한 자연어 처리 태스크에 활용할 수 있는 혁신적인 접근법임을 보여준다. 이는 기존의 언어 모델들이 겪었던 어휘 크기 증가에 따른 계산 복잡성 문제를 크게 개선한 점에서 주목받았다.

CBOW 모델의 주요 목표는 벡터 공간에서 의미적으로 유사한 단어들이 서로 가까운 위치에 있도록 학습하여 단어를 벡터로 변환하는 것이다. 이러한 신경망 기반의 연산은 단어 간의 의미적 관계를 수리적으로 표현할 수 있음을 보여준다. 이는 Word2Vec이 단순히 단어의 빈도나 위치 정보를 반영하는 것을 넘어, 단어 간의 의미적 관계를 확률적으로 학습할 수 있음을 의미한다. 특히, 한국어와 같은 언어 특성을 고려하여 Word2Vec 모델을 최적화한 연구에서도 CBOW와 Skip-Gram 모델이 효과적으로 활용된 사례가 보고되었다[6].

또한, Word2Vec은 매우 빠르고 효율적인 학습이 가능하도록 설계되어 이전의 통계적 기법들보다 학습 속도가 빠르며, 상대적으로 적은 양의 데이터로도 의미 있는 결과를 도출할 수 있다. 이러한 특징 덕분에 Word2Vec은 감성 분석, 기계 번역, 챗봇과 같은 다양한 자연어 처리 응용 분야에서 널리 활용되고 있으며, 여러 학문 분야에서 이 모델에 영감을 받아 유사한 분산 표현 방식을 채택하고 있다. 특히, 다국어 텍스트 분류와 같은 복잡한 작업에서도 효과적으로 활용될 수 있으며, 여러 언어에서 동일한 단어 벡터 공간을 활용한 성능 향상을 보여준 연구도 있다[7].

2022년에 발표된 “On the validity of pre-trained transformers for natural language processing in the software engineering domain”은 직접적인Word2Vec 접근 방식보다도 Word2vec와 유사한 단어 임베딩 모델을 기반으로 여러 신경망 가중치 할당 계층을 추가하는 ELMo 및 BERT와 같은 Transformer모델에 응용된다고 주장한다[8, 9]. 특히, Word2Vec의 하나의 구성 모델인 CBOW 모델은 소셜 미디어 데이터를 기반으로 감성 분석과 같은 작업에서도 효과적으로 활용되고 있다[10, 11].

이러한 CBOW의 중요성을 감안하여 본 논문에서는 Negative Sampling을 적용한 개선된 CBOW모델의 Sigmoid-with-Loss 계층에서 손실을 줄이기 위해 배치 정규화(Batch Normalization) 기법을 적용한다. 그리고 CBOW 모델에 배치정규화 기법을 적용하는 실험을 통해 제안된 방법이 손실함수의 평균 손실값 개선에 효과적임을 검증한다.

아래는 Word2Vec의 개념을 채택한 분산 표현 방식 모델의 연도별 개발 순서와 개발자를 나타낸 표이다[12].

Table 1. Yearly Development Sequence and Developers for models similar to Word2Vec

Year	Model Name	Developer
2013	Word2Vec	Google
2014	Doc2Vec	Google
2014	Paragraph2Vec	Google
2015	Sent2Vec	NIRIA
2016	Node2Vec	Stanford
2017	Tweet2Vec	Carnegie Mellon Univ
2017	StarSpace	Facebook AI Research
2018	Face2Vec	Facebook
2019	Product2Vec	Amazon
2019	Graph2Vec	MIT
2020	Context2Vec	Google
2021	Code2Vec	Technion
2021	CodeBERT	Microsoft Research
2022	Concept2Vec	OpenAI
2022	BioVec	DeepMind

II. Preliminaries

CBOW 모델은 주변 단어(맥락)로부터 표적 단어를 학습하는 신경망 모델로, 주어진 문장에서 맥락들을 입력으로 받아 예측 단어로 표현한다. 초기 가중치로부터 학습을 시작하고, 실제 단어와 예측 단어 간의 차이는 Softmax-with-Loss 계층을 통해 계산된다. 이 차이들의 평균을 평균 손실값이라 하고 이것을 줄이기 위해 역전파 방식을 사용하여 가중치 매개변수를 조정한다. 보다 엄밀한 손실값의 정의는 4.2 Negative Sampling 부분에서 상세히 소개되어 있다. 위에서 언급한 기본 신경망 모델을 아래의 그림으로 나타낸다[13].

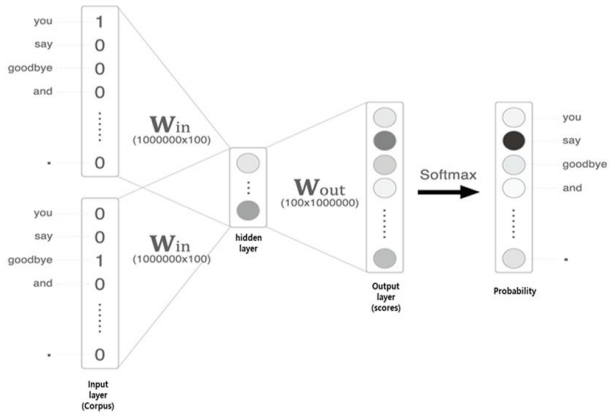

Fig. 1. Schematic underlying neural network model

말뭉치를 맥락과 표적 단어로 분리하여 정수 데이터로 재구성한 뒤, 정수 데이터를 원 핫 벡터로 변환한다. 재구성된 데이터는 원 핫 벡터로 입력되어 다음과 같은 Computational Graph를 따라 위의 신경망 모델이 학습된다. 그림1에서, 맥락인 "you"와 "goodbye"를 입력으로 받아, 맥락을 완성하는 표적 단어인 "say"를 예측하는 방식으로 학습이 이루어진다. 학습되는 과정을 Computational Graph로 좀 더 상세히 살펴보면 다음과 같다[14].

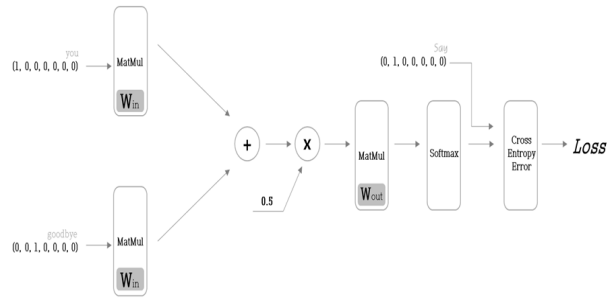

Fig. 2. Computational Neural Network Diagram

입력층에서는 단어 "you"와 "goodbye"를 원 핫 벡터로 표현하며, 이를 가중치 행렬 *Win*와 곱하고, 두 개의 산출된 벡터의 평균을 계산하여 은닉층에 전달한다. 출력층에서는 은닉층의 출력이 가중치 행렬 *W*out와 곱해진 후 Softmax 함수를 통해 점수(확률)로 생성하여 "say"와 유사한 단어를 예측한다. 손실 함수로는 Cross-Entropy 함수를 사용하여 예측 값과 실제 값 “say”와의 오차를 측정하고, 이 오차를 줄이도록 학습한다. 이 구조는 맥락을 기반으로 표적 단어를 예측하는 CBOW의 특성을 잘 보여주며, *Win* 과 *W*out 가중치 행렬의 학습을 통해 예측 값을 산출한다.

III. Problems on the Basic CBOW Model

기본 CBOW 모델은 다음과 같은 3가지 계산상 병목 현상을 유발하기에 이를 해결하는 개선된 CBOW 모델이 필요하다[1, 15].

3.1 Input-to-Hidden Layer Bottleneck

CBOW는 주어진 맥락들의 원 핫 벡터를 입력으로 받는다. 말뭉치의 Size가 100만일 때 입력 원 핫 벡터는 1x100만이 된다. 하지만 은닉층이 100개의 뉴런으로 구성된다면, 100만×100 크기의 입력가중치 행렬 *Win*이 필요하다. 이를 통하여 은닉층으로 정보를 전달하기 위해서 입력 원 핫 벡터와 가중치 행렬 *Win*과의 행렬곱을 실제로 계산을 수행하면 병목 현상이 일어난다.

3.2 Hidden-to-Output Layer Bottleneck

만약 은닉층이 100개의 뉴런으로 구성되어 있다면, 출력가중치 *W*out이 100×100만이 되어야 하므로, 이 두 개의 행렬곱 연산도 계산상 병목현상을 유발한다.

3.3 Computation of the Softmax Layer

출력층 점수를 계산하기 위해서 Softmax에 의한 계산은 100만개 Exponential값의 합을 수행해야 하므로 계산상 병목현상을 유발하게 된다.

IV. The Proposed Scheme

기본 CBOW 모델의 구조에서 발생하는 세 가지 병목현상을 해결하기 위해 다음과 같은 개선 방안을 적용한다[13].

4.1 Adaption of the Embedding Layer

기존 CBOW 모델에서 각 단어를 원 핫 벡터로 변환한 후 가중치 행렬과 곱하는 실제 계산 방식은 연산 비용이 많이 들고 메모리도 많이 소모된다. 원 핫 벡터는 단어의 위치를 표시함으로, 가중치 행렬곱과의 실제 계산은 불필요하다. 원 핫 벡터와 가중치 행렬의 행렬곱을 대체하여 특정 단어에 해당하는 가중치 행렬의 특정 행을 즉시 가져오는 방식이 Embedding이며, 이것을 구현하는 계층을 Embedding 계층이라고 한다[14]. 그러므로 Embedding 계층을 도입하면 실제 행렬곱 연산을 대체하는 효과를 볼 수 있으므로 병목현상이 해결된다. 이는 메모리와 계산량을 절감한다. 이를 통해 입력층-은닉층의 병목현상을 해결한다. 역전파를 포함하는 Embedding 계층 코드는 아래와 같다.

```
class Embedding:
    def __init__(self, W):
        self.params = [W]
        self.grads = [np.zeros_like(W)]
        self.idx = None
    def forward(self, idx):
        W, = self.params
        self.idx = idx
        out = W[idx]
        return out
    def backward(self, dout):
        dW, = self.grads
        dW[...] = 0
        if GPU:
            import cupy
            cupy.scatter_add(dW, self.idx, dout)
        else:
            np.add.at(dW, self.idx, dout)
        return None
```

code. 1

4.2 Negative Sampling

CBOW 기본모델의 계산상 병목현상인 은닉층 뉴런과 가중치 행렬(*W*out)의 곱과 Softmax 계층의 계산을 해결하기 위해 Negative Sampling이 도입된다. Negative Sampling은 다중 분류를 이진 분류로 근사화하는 방식이며, 실제 단어 쌍(정답)과 무작위로 샘플링된 단어 쌍(오답)을 통해 모델을 학습시킨다. 학습 시 정답 쌍에 대해서는 1에 가깝게, 오답 쌍에 대해서는 0에 가깝게 예측하도록 한다. 이를 통해 은닉층-출력층의 병목현상과 Softmax-with-Loss 계층의 병목현상을 해결한다. 즉, Sigmoid-with-Loss 계층을 사용하여 이진 분류로 근사함으로 병목현상을 제거한다[1].

다음 그림은 Negative Sampling을 도식화한 Computation graph를 보여준다.

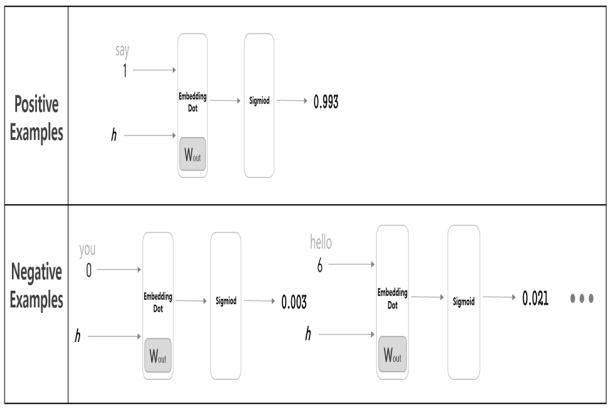

Fig. 3. Computational Negative Sampling Diagram

위의 그림Ⅳ에서 보듯이, Negative Sampling은 Embedding Dot와 Sigmoid with Loss 계층으로 구성되어 역전파를 포함한 실제적인 학습이 이루어진다. 그리고 Embedding Dot 계층은 Embedding 연산과 내적으로 구성되어 있다. 아래의 코드는 Embedding Dot 계층의 코드 구현이다.

```
def forward(self, h, idx):
    target_W = self.embed.forward(idx)
    out = np.sum(target_W * h, axis=1)
    self.cache = (h, target_W)
    return out
```

code. 2

자세히 살펴보면, Embedding Dot 계층에서 Embedding 연산은 target_W로서 구현되며 원 핫 벡터로 구성된 필요한 단어 ID의 집합인 idx에 해당하는 출력 가중치 행렬 W_{out}에서 Embedding한 정보를 추출한다. 그리고 이렇게 추출된 정보 target_W와 은닉층의 100개 뉴런으로 구성된 h와의 내적한 결과(out)를 반환값으로 얻는다.

여기서 target_W의 idx는 단어 ID의 NumPy 배열이며, 배열을 받는 이유는 데이터를 한꺼번에 처리하는 미니 배치 처리를 가정했기 때문이다. 다음으로, Embedding Dot 계층을 통해 처리된 결과값은 활성화 함수 Sigmoid 함수를 사용하여 예측값을 얻는다. 이는 Softmax 함수의 역할을 대체하고, 위에서 언급한 Computational Graph를 완

성한다.

참고로 활성화 함수로 Sigmoid 함수를 사용할 때도 Softmax와 마찬가지로 손실 함수로서 일반적으로 Cross-Entropy 손실 L을 사용한다:

$$L = -(t\log y + (1-t)\log(1-y))$$

Equation. 1. Cross-Entropy Loss Function L

각 미니 배치에 대한 평가는 eval_inetrval = 20으로 설정되어 있으며 손실은 avg_loss = total_loss / loss_count로 평가한다. 개선된 CBOW 모델의 Computational Graph 다음과 같다[13].

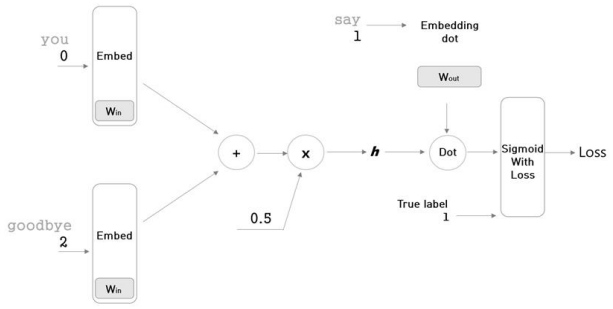

Fig. 4. Computational CBOW Diagram equipped with Negative sampling

4.3 CBOW Model with Batch Normalization

배치정규화(Batch Normalization 또는 BN)는 Ioffe and Szegedy(2015)에 의해 도입된 기법으로, 심층 신경망 훈련을 개선하여 내부 공변량 변화를 줄이는 역할을 한다. 내부 공변량 변화란 훈련 중 파라미터가 업데이트됨에 따라 신경망 활성화의 분포가 변하는 것을 의미하며, 배치정규화는 각 층의 입력을 미니 배치 단위로 정규화 함으로써 이 문제를 해결한다[16]. 배치정규화를 적용하여 성능이 개선된 재검증 논문은 아래의 표에 나타난다[16, 17, 18, 19, 20].

Table 2. Test accuracy in previous studies

	LeNet-5	Ioffe's ANN with BN	ANN with BN	CNN with BN (2018)	CNN with BN (2024)
Author /Year	LeCun et al. /1998	Ioffe & Szegedy /2015	Lee & Rhee /2023	Ji, Chun &Kim /2018	Lee & Rhee /2024
optimizer	SGD	SGD	Adam	-	Momentum
test accuracy	99.05%	95.2%	96.47%	99%	99.2%

다음의 수식은 배치정규화의 수리적 표현이다. 방정식의 왼쪽 항들은 각각 미니 배치의 평균, 분산, 정규화된 데이터를 정의한다.

$$\mu_B = \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$\sigma_B^2 = \frac{1}{m}\sum_{i=1}^{m}(x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \varepsilon}}$$

Equation. 2. Batch Normalization Mathematical Formula

배치정규화를 적용한 CBOW 모델의 Computational Graph는 다음 그림과 같다.

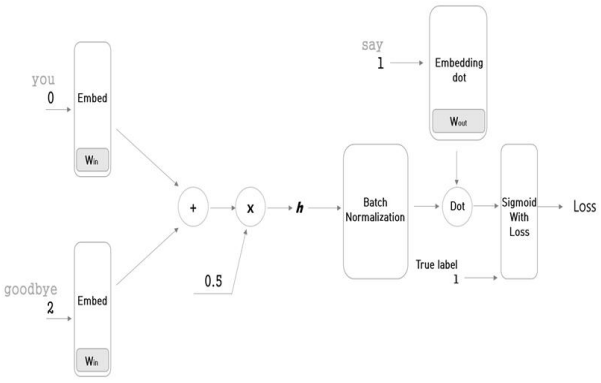

Fig. 5. New Computational CBOW Diagram with Batch Normalization

그림과 같이 입력 가중치가 적용된 결과는 중간 은닉층 벡터 h로 전달되고, 이후에 배치정규화 계층이 추가된다. 다음은 배치정규화를 적용한 CBOW 모델의 구현 코드이다.

```
class CBOW:
    def __init__(self, vocab_size, hidden_size, window_size, corpus):
        self.bn = BatchNormalization()
    def forward(self, contexts, target, train_flg=True):
        h = self.bn.forward(h, train_flg)
    def backward(self, dout=1):
        dout = self.bn.backward(dout)
```

code. 3

미리 역전파를 포함한 수정된 BatchNormalization()[16] 함수를 가져와 저장하고, init()함수, 순전파함수와 역잔파 함수에 배치정규화 코드를 추가하였다.

V. Model Training

5.1 Hyper parameter setting

Optimizer : 모델의 최적화 학습을 위해 Adam을 사용한다. 본 연구에서는 Adam을 기본 Optimizer로 설정하여, 학습 중 손실 값이 더욱 안정적으로 줄어드는 것을 확인하였다.

Learning Rate (lr) : 학습률은 모델의 수렴 속도와 정확성에 큰 영향을 미치므로 최적의 값을 찾기 위해 다양한 학습률을 실험했다. 실험의 결과로서 초기 학습률은 0.001로 설정하였다.

Epoch : 본 모델은 20 epochs에 걸쳐 학습을 진행하였다. 초기 몇 epochs에서는 손실 값이 빠르게 감소하는 것을 확인하였고, 이후 일정 수준에서 수렴하는 경향을 보였다. 추가 실험으로 40 epochs까지 학습을 진행해본 결과, 손실 값이 약간 더 줄어들기는 했으나, 전반적으로 20 epochs와 유사한 수준에서 수렴하였다. 그러므로 본 논문에서는 계산의 시간을 단축시키는 목적 하에 20 epochs으로 학습 횟수를 제한하였다. 참고로 미니 배치의 Size는 100으로 한정하였다.

5.2 Hardware setup

Cupy 활용 및 학습 시간: 학습 속도 향상을 위해 Cupy를 사용하여 GPU 연산을 수행한다. Cupy는 NumPy와 유사한 API를 제공하며, GPU에서 빠른 병렬 연산을 지원하여 모델 파라미터의 업데이트를 가속화할 수 있다. 이를 통해 학습 시간을 크게 단축하였으며, 특히 대용량의 말뭉치를 학습해야 하는 언어모델에서는 CPU 대비 10배 이상 빠르게 학습할 수 있는 이점이 있다. 본 논문의 실험 시간은 대략 40분 정도로 나타냈다.

VI. Experimental Results

본 실험에서는 Window Size = 5인 CBOW 모델을 구축하고, 배치정규화를 적용한 경우와 적용하지 않은 경우의 성능을 비교하였다. 실험 결과, 배치정규화를 적용한 모델은 평균 손실률이 1.25서 0.65로 약 50% 감소하는 것을 확인하였다. 이는 배치정규화가 CBOW 모델 학습 과정에서 발생하는 내부 공변량 이동 문제를 해결하고, 학습 안정성을 향상시켜 손실률 감소에 기여했음을 의미한다.

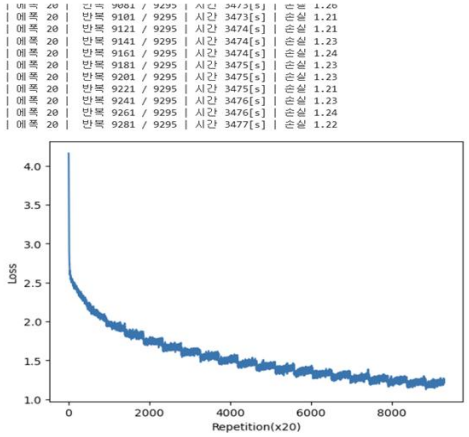

Fig. 6. CBOW Model Results without Batch Normalization

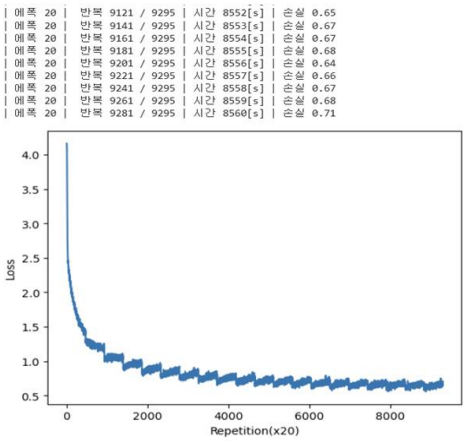

Fig. 7. CBOW Model Results with Batch Normalization

VII. Evaluation

Word2Vec의 CBOW와 같은 단어 임베딩 모델은 주로 단어의 유사도, 유추 관계, 의미관계 분석 등을 통해 성능과 실용성을 검증한다. 예를 들어, "you"라는 단어와 유사한 단어로 "i"나 "we"가 도출되고, "year"에 대해 "month"와 "week" 같은 관련 단어들이 나타나는 것은, CBOW 모델이 문맥을 효과적으로 학습하며 단어 간의 유사성을 잘 반영한다는 것을 보여준다[1, 13].

따라서 본 연구에서는 Batch normalization 을 사용하면 CBOW의 손실값(training loss)이 줄어든다는 것을 실험을 통하여 보였다.

VIII. Conclusions

본 논문에서는 기본 CBOW 모델에 배치정규화를 적용하여 평균 손실값을 개선했다. 실험을 통해 BN이 CBOW 모델의 평균 손실값 개선에 효과적임을 검증하였다. 즉, 평균 손실값이 1.25에서 0.65로 약 50% 감소하는 것으로 확인되었다. 서론에서 언급한 것처럼, CBOW를 포함한 Word2Vec의 손실값을 개선하는 방법을 찾아내는 것은 전이학습을 위한 도구로서 중요성을 시사하는 것이라고 할 수 있다. 그러나 배치정규화는 미니 배치 크기에 의존적으로 성능이 달라 질 수 있으며, 이는 한계점으로 대두된다. 너무 작은 배치 크기에서는 분산이 0에 가까워져 효과가 감소 할 수 있으며 훈련에 영향을 미칠 수 있다.

배치정규화는 실험적으로 효과가 입증된 것일 뿐, 수학적으로 증명된 것은 아니다. Ioffe의 논문에서는 배치정규화가 효과적일 것이라는 heuristic한 주장이 있지만, 이를 수학적으로 증명한 것은 아니다. 특정 조건에서는 배치정규화가 효과를 보일 수 있으나, Skip-Gram 모델에서는 실험적으로 유의미한 결과를 얻지 못하였다.

향후 연구 계획으로는 CBOW를 전이학습으로 적용한 신경망(CNN, RNN, LSTM 등 포함)에서 배치정규화의 효과를 정밀하게 연구 할 것이다. 이러한 연구 방향을 통해 Word2Vec과 유사한 다양한 Embedding 모델에서 평균 손실값의 개선효과를 줄 수 있는 배치정규화와 같은 또 다른 기법들을 탐구할 것이다.

ACKNOWLEDGEMENT

“This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ICAN(ICT Challenge and Advanced Network of HRD) program(IITP-2024-2020-0-01825) supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation)”

REFERENCES

[1] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," Proceedings of the 1st International Conference on Learning Representations (ICLR), 2013. DOI: 10.48550/arXiv.1301.3781

[2] Y. Feng and C. Hu, "A Simple and Effective Usage of Word Clusters for CBOW Model," Proceedings of the International Conference on Natural Language Processing (NLP), 2022. DOI: 2022arXiv220705801F

[3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. In Advances in neural information processing systems (pp. 3111-3119). DOI : 10.48550/arXiv.1310.4546

[4] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," Advances in Neural Information Processing Systems (NeurIPS), pp. 3111-3119, 2013. DOI: 10.48550/arXiv.1310.4546

[5] Unified Patents, "Patent US-9740680-B1," Unified Patents Portal, https://portal.unifiedpatents.com/patents/patent/US-9740680-B1 (2024-11-19 방문).

[6] S. Choi and M. Jang, "Optimizing Word2Vec for Korean Language Processing," Proceedings of the Korea Language Processing Conference, pp. 45–50, Seoul, Korea, November 2016.

[7] L. Zhang and H. Wu, "Word Embedding Techniques for Multilingual Text Classification," IEEE Transactions on Knowledge and Data Engineering, Vol. 34, No. 5, pp. 1234–1245, May 2022. DOI: 10.1109/TKDE.2021.1234567.

[8] J. Chang, M. W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," arXiv preprint, 2018. DOI: arXiv:1810.04805

[9] J. J. von der Mosel, A. Trautsch, and S. Herbold, "On the validity of pre-trained transformers for natural language processing in the software engineering domain," Proceedings of the International Conference on Software Engineering (ICSE), 2021. DOI: 10.1145/3324884.3416532

[10] J. Lee, "Informal Quality Data Analysis via Sentimental Analysis and Word2vec Method," Journal of Korean Society for Quality Management, vol. 45, no. 1, pp. 117-123, 2017. DOI: 10.7469/JKSQM.2017.45.1.117.

[11] Y. Kim, "CBOW-based Sentiment Analysis for Social Media Data," Journal of Computational Linguistics, Vol. 12, No. 3, pp. 123-130, 2021. DOI: 10.1007/s12652-018-1095-6.

[12] J. Lee and S. Park, "A Study on Analyzing Keyword-Centric Social Network Data Using CBOW and Skip-gram," Journal of Korean Computational Studies, Vol. 18, No. 2, pp. 45-57, 2020. DOI: 10.1016/j.jks.2020.07.003

[13] Saito Koki, "Deep Learning from Scratch ②," KADOKAWA, Tokyo, 2018.

[14] Y. Feng and C. Hu, "A Simple and Effective Usage of Word Clusters for CBOW Model," arXiv preprint arXiv:2207.05801, 2022. DOI: 10.48550/arXiv.2207.05801.

[15] J. H. Kim, "Word Embedding-based Research Paper Classification

Technique," Proceedings of the KIPS Conference, pp. 123-130, Seoul, Korea, March 2021.

[16] S. Ioffe and C. Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift”, Proceedings of the International Conference on Machine Learning, Proceedings of Machine Learning Research, vol. 37, pp. 448-456, Feb. 2015, DOI: https://doi.org/10.48550/arXiv.1502.03167

[17] S. Lee, "Improving Test Accuracy on the MNIST Dataset using a Simple CNN with Batch Normalization," Proceedings of the International Conference on Machine Learning (ICML), 2024. DOI: 10.9708/jksci.2024.29.09.001

[18] S. B. Lee and J. S. Rhee, “Applying Batch Normalization to the MNIST Dataset”, Quantitative Bio-Science, vol. 42, no. 2, pp. 133-137, Nov. 2023, DOI: http://doi.org/10.22283/qbs.2023.42.2.133

[19] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition”, Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.

[20] M. G. Ji, J. C. Chun, and N. G. Kim, “An Improved Image Classification Using Batch Normalization and CNN”, Journal of Internet Computing and Services, vol.19, no.3, pp. 35-42, June. 2018, DOI : 10.7472/jksii.20

Authors

Gukbeom Yoon received the B.S. degree in Smart Convergence Security from Busan University of Foreign Studies, Korea, in 2023. Gukbeom Yoon entered the graduate school at Busan University of Foreign Studies in 2023 as a Master's student. His research interests include artificial intelligence and information security.

Jungsoo Rhee received the B.S. degree in Mathematics Education from Kyungpook National University, Korea, in February 1982, the M.S. degree in Mathematics from Kyungpook National University, Korea, in February 1984, and the Ph.D. degree in Mathematics from Florida State University, USA, in August 1993. Dr. Rhee joined the faculty of the Department of Smart Convergence Security at Busan University of Foreign Studies, Korea, in 1994. He is currently a Professor in the Department of Smart Convergence Security at Busan University of Foreign Studies. His research interests include AI/Cybersecurity, Quantum information science, Cryptography, and Fourier Analysis.