

재귀중첩신경망을 리용하여 조선어문장을 분류하기 위한 한가지 방법

심천룡, 리명철

우리는 재귀중첩신경망(Recurrent Convolutional Neural Network)의 구조를 변경하여 조선어문장을 분류하기 위한 방법을 제기하고 실험을 통하여 그 효과성을 검증하였다.

선행연구[1]에서는 LSTM(Long-Short Term Memory)에 기초한 모형을 리용하여 단어의 순서정보를 얻어내는 방법으로 분류와 의미해석성능을 높였으며 선행연구[2]에서는 중첩신경망(CNN)을 리용하여 단어들사이의 국부적관계와 개별적인 단어의 특성을 얻어내었으며 계산의 효율성을 보장하였다. 그러나 선행연구[1]에서 제안한 모형은 문장이 길어질수록 단어들사이의 호상순서관계를 학습하기 힘들어하는 결함이 있으며 선행연구[2]의 모형은 단어들의 순서관계를 효과적으로 반영할수 없는 결함을 가지고있다.

본문에서는 이러한 결함들을 극복하고 조선어문장을 분류하기 위하여 재귀중첩신경망[3]을 리용하였다. 조선어문장의 합리적인 표현을 얻기 위하여 재귀중첩신경망의 구조를 변경하였으며 실험을 통하여 조선어문장분류에서 제안한 망이 종전의 재귀중첩신경망[4], 1차원중첩신경망[2], LSTM[1]보다 분류정확도를 높일수 있다는것을 확인하였다.

1. 조선어문장분류를 위한 망의 구조

일반적으로 본문분류에 재귀중첩신경망을 리용할 때 입력값은 (w_1, w_2, \dots, w_n) 과 같은 하나의 문장이다.

이때 재귀중첩신경망에서는 문장의 단어들에 해당하는 벡토르표현을 쌍방향재귀신경망을 통과한 문맥의 벡토르표현과 련결하고 1차원중첩층을 통과시킨다.

1차원중첩층의 출력값에 최대선택층을 적용하여 해당 문장의 최종적인 벡토르표현을 얻고 전결합층을 거쳐 분류를 진행한다.

본문에서는 길이가 긴 조선어문장들의 표현도 효과적으로 얻을수 있도록 우의 재귀중첩신경망의 구조에서 쌍방향일반재귀신경망을 긴 문장에서 단어순서관계를 학습할수 있는 쌍방향LSTM으로 교체하였으며 창문크기가 1, 3, 5, 9인 여러개의 중첩층들을 리용하여 각이한 창문크기안의 단어들의 순서관계를 동시에 학습할수 있게 하였다.

이때 중첩층들의 활성화함수로는 ReLU(Rectified Linear Unit)[2]를 리용하였다.

또한 최대선택층과 평균선택층을 모두 리용하였으며 효과성을 실험을 통하여 검증하였다.

본문에서 제안한 망의 구조는 그림 1과 같다.

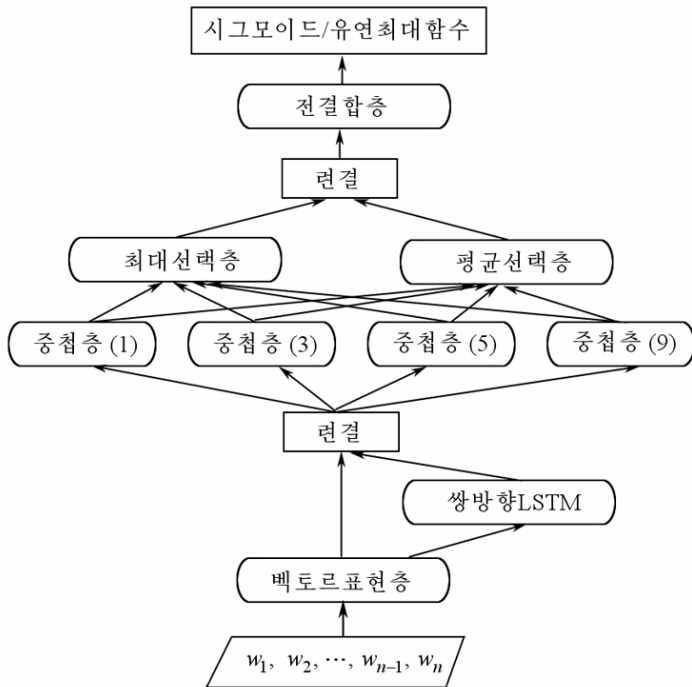


그림 1. 논문에서 제안한 망의 구조

논문에서 제안한 변경된 재귀중첩신경망의 층들은 표 1과 같다.

표 1. 논문에서 제안한 변경된 재귀중첩신경망의 층들

층이름	출력형태	훈련파라미터존재
입력층	[뭉음크기, 문장크기]	×
벡터표현층	[뭉음크기, 문장크기, 벡터표현차원수]	○
쌍방향LSTM	[뭉음크기, 문장크기, 2×LSTM출력차원수]	○
런결층(하단)	[뭉음크기, 문장크기, 벡터표현차원수+2×LSTM출력차원수]	×
중첩층 (1)	[뭉음크기, 문장크기, 중첩층출력차원수]	○
중첩층 (3)	[뭉음크기, 문장크기-2, 중첩층출력차원수]	○
중첩층 (5)	[뭉음크기, 문장크기-4, 중첩층출력차원수]	○
중첩층 (9)	[뭉음크기, 문장크기-8, 중첩층출력차원수]	○
선택층(최대, 평균)	[뭉음크기, 중첩층출력차원수]	×
런결(상단)	[뭉음크기, 중첩층출력차원수×8]	×
전결합층	[뭉음크기, 모임개수]	○

논문에서는 우선 일반적인 재귀신경망대신에 LSTM을 리용하여 길이가 긴 조선어문장에 대하여서도 보다 효과적인 표현을 얻을수 있게 하였다.

논문에서는 다음으로 중첩의 창문크기가 1이던 중첩층을 창문크기가 1, 3, 5, 9인 중첩층들로 교체하여 보다 넓은 범위의 국부적특징을 학습할수 있게 하였으며 활성화함수로 ReLU를 리용하여 그라디언트손실을 줄일수 있게 하였다.

그리고 마지막으로 중첩층의 출력에 최대선택층과 평균선택층을 동시에 적용하여 망의 표현능력을 높이였으며 최대선택층과 평균선택층을 모두 리용할 때와 최대선택층만을 리용할 때의 성능비교실험을 진행하였다.

론문에서 제안한 망의 흐름도는 그림 2와 같다.

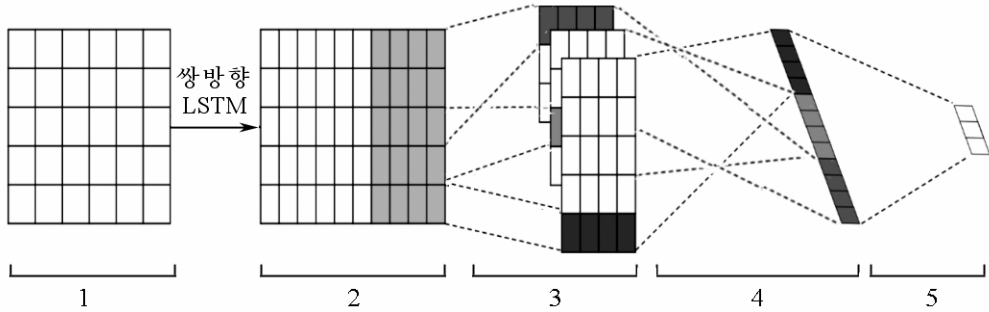


그림 2. 론문에서 제안한 망의 흐름도

- 1- 해당 문장의 표현(단어개수×벡터표현차원수), 2-쌍방향LSTM의 출력과 문장표현의 연결,
3-창문크기가 여러가지인 중첩층의 적용, 4-선택층의 적용과 그 출력들의 연결,
5-전결합층과 유연최대함수의 출력

론문에서는 Adam알고리즘[5]의 모멘트감쇠초파라미터 β_1 과 척도감쇠초파라미터 β_2 를 조절함으로써 모형이 빨리 수렴할수 있게 하여 학습속도를 높이었다.

Adam알고리즘은 다음과 같다.

- ① $m \leftarrow \beta_1 m + (1 - \beta_1) \nabla_{\theta} J(\theta)$
- ② $s \leftarrow \beta_2 s + (1 - \beta_2) \nabla_{\theta} J(\theta) \otimes \nabla_{\theta} J(\theta)$
- ③ $m' \leftarrow (m / (1 - \beta_1^T))$
- ④ $\hat{s} \leftarrow (s / (1 - \beta_2^T))$
- ⑤ $\theta \leftarrow \theta - \eta m' \otimes (\sqrt{\hat{s}} + \epsilon)^{-1}, \epsilon = 10^{-10}$

제안모형의 경우 RMSProp[5]보다 Adam알고리즘이 보다 좋은 훈련성능을 보여주었으며 β_1 을 기정값(0.9)보다 약간 크게 하여 모멘트의 영향을 증가시키고 β_2 를 기정값(0.999)보다 작게 하여 모형의 파라미터갱신속도를 증가시켰다. 즉 론문에서 제안한 모형의 파라미터공간은 그라디언트변화가 일정한 구간 다시말하여 평탄한 구간이 많았으며 이것을 해결하기 위하여 모멘트감쇠초파라미터의 값을 크게 하는 방식을 취하였다. 또한 론문에서 제안한 모형의 파라미터공간에서 파라미터들사이의 그라디언트변화률이 심하지 않으므로 척도감쇠초파라미터의 값을 기정값보다 크게 취하여 파라미터들사이의 변화률을 조절해주어 학습이 빨리 진행되도록 하였다.

2. 실험을 위한 자료기지와 실험결과

론문에서는 조선어문장을 《신문》(보도기사), 《소설》, 《교과서》의 3개 부류로 나누기 위한 망을 구성하고 훈련자료를 준비하였다. 신문부류에서는 로동신문, 평양신문에서 나오는 보도기사자료를 추출하여 리용하였으며 소설부류는 우리 나라 현대소설들을, 교과서부류는 여러가지 자연과학참고서와 교과서들의 본문자료를 리용하였다. 이때 전체 자료를 4 : 1의 비율로 나누어 모형의 훈련과 검사에 리용하였다.

학습은 벡터표현층부터 마지막층까지 모든 층을 일괄적으로 학습시키는 방법으로 진행하였다.

론문에서 훈련에 리용한 모형의 초파라미터들은 표 2와 같다.

최대선택층과 평균선택층을 모두 리용할 때와 최대선택층만을 리용할 때의 실험결과는 표 3과 같고 론문에서 여러가지 중첩층을 동시에 리용할 때의 실험결과는 표 4와 같다.

실험을 통하여 최대선택층과 평균선택층을 동시에 리용하여 모형의 정확도를 향상시킬 수 있다는 것을 검증하였다.

표 2. 훈련에 리용한 초파라미터

초파라미터이름	값
묶음크기	32
단어선택개수 및 벡토르표현의 차원수	100 000, 100
쌍방향LSTM출력차원수	128
중첩층의 출력차원수	128
중첩층의 창문크기	1, 3, 5, 9
Adam알고리즘의 초파라미터설정 (β_1, β_2)	0.95, 0.985

표 3. 최대선택층과 평균선택층을 모두 리용할 때의 실험결과

분류	정확도/%
최대선택층과 평균선택층의 리용	91.8
최대선택층의 리용	90.9

표 4. 여러가지 중첩층을 동시에 리용할 때의 실험결과

분류(창문크기)	정확도/%
1	89.85
9	90.82
1, 3, 5, 9	91.80

변경된 재귀중첩신경망과 종전의 재귀중첩신경망을 비롯한 여러 망들을 리용하여 학습시켜 비교한 결과는 표 5와 같다. 이때 평가는 검사자료기지에서 정확도로 진행하였다.

표 5. 론문에서 제안한 모형과 선행모형들의 비교

모형	정확도/%
변경된 재귀중첩신경망(제안모형)	91.80
재귀중첩신경망(종전의 모형 RCNN)	89.32
1차원중첩신경망(CNN)	88.57
LSTM	88.07

론문에서 제안한 모형을 실험을 통하여 검증함으로써 재귀중첩신경망의 구조를 변경하여 조선어문장분류에 리용하면 종전의 모형들에 비하여 좋은 결과를 얻을 수 있다는 것을 확인하였다.

참 고 문 헌

- [1] Fei Liu et al.; NAACL, 2, 278, 2018.
- [2] Rie Johnson et al.; ACL, 1, 562, 2017.
- [3] Siwei Lai et al.; AAAI, 333, 2267, 2015.
- [4] Peixiang Zhong et al.; arXiv:1902.07867v2, 2019.
- [5] S. J. Reddi et al.; arXiv:1904.09237, 2019.

주체109(2020)년 9월 5일 원고접수

A Method for Korean Sentence Classification Using Recurrent Convolutional Neural Network

Sim Chon Ryong, Ri Myong Chol

We propose a new network construction method for Korean sentence classification by modifying the recurrent convolutional neural network structure and verify the accuracy through experiments.

The results demonstrate that our new method is more effective for Korean sentence classification.

Keywords: Korean sentence classification, recurrent convolutional neural network