

순환 신경망 기반 언어 모델을 활용한 초등 영어 글쓰기 자동 평가

박영기

춘천교육대학교

요 약

작성된 문서의 문법적 오류 교정을 할 때 맞춤법 검사기를 사용하는 것이 일반적이다. 그러나 초등학생들이 작성한 글 중에는 문법적으로는 옳더라도 자연스럽지 않은 문장이 있을 수 있다. 본 논문에서는 동일한 의미를 가진 2개의 문장이 주어졌을 때, 어떤 것이 더 자연스러운 문장인지 자동 판별할 수 있는 방법을 소개한다. 이 방법은 순환 신경망(recurrent neural network)을 이용하여 장기 의존성(long-term dependencies) 문제를 해결하고, 보조 단어(subword)를 사용하여 희소 단어(rare word) 문제를 해결한다. 약 200만 문장의 단일어 코퍼스를 통해 순환 신경망 기반 언어 모델을 학습하였다. 그 결과, 초등학생들이 주로 틀리는 표현들과 그에 대응하는 올바른 표현을 입력으로 주었을 때, 모든 경우에 대해 자연스러운 표현을 자동으로 선별할 수 있었다. 본 소프트웨어가 스마트 기기에 사용될 수 있는 형태로 구현된다면 실제 초등학교 현장에서 활용 가능할 것으로 기대된다.

키워드 : 초등 글쓰기 자동 평가, 순환 신경망, 언어 모델

Automatic Evaluation of Elementary School English Writing Based on Recurrent Neural Network Language Model

Youngki Park

Department of Computer Education, Chuncheon National University of Education

ABSTRACT

We often use spellcheckers in order to correct the syntactic errors in our documents. However, these computer programs are not enough for elementary school students, because their sentences are not smooth even after correcting the syntactic errors in many cases. In this paper, we introduce an automated method for evaluating the smoothness of two synonymous sentences. This method uses a recurrent neural network to solve the problem of long-term dependencies and exploits subwords to cope with the rare word problem. We trained the recurrent neural network language model based on a monolingual corpus of about two million English sentences. In our experiments, the trained model successfully selected the more smooth sentences for all of nine types of test set. We expect that our approach will help in elementary school writing after being implemented as an application for smart devices.

Keywords : Elementary School Writing, Recurrent Neural Network, Language Model

논문투고 : 2016-10-24

논문심사 : 2016-10-26

심사완료 : 2017-01-31

1. 서론

초등 교육을 위해 ICT 기술을 활용하는 다양한 연구가 진행되고 있다. 원어민과 유사한 억양을 얻기 위해 TTS (Text-to-speech) 기술을 활용하거나 영어 학습 앱을 개발하기도 하고[6, 19], 센서를 통해 얻어진 로그를 활용하여 논리적 사고력을 향상시키기도 한다[11]. 이미 많은 초등 교사들이 스마트 기기 및 앱을 이용한 교육을 수행하고 있다는 점을 감안해 보면[21], 이와 같은 연구들은 빠른 시일 내에 실제로 적용될 가능성을 갖 것으로 예상된다.

본 논문에서는 컴퓨터를 이용해 초등학생들의 글쓰기 학습을 자동으로 도와줄 수 있는 보조 도구를 개발하는 것에 초점을 맞춘다¹⁾. 교사가 직접 초등학생들이 쓴 글을 보고 1:1 교정해 주는 방식[18, 22]으로 교육할 수 있다면 가장 좋겠지만, 그러기 위해서는 교사가 상당히 많은 시간을 투자해야 하기 때문에 이와 같은 자동화된 도구가 있다면 도움이 된다. 가장 대표적인 글쓰기 보조 프로그램은 맞춤법 검사기다. 그러나 이 프로그램들의 주목적은 문법적 오류를 교정하는 것이어서 초등학생들의 글에서 나타나는 다양한 오류들을 바로잡기에는 한계가 있다. 다시 말하면, 초등학생이 쓴 글에서 문법적 오류를 모두 수정하더라도 그 문장이 자연스러운 것이라 보장할 수는 없다. 예를 들어 다음과 같은 문장은 최신 한글 맞춤법 검사기²⁾로 검사했을 때 아무런 문제가 없지만, 자연스럽지 않은 문장임을 알 수 있다.

- 철수는 야구 좋아해 친구 하키를 좋아해

그렇다면 컴퓨터가 자동으로 어떤 문장이 자연스러운 문장인지를 검증할 수 있는가? 조금 더 문제를 단순화 하면, 만약 어떤 두 문장이 주어졌을 때, 두 문장 중 더 자연스러운 문장이 어떤 것인지 컴퓨터가 자동으로 분류할 수 있는가? 예를 들어 다음과 같은 두 개의 문장이 있을 때, 사람의 관점에서는 첫 번째 문장보다 두 번째 문장이 더 자연스럽다는 것을 쉽게 알 수 있다. 컴퓨터가 최신의 언어 모델 학습 방법을 활용한다면 이것을 잘할 수 있는지 검증하는 것이 본 논문의 목표이다.

- Tom likes baseball his friend likes hockey
- Tom likes baseball, but his friend likes hockey

논문의 구성은 다음과 같다. 2장에서는 관련 연구로써 ICT 기반 교육 방법을 소개하고, 통계 기반 언어 모델을 기술한다. 통계 기반 언어 모델은 인공지능 및 자연어 처리와 관련된 상용화 시스템에서 가장 많이 사용되고 있는 언어 모델이다. 3장에서는 순환 신경망 기반 언어 모델을 설명하는데, 최근 연구자들에 의해 많이 연구되고 있는 방법이다. 어휘의 개수가 많은 경우에는 메모리 문제가 발생하여 학습하기가 어려운데, 이때 보조단어(subword) 기술을 사용하여 문제를 해결하였다. 4장에서는 자연스럽게 읽은 문장과 자연스럽게 문장을 입력으로 주었을 때, 언어 모델이 자연스럽게 문장을 얼마나 잘 구분해 내는지 실험한다. 5장에서는 현 기술의 상용화 가능성 및 부족한 점, 향후 연구 방향에 대해 고찰한다.

2. 관련 연구

2.1절에서는 ICT 기반 교육 연구들을 소개하고, 2.2절에서는 본 연구의 기술적 배경이 되는 통계 기반 언어 모델을 설명한다.

2.1 ICT 기반 교육

많은 교사들이 스마트 기기 및 앱을 이용한 교육을 경험하거나 그 중요성을 인지하고 있다[21]. 특히 제4차 산업혁명 시대에는 발전된 SW 기술들을 활용한 새로운 교육 방법들이 주목받을 것으로 예상되며, 라즈베리 파이[9] 등을 통해 PC 보급이 활발해진다면 변화의 속도는 더욱 빨라질 것이다.

지금까지 제안된 ICT 기반 교육 방법들은 수학, 과학, 사회, 국어, 영어 등 다양한 과목을 대상으로 하였다. 예를 들어, 수학/과학 과목을 위해 ICT 콘텐츠를 만

1) 본 논문은 2016 한국정보교육학회 하계 학술대회에서 기 발표된 내용을 보완/확장하여 작성되었음

2) <http://speller.cs.pusan.ac.kr/>

든 사례[15], 데이터 로깅 로봇을 이용한 과학 교육[11], 사회 교육을 위한 교수-학습용 시스템[12], TTS (Text-to-speech) 기술을 이용한 영어 교육[6, 19], 영어/국어 교육을 위한 객관식 빈칸 채우기 문제 생성 및 유사 문장 자동 선별[16, 17] 등이 제안되었다. 기초학력을 진단, 보정하는 웹 기반 시스템[8], 진로교육을 위한 웹 기반 인터페이스[13] 등 특정 과목과 직접적인 관련은 없지만 교육적 효과가 있는 사례도 제시되었다.

ICT 기반 교육의 필요성은 인지하지만, 모든 과목에 적합한 것은 아니라고 생각하는 교사들도 있다. Jang과 Chun의 연구 결과에 따르면, 70%의 교사들이 앱을 활용하기에 적합한 과목으로 사회/도덕을 선택한 반면, 영어를 선택한 교사는 46%, 국어를 선택한 교사는 24%에 불과했다[7]. 그러나 이것은 당시에 영어/국어와 같은 언어 과목의 학습을 보조하기 위한 SW가 부재했던 것이 원인일 수 있다. 실제로 교사들이 활용했던 앱 중 사회 과목과 직접적으로 관련된 것이 7개였던 반면 영어, 국어와 직접적으로 관련된 것은 각각 0개, 1개로 나타났다.

2.2 통계 기반 언어 모델

가장 대표적인 통계 기반 언어 모델은 N-gram 모델이다. N-gram 모델에서 어떤 문장이 나타날 확률은, 그 문장에 있는 각 단어가 나타날 확률의 곱으로 계산된다. 이때 각 단어가 나타날 확률은 그 이전에 나타난 N개의 단어에 기반하여 계산된다. 즉, 어떤 문장이 T개의 단어 w_1, w_2, \dots, w_T 로 구성된다고 하자. N-gram 모델은 이 문장이 나타날 확률을 다음과 같이 계산한다.

$$p(w_1, w_2, \dots, w_T) = \prod_{t=1}^T p(w_t | w_{t-(N-1)}, w_{t-(N-2)}, \dots, w_{t-1})$$

예를 들어, 'I saw the red house'라는 문장의 확률은 3-gram Model을 사용하여 아래와 같이 계산될 수 있다³⁾.

$$P("I saw the red house") = P(I | \langle start \rangle) * P(saw | \langle start \rangle, I) *$$

$$P(the | I, saw) * P(red | saw, the) * P(house | the, red) * P(\langle end \rangle | red house)$$

$P(the | I, saw)$ 와 같은 각각의 n-gram 확률 값들은, 우리가 가지고 있는 말뭉치를 이용하여 아래와 같이 계산할 수 있다.

$$P(w_t | w_{t-(N-1)}, w_{t-(N-2)}, \dots, w_{t-1}) = \frac{freq(w_{t-(N-1)}, w_{t-(N-2)}, \dots, w_{t-1}, w_t)}{freq(w_{t-(N-1)}, w_{t-(N-2)}, \dots, w_{t-1})}$$

여기서 $freq(w_1, w_2, \dots, w_t)$ 는 말뭉치에서 순열 w_1, w_2, \dots, w_t 이 나타난 빈도 수를 의미한다. 예를 들어, 'I saw'라는 순열이 우리가 가지고 있는 말뭉치에 10번 나타났고, 'I saw the'라는 순열이 3번 나타났다고 하자. 그러면 $P(the | I, saw)$ 는 0.3으로 계산될 수 있다.

N-gram 모델은 학계와 산업계에서 널리 사용되고 있는 방법이지만, 글쓰기 교정 목적으로 활용하기에는 3가지 면에서 부적절하다. 그 중 2가지는 일반적인 N-gram 모델의 단점으로 지적되고 있는 것으로, 데이터 희소성(data sparsity) 문제와 일반화 능력의 부족(lack of generalization) 문제다[2]. 데이터 희소성 문제는 우리가 가진 말뭉치가 충분히 많지 않기 때문에 발생한다. 예를 들어 'I saw the'라는 순열이 우리가 가진 말뭉치에 없다고 가정해 보자. 그러면 위 3-gram 확률 값을 계산할 수 없게 될 것이다. 만약 말뭉치가 충분히 크다면 이 문제를 해결할 수 있지만, 우리가 사용하는 어휘의 개수는 매우 크기 때문에 큰 말뭉치를 마련하더라도 확률 값을 계산하지 못하는 경우가 자주 발생한다. 예를 들어 Google에서 대규모 말뭉치를 이용하여 만든 N-gram의 경우에도⁴⁾, 'saw the red house' 순열에 대한 N-gram 확률 값을 계산하지 못한다. N-gram의 두 번째 문제는 일반화 능력이 부족하다는 것이다. 예를 들어 우리가 가진 말뭉치에 'chases a dog', 'chases a cat', 'chases a rabbit'이라는 표현이 있다고 가정하자. 만약 사람에게 'chases a llama' 순열이 나타날 확률을 계산하라고 한다면, 우리가 가진 말뭉치에 'chases a 동물'이라는 형태의 표현이 많으므로 이 확률 값을 높게 매길 것이다. 그러나 N-gram 모델은, 만약 'chases a llama'

3) https://en.wikipedia.org/wiki/Language_model

4) <https://books.google.com/ngrams>

라는 표현이 우리 말뭉치에 매우 적게 나타난다면 이 확률 값을 아주 낮게 매길 것이다. N-gram의 3번째 문제는 장기 의존성(long-term dependencies) [5]를 고려하지 못한다는 점이다. 이는 문장의 자연스러움을 판단하고자 했을 때 치명적인 단점이 될 수 있다. 예를 들어 서론에서 예를 든 ‘Tom likes baseball, but his friend likes hockey’ 문장이 나타날 확률을 4-gram 모델을 이용하여 계산한다고 하자. 이때 단어 ‘his’의 확률 값을 계산할 때는 ‘likes’, ‘baseball,’ ‘but’만 활용하고, ‘his’와 ‘her’이 나타날 확률을 결정하는 데에 중요한 역할을 하는 ‘Tom’은 고려하지 않는다. 이 문제를 해결하기 위해서는 N의 값을 키우면 되지만, N의 값이 커질수록 N-gram의 첫 번째 문제인 데이터 희소성 문제가 더 심각해지기 때문에 근본적인 문제 해결은 매우 어렵다.

3. 신경망 기반 언어 모델

3.1절에서는 N-gram 모델의 단점을 극복하기 위한 순환 신경망(recurrent neural network) 기반 언어 모델을 설명하고, 3.2절에서는 순환 신경망에서 사용되는 단어의 수가 많아지는 경우 발생하는 문제를 해결하기 위한 보조 단어 기반 데이터 전처리 기법을 소개한다.

3.1 순환 신경망을 활용한 언어 모델

본 논문에서는 N-gram 모델의 단점을 극복하기 위해 순환 신경망을 활용한 언어 모델[4]을 사용한다. 인코더-

디코더 아키텍처[5]를 단순화한 신경망이라고도 볼 수 있다. 어떤 문장이 T개의 단어 (w_1, w_2, \dots, w_T)로 구성되었다고 하자. 이 단어들은 각각 어휘 사전의 크기(모든 가능한 단어의 개수) |V|만큼의 차원을 가지는 원-핫 벡터(one-hot vector)로 표현된 후, 신경망의 파라미터 행렬에 곱해져 임베딩 벡터(embedding vector) $s_t = f_t(w_t)$ 로 변환된다. 이때 임베딩 벡터의 차원은 원-핫 벡터의 차원보다 일반적으로 훨씬 작은 값이다. 각 s_t 값들은 다음과 같이 순환 신경망의 입력 값으로 사용된다.

$$h_t = \Phi(h_{t-1}, s_t)$$

여기서 h_0 의 값은 모두 0의 값을 가지는 벡터다.

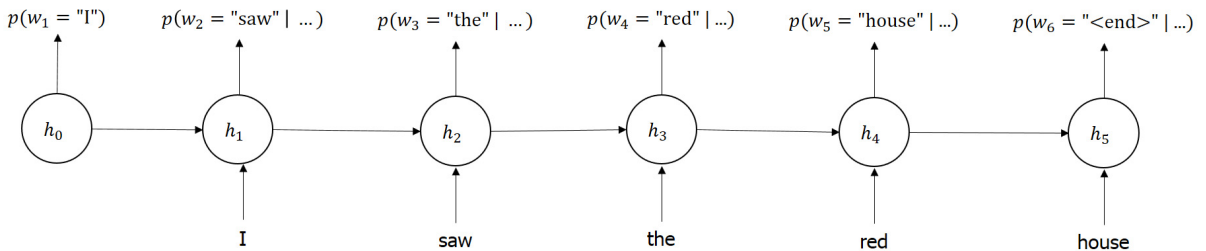
순환 신경망의 h_t 를 구하고 나면, 다음과 같이 t번째 단어 w_t 가 w 일 확률을 다음과 같이 계산할 수 있다.

$$p(w_t = w | w_{t-1}, w_{t-2}, \dots, w_1) = \frac{g_w(h_{t-1})}{\sum_{i=1}^{|V|} g_i(h_{t-1})}$$

예를 들어, N-gram 모델에서 예로 설명한 ‘I saw the red house’가 나타날 확률을 계산하는 순환 신경망은 (Fig. 1)과 같이 나타낼 수 있다.

학습을 위한 목적 함수는 다음과 같이 정의된다. 먼저, 어떤 문장 (w_1, w_2, \dots, w_T)의 로그 확률 값(log-probability)은 다음과 같이 정의된다.

$$\log p(w_1, w_2, \dots, w_T) = \sum_{t=1}^T \log p(w_t | w_{t-1}, w_{t-2}, \dots, w_1)$$



(Fig. 1) A Recurrent Neural Network Based Language Model

5) <https://devblogs.nvidia.com/parallelforall/introduction-neural-machine-translation-gpus-part-2/>

4-gram Model

"Tom likes baseball , but his friend likes hockey"



Recurrent Neural Network based Language Model

"Tom likes baseball , but his friend likes hockey"



(Fig. 2) A Comparison of N-gram and Recurrent Neural Network Based Language Models

M개의 학습 문장이 주어졌다고 할 때, 각 학습 문장의 로그 확률 값을 위와 같이 계산할 수 있다. 이 확률 값들의 평균이 최대화되도록 하는 파라미터 값을 찾는 것이 신경망의 목표 함수가 된다.

위와 같이 순환 신경망을 이용한 언어 모델은 N-gram 모델이 가지고 있는 3가지 단점을 모두 보완한다. 첫 번째로, 4-gram 모델을 사용하면 'saw the red house'에 대한 확률 값을 계산해 내지 못했는데, 순환 신경망을 사용하면 확률 값을 계산해 낼 수 있다. 두 번째로, 만약 우리 말뭉치에 'chases a llama'라는 표현이 매우 적게 나타난다면, N-gram 모델에서는 이 확률 값을 매우 낮게 계산할 것이다. 그러나 순환 신경망 기반 언어 모델에서는 만약 'chases a llama'라는 표현은 많지 않지만 'chases a dog', 'chases a cat', 'chases a rabbit'이라는 표현이 많이 있다면 'chases a llama'의 확률도 높게 계산할 가능성이 높다. dog, cat, rabbit, llama 등은 모두 동물이기에 비슷한 임베딩 벡터를 가지게 될 것이기 때문이다. 세 번째로, (Fig. 2)에 나타난 바와 같이 N-gram 모델에서는 이전 (N-1) 단어만 고려하여 확률 값을 계산하지만, 순환 신경망 기반 언어 모델은 모든 이전 단어를 고려한다는 장점이 있다.

3.2 보조 단어를 이용한 데이터 전처리

3.1절에서 사용한 순환 신경망을 이용한 언어 모델은 N-gram 모델의 단점들을 해결했지만, 몇 가지 개선될 여지가 있다. 첫 번째로, 우리가 쓰는 단어의 개수는 최소한 10만개 이상이기 때문에 신경망의 크기(파라미터 수)가 상당히 커지게 된다. 언어에 따라서는 어휘의 개수가 100만개 이상이 되는 경우도 있기 때문에, 그런 경우에는 GPU 메모리에 신경망을 모두 불러오기 어렵고 학습 속도도 매우 느려지게 된다. 두 번째로, 말뭉치에

존재하지 않는 완전히 새로운 단어가 나타나게 될 경우 확률 값을 계산할 수 없다. 예를 들어 'lower'는 말뭉치에 있지만 'lowest'는 말뭉치에 없을 경우, 'lowest'에 대한 확률 값을 계산하기 어렵다. 또, '104382'와 같은 큰 수는 각각의 숫자 단위로 나누지 않는 이상 말뭉치에 없는 단어일 가능성이 높다.

위 2가지 문제를 해결하기 위해 별도의 전처리 과정을 추가로 수행한다. 단어(word)를 보조 단어(subword)로 나누는 것인데[20], 이 방법에 따르면 만약 어떤 단어가 말뭉치에 자주 나타나지 않는 단어일 경우 2개 이상의 보조 단어로 나누어질 수 있다. 예를 들어, 'lowest'라는 단어가 만약 우리가 가진 말뭉치에 자주 나타나지 않는다면, 'lowest'는 사전에 추가하지 않고 'low', 'est'와 같은 보조 단어만 사전에 추가한다. 마찬가지로 '104382'와 같은 큰 수는 사전에 없을 수 있으나, '1', '0', '4', '3', '8', '2'와 같은 보조 단어로 나누어서 표현한다면 문제를 해결할 수 있다. 단, 이때 각 숫자는 사전에 있어야 한다. 만약 'lowest'와 '104382'가 우리가 가진 말뭉치에서 자주 나타나는 단어라면, 보조 단어로 나누어지지 않고 어휘 사전에 추가될 것이다. 4장에서는 보조 단어 변환 기술을 사용하여 어휘 사전의 크기를 3만개 수준으로 줄여 실험한 결과를 나타내었다.

4. 실험

4.1 실험 환경

4.1.1 학습 데이터

학습을 위해 Europarl 뉴스 데이터[10]를 사용하였다. 해당 데이터는 병렬 코퍼스를 제공하지만, 언어 모델을

만들 때에는 단일어 코퍼스만 사용하였다. 본 실험에서는 200만 문장의 영어 단일어 코퍼스만 사용되었으나⁶⁾, 더 많은 데이터를 활용하였을 때 성능이 어떻게 달라지는지를 실험해 보지는 않았다. 현재 공개되어 있는 영어 단일어 코퍼스의 양은 매우 방대하며, WMT 2016에서 공개한 영어 단일어 코퍼스(100GB 이상)는 현재 학습한 데이터(약 0.1GB)보다 1000배 이상 많은 수준이다. 일반적으로 데이터가 많으면 많을수록 성능이 좋아지기 때문에 더 많고 다양한 단일어 코퍼스로 학습한다면 성능이 지금보다 개선될 것으로 기대하며, 이는 추후 연구로 남긴다. 영어 단일어 코퍼스만 사용하였으므로 영어 모델만 학습하게 되지만, 다른 언어의 모델도 동일한 방식으로 학습이 가능하다.

데이터는 학습되기 전 3.2절에서 설명한 보조 단어로 변환된다. 이에 따라 테스트가 끝난 후에는 보조 단어를 단어로 바꾸어 주는 후처리 작업을 수행해야 한다.

4.1.2 학습 방법

Deep Learning 프레임워크인 Theano를 활용하였고, DL4MT-material⁷⁾에 구현되어 있는 GRU 기반의 순환 신경망을 사용하여 언어 모델을 학습하였다. 어휘의 개수 $|V|$ 는 약 30000이며, 임베딩 벡터의 차원의 수는 500, 순환 신경망의 차원의 수는 1000이다. 단일 GPU를 이용하여 학습하였고 3일 후 수렴하였다.

4.1.3 테스트 데이터

영어 모델을 학습하였으므로 영어 데이터로 테스트할 필요가 있다. 영어권 국가의 초등학생들이 글쓰기를 할 때 자주 실수하는 표현들과 그에 대응하는 올바른 표현들을⁸⁾ 테스트 데이터로 활용하였다. 즉, 자주 실수하는 표현들과 그에 대응하는 올바른 표현들을 각각 입력으로 주었을 때, 어떤 로그 확률 값이 나타나는지를 보는 것이다. 로그 확률 값을 계산하는 방법은 3.1절에서 설명한 것과 거의 동일하다. 단, 길이가 길면 길수록 로그 확률 값이 커지게 되므로, 이를 정규화하기 위해 로그

확률 값을 길이로 나누어 준다. 이 값이 클수록 더 자연스러운 문장임을 뜻한다. 만약 올바른 표현들의 로그 확률 값이 더 높다면, 언어 모델이 사람처럼 자연스러운 표현을 잘 찾는다고 결론내릴 수 있을 것이다.

테스트 데이터로 사용된 초등학생들이 자주 실수하는 표현은 다음과 같이 크게 9가지로 분류된다.

- 고유명사인데 대문자를 사용하지 않은 경우
- 음조상은 문의 특징을 가지고 있지만, 구조상은 문의 특징을 가지지 못하는 언어 형식
- 주절들을 접속사 없이 콤마로 이은 문장
- 수의 일치가 되지 않은 경우
- 복수 명사를 잘못 표현한 경우
- 복수명사의 소유격을 잘못 만든 경우
- 구두점을 잘못 사용한 경우
- 중문을 형성하지 못한 경우
- 연속된 문장에서 콤마를 사용하지 않은 경우

실험에 사용된 테스트 데이터는 <Table 1>에 자세히 정리되어 있다. 이는 영어권 국가의 초등학생들의 사례이지만, 영어가 모국어가 아닌 다른 나라의 초등학생들도 비슷한 실수를 한다는 점을 감안하면[3] 우리나라 초등학생들에게도 유사한 현상이 나타날 것임을 예상할 수 있다.

4.2 실험 결과

<Table 1>은 실험 결과를 나타낸다. 첫 번째 컬럼은 9가지 종류의 문제를 나타내며, 두 번째 컬럼은 어색한 문장(첫 번째 문장)과 그에 대응하는 자연스러운 문장(두 번째 문장)을 나타냈다. 세 번째 컬럼에서는 우리가 학습한 모델이 계산한 로그 확률 값을 나타낸 것으로, 더 높은 값일수록 더 자연스럽다고 판단한 것이다. 예를 들어, 첫 번째 문장 'I saw dr. smith in new York.'의 로그 확률 -7.71402 보다 두 번째 문장 'I saw Dr. Smith in New York.'의 로그 확률 -5.77584 가 더 큰 값이다. 따라서 언어 모델은 두 번째 문장이 더 자연스

6) <http://www.statmt.org/wmt14/training-monolingual-europarl-v7/europarl-v7.en.gz>

7) <https://github.com/nyu-dl/dl4mt-tutorial>

8) <http://www.time4writing.com/writing-resources/writing-problems-elementary/>

<Table 1> Experimental Results Using Recurrent Neural Network Language Model

Problem	Sentence	Log Probability
Proper Nouns Not Capitalized	I saw dr. smith in new York.	-7.71402
	I saw Dr. Smith in New York.	-5.77584
Sentence Fragments	Going to the grocery store for milk.	-6.84363
	I am going to the grocery store for milk.	-5.26493
Run-on Sentences	I like riding my bike after dinner first I have to help with the dishes.	-6.05918
	I like riding my bike after dinner, but first I have to help with the dishes.	-5.36202
Lack of Subject-Verb Agreement	He run every day.	-6.14207
	He runs every day.	-6.09159
Incorrect Noun Plurals	The berrys are ripe.	-7.18419
	The berries are ripe.	-6.64438
Incorrect Plural and Possessive Nouns	My parent's wedding photo is beautiful.	-7.20945
	My parents' wedding photo is beautiful.	-6.40073
Wrong End Punctuation	What are you doing.	-3.47638
	What are you doing?	-3.26836
Not Forming Compound Sentences	Tom likes baseball his friend likes hockey.	-10.2532
	Tom likes baseball, but his friend likes hockey.	-8.53779
Lack of Commas in a Series	Our flag's colors are red white and blue.	-7.06095
	Our flag's colors are red, white, and blue.	-6.41675

리운 문장으로 판단하며, 이는 사람의 생각과 같다. 실험 결과, 언어 모델은 모든 경우에 대해 첫 번째 문장보다 두 번째 문장이 더 자연스럽다고 판단했다. 즉, 모든 경우에 대해 자동으로 자연스러운 문장을 잘 분류해낼 수 있었다.

실험 결과가 고무적이기는 하지만, 실용화를 위해서는 풀어야 할 문제가 남아 있다는 것을 알 수 있다. 그 문제는, 2개의 유사한 문장 중 자연스러운 문장은 잘 구분해낼 수 있지만 완전히 다른 문장에 대해서는 어떤 것이 더 자연스러운지 판단할 수 없다는 점이다. 예를 들어, 'What are you doing.'의 로그 확률 값은 -3.47638이고 'My parents' wedding photo is beautiful.'은 -6.40073인데, 그렇다고 해서 'What are you doing.'이 더 자연스러운 문장이라 보기는 어렵다. 일반적으로 더 많이 나타난 표현들이 많을 경우 확률 값이 높아지게 되는데, 많이 나타난 표현이 아니더라도 자연스러운 표현이 있을 수 있다는 문제가 있다. 이 문제를 해결하는 것은 추후 연구로 남긴다.

5. 결론

본 논문에서는 통계 언어 모델과 순환 신경망 기반

언어 모델을 비교하였다. 그리고 순환 신경망 모델을 이용하여 초등학생들이 잘못 작성한 문장과 올바른 문장 중 어떤 것이 더 자연스러운지 평가하게 했다. 실험 결과, 모든 9가지의 경우에 대해 자연스러운 문장을 잘 판별하는 것을 확인할 수 있었다. 순환 신경망 모델을 이용한 언어 모델이 초등 글쓰기 교육에 도움이 될 수 있다는 가능성을 보인 것이 본 논문이 공헌한 점이다.

향후 연구는 크게 4가지로 나누어 볼 수 있다. 첫 번째는, 더 나은 네트워크 구조를 사용하는 것이다. 단순히 순환 신경망을 쓰는 것이 아니라, 인코더-디코더 아키텍처(encoder-decoder architecture)를 사용하거나, 더 나아가 주의 집중 방법(attention mechanism) [1, 14, 23]을 사용하는 것도 시도해볼 만하다. 예를 들어, (Fig. 2)에서 'his'의 확률 값을 계산할 때는 앞에 'Tom'이라는 단어가 있었다는 것에 주의를 더 집중하는 것이다.

두 번째 향후 연구는, 지금보다 더 많고 다양한 데이터를 활용하는 것이다. 학습에 사용하는 데이터의 수가 더 많아질수록 학습 시간이 오래 걸리기 때문에 많은 데이터를 활용하기 어렵지만, 보통 많은 데이터를 학습할수록 성능이 높아지기 때문에 추가 데이터를 학습해볼 필요가 있다. 단일어 코퍼스의 양은 언어에 상관없이 매우 방대하기 때문에 데이터를 어떻게 수집할지 고민할 필요는 없을 것으로 보인다.

세 번째 향후 연구는, 로그 확률을 더 잘 구하는 방법에 관한 것이다. 현재의 방식은 많이 나타나는 단어에 높은 확률 값을 부여하고, 반대로 고유명사나 숫자와 같이 자주 나타나지 않는 단어인 경우에는 확률 값을 상대적으로 낮게 부여한다. 따라서 로그 확률의 절대적인 값은 문장의 자연스러움과 큰 관련성이 없다. 어떻게 로그 확률 값을 계산해야 이 값이 어느 수치 미만일 때 자연스럽게 읽히는 문장이라 말할 수 있는가? 이 문제를 해결할 수 있다면 글쓴이에게 어색한 문장을 알려주고 다른 문장을 추천해 주는 등의 작업을 수행할 수 있을 것이다.

마지막으로, 본 연구가 실제 교육현장에서 적용되기 위해서는 스마트 기기에서 사용될 수 있는 형태로 구현되는 작업이 필요하다. 라즈베리 파이 등 초소형 PC가 보급될 수 있다면, MS Word의 추가 기능으로 구현되어도 좋다. 학생들은 과제 제출 전 스스로 자신이 작성한 문서를 검토할 수 있고, 교사들은 학생들의 문서를 교정해 주기 전에 사전 검토하는 목적으로 사용할 수 있으리라 기대한다.

참고문헌

- [1] Bahdanau, D., Cho, K. & Bengio, Y. (2015). Neural Machine Translation By Jointly Learning to Align and Translate. In Proceedings of the International Conference on Learning Representations, 1-15.
- [2] Cho, K. (2015). Deep Learning for Machine Translation. DLAMT Winter School.
- [3] Gonzalez, Y., Saenz, L., Bermeo, J. & Chaves, A. (2013). The Role of Collaborative Work in the Development of Elementary Students' Writing Skills. *Profile Issues in Teacher' Professional Development*, 15(1), 11-25.
- [4] Graves, A. (2013). Generating Sequences with Recurrent Neural Networks. arXiv:1308.0850.
- [5] Hochreiter, S. & Schmidhuber, J. (1997). Long Short-term Memory. *Neural Computation*, 9(8), 1735-1780.
- [6] In, J. & Han, J. (2016). The Prosodic Changes of Korean English Learners in Robot Assisted Learning. *Journal of the Korean Association of Information Education*, 20(4), 323-332.
- [7] Jang, S., Chun, S. (2016). A Study on Teachers' Use of Applications in Teaching-Learning Activities. *Journal of the Korea Association of Information Education*, 20(1), 1-12.
- [8] Jeon, S. & Kim, H. (2016). A Study on Improvement of Web-based Diagnosis-Supplement System for Basic Academic Skills, 20(5), 487-498.
- [9] Kim, C. (2014). A Study on the Educational Use of Tiny PC in an Elementary School. *Journal of the Korean Association of Information Education*, 18(1). 101-110.
- [10] Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. MT Summit.
- [11] Lee, J. & Yoo, S. (2014). A Data Logging Smart r-Learning Effect on Students' Logical Thinking. *Journal of the Korean Association of Information Education*, 18(1), 25-33.
- [12] Lee, M. & Ham, S. (2015). The Development and Effectiveness of the Smart System for Supporting Instructional Materials. *Journal of the Korean Association of Information Education*, 19(4), 399-408.
- [13] Lee, Y., Oh, D. & Park, S. (2015). The Web-Based Interface Design for University Students' Activity-Oriented Career Education. *Journal of the Korean Association of Information Education*, 19(3), 345-360.
- [14] Luong, M., Pham, H. & Manning, C. D. (2015). Effective Approaches to Attention-based Neural Machine Translation. arXiv:1508.04025.
- [15] Park, S., Han, K., Lee, D., Shin, B. & Lee, J. (2015). Designing and Developing ICT Contents of Mathematics and Science in Agricultural, Mountain and Fishing Villages. *Journal of the Korean Association of Information Education*, 19(2), 215-224.
- [16] Park, Y. (2016). Automatic Generation of

Multiple-Choice Questions Based on Statistical Language Model. *Journal of the Korean Association of Information Education*, 20(2), 197-206.

- [17] Park, Y. (2016). Automatic Selection of Similar Sentences for Teaching Writing in Elementary School. *Journal of the Korean Association of Information Education*, 20(4), 333-340.
- [18] Pham, T. (2012). A Study on Teaching and Learning Korean Grammars Method Based on Paraphrasing Activities. Master's Thesis, Seoul National University.
- [19] Ryu, M. & Han, S. (2016). Development of Smart Application for English Speaking. *Journal of the Korean Association of Information Education*, 20(4), 367-374.
- [20] Sennrich, R., Haddow, B. & Birch, A. (2016). Neural Machine Translation of Rare Words with Subword Units. In Proceedings of the Annual Meeting of the Association for Computational Linguistics, arXiv:1508.07909.
- [21] Suh, S. & Goh, Y. (2016). Study on School Teachers' Perception of and Usage of SMART Education. *Journal of the Korean Association of Information Education*, 20(2), 139-150.
- [22] Thornbury, S. (2000). *How to Teach Grammar*. Longman.
- [23] Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., Zemel, R. S. & Bengio, Y. (2015). Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. arXiv:1502.03044.

저자소개

박 영 기



2008 KAIST 전자전산학과
전산학전공(공학사)

2010 서울대학교 대학원
컴퓨터공학과(공학석사)

2015 서울대학교 대학원
컴퓨터공학과(공학박사)

2015~2016 삼성전자 종합기술원
전문연구원

2016~현재 춘천교육대학교 컴퓨
터교육과 조교수

관심분야 : 초등컴퓨터교육, 데이
터 마이닝, 인공지능

E-Mail : ypark@cnu.ac.kr