딥러닝을 활용한 저연령층 영어 교육 시스템

김희용*, 장호택*, 이수현*, 이해연*
*금오공과대학교 컴퓨터소프트웨어공학과
e-mail: rgd5262@gmail.com

English Education System for Kids using Deep Learning

Hee-Yong Kim*, Ho-Taek Jang*, Soo-Hyeon Lee*, Hae-Yeoun Lee*
*Dept of Computer Software Engineering,
Kumoh National Institue of Technology

요 약

국제화 시대를 맞이하여 세계 공용어인 영어의 중요성이 부각되고 있다. 특히, 영어 교육의 학습 연령대는 점점 낮아지고 있는 추세이며, 이에 동반하여 저 연령층 영어 교육 콘텐츠가 출시되고 있다. 하지만 현재 저 연령층을 대상으로 출시되는 콘텐츠들은 연령에 맞지 않는 교육 자료를 제시하거나 언어 학습에 필요한 상황적 다양성이 부족한 것이 현실이다. 본 논문에서는 딥러닝을 적용하여 사용자가 원하는 상황을 촬영한 영상에서 대상 연령에 적합한 영어 문장을 생성하고 읽어주는 학습 시스템을 제안한다. 본 시스템을 통하여 저 연령층에 적합한 영어 교육 환경을 제공하고, 저 연령층에게 나타나는 영어 교육의 불균형을 해소하고자 한다.

1. 서론

국제화 시대를 맞이하여 세계 공용어인 영어의 필요성은 날로 부각되고 있으며, 이는 영어 교육 시설 및 시스템의 폭발적인 증가를 일으켰다. 하지만 현재 우리나라의 영어 교육 시설 및 시스템은 중고등 교육에 초점이 맞추어져 있으며 초등 혹은 유아를 위한 영어 교육은 적절한 플랫폼이 마련되어 있지 않거나, 존재 하더라도 고액의 비용이 요구되어 대중성이 부족하다. 영어 유치원, 영어 캠프, 영어 학원을 필두로 한 저 연령층의 영어 교육은 고 연령층이 받는 영어 교육에 비해 비용이 약 30%이상의 더 발생하였으며, 영어 교육을 받는 저 연령층의 비율은 고 연령층에 비해 저조한 것으로 조사되었다[1].

[표 1]은 2010년부터 2016년도까지의 초등학생 영어 교육에 사용된 사교육비의 현황을 나타낸다. 영어 교육의 필요성이 증가하고 있는 반면, 저 연령층에게 사용되는 영어 사교육비는 점차 감소하는 모습을 보인다. 이는 저 연령층에 실시되는 교육의 품질이 만족스럽지 않으며 비용대비 효용성이 높지 않음을 짐작하게 한다.

(표 1) 초등학생 영어 사교육비 현황 (명)

2010	2011	2012	2013	2014	2015	2016
33,476	30,927	26,350	25,918	24,084	22,886	21,631

본 논문에서 제안하는 시스템은 딥러닝 기술을 이용하여 사용자가 제시한 영상에 대하여 저 연령층에 적합한 영어 문장을 생성하는 것이다. 먼저 딥러닝을 기반으로 사용자가 입력한 이미지를 통해 적절한 고수준의 영어문장을 생성한다. 그 후에 LSTM에 기반을 둔 언어모델을 통하여 고수준의 영어문장을 저 연령층에 적합한 저수준의 영어문장으로 변환하여 제공하고 TTS를 통하여 발음을 하도록 한다. 사용자들은 입력한 이미지에 대한 영어 문장을 학습함으로써 큰 비용을 들이지 않고 상황에 맞는 영어를 학습할 수 있을 것이다.

본 논문의 구성은 다음과 같다. 2장에서는 관련 연구를 제시하고, 제안하는 시스템은 3장에서 설명한다. 4장에서는 실험 환경 및 실험 결과를 제시한다. 마지막으로 5장에서는 결론과 개선해야 할 점을 설명한다.

2. 관련 연구

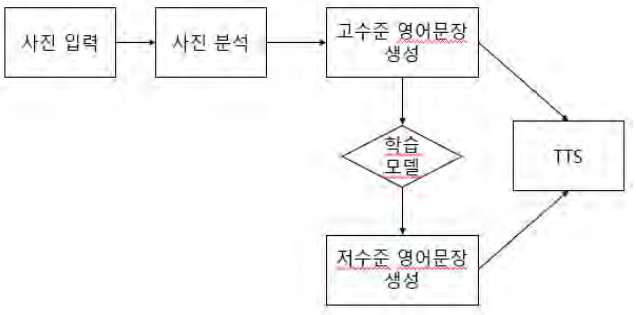
딥러닝을 통하여 언어 모델을 구성하고, 이를 통해 문장의 생성 또는 변환을 시도하는 연구는 다양하게 진행되고 있다. 그 중 김양훈 등은 LSTM 언어모델 기반 한국어 문장 생성 연구에서 주어진 단어 배열에 대한 확률 분포를 정의하고, 말뭉치의 데이터를 분석하여 언어 구조상 적합한 문장을 생성하였다[2]. 이를 위해 LSTM(Long-Short Term Memory) RNN에 기반을 둔 언어 모델을 사용하였다. LSTM 모델은 길이가 긴 입력 데이터를 적용하는 작업에서 기존 RNN보다 좋은 성능을 보인다. 해당 연구에서는 LSTM 모델을 통해 불완전한 한국어 문장 혹은 단어가 주어졌을 시 적합한 문장을 생성하는 시스템을 제안하였다.

3. 제안하는 시스템

본 논문에서 제안하는 시스템은 사용자가 사진을 입력하면 사진에 가장 적합한 고수준의 영어 문장을 생성하고, 그 문장을 저수준의 영어 문장으로 변환하여 출력한다.

사진을 입력하고 문장이 출력되는 프로그램은 Google Tensorflow에서 제공하는 API를 사용한다. 해당 API는 사진을 입력받고 그 사진을 가장 잘 설명하는 문장을 출력한다. 출력된 고수준의 영어 문장은 Seq2Seq 기술을 기반으로 생성된 모델을 통하여 저수준의 영어 문장으로 변환된다. 사용되는 모델은 ImageNet에서 배포하는 학습용 데이터 셋 16만개를 전처리하여 만든 3가지 데이터 셋을 사용하여 학습되었다. 또한 Google TTS API를 이용해 변환된 문장을 음성으로 도출한다.

제안하는 시스템의 알고리즘은 (그림 1)과 같다. 먼저 사용자가 영어 학습을 원하는 상황의 이미지를 시스템에 입력한다. 이후 Google Tensorflow에서 제공하는 API를 통하여 사진의 분석 및 고수준의 영어 문장 생성을 실시한다. 이때 생성되는 문장은 형용사, 부사를 비롯한 수식어들을 포함한 복잡한 구조를 지니게 되므로 저연령층을 위한 교육에는 알맞지 않다. 이에 다양한 문장을 학습시킨 모델을 통하여 고수준의 영어 문장을 구조가 간단한 저수준의 영어문장으로 변환한다. 저수준의 영어 문장은 가장 기본적인 5형식 문법을 기반으로 구성되어 있으며 부가적인 수식어는 포함되어 있지 않다. 마지막으로 사용자는 고수준의 문장과 저수준의 문장을 모두 TTS 기능을 통하여 음성으로 해당 문장을 들을 수 있다.



(그림 1) 제안하는 시스템의 처리 과정

4. 실험 환경 및 실험 결과

시스템의 개발 환경으로는 운영체제는 Ubuntu 16.04 LTS를 사용하고 플랫폼은 Python 2.7을 기반으로 한 Tensorflow 1.0을 사용했다. 하드웨어적 사양은 CPU는 intel i7-7700K, GPU는 Nvidia GTX 1060 6GB를 사용하였고, 메모리는 16GB - Dual Channel을 사용하였다.

개발한 시스템의 UI는 (그림 2)와 같다. 사진을 입력하면, 우측 상단에 고수준의 언어가 표현되고, 이를 변환한 저수준의 언어가 우측 하단에 표시된다. 그리고 TTS 기능을 통하여 소리를 낼 수 있다. 현재는 연구 개발 수준에 적합한 인터페이스를 가지고 있으며, 차후 완성도를 위한 개선이 필요하다.

(그림 2) 개발 시스템의 사용자 인터페이스

4.1 학습 데이터 셋 구성 방법

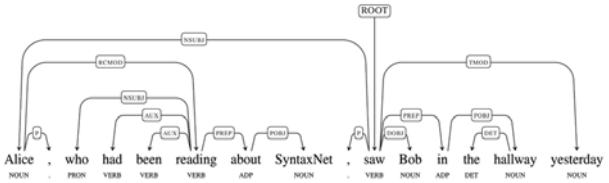
ImageNet에서 기본 제공하는 학습용 데이터 셋 16만개를 기반으로 새로운 데이터 셋을 구성하였다. (표 2)와 같이 3가지 데이터 셋이 존재하며, 의미론적(Semantics), 구문법적(Syntax), 복합적(Complex)으로 기준을 세분화하였다. 의미론적 데이터 셋은 문법은 상이하지만 동일한 의미를 가지는 3가지의 문장을 한가지의 간단한 문장으로 대응시켜 생성하였다. 문법적 데이터 셋은 형태소 분석을 통하여 복잡한 수식어를 제거한 간단한 문장으로 이루어지며 복합적 데이터 셋은 의미론적이고 문법적 데이터 셋이 모두 포함되어 있다.

(표 2) 학습 데이터 셋 구성

데이터 셋	기반 데이터	대응 데이터	갯수
의미론적 데이터셋	복잡한 문장 3개	간단한 문장 1개	7만 쌍
문법적 데이터셋	기존 문장	형태소 전처리를 거친 문장	13만 쌍
복합적 데이터셋	의미론적 및 문법적 데이터셋		20만 쌍

4.2 형태소를 이용한 전처리 데이터 셋 생성 방법

Google에서 지원하는 API 중에서 SyntaxNet은 문맥 구조 분석을 지원한다. (그림 3)과 같은 형식으로 문장요소 간의 관계의 시각화를 지원한다.



(그림 3) SyntaxNet 기술 구조

이를 이용하여 기본적인 문장형식에 필요한 주어, 동사, 목적어 등을 파악하고 부사와 전치사 같은 부가적인 요소를 삭제하는 형식으로 전처리를 진행한다.

4.3 실험 케이스

기반 모델에서 Node와 Layer의 수에 변화를 주어 학습의 깊이와 너비를 조정하였다. Layer의 경우 2, 3 Layer로 구분하였고, Node의 경우 128, 256, 512, 1024개로 구분하여 진행하였다. 입력 데이터는 임의의 문장으로 제시하거나, 학습 데이터 셋에 존재하는 문장으로 구성하여 2가지의 경우가 존재한다. 마지막으로 모델에 사용된 데이터의 학습 횟수를 7500, 15000, 22500으로 나누어 실험을 진행하였다.

4.4 실험 결과

실험 결과 중 값이 0에 수렴하거나 유의미한 결과가 없는 실험케이스는 제외를 하였다. 그 결과 (표 3)과 같이 총 36개의 실험 케이스만이 실질적으로 비교를 할 수 있는 값을 나타내었다.

(표 3) 데이터 완성도 측정 값

Layer - Node	데이터 셋	Learning Number(epochs)		
		7500	15000	22500
Two - 128	Semantics	21.5	24	15
	Syntex	25.5	30	16
	Complex	16	26	9
Two - 256	Semantics	47.5	41	2.15
	Syntex	55	70	46
	Complex	35.5	31.5	15.5
Two - 512	Semantics	10	10	2.5
	Syntex	10	11.5	7.5
	Complex	5.5	5.5	0.5
Three - 256	Semantics	9.5	13	2.5
	Syntex	12	14	4.5
	Complex	7.5	13.5	5

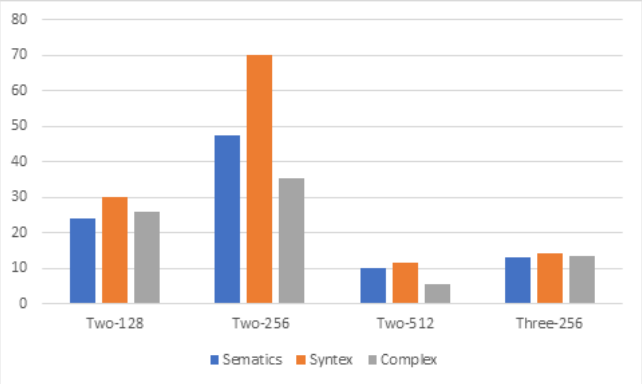
Layer Number와 Node Size가 특정한 값 이상으로 늘어날수록 정확도가 떨어지는데 과적합(Over-Fit)에 따른 문제로 추정된다. Learning Number가 15000일 때, 대부분의 결과가 가장 높은 완성도를 나타내었다. (표 3)에서 학습에 사용된 각각의 데이터 셋의 결과값 중 최댓값을 구하여 Layer-Node를 기준으로 (그림 3)와 같이 그래프로 나타내었다. 전체 실험 결과 중 Two Layer - 256 Node Size에서 완성도가 가장 높았다.

(표 4)에는 입력 영상에 대하여 고수준 표현과 저수준 표현으로 변환한 예들을 확인할 수 있다. 복잡한 표현이 사라지고 간단한 장면에 대한 묘사를 포함하고 있는 것을 볼 수 있다.

5. 결론

본 논문에서는 딥러닝을 활용하여 다양한 사진들을 영어 평문으로 도출하고, 도출된 평문을 저 연령층이 이해하기 쉽도록 저 수준의 문장으로 변환시켜 학습 할 수 있는 시스템을 제안하였다.

현재 개발된 시스템은 자체적인 데이터 셋을 구축하여, 변환된 문장이 정확한지 실험 및 분석이 이루어져야 한다. 하지만 본 시스템은 특별한 비용을 들이지 않고, open API, 자체적인 데이터 셋으로만 이용하여 저 연령층에게 다양한 상황을 접하며 학습 할 수 있다는 점에서 긍정적이다. 향후, 좀 더 적합한 데이터 셋 구축을 통해 완성도를 높이고, 사진이 아닌 동영상으로 확대 가능하며 저 연령층들이 원하는 상황에서 영어를 좀 더 접근하기 쉽고, 흥미를 유발하여 교육 환경에 개선을 할 수 있을 것으로 기대한다.



(그림 3) 수준 변환 완성도

(표 4) 시스템을 통한 변환 결과

	고수준	a group of cars parked on the side of a road.
	저수준	there is a group of cars on side.
	고수준	a room with a table, chairs and a table.
	저수준	there is a room with table.
	고수준	a group of young men playing a game of frisbee.
	저수준	there is a group of men playing game.

참고문헌

[1] 2016 초중고 사교육비조사 결과, http://kostat.go.kr/portal/korea/kor_nw/2/1/index.board?bmode=read&aSeq=359420.

[2] 김양훈, 황용근, 강태관, 정교민. (2016). LSTM 언어모델 기반 한국어 문장 생성. 한국통신학회논문지, 41(5), 592-601.