Journal of the Korea Convergence Society
Vol. 12. No. 1, pp. 57-63, 2021

ISSN 2233-4890 / eISSN 2713-6353
https://doi.org/10.15207/JKCS.2021.12.1.057

발화 의도 예측 및 슬롯 채우기 복합 처리를 위한 한국어 데이터셋 개발

한승규[1], 임희석[2*]
[1]고려대학교 컴퓨터학과 석사과정, [2]고려대학교 컴퓨터학과 교수

Development of Korean dataset for joint intent classification and slot filling

Seunggyu Han[1], Heuiseok Lim[2*]
[1]Master Student, Department of Computer Science and Engineering, Korea University
[2]Professor, Department of Computer Science and Engineering, Korea University

요 약 사람의 발화 내용을 이해하도록 하는 언어 인식 시스템은 주로 영어로 연구되어 왔다. 본 논문에서는 시스템과 사용자의 대화 내용을 수집한 말뭉치를 바탕으로 언어 인식 시스템을 훈련시키고 평가할 때 사용할 수 있는 한국어 데이터셋을 개발하고, 관련 통계를 제시한다. 본 데이터셋은 식당 예약이라는 고정된 주제 안에서 사용자의 발화 의도와 슬롯 채우기를 해야 하는 데이터셋이다. 본 데이터셋은 6857개의 한국어 문장으로 이루어져 있으며, 표기된 단어 슬롯의 종류는 총 7개이다. 본 데이터셋에서 표기된 발화의 종류는 총 5개이며, 문장의 발화 내용에 따라 최대 2개까지 동시에 기입되어 있다. 영어권에서 연구된 모델을 본 데이터셋에 적용시켜 본 결과, 발화 의도 추측 정확도는 조금 하락하였고, 슬롯 채우기 F1 점수는 크게 차이나는 모습을 보였다.

주제어 : 자연어처리, 발화 이해, 발화 의도 예측, 슬롯 채우기, 데이터셋, BERT

Abstract Spoken language understanding, which aims to understand utterance as naturally as human would, are mostly focused on English language. In this paper, we construct a Korean language dataset for spoken language understanding, which is based on a conversational corpus between reservation system and its user. The domain of conversation is limited to restaurant reservation. There are 7 types of slot tags and 5 types of intent tags in 6857 sentences. When a model proposed in English-based research is trained with our dataset, intent classification accuracy decreased a little, while slot filling F1 score decreased significantly.

Key Words : Natural Language Processing, Spoken Language Understanding, Intent Classification, Slot Filling, Dataset, BERT

*This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2020-2018-0-01405) supervised by the IITP(Institute for Information & Communications Technology Planning & Evaluation).
*Corresponding Author : Heuiseok Lim(limhseok@korea.ac.kr)
Received November 26, 2020 Revised December 18, 2020
Accepted January 20, 2021 Published January 28, 2021

1. 서론

1.1 발화 이해 연구

자연언어처리(Natural Language Processing)에서는 문어체 문장과 구어체 문장을 서로 다른 방법으로 다룬다. 문어체와 구어체는 일반적으로는 언어의 전달 매개체로 구분하나, 최근에는 인터넷과 실시간 채팅 시스템의 발달 등으로 그 경계를 매개체로만 구분하기 어려워졌다. 문어체 문장은 주로 서적이나 인터넷 기사, 블로그 내용 등 작성자가 충분한 시간을 들여 작성한 글에서 볼 수 있는 문체라고 할 수 있다. 맞춤법이 잘 지켜져 있고, 글이 명료하게 작성되어 있으며, 정보의 집중도가 높다. 반면 구어체 문장은 주로 사람들의 실제 대화 내용을 받아적은 것이나, 온라인 채팅 프로그램에서 작성되는 문장들이다. 맞춤법이 잘 지켜지지 않으며, 신조어나 오타(받아쓴 문장의 경우 오류)가 자주 등장한다. 문어체 문장에 비해 정보의 집약도가 낮은 것도 특징이다.

문어체 문장은 정보의 집약성 때문에 일반적으로 이를 이용하여 문장 내 개체(Entity)간의 관계를 추출하거나, 더 나아가 지식 데이터베이스를 구축하는 등, 지식과 관련한 연구에 많이 이용된다.

구어체 문장은 이와는 다르게 사람의 일상적인 대화를 이해하거나, 기계가 그러한 대화를 하도록 하는 연구에서 많이 다루어진다. 최근 큰 발전을 이루어 사람들이 많이 사용하는 대화형 가상 비서가 대표적인 예시이다.

본 논문에서 구축한 데이터셋은 구어체 문장에 관한 것으로, 사용자와 대화를 하거나, 사용자의 음성을 인식하여 명령을 처리하는 다양한 시스템에서 그 바탕이 되는 내용 이해를 위한 모델을 학습시킬 때 필요한 것이다.

1.2 한국어 관련 연구

한국에서 자연언어처리를 연구하는 연구자들은 대부분 영어를 연구 기준으로 한다. 이는 영어의 분석 편리성과 연구를 위한 자료 수집의 용이성 때문이다.

한국어는 언어 분류 상 교착어로 분류되는 언어로, 어근과 어미가 결합하여 단어를 이루는 그 특성 때문에 띄어쓰기를 기준으로 단어를 분리할 경우 지나치게 많은 종류의 단어가 나타나 기계학습으로 자연어를 처리하는데 난관이 된다[1]. 이는 굴절어인 영어에서는 거의 발견되지 않는 문제로, 외국에서 개발된 많은 자연언어처리 모델을 한국어에 접목시키는데 어려움을 주고 있다.

영어는 전세계에서 공용어로 사용되고 있다는 점도 영어 관련 연구가 활발한 큰 이유이다. 연구를 위한 자료를 수집하는 것이 매우 용이하기 때문이다. 한국어에 비해 영어의 화자 수가 압도적으로 많기 때문에 관련된 데이터셋을 구하는 것이 어렵지 않다. 특히, 모델의 성능을 평가하기 위한 표준 데이터셋 역시 대부분 영어로 되어 있다.

최근에는 한국어 화자를 대상으로 한 다양한 음성인식 스마트 스피커 등이 개발되면서 한국어 데이터셋을 구축하려는 움직임이 커지고 있지만, 대부분은 기업에서 내부에서만 공유되는 형태로 만들어지고 있다. 국립국어원이나 한국전자통신연구원(ETRI) 등의 국가 기관에서도 공개된 말뭉치 등을 구축하려고 노력하고 있지만, 역시 영어에 비해선 턱없이 양이 부족한 것이 현실이다.

이러한 이유로 한국에서 자연언어처리를 연구하는 연구자들 사이에서도 한국어 관련 연구는 영어를 다루는 논문에 비해 그 수가 상대적으로 적다. 따라서 본 논문에서는 한국어로도 다양한 자연언어처리 연구를 할 수 있는 기반이 되는 양질의 한국어 데이터셋을 구축하였다. 3장에서 데이터셋을 구축하기 위해 원본 대화 내역을 수정 보완한 내역과 그 통계를 제시하였다. 또한, 기존의 영어 데이터셋으로 훌륭한 성능을 보인 모델을 한국어로 다시 학습시킨 결과의 비교를 4장에서 제시한다.

2. 관련 연구

2.1 발화 이해 관련

발화 이해는 오랫동안 많은 연구자들의 관심을 받았으며, 딥 러닝이 자연어처리에 사용된 이후 딥 러닝 구조들이 새로 등장할 때마다 발화 이해에 접목되어 왔다.

합성곱 신경망(Convolutional Neural Network; CNN)을 자연어처리에 접목시켜 문장을 분류하는 모델을 제시한 [2]은 간단한 구조를 통해 높은 성능을 보일 수 있어 많은 주목을 받은 연구이다. 이 연구는 특정 과제를 수행하기 위한 것은 아니나, 평가를 위해 사용한 데이터셋 중 사용자의 발화 의도를 분석하는 데이터셋이 포함되어 있다. 이 논문이 발표된 이후 CNN에 Attention을 도입하는 등 연구자들의 여러 시도가 계속되었다[3].

CNN 이후에는 순환 신경망(Recurrent Neural Network; RNN)을 기반으로 한 머신러닝 모델이 자연어처리에 적극적으로 도입되었다. 특히 LSTM(Long Short-Term Memory) 구조[4]와 이를 양방향으로 진행하는 Bi-LSTM(Bidirectional LSTM) 이 많이 사용되었고, 이를 기반으로 한 발화 이해 모델도 개발되었다. [5]

가 LSTM을 기반으로 발화 이해 연구를 진행한 대표적인 사례이다. [6]은 Bi-LSTM을 기반으로 발화의 주제를 포함한 종합적 발화 이해를 처리하는 모델을 개발한 논문으로, 영어를 이용한 데이터셋을 처리할 때 글자 단위에서도 Bi-LSTM을 적용하여 정확도를 높였다. [7]은 CRF 층을 함께 사용하여 슬롯 채우기의 정확도를 높였다. [8]은 LSTM에 Attention 기법을 도입하여 성능을 개선하였다.

이후 BERT[9]가 등장하며 자연어처리 관련 연구들이 대부분 BERT를 사용하기 시작했다. BERT는 Transformer 구조[10]를 바탕으로 학습된 언어 모델로, 최초 학습에는 일반적인 수준의 장비로는 감당이 힘들 정도로 오랜 시간이 걸린다. 하지만 이렇게 사전 훈련된 모델을 전이학습을 이용하여 쉽게 다른 과제에 접목시킬 수 있다는 장점이 있다. 비교적 일반적인 형태로 학습된 사전 훈련 모델을 감성 분석이나 개체명 인식과 같은 특정 문제 해결에 특화된 모델로 전이학습 시키는 과정을 fine-tuning 이라고 부른다. fine-tuning은 최초 학습에 비해서 상대적으로 적은 시간이 걸리며, BERT를 제외한 모델의 나머지 구조가 복잡하지 않아도 훌륭한 성능을 내고 있어 많은 연구자들의 관심을 받고 있다. [11]이 BERT를 사용하여 발화 의도 예측 및 슬롯 채우기를 수행하는 모델을 제시하였다.

2.2 한국어 자연어처리 관련

BERT가 처음 공개되었을 때에는 영어를 바탕으로 학습되어 영어로 이루어진 데이터셋을 통해 그 성능을 검증하였다. 이후 많은 연구자들이 이를 이용하면서 언어와 관계없이 사용할 수 있는 다국어 BERT를 사전학습하여 공개하기도 하였다. 하지만 다국어 BERT는 일반적으로 한 언어에 한해 학습된 것에 비해 아쉬운 성능을 보여주었고, 한국어도 마찬가지이다. 이에 SKT-Brain에서 KoBERT라는 한국어로 학습된 BERT를 공개하였다 [12]. KoBERT는 한국어 위키백과의 본문과 각종 신문 기사들을 수집해 학습시킨 BERT 사전학습 모델이다. 최초 훈련된 언어만 다를 뿐 기본적인 구조는 동일하다.

본 논문에서 다루는 발화 이해와 정확히 일치하지는 않지만, 문장 단위 분류와 토큰 단위 분석을 진행하는 한국어 연구도 일부 있다. [13]은 RNN을 이용하여 발화 의도를 분류해내는 모델을 만들었고, 영화 및 드라마의 자막 데이터를 이용하여 학습하였다. [14]는 하나의 모델이 개체명 인식과 감성 분석을 포함한 여러 분석을 동시에 진행하는 모델로, 영화 감상평 데이터를 이용하여 BERT를 학습하였다.

3. 데이터셋 개발

3.1 기존 말뭉치 분석

본 논문에서 만들고자 하는 데이터셋은 [15]의 말뭉치를 바탕으로 한다. 해당 논문에서 수집한 문장들은 가상의 대화형 식당 예약 시스템과 사용자가 대화한 내역을 수집한 것이다. 사용자는 시스템에게 식당을 추천받거나, 위치 혹은 음식 종류를 통해 조건에 맞는 식당을 검색한다. 시스템은 예약을 수행하기 위해 방문하는 날짜와 시간, 방문 인원을 확인한다. 본 말뭉치에는 약 306건의 대화가 기록되어 있고, 문장 수로는 총 10300문장이다.

Table 1. Example conversation of original corpus

[시스템]	안녕하세요, 원하시는 지역을 말씀해주세요.
[사용자]	[[홍대 ||위치]]주변 [[일식집 ||음식종류]] 추천해주세요!
[시스템]	[[홍대 ||위치]]주변 예약가능한 [[일식집 ||음식종류]]은 [[혼가츠 ||식당]], [[키움초밥 ||식당]], [[맛있는교토 ||식당]], [[무라 ||식당]], [[하카타분코 ||식당]], [[토끼정 ||식당]] 이 있습니다.
[사용자]	[[토끼정 ||식당]]이요!
[시스템]	몇명이 방문하시겠습니까?
[사용자]	[[4명 ||인원]]!
[시스템]	언제 방문하시겠습니까?
[사용자]	[[3월 1일 ||날짜]] [[12시 30분 ||시간]] 예약해주세요~
[시스템]	[[홍대 ||위치]] [[토끼정 ||식당]] [[3월 1일 ||날짜]] [[오후 12시 30분 ||시간]] [[4명 ||인원]] 예약맞으신가요?
[사용자]	네!
[사용자]	예약이 완료되었습니다. 감사합니다.

이 말뭉치의 문장들 중 다음과 같은 기준을 이용하여 적합하지 않은 문장들을 걸러내었다. 먼저, 발화 내역을 이해하는 데 있어 시스템의 발화 내역은 필요치 않으므로 모두 제거하였다. 또한, 사용자가 채팅 등의 형태로 입력한 것으로 보이는 부분 (오타, 기호 혹은 ㅇㅇ 과 같은 축약 표현)도 모두 제거하였다. 이렇게 제거한 후 남은 문장의 개수는 6857개이다.

3.2 슬롯 태그 수정

[15]에서 슬롯에 부착한 태그는 총 12가지로, 식당 검색 및 예약에 필수로 필요한 정보 여섯 가지와, 부가적인 정보 여섯 가지이다. 부착된 태그들을 분석한 결과, 부가 정보로 분류된 태그들 대부분이 필수 정보에 비해 등장 빈도가 굉장히 낮았다. 또한, 부가 정보의 경우 날짜나 식당과 같이 같은 분류로 묶이는 서로 다른 내용이 아니라, 하나의 똑같은 대상을 설명하는 다른 방식들이 하나의

태그로 분류되어 있는 경우도 있었다.

등장 빈도가 너무 낮은 태그가 여럿 있을 경우 일반적인 기계학습을 위한 데이터셋으로는 적절하지 못하기 때문에 특별한 이유가 없을 경우 제거되는 경우가 많다. 태그가 하나의 분류가 아니라 특정 대상을 가리키는 것 역시 태그의 정의에 부합하지 않기 때문에 삭제해야 한다.

'주차장 유무' 로 기록된 태그는 위의 두 삭제 기준에 모두 부합하는 예시이다. 선별된 문장에서 해당 태그가 등장하는 회수는 50회 미만으로, 비율 상으로도 1%가 되지 않는다. 또한, '주차장 유무' 태그는 '주차장' 혹은 '주차공간' 등 주차장을 지칭하는 여러 단어들에 달려있는 태그로, 슬롯 채우기를 위한 태그에는 적절치 않다. 주차장 유무를 포함하여 전화번호, 주소, 주차 서비스 등의 추가 데이터들이 모두 이에 해당되어 삭제하였다. '예약 도구' 라는 태그는 등장 빈도가 적절히 높았고 전화와 앱 등 다양한 도구들이 언급되었지만, 실제로 이 데이터셋을 이용하여 서비스를 구축할 경우 예약 도구는 이 서비스 자체가 되므로 표기하는 의미가 없기에 삭제하였다.

따라서 필수 태그 여섯가지(위치, 음식종류, 식당이름, 날짜, 시간, 인원)와, 부가 정보 중 필수 태그와 비율이 가장 비슷했던 가격 정보만을 남기고 나머지 태그는 태그 정보를 삭제한 후 일반 글자로 다시 환원하였다. 처리 후 데이터셋 내의 태그 통계 정보를 Table 2.에 정리하였다.

Table 2. Statistics of slot tags in the dataset

슬롯	예시	개수	비율
위치	이태원 근처, 강남역 주변	687	11.79%
음식종류	한식, 양식, 조개구이, 쌀국수	751	12.89%
식당이름	(실제 식당 이름들)	1379	23.51%
날짜	3월 11일, 오늘, 내일, 토요일	794	13.63%
시간	한 시, 점심, 오후	849	14.57%
인원	세 명, 5명	792	13.59%
가격	1인당 15000원, 4만원	584	10.02%
계		5827	100%

3.3 의도 태그 부착

발화 의도 및 슬롯 채우기를 복합적으로 실행하기 위한 데이터셋에는 발화 의도 태그도 부착되어 있어야 하나, 원본 말뭉치에는 해당 정보가 기입되어 있지 않다. 따라서 본 논문에서는 자체적으로 기준을 설정하여 발화 의도를 다섯 가지 종류로 구분하였고, 이를 각 문장에 표기하였다. 각 의도와 예시를 Table 3.에서 볼 수 있다.

일반적인 데이터셋에서는 발화 의도를 한 문장에 하나씩만 설정한다. 하지만 위 표의 '거절' 의 예시와 같이, 거절을 한 다음 다시 예약을 하기 위한 말을 덧붙이는 경우가 말뭉치 안에서 종종 확인된다. 검색에 필요한 정보와 예약에 필요한 정보가 서로 다름에도 식당 검색을 할 때 예약 정보인 인원 수 등을 같이 말하는 경우도 있다. 이와 같은 경우에 대해 반드시 하나의 의도를 고르게 되면 문장의 순수한 의도가 잘못 표현될 수 있다. 따라서 본 데이터셋에서는 이러한 여러 의도가 동시에 담긴 문장에 대해서는 최대 2개의 의도 태그를 부착하였다. 2개의 의도가 부착된 문장의 예시는 Table 4에서 찾을 수 있다.

Table 3. Example of each intent in the dataset

발화의도	예시
검색 (Query)	[[용인||위치]] 지역에 있는 [[이탈리안||음식종류]] 음식 식당 추천해줘
예약 (Reservation)	[[다음주 금요일||날짜]] [[저녁 8시||시간]]에 갈 거야
질문 (Question)	[[자원방래||식당]] 위치 알려줘
승인 (Confirm)	응 맞아
거절 (Deny)	아니 [[명동피자|식당]] 로 바꿀래

Table 4. Example of multi-intent sentence in the dataset

발화의도	예시
검색 및 예약	[[다음주 월요일||날짜]] [[점심||시간]]에 갈 수 있는 [[연남동||위치]] 맛집 추천해줘
거절 및 검색	다른 음식점도 보여줘
거절 및 예약	아니 인원 변경할래

검색과 예약, 질문, 승인문의 경우 단일 의도인 경우가 많았고, 거절의 의도가 담긴 문장에는 대부분 검색 혹은 예약 의도가 동시에 포함되어 있었다. 문장의 의도 를 Table 5에 제시한다.

Table 5. Statistics of intents in the dataset

	(단독)	예약	질문	승인	거절
검색	1824	100	0	1	196
예약	2217		3	4	128
질문	1661			1	2
승인	671				0
거절	49				

4. 실험

4.1 사용 모델

위에서 구축한 데이터셋을 검증하기 위하여, 발화 의도와 슬롯 채우기를 동시에 수행하는 모델을 찾아 실험을 진행한다. 다양한 모델 중에서 선정한 모델은 관련 연구에서 언급했던 [11]에서 제시한 모델로, Fig.1에서 그 구조를 확인할 수 있다. 이 모델은 최근 자연언어처리에서 표준적인 언어 모델로 자리잡은 BERT를 이용한 것으로, 사전 학습된 BERT에 전이 학습 기법을 이용하여 훈련한다. BERT에 입력되는 문장에는 반드시 가장 앞에 [CLS] 토큰이 있으며, 이 토큰이 BERT층을 통과한 후의 값을 바탕으로 문장의 발화 의도를 파악한다. [CLS]를 포함한 문장 전체가 BERT층을 통과한 후의 값들은 각 토큰별로 슬롯 채우기 예측을 하는데 사용된다.

이 모델은 문장을 토큰화 할 때 Sentencepiece를 사용한다. 토큰화하는 과정에서 단어가 여러 토큰으로 나뉘었을 경우, 최종적으로 단어 단위에서 슬롯 채우기를 할 때에는 이 토큰들 중 가장 앞 토큰만을 사용하여 결정한다. 원본 모델은 Loss 역전파를 위하여 cross-entropy loss만을 사용하지만, 여기서는 발화 의도가 최대 2개일 수 있으므로 발화 의도에 한해서 Binary Cross-entropy를 이용한다. 최종 Joint Loss는 두 Loss의 합으로 한다.

위 모델은 영어로 학습된 모델이기에 최초 BERT 논문에서 공개했던 사전 학습 모델을 사용한다. 하지만 이것을 그대로 한국어에 적용하기에는 무리가 있으므로 한국어 데이터셋으로 실험할 때는 이를 KoBERT로 대체하여 실험하였다.

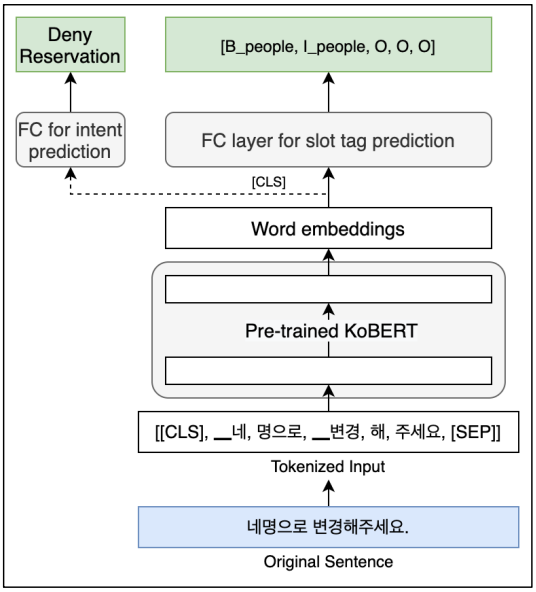


Fig. 1. Architecture of model of [11].

4.2 평가 방법

본 실험에서는 발화 의도 파악과 슬롯 채우기에 서로 다른 평가 지표를 사용한다. 예측한 슬롯을 평가하기 위한 평가 지표는 Macro F1이다. 원론적인 F1 점수는 참/거짓을 예측하는 모델의 성능을 평가할 때 사용된다. Macro F1은 이를 여러 항 가운데서 분류를 수행하는 모델의 성능 평가에도 사용할 수 있도록 확장한 것이다. F1 점수를 계산할 때 사용되는 정밀도(precision)과 재현율(recall)을 각 분류 항마다 계산하고, 모든 분류의 정밀도와 재현율의 평균을 이용하여 F1 점수를 계산한다. 평균값을 이용하여 최종 점수를 계산하기 때문에 분류해야 하는 데이터의 분포가 고르지 않더라도 그 차이를 완화할 수 있어 불균형한 데이터셋의. Macro F1 점수는 토큰들을 단어로 합치지 않은 토큰 상태에서 한 번(Token F1), 단어별로 가장 앞 토큰만을 선택하여 단어로 합친 후 한 번(Word F1) 측정한다.

발화 의도의 성능 평가에는 정확도를 사용한다. 정확도는 모델이 의도의 개수와 종류를 모두 정확하게 추측하였을 경우에만 옳은 것으로 판단하고, 둘 중 하나라도 틀렸다면 틀린 것으로 간주한다. [11]에서 원래 실험하였던 영문 데이터셋과 같이 선택해야 하는 의도가 하나로 고정되어 있을 때는 모델이 내놓는 의도별 확신도(Confiednce) 중 가장 큰 것을 선택한다. 하지만 본 논문에서 구축한 데이터셋에서는 문장에 따라 하나 혹은 두 개의 의도를 예측해내야 하므로, 위와 같은 방법을 사용할 수 없다. 따라서 여기서는 0과 1사이의 확신도에서 0.4가 넘는 모든 의도를 모델이 예측한 의도로 간주하였다.

4.3 실험 결과 및 분석

아래의 Table 6.은 [11]에서 제시했던 원본 영어 데이터셋을 이용한 실험과, KoBERT 및 본 논문에서 구축한 데이터셋을 이용한 실험의 결과를 정리한 것이다.

Table 6. Results of model of [11] in each dataset.

사용 데이터셋	발화의도	슬롯(토큰)	슬롯(단어)
본 논문 구축	95.0	83.7	83.4
Snips	98.6	-	97.0
ATIS	97.5	-	96.1

먼저 전체적으로 본 논문에서 구축한 데이터셋에서 전체적으로 성능이 낮게 나오는 것을 확인할 수 있다. 특히 슬롯 채우기의 F1 점수 차이가 더 크게 벌어졌다. 언어별로 데이터셋 뿐 아니라 한국어를 Sentencepiece 로 토큰화한 것이나 KoBERT를 사용한 등의 차이점도 있지만, 토큰 단위와 단어 단위의 성능 차이가 크게 나지 않는 것으로 보아 Sentencepiece의 영향력은 거의 없다고 할 수 있고, KoBERT 역시 초기 학습만 다를 뿐 모두 fine-tuning 과정을 거치므로 초기 학습의 영향력은 그렇게 크다고 볼 수 있다. 따라서 영어에서 좋은 성능을 보인 모델이라 하더라도 다른 언어에서도 비슷한 수준의 결과를 보인다고 장담할 수 없고, 따라서 언어에 따라서 각 언어에 최적화된 슬롯 채우기 모델을 별도로 구성할 필요가 있음을 알 수 있다.

발화 의도의 정확도 역시 조금 영어 데이터셋에 비해 조금 낮은 값을 보였지만, 문장의 의도를 1개만 골라야 하는 영어 데이터셋에 비해 최대 2개까지 정확하게 맞춰야 하는 과제로 난이도가 상승한 만큼 이는 예측 가능한 성능 하락이라고 할 수 있다.

5. 결론

한국의 자연어처리 연구자들은 영어에 비해 한국어 데이터셋을 구하는 것이 어려워 보통 영어로 된 데이터셋을 이용하여 연구를 해왔다. 본 논문에서는 한국어로 발화 의도 및 슬롯 채우기를 수행할 때 사용할 수 있는 한국어 데이터셋을 제시하였다. 본 데이터셋의 6857개의 문장에는 5가지 발화 의도 중 최대 2개가 부착되었고, 단어별 슬롯의 종류는 7가지 중 하나로 표기되어 있다.

본 논문에서 구축한 데이터셋을 영어로 연구된 모델에 학습시킨 결과, 발화 의도 추측 정확도는 조금 하락하였고 슬롯 채우기 F1 점수는 큰 차이를 보였다. 영어 데이터셋으로 좋은 성능을 낸 모델이 곧바로 한국어 문장에서 좋은 성능을 내지 못하는 만큼 한국어를 잘 처리하기 위한 모델을 따로 설계해야 할 필요가 있고, 이를 위해서는 본 논문에서 제시한 것과 같은 한국어 데이터셋이 필요함을 알 수 있다.

이번에 구축한 데이터셋은 총 문장수가 타 데이터셋에 비해 부족한 측면이 있으므로, 이를 보강하기 위한 추가 수집 및 정제를 수행하는 후속 연구를 계속할 예정이다. 또한, 이번 연구에서 제시한 한국어 데이터셋을 바탕으로 더 높은 성능의 한국어 발화 이해를 수행하는 모델을 제작하는 것도 계획하고 있다.

REFERENCES

[1] S. Yu, N. Kulkarni, H. Lee, & J. Kim. (2017). Syllable-level neural language model for agglutinative language. *arXiv preprint, arXiv:1708.05515*.

[2] Y. Kim. (2014). Convolutional neural networks for sentence classification. *arXiv preprint, arXiv:1408.5882*.

[3] Z. Zhao & Y. Wu. (2016). Attention-based convolutional neural networks for sentence classification. *INTERSPEECH,* 705-709.

[4] S. Hochreiter & J. Schmidhuber, (1997). Long short-term memory. *Neural computation, 9(8)*, 1735-1780.

[5] K. Yao, B. Peng, Y. Zhang, D. Yu, G. Zweig, & Y. Shi. (2014). Spoken language understanding using long short-term memory neural networks. *2014 IEEE Spoken Language Technology Workshop (SLT),* 189-194.

[6] Y. B. Kim, S. Lee, & K. Stratos. (2017). Onenet: Joint domain, intent, slot prediction for spoken language understanding. *IEEE Automatic Speech Recognition and Understanding Workshop(ASRU)*, 547-553.

[7] Z. Huang, W. Xu, and K. Yu. (2015). Bidirectional lstm-crf models for sequence tagging. *arXiv preprint, arXiv:1508.01991*.

[8] B. Liu & I. Lane. (2016). Attention-based recurrent neural network models for joint intent detection and slot filling. *arXiv preprint, arXiv:1609.01454*.

[9] J. Devlin, M. W. Chang, K. Lee, & K. Toutanova. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint, arXiv:1810.04805*.

[10] A. Vaswani, et al. (2017). Attention is all you need. *In Advances in neural information processing systems*, 5998-6008.

[11] Q. Chen, Z. Zhuo, & W. Wang. (2019). Bert for joint intent classification and slot filling. *arXiv preprint, arXiv:1902.10909*.

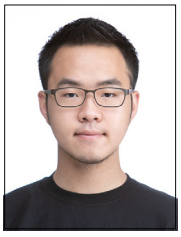
[12] SKT-Brain. (2019). *KoBERT*, GitHub[Online], https://github.com/SKTBrain/KoBERT

[13] J. Oh, S. Jo, Y. Lim, & Y.S. Choi. (2018). Improving Utterance Intent Classification via Hierarchical Attention-based Recurrent Neural Network. *The Korean Institute of Information Scientists and Engineers,* 575-577.

[14] K. Park, S. Na, J. Shin, & Y. Kim. (2019). BERT for Korean Natural Language Processing: Named Entity Tagging, Sentiment Analysis, Dependency Parsing and Semantic Role Labeling. *The Korean Institute of Information Scientists and Engineers,* 584-586.

[15] A. So, K. Park, & H. Lim. (2018). A study on building korean dialogue corpus for restaurant reservation and recommendation. *Annual Conference on Human and Language Technology*, 630-632.

한 승 규(Seunggyu Han) [학생회원]



· 2019년 2월 : 대구경북과학기술원 융복합대학 기초학부(공학학사)
· 2019년 3월 ~ 현재 : 고려대학교 컴퓨터학과 석사과정
· 관심분야 : 자연어처리, 발화 이해
· E-Mail : seunggyu-han@korea.ac.kr

임 희 석(HeuiSeok Lim) [종신회원]

· 1992년 2월 : 고려대학교 컴퓨터학과 (이학학사)
· 1994년 2월 : 고려대학교 컴퓨터학과 (이학석사)
· 1997년 8월 : 고려대학교 컴퓨터학과 (이학박사)
· 2008년 ~ 현재 : 고려대학교 정보대학 컴퓨터학과 교수
· 관심분야 : 자연어처리, 뇌신경 언어 정보 처리
· E-Mail : limhseok@korea.ac.kr