

Error Correction for Korean Speech Recognition using a LSTM-based Sequence-to-Sequence Model

Hye-won Jin*, A-Hyeon Lee*, Ye-Jin Chae*, Su-Hyun Park*, Yu-Jin Kang*, Soowon Lee*

*Student, School of Software, Soongsil University, Seoul, Korea

*Student, School of Software, Soongsil University, Seoul, Korea

*Student, School of Software, Soongsil University, Seoul, Korea

*Student, School of Software, Soongsil University, Seoul, Korea

*Student, School of Software, Soongsil University, Seoul, Korea

*Professor, School of Software, Soongsil University, Seoul, Korea

[Abstract]

Recently, since most of the research on correcting speech recognition errors is based on English, there is not enough research on Korean speech recognition. Compared to English speech recognition, however, Korean speech recognition has many errors due to the linguistic characteristics of Korean language, such as Korean Fortis and Korean Liaison, thus research on Korean speech recognition is needed. Furthermore, earlier works primarily focused on editorial distance algorithms and syllable restoration rules, making it difficult to correct the error types of Korean Fortis and Korean Liaison. In this paper, we propose a context-sensitive post-processing model of speech recognition using a LSTM-based sequence-to-sequence model and Bahdanau attention mechanism to correct Korean speech recognition errors caused by the pronunciation. Experiments showed that by using the model, the speech recognition performance was improved from 64% to 77% for Fortis, 74% to 90% for Liaison, and from 69% to 84% for average recognition than before. Based on the results, it seems possible to apply the proposed model to real-world applications based on speech recognition.

▶ **Key words:** Speech Recognition, Error Correction, Korean, LSTM, Sequence-to-Sequence, Bahdanau Attention

-
- First Author: Hye-won Jin, Corresponding Author: Soowon Lee
 - *Hye-won Jin (hwj0338@soongsil.ac.kr), School of Software, Soongsil University
 - *A-Hyeon Lee (ahyeon0508@soongsil.ac.kr), School of Software, Soongsil University
 - *Ye-Jin Chae (yejin013@soongsil.ac.kr), School of Software, Soongsil University
 - *Su-Hyun Park (ttngus8115@soongsil.ac.kr), School of Software, Soongsil University
 - *Yu-Jin Kang (yoojin424@soongsil.ac.kr), School of Software, Soongsil University
 - *Soowon Lee (swlee@ssu.ac.kr), School of Software, Soongsil University
 - Received: 2021. 06. 23, Revised: 2021. 08. 30, Accepted: 2021. 08. 31.

[요 약]

현재 대부분의 음성인식 오류 교정에 관한 연구는 영어를 기준으로 연구되어 한국어 음성인식에 대한 연구는 미비한 실정이다. 하지만 영어 음성인식에 비해 한국어 음성인식은 한국어의 언어적인 특성으로 인해 된소리, 연음 등의 발음이 있어, 비교적 많은 오류를 보이므로 한국어 음성인식에 대한 연구가 필요하다. 또한, 기존의 한국어 음성인식 연구는 주로 편집 거리 알고리즘과 음절 복원 규칙을 사용하기 때문에, 된소리와 연음의 오류 유형을 교정하기 어렵다. 본 연구에서는 된소리, 연음 등 발음으로 인한 한국어 음성인식 오류를 교정하기 위하여 LSTM을 기반으로 한 인공 신경망 모델 Sequence-to-Sequence와 Bahdanau Attention을 결합하는 문맥 기반 음성인식 후처리 모델을 제안한다. 실험 결과, 해당 모델을 사용함으로써 음성인식 성능은 된소리의 경우 64%에서 77%, 연음의 경우 74%에서 90%, 평균 69%에서 84%로 인식률이 향상되었다. 이를 바탕으로 음성인식을 기반으로 한 실제 응용 프로그램에도 본 연구에서 제안한 모델을 적용할 수 있다고 사료된다.

▶ **주제어:** 음성인식, 오류 교정, 한국어, LSTM, Sequence-to-Sequence, Bahdanau Attention

I. Introduction

음성인식(Speech Recognition)은 사람의 음성언어를 컴퓨터가 해석하여 문자언어로 변환하는 과정을 말하며, STT(Speech-to-Text)라고도 한다. 음성을 통한 전달이 다양한 소통방법 중에서도 가장 일반적으로 사용되는 만큼, IOT 시장 등 많은 분야에서 음성인식 기술을 활발히 활용하고 있다. 하지만 현재의 음성인식 기술은 음성을 완벽히 변환하지 못하고 의도와 다른 오류가 나타나기도 한다. 특히 한국어는 된소리나 연음 등 그 언어적인 특성으로 인하여 아직도 많은 음성인식 오류가 나타나고 있다.

본 연구에서는 한국어 음성인식 오류를 다음과 같은 5가지 유형으로 구분한다. 첫 번째 오류 유형은 고유 명사를 인식하지 못하여 생긴 오류이다. 이는 사전에 등록되어 있지 않은 고유명사를 비슷한 발음의 다른 단어로 인식하여 생긴 오류로, 개체명 인식 등에 대한 연구가 필요하다. 두 번째 오류 유형은 비슷한 발음의 다른 단어로 잘못 인식하는 오류이다. 이와 같은 오류 교정을 위해서는 주위 문맥을 파악하여 오류 교정을 할 필요가 있다. 세 번째 오류 유형은 인식되어야 할 단어가 누락된 오류이다. 이는 음성을 입력하는 상황에서 잡음이 많거나 해당 단어의 소리가 작아서 인식이 되지 않아 생기는 오류이다. 네 번째 오류 유형은 된소리가 아닌 발음을 된소리로 인식하는 오류(이하 된소리 인식 오류)이다. 주로 연결어미 '-고'가 '-꼬'로 변형되는 경우이다. 다섯 번째 오류 유형은 연음 법칙으로 생기는 오류(이하 연음 인식 오류)이다. 앞 음절의 끝소리가 뒤 음절의 첫소리로 올라가면서 '깎아'를 '까까', '걸음'을 '거름' 등으로 인식하는 경우이다.

한국어 음성인식 오류 교정은 일반적으로 음성인식 후처리에 의해서 수행된다. 기존 연구에서는 Levenstein의 편집 거리 알고리즘과 해싱 알고리즘의 조합[1], 품사 정보를 이용한 어절 재결합 기법을 기본으로 음절 bigram 및 4-gram 알고리즘[2] 등을 사용하여 음성인식 오류를 교정하였다. 이 경우 일부 오류는 높은 정확도로 교정이 이루어지지만 위에서 기술한 오류 유형들에 대한 정확한 교정은 어렵다는 문제점이 있다.

본 연구에서는 위에서 기술한 5가지 문맥 기반 오류 유형 중 된소리 인식 오류 및 연음 인식 오류를 문맥 기반 딥러닝을 통하여 교정하는 음성인식 후처리 모델을 제안한다. 제안 모델은 문맥 기반 딥러닝을 위하여 기계 번역에서 널리 사용되고 있는 LSTM을 사용한 Seq2seq(Sequence-to-Sequence)[3] 모델과 Bahdanau Attention[4] 기법을 함께 사용한다.

본 논문의 구성은 2장에서 오류 교정 모델들을 분석하고, 3장에서는 본 연구에서 제안하는 딥러닝 모델을 사용한 문맥 기반 오류 교정 모델을 설명한다. 4장에서는 실험 결과를 제시하고 마지막으로 5장에서는 결론과 향후 연구를 기술한다.

II. Preliminaries

1. Related works

음성인식 후처리 관련 기존 연구는 주로 편집 거리 알고리즘이나 음절 복원 규칙 등을 사용한다. [5]에서는

Levenshtein의 편집 거리 알고리즘[6]을 사용하여 음성인식 후처리를 수행하였다. 기본 아이디어는 Levenshtein의 편집 거리 알고리즘에 인식된 인명과 인명사전의 각 후보 인명들 간의 자소 편집 거리를 baseline으로 하고, 음절 Unigram 편집 거리, 음절 Bigram 편집 거리를 추가적으로 결합하여 오류를 교정하였다. 특히 고유 명사에 특화하여 발음 변이 현상(초성 “ㅇ” 무음화, 자음동화와 비음화, 자음 축약)을 적용하였다. 성능평가 결과, 발음 변이 현상을 적용하였을 때 약 5~10%의 성능 향상을 보였다. 또한 음절 Unigram 편집 거리와 음절 Bigram 편집 거리를 합친 모델의 성능이 발음 변이 적용 후 93%로 정확도가 가장 높은 것으로 평가되었다.

[7]의 연구에서는 음절 복원 규칙과 형태소 분석을 이용하여 음성인식 후처리를 진행하였다. 연속된 음성을 조사, 어미, 어간 사전과 음운 규칙으로 처리할 수 없는 단어들을 위한 예외 사전을 구성하고, 사전을 사용하여 한국어 음운 규칙을 역적용하여 복원을 수행하였으며, 평가결과 음절 단위로는 94%, 음절 단위로는 90%의 복원율이 산출되었다.

III. Context-Sensitive Error Correction Model

본 연구에서는 문맥 기반으로 음성인식 오류를 교정하는 음성인식 후처리 모델을 제안한다. 문맥 기반 오류는 문맥 상 어울리지 않는 어절의 오류와 ‘할꺼다’ 등의 된소리 인식 오류, ‘해씨’ 등의 연음 인식 오류로 정의한다. 문맥 상 어울리지 않는 어절의 오류는 어절 단위로 보았을 때는 오류가 아니지만 문맥 상으로 어울리지 않는 어절의 오류를 의미한다. 이에 대한 예시로 고유 명사를 인식하지 못하여 생긴 경우, ‘같이’, ‘가치’처럼 어절 단위로 보았을 때는 오류가 없지만 문맥에 어울리지 않는 경우, 인식되어야 할 단어가 누락되어 문맥이 어색한 경우가 있다.

이와 같은 문맥 기반 오류 중 된소리 인식 오류 및 연음 인식 오류의 교정을 위하여 본 연구에서 제안하는 모델은 LSTM 기반 Seq2seq 모델과 Bahdanau Attention 기법을 사용한다. Seq2seq는 RNN을 기반으로 한 인코더와 디코더로 구성된다. 인코더는 입력 시퀀스를 문맥이 포함된 컨텍스트 벡터로 변환하여 디코더로 전송한다. 컨텍스트 벡터를 통하여 디코더가 출력 시퀀스를 예측한다는 이러한 특징은 문맥 기반 오류 교정에 적합한 점이다. 또한 기존 RNN 모델과는 달리 인코더에서의 입력 시퀀스 길이와 디코더의 출력 시퀀스 길이가 자유롭다는 장점을 가지고

있어 본 연구의 모델로 채택하였다. Seq2seq는 위와 같은 장점들로 인하여 입력된 시퀀스로부터 다른 도메인의 시퀀스를 출력하는 챗봇(Chatbot)이나 기계 번역(Machine Translation) 등 다양한 분야에서 사용된다.

특히 본 연구가 LSTM을 사용하는 이유는 기존 RNN(Recurrent Neural Network)의 경우 입력 시퀀스의 길이와 출력 시퀀스의 길이의 연관성을 알 수 없을 때 RNN을 적용하기에는 한계가 있기 때문이다. 이를 해결하기 위해서는 RNN을 이용해 입력에 대한 고정된 차원의 벡터를 생성하거나 해당 벡터를 입력으로 제공하여 다른 시퀀스를 결과로 구하는 방법 등이 있지만, RNN은 장기 의존성(Long-term Dependency)을 학습하기 어렵다는 문제점이 있으므로 이를 해결하기 위해 LSTM을 사용한다.

또한 Seq2seq가 가진 정보 손실 및 기울기 소실 문제(Vanishing Gradient Problem)를 보완하기 위해 Bahdanau Attention을 함께 사용한다. Attention은 인코더에서 전체 입력문장 중 예측부분과 연관된 입력을 위주로 하여 디코더의 Time Step마다 다시 참고한다. Attention의 추가는 입력문장의 길이에 맞게 컨텍스트 벡터를 개선하며, 마지막 은닉상태 벡터뿐만 아니라 전체 은닉상태 벡터 또한 전달 받아 성능을 향상시킨다. 이 때 사용하는 Attention 스코어 함수에 따라 다양한 Attention 종류가 있는데, 본 연구에서는 컨텍스트 벡터와 임베딩 된 벡터를 연결(concatenate)하는 방법인 Bahdanau Attention을 적용하여 모델을 개선한다.

Fig 1은 본 연구에서 제안하는 모델의 구조도이다.

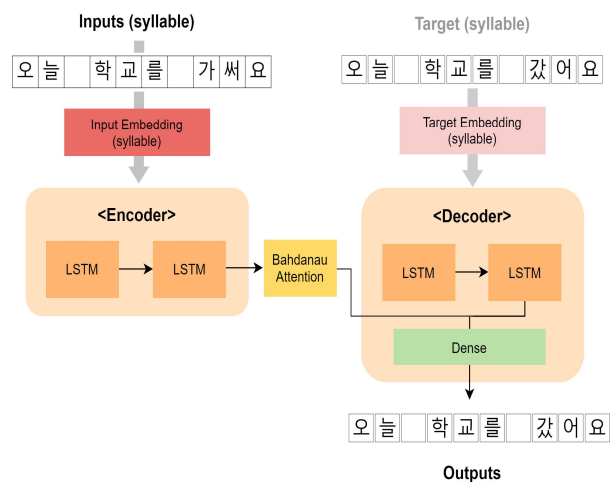


Fig. 1. Structure of the proposed LSTM-based Sequence-to-Sequence Model

Fig 1에서 LSTM 기반의 Seq2seq 모델은 교정해야 하는 문장을 입력받아 음절 단위로 문장을 임베딩하고 인코더(Input Embedding)에 입력한다. 인코더에는 입력 받은

음절 단위로 임베딩된 문장을 2개의 LSTM 레이어를 거쳐 입력 문장의 의미 정보를 출력값으로 내보낸다. 그 후 Bahdanau Attention을 거쳐 해당 값이 디코더로 전달된다. 디코더는 입력 문장에 대응하는 올바른 문장을 음절 단위로 임베딩(Target Embedding) 하여 입력 받는 Teacher Force 방식을 적용하여 Attention 레이어에서 받은 벡터값과 함께 2개의 LSTM을 거쳐 계산되고 Dense 레이어를 통해 가장 높은 확률을 가지는 음절들을 선택해 올바른 문장을 완성한다.

IV. Experiment and Evaluation

본 연구에서는 ‘한국어 특성 기반의 STT 엔진 정확도를 위한 정량적 평가 방법 연구’[8], ‘음성인식 오픈 API의 음성인식 정확도 비교 분석’[9], ‘음성인식 오픈 API 성능 비교’[10]를 참고하여 한글의 오인식률 28%의 성능을 지닌 ETRI 한국어 음성인식기[11]로 실험을 진행하였다. ETRI 한국어 음성인식기는 16kHz Sampling Rate로 녹음된 음성신호에 대해 문자(텍스트)로 제공하고 있다. ETRI 한국어 음성인식기를 통해 음성인식한 결과 Table 1과 같은 음성인식 오류가 발견되었다.

Table 1. Recognition Results by Speech Recognition Error Type

Error Type	Original Texts	Recognition Results
Proper Noun Error	제가 숭실대학교에 지원하게 된 이유는	제가 숭실 대학교에 지원하게 된 이유는
	저는 아토라는 동아리에서	저는 앞으로는 동아리에서
Wrong Word Error	여러 프로젝트 수업이 많은 것이 제가 많은 도움을 받을 수 있겠다라는	여러 프로젝트 수 없이 많은 것이 제가 많은 도움을 받을 수 있겠다라는
	제 좌우명과 제 꿈이 동일한 선상에 있지 않다고	제 좌우명과 제 꿈이 동일한 현상이 있지 않다고
Missing Word Error	마케팅은 판매자, 즉 이익을	마케팅은 판매자 이익을
	내 뒤에 얼른 숨어	내 얼른 숨어
Fortis Error	온라인 스터디 게시판에 보이고, 만들고, 퀴즈를 생성하고,	온라인 스터디 게시판에 보이고 만들꼬 기점을 생성하고
	시골보다 도시가 생활비가 훨씬 많이 들어요	시골보다 도시가 생활비가 훨씬 많이 들어요
Liaison Error	파이썬 공부를 했어요.	파이썬 공부를 해써요
	병원에 입원해 있는 동안 아내가 저를 간호했어요.	병원에 입원해 있는 동안 아내가 저를 간호해써요

본 연구에서는 Table 1의 오류 중 된소리 인식 오류(Korean Fortis Error)와 연음 인식 오류(Korean Liaison Error)를 교정하였다.

학습 및 평가를 위한 데이터 셋은 KSS Dataset: Korean Single speaker Speech Dataset[12]에서 제공한 음성 파일 및 텍스트 원문을 사용하였다. 음성 파일은 44,100kHz로 제공되어 있어 ETRI 한국어 음성인식기 사용을 위해 16kHz Sampling Rate로 Down Sampling 하였다.

실험을 위하여 학습 데이터에서 음성 파일 부분을 ETRI 한국어 음성인식기로 인식한 결과와 텍스트 원문을 Key-Value 형식으로 저장하였으며, 본 연구에서 교정하고자 하는 오류의 데이터만을 선별하여 진행하였다. 또한 해당 데이터를 이용해 비슷한 데이터 셋을 추가적으로 구축하였다. 이를 통해 구축된 데이터는 총 4800개이며 Train Data Set과 Test Data Set은 8:2의 비율로 나누어 실험을 진행했다.

평가 데이터에 대한 ETRI 한국어 음성인식기의 인식률을 [식 1]과 같이 어절 단위로 평가하였고, 평가 결과 약 69%의 인식률을 보였다.

$$\text{문장별 인식률} = \frac{\text{바르게 인식된 어절 수}}{\text{문장내 총 어절 수}} \quad [\text{식 1}]$$

사전 모델 결정 실험은 3장에서 제안한 LSTM 기반 Seq2seq과 Bahdanau Attention을 결합하여 진행하였으며, Learning Rate를 0.001로 하였다.

또한 모델의 성능을 높이고자 최적의 파라미터를 추정하는 실험을 진행하였다. 실험에서 선정한 파라미터는 총 3개로, 배치 크기(Batch Size), 임베딩 크기(Embedding Size), LSTM 레이어 개수(Number of LSTM Layers)를 선정하였다.

Table 2는 사전 모델 결정 실험에서 설정한 파라미터를 정리한 표이다. 먼저, 배치 크기는 한 번에 네트워크로 넘겨져 학습되는 데이터의 개수를 말한다. 일반적으로 32와 64를 사용하며, 실험의 첫 번째 파라미터로 설정하였다. 임베딩 크기는 조밀한 차원에 단어의 의미를 고려한 벡터로 표현할 때의 차원 크기이다. 본 실험에서는 임베딩을 200, 300으로 하여 두 번째 파라미터로 설정하였다. 마지막으로 LSTM 레이어 개수는 인코더와 디코더의 신경망에서 사용되는 셀의 개수이며, 본 실험에서는 2, 3, 4, 5로 레이어 개수를 설정하였다.

Table 2. Parameter Settings for the Experiments

Batch Size	32		64	
Embedding Size	200		300	
Number of LSTM Layers	2	3	4	5

본 연구에서는 [13] 등 철자오류 교정 연구에서 주로 사용하는 Recall, Precision 그리고 F1-score를 평가지표로 사용하였다. [식 2]는 Recall, [식 3]은 Precision, 그리고 [식 4]는 F1-score에 대한 수식이다.

$$Recall = \frac{\text{바르게 교정한 수}}{\text{오류 데이터 수}} \quad \text{[식 2]}$$

$$Precision = \frac{\text{바르게 교정한 수}}{\text{교정한 데이터 수}} \quad \text{[식 3]}$$

$$F1\text{-score} = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \text{[식 4]}$$

Fig 2 ~ 5는 사전 모델 결정을 위한 실험 결과의 일부이다. 각 평가지표는 문장 단위로 계산하였으며, Fig 2와 Fig 3은 파라미터를 기준으로 정리한 그림이고, Fig 4와 Fig 5는 레이어를 기준으로 정리한 그림이다.

평균적으로 가장 높은 성능을 보였던 2레이어의 하이퍼 파라미터 실험 결과를 Fig 2와 Fig 3에 나타내었다. Fig 2는 된소리일 경우 2레이어일 때의 모델 성능이며, 배치 32, 임베딩 300에서 성능이 가장 뛰어났다.

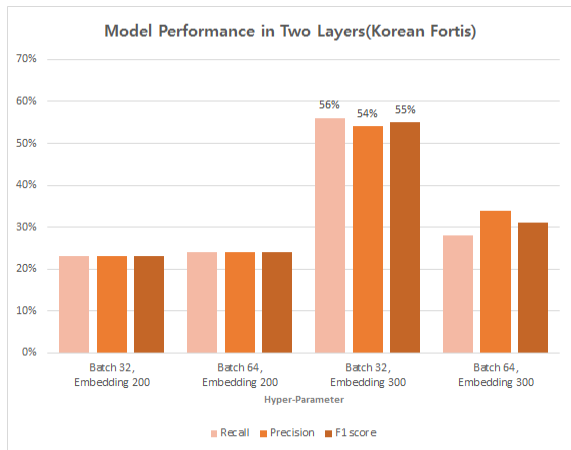


Fig. 2. Model Performance in Two Layers(Korean Fortis)

Fig 3은 연음일 경우 2레이어일 때의 모델 성능이다. 된소리와 같이 배치 32, 임베딩 300에서 성능이 가장 뛰어났다.

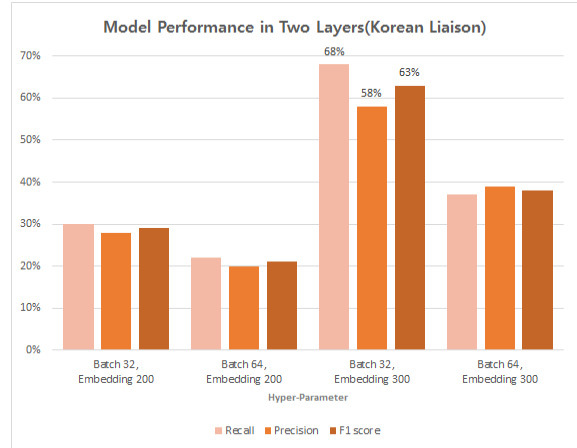


Fig. 3. Model Performance in Two Layers(Korean Liaison)

Fig 4는 된소리일 경우 레이어 개수별 모델 성능이다. 2레이어일 때 된소리의 F1 score값이 55%로 가장 높았다.

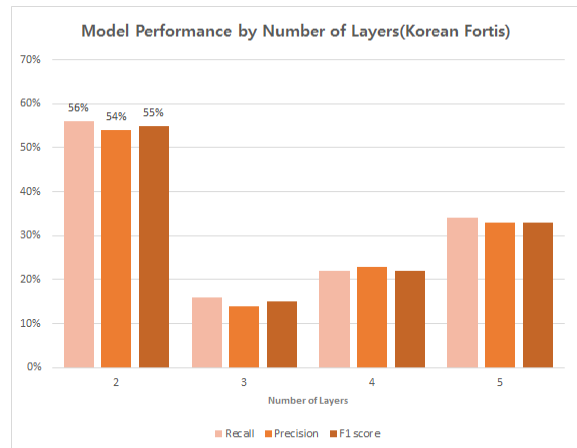


Fig. 4. Model Performance by Layer Count(Korean Fortis)

Fig 5는 연음일 경우 레이어 개수별 모델 성능이다. 2레이어일 때 연음의 F1 score값이 63%로 가장 높았다.

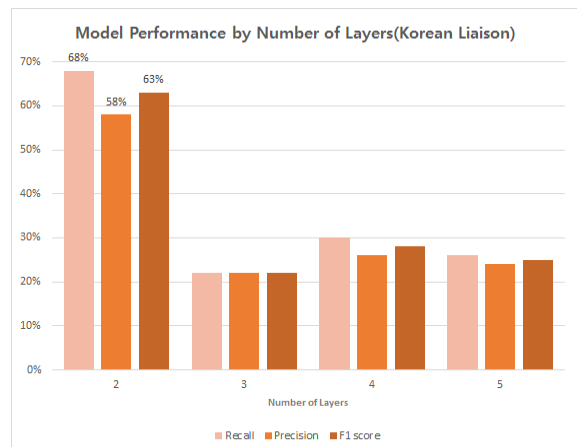


Fig. 5. Model Performance by Layer Count(Korean Liaison)

사전 모델 결정 실험 결과로부터 2레이어, 배치 32, 임베딩 300 파라미터를 가진 모델을 결정하였다.

Table 3은 된소리와 연음의 실험 결과 중 일부의 교정 결과를 표로 나타낸 것이다. 된소리에서는 '문제일찌', '알찌도' 와 같은 오류가 나타났고, 연음에서는 '가씨요', '업씨' 와 같은 오류가 나타났다. 2레이어에서는 위와 같은 오류가 각각 '문제일지', '알지도', '갔어요', '없어'로 알맞게 교정이 된 것을 확인할 수 있다. 하지만 다른 레이어에서는 오류를 교정하지 못한 결과가 나타났을 뿐만 아니라 오류가 나타나지 않은 단어가 변경된 경우도 발생하였다. 따라서 결과 문자열이 입력 문자열의 길이에 대응해서 벗어나는 경우 입력 문장과 같다고 취급하여 교정 결과에 일어날 수 있는 일부 치명적인 오류를 방지할 수 있었다. 본 실험에서 2레이어의 성능이 다른 레이어의 성능보다 우수한 것은 본 연구의 학습 데이터 수가 많지 않아 레이어 수가 늘어남에 따라 오히려 부적절한 교정 규칙을 학습하게 되기 때문인 것으로 사료된다.

Table 3. Correction Results of Proposed Model by Number of Layers

# of Layers	Error Type	Input Sentences	Correction Results
2	Fortis	뭐가 문제일찌 상상이 간다	뭐가 문제일지 상상이 간다
		잘 알찌도 못하면서 어이었다	잘 알지도 못하면서 어이었다
	Liaison	오늘 학교를 가씨요	오늘 학교를 갔어요
		나는 잘못이 업씨	나는 잘못이 없어
3	Fortis	뭐가 문제일찌 상상이 간다	뭐가가치사람은이 상상상다
		잘 알찌도 못하면서 어이었다	너정 만만 못하면서 어이었다
	Liaison	오늘 학교를 가씨요	오늘 학교를 갔어요
		나는 잘못이 업씨	나는 잘못이 없어
4	Fortis	뭐가 문제일찌 상상이 간다	뭐가 문제일지 상상이 간다
		잘 알찌도 못하면서 어이었다	잘 알지도 못하면서 어이었까
	Liaison	오늘 학교를 가씨요	오늘 학교를 갔어요
		나는 잘못이 업씨	나는 잘못이 없어
5	Fortis	뭐가 문제일찌 상상이 간다	뭐 남제일 문제일지 간정
		잘 알찌도 못하면서 어이었다	잘 알지도 못하면서 하이일 다
	Liaison	오늘 학교를 가씨요	오늘 학교를 갔어요
		오늘 별다른 계획 업씨요	오늘 별다른 계획 없어요

Table 4는 사전 모델 결정 실험으로부터 최적의 파라미터를 설정하여 제안한 모델로 음성인식 오류를 교정한 결

과를 표로 나타낸 것이다. 제안 모델을 통하여 음성인식 오류를 교정한 결과, [식 1]에 따른 인식률이 된소리 64%에서 77%, 연음 74%에서 90%, 평균 69%에서 84%로 향상되었다.

Table 4. Speech Recognition Correction Results

Classification	STT Recognition Rate	Recognition Rate after Correction
Fortis	64%	77%
Liaison	74%	90%
Average	69%	84%

V. Conclusion

본 연구에서는 LSTM을 기반으로 한 Seq2seq 모델과 Bahdanau Attention 기법을 사용한 문맥 기반 음성인식 오류 교정 모델을 제안하였다. 특히 다양한 한국어 음성인식 오류 중 된소리 인식 오류와 연음 인식 오류에 대해 교정 실험을 하였다. 문장별로 바르게 인식된 어절을 계산하여 제안 모델 성능을 평가한 결과 평균 69%에서 84%로 음성인식의 성능 향상을 확인할 수 있었다. 하지만 Seq2seq의 특징상 오류와 관련된 데이터가 적거나 없을 경우 완전히 해결되지 않는 오류도 존재하였다.

향후 연구에서는 더 많은 학습 데이터 및 다양한 유형의 데이터 구축을 통해 본 연구가 제안한 방식의 추가적인 성능 향상을 기대할 수 있을 것이다. 또한 본 연구에서 다루었던 된소리 인식 오류와 연음 인식 오류 외 고유 명사 오류 등 다양한 오류 유형들에 대한 교정 연구가 필요하다.

ACKNOWLEDGEMENT

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2021-2018-0-01419) supervised by the IITP(Institute for Information & communications Technology Promotion)

REFERENCES

[1] K. Nam, "A Study on Processing of Speech Recognition Korean Words," JCCT, vol. 5, no. 4, pp. 407-412, Nov. 2019,

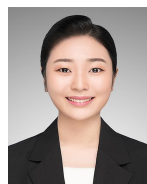
doi:10.17703/JCCT.2019.5.4.407.

- [2] Dong-Hee Lim, Seung-Shick Kang, Du-seong Chang, "Word spacing Error Correction for the Postprocessing of Speech Recognition", Proceedings of the Korean Information Science Society Conference, vol.33, no. 1(B), pp.25-27, June. 2006.
- [3] Ilya Sutskever, Oriol Vinyals, Quoc V. Le, "Sequence to Sequence Learning with Neural Networks", Neural Information Processing Systems, pp.3104-3112, 2014, arXiv:1409.3215.
- [4] D. Bahdanau, K. Cho, Y. Bengio, "Neural machine translation by jointly learning to align and translate", Proc. Int. Conf. Learn. Representations, 2014, 2014. arXiv:1409.0473.
- [5] Ye-jin Kim, Young-min Park, Sang-woo Kang, Sang-keon Jung, Cheong-jae Lee, Jung-yun Seo, "Post-Processing of Speech Recognition Using Phonological Variables and Improved Edit-distance", Annual Conference on Human and Language Technology, pp.9 - 12, Oct. 2014.
- [6] Eiichi Tanaka and Tamotsu Kasai, "Synchronization and Substitution Error-correcting codes for the Levenshtein Metric", IEEE Trans. Information Theory, Vol.IT-22, No.2, pp.156-176, Mar. 1976, doi: 10.1109/TIT.1976.1055532.
- [7] Sang-Hyun Seo, Jae-Hong Kim, Hae-Jin Kim, Mi-Jin Kim, "Post-Processing of Voice Recognition Using Phonologic Rules and Morphologic analysis", Annual Conference on Human and Language Technology, pp.495-499, Oct. 1997.
- [8] So-Yeon Min, Kwang-Hyong Lee, Dong-Seon Lee, Dong-Yeop Ryu, "A Study on Quantitative Evaluation Method for STT Engine Accuracy based on Korean Characteristics", Journal of Korea Academia-Industrial cooperation Society, Vol.21, No 7, pp.699-707, Jul 2020, doi:10.5762/KAIS.2020.21.7.699.
- [9] Seung-joo Choi, Jong-bae Kim, "Comparison Analysis of Speech Recognition Open APIs' Accuracy", Asia-pacific Journal of Multimedia Services Convergent with Art, Humanities, and Sociology, Vol.7, No.8, pp. 411-418, Jul 2017, doi:10.35873/ajm ahs.2017.7.8.038.
- [10] Hyun-Woo Oh, Koen-Nyeong Lee, Dong-Suk Yook, "Performance Comparison Of Open Apis For Speech Recognition", The Journal Of The Acoustical Society Of Korea, 2019.
- [11] ETRI, ETRI Speech Recognition, <https://aiopen.etri.re.kr>.
- [12] Kyubyong Park, KSS Dataset: Korean Single Speaker Speech Dataset, <https://kaggle.com/bryanpark/korean-single-speaker-speech-dataset>.
- [13] Seung-Hyeon Park, "Correction of Korean Spelling Errors Using Cosine Similarity Algorithm", Graduate School of Industrial Technology Convergence, Chosun University, 2016.

Authors



Hye-won Jin is a student in the School of Software, Soongsil University. She is interested in text generation and MLops.



A-Hyeon Lee is a student in the School of Software, Soongsil University. She is interested in Natural Language Processing and Data Analytics.



Ye-Jin Chae is a student in the School of Software, Soongsil University. She is interested in internet and mobile computing, and cloud computing.



Su-Hyun Park is a student in the School of Software, Soongsil University. She is interested in internet and mobile computing, and UX/UI design.



Yu-Jin Kang is a student in the School of Software, Soongsil University. She is interested in system software.



Soowon Lee received his BS degree in 1982 in Computer Science and Statistics from Seoul National University, his MS degree in 1984 in Computer Science from Korea Advanced Institute of Science and

Technology, and his PhD degree in 1994 in Computer Science from University of Southern California, USA. He joined the faculty of the School of Software at Soongsil University, Korea, in 1995. He is interested in data science and artificial intelligence.