음성합성을 위한 텍스트 음역 시스템과 숫자 음역 모호성 처리

박정연○, 신형진, 육대범, 이재성
충북대학교
parkjeongyeon@cbnu.ac.kr, hjshin@cbnu.ac.kr, daebum1994@cbnu.ac.kr, jasonlee@cbnu.ac.kr

Text Transliteration System and Number Transliteration Disambiguation for TTS

Jeong Yeon Park○, Hyeong Jin Shin, Dae Bum Yuk, Jae Sung Lee
Chungbuk National University

요 약

TTS(Text-to-Speech)는 문자열을 입력받아 그 문자열을 음성으로 변환하는 음성합성 기술이다. 그러나 실제 입력되는 문장에는 한글뿐만 아니라 영단어 및 숫자 등이 혼합되어 있다. 영단어는 대소문자에 따라 다르게 읽을 수 있으며, 단위로 사용될 때는 약어로 사용되는 것이므로, 알파벳 단위로 읽어서는 안 된다. 숫자 또한 함께 사용되는 단어에 따라 읽는 방식이 달라진다. 본 논문에서는 한글과 숫자 및 단위, 영단어가 혼합된 문장을 분류하고 이를 음역하는 시스템을 구성하며 word vector를 이용한 숫자 및 단위의 모호성 해소방법을 소개한다.

주제어: Text-to-Speech, 음역, 모호성 해소

1. 서론

TTS(Text-to-Speech)는 문자 텍스트가 입력되었을 때, 이를 자동으로 음성 변환하여 출력해주는 음성합성 기술을 말한다. TTS는 현재 기술의 완성도가 높아짐에 따라 뉴스 읽기, 책 읽기 및 스마트폰, 통신 등에서부터, 금융, 서비스업 등의 다양한 분야에서 활용되고 있으며, 나아가 언어를 배우는 아이들을 위한 시스템, 시각 장애인을 위한 시스템 등 이미 실생활에서 유용하게 활용할 수 있는 연구가 진행되고 있다[1, 2].

TTS를 위해서는 입력된 문자 텍스트의 영문 및 특수문자, 숫자 및 단위를 한글로 음역할 필요가 있다[3]. 이는 한글을 제외한 영단어, 숫자 및 숫자 단위, 특수문자 등을 어떻게 변환하느냐에 따라서 음성합성 기술을 이용한 자연스러운 발음이 가능하게 될 수 있는지를 의미하게 된다. 예를 들어 “10대 아이들은 빠르게 성장한다.” 라는 문장에서 10대는 ‘십대’로 음역되지만, “주차장에 차량 10대가 있다.” 라는 문장에서는 ‘열대’로 음역된다. 또한 “119” 와 “6.25” 는 어떤 수치를 나타내는 숫자인 동시에 화재와 응급 전화번호나 한국 전쟁이란 의미로 사용될 수 있다. 이러한 모호성을 규칙만 사용하여 모든 경우를 해결하기 힘들다. 본 논문에서는 [4]을 활용하여 음역의 모호성을 해소하는 방법을 제안한다.

2. 관련연구

기존의 음역방법에 대한 연구로는 숫자 및 단위를 한글로 음역하기 위한 연구로 숫자 전치어, 문자 수단위, 숫자 후치어 및 분류사 후치어 등의 정보를 이용해 규칙으로 음역하는 [5]의 연구가 있다. 영단어를 한글로 음역하는 연구로는 규칙 기반 모델, 히든 마코프 모델, 최대 엔트로피 마코프 모델 등이 있다[6, 7, 8]. [6]에서는 입력된 알파벳의 앞 뒤 단어를 고려하여 규칙을 구축하고, 규칙에 맞게 음역하도록 하였다. [7]의 연구에서는 학습 데이터를 이용하여 언어 모델과 번역 모델의 확률 연산으로 음역하였다. [8]의 연구에서는 출력될 발음이 앞 뒤 출력을 고려하여 결정되도록 하였다. 또한 숫자와 영단어를 동시에 음역하는 통합시스템을 구축한 [9]의 연구 등이 있다.

특히 숫자는 규칙을 이용하여 음역하였을 경우, 높은 성능을 보이고 있지만, 문맥에 따라 다르게 읽을 수 있는 모호성이 존재한다. 단어의 모호성을 해소하고자 word vector를 이용하여 모호성을 해소하는 연구인 [4]에서는 word vector를 사용하여 모호성이 있는 단어의 문맥 정보를 학습하고, 학습한 문맥정보와 모호한 음역 대상의 문맥정보의 유사도를 계산하여 모호성을 해소한다.

단어의 중의성을 해소하고자 할 때, word vector를 활용하기 때문에 품질 높은 word vector를 활용할 필요가 있다. [10]의 연구에서는 word vector 생성 시, 실질형태소에 “형식형태소 및 태그”, “형식형태소 태그”, “형식형태소의 클러스터”를 추가하여 학습하여 비교하였다.

3. 텍스트 음역 시스템

본 논문에서 소개하는 시스템은 자동어휘생성 프로그램인 Flex[11]를 이용한다. Flex의 정규표현을 작성하고 입력된 혼합문장으로부터 영단어와 숫자 및 단위, 한글과 그 외 기타 단위를 분류하여 음역한다. 표 1은 이를 통해 음역한 예시다.

표 1. 음역의 예시

입력	음역결과
6권만 남았습니다.	여섯권만 남았습니다.
10% 할인 적용되어	십퍼센트 할인 적용되어
CPR을 할 줄 아시나요	씨피알을 할 줄 아시나요
10km 전방	십킬로미터 전방
season2에서는	시즌투에서는
T map 대중교통	티 맵 대중교통

혼합문장에서 분류된 텍스트는 각각의 음역엔진을 통해 한글로 음역하는데, 한글과 기타 단위 등은 음역 등은 진행하지 않고 출력한다. 숫자는 문맥 및 상황에 따라 '일', '이', '삼' 등의 한자어와 '하나', '둘', '셋' 등의 고유어로 음역한다. 또한 숫자와 영어가 혼합되었을 경우 영어발음의 한글로 음역하도록 하였다. 특정 사건을 가리키는 날짜와 기관 등을 나타내는 숫자 등은 사전을 이용하여 음역하도록 한다.

영단어는 기본적으로 사전과 음차표기 모듈을 구축하여 사전에 없는 단어는 음차표기 시스템에서 처리하도록 한다. 음차표기 시스템은 규칙 기반의 음역방법[6]을 사용하며 일부 규칙을 수정 및 추가하여 사용하였다. 전체 시스템의 구성은 그림 1과 같다.

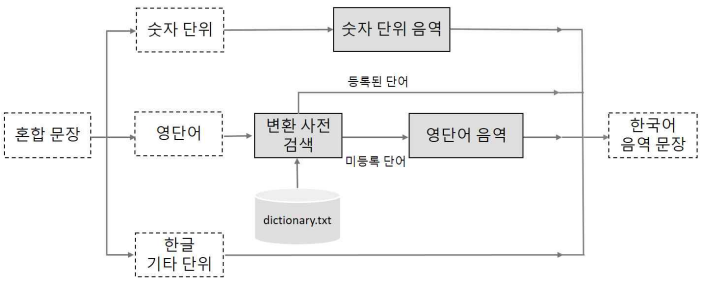

그림 1. 통합 음역시스템 구조

4. 숫자음역 모호성 처리 방법

숫자 단어 음역은 한자어, 고유어, 영어, 고유명사의 4가지의 음역방향(s_i)을 가진다. 국제표준단위를 따르는 단위와 일반적인 숫자는 한자어로 음역된다. 수량이나 순서, 크기등을 나타내는 단위명사가 붙은 숫자는 문맥에 따라 고유어 또는 한자어로 음역되며 영어가 붙은 숫자는 영어발음의 한글로 음역된다. 이 중 "119", "6.25", "10대"와 같은 몇몇 숫자 및 단위는 모호성을 지니고 있다. 이 단어들은 규칙을 이용한 방법과 word vector를 이용하여 문맥의 유사도을 비교한 뒤 높은 유사도의 방향으로 음역하는 방법을 통해 모호성을 해소할 수 있다.

4.1 규칙을 이용한 모호성 해소

한국어에서 숫자를 음역할 때는 주로 숫자와 결합된 숫자 앞뒤에 있는 '~자리', '~개'와 같은 의존명사 등을 보고 음역방향을 결정할 수 있다. 영단어 또는 한글로 쓰인 특정 외래어와 결합된 숫자는 영어발음의 한글로 음역하는 등의 규칙을 구성하여 숫자의 음역방향을 결정할 수 있다[5]. 이 경우 규칙만으로 많은 경우를 한자어와 고유어, 영어발음 한글로 나누어 음역할 수 있다. 다만, 모호성이 있을 때 음역방향의 결정이 어려워진다.

4.2 word vector를 활용한 모호성 해소

숫자의 음역방향이 모호한 모든 문제를 규칙으로 해결하기는 어렵다. 이에 음역의 모호성을 해결하기 위해 [4]에서 제시한 word vector를 이용한 단어 모호성 해소 방법을 적용한다. 즉, 주변 문맥을 학습한 정보를 바탕으로 모호성이 있는 음역 대상의 음역방향을 결정한다.

여기서 모호성을 지닌 임의의 숫자 단어 w가 출현했을 때, 음역방향이 s_i라면 w_{s_i}로 표현한다. 이 때, w_{s_i}는 주변문맥에 있는 윈도우 사이즈 c 이내에 있는 단어들(v_i)의 k차원 word vector 합이다. 이는 식(1)과 같이 표현된다. 이 때, w_{s_i}에 대해 학습할 문장이 n개 있을 때, 이 문장들을 학습한 $\overrightarrow{W_{s_i}}$는 식(2)와 같이 표현된다. 음역이 필요한 모호성 있는 숫자 단어를 $\overrightarrow{q_w}$ 로 표현하였을 때, 식(3)를 이용하여 높은 유사도를 지니는 음역방향 s_i를 선택한다.

학습정보로 사용되는 문맥정보는 명사와 동사, 형용사의 실질 형태소를 활용하며, 본 논문에는 [8]의 모델에서 Skip-gram 방식의 200차원으로 생성한 4종류의 word vector를 사용한다.

$$w_{s_i} = \sum_{i=1}^{2c} Vec(v_i) \quad (1)$$

$$\overrightarrow{W_{s_i}} = \sum_{j=1}^{n} w_{s_{i,j}} \quad (2)$$

$$Sim(\overrightarrow{W_{s_i}}, \vec{q_w}) \equiv \cos(\overrightarrow{W_{s_i}}, \overrightarrow{q_w}) = \frac{\overrightarrow{W} \bullet \overrightarrow{q_w}}{|\overrightarrow{W_{s_i}}| \bullet |\vec{q_w}|} \quad (3)$$

5. 실험 및 평가

5.1 전체 시스템 실험 및 평가

텍스트 음역 시스템의 실험 및 평가를 위하여 다양한 일상대화의 가상시나리오를 구축한 문장[1] 중 숫자 및 단위, 영어를 포함한 음역이 필요한 문장 10,999개를 실

1) 전자부품연구원(KETI) 프로젝트 과정에서 제공받음

험하였다. 그 중 무작위로 1,555 문장을 추출하여 평가 데이터로 사용하였다. 표 2는 규칙만을 이용하여 음역 시스템을 사용하였을 경우의 실험결과다. 평가는 음역 횟수(N) 중 정확히 음역된 횟수(c_i)를 측정하여 정확도를 계산하였다. 여기서 c는 정답 횟수를 뜻하며, i는 무엇을 평가했는지를 의미한다.

$$Accuracy = \frac{c_i}{N} \quad (4)$$

시스템 실험 결과는 표 2와 같이 나타났다.

표 2. 음역 평가 결과(%)

	숫자	영어	전체
정확도	96.15	64.28	74.31

숫자는 모호성처리를 단순 규칙만을 적용하였음에도 높은 결과가 나온 것을 볼 수 있다. 그 중 약 4%p의 오류는 모호성 문제로 인한 문제점이다. 이에 좀 더 정확한 음역을 위하여 word vector를 활용한 모호성 해소를 통해 음역 방향을 적용하고 실험하였다. 영어의 음역문제의 경우에는 숫자 음역에 비해 어려운 문제이므로 다른 연구로 다루도록 한다.

5.2 word vector를 활용한 모호성 해소 평가

word vector를 이용하였을 때 모호성 해결이 가능함을 보이기 위해 모의테스트를 진행하였다. 모의테스트에 사용된 가상데이터는 모호성을 지닌 숫자 단어 '10대', '119', '6.25'의 3가지를 구축하였다. 가상데이터는 각 단어에 대해 2가지 음역방향이 편중되지 않도록 하여 1:1 비율로 구축하였고, 각 단어 당 20개씩 문장을 구축하여 총 60개의 문장을 5-Fold-Test로 평가하였다.

표 3. 모호성 평가 데이터의 일부

숫자 음역	문장
일일구	이제 곧 119가 도착할꺼에요.
백십구	이번 문제의 정답은 119입니다.

평가는 각 숫자 단위가 두 가지 음역방향 중 정확한 음역방향을 결정하는지를 식(4)을 이용하여 정확도를 측정하였으며 단순 규칙과 [10]의 4가지 모델을 사용하여 평가하였다.

규칙 : 규칙을 적용한 숫자 음역
Nul : 실질형태소만으로 생성한 word vector를 활용
Tag : 실질형태소와 형식형태소 태그를 포함하여 생성한 word vector를 활용
Cls : 실질형태소에 형식형태소 클러스터 번호를 포함하여 생성한 word vector를 활용
All : 실질형태소와 형식형태소 및 태그를 포함하여 생성한 word vector를 활용

표 4. 모호성 평가 결과(%)

	규칙	Nul	Tag	Cls	All
정확도	50.0	83.3	83.3	85.0	83.3

모의테스트에서 규칙을 이용한 결과는 50%의 정확도를 보였는데, 이는 음역방향을 한 가지로 지정하여 음역하였고, 현재 가상데이터 구성의 영향을 받기 때문이다. word vector를 이용한 4종류의 성능이 비슷한 것은 평가데이터의 개수가 적기 때문으로 보이지만, 규칙을 이용하였을 때보다 모두 높은 성능을 보이고 있다.

6. 결론

최근 활발히 연구되고 있는 Text-to-Speech 기술을 위해서는 혼합문장에서부터 한글이 아닌 텍스트를 한글 텍스트로 음역할 필요성이 있다. 그러나 한글에서 숫자 및 단위를 발음할 때에는 문맥에 따라 모호성이 존재한다. 본 논문에서는 음역 방향을 시스템으로 제시하고, 규칙을 적용한 숫자 음역 방향과 word vector를 활용하여 주변 문맥정보를 고려한 음역 방향 결정을 제시하였으며, word vector 정보를 사용하였을 때 음역 방향 결정의 모호성을 좀 더 잘 해소할 수 있음을 보였다.

사사(Acknowledgement)

이 논문은 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임(No. 2017R1D1A3B03035676)

참고문헌

[1] Sarah G. Wood, MS1, Jerad H. Moxley, PhD1, Elizabeth L. Tighe, PhD1, Richard K. Wagner, PhD1, "Does use of text-to-speech and related read-aloud tools improve reading comprehension for students with reading disabilities A meta-analysis", Journal of learning disabilities, pp.73-84, 2018.

[2] 이종석, 박기태, 이준우, "시각 및 언어장애인을 위한 음성합성 기술의 현황", 전자공학회지 제41권 3호, pp.28-39, 2014.

[3] 최연주, 정영문, 김영관, 서영주, 김회린, "한국어 text-to-speech(TTS) 시스템을 위한 엔드투엔드 합성고무 방식 연구", 말소리와 음성과학 제10권 제1호, pp.39-48, 2018.

[4] Kang, M. Y., Min, T. H., Lee, J. S.. "Sense Space for Word Sense Disambiguation", 2018 IEEE International Conference on Big Data and Smart computing, pp.669-672, 2018.

[5] 정영임, 윤애선, 권혁철, "임베디드 TTS 시스템을

위한 아라비안 숫자의 문자 변환”, 한국정보과학회 학술발표논문집, pp.442-444, 2005.
[6] 김민정, 홍금원, 박소영, 임해창, “영・한 음차 표기 성능 향상을 위한 음철법 기반 규칙 구축”, 말소리와 음성과학 제1권 제4호, pp.133-144, 2009.
[7] 이재성, 최기선, “정보검색을 위한 외래어 자동표기 모델”, 한국정보관리학회 제4회 학술대회 논문집, pp.17-24, 1997.
[8] 김태일, “최대 엔트로피 모델을 이용한 다국어 정보 검색에서의 영-한 음차 표기 모델”, 서강대학교 석사학위 논문, 1999.
[9] 이경님, 정민화, “한국어 발음열 자동 생성을 위한 형태소 태그 정보 기반의 텍스트 전처리기”, 한국정보과학회 학술발표논문집, pp.199-201, 2001.
[10] 윤준영, 김도원, 민태홍, 이재성, “형식형태소가 한국어 단어 벡터 생성에 미치는 영향”, 한국정보과학회 학술발표논문집, pp.179-183, 2017.
[11] https://www.gnu.org/software/flex/flex.html.