

# 발음 변이 기반 키워드 증강을 통한 한국어 키워드 스포팅 외래어 검출 성능 향상

김선욱, 홍지민, 강주연, 김남수

서울대학교 전기정보공학부 뉴미디어통신공동연구소

{sukim, jmhong, jykang}@hi.snu.ac.kr, nkim@snu.ac.kr

## Enhancing Korean Keyword Spotting for Loanwords through Keyword Augmentation Based on Pronunciation Variation

Seonuk Kim, Ji Min Hong, Ju Yeon Kang, Nam Soo Kim

Department of Electrical and Computer Engineering and INMC, Seoul National Univ.

### 요약

본 연구에서는 한국어 키워드 스포팅에서 외래어 키워드 인식률을 향상시키기 위한 발음 변이 기반 키워드 증강 기법을 제안한다. AI Hub 한국어 외래어 발화 데이터셋을 활용하여 LCS(Longest Common Subsequence) 기반 정렬을 통해 발음 변이를 분석하고, 변이 빈도와 발음 유사도에 기반한 필터링을 수행하였다. 발음 변이 정보를 반영한 키워드 증강 기법은 10개의 외래어 키워드에 대해 hit ratio를 향상시켜, 모델의 재학습 없이도 외래어 검출 성능을 효과적으로 개선할 수 있음을 확인하였다.

### I. 서론

인공지능 기반 음성 신호 처리 기술의 발달은 음성 기반 인간-컴퓨터 인터페이스의 유용성을 크게 증대시키고 있으며, Siri(Siri)와 Bixby 등과 같은 다양한 음성 인공지능 에이전트가 사람들의 일상생활에 널리 채택되고 있다. 이러한 에이전트는 사용자의 음성 명령을 빠르고 정확하게 인식해야 하기 때문에, 음성 인식 기술의 성능이 사용자 경험에 중요한 역할을 한다. 특히, 음성 인식 기술의 한 분야인 키워드 스포팅은 음성 신호에서 특정 키워드의 발화를 실시간으로 검출하는 기술로, 인간과 음성 인공지능 에이전트 간의 상호작용을 구현하는 핵심 기술이다.

일반적으로 딥러닝 기반 키워드 스포팅 시스템은 대규모의 전사된 음성 데이터를 기반으로 지도학습을 통해 훈련된 음성 인식 모델로 구현된다. 사용자의 발화가 입력되면, 음성 인식 모델이 이를 텍스트로 전사하고, 사전에 등록된 키워드와의 일치 여부를 검사하여 검출 여부를 결정한다. 그러나 이 방식은 의미가 동일한 단어가 실제 환경에서 여러 방식으로 발음되는 경우, 미검출 오류(false negative)가 발생할 가능성이 크다. 이는 음성 인식 모델이 주로 표준 발음에 맞춰 학습되기 때문이다.

특히 한국어 외래어의 경우 원어 발음을 한국어 음운 체계에 맞춰 단순화하거나 변형하는 과정에서 발음이 다양하게 변이되어 모델이 학습한 발음과 큰 차이를 보인다. 예를 들어, 'gas'라는 단어는 한국어로 전사될 때 표준 발음인 '가스' 외에도 '까스', '갓스', '가쓰'와 같이 다양한 발음으로 나타날 수 있다. 그러나 키워드 등록 시 '가스'로만 등록된다면, 발화자가 다르게 발음할 경우 키워드 스포팅 시스템은 이를 검출하지 못하고 미검출 오류를 발생시킨다.

이러한 문제를 해결하는 일반적인 접근으로 파인튜닝(fine-tuning)이 사용된다. 파인튜닝은 사전 학습된 모델에 새로운 데이터를 추가 학습시켜 특정 도메인이나 환경에 맞게 성능을 최적화하는 기법이다. 본 문제의 경우, 외래어 발음의 변이를 학습 데이터에 반영하여 모델이 다양한 발음 변형에도 대응할 수 있도록 인식 범위를 확장하는 방식으로 적용할 수 있다. 그러나 이 방법은 모델 학습에 많은 비용이 소요되며, 모델이 외래어

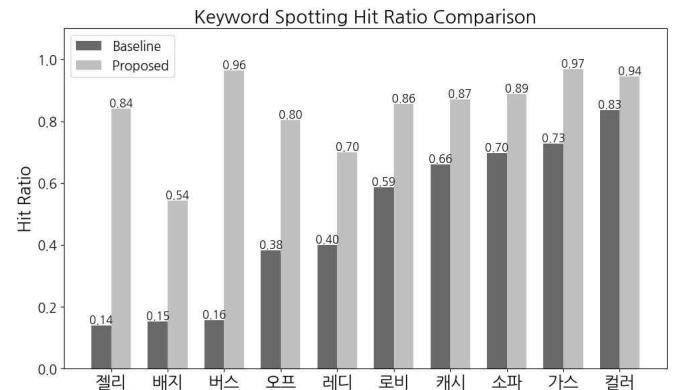


그림 1: 한국어 외래어 키워드 Hit Ratio

발음 변이에 최적화되는 과정에서 기존 한국어 키워드의 인식률이 저하될 수 있는 문제가 존재한다.

본 연구에서는 이러한 문제를 해결하기 위해 발음 변이 기반 키워드 증강 기법을 제안한다. AI Hub 한국어 외래어 발화 데이터셋[1]을 활용하여 외래어 발화와 전사문을 LCS(Longest Common Subsequence) 알고리즘[2]을 통해 정렬하고, 발음 변이 정보를 추출하였다. 추출된 변이 정보는 상위 빈도수를 기준으로 1차 필터링한 뒤, Python의 PanPhon[3] 라이브러리를 사용하여 계산된 발음 유사도를 바탕으로 2차 필터링을 수행하였다. 이후, 최종적으로 선별된 발음 변이 정보를 기반으로 외래어 키워드를 여러 개로 증강하여 RNN-T(Recurrent Neural Network Transducer)[4] 기반 Open Vocabulary 키워드 스포팅 시스템에 등록하였다.

이러한 기법을 검증하기 위해 10개의 외래어 키워드를 대상으로 키워드 스포팅 실험을 진행하였으며, hit ratio를 바탕으로 성능을 평가한 결과, 기존 모델 대비 유의미한 성능 개선이 확인되었다. 이로써, 제안된 기법이 모델 구조의 수정이나 추가 학습 없이도 한국어 키워드 스포팅 시스템의 외래어 인식 문제를 효과적으로 해결할 수 있음을 확인하였다.

한국어 외래어의 발음 변이에 관한 연구로는 류혁수, 나민수, 정민화 (2015)[5]의 연구가 있다. 이 연구에서는 한국인의 외래어 발화에서 나타나는 발음 변이 패턴에 대해 분석하고, 이를 바탕으로 자소-음소 변환 성능을 개선하는 방법을 제안하였으나, 특정 도메인에 한정된 발음 데이터를 사용하였으며, 실제 음성 인식 시스템에서의 성능 검증이 이루어지지 않았다는 한계가 있다. 반면, 본 연구는 AI Hub 한국어 외래어 발화 데이터셋[1]을 활용하여 다양한 발화 환경에서의 발음 변이를 분석하고, 이를 반영한 키워드 증강 기법을 통해 Open Vocabulary 키워드 스포팅 성능을 실질적으로 개선함으로 그 유효성을 입증하였다는 점에서 차별점이 있다.

## II. 본론

### 1. RNN-T 기반 Open Vocabulary 키워드 스포팅

본 연구에서는 한국어 발화의 외래어 키워드 검출을 위해 RNN-T[4] 구조를 기반으로 설계된 Open Vocabulary 키워드 스포팅 모델을 사용하였다. 해당 모델은 먼저 여러 층의 1차원 CNN(Convolutional Neural Network)을 통해 입력된 음성 신호를 저차원의 특징 벡터로 변환하고, 이를 RNN-T에 전달하여 발화 내용을 텍스트로 전사한다. 이후, 키워드 스포팅 시스템은 변환된 텍스트 결과에서 등록된 키워드를 검출한다. 모델 학습은 ksponspeech 데이터셋[6]을 활용하였으며, 평가는 zeroth-korean 데이터셋[7]을 활용하였다. 학습 과정에서 전사된 한국어 발화 텍스트는 국제 음성 기호(IPA, International Phonetic Alphabet) 시퀀스로 변환하여 발음 정보를 정교하게 반영하였고, 이후 BPE(Byte Pair Encoding)을 사용하여 500개의 서브워드 단위로 토큰화를 수행하였다.

### 2. LCS 기반 정렬 및 발음 변이 추출

본 연구에서는 한국어 외래어 발화와 대응 전사문 간의 발음 변이 분석을 위해 LCS 알고리즘[2] 기반 정렬 기법을 적용하였다. 이 기법은 두 시퀀스 간 최장 공통 부분을 찾고, 이를 기반으로 두 시퀀스를 정렬하여 토큰 단위로 변이를 추정한다. 실험에서는 실제 발화 시퀀스와 음성 인식 모델의 예측 시퀀스에 대해 정렬을 수행하였으며, 치환된 구간의 길이가 동일한 경우만 발음 변이 정보를 추출하고 빈도를 누적하였다.

### 3. 변이 빈도 및 발음 유사도 기반 필터링

추출한 한국어 외래어 키워드의 발음 변이 중 유의미한 정보를 선별하기 위해 본 연구에서는 각 변이 토큰 쌍의 빈도와 발음 유사도를 기반으로 필터링을 진행하였다. 먼저 빈도수 기반으로 변이 쌍을 정렬하고, 상위 일부의 자주 등장하는 변이만 유의미한 발음 변이로 간주하여 1차 필터링을 진행하였다. 이후 1차 필터링된 변이 쌍에 대해 IPA 시퀀스를 조음 기반 특징 벡터로 변환하는 기능을 제공하는 PanPhon[3] 라이브러리를 사용하여 두 발음 간의 차이를 특징 벡터 간의 거리로 정량화하였다. 이 거리가 특정 값 이내에 있는 경우만을 유의미한 발음 변이로 간주하여 최종적으로 2차 필터링을 수행하였다.

### 4. 평가

본 연구에서는 AI Hub 한국인 외래어 발화 데이터셋[1] 중, 스튜디오에서 수집된 학습용 데이터를 사용하여 외래어 발음의 변이 패턴을 분석하였고, 그 결과를 사용하여 키워드 증강을 수행한 후 키워드 스포팅 모델에

적용하였다. 모델의 성능 평가는 AI Hub 한국인 외래어 발화 데이터셋[1] 중, 음성수집도구에서 수집된 검증용 데이터를 활용하여 이루어졌다. 평가 결과는 그림 1에 요약되어 있으며, 다양한 hit ratio를 보이는 한국어 외래어 키워드 10개를 대상으로 키워드 스포팅을 테스트하였다. 실험 결과, 모든 외래어 키워드에 대해 hit ratio가 최소 0.11에서 최대 0.8까지 절대적으로 증가하는 개선 효과를 보였으며, 이를 통해 제안된 키워드 증강 기법이 외래어 검출 성능 향상에 효과적임을 확인할 수 있었다.

## III. 결론

본 연구에서는 한국어 외래어 키워드의 인식률을 향상시키기 위해 발음 변이 기반 키워드 증강 기법을 제안하였다. 이를 위해 AI Hub 한국어 외래어 발화 데이터셋[1]에서, LCS 알고리즘[2] 기반 정렬 기법을 통해 발음 변이 정보를 분석한 뒤, 변이 빈도와 PanPhon[3]을 활용한 발음 유사도 기반 필터링을 거쳐 유의미한 변이 쌍을 선별하였다. 이렇게 추출된 변이 정보를 바탕으로 키워드 증강을 수행하였으며, 제안된 기법을 적용한 모델은 10개의 외래어 키워드에 대해 hit ratio가 최소 0.11에서 최대 0.8까지 절대적으로 상승하는 성능 향상을 보였다. 이는 모델의 재학습 없이도 외래어 키워드 검출 성능을 효과적으로 높일 수 있음을 나타낸다.

다만, 본 연구에서는 발음 변이를 허용하는 과정에서 생길 수 있는 false detection에 대한 정량적 분석이 이루어지지 않았다. 향후 연구에서는 키워드 증강 기법의 검출 신뢰도에 대한 평가가 요구되며, 이를 바탕으로 실제 발화 환경에서도 신뢰도 높은 키워드 검출이 가능할 것으로 기대된다.

## ACKNOWLEDGMENT

이 논문은 2025년도 BK21 FOUR 정보기술 미래인재 교육연구단에 의해 지원되었음.

## 참 고 문 헌

- [1] 한국지능정보사회진흥원, "한국인 외래어 발화," AI Hub, 2020. [Online]. Available: <https://aihub.or.kr/aidata/30706>.
- [2] D. S. Hirschberg, "Algorithms for the Longest Common Subsequence Problem," J. ACM, vol. 24, no. 4, pp. 664 - 675, Oct. 1977.
- [3] D. R. Mortensen, P. Littell, A. Bharadwaj et al., "PanPhon: A Resource for Mapping IPA Segments to Articulatory Feature Vectors," Proc. 26th Int. Conf. Comput. Linguistics (COLING), Osaka, Japan, Dec. 2016, pp. 3475 - 3484, doi: 10.18653/v1/C16-1328.
- [4] A. Graves, "Sequence Transduction with Recurrent Neural Networks," Proc. ICML Representation Learning Workshop, Edinburgh, U.K., 2012. Available: <https://arxiv.org/abs/1211.3711>.
- [5] 류혁수, 나민수, 정민화, "한국인 화자의 외래어 발음 변이 양상과 음절 기반 외래어 자소-음소 변환," 말소리와 음성과학, vol. 7, no. 3, pp. 139 - 149, 2015.
- [6] J.-U. Bang et al., "KsponSpeech: Korean Spontaneous Speech Corpus for Automatic Speech Recognition," Appl. Sci., vol. 10, no. 19, Art. no. 6936, Oct. 2020, doi: 10.3390/app10196936.
- [7] OpenSLR, Speech Corpus, no. 40, "Zeroth-Korean Speech Corpus," 2017. Available: <https://openslr.org/40>.