

모바일 기기를 위한 사용자 적응형 음성인식 후처리*

김영진[○] 김은주 김명원

송실대학교 컴퓨터학과

{liebulia, blue7786}@ssu.ac.kr, mkim@comp.ssu.ac.kr

User Adaptive Post-Processing of Speech Recognition for Mobile Device

Young Jin Kim[○] Eun Ju Kim Myung Won Kim

Dept. of Computer Science, Soongsil University

1. 서 론¹⁾

최근 PDA(personal digital assistants)나 스마트 폰(smart phone)과 같은 모바일 기기는 통신, 영화, GPS(global positioning system)등 여러 가지 기능들을 사용할 수 있도록 다중화 되고 있으며 휴대하기 편하도록 소형화 되고 있다. 따라서 모바일 기기에는 사용자의 편의성을 제공하기 위한 사용자 친화적 인터페이스가 필요하게 되었고 이러한 기술로 음성인식이 연구 중이나 현재 음성인식 기술은 잡음이 없는 환경에서 87%~97%의 인식률만을 보인다. 그러나 모바일 기기는 사용자가 휴대하고 다니면서 어디서나 사용 가능함으로 항상 다양한 잡음에 노출된다. 따라서 잡음에 민감한 현재의 음성인식 기술이 모바일 기기에 사용되었을 경우 인식률은 현저하게 떨어지게 된다.

이러한 단점을 보완하기 위해 음성인식 결과를 효과적으로 보정할 수 있는 음성인식 후처리 방법이 필요하다. 이러한 음성인식 후처리 방법에는 오류 패턴(error pattern) 정보를 이용한 후처리[1]와 의미(semantic) 정보를 이용한 후처리[2], 문맥(context) 정보를 이용한 후처리[3], 발화 순차 패턴(sequence pattern)[4] 정보를 이용한 후처리 등이 있다. 그러나 이러한 기존 음성인식 후처리 방법들은 단지 인식기의 최종 인식 단어만을 이용하여 인식기와 독립적으로 후처리를 적용할 수는 있으나 주변 잡음에 의해 연속적으로 인식기의 오인식이 발생하는 경우 후처리 결과를 신임할 수 없게 된다.

본 논문에서는 주변 잡음 환경에서 인식기의 인식률을 높이기 위하여 사용자의 특성 정보를 학습하고 후처리에 이용하는 사용자 적응형 후처리 방법을 제안한다. 이 방법은 사용자가 모바일 기기에 음성으로 명령을 내릴 때 자주 발생하게 되는 주변 잡음이나 사용자의 특별한 발화 특성 패턴을 학습하고, 학습된 정보를 이용하여 올바른 인식을 하도록 수정함으로써 인식률을 향상시키는 방법이다.

2. 사용자 적응형 후처리

본 논문에서 음성인식 시스템은 음성 인식기와 사용자 적응형 후처리로 구성된다. 인식기는 사용자가 발화한 음성 신호에서 음성특징을 추출하고 HMM(hidden Markov model)[5, 6]을 이용하여 일차적인 인식과정을 수행한다. 사용자 적응형 후처리는 인식기의 인식 결과를 받아 후처리 적용여부를 확인하고 후처리가 필요한 경우 적용하여 최종 결과를 도출한다.

사용자 적응형 후처리에서는 학습을 위해 음성 인식기의 인식 결과로부터 두 가지의 사용자 특성 정보를 추출한다. 그중 하나는 인식기가 인식한 결과 단어로써 사용자에게 요청한 발화 단어와 이를 인식한 음성 인식기의 결과 단어를 쌍으로 구성한 것이다. 모델 생성은 이렇게 구성된 데이터를 기반으로 하여 어떤 단어가 어떤 단어로 일정수준 이상 오인식 되는지 분석하여 생성된다. 즉, 모델 선정은 어떠한 단어로 오인식 되었는지를 가지고 정하며 출력 단어는 어떠한 단어가 오인식 되는지를 가지고 정한다.

다른 하나의 특성 정보는 인식기 인식 결과에 영향을 미치는 모든 단어들의 likelihood로써 이를 이용하여 전에 생성된 모델을 학습한다. 이때 모델은 다층 퍼셉트론(multilayer perceptron)[7, 8]을 사용하고 학습은 역전파 알고리즘(back-propagation algorithm)을 사용한다. 이러한 신경망 모델들은 적응이 필요한 단어에

* 본 연구는 서울시 산학연 협력사업(10581cooperateOrg39111)의 연구비 지원으로 수행하였습니다.

대해서만 생성되고 수정이 필요한 모델만을 수정하면서 사용자의 모바일 기기 사용 환경과 발화 특성을 최대한 반영한다.

3. 실험 및 결과

본 실험에서 음성 데이터는 16,000kHz, 16bit, 모노(mono)로 설정하였으며 주변 잡음이 정제되지 않은 일반 잡음 환경에서 마이크로폰(microphone)을 이용하여 녹음하였다. 녹음된 웨이브(wave) 데이터에서 음성 특징 추출은 MFCC(Mel frequency cepstral coefficient)를 사용하였고 인식기는 Cambridge 대학에서 HMM을 기반으로 개발한 HTK(HMM tool boot)[9]를 이용하였다.

실험에 사용된 단어는 모바일 환경에서 많이 사용되는 32개의 명령어를 선별하여 사용하였고 데이터는 3명의 사용자가 각각 단어 당 500회씩 발화하여 구축하였다. 이중 각 단어 당 150개의 데이터는 인식기를 학습하는데, 단어 당 50개는 인식과정 전체를 평가하는데 사용하였다. 그리고 나머지 데이터는 인식률이 떨어지는 단어들에 대해서만 50개씩 추가하면서 후처리 학습에 사용하였다. 단 실험에 사용된 모든 단어를 실험에 적용하기 위하여 단어 당 50개씩의 데이터를 가지고 초기 모델을 생성하였다.

모델을 생성하는 기준은 음성인식기의 인식률을 고려하여 각 단어마다 오인식이 3% 이상 발생하는 단어에 대해서만 생성하였으며 모델 별 출력단어 선정 역시 같은 기준을 적용하였다. 학습에 사용된 다층 퍼셉트론의 경우 출력 노드의 수는 모델 마다 가변적으로 적용되어 2개에서 6개 까지 적용되었고 입력 노드의 수는 32개로 하였다. 그 결과 90.2%였던 음성 인식기 평균 인식률이 적응형 후처리 적용 후 평균 94.2%로 약 4% 향상되었고 평균 41.0%의 오류 수정률을 보였다.

따라서 사용자의 발화 특성은 인식기의 인식 결과에 반영이 되고 이러한 인식 결과인 likelihood 패턴을 학습함으로써 인식기의 오류를 수정할 수 있음을 보였다. 또한 사용자 적응형 후처리는, 사용자의 계속적인 사용으로 오인식이 자주 발생하는 단어를 수정해 감으로서 점점 더 높은 인식률을 보일 수 있다는 장점을 가지고 있다.

4. 참고문헌

- [1] Satoshi Kaki, Eiichiro Sumita, and Hitoshi Iida, "A method for correcting speech recognition using the statistical features of character co-occurrence," International Conference On Computational Linguistics, vol. 1, pp.653-657, 1998.
- [2] Minwoo Jeong, Byeongchang Kim, Gary Geunbae Lee, "Semantic- oriented error correction for spoken query processing," Automatic Speech Recognition and Understanding, IEEE, pp.156-161, 2003.
- [3] Myung Won Kim, Joung Woo Ryu, Eun Ju Kim, "Speech Recognition with Multi-Modal Features Based on Neural Networks," International Conference on Neural Information Processing (ICONIP), LNCS 4233, pp.797-806, 2006.
- [4] 송원문, 김명원, "문맥 및 사용 패턴 정보를 이용한 음성인식 후처리", 정보처리학회논문지 제13-B권 제5호, pp.553-560, 2006.
- [5] Deller, Hansen, Proakis, "Discrete-time processing of speech signals," IEEE PRESS, pp.677-744, 2000.
- [6] M. Ostendorf, "From HMM's to segment models: a unified view of stochastic modeling for speech recognition," IEEE SPA. pp.360-378, 1996.
- [7] Jiawei Han, Micheline Kamber, Data Mining concepts and techniques, Morgan Kaufmann Publishers, pp.303-311, 2001.
- [8] Tom M. Mitchell, Machine learning, McGraw-hill international editions, pp.81-127, 1997.
- [9] Steve Young, Gunnar Evermann, Mark Gales, Thomas Hain, Dan Kershaw, Gareth Moore, Julian Odell, Dave Ollason, Dan Povey, Valtcho Valtchev, Phil Woodland, The HTK book (for HTK version3.3), Cambridge University Engineering Department, 2005.