문맥 및 사용 패턴 정보를 이용한 음성인식의 성능 개선

송 원 문[†] · 김 명 원[††]

요 약

최근 음성인식에서는 잡음환경에서 좀 더 신뢰성 있는 결과를 얻기 위해 인식 결과 도출 단계에서 여러 가지 정보의 내용들을 융합하거나 이전 인식 결과의 후처리를 통하여 성능을 향상시키는 방법들이 연구되고 있다. 본 논문에서는 잡음 환경에서의 인식률 하락을 보완하기 위해 개인 모바일 기기를 위한 음성 명령어 인식에서 사용자의 사용패턴과 문맥 정보를 사용하는 방법을 제안한다. 기본 인식 결과를 보정하기 위해서 현재 명령어를 발화하기 이전에 사용자가 사용한 순차적 명령어 패턴을 사용하였다. 또한 문맥 정보를 위해서는 사용중인 기기의 현재 기능과 발화된 명령어간의 연관성을 사용하였다. 실험을 통해 제안한 방법이 기본 인식 시스템에서 발생한 오인식의 약 50%를 수정하였음을 보였으며 이로써 제안한 방법의 타당성을 검증하였다.

키워드 : 음성인식, 후처리, 문맥 정보, 사용자 정보, 순차적 사용 패턴

Performance Improvement of Speech Recognition Using Context and Usage Pattern Information

Won Moon Song[†] · Myung Won Kim[††]

ABSTRACT

Speech recognition has recently been investigated to produce more reliable recognition results in a noisy environment, by integrating diverse sources of information into the result derivation-level or producing new results through post-processing the prior recognition results. In this paper we propose a method which uses the user's usage patterns and the context information in speech command recognition for personal mobile devices to improve the recognition accuracy in a noisy environment. Sequential usage (or speech) patterns prior to the current command spoken are used to adjust the base recognition results. For the context information, we use the relevance between the current function of the device in use and the spoken command. Our experiment results show that the proposed method achieves about 50% of error correction rate over the base recognition system. It demonstrates the feasibility of the proposed method.

Key Words : Speech Recognition, Post-processing, Context Information, User Information, Sequential Usage Pattern

1. 서 론

최근 컴퓨팅 환경이 더욱 다양하고 복잡해지면서 이를 사용하는 사용자를 위하여 좀 더 쉽고 인간 친화적인 접근을 할 수 있도록 하는 HCI(human computer interaction)와 같은 연구들이 많이 진행되고 있다. 음성인식은 이러한 목적을 가진 연구로서 사용자가 기존의 문자 입력이나 버튼을 누르는 등의 행동으로 컴퓨터와 상호작용을 하던 개념에서 벗어나 인간사이의 대화처럼 음성으로 컴퓨터와 상호작용을 할 수 있도록 하는 것이다. 이러한 목적의 음성인식은 모바일 기기, 자동차 네비게이션, 음성기반 검색 시스템, 자동응답 시스템 등의 여러 분야에 활발히 적용되고 있다.

음성인식 알고리즘은 다른 알고리즘들에 비해 특히 인식률이 좋은 HMM(hidden Markov model) 알고리즘이 주로 사용되고 있다. 하지만 음성신호에 포함되는 다양한 잡음과 현재 상태는 바로 이전 상태에 의해서만 영향을 받는다는 HMM 알고리즘의 기본 가정 때문에 주변 환경에 의해 인식률이 민감하다는 단점이 있다[1, 2]. 이러한 문제점을 해결하여 다양한 잡음 환경에서 음성의 인식률을 높이기 위한 방법으로 음성이 아닌 다른 여러 가지 정보들을 사용하여 음성의 인식 결과를 조정하는 음성인식 후처리 방법들이 연구되고 있다[3, 4, 5].

기존 후처리 방법으로는 단어의 오인식 패턴이나 단어를 포함하는 블록의 오인식 패턴을 이용한 후처리 방법[3], 단

※ 본 연구는 숭실대학교 교내연구비 지원으로 이루어졌음.
† 준 회 원 : 숭실대학교 대학원 컴퓨터학과 박사과정
†† 정 회 원 : 숭실대학교 컴퓨터학부 교수
논문접수 : 2006년 8월 17일, 심사완료 : 2006년 9월 14일

어의 어휘 및 의미적 범주를 이용한 후처리 방법[4] 등이 있다. 이러한 방법들은 단어의 발음 및 어휘적 특성이나 의미적 범주를 사용한 일반적인 접근 방법으로써 음성인식 후처리의 응용분야에 쉽게 적용될 수 있다. 하지만 모바일 환경에서와 같이 사용자가 상황에 따라 특정한 패턴으로 발화하는 경우나 개인 정보에 의해서 발화 내용이 틀려지는 경우에 사용자의 발화 패턴이나 개인 정보를 고려하지 못하므로 인식률이 저하된다. [5]에서는 이러한 점을 극복하고자 사용자의 발화 패턴을 활용한 후처리를 제안하였으나 발화 패턴을 추출하고 후처리에 적용하는데 신경망을 사용함으로써 발화 패턴에 대한 신뢰도를 구체적 수치로 표현하기 힘들다는 단점이 있다.

본 논문에서는 사용자 발화 순차 패턴 정보와 기기의 수행중인 기능 정보를 사용자의 상황 정보로 정의하고 이러한 정보를 이용하여 음성의 인식률을 향상시키는 상황 정보 후처리 방법을 제안한다. 또한 상황 정보 후처리의 효율성을 높이기 위하여 인식기의 인식 결과에 대한 신뢰성을 판단하여 후처리를 적용하는 방법을 제안한다. 상황 정보 후처리에서는 인식 결과에 대한 신뢰성이 떨어진다고 판단되면 제안한 두 가지 상황 정보와 인식기의 인식 결과를 융합하여 주변 환경의 잡음에 강인한 새로운 음성인식 결과를 도출한다.

본 논문의 2장에서는 지금까지 국내외 연구들에서 제안된 주요 후처리 방법에 대하여 간단히 소개하고 3장에서는 본 논문에서 제안하는 상황 정보 후처리 방법에 대해서 기술한다. 4장에서는 제안한 방법에 대한 실험 결과를 기술하고 분석하여 타당성을 검증하며 5장에서는 결론을 맺는다.

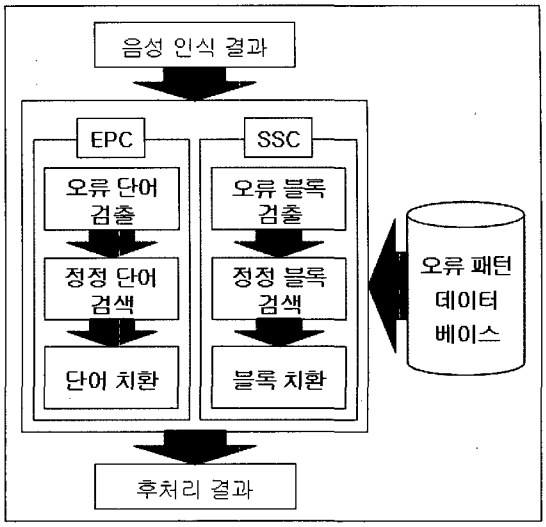
2. 음성인식 후처리 방법

음성의 인식률 향상을 위한 음성인식 후처리 방법에 대해서는 이미 국내외에서 많은 연구들이 진행되고 있다. 2장에서는 최근의 몇 가지 후처리 연구들에 대해 간단히 기술한다.

2.1 오류 패턴 비교

[3]에서는 미리 구축된 오류 패턴 데이터를 이용한 EPC (error-pattern correction)와 SSC(similar-string correction)의 두 가지 방법을 사용하여 음성인식의 오류를 수정하는 후처리 방법을 제안하였다. 이 방법은 음성의 오인식은 일정한 유형을 가지고 발생하며 무작위로 발생하는 오인식 유형은 거의 없다고 가정한다. 따라서 훈련 데이터를 통해 미리 오인식 패턴을 모은 후 이를 오류 패턴 데이터베이스로 구축하여 후처리에 이용한다. 오류 패턴 데이터베이스는 오인식이 잘되는 부분과 그 부분의 오인식 형태를 쌍으로 묶어 구축한다. 이때 EPC를 위해서는 오류 패턴의 단위를 단어로 하고 SSC를 위해서는 오류 패턴의 단위를 목표 단어와 그 단어의 앞뒤 단어까지 함께 묶은 블록으로 한다.

음성이 발화 되면 후처리 단계에서는 인식기를 통해 인식된 결과와 미리 구축된 오류 패턴 데이터베이스를 비교하여 오인식 예상부분을 검출해 낸다. 이후 예상된 오인식 단어를 오류 패턴 데이터베이스상의 원래의 단어나 블록으로 치환해 주면 오류가 정정된다. 이러한 후처리 시스템의 구조는 (그림 1)과 같다.

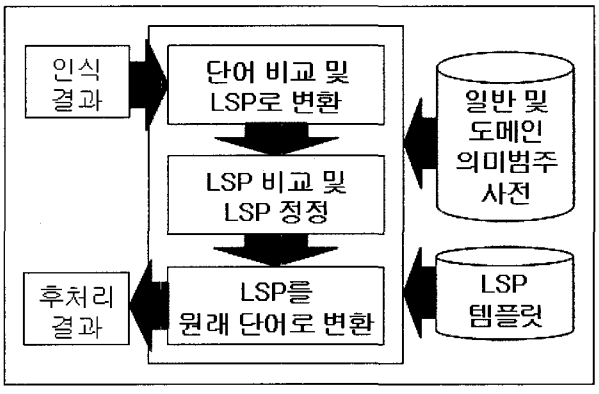


(그림 1) 오류 패턴 비교를 이용한 후처리 구조도

[3]에서는 EPC와 SSC의 두 가지 방법을 조합한 후처리 방법을 사용함으로써 음성인식의 오류율이 8.5% 감소되었다. 이와 같은 방법은 단어의 오인식 유형이 특정한 패턴으로 발생되거나 전후에 같이 나온 단어에 의해 오인식이 일어나는 경우에 높은 인식률을 기대할 수 있다. 그러나 오류 패턴 데이터베이스 자체가 오인식 단어와 정정될 단어의 쌍, 그리고 이들이 훈련 데이터에서 나타난 횟수로만 이루어져 있어 오류 패턴의 종류가 다양한 단어나 블록의 경우에는 적용할 오류 패턴에 대한 신뢰도를 예측할 수 없다.

2.2 어휘 의미 패턴

[4]에서는 단어의 의미적 정보를 고려한 후처리를 위해 LSP(lexico-semantic pattern, 어휘의미 패턴)를 제안하였다. LSP란 연속 음성인식에서 발화된 문장을 단어별로 각각 어휘 및 의미정보를 포함한 특정 범주 정보로 대치하여 연결한 스트링이며 후처리에 사용될 LSP는 훈련 데이터를 통하여 템플릿 데이터로 미리 구성되어진다.



(그림 2) LSP를 이용한 후처리 구조도

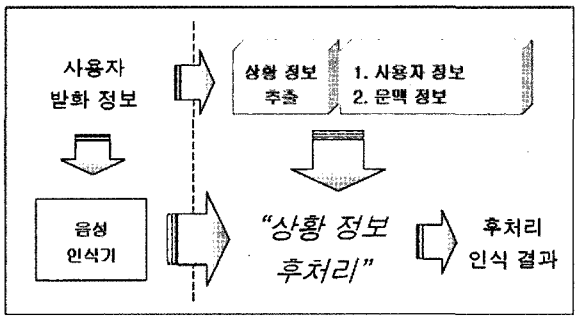
후처리 적용시에는 먼저 발화한 문장의 인식결과에 대하여 단어들을 각각의 범주 정보로 대치하여 LSP로 바꾼 후 미리 구성된 LSP 템플릿 데이터와 비교한다. 이때 LSP 템플릿 데이터 중에서 인식결과 LSP와 가장 유사한 LSP를 선택하며 인식결과 LSP를 선택된 LSP로 바꾸어 먼저 의미적 오류를 수정한다. 마지막으로 수정된 LSP내의 각 범주 정보들을 실제 단어로 바꾸는 어휘적 오류 수정을 통하여 최종 인식결과를 도출한다. 이러한 후처리 시스템의 구조는 (그림 2)와 같다.

(그림 2)에서 일반 의미범주 사전은 특정 도메인에 상관없이 일상적으로 많이 쓰이는 단어들에 대하여 단어의 의미적 범주를 지정해 놓은 데이터이며 도메인 의미범주 사전은 음성인식을 적용할 도메인에 속하는 특정 단어들에 대하여 단어의 의미적 범주를 지정해 놓은 데이터이다.

LSP를 이용한 후처리 방법은 기존의 어휘적, 신호적 단계의 후처리에서 한 단계 나아가 의미적인 정보를 이용하여 후처리를 구현했다는데 의의가 있다. 그러나 LSP는 단어 자체에 대한 의미정보를 이용하기 때문에 특정 사용자에 대한 사용 패턴 정보를 반영하지 못한다. 또한 잡음에 의하여 문장에서 여러 단어가 오인식될 경우 LSP 역시 잘못된 구성을 가지게 되어 정확한 어휘의미 단계의 정정이 불가능하다.

3. 문맥 및 사용자 정보를 이용한 음성인식 후처리

이 장에서는 잡음환경에 강인한 후처리를 위해 사용자의 상황 정보를 이용하여 음성의 인식률을 향상 시키는 상황 정보 후처리 방법을 제안한다. 상황 정보로는 사용자 정보로 정의되는 사용자 발화 패턴 정보와 기기의 문맥 정보로 정의되는 기기의 기능 정보의 두 가지 정보를 이용한다. 상황정보 후처리는 사용자의 발화 정보로부터 사용자 정보와 문맥 정보를 추출하고 추출된 정보를 발화 음성의 인식 결과와 융합하여 새로운 인식 결과를 도출한다. 이와 같은 상황 정보 후처리의 구조는 (그림 3)과 같다.



(그림 3) 상황 정보 후처리의 구조도

3.1 사용자 발화 순차 패턴 정보의 구성과 활용

본 논문에서는 잡음환경에서의 인식률 향상을 목적으로 음성 신호가 아닌 추가 정보를 음성인식 후처리에 이용하기 위하여 사용자의 행동 정보인 사용자 발화 순차 패턴을 사용하였다. 이것은 잡음환경에 민감한 신호위주의 처리에서 탈피한 후처리 방법이다. 또한 모바일과 같은 개인 기기의 사용 환경에서 사용자 발화 순차 패턴을 적용한 후처리는 개인의 행동 정보를 적용한 개인화 후처리이다.

후처리에 적용된 사용자 발화 순차 패턴의 추출에는 패턴의 빈도와 신뢰성을 적절히 반영하기 위하여 PrefixSpan 순차 패턴 추출 알고리즘[6]을 사용하였다. PrefixSpan 순차 패턴 추출 알고리즘은 모바일 환경과 같은 크지 않은 데이터 내에서 순차 패턴을 찾는데 효율적인 알고리즘이다.

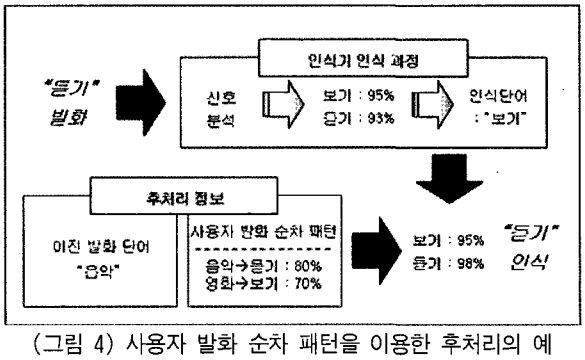
사용자 발화 순차 패턴은 사용자가 기기를 사용한 기록으로부터 추출하여 <표 1>과 같은 예로 구성되며 첫 번째 예는 사용자가 "음악"이라는 단어를 발화한 후에는 "듣기"라는 단어를 발화할 확률이 80%라는 것을 의미한다. 여기서 순차 패턴 규칙 "$A \rightarrow B$"(A이면 B이다)에 대한 신뢰도는 다음 (식 1)과 같이 구하며 이는 선행 조건 A가 발생한 경우에 B가 발생할 확률을 의미한다[7, 8]. 사용자 발화 순차 패턴을 통한 실제 후처리의 예는 (그림 4)와 같다. (그림 4)에서 사용자는 "듣기"라는 단어를 발화하였으나 인식기는 "보기"라고 인식하였다. 그러나 인식기의 계산 결과, 발화 단어가 "보기"일 확률과 "듣기"일 확률이 비슷하게 계산되어 후처리를 적용하였다. 후처리 과정에서는 이전에 발화된 "음악" 단어 및 <표 1>과 같은 사용자의 발화 순차 패턴 데이터 정보를 활용하여 현재 상황에 "보기"보다 "듣기"를 발화했을 가능성이 더 높다고 판단하였으며 따라서 최종 인식결과를 "듣기"로 정정하였다.

〈표 1〉 사용자 발화 순차 패턴의 예

사용자 발화 순차 패턴		순차 패턴의 신뢰도
선행(조건) 단어	발화(결과) 단어	
음악	듣기	80%
영화	보기	70%

$$A \rightarrow B \text{ 의 신뢰도}(\%) = \frac{A\text{와 } B\text{를 동시에 포함하는 경우의 수}}{A\text{를 포함하는 경우의 수}} \times 100$$

(식 1)



(그림 4) 사용자 발화 순차 패턴을 이용한 후처리의 예

3.2 문맥 정보 구성과 활용

잡음에 의한 인식률 하락을 극복하고 현재 상황에 맞는 음성인식을 위하여 본 논문에서는 현재 기기에서 수행중인 기능과 발화 명령어간의 연관성으로 정의한 문맥 정보를 후처리에 적용하는 방법을 제안한다. 모바일 기기와 같은 환경에서 명령어는 명령 수행의 목표나 해당 명령이 적용될 수 있는 특정 기능을 가진다. 따라서 이전에 사용된 명령어들을 고려하여 현재 사용중인 기기의 기능을 결정할 수 있으며 결정된 기능은 사용자가 발화한 현재 명령어의 기능과 동일하거나 비슷할 가능성이 크다.

본 논문에서는 인식 가능한 명령어들이 속한 기능을 정의한 후 각 기능들에 대한 상호 연관성 정도를 의미하는 가중치로 문맥 연관성을 정의하고 후처리에 사용하였다. 문맥 연관성 값은 명령어에 대한 문맥적 연관성이 있는 경우 해당 명령어들을 같이 사용할 가능성이 있으므로 '0.5'로 50% 정도 양의 상관관계를 가지도록, 명령어에 대한 문맥적 연관성이 없는 경우 해당 명령어들을 같이 사용할 가능성이 거의 없으므로 '-0.5'로 50%정도 음의 상관관계를 가지도록 하였다. 또한 문맥 연관성 값을 정의하기 모호한 경우에는 값을 '0'으로 하여 상관관계를 표현하지 않았다. 예를 들어 '음악', '영화' 및 '탐색'에 대한 기능이 가능하고 "음악", "영화", "듣기", "보기" 및 "찾기"의 5가지 명령어를 인식할 수 있다면 이에 대한 기능별 명령어 목록은 <표 2>와 같이, 특정 기능 '음악'과 다른 명령어간의 문맥 연관성은 <표 3>과 같이 구성된다.

문맥 연관성을 이용하면 현재 기기의 상황에 맞는 후처리가 가능하다. 예를 들어 이전에 "음악" 명령어를 사용하였을 경우 <표 2>를 참고하면 문맥 정보인 현재 기기의 기능은 '음악'으로 결정된다. 그리고 이때 (그림 5)의 예에서와 같이 발화된 음성의 인식결과는 "보기"이나 인식기의 계산 결과 "보기"와 "듣기"일 확률에 큰 차이가 없다면 후처리에서는 <표 3>을 참고하여 현재 기기의 기능인 '음악'과 더 높은 문맥 연관성을 가지는 "듣기"를 인식 결과로 정정한다.

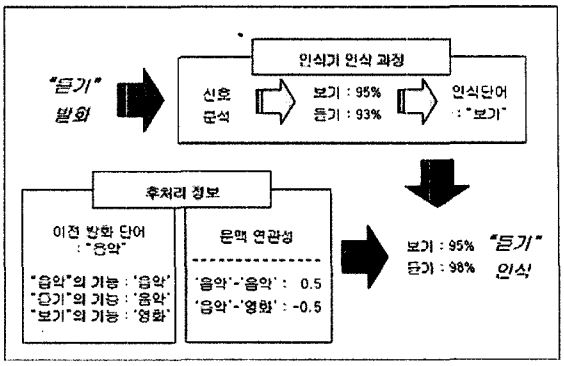
기기의 현재 기능정보는 가장 최근 사용한 n개의 단어를 고려하여 결정하되 단어의 사용 순서에 따른 가중치를 표현

〈표 2〉 기능별 단어 목록

기능 분류	명령어 목록
음악	음악, 듣기
영화	영화, 보기
탐색	찾기

〈표 3〉 '음악'기능과 다른 명령어간의 문맥 연관성 값

명령어(기능) / 기능	듣기 (음악)	영화 (영화)	보기 (영화)	찾기 (탐색)
음악	0.5	-0.5	-0.5	0



(그림 5) 문맥 정보를 이용한 후처리의 예

하기 위해 [9]에서 제안된 (식 2)를 사용한다. (식 2)에서 $W(t)$는 고려된 n개의 단어 중에서 현재로부터 t번째 이전에 사용된 단어의 가중치를 의미한다. 이와 같은 지수함수를 이용한 가중치 계산 방법은 시간에 따라 발생한 아이템에 발생 순서에 따른 가중치를 줄 때 효과적이며 가장 많이 사용하고 있는 방법이다[9, 10]. 현재의 기능을 설정할 때는 단어별로 계산된 가중치를 기능별로 합산하여 가장 높은 값을 가지는 기능을 선택한다.

$$W(t) = e^{-\frac{1}{n} \cdot t} \qquad \text{(식 2)}$$

3.3 사용자 정보와 문맥 정보를 이용한 후처리

본 논문에서는 후처리를 위한 음성인식 시스템에 HTK(HMM tool-kit) 음성인식 시스템[11, 12]을 사용하였다. HTK는 HMM알고리즘을 사용하여 인식 가능한 각각의 단어에 대해 발화된 음성과 같을 확률을 계산한다. 이때 HTK는 연속적 확률 계산에 의한 결과의 언더플로우(under-flow)를 방지하기 위해 로그함수를 사용하며 최종적으로 계산된 결과를 가능성(likelihood) 값으로 표현한다. 따라서 가능성 값이 높을수록 발화된 음성과 같을 확률이 높으며 HTK의 인식결과는 최대의 가능성 값을 가지는 단어로 선정된다.

상황 정보 후처리에서는 발화한 음성에 대하여 HTK를 통해 계산된 가능성 값이 높은순으로 정렬된 *n-best* 인식결과를 구성한다. 이때 발화 단어와 같을 가능성이 적은 단어를 후보에서 미리 배제하기 위해 인식 가능한 모든 단어에 대해 계산된 최대 가능성 값과 최소 가능성 값의 중간값 이상의 가능성 값을 가지는 단어만을 *n-best* 인식결과로 구성한다. 음성이 인식되면 후처리를 위해 이전에 사용된 명령어를 확인하여 *n-best* 인식결과 중에서 사용자 발화 순차 패턴상 결과로 예상되는 단어들을 찾는다. 이후 HTK를 통해 인식된 단어와 사용자 발화 순차 패턴상 결과 단어들을 후처리 보정 대상 단어로 선정하고, 각 단어에 대해 HTK로부터 계산된 가능성 값들과 이전에 사용된 단어들을 고려한 순차패턴 신뢰도 및 문맥 연관성 값 등 세 가지를 동시에 고려하여 각 단어들에 대한 후처리 보정값을 산출한다. 후

처리 보정 대상 단어들 중에서 특정 단어를 W_i라 하면 단어 W_i에 대한 후처리 보정값 S_{W_i}는 다음의 (식 3)과 같이 산출한다.

$$S_{W_i}=(MAX(L_{n-best})-MIN(L_{n-best}))\times(R_{W_i}+C_{W_i}) \quad \text{(식 3)}$$

단, 단어 W_i가 인식기의 인식 단어이면서
순차패턴상 단어로 존재하지 않으면, $C_{W_i}=1-\frac{|N-BEST|}{|WORDS|}$

위 식에서 L_{n-best}는 구성된 *n-best* 인식결과 단어들의 가능성 값들을 의미하며 $MAX(L_{n-best})$와 $MIN(L_{n-best})$는 각각 L_{n-best} 중에서 최대값과 최소값을 의미한다. 또한 R_{W_i}는 기기에서 현재 사용중인 기능과 단어 W_i 와의 문맥 연관성 값을, C_{W_i}는 사용자가 이전에 사용한 단어와 단어 W_i로 이루어진 순차 규칙의 신뢰도를 의미한다. 따라서 $MAX(L_{n-best})-MIN(L_{n-best})$는 후처리 대상 단어의 가능성 값을 얼마만큼 보정할지에 대한 기준값을 의미하며 $R_{W_i}+C_{W_i}$는 단어 W_i에 적용할 기준값의 상황 정보 가중치를 의미한다. 그런데 이때 W_i가 인식기 결과 단어이면서 이전에 사용된 단어와 함께 이루어진 순차 패턴이 존재 하지 않을 경우 가중치에 적용할 순차 규칙의 신뢰도 C_{W_i}가 존재하지 않게 된다. 이 경우에는 순차 규칙의 신뢰도 대신 인식 가능한 전체 단어 개수에 대해 후보로 구성된 *n-best* 인식결과의 개수를 고려하기 위해 $1-\frac{|N-BEST|}{|WORDS|}$를 적용하였다.(단, *N-BEST*는 *n-best* 인식 결과 집합을 의미하며 *WORDS*는 인식 가능한 전체 단어 집합을 의미한다.) (식 3)을 통해 산출된 보정값을 이용하여 각 후처리 보정 대상 단어 W_i 에 대해 새롭게 계산되는 가능성 값 L'_{W_i}는 다음의 (식 4)와 같다.

$$L'_{W_i}=L_{W_i}+S_{W_i} \quad \text{(식 4)}$$

단, L_{W_i} 는 단어 W_i에 대하여 인식기에서 계산된 원래의 가능성 값

최종 후처리 결과 단어는 (식 4)를 통해 보정된 모든 후처리 대상 단어 중에서 가장 높은 가능성 값을 가지는 단어로 선정된다.

3.4 효율적인 후처리 적용을 위한 방법

본 논문에서는 상황 정보 후처리를 좀 더 효율적으로 적용하기 위하여 음성인식기의 인식 결과에 대한 신뢰성을 판단한 후 순차적 결합 기법을 사용하여 상황 정보 후처리를 적용하는 방법을 제안한다. 순차적 결합 기법이란 여러 가지 결과에 순차적 우선순위를 주어 최종 결과를 결정하는 방법이다. 따라서 이전의 결과를 우선적으로 채택하여 결과로 결정하되 임계값 등에 의해 확실한 결론을 내리기 어려울 때 다른 방법을 통해서 결과를 예측 한다[5, 13].

음성인식기 인식 결과의 신뢰성은 음성인식기를 통해 구성된 *n-best* 결과중 발화 음성과 같을 확률이 가장 높은 첫 번째 단어의 가능성값과 두 번째로 발화 음성과 같을 확률이 높은 단어의 가능성값의 차이로 판단하며 이 차이가 실험을 통해 결정된 임계값보다 작으면 음성인식기가 결과를 확정적으로 결정하지 못하였다고 판단하고 상황 정보 후처리를 적용한다. 음성인식기 인식 결과의 신뢰성 판단을 위한 임계값 설정은 4.1절에서 설명한다.

순차적 결합 기법을 통한 상황 정보 후처리 적용에서는 가장 먼저 순차 패턴상 결과단어들이 음성인식기로부터 구성된 *n-best* 인식결과 내에 있는지 확인한다. 이전에 사용된 단어에 대한 순차 패턴이 존재하더라도 패턴상 결과 단어가 *n-best* 인식결과 내에 없다면 그 단어는 음성인식기로부터 인식될 가능성이 거의 없는 것으로 판단하여 후처리 보정 대상에서 제외한다. 만약 이전에 사용된 단어에 대한 순차 패턴이 존재 하지 않으면 순차 패턴에 의한 후처리가 이루어질 수 없으므로 이때에는 후처리 보정을 하지 않고 음성인식기의 인식 결과를 최종 후처리 인식 결과로 결정한다. 순차 패턴상 결과단어가 하나이고 *n-best* 인식결과 내에 존재 하면서 가장 높은 가능성값을 가져 음성인식기 인식 결과와 동일한 경우에도 후처리 보정을 하지 않고 최종 후처리 인식 결과를 음성인식기의 인식 결과로 결정한다. 왜냐하면 후처리 보정 대상으로 선정되는 단어가 하나이기 때문이다.

이와 같은 과정을 요약하면 첫째, 음성인식기의 결과를 신뢰할 수 없고 둘째, 순차 패턴상 결과 단어가 음성인식기로부터 구성된 *n-best* 인식결과 내에 하나 이상 존재하며 셋째, 순차 패턴상 결과 단어가 하나일 경우에는 음성인식기의 결과 단어와 다를 때에만 후처리 보정을 한다. 이러한 순차적 결합 방법을 사용함으로써 후처리 이전의 결과를 적절히 신뢰하며 또한 불필요한 후처리 시간을 줄일 수 있다.

4. 실험 결과 및 분석

4장에서는 기존의 음성인식 시스템에 본 논문에서 제안한 상황 정보 후처리 적용 실험을 하고 결과를 분석한다. 실험에서는 논문에서 가정한 개인용 모바일 기기 사용 환경을 위하여 차세대 PMP(portable multimedia player) 기기를 가정하고 10개의 수행 가능한 기능과 32개의 인식 가능한 단어를 설정하였으며 이에 대한 내용은 <표 4>와 같다.

잡음환경을 위해서 일상적인 사무실 환경에서 녹음한 음성 데이터를 이용하여 음성인식기를 학습시키고 실험하였으며 음성데이터의 발화자는 가정한 개인 모바일 기기 환경에 따라 단일 발화자로 하였다. 음성인식기는 각 단어별로 150회씩 발화한 총 4800개의 음성 데이터를 학습에 사용하여 단어별로 50회씩 인식 테스트를 하였을 때 90%정도의 인식률을 보일 수 있도록 하였다.

〈표 4〉 기기의 기능과 각 기능에 속한 명령어 목록

기능명	해당 단어 목록
음악	음악, 듣기
영화	영화, 보기
폴더이동	폴더, 상위, 하위
플레이	시작, 다음, 정지, 종료
소리	소리, 크게, 작게
네비게이션	지도, 출발지, 도착지, 확대, 축소, 거리, 위치
탐색	찾기, 삭제
방향이동	위, 아래, 오른쪽, 왼쪽
확인여부	예, 아니오, 확인, 취소
시간	시간

후처리에 필요한 첫 번째 정보인 사용자 발화 순차 패턴은 실제 사용자의 기록을 바탕으로 구성하였다. 앞서 설정한 기능과 명령어를 포함하는 모바일 기기를 가정하여 사용자로부터 사용 정보들을 기록한 3개의 로그데이터를 생성하였으며 각 데이터의 내용을 순차 패턴과 관련하여 요약하면 <표 5>와 같다. 후처리에 필요한 두 번째 정보인 기능 연관성은 32개 단어에 대하여 앞서 설정한 10개의 기능과 3.2절에서 제안한 3종류의 연관성을 바탕으로 <표 6>과 같이 구성하였다.

〈표 5〉 각 사용자별 데이터 요약

데이터 정보 / 사용자	추출된 패턴의 수 (신뢰도 50% 이상인 패턴의 수)	데이터 내 패턴의 비율(신뢰도 50% 이상인 패턴의 비율)
로그데이터-1	24 개 (13 개)	66.66% (36.11%)
로그데이터-2	29 개 (12 개)	70.73% (29.27%)
로그데이터-3	44 개 (10 개)	57.89% (13.17%)

〈표 6〉 각 기능들의 문맥 연관성
(+:양의 연관성, -:음의 연관성, 0:연관성 없음)

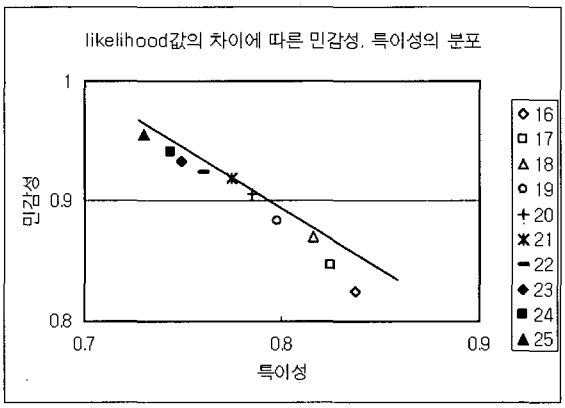
기능 / 기능	음악	영화	폴더이동	플레이	소리	네비게이션	탐색	방향이동	확인여부	시간
음악	+									
영화	-	+								
폴더이동	+	+	+							
플레이	+	+	0	+						
소리	0	0	-	-	+					
네비게이션	-	-	-	-	-	+				
탐색	0	0	-	-	+	-	+			
방향이동	-	-	-	-	+	+	0	+		
확인여부	0	0	+	+	-	0	+	0	+	
시간	0	0	0	0	0	0	0	0	0	+

4.1 후처리 적용

음성인식기의 결과에 무조건적으로 후처리를 적용하는 것은 불필요한 후처리를 증가시킴으로써 후처리 효율을 저하시키는 원인이 되며 또한 잘못된 후처리 발생의 원인이 될 수도 있다. 따라서 음성인식기의 결과를 분석하여 적절한 시점에 후처리를 적용하는 것은 음성인식기의 결과를 어느 정도 신뢰하는 의미를 가지며 불필요한 후처리를 피하여 최대의 후처리 효과를 보이도록 한다. 이를 위하여 본 논문에서는 HTK 음성인식기가 계산한 가능성 값을 분석하여 적절한 후처리 적용 시점을 판단하는 방법을 제안한다.

음성인식기 결과에 대한 신뢰성 판단 실험을 위해 HTK로부터 인식기 학습 데이터에 대한 *n-best* 인식결과와 각 단어에 대한 가능성 값들을 추출하였다. 그리고 *n-best* 인식결과 중에서 발화 음성과 같을 확률이 가장 높은 첫 번째 단어의 가능성 값과 두 번째로 발화 음성과 같을 확률이 높은 단어의 가능성 값의 차이에 대하여 인식기의 인식결과가 맞았는지 여부를 분석하였다. 분석 결과, 일반적으로 가능성 값의 차이가 20전후일 때 인식결과가 틀리는 경우가 많았으며 차이가 클 때는 거의 대부분 인식기의 결과가 발화한 음성과 맞았다. 따라서 본 실험에서는 3.4절에서 제안한 순차적 결합에서 음성인식기의 인식결과 신뢰성 판단을 위한 정확한 임계값을 결정하기 위하여 가능성 값의 차이를 20±5의 범위 내에서 인식기 결과가 맞았는지 여부에 대하여 후처리를 적용했는지 여부를 측정하였다. 결과에 대한 민감성(sensitivity)과 특이성(specificity)[7, 8]을 계산하여 그래프로 그린 결과는 (그림 6)과 같다.

민감성은 후처리가 제때 적용되었는가에 대한 판단 기준으로 인식결과가 틀린 경우에 대해 얼마나 후처리를 적용했는지를 의미하며 (식 5)와 같이 계산된다. 또한 특이성은 불필요한 후처리를 하지 않았는가에 대한 판단 기준으로 인식결과가 맞은 경우에 대해 얼마나 후처리를 적용하지 않았는지를 의미하며 (식 6)과 같이 계산된다. 따라서 민감성과 특이성은 음의 상관성을 가지며 이는 민감성이 높을수록 높은 후처리 효율성을 기대할 수 있으나 불필요한 후처리가 많아지며 특이성이 높을수록 불필요한 후처리는 적어지나 후처



(그림 6) 가능성 값의 차이에 따른 민감성과 특이성의 분포

리의 효율성도 떨어짐을 의미한다. 따라서 민감성과 특이성을 적절하게 만족시키는 시점에서의 가능성 값 차이를 후처리 적용 임계값으로 선정하여 후처리의 비용 및 성능 등을 고려하여야 할 것이다.

$$\text{민감성} = \frac{\text{인식결과가 틀리고 후처리를 적용한 경우의 횟수}}{\text{인식결과가 틀린 경우의 횟수}} \quad \text{(식 5)}$$

$$\text{특이성} = \frac{\text{인식결과가 맞고 후처리를 적용하지 않은 경우의 횟수}}{\text{인식결과가 맞은 경우의 횟수}} \quad \text{(식 6)}$$

본 실험에서는 민감성과 특이성을 동시에 가장 잘 만족시키는 시점을 찾기 위해 민감성과 특이성의 합을 최대로 하는 가능성 값 차이를 음성인식기 인식 결과 신뢰성 판단을 위한 임계값으로 선택하였다. 실험 결과 (그림 6)에서 보는 바와 같이 가능성 값의 차이가 '21'일 때 민감성과 특이성의 합이 최대가 되었다. 따라서 음성인식기로부터 구성된 *n-best* 인식결과 중에서 발화 음성과 같을 확률이 가장 높은 첫 번째 단어의 가능성 값과 두 번째로 발화 음성과 같을 확률이 높은 단어의 가능성 값의 차이가 '21'이하일 때 음성인식기의 인식 결과를 신뢰할 수 없다고 판단하여 후처리를 수행한다.

4.2 성능 평가

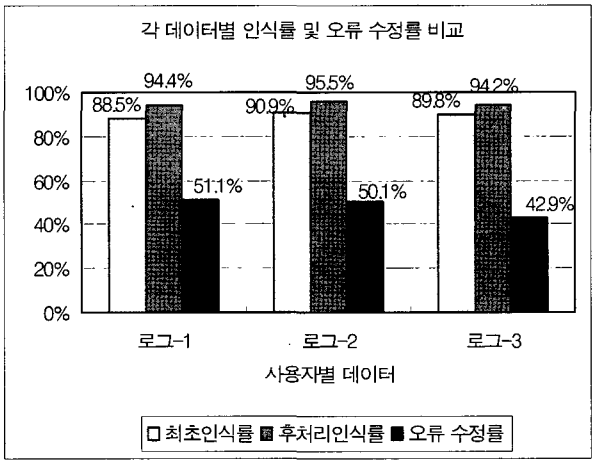
본 절에서는 각기 다른 사용자 발화 패턴 데이터별로 상황 정보 후처리를 적용한 실험에 대한 결과를 기술한다. 실험에서는 서로 다른 음성인식기를 사용한 기존의 후처리 방법들을 객관적으로 비교하기 위해 오류 수정률을 비교하였으며 오류 수정률은 다음의 (식 7)을 이용하여 산출하였다 [3, 4]. 단, E_{SR}은 후처리 적용전의 오인식률을, E_{PP}는 후처리 적용후의 오인식률을 의미한다.

$$\text{오류 수정률} = \frac{E_{SR} - E_{PP}}{E_{SR}} \quad \text{(식 7)}$$

<표 5>에서와 같은 서로 다른 3명의 발화 패턴 데이터를 이용한 후처리 결과는 (그림 7)과 같다. 제안한 방법의 특성상 후처리가 순차 패턴의 포함도에 민감함에도 불구하고 <표 5>의 데이터의 패턴 포함비율과 (그림 7)의 오류 수정률이 완전한 정비례관계를 가지지 않는 이유는 추출된 모든 사용자 발화 순차 패턴이 오류 수정에 도움을 주는 것이 아니기 때문이다. 사용자의 발화 패턴으로 추출되었더라도 발화 단어가 원래 인식률이 좋은 단어인 경우나 3.4절의 후

〈표 7〉 오인식 정정률 비교

음성인식 후처리 방법	최대 오류 수정률
어휘의미 패턴 후처리[4]	36.7%
오류 패턴구문 정정 후처리[3]	8.5%
상황 정보 후처리	51.1%



(그림 7) 각 데이터별 인식률 및 오류 수정률

처리 적용 절차에 의해 음성인식기의 인식 결과가 오인식임에도 높은 신뢰성을 가지는 것으로 판단되는 경우에는 사용자 발화 패턴이 실제 후처리에 도움을 주지 못하였다. 그러나 결과에서 볼 수 있듯이 최초 본 논문에서 가정한 개인 모바일 기기 사용 환경과 같이 사용자의 발화 패턴이 이전의 사용 패턴과 유사할수록 더 높은 오류 수정률을 기대할 수 있다.

<표 7>은 기존에 제안된 오류 패턴구문 정정 후처리 방법[3], 어휘의미 패턴 후처리 방법[4]과 본 논문에서 제안한 상황 정보 후처리 방법을 비교한 결과이다. 제안된 방법들의 특성상 음성인식에 대한 도메인과 후처리를 적용한 조건이 각기 다르기 때문에 절대적인 비교는 불가능 하지만 상황 정보 후처리 방법에서는 (그림 7)과 같이 최대 51.1%까지 인식기의 오류를 수정함으로써 오류패턴구문 정정 후처리나 어휘의미패턴 후처리 등의 다른 후처리 방법들에 비해서 오류 수정률이 우수함을 확인하였다.

5. 결론 및 향후 연구

본 논문에서는 잡음환경과 개인 모바일 기기 환경의 음성인식에서 신호처리 위주 인식의 한계를 극복하고 인식률을 향상시키기 위하여 사용자 정보와 문맥 정보를 현재의 상황 정보로 정의하고 이를 이용한 음성인식 후처리 기법을 제안하였다. 사용자 정보는 사용자의 발화 패턴으로 정의하였으며 문맥 정보는 현재 기기에서 수행중인 기능으로 정의하였다. 이러한 두 가지 상황 정보는 사용자 행동 패턴과 현재 기기의 기능 정보를 가지는 고급 지식(high-level knowledge)으로서 이를 적용한 후처리 역시 사용자 정보와 문맥 정보를 내포한 고급 후처리(high-level post-proces sing)이다. 또한 논문에서는 후처리 효율을 높일 수 있도록 상황 정보 후처리를 적용하기 이전에 사용자 적응형 후처리를 먼저 적용하고 인식기와 사용자 적응형 후처리의 신뢰성을 판단하는 방법을 제안함으로써 불필요한 후처리를 최대한 줄

이고자 하였다. 실험을 통하여 후처리에서 최대 51%까지의 인식기의 오류를 수정하여 제안한 후처리 방법이 음성의 인식률 향상에 상당한 기여를 하였음을 확인하였다. 향후에는 여러 상황에 대한 문맥 정보의 다양한 적용기법과 효과적인 융합 기법을 연구하여 사용자 정보가 효과적이지 못할 경우에도 인식률을 향상시킬 수 있도록 할 것이다.

참 고 문 헌

[1] M. Ostendorf, "From HMM's to segment models: a unified view of stochastic modeling for speech recognition," Speech and Audio Processing, IEEE, Vol.4, pp.360-378, 1996.

[2] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE, Vol.77, No.2, pp.257-286, 1989.

[3] 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.

[4] Minwoo Jeong, Byeongchang Kim, Lee, G.G., "Semantic-oriented error correction for spoken query processing," Automatic Speech Recognition and Understanding, IEEE, pp.156-161, 2003.

[5] Myung Won Kim, Joung Woo Ryu, Eun Ju Kim, "Speech recognition by integrating audio, visual and contextual feature based on neural networks," International Conference on Natural Computation, LNCS 3614, pp.155-164, 2005.

[6] J. Pei, J. Han, B. Mortazavi-Asl, H. Pinto, Q. Chen, U. Dayal and MC. Hsu., "PrefixSpan: mining sequential patterns efficiently by prefix-projected pattern growth," International Conference on Data Engineering, pp.215-224, 2001.

[7] Jiawei Han, Micheline Kamber, 'Data mining: concepts and techniques', Morgan Kaufmann Publishers, Academic Press, 2001

[8] Richard J. Roiger, Michael W. Geatz, 'Data mining: a tutorial-based primer', Addison Wesley, Peardon Education, Inc., 2003.

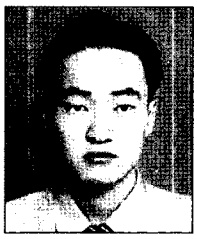
[9] Yi Ding, Xue Li, "Time weight collaborated filtering," Proceedings of the 14th ACM international Conference on Information and Knowledge Management, pp.485-492, 2005.

[10] C. C. Aggarwal, J. Han, J. Wang, and P. S. Yu. "A framework for projected clustering of high dimensional data streams," Conference on Very Large Data Bases, pp.852-863, 2004.

[11] Steve Young, et., 'The HTK book(Version 3.1)', Cambridge University Engineering Department, 2001.

[12] HTK Speech Recognition Toolkit, http://htk.eng.cam.ac.uk/, Cambridge University Engineering Department.

[13] 도영아, 김종수, 류정우, 김명원, "협력적 추천을 위한 사용자와 항목 모델의 효율적인 통합 방법", 한국정보과학회 논문지:소프트웨어 및 응용, Vol.30, No.6, pp.542-549, 2003.



송 원 문

e-mail : gtangel@ssu.ac.kr

2004년 한신대학교 전자계산학과 졸업

2006년 숭실대학교 대학원 컴퓨터학과 (공학석사)

2006~현재 숭실대학교 대학원 컴퓨터학과 박사과정

관심분야 : 신경망, 음성인식, 데이터마이닝, 개인화

김 명 원

e-mail : mkim@comp.ssu.ac.kr

1972년 서울대학교 응용수학과 졸업

1981년 University of Massachusetts (Amherst) Computer Science 석사학위 취득.

1986년 University of Texas (Austin) Computer Science 박사학위 취득

1975년~1978년 한국과학기술연구소 연구원

1982년~1985년 Institute for Computing Science & Computer Application (Univ. of Texas) 연구원

1985년~1987년 AT&T Bell Labs. (Naperville) 연구원

1987년~1994년 한국전자통신연구소 책임연구원

1992년~1993년 한국신경회로망 연구회 회장

1993년~1995년 IEEE Neural Network Council 한국지부장

1993년~1995년 정보과학회 뉴로컴퓨팅연구회 위원장

1994년~현재 숭실대학교 컴퓨터학부 교수

1998년~2000년 한국인지과학회 부회장

2000년~2001년 미국 IBM T.J Watson 연구소 방문과학자

2001년~2002년 한국뇌학회 회장

2002년~2003년 숭실대학교 정보지원처장

2004년~2006년 숭실대학교 정보과학대학원장

관심분야 : 신경회로망, 퍼지시스템, 진화알고리즘, 패턴인식, 자동추론, 기계학습, 데이터마이닝, creativity engineering 등