사용자 발화 순차패턴을 이용한 음성인식 후처리

송원문°, 김은주, 김명원
숭실대학교 컴퓨터학부
{gtangel°, blue7786}@ssu.ac.kr, mkim@comp.ssu.ac.kr

Post-Processing of Speech Recognition Using User Utterance Sequential Pattern

WonMoon Song°, EunJu Kim, MyungWon Kim
School of Computing, Soongsil University

요 약

최근 음성인식 분야 에서는 발화된 음성의 단순한 신호 처리위주의 인식 결과로부터 좀 더 신뢰할 수 있는 결과를 얻기 위하여 여러 가지 후처리 기법들이 연구 되고 있다. 본 논문에서는 개인 사용자를 위한 음성 명령어 인식 환경에서 사용자의 발화 정보를 후처리에 적용함으로써 사용자 정보를 고려한 음성인식 후처리 기법을 제안한다. 먼저 이전에 사용했던 음성 명령어들로부터 명령어 발화 순차 패턴 규칙을 추출한 후 사용자가 사전에 발화한 명령어를 바탕으로 구성된 순차 패턴을 비교 하여 순차 규칙상 얻어 질수 있는 단어를 결정한다. 이렇게 얻어진 단어를 고려하여 음성인식기 인식단어 후보들의 확률값을 적절히 보정한 후 최종 인식 단어를 재결정한다. 이러한 과정에서 적절한 보정을 위하여 발화 순차 패턴의 신뢰도와 인식기의 결과단어를 고려한 보정 방법을 제안한다. 실험을 통하여 제안한 후처리를 이용한 음성인식이 HMM을 이용한 기본 음성인식에 비해 오류율을 15%이상 낮추어 인식률에 상당한 기여를 하였음을 확인할 수 있다.

1. 서 론

최근 컴퓨팅 환경이 더욱 다양하고 복잡해지면서 이를 사용하는 사용자를 위하여 사용자가 좀 더 쉽고 인간 친화적인 접근을 할 수 있도록 하는 많은 연구들이 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 자동차 네비게이션, 음성기반 검색시스템, 자동응답시스템 등의 여러 분야에 활발히 적용되고 있다. 이러한 음성인식의 핵심 알고리즘은 다른 모델에 비해 특히 성능이 좋은 HMM(Hidden Markov Model) 알고리즘이 주로 사용되고 있지만[1][2] 기본적으로 음성의 신호처리 단계에서 생기는 한계점과 HMM의 세 가지 가정 때문에 실제로는 인식률이 크게 좋아지지 않는 문제점이 있다[2]. 이러한 문제점을 해결하여 인식률을 높이기 위한 방법으로 인식결과에 후처리를 하여 인식기의 인식결과를 조정하는 음성인식 후처리에 대한 여러 가지 연구들이 진행되고 있다.

이미 제안된 방법으로는 인식 단어의 오류 패턴이나 단어를 포함하는 블록의 오류 패턴을 이용한 후처리 방법[3], 단어의 어휘적, 의미적 카테고리를 이용한 후처리 방법[4] 등이 있으며 이러한 방법들은 단어의 발음 및 어휘적 특성이나, 의미적 카테고리를 사용한 일반적인 접근 방법으로써 음성인식 후처리의 응용분야에 쉽게 적용될 수 있다. 하지만 이러한 단계에서의 접근은 사용자가 상황에 따라 특정한 패턴으로 발화하는 경우나 개인 정보에 의해서 발화 내용이 틀려지는 경우에 사용자의 발화패턴이나 정보를 고려하지 못하므로 인식률 향상의 저해 요인이 된다.

따라서 본 논문에서는 개인 사용자에 대한 음성인식 환경을 배경으로 단어의 어휘나 의미단계를 넘어서는 고급지식(high-level knowledge)인 사용자 정보를 활용하여 음성인식 후처리를 하는 방법을 제안하며 이를 위하여 사용자 발화 순차 규칙을 적용한다. 제안한 방법에서는 먼저 사용자가 이전에 사용했던 음성 명령어들에서 추출될 패턴의 지지도와 신뢰도를 제한하여 신뢰성 있는 순차적 발화 패턴을 추출한 후 명령어 발화 순차 패턴 규칙을 구성한다. 이후 사용자가 발화한 음성에 대하여 음성인식기의 인식단어 후보들 중에서 최종 인식단어를 결정하기 전에 사전에 발화된 명령어에 대한 순차 패턴을 찾는다. 최종적으로 찾은 순차 패턴의 신뢰도를 적절히 고려하여 새로운 인식결과를 도출해 낸다.

2절에서는 지금까지 국내외 연구들에서 제안된 주요 후처리 방법에 대하여 간단히 소개하고 3절에서는 본 논문에서 제안하는 사용자 발화 순차 패턴을 이용한 후처리 기법에 대해서 기술한다. 4절에서는 제안한 방법에 대한 실험 결과를 기술하고 분석하여 타당성을 검증하며 5절에서는 문제점에 따른 향후 연구 과제에 대해서 검토 한다.

2. 관련 연구

2.1 오류 패턴 비교(Error-Pattern Matching)

Satoshi Kaki[3]는 미리 구축된 오류 패턴 데이터를 이용한 EPC(error-pattern correction)와 SSC(similar-string correction)의 두 가지 방법을 사용하여 음성인식 오류를 수정하는 후처리 방법을 제안하였다. 먼저 EPC방법에서는 훈련 데이터를 통해 미리 구축된 오류인식이 잘되는 단어와 그 단어에서 발생한 오류 형태의 쌍으로 이루어진 오류 패턴 데이터베이스를 이용해서 인식결과 내에서 오류로 예상되는 부분을 추출하여 에러를 보정하게 된다. 이러한 방법은 미리 구축된 에러 패턴 데이터베이스에 상당히 민감하기 때문에 오류로 예상되는 부분을 추출할 때 몇 가지 조건을 주어 이를 보완하며 고립단어 인식에 사용될 수 있다. 두 번째 방법인 SSC방법에서는 훈련데이터를 통해 구성된 데이터의 오류 인식 단어 대신 오류 인식 블록을 사용한다. 이는 단어의 오류가 독립적으로 일어나는 것이 아니고 전후에 같이 나온 단어에 의해 오류가 일어나는 경우가 많음에 착안한 방법이며 연속음성 인식에 쓰일 수 있다. 이러

한 두 가지 방법을 조합하여 사용함으로서 [3]에서는 인식기의 오류율이 8.5% 감소되었다.

이 방법은 에러 패턴에 대한 데이터가 정확하고 응용 도메인이 적어 데이터가 적게 구성될 경우 효율적인 후처리를 기대할 수 있다. 그러나 후처리 기준에 단어나 단어를 포함하는 블록 자체에 대하여 인식기에서 측정된 오인식 패턴들만을 사용하므로 특정 사용자에 대한 정보를 고려해야 하는 개인사용자를 위한 음성인식 환경에서는 강건한 모델이라고 할 수 없다.

2.2 어휘 의미 패턴(LSP:Lexico-Semantic Pattern)

정민우[4]는 기존의 대부분의 후처리 방법이었던 어휘 중심적 접근 방법만이 아닌 어휘 및 의미적인 정보를 모두 고려하여 인식결과를 수정하기 위해 LSP를 제안함으로써 인식 단어에 대한 의미적인 후처리를 시도 하였다. LSP는 연속음성 인식에서 발화된 문장을 단어별로 각각 어휘 및 의미정보를 포함한 특정 스트링으로 대치한 것이라고 볼 수 있으며 사용될 LSP는 훈련 데이터를 통하여 미리 구성되어 진다. 실제 후처리는 발화한 문장에 대한 인식결과를 인식된 단어들에 맞는 LSP로 바꾼 후 이를 미리 구성된 LSP들과 비교 한다. 이때 미리 구성된 LSP중 인식결과 LSP와 제일 유사한 것이 선택되며 인식결과 LSP를 선택된 LSP로 바꾸어 먼저 의미적 오류를 수정 한다. 이후 수정된 LSP내의 각 스트링들을 실제 단어로 바꾸는 어휘적 오류 수정을 통하여 인식결과를 도출한다. 이 방법으로 [4]에서는 인식률이 약 8%정도 향상 되었다.

이 방법은 의미적인 후처리를 구현했다는데 의의가 있다. 그러나 LSP역시 단어 자체에 대한 의미정보를 이용하고 있으므로 특정 사용자에 대한 발화 패턴이나 개인 정보 등의 특성을 고려하지 못한다.

3. 사용자 발화 패턴을 이용한 후처리

이번 절에서는 사용자 발화 패턴 마이닝에 대한 간단한 방법을 설명하고 본 논문에서 제안한 방법인 추출된 순차 패턴 규칙과 각 규칙에 따른 신뢰도를 이용하여 음성인식기 결과를 보정한 방법에 대해 설명한다.

3.1 사용자 발화 순차 패턴 마이닝

음성인식의 후처리를 위한 사용자의 발화 순차 패턴 마이닝에는 PrefixSpan알고리즘을 사용하였다. PrefixSpan 알고리즘은 명령어 환경에서와 같이 크지 않은 데이터베이스에서 작은 길이의 순차 패턴을 마이닝할 때 다른 알고리즘에 비해 성능이 좋으며[5] 또한 일반적인 순차 패턴 마이닝에 최근 가장 보편적으로 사용되고 있다[6].

본 논문에서는 사용자가 이전에 사용했던 순차적 명령어를 기록한 데이터베이스로부터 명령어 순차 패턴을 마이닝 하였다. 이때 최소지지도와 최소신뢰도를 설정하여 의미 있는 규칙들이 생성될 수 있도록 하였다. 이렇게 마이닝된 순차 규칙들은 사용자의 특성과 개인 정보를 반영하고 있는 고급 지식(high-level knowledge)으로써 이를 음성인식의 후처리에 반영하는 것은 음성인식의 결과가 사용자의 특성과 정보를 가지고 있는 결과이다.

[표-1] 마이닝된 사용자 발화 순차 패턴의 예

사용자 발화 순차 패턴	신뢰도
음악 → 듣기	0.8(80%)
네비게이션 → 찾기	0.7(70%)

마이닝된 순차 패턴은 [표-1]과 같은 식으로 구성되며 첫 번째 예는 사용자가 "음악"이라는 명령을 발화한 후에는 "듣기"라는 명령을 발화할 확률이 80%라는 것을 의미한다.

3.2 순차 패턴을 이용한 후처리 방법

3.1절에서 얻어진 사용자 발화 순차 패턴을 음성인식의 후처리에 이용한다. 먼저 사용자가 어떠한 단어를 발화 하였을 때 사전에 발화한 단어가 포함된 순차 패턴 규칙을 찾는다. 이 후 패턴에 따른 결과로 선정될 수 있는 단어가 인식기 결과 후보 단어들 중에 포함되어 있을 경우 해당 규칙의 신뢰도를 이용하여 후보단어들의 확률값을 보정한다. 이때 인식기로부터 산출된 확률값을 유지하고 확률값들의 차이를 고려하여 적당히 보정하기 위하여 해당 단어 W에 대한 보정값($W(V)$)은 다음과 같이 산출 하였다.

$$W(V) = (MAX(L_{set}) - MIN(L_{set})) * W(C) \quad (1)$$

식에서 $W(C)$는 보정될 단어 W를 결과로 가지는 순차 규칙의 신뢰도를 의미하며 $MAX(L_{set})$과 $MIN(L_{set})$은 각각 인식기의 후보단어 확률값들중 최대값과 최소값을 의미한다. 이러한 보정값을 이용하여 단어 W에 대하여 새롭게 계산되는 확률값($W(L')$)은 다음과 같다.

$$W(L') = W(L) + W(V) \text{(단, } W(L) \text{은 원래의 확률값)} \quad (2)$$

인식기가 산출한 단어의 확률값을 계산된 새로운 확률값으로 대치하여 결과 후보들 중에서 최대의 확률값을 가지는 단어를 재선정 하였다. 이러한 과정을 통하여 최종적으로 인식된 단어는 단순히 신호처리만을 거친 단계의 인식결과가 아닌 사용자의 발화 패턴과 특성을 내포한 결과이다.

4. 실험 결과

후처리 효율을 높이기 위해서는 음성인식기의 인식 결과에 대한 신뢰도가 낮은 경우에만 후처리를 적용하여 불필요한 처리를 줄여야 한다. 이를 위하여 4.1절에서는 후처리 적용 시점을 결정하기 위해 음성인식기로부터 계산된 확률값을 분석한 실험결과를 기술하였다. 4.2절에서는 4.1절의 결과를 바탕으로 후처리를 적용해야 한다고 판단되었을 경우 3.2절에서 제안한 식 (1)과 (2)를 이용한 보정방법으로 순차 규칙의 평균 신뢰도를 80%~ 50%까지 변화시켜가며 인식기의 인식률을 실험한 결과를 기술하였다.

실험을 위하여 HMM을 이용한 기본 음성인식기[7](ezCSR)를 사용하였으며 명령어라고 가정한 총 41개의 단어에 대하여 일상적인 잡음 환경에서 한명의 발화자가 각 단어를 10번씩 발화한 총 410개의 데이터를 사용하였고 순차 규칙은 [표-1]에서와 같은 예를 사용하였다.

4.1 후처리 적용시점 판단 실험

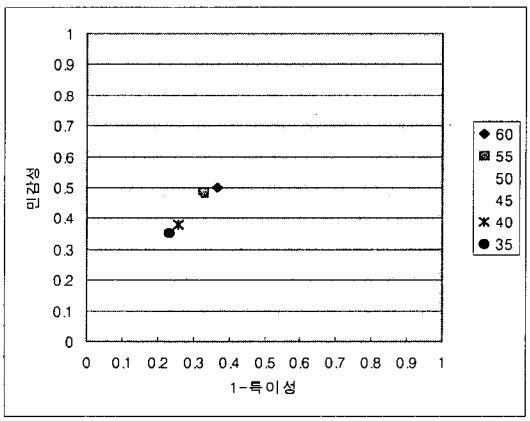
본 논문에서 사용한 인식기와 같이 HMM을 이용한 음성인식 방법들은 기본적으로 발화한 음성신호에 대하여 인식 가능한 단어들의 확률값을 계산하여 최대의 확률값을 가지는 단어를 인식결과로 결정한다[1][2][7]. 이렇게 결정되는 인식기의 결과에 매번 후처리를 적용하는 것은 인식의 속도 저하와 인식률 향상의 저해 요인이 된다. 따라서 인식기의 결과를 분석하여 적절한 시점에 후처리를 적용하는 것은 인식기의 결과(인식률)를 어느 정도 신뢰하는 의미를 가지며 또한 불필요한 후처리를 하지 않아 최대의 인식률 향상을 기대 할 수 있는 방법이다.

적절한 후처리 적용시점 판단을 위한 실험을 위하여 사용자가 어떠한 단어를 발화했을 때 인식 가능성이 높은 단어(후보단어)에 대해서 인식기가 계산한 최종 확률값들을 추출하였다.

이 후 확률값들중 인식기 결과단어에 대한 확률값(확률값들중 최대값)과 두 번째로 가능성이 있는 단어에 대한 확률값(확률값들중 두 번째로 높은값)을 비교 하여 그 차이(후처리 적용 임계값)와 인식률의 상관성을 분석하였다. 일반적으로 그 차이가 50전후(차이가 별로 없을 때) 일 때 인식결과가 틀리는 경우가 많았으며 차이가 큰 경우는 거의 인식기의 결과가 발화한 단어와 맞았다.

따라서 본 실험에서는 후처리 적용을 위한 정확한 임계값을 선정하기 위하여 임계값을 30에서 50까지 5씩 증가 시켜 가면서 인식기 결과가 맞았는지 여부에 대하여 후처리를 적용했는지 여부를 측정하였다. 그 결과에 대한 특이성(인식결과가 틀린 경우에 후처리를 한 경우의 확률)과 민감성(인식기의 결과가 맞은 경우에 후처리를 하지 않은 경우의 확률)을 계산하여 도표로 그린 결과는 [그림-1]과 같다. 민감성과 특이성이 클수록 좋은 임계값이라고 할 수 있으며 그래프로 표현하였을 때 특이성을 1-특이성 값으로 변환하여 그래프 상에서 왼쪽 상위에 있는 값일수록 좋은 값임을 나타낸다.

[그림-1]로부터 후처리를 적용했을 때 불필요한 후처리의 횟수를 가능한 줄여 후처리의 극대화를 기대 할 수 있는 시점의 임계값이 55임을 알 수 있다. 따라서 4.2절에서는 임계값이 55인 경우에 사용자의 발화 순차 규칙을 이용하여 인식기의 결과에 후처리를 적용한 실험결과를 보인다.



[그림-1] 임계값에 따른 민감성과 1-특이성의 분포

4.2 후처리 적용 후 최종 인식 결과

4.1절에서의 실험 결과를 통해서 인식기 후보결과의 확률값에 대한 임계값이 55일 때 후처리를 적용해야 함을 알 수 있다. 따라서 본 절에서는 인식기의 후보결과들에 대한 확률값들 중 최대값과 그 다음값의 차이가 55이하인 경우에 3.4절에서 제안한 보정방법을 이용하여 후처리를 하여 인식결과를 재조정하는 실험을 해 보았다. 또한 순차 규칙의 신뢰도와 비교하여 좀 더 객관적인 인식률을 알아보기 위하여 후처리에 적용된 순차 규칙의 평균 신뢰도를 50%에서 80%까지 10%씩 변경해 가며 오인식률을 확인해 보았으며 결과는 [표-2]와 같다.

실험 결과를 통해서 후처리에 적용된 순차 규칙의 평균 신뢰도가 50%에서 80% 사이일 때 본 논문에서 제안한 후처리 방법의 오인식률이 후처리를 하지 않은 인식기의 오인식률에 비해 최대 16%정도 낮아져 실제 인식률이 향상되었음을 확인할 수 있다.

[표-2] HMM 기반 인식기에 대한 후처리 전과 후처리 후의 오인식률 비교

인식결과 추출 기준		오인식률
HMM Baseline ASR		39%
순차 규칙을 이용한 후처리	평균 신뢰도 50%	26%
	평균 신뢰도 60%	24%
	평균 신뢰도 70%	23%
	평균 신뢰도 80%	23%

5. 결론 및 향후연구

음성인식의 인식률을 높이기 위해서는 음성 신호를 분석하여 인식결과를 얻어내는 기본 단계를 넘어서 인식기의 인식결과에 후처리를 적용하는 것이 필수 불가결 하다. 본 논문에서는 사용자 발화 패턴이라는 개인의 특성과 정보를 가지는 고급지식(high-level knowledge)을 음성인식의 후처리에 적용하는 방법을 제안하였다. 이렇게 후처리된 음성인식의 결과는 개인의 특성과 정보를 그대로 포함하고 있는 결과이다. 실험을 통해서 제안한 후처리 방법의 타당성을 검증 하였고 낮아진 오인식률을 통해 인식성능이 향상되었음을 보였다.

향후 연구에서는 각 단어에 대한 발화수와 발화자를 늘려서 실험 결과의 타당성을 높이는 연구가 진행되어야 할 것이다. 또한 후처리 보정방법에 인식기의 인식결과에 대하여 순차 규칙의 신뢰도가 더욱 적절하게 반영되는 의미를 부여하기 위해 인식기 결과 단어나 순차 규칙의 결과 단어에 대한 인식기 인식률을 반영하는 보정방법에 대한 연구를 진행할 것이다.

6. 참고 문헌

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

[2] L. Rabiner, "A tutorial on hidden Markov models and selected applicationsin speech recognition," Proc. 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.", COLING-ACL, pp.653-657, 1998

[4] Minwoo Jeong, Byeongchang Kim, Lee, G.G., "Semantic-oriented error correction for spoken query processing", ASRU IEEE Workshop on, pp.156-161, 2003

[5] 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." ICDE, pp.215-224, 2001

[6] 이순신, 김은주, 김명원, "다차원 순차패턴 마이닝을 위한 효율적 알고리즘", 한국정보과학회 2004 추계학술대회, VOL. 31 NO. 02 pp.0214 ~ 0216, 2004

[7] 권오욱, 박준, 황규웅, "의사 형태소 단위 대어휘 연속음성 인식기 개발", 제 15회 음성통신 및 신호처리 워크샵 논문집, pp.320-323, 1998