

# 음성 질의 처리를 위한 의미 기반 오류 수정

정민우\* 김병창\*\* 이근배\*

\* 포항공과대학교 컴퓨터공학과 자연언어처리연구실

\*\* 위덕대학교 컴퓨터멀티미디어학부

## Semantic-oriented Error Correction for Spoken Query Processing

Minwoo Jeong\*, Byeongchang Kim\*\*, Gary Geunbae Lee\*

\* NLP Lab, Dept. of CSE, POSTECH

\*\* Division of Computer and Multimedia, UIDUK Univ.

stardust@postech.ac.kr, bckim@uiduk.ac.kr, gblee@postech.ac.kr

### Abstract

Voice input is often required in many new application environments such as telephone-based information retrieval, car navigation systems, and user-friendly interfaces, but the low success rate of speech recognition makes it difficult to extend its application to new fields. Popular approaches to increase the accuracy of the recognition rate have been researched by post-processing of the recognition results, but previous approaches were mainly lexical-oriented ones in post error correction. We suggest a new semantic-oriented approach to correct both semantic level and lexical errors, which is also more accurate for especially domain-specific speech error correction. Through extensive experiments using a speech-driven in-vehicle telematics information application, we demonstrate the superior performance of our approach and some advantages over previous lexical-oriented approaches.

### I. 서론

음성 인터페이스는 인간에게 가장 친숙한 사용자 인터페이스로서 음성 기반 정보 검색, 차량 내 내비게이션 시스템, 유비쿼터스 컴퓨팅 등의 응용에 유용하게

사용될 수 있다. 음성 입력은 공간적인 제약으로 키보드 사용이 어렵거나, 사용자의 편의성이 필요한 많은 응용 시스템과의 결합이 시도되고 있다.

하지만 현재 음성 인식의 낮은 성능으로 인해 응용 시스템으로 실용적인 활용이 어려운 실정이다. 그렇기 때문에 음성 인식 성능을 향상시키기 위한 많은 연구들이 수행되어왔다. 그러나 대부분의 음성 인식 성능 향상 기법은 신호 처리에 기반을 두고 있으며, 응용 시스템과의 결합을 고려하지 않고 있다. 언어 처리 레벨에서의 음성 인식 향상에 관한 연구로는 음성 인식 후처리에 관한 연구가 수행되고 있다.

후처리에 관한 기존 연구는 대부분 통계적인 방법을 통해 단어를 복원하거나, 오류 패턴을 수집하여 이용한다. 이러한 시스템들의 성능은 음성 인식 결과나 수집된 오류 패턴에 의존적이게 된다. 또한 풍부한 양의 오류 패턴이 요구되어, 비용적인 문제가 발생하게 된다.

본 논문은 언어 처리 수준에서의 음성 인식 향상 방법으로서, 음성 질의 처리를 위한 의미 기반 음성 인식 오류 수정을 제안한다. 의미 기반 오류 수정 방법에서는 어휘 정보뿐 아니라, 구문 및 의미 정보와 같은 상위 레벨의 언어 지식을 이용한다. 의미 기반 방법은 환경에 종속적인 오류 패턴에 덜 의존적이며, 의미적인 오류에 보다 강건하게 된다.

## II. 관련 연구

음성 질의 처리의 가장 큰 문제는 낮은 음성 인식률로 인해 정보 검색 및 질의응답 시스템 등의 응용 시스템의 성능이 저하되는 것이다. 오류가 포함된 질의는 응용 시스템의 성능을 저하시키는 주된 원인이 된다. 이에 따라 인식 오류가 응용 시스템에 미치는 영향을 분석하기 위해서 음성 인식기와 정보 검색 시스템을 결합하여 시스템의 성능 저하 정도를 분석한 연구가 수행되었다 [1, 2, 3]. 이러한 연구에 의하면, 음성 질의에 포함된 오류로 인해 정보 검색 시스템의 성능이 뚜렷하게 저하되며, 단순히 음성 인식기와 응용 시스템을 결합하는 것만으로는 만족할 만한 성능을 낼 수 없다는 것을 알 수 있다.

이러한 문제를 해결하기 위해서 후처리에 관한 여러 가지 연구가 수행되어 왔다. Ringger 는 TRAIN-95 도메인에 대하여 음성 대화 시스템을 구축하기 위해 Noisy Channel Model 에 기반을 둔 음성 인식 후처리 기법을 제안하였다 [4]. 이 방법은 인식기가 사용되는 발화 환경, 화자, 잡음 등의 특성을 Noisy Channel Model을 통해 모델링하고, 이를 이용하여 인식 오류 단어를 통계적으로 수정한다. 하지만 Noisy Channel Model 은 많은 양의 학습 데이터 즉 인식 스크립트를 요구한다.

또한 Kaki 는 음성 인식 결과로부터 오류 패턴을 수집하여 이를 오류 후처리에 사용하였다 [5]. 오류 패턴은 마찬가지로 많은 양의 인식 스크립트를 통해, 오류 단어의 앞뒤 문맥 정보를 이용하거나 휴리스틱을 이용하여 수집한다. 이러한 방법은 음성 인식 오류의 유형이 특정한 패턴을 가지고 나타난다는 가정에서 출발하였으나, 환경이나 화자가 달라지면 패턴이 달라진다는 문제점을 가지게 된다. 하지만 이 방법은 올바른 단어 또한 오류로 인식하게 되는 문제점을 갖는다.

이러한 대부분의 이전 연구들은 어휘 정보에 바탕을 둔 오류 수정 방법을 이용하며, 화자나 발화 환경에 종속적이다. 또한 많은 양의 인식 스크립트를 이용하므로 많은 비용이 들게 된다.

## III. 음성 인식 오류 수정

음성 질의 처리를 위해서는 정확한 음성 인식이 요구된다. 특히 질의 처리를 위해서는 도메인에 특정한 개체명(Named Entity)의 인식이 정확하여야 하는데, 음성 인식기의 특성상 특정 개체명보다는 일반 단어의 인식률이 높게 나타난다 [3]. 그러므로 본 논문에서 제안하는 의미 기반 오류 수정 방법에서는 이러한 특정

도메인의 개체명에 대한 지식을 이용하여, 음성 질의에 포함된 오류를 수정한다.

### 1. 도메인 지식 구축

의미 기반의 음성 인식 오류 수정을 위해서는 도메인 지식 구축이 선행되어야 한다. 도메인 지식은 후보가 되는 질의들을 LSP(Lexico-Semantic Pattern)로 변환시킨 질의 템플릿 데이터베이스와, 이 변환에 사용되는 의미 카테고리 사전들로 구성된다. LSP는 정보 추출, 질의응답 시스템에 적합한 어휘 정보와 의미 정보로 이루어진 추상화된 문장의 패턴으로 정의할 수 있다 [6]. 표 1은 LSP로 의미 추상화된 템플릿의 예제이다. 템플릿 데이터베이스는 후보 질의들을 의미적으로 추상화시킨 도메인 지식으로, 추상화시킬 대상이 되는 어휘들은 의미 카테고리 사전을 참조하여 구성된다. 또한, 의미 카테고리 사전은 도메인 사전과 온톨로지 사전으로 구성되며, 도메인 사전은 특정 도메인에 종속적인 개체명의 의미를 포함하며, 온톨로지 사전은 일반적인 단어의 의미를 포함하게 된다.

Phrase	LSP
Reading trainer	%hobby @position
Fairy tale trainer	
Fairy tale oral narrator	
Recreation coach	

표 1 의미 추상화된 템플릿 예제

### 2. 의미 기반 오류 수정

그림 1은 구축된 도메인 지식을 이용하여, 오류가 포함된 음성 질의를 의미 기반 음성 인식 오류 수정을 통해 수정하는 과정을 나타낸다.

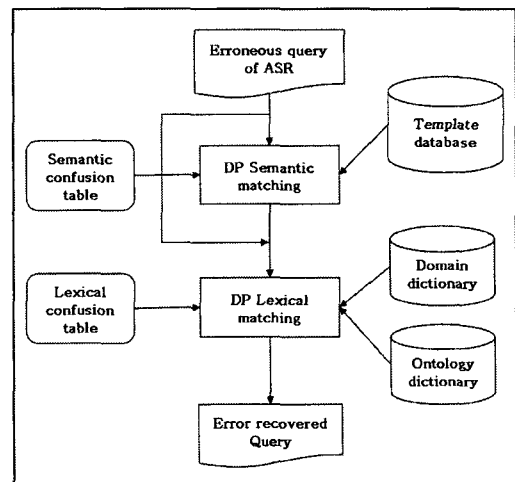


그림 1 의미 기반 오류 수정 과정

의미 기반 오류 수정 과정은 크게 의미 복원 단계와 어휘 복원 단계를 거치게 된다. 의미 복원 단계에서는

먼저 오류를 포함하는 음성 인식 결과를 LSP로 변환하는 과정을 거친다. 오류가 포함된 LSP는 미리 구축된 템플릿 데이터베이스에서 가장 유사한 템플릿으로 교체된다. 이때 가장 유사한 템플릿을 찾는 방법으로는 Minimum Edit Distance를 이용한 동적 검색(Dynamic Programming: DP) 기법을 사용한다 [7]. Minimum Edit Distance는 철자 수정, OCR 오류 수정 및 DNA 서열 분석 등에 사용되는 동적 프로그래밍 기법으로, 두 단어 사이의 최소 거리(minimum distance)는 최소의 삽입, 삭제, 대치되는 글자의 수로 정의된다. 의미 복원에서 이를 적용하기 위해서 의미 혼동 테이블(confusion table) 혹은 혼동 확률(confusion probability)을 이용하여 의미 단위의 최소 거리를 계산한다.

이 단계를 거친 후, 다음으로 어휘 복원 단계를 수행한다. 의미 복원 단계를 통해 복원된 의미 태그와, 오류가 포함된 질의의 오류 어휘가 어휘 복원 단계에서 이용된다. 오류가 포함된 질의와 복원된 의미 태그를 동적 프로그래밍 기법을 통해 정렬하고, 오류 어휘로 검출된 어휘는 올바른 어휘로 복원시킨다. 이때 같은 의미를 지닌 많은 어휘 후보들이 생길 수 있으며, 오류 어휘와 가장 유사한 어휘를 선택하기 위해 어휘 단위에서의 Minimum Edit Distance를 이용한다. 어휘 복원 단계에서는 의미 복원과 유사하게 어휘 혼동 테이블(confusion table)을 이용하여, 가장 유사한 어휘로 복원하게 된다. 그림 2는 차량 내 네비게이션 시스템의 음성 질의에 대한 의미 기반 오류 수정의 예이다.

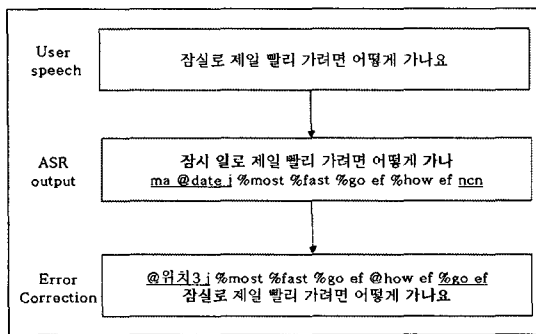


그림 2 의미 기반 오류 수정 예

#### IV. 실험

음성 인식 오류 수정을 통한 음성 인식을 향상을 평가하기 위해 텔레매틱스 서비스를 위한 차량 내 네비게이션 관련 음성 질의 462 문장을 이용하였다. 또한 실험에 사용된 음성 인식기는 LG 대화체 음성 인식기<sup>1)</sup>와 ByVoice 2.0<sup>2)</sup>을 이용하였다.

또한 기존의 Noisy Channel Model 과의 비교 성능

평가를 위해 SLM toolkit을 이용하여 언어 모델을 구축하고 [8], GIZA toolkit을 이용하여 채널 모델을 구축하였다 [9]. 의미 기반 오류 수정 모델을 위해서 436개의 템플릿을 수집하였고, 3,195 엔트리의 도메인 사전과 13,154 엔트리의 온톨로지 사전을 구축하였다.

표 2는 음성 인식기(baseline), Noisy Channel Model, 의미 기반 오류 수정 방법의 인식 정확도를 비교한 결과를 보여준다. 차량 내 네비게이션 도메인에서의 Baseline 시스템의 성능은 약 79%~81% 정도임을 볼 수 있다.

	LG 인식기	ByVoice
Baseline	79.51%	81.25%
Noisy-Channel	85.79%	84.57%
의미 기반	87.03%	86.93%

표 2 인식 정확도 비교

결과를 통해 음성 인식 결과의 후처리를 통해 인식이 향상됨을 알 수 있다. Baseline 시스템과 비교하여 볼 때, Noisy Channel Model 은 각각 6.2%, 3.3%, 의미 기반 오류 수정 방법은 각각 7.5%, 5.7% 성능이 향상되었다.

표 3은 음성 인식 결과와 후처리를 통한 오류 복원 결과가 음성 질의 처리에 어떠한 영향을 미치는지를 알아보기 위하여, 정보 검색이나 질의응답의 성능에 중요한 영향을 끼치는 개체명에 대한 인식 오류율을 비교한 결과이다. 여기에서 WER(Word Error Rate)은 일반 단어와 개체명을 포함한 인식 오류율을 나타내며, TER(Term Error Rate)은 개체명의 오류율을 나타낸다.

	LG 인식기		ByVoice	
	WER	TER	WER	TER
Baseline	20.49%	56.17%	18.75%	64.14%
Noisy Channel	14.21%	23.32%	15.43%	30.03%
의미 기반	12.97%	19.34%	13.07%	27.02%

표 3 WER과 TER의 비교

결과에서 볼 수 있듯이 의미 기반 오류 수정을 통해 음성 인식이 향상되었으며, 특히 질의 처리에 필요한 개체명의 인식을 향상이 두드러짐을 알 수 있다. 비교 실험을 통해 개체명 인식률에 관해 의미 기반 오류 수정 방법은 Baseline 시스템과 비교하여 각각 36.8%, 37.12%의 두드러진 성능 향상을 보여준다. 이

1) LG전자기술원 (<http://www.lgcit.com>)

2) 보이스텍 (<http://www.voicetech.co.kr>)

를 통해 의미 기반 오류 수정 방법이 기존의 어휘 기반의 후처리 시스템에 비해 질의 처리를 수행하는 응용 시스템에 더 적합하다고 할 수 있다.

## V. 결론

본 논문은 음성 인터페이스가 특정 응용 시스템에 적합하게 결합되어 사용될 수 있도록 음성 질의 처리를 위한 의미 기반의 오류 수정 기법을 제안하였다. 의미 기반 오류 수정은 다른 방법과 비교하여 다음과 같은 이점을 갖는다. 첫째, 음성 인식기와 독립되어 빠르고 쉽게 구축되고, 구현이 용이하다. 도메인 지식에 이용되는 의미 카테고리 사전은 음성 인식 시스템에 독립되어 구축되며, 템플릿 데이터베이스는 대상 도메인으로부터 자동적으로 수집된다. 이전 기법들과 같이 많은 양의 인식 스크립트 없이도 구축이 가능하며, 비용 절감이 가능하다. 둘째, 화자 및 발화 환경에 독립적이다. 특정 화자나 잡음 환경에 종속되지 않는 도메인 지식을 이용한 언어 처리 기술을 통해 인식 결과에 포함된 오류를 수정한다. 셋째, 의미 추상화된 템플릿은 오류 패턴 데이터베이스의 크기를 줄이며, 의미 정보를 통해 의미적인 오류를 수정할 수 있다. 기존의 방법들은 단지 어휘 정보만을 이용하기 때문에 의미적인 오류를 해결할 수 없다. 특히 제안하는 의미 기반 오류 수정은 특정 도메인에서 음성 질의 처리를 수행하는데 도움을 준다.

## VI. 감사의 글

본 논문은 산업자원부 중기거점 기술개발사업의 지원으로 수행되었습니다.

## 참고문헌

[1] J. Barnett, S. Anderson, J. Broglio, M. Singh, R. Hudson, and S.W. Kuo, "Experiments in spoken queries for documents retrieval," in *proceedings of Eurospeech*, volume 3, pages 1323-1326, 1997.

[2] F. Crestani, "Word recognition errors and relevance feedback in spoken query processing," in *proceedings of the 2000 Flexible Query Answering Systems Conference*, pages 267-281, 2000

[3] Atsushi Fujii, Katunobu Itou, Tetsuya Ishikawa, "Speech-driven text retrieval: Using target IR collections for statistical language model adaptation in speech recognition," in *A. R. Cohen, E. W. Brown, and S. Srinivasan, editors, Information Retrieval Techniques for Speech Application (LNCS 2273)*, pages 94-104, 2002.

[4] Eric K. Ringger and James F. Allen, "A fertility model for post correction of continuous speech recognition," *ICSLP'96*, pp. 897-900, 1996.

[5] Satoshi Kaki, Eiichiro Sumita, and Hitoshi Iida, "A Method for Correcting Speech Recognition Using the Statistical features of Character Co-occurrence," *COLING-ACL'98*, pp.653-657, 1998.

[6] Geunbae Lee, Jungyun Seo, Seungwoo Lee, Hanmin Jung, Bong-Hyun Cho, Changki Lee, Byung-Kwan Kwak, Jeongwon Cha, Dongseok Kim, JooHui An, Harksoo Kim, and Kyungsun Kim, "SiteQ: Engineering High Performance QA System Using Lexico-Semantic Pattern Matching and Shallow NLP," in *proceedings of the 10th Text Retrieval Conference (TREC-10)*, Washington D.C., 2001.

[7] Robert A. Wagner and Michael J. Fischer, "The string-to-string correction problem," *Journal of the ACM*, 21(1):168-173, 1974.

[8] R. Rosenfeld, "The CMU Statistical Language Modeling Toolkit and its use in the 1994 ARPA CSR Evaluation," In *ARPA Spoken Language Technology Workshop*, 1995.

[9] Y. Al-Onaizan, J. Curin, M. Jahr, K. Knight, J. Lafferty, D. Melamed, F. J. Och, D. Purdy, N. A. Smith, and D. Yarowsky, "Statistical machine translation, Technical report," *John Hopkins University Summer Workshop*, 1999.