

음성 자연어 처리를 위한 대화 관리 시스템[†]

포항공과대학교 정민우 · 은지현 · 이청재 · 정상근 · 이근배

1. 개요

음성 언어 인터페이스(Spoken Language Interface)는 기존의 제한된 인터페이스(예 Keyboard or Graphic User Interface)를 뛰어넘어 보다 사람에게 친숙하고 자연스러우며, 공간의 제약을 받지 않는 유동적인 방법으로 컴퓨터와 정보를 주고받기 위한 대체 인터페이스이다. 음성 언어 인터페이스는 텔레매틱스, 홈네트워킹, 모바일 서비스, 로봇틱스 등과 같은 분야에 적용될 수 있으며(그림 1), 21세기 유비쿼터스 컴퓨팅 시대를 위한 필수적인 지능형 인터페이스로 활발히 연구되고 있다[1]. 음성 대화 관리 시스템(Spoken Dialogue Management System)은 보다 효과적인 음성 인터페이스를 제공하기 위해 사람과 컴퓨터 사이에서 매개체 역할을 하게 된다. 예를 들어 디지털 TV와의 정보를 교환하기 위해 그림 2와 같은 시나리오를 생각해 볼 수 있다.

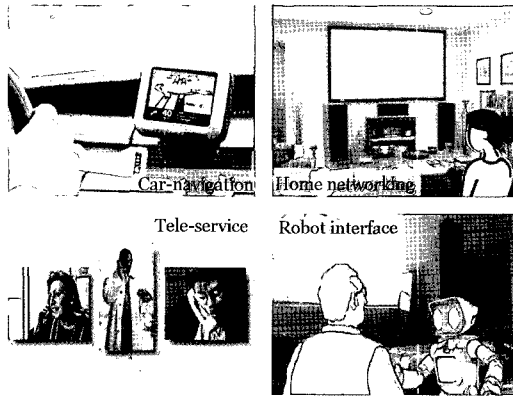


그림 1 음성 대화관리 시스템의 응용분야 예

음성 대화 관리 시스템은 일반적으로 음성 인식(Speech Recognition) 및 언어 모델 적응(Language Model Adaptation), 음성 언어 이해(Spoken Language Understanding), 대화 관리(Dialogue Management) 및 확인 대화(Clarification and

Verification), 그리고 음성 합성(Text-To-Speech Synthesis)과 같은 일련의 과정을 포함하는 프레임워크를 제공한다(그림 3).

System	무엇을 도와드릴까요?
User	수능 영어 강의 찾아줘.
System	총 2개의 결과를 찾았습니다. 9번, 수능 영문법 고급 8시. 2번, 핵심마무리 문법 10시입니다.[목록Display] 선택하세요.
User	2번 채널로 돌려줘. 그리고 내일 서울 지역 날씨 좀 알아 봐 줄래?
System	채널 전환합니다.[2번 채널로 전환] 내일 서울 지역 날씨는 대체로 맑음 입니다. 더 자세한 날씨 정보를 원하십니까?
User	그래, 자세히 알려줘

그림 2 음성 대화 인터페이스의 예

본 고에서는 음성 자연어 처리를 위한 대화 관리 시스템을 구성하는 오류 수정 언어 모델 적응, 강건한 음성 언어 이해, 상황 기반 대화 관리, 그리고 3단계 검증 및 확인 대화의 4가지 기술을 소개한다.

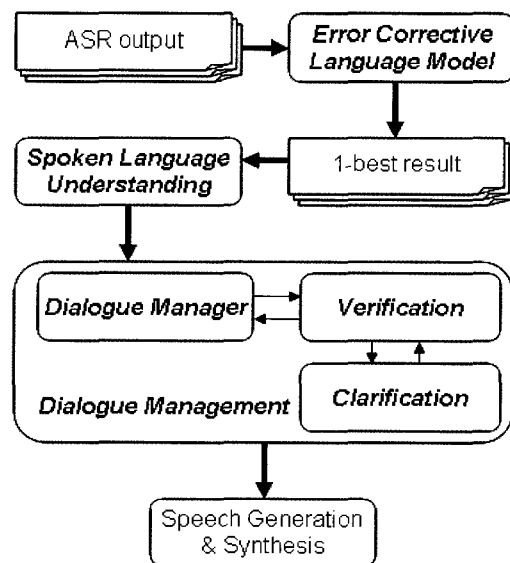


그림 3 전체 대화 관리 시스템 구조

[†] 이 연구는 산업자원부 지원으로 수행하는 뇌신경 정보학 산업과 21세기 프론티어 연구 개발사업(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

2. 언어 모델 적응

2.1 음성 인식을 위한 언어 모델 적응

음성 언어 인터페이스는 낮은 음성 인식률로 인해 실제 환경에서 활용되기 어렵다. 이러한 인식률 저하는 인식 데이터가 학습된 데이터와 다를 때 특히 두드러지며, 이를 해결하기 위해서는 음성 인식기를 인식 도메인에 적응하는 과정이 필요하게 된다. 연속 음성 인식은 통계적인 결정 과정으로 해석할 수 있으며, 필요한 모델은 크게 음향 모델과 언어 모델로 나눌 수 있다. 인식 과정은 주어진 음성 신호(a)에 대해, 확률 $Pr(w|a)$ 을 최대화하여 최적의 단어열(w)을 찾는 과정으로 볼 수 있다.

이때 언어 모델은 학습 환경과 인식 환경이 같다면 측정된 파라미터가 좋은 성능을 보장하지만 인식 환경이 달라져, 어휘, 구문, 토픽 등의 언어 현상이 변화되면, 기존에 학습된 언어 모델은 더 이상 좋은 모델을 제시한다고 할 수 없다. 이러한 문제를 해결하기 위해 다양한 언어 모델 적응 기법이 연구되어 왔다. 기존의 언어 모델 적응 기법들은 관측된 데이터를 적응시키기 위해 단순한 후보 재평가(n -best re-scoring) 기법에 의존하고 있다. 하지만 인식기의 후보 결과에 반드시 옳은 가설이 존재하리라는 보장이 없기 때문에 이를 위해 오류 수정 언어 모델 적응 기법이 제안되었다 [6].

2.2 관련 연구

음성 자연어 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. 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 후 처리에 관한 여러 가지 연구가 수행되어 왔다. 대표적인 연구에는 통계적으로 오류를 수정하는 잡음 채널 모델에 기반을 둔 음성 인식 후 처리 기법이 있다[17]. 이 방법은 인식기가 사용되는 발화 환경, 화자, 잡음 등의 특성을 잡음 채널 모델을 통해 모델링하고, 이를 이용하여 인식

오류 단어를 통계적으로 수정한다. 또한 많은 양의 인식 스크립트를 통해, 오류 단어의 앞뒤 문맥 정보를 이용하거나 휴리스틱을 이용하여 오류 패턴을 수집, 이를 이용하는 방법이 있다[9]. 이러한 방법은 음성 인식 오류의 유형이 특정한 패턴을 가지고 나타난다는 가정에서 출발하였으나, 환경이나 화자가 달라지면 패턴이 달라진다는 문제점을 가지게 되며, 올바른 단어 또한 오류로 인식하게 되는 문제점을 갖는다. 이러한 대부분의 이전 연구들은 어휘 정보에 바탕을 둔 오류 수정 방법을 이용하며, 화자나 발화 환경에 종속적이다.

2.3 오류 수정 언어 모델 적응

앞서 소개한 연구들의 단점을 극복하기 위해 오류 수정 기법과 언어 모델 적응 기법을 융합한 새로운 적응 기법을 소개한다[6]. 오류 수정 언어 모델(Error Corrective Language Model) 적응 기법은 그림 4와 같이 요약될 수 있다. 음성 인식기의 언어 모델은 대량의 배경 데이터로부터 학습되며, 적응 데이터는 채널 적응을 위한 음성 인식 결과와 언어 자질을 포함한 텍스트로 나뉜다. 여기서 모델 적응 문제는 입력 단어열에서 확률 $Pr(w'|w)$ 을 최대화 시키는 최적의 수정된 단어열 w' 를 찾는 과정으로 볼 수 있다.

이때 이 모델은 잡음 채널 모델을 따르게 되며, 채널 모델과 언어 모델로 구성된다. 이는 최대 후험 확률(Maximum A Posteriori) 적응 혹은 오류 최소화를 위한 최적화 모델로 해석될 수 있다. 채널 모델은 인식 환경의 채널 특성을 반영하며 이를 통해 언어 모델이 도메인, 잡음 환경, 또는 화자에 적응이 된다고 볼 수 있다. 또한 단순 매핑 모델의 한계를 극복하기 위해서 통계적 기계 번역의 개념을 빌려 다대다 매핑 모델을 도입하여 모델의 일반성을 높였다.

또한 이때 다양한 언어 정보를 이용하기 위해서 재평가(RERANKER) 모듈에서는 의미 기반의 언어 모델을 이용하여 n -개의 후보들로부터 최적의 문장을 찾아내게 된다. 이러한 두 단계의 과정을 거쳐 적응된 모

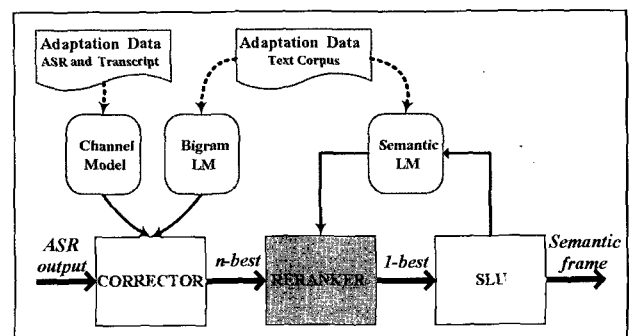


그림 4 오류 수정 언어 모델 적응 구조

델은 음성 언어 이해와 대화 관리 시스템에게 좀더 정확한 입력을 제공하게 된다[6].

3. 음성 언어 이해

3.1 음성 언어 이해의 문제점

음성 언어 이해를 위한 연구는 1980년대 후반부터 광범위하게 진행되어 왔다. 음성 언어 이해는 음성 인식과 자연 언어 이해라는 주요 기술의 결합을 필요로 한다. 그러나 음성 인식과 자연 언어 이해의 결합이 단순한 문제는 아니다. 문어체 언어와 달리, 음성 언어는 사람들이 다른 사람과 대화할 때 깊이 생각하지 않고 즉석으로 뭔가를 말하기 때문에 문법적으로 불완전한 문장 구조를 가지고 있기 때문이다. 따라서 음성 문장의 구문론적, 의미론적 가능성을 설명하기 위해서 특정한 형태의 규칙을 적용하게 되면, 사람들이 실제로 사용하는 온전하지 않은 형태의 음성 문장이 주어졌을 때 혼란을 야기하게 된다. 게다가 자연 언어가 가지고 있는 다양한 모호성, 문법상의 오류, 잡음이 심한 발화 환경 등을 고려하면 이러한 규칙 기반의 접근 방식은 더욱 복잡한 문제가 된다. 또한 음성 인식기의 결과는 연속 음성을 인식할 때 휴지, 재시작, 반복, 수정 발화 뿐 아니라 잘못된 인식 결과도 포함하고 있다.

3.2 관련 연구

이러한 음성 언어 이해를 다루기 위한 연구가 시작된 이후로, 규칙 기반의 방법론과 통계적 방법론이 제안되었다. 이들 방법은 성능과 효율성의 측면에서 일장 일단이 있지만, 현재는 성능과 효율성 모두를 얻을 수 있는 통계적 방법론 방향으로 연구가 진행되고 있다.

앞서 언급한 대로 특정한 형태의 규칙을 적용해서 음성 언어를 이해하는 규칙 기반의 방법은 다양한 형태의 음성 문장이 주어졌을 때 이해도가 떨어진다는 단점이 있다. 이러한 단점이 나타나는 이유는 대부분 특정 영역에서 이루어지는 음성 문장의 구문론적인 분석이 대부분 완전하지 않고, 많은 오류와 모호성을 포함하고 있기 때문이다. 또한 이러한 방법론을 사용하는 시스템은 일반적으로 손수 고안된 의미 단계의 문법과 이를 다루기 위한 강인한 의미 분석기를 사용하고 있다. MIT의 TINA[18], CMU의 PHOENIX[20] 같은 시스템이 강인한 의미 분석기를 채택한 예이다. 이 시스템들은 성능 면에서 뛰어난 이해도를 보이고 있지만, 다양한 영역에 적용될 경우 각 영역에 따라 개별적인 규칙들이 요구된다. 이를 구축하기 위해서는 전혀 다른 규칙 제작 과정을 거쳐야 하기 때문에 규칙을 사

용하는 것은 개발 비용 측면에서 상당한 단점으로 작용한다. 또한 시스템이 예상하는 사람들의 문법 환경과 이용자들이 실제 사용하는 문법 환경이 다를 경우 시스템의 예상 규칙이 제대로 동작하지 않을 수도 있다.

이와 다른 접근 방식으로, 음성 문장을 의도된 의미 구조에 직접적으로 맞추기 위한 통계적 방법론이 있다. 규칙 기반의 방법론에서 사용된 특정 규칙은, 훈련 데이터로부터 자동으로 학습된 통계적인 모델로 대체된다. 통계적 방법은 적절한 훈련 데이터가 주어지기만 하면 새로운 조건(새로운 작업 또는 새로운 언어)에서도 적절하게 적용될 수 있다는 점에서 매우 흥미롭다. 이 방법은 언어 이해 정확도를 저하시키지 않으면서도 다양한 도메인간의 이식성이 뛰어나고 구축 비용 면에서도 앞의 방법보다 효율적이다. 통계적 모델 기반의 음성 언어 이해 시스템에는 AT&T의 CHRONUS [13], BBN의 Hidden Understanding Model[16], Cambridge 대학의 Hidden Vector State Model [5] 등이 있다. 이러한 시스템들은 주로, 기존 음성 인식과 자연 언어 처리 분야에서 성공적으로 적용되었던 은닉 마르코프 모델(Hidden Markov Model)에서 영감을 얻어 개발되어 왔다.

3.3 강건한 음성 언어 이해(Robust Spoken Language Understanding Model)

음성 언어 이해의 목적은 음성 인식의 결과를 분석하고, 이를 대화 처리 모듈이 처리할 수 있는 형태의 의미 구조로 할당하는 역할을 수행하는 것이다. 때문에 음성 인식의 한계로 인한 여러 가지 제약 조건을 극복하기 위해서, 미리 정의된 의미 구조를 바탕으로 필수적인 요소만을 추출하는 것을 목적으로 하는 개념 집어내기(Concept Spotting) 방식 혹은 시퀀스 레이블링(Sequential Labeling; CRF)[10] 의 언어 이해가 더욱 적합한 방식이다[3]. 이러한 방식은 미리 정의된 개념 구조인 슬롯(slot)에만 관심을 가지기 때문에, 언어를 완전히 이해하기보다 부분적으로 이해하는 방식(Partial Understanding)을 취한다. 이 방식은 부분적인 언어 이해 방식이지만 특정 영역의 언어 이해를 위해 적절하게 설계된 슬롯이 있기 때문에 각 슬롯의 값으로부터 언어 이해에 필수적인 정보를 얻을 수 있다는 장점이 있다. 또한 입력 문장을 의미 구조에 할당하는 것을 분류(Classification) 문제로 다루고 있는데, 단순 분류기가 아닌 복수 분류기를 이용하여 이를 바탕으로 투표(Voting)하여 최종 할당이 이루어지는 방식을 취하고 있다[3]. 단순 분류기를 기반으로 하는 것과 달리 복수 분류기를 바탕으로 최종 선택을 할 경

우 분류기의 평균적인 선택을 고려할 수 있기에 보다 강건한 선택을 기대할 수 있다는 장점이 있다. 그림 5는 이러한 음성 언어 이해 모델을 도식화한 것이다.

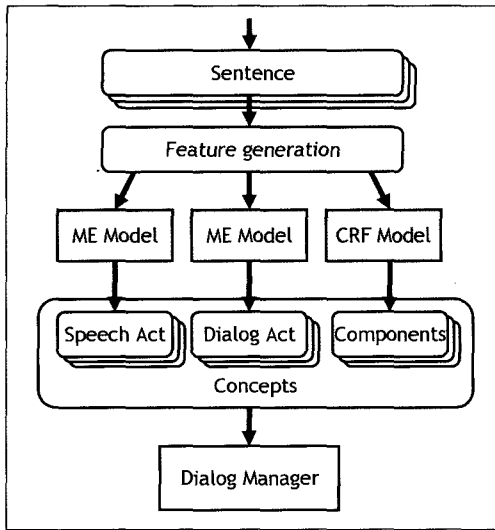


그림 5 강건한 음성 언어 이해 모델 구조

4. 대화 관리

4.1 대화 관리 시스템(Dialogue Management System)

대화 관리는 음성 자연어 처리를 위한 대화 시스템에서 중추적인 역할을 하는 부분으로 사용자의 음성을 받아 의미를 추출하고 사용자에게 필요한 정보를 제공하기 위해 외부 지식 자원들을 연결하여 시스템 발화를 생성하는 전반적인 대화 흐름을 제어하는 부분이다. 이번 장에서는 현재 사용자의 발화를 바탕으로 적절한 시스템 발화를 생성하는 대화 관리 시스템을 구축하기 위한 기존의 방법과 현재의 방법을 비교 제시 한다.

4.2 관련 연구

기존의 대화 관리 시스템의 대부분은 유한 상태 기반 방식(Finite State-Based Model)을 바탕으로 개발되었다[14]. 현재 대화 상태를 특정 상태로 사상하여 그 상태에서 적절한 대화 진행을 하고 다음 상태를 예측하는 방식이다. 이것은 실제로 고정된 상태 전이 모델 위에서 대화를 진행하기 때문에 특정 도메인의 목적에 맞는 상태 전이 네트워크를 디자인하여 빠른 대화 관리 시스템을 구축할 수 있다. 이 모델은 전형적으로 대화 형태가 고정되어 있는 간단한 대화 시스템에 많이 사용된다. 그러나 사용자 주도의 대화가 주로 이루어져서 사용자의 대화 자유도가 낮다는 단점이 있다. 또한 대화 문법에 의해 다음에 할 수 있는 사

용자의 발화가 제한이 되거나 도메인이 바뀌면 모델 자체를 바꿔야 하기 때문에 도메인 간의 확장성이 떨어지는 단점들을 가지고 있다.

좀 더 자연스러운 대화 처리를 위해 고안된 방식이 프레임 기반 대화 모델(Frame-Based Dialogue Model)이다[12]. 이 방식은 사용자의 발화에서 필요한 정보를 추출하여 프레임을 채워 넣어 그 정보를 바탕으로 대화를 진행하는 방식이다. 발화에서 우리가 필요한 것은 발화의 의도와 사용자에게서 제공되는 새로운 정보의 값들이다. 이러한 값들을 사용자 발화에서 의미 파서나 기계 학습 등을 이용하여 추출하여 대화 관리에 이용하는 것이다. 그러나 실제 대화 관리는 이러한 프레임을 기반으로 한 상태 전이나 수작업으로 만든 규칙에 의존하는 경우가 많아 유한 상태 모델과 같은 단점을 가지게 된다. 이러한 방법은 사용자 주도 대화를 통한 데이터베이스 검색이나 여행 예약이나 안내 시스템 등에 많이 쓰인다.

또한 최근에는 좀 더 복잡한 도메인을 처리하기 위해 계획 기반 모델(Plan-Based Model)이 연구되고 있다[2]. 이것은 사용자와 컴퓨터 간의 대화 목적을 정하고 서로 주도권을 주고 받으면서 그것을 달성하기 위한 가장 최적화된 대화 계획을 적용하여 대화를 진행하는 방식이다. 그러나 이것은 자유도가 높은 반면에 처리하고자 하는 도메인의 복잡성이 높기 때문에 대화 관리 성능이 떨어져서 상용성이 떨어진다.

4.3 상황 기반 대화 관리(Situation-Based Dialogue Management)

기존의 대화 관리 시스템들이 갖는 단점을 극복하고, 좀더 유연하고 강건한 음성 대화 관리를 위해 POSSDM(POSTECH Situation-Based Dialogue Manager)이라는 대화 관리 시스템을 구현/개발 하였다[11]. POSSDM에서 제안하는 방법은 기존의 연구 방식들의 장점을 취하고 단점을 해소하기 위해 대화 예제를 이용한 상황 기반 대화 관리 기술이다. 상황 기반 대화 관리 시스템의 기본적인 아이디어는 프레임 기반의 대화 모델을 취하되, 좀 더 유연한 대화를 진행할 수 있도록 하며 규칙을 대화 예제를 통하여 자동으로 학습을 하여 도메인 확장성을 증대시키는 것이다.

상황 기반 대화 관리라는 것은 현재의 대화 상황에 따라서 특정 상태 전이 규칙에 얽매이지 않고 어느 상태로든지 전이될 수 있다는 것을 의미한다. 여기서 대화 상황이라는 것은 사용자의 의도, 현재 프레임의 내용, 그리고 슬롯이 채워진 정도에 대한 담화 기록 등을 이용하여 정의한다. 이것은 특정한 문법이나 상태 전이

없이 좀 더 자유로운 사용자의 발화를 관리할 수 있으며 도메인에 대한 특성화된 방법이 없어서 새로운 도메인에 적용하기가 쉬운 장점이 있다. 또한, 시스템을 객체 기반 방식으로 구현하여 도메인 의존적인 모듈과 비의존적인 모듈을 나누어서 일반적인 클래스에서 상속을 받아서 도메인 의존적인 모듈을 쉽게 구현할 수 있도록 디자인 하였다[11].

4.4 대화 예제 기반 대화 모델링(Example-Based Dialogue Modeling)

상황 기반 대화 관리에 적합한 대화 모델링을 하기 위한 방법론으로 대화 예제를 이용한 대화 모델링 방법을 이용한다. 대화 예제를 이용한 대화 모델링은 규칙 기반 모델에서 필요한 규칙 코딩 작업을 최소화 하기 위해 대화 말뭉치로부터 현재 대화 상황에 맞는 대화 예제를 찾아서 적절한 시스템 발화를 할 수 있도록 한다. 그러므로 대화 말뭉치로부터 대화 예제 데이터베이스를 만들고 그것을 검색하여 시스템 응답을 찾는 방식이다. 대화 예제 기반 모델링은 규칙 기반 대화 모델링의 단점인 규칙 학습의 인력 비용을 간소화하기 위해 대화 말뭉치로부터 자동적으로 규칙을 학습하여 시스템 발화를 생성한다. 이는 대화 관리 시스템을 구축하는데 좀 더 효율적이고 쉬운 도메인 확장이 가능한 장점이 있다.

대부분 대화의 상황을 대화 예제를 통해 시스템 발화를 결정할 수 있지만, 대화 예제가 없는 경우 등을 처리하기 위해 미리 만들어 놓은 메타 규칙도 이용하였다. 그림 6은 대화 예제 기반의 대화 모델링의 전반적인 전략을 도식적으로 보여준다.

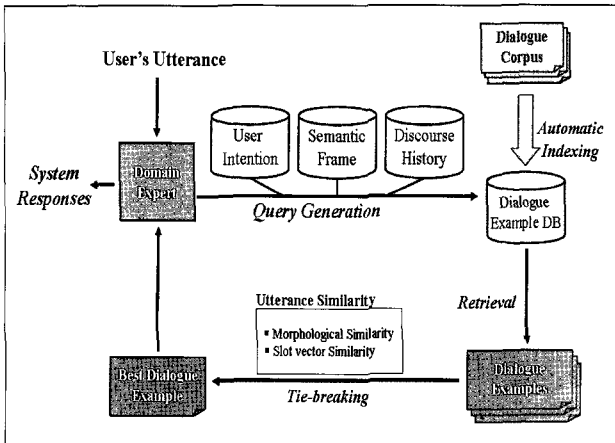


그림 6 대화 예제 기반 대화 모델링 전략

4.5 색인과 질의

대화 모델링을 위해서는 우선 대화 말뭉치로부터 자동으로 대화 예제 데이터베이스(Dialogue Example

Database)를 만드는 과정이 요구된다. POSSDM에서는 말뭉치의 태깅을 최소화하기 위해 음성 언어 이해 모델을 훈련하기 위한 말뭉치를 재사용 했으며 여기에 담화 기록 벡터(Discourse History Vector)와 시스템 응답(System Response)만을 추가적으로 부착하였다. 여기서 담화 기록 벡터는 프레임의 슬롯이 채워졌는지 아니지를 나타내는 이진 벡터로 표시한다. 이러한 부착 정보들은 대화 예제 데이터베이스의 색인 키로써 이용되고 이것을 가지고 대화 예제 데이터베이스에 질의를 던져 적절한 시스템 응답을 찾아낸다. 대화 예제를 찾기 위한 질의에서 현재 대화 상황을 바탕으로 검색을 하여 찾을 때 항상 전체 일치(Exact Match)가 가능하지 않을 수 있다. 그러므로 부분 일치(Partial Match)를 위해서 제약을 완화하여 사용자의 의도(Dialog Act and Main Action)만을 가지고 대화 예제를 검색한다. 이것은 실제 시스템 응답은 현재 상황의 사용자 의도에 주로 의존하기 때문이다.

4.6 발화 유사성(Utterance Similarity)

대화 예제 데이터베이스에서 현재 상황에 적합한 대화 예제들을 가져올 때, 여러 개의 예제 후보들이 생긴다. 이러한 것들 중에서 가장 적합한 예제를 찾기 위해서 발화 유사성(Utterance Similarity)을 정의한다. 발화 유사성은 담화 기록 유사성(Discourse History Similarity)과 어휘-의미 유사성(Lexico-Semantic Similarity)의 선형 보간(Linear Interpolation)으로 정의한다. 담화 기록 유사성은 담화 기록 벡터의 코사인 계수(Cosine Measure)로 나타내며, 어휘-의미 유사성은 추출된 슬롯의 값을 미리 정의된 슬롯 이름으로 대체하여 대화 예제의 사용자 발화와 현재 사용자 발화의 유사성을 측정하였다. 그림 7은 어휘-의미 유사성의 예제이다.

User Utterance	그럼 SBS 드라마는 언제 하지? Then, when do the SBS dramas start?
Component Slots	[channel = SBS, gene = 드라마]
Lexico-Semantic Input	그럼[channel][gene]는 언제 하지

그림 7 어휘-의미 유사성 예제

4.7 시스템 발화 생성

시스템 발화를 생성하기 위해서는 대화 관리 시스템에서 결정된 추상적인 시스템 발화 개념을 자연어 생성 모듈을 이용한다. 일반적으로 규칙 기반이나 템플릿 기반의 방식으로 언어를 생성한다. 그러나 이것은 부자연스러운 발화를 생성하고 수작업이 요구되기 때문에 최근에는 기계 학습과 웹 마이닝 기술을 이용한 다양

한 방식으로 접근하고 있다.

5. 확인 대화 및 검증

5.1 확인 대화(Clarification Dialogue)

확인 대화(Clarification Dialogue)란 사람과 사람 혹은 사람과 컴퓨터 사이에서의 대화에서 전달된 정보가 적절치 못할 경우, 즉 불분명 혹은 불충분 하거나 혼동의 여지가 많은 정보를 대화를 통해서 재확인 해 나가는 대화 유형을 말한다. 다음은 사람들간의 대화에서 확인 대화를 하는 예이다.

철수 : 영화야 ~~ 먹었어? (~~ 는 주변의 소음으로 전달안된 단어) 영희 : 뭐 먹었냐? ('~~'에 대한 재확인을 시도함)

확인 대화는 정보를 재확인하는 모든 대화에 해당하지만, 그 종류를 대화의 흐름에 따라 다음과 같이 4가지로 분류할 수 있다.

a. 담화 수준 확인 대화 담화수준에서의 확인 대화가 필요한 경우를 말한다. 즉 바로 전의 대화에서는 '주식' 관련 대화를 진행하고 있는데, 다음 대화 턴에 입력된 정보가 '스포츠'와 관련된 경우 이를 재확인 할 필요가 있다.
b. 관계 수준 확인 대화 입력된 정보의 관계수준에서의 확인 대화가 필요한 경우를 말한다. 예를 들어 "최홍만이 나오는 농구 언제 해?" 라는 요청이 온 경우 두 가지 정보 즉 '최홍만'과 '농구'는 관계가 적거나 없는 경우이므로 이 경우에는 두 가지 정보 중 한가지를 재확인해야만 한다.
c. 이해 수준 확인 대화 입력된 문장에 대해서 제대로 이해를 못하여 발생하는 잘못된 정보를 재확인하는 경우가 이에 해당한다. 예를 들어 "KBS 해신 어디서 해?"를 제대로 이해하려면, 채널은 'KBS', 프로그램 명은 '해신'으로 이해해야 하는데, 이 출연자 이름은 '해신'으로 이해하는 경우, 대화진행에 문제가 생기게 된다.
d. 채널 수준 확인 대화 대화의 전달이 이루어지는 채널의 잡음 때문에 발생하는 잘못된 정보를 재확인하는 과정이다. 이 경우는 확인 절차가 음성 채널 수준에서 이루어지므로 음성 인식 수준의 확인 대화라고 생각할 수 있다.(예: 해신 보자 → 배신 보자)

확인 대화는 그 수준과 목적에 따라 다양하게 분류될 수 있겠지만, 어떤 확인 대화이든지 반드시 다음의 "불분명한 정보의 파악", "확인 대화 전략의 수립과 대화 수행"이라는 두 가지 사항을 고려해야만 한다. 즉

무엇이 잘못된 정보인지 혹은 불충분한 정보인지 파악하고 그 정보를 어떤 식으로 효과적으로 재확인하는가가 확인 대화의 주요 과정이다.

5.2 관련 연구

다양한 레벨에서의 확인 대화 과정이 가능하지만, 최근에 가장 많이 이루어지는 확인 대화에 관한 연구는 음성 환경에서의 부정확한 음성 인식 결과로 인한 불분명한 정보의 전달, 그리고 그로 인한 시스템의 오동작을 사전에 감지하고 대화를 통해 불분명한 정보를 재확인해 나가는 연구에 집중되어 있다. 즉 앞에서 제시한 분류에서 이해 수준/채널 수준의 확인 대화에 집중되어 있다[4, 7].

이러한 음성 환경에서의 확인 대화에 초점을 맞추어 위의 두 가지 태스크를 다시 생각해 보면, 음성 인식 및 음성 언어 이해 과정에서 생기는 오류 정보 검출(정보 확인)과 오류 정보를 효과적으로 고치는 대화 전략 "확인 대화 전략 수립"의 두 가지로 생각할 수 있다.

정보 확인에 관련된 연구는 대부분 단어 오류 신뢰도(word error confidence)에 집중되어 있다. 음성 인식의 단위인 '단어' 레벨에서의 음성이 얼마나 잘 인식되었나를 살펴 보며, 그것을 효과적으로 분석하기 위한 자질들 역시 많이 연구되었다.

확인 대화 전략 수립에 관한 연구는 음성인식의 오류를 재확인 하는 수준 정도에만 이루어져 있다. 즉 정보 확인 단계에서 음성 인식의 신뢰도 수준을 결정해 주면 그 신뢰도 점수를 이용해 담화 상태 전이 모델에 적용하거나, 발화 자체를 다시 요청, 담화기록을 저장할 때 특정정보의 신뢰도 점수에 반영하는 방법 등으로 확인 대화를 구현하는 연구가 이루어져 왔다[15, 19].

5.3 3단계 정보 확인 및 대화 전문가(3-Phase Verification and Clarification Expert)

확인 대화를 좀 더 정확하고 체계적으로 풀기 위해 3단계 정보확인 과 확인 대화 전문가(Clarification Expert)를 제안한다[8]. 기존의 연구에서는 단순히 음성 인식 측면에서만 그 음성 인식 단위나 문장 전체에 대한 오류 검출이 이루어졌다. 이에 비해 3단계 정보 확인 방법은 오류 검출 수준이 음성 인식 단위 수준 검출, 정보 보존율에 의한 문장 수준 검출, 3) 대화 정보 단위 수준의 검출 등의 3단계로 나누어져 체계적으로 이루어진다. 즉, 우선 음성 인식 단위 수준에서의 오류 검출이 이루어지고 이러한 정보가 다음 단계인 문장 수준 검출로 건너지게 된다. 이때, 정보 보존율에 의거하여 음성인식 수준과 음성 언어 이해 수

준 모두에 걸쳐 얼마큼 그 문장이 믿을 만한지를 판단하고, 그러한 정보 보존율이 다음 단계인 대화 정보 단위로 건너지게 되어 대화 진행에 필요한 정보 단위 수준에서 그 정보가 음성 인식 측면에서나 이해 측면, 그리고 관계 측면에서 얼마나 잘 전달되었는지를 검증하게 된다.

기존의 연구에는 확인 대화를 위한 전략 수립을 위한 방법론이 부족하였다. 이에 확인 대화도 하나의 대화 패턴으로 기존의 전문가 방식의 대화 모델처럼 하나의 확인 대화 전문가로 구성하였다[8, 11]. 확인 대화 전문가의 장점은 대화 관리 시스템의 전반적인 시스템 구조상에서 확인 대화 전문가가 자연스럽게 하나의 모듈로 자리잡을 수 있으며, 확인 대화에 필요한 정보를 기존의 각 전문가와 공유하게 함으로서 모든 정보를 이용하여 다양한 수준에서 확인 대화 전략을 수립할 수 있게 한다는 것이다. 그림 8은 이러한 3단계 정보 확인 및 대화 전문가 방법의 흐름을 표현한다.

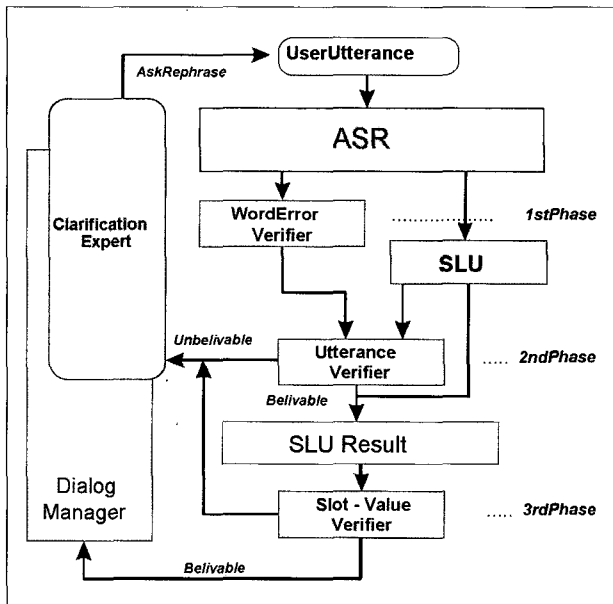


그림 8 3단계 정보 확인 및 대화 확인 전문가

6. 결 론

본 고에서는 차세대 음성 인터페이스를 제공하고 음성 자연어 처리를 위한 대화 관리 시스템을 소개하였다. 소개된 시스템은 오류 수정 언어 모델 적용, 강건한 음성 언어 이해, 상황 기반 대화 관리, 그리고 3단계 검증 및 확인 대화의 4가지 기술이 융합되어 보다 강건하고 음성 자연어 처리에 적합한 프레임워크를 제공한다. 앞으로는 이러한 각각의 기술을 발전시켜 나감과 동시에, 대화 관리 시스템 안에서 시너지 효과를 낼 수 있는 연계성을 찾는 연구가 진행 되어야 할 것이다.

또한 다양한 응용 분야에 실용적이고 효과적으로 음성 대화 관리 시스템이 적용될 수 있는지를 검증해 나갈 계획이다.

참고문헌

- [1] J. Allen, D. Byron, M. Dzikovska, G. Ferguson, L. Galescu, and A. Stent, Toward conversational human-computer interaction, *AI Magazine*, vol. 22, no. 4, pp 27-37, 2001.
- [2] J. Chu-Carroll, MIMIC: An Adaptive Mixed Initiative Spoken Dialogue System for Information Queries, *Applied Natural Language Processing Conference*, pp. 97-104, 2000.
- [3] J. Eun, M. Jeong, G. G. Lee, A Multiple Classifier-based Concept-Spotting Approach for Robust Spoken Language Understanding, *Interspeech 2005-Eurospeech*, 2005.
- [4] T.J. Hazen, J. Polifroni, and S. Seneff, Recognition confidence scoring and its use in speech language understanding systems, *Computer Speech and Language*, vol. 16, no. 1, pp. 49 - 67, 2002.
- [5] Y. He, and S. Young, Semantic Processing using the Hidden Vector State Model, *Computer Speech and Language*, Vol. 19, No. 1, pp. 85-106, 2005.
- [6] M. Jeong, J. Eun, S. Jung, and G. G. Lee, An error-corrective language-model adaptation for automatic speech recognition, *Interspeech2005-Eurospeech*, 2005.
- [7] H. Jiang, Confidence measures for speech recognition, *Speech Communication*, vol. 45, no. 4, pp. 455 - 470, 2005.
- [8] S. Jung, C. Lee, G. G. Lee. Three phase verification for spoken dialog clarification. To appear in *IUI 2006 Proc.*, 2006.
- [9] S. Kaki, E. Sumita, and H. Iida, A Method for Correcting Speech Recognition Using the Statistical features of Character Co-occurrence, *COLING-ACL'98*, pp.653-657, 1998.
- [10] J. Lafferty, A. McCallum, and F. Pereira, *Conditional Random Fields: Probabilistic*

Models for Segmenting and Labeling Sequence Data, ICML, 2001.

- [11] C. Lee, S. Jung, J. Eun, M. Jeong, G. G. Lee. Example and situation based dialog management for spoken dialog system. IEEE Workshop on ASRU, 2005.(Demo presentation)
- [12] O. Lemon, A. Gruenstein, A. Battle, and S. Peters, Multi-tasking and Collaborative Activities in Dialogue Systems, SIGDIAL Workshop, pp. 113-124, 2002.
- [13] E. Levin, and R. Pieraccini, CHRONUS, the next generation, The DARPA Speech and Natural Language Workshop, 1995, pp. 269-271.
- [14] M. McTear, Modelling Spoken Dialogues with State Transition Diagrams: Experiences with the CSLU Toolkit, ICSLP, vol. 4, pp. 1223-1226, 1998.
- [15] M. McTear, I. O'Neill, P. Hanna, and X. Liu, Handling errors and determining confirmation strategies - an object-based approach, Speech Communication, vol. 45, no. 3, pp. 249 - 269, 2005.
- [16] S. Miller, M. Bates, R. Bobrow, R. Ingria, J. Makhoul, and R. Schwartz, Recent progress in hidden understanding models, The DARPA Speech and Natural Language Workshop, 1995, pp. 276-280.
- [17] E. K. Ringger and J. F. Allen, A fertility model for post correction of continuous speech recognition, ICSLP'96, pp. 897-900, 1996.
- [18] S. Seneff, TINA: A Natural Language System for Spoken Language Applications, Computational Linguistics, 1992, 18.1, pp. 61-86.
- [19] F. Torres, L.F. Hurtado, F. Garcia, E. Sanchis, and E. Segarra, Error handling in a stochastic dialog system through confidence measures, Speech Communication, vol. 45, no. 3, pp. 211 - 229, 2005
- [20] W. Ward, and B. Pellom, The CU Communicator System, IEEE Workshop on ASRU, Keystone, Colorado. 1999.

정민우



1999. 3~2003. 2 전북대학교 컴퓨터공학과 (학사)
 2003. 3~2005. 2 포항공과대학교 컴퓨터공학과(석사)
 2005. 3~현재 포항공과대학교 컴퓨터공학과 박사과정
 관심분야: 언어 모델, 언어 이해
 E-mail : stardust@postech.ac.kr

은지현



1998. 3~2002. 2 포항공과대학교 컴퓨터공학과(학사)
 2004. 3~2006. 2 포항공과대학교 컴퓨터공학과(석사)
 관심분야: 언어 이해, 구문 분석
 E-mail : tiger@postech.ac.kr

이청재



2000. 3~2004. 8 포항공과대학교 컴퓨터공학과(학사)
 2004. 9~현재 포항공과대학교 컴퓨터공학과 박사과정
 관심분야: 대화 모델
 E-mail : lcj80@postech.ac.kr

정상근



1999. 3~2004. 2 포항공과대학교 컴퓨터공학과(학사)
 2004. 3~2006. 2 포항공과대학교 컴퓨터공학과(석사)
 2006. 2~현재 포항공과대학교 컴퓨터공학과 박사과정
 관심분야: 대화 모델, 확인 대화
 E-mail : hugman@postech.ac.kr

이근배



1984 서울대학교 컴퓨터공학과(학사)
 1986 서울대학교 컴퓨터공학과(석사)
 1991 UCLA 컴퓨터학과(박사)
 1991. 3~1991. 9 UCLA 연구원
 1991~1996 포항공과대학교 조교수
 1997~2003 포항공과대학교 부교수
 2000~2001 미국 Stanford CSLI 연구원
 2004~현재 포항공과대학교 정교수
 관심분야: 자연언어 처리, 음성인식, 정보 검색, 바이오 인포매틱스
 E-mail : gblee@postech.ac.kr