음성기반의 확인대화를 위한 3 단계 검증 방법

정상근 이청재 이근배
포항공과 대학교 컴퓨터 공학과
{ hugman, lcj80, gblee }@postech.ac.kr

THREE PHASE VERIFICATION FOR SPOKEN DIALOG CLARIFICATION

Sangkeun Jung Cheongjae Lee Gary Geunbae Lee
Department of Computer Science and Enginerring
Pohang Universitiy of Science and Technology

요 약

음성 기반의 대화 처리는 음성인식 오류, 음성언어이해 오류, 대화 처리 오류 등 많은 오류를 수반하며 진행 된다. 이러한 오류들로 인해 시스템이 인지하는 사용자의 의도는 최초의 사용자 의도와는 그 차이가 벌어지게 된다. 사람과 사람사이의 대화에서도 역시 이러한 의도의 차이는 발생하게 되며, 사람은 이러한 의도의 차이를 확인대화를 통해서 좁혀가게 된다. 본 연구는 이러한 사람과 사람사이의 확인대화를 음성대화시스템에 적용하는 방법론에 대해 논한다. 확인대화의 신뢰도와 오류수정율을 높이기 위하여 오류감지 단계를 3단계로 나누고 그 3단계의 오류 정보를 이용하여, 확인대화전문가 시스템을 통한 다단계 오류 수정 방법을 보인다.

1. 소개

확인 대화란 사람과 사람사이의 혹은 사람과 컴퓨터사이의 의도 이해차를 줄이기 위한 대화 종류의 하나이다. 다음 예제는 티비 안내 시스템에서의 전형적인 확인대화의 형태를 보여준다.

예제
- 사용자 : 드라마 해신 좀 보자.
- 시스템 : 다큐멘터리 '해전' 이 현재 MBC에서 방영중에 있습니다. 보시겠습니까?
  (해신이 해전으로 음성 인식됨)
- 사용자 : 드라마 해신 보자.
- 시스템 : 지금 현재 2개의 드라마가 방영중입니다. 해신이 KBS에 내 이름은 김삼순이 MBC에 방영중입니다.
- 사용자 : 해신 KBS

위의 예제에서 보다시피, 음성대화는 반드시 오류를 수반하게 되며 성공적인 확인대화는 이러한 오류를 실제 대화를 진행하기 전에 미리 감지하고 오류의 종류에 따라서 대화를 통한 오류 수정을 이끌어 낼 수 있어야 한다. 즉 확인대화를 효과적으로 수행하는 것은 다음의 두가지 문제를 효율적으로 해결하는 것으로 접근할 수 있다.

- 수정되어야 할 대상 감지 및 오류 판별 문제
- 효율적인 오류 수정 문제

위의 첫 번째 문제는 일반적으로 신뢰도 확인(Beliefe Confirmation)라 불리우며, 많은 연구들이 행해져 왔으며 그 대부분의 연구는 사용자의 발화가 음성인식 되었을때 그 음성인식이 얼마나 믿을 수 있는지를 판별 하는 방식으로 연구되어져 왔다. 기존의 많은 연구 결과를 통해서 인식기 수준에서의 인식 판정 기준 및 자질들은 많이 소개되었으며, [2][3]에 잘 정리되어 있다.

최근에는 단순히 인식기 수준의 정보뿐만이 아니라 문맥의미 정보까지 살펴봄으로서 오류 판정을 하기도 하며 [1], 음성언어이해 단계를 거쳐서 그 이해 정보를 이용하여 판정하기도 한다[9]. 이러한 다양한 최근의 오류 판정 경향은 [4]에 잘 조사되어 정리되어 있다.

기존의 확인대화를 구현한 많은 연구들은 위의 신뢰도 확인의 결과를 이용하여 확인대화를 시도하였다. [9]에서는 위의 신뢰점수를 대화 모델의 전이 확률에 적용함으로서 확인대화 상태로 전이되도록 하는 방식을 취했고, [5] 역시 위의 [3]의 신뢰도 측정 방법론을 그대로 적용하여 대화모델에서 사용하는 각 정보의 신뢰 수준을 정하도록 하였다.

* 본연구는 산업자원부 중기거점 지원 "자동차용 음성 HMI"과제의 지원을 받아 수행되었음.

기존의 확인대화 연구들의 한계점중 하나는, 확인대화가 수정하고자 하는 대상이 각 단어 수준에 불과하다는 것과 음성언어의 오류는 음성인식의 오류에서 뿐만이 아니고 음성언어이해에서도 발생함에도 불구하고, 오류의 대상을 단지 음성인식에 국한하였다는 점에 있다[9][5][3]. 또한 음성언어이해의 오류 또한 감지하고 수정하고자 하였던 연구에서도 음성인식 오류판별과 음성언어이해 오류판별을 체계적으로 분리하지 않고 하나로 묶음으로서 그 신뢰성을 떨어뜨렸다[7].

본 연구의 3단계 검증 방법론은 위의 한계점들을 극복할 수 있도록 고안되었다. 먼저 오류의 의미를 단지 음성인식의 오류가 아닌 음성언어이해의 오류까지 포함된 포괄적 오류로 정의하고 판정을 함으로서 오류 수정의 범위를 넓혔으며, 체계적으로 나누어진 ***단어단위 음성인식 신뢰도 계산***, ***문장단위의 포괄적 오류판정***, ***대화 정보의 포괄적 오류판정*** 단계를 거침으로서, 각 오류 판별의 신뢰도를 높일 수 있었다. 또한 이러한 3단계의 검증은 보다 풍부한 오류 정보를 대화시스템에 넘겨줌으로서 단어 단위의 부분 수정 뿐만 아니라 문장 전체에 대한 수정 지침 정보를 제공할 수 있게 되는 장점이 있다

본 논문은 다음과 같은 차례로 구성되어 있다. 2절에서는 이번 연구에 사용된 대화 시스템인 상황기반 대화 관리 시스템에 대해 소개하고, 3절에서는 3단계 검증법에 대한 설명을 하며 4절에서는 본연구의 대화수정 방법론에 대해 논하고 5절에서는 오류판정 실험 결과들과 확인대화 실험의 결과에 대해 기술하고 분석하도록 한다.

2. 상황기반 대화 관리 시스템

McTear의 객체지향적 대화시스템 [6] 에 영향을 얻어, 본 연구진은 기존의 대화시스템의 한계를 극복할 수 있는 대화 시스템을 개발하였으며, 이는 다음의 두 대화 모델 철학에 바탕을 두었다.

- 대화 관리는 현재 대화의 상황에 기반을 두어야 하며 이는 제한된 상태전이 네트웍에 기반하지 않아야 한다.
- 특정 도메인의 대화를 전담하는 ***대화전문가***를 두어, 여러 도메인의 대화를 자연스럽게 드나들 수 있도록 한다.

기존의 대부분의 대화 시스템은 상태전이 기반의 대화 시스템 이었다. 즉, 미리 정해져 있는 상태 전이 넷트웍 안에서 각 상태들 사이를 전이 하면서 그 상태에 맞는 대화를 이끌어 나가는 방식이었다. 이러한 방식은 빠른 시스템 개발과 간단한 대화관리에 있어서는 매우 효율적이지만, 자연어 음성 인식을 기반으로 하고 대화의 형태가 매우 다양해지는 경우에는 고정된 넷트웍으로 인해 다양한 대화를 처리할 수 없게 된다. 또한 상태 전이 넷트워크는 사람에 의해서 만들어 지고, 고정 되었기 때문에 만약 전혀 다른 도메인의 대화를 처리하려고 하면 처음부터 이러한 넷트워크를 다시 개발해야 하는 단점이 있다. 이러한 단점을 극복하기 위해서 우리는 상태전이 방식에서 벗어나 현재 대화가 처해있는 상황 자체를 파악하고 그 정보에 의거하여 대화를 관리하는 방식을 취하였다. 여기서 ***상황*** 이란 다음과 같이 정의된다.

- 상황 이란 다음의 정보를 포함하는 현재 대화의 정보로 정의 된다.
  (ㄱ) 사용자 발화 내용 및 발화 의도
  (ㄴ) 의미 슬랏과 그 값
  (ㄷ) 대화의 기록 (History)
  (ㄹ) 전 대화에서의 시스템의 의도
  (ㅁ) 데이터베이스에의 질의와 그 결과

현 단계에서의 사용자의 발화에 대한 적절한 대화를 처리하기 위해서 위의 모든 정보를 동시에 고려하여 대화를 처리하도록 설계하였다. 각 단계에서의 대화 전략 결정은 세 가지 종류의 규칙에 의해서 결정된다.

- ***상황-전략 규칙*** : 현재 상황에 맞는 적절한 시스템의 의도를 규칙으로 정리
- ***제약이완 규칙*** : 데이터 베이스 질의에 적합한 결과가 없을때 질의 제약을 이완하는 규칙
- ***프레임 재시작 규칙*** : 하나의 대화를 마무리 하고 새로운 대화 토막을 시작하는 기준을 규칙화 함

McTear의 연구 처럼 [6], 우리 역시 각 도메인 대화를 전담하는 ***대화전문가*** 시스템 방식을 취하였다. 예를 들어 티비 안내 관련 대화의 경우에는 티비전문가가 대화를 관리하며 영화 안내 관련 대화의 경우에는 영화전문가가 대화를 전담하는 방식을 취하였다. 이러한 각 대화 전문가는 각각의 상황관련 규칙을 가짐으로서 각 도메인에 적합한 대화를 진행해 나갈 수 있다.

이러한 대화전문가 방식의 큰 장점은, 특정 도메인 대화를 잘 처리하게 해주는 것 뿐만 아니라 확인대화 역시 체계적으로 구현되도록 해준다. 즉 확인대화 역시 특정한 대화 패턴의 종류이므로 이러한 대화 패턴을 전담하는 대화전문가를 구성하게 되면 그것이 바로 확인대화를 설계하는 방식이 될 수 있는 것이다. 이러한 아이디어에 기반하여 우리는 확인 대화를 전담하는 ***확인대화전문가*** 를 만들어 효율적이고 체계적인 확인대화를 전담하도록 하였다.

그림 1 은 우리가 개발한 상황기반 대화 시스템에 확인 대화가 가능하도록 확인대화전문가를 연결한 모습을 보이고 있다. 각 모듈에 대해서 살펴보면,

- **대화관리자** : 대화 관련 하부 모듈을 통제하며 음성인식이나 키보드를 비롯한 외부 입력 시스템과 음성합성기나 모니터를 비롯한 외부 출력 시스템과의 통신을 담당한다.
- **대화기록** : 두가지의 정보 기록으로 나누어 진다.
  - **대화프레임** : 현재 대화토막 전체에 대한 의미 정보 기록
  - **담화기록** : 사용자의 발화를 분석한 정보를 기록하며 그 신뢰도 및 상태에 대해서 기록 한다.
- **담화관리자** : 담화관리자는 대화전문가에게 상속됨으로서 대화를 처리 할 수 있으며, 기초적인 대화 패턴에 대해서 처리를 한다. (예 : 정보가 부족한 슬랏은 좀 더 구체화 되도록 시스템 발화를 정보 부족 슬랏의 구체화로 조정하는 것)
- **대화전문가(도메인전문가)** : 특정 도메인의 대화를 전담하여 처리함. 대화 전문가 자체는 도메인에 국한되지만 담화관리자를 상속함으로서 기초적인 대화 패턴을 처리 할 수 있다. 각 대화전문가는 자신만의 대화기록, 대화프레임, 상황관련 규칙을 가지고 있다.

담화관리자와 대화전문가는 서로 보완적인 관계를 가지고 있다. 담화 관리자는 그 자체로는 대화를 처리 할 수 없지만, 대화전문가에게 일반적인 대화 패턴을 상속시킴으로서 대화전문가가 특정 도메인의 대화를 처리함과 동시에 매우 기초적인 대화처리 역시 가능하게 한다.

대화시스템의 음성입력을 위하여 HTK(Hidden Markov Toolkit) 를 티비 안내 도메인에 맞게 훈련하고 3단계 검증법에 사용되게 될 정보를 제공할 수 있도록 수정하여 사용하였다.

3. 3단계 검증법

수정되어야 할 대상을 선택하고 그 오류의 신뢰도를 측정하기 위해서 3단계를 거쳐서 판별을 해나가도록 하였다. 첫 번째 단계는 ***단어단위 음성인식 신뢰도 계산***

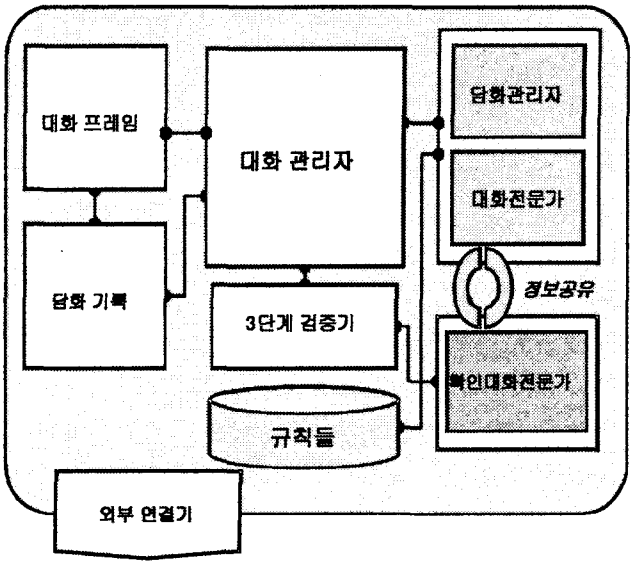

그림 1 상황기반 대화 시스템

으로서 음성인식된 결과물의 각 단어에 대해서 얼마나 신뢰할 수 있는지 확인하는 단계이며, 두 번째 단계는 ***문장단위의 포괄적 오류판정*** 이다. 이 단계에서의 포괄적 오류란, 단순히 음성인식의 오류만 고려하지 않고 음성인식의 오류와 음성언어이해를 동시에 고려하여 살펴본다는 의미이다. 마찬가지로 세 번째 단계에서도 포괄적인 오류 판정을 하게 되는데 세 번째 단계의 ***대화정보 단계의 포괄적 오류판정*** 란 음성인식의 결과물에 대해서 음성언어이해 모듈이 중요한 정보를 분석하고 추출하게 되는데 이러한 중요 정보를 ***대화정보***라 하고 그 대화정보 단위에서 음성인식 및 음성언어이해를 모두 고려하여 오류를 판정하는 것을 말한다. 각 단계의 판정 대상과 간단한 예제는 다음과 같다.

<사용자 발화 - 음성 인식 결과 >
케이비에스 드라마 해신 보고 싶다
엠비씨 드라마 배신 보고 싶다
< 단어단위 음성인식 신뢰도계산 >
엠비씨(29.24%) 드라마(87.32%) 배신(65.2%)
보고(78.32%) 싶다(54%)
< 문장단위의 포괄적 오류판정 >
엠비씨 드라마 배신 보고 싶다 : 40%
< 대화정보의 포괄적 오류 판정 >
채널 - 엠비씨 : 65%
장르 - 드라마 : 99%
이름 - 배신 : 45%

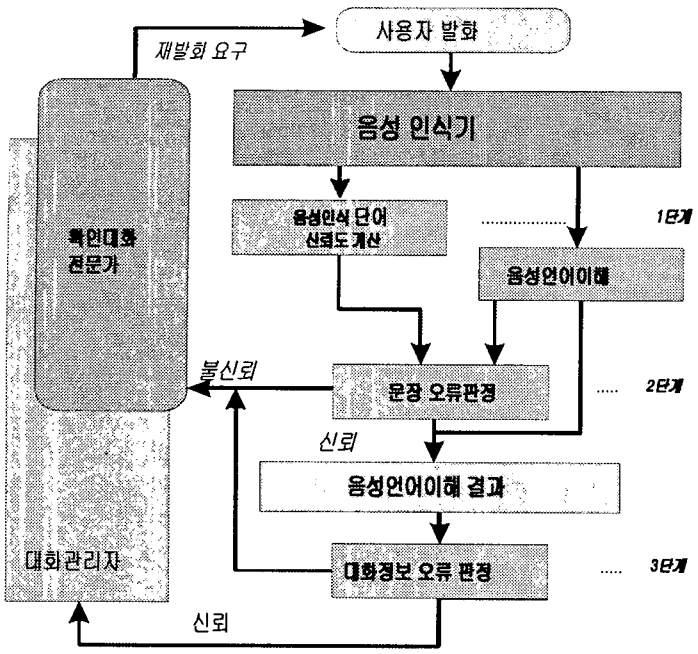

그림 2 3단계 검증의 흐름

3-1 단어단위 음성인식 신뢰도 계산

단어 단위의 음성인식은 전통적인 신뢰도 측정과 같은 검증이다[2][3][4]. 이 단계에서는 음성인식의 결과물에서도 특히 각 단어에 대해서 그 단어가 얼마나 신뢰할 수 있는지 검증한다. 본 연구에서의 이 단계는 곧바로 확인대화에서 사용되지 않고 다만 다음 단계인 **문장단위의 포괄적 오류판정**과 **대화정보 단계의 포괄적 오류 판정** 에 보다 정확한 신뢰도 점수를 제공하기 위해서 먼저 검증된다.

우리는 기존 연구 결과에서 신뢰도 측정에 좋다고 알려진 몇 개의 자질들을 최대 엔트로피 모델[9]을 이용하여 신뢰도 계산을 하였으며 검증시에 사용된 클래스와 자질들은 다음과 같다.

- ***클래스*** : 신뢰 / 거부
- ***평준화된 음향 점수*** : 각 음성인식 격자에서의 노드 하나의 음향점수를 프레임 길이로 평준화함
- ***단어명*** : 단어 글자 자체
- ***언어모델점수*** : Trigram 언어 모델 점수 P(W2|W0,W1)
- ***N-Best 순수도 (N-best purity)***: N-best 의 각 결과물을 살펴 현재 단어가 같은 위치에 몇 번 나타났는가 하는 비율
- ***프레임길이*** : 현재 단어의 상대적 시간 길이

3-2 문장단위의 포괄적 오류판정

확인대화의 가장 최우선 목적은 사용자의 의도와 시스템이 이해한 사용자의 의도사이의 의도차이를 줄여나가는 것에 있다. 이는 사용자의 의도가 얼마나 최종 대화시스템에게 손실없이 전달되는지 파악 하는것과 매우 밀접하다. 다시 말하면, 사용자의 의도가 언어 형태를 갖추어 음성형태를 가지고 음성 채널을 거쳐서 음성인식기에게 인식되고, 그 음성 인식 결과가 다시 음성언어이해 모듈에 의해서 이해되는 이 모든 과정을 거치는 동안 사용자 의도정보의 손실이 얼마나 적었느냐를 계산하고 그 결과에 의거하여 확인대화 전략을 세워나가는 것이 자연스러운 접근이다. 이를 위해 우리는 다음과 같은 논리를 통해 ***정보보존율***을 정의하였다.

정보보존율 = 사용자의 정보(의도)가 얼마나 손실없이 대화시스템에게 전달되는지를 정량화
= 시스템이 이해한 사용자의 정보를 얼마나 믿을 수 있는가 정량화
= P(***신뢰*** | ***시스템에 인지된 사용자 정보***)

정보보존을 잘한다는 것은 바꿔 말하면 그 정보를 전달(음성인식)하고, 해독하여 이해(음성언어이해)할 때의 신뢰도가 얼마나 되느냐와 같은 문제가 된다. 즉 다시 말하면 다음과 같다.

정보보존율 ~ < 음성인식의 신뢰도 >
~ < 음성언어이해의 신뢰도 >
~ < 그 외의 정보 >
('~' 는 관계가 있음을 의미 함)

이와 같은 논리에 의거하여, 여러 가지 점수와 정보들을 **Logistic Regression** 을 통해 정보보존율을 계산하였다. 사용된 자질과 점수는 다음과 같다.

- **평준화된 음향점수의 평균 :** 문장의 모든 단어의 평준화된 음향점수의 평균
- **언어모델점수의 평균 :** 문장의 모든 단어의 언어모델 점수의 평균
- **N-best 순수도의 평균 :** 문장의 각 단어의 N-best 순수도의 평균
- **음성언어이해점수의 평균** : 문장에서 음성언어이해 모듈이 추출한 각 정보에 대한 음성언이해점수의 평균
- **단어단위 음성인식신뢰도의 평균 :** 1단계에서 검증한각 단어의 음성인식신도들의 평균
- **단어 수 :** 문장에 나타난 단어 수
- **예측된 단어오인식률 :** 1단계 검증에서 음성인식오류라고 판정된 단어의 비율

위와 같은 자질들을 사용하여 계산한 정보보존율이 계산되었으면, **그림 2**에 보인것처럼 그 문장의 정보보존율이 기준점 보다 낮게 나올 경우 불신뢰 판정을 내리게 된다. 그리고 그러한 불신뢰 문장은 바로 확인대화전문가

에게 보내져서 문장 전체의 재발화를 요구토록 한다. 이에 비해 기준점보다 높게 나와 정보보존이 충분히 잘 되어 대화시스템에게 전달될 만한 문장에 대해서는 다음단계인 대화정보의 오류 판정 단계로 넘어가 문장전체는 믿을만 하지만, 대화정보 하나 하나 단위에서는 얼마만큼 믿을 수 있는지 다시 검증하는 단계로 넘겨지게 된다.

3-2 대화정보의 포괄적 오류 판정

이 단계에서의 검증은, 음성인식에 의해서 인식되고 다시 음성언어이해 모듈에 의해서 이해된 문장이 충분히 정보보존이 잘되어 전달되었다고 생각되어 졌을 때 실행된다. 즉, 문장 전체로 보아서는 정보보존이 잘되었다고 생각되지만 각 대화정보*(예 : 채널-엠비씨)* 에 대해서 보다 세밀하게 음성인식측면에서나 언어이해측면에서 과연 신뢰할만한지 다시 검증 하는 단계이다.
따라서 검증에 사용되는 자질들은 1단계의 ***단어단위 음성인식 신뢰도 계산***에서 사용된 자질을 그대로 사용하며 추가된 자질은 음성언어이해 모듈의 ***이해점수***를 추가하여 최대엔트로피 모델을 이용하여 오류를 판정하도록 하였다.

- **언어이해점수** : 해당 대화정보를 음성언어이해모듈이 추출하였을 때의 점수

이렇게 계산된 점수는 그 신뢰/불신뢰 여부에 따라서 신뢰할만하다 생각되는 경우는 바로 대화관리자에게 입력 정보로 넘겨지지만, 불신뢰의 경우는 확인대화전문가에게 넘겨져 특별히 현재 대화정보가 믿을만 하지 못하므로 사용자로 하여금 재발화하도록 하는 시스템발화를 만들게 한다.

4 대화 수정 방법론

3단계의 문장 및 단어 그리고 대화정보 검증을 거쳐서 그 결과는 확인대화전문가에게 보내지게 된다. 확인대화전문가는 특정 도메인의 대화를 전담하는 대화전문가(도메인전문가) 와 마찬가지로 확인대화가 필요한 경우 관련된 대화를 전담하여 처리하는 대화전문가라고 할 수 있다. 그러나 일반적인 대화전문가와는 다음과 같이 몇가지 차이점이 있다.

- 확인대화전문가는 대화전문가가 없이는 활성화 될 수 없다. 즉 대화전문가에 보완적으로 활동하는 부-대화전문가이다.
- 확인대화전문가는 도메인에 독립적인 대화전문가이다.
- 확인대화전문가는 대화전문가의 모든 정보에 접근할 수 있다.

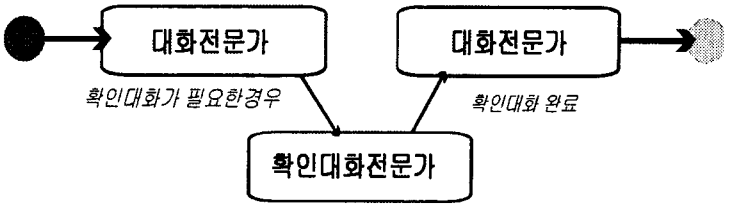

그림 3 대화전문가의 전환

그림 3에서 보이는 것처럼, 3단계 검증 단계에서 확인대화가 필요함이 전해오면 대화관리자는 현재 대화 전문가를 잠시 정지시키고 확인대화 전문가에게 주도권을 넘겨주며 이때 현재 확인대화가 필요한 대화전문가의 대화기록, 대화프레임, 담화 기록들을 모두 확인대화전문가가 사용할 수 있도록 해준다. 확인 대화 전문가는 3단계 검증단계에서 보내온 수정해야할 대상과 그 신뢰/불신뢰 점수, 그리고 여러 가지 사용가능한 정보를 조합하여 그에 맞는 확인대화전략을 정하게 된다.

이때 보다 효율적이고 체계적인 대화전략을 정하기 위해서 확인대화전문가는 다음과 같은 세가지 기본 원칙에 의거하여 전략을 결정한다.

- 원칙 1 : 대화정보 사이에는 의존성이 존재한다.
- 원칙 2 : 대화정보들 사이에도 우선순위가 있다.
- 원칙 3 : 대화정보들 사이의 관계는 능동적으로 계속 변한다

대부분의 대화정보 사이에는 매우 강한 의존성이 존재한다. 예를 들어 '해신' 은 언제나 'KBS'에서 방영된다. 즉, 대부분의 프로그램 이름과 채널사이에는 강한 의존성이 있으며 이를 고려하지 않는다면, 다음과 같은 사용자발화를 확인대화 할 때 불필요한 단계를 거치게 된다.

사용자 발화	KBS 드라마 해신 보자
음성인식결과	MBC 드라마 배신 보자
수정해야할 대상	배신 / MBC

즉, '해신'은 항상 'KBS' 의존 된다는 것을 모른다면 확인대화전문가는 '배신' 과 'MBC' 두 개에 대해서 사용자에게 두 번 재발화 요청을 내려야 하지만 이와 같은 의존성을 알고 있다면 '배신' 이 '해신'으로 성공적으로 재조정이 되는 순간 바로 채널을 KBS라고 사용자에게 물어보지 않고 수정할 수 있다. 즉 위의 **원칙 1**과 **원칙 3**을 잘 고려하면 효율적으로 확인대화를 구현할 수 있다.

그림 4 티비가이드 도메인에서의 대화정보의 범위

위의 예제에서 두 개의 수정대상이 있을 때(배신/MBC) 확인대화전문가는 무엇이 먼저 수정되어야 할지 결정하게 된다. 이때 **원칙 2** 에 따라 대화정보 사이의 우선순위를 고려하여 무엇이 먼저 수정되어야 하는지 결정하게 된다. **그림 4** 는 티비가이드 도메인에서의 대략적인 대화정보의 범위를 보였다. 그림에서 보듯이 만약 출연자가 결정되면, 프로그램, 채널 그리고 장르등의 대화수정이 필요 없이 현재 시간대에 방송되는 특정출연자가 나오는 프로그램을 몇 개 이내로 한정할 수 있다. 즉 출연자>프로그램>채널>장르 순으로 우선순위를 주고 이와 함께 포괄적 신뢰도를 동시에 고려함으로서 효율적인 확인대화가 이루어지도록 하였다.

5 실험 결과 및 분석

1 주일 분의 티비스케쥴에 대하여 음성인식을 입력으로 하는 실험을 하였다. 이때 사용된 음성인식기의 단어오인식률은 실험실 환경에서 10.2 % 였다.

표1은 1단계의 단어단위 음성인식 신뢰도 측정에 사용된 최대엔트로피 기반의 신뢰도 측정단계의 정확도를 나타 낸 것이다. 잘못 인식된 단어를 잘 인식되었다고 하는 경우(False Accept Rate)가 10% 안쪽으로 낮게 나왔으므로, 1단계의 단어단위 인식신뢰도는 상당히 믿을만하며, 2단계와 3단계 검증단계에 신뢰할만한 정보를 제공함을 알 수 있다.

표2는 2단계에서 계산해낸 정보보존율에 의거하여 신뢰한 문장, 불신뢰한 문장 그리고 모든 문장에 대해서 각각 음성언어이해 모듈로 정확도를 측정하고 그것의 정확도를 측정하였을 때의 모습이다. 이를 바탕으로 보면, 정보보존이 잘 되었다고 판단된 문장은 실제로도 음성언어이해 모듈에 의해서 성공적으로 이해됨을 알 수 있고 불신뢰된 문장은, 음성언어이해모듈을 거쳐서 그 결과를 내놓는다고 해도 성공적인 대화 처리가 힘들만큼 이해가 잘 안됨을 알 수 있다. 이 결과를 보면 우리의 정보보존율에 의거한 신뢰/불신뢰 판정이 불필요한 대화 턴수를 줄이게 함을 예측 할 수 있고 그 결과는 **표3** 의 평균대화 턴수의 감소에서도 확인 할 수 있다.

표3은 음성인식기와 대화시스템만을 가지고 실험을 한 경우와 3단계 검증법과 확인대화전문가를 사용하여 실험

정답 \ 예측	Correct	Error
Correct	95.14%	4.59%
Error	10.3%	89.70%
False Alarm Rate	8.04%	
False Accept Rate	5.94%	

표1 단어단위 음성인식오류판정의 Confusion Matrix

	음성언어이해 정확도
정보보존신뢰 문장	93.40%
정보보존불신뢰 문장	31.07%

표2 정보보존 되었다고 판단된 문장과 일반 문장에 대한 음성언어이해 정확도 측정

	음성인식 Baseline	3단계검증을 거친 확인대화
대화 성공률	0.77	0.96
평균 대화 턴수	9.44	8.96
정보보존 불신뢰 발생률	-	0.15
대화정보 불신뢰 발생률	-	0.07
오류 수정률		
문장재발화	-	0.72
단어 재발화	-	0.44

표 3 확인대화 실험 결과

을 한 경우의 정보를 정리해 놓은 표이다. 총 6명의 실험자가 10개의 시나리오에 대해서 각각 발화를 하였으며 각 사용자 발화와 시스템 발화에 대해서 사용자가 수정이 잘 되었는지를 판단한다. 두 개 시스템에서 가장 차이가 나는 부분으로 평균대화 턴수의 감소를 들 수가 있다. 여기서 대화 턴수는 사용자의 발화 문장 개수와 시스템의 물리적 행동 갯수를 합친 수를 말한다. 물리적 처리란 데이터베이스에의 질의행동이나 티비의 채널을 바꾸는 등 실제로 행동을 하는 경우를 말한다.

음성인식 Baseline실험의 경우 시스템이 사용자가 원하

는 바와 다르게 행동할 경우 의도하는 바가 제대로 인지될때까지 반복을 하게 하며 3번까지 재발화를 허용하고 3번안에 제대로 처리가 안되는 경우 대화 실패로 처리하게 된다. 평균 대화 턴수의 감소는 바꿔 말하면 3단계 검증을 거친 확인대화가 불필요한 시스템의 물리적 행동을 줄여주는 것을 의미하고 이는 사용자의 시스템에 대한 편의성을 높인다는 것을 의미한다.

또한 문장재발화가 필요하여 문장재발화를 통해 대화수정을 시도하였을때 그 성공률이 단어의 재발화보다 훨씬 높은 성공률을 보임을 알 수 있다. 이는 여러 가지면에서 해석할 수 있는데, 음성인식 자체가 문장으로 재발화할 경우 훨씬 잘 되고, 단어단위의 재발화할 때는 단어단위만 수정되지만 문장단위의 재발화 경우엔 다른 대화정보가 들어오므로 보다 확실하게 확인대화가 이루어지는 것으로 생각해 볼 수 있다.

6 결론

본 논문에서 우리는 사람과 사람사이의 확인대화를 음성기반대화시스템에 구현하였다. 확인대화문제는 필연적으로 수정대상의 감지와 오류 판정 문제 그리고 그것을 효율적으로 수정하는 문제로 나뉠 수 있다. 수정대상의 감지를 위하여 사용자의 의도 정보가 얼마나 정확히 시스템에 전달되었는가의 측정을 위해 음성인식과 음성언어이해 모두를 고려한 정보보존율을 제안하였다. 또한 문장으로서는 신뢰 가능하지만 부분적으로 잘못 전달된 단어를 찾기 위해 대화정보단위의 세밀한 신뢰도 측정을 통해 단어단위의 확인대화를 가능하게 하였다. 다양한 실험을 통하여 3단계의 검증이 수정할 대상을 높은 신뢰도를 가지고 찾아내며, 확인대화전문가는 이를 효율적으로 수정함을 확인 할 수 있었다.

참고 문헌

[1 S.J. Cox and S. Dasmahapatra, "A semantically-based confidence measure for speech recognition," in *Proceedings of The ICSLP 2000*, Beijing, 2000.

[2] 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, January 2002.

[3] T. J. Hazen, T. Burianek, J.Polifroni, and Stephanie Seneff, "Recognition confidence scoring for use in speech understanding systems,"in *Proceedings of the ISCA ASR2000 Tutorial and Research Workshop*, Paris, September 2000.

[4] H. Jiang, "Confidence measures for speech recognition," *Speech Communication*, vol. 45, no. 4, pp. 455–470, April 2005.

[5] 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, March 2005.

[6] I. O'Neill, P. Hanna, X. Liu, D. Greer and M. McTear "Implementing advanced spoken dialogue management in Java," *Speech Communication*, vol. 54, no. 1, pp. 99–124, January 2005.

[7] T. Paek & E. Horvitz. Conversation as action under uncertainty. Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence, 2000, pp. 455-464.

[8] A. Ratnaparkhi "Maximum Entropy Models for Natural Language Ambiguity Resolution", Ph.D. Dissertation. University of Pennsylvania. 1998

[9] 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, March 2005.