

텔레매틱스에서 효율적인 음성 대화 시스템을 위한 통계적 대화 관리

A Stochastic Dialog Management for Efficient Spoken Dialog Systems in Telematics

유 제 만

(광운대학교, 석사과정)

최 기 호

(광운대학교, 교수)

Key Words : 대화 관리 시스템, 통계적 대화 관리, Dialog Slot

목 차

- I. 서 론
- II. 시스템 구성
- III. Corpus Labeling
 - 1. User Dialog Acts
 - 2. System Dialog Acts
- IV. 통계적 대화 관리
- V. 구현 및 결과
- VI. 결 론
- 참고문헌

I. 서론

최근 음성인식 기술의 발달과 중요성 등으로 인해 다양한 범위로 응용범위가 확대되고 있다. 이러한 분야 중에 최근에 대두되고 있는 분야가 텔레매틱스 분야이다. 지능형 자동차 기술의 급격한 발달로 점차 차량 환경에서 사용자와 시스템 간의 상호작용을 필요로 하는 음성 대화 시스템의 중요성이 나날이 증가하고 있다.[1][2]

음성 대화 시스템은 차량 내에서 운전자가 운전을 하는 도중에 차량 인터페이스 제어를 위해 한눈을 파는 사이 교통사고의 위험에 노출되기 때문에 안전운행을 위한 수단으로서 필요성이 제기된다. 하지만 음성 대화 시스템의 사용과 역할에 있어 이에 대한 충분한 연구가 행해지지 않고 있다.

음성 대화 시스템은 일반적으로 상호작용을 위해 몇 개의 모듈로 구성되어 있다. 이 모듈은 음성 인식기, 언어 이해 모듈, 대화 관리자, 응답 생성기, 음성 합성기이다. 이러한 모듈은 도메인의 목적, 도메인의 응답 범위, 모듈의 능력에 따라 가장 편리한 모듈을 선택 할 수 있다. 이러한 모듈을 구성하는 가장 성공한 방법 중의 하나는 통계적인 방법의 사용이라고 할 수 있다(음성 인식, 음성 이해)[3][4][5].

이 논문에서는, 사용자와 시스템 간의 상호작용에서 효율적인 대화 관리와 예측 불가능한 상황(음성인식 오류, 예상외의 문장 발화, 인식과 이해 과정상의 오류)을 해결하기 위한 방법을 제안한다. 따라서, 본 연구에서는 음성 대화 시스템에서 효율적인 대화 관리를 위해 통계적인 방법을 이용한 대화 관리 방법을 제안한다. 이 방법은 사용자와 시스템 사이에 효율적인 상호 작용을 위해 대화 관리자를 사용한다.

본 논문은 대화를 분석하고 판단하고 적절한 응답을 생성하기 위한 대화 관리를 위해 Dialog Acts(User Dialog Acts와 System Dialog Acts)에 기반을 두고 모델화 한다.

이 시스템은 HTK(Hidden Markov Toolkit)를 통해 획득된 발화 정보만을 사용하고, 제안된 대화 관리 방법을 이용한 시스템은 차량 내의 대화를 통한 제어와 노선안내, 교통정보안내 도메인을 대상으로 한다.

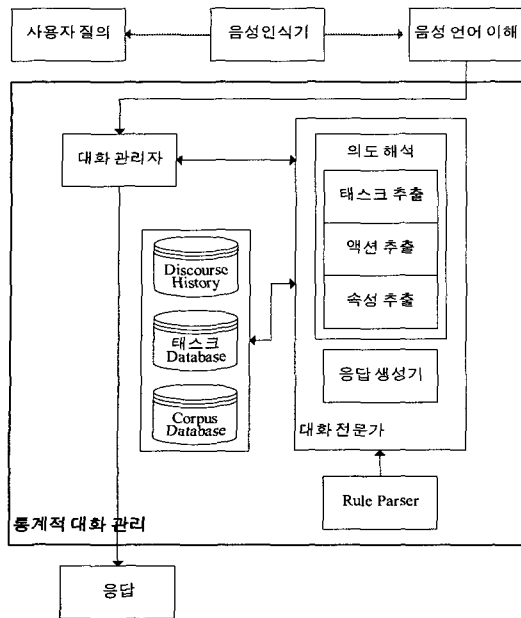
II. 시스템 구성

본 논문에서는 음성 인식기와 음성 언어 이해를 통한 발화 정보는 대화 관리가 적절한 응답을 생성 할 수 있도록 통계적으로 구성되었다. 이러한 통계적 대화 관리는 크게 대화 관리자, 대화 전문가, Domain DB, Corpus DB, Rule Parser, Discourse History로 구성되었다. <그림 1>은 통계적 대화 관리의 전체적인 시스템을 나타낸다.

대화 관리자는 대화 처리를 위한 시스템에서 중추적인 역할을 하는 부분이다. 사용자에게 필요한 정보를 제공하기 위해 외부 지식 자원들을 연결하여 시스템 발화를 생성하는 전반적인 대화 흐름을 제어하는 부분이다.

Discourse History는 하나의 대화 세션을 저장하는 프레임이다. 사용자는 시스템과의 대화에서 자연스러운 대화로 인해 생략 현상이 빈번하게 발생한다. 이로 인해 시스템이 사용자의 의도를 파악하는 것은 쉽지 않다. 따라서 이러한 문제점을 해결하기 위해 Discourse History를 가진다.

대화 전문가는 사용자와 시스템간의 대화를 위한 지식을 가지고 있으며, Rule parser와 연동하여 질의를 분석하고 응답을 생성하는 언어 생성 모듈을 가지고 있다.



<그림 1> 시스템 구조

시스템은 도메인에 대한 정보를 제공하는 Domain DB를 가지고 있다. 그리고 음성 언어 이해의 출력 결과와 User Dialog Acts, System Dialog Acts의 정보를 가지고 있는 Corpus DB를 가지고 있다.

제한된 시스템의 Domain DB는 사용자 질의를 명확히 처리하기 위하여 베이지안 네트워크로 구성하였다.

또한 통계적 대화 관리는 사용자 발화를 입력 받아 음성 인식기와 음성 언어 이해를 통해 출력된 결과를 사용한다. Labeling된 Corpus를 통계적 대화 관리에 적용하여 사용자와 시스템 간에 효율적인 대화를 할 수 있도록 구성하였다.

III. Corpus Labeling

사용자와 시스템의 행위를 정의하여 통계적 대화 관리를 효율적으로 하기 위해 사용자와 시스템에 대한 대화 행위 labeling을 하였다. 이것은 상호작용을 명확하게 하여 대화를 원활하게 할 수 있다. 본 연구에서는 사용자와 시스템의 대화를 기술하기 위해 Dialog Acts 집합을 정의한다.

1. User Dialog Acts

사용자의 대화는 User Dialog Acts에 대한 labeling으로 표현된다. 이것은 사용자 발화를 의미하는 고전적인 프레임의 표현이다. User Dialog Acts의 하나 또는 그 이상의 정보는 사용자 발화의 의도를 표현한다. 사용자 발화는 User Dialog Acts에 대한 정보를 포함하기 위해 labeling으로 나타낸다. User Dialog Acts는 <표 1>에서 나타낸다.

음성 언어 이해 모듈은 인식 과정을 통해 제공된 문장을 분석하고, 대화 전문가를 통해 문장에 대한 labeling을 생성한다.

User Dialog Acts는 총 3부분으로 구성되어있다.(3 Tasks, 8 Actions, 6 Attributes)

1. Tasks : 대화 관리 시스템을 위한 도메인을 나타낸다.
2. Actions : 검색, 수정, 켜기, 끄기, 열기, 닫기, 확인, 부정과 같은 것의 개념을 표현하여 사용자가 할 수 있는 행동을 표현한다.
3. Attributes : 태스크에 관련하여 제공되는 속성이다.

<표 1> User Dialog Acts

Tasks	Actions	Attributes
Traffic	Search	Origin
Navigation	Modify	Destination
Service	Confirmation	Heater
	Negative	Air conditioner
	On	Window
	Off	Audio
	Open	
	Close	

입력 문장의 의미적 해석에 대한 예는 <표 2>에서 보여준다.

<표 2> 입력 문장에 대한 의미 해석의 예

Input Sentence	: 서울에서 부산가는 길을 알려주세요.
DS	: 010-10000000-110000
Semantic Interpretation	:
Task	- Navigation
Actions	- Search
Attributes1	- Origin : 서울
Attributes2	- Department : 부산

2. System Dialog Acts

System Dialog Acts를 2단계 레벨로 정의하여 <표 3>에 보인다. 첫 번째 레벨은 도메인에 대해서 독립적으로 행해지는 일반적인 행위를 기술한다. 두 번째 레벨은 사용자와 시스템 사이의 대화에 관련된 속성들로 구성한다.

<표 3> System Dialog Acts

First Level	Second Level
Opening	Search
Closing	Modify
Un-defined	Confirmation
Not-Understood	Negative
Waiting	On
New-Query	Off
Acceptance	Open
Rejection	Close
Question	Null
Confirmation	
Answer	

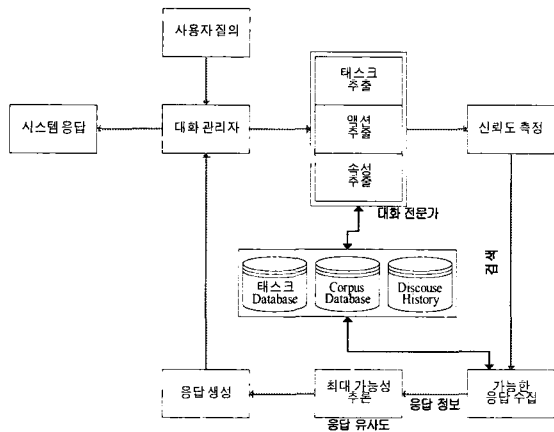
이것은 User Dialog Acts와 같이, 하나 또는 그 이상의 속성

들로 labeling한다. <표 3>은 System Dialog Acts를 나타낸다.

IV. 통계적 대화 관리

일반적인 상호작용에서 Dialog Acts의 양을 고려할 때, 그 크기가 작으면 잘 훈련된 모델을 얻을 수 있다. 그러나 그 모델로 대화를 완벽하게 하는 데는 충분하지 않다. 또한 도메인에 대한 완벽한 모델을 구현하기 위해 추가적으로 많은 양의 수작업을 필요로 한다.

따라서, 사용자와 시스템간의 대화를 효율적으로 하기 위하여 Dialog Acts를 사용한 통계적 대화 관리를 제안한다. 제안된 대화 관리는 일반적인 의도뿐만 아니라 사용자나 시스템이 원하는 정보에 대해 적절하게 대처 할 수 있도록 설계되었다. <그림 2>는 통계적 대화 관리를 위해 설계된 구조이다.



<그림 2> 통계적 대화 관리

제안된 통계적 대화 관리의 형식적 기술은 아래와 같다.

1. 질의 정보 : 음성 언어 이해를 통해 분석된 문장의 정보는 대화 전문가와 Domain DB를 통해 도메인 정보(Task), 사용자의 의도(Actions), 질의 속성(Attributes)을 추출하여 labeling을 한다.

2. Labeling된 정보를 통해 Domain DB, Corpus DB와 Discourse History를 검색하여 응답에 필요한 정보를 수집한다. 이 때, 질의에 대한 응답 가능한 모든 정보를 수집한다.

3. 수집된 응답 가능한 정보를 이용하여 최대 가능성 추론(Maximum Likelihood Estimation)을 통해 최적의 응답을 생성한다.

최적의 응답을 추론하기 위해, 시간 i 에서, 사용자의 대화를 U_i 라 하고, 시스템의 대화를 A_i 라 할 때, 대화열은 다음과 같이 정의 할 수 있다.

$$(A_1, U_1), \dots, (A_i, U_i), \dots, (A_n, U_n)$$

A_1 은 시스템의 인사말이고, U_n 은 사용자의 마지막 대화이다.

시간 i 의 대화열의 상태 (A_i, U_i) 를 S_i 라 한다. 최적 시스템 응답 \hat{A}_i 을 찾는다.

$$\hat{A}_i = \underset{A_i \in A}{\operatorname{argmax}} P(A_i | S_1, \dots, S_{i-1})$$

만약 발화가 정확하지 않을 경우, 더 진보된 방법이 필요하다. 이를 위해, DS(Dialog Slot)를 이용한 거리 측정을 제안한다.

DS는 대화 행위에 대한 정보를 포함하는 데이터 구조로 정의한다. DS의 각 Slot은 신뢰도 측정으로 value(0, 1, 2) 중 하나로 채워지게 된다.

0 : 속성의 값이 주어지지 않는다.

1 : 속성의 값이 주어진 경계 값(0과 1사이의 값) 보다 낮은 신뢰도 점수에 나타내어진다. 신뢰도 점수는 확인 발언에 의해 높일 수 있다.

2 : 속성의 값은 주어진 경계 값보다 높은 신뢰도 점수를 가진다.

신뢰도 측정은 HTK에서 제공되는 결과 값을 사용한다.

DS는 현재의 발화 정보와 Discourse History를 통해서 제공된다. DS는 시간 i 일 때 S_i 에 대응한다.

함수 f 는 현재 DS와 DS'의 거리를 계산하는 함수이다. 따라서, 아래와 같이 정의한다. DH는 Dialog History이다.

그러므로, 이것을 다음과 같이 정의한다.

$$d(DS, DS') = \underset{ds' \in DH}{\operatorname{argmin}} f(ds, ds'_k), k = 1, 2, \dots, n$$

게다가, 추론된 DS'는 Discourse History에 업데이트된다. 이러한 방법으로 추론된 결과에 의존하여 응답을 생성한다.

V. 구현 및 결과

이 시스템은 HTK를 통해 획득된 발화 정보만을 사용하고, 차량 내의 대화를 통한 제어와 노선안내, 교통정보안내 도메인을 대상으로 한다. 시스템은 대화로 구성된 10개의 시나리오를 수행하는 대화 관리의 행동으로 평가했다. 5명의 사용자에게 의해 시나리오 당 평균 4.4개의 대화가 수행되었다. 음성 인식기는 HTK를 사용하였다. 또한 DS의 체계화를 위한 신뢰도 측정의 경계 값은 0.5이다.

대화 시스템은 텍스트 입력과 음성 입력으로 평가하였다.

Dialog Success Rate(%)는 대화가 성공적으로 이루어진 경우의 백분율이다. 대화가 진행된 시나리오에서 사용자는 시스템으로부터 하나 또는 그 이상의 정보를 획득해야 한다. 대화 성공은 시스템으로부터 정확한 정보를 획득 했는지 아닌지에 의존한다. <표4>에서 대화 성공률은 음성 입력의 경우 76%, 텍스트 입력일 경우 84.7%의 성공률을 보였다.

대화관리자 시스템 사용자 음성 언어 이해	A_1 (Opening : Null) 차량 대화 시스템에 오신 것을 환영합니다. 무엇을 도와드릴까요? 서울에서 부산가는 길을 알려주세요. 서울/NNP+에서/JKB+부산/NNP+가/XSN+는/JX+길/NNG+을/JKO+알리/VV+어/EC+주/VX+시/EP+어요/EF+./SF System : (Opening : Null) \hat{A}_2 (Answer : Search)(New-Query)	DS : 100-0000000-000000 DS : 010-10000000-110000
대화관리자 시스템 사용자 음성 언어 이해	A_2 (Answer : Search)(New-Query) 길을 안내해 드리겠습니다.(Display) 더 필요한 것이 있습니까? 교통상황을 말씀해 주세요. 교통/NNG+상황/NNG+을/JKO+말씀/NNG+하/XSV+아/EC+주/VX+시/EP+어요/EF+./SF System : (New-Query : Search) \hat{A}_3 (Answer : Search)(New-Query)	DS : 010-10000000-110000 DS : 100-10000000-110000
대화관리자 시스템 사용자 음성 언어 이해	A_3 (Answer : Search)(New-Query) 서울에서 부산가는 교통상황을 알려드리겠습니다.(Display) 더 필요한 것이 있습니까? 아니요. 감사합니다. 아니/VCN+요/EF+./SF 감사/NNG+하/XSV+입니다/EF+./SF System : (Closing : Null) \hat{A}_4 (Closing : Null)	DS : 100-10000000-110000 DS : 100-10000000-110000
대화 관리자 시스템	A_4 (Closing : Null) 서비스를 이용해 주셔서 감사합니다.	DS : 100-10000000-110000

<표 3> 제안된 대화 관리의 수행 과정

음성입력과 텍스트입력의 경우 8.7%의 차이를 보였으며, 이 차이는 음성인식기의 오류로 발생한다. <표 3>은 제안된 대화 관리의 수행 과정을 보여주는 것이다.

<표 4> 통계적 대화 관리에 대한 실험 결과

평가	음성 입력	텍스트 입력
사용자 수	5	
시나리오 수	10	
시나리오 당 사용자 평균 발화 수	4.4	
대화 성공률	76%	84.7%

시스템이 사용자가 대화를 끝내고자 하는 것을 찾지 못하는 경우가 몇 번 발생 되었는데 이것은 단지 불확실한 상황이 DS의 거리에 의존하기 때문이다. 이와 같이 Closing 동작이 수행되지 못하는 경우 이에 대해 추론 방법의 보완이 필요하다고 본다.

VI. 결 론

본 논문에서, 텔레매틱스 환경의 음성 대화 시스템에서 효율적인 대화 관리를 위한 통계적 방법을 제안했다. 사용자와 시스템이 효율적으로 대화를 수행하고 적절한 응답을 생성하기 위하여 사용자와 시스템에 대한 Dialog Acts를 정의하였다.

또한 음성인식의 신뢰도와 DS 거리 측정을 이용함으로써 불확실한 발화를 좀 더 명확하게 하여 적절한 응답을 생성할 수 있도록 하였다. 그리고 대화 관리는 DS를 이용함으로써

대화의 내용을 좀 더 간략히 정리하여 분석할 수 있도록 설계하였다.

향후, 한국어의 다양한 어순에 적용하여 응답을 생성할 수 있는 대화 관리의 설계와 불확실한 대화에 대해 좀 더 명확하게 의도를 분석할 수 있는 대화 관리에 대한 연구가 필요하다고 본다.

참 고 문 헌

- [1] Cheongjae Lee, Sangkeun Jung, Jihyun Eun, Minwoo Jeong, G.G. Lee, "A Situation-Based Dialogue Management using Dialogue Examples," *IEEE International Conference on Acoustics, Speech, and Signal Processing(ICASSP), May 2006*
- [2] 유제만, 차명호, 최기호, "텔레매틱스에서 베이지안 네트워크를 이용한 대화 관리 시스템," 06년 추계 ITS 학회, pp. 286-289, 2006
- [3] W. Minker, A. Waibel, and J. Mariani, "Stochastically-Based Semantic Analysis," Kluwer Academic Publishers, Boston, 1999.
- [4] Yulan He and S. Young, "A data-driven spoken language understanding system," in Prof. of. ASRU, 2003, pp. 583-588.
- [5] E. Segarra et al., "Extracting Semantic Information Through Automatic Learning Techniques," *International Journal on Pattern Recognition and Artificial Intelligence*, vol. 16, no. 3, pp. 301-307, 2002.