

대화 예제를 이용한 상황 기반 대화 관리 시스템*

이청재(POSTECH), 정상근(POSTECH), 이근배(POSTECH)

<차 례>

- | | |
|---------------------|---------------|
| 1. 서론 | 4.2.1. 색인과 질의 |
| 2. 기존의 대화 모델 | 4.2.2. 발화 유사성 |
| 3. POSSDM 시스템 구조 | 5. 실험 및 결과 |
| 4. 대화 모델 | 5.1. 실험 환경 |
| 4.1. 규칙 기반 대화 모델 | 5.2. 대화 실험 |
| 4.2. 대화 예제 기반 대화 모델 | 6. 결론 |

<Abstract>

A Situation-Based Dialogue Management with Dialogue Examples

Cheongjae Lee, Sangkeun Jung, Gary Geunbae Lee

In this paper, we present POSSDM (POSTECH Situation-Based Dialogue Manager) for a spoken dialogue system using a new example and situation-based dialogue management technique for effective generation of appropriate system responses. Spoken dialogue system should generate cooperative responses to smoothly control dialogue flow with the users. We introduce a new dialogue management technique incorporating dialogue examples and situation-based rules for EPG (Electronic Program Guide) domain. For the system response inference, we automatically construct and index a dialogue example database from dialogue corpus, and the best dialogue example is retrieved for a proper system response with the query from a dialogue situation including a current user utterance, dialogue act, and discourse history. When dialogue corpus is not enough to cover the domain, we also apply manually constructed situation-based rules mainly for meta-level dialogue management.

* Keywords: Spoken dialogue system, Dialogue management, Dialogue modeling.

* 본 연구는 산업 자원부 21C 프론티어 연구 “인간생활자원 지능로봇” 과제의 지원을 받아 수행되었음.

1. 서론

음성 대화 시스템(Spoken Dialogue System)은 자연어를 이용하여 컴퓨터와 의사소통을 하는 차세대 인터페이스이다. 음성 대화 시스템은 많은 분야에서 이용되고 있으며 21세기에 유비쿼터스 시대를 위한 필수적인 지능형 인터페이스로 활발히 연구되어지고 있다[1]. 음성 대화 시스템은 일반적으로 음성 인식 (Speech Recognition), 음성 언어 이해 (Spoken Language Understanding, SLU), 대화 관리 (Dialogue Management), 음성 합성 (Text-To-Speech Synthesis, TTS) 등의 모듈로 구성된다. 이번 논문에서 발표하고자 하는 것은 현재 사용자의 발화를 바탕으로 적절한 시스템 발화를 생성하는 대화 관리 시스템의 방법론을 제시하고자 한다. 즉, 대화 관리는 음성 대화 시스템에서 중추적인 역할을 하는 부분으로 사용자의 음성을 받아서 의미를 추출하고 사용자에게 필요한 정보를 제공하기 위해 외부 지식 자원들을 연결하여 시스템 발화를 생성하는 전반적인 대화 흐름을 제어하는 부분이다.

2. 기존의 대화 모델

대부분의 기존 대화 관리 시스템의 유한 상태 기반 방식 (Finite State-Based Model)을 바탕으로 개발되어졌다[2]. 현재 대화 상태를 특정 상태로 사상을 하여 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 대화 시스템에 많이 사용된다. 그러나 사용자의 현재 발화에 의해 다음 턴에 사용자의 발화가 제한이 되거나 도메인이 바뀌면 모델 자체를 바꿔야 하기 때문에 도메인 간의 확장성이 떨어지는 단점을 가지고 있다. 이를 개선하기 위해 제시된 방식이 프레임 기반 대화 모델 (Frame-Based Dialogue Model)이다[3]. 이 모델은 자연스러운 대화 처리를 위해 고안되었으며 사용자의 발화에서 필요한 정보를 추출하여 의미 프레임을 채우고 그 정보를 바탕으로 대화를 진행하는 방식이다. 그러나 실제 대화 관리는 이러한 의미 프레임을 이용한 상태 전이나 수작업으로 만든 규칙에 의존하는 경우가 많아 유한 상태 모델과 같은 단점을 가지게 된다. 또한 최근에는 복잡한 도메인을 처리하기 위해 계획 기반 모델 (Plan-Based Model)이 연구되어지고 있다[4]. 이것은 사용자와 컴퓨터 간의 대화 목적을 정하고 그것을 위해 다양한 대화 계획 스키마를 적용하여 대화를 진행하는 방식이다. 그러나 이것은 사용자의 발화 자유도가 높고 어려운 작업을 위한 도메인을 처리하기 때문에 대화 관리 성능의 저하로

상용성이 떨어진다는 단점이 있다.

최근에는 단순히 특정 도메인을 위한 음성 대화 시스템을 벗어나 도메인 확장성을 향상시키기 위한 객체 기반의 구현 방식이 연구되어졌다[5]. 이것은 대화 시스템의 각 모듈을 객체로 구현하여 도메인에 무관한 공통적인 처리 부분과 도메인에 관련된 처리 부분을 나누어 도메인 확장성을 향상시킨 것이다. 그러나 이 시스템도 대화 관리를 위해 수작업으로 많은 규칙을 인코딩해야 하기 때문에 실용적인 대화 관리 시스템 개발이 힘들다.

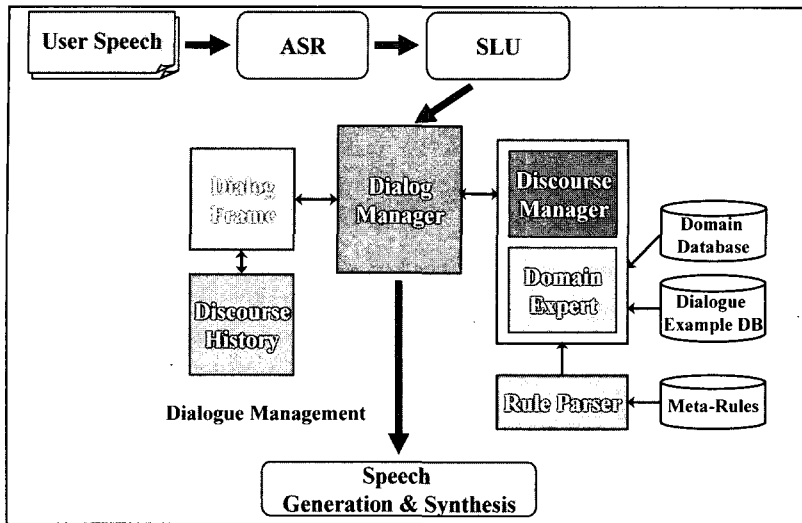
이 논문에서는 이러한 기존의 연구 방식들의 장점을 취하고 단점을 해소하기 위해 대화 예제를 이용한 상황 기반 대화 관리 기술을 제안한다. 상황 기반 대화 관리 시스템의 기본적인 아이디어는 프레임 기반의 대화 모델을 취한다. 그리고 상황 기반 방식의 대화 관리로 좀 더 유연한 대화를 진행할 수 있도록 하며 대화 모델에 필요한 규칙을 대화 예제를 통하여 자동으로 학습을 하여 수작업을 간소화하는 방법을 제시한다.

3. POSSDM 시스템 구조

본 연구에서 우리는 POSSDM (POSTECH Situation-Based Dialogue Manager)이라는 대화 관리 시스템을 개발하였다. 본 대화 관리 시스템 구현의 목적은 실용적이며 사람과 컴퓨터 간의 자연어를 이용한 유연한 대화 처리를 할 수 있고 다양한 응용 분야에 적용을 할 수 있도록 도메인 확장성을 향상시키는 것이다. 2장에서 언급한 기존 대화 시스템의 한계를 극복할 수 있는 대화 시스템을 개발하기 위한 방법론으로 McTear의 객체 지향적 대화 시스템을 응용하여 한국어 대화 시스템에 적용을 하고 대화 예제를 이용한 대화 모델링 방법을 제시한다.

전반적인 POSSDM의 구조는 <그림 1>과 같다. 여기서 대화 관리자 (Dialog Manager)는 필요한 도메인 전문가 (Domain Expert)를 호출해 SLU에서 만들어진 의미 프레임과 외부 도메인 데이터베이스, 대화 예제 데이터베이스를 연결하여 현재 사용자의 상황에 맞는 적절한 시스템 발화를 해준다. 이 논문에서 정의된 “상황 (Situation)”이라는 것은 현재 사용자의 발화와 의도, 의미 프레임, 담화 이력 등을 포함하는 전반적인 현재 대화 상태를 말한다. <그림 1>에서 보면 각각의 도메인 전문가 (Domain Expert)는 각 도메인 대화를 전담하고 현재 상황에 맞는 규칙을 파싱하여 시스템 발화를 생성한다.

여기서 상황 기반 대화 관리 시스템 (Situation-Based Dialogue Manager)은 유한 상태 기반의 상태 전이를 지양하고 현재 대화 상황에서 자유로운 상태 전이를 허락한다. 그러므로 현재 대화 턴에 규제된 대화 진행 규칙이 없고 사용자의 발화를 입력 받아 현재 대화 상황을 분석하게 된다.



<그림 1> POSSDM의 시스템 구조

4. 대화 모델링 (Dialogue Modeling)

4.1. 규칙 기반 대화 모델링

규칙 기반 대화 모델링에서는 개발자가 미리 지식 작업을 하여 일관성 있고 적용성이 큰 규칙을 인코딩을 해야 한다. POSSDM도 초기에는 이러한 작업을 하기 위해 다음과 같은 3가지 규칙을 디자인하였다.

- 상황-전략 규칙 : 현재 상황에 맞는 적절한 시스템의 의도를 규칙으로 정리.
- 제약-이완 규칙 : 데이터베이스 질의에 적합한 결과가 없을 때 질의 제약을 이완하는 규칙.
- 프레임-재시작 규칙 : 하나의 대화를 마무리 하고 새로운 대화 토막을 시작하는 기준을 규칙화 함.

그러나 효과적인 상황 기반 대화 관리를 위해서는 수작업에 의한 규칙 작성이 필요하다. 이러한 규칙이 일관성이 없거나 잘못된 경우에는 적절한 대화 진행을 하지 못하는 경우가 있다. 즉, 대화 관리 시스템에는 수많은 대화 상태가 존재한다. 이러한 대화 상태를 모두 처리할 수 있는 규칙을 수작업을 통해 만드는 것은 불가능하고 도메인이 바뀌는 경우에는 다시 디자인을 해야 하는 비용이 있다. 가령, 본 논문에서 적용한 TV 가이드 도메인에서 사용자의 의도와 의미 프레임의

조합만으로도 이론적으로 2^{13} 개의 대화 상태가 존재한다. 이러한 대화 상태를 모두 고려하여 규칙으로 처리하기에는 힘들다. 그러므로 본 논문에서 이러한 단점을 극복하기 위하여 대화 예제를 이용한 대화 모델링을 제안한다.

4.2. 대화 예제 기반 대화 모델링 (Example-Based Dialogue Modeling)

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.

4.2.1. 색인과 질의 (Indexing and Searching)

우선, 대화 모델링을 위해 대화 말뭉치로부터 자동으로 대화 예제 데이터베이스(Dialogue Example Database)를 만든다. 우리는 말뭉치의 태깅을 최소화하기 위해 SLU 모델을 훈련하기 위한 말뭉치를 재사용하였다. 즉, 프레임 기반의 대화 관리 시스템을 구축하기 위해서는 기본적으로 사용자의 발화에 대한 의미 프레임 태깅 작업이 필요하다. 여기서는 대화 모델링을 위한 추가적인 태깅 작업으로 담화 이력 벡터(Discourse History Vector)와 시스템 응답(System Response)만을 추가적으로 태깅하였다. 여기서 담화 이력 벡터는 현재 대화 상황까지 의미 프레임의 슬롯이 채워졌는지 아니지를 나타내는 이진 벡터로 표시한다.

이러한 태깅 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 검색한다면 항상 전체 일치(Exact Match)가 가능하지 않을 수 있다. 그러므로 부분 일치(Partial Match)가 가능하도록 제약을 완화하여 사용자의 의도(Dialog Act and Main Action)만을 가지고 대화 예제를 검색한다. 이것은 실제 시스템 응답이 현재 상황의 사용자 의도에 주로 의존하기 때문이다.

4.2.2. 발화 유사성 (Utterance Similarity)

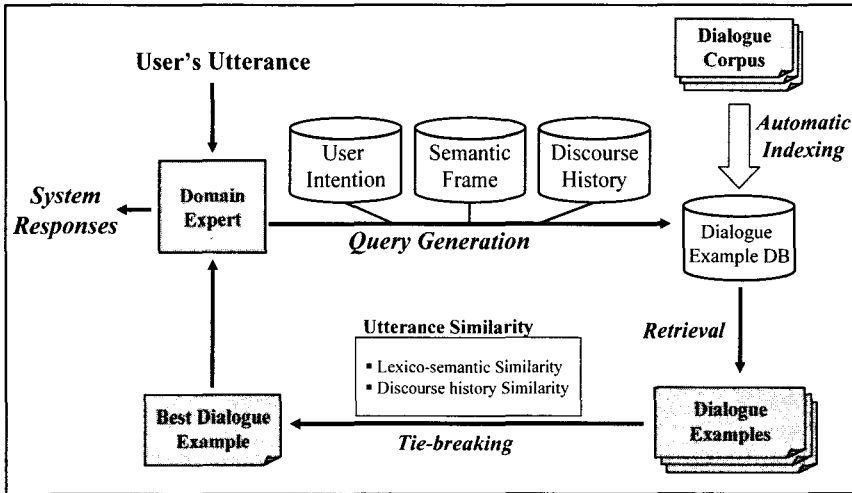
대화 예제 데이터베이스에서 현재 상황에 적합한 대화 예제들을 가져올 때, 여러 개의 예제 후보들이 생긴다. 이러한 것들 중에서 가장 적합한 예제를 찾기 위해서 발화 유사성(Utterance Similarity)을 정의하였다. 발화 유사성은 담화 이력 유사성(Discourse History Similarity)과 어휘-의미 유사성(Lexico-Semantic Similarity)의

선형 보간(Linear Interpolation)으로 정의하였다. 담화 이력 유사성은 담화 이력 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 2>는 어휘-의미 유사성의 예제를 보여준다.

User Utterance	그럼 SBS 드라마는 언제 하지?
Component Slots	[channel = SBS, genre = 드라마]
Lexico-Semantic Input	그럼 [channel] [genre] 는 언제 하지

<그림 2> 어휘-의미 유사성 예제

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 3>은 대화 예제 기반의 대화 모델링의 전반적인 전략을 나타낸다.



<그림 3> 대화 예제 기반 대화 모델링 전략

5. 실험과 결과

5.1. 실험 환경

본 연구에서는 TV 가이드 대화 말뭉치를 이용하여 대화 예제 데이터베이스를

구축하였다. 이 대화 말뭉치는 380개의 사용자 발화로 이루어진 88개의 한국어 TV 가이드 대화로 구성이 되어 있다. 각 사용자 발화는 SLU 모델을 위한 태깅이 되어져 있고 여기에 대화 예제 기반 대화 모델을 구축하기 위해 담화 이력 벡터와 시스템 응답 태깅을 하였다. POSSDM의 음성 인식기는 HTK를 기반으로 TV 가이드 도메인의 언어 모델로 훈련이 되어 있으며 이번 실험 환경에서 단어 오인식률(WER)은 15.3%이다. 그리고 POSSDM의 SLU 모듈은 TV 가이드 도메인을 대해서 <표 1>과 같은 성능을 보인다[6]. 여기서 화행 (Dialog Act)는 일반적으로 문형 정보를 나타내며 상황에 따라서는 반드시 문형과 일치하지 않다. 그리고 행위 (Main Action)는 실제 사용자가 그 도메인 내에서 원하는 행위를 나타내는 의미적인 정보이다. 구성성분 요소(Component Slot)은 실제 사용자 발화에 나타나는 도메인에 대한 개체 정보이다. 이것은 슬롯과 그것에 해당하는 값을 추출하게 된다. <그림 2>의 사용자 발화를 예로 들어서 설명을 하면 화행은 “wh-question”이 되고, 행위는 “search_start_time”이 된다. 그리고 사용자가 말한 구성성분 요소는 채널(Channel)과 장르(Genre)에 대응하는 값으로 “SBS”와 “드라마”가 된다.

<표 1> SLU 성능

Slot Type	Textual Input (WER 0.0%)	Spoken Input (WER 15.3%)
Dialogue Act	95.33	85.34
Main Action	93.50	81.78
Component Slot	90.85	80.12

5.2. 대화 실험

우선, 대화 예제 기반의 대화 모델링의 성능을 측정하기 위해 대화 예제 일치 비율(Example Matching Rate, EMR)과 그것에 따른 턴 성공 비율(Success Turn Rate, STR)을 측정하였다. EMR은 사용자 발화 입력에 대한 대화 예제 일치의 평균 일치 비율로 전체 일치와 부분 일치로 나누어서 측정하였다. 그리고 그것에 따라 시스템 발화가 적절한지를 체크하기 위해 STR을 측정하였다. 이것을 실험하기 위해 5명의 사용자에게 TV 가이드 도메인에 대한 10개의 임의의 텍스트 입력을 넣도록 하였으며 각 턴마다 시스템 발화의 적절성을 체크하도록 하였다. <표 2>는 각 예제 일치 분류에 따라 EMR과 STR을 나타내었다. 실험 결과를 보면 대부분의 대화는 전체 일치나 부분 일치로 진행이 가능하며 원만한 대화 진행이 되는 것을 알 수 있다.

<표 2> EMR과 STR

Example Match Type	EMR	STR
Exact Match	0.42	0.90
Partial Match	0.52	0.73
No Example	0.06	0.33

또한, POSSDM의 대화 관리 시스템의 성능을 측정하기 위해 텍스트 입력과 음성 입력을 둘 다 고려하여 PARADISE 성능 측정법을 바탕으로 평가를 하였다[4]. 우선 사용자 만족도 (User Satisfaction)을 음성 인식률 (Mean Recognition Accuracy, MRA)와 대화 성공률(User Perception of Task Completion Rate, TCR), STR의 값들을 1/3의 가중치를 주어서 선형 보간법을 이용하여 정의하였다. 우선 5명의 사용자에게 5개의 다른 TV 가이드 시나리오를 주고 그것을 수행하도록 하였다. 그리고 각각의 턴과 대화에 대해서 시스템 발화의 정확성과 성공 유무를 체크하도록 하였다. <표 3>은 POSSDM의 대화 성능 평가를 나타내었다.

<표 3> POSSDM의 대화 성능

Evaluation	Textual Input	Spoken Input
TCR	0.92	0.76
STR	0.88	0.65
MRA	1.00	0.85
User Satisfaction	0.93	0.75

6. 결론

본 논문에서는 상황 기반 대화 관리 시스템과 대화 예제를 이용한 대화 모델링 방법을 제안하였다. 대화 예제를 통하여 대화 관리를 위한 규칙을 자동으로 대화 코퍼스로부터 만들어 효율적이고 실용적인 음성 대화 시스템 개발 방법을 제안했다. 본 연구에서는 TV 가이드라는 하나의 영역에 대한 대화 관리 시스템을 구축하였지만 본 연구에서 제시한 방법으로 다른 도메인을 위한 대화 관리 시스템도 쉽게 개발할 수 있을 것으로 본다. 향후 계획은 이러한 방법론을 다른 도메인에 적용을 하여 여러 개의 도메인을 처리할 수 있는 음성 대화 시스템 개발에

이용할 것이다.

참 고 문 헌

- [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] McTear, "Modelling spoken dialogues with state transition diagrams : Experiences with the CSLU toolkit," in *Proceedings of the 5th International Conference on Spoken Language Processing*, Vol. 4, pp. 1223-1226, 1998.
- [3] O. Lemon, A. Gruenstein et al., "Multi-tasking and collaborative activities in dialogue systems," in *Proceedings of the 3rd SIGDIAL Workshop on Discourse and Dialogue*, pp. 113-124, 2002.
- [4] J. Chu-Carroll, "MIMIC: An adaptive mixed initiative spoken dialogue system for information queries," in *Proceedings of the 6th Applied Natural Language Processing Conference*, pp. 97-104, 2000.
- [5] I. O'Neil, P. Hanna et al., "Implementing advanced spoken dialogue management in Java", *Speech Communication*, Vol. 54, no. 1, pp 99-124, January 2005.
- [6] J. Eun, C. Lee, and G. Lee, "An Information extraction approach for spoken language understanding," in *Proceedings of the 8th International Conference on Spoken Language Processing*, pp. 2145-2148, 2004.
- [7] D. J. Litman and S. Pan, "Empirically evaluating an adaptable spoken dialogue system", in *Proceedings of the 7th International Conference on User Modeling*, pp. 55-64, 1999.

접수일자: 2005년 11월 23일

게재결정: 2005년 11월 23일

▶ 이청재(Cheongjae Lee)

주소: 790-784 경상북도 포항시 효자동 포항공과대학교 컴퓨터공학과

소속: 포항공과대학교(POSTECH) 지능소프트웨어 연구실

전화: 054) 279-5581

E-mail: lcj80@postech.ac.kr

▶ 정상근(Sangkeun Jung)

주소: 790-784 경상북도 포항시 효자동 포항공과대학교 컴퓨터공학과

소속: 포항공과대학교(POSTECH) 지능소프트웨어 연구실

전화: 054) 279-5581

E-mail: hugman@postech.ac.kr

▶ 이근배(Gary Geunbae Lee) : 교신저자

주소: 790-784 경상북도 포항시 효자동 포항공과대학교 컴퓨터공학과

소속: 포항공과대학교(POSTECH) 지능소프트웨어 연구실

전화: 054) 279-5581

E-mail: gblee@postech.ac.kr