

강건한 음성 대화 시스템을 위한 담화분석 기술

(Discourse Analysis for Robust Spoken Dialogue System)

이 충 희 [†] 오 효 정 ^{**}
(Chung-Hee Lee) (Hyo-Jung Oh)

장 명 길 ^{***} 서 영 훈 ^{****}
(Myung-Gil Jang) (Young-Hoon Seo)

요약 지시대명사와 같은 조용어(anaphora)의 본래 단어나 구를 선행사라고 지칭하며, 음성 대화 중에는 선행사에 대한 생략과 대용어 사용이 빈번히 발생한다. 또한 언어 현상들은 문맥을 보지 않으면 이해될 수 없는 것들이 많다는 것이 담화분석의 기본 가정이므로, 생략 및 대용어 복원은 담화분석에서 매우 중요한 역할을 한다. 본 논문에서는 대용어와 생략어 복원에 기반해서 대화 레벨에서의 강건성을 향상시킨 음성 기반 대화 시스템을 제안한다. 제안된 방법의 적절성과 효과는 TV 도메인에서 평가되었다.

키워드 : 담화분석, 대용어복원, 생략어복원

Abstract Elliptical and anaphoric utterances occur frequently during spoken dialogue. Because discourse analysis rests on the basic premise that linguistic items cannot be understood without reference to the context, ellipsis and anaphora resolution plays an important role in

· 이 논문은 2010 한국컴퓨터종합학술대회에서 '강건한 음성 대화 시스템을 위한 담화분석 기술'의 제목으로 발표된 논문을 확장한 것임

[†] 정 회 원 : 한국전자통신연구원 지식마이닝팀 선임연구원
forever@etri.re.kr

^{**} 비 회 원 : 한국전자통신연구원 지식마이닝팀 선임연구원
ohj@etri.re.kr

^{***} 비 회 원 : 한국전자통신연구원 지식마이닝팀장
mgjang@etri.re.kr

^{****} 종신회원 : 충북대학교 컴퓨터공학과 교수
yhseo@chungbuk.ac.kr

논문접수 : 2010년 8월 12일

심사완료 : 2010년 9월 16일

Copyright©2010 한국정보과학회 : 개인 목적이나 교육 목적인 경우, 이 저작물의 전체 또는 일부에 대한 복사본 혹은 디지털 사본의 제작을 허가합니다. 이 때, 사본은 상업적 수단으로 사용할 수 없으며 첫 페이지에 본 문구와 출처를 반드시 명시해야 합니다. 이 외의 목적으로 복제, 배포, 출판, 전송 등 모든 유형의 사용행위를 하는 경우에 대하여는 사전에 허가를 얻고 비용을 지불해야 합니다.

정보과학회논문지: 컴퓨팅의 실제 및 레터 제16권 제10호(2010.10)

discourse analysis. In this paper, we present a spoken dialogue system improving the robustness at dialogue level based on discourse analysis, such as anaphora and ellipsis resolution. The applicability and effectiveness of the proposed method is evaluated in the TV domain.

Key words : Ellipsis resolution, Anaphora resolution, Discourse Analysis

1. 서론

대화에서는 같은 단어가 다양한 방법으로 언급되고 효율성 측면에서 사용자들은 이전 문맥에서 언급된 단어나 구는 반복하지 않는 경향이 있기 때문에 지시어 복원은 음성 대화에서 중요한 이슈이다. 또한 한국어에서는 영어와 달리 생략이 매우 빈번히 발생하므로 생략어 복원 또한 음성 대화시스템을 위해서는 필수 요소이다.

실제 환경에서 대화 시스템 강건성을 향상시키기 위해서 많은 이론과 접근이 이루어졌다. [1]은 다양한 도메인에서의 사용자 요구를 다룰 수 있는 멀티도메인 음성대화시스템을 개발하였는데, 사용자 입력에 오류가 있는 경우에도 알맞은 행동을 할 수 있도록 설계되었다. [2]는 음성인식 오류가 존재할 때 대화의 강건한 해석에 대한 이슈를 다루고 있고, 통계적 오류 후처리, 구문과 의미에 기반한 강건한 구문분석, 그리고 대화 문맥의 광범위한 사용을 복합적으로 활용한다. [3]은 음성인식, 구문분석, 대화 레벨에서의 강건성 향상 방법을 기술하였다. 이런 연구들은 주로 음성인식 오류에 초점을 맞춰서 강건성을 다루고 있고, 일부 시스템은 대화 레벨에서의 강건성 향상을 이루었다.

본 논문은 생략어 및 대용어 복원을 통해 대화 레벨에서 강건성을 향상시킨 음성기반 시스템을 제안한다. 2장에서 음성기반 대화 시스템의 전체적인 구성을 알아보고, 3장에서 대화 관리 모델을 설명한다. 4장과 5장에서는 대용어와 생략어 복원 방법에 대해서 자세히 설명하고, 6장에서 실험 결과를 보고 7장에서 요약한다.

2. 음성 기반 대화 시스템

우리가 제안하는 TV 도메인에서의 음성기반 대화 시스템은 그림 1과 같다.

자동음성인식을 위해 ESTk(ETRI Speech Toolkit, [4])를 사용하였고, 음성 언어 이해를 위한 개념 추출 접근 방법[5]을 사용해서 대화 관리에 필요한 필수 요소들을 추출한다. 의미 표현은 프레임 형태의 슬롯 구조를 가짐으로써 목표 지향적인 화행 및 대화 관리를 위한 작업 구조를 직접적으로 반영한다. 본 논문은 그림 1의 4개 파트 중에서 입력 파트 이후의 담화 분석 과정을 포함하는 제어 파트에 초점을 맞춰서 설명한다.

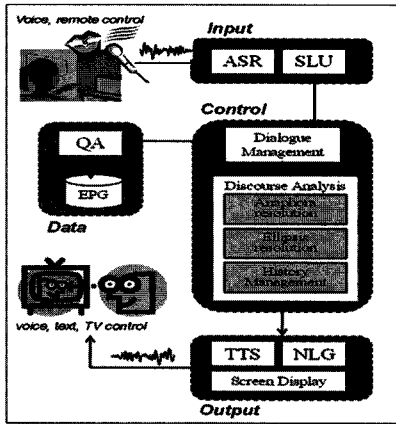


그림 1 음성기반 대화시스템 구성도

3. 대화 관리

3.1 대화 모델링

우리는 대화 모델링을 위해서 내부 데이터 구조로 프레임 기반 의미 표현을 이용하였다.

그림 2는 "오늘 나이트라인은 언제 하지"라는 사용자 발화에 대한 프레임 정보를 보여준다. EPG 도메인을 위해서 프로그램, 채널, 장르, 날짜 등의 20개의 프레임 슬롯을 만들었다. 음성 언어 이해 기술(spoken language understanding: SLU)로부터 추출된 정보는 슬롯-값 쌍으로 표현되고, 프레임은 CUR와 PREV 등의 History 상속 태그와 같은 담화 분석 결과도 포함한다. History 상속 태그는 해당 슬롯 값이 이전 문맥으로부터 상속되었는지 여부를 나타낸다. Action 태그는 시스템이 처리할 다음 동작과 관련된 정보를 나타낸다. 그림 2에서 User2의 action 태그는 사용자가 해당 프로그램의 시작 시간을 알고 싶어 한다는 정보를 가리킨다("?" 태그로 표현됨).

<ul style="list-style-type: none"> User1: SBS로 돌려봐 <table border="1"> <thead> <tr> <th>Action</th> <th>History</th> <th>Slot</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td></td> <td>CUR</td> <td>channel</td> <td>SBS</td> </tr> </tbody> </table>	Action	History	Slot	Value		CUR	channel	SBS												
Action	History	Slot	Value																	
	CUR	channel	SBS																	
<ul style="list-style-type: none"> System1: SBS 8시 뉴스입니다. <table border="1"> <thead> <tr> <th>Action</th> <th>History</th> <th>Slot</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>*</td> <td>CUR</td> <td>program</td> <td>8 news</td> </tr> <tr> <td>*</td> <td>PREV</td> <td>channel</td> <td>SBS</td> </tr> </tbody> </table>	Action	History	Slot	Value	*	CUR	program	8 news	*	PREV	channel	SBS								
Action	History	Slot	Value																	
*	CUR	program	8 news																	
*	PREV	channel	SBS																	
<ul style="list-style-type: none"> User2: 오늘 나이트라인 언제 하지? <table border="1"> <thead> <tr> <th>Action</th> <th>History</th> <th>Slot</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>?</td> <td>CUR</td> <td>start time</td> <td></td> </tr> <tr> <td></td> <td>CUR</td> <td>date</td> <td>2007.06.29</td> </tr> <tr> <td></td> <td>CUR</td> <td>program</td> <td>Night-Line</td> </tr> <tr> <td></td> <td>PREV</td> <td>channel</td> <td>SBS</td> </tr> </tbody> </table>	Action	History	Slot	Value	?	CUR	start time			CUR	date	2007.06.29		CUR	program	Night-Line		PREV	channel	SBS
Action	History	Slot	Value																	
?	CUR	start time																		
	CUR	date	2007.06.29																	
	CUR	program	Night-Line																	
	PREV	channel	SBS																	
<ul style="list-style-type: none"> System2: SBS에서 오늘 밤 12시에 방송될 예정입니다. <table border="1"> <thead> <tr> <th>Action</th> <th>History</th> <th>Slot</th> <th>Value</th> </tr> </thead> <tbody> <tr> <td>*</td> <td>CUR</td> <td>start time</td> <td>24:00</td> </tr> <tr> <td></td> <td>PREV</td> <td>date</td> <td>2007.06.29</td> </tr> <tr> <td></td> <td>PREV</td> <td>program</td> <td>Night-Line</td> </tr> <tr> <td>*</td> <td>PREV</td> <td>channel</td> <td>SBS</td> </tr> </tbody> </table>	Action	History	Slot	Value	*	CUR	start time	24:00		PREV	date	2007.06.29		PREV	program	Night-Line	*	PREV	channel	SBS
Action	History	Slot	Value																	
*	CUR	start time	24:00																	
	PREV	date	2007.06.29																	
	PREV	program	Night-Line																	
*	PREV	channel	SBS																	

*: *Italic words indicate key concepts such as channel, program, etc.*

그림 2 대화 및 의미 프레임 예제

3.2 담화 분석

담화 분석 과정은 3개로 구분된다: 1) 대용어 복원 파트, 2) 생략어 복원 파트, 그리고 3) 히스토리 관리 파트이다.

대용어는 이전에 언급된 개념, 즉 선행사를 가리키는 단어나 구를 말하며, 우리는 CRF 분류기를 이용해서 복원 작업을 수행한다. 4장에서 대용어 복원 과정이 자세히 기술된다. 생략어가 포함된 말하는 TV 도메인에서 자주 발생하는데 그 이유는 사용자들이 이전에 언급했던 내용을 반복하기 싫어하기 때문이다. 본 논문에서는 이런 생략어 복원을 이전 포커스 프레임의 상속 여부로 다룬다. 이 부분에 대해서는 5장에서 자세히 설명하겠다.

히스토리 관리 기능은 대화 중에 나타난 중요한 정보를 관리하는 기능을 수행하며, 대상 정보로는 화행, 개념열, 대용어정보, 객체 등이 있다. 객체 관리의 히스토리 관리의 중요 작업 중 하나로서 TV 도메인에서는 TV 프로그램을 객체로 정의해서 객체 스택에서 유지, 관리한다. 그림 3은 객체 관리 예를 보여준다.

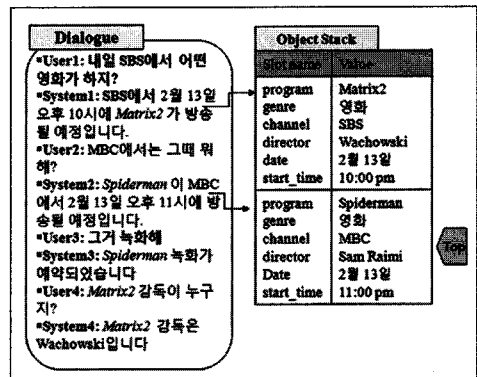


그림 3 객체관리 예제

4. 통계 학습 기반 대용어 복원

대용어 복원은 대용어 표현 인식과 선행사 할당 과정으로 구성된다. 대용어 표현 인식은 대용어 표현의 경계를 인식하고 대용어 태그를 할당하는 것으로 이루어진다. 우리는 600셋의 대화 시나리오를 분석해서 7개의 대용어 태그를 정의하였다(표 1).

- 대용어 표현 인식을 위해 사용된 자질은 다음과 같다.
- 형태소 및 품사: 현재 발화에 나타난 형태소 및 품사정보 (e.g. "[SBS/mn]에서 [하/pv]는 [거/nb]")
 - 접미사: 특정 대용어 표현 인식에 유용한 일부 접미사 (e.g. "4번째", "2번",.....)
 - 사전: 모호성이 없어서 사전에 의해 직접적으로 다루어질 수 있는 표현들 (e.g. N_program_this : "지금 보고 있는 프로")
 - 현재 발화의 화행 (CDA): 현재 발화의 화행 정보

표 1 대용어 태그 및 예제

Tag	Example
N_program	이전에 나타난 TV 프로그램 (Ex. <u>그 프로그램</u> 녹화해)
N_program_this	사용자가 현재 보고 있는 프로그램 (이전 발화에서도 언급된 적이 없다) (Ex. 이게 뭐지)
N_channel	이전에 나타난 채널 (Ex. 저녁에 <u>그 채널</u> 에서 어떤 프로그램이 하지)
N_channel_this	사용자가 현재 보고 있는 채널 (이전 발화에서도 언급된 적이 없다) (Ex. 오늘 여기서 CSI가 하나)
N_date	이전에 나타난 날짜 (Ex. <u>이날</u> 어떤 영화가 하지)
N_time	이전에 나타난 시간 (Ex. <u>그때</u> MBC에서 뭐가 하지)
N_list	검색된 프로그램 리스트 중 1개 프로그램 (Ex. <u>새 번째 거</u> 자세히 보여 줘)

로 화행은 사용자 의도를 나타내는 것으로, 우리는 TV 도메인에서 사용할 수 있는 것으로 45개를 정의하였다 (e.g. search_program: "오후 2시 이후에 SBS에서 뭐가 해")

- 이전 발화의 화행 (PDA): 이전 발화의 화행

아래 예제는 특정 대화의 예이며, 각 발화에서 추출된 자질들과 대용어 표현 인식 결과를 보여준다.

-User1: 오늘 골프채널에서 뭐 하지?
(DA: search_program)

-System1: 15 개의 프로그램이 검색되었습니다

- 1) PGA Tour golf school (J Golf)
- 2) 2007 LPGA HANA bank championship (J Golf)
- 3) Live lesson 70 (J Golf)
- 4) 2007 PGA Price Electronics Open (SBS Golf)
- 5) Woman in golf (X Golf)

-User2: SBS 골프 채널에서 하는 거 알람 설정해
(DA: alarm)

Features:

- 형태소 & 품사: SBS/nn Golf/nn 채널/nn+에서/jo 하/pv+는/em 거/nx 알람/nn 설정/nn+하/xvn+어/em
- CDA: alarm
- PDA: search_program

대용어 표현 인식 결과:

- [SBS 골프 채널에서 하는 거:N_program] 알람 설정해

대용어 표현 인식의 두 번째 단계는 인식된 대용어에 알맞은 태그를 할당하는 것이다. 입력 파트의 SLU 모듈 중 한 개인 개념열 인식기는 일종의 개체명 인식기이며 19개의 개념열을 인식한다. 개념열은 대화의 핵심을 표현하기 때문에 대부분의 사용자는 그것을 반복하게 되고 대용어로 표현한다. 그래서 우리는 자주 사용되는 4가지 개념열을 선행사 후보로 사용한다(PROGRAM, CHANNEL, DATE, TIME).

대화 중 발생하는 모든 선행사는 그들의 유형에 따라

특정 대용어 스택에 저장되며, 선행사 할당 모듈은 인식된 대용어에 가장 알맞은 선행사를 스택을 이용해서 할당한다. 대용어 표현은 2 가지 유형으로 구분된다.

- 대명사: "그거", "저거", "이거"

- 명사구: "그 SBS 프로그램", "오후 3시에 하는 거", "MBC에서 하는 거", "스포츠 채널에서 오후 2시 10분에 하는 프로그램" 등

사용자들은 대명사 형태의 대용어를 사용할 때는 주로 가장 최근의 선행사를 가리키는 경우가 많다. 그러므로 대명사 유형에 대해서는 straightforward preference 전략을 사용한다. 즉, 선행사로 가장 최신 후보를 지정한다. 예를 들어, 그림 3의 System3까지 대화를 한 경우에 2개의 선행사가 존재하며(Matrix2와 Spiderman), 스택에 저장된 순서에 따라 User3 발화에 나타난 대용어 "그거:N_program"는 Spiderman으로 복원된다. 반면에 명사구로 된 대용어는 때때로 해당 표현 안에 추가 정보를 가지고 있다. 예를 들면, "오후 6시 이후에 SBS에서 하는 프로그램"이라는 표현은 "N_program" 태그를 가지며, 표현 내부에 2개의 제약 조건으로 "SBS"와 "오후 6시 이후"를 가지므로 이런 경우에는 시스템이 6시 이후에 하는 SBS 채널에서 하는 프로그램으로 선행사를 먼저 필터링하고, 남은 후보들 중에서 가장 최근에 언급된 것으로 대용어를 복원한다.

5. 문맥 상속에 기반한 생략어 복원

대화 관리기는 포커스 프레임에 기반해서 시스템 동작을 결정한다. 대부분의 생략된 선행사는 이전 문맥, 즉 이전 포커스 프레임에 나타난다. 그러므로 우리는 생략어 복원 문제를 문맥 상속 문제로 다루며, 상속 문제에 대한 해결방법을 연구하였다. 예를 들어, 그림 2의 User2 발화 "오늘 나이트라인 언제 하지는 생략어가 포함된 문장이다. 완벽한 문장인 "SBS에서 오늘 나이트라인 언제 하지를 만들기 위해서는 "SBS" 채널 정보가 상속되어야 한다.

생략어 복원 과정은 3 단계로 구성된다.

- 1단계: 생략어 존재 확인

- 2단계: 프레임 슬롯 검토

- 3단계: 필수 슬롯이지만 값이 없는 슬롯의 값 생성
1단계에서 포커스 프레임의 이전 슬롯 값의 상속 여부를 결정하며, 이를 위해 "clear"와 "not clear"의 이전 분류를 위해 ME 모델을 적용하였다. "not clear" 태그는 현재 발화에 생략어가 있다는 것을 나타내며, 이런 경우에 시스템은 이전 포커스 프레임으로부터 슬롯 값들을 상속한다.

ME 모델에서 사용되는 5가지 자질은 다음과 같다.

- 형태소 & 품사: 4장과 동일

- 개념열 [6]: 일종의 개체명 (e.g. 프로그램, 채널, 시간, 날짜,.....)
- 대용어: 현재 발화에 나타난 대용어 태그
- CDA: 4장과 동일
- PDA: 4장과 동일

ME 모델과 CRF 모델을 비교해 보면, CRF는 이전 관찰 결과가 모델에 반영되지만, ME는 각 상태를 독립적으로 고려해서 결정한다. 대용어는 이전 결과가 현재에 영향을 주지만 상속은 이전 결과가 현재 결과에 영향을 주지 않으므로 각각 CRF와 ME를 사용하였다.

2단계는 1단계 결과가 "not clear" 인 경우에만 동작하며, 포커스 프레임의 상속된 값들 중에서 불필요한 것들을 제거하는 것이다. 대화 시스템은 포커스 프레임의 슬롯 값들에 따라 다음 행동을 결정한다. 예를 들어 다음 행동이 TV 프로그램을 DB로부터 찾는 거라면, QA 모듈은 포커스 프레임에 기반해서 DBQ를 생성한다. 사용자 의도와 관련 없는 정보가 프레임에 있다면 검색결과에 오류가 발생하므로, 2단계는 매우 중요한 역할을 한다. 다음 대화는 2단계가 필요한 경우를 보여준다.

User1: "지금 MBC에서 뭐해?"
 System1: "MBC에서 주몽이 하는 중입니다."
 User2: "EBS에서는?"
 System2: "EBS에서는 텔레토비가 방송 중입니다."

System1의 포커스 프레임은 주몽 프로그램과 관련된 여러 슬롯 값으로 [program:주몽], [channel:MBC], [start_time:now], [end_time:xxx], [program code:P0010], [actor:xxx], [director:xxx]를 가지고 있다. User2에서 1단계 결과는 "not clear"이며, 그에 따라 System1 포커스 프레임의 모든 슬롯들이 상속되고 User2 발화에 나오는 "EBS" 채널 정보만 업데이트 된다. User2 발화 의도에 의하면, channel과 start_time 슬롯만 필수이고 나머지 슬롯은 유해한 정보이므로 2단계에서 제거된다. 2단계는 23개 규칙에 기반해서 동작하며, 규칙들은 대화 시나리오 집합에 기반해서 수작업으로 구축됐다. 다음 예제는 규칙 1개를 보여준다.

Rule1
Action
 - clear all inherited slots except program and genre slot
Condition
 - Dialogue Act: search_program
 - Current concept sequence
 - TIMES is only exist

마지막 단계인 3단계는 대화 중에 나타나지 않지만 필수적인 슬롯 정보를 새로 생성한다. 우리는 필수 슬롯을 올바른 시스템 동작을 위해서 꼭 있어야 하는 슬롯으로 정의하였다. 필수 슬롯은 도메인과 화행에 의존적인데, TV 도메인에서는 date, start_time, channel, program 등이 있다. 만약 필수 슬롯이 이전 포커스 프레임에 있다면, 그것들은 1단계와 2단계에서 상속된다. 이전 프

임에 없는 경우에는 3단계에서 새로 생성되어야 하며, 3단계는 규칙에 기반해서 이루어진다. 대화에 안 나타나는 필수 슬롯은 대부분 TV 제어 명령과 관련된 것이다. 다음 대화는 3단계가 동작하는 예를 보여준다.

- User1: MBC에서 뭐 해
- System1: 뉴스데스크가 MBC에서 방송 중입니다
- User2: 채널 돌려

User2의 화행은 "change_channel"이고, 필수 슬롯으로 channel과 start_time이 필요하다. 즉, "채널 돌려"의 완전한 문장은 "MBC로 채널 돌려, 지금"이다. 그에 따라, 시스템은 이전 포커스 프레임에서 "channel:MBC"를 상속하고(1,2단계), "지금"을 위해 start_time 슬롯을 새로 생성해야 한다(3단계). 우리는 3단계를 위해 휴리스틱 정보를 기반으로 26개 규칙을 만들었고, 시스템을 개발하면서 점진적으로 확장하는 중이다. 다음은 규칙 1개의 예이다.

```
IF
  dialogue act = "show_next_page",
  there is not date and start_time
THEN
  Generate a new slot [start_time: "now"]
```

6. 실험 및 평가

제한한 방법의 효과를 검증하기 위해서 표 2에 보이는 평가셋에 기반해서 실험하였다.

표 2 평가셋

Evaluation set	ES1(freestyle)	ES2(restricted)
# of scenarios	225	100
# of utterance	2,282	642

ES1은 일반 사용자를 대상으로 제약 없이 자유롭게 만들어졌으며, out-of-grammar(OOG)와 out-of-vocabularies(OOV)가 존재하고 심지어 out-of-function(OOF)까지 존재한다. 반면 ES2는 상용화를 위한 만들어진 260개 발화 패턴에 기반해서 제한된 발화로 이루어졌다. 이 패턴들은 사용자로부터 수집된 말뭉치에서 일반적으로 가장 많이 사용되는 발화에 기반하였으므로 TV 도메인에서 나타날 수 있는 대부분의 발화를 커버할 수 있다.

실험은 대용어와 생략어 복원 시스템의 효과를 평가하기 위한 것으로, 평가 척도는 F-measure를 사용하였다.

표 3 대용어 및 생략어 복원 성능평가 결과

	Test set	Precision	Recall	F-score
Anaphora resolution	ES1	89.61%	81.18%	85.19%
	ES2	94.0%	94.95%	94.47%
Ellipsis resolution	ES1	89.8%		89.8%
	ES2	99.42%		99.42%

대용어 복원 평가는 7개 대용어 표현 인식률과 선행 사 할당의 정확성을 평가하였으며, 2개가 모두 맞는 경우에만 맞춘 것으로 고려하였다. 생략어 복원 평가는 생략어 존재 인식, 불필요한 슬롯 제거, 그리고 필수슬롯 생성의 성능을 평가하는 것으로, 평가는 시스템이 생성한 프레임과 정답 프레임의 일치 여부로 판단하였다. 표 3은 평가 결과를 나타낸다.

평가는 ASR 결과가 아니라 수작업으로 필사한 발화를 이용하였다. 즉, 음성인식 오류가 포함되어 있지 않다. 예측한 대로 ES2 성능이 ES1보다 높은 성능을 보였다. 이전 연구에서는 Tetreault and Allen[7]가 담화 분할 정보를 이용한 대명사 복원 알고리즘을 평가해서 68.6% F-score를 보였다. Nielsen[8]은 동사구 생략 인식 시스템을 개발하였고 72% F-score를 얻었다. 도메인과 언어, 그리고 시스템 기능이 다르기 때문에 본 논문에서 제안한 시스템과 이전 연구를 직접 비교하는 것은 어렵다. 동일한 데이터에 대한 합리적인 baseline과 성능이 비교된 것은 아니지만 [7]과 [8]에 비해서 표3의 결과는 제안된 시스템이 유용하고 실제 대화 시스템에서 적용될 수 있음을 보여준다.

우리는 또한 음성 기반 대화 시스템을 평가하였다. 평가 척도로는 3가지 척도를 사용하였다 1) 문장 인식율(Sentence Recognition Ratio:SRR), 2) 대화 성공률(Dialogue Success Ratio:DSR), 3) 작업 성공률(Task Completion Ratio:TCR). 여기서 SRR은 사용자 발화 중 정확히 인식된 발화의 비율로서 음성인식 정확률에 대한 평가 척도이다. DSR은 사용자 발화에 대한 시스템의 정확한 응답 비율로 시스템 동작의 정확성을 평가하는 척도이며, 검색된 프로그램의 정확성이나 올바른 TV 제어에 대한 성능을 평가한다. TCR은 특정 목적이 주어진 상황에서 시스템이 최종적으로 목적을 달성하였는지를 평가한다. 사용자 만족도(User Satisfaction Ratio:USR)는 SRR, DSR, TCR의 평균값이며, 시스템 성능의 최종척도로 사용하였다. 표 4는 그 결과를 보여준다. 예상대로 ES1의 경우에 OOG와 OOV 문제 때문에 SRR의 성능이 매우 낮았다. DSR과 TCR 평가에서는 대화 시스템만을 독립적으로 평가하기 위해서 음성인식 오류를 모두 빠르게 수정하였다. 과거 연구에 비교했을 때, JUPITER[9]는 2,000개 단어를 사용해서 약 80% 발화에 대해서 정답을 제시하였고, 그에 비해 우리는 USR로 ES1의 경우에 73%, ES2의 경우에 90.4% 성능

을 보였다. 이 결과로부터 제안된 대용어 복원 및 생략어 복원 기술이 실제로 음성 대화 시스템의 성능에도 매우 긍정적인 효과를 주고, 실제 상용화 목적으로도 활용할 수 있음을 보여 주었다.

7. 결론 및 향후 연구

본 논문은 대용어 및 생략어 복원에 기반해서 성능이 개선된 대화 시스템을 제안하였다. 생략어와 대용어가 존재하는 발화는 음성 대화에서는 자주 발생하기 때문에 대화 시스템이 제대로 동작하기 위해서는 생략어와 대용어 복원은 필수적이고, 평가 결과에 의해 그것이 증명되었다.

다음 연구로는 TV 도메인이 아닌 다른 영역에서의 대화 시스템에 적용해 볼 예정이며, 추가적인 연구로 담화 분석 과정 중에 음성 인식 오류를 복원하는 방법에 대해서 연구할 예정이다.

참고 문헌

- [1] M. Maragoudakis, A. Thanopoulos, and N. Fakotakis, "MeteoBayes: effective plan recognition in a weather dialogue system," *IEEE Intelligent Systems*, vol.22, pp.67-77, 2007.
- [2] K. Komatani, N. Kanda, M. Nakano, et al., "Multi-domain spoken dialogue system with extensibility and robustness against speech recognition errors," *Proc. 7th SIGdial Workshop on Discourse and Dialogue*, 2006.
- [3] J. F. Allen, B. W. Miller, F. K. Ringger, et al., "A robust system for natural spoken dialogue," *Proc. 34th annual meeting on Association for Computational Linguistics*, pp.62-70, 1996.
- [4] S. J. Goldwater, E. O. Bratt, J. M. Gawron, et al., "Building a robust dialogue system with limited data," *Proc. ANLP/NAACL 2000 Workshop on Conversational systems*, vol.3, pp.61-65, 2000.
- [5] J. Park, S. Lee and S. Kim, "Keyword spotting for far-field speech input by categorical fillers and speech enhancement," *Proc. 22nd Speech Communication and Signal Processing*, 2005.
- [6] C. Lee, J. Eun, M. Jeong, et al., "A multi-strategic concept-spotting approach for robust understanding of spoken Korean," *ETRI journal*, vol.29, no.2, pp.179-188, 2007.
- [7] J. R. Tetreault and J. F. Allen, "Dialogue structure and pronoun resolution," *Proc. DAARC*, 2004.
- [8] L. A. Nielsen, "Verb phrase ellipsis detection using automatically parse text," *Proc. COLING*, pp.1093-1099, 2004.
- [9] V. Zue, S. Seneff, J.R. Glass, et al., "JUPITER: A telephone-based conversational interface for weather information," *IEEE Trans. Speech and Audio Processing*, vol.8, no.1, pp.85-96, 2000.

표 4 음성 대화 시스템 실험 결과

	SRR	DSR	TCR	USR
ES1	53.5%	77.8%	87.7%	73.0%
ES2	78.3%	96.3%	96.7%	90.4%