강건한 음성 대화 시스템을 위한 담화분석 기술

이충희○ 오효정 장명길 서영훈**
한국전자통신연구원 지식마이닝연구팀
**충북대학교 전자정보대학 컴퓨터공학과
{forever,ohj,mgjang}@etri.re.kr, **yhseo@chungbuk.ac.kr

Discourse Analysis for Robust Spoken Dialogue System

ChungHee Lee○ HyoJung Oh MyungGil Jang YoungHoon Seo**
Electronics and Telecommunications Research Institute (ETRI)
**Chungbuk National University

1. 서 론

지시대명사와 같은 조응어(anaphora)의 본래 단어나 구를 선행사라고 지칭하며, 음성 대화 중에는 선행사에 대한 생략과 대용어 사용이 빈번히 발생한다. 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 개선된 대용어와 생략어 복원에 기반해서 대화 레벨에서의 강건성을 향상시킨 음성 기반 대화 시스템을 제안한다.

2. 본 론

대화 모델링을 위해서 우리는 내부 데이터 구조로 프레임 기반 의미 표현을 이용하였고 [1], 프레임은 Action, History, Slot, Value로 구성된다. 사용자 질문의 언어분석 결과로부터 추출된 정보는 slot-value 쌍으로 표현되며 TV 도메인을 위해서 프로그램, 채널, 장르 등의 20개 프레임 slot을 만들었다. History는 해당 slot value의 이전 문맥으로부터 상속 여부를 나타내는 CUR과 PREV 값, 이전/현재 발화에는 없지만 필요해서 새로 생성된 DEF 값을 가진다. Action 태그는 시스템이 처리할 다음 동작과 관련된 정보를 나타내며 ‘ ?’ 와 ‘ *’ 값을 가진다. ‘ ?’ 는 시스템이 찾아야 할 정보를 나타내고, ‘ *’ 는 시스템이 사용자에게 발성할 정보를 나타낸다. 예를 들어 “ 오늘 나이트라인 언제 하지?” 발화의 Action 태그는 ‘ start_tme’ slot에 ‘ ?’ 값이 주어지며, 사용자 프레임을 통해 검색된 결과로 “ SBS에서 오늘 밤 12시에 방송될 예정입니다” 를 발성하기 위해서 시스템 결과 프레임에는 ‘ start_time’ 과 ‘ channel’ slot의 Action 태그로 ‘ *’ 가 붙여진다.

대용어 복원은 대용어 표현 인식과 선행사 할당 과정으로 구성된다. 대용어 표현 인식 과정은 대용어 표현의 경계를 인식하고 대용어 태그 중 1개를 할당하는 작업이며, 약 600셋의 대화 시나리오를 분석해서 7개의 대용어 태그를 정의하였다(N_program,N_program_this,N_channel,N_channel_this,N_date,N_time,N_list). 대용어 표현 인식은 다음 자질을 이용해서 CRF 분류기를 만들어 적용했다.

- 형태소 및 품사: 현재 발화에 나타난 형태소 및 품사정보 (e.g. "[SBS/nn]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(45개) (e.g. search_program: "오후 2시 이후에 SBS에서 뭐가 해?")
- 이전 발화의 화행 (PDA): 이전 발화의 화행정보로, 이전 문맥을 고려하기 위해서 필요함

대용어 표현 인식의 두 번째 단계는 인식된 대용어에 알맞은 태그를 할당하는 것으로, 대용어 표현으로 자주 사용되는 4가지 개념열[2]을 선행사 후보로 사용한다(PROGRAM, CHANNEL, DATE, TIME). 대화 중 발생하는 모든 선행사는 그들의 유형에 따라 특정 대용어 스택에 저장되어 선행사 할당 시에 사용된다. 대용어 표현은 2가지 유형으로 구분된다; 1) 대명사 ("그거", "저거", "이거" 등), 2) 명사구 ("그 SBS 프로그램", "오후 3시에 하는 거", “ MBC에서 하는 거” 등). 사용자들은 대명사 대용어를 사용할 때는 주로 가장 최근의 선행사를 가리키는 경우가 많으므로 대명사 유형에 대해서는 straightforward preference 전략을 사용해서 가장 최근 선행사로 복원한다. 명사구로 된 대용어는 “ 오후 6시 이후에 SBS에서 하는 프로그램” 과 같이 표현 안에 추가 정보를 가지고 있으므로, 추가 정보를 제약 조건으로 사용해서 선행사를 복원한다. 즉, 앞의 발화는 6시 이후에 하는 SBS 채널에서 하는 프로그램 스택에 있는 선행사 후보 중에서 가장 최근에 언급된 것으로 대용어를 복원한다.

생략어 복원은 현재 발화에 생략된 선행사를 복원하는 문제이며, 대부분의 선행사는 이전 문맥, 즉 이전 포커스 프레임에 나타난다. 그러므로 우리는 생략어 복원 문제를 문맥 상속 문제로 다루어 해결하였다. 사용자가 User1:“ 오늘 10시에 MBC에서 뭐해?” → User2:“ 11시에는?” 과 같이 말한 경우, User2의 완벽한 문장은 “ *오늘* 11시에 *MBC에서* 뭐해?” 이며, 이전 프레임에서 “ 오늘” , “ MBC” 정보가 상속되어야 한다. 복원 과정은 3단계로 구성된다; 1) 생략어 존재 확인, 2) 프레임 슬롯 검토, 3) 필수 슬롯이지만 값이 없는 슬롯의 값 생성. 1단계에서 포커스 프레임의 이전 슬롯 값의 상속 여부를 결정하며, 이를 위해 “ clear” 와 “ not clear” 의 이진 분류를 위해 5가지 자질을(형태소&품사, 개념열, 대용어, CDA, PDA) 이용한 ME 모델을 적용하였다. “ not clear” 태그는 현재 발화에 생략어가 있다는 것을 나타내며, 이런 경우에 시스템은 이전 포커스 프레임으로부터 슬롯

값들을 상속한다. 2단계는 1단계 결과가 “ not clear” 인 경우에만 동작하며, 임무는 상속된 값들 중에서 불필요한 것들을 제거하는 것이다. 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 [program:주몽], [channel:MBC], [start_time:now], [end_time:xxx], [program code:P0010], [actor:xxx], [director:xxx]과 같이 주몽 프로그램과 관련된 여러 슬롯 값을 가지고 있다. User2에서 생략어 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 제거 되야 한다. 2단계는 23개 규칙에 기반해서 동작하며, 규칙들은 대화 시나리오에 기반해서 수작업으로 구축됐다. 3단계는 대화 중에 나타나지 않지만 필수적인 슬롯 정보를(‘ 필수 슬롯’) 새로 생성한다. 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 프레임에 없는 경우에는 3단계에서 새로 생성되어야 하며, 3단계는 규칙에 기반해서 이루어진다. 다음 대화는 3단계가 동작하는 예를 보여준다.

- User1: "MBC에서 뭐 해?” → System1: "뉴스데크스가 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 만들었다. 2단계와 3단계 규칙의 예는 다음과 같다

2 단계 규칙	3 단계 규칙
❑ Action ✓ clear all inherited slots except program and genre slot ❑ Condition ✓ Dialogue Act: search_program ✓ Current concept sequence ◆ TIMES is only exist	IF dialogue act = "show_next_page", there is not date and start_time THEN Generate a new slot [start_time: "now"]

제안한 방법의 효과를 보기 위해서 2개의 평가셋에 기반해서 실험하였다. 일반 사용자를 대상으로 제약 없이 자유롭게 만들어진 평가셋1(ES1)은 2,282개 발화로 구성되었고, out-of-grammar, out-of-vocabularies, 그리고 out-of-function가 존재한다. 반면, 평가셋2(ES2)는 일반적으로 가장 많이 사용되는 발화 패턴 260개에 기반해서 제한된 발화로 이루어졌다. 대용어 복원 평가는 7개 대용어 표현 인식률과 선행사 할당의 정확성을 평가하였으며, 2개가 모두 맞는 경우에만 맞춘 것으로 고려하였고, 생략어 복원 평가는 시스템이 생성한 프레임과 정답 프레임의 일치 여부로 판단하였다. F-measure 평가결과, 대용어복원은 ES1은 85.19%, ES2는 94.47%, 생략어복원은 ES1은 89.8%, ES2는 99.42% 성능을 보였다. 이전 연구에서는 Tetreault and Allen [3]가 담화 분할 정보를 이용한 대명사 복원 알고리즘을 평가해서 68.6% F-score를 보였고, Nielsen [4]은 동사구 생략 인식 시스템을 개발해서 72% F-score를 얻었다. 또한 음성 기반 대화 시스템을 3 가지 평가 척도를 이용해서 평가하였다; 1) 문장 인식율(Sentence Recognition Ratio:SRR), 2) 대화 성공률(Dialogue Success Ratio:DSR), 3) 작업 성공률(Task Completion Ratio:TCR). 대화 시스템의 최종평가는 SRR, DSR, TCR 3개의 평균값인 사용자 만족도(User Satisfaction Ratio:USR)로 계산하였다. 평가결과로 ES1에서 SRR 53.5%, DSR 77.8%, TCR 87.7%, USR 73.0%를 얻었고, ES2에서 SRR 78.3%, DSR 96.3%, TCR 96.7%, USR 90.4%를 보였다.

3. 결 론

본 논문에서는 대용어 복원을 위해서 CRF 분류 모델을 사용해서 상당한 성능 향상을 얻을 수 있었고, 생략어 복원은 이전 포커스 프레임의 상속 문제로 다뤄서 ME 모델을 적용해서 좋은 결과를 얻었다. 또한 실제 대화시스템에 적용해서 성능평가 결과로 USR 90.4% 성능을 보였는데, 기존 연구인 JUPITER[5]가 2,000개 단어를 사용해서 약 80% 발화에 대해서 정답을 제시한 것에 비해 높은 성능을 보였다. 그러므로 제안된 대용어 복원 및 생략어 복원 기술이 실제로 음성 대화 시스템의 성능에도 매우 긍정적인 효과를 주고 실제 상용화 목적으로도 활용할 수 있음을 확인하였다.

참고문헌

[1]J. C. Carroll, "MIMIC: An adaptive mixed initiative spoken dialogue system for information queries," Proc. 6th Applied NLP Conference, pp.97-104, 2000.

[2]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.

[3]J. R. Tetreault and J. F. Allen, "Dialogue structure and pronoun resolution", Proc. DAARC, 2004

[4]L. A. Nielsen, "Verb phrase ellipsis detection using automatically parse text", Proc. COLING, pp. 1093-1099, 2004.

[5]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.