

목적지향 대화에서 화자 의도의 통계적 예측 모델

(A Statistical Prediction Model of Speakers' Intentions in a
Goal-Oriented Dialogue)

김 동 현 [†] 김 학 수 ^{**} 서 정 연 ^{***}
(Donghyun Kim) (Harksoo Kim) (Jungyun Seo)

요 약 사용자 의도 예측 기술은 음성인식기의 탐색 공간을 줄이기 위한 후처리 방법으로 사용될 수 있으며, 시스템 의도 예측 기술은 유연한 응답 생성을 위한 전처리 방법으로 사용될 수 있다. 이러한 실용적인 필요성에 따라 본 논문에서는 화행과 개념열의 쌍으로 일반화된 화자의 의도를 예측하는 통계 모델을 제안한다. 단순한 화행 n-그램 통계만을 이용한 기존의 모델과는 다르게 제안 모델은 현재 발화까지의 대화 이력을 다양한 언어 레벨의 자질 집합(화행과 개념열 쌍의 n-그램, 단서 단어, 영역 프레임의 상태 정보)으로 표현한다. 그리고 추출된 자질 집합을 CRFs(Conditional Random Fields)의 입력으로 사용하여 다음 발화의 의도를 예측한다. 일정 관리 영역에서 실험을 수행한 결과, 제안 모델은 사용자의 화행과 개념열 예측에서 각각 76.25%, 64.21%의 정확률을 보였다. 그리고 시스템의 화행과 개념열 예측에서 각각 88.11%, 87.19%의 정확률을 보였다. 또한 기존 모델과 비교하여 29.32% 높은 평균 정확률을 보였다.

키워드 : 의도 예측, 화행 예측, 개념열 예측

Abstract Prediction technique of user's intention can be used as a post-processing method for reducing the search space of an automatic speech recognizer. Prediction technique of system's intention can be used as a pre-processing method for generating a flexible sentence. To satisfy these practical needs, we propose a statistical model to predict speakers' intentions that are generalized into pairs of a speech act and a concept sequence. Contrary to the previous model using simple n-gram statistic of speech acts, the proposed model represents a dialogue history of a current utterance to a feature set with various linguistic levels (i.e. n-grams of speech act and a concept sequence pairs, clue words, and state information of a domain frame). Then, the proposed model predicts the intention of the next utterance by using the feature set as inputs of CRFs (Conditional Random Fields). In the experiment in a schedule management domain, The proposed model showed the precision of 76.25% on prediction of user's speech act and the precision of 64.21% on prediction of user's concept sequence. The proposed model also showed the precision of 88.11% on prediction of system's speech act and the precision of 87.19% on prediction of system's concept sequence. In addition, the proposed model showed 29.32% higher average precision than the previous model.

Key words : Intention prediction, speech act prediction, concept sequence prediction

· 이 연구(논문)는 지식경제부 지원으로 수행하는 21세기 프론티어 연구개발사업
(인간기능 생활지원 지능로봇 기술개발사업)의 일환으로 수행되었습니다.

논문접수 : 2008년 3월 19일

심사완료 : 2008년 8월 29일

† 정 회 원 : (주)다이캐스트 연구소 연구원
kdh2007@sogang.ac.kr

** 정 회 원 : 강원대학교 컴퓨터정보통신공학 교수
nlpdrkim@kangwon.ac.kr
(Corresponding author)

*** 종신회원 : 서강대학교 컴퓨터공학과 교수
seojy@sogang.ac.kr

Copyright©2008 한국정보과학회: 개인 목적이나 교육 목적의 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 : 소프트웨어 및 응용 제35권 제9호(2008.9)

1. 서론

목적지향 대화시스템은 호텔 예약이나 길안내와 같이 특정한 목적을 달성하기 위해서 이루어지는 대화를 이해하고 적절한 응답을 찾아 제시해 주는 지능형 소프트웨어를 말한다. 이러한 목적지향 대화시스템을 구현하기 위해서는 각 발화에 숨겨진 화자의 의도를 찾아내는 의도 분석 기술과 적절한 시스템 응답을 만들어내기 위한 의도 예측 기술이 모두 필요하다. 그러나 화자의 의도와 관련된 기존의 많은 연구들은 계획 추론 규칙[1]이나 통계 기반의 기계학습[2,3]을 이용하여 입력된 발화를 특정한 형태의 문자열로 일반화하는 분석기술에만 초점을 맞추어 왔다. 반면에 의도예측에 관한 연구는 다음과 같은 활용 가능성에도 불구하고 폭넓게 이루어지지 않고 있다.

- 사용자 음성인식: 사용자의 의도를 예측할 수 있다면 음성인식 후보군을 제약함으로써 탐색 공간을 줄이거나 인식 성공률을 높일 수 있다. 예를 들어, 시스템의 발화 “내일 몇 시에 만나기로 했나요?”에 대한 사용자의 발화는 약속 시간에 관한 것일 가능성이 높다. 그러므로 음성인식기의 탐색 공간을 약속 시간에 관한 발화로 제약하거나 약속 시간에 관한 발화에 가중치를 부여한다면 인식 성능을 높일 수 있다.
- 시스템 응답생성: 기존의 목적지향 대화시스템에서 응답생성 방법은 대부분 시스템 설계자의 지식에 의존한 수동 규칙들을 사용했다. 그러나 시스템의 의도를 예측할 수 있다면 응답 생성시에 적절한 의미구조를 선택하게 함으로써 유연한 응답을 생성할 수 있다. 예를 들어, 사용자의 발화 “전화번호는 010 123에 4567입니다.”에 대한 시스템의 발화는 다른 질문을 하거나 정보를 제공하기 보다는 “010 123에 4567이요?”와 같이 확인을 요구하는 것이 바람직하다. 그러므로 응답생성기의 후보 문장을 확인 요구에 대한 것으로 제약한다면 사용자 발화가 가지는 특색에 따라 유연한 응답을 생성할 수 있다.

본 논문에서는 상기한 목적으로 활용 가능한 통계 기반의 화자 의도 예측 모델을 제안한다. 제안 모델은 이전과 현재 발화의 의도, 현재 발화에 포함된 단서 단어, 현재까지의 영역 프레임 상태 정보와 같이 다양한 레벨의 언어적 지식을 입력 자질로 사용함으로써 정확률의 향상을 꾀한다.

본 논문의 구성은 다음과 같다. 2장에서는 화자 의도 예측에 관한 기존 연구를 살펴본다. 3장에서는 통계적으로 의도를 예측하는 모델과 필요한 입력 자질들을 제안한다. 4장에서는 실험 데이터를 설명하고 실험 결과를 분석한다. 마지막으로 5장에서 결론을 내린다.

2. 관련 연구

의도 예측에 관한 연구는 크게 사용자 의도 예측과 시스템 의도 예측으로 나누어진다. 사용자 의도 예측에 관한 대표적인 연구로는 Smith[4]와 Reithinger[5]가 있다. Smith는 대화 진행 상태별로 사용자가 다음에 발화 가능한 다수의 문장들(또는 주요 단어들)을 미리 정의해 두고, 음성인식된 후보 문장들과 현재 대화 진행 상태에서 다음에 출현 가능한 문장들을 어휘적으로 비교하여 가장 유사한 문장을 선택하는 방법을 사용하였다. 이러한 방법은 매우 제한된 영역에서 비교적 간단한 방법으로 높은 성능을 얻을 수 있지만 영역이 조금만 넓어지면 사용자 발화의 예측과 수집이 매우 어려워진다는 단점이 있다. Reithinger는 음성인식에서의 오류를 줄이기 위하여 Speech-to-Speech 번역 시스템인 VERBMO-BIL[6]에서 통계적인 화행 n-그램 정보를 이용하여 화행을 예측하는 실험을 하였다. 식 (1)은 Reithinger가 제안한 화행 예측 확률식이다.

$$P(SA_n | SA_{n-1}, SA_{n-2}) = q_1 f(SA_n) + q_2 f(SA_n | SA_{n-1}) + q_3 f(SA_n | SA_{n-1}, SA_{n-2}) \tag{1}$$

식 (1)에서 SA_n 은 예측하고자 하는 n번째 발화의 화행을 의미하며, SA_{n-1} 과 SA_{n-2} 는 예측 대상의 바로 이전 발화의 화행과 그 이전 발화의 화행을 의미한다. f 는 학습말뭉치에 나타난 해당 화행의 상대 빈도수를 의미하며, q 는 합이 1이 되는 가중치 인자이다. 식 (1)에서 보는 것과 같이 Reithinger의 방법은 매우 간단하지만 화행 예측 정보가 음성인식기의 탐색 복잡도를 19~60%의 감소시킬 수 있다는 가능성을 보였다. 그러나 단순한 화행 n-그램 정보만을 문맥으로 사용하여 높은 화행 예측 정확률을 얻지는 못했다.

시스템 의도 예측에 관한 연구는 주로 시스템 응답 생성 연구의 일부로 취급되어 왔으며, 대화 문법에 기반한 방법[7]과 계획에 기반한 방법[8]이 대표적이다. 대화 문법에 기반한 방법은 대화에서 나타나는 인접쌍의 원리에 바탕을 두고 있다. 즉, 질문 뒤에는 대답이 나오고 제안 뒤에는 승낙 또는 거절이 나온다는 사실을 이용하여 상대방의 의도를 예측하는 것이다. 대화 문법에 기반한 방법은 원리가 명쾌하고 단순하기 때문에 구현이 쉽다는 장점이 있다. 그러나 수작업으로 만들어진 제한된 규칙을 이용하기 때문에 예상치 못한 사용자 반응에 매우 취약하다는 단점이 있다. 이러한 문제점을 해결하기 위하여 계획에 기반한 방법이 연구되었다. 오종진[8]은 그동안 인식 관점에서 연구되었던 계획 기반 대화 모델을 변형하여 발화 생성에 이용하는 방법을 제안하였다. 계획에 기반한 방법은 대화 문법에서 나타나는 응답의

유연성 문제를 극복할 수 있지만, 목적지향 대화와 같이 비교적 단순한 영역에서는 시스템이 불필요할 정도로 복잡해진다는 단점이 있다.

본 논문에서는 기존의 사용자 의도 예측 방법과 시스템 의도 예측 방법들에서 나타나는 일련의 문제들을 해결하기 위하여 다양한 자질들의 통계적 정보를 이용하여 화자의 의도를 예측할 수 있는 새로운 모델을 제안한다. 제안 모델은 Reithinger 모델의 낮은 정확도를 향상시키기 위하여 화행 n -그램 정보와 함께 단서 단어와 영역 프레임 정보를 문맥 정보로 사용한다. 그리고 대화 문법 모델의 응답 유연성을 향상시키고 계획 모델의 구현 복잡성을 조금이나마 줄이기 위한 대안으로 통계 기반의 방법을 사용한다.

3. 통계적 의도 예측

3.1 의도의 일반화

목적 지향 대화에서 각 발화에 내포된 화자의 의도는 포괄적인 담화 의도를 나타내는 화행과 영역 종속적인 세부 의미를 나타내는 개념열(concept sequence)의 쌍으로 일반화될 수 있다[9]. 표 1은 일정 관리 영역에서 수집한 대화 말뭉치를 분석하여 정의한 화행을 보여준다.

표 2는 일정관리 영역에서 데이터베이스 연산을 모델링하여 정의한 개념열의 구성을 보여준다. 본 논문에서는 2가지 테이블, 4가지 연산자, 8가지 필드를 대상으로 3층 구조의 개념열 부착 방법[10]에 따라 53개의 개념

표 1 화행의 종류

화행	설명
greeting	대화 서두의 인사말
expressive	대화 후미의 인사말
opening	실제 대화의 시작
ask_ref	WH-질문
ask_if	YN-질문
response	응답
request	행위를 요청
ask_confirm	이전 발화의 확인
confirm	확인 발화의 응답
inform	정보 제공
accept	호응

표 2 개념열의 구성

테이블명	연산자명	필드명
timetable	insert	agent
alarmtable	delete	date
	select	day_of_week
	update	time
		person
		place
		content
		field

열을 정의하여 사용한다.

3.2 의도 예측 모델

n 개의 발화로 구성된 대화에서 i 번째까지의 발화가 입력되었을 때, 다음 발화에 대한 의도 예측 모델은 화행과 개념열이 서로 독립이라고 가정하면 식 (2)와 같이 근사된다.

$$P(I_{i+1}|U_{1,i}) \approx P(SA_{i+1}|U_{1,i})P(CS_{i+1}|U_{1,i}) \quad (2)$$

식 (2)에서 $U_{1,i}$ 는 대화의 시작에서 i 번째까지의 발화를 의미하며, I_{i+1} 은 $i+1$ 번째 발화에 내포된 화자의 의도를 의미한다. 그리고 SA_{i+1} 과 CS_{i+1} 는 각각 $i+1$ 번째 발화에 대한 화행과 개념열을 의미한다.

식 (2)에서 $U_{1,i}$ 을 직접 계산하는 것은 현실적으로 불가능하다. 그러므로 I_{i+1} 이 i 번째 발화에 대한 다양한 관찰로부터 유추될 수 있다고 가정하고, 이것을 $i+1$ 번째 발화에 대한 자질집합 FS_{i+1} 라고 하면 식 (3)과 같다.

$$P(I_{i+1}|U_{1,i}) \approx P(SA_{i+1}|FS_{i+1})P(CS_{i+1}|FS_{i+1}) \quad (3)$$

다양한 자질 값이 주어졌을 때 지정된 출력의 조건부 확률값을 나타내는 식 (3)은 식 (4)와 같이 CRFs(Conditional Random Fields)를 이용하여 효과적으로 계산될 수 있다[11].

$$P_{CRF}(SA_{1,i+1}|FS_{1,i+1}) = \frac{1}{Z(FS)} \exp\left(\sum_{j=1}^{i+1} \sum_k \lambda_k F_k(SA_j, FS_j)\right)$$

$$P_{CRF}(CS_{1,i+1}|FS_{1,i+1}) = \frac{1}{Z(FS)} \exp\left(\sum_{j=1}^{i+1} \sum_k \lambda_k F_k(CS_j, FS_j)\right) \quad (4)$$

식 (4)에서 $Z(FS)$ 는 자질집합에 대한 정규화 요소이며, $F_k(SA_j, FS_j)$ 와 $F_k(CS_j, FS_j)$ 는 j 번째 발화에 대한 화행 예측과 개념열 예측을 위한 자질 함수로서 통계 모델 학습 시에 얻어진 k 개의 자질들 중에서 해당 자질이 있으면 1을, 그렇지 않으면 0의 값을 가진다.

3.3 자질 집합

본 논문에서는 화자의 의도를 예측하기 위한 통계모델의 입력으로 다음과 같은 3가지 종류의 자질을 사용한다.

- 화행과 개념열 쌍에 대한 n -그램: 대화는 서로 연관된 발화들의 집합이므로 현재 발화는 그 이전의 발화나 다음 발화와 밀접한 관계를 가진다. 따라서 다음 발화의 의도를 예측하기 위해서 현재까지의 발화 정보가 매우 중요함을 직관적으로 알 수 있다. 본 논문에서는 다음 발화의 의도를 예측하기 위해서 현재 발화와 이전 발화의 화행과 개념열 쌍들을 입력 자질로 사용한다. 예를 들어, 표 3과 같은 대화에서 2번째 발화까지 진행된 상태라고 가정했을 때, 3번째 발화의 의도 'response & timetable_select_date'를 예측하기 위해서 제안 모델은 현재 발화의 의도 'ask_ref & timetable_select'와 이전 발화의 의도 'request &

표 3 대화 예제

화자	발화	화행	개념열
사용자	일정 확인 좀 해줄래?	request	timetable_select
시스템	언제 일정을 확인할까요?	ask_ref	timetable_select
사용자	다음달 일정 좀 알려줘.	response	timetable_select_date

timetable_select'를 식 (4)의 입력 자질(자질 함수에 대한 입력 값)로 사용한다.

- 단서 단어: 발화에 포함된 특정한 단어들은 다음 발화의 의도 예측에 중요한 단서를 제공한다. 예를 들어, 표 3에서 보듯이 '일정', '확인', '?'와 같은 어휘는 다음 발화 "언제 일정을 확인할까요?"와 매우 밀접한 연관 관계를 가진다. 그러므로 본 논문에서는 발화에 포함된 '어휘/품사'와 '품사-품사' 중에서 정보량이 많은 것들을 추출하여 의도 예측의 입력 자질로 사용한다. 정보량 측정은 문서 분류 문제에서 좋은 결과를 보이고 있는 χ^2 통계량을 이용한다[10,13]. 먼저, 화행과 개념열이 부착된 학습말뭉치를 형태소 분석한 후, 각 문장을 화행별, 개념열별로 범주화한다. 그리고 형태소 분석된 문장으로부터 '어휘/품사'와 '품사-품사'를 추출한 후, 식 (5)와 같은 χ^2 통계량을 계산한다. 마지막으로 '어휘/품사'와 '품사-품사'를 χ^2 통계량에 따른 내림차순으로 정렬한 후, 실험에 의해서 상위 n개를 단서 단어로 채택한다.

$$\chi^2(f, c) = \frac{(A+B+C+D) \times (AD - CB)^2}{(A+C) \times (B+D) \times (A+B) \times (C+D)} \quad (5)$$

식 (5)에서 f 는 자질(어휘/품사 또는 품사-품사)을 의미하며, c 는 화행 또는 개념열 범주를 의미한다. A 는 c 에 속해 있는 발화 중에서 f 를 포함하는 발화 수, B 는 c 이외의 범주에 속해 있는 발화 중에서 f 를 포함하는

발화 수, C 는 c 에 속해 있는 발화 중에서 f 를 포함하지 않는 발화 수, D 는 c 이외의 범주에 속해 있는 발화 중에서 f 를 포함하지 않는 발화 수를 의미한다.

단서 단어 집합이 결정되면, 입력 발화에 포함된 모든 '어휘/품사'와 '품사-품사'를 추출하고 그 중 단서 단어 집합에 포함되어 있는 것만을 식 (4)의 입력 자질로 사용한다. 예를 들어, "다음달 일정 좀 알려줘"라는 발화에 대하여 '일정/ncn 확인/ncn 좀/ma 하/pv 어/ef 주/pv px 르래/ef ?/sf'와 'ncn-ncn, ncn-ma, ma-pv, pv-ef, ef-px, px-ef, ef-sf'를 추출한 후, 단서 단어 집합에 포함되어 있는 '일정/ncn, 확인/pv, ?/sf'만을 선택하여 식 (4)의 입력 자질로 사용한다.

- 영역 프레임의 상태 정보: 목적 지향 대화에서 화자들은 해당 영역에서 서로에게 공유된 지식(즉, 대화 이력)을 바탕으로 대화를 이끌어 나가며, 이러한 지식들은 유한상태 모델(finite-state model), 프레임 기반 모델(frame-based model)[12], 계획 기반 모델(plan-based model)[1] 등으로 표현될 수 있다. 본 논문에서는 일정 수준의 대화 유연성을 보장하면서 구현하기 쉬운 프레임 기반 모델의 영역 프레임(domain frame) 정보를 대화 이력으로 사용한다. 영역 프레임 정보는 슬롯 수정 정보와 슬롯 검색 정보로 나뉜다. 슬롯 수정 정보는 어떤 슬롯이 채워진 것이고 어떤 슬롯이 비어있는가 하는 것을 나타낸 이진 정보이며,

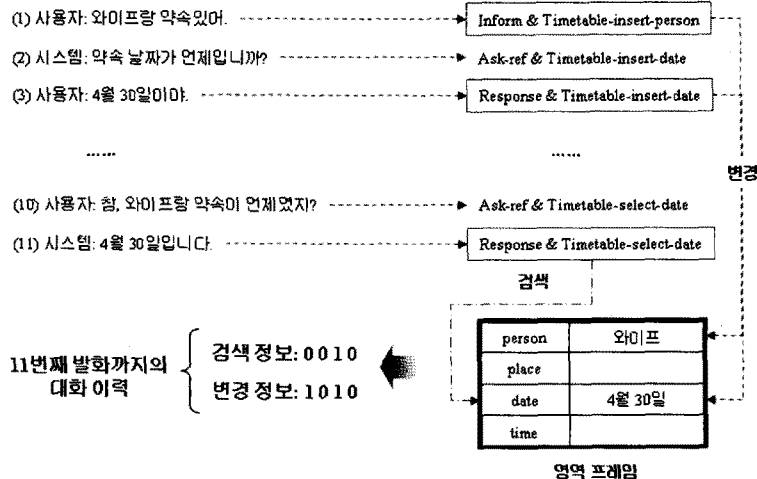


그림 1 영역 프레임의 상태 정보

슬롯 검색 정보는 어떤 슬롯이 한번 이상 참조된 것이고 어떤 슬롯이 아직 참조되지 않은 것인가를 나타내는 이진 정보이다. 프레임 기반 모델에서 슬롯의 수정이나 검색은 대화 관리자에 의해서 이루어진다. 그림 1은 예제를 기반으로 영역 프레임의 상태 정보를 보여준다. 그림 1에서 보듯이 대화 관리자는 발화 (1)의 의도 'inform & timetable_insert_person'과 발화 (3)의 의도 'response & timetable_insert_date'를 분석하여 사용자의 발화로부터 정보를 추출하여 'person' 슬롯과 'date' 슬롯을 채워 넣는다. 그리고 발화 (11)의 의도 'response & timetable_select_date'에 따라 'date' 슬롯으로부터 정보를 얻어 사용자에게 제공한다. 이러한 대화 관리자를 실제 구현하는 것은 매우 복잡할 뿐만 아니라 본 논문의 주제와도 맞지 않는다. 그러므로 본 논문에서는 각 슬롯의 수정이나 검색과 관련된 의도들을 미리 정의해 두고, 해당 의도가 나타나면 슬롯의 값을 '1'로 설정하는 방법으로 슬롯 수정 정보와 슬롯 검색 정보를 추출한다. 이러한 방법에 따라 그림 1의 예에서 발화 (11)이 끝난 후에 슬롯 수정 정보는 '1 0 0 1'의 값을, 슬롯 검색 정보는 '0 0 1 0'의 값을 가지며 식 (4)의 입력으로 사용된다.

4. 실험 및 평가

4.1 실험 데이터

의도 예측 모델의 성능을 평가하기 위하여 일정관리 영역에서 Wizard-Of-Oz 방식으로 시뮬레이션한 대화 말뭉치를 수집한 후, 수동으로 화행과 개념열을 부착하였다. 수집된 말뭉치는 일정 추가, 삭제, 변경과 관련된 내용을 포함하고 있다. 말뭉치의 구성은 956개의 대화 (21,336개의 발화)로 구성되며, 대화당 평균 발화의 수는 22.32개이다. 표 4는 화행과 개념열이 부착된 말뭉치의 일부를 보여준다.

표 4에서 '/ID/'는 발화의 고유 번호로써 대화 번호와 발화 번호로 구성된다. '4-5'는 4번 대화의 5번째 발화

를 의미한다. '/SP/'는 화자 정보로써 'User'와 'System'으로 구성된다. '/UT/'는 발화이고, '/SA/'는 화행, '/CS/'는 개념열을 나타낸다. 화행과 개념열 부착은 대학원에서 자연어처리를 전공하는 5명의 석사과정 학생이 수행하였으며, 일관성 유지를 위하여 1명의 박사과정 학생이 최종 검토하였다.

실험은 학습데이터와 평가데이터의 비율을 4:1로 하여 5배 교차검증(5-fold cross validation)은 수행하였다. 단어 단어는 시스템 의도 예측시에 100개, 사용자 의도 예측시에 1,000개를 사용하였다. CRFs의 내부 설정 인자로 추정알고리즘은 L-BFGS를 이용하였으며[14], 회소 데이터 문제를 위한 평탄화 요소는 Gaussian Prior를 이용하였다[11]. Gaussian Prior의 값은 10으로 설정하였으며, 훈련 반복 회수는 30으로 설정하였다.

4.2 실험 결과

표 5는 자질들을 개별적으로 사용했을 때의 화행 예측 정확률을 보여준다. 표 5에서 보듯이 화행 예측에는 '현재 <화행, 개념열>'이 가장 중요한 역할을 하였다. 이것은 목적지향 대화는 주로 요구와 응답 쌍으로 이루어져 있으며, 이것들은 서로 긴밀한 관계를 이루고 있기 때문인 것으로 생각된다. 그리고 사용자 화행 예측 정확률이 시스템 화행 예측 정확률보다 상대적으로 낮은 이유는 시스템 발화에 대응되는 사용자 발화의 다양성이 훨씬 크기 때문인 것으로 생각된다.

표 6은 자질들을 개별적으로 사용했을 때의 개념열 예측 정확률을 보여준다. 표 6에서 보듯이 사용자 개념열 예측에는 단어 단어가 가장 중요한 역할을 하였다. 이것은 프레임 기반 대화 모델의 특성상 미리 순서가

표 5 자질별 화행 예측 정확률

사용 자질	시스템 화행 예측 정확률(%)	사용자 화행 예측 정확률(%)
현재 <화행, 개념열>	83.25	72.89
이전 <화행, 개념열>	73.04	64.28
단서 단어	76.51	72.01
영역 프레임 정보	63.44	49.03

표 4 화행과 개념열이 부착된 말뭉치의 예

태그	값	태그	값
/ID/	4-5	/ID/	4-7
/SP/	User	/SP/	User
/UT/	약속 시간이 몇 시지?	/UT/	장소는 어디야?
/SA/	Ask_ref	/SA/	Ask_ref
/CS/	Timetable_select_time	/CS/	Timetable_select_place
/ID/	4-6	/ID/	4-8
/SP/	System	/SP/	System
/UT/	11시 30분입니다.	/UT/	코엑스홀입니다.
/SA/	Response	/SA/	Response
/CS/	Timetable_select_time	/CS/	Timetable_select_place

표 6 자질별 개념열 예측 정확률

사용 자질	시스템 개념열 예측 정확률(%)	사용자 개념열 예측 정확률(%)
현재 <화행, 개념열>	82.40	56.26
이전 <화행, 개념열>	52.43	52.80
단서 단어	66.35	59.40
영역 프레임 정보	37.68	49.03

정해져 있는 시스템 발화와는 다르게 이전 발화에 포함된 새로운 정보(예: 새로운 어휘 출현 등)에 따라 다양하게 변할 수 있는 사용자 발화의 특성 때문인 것으로 보인다. 그리고 개념열 예측 정확률이 화행 예측 정확률보다 전반적으로 낮은 이유는 개념열 범주의 개수가 화행 범주의 개수보다 훨씬 많기 때문인 것으로 보인다.

표 7은 서로 다른 자질들의 조합을 사용했을 때의 화행 예측 정확률을 보여준다. 표 7에서 보듯이 '현재 <화행, 개념열>'과 다른 자질들을 결합했을 경우에 대부분 3~4% 정도 정확률이 향상되었다. 그러나 사용자 화행 예측시에 '현재 <화행, 개념열>'과 '이전 <화행, 개념열>'을 결합했을 경우에는 정확률이 약간 하락하였다. 이것은 자질들 사이의 간섭 현상 때문인 것으로 보인다.

표 8은 서로 다른 자질들의 조합을 사용했을 때의 개념열 예측 정확률을 보여준다. 표 8에서 보듯이 '현재 <화행, 개념열>'과 다른 자질들을 결합했을 경우에 1~4% 정도 정확률이 향상되었다. 시스템 개념열 예측시에 '현재 <화행, 개념열>'과 '단서 단어'를 결합했을 경우에 가장 낮은 정확률 향상을 보였다. 이것은 시스템의 발화가 미리 정해진 순서에 따라 결정됨으로써 두 자질들 사이에 중복된 정보가 많기 때문인 것으로 보인다.

표 7과 표 8에서 보듯이 사용자 화행 및 개념열 예측

이 시스템 화행 및 개념열 예측보다 훨씬 낮은 정확률을 보였다. 이것은 사용자 발화의 다양성에 기인한 것으로 보인다. 그러나 사용자 의도 예측을 수행한 이유가 음성인식 단계에서의 탐색 공간을 줄이고 정확률을 높이기 위한 것이기 때문에 상위 n개 후보를 고려해볼 가치가 있다. 예를 들어, 상위 5개를 고려했을 때 정확률이 100%라면 화행은 12개 중에서 5개만을 고려하면 되고, 개념열은 53개 중에서 5개만을 고려하면 되기 때문에 탐색공간을 줄이면서 정확률을 향상시키는 효과를 얻을 수 있게 되는 것이다. 표 9는 상위 n개를 고려했을 때 사용자 화행과 개념열 예측에 대한 정확률을 보여준다. 표 9에서 보듯이 화행은 상위 5개를 정답으로 간주했을 때 거의 100%에 가까운 정확률을 보이고, 개념열은 상위 15개를 정답으로 간주 했을 때 98% 정도의 정확률을 보였다.

제안 모델의 성능을 평가하기 위하여 동일한 학습데이터와 평가 데이터를 이용하여 Reithinger(1995)[5]와 비교하였다. 2장 관련 연구의 식 (1)에서 보는 것과 같이 Reithinger(1995)는 화행 예측에 대한 연구만을 진행했기 때문에 개념열 예측에 대한 결과를 포함하고 있지 않다. 그러므로 본 논문에서는 식 (1)을 개념열 예측에 그대로 적용하여 식 (6)과 같이 확률 모델을 만들고 제안 시스템과 비교하였다.

$$P(CS_n | CS_{n-1}, CS_{n-2}) = q_1 f(CS_n) + q_2 f(CS_n | CS_{n-1}) + q_3 f(CS_n | CS_{n-1}, CS_{n-2}) \quad (6)$$

식 (6)에서 CS_n 은 예측하고자 하는 n번째 발화의 개념열을 의미하며, CS_{n-1} 과 CS_{n-2} 는 예측 대상의 바로 이전 발화의 개념열과 그 이전 발화의 개념열을 의미한다. f 는 학습말뭉치에 나타난 해당 개념열의 상대 빈도

표 7 조합된 자질에 따른 화행 예측 정확률

사용 자질	시스템 화행 예측 정확률(%)	사용자 화행 예측 정확률(%)
현재 <화행, 개념열>+단서 단어	86.29	74.78
현재 <화행, 개념열>+이전 <화행, 개념열>	87.31	72.80
현재 <화행, 개념열>+영역 프레임 정보	86.70	74.90
모든 자질	88.11	76.25

표 8 조합된 자질에 따른 개념열 예측 정확률

사용 자질	시스템 개념열 예측 정확률(%)	사용자 개념열 예측 정확률(%)
현재 <화행, 개념열>+단서 단어	83.30	61.62
현재 <화행, 개념열>+이전 <화행, 개념열>	86.56	62.62
현재 <화행, 개념열>+영역 프레임 정보	86.07	61.10
모든 자질	87.19	64.21

표 9 상위 n개에 대한 사용자 화행 및 개념열 예측 정확률

	Top-1	Top-3	Top-5	Top-7	Top-9	Top-11	Top-13	Top-15
화행 예측 정확률(%)	76.25	94.80	99.24	99.90	99.98	100	100	100
개념열 예측 정확률(%)	64.21	82.40	89.47	93.59	95.79	96.84	97.61	98.05

표 10 화행 및 개념열 예측 정확률 비교

		Reithinger(1995)	제안 모델
시스템 발화	화행 예측 정확률(%)	43.37	88.11
	개념열 예측 정확률(%)	68.06	87.19
사용자 발화	화행 예측 정확률(%)	37.59	76.25
	개념열 예측 정확률(%)	49.48	64.21

수를 의미하며, q 는 합이 1이 되는 가중치 인자이다. 표 10은 제안 모델을 Reithinger(1995)와 비교한 결과를 보여준다. 표 10에서 보듯이 제안 모델이 Reithinger(1995)보다 모든 경우에서 훨씬 높은 정확률을 보였다.

제안 모델이 Reithinger(1995)보다 높은 정확률을 보인 것은 입력 자질의 차이에 기인한 것으로 생각된다. 다음은 Reithinger가 사용하지 않은 단서 단어 자질과 영역 프레임의 상태 정보 자질이 의도 예측에 영향을 미친 예를 보여준다.

- 단서단어: 일정 관리 영역에서 시간과 장소는 매우 중요한 정보이기 때문에 대화 참여자는 해당 정보를 포함하는 단어가 언급되었을 경우에 그 사실을 다시 확인하려는 경향이 있다.

사용자: 몇 시에 홍길동과 약속이 있지? <- 약속 시간을 확인하기 위한 질문
 시스템: 6시에 약속이 있습니다. <- 약속 시간을 알려주기 위한 답변
 사용자: 6시? <- 약속 시간을 재확인하기 위한 질문

- 영역 프레임의 상태 정보: 목적 지향 대화에서 대화 참여자는 주어진 일을 완성하기 위해서 필요한 정보를 얻어내려는 경향이 있다. 그리고 대화의 효율성을 위해서 이미 얻어진 정보를 중복해서 요청하지 않으려는 경향이 있다. 즉, 영역 프레임에 비어있는 슬롯은 적극적으로 채우려고 하며, 특별한 이상이 없는 한 이미 채워져 있는 슬롯을 다시 채우려고 하지는 않는다.

사용자: 내일 압구정에서 홍길동씨와 약속이 있어. <- 날짜, 장소, 시간 슬롯이 채워짐
 시스템: 몇 시에 약속이 있으십니까? <- 시간 슬롯을 채우기 위한 질문
 사용자: 오후 6시. <- 시간 슬롯이 채워짐

5. 결론

본 논문에서는 목적지향 대화에서 통계를 이용하여 다음 발화의 의도를 예측하는 모델을 제안하였다. 제안 모델은 화행과 개념열의 n -그램, 단서 단어, 영역 프레임의 상태 정보를 이용하여 의도 예측에 필요한 대화 이력을 근사하였다. 그리고 근사된 정보를 자연어처리 분야에서 최근에 높은 성능을 보이고 있는 CRFs의 입력으로 사용하여 통계적으로 의도를 예측하였다. 일정 관리 영역에서 실험을 수행한 결과, 제안 모델은 시스템

화행과 개념열 예측에서 각각 88.11%, 87.19%의 정확률을 보였으며, 사용자 화행과 개념열 예측에서 각각 76.25%, 64.21%의 정확률을 보였다. 제안 모델을 시스템 의도 예측에 활용하면 통계적으로 적합한 의도를 찾아냄으로써 시스템 응답의 유연성을 높일 수 있고, 사용자 의도 예측에 활용한다면 음성인식에서의 탐색공간을 줄임으로써 음성인식의 효율성을 높일 수 있을 것으로 기대된다.

참고 문헌

- [1] Lambert, L. and Caberry, S., "A Tripartite Plan-based Model of Dialogue," Proceedings of ACL, pp.47-54, 1991.
- [2] Langley, C. "Analysis for Speech Translation Using Grammar-based Parsing and Automatic Classification," Proceedings of the ACL Student Research Workshop, 2002.
- [3] 은종민, 이성욱, 서정연, 지지벡터기제를 이용한 한국어 화행분석, 한국정보처리학회 논문지, Vol.12B, No.3, pp.365-368, 2005.
- [4] Smith, R. W. and Hipp, D. R., Spoken Natural Language Dialogue Systems: A Practical Approach, Oxford University Press, 1994.
- [5] Reithinger, N., "Some Experiments in Speech Act Prediction," Proceedings of Empirical Methods in Discourse Interpretation and Generation, 1995.
- [6] Wahlster, W. "Verbmobil-Translation of Face-to-Face Dialogs," Proceedings of MT Summit IV, 1993.
- [7] 김용재, 데이터베이스 검색을 위한 한국어 대화 인터페이스 시스템의 설계, 석사학위논문, 서강대학교, 1997.
- [8] 오종건, 작업수행영역에서 계획에 기반한 대화 시스템의 설계, 석사학위논문, 서강대학교, 1999.
- [9] Levin, L., Langley, C., Lavie, A., Gates, D., Wallace, D., and Peterson, K., "Domain Specific Speech Acts for Spoken Language Translation," Proceedings of 4th SIGdial Workshop on Discourse and Dialogue, 2003.
- [10] Lee, H., Kim, H., and Seo, J., Efficient Domain Action Classification Using Neural Networks, Lecture Notes in Computer Science, Vol.4233, pp.150-158, 2006.
- [11] Lafferty, J., McCallum, A., and Pereira, F., "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data", Proceedings of ICML, pp.282-289, 2001.

- [12] Goddeau, D., Meng, H., Polifroni, J., Seneff, S. and Busayapongchai, S., "A Form-based Dialogue Manager for Spoken Language Applications," Proceedings of International Conference on Spoken Language Processing, pp.701-704, 1996.
- [13] Yang, Y. and Pedersen, J. O., "A Comparative Study on Feature Selection in Text Categorization," Proceedings of ICML, 1997.
- [14] Fei, S. and Pereira, F., "Shallow Parsing with Conditional Random Fields," Proceedings of HLT and NAACL, 2003.



김 동 현

2005년 서강대학교 컴퓨터공학과 학사
2007년 서강대학교 컴퓨터공학과 석사
2007년~현재 (주)다이렉스트 선임연구원. 관심분야는 대화 처리, 자연어 검색

김 학 수

정보과학회논문지 : 소프트웨어 및 응용
제 35 권 제 5 호 참조

서 정 연

정보과학회논문지 : 소프트웨어 및 응용
제 35 권 제 1 호 참조