멀티 도메인 대화시스템을 위한 도메인 결정 기술

이인재, 김경덕, 김석환, 이동현, 최준휘, 이근배
포항공과대학교 컴퓨터공학과
{lij1984, getta, megaup, semko, chasunee, gblee}@postech.ac.kr

A Domain Selection for Multi-Domain Dialog System

Injae Lee, Kyungduk Kim, Seokhwan Kim, Donghyeon Lee, Junwhi Choi, Gary Geunbae Lee
Department of Computer Science and Engineering, POSTECH

요 약

본 논문은 다중 도메인 대화 시스템에서 사용자의 발화에 가장 적합한 도메인을 결정하는 방법에 관하여 논한다. 다중 도메인 대화 시스템 구축 시, 도메인의 확장성 재고 및 각 도메인 별 특성의 효과적인 적용을 위해, 각 도메인 별 단일 도메인 대화 전문가를 구축하고, 다수의 도메인 대화 전문가들을 통합함으로써, 다양한 도메인을 처리할 수 있는 대화 시스템을 구축할 수 있다. 이 때, 자연스러운 대화 처리를 위해서 주어진 사용자의 발화에 가장 적합한 도메인을 결정하는 기술이 필요하다. 본 논문에서는 사용자 발화의 의도 분석 결과 및 이전 사용자 발화 도메인 정보를 이용하는 구축된 분류 모델에 기반한 도메인 결정 방법을 제안하고, 실험 결과를 통해 제안된 방법의 효과를 검증한다.

주제어: 대화 시스템, 다중 도메인, 도메인 결정

1. 서론

자연어 대화 인터페이스는 인간에게 가장 자연스럽고 효율적이며 별도의 학습을 필요로 하지 않고, 다양한 기기 및 환경에 폭넓게 적용될 수 있다는 점에서 차세대 인터페이스로 각광을 받고 있다. 음성 대화 시스템은 자연어를 이용하여 컴퓨터와 의사 소통 하는 시스템으로 지능형 인터페이스로서 활발히 연구되어 지고 있다.

대화 시스템은 음성 인식, 자연어 이해, 대화 관리, 자연어 생성, 음성 합성 등의 다양한 기능들의 조합으로 구성되며, 각각의 기능들을 구현하기 위해서는 규칙이나 학습 데이터 등의 자원들이 요구된다. 하지만 이러한 자원들을 구축하기 위해서는 많은 비용이 소요되고, 이로 인해 자연스러운 대화 시스템 구현을 위한 요구 자원 확보를 위해 초창기 대화 시스템들은 날씨 안내[1]나 여행 계획[2]등 주로 작은 범위의 한정된 단일 도메인 상에서 개발되어 왔다.

하지만 실생활에서 인간과 인간 사이의 대화는 한정된 단일 도메인 상에서만 이루어지지 않으며, 대화의 흐름 변화에 따라 도메인 전환이 이루어지기도 한다. [표 1]은 다중 도메인 대화의 예제이다.

[표 1]과 같은 다중 도메인 대화를 기존의 단일 도메인 대화 시스템을 이용하여 구현하는 데에는 한계가 있다. 그 첫 번째 이유로는 대화 시스템을 구성하는 개별 기능들의 성능 저하 가능성을 고려할 수 있다. 각각의 도메인 상의 대화는 어휘, 대화 양상, 컨텐츠 등 다양한 측면에서 서로 다른 특성을 지닌다. 그러므로 높은 성능을 얻기 위해서는 각 도메인에 특화된 세부 기능 구현이 요구된다.

또한 단일 도메인 대화시스템으로는 도메인 확장 상의 한계가 존재한다. 여러 도메인을 하나의 모델을 이용하여 구축하였을 때, 만약 새로운 도메인을 추가하거나 특정 도메인을 삭제하고자 한다면, 시스템 전반에 걸쳐 많은 수정이 요구된다.

표 1 다중 도메인 대화 예제 (U: User, S: System)

	발화	도메인
U	지금 TV에서 뭐하지?	TVguide
S	MBC에서 뉴스를, ... 방영중입니다.	TVguide
U	MBC 프로그램 녹해해줘	TVguide
S	알겠습니다.	TVguide
U	오늘 서울에 비 올까?	Weather
S	오늘 서울의 강수확률은 ~ 입니다	Weather
U	근처 맛집 좀 알려줘	Resfinder
S	한식, 일식, 중식, 양식 중 어떤 것으로 알려 드릴까요?	Resfinder
U	중식으로 알려줘	Resfinder
S	중식으로는 부천성이 가장 많은 추천을 받았습니다.	Resfinder
U	거기 전화번호 알려줘	Resfinder
S	부천성 전화번호는 ~입니다.	Resfinder
U	최신가요 100곡 틀어줘	Songfinder

이와 같은 문제를 해결하기 위해, 각각의 도메인에 특화되어 해당 도메인의 특성을 효과적으로 반영하는 도메인 전문가를 구현하고, 입력된 사용자 발화에 대한 도메인을 결정한 뒤, 적합한 도메인 전문가를 통해 대화를 처리하도록 하는 다중 도메인 대화 시스템이 개발되었다[3]. 높은 성능의 다중 도메인 대화 시스템을 구축하기 위해서는 사용자의 입력을 적합한 도메인으로 분류하는 과정에서 높은 정확도와 함께 자연스러운 도메인 전환이 요구된다.

2. 관련연구

다중 도메인 대화 시스템을 위한 도메인 결정 방법과 관련된 기존 연구는 다음과 같다.

초기 접근 방법으로, 주어진 발화의 각 도메인 별 음성 인식 모델 점수에 기반하여, 가장 높은 점수를 얻은 도메인을 선택하는 도메인 결정 방법에 관한 연구가 진행되었다[4]. 하지만 이 방법의 경우, 대화의 흐름을 고려하지 않고 단일 발화만을 기준으로 도메인을 결정하며, 음성 인식 모델 점수 이외의 자질들이 반영되기 어렵기 때문에, 도메인 결정 성능상의 한계를 지닌다.

도메인 결정 과정에서 대화의 흐름을 반영하지 않고 단일 발화만을 고려하는 경우, 도메인 결정 오류로 인해 의도치 않은 도메인 전환이 일어날 수 있다. 이 문제를 해결하기 위해, 도메인 별 모델 점수 계산 시에 이전 발화의 도메인에 가중치를 주는 방법[5]과, 한 도메인에서의 대화가 종료되기 전까지는 도메인 전환을 허용하지 않는 방법[6]이 제안되었다. 하지만 이 방법들은 여전히 현 발화의 도메인 결정 과정에서 고려될 수 있는 다양한 요인들을 반영하는데 한계를 지니며, 이전 발화의 도메인 결정 결과에 지나치게 높은 의존도를 보임으로써, 단일 발화에 대한 도메인 결정 오류가 발생할 경우, 뒤에 이어지는 대화 전체에 악영향을 미칠 수 있다.

도메인 결정에 영향을 미치는 다양한 자질들을 효과적으로 반영하기 위해, 도메인 결정 문제를 통계적 분류 기법으로 해결하고자 하는 연구가 진행되었다[7]. 이 연구에서는 주어진 발화에 포함된 단어의 표층 정보 뿐만 아니라 품사, 화행, 키워드 등의 다양한 자질들을 반영함으로써, 도메인 결정 성능의 향상을 꾀하였다. 하지만 이 방법론 역시 현재 사용자 발화에 대한 자질들만을 이용하여 도메인을 결정함으로써 잦은 도메인 변경이 발생할 수 있다는 문제점을 지닌다.

본 연구에서는 기존에 적용되었던 다양한 자질들과 더불어 이전 발화의 도메인을 함께 고려함으로써 무분별한 도메인 전환을 지양하고, 도메인 분류 성능을 향상시킬 수 있는 도메인 결정 방법을 제안한다.

3. 다중 도메인 대화시스템 구조

[그림 1]은 다중 도메인 대화 시스템의 전체적인 구조를 나타낸다.

주어진 사용자 음성 발화를 음성 인식 및 문자 변환을 통하여 텍스트로 변환한 뒤, 일반적 자연어 이해 모듈(General SLU)을 통하여 해당 발화에 대한 자질들을 추출한다. 도메인 결정 모듈에서는 자연어 이해 모듈에서 추출된 자질들 및 이전 발화의 도메인 정보를 이용하여 사용자의 발화가 어떤 대화 시스템에서 처리되어야 할지 결정한다. 도메인이 결정된 후, 해당 도메인에 특화된 자연어 이해 모듈을 이용하여 도메인 특화 자질들을 추출하고, 그 결과들을 바탕으로 도메인 전문가가 이전 발화 정보, 대화 예제 DB, 컨텐츠 DB등을 이용하여 사용자발화에 대한 적절한 시스템 응답을 생성한다.

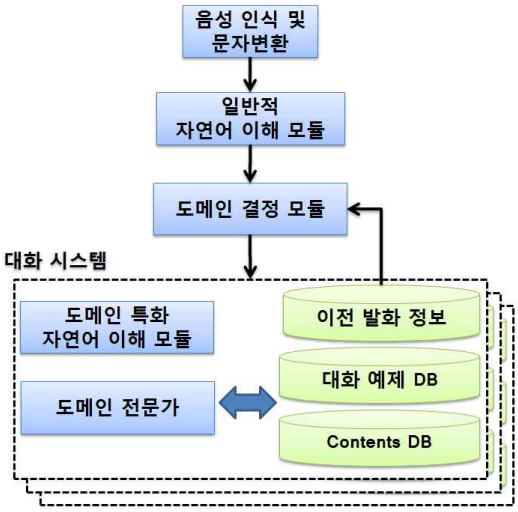

그림 1 다중 도메인 대화 시스템 구조

4. 도메인 결정 방법

도메인 결정을 하기 위해서 통계적 분류 모델을 기반하였으며, 여기서는 최대 엔트로피 모델을 사용하였다. 도메인 결정을 위해 사용된 자질들은 [표 2]와 같다.

단어는 음성 인식 및 문자 변환을 통하여 얻어진 문장에서 공백을 기준으로 구분된 각각의 단어를 의미한다.

표 2 도메인 결정을 위해 사용되는 자질들

자질	설명
단어	사용자 발화에 포함된 단어
화행(Dialog Act)	사용자 발화의 의도
이전 발화의 도메인정보	이전 발화가 속한 도메인

화행은 문장에 대한 일반적인 특성을 의미한다. 이전 발화에 대한 도메인 정보는 사용자의 발화 바로 이전의 발화가 처리된 도메인 정보를 의미한다. 해당 내용을 자질로 사용함으로써 도메인 결정 성능을 향상시킬 수 있다.

5. 실험 및 결과

5.1 대화 말뭉치

4개의 도메인(TVguide, Weather, Resfinder, Songfinder)이 사용되었다. 'TVguide' 는 기본적인 TV 기능 동작 및 방영 프로그램 검색/추천 등을 수행하는 도메인이며, 'Weather' 는 날씨 정보 검색, 'Resfinder' 는 사용자 주변 음식점 추천, 'Songfinder' 는 노래, 앨범, 가수 등의 정보를 검색/추천하기 위한 도메인이다. 도메인 결정 실험 진행을 위해 주어진 말뭉치를 임의의 순서로 나열한 뒤, [표 3]의 분포에 따라 학습 데이터와 평가 데이터를 구성하였다.

5.2 도메인 결정 실험 진행

통계적 분류 모델을 기반한 최대 엔트로피 모델을 이용하여 도메인들을 구분하였으며, 이 때 사용된 단어 정

표 3 도메인 별 사용된 학습, 평가 말뭉치 수

도메인	학습 발화	평가 발화
TVguide	464	97
Songfinder	30	11
Resfinder	77	26
Weather	35	17
Total	606	151

보, 화행(Dialog Act) 정보, 이전 발화의 도메인 정보 등의 자질들을 사용하였다. 사용한 화행 정보와 도메인 정보 자질 집합은 [표 4]와 같다.

표 4 사용된 화행 정보 및 도메인 자질정보

화행	도메인
wh_question	TVguide
request	Songfinder
hope	Resfinder
yn_question	Weather
statement	

주어진 사용자 발화에 대해 일반적 자연어 이해 모듈에서 최대 앤트로피 모델을 사용하여 화행 자질을 추출하고, 이전 발화의 도메인 결정 결과와의 조합을 통해 자질 벡터를 구성한다. [표 5]예제는 도메인 분류를 위한 자질 추출의 예제이다.

표 5 도메인 분류 위한 자질 추출 예

사용자 발화	화행	이전발화도메인
지금 TV에서 뭐하지?	wh_question	None
MBC프로그램 녹화해줘	request	TVguide
오늘 서울에 비 올까?	yn_question	TVguide
근처 맛집 좀 알려줘	request	Weather
중식으로 알려줘	request	Resfinder
거기 전화번호 알려줘	request	Resfinder
최신가요 100곡 틀어줘	request	Resfinder

도메인 결정 성능 평가를 위해 평가 데이터의 실제 도메인 정보와 도메인 분류 모델을 통해 결정된 도메인을 비교하여 정확도(accuracy)를 측정하였으며, 각기 다른 자질 조합에 따른 평가 결과는 [표 6]과 같다.

표 6 도메인 결정 평가 결과(정확도)

자질 조합	정확도(일치발화/전체발화)
단어	92.05% (139/151)
단어, 화행	93.37% (141/151)
단어, 이전발화도메인	95.36% (144/151)
단어, 화행, 이전발화도메인	96.02% (145/151)

실험 결과, 단어 자질만 사용했을 때에 비해, 화행 자질과 이전 발화의 도메인 자질을 추가하였을 때 각각 성능 향상을 보였으며, 화행 자질과 이전 도메인 자질을 함께 적용한 경우, 가장 높은 성능을 보임을 확인할 수 있었다.

6. 결론

이 논문에서는 다중 도메인 대화 시스템에서 자연스러운 시스템 응답을 생성하기 위해 사용자의 의도에 맞는 대화 시스템으로 사용자의 발화를 분류하기 위한 도메인 결정 방법에 대하여 소개하였다. 이는 다중 도메인 대화 시스템에서 사용자의 발화를 인지하여 가장 적합한 응답을 제공할 수 있는 도메인 전문가에게 발화를 분류할 뿐 아니라, 이전 발화의 정보를 고려함으로서 무분별한 도메인 전환을 방지하고 기존의 연구보다 자연스러운 시스템 응답을 생성할 수 있다. 발화를 가장 적합한 도메인 전문가로 분류하기 위해서 단어 정보 뿐만 아니라, 문장에 대한 화행 정보, 이전 발화의 도메인 정보를 이용함으로써 좀 더 정확한 발화 분류를 할 수 있다. 향후 연구 계획으로는, 더 다양한 자질들을 반영함으로써 기존 방법을 통해서 정확히 분류할 수 없었던 발화에 대한 처리가 가능한 방법에 대한 연구를 진행할 예정이다.

7. Acknowledge

"본 연구는 지식경제부 및 정보통신산업진흥원의 대학 IT연구센터 육성지원사업의 연구결과로 수행되었음" (NIPA-2011-(C1090-1121-0008))

참고문헌

[1] Zue, V., Seneff, S., Glass, J., Polifroni, J., Pao, C., Hazen, T., Hetherington L., 2000. JUPITER: a telephone-based conversational interface for weather information. IEEE Trans. Speech Audio Process. 8, 85-96.

[2] Walker, M., Aberdeen, J., Boland, J., Bratt, E., Garofolo, J., Hirschman, L., Le, A., Lee, S., Narayanan, S., Papineni, K., Pellom, B., Polifroni, J., Potamianos, A., Prabhu, P., Rudnicky, A., Sanders, G., Seneff, S., Stallard, D., Whittaker, S., 2001. DARPA communicator dialog travel planning systems: the June 2000 data collection. In: Proc. Eur. Conf. on Speech, Communication and Technology, pp. 1371-1374.

[3] O'neill, I., Hanna, P., Liu, X., Greer, D., McTear, M., 2005. Implementing advanced spoken dialogue management in Java. Speech Comm. 54, 99-124.

[4] T. Isobe, S. Hayakawa, H. Murao, T. Mizutani, K. Takeda, and F. Itajura. 2003. A study on domain recognition of spoken dialogue systems. In Proc. Eurospeech-2003, pages 1889-1892.

[5] Bor-shen Lin, Hsin-ming Wang, and Lin-shan Lee. 1999. A distributed architecture for cooperative spoken dialogue agents with coherent dialogue state and history. In Proc. ASRU-99.

[6] Ian O'Neill, Philip Hanna, Xingjun Liu, and Michael McTear. 2004. Cross domain dialogue modelling: an object-based approach. In Proc. Interspeech-2004(ICSLP), pages 205-208.

[7] Cheongjae Lee, Sangkeun Jung, Seokhwan Kim, and Gary Geunbae Lee. 2009. Example-based dialog modeling for practical multi-domain dialog system. *Speech Communication*, 51(5):466-484.