

사람과 서비스 로봇의 상호주도형 의사소통을 위한 계층적 베이지안 네트워크

송윤석⁰, 홍진혁, 조성배
연세대학교 컴퓨터과학과

corlary@sclab.yonsei.ac.kr, hjinh@sclab.yonsei.ac.kr, sbcho@csai.yonsei.ac.kr

Hierarchical Bayesian Networks for Mixed-Initiative Interaction between Human and Service Robot

Youn-Suk Song⁰, Jin-Hyuk Hong and Sung-Bae Cho
Dept. of Computer Science, Yonsei University

요 약

서비스 로봇은 일상생활에서 사람들의 업무를 보조한다. 이때, 효과적인 서비스를 위해서는 사람과 로봇 사이의 상호작용이 매우 중요하다. 대화는 사람과 로봇이 보다 유연하고 풍부한 의사전달을 하는데 도움을 준다. 전통적인 로봇 연구에서는 명령과 같은 간단한 질의 등을 처리하는 것이 의사소통의 전부였으나, 실제 사람들 사이의 대화에서는 배경 지식이나 대화의 문맥 등에 의해 중요한 정보가 대화에서 생략되기도 한다. 이런 상황은 여러 불확실성을 포함하게 되는데 대화의 문맥이나 불확실성을 다루는 것이 필요하다. 본 논문에서는 ‘상호-주도’ 방식을 통해 사람이 쓰는 일상 대화를 계층적 베이지안 네트워크를 이용하여 처리하는 방법을 제안한다. 실제 로봇의 시뮬레이션 환경은 제안하는 방법의 유용함을 보여주었다.

1. 서론

과거 자율 로봇은 단순히 사람의 지시를 받는 정교한 도구로 여겨졌고 사람과 로봇의 의사소통은 단순하게 이루어졌다[1]. 하지만 최근 집과 사무실에서 서비스 로봇의 사용이 중요해져 감에 따라 좀더 사람에게 친숙한 의사 소통 방법이 요구되고 있다. 자연 언어를 통한 의사 소통은 사람과 로봇의 상호 작용에서 효과적인 방법으로 최근 주목 받고 있다[2,3,4]. 그러나 일상 대화 속에서는 종종 “너의 학교 이름이 무엇이니?” 보다는 “이름이 뭐니?”처럼 구체적인 단어가 생략되거나 모호한 표현이 사용되기도 한다. 사람들은 대화 속에서 배경 지식이나 문맥을 사용하기 때문에 이런 내용들을 이해한다[5,6]. 따라서 자연스러운 대화를 위해서는 로봇이 대화에서의 모호함과 불명확함을 처리하는 것이 필요하다. 최근 Human-Computer-Interaction (HCI) 분야에서는 사람과 로봇이 서로 도와나가며 문제를 해결하는 상호-주도적 접근 방법이 연구되고 있다. 본 논문에서는 계층적 베이지안 모델을 이용한 상호-주도적 접근 기술을 구현하고 홈 서비스를 위한 지능형 로봇과의 의사 소통 방법을 제안하고자 한다.

2. 배경

2.1 상호 주도 방식

HRI(Human-Robot-Interaction)에서는 자연언어를 통한 다양한 의사소통 방법이 연구되어 왔지만 대부분 실제 대화가 아닌 단순한 명령 수준에서 이루어졌다. HCI 분야에서도 여러 종류의 대화 처리기법들이 연구되어 왔는데[7,8,9,10] 그 중 전통적인 대화 모델인 시스템-주도형과 사용자-주도형 모델이 널리 사용되었다. 시스템-주도형은 완전히 짜여진 방식에 따라 절차적인 순서를 사용자에게 제공해주는 반면, 사용자-주도형은 사용자가 원하는 방향으로 행동 및 서비스를 결정하게 된다. 두 가지 방법은 모두 완수율이나 사용자들 돕는 부분에서 한계점을 가진다[11].

상호주도 모델은 시스템과 사용자가 협력을 통해서 점진적으로 문제를 풀어가는 방법이다.[12] Meng은 신뢰

네트워크와 후진 추론을 사용해서 상호 주도 모델을 제안했었고[13] Bechat은 음성 인식과 언어 이해 사이에 두 단계의 프로세스를 만들고 간격을 연결 짓는 방법을 시도하기도 했었다[14]. 계층적 베이지안 방식은 대화의 문맥과 자연언어의 모호함을 처리하기 위해서 시도되어 왔다[15].

2.2 계층적 베이지안 네트워크

자연 언어와 상황으로부터 사용자의 의도를 이해하기 위해서는 질의와 상황을 특정 도메인에 대한 의미 요소(primitive-semantics)로 분석해서 단서(evidence)를 얻고 이를 바탕으로 로봇의 서비스 목표(service-goal)를 추론한다. 시스템은 제한된 도메인 안에서 동작하므로 이런 의미 요소와 서비스 목표 수는 제한적이다. 특별히 상호-주도적 의사 소통을 위해서 의미 요소와 서비스 목표 사이에 중간 목표(sub-goal)층을 넣는다. 중간 목표는 한 단계 더 추론을 진행시키기 위해서 사용된다. 한 번에 서비스 목표를 결정하기 어려운 경우 중간-목표가 평가 되고 상호-주도적인 대화를 통해서 다시 서비스 목표를 정하게 된다. 그림1에 이런 계층적 베이지안 네트워크의 간략한 구조가 나와있다.

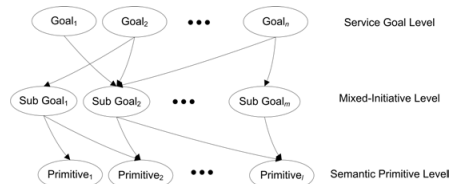


그림1. 상호-주도를 위한 계층적 베이지안 네트워크

목표 추론은 중간 목표 추론과 서비스 목표 추론으로 수행된다. 중간 목표는 의미 요소로부터 추론되고 서비스 목표는 중간 목표로부터 추론되게 된다. 본 논문에서는 다음 식을 바탕으로 각 목표의 확률 값을 추정한다.

$$\begin{aligned}
 P(G_i = 1 | SG) &= \frac{P(SG | G_i = 1)P(G_i = 1)}{P(SG)} \\
 &= \frac{\prod_{k=1}^M P(SG_k | G_i = 1)P(G_i = 1)}{\prod_{k=1}^M P(SG_k | G_i = 0)P(G_i = 0) + \prod_{k=1}^M P(SG_k | G_i = 1)P(G_i = 1)} \\
 P(SG_i = 1 | Pr) &= \frac{P(Pr | SG_i = 1)P(SG_i = 1)}{P(Pr)} \\
 &= \frac{\prod_{k=1}^L P(Pr_k | SG_i = 1)P(SG_i = 1)}{\prod_{k=1}^L P(Pr_k | SG_i = 0)P(SG_i = 0) + \prod_{k=1}^L P(Pr_k | SG_i = 1)P(SG_i = 1)}
 \end{aligned}$$

3. 상호 주도 방식 서비스 로봇

3.1 홈 서비스 환경

본 논문에서는 제안하는 방법을 이용하여 대화형 가정용 서비스를 시뮬레이션 해보았다. 서비스 로봇은 불을 켜거나 끄거나 창을 열거나 닫거나 오디오, TV, 에어컨을 켜고 끈다. 이들 서비스는 우선 사용자의 명령으로 시작되고 서비스 로봇은 필요한 정보를 추출한다. 환경에 민감한 서비스 추론을 위해서 환경 정보를 추가로 고려한다. 그림 2는 본 논문에서 구성한 시뮬레이션 환경을 보여준다.



그림2. 제안된 방법의 기능을 위해서 구성된 집 환경

3.2 상호 주도 방식에 기초한 서비스 결정

서비스 로봇은 2단계로 서비스를 결정한다. 첫 번째 단계에서는 서비스 종류를 결정한다. 로봇이 사용자로부터 “에어컨을 켜라” 라는 질의를 받으면 (에어컨, 켜라) 와 같은 의미 요소 노드에 ‘true’ 값을 할당 하고 ‘에어컨을 켜라’ 가 서비스 종류로 결정 된다. 하지만 “너무 덥다” 와 같은 불명확한 표현을 받았을 때는 서비스 종류를 직접 결정하기가 어려워진다. 이런 경우 로봇은 상호-주도 방식에 기반해서, 사용자의 의도를 명확히 이해하기 위해 추가적인 정보 요청을 하게 된다. “덥다” 라는 표현에 대해서 로봇은 현재 환경을 고려한 다음 “창문을 열까요? 아니면 에어컨을 켤까요?” 와 같은 질문을 사용자에게 제공한다. 그림 3은 서비스 종류를 결정하기 위해 설계 된 계층적 베이지안 네트워크의 모습을 보여준다.

두 번째 단계에서는 대상을 결정한다. 전 단계에서 창문을 여는 서비스가 결정되었다면 로봇은 어떤 창문을 열어야 할지를 추론하게 된다. 이 때 사용자의 표현이 “열 방에 있는 창 문을 열어라” 처럼 동시에 여러 개의 대상을 가리키게 될 수도 있으며 로봇은 상호-주도 방식에 의해 ‘열 방’ 에 해당되는 방 이름으로 사용자와 대화를 하게 되며 정확한 방 정보를 수집한 후 서비스를 수행한다.

3.3 지능형 서비스를 위한 컨텍스트 모델링

본 논문에서는 사용자의 의도를 파악하고 보다 지능적인 서비스를 위해서 환경 정보를 이용한다. 이러한 환경 정보는 사용자의 모호한 표현에 대한 보다 적절한 서비스를 결정하는데 도움을 준다. 본 논문에서 사용된 컨텍스트 정보는 다음 표1과 같다.

표1. 환경 속성

속성	값
밝기	밖이 더 밝다, 안이 더 밝다
온도	밖이 더 덥다, 안이 더 덥다
소음	밖이 더 시끄럽다, 안이 더 시끄럽다.
창문 상태	열려있다. 닫혀있다.
오디오 상태	켜져있다. 꺼져있다.
TV 상태	켜져있다. 꺼져있다.
에어컨 상태	켜져있다. 꺼져있다.

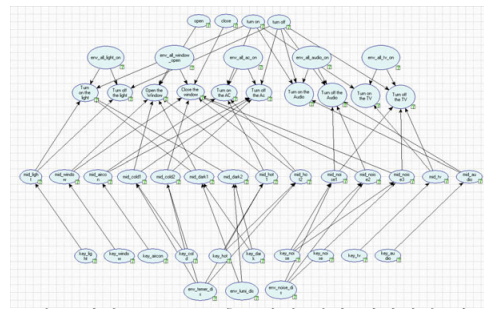


그림3. 서비스 종류를 추론하기 위한 베이지안 네트워크

4. 실험 결과

4.1 실험 환경

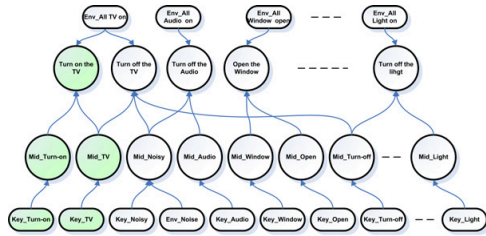


그림4. 실제 응용을 위한 실험 환경

캐페라를 이용하여 본 논문에서는 그림 4와 같은 시뮬레이션 환경을 설계하였다. 환경은 방 6개, 불 6개, 창 6개, TV 3대, 오디오 3대 그리고 에어컨 1대로 구성되어 있다. 서비스 종류는 (불 켜기/끄기), (창 닫기/열기), (TV 켜기/끄기), (오디오 켜기/끄기), (에어컨 켜기/끄기) 가 있다.

4.2 서비스 수행 시나리오

‘TV를 켜라’ 서비스가 선택되는 과정이 그림 5에 나와있다. 서비스 종류를 선택하는 과정 후 다시 구체적인 대상인 TV를 결정하는 추론이 이어진다. 그림5에서 보듯이 첫 단계에서는 (TV, 켜라)를 통해 의미 노드가 ‘true’로 세팅되고 중간 목표 노드의 확률 값과 서비스 목표 노드의 확률 값이 정해진다. 최종적으로 ‘TV를 켜라’ 라는 서비스 종류가 결정되면 두 번째 단계를 추론하게 된다. 이 때 로봇은 환경 정보와 사용자가 제공한 정보를 통해서 질문을 하거나 대상이 되는 TV를 결정하고 서비스를 수행하게 된다. 그림6은 사용자가 ‘F방’이라는 응답을 하고 F방에 있는 TV를 켜는 서비스가 선택되는 과정을 보여준다.



ALL_0n	YES	NO	Turn on the TV
M_Id_TV	YES	NO	YES
Turn-on	YES	NO	YES
YES	0.1	0.1	0.9
NO	0.9	0.9	0.1

그림5. 서비스 종류가 추론되는 과정

Env_0n	YES	NO	Turn on the TV in Room
M_Id_Room F	YES	NO	YES
YES	0.1	0.1	0.9
NO	0.9	0.9	0.1

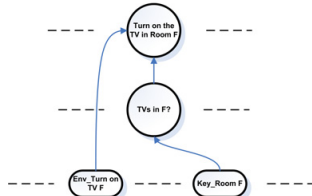


그림6. 대상이 추론되는 과정

4.3 모호한 상황 해결의 대화 시나리오

사용자: 여기가 너무 시끄럽다.
 로봇: TV를 끌까요, 오디오를 끌까요?
 ->의미 요소(시끄럽다)
 ->서비스 추론(중간 목표: 시끄럽다)
 ->상호-주도 대화 결과
 사용자: 옆방에 있는 오디오 말이야.
 로봇: 옆방이 어디죠? C번방인가요 F번방 인가요?
 ->의미 요소(오디오, 옆방)
 ->서비스 대상 추론(중간 목표: 방B 옆)
 ->컨텍스트(현재 B번 방에 위치)
 컨텍스트(오디오, 방A, 꺼짐)
 컨텍스트(오디오, 방B, 꺼짐)
 컨텍스트(오디오, 방F, 꺼짐)
 ->상호 주도 대화
 사용자: C번 방 부탁해.
 로봇: C번 방에 있는 오디오를 끄겠습니다.
 ->이전 의미 요소(시끄럽다)
 ->의미 요소(오디오, 방C)
 ->서비스 기능 추론(오디오를 꺼라)
 ->서비스 대상 추론(C번 방 오디오))
 ->서비스: C번 방에 있는 오디오를 꺼라

대화1. 모호한 정보로부터 서비스를 제공하는 예

위 대화는 제안된 방법의 상황 인식 및 서비스 추출에서의 유용함을 보여준다. 일상 생활에서 발생할 수 있는 사용자의 모호한 표현을 상호-주도 방식을 통해 더 유연하고 정확하게 풀어가게 된다.

5 결론

본 연구의 동기는 보다 원활하게 로봇과 대화하는 효과적인 인터페이스를 만들어 보는데 있다. 이를 위해서 상호-주도 방식으로 그 서비스 하는 로봇을 설계하였고 시뮬레이션을 통해서 그 유용함을 보였다. 향후에는 보다 다양한 센서와 모듈을 이용하여 실제 환경을 적절히 모델링하며 지능적 서비스 수행이 가능하도록 하고자 한다.

감사의 글

이 연구는 과학기술부가 지원한 뇌과학 연구 프로그램에 의해 지원되었음.

참고문헌

- [1] V. Zue and J. Glass, "Conversational interfaces: Advances and challenges," *Proc. of the IEEE, Special Issue on Spoken Language Processing*, vol. 88, no. 8, pp. 1166-1180, 2000.
- [2] V. Zue and J. Glass, "Conversational interfaces: Advances and challenges," *Proc. of the IEEE, Special Issue on Spoken Language Processing*, vol. 88, no. 8, pp. 1166-1180, 2000.
- [3] H. Prendinger and M. Ishizuka, "Let's talk! Socially intelligent agents for language conversation training," *IEEE Tans. on Systems, Man and Cybernetics, Part A*, vol. 31, no. 5, pp. 465-471, 2001.
- [4] D. Sanford and J. Roach, "A theory of dialogue structures to help manage human-computer interaction," *IEEE Trans. on Systems, Man and Cybernetics*, vol. 18, no. 4, pp. 567-574, 1988.
- [5] H. Meng, et al., "The use of belief networks for mixed-initiative dialog modeling," *IEEE Trans. On Speech and Audio Processing*, vol. 11, no. 6, pp. 757-773, 2003.
- [6] C.Sammut, "Managing context in a conversational agent," *Electronic Trans. on Artificial Intelligence*, vol. 5, pp. 189-202, 2001.
- [7] E. Horvitz and T. Paek, "A computational architecture for conversation," *Proc. of the 7th Int. Conf. on User Modeling*, pp.201-210, 1999.
- [8] V. Zue, et al., "JUPITER: A telephone-based conversational interface for weather information," *IEEE Trans. On Speech and Audio Processing*, vol. 8, no. 1, pp. 85-96, 2000.
- [9] F. Rosis, et al., "From Greta's mind to her face: modelling the dynamics of affective states in a conversational embodied agent," *Int. J. Human-Computer Studies*, vol. 59, no. 1-2, pp. 81-118, 2003.
- [10] W. Clancey, "Roles for agent assistants in field science: Understanding personal projects and collaboration," *IEEE Trans. On Systems, Man, and Cybernetics-Part C: Applications and Reviews*, vol. 34, no. 2, 125-137, 2004.
- [11] E. Horvitz, et al., "The lumiere project: Bayesian user modeling for inferring the goals and needs of software users," *Proc. of the 14th Conf. on Uncertainty in Artificial Intelligence*, pp. 256-265, 1998.
- [12] E. Horvitz, "Uncertainty, action, and interaction: In pursuit of mixed-initiative computing," *IEEE Intelligent Systems*, vol. 14, no. 5, pp. 17-20, 1999.
- [13] H. Meng, et al., "The use of belief networks for mixed-initiative dialog modeling," *IEEE Trans. On Speech and Audio Processing*, vol. 11, no. 6, pp. 757-773, 2003.
- [14] F. Bechet, et al., "Detecting and extracting named entities from spontaneous speech in a mixed-initiative spoken dialogue context: How may I help you?" *Speech Communication*, vol. 42, no. 2, pp. 207-225, 2004.
- [15] J.-H. Hong and S.-B. Cho, "A two-stage Bayesian network for effective development of conversational agent," *Lecture Notes in Computer Science*, vol. 2690, pp. 1-9, 2003.