

POMDP 기반 사용자-로봇 인터랙션 행동 모델

POMDP-based Human-Robot Interaction Behavior Model

김 종 철*
(Jong-Cheol Kim^{1,*})

¹Institute of Convergence Technology, KT

Abstract: This paper presents the interactive behavior modeling method based on POMDP (Partially Observable Markov Decision Process) for HRI (Human-Robot Interaction). HRI seems similar to conversational interaction in point of interaction between human and a robot. The POMDP has been popularly used in conversational interaction system. The POMDP can efficiently handle uncertainty of observable variables in conversational interaction system. In this paper, the input variables of the proposed conversational HRI system in POMDP are the input information of sensors and the log of used service. The output variables of system are the name of robot behaviors. The robot behavior presents the motion occurred from LED, LCD, Motor, sound. The suggested conversational POMDP-based HRI system was applied to an emotional robot KIBOT. In the result of human-KIBOT interaction, this system shows the flexible robot behavior in real world.

Keywords: POMDP, human-robot interaction, KIBOT robot

I. 서론

대화시스템은 인간과 시스템간의 의사 소통의 매개체 역할로 다양한 영역에서 적용되고 있다. 특히 통계학적 확률에 기반한 부분관측 마르코프 의사결정 과정(POMDP: Partially Observable Markov Decision Process)는 실생활에서 많이 경험하는 불확실성(uncertainty)에 대한 대처 능력으로 많은 분야에서 활발히 응용되고 있다[1,2]. 산업응용 분야에서 기계 유지보수나 구조 결함 검사, 과학응용 분야에서는 자율이동 로봇이나 머신 비전, 상업응용 분야에서는 마케팅, 군사응용 분야에서는 움직이는 목표물 분류 등에 연구되고 있다.

서비스로봇 분야에서 기존에 사람이 간접적으로 조작에 개입하거나 미리 입력된 예정된 행동을 수행하게 하였지만, 점차 사람과 로봇 간의 상호작용성(HRI: Human Robot Interaction) 연구에 중점을 두고 있다[3-7]. 현재 다양한 HRI 방법이 연구되고 있지만, 사람이 로봇에 살아있는 애완동물처럼 친근감을 갖기 위해서는 마치 서로가 대화하는 것처럼 커뮤니케이션 하는 것이 핵심이다. AIBO 로봇은 애완동물인 강아지의 행동 패턴을 롤기반으로하여 유년기로부터 장년기의 성장 개념을 도입하였고, 센서 등의 사용자의 입력에 성장 단계별로 상황에 맞는 기쁨, 화남 등의 다양한 행동을 표현하게 된다 [4]. PARO는 바다표범 로봇으로 쓰다듬어 주거나 먹이를 주는 등의 사용자 행동에 대해서 감성적인 인터랙션을 통해서 인지증 증상을 개선시킬 수 있다[5]. 또 PaPeRo 로봇은 가정용 서비스 로봇으로 음성인식엔진을 통해 다양한 커뮤니케이션이 가능하고, 사람의 성격을 가진 감성적 캐릭터로 행동을 표현할 수도 있다[6].

국내에서도 유아용 에듀테인먼트 키봇(KIBOT)로봇을 대상

으로 HRI 연구가 활발히 진행되고 있다[8-11]. KIBOT은 취학 전의 아동을 대상으로 하며, RFID 태그가 부착된 엄마, 아빠 카드로 영상통화를 걸 수 있고, 동화책 읽어주기나 낱말카드 놀이 등을 제공한다. 또 스스로 장애물을 회피하면서 자율주행이 가능한 기능도 탑재되어 있다[8,11]. KIBOT에 살아있는 생명체처럼 생동감 있는 행동을 발현하기 위해서 KIBOT의 마이크에서 들어오는 주위의 소리를 통해서 사용자의 감성을 추정하고 행동을 표현하는 연구가 진행되어 왔다[9]. 또, KIBOT은 사용자와의 상황에 맞는 다양한 인터랙션을 제공하기 위해서 상황별 시나리오를 스크립트화하여 다양한 행동을 표현하기 위한 연구도 진행되어 왔다. 이러한 스크립트 방식은 롤기반과 달리 발생할 상황별로 스크립트화하여 사용자와 로봇간액션과 리액션으로 상호작용할 수 있는 장점이 있다. 스크립트 방식은 정해진 상황에 따른 행동을 규정하기 때문에 상황이 애매모호한 경우에는 전혀 엉뚱한 로봇행동을 표출할 수 있고, 다양한 행동을 분류하는 시나리오를 사전에 전부 작성해야 하는 특징이 있다. 이러한 방식은 사용자의 특정 입력에 로봇의 출력을 사전에 정의하기 때문에 로봇의 행동이 단순 반복적으로 되풀이될 수 있어서, 사용자와 커뮤니케이션하는 인터랙션이 아니라 지시한 바를 출력하는 단순 작업이 이루어지는 단점이 있다. 또 로봇으로 입력되는 정보에 센서 오류나 동작 오류, 상황 인식 오류 등이 발생할 경우에 전혀 대처하지 못하고, 사용자의 입력을 오관하여 전혀 기대하지 않는 행동을 출력함으로써 사용자와의 커뮤니케이션을 저해하는 문제점이 있다. 이와 같은 문제점을 해결하기 위해서 사용자의 입력 및 주변 상황 인식에 대해서 오류가 있다고 할지라도 상황에 적합한 다양한 로봇의 행동을 발현할 수 있는 통계학적 학습추론 방법인 POMDP를 도입하였다.

본 논문에서는 사용자와 로봇간 상황에 맞는 상호작용적 유연한 커뮤니케이션 수행을 목적으로 POMDP를 이용한 사용자와 로봇간의 행동중심의 대화형 인터랙션 시스템을 제안한다.

* Corresponding Author

Manuscript received February 15, 2014 / revised March 15, 2014 / accepted March 30, 2014

김종철: Service Laboratory(jong-cheol.kim@kt.com)

※ 본 논문은 음성학회 봄학술대회에서 초안이 발표되었음.

제안된 사용자-로봇 인터랙션 시스템은 사용자의 입력에 대한 센서 오차를 미리 반영하여 사용자 입력이 애매모호한 경우에도 로봇이 한가지 행동만 발현하는 것이 아니라 확률에 기반한 몇 가지 행동을 표현할 수 있다. 또 사전에 모든 시나리오를 작성할 필요 없이 크게 예측가능한 시나리오를 기반으로 상호작용 행동모델을 학습시키면 된다. 제안된 사용자-로봇의 상호작용 시스템은 사용자가 로봇을 사용할 때 로봇에 탑재된 센서 정보나 로봇서비스의 사용이력이 입력이 되고, 로봇 행동이 출력으로 된다. 로봇 행동은 모션으로 나타나는 불빛, LCD, 사운드와 함께 움직임으로 표출된다.

본 논문의 구성은 다음과 같다. II 장에서는 제안된 시스템이 적용될 KIBOT의 하드웨어와 소프트웨어의 구성을 소개한다. III 장에서는 POMDP 기반 사용자-로봇 상호작용 시스템에 대해서 자세히 설명한다. IV 장에서는 제안된 시스템에 대한 다양한 행동 발현 시뮬레이션과 실험 결과를 조사한다. V 장에서는 제안된 시스템에 대한 결과를 요약하도록 한다.

II. 유아용 에듀테인먼트 KIBOT 로봇

유아용 에듀테인먼트 KIBOT은 유아들에게 유익하면서도 재미있게 친구같이 놀아주는 원숭이 캐릭터의 로봇이다. 크기는 가로, 세로, 높이는 각각 20cm, 14 cm, 23 cm이고, 무게는 약 2kg이다. KIBOT은 전후좌우 이동이 가능하며, 상하좌우 목도 움직일 수도 있다. 감성적 표현을 위해서 머리, 이마, 입과 양쪽 볼에 LED가 도입되었다. 또 몸통 정면에는 3.5인치의 LCD가 위치하여 다양한 어플리케이션 정보를 시각적으로 제공해 준다. 외부환경을 인식하고, 사용자와의 인터랙션을 위해서 터치감지, 거리감지, 바닥감지, 소음감지, 열감지, 기울기감지, 조도감지 등의 센서가 탑재되었다. 유아에게 사



(a) The sensors position in front.



(b) The sensors position in back.

그림 1. KIBOT 로봇 외형과 센서 위치.

Fig. 1. The KIBOT and sensors.

용하기 쉬운 인터페이스를 제공하기 위해서 RFID 태그를 감지할 수 있는 RFID 리더기가 코 안쪽에 부착되었다. 그림 1은 본 논문에서 소개하고자 하는 유아용 에듀테인먼트 KIBOT 로봇의 외형과 각종 센서를 보여준다.

표 1은 KIBOT의 하드웨어 사양을 보여준다. CPU는 ARM11 프로세서를 사용하였고, 모터 제어 및 센서값 처리를 위해서 8비트 마이크로컨트롤러를 이용하였다. 또, 바퀴, 머리, 꼬리 구동을 위해 모터를 사용하였다. 영상통화를 위해서 카메라가 탑재되었으며, 무선네트워크용 무선 USB 모듈도 지원한다. 충전은 리튬이온 배터리를 이용하여 충전하며, 별도의 충전 크레들을 제공한다. KIBOT의 머리, 양쪽 귀, LCD 양쪽에는 버튼이 위치하여 사용자 입력을 받는다.

KIBOT의 소프트웨어 구조는 그림 2에서 보여준다. OS로는 임베디드 리눅스를 사용하고, 로봇의 각종 기능들을 관장하는 로봇 메인프로그램이 존재하며, 사용자와의 상호작용에 의한 로봇의 모션은 로봇의 HRI라이브러리를 사용하도록 설계되었다. 플래시 엔진, 영상통화 엔진과 마찬가지로 HRI 엔진도 별도로 동작하도록 개발 되었다. HRI 엔진은 POMDP 기반 대화형 상호작용 학습엔진을 통해서 미리 학습된 정책 파일과 행동모델 파일을 참조하여 사용자 입력에 따른 다양한 행동을 발현하게 된다.

표 1. KIBOT 로봇의 H/W 사양.

Table 1. The Specification of KIBOT.

항목	내용
CPU	ARM11(S3C6410-667MHz) 2 x ATmega128
Display	3.5 inch TFT LCD(320x240), 16.7M color
Actuator	BLDC motors for wheel, head, tail
Sensor	IR sensors for distance, bottom measure Touch, Tilt, PIR, Noise, Light sensors
Network	USB Dongle Wireless LAN (802.11b/g)
Others	2.0 M pixel CMOS camera RFID reader (13.56 MHz) Li-ion rechargeable battery
Size Weight	20cm x 14 cm x 23 cm (L x W x H) Approximately 2.0Kg

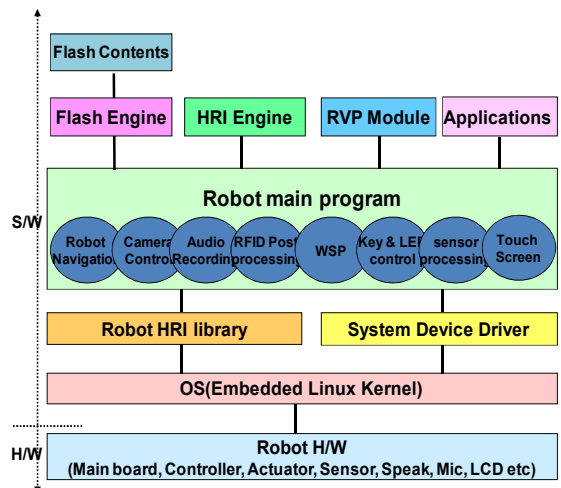


그림 2. KIBOT 로봇의 소프트웨어 구조.

Fig. 2. The S/W structure of KIBOT.

III. POMDP 기반 사용자-로봇 인터랙션 행동 모델

본 장에는 KIBOT과 사용자간의 상호작용 행동표현을 위한 모델링 방법을 설명한다. 우선 POMDP의 일반적인 개념을 소개하고, 학습에 사용된 Symbolic HSVI 알고리즘을 설명한 후, POMDP 기반 사용자-로봇 인터랙션 행동 모델을 상세히 설명하도록 한다.

1. POMDP 모델

POMDP는 부분 관측 마르코프 의사 결정 프로세스로서, 현재의 상황에서 미래의 입력을 예측하여 최적의 행동을 하기 위한 통계적 기반 의사 결정 프로세스이다. POMDP는 실제의 외부 상황을 전부 관찰할 수 없기 때문에 정확히는 모른다고 가정한다. 오차가 있는 사용자 입력에 대해서 측정 가능한 센서로 관측하고, 사용자-로봇 인터랙션 시나리오 기반으로 행동모델을 설계하고, 학습을 통해 시스템 행동에 대한 보상을 최대화 할 수 있도록 훈련시켜 통계기반 최적의 정책을 도출해 낸다.

POMDP는 부분적으로 관측할 수 있는 불확실한 상황을 고려하여 기본적으로 다음의 구성 요소를 통해 모델링 된다 [12-14].

<p>{S,A,T,R,O,Z}</p> <p>-S: 대화의 실제 상태(state) s의 집합</p> <p>-A: 시스템이 취할 수 있는 행동(action) a의 집합</p> <p>-T: 상태 s의 변이 확률 분포</p> <p>-R: 시스템의 행동에 따른 보상값의 기대치 $r(s,a)$의 확률</p> <p>-O: 관측치 o의 집합</p> <p>-Z: 실제 상태에 대한 관측치의 확률 분포 $P(o' s',a)$</p>
--

시점 t의 상태 s의 분포 확률을 $b_t(s)$ 라고 하면, 시점 t에서 시스템 행동을 취함으로써 얻을 수 있는 보상의 기대치는

$$\rho(b_t, a_t) = \sum_{s \in S} b_t(s) r(s, a_t) \tag{1}$$

로 표시될 수 있으며, 다음으로 표현되는 누적 보상값을 최대화 할 수 있도록 최적의 시스템 행동을 도출하는 것이 궁극적인 목표가 된다.

$$\sum_{t=0}^{\infty} \gamma^t \rho(b_t, a_t) = \sum_{t=0}^{\infty} \gamma^t \sum_{s \in S} b_t(s) r(s, a_t) \tag{2}$$

POMDP 모델을 풀기 위한 최적화 알고리즘으로는 Exact Method, Perseus, HSVI, PBVI 등 많은 연구 결과가 있지만, 본 논문에서는 행렬연산에 대한 연산시간을 줄이는데 효율적인 Symbolic HSVI (Heuristic Search Value Iteration) 알고리즘을 기반으로 한다[13]. Symbolic HSVI 방식은 가치함수(Value Function)를 계산해가면서 그 오차를 줄이는데 가장 중요한 요소인 신뢰점(Belief Point)를 우선적으로 탐색해 가는 HSVI 방식에서, 트리구조를 적용하여 ADD (Algebraic Decision Diagram) [13]를 활용하여 대수적 연산을 적용한 방식이다.

또한, POMDP 모델을 간결하게 표현하기 위해서 문제 집합의 상태를 작은 하부 구성 요소로 인수분해 하는 Factored POMDP [15] 방식을 적용하였다. 궁극적인 목표가 뚜렷한 대화

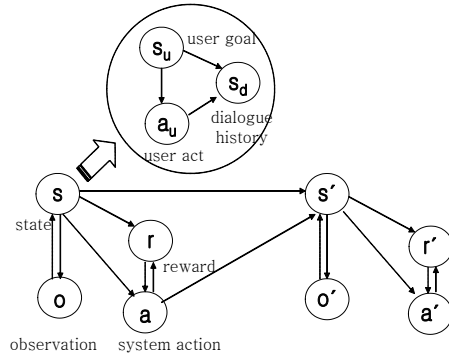


그림 3. Symbolic HSVI 알고리즘이 적용된 모델.
Fig. 3. The model of symbolic HSVI algorithm.

에의 적용 예를 든다면, 사용자의 궁극적 목표 및 의도인 s_u , 현재 사용자의 행동 유형인 a_u , 사용자 입장에서의 대화의 현재 상태를 나타내는 s_d 으로 인수분해 할 수 있다.

$$b'(s'_u, s'_d, a'_u) = k \cdot p(o'|a'_u) \cdot p(a'_u|s'_u, a) \times \sum_{s_u \in S_u} p(s'_u|s_u, a) \times \sum_{s_d \in S_d} p(s'_d|a'_u, s_d, a) \times \sum_{a_u \in A_u} b(s_u, s_d, a_u) \tag{3}$$

이 때의 전체 시스템의 상태 변화도는 그림 3과 같다.

2. POMDP 기반 사용자-로봇 인터랙션 행동 모델링 모델

본 논문에서 제안하는 POMDP 기반 사용자-로봇간 인터랙션 시스템의 구성도를 그림 4에 보여준다. KIBOT의 센서 입력과 서비스 사용이력이 사용자 입력으로 센싱 모듈에서 처리된다. 사용자-로봇 상호작용 시나리오로부터 행동 모델과 보상정책을 설계하여 POMDP 기반 학습기로 학습한다. 학습된 행동모델 파일과 정책파일을 KIBOT에 탑재된 대화형 상호작용 엔진을 통해 사용자 입력을 받아서 최적의 로봇행동 이름을 출력한다. KIBOT 로봇 메인 소프트웨어에서는 상호작용 엔진에서 받은 로봇행동 이름을 HRI 시뮬레이터로부터 미리 생성한 실제 로봇의 모션과 연동하여 표현하게 한다.

POMDP 기반 인터랙션 시스템을 구성하기 위해서는 사용자-로봇간 인터랙션을 위한 행동모델을 설계하여야 한다.

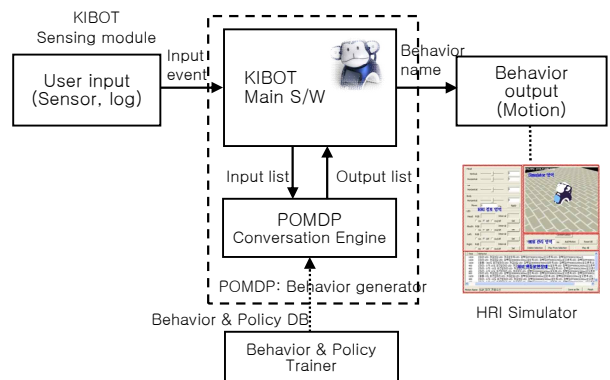


그림 4. POMDP기반 사용자-로봇 인터랙션 행동 표현 시스템.
Fig. 4. POMDP-based HRI behavior system.

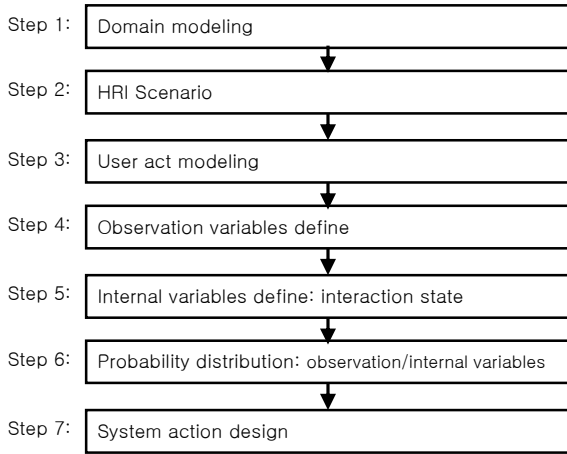


그림 5. POMDP기반 사용자-로봇간 인터랙션을 위한 행동 모델 설계 절차.

Fig. 5. The behavior modeling process for POMDP-based HRI.

그림 5는 이런 행동모델의 설계 절차를 보여준다.

Step1: 사용자의 입력을 받아 최적의 로봇의 행동을 표출하는 HRI 시스템 영역으로 도메인을 모델링한다.

Step2: 사용자와 로봇간의 실제 사용에 있어서 발생할 만한 사용 시나리오를 정의한다. 본 KIBOT에서는 터치, 사람감지, 큰 소음, 흔들기, 서비스 추천에 대한 수락여부 등의 사용자 입력에 대해서 로봇의 출력행동의 시나리오를 작성하였다. 로봇의 출력행동은 기뻐하기, 무서워하기 등의 감성행동이 해당된다.

Step3: 수집된 시나리오로부터 도메인 영역에서 사용자의 행동을 모델링한다. 표 2는 정의된 사용자 행동을 보여준다.

Step4: 관측 변수를 정의한다. 관측 변수는 사용자의 행동 (User Act)으로 인해 KIBOT이 인식할 수 있는 관측의 종류와 가질 수 있는 값을 고려하여 설계한다. 표 3은 KIBOT의 관측 변수를 보여준다. 본 과정은 POMDP 모델 {S,A,T,R,O,Z}에서 관측치 O에 해당한다.

Step5: 내부 변수와 내부 상태를 정의한다. 내부 변수는 실제로 현재 인터랙션의 상태를 나타내는 팩터들로 구성한다. 위에서 설계된 사용자의 행동도 내부 변수에 해당되며, 그 외 시스템의 현재 상태, 히스토리, 기타 행동을 결정하는데 고려해야 할 요소가 되는 부분을 모두 추출하여 설계한다. KIBOT은 사용자에 대한 신뢰 상태에 따라 적절한 행동을

표 2. 사용자 행동 정의.

Table 2. The definition of user behavior.

행동 유형	설명
noexist	없음
close	가까이에 존재
up	들어올림
shake	공중에서 흔들
pat	쓰다듬음
shout	소리침
accept	서비스 제안 승낙
reject	서비스 제안 거절
nointer	인터랙션 없음

표 3. KIBOT의 관측변수.

Table 3. The observation variables of KIBOT.

변수명	설명	값
heat	열	on off
touch	터치	on off
tilt	기울어짐	on off
bottom	바닥감지	on off
sound	소리감지	on off
service	서비스사용	yes no short long
shist	서비스사용이력	low high

표 4. 설계된 내부 변수.

Table 4. The internal variables.

변수명	설명	값 및 보상정책
user	사용자의 행위	noexist, close, up, shake, pat, shout, accept, reject, nointer
kibot	- 로봇의 심리적 친숙/ 안정감의 상태 - 경계함(0,1) → 익숙함(2,3) → 신뢰함(4,5) - 터치 등 높아줄수록 증가	L0, L1, L2, L3, L4, L5
hist	사람의 서비스 사용 이력	low high

보여주어야 할 필요가 있으므로 그 신뢰도(혹은 친숙함)의 정도를 내부 변수로서 관리할 수 있도록 한다. 내부 변수의 실제 값을 유추하는데 관측 변수의 값이 활용되며, 내부 변수는 행동의 결정에 필수적으로 고려되어야 할 요소 위주로 설계된다. 표 4는 설계된 내부 변수를 보여준다. 또 POMDP 모델 {S,A,T,R,O,Z}에서 관측치 S에 해당한다.

Step6: 현재의 KIBOT 상태, 관찰된 관측치, 시스템의 이전 행동 유형에 따라 다음의 상태가 어떻게 변화할 지, 혹은 사용자가 다음에 어떤 행동을 취할지에 대한 상태 천이와 초기의 전체 상태 간 확률 분포를 모델링한다. 확률 분포 값은 시나리오 자료 등을 활용하여 통계적으로 추출한다. 표 5는

표 5. 설계된 내부 변수의 확률 분포.

Table 5. The probability distribution of internal variables.

Heat (user'	(noexist	(heat'	(on (0.00)	(off' (1.00)))
	(close	(heat'	(on (0.80)	(off' (0.20)))
	(up	(heat'	(on (0.90)	(off' (0.10)))
	(shake	(heat'	(on (0.90)	(off' (0.10)))
	(pat	(heat'	(on (0.90)	(off' (0.10)))
	(shout	(heat'	(on (0.05)	(off' (0.95)))
	(accept	(heat'	(on (0.90)	(off' (0.10)))
Touch (user'	(noexist	(touch'	(on (0.00)	(off' (1.00)))
	(close	(touch'	(on (0.00)	(off' (1.00)))
	(up	(touch'	(on (0.90)	(off' (0.10)))
	(shake	(touch'	(on (0.90)	(off' (0.10)))
	(pat	(touch'	(on (0.90)	(off' (0.10)))
	(shout	(touch'	(on (0.05)	(off' (0.95)))
	(accept	(touch'	(on (0.90)	(off' (0.10)))
(nointer	(touch'	(on (0.00)	(off' (1.00)))	
(reject	(touch'	(on (0.05)	(off' (0.95)))	

사용자 행동에 관한 관측변수 중 사람을 감지하는 열감지(Heat)와 터치감지 센서(Touch)에 대한 확률분포의 예를 보여준다. 예로, 열감지는 사람이 없을 때(noexist)나 소리치거

(shout)나 서비스 거절(reject), 사용자 인터랙션이 없을 때(nointer)에 off 될 확률이 95% 이상으로 높다. 표 3에서의 관측 변수의 확률분포는 KIBOT의 센서 특징을 고려하여 확률값을 설정하였다.

표 6. KIBOT 로봇의 상태 천이.

Table 6. The state transition of KIBOT.

현재 로봇의 친숙도	사용자의 행동	로봇의 친숙도 변화
L0	noexist	L0
	close	L0
	up	L1
	shake	L0
	pat	L1
	shout	L0
	accept	L1
	reject	L0
	nointer	L0
L1	noexist	L1
	close	L1
	up	L2
	shake	L0
	pat	L2
	shout	L1
	accept	L2
	reject	L0
	nointer	L0
L2	noexist	L2
	close	L2
	up	L3
	shake	L1
	pat	L3
	shout	L2
	accept	L3
	reject	L1
	nointer	L1
L3	noexist	L3
	close	L3
	up	L4
	shake	L2
	pat	L4
	shout	L3
	accept	L4
	reject	L2
	nointer	L2
L4	noexist	L4
	close	L4
	up	L5
	shake	L3
	pat	L5
	shout	L4
	accept	L5
	reject	L3
	nointer	L3
L5	noexist	L5
	close	L5
	up	L5
	shake	L4
	pat	L5
	shout	L5
	accept	L5
	reject	L4
	nointer	L4

POMDP 모델 {S,A,T,R,O,Z}에서 Z에 해당하며, 관측치의 확률 분포 $P(o'|s',a)$ 라고 할 수 있다.

표 6은 KIBOT의 내부변수(kibot)의 상태 천이를 보여준다. 사용자 행동 중 pat, accept, up시에는 친숙도가 올라가고 shake, reject, nointer시에는 신뢰도(친숙도)가 낮아지도록 설계되었다. 이와 같은 상태 천이는 그림 2에서 말한 대화 이력(dialogue history)으로 볼 수 있다. 또 POMDP 모델 {S,A,T,R,O,Z}에서 T에 해당된다.

Step7: 마지막 단계에서는 시스템의 행동을 설계한다. 본 인터랙션에서는 KIBOT의 행동 변수를 표 7에 보이는 바와 같이 출력한다. 로봇의 행동 변수는 그림 2에서 사용자 목표(user goal)라고 볼 수 있고, POMDP 모델 {S,A,T,R,O,Z}에서 A에 해당된다.

그리고 사용자와 로봇의 사용 시나리오로부터 KIBOT 로봇이 취할 수 있는 행동과 보상 정책(reward)을 도출한다. 보상은 특정 상태에서 유도하고자 하는 행동, 혹은 바람직한 행동을 검토하고 이에 대해서는 보상을 부여하고, 특별히 금지할 행동을 고려하여 패널티를 부여하도록 설계한다. 또한, 이 행동을 했을 때, 사용자 행동 등 외부 자극의 상태 천이를 선택하여 결정하였다. 본 과정은 POMDP 모델 {S,A,T,R,O, Z}에서 R에 해당된다.

상기에서 기술한 사용자-로봇간 인터랙션을 위한 행동모델을

표 7. KIBOT 로봇 시스템의 행동 설계.

Table 7. The behavior modeling of KIBOT.

행동	비고	보상 정책
patted	쓰다듬 받기	친숙도1이하에서 터치를 받았을 때
pleased1	기뻐하기1	친숙도2~3에서 터치를 받았을 때
pleased2	기뻐하기2	친숙도4~5에서 터치를 받거나, 서비스를 실행하였을 때
urge1	조르기1	친숙도2~3의 서비스사용이력이 높은 상태에서 일정시간 인터랙션이 없을 때
urge2	조르기2	친숙도4~5의 서비스사용이력이 높은 상태에서 일정시간 인터랙션이 없을 때
attention	관심끌기	친숙도1이하에서 인터랙션이 없을 때
glad	반가워하기	친숙도1이하에서 사람을 감지했을 때
cute	재롱부리기	친숙도4~5에서 사람을 감지했을 때
beware	경계하기	친숙도4~5에서 큰소음을 감지했을 때
scare1	무서워하기1	친숙도2~3에서 공중으로 들어올려지거나 큰소음을 감지했을 때
scare2	무서워하기2	친숙도1이하에서 공중으로 들어올려지거나 큰소음을 감지했을 때
dislike	싫어하기	친숙도2~3에서 기울여졌을 때
warn	경고하기	친숙도1이하에서 기울여졌을 때
elated	신나하기	친숙도4~5에서 공중에 올려졌을 때
sad	슬퍼하기	서비스제안이 거절되었을 때
recomm	서비스 제안	친숙도2~5의 서비스사용이력이 낮은 상태에서 인터랙션이 없을 때
thank	감사하기	도움요청이 받아들여졌을 때
quit	끝내기	사람이 없다고 판단될 때

이용하여 POMDP의 Symbolic HSVI 알고리즘을 활용하여 학습하게 된다. Symbolic HSVI 알고리즘은 훈련 시킬 때, 수집된 포인트의 상위 한계(upper bound)와 하위 한계(lower bound)의 차이 목표 값을 지정하는 양의 정수의 목표폭값(GoalWidth)을 지정하게 된다. 이 값이 적으면 적을수록 그 목표치에 부합하는 정책을 산출하기 위하여 더 많은 사용자 시뮬레이션과 훈련이 시행되게 되며, 정책은 보다 많은 수의 정책을 포함하게 된다. 반면, 과도하게 훈련이 되어 편향적인 행동을 추론하게 될 가능성과 훈련 시간이 과도하게 소요될 수 있으며, 모델링에 따라 에러(out of memory, runtime)가 발생하는 경우도 있다. 본 논문에서는 경험적으로 목표폭값(GoalWidth) = 25로 설정하였다.

학습기로부터 학습된 정책파일과 행동모델 파일을 이용하여 KIBOT의 소프트웨어에 탑재된 대화형 인터랙션 엔진에 의해서 사용자 입력이 들어오면 로봇 모션으로 표현하게 된다.

IV. 시뮬레이션

장에서 사용자와 KIBOT의 행동 모델에서 입력변수는 센서 입력과 서비스 이력 등 7개 (heat, touch, tilt, bottom, sound, service, service history), 출력 변수는 KIBOT의 단위 행동 18개 (patted, pleased, urge 등)로 설정되었다. 또 내부 변수로 사용자의 입력을 예측하는 상태변수 9개, 로봇의 내적 성격상태 L0-L5 단계가 있다. L0와 L1은 경계감, L2와 L3은 익숙함과 L4와 L5는 신뢰감을 나타낸다. POMDP 학습은 HRI 시나리오를 기본으로 로봇이 해당 행동 혹은 유사한 행동을 나타내면 긍정적인 보상(+, reward)을, 상황과 맞지 않거나 부적절한 행동 출력 시엔 부정적인 보상(-, penalty)을 한다. 예를 들어, 열감지(heat)가 되었을 때, 반가워하거나(glad), 기뻐하는 행동에 보상을 해주고, 싫어하거나(dislike), 무서워하는 행동에 벌칙을 준다. 상황에 따른 맞지 않은 행동을 금칙행동이라고 한다. 이와 같이 보상의 기준은 사용자가 로봇 사용에서 일반적인 HRI를 기준으로 임의적으로 정하였다. 또, 사용자와 지속적으로 긍정적인 인터랙션을 함으로써 KIBOT 로봇의 성격은 신뢰도가 높은 상태로 발전된다.

본 논문에서는 제안한 사용자-로봇 인터랙션의 행동모델이 실제 상황을 고려한 다양한 상황에도 유연한 행동을 표출해 낼 수 있을까를 시뮬레이션을 통해서 조사하였다. 행동 모델의 초기 확률분포는 사용자 변수 중 close 상태가 100%, kibot 상태는 L0이 100%, 서비스 사용 이력(hist)의 Low가 100 %로 설정하였다.

시뮬레이션은 1,000번의 시도(trial)로 조사되었고, 인터랙션은 시스템 행동이 quit로 끝날 때 종료되는 것으로 하였다. 도출된 정책값(policy)은 36으로 조사되었다. 시뮬레이션 결과에서 발견되는 행동의 수로 행동의 다양성을 살펴보았다. 표 8에서는 실험으로 발견되는 행동의 다양성을 보여준다.

사용자 행동에 대해서 KIBOT이 3개 이상의 행동을 발견할 확률이 높은 것은 그만큼 다양한 행동을 표출하는 것을 의미한다. 한 예로, KIBOT의 상태가 L1 상태일 때, 사용자의 입력이 pat으로 들어 온다면, KIBOT 행동은 pattd = 79.7%, glad = 9.5%, pleased1=8.3%, 기타 = 2.6% 로 발견한다는 것을 의미한다.

표 8. KIBOT 로봇의 행동의 다양성.

Table 8. The variation of KIBOT behavior.

발현 되는 행동의 수	비율
1개의 행동만 발현	28.00 %
2개의 행동만 발현	9.75 %
3개 이상의 행동 발현	62.18 %

표 9. POMDP와 MDP 교차 실험 결과.

Table 9. The experiment result of POMDP and MDP.

사용자 모델	보상정책 1		보상정책 2	
	POMDP	MDP	POMDP	MDP
제한된 행동 모델	402.48	347.34	40.89	16.67
Active User 모델	398.74	367.70	36.26	17.16
Passive User 모델	305.86	265.59	28.91	1.95

또, 훈련에 사용된 시나리오 외의 사용자 행동패턴에 대해서도 조사하였다. 첫 번째는 제한된 시나리오를 이용한 모델, 두 번째는 로봇의 서비스 추천에 승락하는 등 보다 적극적으로 로봇을 가지고 노는 적극적인 사용자(Active User) 모델, 그리고 세 번째는 서비스 추천에 거절하는 소극적으로 로봇을 사용하는 소극적 사용자(Passive User) 모델에 대한 조사이다. 또, 센서의 오류 확률이 전혀 없다고 가정된 MDP (Markov Decision Process) 기반 정책 훈련과도 비교하였다. 보상(reward) 정책으로는 2가지로 살펴보았다. 보상정책1은 훈련에 적용한 보상 정책의 부여 원칙을 그대로 적용하고, 보상정책 2는 금칙행동 및 다양한 발현에 따른 보상하였다. 상황에 따른 금칙행동 리스트를 정의하고 발현 시 벌칙(penalty) -10, 반대의 상황에서(내부 KIBOT 상태) 추가적인 다양한 행동 발현 시 보상(reward) +5하여 조사하였다.



(a) The pattd behavior on touch.



(b) The wam behavior on shake.

그림 6. POMDP기반 사용자-로봇 인터랙션 행동 표현. Fig. 6. The KIBOT behavior of POMDP-based HRI.

각각의 사용자 모델과 POMDP, MDP 기반 행위 추론기간 교차 실험을 시행하였으며, 상황에 따라 적절하고 다양한 행동을 보이는가? 아니면 금칙행동을 보이는가?에 대해서 조사를 하였다. 표 9에서는 3가지 모델에 대해서 보상정책 1, 2로 POMDP와 MDP 각각의 보상값(+5), 벌칙값(-10)의 합계를 보여준다. 보상정책1과 2에 대해서 오류 상황에 탄력적으로 대응하는 POMDP가 MDP보다 3가지 사용자 모델에 대하여 높은 값을 나타낸다. 이것은 POMDP가 상황에 따라 적절하고 다양한 행동을 보여주었음을 의미한다. 또 보상정책1이 벌칙을 부여한 보상정책 2에 비해서 높은 값을 보여준다.

제안된 POMDP 기반 사용자-로봇 인터랙션 행동 모델은 실제 KIBOT에 적용하여 테스트되었다. 표 7에 제시한 로봇 행동은 [10]에서 제안한 로봇 모션생성 시뮬레이터를 이용하여 각 행동에 대한 느낌을 잘 표현하도록 모션을 생성하였고, LCD에도 각 행동의 느낌을 살려주도록 이모티콘으로 표시하였다. 그림 6은 제안된 POMDP 기반 사용자-로봇 인터랙션 행동 모델이 적용되어 KIBOT에서 시스템 행동으로 구현되는 예를 보여준다.

V. 결론

본 논문에서는 사용자와 로봇의 상황에 맞는 상호작용적 유연한 커뮤니케이션을 위하여, 통계학적 확률에 기반한 POMDP를 이용하여 사용자-로봇 인터랙션 행동 모델을 제안하고 KIBOT에 적용하여 실험하였다. POMDP는 주로 대화시스템 등 다양한 응용 분야에 활용되고 있으며, 로봇에는 내비게이션 연구에 주로 적용되었다. 본 논문에서는 POMDP가 대화시스템에 강점이 있는 것에 주목하고, 사용자-로봇간의 인터랙션도 커뮤니케이션과 같은 개념이라는 것에 착안하여 사용자-로봇간 인터랙션에 적용하여 상황에 따른 유연한 행동을 표출하도록 제안하였다. 제안된 POMDP를 이용한 인터랙션 시스템은 다양한 로봇 행동생성으로 사용자에게 로봇이 살아 있는 생명체처럼 느끼고 유쾌하고 재미있게 상호교감할 수 있도록 하였다.

본 논문에서는 사용자 입력이나 로봇 행동을 몇 가지 상황으로 하여 모델링하였지만, 향후 실제 상황에 맞는 다양한 시나리오 발굴로 많은 사용자 입력과 유연한 KIBOT의 행동으로 확대해 나갈 계획이다.

REFERENCES

- [1] C. Boutilier and D. Poole, "Computing optimal policies for partially observable decision processes using compact representations," *Proc. AAAI*, pp. 1168-1175, 1996.
- [2] A. Green and K. Eklundh, "Designing for learnability in human-robot communication," *IEEE Trans. Electron.*, vol. 50, no. 4, pp. 644-650, 2003.
- [3] C. Breazeal, "Social Interaction in HRI: The robot view," *IEEE Trans. on Systems, Man, and Cyber-Part C.*, vol. 34, no. 2, pp. 181-185, 2004.
- [4] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa, "An ethological and emotional basis for human-robot interaction," *Robotics and Autonomous Systems*, vol. 42, pp. 191-201, 2003.
- [5] K. Wada, T. Shibata, T. Saito, and K. Tanie, "Robot assisted activity at a health service facility for the aged for ten weeks: an

interim report of a long-term experiment," *Proc. of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, vol. 220, no. 8, pp. 709-715, 2006.

- [6] <http://www.nec.co.jp/products/robot/>
- [7] H. R. Kim, Y. M. Kim, J. C. Park, J. S. Park, T. W. Kang, and D. S. Kwon, "Design of reactive emotion process for the service robot," *Journal of Korea Society*, vol. 2, no. 2, 2007.
- [8] Y.-C. Kim, H.-T. Kwon, W.-C. Yoon, and J.-C. Kim, "Designing emotional and interactive behaviors for an entertainment robot," *Proc. of Int'l Conf. on Human Computer Interaction, LNCS 5611*, pp. 321-330, Jul. 2009.
- [9] J. C. Kim and G. H. Park, "Sound-based emotion estimation and growing HRI system for an edutainment robot," *Journal of Korea Society*, vol. 5, no. 1, pp. 7-13, 2010.
- [10] G. H. Park, J. C. Kim, and H. J. Ann, "Development of videophone-based application services for KT Mon-e (KT Edutainment Robot for Young Children)," *Journal of Korea Society*, vol. 5, no. 2, pp. 93-101, 2010.
- [11] J. C. Kim and H. H. Kim, "An edutainment Mon-E robot for young children," *Journal of Korea Society*, vol. 6, no. 2, pp. 147-155, 2011.
- [12] <http://pomdp.org/>
- [13] H. Sim, K. E. Kim, J. H. Kim, D. S. Chang, and M. W. Koo, "Symbolic heuristic search value iteration for factored POMDPs," *Proc. AAAI*, pp. 1008-1093, 2008.
- [14] J. W. Sung, J. H. Eun, H. J. Kim, and D. S. Jang, "Statistical dialog management policy using POMDP POMDPs for multi multi-goal domains," *The Proc of Korea Computer Conference*, vol. 36, no. 1(B), pp. 126-131, 2009.
- [15] C. Boutilier and D. Poole, "Computing optimal plokcies for partially observable decision processes using compact representations," *Proc. of AAAI*, 1996.



김 종 철

2000년 창원대학교 제어계측공학과(공학사). 2002년 포항공과대학교 전자전기공학과(공학석사). 2007년 일본 Keio University, Integrated Design Engineering (공학박사). 2007년~현재 (주)KT 융합기술원 Service연구소 팀장. 관심분야는 지능로봇, 인간로봇상호작용, 로봇비전, 기계학습, 지능제어.