

POMDP 통합을 기반으로 한 다기능 로봇 의사결정 최적화에 관한 연구

허영, 윤제진, 김양중*

한국공학대학교

{gjdud852, 2022611520, zeroplus*}@tukorea.ac.kr

A Study on the POMDP Integration based Decision-making Optimization for Multi-function Robot

Young Heo, Jaejin Yoon, Yangjung Kim*

Tech University of Korea

요약

최근에 더욱 다양화된 응용에 따른 빠른 대응이 소프트웨어 요구사항에도 반영되어 소프트웨어 산업변화에 맞춰 진화하고 있다. 이러한 흐름에 맞추어 로봇의 기능을 확장하고 시스템의 안정성을 유지하기 위해서 많은 기술적인 제약과 어려움이 수반된다. 이에, 본 논문은 다기능 로봇의 구성과 기능 확장에 따른 의사결정 최적화에 대한 전략으로 각 로봇의 기능 수행 단위를 POMDP로 의사결정 및 기능 수행을 하도록 구성하고, 제안하는 '계층화 POMDP 알고리즘'을 통해서 계층화된 의사결정 시스템을 구성하여 다기능 로봇의 시스템 구성 및 의사결정 최적화 전략을 제시하고자 한다. 이리 구성된 다기능 로봇은 하나의 단일 POMDP로써 의사결정을 만드는 다기능 로봇 시스템보다 확장성과 복잡성, 실용성의 측면에서 보다 우수한 특성을 갖게 된다. 이를 통해 로봇 단위기능의 모듈식 구성 및 조합으로 다기능 로봇을 쉽고 빠르게 구성 및 개발하여 로봇의 개발 생산성을 향상할 수 있다.

I. 서론

최근에 소프트웨어 개발 산업은 빠르게 변화하는 요구사항에 신속하게 대응하도록 방법론적 특성 또한 발전해 왔다. 대표적인 사례로, 개발 방법론 중 하나인 '애자일(Agile) 개발 방법론'은 요구사항의 추가 및 변경에 따라 기능을 추가하면서도 지속적으로 서비스를 제공하는 방법론을 제시하여 최근 소프트웨어 개발 산업에서 가장 보편적인 개발 방법론으로 자리매김하였다.

'로봇 시스템'에서 이러한 동향에 발맞추어 빠르게 로봇 기능을 확장하고 조합하는 것은 소프트웨어 산업과 다르게 여러 가지 제약이 따르게 된다. 이에, 가장 대표적인 제약은 '상태의 조합'과 '일관된 기능 수행'에 대한 어려움이다. 예를 들어 '쇼핑몰' 플랫폼을 개발하며 '장바구니담기' 기능과 '상품선택' 기능을 조합하는 일에 반해, '청소 기능'을 수행하는 로봇에 '사용자 맞춤 기능'을 추가하여 조합하는 일은 보다 많은 개발 부담과 비용이 따르게 된다. 그 이유는 쇼핑몰의 경우 각 기능 간 조합하는 상태 정보가 적기 때문에 기능의 확장 및 조합이 쉽지만, 로봇의 각 단위 기능은 의존하는 상태정보가 상이하면서 집약적인 시스템 구성으로 인해 조합하는 상태정보가 많기 때문이다. 이런 상황은 로봇의 각 단위 기능이 단순히 독립적으로만 수행하는 것이 아니라, 상호 연동되어 복잡한 환경 속에서 적절히 반응하여 수행되어야 할 수 있어야 함을 보여준다.

본 논문은 로봇의 각 단위기능을 POMDP(Partially Observable Markov Decision Processes)로써 의사결정을 이루어 수행되도록 최적화하고, 그 기능에 대한 조합 시 '계층화 POMDP 알고리즘'을 통해 로봇의 기능을 손쉽게 확장하면서 동시에 일관된 의사결정을 이루어 기능의 수행을 적절히 해낼 수 있는 설계전략을 제시한다. 이 설계를 통해 다기능 로봇이 여러 불확실한 환경 속에서도 효과적으로 의사결정을 이루어 일관된 기능을 수행할 수 있도록 하며, 동시에 모듈식 설계를 통한 로봇의 단위 기능의 조합을 효과적으로 이루는 것을 목표로 한다.

II. 본론

1. 단위 기능 POMDP 구성

하나의 기능을 수행하는 단위 시스템에 대해서, POMDP를 이용한 의사결정을 이루고 기능 수행하도록 최적화를 위해 필요한 변수는 다음과 같다.

[표 1] POMDP 모델링을 위한 변수

Factor	Notation	Description
State Space	S	The set of all possible states of the system.
Action Space	A	The set of all actions that the agent can take.
Observation Space	O	The set of all possible observations that the agent can perceive.
Transition Probabilities	$T(s, a, s')$	The probability of transitioning to state s' from state s when action a is taken: $T(s, a, s') = P(s_{t+1} = s' s_t = s, a_t = a)$
Observation Probabilities	$Z(s', a, o)$	The probability of observing o given that action a was taken and the system is in state s' : $Z(s', a, o) = P(o_t = o s_{t+1} = s', a_t = a)$
Reward Function	$R(s, a, s')$	The reward received for transitioning from state s to state s' , given that action a was taken: $R(s, a, s') = \text{Reward}(s, a, s')$
Discount Factor	γ	The factor used to discount future rewards to their present value, typically within the range $0 \leq \gamma \leq 1$.

이에, 로봇의 각 단위기능을 추상화하여 POMDP에 의한 의사결정을 도출하도록 최적화하며, 이 단위기능을 기반으로 본 논문에서 제안하는 HPA(Hierarchical POMDP Algorithm)으로 확장해 탁월한 다기능 로봇의 의사결정의 최적화 방안을 제시하고자 한다.

2. 제안하는 HPA(Hierarchical POMDP Algorithm) 개요

제안하는 HPA의 목표는 POMDP로 수행되도록 최적화된 로봇의 각 단위 기능을 조합하여, 로봇 시스템 전반적인 의사결정을 이룰 수 있는 상위 레벨의 POMDP를 구성하는 것이다. 이에 알고리즘은 다음과 같이 설명될

수 있다.

먼저 상위 POMDP 정책과 하위 POMDP 정책을 초기화한다. 그리고 각 반복 과정에서 정책의 변경 여부를 '상태 플래그'를 사용하여 추적한다. 또, 각 상위 POMDP 상태에 대해 가능한 모든 행동을 고려하여, 그 행동을 취했을 때의 기대되는 총 보상을 계산하는 Q 함수를 갱신한다. 계산된 Q 값에서, 가장 높은 값을 제공하는 행동을 최적의 행동으로 선택하고, 현재 정책과 비교하여 변경이 필요한 경우 정책을 갱신한다. 그리고 각 하위 POMDP 상태에서, 가능한 모든 행동에 대해 각 행동 후 관측될 수 있는 모든 결과에 대한 확률을 고려하여 Q 값을 재계산한다. 이 과정을 반복하면서 더 이상 정책에 변화가 없으면 '상태 플래그'가 False로 설정되며 반복을 종료한다. 이러한 과정을 통해서, 최적화된 상위 레벨의 POMDP 정책을 생성한다. 그림 1은 계층화 POMDP 알고리즘의 의사코드를 구현한 내용이다.

Algorithm 1 HPA(Hierarchical POMDP Algorithm)

```

function HPA( $S, A, O, T, Z, R, \gamma$ )
  Initialize high-level policy  $\pi^H$ 
  Initialize low-level policies  $\pi^L$ 
  Initialize value functions  $V^H, V^L$ 
  repeat
    change  $\leftarrow$  False
    for each high-level state  $s \in S$  do
      for each action  $a \in A$  do
         $Q[s, a] \leftarrow \sum_{s' \in S} T(s, a, s') (R(s, a, s') + \gamma V[s'])$ 
         $best\_a \leftarrow \arg \max_a Q[s, a]$ 
        if  $\pi^H[s] \neq best\_a$  then
           $\pi^H[s] \leftarrow best\_a$ 
          change  $\leftarrow$  True
        end if
      end for
    end for
    for each low-level state  $s \in S$  do
      for each action  $a \in A$  do
         $Q[s, a] \leftarrow \sum_{o \in O} Z(s, a, o) \sum_{s' \in S} T(s, a, s') (R(s, a, s') + \gamma V[s'])$ 
         $best\_a \leftarrow \arg \max_a Q[s, a]$ 
        if  $\pi^L[s] \neq best\_a$  then
           $\pi^L[s] \leftarrow best\_a$ 
          change  $\leftarrow$  True
        end if
      end for
    end for
  end for
  if change = False then
    break
  end if
until convergence
return  $\pi^H, \pi^L$ 
end function

```

그림 1. 계층화 POMDP 알고리즘의 의사코드

3. 시스템 아키텍처

이 과정에 따라, 시스템이 모델링이 된 아래의 그림과 같은 계층형태의 모듈을 구성하게 된다. 상위 POMDP의 모듈에서는 전략적 결정을 내리고, 하위 POMDP에서는 이러한 결정을 구체적인 행동으로 변환하여 동작을 수행하게 된다.

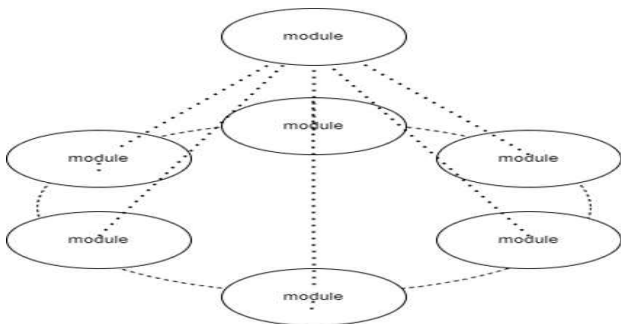


그림 2. 제안하는 HPA 적용에 따른 로봇 시스템 모듈 구성도

이러한 계층화된 접근 방식을 통해 환경 상태 및 목표의 변화에 대응하여 의사결정 프로세스를 동적으로 조정하고 최적화할 수 있게 되고, 복잡하고 동적인 상황에서 최적화할 수 있다.

4. 단일 POMDP 구성의 의사결정 시스템과 비교 분석

계층화 POMDP에 따른, 다기능 로봇 시스템은 복잡한 여러 기능적 의사결정 작업을 관리하는데 있어, 단일 POMDP로 다기능 의사결정을 수행하는 로봇 시스템보다 우월한 성능을 발휘할 수 있다. 모든 기능을 단일 POMDP 구성으로 의사결정을 수행하도록 통합하면 계산의 복잡성이 증가하고 유연성이 감소하게 된다. 반면에 계층적 POMDP는 의사결정을 상위 하위 계층으로 분리하여 유지함으로, 이에 따른, 계산 오버헤드 및 복잡성을 줄이면서 각 단위기능을 독립적으로 최적화하고 확장할 수 있도록 해 기능적 확장성을 높일 수 있다. 또한 계층적 POMDP 다기능 로봇 시스템은 기능 모듈성을 가져 병렬처리를 용이하게 하여, 시스템 처리량과 응답 시간을 크게 향상시킬 뿐만 아니라, 특정 단위 기능에 대한 변경 사항을 격리하여 시스템 유지 관리 및 업그레이드를 단순화시키는 장점을 갖고 있다.

결과적으로 제안하는 HPA 알고리즘을 이용하면, 다기능 로봇 시스템은 목표나 운영 환경의 변화를 신속하게 수용할 수 있으며, 보다 적응력이 뛰어난 설계를 제공하여 시스템의 유연성과 빠른 확장성을 갖도록 할 수 있다.

III. 결론

본 논문에서는, 계층화 POMDP를 이용해 로봇의 각 단위기능을 조합하여, 각 기능이 의존하는 상태를 적절히 조합하고, 일관된 의사결정에 따라 기능을 수행하는 다기능 로봇의 최적화 방안을 제안하였다. 향후 연구로는, 본 논문에서 제시한 전략으로 구성된 다기능 로봇에 대해 사용자에 설문 등 의사결정 및 수행에 대한 타당성을 정량적으로 측정할 수 있는 새로운 평가 기준을 개발 및 도입하여, 다양한 설정 및 조건에서 그 성능에 대한 평가를 진행하고자 한다.

ACKNOWLEDGMENT

본 연구는 고용노동부 및 한국산업인력공단의 '2024년 고속련 마이스터 사업'의 지원을 받음

참 고 문 헌

- [1] The Social Robot Architecture: A Framework for Explicit Social Interaction. Brian R. Duffy, Mauro Dragone, Gregory.M.P. O'Hare
- [2] Partially Observable Markov Decision Processes in Robotics: A Survey. Mikko Lauri, David Hsu and Joni Pajarinen
- [3] 두개의 목적함수를 가지는 다목적 최적설계를 위한 적응 가중치법에 대한 연구. 김일용 한국정밀공학학회지 제21권 제9호 (2004년 9월)