Factored POMDP를 이용한 가상군의 자율행위 모델링 사례연구

이강훈 ○ 임희진 김기응
한국과학기술원 전산학과
khlee@ai.kaist.ac.kr, hlim@ai.kaist.ac.kr, kekim@cs.kaist.ac.kr

A Case Study on Modeling Computer Generated Forces based on Factored POMDPs

Kanghoon Lee ○ Heejin Lim Kee-Eung Kim
Department of Computer Science, KAIST

요 약

가상군의 자율행위 모델링은 전장모의모델링 시스템의 성능을 결정하는 주요한 요소이다. 불확실한 상황을 확률적으로 고려하여 최적의 의사결정을 가능하게 하는 POMDP (partially observable Markov decision process) 모델은 가상군의 자율행위 모델링에 있어서 매우 자연스러운 프레임워크이다. 그러나 POMDP 모델의 높은 계산복잡도로 인한 최적 행동정책 계산의 어려움은 POMDP 모델을 이용한 가상군의 자율행위 모델링을 저해하는 요소이다. 본 논문에서는 대규모 가상군의 자율행위 모델링을 위해 factored POMDP 모델을 이용한다. 그리고 "Hasty Defense" 사례연구를 통해 그 효과를 확인한다.

1. 서론

가상군의 자율행위 모델링은 전장모의 시뮬레이션 시스템의 성능을 결정하고 나아가 이를 이용한 아군 전략의 품질 및 훈련 효과를 결정하는 핵심 요소이다. 군인들이 전장에서 자율적으로 취하는 행동을 컴퓨터로 모의한 것을 CGF (computer generated force)라고 한다. 자율적이고 현실적으로 행동하도록 모의된 CGF를 전장모의 시뮬레이션 시스템에서 가상군으로 사용함으로써, 작전계획작성, 작전분석 및 훈련을 현실적으로 수행할 수 있다. 그 결과, 고도화된 작전계획을 세울 수 있으며 모의훈련의 효과를 증진시킬 수 있다.

가상군의 자율행위를 모델링하는 기존 방법론으로는 규칙기반 시스템 (rule-based system)을 이용하는 것이 전통적이다. 하지만, 규칙기반 시스템은 전문가의 지식을 활용하여 가상군의 행동을 결정하는 시스템으로 자율행위를 결정하는데 필요한 고려사항이 많아질수록 규칙작성이 힘들어지고, 불확실한 전장상황을 고려하기 어렵다. 이러한 측면에 있어서 규칙기반 시스템을 이용하는 기존의 방법론들은 한계점이 존재한다.

부분관찰 마코프 의사결정과정 (partially observable Markov decision process; POMDP) [1] 은 불확실한 혹은 부분적으로 관찰 가능한 상황을 확률적으로 고려하여 순차적 의사결정을 하는 문제를 모델링할 수 있는 인공지능 및 기계학습 방법론이다. POMDP는 최적행동 정책을 계산함에 있어 규칙기반 시스템에 비하여 전문가의 지식에 대한 의존도가 낮다. 그리고 불확실한 전장상황에 대비한 행동정책을 계산해낼 수 있다. 그러므로 POMDP 모델을 이용하여 가상군의 자율행위를 모델링하는 것은 매우 자연스럽다고 볼 수 있다.

그러나 POMDP의 높은 계산복잡도는 가상군의 자율행위 모델의 활용을 어렵게 한다. Factored POMDP [2]는 POMDP 문제를 문제의 구조적 특징을 이용하여 간결하게 표현하는 방법이다. 이는 RDDL (relational influence diagram language) [3] 등의 언어로 기술할 수 있으며 symbolic HSVI (heuristic search value iteration) [4] 등의 알고리즘으로 효율적으로 최적 행동정책을 계산할 수 있다.

본 논문에서는 "Hasty Defense" 시나리오 [5]의 사례연구를 통해 대규모 가상군의 자율행위 모델링을 위한 factored POMDP 모델을 제안한다. 이를 위해 RDDL 언어를 이용하여 factored POMDP 모델을 서술하고 symbolic HSVI 알고리즘을 사용하여 최적 행동정책을 계산한다. 그리고 가상군의 자율행위 모델링에 있어 POMDP 모델의 효과를 확인한다.

2. 사례연구: Hasty Defense 시나리오

Hasty defense 시나리오 [5]는 남하하는 적군 탱크

[1] 본 연구는 방위사업청과 국방과학연구소의 지원으로 수행되었습니다. (UD110006MD)

부대를 사전에 인지하고 적군을 지연시키기 위한 아군 탱크 소대의 방어 시나리오다. 그림1에서 적군여단은 지도의 북쪽을, 아군여단은 지도의 남쪽을 점거하고 있다. 적군 선봉대대는 제1전투진지 (BP1)를 향해 진군하며 적군 선봉소대는 "road", "trail", "cross-country" 3가지 경로 중 한 곳에서 출현한다고 가정한다. 아군은 적군의 위협을 막기 위해 A중대(Company A)의 제1 탱크 소대 (1st tank platoon)를 제1전투진지로 보내 적군의 진군을 늦추려 한다. 탱크 소대는 2개의 분대로 이루어지고, 각 분대는 2대의 탱크로 이루어진다. 아군의 목표는 적군 선봉소대의 예상 경로를 고려하여 제1소대를 제1전투진지인 3개의 고지(hill647, 725, 808) 중 최적의 위치에 배치하고 적군 선봉소대와 교전 및 적군 전개 후 A중대로의 퇴각이다.

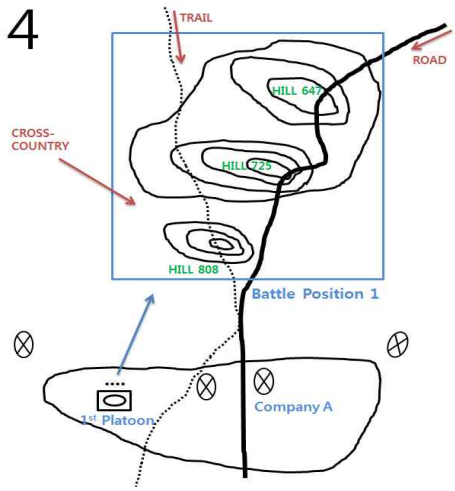


그림 1 Hasty Defense 시나리오 [5]

2.1 적군 예상 경로에 따른 대응 작전 교리

Hasty Defense 시나리오에서의 실제 작전 교리는 적군의 남하 방향에 따라 아래의 세 가지로 나뉜다.

Road & trail: 적군의 예상 경로가 각각 50%의 확률로 road 혹은 trail로 온다는 사전 정보가 있을 경우, 제1소대는 분대 별로 hill647과 hill725에 배치한다. 출현하는 적군 선봉소대를 각 분대가 서로 다른 위치에서 지원하며 교전을 실시하고 퇴각한다.

Road: 적군의 예상 경로가 road라는 사전 정보가 있을 경우, 제1소대는 hill647을 점거한 후 road를 감시한다. 소대의 모든 탱크가 출현하는 적군 선봉소대를 공격한다.

Cross-country: 적군의 예상 경로가 cross-country라는 사전 정보가 있을 경우, hill725를 점거 후 다이아몬드 형태로 사방을 주시한다. 적군 선봉소대가 출현하면 3대의 탱크는 적군과 교전하고 나머지 한 대의 탱크는 후방을 경계한다.

3. 사전 연구

3.1 POMDP 모델과 Factored POMDP 모델

POMDP 모델 [1]은 $<S, A, Z, T, O, R, b_0>$으로 정의된다. S 는 환경상태의 집합, A는 가능한 행동의 집합, Z는 관찰값의 집합, T는 환경상태 전이함수, O는 관찰함수, R은 보상함수, b_0는 초기 환경상태 확률분포이다.

Factored POMDP [2]는 POMDP 모델을 간결하게 표현하기 위한 방법으로서, 환경상태, 행동, 관찰값 등을 문제의 구조적 특징에 기반해 확률 변수로 나눠 표현한다. 흔히 쓰이는 데이터 구조로는 ADD (algebraic decision diagram) [4]가 있다. ADD는 환경상태 및 관찰값을 트리 형태의 구조로 표현함으로써, ADD 사이의 사칙연산 및 내적 등의 행렬연산을 효율적으로 할 수 있다.

3.2 RDDL 언어

RDDL [3]은 factored POMDP 모델을 기술함에 있어 환경상태, 관찰, 행동, 보상함수를 상호관계 영향도 (influence diagram)를 만들어 표현한다. 따라서 직관적으로 이해할 수 있으며 ADD 구조의 표현력을 지니고 있는 서술 언어이다. RDDL은 "domain", "non-fluent", "instance" 세 부분으로 나뉜다. "domain"은 환경상태, 관찰, 행동, 보상함수의 상호관계, "non-fluent"는 상호 관계에 필요한 상수값 및 변수값, "instance"는 모델의 초기 환경상태 등을 정의한다.

4. Hasty Defense 시나리오의 Factored POMDP 모델링

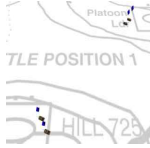
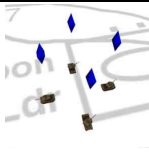
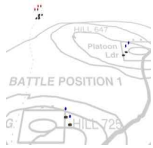
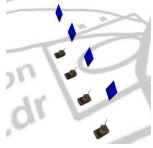
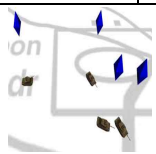
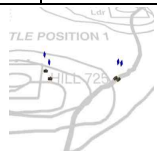
본 논문의 목표는 factored POMDP 모델을 이용한 가상군의 자율행위 모델링의 적합성을 보이는 것이다. Hasty defense 시나리오를 이용하여 factored POMDP 모델에서 계산된 최적 행동정책이 군사 전문가들에 의해 수립된 실제 작전 교리와 일치함을 확인하는 것으로 이를 보이고자 한다.

4.1 Hasty Defense 시나리오의 3단계 분류

Hasty Defense 시나리오는 아군의 역할에 따라 준비단계, 교전단계, 퇴각단계로 분류할 수 있다. 준비단계는 적군 선봉소대의 예상 경로에 따른 제1전투진지에서의 제1소대 점거 방식을 결정하는 단계로 그림1에서 제1소대가 hill647, 725, 808을 점거 후 적을 기다리게 된다. 교전단계는 관찰된 적군 선봉소대에 따라 교전 전략을 결정한다. 즉, 적군의 출현위치 및 교전상황에 따라 제1소대의 공격전략 및 퇴각여부를 결정한다. 퇴각단계는 적군부대를 전개시킴으로써 남하를 일시 정지시키거나 제1소대가 위험할 경우 A중대로 돌아오는 상황으로 구성된다.

Factored POMDP 모델을 이용하여 계산하는 제1소대의 전략은 준비단계에서의 점거할 고지의 위치 결정, 교전단계에서의 정찰전략, 공격전략, 퇴각전략으로 구성된다. 정찰전략에서는 대형에 따라 적

정찰 확률이 결정된다. 공격전략에서는 소대의 배치와 대형에 따라 적 공격확률 및 아군 피격확률이 결정된다. 퇴각전략에서는 대형에 따라 아군 피격확률이 결정된다. 전략에 따른 각 대형과 배치는 아래 표와 같다.

정찰전략	 다이아몬드대형	 분대분리대형	 선형대형
공격전략	 3기공격 1기후방경계	 1분대공격 2분대지원	 4기공격
퇴각전략	 3기퇴각 1기엄호	 1분대퇴각 2분대엄호	

4.2 RDDL 모델링

RDDL의 domain은 state-fluent, observ-fluent, act-fluent로 이루어진 parameterized variable과 이들사이의 상호관계 영향도를 지니는 cpf (conditional probability function)로 구성된다. cpf 및 non-fluent, instance는 앞의 시나리오 설명에 따라 그대로 정의되므로 생략하도록 한다.

State-fluent	현재 단계 (준비, 교전, 퇴각, 종료) 제1소대의 생존 여부 및 위치 제1소대의 행동 전략(정찰, 공격, 퇴각) 적군선봉소대의 생존 여부 및 접근 위치 적군의 전개 여부
Observ-fluent	제1소대의 생존 여부 적군선봉소대의 생존 여부 및 접근 위치 적군의 전개 여부
Act-fluent	다음 단계로의 전이 제1소대가 사용할 전략 선택 제1소대의 점거할 고지 선택 교전수행

위의 state-fluent는 아군과 적군의 현재상태를 나타낸다. Observ-fluent는 현재상태의 불확실한 관찰값을 나타내고 act-fluent는 가상군의 행동을 나타낸다.

5. 실험결과

실험에서는 앞서 정의하였던 factored POMDP 모델을 사용하였으며 최적 행동정책 계산을 위해 symbolic HSVI [3] 알고리즘을 사용하였다. 아래는 적이 road 혹은 trail에서 올 것이라는 사전 정보를 가지고 있을 때 계산된 최적 행동정책에 따른 결과이다.

<준비단계> 제1소대는 “hill647”과 “hill725”를 점령
제1소대는 “분대분리대형”을 유지
<교전단계> [“Road”에서 적군선봉소대 발견]
제1소대는 “1분대공격 2분대지원” 전략 수행
적군선봉소대의 제1소대 공격
[적군 전개 관찰]
제1소대는 “1분대퇴각 2분대엄호” 전략 수행
적군선봉소대의 제1소대 공격
<퇴각단계> 제1소대 생존귀환 및 적군 전개시킴
→ 미션성공 및 적군전개

위 실험 결과, 앞서 정의한 factored POMDP 모델의 최적 행동정책은 전문가들에 의해 수립된 시나리오의 실제 작전교리와 일치함을 확인하였다. 또한, 적군의 예상 경로가 cross-country라는 사전 정보가 있을 경우 역시 factored POMDP 모델의 최적 행동정책이 실제 작전교리와 일치함을 확인할 수 있었다.

6. 결론

가상군의 자율행위 모델링에 있어 POMDP 모델을 이용한 방법은 전문가의 도움 없이 불확실한 상황에서도 강인한 행동정책을 계산할 수 있다. Hasty Defense 사례연구를 통해서 factored POMDP 모델의 최적정책이 실제 작전교리와 일치함을 실험적으로 확인하였다. 향후에 각 객체가 독립적으로 행동정책을 계산할 수 있는 decentralized 모델을 이용한 각 소대의 자율적 행위 모델링에 대한 연구가 필요하다.

참고문헌

[1] L. P. Kaelbling et. al., Planning and Acting in Partially Observable Stochastic Domains, *Artificial Intelligence*, 101:99-134, 1998

[2] C. Boutilier and D. Poole, Computing Optimal Policies for Partially Observable Decision Processes using Compact Representations, *In Proc. of AAAI*, 1996

[3] S. Sanner, Relational Dynamic Influence Diagram Language (RDDL) : Language Description, http://users.cecs.anu.edu.au/~ssanner/IPPC_2011/RDDL.pdf, 2011

[4] H. Sim et. al., Symbolic Heuristic Search Value Iteration for Factored POMDPs, *In Proc. of AAAI*, 2008

[5] R. W. Pew and A. S. Mavor, Modeling Human and Organizational Behavior, *National Academy Press*, pp. 20-32, 1998