

지형정보 기반 조난자 행동예측을 위한 마코프 의사결정과정 모형

손진호
국방대학교 국방과학학과
(thswlsgh0112@naver.com)

김수환
국방대학교 국방과학학과
(ksuhwan@korea.kr)

유사시 중심 깊숙한 곳에서 적을 타격하는 임무를 수행하는 항공기의 경우 격추될 위험에 항상 노출되어 있다. 현대전의 핵심 전투력으로써 최첨단의 무기체계를 운용하는 공중군무 요원은 양성하는데 많은 시간과 노력, 국가 예산이 소요되며 그들이 가진 작전 능력과 군사기밀이 매우 중요하기에 공중군무 요원의 생존은 매우 중요한 문제이다. 따라서, 본 연구에서는 적지에서 비상탈출한 조난자가 장애물을 피해 목표지점까지 도피·탈출을 시행할 경로를 예측하는 경로 문제를 연구하였으며 이를 통해 비상탈출한 조난자의 무사 생존 가능성을 높이고자 하였다. 본 연구 주제와 관련된 기존 연구들은 경로 문제를 네트워크 기반 문제로 접근하여 TSP, VRP, Dijkstra 알고리즘 등으로 문제를 변형하여 최적화 기법으로 접근한 연구가 있었다. 본 연구에서는 동적 환경을 모델링 하기에 적합한 MDP(마코프 의사결정과정)를 적용하여 연구하였다. 또한 GIS를 이용하여 지형정보 데이터를 추출하여 활용함으로써 모형의 객관성을 높였으며, MDP의 보상구조를 설계하는 과정에서 기존 연구 대비 모형이 좀 더 현실성을 가질 수 있도록 보다 상세히 지형정보를 반영하였다. 본 연구에서는 조난자가 지형적 이점을 최대한 이용함과 동시에 최단거리로 이동할 수 있는 경로를 도출하기 위하여 가치 반복법 알고리즘, 결정론적 방법론을 사용하였으며 실제 지형정보와 조난자가 도피·탈출 과정에서 만날 수 있는 장애요소들을 추가하여 모형의 현실성을 더하고자 하였다. 이를 통해 조난자가 조난 상황에서 어떠한 경로를 통해 도피·탈출을 수행할지 예측해 볼 수 있었다. 본 연구에서 제시한 모형은 보상구조의 재설계를 통해 여러 가지 다양한 작전 상황에 응용이 가능하며 실제 상황에서 조난자의 도피·탈출 경로를 예측하고 전투 탐색구조 작전을 진행시키는 데 있어 다양한 요소가 반영된 과학적인 기법에 근거한 의사결정 지원이 가능할 것이다.

주제어 : MDP(마코프 의사결정과정), 지형정보체계, 최적 경로, 조난자, 보상구조

논문접수일 : 2022년 12월 27일 논문수정일 : 2023년 3월 27일 게재확정일 : 2023년 4월 17일

원고유형 : Regular Track 교신저자 : 김수환

1. 개요

전쟁 수행의 핵심 전투력으로써 최첨단의 무기체계를 운용하는 공중군무 요원의 무사 생존은 개인의 문제를 넘어 전쟁의 승패를 좌우하는 항공 우주작전 수행에 있어 매우 중요하다. 공중군무 요원은 현대전 수행에 가장 중요한 핵심 인력으로 양성에 많은 시간과 노력, 막대한 국가

예산이 소요된다. 금전적인 부분을 떠나 각국이 공중군무 요원의 생존에 힘을 쏟는 이유는 그들이 지닌 작전 능력과 군사 기밀이 매우 중요하며 생존 여부가 군의 사기와 국민들의 애국심에 직결되기 때문이다.

예기치 못한 적의 공격으로부터 격추된 상황을 가정해 보았을 때, 적지에서 선불리 아군에게 구조요청을 시도하는 것은 자신의 위치를 적에

게 노출시킬 수 있어 조난 직후에는 선부른 구조 요청보다는 적으로부터 은신하고 도피·탈출을 수행하며 적절한 타이밍에 구조 신호를 송출하여 아군의 구조전력이 투입될 수 있도록 하는 것이 중요하다. 만약 적지에 조난된 아군이 어떠한 경로로 도피·탈출을 수행할지 미리 예측하고 대응할 수 있다면 이를 통해 구조 기회를 극대화하고 무사 생존의 가능성을 증대시켜 아군의 전투력 손실을 최소화할 수 있을 것이다.

본 연구에서는 이 모형을 구현하기 위하여 MDP(Markov Decision Process: 마코프 의사결정 과정)로 대표되는 강화 학습 방법론을 적용하였다. 강화 학습은 머신러닝의 또 다른 종류로서 행동심리학에서 영감을 받았으며, 어떤 환경 안에서 정의된 에이전트가 현재의 상태를 인식하여 선택 가능한 행동 중 보상을 최대화하는 행동 혹은 행동 순서를 선택하는 방법이다. 이러한 강화 학습의 방법론은 국방 분야에 적용하기에 매우 적절한데, 현재 주어진 환경에서 보상을 최대화하는 행동 또는 행동 순서를 선택하도록 하는 강화 학습의 방법론이 군사 집단이 가지고 있는 전술과 전략을 가지고 특정 목적을 위하여 움직이는 행동 특성에 매우 부합하기 때문이다.

본 연구에서는 특정 목표를 가지고 움직이는 에이전트의 움직임을 예측하기 위해 에이전트가 목적 달성을 위하여 보상(reward) 값을 극대화하는 방법으로 행동할 것이라고 가정하고 이를 모델링 하였다. 에이전트의 목적은 적에게 발각되지 않고, 체력을 최대한 아끼는 경로로, 최단 경로로 목적지까지 이동하는 것이다. 이러한 목적을 달성하기 위한 보상구조를 설계하기 위해 GIS(Geographic Information System: 지형정보체계)를 통해 지형정보를 수치화하여 활용하였다. 이를 통해 조난자가 탈출 지점까지 이용할

것이라 예측되는 경로를 구할 수 있었고, 추가로 감시초소나 수색 병력과 같은 장애물이 존재하는 환경에서도 예상되는 이동 경로를 구할 수 있었다.

본 논문의 구성은 다음과 같다. 2장에서는 연구의 이론적 배경이 되는 MDP와 관련 연구 현황에 관해 설명하고, 3장에서는 MDP를 기반으로 한 경로 예측 모형에 대해 소개한다. 4장에서는 MDP 모형이 도출한 결과를 정리하고 5장에서는 본 연구의 한계점과 향후 연구 방향을 제시하였다.

2. MDP 개념 및 관련 연구 현황

MDP는 확률을 기반으로 현재 내린 결정이 주는 즉각적인 보상과 시스템 전체에 미치는 영향을 종합적으로 고려하여 의사결정 문제를 해결하는 방법이라고 정의할 수 있다. MDP를 통한 의사결정은 당장의 보상뿐 아니라 에이전트가 최종적으로 달성하고자 하는 목표까지 고려할 수 있다는 점에서 현실 세계에서의 의사결정 과정을 잘 반영하고 있다고 할 수 있다.

본 연구에서는 MDP를 활용하여 모델링 하는 과정에서 GIS를 이용하여 보상구조를 설계하였기 때문에 먼저 기존에 GIS를 활용한 연구들에 대해 살펴보았다. GIS를 활용한 기존 연구들은 크게 입지 선정 문제와 최적 경로 문제로 나누어 볼 수 있었다. 먼저 입지 선정에 관련된 연구는 대표적으로 김영화, 김수환(2015)과 조성진, 이문걸(2019)의 연구를 예로 들 수 있다. 이 연구에선 지형정보의 속성정보를 통해 감시자산의 탐지 확률 지도를 작성하고 이를 통해 감시자산의 최적 배치 후보지를 선정하는데 활용하였다.

본 연구와 관련된 최적 경로를 찾는 문제는 결정론적(Deterministic) 정책 문제와 확률론적(Stochastic) 정책 문제로 세분화하여 볼 수 있었다. MDP에서 전이 확률과 보상은 각 상태(s)와 행동(a)에 따라 달라지는데 이때 정책 $\pi(a | s) = 1$ 일 때, 이를 결정론적 정책이라 하며 정책 $\pi(a | s)$ 가 확률값을 따를 때 이를 확률론적 정책이라 한다. 이를 바꿔 이야기하면 어떤 상태(s)에서 행동(a)를 취할 때 도출되는 정책이 항상 일정한 경우 결정론적 정책, 정책이 확률값에 따라 다르게 도출되는 경우에 확률론적 정책이라 할 수 있다. 먼저 결정론적 정책을 사용한 연구로는 손흥규 등(2014), 신내호 등(2009), 최용찬, 박성수(2020)의 연구가 있었다. 손흥규 등(2014), 신내호 등(2009)은 해당 연구에서 지형정보와 수치 고도 모형을 활용하여 열상 감시장비(TOD)의 탐지율이 낮은 지역을 도출하고 이를 A^* , Dijkstra 알고리즘을 사용하여 최적 경로분석을 실시하였다. 최용찬, 박성수(2020)는 지형정보 데이터를 활용하여 탐지 확률 지도를 생성하고 MDP 기반 모델링, 행위자 기반(ABM) 모형 구축을 통해 피탐 확률이 가장 낮은 정찰 무인기의 비행경로를 제시하였다. 확률론적 정책을 사용한 연구로는 김동호 등(2011)와 진효원 등(2015)이 있다. 김동호 등(2011)은 무인기의 임무 할당 및 정찰 경로 선택에 POMDP(Partially Observable MDP)가 적용된 연구를 수행하였다. POMDP는 기본적으로 MDP의 방법론을 사용하지만 에이전트가 주어진 환경에 대해 시야가 제한되는 상황을 의미한다. 그렇기 때문에 에이전트는 흩어져 있는 정보들을 종합해야 하기 때문에 연산 과정이 MDP보다 다소 복잡하다. 해당 연구에서는 보상구조를 설계함에 있어 대공미사일에 의해 격추될 경우 감점, 임무를 완수하지 못할 경우 매시간당 감점 등을 구조화하여 효과

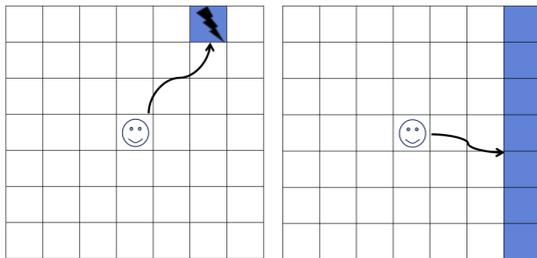
적인 무인기 임무 경로를 제시하였다. 진효원 등(2015)은 지형정보를 MDP의 보상구조를 활용하였으며 침투자의 확률적 위치 추정과 이에 대응하기 위한 봉쇄선 위치 선정에 대한 분석을 실시하였다. 이를 위하여 확률론적 정책을 적용해 침투한 적이 시간의 흐름에 따라 특정 위치에 존재할 확률을 계산하고 이를 봉쇄작전을 위한 의사결정 방법론으로 제시하였다.

본 연구에서는 조난자의 도피·탈출 경로를 예측하기 위하여 MDP를 사용하였고 이 중 보상구조 설계에 있어 GIS를 통해 획득한 지형정보를 활용하였다. 또한 결정론적 정책과 가치 반복법 알고리즘을 적용하였으며 기존 연구 대비 모형의 현실성을 높이기 위해 다음과 같은 점들을 보완하였다. 첫째, 실제 조난 상황을 반영하기 위해 조난자의 구조 지점이 사전에 정해져 있는 경우와 구조 지점이 사전에 정해져 있지 않은 경우 두 가지에 대해 모델링을 실시하였다. 실제 항공작전 수행 시에는 구조 지점이 사전에 정해져 있는 것이 일반적이나, 작전 상황상 사전에 약속된 지점을 이용할 수 없는 경우에는 해안가 방향으로 도피·탈출을 수행하도록 훈련받는다. 이러한 점을 고려해 두 가지 경우에 대해 모델링을 실시하였으며 이를 통해 조난자가 상황에 따라 어떤 탈출 경로를 선택할지에 대해 좀 더 상세하게 알 수 있었다. 둘째, 기존 연구에 반영되지 않았던 지형정보를 보상구조 설계 시 추가 반영하였다. 기존에 반영되지 않았던 강, 하천, 절벽 등과 같은 지형적 요소들을 추가하였으며 기존에 적용된 변수들도 세분화하여 보상구조에 반영하였다. 이러한 요소들이 보상구조에 새로이 반영되면서 모형의 현실성을 더욱 높일 수 있었다.

3. MDP 기반 경로 예측 모형

3.1 문제 정의 및 가정사항

본 연구는 적 지역에서 조난되어 고립된 아군이 도피·탈출을 위하여 이동하는 경로를 MDP를 이용하여 예측하는 문제이다. 일반적으로 군에서 항공작전을 수행할 때에는 조난 상황에 대비하기 위하여 사전에 구조 지점을 선정하고 구조 요청 시 사용할 음어 등을 미리 정한다. 다만 적 지상군의 활동이나 제공권의 미확보 등 작전 상황상 사전에 약속된 지점을 구조 지점으로 활용할 수 없는 경우에는 해안가 방향으로 도피·탈출을 수행하도록 되어 있다. 이러한 이유로 본 연구에서는 사전에 약속된 구조 지점이 있는 경우와 해안가 방향으로 도피·탈출을 수행하는 두 가지 경우에 대해 모델링을 실시하였다.



구조지점이 정해져 있는 경우 해안가로 도피하는 경우

〈Figure-1〉 Two experimental Models

실제 상황에서 적지에 고립된 아군 조난자의 생사 유무가 확인되고 구조활동이 가능한 경우에는 전투 탐색구조 전력이 조난자를 구출하기 위해 출동하게 된다. 전투 탐색구조 전력은 기존의 탐색구조 전력에 적의 공중 및 지상 위협으로부터 아군 전력을 보호하기 위한 전투기, 공격기, 공격 헬기와 같은 엄호 전력이 추가된 개념이다. 본

연구는 조난자가 도피·탈출을 수행하는 데 있어 가장 용이한 경로, 지역이 어디인지를 예측하는 모형이며 모델링을 위하여 가정한 내용은 다음과 같다.

가정 1 : 비상탈출한 조난자는 60×60 셀의 중심에 위치하며 약속된 탈출 지점이나 해안가 방향으로 도피하는 것을 목표로 한다. 틱(tick)마다 사각형의 셀 중심에서 중심으로 이동한다고 가정한다.

가정 2 : 생존 교범상에는 생존 확률을 높이기 위하여 되도록 주간에는 엄폐하고 야간에 이동하는 것을 추천하나, 모델링의 편의를 위하여 조난자는 사전에 약속된 구출 지점이나 해안가를 향하여 끊임없이 이동한다고 가정한다.

가정 3 : 조난자는 훈련받은 대로 탈출 과정에서 적에게 발각되지 않도록 최대한 은신하고, 체력 소모를 최소화할 수 있으며 사전에 훈련받은 지형을 선택(6~7부 능선) 하여 이동한다고 가정한다.

가정 4 : 조난자는 도피·탈출 과정에서 지형적으로 불리한 요소는 돌파할 수 있으나 감시 초소, 수색 병력 등으로 가정된 장애물은 돌파할 수 없으며 우회해야 한다.

3.2 MDP 모델링을 위한 가정 사항

본 장에서는 3.1.에서 정의한 문제에 대해 MDP 기반 모델링을 실시한다. 실제 환경이라면 조난자의 탈출 과정은 연속적으로 이루어질 것이지만 이를 이산화하여 문제를 단순화하였다. 또한 이를 계산하는 과정에서 가치 반복법 알고리즘을 활용하였다.

MDP 모델링을 위한 상태(State), 행동(Action), 전이 확률(Transition Probability), 보상(Reward), 감가

인수(Discount Factor)를 다음과 같이 정의한다.

상태(State)는 조난자가 이동할 수 있는 모든 셀은 보통 상태(Normal State), 조난자가 도달하고자 하는 구조지점이나 해안가 방향을 목적 상태(Object State)로 정의했다.

행동(Action)은 각 셀에서 조난자가 이동할 수 있는 모든 방향을 의미한다. 본 연구에서는 상하좌우에 대각선 방향을 포함하여 총 8개 방향으로 이동할 수 있다고 설정하였다.

보상(Reward)은 조난자가 어떠한 경로로 이동할지를 결정하는데 가장 큰 영향을 주는 요소이다. 조난자의 목표는 조난자가 원하는 목표 상태에 도달할 때까지 최소의 이동거리와 지형적 요소로부터 최소의 감점을 받는 최적의 경로를 찾는 것이다. 조난자는 각각 일반 상태와 목적 상태에서 보상을 얻을 수 있다. 일반 상태에서 얻을 수 있는 보상 값은 지형정보를 기반으로 설계한 보상 값을 받도록 하였으며 이 과정에서 상용 프로그램인 ArcGIS 10.5와 Q-GIS 3.22.7을 활용하였다.

본 연구에서 적용한 보상 값은 충분히 많은 실험을 통해서 설정하였다. 실험을 위해서 각 디지털 지형 조건에서 조난자가 받을 수 있는 보상 값을 도피·탈출에 유리한 요소와 불리한 요소로 구분하여 점수를 부여하였다. 이때 유리한 요소에는 양의 보상 값을 부여하고, 불리한 요소에는 음의 보상 값을 부여하였으며, 유·불리한 요소를 특정할 수 없는 지형에 대해서는 0의 보상 값을 부여하였다.

실험의 목적은 실제 전장 환경에서 조난자의 도피·탈출 경로를 가장 잘 설명할 수 있는 보상 값을 찾는 것이다. 실험에 적용한 보상 값의 범위는 -5점 ~ +5점 사이의 값을 적용하였으며, 보상 값 사이의 간격은 0.5점을 적용하였다. 충분한 실험을 통해서 얻는 최적의 보상 값 범위는 -3점 ~ +2점이고 보상 값간의 간격은 +1점

로, 세부 결과는 <Table-1>과 같다.

<Table-1> Scoring criteria

Point	Criteria
-3	반드시 회피해야 하는 불리한 요소
-2	회피하는 것이 추천되나, 어쩔 수 없다면 돌파해야 하는 요소
-1	불이익을 감수할 수 있는 정도의 요소
0	유리한지 불리한지 판단할 수 없는 요소
2	적극적인 이용이 추천되는 요소
1	이용하면 장점이 있는 요소

보상 값을 0.5점 단위로 세분하는 것은 실험 결과에 크게 영향을 미치지 못했으며, <Table-1>과 같이 각 요소별 점수 차이를 1점 차이의 정수 형태로 부여하여 실험 결과를 해석하기 용이한 장점을 얻을 수 있었다.

<Table-1>의 점수 배점 기준에 따라 추출한 지형정보를 각각 은신, 체력, 능선의 세 가지 요소로 구분하여 각각의 보상 점수를 부여하였다. 은신 요소는 조난자가 적에게 발각되지 않고 이동하기 위해서 필요한 요소로써 요소별 배점은 <Table-2>와 같다.

<Table-2> Points awarded according to stealth factor

Factor	Point
도로(고속도로)	-2
도로(국도)	-1
도로(지방도)	-1
비포장도로	-1
건물/인구밀집지	-3
경작지/논	+1
초지	+1

<Table-2>에 나타난 바와 같이 사람이나 차량의 왕래가 적을수록 발각될 가능성이 낮아지기 때문에 도로와 건물/인구 밀집지에 대해서는 음수의 보상 값을 적용하였으며, 어느 정도 수풀이 우거져 은신하기 용이한 지형인 경작지, 초지 등에 대해서는 양수의 보상 값을 적용하였다.

체력 요소는 조난자가 도피·탈출을 위한 이동 과정에서 체력을 아끼는데 필요한 요소로써 지형의 형태와 경사도에 대한 보상 값이다. 지형의 형태에 따른 배점은 <Table-3>과 같다.

<Table-3> Points awarded according to health factor associated with topographic shape

Factor	Point
절벽/급사면	-2
늪/습지	-2
급류	-3
저수지/호수	-3
강/하천	-3
도랑	-1
산길	+1

<Table-3>에서 절벽이나 늪, 강과 같이 돌파하는데 체력의 소모가 크게 발생할 것이라고 예상되면 음수의 보상 값을, 산길이 나 있어 체력을 아낄 수 있다면 양수의 보상 값을 부여하였다. 또한 지형의 경사도에 따라 체력의 소모가 다를 것이기에 각 지형의 경사도에 따라 <Table-4>와 같이 점수를 배정하였다.

<Table-4> Points awarded according to health factor associated with the slope of each terrain factor

0°-10°	10°-20°	20°-30°	30°-40°
1	0	-1	-2

능선 요소는 조난자가 도피·탈출 과정에서 몇부 능선을 이용할 것인지에 대한 요소로 배점은 <Table-5>와 같다.

<Table-5> Points awarded according to ridge factor

Ridge	Point
6부	+2
5, 7부	+1
4, 8부	0
3, 9부	-1
2, 10부	-2
1부	-3

<Table-5>는 생환 교범에 조난자는 이동 중에 발각될 가능성이 높은 지평선(1부 능선)을 피하고 체력을 아낄 수 있는 5~7부 능선을 이용하도록 되어 있어 이를 근거로 배정된 점수이다.

보통 상태에서 각 셀이 지형정보로 얻을 수 있는 최종 보상 값은 은신, 체력, 능선 3가지 요소들의 총합으로 계산된다. 즉, 연구 지역으로 설정한 60×60의 셀들이 관련된 지형 요소를 가지고 있다면 이에 해당하는 보상 점수를 부여받게 되며 각 셀은 각자가 가지고 있는 지형정보에 따라 상이한 보상 값을 가지게 된다.

보통 상태 이외에도 목적 상태에 도달한 경우에도 보상값을 받도록 되어 있는데, 첫 번째 모형의 경우에는 약속된 구조 지점, 두 번째 모형의 경우에는 해안가 방향(셀의 맨 모서리 지점)에 도달하면 보상값을 받도록 하였다. 시험을 진행 시켜보았을 때 모형의 크기에 비해 지나치게 큰 보상 값을 부여하게 되면 에이전트가 지형 요소의 보상 값을 고려하지 않고 직선 최단거리로 이동해버리는 왜곡 현상을 발견할 수 있었다. 이는 지형적인 유리함을 이용하기보다 최단거리로 이

동하여 목적 상태의 보상 값을 얻는 것이 보상 값을 가장 높게 얻을 수 있는 방법이기 때문에 나타나는 왜곡 현상으로 모형의 크기에 맞추어 예상되는 step의 약 1.5배로 지정하였을 경우 모형이 정상 작동함을 확인할 수 있었다. 이렇게 일반 상태의 보상 값과 목적 상태의 보상 값을 모두 합산한 결과는 60×60 크기의 셀로 Excel 파일로 작성하여 활용하였다.

전이 확률(Transition Probability)은 현재 상태에서 다음 상태로 전이될 확률을 의미한다. MDP 환경에서는 어떠한 행동(Action)을 취했을 때 그 행동대로 상태가 전이될 확률을 의미한다. 본 연구에서 필요한 것은 하나의 최적 정책 π 이므로 결정론적 정책($\pi(a | s) = 1$, 즉 전이 확률이 1)을 통해 최적 경로를 도출하였다.

감가 인수(Discount Factor)은 0~1 사이의 값을 갖는데, 1에 가까울수록 미래에 받을 보상이 현재 받을 수 있는 보상의 가치만큼 크다고 평가하는 것이고 0에 가까울수록 미래에 받을 보상이 현재에 받을 수 있는 보상의 가치보다 낮다고 평가하는 것이다. 본 연구에서는 조난자가 가장 우선하는 가치는 목적 상태에 도달하는 것(구조 지점에 도달하는 것)을 목표로 할 것이라고 보고 0.9를 적용했다.

4. 실험결과 및 분석

4.1 지형정보를 반영한 기본 모형

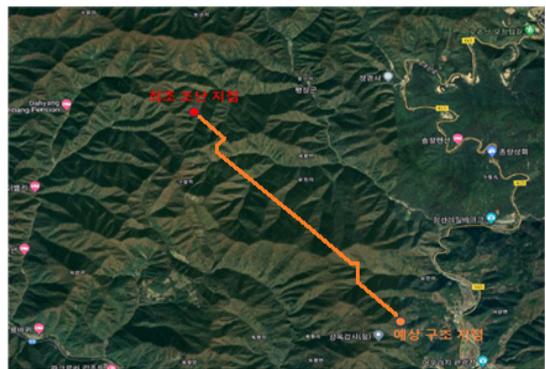
본 연구를 위해 강원도 인근의 12km × 12km 크기의 지역을 연구 지역으로 선정하였다. 이 지역은 도심과 이격 되어 있고 본 연구에서 반영하고자 하는 다양한 지형 요소들을 가지고 있어 연구에 적합한 지형이라고 판단하였다. 조난자는 지도의

중심인 적색 지점에 최초로 조난되었다고 가정하였고 첫 번째 모형은 약속된 예상 구조 지점으로, 두 번째 모형은 우측 끝으로 이동한다고 가정하였다.



〈Figure 2〉 Study area

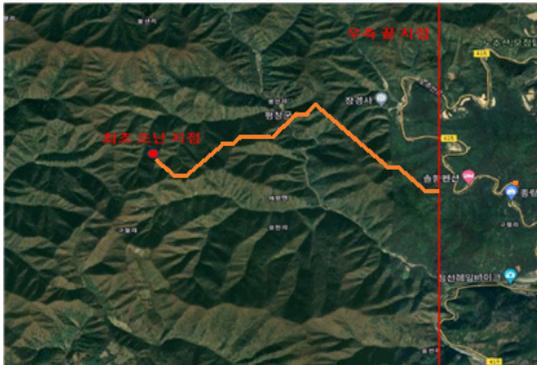
본 모형을 구현하기 위해 AMD Ryzen5 2.4Ghz, RAM 16GB PC를 사용하였으며 프로그램은 Python 3.10 버전을 사용했다. 먼저, 구조 지점이 사전에 정해져 있는 1번 모형에 대해 예상 도피·탈출 경로를 도출하였다. 60×60 셀의 보상구조로 작성된 Excel 데이터를 기반으로 예상 이동 경로를 도출하였고 <Figure-3>과 같이 Q-GIS를 이용하여 Google 맵 상에 예상 이동 경로를 시현시켰다.



〈Figure 3〉 Expected movement of Model 1

1번 모형의 경우에는 지형적으로 다소 불리하더라도 목표지점까지 직선으로 이동하려는 경향이 있었다. 이는 지형적 유리함을 이용하기 위하여 우회하는 것보다 최단거리로 이동하는 것이 보상 값을 극대화할 수 있는 방법이기 때문이다. 실제 상황에서도 조난자는 지형적 유리함을 위하여 우회하기보다는 구조 지점을 향해 최단거리로 이동하는 것을 선호할 것이며 이러한 심리가 모형에 반영되었다고 볼 수 있다.

두 번째로 해안가 방향(본 모형에서는 우측 끝)으로 도피·탈출을 수행하는 2번 모형에 대해 모델링을 실시하였다. 우측 끝 셀을 모두 목적 상태로 설정하였으며 Google 맵을 통해 이동경로를 시현시켰다.



<Figure 4> Expected movement of Model 2

2번 모형의 경우에는 경사도가 완만한 능선을 이용하며 불리한 지형을 최단거리로 돌파하는 등 첫 번째 모형보다 도피·탈출 과정에서 지형적 유리함을 이용하려는 경향이 있음을 확인할 수 있었다. 이는 조난자가 첫 번째 모형 대비 경로 선택의 자유도가 더 높기 때문으로 볼 수 있는데, 첫 번째 모형은 한 지점의 목적 상태로 이동해야 했지만 두 번째 모형은 우측 끝 전체 60개

셀이 목적 상태이기에 그중 한곳만 도달하면 되는 모형으로 경로 선택의 폭이 넓어지기 때문이다. 1번, 2번 모형 모두 목표지점까지 도달하는데 최소 step의 수인 30 step 만에 목표지점에 도달하는 것을 확인할 수 있었다. 다만, 최소 step으로 목표지점에 도달하였다고 하여 이를 최단거리로 볼 수는 없다. 조난자는 셀에서 셀로 이동할 때 직선으로 이동할 때보다 대각선으로 이동할 때 더 먼 거리를 이동해야 하기 때문이다.

4.2 기본 모형의 확장

본 장에서는 모형의 현실성을 보다 높이기 위하여 조난자가 도피·탈출 과정에서 마주칠 수 있는 수색 병력이나 감시초소와 같은 제약조건을 추가하였다. 이와 같은 장애물은 첫 번째 모형을 통해 도출해 낸 조난자의 예상 이동 경로와 평지의 일부 길목 등에 위치할 수 있다고 가정하였으며 소총의 유효사거리(500M)를 감안하여 장애물마다 셀의 중심을 포함하여 3×3의 영역을 감시할 수 있다고 가정하였다. 장애물로 설정된 구역에는 보상구조 설계 시 Big-M 방법을 준용하여 Penalty 값으로 -50의 보상 값을 부여하였다. 이를 통해 에이전트가 해당 지역을 회피하여 이동하도록 하였다. 또한 장애물의 위치는 수시로 변경될 수 있는데, 본 연구에서는 셀의 보상구조를 변경하여 장애물의 위치는 즉각적인 변경이 가능하다. 장애물의 위치는 지도상에 적색으로 표현하였다.

1번, 2번 모형과 같이 Excel 데이터를 기반으로 예상 이동경로를 도출하였고 Q-GIS를 이용하여 Google 맵 상에 이동 경로를 시현시켰다. <Figure 5>는 1번 모형에 장애물을 반영한 확장된 모형에 대한 결과이다. <Figure 5>에 보이는 바와 같이 기존 모형 대비 험준한 경사로를 직선

으로 돌파하는 등 불리한 지형 요소를 일부 감수 하면서 장애물을 회피하는 움직임을 볼 수 있었다.



<Figure 5> Extended model of Model 1

<Figure 5>의 결과는 장애물을 회피하는 과정에서 발생한 불가피한 부분으로, 기본 모형과 유사하게 장애물을 회피하며 최대한 직선으로 이동하려는 유사한 경향성을 확인할 수 있었다.

2번 모형의 확장된 모형의 결과는 <Figure 6>과 같다,



<Figure 6> Extended model of Model 2

<Figure 6>는 2번 모형에 장애물을 반영한 확장 모형으로 1부 능선이나 도로와 같은 다소 불리

한 지형을 직선으로 돌파하는 움직임을 볼 수 있었으나, 첫 번째 모형의 확장 모형 대비 지형적 유리함을 이용하면서도 장애물을 회피하는 움직임을 볼 수 있었다. 이는 기본 모형에서와 같이 경로 선정의 자유도가 높기 때문으로 볼 수 있다. 확장 모형 또한 첫 번째 확장 모형은 31step, 두 번째 모형은 30step에 목적 상태에 도달하여 최소 step인 30step에 거의 근접한 것을 볼 수 있었다.

5. 결론 및 시사점

5.1 연구 결과 및 연구 발전 방향

본 연구의 목적은 유사시 적지에 조난된 조난자의 구조 확률을 높이기 위하여 조난자가 어떠한 경로로 도피·탈출을 수행할지 예측하는 것이다. 조난자는 목표지점으로 이동하고자 하는 목적을 달성하기 위하여 자신에게 주어진 보상 값을 가장 크게 받는 경로로 이동할 것이라고 가정하고 이를 MDP를 통해 모델링 하였다. 이를 위해 GIS를 통해 얻은 실제 지형정보를 보상구조에 반영하였으며 가치 반복법 알고리즘과 결정론적 방법론을 활용하여 예상 경로를 도출하였다. 본 연구가 갖는 의의는 다음과 같다.

첫째, 본 연구에서는 MDP 모형에서 중요한 문제 중 하나인 보상구조를 설계하는 데 있어 실제 지형공간 데이터를 사용함으로써 논리적인 근거를 제시하였다. 또한 기존 연구 대비 모형의 현실성을 높일 수 있도록 보다 상세한 지형정보를 반영하였으며 도피·탈출 과정에서 만날 수 있는 장애 요소를 추가하였다.

둘째, 본 연구에서 제시한 모형은 보상구조의 재설계를 통해 다양한 작전 상황에 응용이 가능

하며 실제 상황에서 조난자의 도피·탈출 경로를 예측하고 전투 탐색구조 작전을 시행하는 데 있어 다양한 요소가 반영된 과학적인 기법에 근거한 의사결정 지원이 가능하다는 장점이 있다.

본 연구의 한계점은 다음과 같다. 첫째, 본 문제에 접근함에 있어 결정론적인 방법론보다 확률론적인 방법이 보다 합리적이라고 볼 수 있는 여지가 있다. 최선의 방책이 있다고 하더라도 모든 사람이 특정 상황에서 100% 동일하게 행동하지 않을 것이며, 적진에 고립되어 모든 행동에 제약이 따르는 상황에서 최선의 방책을 알고 있다 하더라도 이를 그대로 행동으로 옮기기 어려울 것이기 때문이다. 둘째, 본 연구에서 활용한 MDP 방법론은 주어진 환경에 대해 모두 알고 있다고 가정(Model-Based) 하고 해를 도출하는 기법의 연구로서 상태와 행동이 커질수록 해를 도출하기 어려워진다는 단점이 있으며 조난자가 실시간으로 마주칠 수 있는 여러 환경적인 요인들을 모두 반영할 수 없다는 한계가 있다. 따라서 현재 활발히 연구되고 있는 신경망 기법과 강화학습 기법을 결합한 Deep RL(Deep Reinforcement Learning)을 통한 연구가 진행된다면 조난자가 도피·탈출 과정에서 마주칠 수 있는 다양한 환경적 요소들을 실시간으로 반영할 수 있는 보다 현실적인 모형의 구축이 가능할 것이다.

참고문헌(References)

공군본부 편. (2018). 공군 지침서 3-15-2 공중근무자 생활법. 국군인쇄창.

공군본부 편. (2022). 공군 교범 3-7 전투탐색구조작전. 국군인쇄창

국방지형정보단 편. (2011). 군 GIS 길라잡이. 국군

인쇄창.

권기덕, 김인철. (2007). 효율적인 멀티 에이전트 강화학습을 위한 RBFN 기반 정책 모델. 한국지능정보시스템학회 추계 학술대회 논문집, 294-302.

김동호, 이재송, 최재득, 김기웅. (2011). 복수 무인기를 위한 POMDP 기반 동적 임무 할당 및 정찰임무 최적화 기법. 정보과학회지, 39, 453-463.

김영화, 김수환. (2015) MGIS 및 유전자 알고리즘을 활용한 정보자산 최적배치에 관한 연구. 대한산업공학회지, 41(4), 396-407.

노승은. (2020). 바닥부터 배우는 강화학습. 영진닷컴.

손흥규, 허준, 방수남. (2014). 지형공간정보를 활용한 최적침투경로 분석 기법. 대한공간정보학회 학술대회, 11-22.

신내호, 오명호, 최호림, 정동운, 이용웅. (2009). 지형공간정보 기반의 침투위험도 예측 모델을 이용한 최적 침투지역 분석. 한국군사과학기술학회지, 12(2), 199-205.

우창윤, 김도희, 배혜림. (2022). 강화학습을 이용한 신발 생산계획 최적화 연구. 한국지능정보시스템학회 추계학술대회 초록집, 33-34.

육군본부 편. (2015). 야전 교범 15-3-9 야간 작전. 국군인쇄창.

윤정미. (2009). ArgGIS 이해와 활용. 기문당.

이용원, 양혁렬, 김건우, 이영무, 이의령. (2020). 파이썬과 케라스로 배우는 강화학습. 위키북스.

이현재, 이홍주. (2020). 강화학습 주식투자 수익률 비교에 대한 연구. 한국지능정보시스템학회 추계학술대회 초록집, 69-70.

이희연. (2011). GIS 지리정보학. 법문사.

조성진, 이문걸. (2019). 지형공간정보 및 경로탐색

- 알고리즘을 이용한 최적 강하지역 선정 모형. *경영과학*, 36(3), 31-45.
- 진효원, 김수환, 정치정, 이문걸. (2015). 폐쇄공간에서의 에이전트 행동 예측을 위한 MDP 모형. *한국경영과학회지*, 40(3), 63-72.
- 최용찬, 박성수. (2022) 강화학습 기반 적대적 위협 환경하에서의 정찰드론 경로 계획. *차세대융합기술학회논문지*, 624-631.
- 한동윤, 이동현, 백민석, 반재필. (2022). 아군 충돌 회피를 위한 강화학습 기반 지대공 유도 미사일 제어 방안. *한국정보기술학회 추계 종합 학술대회 논문집*, 742-745.
- Rüdiger Thiede 외 3 인. (2014) 실전 Q-GIS 무작정 따라하기. *한국오픈소스 GIS 포럼*.
- Bonet, B and H Geffner, "Solving Stochastic Shortest-Path Problem with RTDP", Technical report, University of California, Los Angeles. (2002)
- Puterman M. L. *Markov Decision Process discrete stochastic* (2005) - 38 - dynamic programming, WILEY.
- Sutton, Richard S., and Andrew G. Barto. *Introduction to reinforcement learning*. (1988) Vol. 2. No.4. Cambridge: MIT press.
- Schaefer, A. J., M. D. Bailey, S. M. Schechter and M. S. Roberts. *Modeling Medical Treatment using Markov Decision Process*, (2004) *Handbook of Operations Research/Management Science Applications in Health Care*. Kluwer Academic Publisher.
- Swarup, S., G. Eubank and M. V. Marathe, *Computational Epidemiology as Challenge Domain for Multiagent Systems*, *Proceeding of International Conference on Autonomous Agents and Multi-agent System*. (2014) International Foundation for Autonomous Agents and Multi-agent System.
- Yu, X.-H. and W.W. Recker *Stochastic Adaptive Control Model for Traffic Signal Systems*, *Transportation Research Part C*, V, 14, (2006) 4:263-282

Abstract

MDP(Markov Decision Process) Model for Prediction of Survivor Behavior based on Topographic Information

Jinho Son* · Suhwan Kim**

In the wartime, aircraft carrying out a mission to strike the enemy deep in the depth are exposed to the risk of being shoot down. As a key combat force in modern warfare, it takes a lot of time, effort and national budget to train military flight personnel who operate high-tech weapon systems. Therefore, this study studied the path problem of predicting the route of emergency escape from enemy territory to the target point to avoid obstacles, and through this, the possibility of safe recovery of emergency escape military flight personnel was increased. based problem, transforming the problem into a TSP, VRP, and Dijkstra algorithm, and approaching it with an optimization technique. However, if this problem is approached in a network problem, it is difficult to reflect the dynamic factors and uncertainties of the battlefield environment that military flight personnel in distress will face. So, MDP suitable for modeling dynamic environments was applied and studied. In addition, GIS was used to obtain topographic information data, and in the process of designing the reward structure of MDP, topographic information was reflected in more detail so that the model could be more realistic than previous studies. In this study, value iteration algorithms and deterministic methods were used to derive a path that allows the military flight personnel in distress to move to the shortest distance while making the most of the topographical advantages. In addition, it was intended to add the reality of the model by adding actual topographic information and obstacles that the military flight personnel in distress can meet in the process of escape and escape. Through this, it was possible to predict through which route the military flight personnel would escape and escape in the actual situation. The model presented in this study can be applied to various operational situations through redesign of the reward structure. In actual situations, decision support based on scientific techniques that reflect various factors in predicting the escape route

* Department of Military Science, Korea National Defense University

** Corresponding Author: Suhwan Kim

Department of Military Science, Korea National Defense University

1040, Hwangsanbeol-ro, Yangchon-myeon, Nonsan-si, Chungcheongnam-do, 33021, Korea

Tel: +82-41-831-5372, Fax: +82-2-748-7588, E-mail: ksuhwan@korea.kr

of the military flight personnel in distress and conducting combat search and rescue operations will be possible.

Key Words : MDP, GIS, Optimal Path, Distress, Reward Structure

Received : December 27, 2022 Revised : March 27, 2023 Accepted : April 17, 2023

Corresponding Author : Suhwan Kim

저 자 소개



손진호

현재 국방대학교 국방과학학과에서 석사학위를 취득하고 공군 방공관제사령부에 재직 중이며 한서대학교 항공교통물류학과에서 학사학위를 취득했다. 주요 관심분야는 강화 학습, 인공지능 등이다.



김수환

현재 국방대학교 국방과학학과 교수로 재직 중이다. 육군사관학교에서 학사, KAIST 산업 및 시스템공학과에서 석사를 취득하고, 미 Texas A&M 대학교에서 산업공학 전공 박사학위를 취득하였다. 한국국방경영분석학회 이사, 한국군사학논집 편집위원을 역임하였다. 주요 관심분야는 조합 최적화, 비용추정, 인공지능 등이다.