Journal of the Korean Institute of Industrial Engineers
Vol. 42, No. 1, pp. 65-72, February 2016.
ISSN 1225-0988 | EISSN 2234-6457
http://dx.doi.org/10.7232/JKIIE.2016.42.1.065

<Application Research>

MDP에 의한 컬링 전략 선정

배기욱 · 박동현 · 김동현 · 신하용†

KAIST 산업 및 시스템공학과

Markov Decision Process for Curling Strategies

Kiwook Bae · Dong Hyun Park · Dong Hyun Kim · Hayong Shin

Department of Industrial and Systems Engineering, KAIST

Curling is compared to the Chess because of variety and importance of strategies. For winning the Curling game, selecting optimal strategies at decision making points are important. However, there is lack of research on optimal strategies for Curling. 'Aggressive' and 'Conservative' strategies are common strategies of Curling; nevertheless, even those two strategies have never been studied before. In this study, Markov Decision Process would be applied for Curling strategy analysis. Those two strategies are defined as actions of Markov Decision Process. By solving the model, the optimal strategy could be found at any in-game states.

Keywords: Sports, Curling, Markov Decision Process, Strategy, Aggressive, Conservative

1. 서 론

스포츠에서 전략은 경기의 진행 방향을 결정하는 요소이다. 전략 선택은 경기 시작 전이나 경기 중에 발생할 수 있으며 팀 혹은 선수는 선택할 수 있는 여러 전략 중 경기에서 이기기 유리한 전략을 선택하고자 한다. 1900년대 후반, 2000년대부터 테니스, 야구, 미식축구 등 다양한 스포츠 종목에서 산업공학에서 사용하는 다양한 기법인 동적 계획법, 확률 모형, 최적화 기법 등을 활용하여 경기에서 이기기 위한 최적의 전략을 찾는 연구가 지속적으로 이루어져 왔다(Wright, 2009).

흔히 '빙판 위의 체스'라고 불리는 컬링은 체스에 비유될 만큼 전략의 종류가 다양하고 또한 전략 선택이 경기 승패에 결정적인 영향을 미친다. 컬링 경기를 이기기 위해선 현재 놓여진 상황마다 최선의 전략을 선택해서 경기해야 한다. 그러나 컬링에서 최적의 전략 선택에 대한 연구는 아직까지 많이 이루어지지 않아 각 팀마다 경험이나 직관에 의한 전략 선택에 의존하는 것이 현실이다.

서론에서 컬링 규칙과 전략, 기존에 있던 관련 연구에 대해 먼저 살펴보고 본론에서는 전략 선택을 포함하는 컬링의사 결정 모델을 정의하고자 한다. 마지막으로 결과 부분에서는 본문에서 정의한의사 결정 모델을 예시를 통해 품으로써 경기 중 상황 별로 최적의 전략을 찾을 수 있다는 것을 보이려고 한다. 본 연구에서 개발한 컬링 의사 결정 모델과 실제 데이터를 활용한다면 실전에서도 항상 최적의 전략을 찾을 수 있을 것으로 기대한다.

1.1 컬링 규칙과 전략

컬링 전략 분석을 위한 모델을 정의하는데 있어서 규칙부터 알아보고자 한다. 컬링 규칙은 캐나다 컬링 협회에서 제공하는 자료를 참고하였다(CCA, 2014). 컬링 경기는 컬링 전용 빙상에서 진행되며 선수들은 서로 자기 팀의 스톤을 던져 빙상에 있는 과녁(하우스)의 중심 가까이 놓기 위해 경쟁하는 스포츠이다. 컬링 경기는 총 10개의 엔드로 구성되어 있으며 각 엔드마다 빙판 위에서 하우스를 향해 양 팀이 번갈아 가며 총 16개의 스톤을 던지고, 가장 중심에 가까이 위치시킨 팀이 그 엔드에서 승리하게 된다. 엔드를 승리하게 되면 승리한 팀이 하우스 안에 상대편 스톤보다 더 중심에 가까이 위치시킨 자신

† 연락저자 : 신하용 교수, 34141 대전광역시 유성구 대학로 291 KAIST 산업 및 시스템공학과, Tel : 042-350-3124, Fax : 042-350-3110, E-mail : hyshin@kaist.ac.kr
2015년 9월 1일 접수; 2015년 10월 29일 1차 수정본 접수; 2015년 12월 12일 2차 수정본 접수; 2015년 12월 22일 게재 확정.

스톤의 개수 만큼 점수로 가져간다. 아래 그림은 한 예시이다. 노란색 스톤 2개가 빨간색 스톤보다 하우스 중심에 가까이 있으므로 노란색 팀이 2점을 얻게 된다.

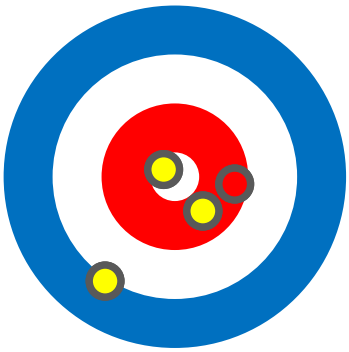

Figure 1. An Example of Scoring in Curling

컬링 경기에선 후공의 기회가 이전 엔드 결과에 따라 결정된다. 이전 엔드에서 점수를 얻지 못한 팀이 후공의 기회를 가져간다. 만약 두 팀 다 하우스 안에 스톤을 넣지 못해 점수를 얻지 못한다면 그 엔드는 'Blank End'가 되어 후공 기회를 가지고 있던 팀이 다음 엔드에서도 가지게 된다. 후공의 기회를 가지고 있는 팀은 상대방의 투구를 보고 자신의 전략을 세울 수 있다는 점과 마지막 스톤을 통해 엔드를 결정지을 수 있다는 점에서 후공이 아닌 팀보다 점수를 얻기가 더 유리하다. 그러므로 후공의 여부는 전략을 결정하는데 있어서 중요한 요소가 된다.

컬링 전략은 엔드를 시작할 때 엔드 운영을 결정하는 전략부터 투구 하나하나 어떻게 던질지 결정하는 전략까지 그 종류가 다양하고 또한 전략을 결정해야 하는 상황이 많이 발생한다(Curltech, Online). 그 중 가장 일반적으로 받아들여지는 전략으로는 엔드를 수비적으로 경기할지, 공격적으로 경기할지 결정하는 전략이 있다(Kostuk *et al.*, 2001). 수비적인 전략은 현재 점수 차이를 안전하게 유지하고자, 공격적인 전략은 점수를 잃을 위험을 감수하더라도 가능한 많은 점수를 얻고자 하는 전략이다. Curling Canada에서 제공하는 전략 가이드라인에 두 전략에 대해 자세히 소개가 되어 있다(Curling Canada, Online).

수비적인 전략, 공격적인 전략 모두 엔드를 시작할 때 후공 여부에 따라서 다르게 나타난다. 먼저, 후공일 때 수비적인 전략은 하우스 안에 있는 스톤의 수를 최대한 줄이면서 상대방에게 점수를 주지 않으려는 전략이고, 반대로 공격적인 전략은 하우스 안에 많은 스톤들을 위치시켜서 위험을 감수하더라도 한 번에 큰 점수를 얻으려는 전략이다. 1점을 얻을 수 밖에 없는 상황으로 경기가 흘러간다면 1점을 얻기보다 그 엔드를 'Blank End'로 만들고 다음 엔드에서 다시 큰 점수를 얻고자 한다. 후공이 아닐 때 수비적인 전략은 상대방이 최대 1점만 얻고 더 큰 점수는 얻지 못하도록 하는 전략이고 공격적인 전략은 적은 기회이긴 하지만 점수를 따고자 하는 전략이다.

경기 후반, 상대방보다 점수가 뒤지고 있고 후공인 경우공격적인 전략을 선택해야 한다는 것이 명확하지만, 예를 들어, 7엔드에 동점일 때 후공인 상황처럼 어느 전략을 선택하는 것이 더 좋다고 확실하게 말 할 수 없는 경우도 많다. 따라서, 컬링 경기 중 펼쳐지는 상황마다 두 전략 중 어떤 전략을 선택하는 것이 경기를 이기기 유리한 최적의 전략인지 찾는 연구는 가까이는 경기 승리부터 멀리는 컬링 전략의 고도화 등 컬링 발전에도 영향을 미칠 것으로 기대한다.

수비적인 전략과 공격적인 전략이 컬링에 있어서 일반적으로 받아들여지고 경기의 방향을 결정하는 중요한 전략인데 반해컬링 전략과 관련 된 연구 중 이 두 전략을 비교하는 연구는 없었다. 전략을 비교한 연구 외에 Kostuk *et al.*(2001)의 논문을 보면 컬링을 마크로프 연쇄 모형으로 모델링 하여 분석하는 연구를 진행했었다.

본 연구에서는 컬링 경기 중 펼쳐지는 상황마다 수비적인 전략과 공격적인 전략 둘중 어느 전략이 최적의 전략인지 찾기 위한 컬링 모델을 만들고자 한다. Kostuk *et al.*(2001)의 연구에서 활용된 마르코프 연쇄 모형에서 더 나아가 전략 선택도 포함하는 수학적 모형을 활용하여 컬링경기를 모델링 한다면 그 모형을 품으로써 상황마다 최적의 전략을 알 수 있을 것으로 기대한다. 먼저, 관련 연구에서 Kostuk *et al.*(2001)의 논문을 살펴본 후, 본문에서는 이를 발전시켜 공격적인 전략과 수비적인 전략을 action으로 정의하는 Markov Decision Process를 활용해 엔드마다 전략을 선택하는 컬링 모델을 만들고자 한다.

1.2 관련 연구

Kostuk *et al.*(2001)에 실린 연구는 컬링과 야구 모두 '투구'라는 discrete한 event로 경기가 진행된다는 점에서 착안해야구를 마르코프 연쇄 모형으로 모델링 한 Bukiet *et al.*(1997)의 논문을 참고하여 컬링을 마르코프 연쇄 모형으로 모델링하였다.

Kostuk *et al.*(2001)의 연구는 매 엔드의 시작을 state로 정의하였고 상대팀과의 점수 차이, 후공의 여부로 state를 표현하였다. 그리고 엔드에서 얻은 점수로 다음 엔드로의 transition이 결정된다. 예를 들어, 현재 state가 1점 앞선 후공이라면 해당 엔드에서 2점을 얻었을 때 다음 state는 3점 앞선 선공으로 결정된다. 그는 1985년부터 1997년까지 누적된 Canadian Championship 데이터로부터 후공을 가진 팀이 각 엔드마다 몇 점을 얻었는지 분포를 얻었고 이는 다음 <Table 1>과 같다. <Table 1>을 보면 2엔드에서 후공인 팀이 1점을 얻은 횟수는 전체 경기 수 중에 322번 있었고 이러한 분포표로부터 state 간의 transition probability를 정의하였다.

Kostuk *et al.*(2001)은 마르코프 연쇄 모형을 가정함으로써 모델의 transition, 즉, 컬링 경기의 진행이 현재 state에만 의존하는 확률 모형으로 컬링 경기를 나타내었다. 그러나 실제 컬링 경기는 진행 도중에 수비적인 전략, 공격적인 전략처럼 전략을 선택하고 전략마다 얻게 되는 점수의 분포가 다르게 나타난다.

본 연구에선 transition이 현재 state 뿐만 아니라 전략의 선택

Table 1. Frequency Table of Scroes at Each END with Hammer

END	-6	-5	-4	-3	-2	-1	0	1	2	3	4	5	6
1			1	4	25	104	251	322	156	32	7		
2			2	6	53	133	184	322	162	32	6	2	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10			1	8	19	92	22	183	70	10	4	1	1
11					4	15	4	85	7	2	1		
12					1	3							
Totals	0	4	23	92	434	1371	1523	2865	1603	407	78	18	3

에도 의존하는 Markov Decision Process(이하 MDP) 모델을 도입해서 컬링 경기를 나타내어 경기 진행 중 최적의 전략을 모델을 품으로써 찾을 수 있을 것으로 기대한다.

2. MDP(마르코프 의사 결정 모델)

MDP는 각 단계에서 선택한 action에 따라 state transition probability가 달라지는 상황에서 최적의 action 선택을 다루는 모형이다. 즉, MDP에서는 각 state마다 action을 선택 할 수 있고 다음 state로 갈 확률 transition probability는 현재 state와 선택한 action에 의존한다. Action을 취하고 state transition이 발생할 때, state와 action에 따라 reward도 발생한다. State, action, transition probability, reward를 알고 있으면 각 state에서 연속되는 transition을 통해 최종적으로 얻을 것으로 기대되는 총 reward를 최대화하는 action의 조합을 찾을 수 있다.

다음의 그림은 MDP의 한 time step에서 이루어지는 action 선택과 상태 전이를 도식적으로 나타낸 것이다. 이 그림에서 직사각형은 state를 마름모는 action 선택을 원은 확률적 상태전이를 나타낸다.

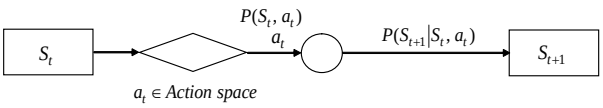

Figure 2. Diagram of One Step Transition in MDP

t시점에서 state, S_t로부터 연쇄적으로 transition이 발생할 때, S_t에서 총 얻을 것으로 기대되는 reward의 총합의 최댓값을 value라고 정의하고 $V_t(S_t)$로 표현한다. 이는 Bellman's equation 형태로도 나타낼 수 있다(Powell, 2007).

$$V_t(S_t) = \max_{a_t}\{R(S_t,a_t) + \gamma E[V_{t+1}(S_{t+1})|S_t]\}$$

$R(S_t, a_t)$는 S_t에서 action a_t를 선택해서 발생하는 transition을 통해 얻을 수 있는 reward의 기댓값을 의미하고, $E[V_{t+1}(S_{t+1})|S_t]$는 S_t에서 transition 된 다음 state S_{t+1}에서 얻을 것으로 기대되는 value이다. Value는 Bellman's equation을 backward induction으로 풀어서 구할 수 있다.각 state에서 value를 구하면 그 때 a_t, 즉 state에서 optimal action 또한 알 수 있다.

3. 컬링 전략에 대한 마르코프 의사 결정 모델

State, action, reward, transition probability를 모두 아는 경우, Bellman's equation을 풀 수 있으므로 각각 state에서 value와 최적의 전략을 찾을 수 있다. 따라서 컬링 경기를 수비적인 전략과 공격적인 전략을 선택할 수 있는 MDP로 모델링하고 풀기 위해선 state, action, reward, 그리고 transition probability부터 정의해야 한다. 각 엔드 시작 할 때 공격적인 전략과 수비적인 전략 중 하나를 선택하는 상황에 맞도록state와 action, transition을 정의해야 한다.

3.1 States

위에서 언급한 Kostuk *et al.*(2001)의 연구에선 t엔드가 시작할 때 상대방과의 점수 차이, 후공의 유무를 state S_t로 정의하였다. 전략을 선택하는 시점이 엔드의 시작이므로 본 연구에서도 동일한 state를 적용하였다. 이를 수식으로 나타내면 다음과 같다.

$$S_t = (d_t, h_t)$$
$$t \in \{1, \cdots, 12\},\ d_t \in \text{정수},\ h_t \in \{-1, 1\}$$

컬링 경기는 10엔드까지 있지만 10엔드 종료 시 동점일 경우, 연장으로 11엔드를 진행하기 때문에 t는 1부터 최대 12까지의 값을 가진다.

d_t는 기준 팀에서 상대팀의 점수를 뺀 차를 의미한다. $d_t = 0$이라면 두 팀은 t엔드 시작할 때 동점이다라는 의미다. h_t는 기준 팀의 후공 여부를 나타낸다. t엔드 시작할 때 후공이면 $h_t = 1$, 후공이 아니면 $h_t = -1$이다.

3.2 Actions

우리가 모든 state에서 취할 수 있는 action은 공격적인 전략과 수비적인 전략이 있다. 공격적인 전략을 *agg*라고 정의하고 수비적인 전략을 *con*라고 정의하면 선택할 수 있는 action들의 집합은 다음과 같이 나타낼 수 있다.

$$Action\ Space = \{agg,\ con\}$$

3.3 Transition Probability

Kostuk *et al.*(2001)의 논문은 t엔드에서 점수 분포를 통해 state 간의 transition을 표현할 수 있다.하지만 MDP 모델은 transition이 현재 state와 전략에 의존하므로 전략에 따른 확률 분포도 알아야 한다. 하지만 t엔드에서 전략에 따른 점수 확률 분포를 알 수 있는 데이터가 없기에 Kostuk *et al.*(2001)의 논문에 실린 데이터를 최대한 활용하고 컬링 전략의 가이드라인을 통해 알 수 있는 전략 별 특징들을 transition probability를 가정할 때 반영하였다. 제 3.3절에서는 확률 분포를 어떻게 가정하였는지 예시를 통해 그 과정을 기술하였고 결과 부분에서는 다른 확률 분포로 모델을 풀면서 확률 분포가 달라짐에 따라 optimal action이 어떻게 달라지는지 sensitivity analysis를 진행하였다.

실제 경기에서 엔드의 점수 분포에 영향을 주는 요소로는 전략, 후공 여부, 남은 엔드의 수, 현재 점수 차이 등이 있다. 남은 엔드의 수와 현재 점수 차를 고려하여 전략 선택이 이루어진다고 가정하였다. 다시 말해, 점수 확률 분포는 엔드에 무관하고 후공 여부와 전략에 의존한다고 가정하였다. 데이터를 통해 후공 여부에 따른 점수 확률 분포를 얻고 가이드라인을 통해 알 수 있는 전략마다의 특징을 갖는 점수 확률 분포를 가정하였다.

(1) 후공 여부에 따른 점수 확률 분포

제 1.3절의 <Table 1>을 다시 한 번 살펴보면 마지막 행이 후공 일 때 얻는 점수에 대한 marginal probability이다. 후공 일 때 x_t점을 얻을 확률을 $f(x_t)$라고 할 때, $f(x_t)$의 확률 분포는 <Table 2>와 같다.

후공이 아닐 때 점수 분포 데이터는 따로 정리되어 있지 않지만 후공일 때 데이터를 활용하여 후공이 아닐 때 점수 확률 분포도 알 수 있다. 후공 일 때 1점을 잃었다는 것은 상대편은 후공이 아닐 때 1점을 얻었다는 것과 동일하기 때문에 후공이 아닐 때 x_t점을 얻을 확률을 $\bar{f}(x_t)$라고 하면 $\bar{f}(x_t)$는 후공 일 때 x_t점을 잃을 확률과 같다.

$$\bar{f}(x_t) = f(-x_t)$$

(2) Action 별 점수 확률 분포

후공 일 때 수비적인 전략의 가이드라인을 보면 상대방에게 점수를 주지 않으려는 전략이기 때문에 0점 혹은 1점을 얻을 확률이 높다. 하지만 상대방도 같이 경기를 하고 예외가 발생하기 때문에 0점 1점 얻는 것을 제외한 다른 확률도 존재한다. 수비적인 전략을 택했을 때 후공의 점수 확률 분포를 $f_{con}(x_t)$라고 하였을 때, 예시로 1점을 얻을 확률이 0.6, 0점을 얻을 확률이 0.2, 2점을 얻거나 1점을 잃을 확률을 각각 0.1인 확률 분포를 가정하였다.

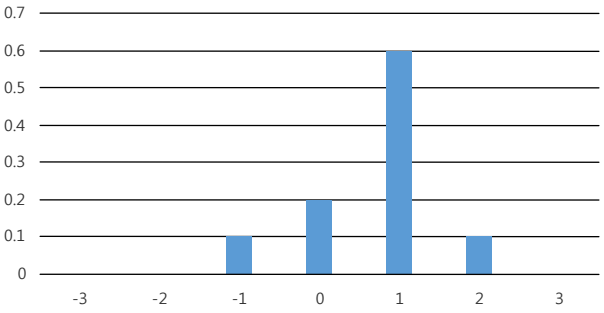

Figure 3. Probability Distribution of the Scores with Conservative Strategy(with Hammer)

Table 4. Probability Distribution of the Scores with Conservative Strategy(with Hammer)

x_t	-3	-2	-1	0	1	2	3
$f_{con}(x_t)$	0	0	0.1	0.2	0.6	0.1	0

후공 일 때 공격적인 전략은 수비적인 전략보다 점수를 잃거나 큰 점수를 얻을 확률이 높다. 이를 위에서 가정한 수비적인

Table 2. Probability Distribution of the Scores with Hammer

x_t	-3	-2	-1	0	1	2	3
$f(x_t)$	0.0143391	0.051693	0.163031	0.183034	0.334980	0.192312	0.060609

Table 3. Probability Distribution of the Scores of Teams Without Hammer

x_t	-3	-2	-1	0	1	2	3
$\bar{f}(x_t)$	0.060609	0.192312	0.334980	0.183034	0.163031	0.051693	0.014339

Table 5. Probability Distribution of the Scores with Aggressive Strategy(with Hammer)

x_t	-3	-2	-1	0	1	2	3
$f_{agg}(x_t)$	0.02867	0.10338	0.22606	0.16606	0.06996	0.28462	0.12121

전략의 점수 확률 분포와 후공의 점수 확률 분포를 통해서 구하고자 한다. $P(a_t = con)$를 수비적인 전략을 선택할 확률, $P(a_t = agg)$를 공격적인 전략을 선택할 확률이라고 하면 이 둘은 서로 상호배반이면서 합은 1이기 때문에 $f(x_t)$와 $f_{agg}f(x_t)$는 Law of total probability에 의해 다음과 같이 나타낼 수 있다.

$$f(x_t) = f_{con}(x_t)P(a_t = con) + f_{agg}(x_t)P(a_t = agg),$$
$$f_{agg}(x_t) = \frac{(f(x_t) - f_{con}(x_t)p(a_t = con))}{p(a_t = agg)}$$

위 식으로부터 공격적인 전략을 선택하였을 때 후공의 점수 확률 분포 $f_{agg}f(x_t)$를 구할 수 있다. $f_{agg}f(x_t)$로 가정한 여러 확률 분포 중 $P(a_t = con) = P(a_t = agg) = 0.5$의 경우를 예시로 보면, $f_{agg}f(x_t)$는 <Table 5>와 같다.

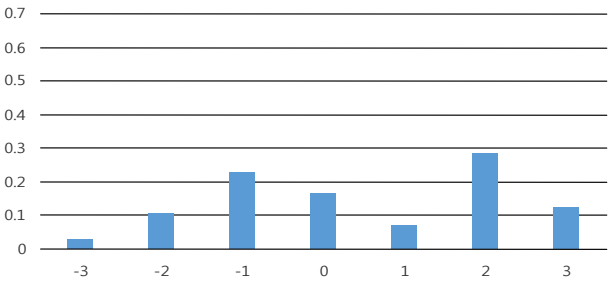

Figure 4. Probability Distribution of the Scores with Aggressive Strategy(with Hammer)

후공이 아닐 때 수비적인 전략과 공격적인 전략의 점수 확률 분포도 마찬가지로 구할 수 있다. 후공이 아닐 때 수비적인 전략의 점수 확률 분포, $\overline{f_{con}}(x_t)$는 <Table 6>에 나타내었고, 후공일 때와 마찬가지로 law of total probability를 활용하여 후공이 아닐 때 공격적인 전략의 점수 확률 분포 $\overline{f_{agg}}(x_t)$도 <Table 7>과 같이 가정하였다.

Table 6. Probability Distribution of the Scores with Conservative Strategy(without Hammer)

x_t	-3	-2	-1	0	1	2	3
$\overline{f}_{con}(x_t)$	0	0.1	0.6	0.2	0.1	0	0

(3) Action 별 점수 확률 분포와Transition Probability

후공 여부와 전략에 따른 점수 확률 분포를 가정했고 이 확률 분포로 transition probability를 구할 수 있다.현재 S_t에서 action a_t를 선택하였고 x_t의 점수를 얻어서 S_{t+1}로 transition이 일어났다면 S_{t+1}에서 점수 차이, d_{t+1}는 $d_{t+1} = d_t + x_t$이 된다. 즉, $x_t = d_{t+1} - d_t$이다. 그러므로 a_t를 선택 했을 때 $x_t = d_{t+1} - d_t$의 점수를 얻을 확률이 S_t에서 a_t를 선택했을 때 S_{t+1}로 transition할 확률이 된다. 이 때, t엔드에서 후공 여부, h_t에 따라 확률이 달라진다.

$$P(S_{t+1}|S_t, a_t) = \begin{cases} f_{a_t}(d_{t+1} - d_t), & if\ h_t > 0 \\ \overline{f}_{a_t}(d_{t+1} - d_t), & if\ h_t < 0 \end{cases}$$

$x_t > 0$일 때 기준 팀은 t엔드에서 점수를 얻었으므로 t+1엔드는 상대방이 후공의 기회를 가져간다. 따라서 $h_{t+1} = -1$이다. 만약 $x_t = 0$이라면 t엔드는 'Blank End'이므로 후공은 그대로 유지된다. 이 경우, $h_{t+1} = h_t$이다. 마지막으로 $x_t < 0$이면 기준 팀이 다음 엔드에 후공을 가져가므로 $h_{t+1} = 1$이다.

3.4 Reward

컬링 경기에서 전략 선택은 경기가 끝났을 때 이기는 것을 목적으로 한다. 경기 중간 transition을 통해서 경기의 승패가 결정되지 않기 때문에 경기 중간 state 간의 transition에서 reward는 0으로 가정하였다.

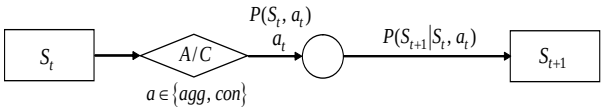

Figure 5. Transition at S_t of Curling Markov Decision Process (t = 1, ⋯ , 10)

<Figure 5>는 컬링 MDP 모델의 t엔드에서transition을 다이어그램으로 표현한 것이다. t엔드가 시작할 때, 즉, S_t에선 공격적인 전략, 수비적인 전략 중 하나를 선택해야 하는 의사 결정 상황이 발생하고 선택한 전략 a에 따라 다음 state S_{t+1}로 전이하는 transition probability, $P(S_{t+1}|S_t, a)$가 결정된다. 현재 S_t에

Table 7. Probability Distribution of the Scores with Aggressive Strategy(without Hammer)

x_t	-3	-2	-1	0	1	2	3
$\overline{f}_{agg}(x_t)$	0.12121	0.28462	0.06996	0.16606	0.22606	0.10338	0.02867

서 action을 선택하여 transition이 일어날 때, reward인 $R(S_t, a)$가 발생하지만 컬링의 경우, reward는 0이라고 가정하였기에 $R(S_t, a) = 0$이다. 경기가 시작할 때인 S_1부터 경기가 끝날 때까지 transition이 연쇄적으로 발생한다.

4. 컬링 MDP 모델의 해

4.1 Value function의 초기값 설정

컬링 경기의 MDP를 풀기 위해선 각 state의 value function 초기 값을 정의해주어야 한다. 그리고 경기가 종료되는 state부터 backward induction으로 문제를 풀면 각 state에서 얻을 것으로 기대되는 value와 그 때 optimal policy를 찾을 수 있다.

컬링 경기는 정해진 10엔드까지 진행 했을 때 점수가 앞서는 팀이 승리하면서 경기가 종료되지만 만약 두 팀의 점수 차이가 없다면 연장 엔드를 진행한다. 연장 엔드는 점수를 1점을 얻는 팀이 승리하게 되므로 전략의 특성이 이전 엔드와는 다르다. 따라서 위에서 정의한 MDP 모델을 따르지 않고 따로 value function을 구하였고 10엔드가 끝났을 때 value function의 값을 정의하였다.

(1) 경기 종료 시점에서Value function

컬링은 10엔드까지 경기를 했을 때, 점수를 앞서고 있는 팀이 승리하고 경기가 종료된다. 만약 동점일 경우, 연장으로 11엔드를 진행하여 한 팀이 이길 때까지 경기가 이루어지기 때문에 10엔드 이후에 $d_t \neq 0$이여야 모델의 transition이 종료된다. Value function은 승률로 정의하였다. 10엔드 이후 점수가 앞서면 경기에서 이긴 것이기 때문에 value는 1, 점수가 뒤지면 반대로 0으로 정의하였다.

$$V_t(S_t = (d_t, h_t)) = \begin{cases} 1, & d_t > 0 \\ 0, & d_t < 0 \end{cases} \quad if \;\; t > 10$$

(2) 연장 엔드에서Value function

10엔드가 끝났을 때 동점인 상황의 $S_{11} = (0, 1)$, $S_{11} = (0, -1)$의 value는 경기 종료 시점이 아니기 때문에 따로 정의해야 한다. 11엔드는 연장 엔드로 먼저 1점을 내는 팀이 경기에서 승리하게 되는 특성 상 transition probability가 전략과 관계없이 현재 state에만 의존한다고 가정하였다. 이 때 transition probability는 <Table 1>의 11엔드 점수 확률 분포로 정의하였다.

$$\begin{aligned} V_{11}(S_{11} = (0, h_{11})) &= E[\, V_{12}(S_{12}) | S_{11}] \\ &= \sum_{S_{12}} P(S_{12} | S_{11})\, V_{12}(S_{12}) \end{aligned}$$

위의 과정을 통해 10엔드가 끝났을 때 모든 state에서 value function을 정의하였다. 10엔드가 끝난 state의 value인 $V_{11}(S_{11})$부터 컬링 MDP 모델의 Bellman's equation을 backward induction으로 품으로써 모든 state에서의 value와 optimal action을 구할 수 있다.

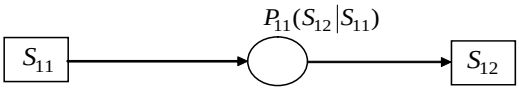

Figure 6. Diagram of transition at 11END

Table 8. Probability distribution of the scores at 11 END and 12 END

	End	-3	-2	-1	0	1	2	3
	11	0	0.03390	0.12712	0.03390	0.72034	0.05932	0.02542
12	0	0.25	0.75	0	0	0	0	

4.2 컬링 MDP 모델의 Bellman's Equation

$V_{10}(S_{10})$부터는 Bellman's equation을 backward induction으로 풀어서 value와 optimal action을 구할 수 있다. 컬링 모델은 reward가 없기 때문에 모든 state에서 $C(S_t, a_t) = 0$ 이고 $\gamma = 1$ 이므로 Bellman's equation을 컬링 모델에 맞게 수정하면 아래 식과 같아진다.

$$\begin{aligned} V_t(S_t) &= \max_{a_t} \{ R(S_t, a_t) + \gamma E[\, V_{t+1}(S_{t+1}) | S_t] \} \\ &= \max_{a_t} \{ E[\, V_{t+1}(S_{t+1}) | S_t] \} \\ &= \max_{a_t} \left\{ \sum_{S_{t+1}} P(S_{t+1} | S_t, a_t)\, V_{t+1}(S_{t+1}) \right\} \end{aligned}$$

이 때, 선택된 action이 S_t에서 optimal action이 되고 $\Pi_t(S_t)$로 나타낸다.

$$\Pi_t(S_t) = \underset{a}{\operatorname{argmax}} \left\{ \sum_{S_{t+1}} P(S_{t+1} | S_t, a)\, V_{t+1}(S_{t+1}) \right\}$$

t = 10부터 t = 1까지 모든 state에서 value function과 optimal action을 구할 수 있다.

5. 결 과

컬링 마르코프 의사 결정 모델의 state, action, transition probability, reward 모두 정의하였고 초기 value function도 설정하였으므로 본문에 있는 예시로 문제를 풀어보았다. <Table 9>와 <Table 10>에 각 엔드, 각 state마다 최대의 value를 가지게 하는 optimal action을 정리하였다.

점수가 뒤지고 있을 때 공격적인 전략을 선택하고 점수가 앞서고 있을 때 수비적인 전략을 선택하는 것은 후공 여부에 관계없이 동일하였다. 그러나 동점일 경우, 후공 일 때는수비

Table 9. Optimal Action of Each States(with Hammer)

	$\Pi_1(S_1)$	$\Pi_2(S_2)$	$\Pi_3(S_3)$	$\Pi_4(S_4)$	$\Pi_5(S_5)$	$\Pi_6(S_6)$	$\Pi_7(S_7)$	$\Pi_8(S_8)$	$\Pi_9(S_9)$	$\Pi_{10}(S_{10})$
d = -2		AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
d = -1		AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
d = 0	CON	CON	CON	CON	CON	CON	CON	CON	CON	CON
d = 1		CON	CON	CON	CON	CON	CON	CON	CON	CON
d = 2		CON	CON	CON	CON	CON	CON	CON	CON	CON

Table 10. Optimal Action of Each States(without Hammer)

	$\Pi_1(S_1)$	$\Pi_2(S_2)$	$\Pi_3(S_3)$	$\Pi_4(S_4)$	$\Pi_5(S_5)$	$\Pi_6(S_6)$	$\Pi_7(S_7)$	$\Pi_8(S_8)$	$\Pi_9(S_9)$	$\Pi_{10}(S_{10})$
d = -2		AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
d = -1		AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
d = 0	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
d = 1		CON	CON	CON	CON	CON	CON	CON	CON	CON
d = 2		CON	CON	CON	CON	CON	CON	CON	CON	CON

Table 11. Probability Distribution of the Scores with Conservative Strategy(with Hammer)

x_t	-3	-2	-1	0	1	2	3	Expected Score
example1	0	0	0.1	0.25	0.55	0.1	0	0.65
example2	0	0	0.1	0.2	0.6	0.1	0	0.7
example3	0	0	0.1	0.18	0.62	0.1	0	0.72
example4	0	0	0.1	0.16	0.64	0.1	0	0.74
example5	0	0	0.1	0.15	0.65	0.1	0	0.75

적인 전략을 선택하는 것이 유리한 반면, 선공일 때는 공격적인 전략을 선택하는 것이 더 유리하다는 결과가 나왔다. 이 모델을 활용하여 경기 진행 중 현재 상황에 해당하는 state부터 항상 최적의 전략을 따라 경기 끝까지 진행할 수 있다.

5.1 Sensitivity Analysis

전략에 따른 확률 분포가 제 3.3절에서 가정한 예시일 때 optimalaction이 <Table 9>, <Table 10>과 같이 나온다는 것을 확인하였다. 확률 분포가 제 3.3절과 달라질 때 optimal action에는 어떤 변화가 있는지 sensitivity analysis를 추가로 진행해보았다.

수비적인 전략을 선택 했을 시 확률 분포를 변화시켰을 때 optimal action은 어떤 차이를 보이는지 example 5개로 비교해보았다. Example 1에서 example 5로 갈수록 0점이 발생할 확률은 줄고, 1점을 얻을 확률은 증가한다. 즉, example 5이 example 1보다 후공 일 때 수비적인 전략을 선택하면 확실하게 1점을 얻을 확률이 증가한다.

각각의 확률 분포의optimal action을 구하여 결과를 비교해보았을 때, 선공일 때 동점일 경우를 제외하고는 전부 다 동일했다. <Table 12>는 선공일 때 동점인 경우, 각각 example의 결과를 비교한 표이다.

Example 3까지는 선공일 때 동점이면 엔드에 상관없이 공격적인 전략을 선택하였다. 그러나, example 4부터 경기 초반 홀수 엔드일 경우 수비적인 전략이 optimal action이 되고, example 5에선 수비적인 전략이 optimal action인 엔드가 더 늘어났다. 즉, 선공일 때 수비적인 전략이 확실하게 1점을 잃을 확률이 증가할수록 수비적인 전략이 우세한 영역이 점점 늘어나는 것을 볼 수 있다.

선공일 때 수비적인 전략을 선택하면 높은 확률로 1점을 잃지만 동시에 다음 엔드의 후공의 기회도 가져온다. 후공 일 때 수비적인 전략의 1점을 낼 확률이 증가하고 0점을 낼 확률이 감소하면서 공격적인 전략은 상대적으로 1점을 얻을 확률이 줄어들고 0점을 얻을 확률이 올라간다. 그러므로 후공 일 때 공격적인 전략을 선택했을 때 후공의 기회를 상대방에 넘겨주지 않거나 큰 점수를 얻게 될 확률이 높아지므로 경기에서 더 유리해진다.

따라서, 확률 분포에 따라 선공일 때 동점이면 1점을 내주고 다음 엔드에서 공격적인 전략을 선택하였을 때 주도권을 가지는 것의 우세함 정도가 달라지기 때문에 <Table 12>의 결과와 같이 수비적인 전략을 선택하는 영역이 달라지는 결과를 볼 수 있다.

Table 12. Optimal Action when d = 0 without Hammer

	$\Pi_1(S_1)$	$\Pi_2(S_2)$	$\Pi_3(S_3)$	$\Pi_4(S_4)$	$\Pi_5(S_5)$	$\Pi_6(S_6)$	$\Pi_7(S_7)$	$\Pi_8(S_8)$	$\Pi_9(S_9)$	$\Pi_{10}(S_{10})$
example 1	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
example 2	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
example 3	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG	AGG
example 4	CON	AGG	CON	AGG	CON	AGG	AGG	AGG	AGG	AGG
example 5	CON	CON	CON	AGG	CON	AGG	CON	AGG	AGG	AGG

Table 13. Optimal Action of Each States with Hammer (100 Actions)

	$\Pi_1(S_1)$	$\Pi_2(S_2)$	$\Pi_3(S_3)$	$\Pi_4(S_4)$	$\Pi_5(S_5)$	$\Pi_6(S_6)$	$\Pi_7(S_7)$	$\Pi_8(S_8)$	$\Pi_9(S_9)$	$\Pi_{10}(S_{10})$
d = -2		100	100	100	100	100	100	100	100	100
d = -1		100	100	100	100	100	100	100	100	100
d = 0	0	0	0	0	0	0	0	0	0	0
d = 1		0	0	0	0	0	0	0	0	0
d = 2		0	0	0	0	0	0	0	0	0

6. 중간 단계 전략

모델이 수비적인 전략과 공격적인 전략만 action으로 갖는 것이 아니라 그 사이에 중간 단계 전략들도 갖도록 하였다. 그리고 중간 단계 전략들의 transition probability의 경우, 가장 공격적인 전략과 가장 수비적인 전략의 선형 조합의 형태로 나타내었다. N = 100일 때, 컬링 MDP 모델을 풀어보았다.

$$Action\ Space = \{a_0, a_1, \cdots, a_n, \cdots, a_N\},$$
$$a_0 : The\ most\ Conservation,$$
$$a_N : The\ most\ Aggressvie$$
$$P(S_{t+1}|S_t, a_n) = (1-\lambda_n) \cdot P(S_{t+1}|S_t, a_0) + \lambda_n \cdot P(S_{t+1}|S_t, a_N),\ n = 0, \cdots, N,\ \lambda_n = \frac{n}{N}$$

<Table 13>에 후공인 경우 결과만 결과를 나타내었는데 최적의 중간 단계 전략이 존재할 것이라는 직관과 다르게 가장 공격적인 전략과 가장 수비적인 전략만 최적의 전략으로 선택된 것을 볼 수 있다(선공일 경우에도 마찬가지였다). 중간 단계 전략 연구를 통해 컬링 모델은 공격적인 전략과 수비적인 전략 두 가지만 고려하면 됨을 알 수 있었다. 본 연구에서 세운 모델로 수비적인 전략과 공격적인 전략만 비교하는 것에서 그치지 않고 중간 단계 전략을 비교하는 것처럼 모델의 변형을 통해 다른 전략 연구에도 활용할 수 있을 것으로 기대 한다.

7. 결 론

본 연구에선 컬링의 여러 전략 중 가장 일반적인 전략인 공격적인 전략과 수비적인 전략을 비교할 수 있는 모델을 세웠다. 이 컬링 전략 분석 모델은 기존 컬링 연구에 활용되던 모델이 아닌, 전략 선택을 포함하는 MDP를 활용하여 개발하였고 또한 이 모델을 풂으로써 최적의 전략을 분석할 수 있음을 보였다.

Sensitivity analysis를 통해 확률 분포가 달라짐에 따라 최적의 전략이 달라지는 경우가 발생하는 것 또한 확인하였다. 따라서 공격적인 전략과 수비적인 전략일 때 실제 점수 분포 데이터를 수집하여 본 연구에서 세운 모델을 푼다면 실제 컬링 경기에서 상황마다 최적의 전략을 찾을 수 있을 것이다. 더 나아가 모델의 변형하여 컬링에서 두 전략을 제외한 중간 단계 전략들은 최적의 전략이 될 수 없고 두 전략만이 최적의 전략이 될 수 있음을 보였다. 이로부터 모델의 변형을 통해 공격적/수비적인 전략뿐만 아니라 다른 다양한 전략 분석에도 활용할 수 있을 것으로 기대한다. 또한 본 연구에서 개발한 컬링 전략 분석 모형은 상대방의 전략 선택은 포함하지 않았지만 상대방의 전략 선택도 포함하는 모델을 개발한다면 추후 좋은 연구 주제가 될 것으로 생각한다. 이번 연구를 통해 새로운 컬링 전략 분석 모델을 제시함으로써 컬링 전략 연구의 기초가 될 것으로 기대한다.

참고문헌

Bukiet, B., Harold, E. R., and Palacios, J. L. (1997), A Markov chain approach to baseball, *Operations Research*, **45**(1), 14-23.

CCA (2014), Rules of Curling for General Play, *Canadian Curling Association.*

Curling Canada (Online), 4 rock free guard zone strategy (available from URL http://www.curling.ca/wp-content/uploads/2010/01/4rockstrat.pdf).

Curltech (Online), Curling Manual (available from URL http://www.curlingschool.com/manual2007/Section7.html).

Kostuk, K. J., Willoughby, K. A., and Saedt, A. P. (2001), Modelling curling as a Markov process, *European Journal of Operational Research*, **133**(3), 557-565.

Powell, W. B. (2007), Approximate Dynamic Programming : Solving the curses of dimensionality, New York : Wiley, 2007.

Wright, M. B. (2009), 50 years of OR in sport, *Journal of the Operational Research Society*, S161-S168.