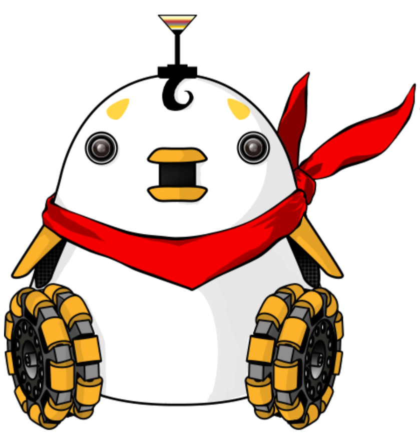
人工知能
第4章 探索（３）：
ゲームの理論

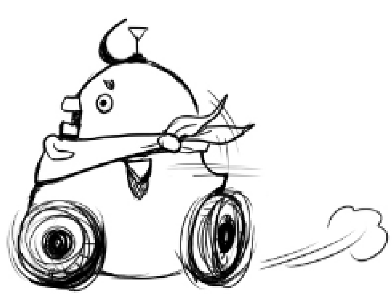


STORY　ゲームの理論

- ホイールダック２号は一つ誤解をしていた．迷路の中ではとにかくまっすぐゴールに向かえばよいわけではない．迷路にはホイールダック２号を邪魔しようとする敵がいる．これとぶつかると何かと面倒である．敵はホイールダック２号にぶつかれば報酬を得ることができ，罠ではダメージを受ける．敵は敵でがんばる中で，ホイールダック２号はその行動を先読みしながら迷路を抜けなければならない．
- 相手の行動が自らの行動の結果に影響を与え，自らの行動が相手の行動の結果に影響を与える．ホイールダック2号は今まさにゲーム的状況の中にあったのだ．

仮定　ゲームの理論

- ホイールダック２号は自分と敵の行動に対する利得を知っている.
- ホイールダック２号は自らの行動に対する結果を確実に予測できるものとする.
- 敵は合理的に行動するものとする.
- ホイールダック２号も敵も物理的につながっている場所・状態には意図すれば確定的に移動することができるものとする.

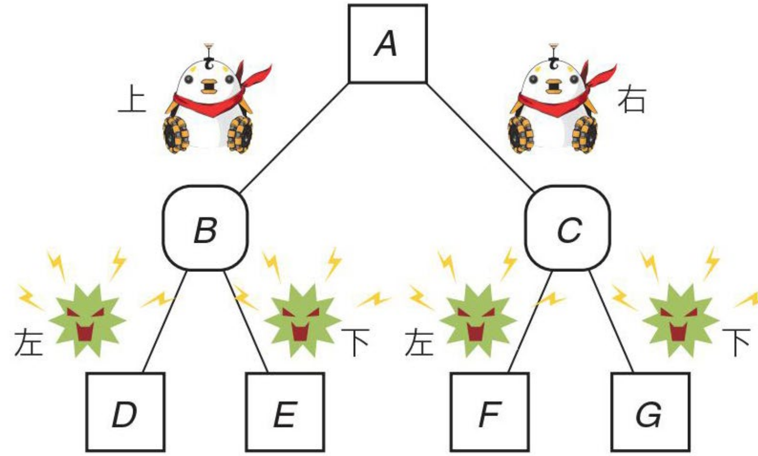


Contents

4.1.1 はじめに

- ホイールダック２号はどうすべきか？

図 4.2　敵をかわしてゴールに向かうタスク



ホイールダック2号	$C_{\text{ホ}}$	0	$D_{\text{ホ}}$	$D_{\text{ホ}}+C_{\text{ホ}}$
敵	$C_{\text{敵}}$	$D_{\text{敵}}$	0	$D_{\text{敵}}+C_{\text{敵}}$

図 4.3　図 4.2 のゲームの展開

表 4.1　図 4.2 のゲームの利得行列

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	$C_{\text{ホ}},\ C_{\text{敵}}$	$0,\ D_{\text{敵}}$
ホイールダック 2 号が右へ	$D_{\text{ホ}},\ 0$	$D_{\text{ホ}}+C_{\text{ホ}},\ D_{\text{敵}}+C_{\text{敵}}$

利得行列

- プレイヤーが二人の場合には，各プレイヤーの行動を行列の行と列に書き，それぞれの交わるセルに，それぞれのプレイヤーが得る利得を書いたものを**利得行列**と呼ぶ.
- 一般の行列とは異なり，**双行列**(bimatrix)と呼ばれる.

プレイヤー２
の行動

表 4.1　図 4.2 のゲームの利得行列

プレイヤー１
の行動

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	$C_{\text{ホ}}$, $C_{\text{敵}}$	0, $D_{\text{敵}}$
ホイールダック 2 号が右へ	$D_{\text{ホ}}$, 0	$D_{\text{ホ}} + C_{\text{ホ}}$, $D_{\text{敵}} + C_{\text{敵}}$

左がプレイヤー１，右がプレイヤー２の利

4.1.2 ケース1：
敵はホイールダック2 号を捕まえたい

- ホイールダック2 号にとっては上に移動すると敵にぶつかってしまうが，まだ利得が–5 で済むため，<u>上への移動を選ぶのが最適な選択</u>となる.

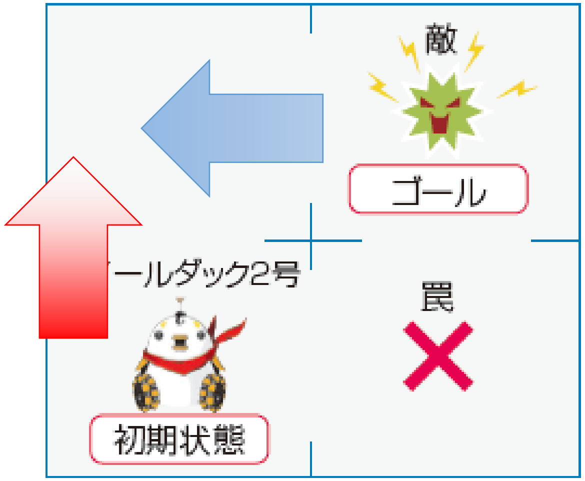


図 4.2　敵をかわしてゴールに向かうタスク

表 4.2　ケース 1 ： $C_{ホ} = -5, C_{敵} = 3, D_{ホ} = D_{敵} = -2$ の場合の図 4.2 のゲームの利得行列

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	$-5,\ 3$	$0,\ -2$
ホイールダック 2 号が右へ	$-2,\ 0$	$-7,\ 1$

4.1.3 ケース2：少しだけ敵のモチベーションが下がったら？

- ホイールダック2 号は右に行くことで×印で–3 のダメージを受けるだけで，ゴールにたどり着けるのである．
- 少しの利得の違いでとるべき行動が変化する．

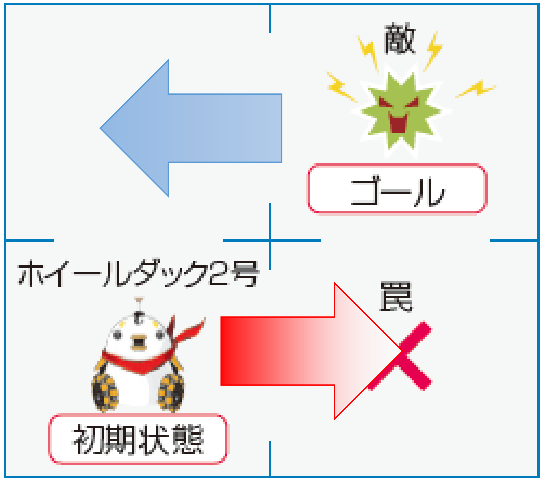


図 4.2 敵をかわしてゴールに向かうタスク

表 4.3 ケース 2 ： $C_{\text{ホ}} = -5, C_{\text{敵}} = 2, D_{\text{ホ}} = D_{\text{敵}} = -3$ の場合の図 4.2 のゲームの利得行列

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	-5, 2	0, -3
ホイールダック 2 号が右へ	-3, 0	-8, -1

主体の意思決定が混ざり合って状況が決定する系を<u>**ゲーム**</u>と呼ぶ

Contents

- ◻4.1 利得と回避行動
- ◻4.2 標準型ゲーム
- ◻4.3 展開型ゲーム
- ◻4.4 ゲームAI の実践的開発に向けて*

4.2.1 はじめに/ 標準型ゲーム

■ゲーム理論

- 複数のプレイヤの意志決定を扱う理論
- 1944年 フォン・ノイマン，モルゲンシュテルン「ゲームの理論と経済行動」

■基本的な用語の定義

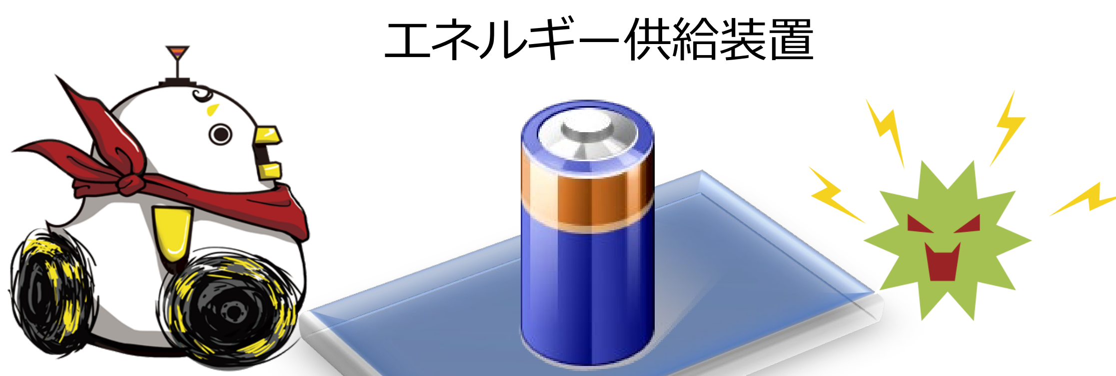
- **プレイヤ**　意志決定を行なう個々の主体．複数存在する．
- **行動**　プレイヤの選択．戦略と呼ぶこともある．探索の作用素に相当．
- **利得**　プレイヤの行動の組み合わせに対して定義される数値．結果に対する各プレイヤの効用を示す．大きいほうがより嬉しいとする．
- **合理的**　各プレイヤは自分の利益を最大化しようと最大限の努力をする．利己的(selfish) と呼ぶこともあるが，ニュアンス的に誤解がある．英語ではrational.
- **均衡**　合理的な意思決定の結果として，自ずと決まる全プレイヤーの行動の落ち着く先.

4.2.4 支配戦略均衡

- 相手の行動が何であろうが，その行動をとった方が高い利得を得られる行動を**支配戦略** という．
- ゲームの状態は**支配戦略均衡**に至る．

表 4.4　支配戦略均衡を持つ利得行列

	敵は エネルギー供給装置へ	敵は 普通に休む
ホイールダック 2 号は エネルギー供給装置へ	4，4	8，1
ホイールダック 2 号は 普通に休む	1，8	1，1



4.2.5 ナッシュ均衡

- **ナッシュ均衡**：行動の組(ホイールダック2 号の行動, 敵の行動) が互いに相手の行動に対する最適な反応となっている

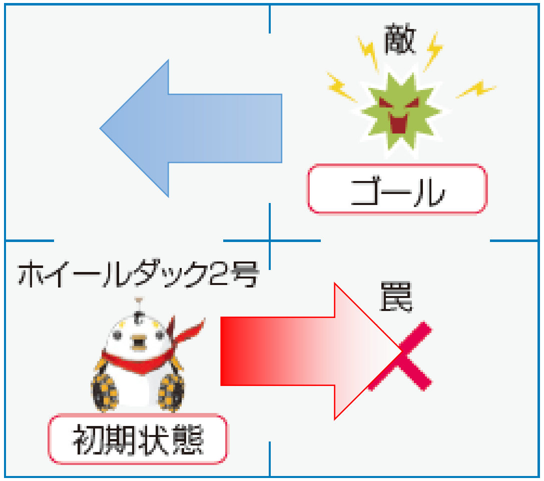


図 4.2　敵をかわしてゴールに向かうタスク

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	−5，2	0，−3
ホイールダック 2 号が右へ	−3，0	−8，−1

4.2.6 囚人のジレンマ

- ナッシュ均衡は必ずしも全体として良い状態に至るわけではない.

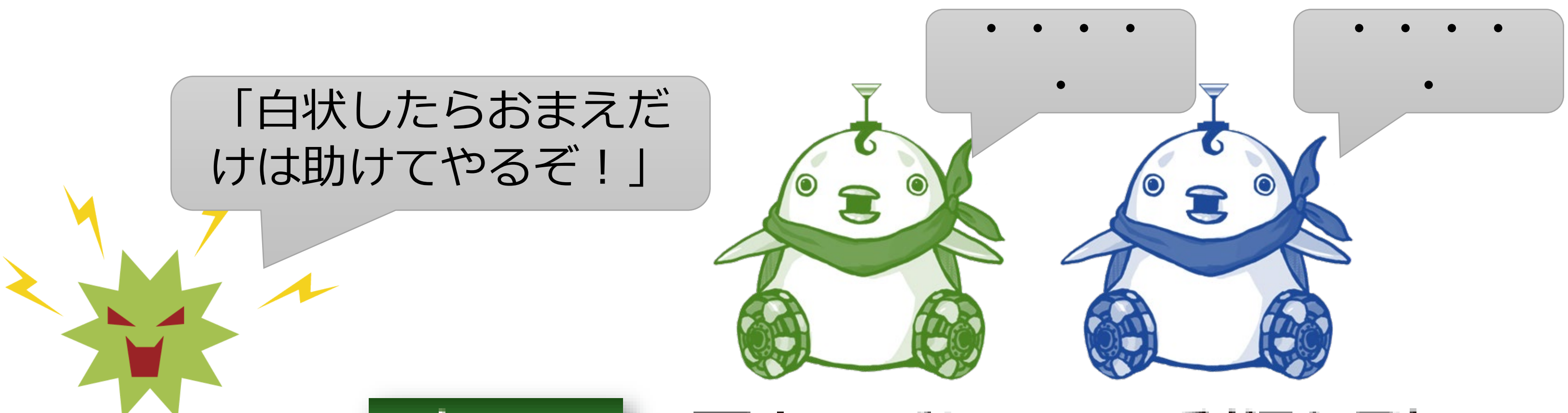
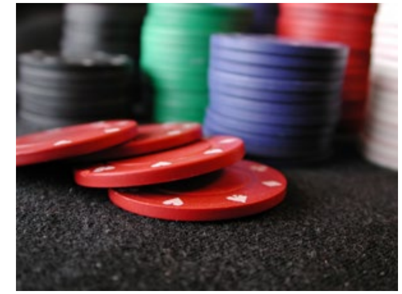


表4.5 囚人のジレンマの利得行列

囚人1＼囚人2	自白	黙秘
自白	−3,　−3	5,　−5
黙秘	−5,　5	3,　3

4.2.7 ゼロサム・ゲーム



- ゼロサム・ゲームはプレイヤーの利得の総和が0 になるゲームであり，特にプレイヤーが2 名の場合はプレイヤー1 の利得をr とすると，プレイヤー2 の利得は$-r$ となることになる．

表 4.6　賭けジャンケンの利得行列：ゼロサム・ゲームの例

	プレイヤー 2：グー	プレイヤー 2：チョキ	プレイヤー 2：パー
プレイヤー 1：グー	0	100	−100
プレイヤー 1：チョキ	−100	0	100
プレイヤー 1：パー	100	−100	0

双行列で書く必要がない

4.2.8 ミニマックス戦略

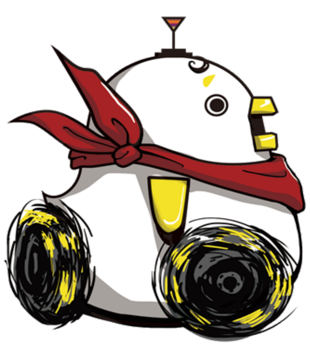
- 相手がプレイヤーの利得を最小化してくることを前提としながら，自らの利得を最大化する戦略
- マックスミニ戦略とミニマックス戦略が同じ利得を実現すればナッシュ均衡となる．

表 4.7 ゼロサム・ゲームでの利得行列とミニマックス戦略

ホイールダック２号 \ 敵	b_1	b_2	b_3	b_4
a_1	7	2	5	1
a_2	2	2	3	4
a_3	5	3	4	4
a_4	5	2	1	6

演習問題4-1 ナッシュ均衡

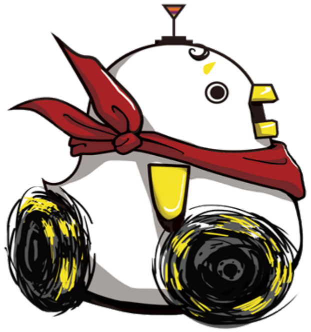
1. 以下の利得行列に対してホイールダック２号がミニマックス戦略を取った歳に至る均衡を求めよ.
2. 1.で求めた戦略で至った状態がナッシュ均衡であることを確認せよ.
3. 支配戦略戦略均衡ではないことを示せ。



ホイールダック２号＼敵	B1	B2	B3
A1	1	2	1
A2	2	3	3
A3	1	8	4

演習問題4-1 ナッシュ均衡 解答

1. 以下の利得行列に対してホイールダック２号がミニマックス戦略を取った歳に至る均衡を求めよ.
2. 1.で求めた戦略で至った状態がナッシュ均衡であることを確認せよ.
3. 支配戦略戦略均衡ではないことを示せ。

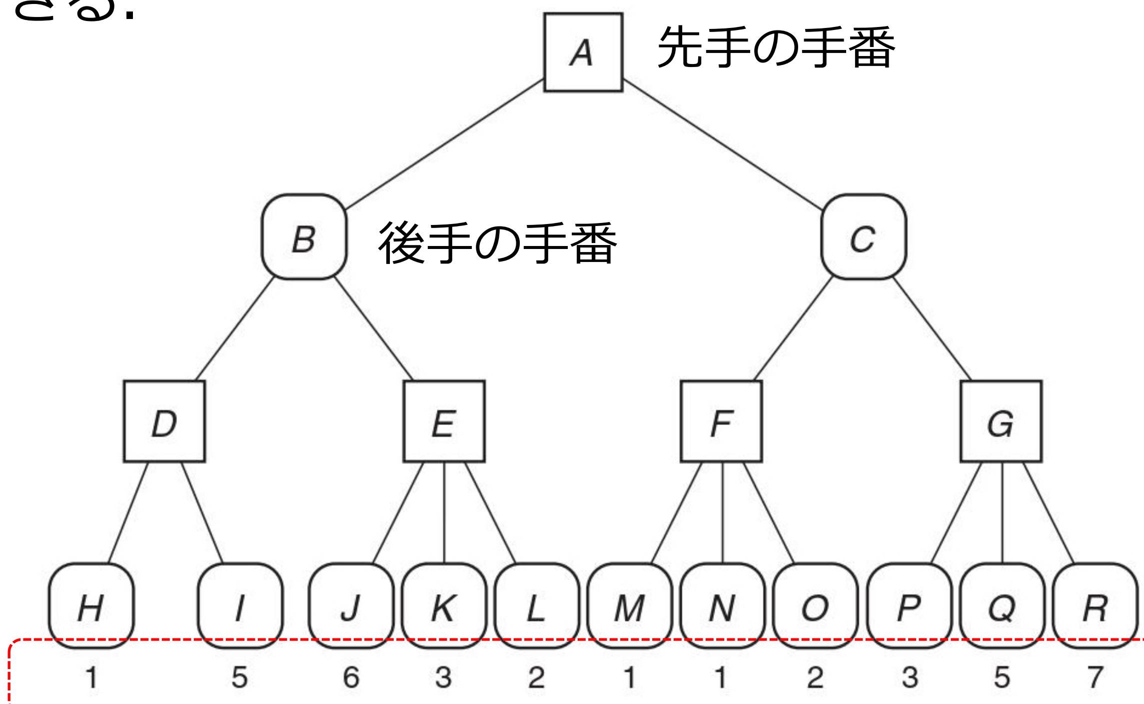


ホイールダック２号＼敵	B1	B2	B3
A1	1	2	1
A2	2	3	5
A3	1	8	4

Contents

4.3.1 展開型ゲーム

- 実際のゲームは一度きりの意思決定ではなく，多段階の意思決定を含む．このようなゲームを展開型ゲームという．
- オセロやチェスなど多くのゲームは展開型ゲームでモデル化できる．
- **ゲーム木**で表現できる．

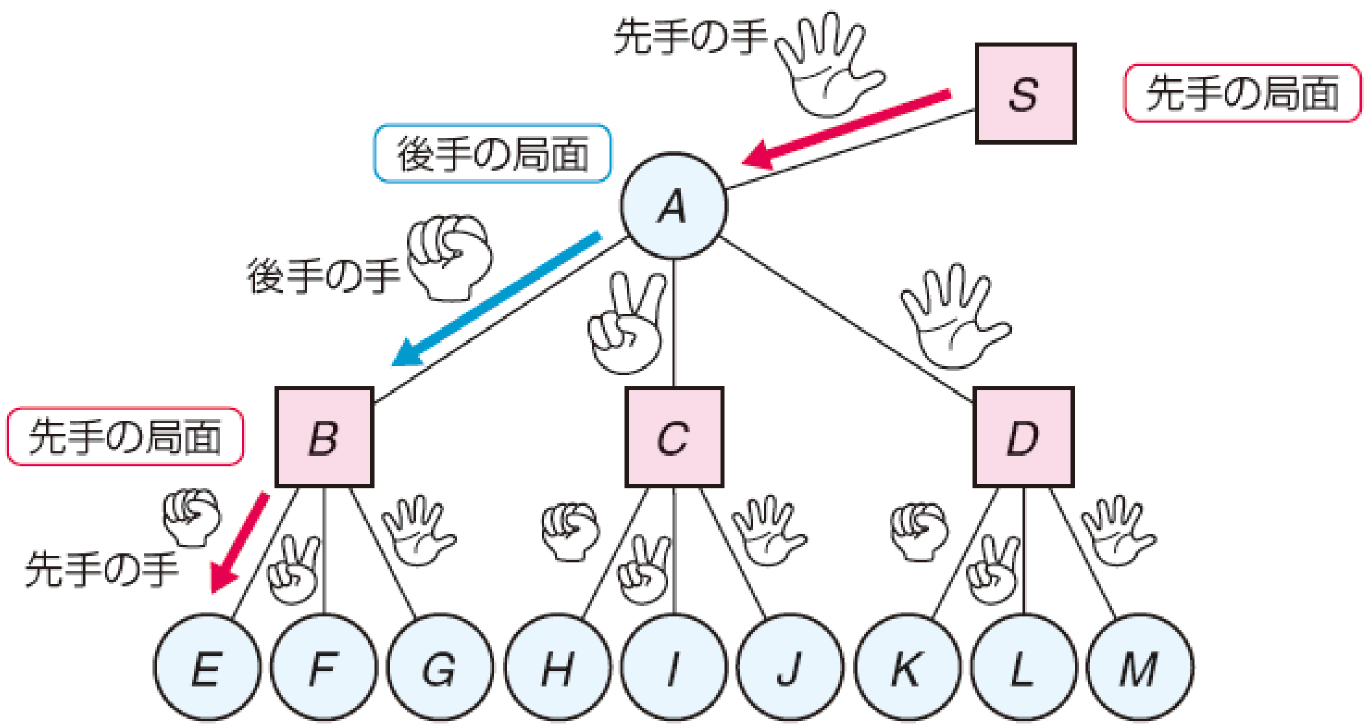


最終的に先手が得る利得

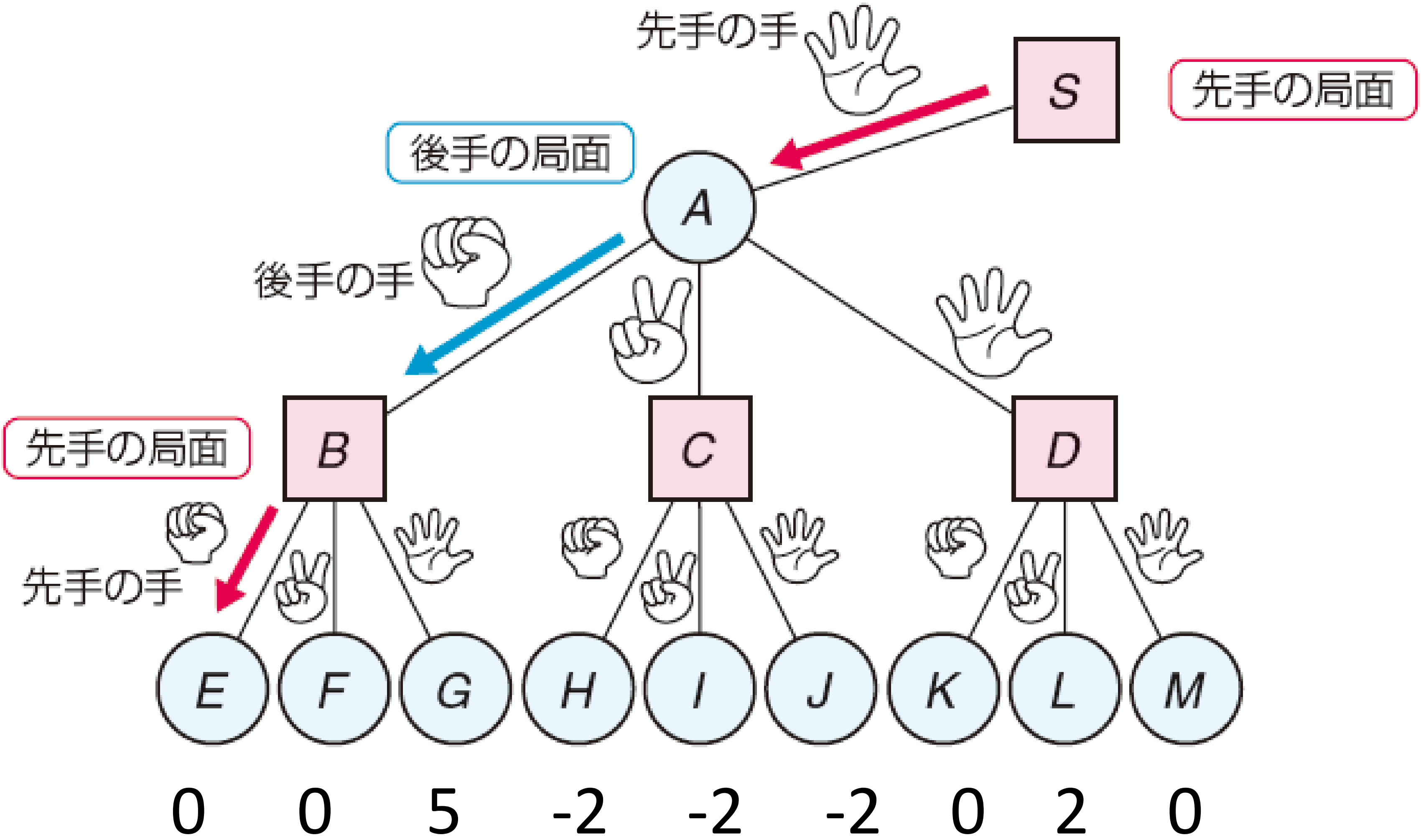
図 4.4 ゲーム木の例

演習問題4-2 交互ジャンケン

1. 順番にジャンケンを出すゲームをする．相手に勝つ手を出すと，自分にその指の本数分加点される．（負けた場合，あいこの場合得点の変化はない．）
自分がまず初期状態からパーを出した状態からスタートし，**初期状態->相手->自分** の一往復で終了する際のゲーム木は以下のようになる．ゲーム木の葉ノードに評価値を記入せよ．ただし評価値は**評価値 = 自分の得点 – 相手の得点**
とする．



演習問題4-2 交互ジャンケン 解答



4.3.3 ミニマックス法

- 先手は後手が「最も先手が低い利得になる」ような手をとるであろうことを前提にして，自らが高い利得を得られる行動をとる．

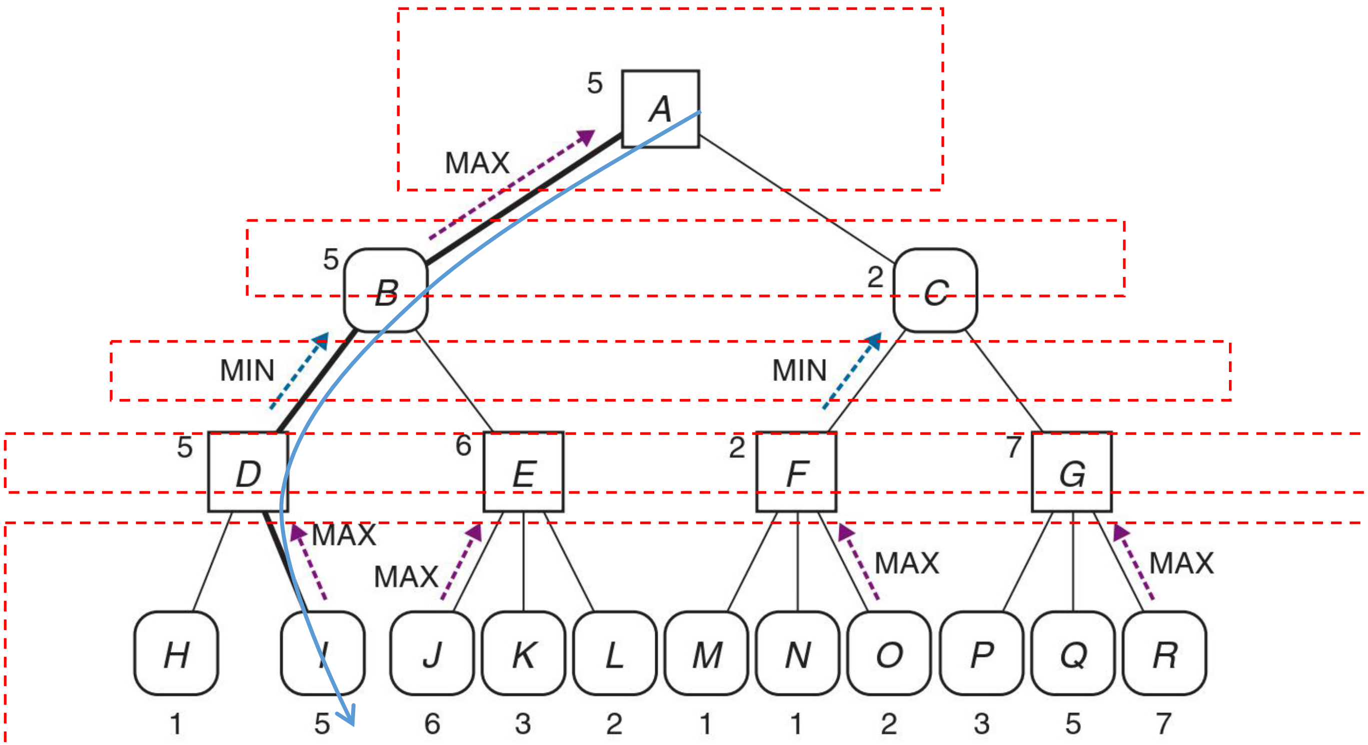
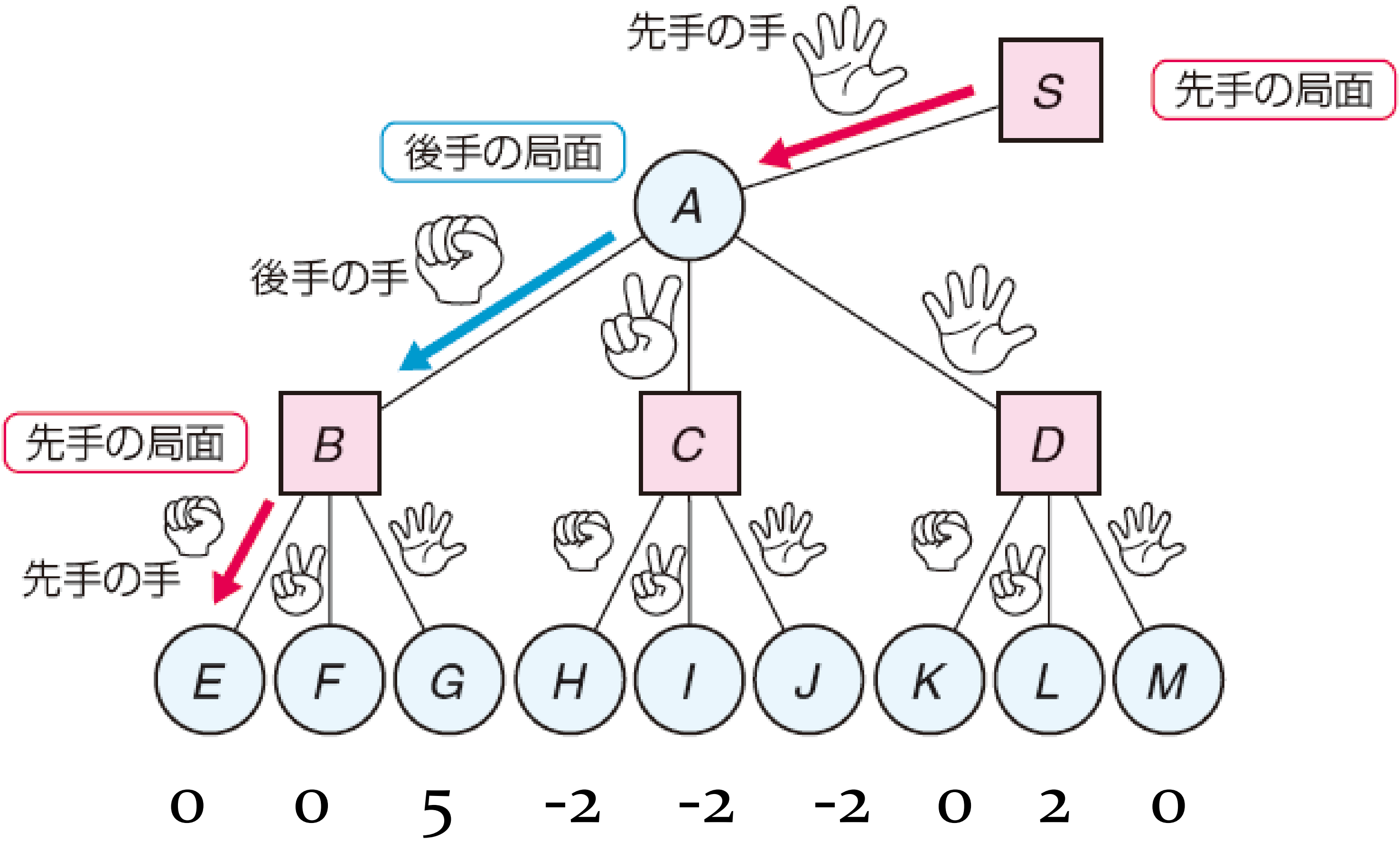


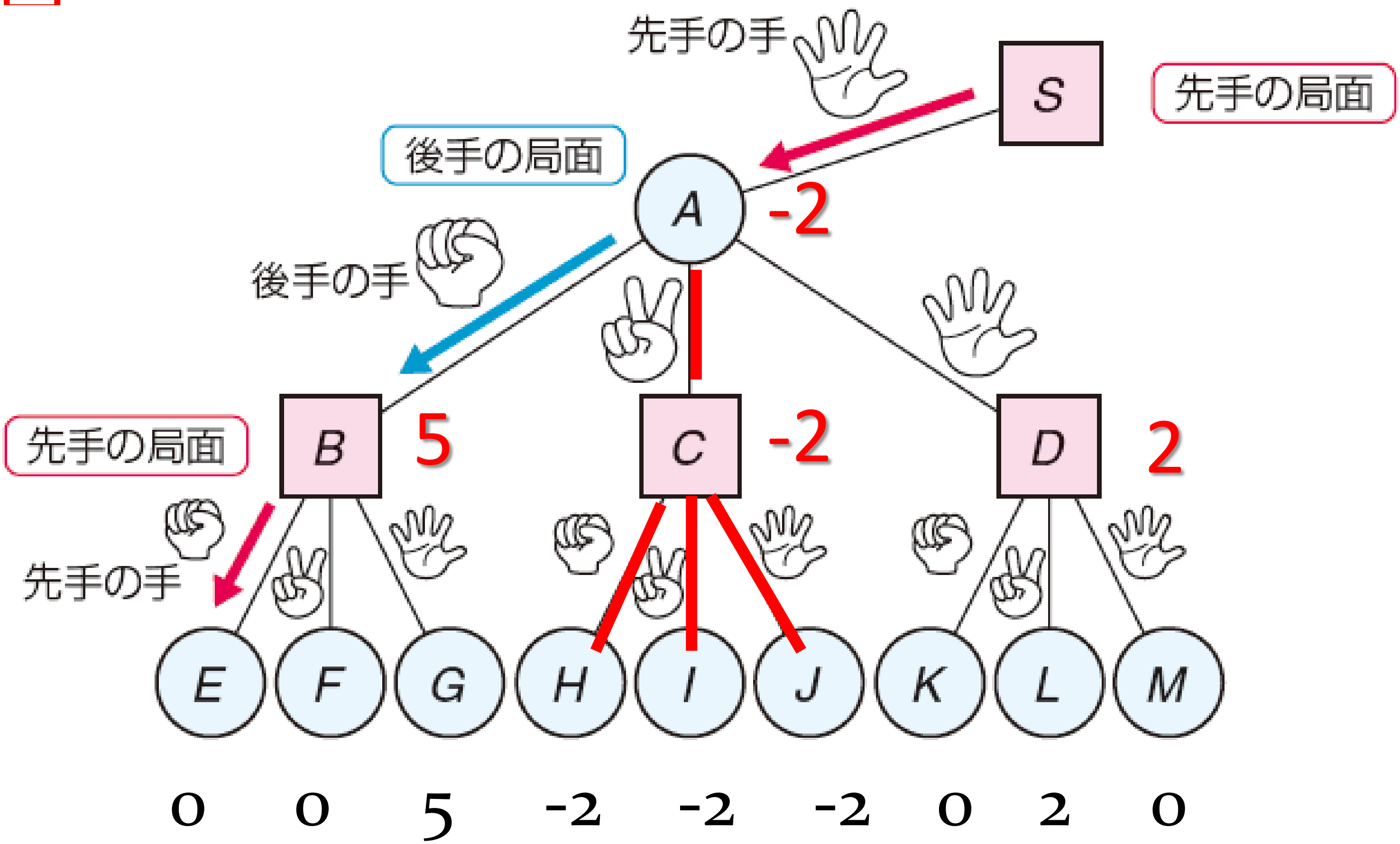
図 4.5 ミニマックス法の解探索例

演習問題4-2 交互ジャンケン min-max法



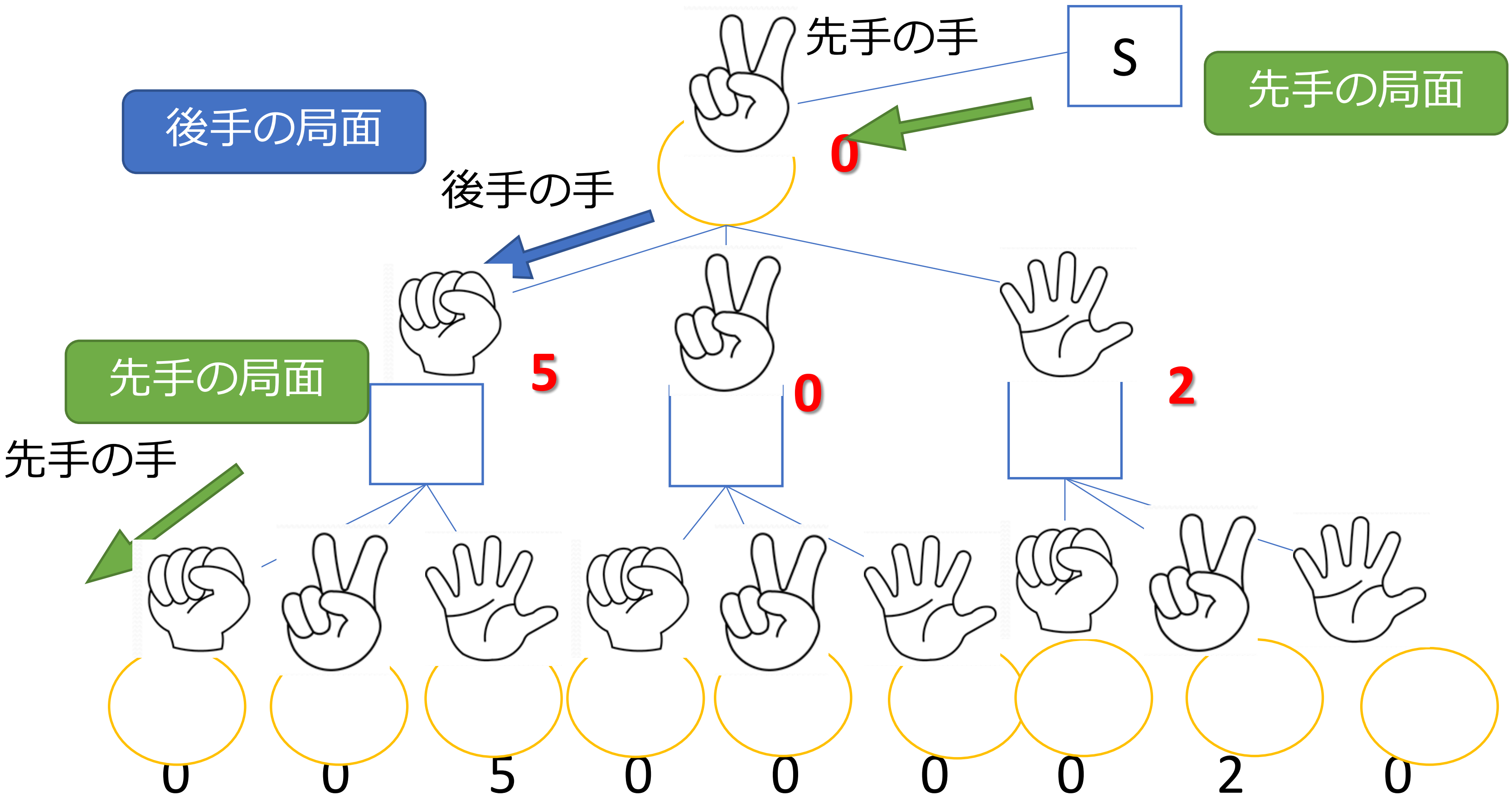
1. 4-1で考えた，交互ジャンケンについてmin-max法を適用し，各ノードの評価値を得よ
2. このゲームに先手は勝つことができるか？
3. このゲームにおける最後の先手の最良の手を述べよ．
4. もし，最初の一手を先手が選ぶことが出来れば先手は勝つことが出来るか

演習問題4-2 交互ジャンケン min-max法
解答



1. 4-1で考えた，交互ジャンケンについてmin-max法を適用し，各ノードの評価値を得よ　**上記の通り**
2. このゲームに先手は勝つことができるか？　**無理**
3. このゲームにおける最後の先手の最良の手を述べよ．　**何出しても同じ**

演習問題4-2 交互ジャンケン min-max法
解答



4. もし，最初の一手を先手が選ぶことが出来れば先手は勝つことが出来るか？

無理．ただし，同点には持っていける．

チョキと，グーの場合のゲーム木を考えればいい，

4.3.4 アルフファ・ベータ法

- ミニマックス法では盤面の局面を先読みすればするほど，良い手を選択し，ゲームを有利に進めていける．
- しかし，探索しすぎるとゲーム木の探索空間が膨大になる．
- ミニマックス法の性質を生かして，不必要な探索を避ける（サボる）ことができる．

pruning = 枝刈り

- アルファ・ベータ法（αβ pruning）
  - βカット(β pruning)　評価値最小化局面の枝刈り
    - 後手がわざわざ評価値の大きな手を打たないことを利用して，**先手の行動（作用）をカット**
  - αカット(α pruning)　評価値最大化局面の枝刈り
    - 先手がわざわざ評価値の小さな手を打たないことを利用して，**後手の行動（作用）をカット**

βカット (β pruning)

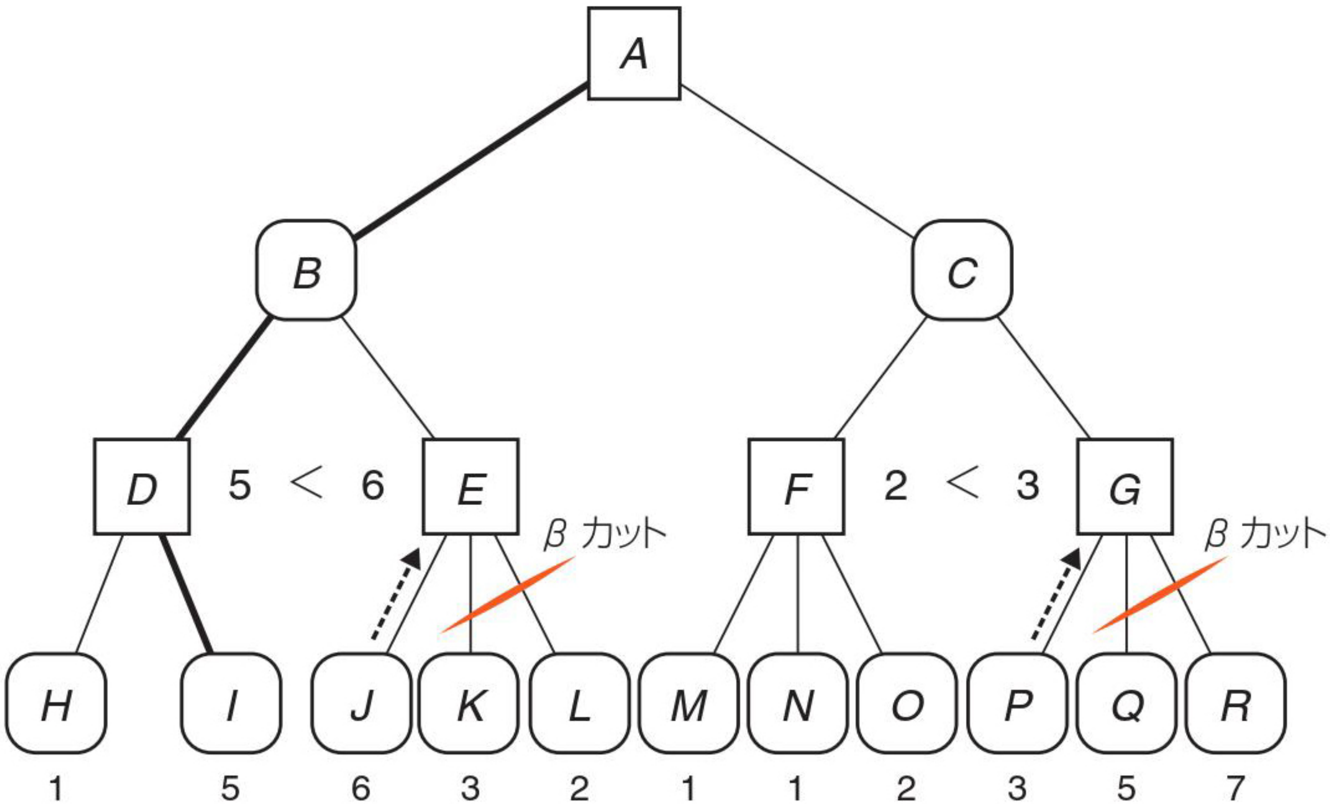


図 4.6 β カットの例

αカット (α pruning)

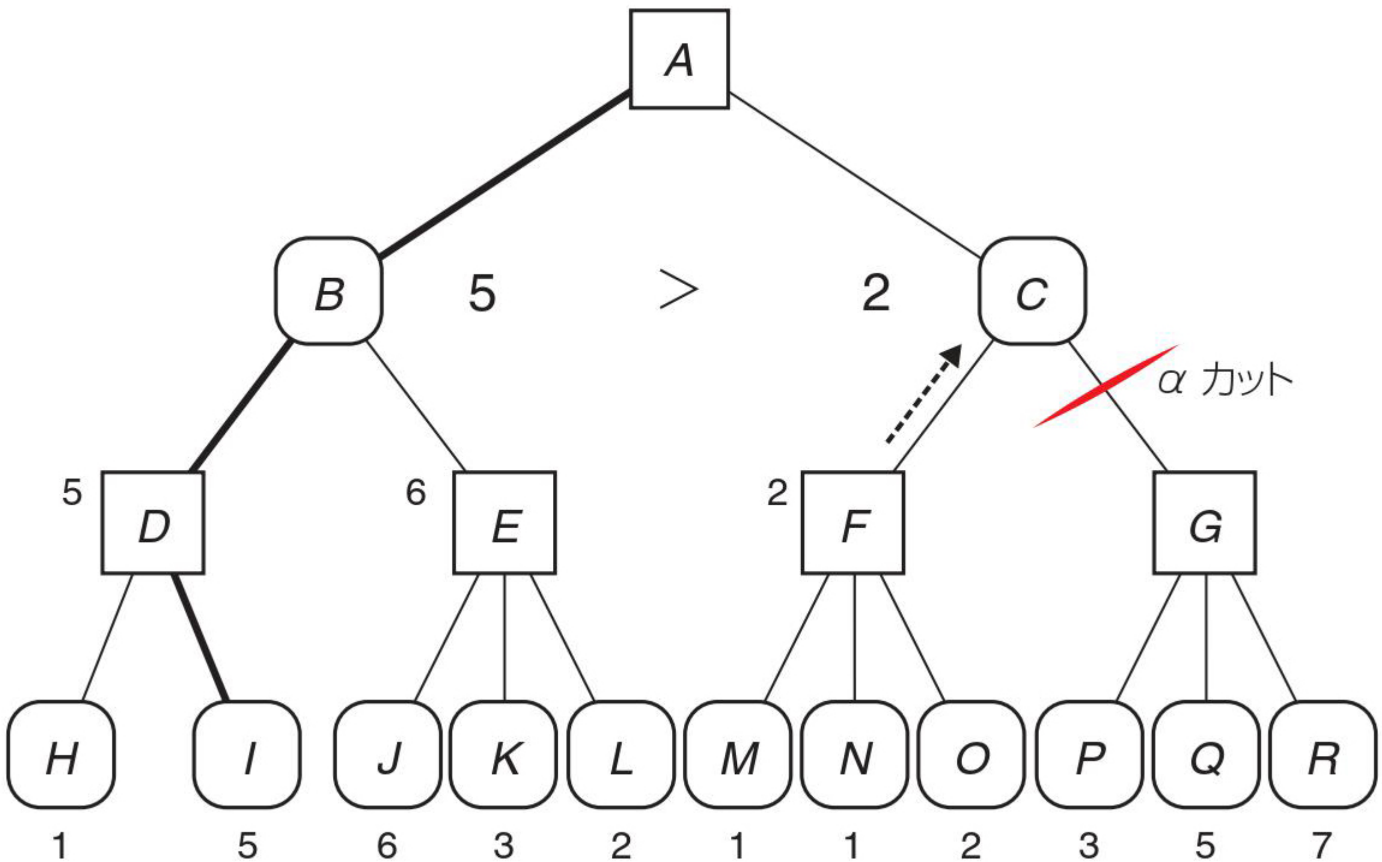
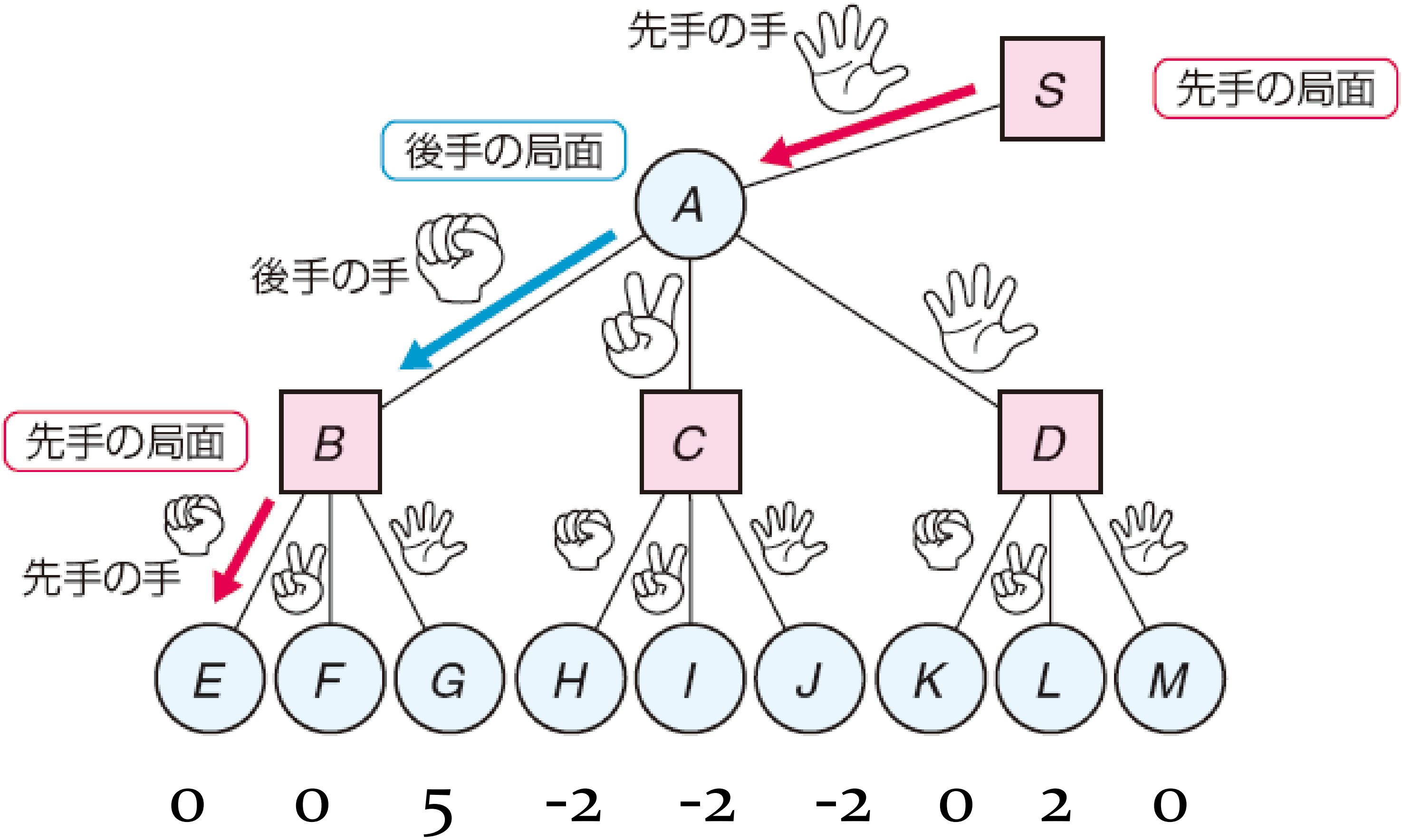


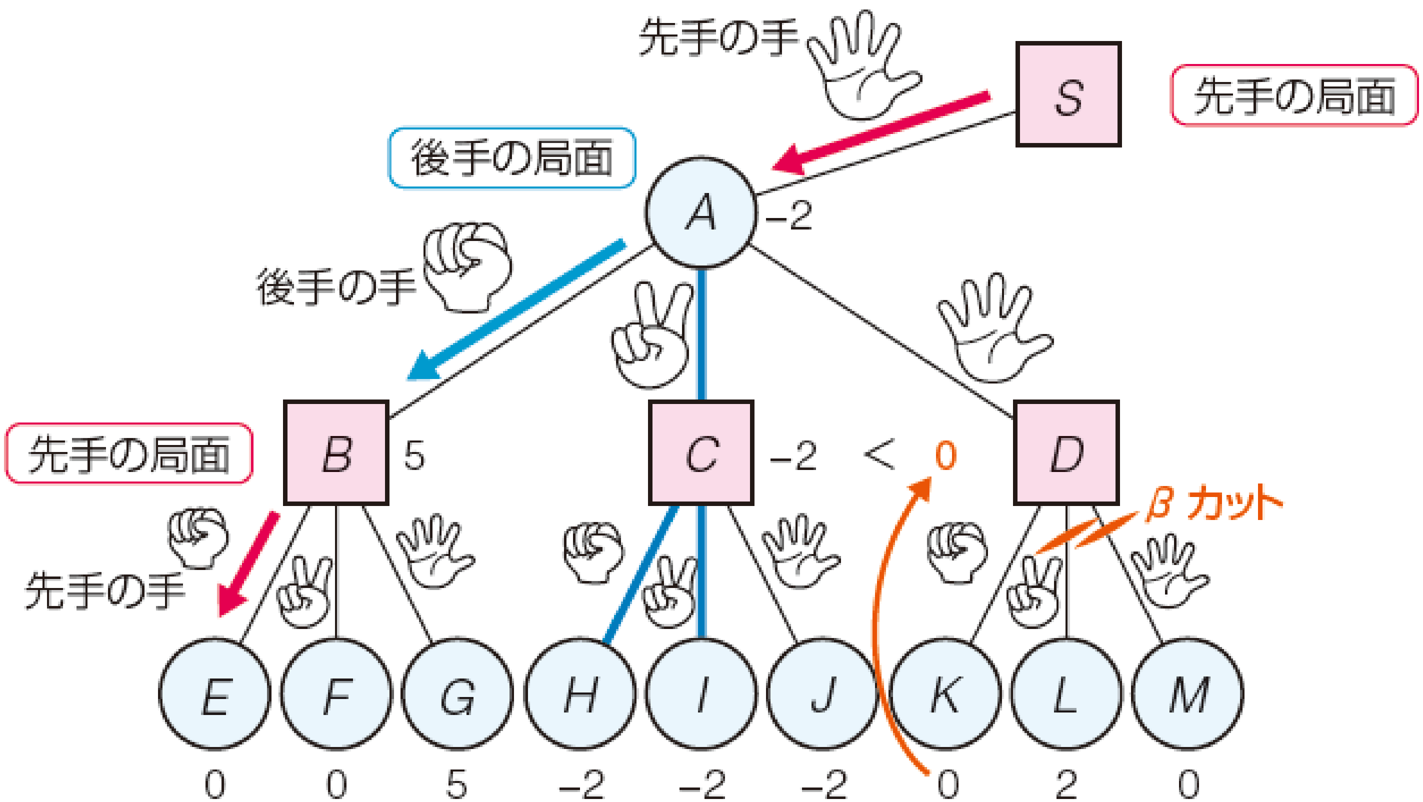
図 4.7 α カットの例

演習問題4-3 交互ジャンケン αβカット



1. 4-2 で考えた，交互ジャンケンについてαカット，もしくはβカットをする所はあるか？あるとすれば何処で生じるか？答えよ.

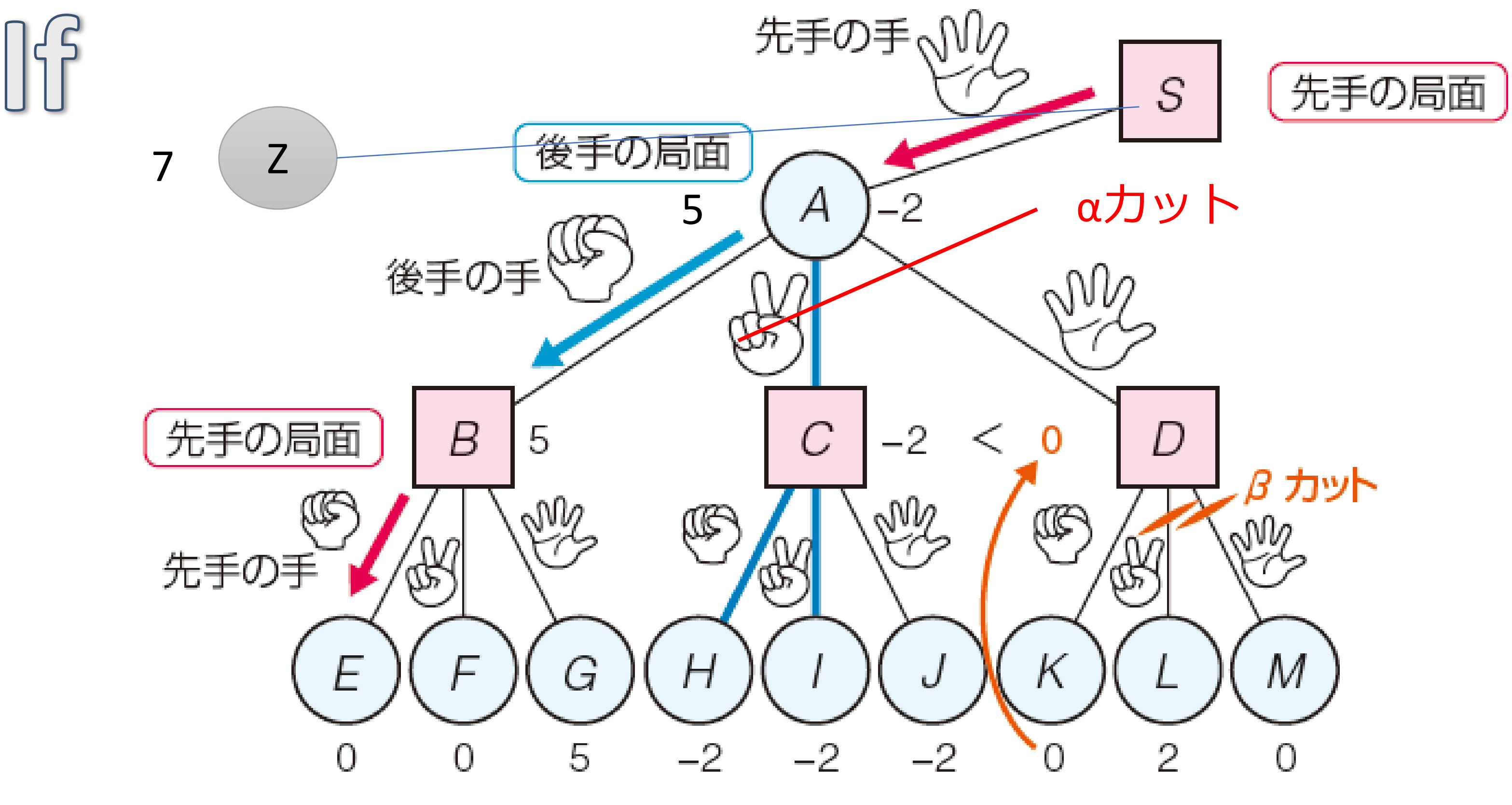
演習問題4-3 交互ジャンケン αβカット 解答



アルファベータ法を適用した交互ジャンケン

1. 4-2 で考えた，交互ジャンケンについてαカット，もしくはβカットをする所はあるか？あるとすれば何処で生じるか？答えよ．
   - αカット　ない．
   - βカット　ある，上記．

演習問題4-3 交互ジャンケン αβカット 解答



アルファベータ法を適用した交互ジャンケン

- たとえばもし，初めの分岐で左側にzというノードが存在し，その評価値が7と求まっていた場合を考えると，αカットが存在します．

Contents

現実のゲーム

- 初手黒が置き，その後，交互に置いていく事を考えると，オセロの状態空間の大きさ（葉ノードの数）はどれくらいになるか？概算を求めよ.
- 盤面は 8×8　である．小さめの近似として一回に置ける手は平均して5箇所程度であるとしてよい.

$\log_{10}(5^{64})$≒64×0.7≒45・・・10の45乗個の葉ノード
１葉ノードあたりのメモリを 1byte としても，
10^{42}KB=10^{39}MB=10^{36}GB=10^{33}TB
=10^{30}PB=10^{27}EB=10^{24}ZB=10^{21}YB……. 無理！
ちなみに東洋の命数法を使えば，１０載バイトです.
=>単純なミニマックス法などではなく，効率的な解法を作る必要がある.

ところで**Alpha Go** はどうやったか？

- 状態空間構成
  - Deep learning (**ニューラルネットワーク**)による 自動構成
  - 相互対戦の**強化学習**による漸次的構成
- 状態価値の推定
  - **Imitation learning（模倣学習）**による人間のプロの手に基づく状態評価の事前学習
- 最良解の効率的探索
  - **モンテカルロ木探索(MCTS)**による効率的探索
    ※**確率**的探索法

ゲーム木の考え方が根底にある.

第4章のまとめ

- プレイヤー，利得行列や合理的な行動といったゲーム理論における基本用語を学び，ゲーム理論の対象となるゲームとは何かを知った.
- 支配戦略均衡やナッシュ均衡といった標準型ゲームにおける均衡概念の基礎について学んだ.
- 展開型ゲームとそのゲーム木による表現について学んだ.
- 展開型ゲームがゼロサム・ゲームであった場合について効率的に解を求めるミニマックス法を学んだ.
- ミニマックス法において解の探索を効率化するα カットとβカットについて学んだ.