人工知能

第 3 章　探索 (3)：ゲームの理論

山田　敏規

2019 年 6 月 19 日

STORY

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．敵がどのように行動するのかを先読みしながら迷路を抜けなければならない．

仮定

- 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
- 敵は合理的に行動する
- 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

参考図書

- "ゲーム理論," 岡田　章著, 有斐閣, 1996

本日の内容

1. ホイールダック 2 号の迷路探索 (復習)
2. 利得と回避行動
3. 標準型ゲーム
4. 展開型ゲーム
   - ミニマックス法
   - アルファベータ法
5. 小レポート

辺のコストと予測評価値の設定

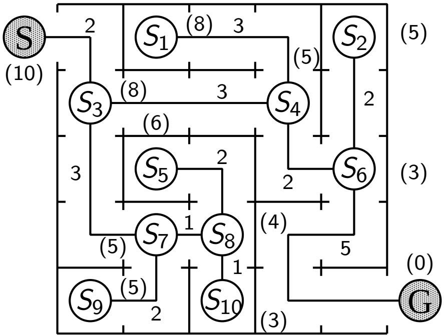


図 3.3 コスト付きのグラフ

- コスト ＝状態間の移動にかかるマス数
- 予測評価値 $\hat{h}(s) = G$ までのマンハッタン距離

注: マンハッタン距離
＝壁が存在しないときの状態間の移動にかかるマス数

A^* アルゴリズムでの探索

オープンリスト

クローズドリスト

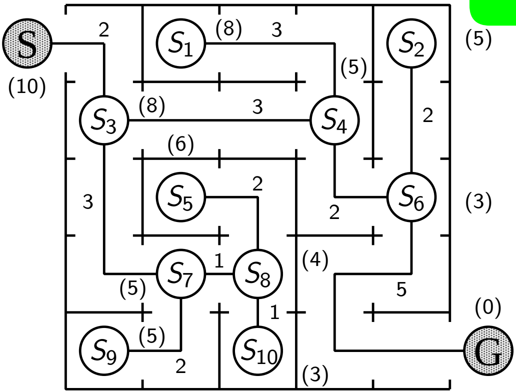
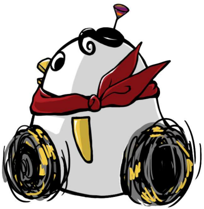
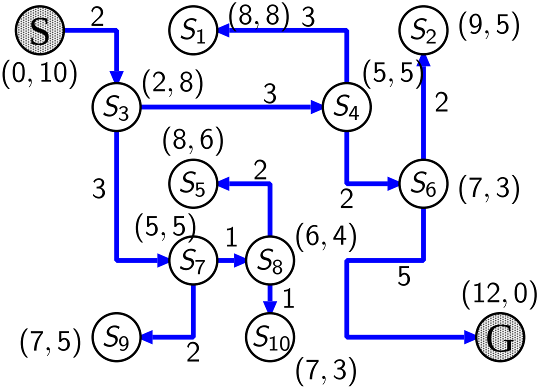


図 3.3 コスト付きのグラフ



A^* アルゴリズムで迷路を抜けてみよう！

A^* アルゴリズムでの探索



A^* アルゴリズムでの探索

ステップ	オープンリスト	クローズドリスト
1	$S(10)$	
2	$S_3(10)$	$S(10)$
3	$S_4(10), S_7(10)$	$S(10), S_3(10)$
4	$S_7(10), S_6(10), S_1(16)$	$S(10), S_3(10), S_4(10)$
5	$S_6(10), S_8(10), S_9(12), S_1(16)$	$S(10), S_3(10), S_4(10), S_7(10)$
6	$S_8(10), G(12), S_9(12), S_2(14)$ $S_1(16)$	$S(10), S_3(10), S_4(10), S_7(10)$ $S_6(10)$
7	$S_{10}(10), G(12), S_9(12), S_2(14)$ $S_5(14), S_1(16)$	$S(10), S_3(10), S_4(10), S_7(10)$ $S_6(10), S_8(10)$
8	$G(12), S_9(12), S_2(14), S_5(14)$ $S_1(16)$	$S(10), S_3(10), S_4(10), S_7(10)$ $S_6(10), S_8(10), S_{10}(10)$
9	$S_9(12), S_2(14), S_5(14), S_1(16)$	$S(10), S_3(10), S_4(10), S_7(10)$ $S_6(10), S_8(10), S_{10}(10), G(12)$

本日の内容

1. ホイールダック 2 号の迷路探索 (復習)
2. 利得と回避行動
3. 標準型ゲーム
4. 展開型ゲーム
   - ミニマックス法
   - アルファベータ法
5. 小レポート

はじめに

ホイールダック２号は
どうすべきか?

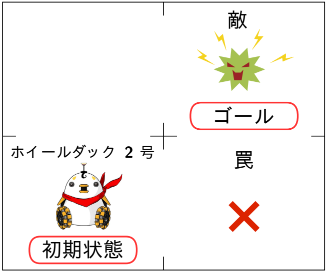
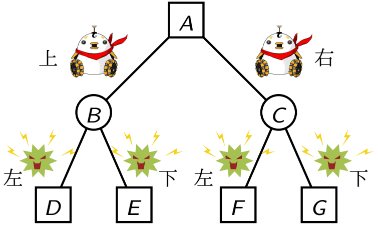


図 **4.2** 敵をかわしてゴールに向かうタスク



ホイール ダック 2 号	C_W	0	D_W	$D_W + C_W$
敵	C_E	D_E	0	$D_E + C_E$

図 **4.3** 図 **4.2** のゲーム展開

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_W, C_E	0, D_E
ホイールダック 2 号が右へ	D_W, 0	$D_W + C_W$, $D_E + C_E$

利得行列

プレイヤーが二人の場合，

- 各プレイヤーの行動を行列の行と列に書く
- 各セルにそれぞれのプレイヤーが得る利得を書く

- 一般の行列とは異なり，**双行列 (bimatrix)** と呼ばれる

表 4.1 図 4.2 のゲームの利得行列

プレイヤー 2 の行動

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	C_W, C_E	$0, D_E$
ホイールダック 2 号が右へ	$D_W, 0$	$D_W + C_W, D_E + C_E$

⇑ プレイヤー 1 の行動

⇑ プレイヤー 1 の利得，プレイヤー 2 の利得

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 号が上
  ⇒ 敵は左
  ⇒ ホイールダック 2 号の利得は −5
- ホイールダック 2 号が右
  ⇒ 敵は下
  ⇒ ホイールダック 2 号の利得は −7

ホイールダック 2 号にとっては
上への移動を選ぶのが最適な選択

仮定: ホイールダック 2 号が先に行動

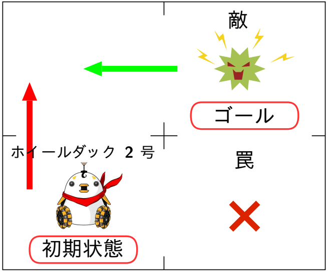


図 4.2 敵をかわしてゴールに向かうタスク

$C_W = -5, C_E = 3, D_W = D_E = -2$	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

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 号が上
  ⇒ 敵は左
  ⇒ ホイールダック 2 号の利得は −5
- ホイールダック 2 号が右
  ⇒ 敵は左
  ⇒ ホイールダック 2 号の利得は −3

ホイールダック 2 号にとっては
右への移動を選ぶのが最適な選択

ホイールダック 2 号が先に行動

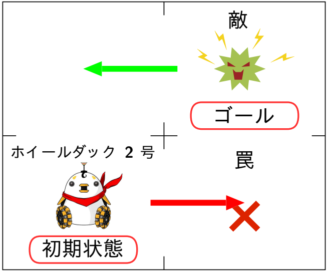


図 4.2 敵をかわしてゴールに向かうタスク

$C_W = -5, C_E = 2, D_W = D_E = -3$	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

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 号が上
  ⇒ 敵は左
  ⇒ ホイールダック 2 号の利得は −5

多くの主体の意思決定が混ざり合って状況が決定する系を**ゲーム**と呼ぶ.

ホイールダック 2 号にとっては
右への移動を選ぶのが最適な選択

ホイールダック 2 号が先に行動

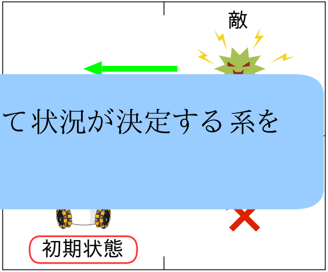


図 4.2 敵をかわしてゴールに向かうタスク

$C_W = -5, C_E = 2, D_W = D_E = -3$	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

本日の内容

ゲーム理論

- 複数のプレイヤーの意思決定を扱う理論
- "Theory of Games and Economic Behavior"
  (ゲームの理論と経済行動)
  J. von Neumann and O. Morgenstern, 1944

基本的な用語の定義

- **プレイヤー:** 意思決定を行う個々の主体
- **行動:** プレイヤーが取ることのできる選択肢
- **戦略:** プレイヤーどの行動を選ぶかを決定すること
- **利得:** プレイヤーの行動の組合せに対して定義される数値
- **合理的:** 各プレイヤーは自分の利益を最大化する
- **均衡:** 合理的な意思決定の結果としての全プレイヤーの戦略の落ち着く先

標準型ゲーム

- (ここでは) プレイヤーの数を 2 に限定
- 標準型 2 人ゲームは利得行列 (双行列) で定義される
- 全てのプレイヤーが 同時に戦略を決定，行動を起こす
  - **純 (粋) 戦略 (pure strategy):** 行動から 1 つを決定的に選択
  - **混合戦略 (mixed strategy):** 行動から 1 つを確率的に選択

表 **4.1** 図 **4.2** のゲームの利得行列

プレイヤー **2** の行動

	敵は左へ	敵は下へ
ホイールダック 2 号が上へ	C_W, C_E	$0, D_E$
ホイールダック 2 号が右へ	$D_W, 0$	$D_W + C_W, D_E + C_E$

⇑ プレイヤー **1** の行動

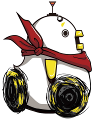
⇑ プレイヤー **1** の利得，プレイヤー **2** の利得

支配戦略均衡

- **支配戦略:** 他のプレイヤーの戦略に関わらず，最も高い利得が得られる戦略
- **支配戦略均衡:** 全てのプレイヤーが支配戦略を取ること

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



ナッシュ均衡

定義 (ナッシュ均衡)

他のプレイヤーが戦略を変えない限り，どのプレイヤーも戦略を変える動機を持たない戦略の組

定理

支配戦略均衡はナッシュ均衡である

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

ナッシュ均衡

定義 (ナッシュ均衡)

他のプレイヤーが戦略を変えない限り，どのプレイヤーも戦略を変える動機を持たない戦略の組

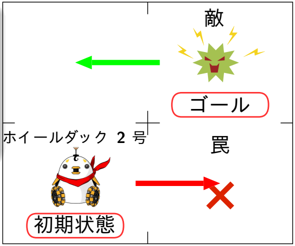


図 4.2 敵をかわしてゴールに向かうタスク

- 敵は左へ移動することが支配戦略
- このとき，ホイールダック 2 号は右への移動を選ぶ

$C_W = -5, C_E = 2, D_W = D_E = -3$	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	$-3,\ 0$	$-8,\ -1$

ナッシュ均衡

定義 (ナッシュ均衡)

他のプレイヤーが戦略を変えない限り，どのプレイヤーも戦略を変える動機を持たない戦略の組

- (上, 右) $= (\frac{1}{6}, \frac{5}{6})$, (左, 下) $= (\frac{7}{10}, \frac{3}{10})$ がナッシュ均衡
- 純戦略でのナッシュ均衡は存在しない

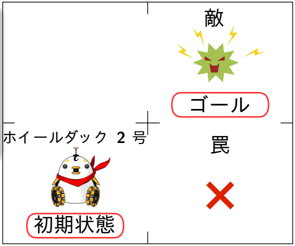


図 **4.2** 敵をかわしてゴールに向かうタスク

$C_W = -5, C_E = 3, D_W = D_E = -2$	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	$-2, 0$	$-7, 1$

2×2 双行列ゲームのナッシュ均衡の求め方

ホイールダック 2 号の利得

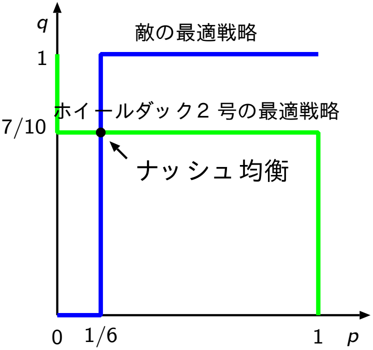
上 $= -5q$

右 $= -2q - 7(1 - q) = 5q - 7$

$q < 7/10$ のとき $-5q > 5q - 7$
$\Rightarrow$ WD2 号は上へ $(p = 1)$

$q = 7/10$ のとき $-5q = 5q - 7$
$\Rightarrow$ WD2 号は任意 (任意の p)

$q > 7/10$ のとき $-5q < 5q - 7$
$\Rightarrow$ WD2 号は右へ $(p = 0)$



		q	$1-q$
	$C_W = -5, C_E = 3, D_W = D_E = -2$	敵は左へ	敵は下へ
p	ホイールダック 2 号が上へ	$-5, 3$	$0, -2$
$1-p$	ホイールダック 2 号が右へ	$-2, 0$	$-7, 1$

2×2 双行列ゲームのナッシュ均衡の求め方

$\boxed{\text{敵の利得}}$

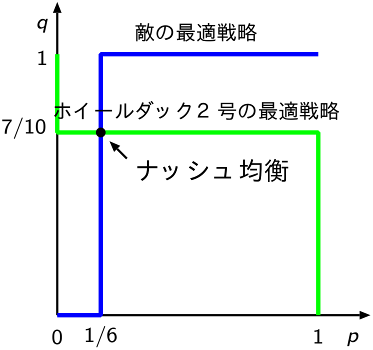
左 $= 3p$

下 $= -2p + (1-p) = -3p + 1$

$p < 1/6$ のとき $3p < -3p + 1$
$\Rightarrow$ 敵は下へ $(q = 0)$

$p = 1/6$ のとき $3p = -3p + 1$
$\Rightarrow$ 敵は任意 (任意の q)

$p > 1/6$ のとき $3p > -3p + 1$
$\Rightarrow$ 敵は左へ $(q = 1)$



		q	$1-q$
	$C_W = -5, C_E = 3, D_W = D_E = -2$	敵は左へ	敵は下へ
p	ホイールダック ２号が上へ	$-5, 3$	$0, -2$
$1-p$	ホイールダック ２号が右へ	$-2, 0$	$-7, 1$

囚人のジレンマ

ナッシュ均衡は全体を良くするとは限らない!

どちらも自白することが支配戦略均衡

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	$-3, \quad -3$	$5, \quad -5$
黙秘	$-5, \quad 5$	$3, \quad 3$

ゼロサム・ゲーム

定義 (ゼロサム・ゲーム)

プレイヤーの利得の総和が 0 となるゲーム

プレイヤーが 2 人の場合

- 利得行列にはプレイヤー 1 の利得を記入
- プレイヤー 1 の利得が $r \Rightarrow$ プレイヤー 2 の利得は $-r$

表 **4.6** 賭けジャンケンの利得行列：ゼロサム・ゲームの例

	P2: グー	P2: チョキ	P2: パー
P1: グー	0	100	-100
P1: チョキ	-100	0	100
P1: パー	100	-100	0

マックスミニ戦略

WD\ 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

ホイールダック 2 号が

- a_1 を選択 ⇒ b_4 のとき利得 1
- a_2 を選択 ⇒ b_1 のとき利得 2
- a_3 を選択 ⇒ b_2 のとき利得 3
- a_4 を選択 ⇒ b_3 のとき利得 1

したがって，a_3 が良さそう
⇒ マックスミニ戦略

WD\ 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

敵が

- b_1 を選択 ⇒ a_1 のとき利得 -7
- b_2 を選択 ⇒ a_3 のとき利得 -3
- b_3 を選択 ⇒ a_1 のとき利得 -5
- b_4 を選択 ⇒ a_4 のとき利得 -6

したがって，b_2 が良さそう
⇒ マックスミニ戦略

ホイールダック 2 号が a_3 を，敵が b_2 を選択するのがナッシュ均衡

マックスミニ戦略

WD\ 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

WD\ 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

定理

プレイヤー 1 のマックスミニ戦略とプレイヤー 2 のマックスミニ戦略の組は，それぞれ得られる利得の和が 0 であるとき，ナッシュ均衡である．

注

上記定理の条件は純戦略では達成できないかもしれないが，混合戦略まで考えれば必ず達成できる．

⇒ マックスミニ戦略

ホイールダック 2 号が a_3 を，敵が b_2 を選択するのがナッシュ均衡

演習問題 1

以下の利得行列 (双行列) に対するナッシュ均衡を全て求めよ．

	b_1	b_2
a_1	2, 3	3, 1
a_2	4, 2	1, 4

演習問題１の解答

プレイヤー１の利得

- a_1: $2q + 3(1-q) = 3 - q$
- a_2: $4q + 1(1-q) = 1 + 3q$

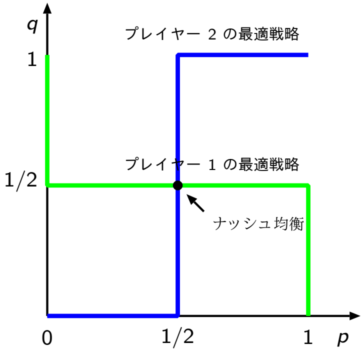
$q = 1/2$ で戦略変更

プレイヤー２の利得

- b_1: $3p + 2(1-p) = 2 + p$
- b_2: $p + 4(1-p) = 4 - 3p$

$p = 1/2$ で戦略変更

		q	$1-q$
		b_1	b_2
p	a_1	2, 3	3, 1
$1-p$	a_2	4, 2	1, 4



$$(p, q) = \left(\frac{1}{2}, \frac{1}{2}\right)$$

本日の内容

1. ホイールダック 2 号の迷路探索 (復習)
2. 利得と回避行動
3. 標準型ゲーム
4. 展開型ゲーム
   - ミニマックス法
   - アルファベータ法
5. 小レポート

展開型 2 人完全ゲーム

- 2 人で行う
- 交互に手を指す
- 状態が完全に分かっている
- 2 人の何れかが勝つか，引き分けとなる

例：

- チェス
- 将棋
- 囲碁

展開型ゲームにおける基本戦略

次のような手順を探す：

- 自分に都合がよい (＝自分の評価値が高い)
- 相手に都合が悪い (＝相手の評価値が低い)

仮定

実際には相手側の評価値は分からないので、

$$\textbf{(相手の評価値)} = - \textbf{(自分の評価値)}$$

と仮定する

ゲーム木 (＝状態空間内の木) を探索

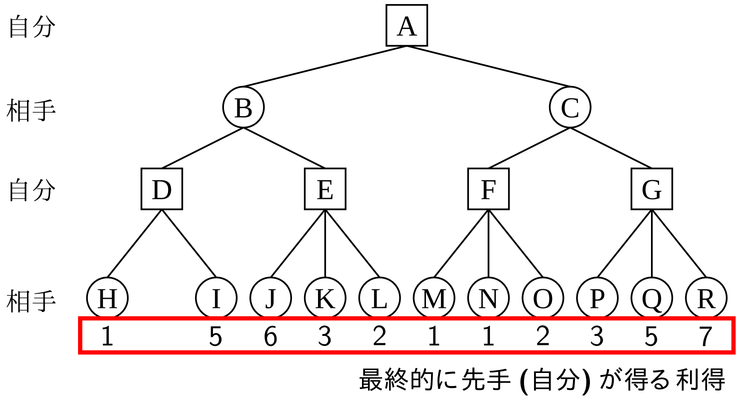
- minimax 法 (ミニ・マックス法)
- α-β 法

評価値の例

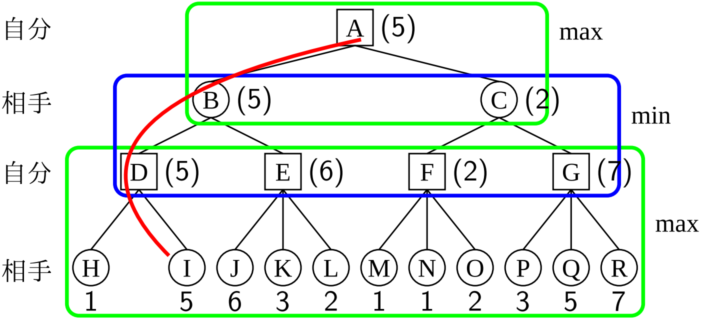
第 3 期叡王戦第 3 局 (ニコ生) より

https://www.youtube.com/watch?v=4jo9OoHPNfQ

ゲーム木


自分
相手
自分
相手
A
B C
D E F G
H I J K L M N O P Q R
1 5 6 3 2 1 1 2 3 5 7
最終的に先手 (自分) が得る利得

ミニマックス法



- ここでは，3 手先の評価で次に打つ手を考える
- 相手の手番は最小値，自分の手番は最大値を選ぶと仮定

ミニマックス法

Algorithm minimax(s, p)

1. **if** $p =$ 自分 **then**
2. payoff $\leftarrow -\infty$
3. **for** $s' \in \mathrm{child}(s)$ **do**
4. $(\mathrm{tmp}, \mathrm{tmp2}) \leftarrow \mathrm{minimax}(s', \text{相手})$
5. tmp $>$ payoff ならば payoff $\leftarrow$ tmp, act $\leftarrow s'$
6. **end for**
7. **return** (payoff, act)
8. **enf if**

注: アルゴリズムは基本的に深さ優先探索を用いている

ミニマックス法

Algorithm minimax(s, p)

```
if p = 相手 then
    if s が葉 then payoff ← s の評価値 else payoff ← +∞
    for s' ∈ child(s) do
        (tmp, tmp2) ← minimax(s', 自分)
        tmp < payoff ならば payoff ← tmp, act ← s'
    end for
    return (payoff, act)
enf if
```

注: アルゴリズムは基本的に深さ優先探索を用いている

minimax 法の注意点

葉の評価値はどのように決められているか？

- 盤面のみから数値化する関数によって取得
- この関数は機械学習によって獲得

なぜ 1 手目で上記の評価値を使わないのか？

- 盤面の評価値とミニマックス法による評価値は別物
- ミニマックス法による評価の方が多くの盤面を検討している分，より良い評価値が得られると考えられる

アルファベータ法

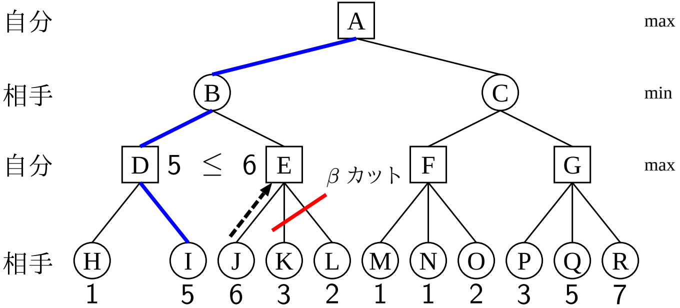
- 盤面の局面を先読みすればするほど良い手を選択し，ゲームを有利に進められる
- 探索の効率化 $\Rightarrow$ 不必要な探索を避ける

アルファベータ法 (α-β 法)

- β カット
  後手が評価値の高い手を打たないことを利用した
  先手の行動のカット
- α カット
  先手が評価値の低い手を打たないことを利用した
  後手の行動のカット

アルファベータ法

左から探索と仮定



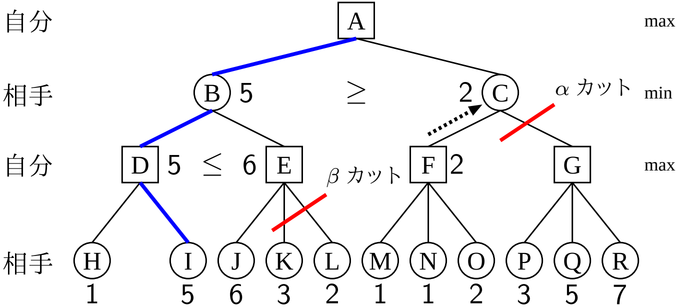
J を探索したとき，

- D の評価値 $= 5$, E の評価値 ≥ 6

局面 B において相手は E を選択しないことが確定

アルファベータ法

左から探索と仮定



O を探索して F の評価値が確定したとき,

- B の評価値 $= 5$, C の評価値 ≤ 2

局面 A において自分は C を選択しないことが確定

アルファベータ法

Algorithm alphabeta(s, p, bound)

1. **if** p = 自分 **then**
2. payoff $\leftarrow -\infty$, $\underline{\beta \leftarrow \text{bound}}$
3. **for** $s' \in \text{child}(s)$ **do**
4. $(\text{tmp}, \text{tmp2}) \leftarrow$ alphabeta(s', 相手, payoff)
5. tmp $>$ payoff ならば payoff $\leftarrow$ tmp, act $\leftarrow s'$
6. <u>payoff $\geq \beta$ ならば for-loop を強制終了</u>(β カット)
7. **end for**
8. **return** (payoff, act)
9. **enf if**

注: アルゴリズムは alphabeta(初期状態, 自分, $+\infty$) で開始する

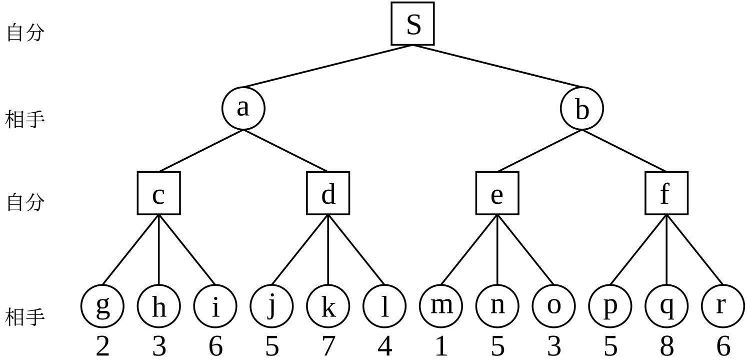
アルファベータ法

Algorithm alphabeta(s, p, bound)

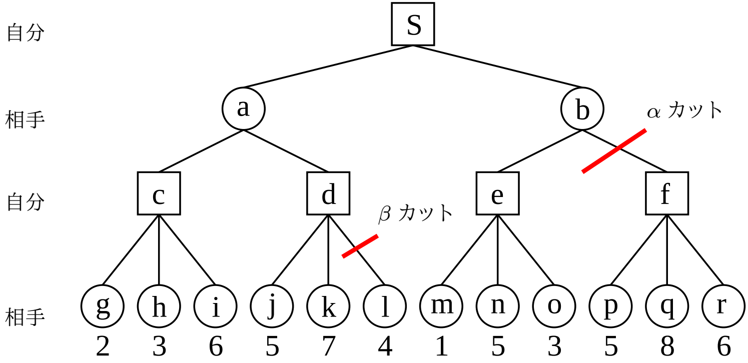
if $p =$ 相手 **then**
    if s が葉 **then** payoff $\leftarrow s$ の評価値
    else payoff $\leftarrow +\infty$, $\underline{\alpha \leftarrow \text{bound}}$
    for $s' \in$ child(s) **do**
        (tmp, tmp2) $\leftarrow$ alphabeta(s', 自分, payoff)
        tmp $<$ payoff ならば payoff $\leftarrow$ tmp, act $\leftarrow s'$
        $\underline{\text{payoff} \leq \alpha \text{ ならば for-loop を強制終了}}$**(**$\alpha$ カット **)**
    end for
    return (payoff, act)
enf if

演習問題 2

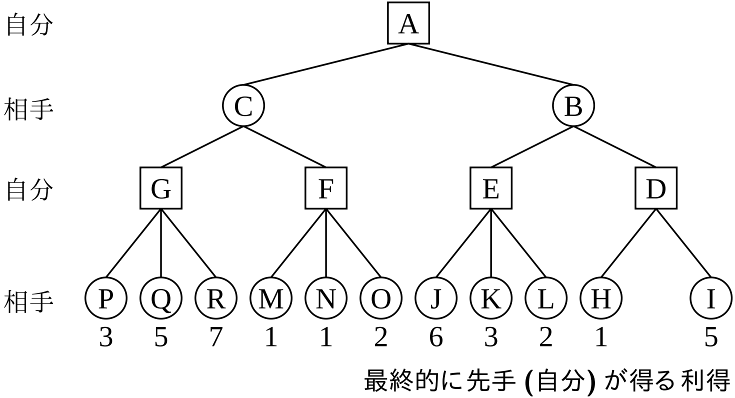
以下のゲーム木に α-β 法を適用し，α カットされる場所と β カットされる場所を示せ.



演習問題２の解答



最善手の重要性



- 悪手順に並べると α カットも β カットも発生せず
- 最善手がわかっていると多くの α カットや β カットが発生

本日の内容

1. ホイールダック 2 号の迷路探索 (復習)
2. 利得と回避行動
3. 標準型ゲーム
4. 展開型ゲーム
   - ミニマックス法
   - アルファベータ法
5. 小レポート

小レポート

1. 以下の利得行列 (双行列) が与えられたとき，ナッシュ均衡を全て求めよ．

(1)

	b_1	b_2
a_1	3, 4	4, 1
a_2	1, 2	3, 5

(2)

	b_1	b_2
a_1	3, 4	1, 1
a_2	1, 2	3, 5

2. 次のスライドのゲーム木に α-β 法を適用し，α カットされる場所と β カットされる場所を示せ．

小レ ポート (続き)