

局面評価の学習を目指した 探索結果の最適制御

東北大学院・理 化学専攻 保木邦仁

力任せの探索は簡単・高性能！

Minimax 法	—————	$(80)^n$
Minimax 法 + beta cut	—————	$(\sqrt{80})^n = (8.9)^n$
Minimax 法 + beta cut + null move pruning や hash cut		$\left(\frac{1}{4}\sqrt{80}\right)^n = (2.23)^n$
Minimax 法 + beta cut + null move pruning や hash cut + Futility pruning		$\frac{1}{4}\left(\frac{1}{4}\sqrt{80}\right)^n = \frac{1}{4}(2.23)^n$

1秒に50万局面探索

→ 1秒で18手先まで読める

実際は静止探索や move reordering の不備により効率が落ちる。
思考時間は序盤 3^n , 終盤 5^n に比例する程度

Bonanza 探索アルゴリズム概要1

通常の探索

基本的に全幅探索+静止探索. 用いる枝刈りは

- Beta cut
- Null move pruning
- Futility pruning
- 2, 8段目の香, 飛角不成りは考えない

- Bitboard
- 反復深化
- Aspiration search. ウィンドウ幅は歩の交換値2つぶん
- PV search
- 延長王手 (1手), one reply (0.5手), リキャプチャー (0.5手)
- 指し手の順序並び替え
 - Transposition table (静止探索では用いない)
 - 再帰的反復深化(PV node で hash move が手に入らない場合)
 - 一手で詰む王手
 - Static Exchange Evaluation (SEE)
 - Killer moves
 - History heuristics
- 専用の詰め探索ルーチンを使わない.

Bonanza 探索アルゴリズム概要2

静止探索

手番を持つ側は手を指すか stand pat を返すか選択

静止探索を開始してから深さ7段目まで SEE の下で駒損しない以下の手を探索

- ・取る手
- ・成る手
- ・王の移動
- ・一手詰みの王手

8段目以降は歩を取る成らない手を除いた駒損しない駒を取る手を生成

SEE による指し手の順序並び替え

静的評価関数

駒二つの組み合わせを選び, 位置等対応する特徴ベクトル要素を足し合わせる

全特徴ベクトル要素は棋譜から機械学習により求めた

1万を超える全ベクトル要素数を探索アルゴリズムと辻褃が合うよう手動で調整し, minimax 型探索の性能を向上させるのは不可能. 機械学習による自動調整が不可欠

局面評価の学習

- TD-Gammon (G. Tesauro)
思考部の学習
Temporal difference + neuron network
- Logistello (M. Buro)
辺のパターンの重み学習
最小二乗法
- GPS 将棋 (金子ら)
序盤の駒組みを棋譜から学習
親子, 兄弟モデルを用いた線形回帰

最適制御理論

最大(小)化問題として力学系の制御方針を推論する

- ラグランジュ形式の解析力学
- パルス整形による化学反応制御
- 最小燃料のロケット弾道
- 工場の消費電力
- 池の魚に与える餌

$$J = \int_0^T l(x, u, t) dt$$

t : 時間に関する数(離散も可)

$x(t)$: 系の状態

u : 制御変数

t を手数, $x(t)$ を minimax 探索の指し手で発展していく局面, u を特徴ベクトルとみなし, 最適制御理論に基づいた将棋プログラムの機械学習をおこなう

Minimax 探索結果の最適制御

最適制御法に基づき、棋譜の指し手と minimax 探索が良く一致する特徴ベクトル v を求める

指し手の一致度を測る目的関数 J' を以下のように設計する

$$J'(P_0, P_1, \dots, P_{N-1}, v) = \sum_{i=0}^{N-1} l(P_i, v)$$

P_i : サンプルされた棋譜中の一局面

$l(P_i, v)$: 全合法手の評価値の違いの度合いを測る

$$l(P, v) = \sum_{m=1}^M T[\xi(p_m, v) - \xi(p_{m=0}, v)]$$

p_m : 局面 P を合法手 m で進めた子局面

M : 局面 P での合法手の数

$m=0$: 棋譜中で実際に指された手

$\xi(p_m, v)$: minimax 探索の評価値

$T(x)$: 評価値の差を、棋譜の指し手との一致度に変換する関数

$T(x)$ の関数形と役割 1

一価の単調増加関数。 $|x|$ が大きい領域で傾きが小さく、 $x=0$ 付近で傾きが大きくなる 1 階微分可能なもの。 $x=0$, $x < 0$, $x > 0$ の意味は...

$T(x)$ を階段型関数にとると、目的関数 J' は「サンプルされた局面中、棋譜で指された手よりも良く判断してしまった合法手の数」となる。

強いプレイヤーと同じ手を指す minimax 関数の設計

↓
目的関数に停留値を与える
特徴ベクトル v の求解

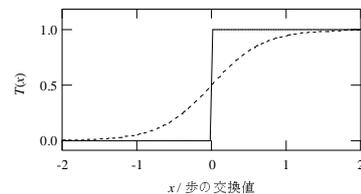


図1: $T(x)$ の関数形。(実線) 階段型関数 (破線) 計算で実際に用いられたもの

$T(x)$ の関数形と役割 2

$T(x)$ が傾きを持つ領域の幅は、目的関数が指し手の善悪判断を行う分解能に相当

- ・ 分解能が良すぎると…

目的関数の滑らかさが失われる.

傾きを持つ領域内にあるサンプル中の合法手が減少し、学習データが不足.

- ・ 分解能が悪すぎると…

目的関数の大きさと、強い人の指し手との一致度の関係が失われる.

悪い手をさらに悪く、良い手をさらに良く評価する方針で最適化されてしまう.
このように、minimax 探索結果を変えない調整は適当ではない.

Minimax 探索の深さが十分ではなく、駒の損得・詰みが全く認識できない局面も学習データとして使ってしまう.

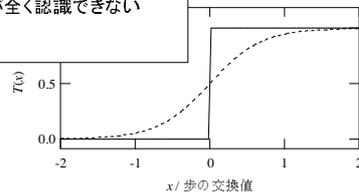


図 1: $T(x)$ の関数形. (実線) 階段型関数
(破線) 計算で実際に用いられたもの

拘束条件の導入

自明解 $v=0$ や駒割り等が定数倍変化した別解の除去

$$J''(P_0, \dots, v) = J'(P_0, \dots, v) + \lambda [M_1(v) - M_0]$$

λ はラグランジュの未定乗数. $M_1(v)$ は駒割りに関する特徴ベクトル要素の大きさ等に相当し、これを定数 M_0 に拘束する.

ラグランジュ未定乗数 λ を導入する利点

p 個の変数に対する q 個の拘束条件から $p-q$ 個の独立な変数を求めることは可能だが、この作業は一般に困難

λ に適当な値を設定し、目的関数の導関数が 0 になる p 個の変数を求めると、独立な変数の組を求めることなく拘束条件を課すことが可能.

ペナルティの導入

- ・ 駒割りの占める割合を出来るだけ高く保つ
- ・ 過学習を回避
- ・ 目的関数の極小点を減らす

最終的な目的関数の表式

$$J(P_0, \dots, \mathbf{v}) = \sum_{i=0}^{N-1} l(P_i, \mathbf{v}) + \lambda [M_1(\mathbf{v}) - M_0] + w M_2(\mathbf{v})$$

w はペナルティの重みを決めるパラメタ,
 $M_2(\mathbf{v})$ は特徴ベクトル \mathbf{v} の大きさを表す関数

最適化の数値的手法 1

ベクトルの要素数が非常に多いため、目的関数の勾配を求めて最適化を行う

$$\begin{aligned} \nabla_{\mathbf{v}} J(P_0, \dots, \mathbf{v}) &= \sum_{i=0}^{N-1} \sum_{m=1}^{M_i-1} \frac{dT(x)}{dx} \nabla_{\mathbf{v}} [f(p_{i,m}^{\text{leaf}}, \mathbf{v}) - f(p_{i,m=0}^{\text{leaf}}, \mathbf{v})] \\ &\quad + \lambda \nabla_{\mathbf{v}} M_1(\mathbf{v}) + w \nabla_{\mathbf{v}} M_2(\mathbf{v}) \end{aligned}$$

minimax 探索の結果としての最善応手列が \mathbf{v} 近傍で単一と仮定し以下の関係を用いた

$$\nabla_{\mathbf{v}} \xi(p_{i,m}, \mathbf{v}) = \nabla_{\mathbf{v}} f(p_{i,m}^{\text{leaf}}, \mathbf{v})$$



興味深いことに、TD-leaf でも...

最適化の数値的手法 2

目的関数が十分滑らかではない

- ・ v を更新すると最善応手系列が変わる
- ・ $T(x)$ の幅の中にあるサンプル数が少ない

よって、2次収束の性質をもつ手法は上手く働かない

$$v_i^{\text{new}} = v_i^{\text{old}} - h \text{sign} \left[\frac{\partial J(P_0, \dots, v)}{\partial v_i} \right]$$

- ・ 初期特徴ベクトル要素と h を整数にとる
- ・ h は初めは粗く、徐々に小さく

計算の詳細 1

静的評価関数で考慮する局面の特徴

- 駒割
- 王と他の駒2つの位置
- 王と王に隣接した味方の駒とその他の味方の駒3つの位置
- 隣接しあった駒2つの位置関係
- 竜馬飛角桂香の利き上にある駒の種類
- 竜馬飛角香が動ける枡の数
- pin analysis. ピンされている駒の種類方向王との距離を考慮
- 角と同じ色の枡にいる味方の歩の数
- 歩桂銀が前進できるか
- 竜飛香の前・後の歩
- 王の周囲25枡の利きの配置

プロ棋士の棋譜3万局と将棋クラブ24の棋譜3万局(主に入玉)
を用いて静的評価関数のパラメタ約1万を調整

$\xi(p, v)$ には一手読み+静止探索の Bonanza を使用

計算の詳細 2

- ・ 拘束条件は飛角金銀桂香歩の駒割りの和
- ・ ペナルティーは以下のように課す

$$M_2(\mathbf{v}) = \sum_i A_i(\mathbf{v}) v_i^2$$

$$A_i(\mathbf{v}) = \sum_{i=0}^{N-1} \sum_{m=1}^{M_i-1} \left| \frac{dT(x)}{dx} \frac{\partial}{\partial v_i} [f(p_{i,m}^{\text{leaf}}, \mathbf{v}) - f(p_{i,m=0}^{\text{leaf}}, \mathbf{v})] \right|$$

出現頻度の高い特徴には強いペナルティー

例) 8八王一金の位置に対する得点

ペナルティーなし

1八金	-200
3八金	-150

A_i なし

1八金	-150
3八金	-149

A_i あり

1八金	-150
3八金	-100

結果

駒割 (交換値) 歩 と 香 桂 成香 成桂 成銀 銀 金 角 馬 飛 竜
106 272 279 304 323 363 415 428 527 617 698 700 854

持ち駒の数

歩 : 27 33 21 6 -8 -17 -23...
 香 : 28 39 51 63
 桂 : 22 12 -15 -48
 銀 : 37 28 -2 -51
 金 : 31 21 -4 -39
 角 : 28 9
 飛 : 59 45

角・馬の動ける升の数

角 : -55 -25 -7 0 8 14 6 9
 馬 : -28 -12 -3 1 8 10 16 11

香・飛・竜縦方向に動ける升の数

香 : x -7 2 6 18 25 27 24
 飛 : -61 -43 -22 -9 -2 9 12 17
 竜 : -45 -26 -17 -13 -1 1 10 8

飛・竜横方向に動ける升の数

飛 : -72 -47 -18 4 10 28 21 27
 竜 : -35 -23 -16 -10 -4 4 11 12

王が8八にいる時の味方の金・と

-67 -60 -45 -27 -30 -51 -61 -73 -67
 -57 -59 -28 -4 6 -13 -48 -82 -62
 -18 -26 -26 0 14 -5 -27 -51 -40
 -57 -35 -20 3 -7 -8 -15 -49 -64
 -55 -36 -13 -18 -13 -23 -17 -35 -77
 -57 -16 -5 -8 -12 -25 -37 -54 -68
-34 -3 13 19 -25 -39 -55 -73 -92
-97 K 30 -16 -6 -51 -36 -85-150
-40 34 -49 23 -40 -17-109-123-118

王が9九にいる時の味方の金・と

-69 -45 -17 8 -13 -38 -70 -79 -75
 -21 -55 11 29 22 -12 -55 -80 -75
 -12 -29 -8 28 37 0 -36 -56 -65
 14 -35 -7 33 9 -5 -33 -56 -83
 -81 -49 -9 -11 -14 -25 -39 -56-101
 -59 -25 2 -28 -37 -57 -63 -56 -72
 14 36 17 -19 -49 -74 -80-108 -75
45 27 18 1 -32 -91-123-127-138
 K 40 31 0 -14 -56-139-133-116

王が8八にいる時の敵の金・と

-106 -25 -5 2 -42 -11 -20 -41 -67
 -53 -7 10 -4 5 -3 8 -18 -72
 -62 -26 -18 9 -21 4 -11 -28 -85
 -50 -14 9 -9 -9 -20 -13 -28 -86
 -12 24 25 4 -20 -27 -30 -47 -63
 110 186 144 49 2 -29 -28 -36 -40
 450 450 450 149 27 -12 -22 -25 -34
 450 K 450 156 15 -24 -40 -31 -44
 112 450 212 63 -19 -52 -69 -59 -72

王が9九にいる時の敵の金・と

-52 -8 15 7 -47 -34 -32 -57-177
 -54 14 30 5 -2 -16 6 -49 -98
 -80 -11 -1 12 -17 -1 -22 -31 -79
 -58 -9 -8 -10 -25 -30 -28 -40-130
 -26 -7 -6 -8 -32 -46 -42 -85-182
 128 120 35 -15 -7 -57 -56 -62-116
 320 271 206 65 11 -18 -45 -36 -68
 318 444 214 106 14 -31 -47 -55 -77
 K 448 207 51 -8 -61 -91 -99 -96

王が8八にいる時の味方の歩

x x x x x x x x x
 -1 -39 -44 -52 -36 -49 -45 -67 -47
 16 -31 -42 -13 0 -3 -32 -30 5
21 21 15 21 25 21 14 23 14
 2 -12 1 4 4 -6 -1 -5 4
 7 -12 2 11 0 -5 0 -8 2
 -5 7 3 0 -3 -3 -6 3 2
 62 K 48 10 -3 -19 -34 -27 -79
 37 -97 15 13 -5 -23 -31 -24 -60

王が9九にいる時の味方の歩

x x x x x x x x x
 -48 -86 -48 11 -7 -40 -35 -49 -35
 28 14 -16 12 19 0 -60 -25 -10
58 29 39 26 38 22 8 24 -1
 -7 -6 7 2 -5 -6 -5 -5 -4
 -5 -13 0 3 0 -11 4 -16 -1
 23 9 2 2 -8 2 -8 -1 8
 72 -3 40 26 -10 -37 -49 -40 -99
 K 36 67 33 -10 -39 -49 -56 -90

王が8八にいる時の敵の歩

-48 -111 -57 -38 -21 -25 -29 -60 -56
 8 -52 -41 -6 -3 -2 -24 -24 2
 -8 -3 -7 -7 -2 -1 -4 1 1
 3 -9 -3 -1 -1 -1 4 -13 1
 19 4 4 15 -3 -2 0 -11 -19
86 156 82 31 27 11 3 17 -20
115 0 77 50 4 -13 -18 -21 -80
83 K 84 -19 -4 -39 -25 -29 -86
 x x x x x x x x x

王が9九にいる時の敵の歩

-101 -97 -49 -4 0 -11 -26 -48 -69
 -24 -81 -25 -2 7 -2 -21 -32 -27
 -30 -9 -13 -16 -10 -5 -10 2 3
 -18 -5 1 -8 -4 -12 -9 -24 -8
 1 8 -4 0 -3 -12 -19 -20 -38
72 108 31 44 15 10 -13 7 -76
200 199 76 53 3 -16 -35 -43 -136
0 200 111 36 -4 -54 -32 -35 -87
 K x x x x x x x x x

88王の時の味方の駒2つの位置

金の位置の得点

+8	-1	+1	-7	+1	+1
王	銀	<u>-33</u>	王	金	+1
+5	-14	<u>+1</u>	-2	+9	<u>-96</u>

歩の位置の得点

+1	+1	-2	+1	+1	-1
王	銀	<u>-16</u>	王	金	+2
+72	+19	<u>+52</u>	<u>-107</u>	-6	<u>-42</u>

78王の時の味方の駒2つの位置

金の位置の得点

+4	+6	+23	-1	+9	+18
王	銀	<u>-12</u>	王	金	+10
+18	+4	<u>+33</u>	+2	-18	<u>-28</u>

歩の位置の得点

+0	+0	+0	+12	+6	+4
王	銀	+4	王	金	+12
+32	+21	<u>+32</u>	+25	+4	<u>-42</u>

問題点

王が6-にいる時の敵の飛

-69,	58	350	0	2	176	64	107	-119
126,	132	350	-80	92	150	181	-15	172
97,	48	101	-146	57	-141	0	128	56
-222	-137	90	-74	75	99	-135	-129	-400
-359	-112	94	-32	-225	15	-303	-281	-56
312	-315	-128	-400	344	<u>-382</u>	39	153	<u>-303</u>
-231	-400	-400	-37	-77	<u>-400</u>	<u>-400</u>	<u>-400</u>	39
-2	-92	66	15	105	<u>-369</u>	<u>-400</u>	<u>-205</u>	<u>-95</u>
77	-200	-155	-217	-81	51	<u>-187</u>	<u>-212</u>	<u>-63</u>

