

世界コンピュータ将棋選手権 初制覇までの歴史

金子 知適

東京大学 総合文化研究科

2009年9月11日

概要: GPS 将棋と最近のコンピュータ将棋

- 話者は誰? – GPS 将棋の開発者の一人
- GPS 将棋とは? – 今年の優勝プログラム
- 今日の内容は?
テーマ1 歴史:
 - GPS 将棋の開発経緯と使われている技術
 - 「かんたんレシピ」+ 将棋への対応強化
 - 人間の棋譜から何を学んでいるか?テーマ2 過去から未来へ:
 - プロまで後何歩?

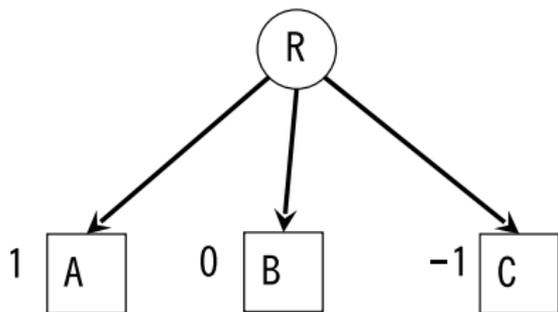
GPS 将棋の歴史

	2003	2004	2005	2006	2007	2008	2009
順位	42	21	8	10	10	12	1
行数	5,974	40,842	65,735	75,585	84,435	92,778	10,1708
改良	初参加	探索 詰探索 (学習)	静止探索 評価関数	探索	評価関数	探索	学習
優勝	IS 将棋	YSS	激指	Bonanza	YSS	激指	GPS 将棋

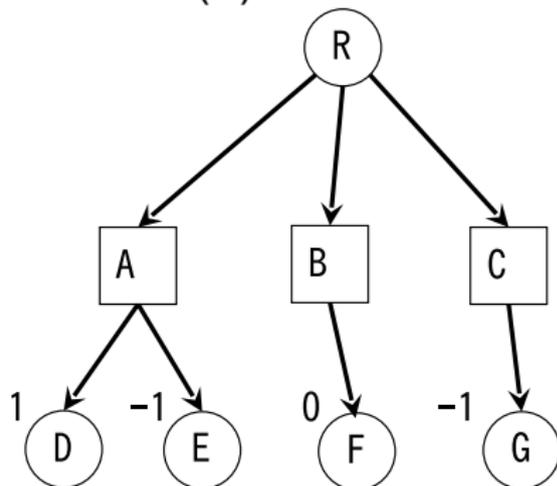
- 2003 年: 研究用プログラムとして開始．ソースコード公開．
- ソースコード (中心部分) の行数は順調に伸びた
↔ 成績の伸びは順調ではなかった
- 2008 年-2009 年: **評価関数の学習**以外はほとんどそのまま
→もっと早くかんたんレシピを知っていれば!

ゲーム木探索 (復習)

評価値: 勝ち (1), 負け (-1), 引き分け (0)

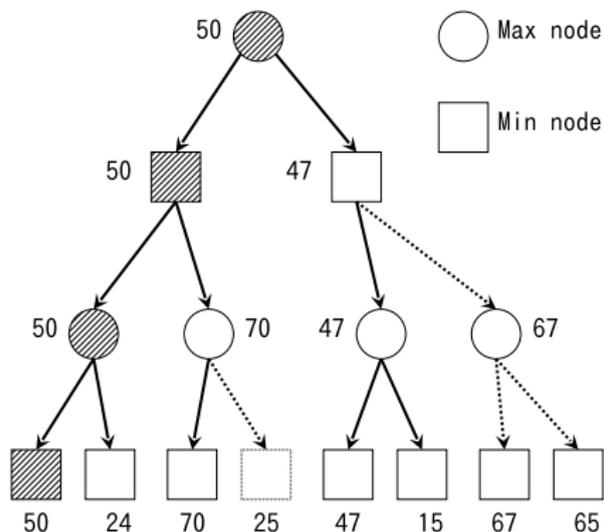


1手の探索



2手の探索

ゲーム木探索と評価関数



評価値:

大きい=先手勝ちや
すい

評価関数:

局面 → 評価値

理想: 勝ち (1), 負け
(-1), 引き分け (0)

現実: 試行錯誤

評価関数の学習 (かんたんレシピ復習)

準備:

- プログラムが評価項目を考える
 - 従来は? 駒の枚数, 持駒の枚数, 玉からの距離...
 - Bonanza は? 玉と玉と駒 ($81 \cdot 80 \cdot 80 \cdot 26$), 玉と駒と駒 ($81 \cdot 80 \cdot 26 \cdot 79 \cdot 26$)
- 棋譜を用意する:
 - プロ棋士の棋譜の方がたぶん良い

本番: 棋譜の手を選ぶように繰り返し重みを調整する

- ① 浅い探索 (1手+静止探索4手) を行い
比較用の局面のペア $a_i > b_i$ を作る
(a_i, b_i は局面)
- ② $f(a_i) > f(b_i)$ となるように重みを調整する
(f は評価関数. 後手番なら不等号反転)

GPS 将棋の評価項目 (> 100 項目) の一部

序中終盤共通の値

PieceEvalComponent 駒の価値

PiecePair 2 駒の関係。駒の座標に関係ないものと、X 座標固定と Y 座標固定

King25EffectAttack 玉の 25 近傍の攻撃側の利きの数 (max 127) とそこに利きを付けてる駒 (max 16) の組み合わせ

King25EffectAttackY King25EffectAttack を玉の Y 座標別に

ProgressBonus2 進行度による先後の進行度 (0 から 15 に正規化) の組み合わせ

ProgressBonusAttackDefense 各玉の、攻撃側の進行度と、防御側の進行度の値の組み合わせ

ProgressBonusAttackDefenseAll 両方の玉の、攻撃側の進行度と、防御側の進行度の値の組み合わせ

序中終盤別の値

PieceStand 持駒の種類別の枚数

Pin 動くと王手になってしまう駒を種類と玉からの相対位置別に

King25EffectEach 玉の 25 近傍の各枰で、どちらの駒があるか空白か、と、利きで勝ってるか負けるか同じかの組み合わせ

PawnDropDefense 歩を打てる筋の自玉との X 軸の相対位置

PawnDropAttack 歩を打てる筋の相手玉との X 軸の相対位置

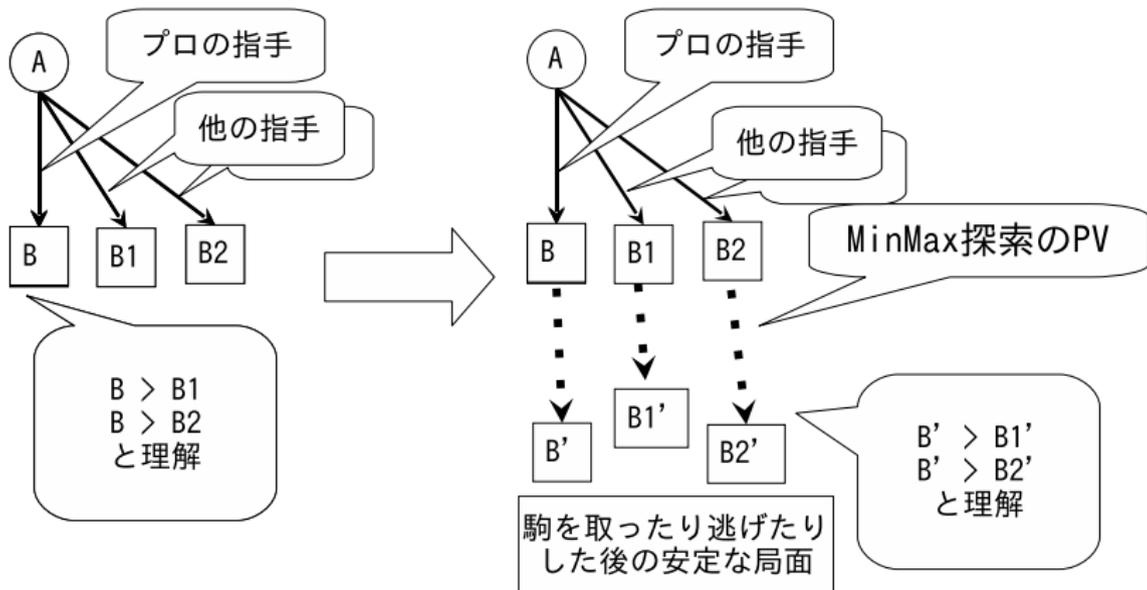
NoPawnOnStand 歩切れでかつ歩の枚数で負けているときの点

GoldRetreat 下がれない金を金の Y 座標別に

SilverRetreat 下がれない銀を銀の Y 座標別に

KnigtAdvance 跳べない桂を桂の Y 座標別に

棋譜と同じ手を指させる: 局面の比較



学習の枠組み

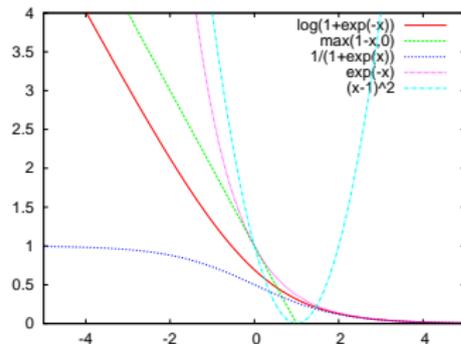
- 教師 $a_i > b_i \leftrightarrow f(a_i) > f(b_i)$
 f : 評価関数, i : ペアの番号, a_i, b_i : 局面
- 最小化する関数

$$\sum_i L(f(a_i) - f(b_i))$$

- 関数 L の候補

$\log(1 + e^{-x})$ (GPS 将棋)

$\frac{1}{1+e^x}$ (Bonanza)



関数 L の背景: ロジスティック回帰

入力:

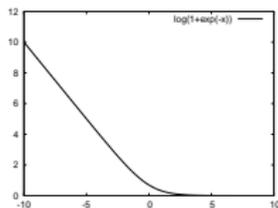
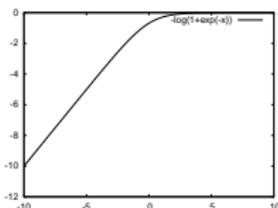
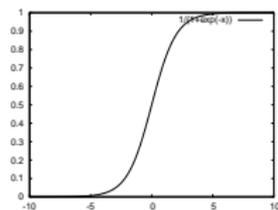
- 教師 i : $a_i > b_i$
- 尤度 i : $\frac{1}{1+e^{-(f(a_i)-f(b_i))}}$

尤度最大化:

- 全体の尤度: $\prod_i \frac{1}{1+e^{-(f(a_i)-f(b_i))}}$
- 対数尤度:
 $\sum_i -\log(1 + e^{-(f(a_i)-f(b_i))})$

L を最小化:

- $\sum_i \log(1 + e^{-(f(a_i)-f(b_i))})$



重みの更新

- 1 各パラメータ w_k について勾配を計算する

$$\frac{\partial}{\partial w_k} \sum_i L(f(a_i) - f(b_i))$$

- 2 勾配に基づき各 w_k を更新する
 - 勾配の符号に基づき固定幅で移動 (Bonanza)
 - 勾配と慣性項 (GPS 将棋)
 - Stochastic Meta Descent (Schraudolph 2002)
 - ヒューリスティック

留意点: 大本の問題は微分できない

- 理由: 探索木が変わる可能性があるため
- 現実: 今のところうまくいっている

- 汎化能力の重要性
 - 現実: 訓練例 (ペア) で出現する局面 \neq 実際に対局するときに評価する局面
 - 目的: 対局するときに正しく評価すること
→ 訓練例だけ丸暗記するのではダメ
- 重みに対するペナルティを含めて最小化: 貢献の少ない重みを 0 にする
 - $\sum_i L(f(a_i) - f(b_i)) + \lambda \sum |w|$
 - $\sum_i L(f(a_i) - f(b_i)) + \lambda \sum |w|^2$
- 出現棋譜数 (\neq 出現ペア数) の少ない特徴を無視

学習の進み具合を示す指標

- 不一致度:

$$\frac{1}{\text{局面数}} \sum_{\text{局面 } i} \sum_{\text{合法手 } j,i} T(\xi(\text{指手 } j,i) - \xi(\text{棋譜の手}))$$

- ξ : 最善応手手順後の局面の評価値,
- $T(x) = 1/(1 + \exp(-3x/100))$ (100 は歩の点数)
- 一致率:

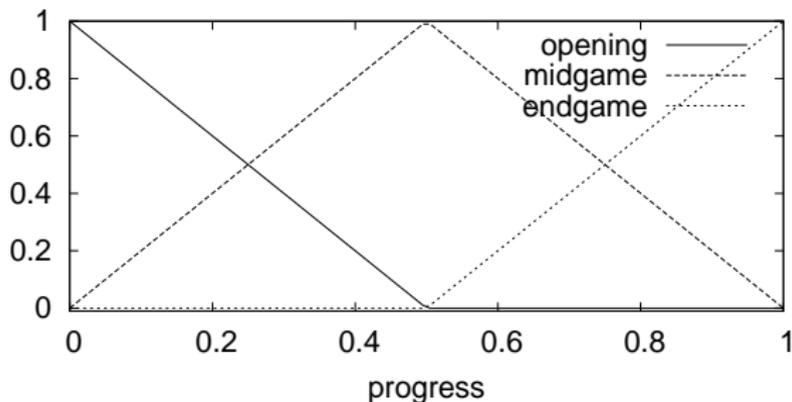
$$\frac{\text{選んだ手が棋譜の手と一致した局面数}}{\text{全局面数}}$$

評価関数の正確さの評価

評価関数	駒割+ α	2008年秋	2009年5月	8月
不一致度	≈ 12	4.0	2.5-3.0	2.0-2.3
一致率			30-35%	$\approx 45\%$
対戦	floodgate 自動対戦のレーティング			
Bonanza	-	-160	-39	-135
2008年版	-	+86	+288	+283
問題集	ラクラク次の一手 基本手筋集 216問			
1	-	171	175	168
2	-	185	182	184

- 昨年 から 今年 への改良は、不一致度やレーティングに反映
- その後の改良の効果は不明確

序盤・中盤・終盤の評価関数の混ぜ合わせ



- 序盤・中盤・終盤の評価関数をそれぞれ用意
- $[0, 1]$ の範囲を取る進行度を定義
- 進行度に応じて序盤と中盤，もしくは中盤と終盤の評価関数の内分を評価値とする
- 学習では一度に重みを調整

GPS 将棋と Bonanza の構成の比較

	GPS 将棋	Bonanza
探索	実現確率 (激指風)	全幅探索
静止探索	(YSS の末端風)	取る手・成る手
評価関数	棋譜から学習	棋譜から学習
評価項目	将棋の知識 100 種類くらい	3 駒の関係 1 種類
詰探索	df-pn (長井 1999)	3 手読み
行数	101,708	29,501

- GPS 将棋: 棋譜・プログラマの将棋の知識の活用
- Bonanza: シンプルな作り, 将棋の知識は最小限
- 強さは? – ほぼ互角
Bonanza: 中盤指向, GPS 将棋: 終盤指向

棋譜から学ぶ・プログラマが教える

● 棋譜から学習

- 評価関数のパラメータ
- 実現確率 (有望な手を深く，そうでない手を浅く探索)
 - 深く: 王手をかけながら金を取る，歩で叩かれたら取る
 - 浅く: そっぽに角を捨てる，一段目に金を打つ
- 詰める確率: 詰探索を行う価値があるか

● プログラマが教える

- 末端の探索
 - 駒を取る，成る，両取りをかける，金銀を歩で叩く，飛車角に利きをつける (ただし n 手まで，指手が多い時は盤上の駒が移動する手を優先させ，持駒の金銀を使う指手は優先度を下げる)
- 評価項目: 何に点数をつけるか
 - 100 項目以上

実現確率を利用した探索

実現確率を利用した探索 (鶴岡ら 2002)

- 指手に確率つけて打ち切り基準まで探索
- 旧来のヒューリスティックの置き換え
 - 前向き枝刈り
 - 指手の並び替え
 - 深さ調整, 探索延長 (例: 王手延長, 取り返し延長)

GPS 将棋での実現

- 指手の“強さ”を求める (評価関数の学習同様)
- “強さ”と棋譜の指手の対応から二つの確率に変換
 - 前向き枝刈り用の確率
 - 深さ調整用の確率

未来へ：プロ棋士の強さまで後何歩？

どこまで強くなったか？

- GPS 将棋はまだ本格的な対局なし
- 他のプログラムはアマチュアトップと良い勝負
 - 激指 - 清水上徹アマ名人
 - 棚瀬将棋 - 加藤幸男朝日アマ名人
 - 激指 - 清水上徹前アマ名人
 - 棚瀬将棋 - 加藤幸男前朝日アマ名人
 - 激指 - 稲葉聡アマ準名人
 - 合議システム - 谷崎生磨学生準名人

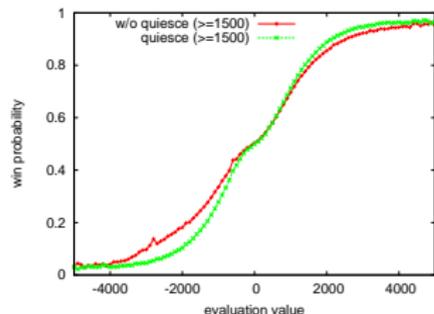
プロのレベルに達するには何が必要か？

- 棋譜の統計データにみるプロとアマの違い
- プロ棋士の棋譜を GPS 将棋に検討させると？

評価関数の正確さの評価: Evaluation Curve

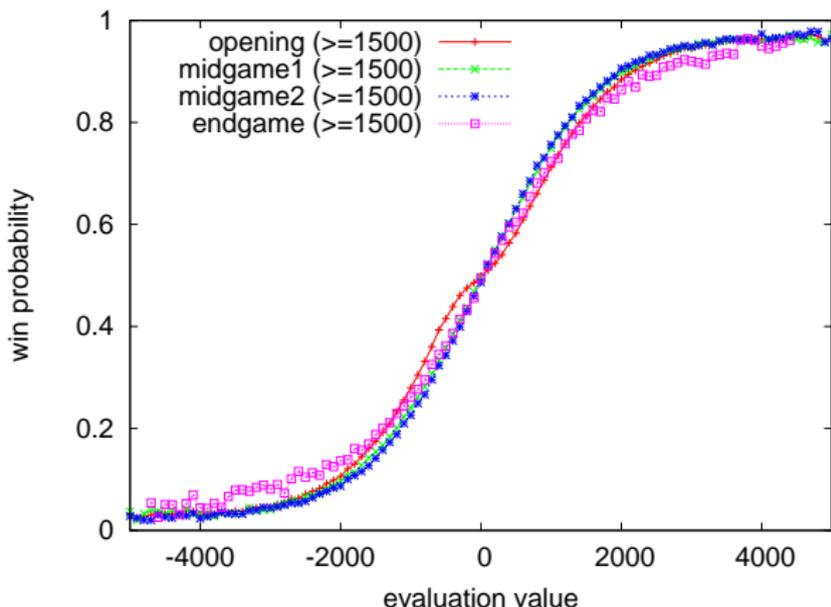
(竹内ら (2007) AAI)

- 動機: 評価関数の正確さを細かく測りたい
c.f. 大雑把な計り方: 対戦して勝った方が良い, 問題集を沢山解いた方が良い etc.
- アイデア: 棋譜の勝敗 \approx 局面の優劣 \approx 評価値



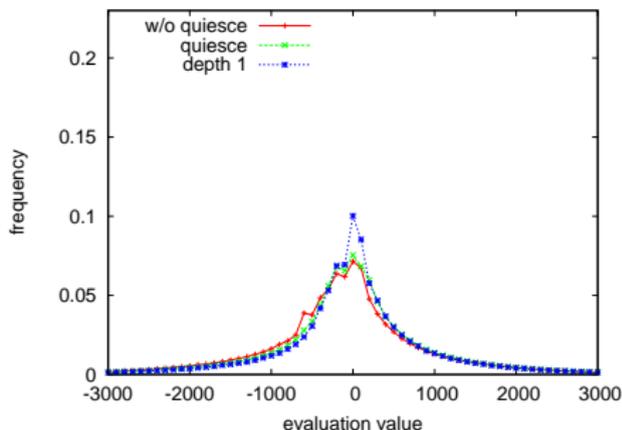
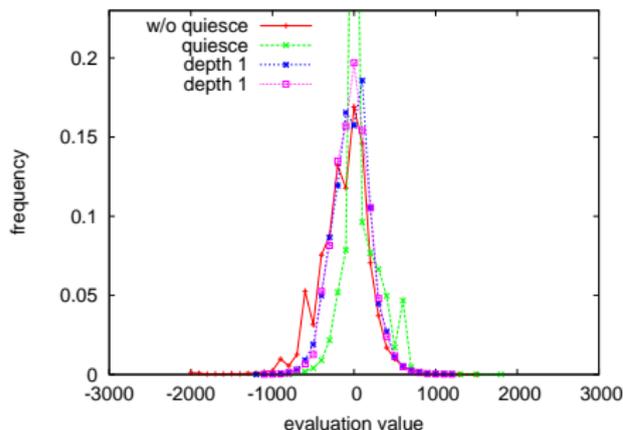
- 横軸: 評価値
- 縦軸: 棋譜上の勝率 (その評価値を持つ局面)
- 望ましい形状:
右肩上がり (c.f., 水平は役に立たない), 傾きが急, 凹凸がない

評価関数の正確さ: ゲームの進み具合



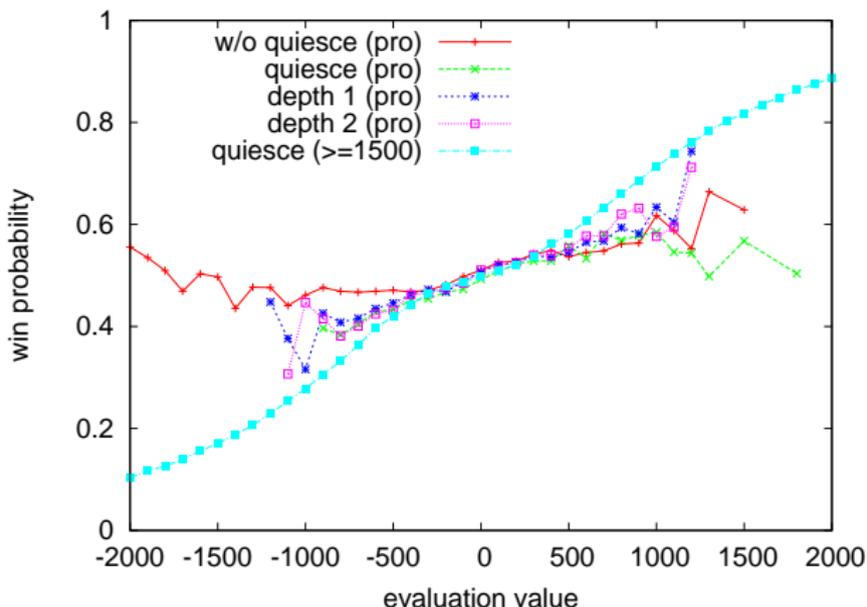
- 正確さ: 中盤 > 序盤 > 終盤
 - 特に序盤の ± 300 点くらいは信頼性が低い
- 棋譜: 将棋倶楽部 24 万局集 (どちらかのレート ≥ 1500)

プロの棋譜の特徴: 頻度の比較



- プロ (左) の方が、0 点近辺に集中
→ きっと互角に近いところで戦っている
- 静止探索すると右にずれる
→ 一手指すと良く見える？

プロの棋譜の特徴: Evaluation Curve



- 水平に近い i.e., 情報量が少ない
- 静止探索や深さ 1 の探索をしてもあまり改善がない

プロの指手を予想する!

<http://twitter.com/gpsshogi>

Game
Programming
Seminar
SHOGI

gpsshogi

[(125) ▲7六桂] 510424 △7
八馬▲同玉△4五角▲5六銀
△同角▲同歩△8七銀▲同玉
△7五桂▲同銀 (45.618 sec)

約2時間前 Perl Net::Twitterで

[(125) ▲7六桂] * 王手を回避して、後手玉は▲2三龍
(以下)の詰めろ。

約2時間前 Perl Net::Twitterで

[(124) △5四角] 509304 ▲6五桂△7五桂▲同銀△6
五角▲同歩△8六歩▲同金△8八金▲同玉△7六桂
▲9八玉 (10.031 sec)

約2時間前 Perl Net::Twitterで

[(124) △5四角] * 応手によっては詰がある。

約2時間前 Perl Net::Twitterで

[(123) ▲7七金] 507802 △7五桂▲同銀△5四角▲6
五桂△8六歩▲同銀△7六金▲9八玉 (4.641 sec)

約2時間前 Perl Net::Twitterで

[(123) ▲7七金] * 王手を回避して、後手玉は▲2三龍
(以下)の詰めろ。

http://twitter.com/gpsshogi_en

Game
Programming
Seminar
SHOGI EN

gpsshogi_en

[(125) N*7f] 510424 6ix7h
8gx7h 5d4e S*5f 4ex5f
5gx5f S*8g 7hx8g N*7e
8fx7e (45.618 sec)

約2時間前 Perl Net::Twitterで

[(125) N*7f] * Gote is in threatmate (3cx2c).

約2時間前 Perl Net::Twitterで

[(124) B*5d] 509304 N*6e N*7e 8fx7e 5dx6e
6fx6e P*8f 7gx8f G*8h 8gx8h N*7f 8h9h (10.031
sec)

約2時間前 Perl Net::Twitterで

[(124) B*5d] * Sente is in danger.

約2時間前 Perl Net::Twitterで

[(123) 6gx7g] 507802 N*7e 8fx7e B*5d N*6e P*8f
7ex8f G*7f 8g9h (4.641 sec)

約2時間前 Perl Net::Twitterで

[(123) 6gx7g] * Gote is in threatmate (3cx2c).

意外と当たる

連続一致数	手数 ≥ 35		手数 ≥ 80	
	頻度	累積%	頻度	累積%
≥ 5	23	20.7	9	13.6
4	30	27.0	12	18.1
3	37	33.3	13	19.7
2	47	42.3	18	27.2
1	64	57.6	28	42.4
0	111	100.0	66	100.0

- 一致率は50%を越えている。(30-180秒)
- 終盤が当たりやすさは棋譜次第
- 当たらない手の「悪さ」の改善が課題:
「悪い手を指すとすぐに負けてしまうゲーム」

判断を誤る事例：後もう少し？



2009年7月23日
 第22期竜王戦決勝トーナメント
 先手: 松尾歩 七段
 後手: 豊島将之 五段

渡辺竜王「3五金には5八角成の一手。同飛4七馬(*1)4三步成5八馬3二銀...同飛は詰みだから(*2)1一玉の一手に3三桂(*3)で、同銀同と3二飛同とで受けなし。後は先手玉が詰むかどうかだけど...」

GPS将棋「5八角成同飛4七馬2八飛以下-312点」(後手有利)
 (*1)4三步成は、飛車取りを放置して後手玉に迫る手。(*2)の展開が先手勝ちだと分からないと飛車を取られて不利と判断。
 →あと1分+1手くらい

判断を誤る事例：当分無理？

	9	8	7	6	5	4	3	2	1	
三歩	皇	将	遊		变	王			皇	一
歩							变	将		二
駒	歩		将	歩	歩	歩	将		歩	三
										四
		歩		馬						五
			歩				飛			六
	歩		桂	歩	歩	歩	歩	歩	歩	七
			金		王		金	銀		八
	香		銀					桂	香	九
										持駒
										歩

8五歩まで

2009年7月17日
 第80期棋聖戦第5局
 先手: 木村一基 八段
 後手: 羽生善治 棋聖

後手の指手: 3一玉
 GPS将棋の読筋: 5二金
 (プロの評価: 1手パスで悪手)

梅田望夫最終局観戦記(4)「5二金なら負けちゃうね」

<http://kifulong.shogi.or.jp/kisei/2009/07/4-5-e9b0.html>

改善の難しさ:

- 直前の指手は新手
 →定跡を暗記する問題ではない
- この1手パスが形勢を悪くするのは何手も後
 →そこまで探索が届かない

おわりに

コンピュータ将棋はどこまで来たか？

- プロの棋譜を理解し始めつつある段階
- →これからが面白いところ

宣伝: GPS 将棋

- ソースコードとバイナリ (Linux, Microsoft Windows 版) 公開中

<http://gps.tanaka.ecc.u-tokyo.ac.jp/gpsshogi/>

- コンピュータ将棋対局場 floodgate 24 時間対局中

<http://wdoor.c.u-tokyo.ac.jp/shogi/floodgate/>

宣伝: Game Programming Workshop

- 2009 年 11 月 13 日 (金) - 11 月 15 日 (日) 箱根
- <http://sig-gi.c.u-tokyo.ac.jp/gpw/2009/>

評価関数の学習:

- 局面評価の学習を目指した探索結果の最適制御. 保木. 第 11 回ゲームプログラミングワークショップ (GPW), 78–83, 2006.
- 将棋の棋譜を利用した, 大規模な評価関数の調整. 金子, 山口. 第 13 回 GPW, 152-159, 2008.
- Visualization and Adjustment of Evaluation Functions Based on Evaluation Values and Win Probability. S. Takeuchi, T. Kaneko, K. Yamaguchi and S. Kawai. AAAI07, 858–863, 2007.
- Comparison training of chess evaluation functions. G. Tesauro. In *Machines that Learn to Play Games*, 117–130, 2001.

詰将棋 (df-pn):

- df-pn アルゴリズムと詰将棋を解くプログラムへの応用. 長井, 今井. 情報処理学会論文誌 43(6):1769–1777, 2002.

実現確率:

- Game-tree Search Algorithm based on Realization Probability. Y. Tsuruoka, D. Yokoyama, and T. Chikayama. *ICGA Journal*, 25(3):145–153, 2002.