## 将棋プログラムにおける棋譜を利用した囲いの評価\*

# 東京大学総合文化研究科 金子 知適†

## 東京大学情報基盤センター 田中 哲朗<sup>‡</sup>

## 概要

将棋プログラムで必要となる囲いの良さの評価を , プログラマの将棋の知識を用いずに実現するために , 棋譜の分析を行ない , 玉の位置と周りの駒の配置で評価する方法を提案する . 実際に 24 万局の棋譜を分析した結果 , 玉の位置だけでも勝率は有意に変わることが分かった . また , 玉の周囲の 5×5 の領域を切り出すという単純な方法により , 良く知られた囲いを得ることができた . また , 得られた囲いの有用性を , 出現回数と勝率により判定することができた .

## 1 はじめに

将棋プログラムを開発するためには、探索アルゴリズムの実装と、形勢を判断する評価関数が必要である、評価関数では駒の損得が最も重要で、強いプログラムで用いられている駒の価値 [4] を用いればある程度までは強くすることができる。しかし、より強くするためには、駒の損得以外の将棋に関する知識が必要である。本稿では、そのような知識の中で特に玉の囲い方について注目する。

将棋では玉を守るために周囲に配置された駒の配置を囲いと呼び,良い囲いや悪い囲いがある.また囲いの効果は相手の陣形にも依存する.囲いに関する知識が必要な理由は,将棋プログラムは,序盤で駒の損得が無いまま形勢に差がつくことが多い¹ためである.そのような例として,図1に第13回世界コンピュータ将棋選手権²の1次予選の1局面を掲載する.先手番が,著者らが開発に参加しているプログ

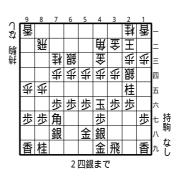


図 1: 囲いを知らない将棋プログラムの序盤

ラム [10] であり、定跡をはずれた後の指し方を教えていないため珍しい形になっている.これに対して、後手番のプログラムはきちんと矢倉に囲い攻撃体制を整えていて、後手有利である.

囲いをはじめ序盤の指し方に関しては,従来,プログラマが持つ将棋の知識をプログラムに翻訳することにより行われてきた.しかし,将棋プログラムをプログラマよりも強くするためには,何らかの方法で自動的に知識を獲得させることが必要である.

そこで本稿では,棋譜から自動的に獲得可能なデータに基づいた囲いの評価方法として,(1) 玉の位置に基づいておおまかに評価し,(2) 玉の周り  $5 \times 5$  の領域の駒の配置によって細かく評価する.という方法を提案する.評価の基準は棋譜を分析して,出現頻度や勝率をもとに設定する.実際に,24 万局の棋譜を分析したところ,玉の位置で勝率は有意に変わるという結果を得た.また玉の周りの駒の分析では,良く知られた囲いを抽出することができた.

## 2 関連研究

囲いのデータを自動的に作らない場合には,あらかじめプログラマが囲いを登録する必要がある.登

<sup>\*</sup>Analysis of Castles in Shogi Records

<sup>&</sup>lt;sup>†</sup>Tomoyuki Kaneko (kaneko@graco.c.u-tokyo.ac.jp), Graduate School of Arts and Sciences, The University of Tokyo

<sup>&</sup>lt;sup>‡</sup>Tetsurou Tanaka, Information Technology Center, The University of Tokyo

<sup>「</sup>序盤と比較して,中盤以降,戦いが始まると駒の損得の計算だけでもそれなりに戦える.

<sup>&</sup>lt;sup>2</sup>http://www.computer-shogi.org/wcsc13/

表 1: 玉の位置の出現回数

	9	8	7	6	5	4	3	2	1	
Ì	78	331	249	226	188	187	211	204	67	六
								1266		
								119945		八
	26638	7703	16489	19317	33275	2363	13641	954	23925	九

表 2: 玉の位置と勝率

9	8	7	6	5	4	3	2	1	
0.449	0.459	0.522	0.385 ***	0.394 **	0.428 *	0.431 *	0.407 **	0.418	六
0.433 **	0.515 **	0.430 ***	0.421 ***	0.447 **	0.417 ***	0.451 **	0.458 **	0.393 ***	七
0.507	0.515 ***	0.506 **	0.477 ***	0.502	0.476 ***	0.505 .	0.508 ***	0.398 ***	八
0.525 ***	0.510.	0.478 ***	0.481 ***	0.483 ***	0.406 ***	0.474 ***	0.413 ***	0.502	九

録した囲いにうまく囲わせる方法としては , YSS の 落とし穴方式 [7] が有名である他, 文献 [1] などの研 究がある「激指」では「玉は端にいる方が良い」な どの汎用性のある規則から囲わせているが,囲いの 形を教えておく方法との優劣は微妙だと述べられて いる[5]. 囲いに近い評価項目としては,中盤以降の 玉の危険度があり,周囲の利きの制圧度や自由度を 測る方法が用いられている [8,3].

一方,棋譜データから獲得可能な将棋に関する囲 い以外の知識としては,定跡や探索延長の指標(実 いた領域を囲いとするという条件も有力であった. 現確率 [2]) がある. 定跡については単純な方法であ る程度うまく行くが,定跡をはずれた後に部分一致 を利用して指す方法などについては研究の余地があ る.他の研究として,著者らは最近2つの駒の関係 に着目し,棋譜に表れた形を良いと認識するような 序盤用の評価関数の学習に成功している[9].

#### 棋譜の分析方法

提案する手法では,囲いを玉の位置と周囲の駒の 配置によって評価する、評価に必要なデータを得る ために,以下に述べるように棋譜を分析した.

用いたデータは将棋倶楽部243で指された24万棋 譜である [6]. データのサイズは 122 メガバイトで あった.様々な棋力のプレイヤが指した棋譜である ため、目的によっては分析の前にデータの洗浄が必 要と思われるが,今回はエラーで読めなかった1棋 譜を除いて全ての棋譜を分析の対象とした.

方針としては,棋譜からあるタイミングで局面を 選び,その時点での玉の位置やその周囲5×5の領域 を分析した.局面を選ぶタイミングとしては,囲い がほぼ完成していることと, まだあまり崩されてい ないことが重要である.本稿では,最初に駒が成っ て相手に取り返されなかった時点とした.発表の時 点では他の基準との比較検討はしていないが、その 後試したところ、棋譜の中で最も長く同じ状態が続

通常は1つの棋譜から先手と後手の2つの囲いが 抽出される.後手の囲いは座標を点対称に反転させ て先手の囲いと同様に扱った、24万局の中で約4.300 局は、そのような条件に到達しなかったか、それま でに反則があった4などの理由で使用できなかった.

分析としては,囲い毎に出現頻度や勝率を調べた. 棋譜の勝ち負けの判定は、棋譜をならべながら詰将 棋プログラムを 1 局面あたり 90 ノードの制限で動 かして、詰みを見つけたらそのプレイヤの勝ち、そ うでなければ終局時に手番だった人が投了したとみ なした.また,囲いの効果は相手の陣形にも依存す るため,互いの玉形による勝率も調べた.

## 玉の位置の出現頻度と勝率

まず, 玉の位置の分析結果を報告する. 玉の位置 の出現回数を表1に示す.まれに先に進んだ棋譜もあ

<sup>3</sup>http://www.shogidojo.com/

<sup>4</sup>王手をかけられて回避しなかった棋譜が多かった.

表 3: 互いの玉の位置と勝率

位	置	出現回数	勝率	有意性
勝	敗			
2/\	6九	2037	0.598	***
2八	6八	2106	0.586	***
2八	5九	2898	0.578	***
2八	3九	1126	0.575	***
9九	3九	1294	0.573	***
8八	7九	3340	0.569	***
8八	6九	1836	0.568	***
8八	7八	1347	0.561	***
7八	5九	1515	0.560	***
2八	7九	2644	0.559	***
7八	4 <i>/</i> \	1887	0.555	***
3八	5九	1792	0.550	***
3八	3九	1121	0.550	***
3八	4 <i>/</i> \	1078	0.548	**
7八	3九	3524	0.535	***
7八	3八	5422	0.532	***
5八	6九	2409	0.527	**
6九	5九	1982	0.525	*
9九	2/\	15484	0.522	***
9八	2八	4227	0.519	*
8七	2八	5381	0.519	**
9九	1九	6426	0.517	**

(勝率 0.5 未満は対称. 出現回数 1000 以上)

るものの,頻度が高いのは8,9段目であり,ほぼ常 識的な結果を得ることができたと思われる.

次に,玉の位置と勝率の関係を表2に掲載する.図 中の記号は勝率 0.5 に対する適合度の  $\chi^2$  検定の結果 '.' ≤ 0.1 ≤' ' ≤ 1 である (以下全て同じ) . いくつか の位置は0.1%の基準でも勝率が5割では無いという 結果が得られている. 有利と思われる位置は2八,8 八,9九の3個所であるが,不利と思われる場所は居 玉 (5九) をはじめ多数ある.また,居飛車穴熊 (9九) は勝率が高い一方で振飛車穴熊(1九)はそうでもな 種類の囲いが求められた. いことなどが読み取れる.

意な差がある位置のペアを表3に掲載する.上位につ について報告する.その過程で,5×5の領域はやや いて、勝ちやすい玉の位置である2八は美濃囲いと 広すぎる傾向が観察されたため、大きさの微調整に 思われる.それに対して,負けやすい位置の6八はついて実験した結果を最後に報告する.

表 4: 玉の位置と囲いの種類

9	8	7	6	5	4	3	2	1	
	2								六
	555	5		3		1	24		七
357	3074	4726	458	373	557	2101	8254	10	八
2055	421	807	676	1540	24	883	7	1988	九

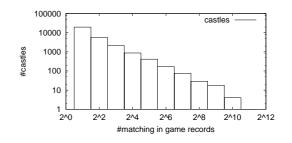


図 2: 囲いの出現回数のヒストグラム

船囲いの途中で,5九は居玉と予想される.6九に関 しては著者らの棋力では解釈が難しいが, まだ棋譜 を調べられていない.

結論として,前例が少ない位置は避け,多く実戦 例があるなかでは相手を見ながらなるべく勝率が高 い場所に囲うことで, 出現頻度と勝率を利用してあ る程度まともにな囲いを組むことができると言える. 例えば,図1の先手玉のような極端に悪い囲いを避け るにはこれだけでも十分である.

#### 5 囲いの分析

続いて, 囲いの分析結果について報告する. 囲い の表現方法は,実戦でのマッチングの容易さを考え て, 玉を中心に含む 5×5 のマスの駒の並び (24 近 傍)とした、玉が端にいるときは領域が将棋盤の外に はみ出さないように調整した.敵の駒に関しては今 回は重視せず、駒の種類を区別せずにまとめて扱っ た.棋譜からそのような領域を切り出した中で,2回 以上出てきた形を囲いとして残したところ,約3万

初めに,得られた囲いを紹介し,続いて囲いが出 続いて,互いの玉の位置と勝率の関係で勝率に有 現した回数と,囲いが用いられたときの勝率の分析

表 5: 囲いと勝率 (出現回数 200 以上)

囲いID	出現回数	勝率	有意性
18478	270	0.622	***
3267	270	0.615	***
12653	205	0.595	**
4581	361	0.579	**
679	472	0.574	**
7821	201	0.572	*
1457	257	0.568	*
8905	513	0.565	**
13282	274	0.562	*
13428	528	0.559	**
13596	712	0.555	**
13605	1510	0.545	***
13446	927	0.538	*
1222	1962	0.538	***
1232	2274	0.533	**
13612	1639	0.528	*
-1	281390	0.484	***
1636	568	0.452	*
2373	403	0.449	*
1773	791	0.439	***
1730	236	0.432	*
2219	413	0.424	**

#### 5.1 出現頻度と勝率

玉の位置毎に,最も頻度が高かった囲いを付録Aに 掲載する . 括弧内の数字が頻度である . 敵の駒は「 で表した.穴熊や美濃囲い,矢倉など,出現頻度が ある程度高い(≥30)ものは自然な囲いが多い.従っ て,出現頻度は囲いの良さを測る上で有効であると 言える.一方,玉の位置が3八や4八の囲いなど,急 戦と思われる囲いも抽出されている.これらを他の 固い囲いと区別するには別の尺度が必要である.

次に,囲いの出現頻度のヒストグラムを図2に示 す. 横軸が出現数で, 縦軸がそのような囲いの種類 である.両軸ともログスケールである.出現頻度が 高い囲いは種類が少なくなっている.特に200回以 上出現した囲いは 62 種類と予想よりもずっと少な かった.原因としては囲いの大きさが広すぎて囲い が細分化してしまった可能性が考えられる.

表 6: 互いの囲いと勝率

勝	負	出現回数	勝率	有意性
18478	-1	208	0.639	***
1232	-1	627	0.609	***
4581	-1	234	0.607	**
1374	-1	244	0.582	*
1222	-1	572	0.563	**

(勝率 0.5 未満は対称,出現回数 200 以上)

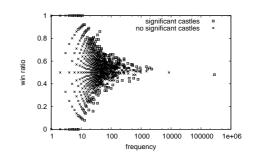


図 3: 囲いの出現頻度と勝率 (5×5)

の勝率を調べた.表5に結果を掲載する.また ID に 対応する実際の囲いを付録Bに掲載した.表におい て, ID が-1 というのは,抽出された囲いの中に棋譜 の囲いが存在しなかった場合に対応する. 本稿では 2回以上表れたものは囲いに採用したため, ID が-1 のものはユニークな囲いである.

勝率の上位は矢倉と美濃囲いであるが,全体とし てまともな囲いである.従ってこの範囲では,勝率 に基づいて囲いの良さに差をつける必要はあまりな 」 いと思われる、勝率を分析するには十分な出現回数 が必要であるため,まずは出現回数が多い囲いを増 やすことが課題である.例えば,9九に玉が出現する 回数 (表1) やその時の勝率 (表2) を考えると, 穴熊は 勝率が良い囲いと予想されるが、このリストには現 れなかった、玉が隅にいるため、他の囲い以上に細 分化が進んで出現回数が200回に満たなかった可能 性が考えられる.実際に玉の位置ごとの囲いの種類 (表4) を見ると, 穴熊 (9九) の囲いが 2千種類と非常 に多くなっている.

さらに,相手の囲いを考慮した時の囲いの勝率を 調べて,表6に掲載した.残念ながら独特の囲いをす ると(ID-1)負けやすいという結果しか得られなかっ た.囲いと囲いの関係をとると,十分な出現回数を 続いて,200回以上出現した囲いを対象に,囲い 得ることが難しいため,囲いと相手玉(あるいは飛

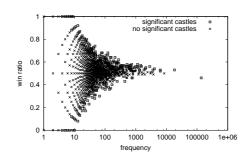


図 4: 囲いの出現頻度と勝率 (5×5 中心の調整なし)

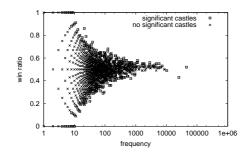


図 5: 囲いの出現頻度と勝率 (3×3)

車)の位置の分析を行うことが有効かもしれない.

#### 5.2 囲いに含める領域の微調整

これまでの結果で,大きさ $5 \times 5$ を全て使うと,囲 いが細分化されすぎてしまい、十分な分析が難しい ことが分かった、そこで、囲いの範囲を狭めること で十分な出現数を確保することを試みた、

#### 5.2.1 囲いの大きさの変更

まず、中心の調整をしなかった場合と、大きさを  $3 \times 3$  にした場合に得られる囲いの性質を報告する . 居る囲いを選びだし , 各マスについて駒の出現頻度 中心の調整をしなければ, 玉が端に近付いた時に囲 いが狭くなるという効果がある.もとの手法も含め, ついて頻度が高い駒から順に左から並べている.つ 3種類の手法で得られたの囲いについて、勝率と出 まり一番左に並んだ駒を集めたものが、もっともよ 現回数の散布図を図3,4,5に掲載する.横軸が出現 回数で縦軸が勝率である.中心の調整をしなかった 場合と大きさを  $3 \times 3$  にした場合のどちらの場合も , ある . 従って , 7八が金か空白かどうかで囲いを分け 狙い通り出現数が多い囲いが増えていて、また勝率 ることが有効であると言える.同様に9六,6六,5 が 0.5 から離れた囲いも増えている.

表 7: 有意差のある囲いの種類と局面の割合

種類		勝率₫	)有意差	囲い無し		
		あり	なし			
$3 \times 3$	種類	643	14099			
	割合	0.32	0.63	0.05		
$5 \times 5$	種類	897	27752			
(調整なし)	割合	0.18	0.53	0.29		
$5 \times 5$	種類	747	28156			
(調整あり)	割合	0.05	0.36	0.59		

す「囲いなし」とはユニークな形の場合である.予 想通り,囲いを小さくしたり中心の調整をしないほ うが,囲いなしと分類される棋譜が少なくなってい る.評価関数に組み込む上では,囲いなしと分類さ れるものは少ない方が良いため,大きな改善である. また勝率に有意差のある囲いの割合も増えている.

以上から,囲いの大きさを小さくすることは有効 であると結論できる.一方で,玉の周囲3マスという のは狭すぎる場合も考えられる.例えば,付録Aの7 八の囲いから 8八角成 同玉と進んだ局面を考え ると,5筋の金銀は囲いからはみでてしまう.そこで, 一律に小さくするのではなく,5×5の領域から不要 なマスを削るアプローチについて,次に報告する.

### 5.2.2 駒の頻度に基づく領域の変更

議論の中で,特徴量に基づく分析の方が,囲いの 細分化の問題もなく有効であろうというコメントを いただいた、そこで特徴量の導入の準備として、発 表後に,囲いに含まれる駒の頻度の分布を調べた.題 材としては,細分化が顕著だった穴熊を用いた.具 体的には,棋譜から抽出した囲いの中で玉が9九に を調べた、結果を表8に掲載する、それぞれのマスに くある囲いである.

表を見ると7八に金が居るかどうかは半々程度で 六の歩をついておくかどうかも場合によるようであ 続いて,勝率に有意差のある囲いの種類と,棋譜 る.一方で,5五,6五など「五」の段はほとんど空 に表れた領域がどの囲いに分類されたかを表7に示 白であり,駒があっても敵の駒である.このような

表 8: 玉が 9九に居る場合 (穴熊) の他の駒の分布

表 6. 立かがたに占る場合 (八衆) の他の側のカガ														
位置	駒	頻度												
9九	玉	7603												
8九	桂	7573	空白	30										
7九	金	5992	空白	1258	銀	337	角	16						
6九	空白	6789	金	810	飛	2	敵駒	2						
5九	空白	7435	金	112	角	32	敵駒	11	銀	6	飛	5	步	2
9八	香	7591	空白	12										
8八	銀	7018	金	328	空白	245	角	12						
7八	金	3961	空白	3596	飛	26	銀	20						
6八	空白	5489	角	977	金	726	銀	360	飛	45	步	6		
5八	空白	6826	金	616	飛	100	步	33	敵駒	18	銀	10		
9七	步	5858	空白	1735	銀	6	香	4						
8七	步	7176	銀	392	空白	33	金	2						
7七	空白	4361	角	2032	銀	901	金	287	桂	18	步	4		
6七	步	3965	空白	1786	金	1781	銀	65	敵駒	6				
5七	空白	4973	銀	1359	步	1153	金	46	角	39	敵駒	18	飛	15
9六	空白	5876	步	1723	敵駒	4								
8六	空白	6642	角	512	步	422	銀	21	飛	2	金	2	桂	2
7六	步	7331	空白	192	敵駒	65	金	13	銀	2				
6六	空白	4331	步	2106	銀	747	敵駒	306	角	74	金	37	飛	2
5六	步	4391	空白	2318	銀	470	敵駒	386	飛	29	金	9		
9五	空白	5241	敵駒	2344	步	18								
8五	空白	7323	敵駒	271	步	5	桂	4						
7五	空白	7518	步	50	敵駒	27	銀	6	角	2				
6五	空白	5379	敵駒	1670	步	524	銀	30						
5五	空白	6378	敵駒	690	步	421	銀	68	角	40	飛	6		

マスは囲いには含めない方が自然である.また,位 囲いの有用性を判定することができた. 置によっては頻度が100に満たない駒も数多くある. これらの駒が囲いを細分化してしまっていると考え られるため,これらの駒を無視し,空白として扱う ことで似た囲いをまとめることが有望である.

## 6 まとめ

将棋プログラムに必要である囲いの知識を棋譜か ら得ることを提案し,実際に24万局の棋譜を分析し た. その結果, 玉の位置により勝率は有意に変わる ことが分かり,また玉の周りの $5 \times 5$ の領域を切り 出すという単純な方法により将棋の主な囲いを抽出 することに成功した. さらに出現頻度と勝率により

今後の課題として,似たような囲いをまとめたり, 囲いとして認識されなかったものの数を減らす必要 があることが分かったが,予備的な実験結果から,囲 いの大きさを使い分けることや,頻度に応じて駒を 無視することで対策できると考えられる.実際に強 い将棋プログラムを作る上では,囲いを組んだり,相 手の囲いを崩す(攻める)手順も重要である.これに ついても,本稿の棋譜の分析を応用し,玉の周囲の 駒の配置の変化を分析することが有効と考えている.

## 議論

沢山のコメントをいただき感謝いたします.

Q. (東京大学 岡野原) 編集距離のように, 例えば歩の 位置が違っても柔軟にマッチングを行なう方が良い のではないか.

A. 今後の課題として取り組んでみたい. 定跡で類似 局面を扱う研究などを参考にできると思われる.

- Q. (岡野原) 出現頻度や勝率だけでは,強い人が使った囲いが必要以上に高く評価される危険がないか. A. たしかに,その可能性は充分考えられる.評価関数に組み込む時には注意が必要である.
- Q. (東京大学 鶴岡) 強い人同士の対局では,勝率に囲いが大きく影響することはないのではないか. (下手な人の場合は囲いで差が出そうだが)

A. たしかに! そうすると, 勝率の高い囲いを探すよりも5割程度の無難な囲いを集める方が現実的か. Q. (東京大学近山)下手な人が途中まで指して,後でうまい人が指した棋譜があると途中の時点での形勢が分かって良いのではないか.また, exact なパターンよりも特徴量で解析した方が面白そう.

A. 特徴量の導入は是非やりたい. 勝率に関しては途中の局面からコンピュータ同士に対戦させることを考えている. しかし, それなりに強いプログラムを作る必要があり, また計算時間もかかるため, まだ実験には到っていない.

Q.(近山)プレイヤーのレーティング等のメタデータ は利用しているか?

A. まだ, 利用できていない.

- Q. (東京大学 横山) 囲いを判定するタイミングは,現在の「成った駒が取られない」というので良いか? A. 早過ぎる場合も遅過ぎる場合もあるようだ.他に良い基準がないか今後検討したい.
- Q. (近山) 急戦と持久戦の区別は囲いの組み合わせを 見ればわかるのではないか?
- A. 今後検討したい.要した手数等いろいろな基準が考えられそうである.
- Q. (田中二郎) プログラムを作ることで将棋は上達したか?

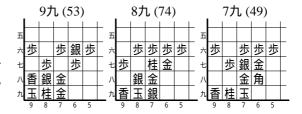
A. 小学校の時から指していないので,分からない. **A** 最近テレビ等を見て勉強を始めている.20年前には 無かった囲いもあって面白い.

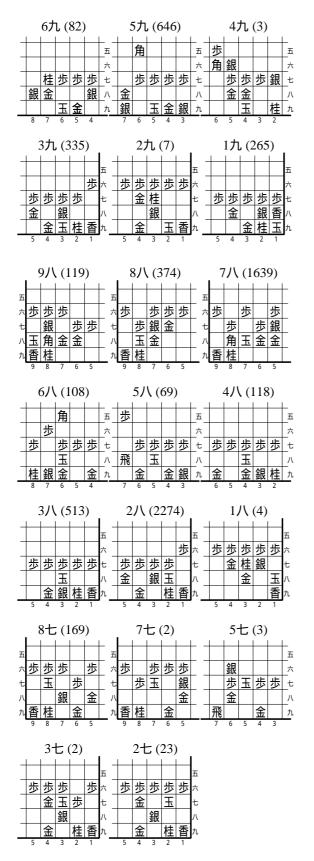
Q. (明大 石畑) 24万局はどの程度強いプレイヤか? A. 強い人は入っているが, 比率はまだ調べていない. 王手に気がつかないような棋譜が約1,000局あるが, 誤差の範囲内と思っている. 投了と詰みがあるかの 逆転は実験の範囲ではそれほど割合が高くない.

## 参考文献

- [1] J. Grimbergen, R. Rollason. Using castle and assault maps for guiding opening and middle game play in shogi. In *The 6th Game Programming Workshop (GPW2001)*, pp. 102–109, 2001.
- [2] Y. Tsuruoka, D. Yokoyama, and T. Chikayama. Game-tree search algorithm based on realization probability. *ICGA Journal*, 25(3):145–153, 2002.
- [3] 柿木. 将棋プログラム K3.0 の思考アルゴリズム. 松原(編), コンピュータ将棋の進歩, 第1章, pp. 1-22. 共立出版, 1996.
- [4] 鶴岡. 将棋. 情報処理, 44(9):900-904, 2003. 特集 ゲーム情報学.
- [5] 鶴岡. 将棋プログラム「激指」. 松原(編),アマ4段を超えるコンピュータ将棋の進歩4,第1章, pp. 1-17. 共立出版, 2003.
- [6] 久米. 将棋倶楽部 24 万局集. ナイタイ出版, 2002.
- [7] 山下. YSS-そのデータ構造, およびアルゴリズムについて. 松原(編), コンピュータ将棋の進歩2,第6章, pp. 112-142. 共立出版, 1998.
- [8] 金沢. 金沢将棋のアルゴリズム. 松原(編),コンピュータ将棋の進歩3,第2章,pp. 15-26. 共立出版,2000.
- [9] 金子, 田中, 山口, 川合. 駒の関係を利用した将棋の評価関数. 第8回ゲームプログラミング ワークショップ, Nov. 2003. (to appear).
- [10] 田中, 副田, 金子. 組合わせて利用可能な高速将棋ライブラリの作成. 第8回ゲームプログラミング ワークショップ, Nov. 2003. (to appear).

## A 頻度が高かった囲い





## B 勝率が偏った囲い

