

将棋における棋譜データベースからの棋士の特徴抽出

登坂 紘介[†], 松原 仁^{††}

[†] 公立はこだて未来大学大学院 ^{††} 公立はこだて未来大学
本研究ではトップクラス棋士である羽生善治氏と他のプロ棋士とを棋譜を元に比較することにより、客観的な棋風の差異を見つけることを試みた。使用するデータは日本将棋連盟より発売されている「将棋年鑑」、「羽生善治実戦集」から約4000局のデータを抽出した。今回の研究では定跡や一般的に評価されているその人の棋風といったものは一切知識として含まず、実際に羽生氏がどのような手を実戦で指したかの統計を取り、他の棋士と統計学上どれほど信頼できる差がどのような項目にあるのかについて調べた。その結果、いくつかの興味深い知見が得られた。例えば、羽生氏は龍の使用頻度が他の棋士に比べて多い、などである。

Feature extraction of players from game records in Shogi

Kohsuke Tosaka[†] Hitoshi Matsubara^{††}

[†] Future University Hakodate

In this paper, we have investigated objective differences of playing style of Shogi by comparing Mr. Habu Yoshiharu's game records with other players' game records. The data are 4000 games extracted from "Shogi Nenkan" and "Habu Yoshiharu Jissensyu". In our comparison, we don't use any Shogi knowledge or any players' characteristics. We have got several interesting results. For example, Mr. Habu plays more Ryu moves than other players.

1 はじめに

近年、ゲームプログラムの成長には目を見張るものがあり、将棋にいたってはアマトップクラスと戦えるほどにもなった。¹⁾ このようなゲームプログラムの成長の背景には高度な探索アルゴリズムと、より効率的なヒューリスティックを与えるための評価関数の開発が大きい。しかし、評価関数はプログラマの経験や一般的に言われている定跡などをもとに作成されていることが多く、どうしてもプログラマの恣意的要素が多く入ってしまう。また、複雑なゲームや必勝法が確立していないゲームでは、相手の取る戦略や癖によってこちらの手もある程度変えなくては行けないが、対戦相手の情報をあらかじめ入力する形式のプログラムというものは少ない。今回の研究では、棋譜ファイルを客観的に評価することにより、将棋に対して新たな知見が発見できるか、また、棋士ごとに差異が見つけれられるのであればそれはどのようなとこ

ろに見られるものであるのかについて統計を用いて調べ、考察を行った。

2 関連研究

ゲーム研究において、棋譜から評価関数の作成をする、強化学習を行うという研究例は多い。将棋においても同じことが言え、コンピュータ将棋ソフト「激指」は羽生善治氏の棋譜データを分析することで指し手による遷移確率を割り出し、そこから評価関数を作成し実現打ち切り確率アルゴリズムという盤面の実現確率によって探索をある程度制約することによって、より効率的な探索を行うことが出来るアルゴリズムを開発した。²⁾ 激指はこのアルゴリズムを搭載することでアマチュア竜王戦でベスト16位に入るなどの活躍をした。³⁾ また、棋譜データベースというものがゲームをプレイする際にどれほど有効であるかについての例としては、IBMが開発したチェス用のスーパーコンピュータ DEEPBLUE が挙げられる。⁴⁾ DEEPBLUE は

膨大な棋譜データベースを用いることで序盤で自分にとって有利な形で試合を進めることが出来るようなアルゴリズムを搭載していた。中盤以降では盤面上の変化が激しくなるためデータベースだけでは想定できないケースが出てくるため、探索による手法を多く取ることとなった。それでも序盤を優勢に進めることは中盤以降の試合運びをより有利に進めることが出来ることに繋がる。実際に DEEPBLUE は膨大な棋譜データベースを用いることによって、探索速度を増すことに成功している。

3 棋譜データベースについて

今回用いるデータベースは日本将棋連盟から発売されている「将棋年鑑 CD-ROM 版」1999～2005 に収録されている棋譜の中で、女流戦、アマ戦を除いたものと、同じく日本将棋連盟から発売されている「羽生善治実戦集 七冠達成の譜」から作成した。総局数は 4037 局で、データの内容は日本将棋連盟が定義している棋譜形式に準じている。データベースでは実際の手順などが書かれているデータファイルとデータファイルを管理するためのマスタファイルの二つに分ける。データファイル、マスタファイルの仕様は以下の通りである。

● マスタファイルの持つデータ

- 対局日
- 対局場所
- 先手番対局者
- 後手番対局者
- 先手番考慮時間
- 後手番考慮時間
- 持ち時間
- 終局までの手数
- 戦型
- 棋譜番号 (データファイルごとのユニークな整数値)

● データファイルの持つデータ

- 棋譜番号
- 指された手の順番 (一手目であれば、1)

- 指された手 (以下の要素により構成)

- * 指された場所 (半角数字で記述、2 六歩であれば「26」と記述)
- * 駒の種類 (ローマ字で記述、歩であれば「fu」と記述)
- * 指すまでの考慮時間

マスタファイル、データファイル共にカンマ区切りのテキストファイル形式である。また、データファイルのユニークな番号はマスタファイルに格納され、これがインデックスナンバーとなる。

以上のマスタファイルを 1 つと、約 4000 のデータファイルを使用することでデータマイニングを行う。なお、この棋譜データベースの構築についてであるが、このデータベースは CD-ROM から一度データ内容を印刷し、そこからアルバイトを雇って手作業で電子データとして入力させたものである。

4 統計

棋譜データベースを人間が眺め、そこから主観的にデータの内容を判断したのでは折角のデータの客観性が失われてしまう恐れがある。そこで、棋譜データベースに対してデータマイニングを行い、具体的にどのようなデータが、羽生氏とその他の棋士の間で有意な差を持っているのかについて調べた。主に統計学を用いることによって、データの中からどのような結果が導き出せるのか、また、導き出された結果が具体的にどれほど信頼に足るのかを統計学を用いて求めた。

4.1 検定

羽生氏がプロとして活躍した期間のプロ棋戦すべての棋譜を、羽生氏の全棋譜とその他の棋士の棋譜という二つの母集団に分ける。この二つの母集団の間に差があるかどうかを、統計的検定 (t 検定) を用いて調べた。なお、検定に用いる抽出標本は今回の棋譜データベースそのものである。本来の棋譜の母集団というものは羽生氏がプロとして長年指してきた全ての棋譜、全てのプロ棋戦の棋譜を指す。この母集団の中から今回は棋譜データベースに当たる部分を抽出し、二つの標本を比べることで羽生氏とそれ以外の棋士の棋戦全てという母集団の二つがどれほど乖離、または一致しているかについて調べた。なお、実際の検定においては棋譜の全データを一度に比較するのではな

く、ある条件を設け、棋譜の中から条件に合う集合を取り出す。例えば条件として「7六歩という手を指した回数」について統計をとるということをするれば、母集団全ての棋譜から「7六歩を指した」という項目を抽出し、その数をカウントする。これを羽生氏の棋譜集合とその他の棋士の棋譜集合に対して行う。このように大きな母集団の中から、特定の項目を羽生氏とその他の棋士の棋譜集合の中から取り出し、取り出した標本を比較することで二つの母集団に対して条件ごとに検定を行う。

4.2 仮説検定

二つの集団の検定には統計的仮説検定を用いる。まず「二つの集合は差異がない」という帰無仮説と「羽生氏の方が条件にあった行動をよくとる（とらない）」という対立仮説を用意する。次に、帰無仮説を棄却するにたる統計量がどれほどかを検証する。そうして出てきた統計量から、実際に帰無仮説が真である場合、標本抽出した標本と同じ位差異のある統計結果がでる確率（P値）はいくらかであるかを検定統計量を基に調べる。求めたP値を、有意水準として定義した統計量 α を打ち切り点とする事で棄却を行う。もしも棄却された場合、同時に成立するような対立仮説を設定し、羽生氏とその他の棋士との差がどのようなものであるかを示す。棄却されなかった場合は「二つの集合に差異は存在しないといえなくはない」（なにも言えない）という結論となる。

5 結果

検定を行った条件は、「駒の出現回数」「指した駒の位置」「終局までの手数」「n-gram 統計を用いた手の組み合わせごとの出現回数」についてである。それぞれの条件に対してデータベースの中から対局者の中に羽生氏が含まれるものを検索し、羽生氏が先手後手のどちらかを見て、羽生氏の手番の手を抽出する。このようにして、その他の棋士もそれぞれの手を抽出していき、最終的に一局に付き二つの指し手データが作られる。

5.1 駒の出現数

羽生氏とその他の棋士の棋譜を比べたところ、指し手における各駒の出現回数は表1のようになった。この総出現数についてそれぞれの駒が有意な

表1 各駒の総出現数

指された駒名	羽生氏の手数	他の棋士の手数
歩	19283	117535
と	642	3543
桂	3971	24591
成桂	181	948
香	1599	10304
成香	92	475
銀	10308	61569
成銀	124	771
金	6657	40326
飛車	6200	36790
龍	1026	5383
角	5881	35027
馬	1185	6841
玉	4816	31436

差を持つかを調べる。統計手法としては、それぞれの母集団（羽生氏の今までの全てのプロ棋戦の集合、羽生氏以外の今までのプロ棋戦で、羽生氏がプロ棋士として活躍していた期間の全ての集合）の平均の差が等しいことを帰無仮説、多く指している（少なく指している）ことを対立仮説として、それぞれの検定統計量（ z ）を求めた。結果は表2の通りである。「 z 」の記号がついているものは有意な差があったものである。

ここで、有意水準を5%として、 $\alpha = 1.96$ とすると、表2から「羽生氏が他の棋士に比べて使用する頻度が等しいとは言えない駒」は「と、香、龍、玉」となる。しかし、これらの駒のうち「と」については抽出標本数が1000以下と少ないので、ここでは「香、龍、玉」が有意な差を持っているものとして考える。これらの駒以外は「羽生氏が他の棋士に比べて使用する頻度が等しいとは言えない」（なにも言えない）ということになり、対立仮説の証明は出来ない。次に帰無仮説の棄却が対立仮説の証明となるような対立仮説を設ける。たとえば、香であれば「羽生氏が他の棋士に比べて香の使用割合が低い」と対立仮説を設ければ、同じように有意水準5%として判断して z が -1.96 より低い（ $P < 0.05$ ）ので、この仮説は証明される。こうして対立仮説の証明をそれぞれお

表 2 各駒の検定統計量(羽生氏 - 他棋士)

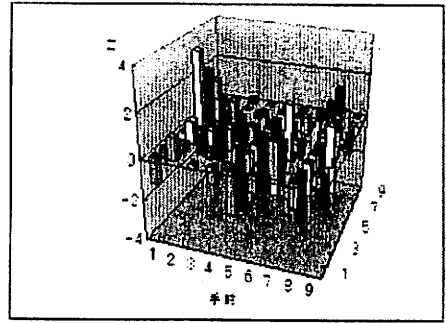
駒名	検定統計量 (z)
歩	-0.890392911
と	2.126394902
桂	-1.316058602
成桂	1.714260247
香	-2.368779849
成香	1.333094264
銀	1.492173233
成銀	-0.267873583
金	0.037267454
飛車	1.611467513
龍	4.063662562
角	1.291714175
馬	1.535177809
玉	-5.140555614

こなうと、最終的に今回の検定では「羽生善治氏と他の棋士との駒の使い方に着目した場合、羽生氏が香、玉の使用頻度が少なく、龍の使用頻度が高いことがそれぞれの駒において 95 %以上信頼できる」という結論となる。

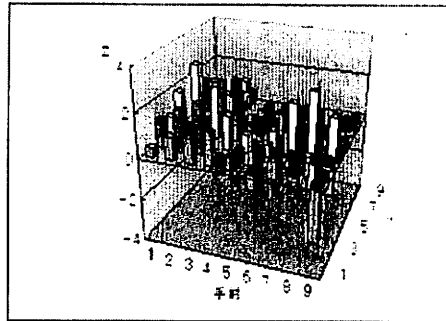
5.2 指した駒の位置

5.1 章において各駒がどれほどの割合で指されているか、またどのような駒が羽生氏とその他の棋士で差異が存在するかについて検定を行った。ここからさらに条件を細かく設定し、各駒がどのような位置に指されることが多いのかについて羽生氏とその他の棋士とを比べる。検定の手法は 5.1 章と変わらない。一例としては「羽生氏が今まで指した全ての手の中で 2 六歩が指される回数」と「他の棋士が今まで指した全ての手の中で 2 六歩が指される回数」について帰無仮説と対立仮説を設けて検定を行い、差が羽生氏とその他の棋士で存在するのかを求める。例のような作業を、母集団が少ない駒(成桂、成香、成銀、と)以外の駒全てにおいて検定を行う。図 1 は羽生氏の駒の使い方が他の棋士のそれと比べてどの位置で多く、どの位置で少ないかを示したものである。図の見方としては、高さが z、手前と書いてあるのが盤面において自分の座っている位置、奥行きが相手までの距離を表す。なお、後手と先手では盤面が入れ替わ

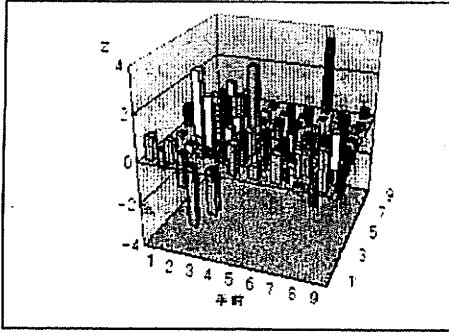
る。そこで、統計を取る際に後手番の盤面の時は盤面を逆向きとして扱うことで、先手番と同じようにした。高さ軸 z が 2 より高ければ、羽生氏は他の棋士と比べてよりその位置に駒を使用し、-2 より低ければ羽生氏は他の棋士と比べてその位置に駒を使用することが少ないといえる。検定統計量はあくまで「羽生氏が他の棋士と比べてその座標に駒を使用することが多い(少ない)と信頼できる目安」であるので、z の絶対値が高ければ高いほど、信頼はしやすくなるが、z が大きいほど(小さいほど)羽生氏がより多く(少なく)その位置に指すということにはならない。



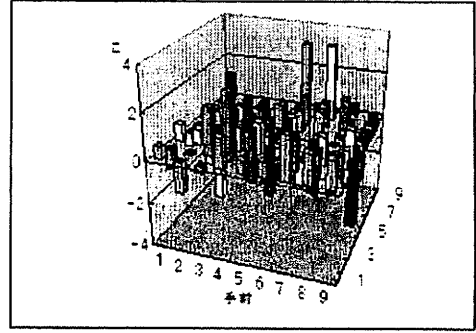
(a) 歩の使用頻度



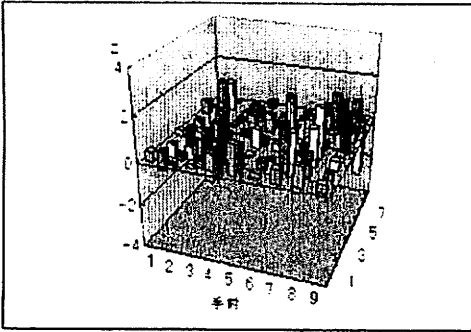
(b) 飛車の使用頻度



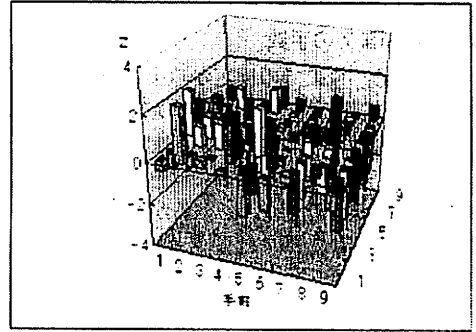
(c) 角の使用頻度



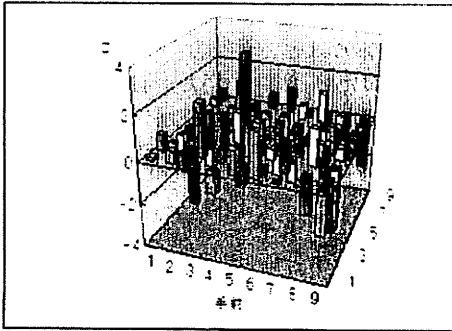
(g) 金の使用頻度



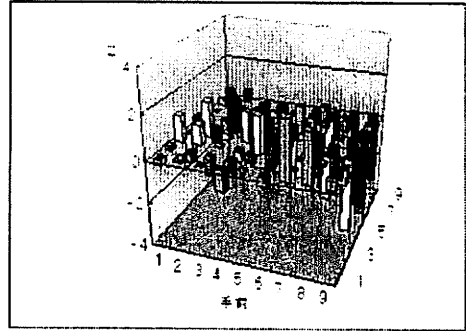
(d) 香車の使用頻度



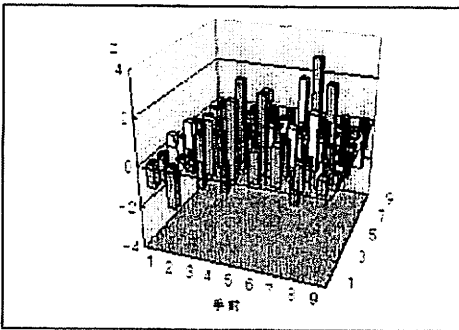
(h) 竜の使用頻度



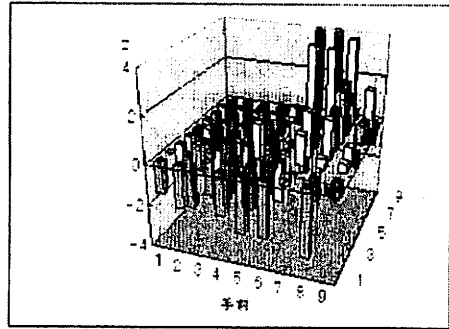
(e) 桂馬の使用頻度



(i) 馬の使用頻度



(f) 銀の使用頻度



(j) 玉の使用頻度

図 1 各駒の使用頻度

各駒ごとに特徴的な指し手が見られる。特徴的なものとしていくつか挙げると、「歩」は左辺中央よりも多く指されており、右辺ではあまり指されていないことが分かる。「銀」については手前側によく指している事が分かる。「玉」については非常に特徴的で、盤面手前側にあまり指さず、盤面相手側に指すことが多いという結果となった。これは、羽生氏が他の棋士に比べてその場所に駒を指すことが多いということであるが、玉自体相手側に指されることが少ないので、羽生氏が他の棋士に比べて少しでも多く指せばその結果が顕著に現れるためであると考えられる。

5.3 終局までの手数

ここでは、今までの検定から少し離れてマスタファイルから羽生氏がどれほどの手数で終局となっているかについて調べる。局数と手数を羽生氏と他の棋士で統計を取ることににより、羽生氏がどれほどの手数で勝負を終えているかについて調べた。ここでは勝敗については要素として含まず、終局までの手数についてのみ検定を行う。なお、千日手、持将棋といったものも検定に含める。あくまで棋戦全てにおける終局までの手数について、羽生氏とその他の棋士とで検定を行うためである。まず、終局までの手数の平均値を求めたところ、羽生氏は110手、その他の棋士は114手であった。これが有意であるかどうかについて調べたところ、 $z = -2.16149855$ となり、「羽生氏は他の棋士と比べて終局までの手数が少ない」ということが95%以上の信頼を持って言える。

5.4 n-gram 統計を用いた手の組み合わせごとの出現回数

羽生氏の指した手、終局までの手数について検定をおこない、羽生氏が具体的にどのような点において他の棋士と差が見つけられたかについて知識発見を行った。ここで、n-gram 統計から羽生氏が指した手の中でどのような手が少ないかについて統計を取る。今までの統計では羽生氏の一手ごとについて統計を取ったものから知識発見を行ったが今回は羽生氏の手をN手ごとに制限して統計をとることによって、一手しか見ることで分らなかった羽生氏の「指し手」についてより詳しく調べることが出来る。

統計手法としては、N手に制約した手から、羽生氏が他の棋士に比べて優位に差がある手の組み合

わせのみを抽出した。今回は有意水準を5%、標本数が100以上のものに限っている。羽生氏が先手番の場合と後手番の場合の2つのパターンにおいてNを2から7の計12通りの統計を行った。 $n = 2$ の結果は表3である。 $n = 2$ を見るだけでnの数を多くしなくても分かることがある。例えば、後手番の場合では「▽3四歩▲2六歩」、「▲2六歩▽4四歩」に対して検定統計量が有意に大きく($P < 0.01$)、その他の棋士よりも羽生氏が多く使用している手であるということが証明されているが、この二つの手の組み合わせには共通項があり、共通項によってこれらの手をさらに組み合わせることで「▽3四歩▲2六歩▽4四歩」という $n = 3$ の手を構成することが出来る。実際に $n = 3$ で実験してみたところ「▽3四歩▲2六歩▽4四歩」の手は $z = 3.459864189$ となり、99%以上の確率で羽生氏が(後手番で)その他の棋士よりも多く指しているということが信頼できる。このようにnの数を増やしていった低次元のnから高次のnを容易に推察することが可能である。nを増やしていくと最終的に一番長い手として「▲2六歩▽8四歩▲2五歩▽8五歩▲7八金▽3二金▲2四歩▽同歩▲同飛」が羽生氏が先手番の場合で $z = -2.152877995$ 、後手番の場合で $z = -2.842090553$ という有意な手として抽出された。この手は定跡と思われるが、羽生氏はこの手を他の棋士よりも使用する頻度が先手、後手問わず少ないということが95%以上信頼出来た。

6 考察

5章において、羽生氏が具体的にどのような点において他の棋士と異なるかについて検定を行った。これらの結果からどのようなことがいえるかについて考察を行う。

6.1 駒の動きと使用頻度から

駒の動きと、使用頻度の二つを比べて分かることとして玉についてがある。羽生氏は玉を他の棋士に比べて動かさないにも関わらず、多くの回数動かさなくては進むことの出来ない敵陣において指される確率が高く、自陣において指される回数が少ない。つまり、羽生氏は入玉の頻度が高いということが言える。また、玉を動かすことが少ないので、この二つと玉は攻めには使わないということを前提条件とすると、羽生氏は玉を逃がす割合が低い、または普段は玉をあまり動かさず、逃げる場合は入玉することが多いということが考察できる。だが、今回の検定ではあくまで「他の棋士に比べて有意な差があるかどうか」についてしか検定を行うことが出来ない（はっきりとした差を言うためには羽生氏がプロ棋士になってからの全プロ棋士の棋譜が必要となる）羽生氏がどれほどの割合で入玉しているのか、また、どれほど他の棋士と比べて玉を動かさないのかについてまでは分からない。

6.2 終局までの手数から

次に、終局までの手数についてであるが、これは他の棋士よりも短い。つまり終局が早いということが証明された。これが羽生氏の実力とどのように関係するのかまではデータから読み取ることが出来ないが、詰めまでの時間が短いのか、詰められるまでの時間が短いのかのどちらかと考えた場合、羽生氏の成績を考えれば前者であることが推察される。よりメタな考察を行えば、これらのことから別の知識発見、または新たな仮説を立てるなどが出来ることが期待される。メタ的な考察とは、定跡や本来言われている羽生氏の指し手の特徴などから、仮説を立てるなどのことである。

6.3 n-gram 統計から

n-gram 統計による羽生氏の指し手についてであるが、今回の統計では定跡という概念も、戦型という概念も含んでいないが羽生氏があまり指さない定跡というものを検出できた。これは、羽生氏

表 3 $n = 2$

有意な手	先手 (x)	後手 (x)
▲7六歩 ▽8四歩	-4.170814	-4.3052032
▽8四歩 ▲6八銀	-4.777335	-4.5691045
▲6八銀 ▽3四歩	-5.1397186	-4.758668
▲5六歩 ▽5四歩	-4.3687881	-3.3285912
▽3四歩 ▲6六歩	2.9248011	有意差なし
▽6二銀 ▲5六歩	-2.1926917	有意差なし
▲4八銀 ▽4二銀	-4.5935145	有意差なし
▲9六歩 ▽9四歩	-2.3547227	有意差なし
▲2四歩 ▽同歩	-3.3282030	有意差なし
▲3五歩 ▽同歩	-2.083220	有意差なし
▲7六歩 ▽3四歩	3.5508095	4.8298281
▽3二金 ▲7八金	-2.2014950	有意差なし
▲2四飛 ▽2三歩	-3.8634441	-3.6270288
▽3二金 ▲2四歩	-2.1354179	-2.8254481
▽2四歩 ▲同飛	-2.1814203	-2.6383207
▽8五歩 ▲7七角	2.4575568	有意差なし
▲2五歩 ▽3三角	有意差なし	2.299052
▲7八金 ▽3二金	有意差なし	-2.7830394
▽3四歩 ▲2六歩	有意差なし	3.221087
▲2六歩 ▽4四歩	有意差なし	3.6512106
▽8六歩 ▲同歩	有意差なし	-2.4522570
▲8六歩 ▽8六飛	有意差なし	-2.3720822
▲2六歩 ▽8四歩	有意差なし	-2.5934290
▽8四歩 ▲2五歩	有意差なし	-2.3991501
▲2五歩 ▽8五歩	有意差なし	-2.7927623
▽8五歩 ▲7八金	有意差なし	-2.6847005

と比較するデータが全ての棋士であることによる。全ての棋士のデータということは様々な戦略を取る棋士のデータが平均されるので、なんらかの信念を持ってプレイをすればそこにデータとの齟齬が生まれ、検定によってその齟齬を抽出することが出来る。

7 終わりに

7.1 まとめ

具体的なデータから客観的な棋風の差異を見つけるために羽生善治氏に対して様々な角度からの分析を行った。その中で羽生氏と他の棋士との間でいくつかの差異を発見することに成功した。ここから羽生氏をモデル化することに特化した評価関数やヒューリスティックの作成を行うことも展望としては考えられる。しかし、人間がこうしたモデルを持った場合、一番使用される可能性が高いのは「羽生氏に勝つ」ために使用することではないかと考えられる。特定の相手モデルから作られた戦略がその相手に対して使用することによりどれほどの効果を出すかまでは分からないが、将棋を解くのではなく相手に勝つためのヒューリスティックを構築することへの第一歩となるのではないだろうか。

7.2 今後の展望

今後の課題としては、今回のデータマイニングの項目抽出自体は恣意的な要素を含んでいたものなので、棋譜に書かれている内容の相関性などに関係なく、全探索的なデータマイニングを行うような機構をもたせて、より客観的に羽生氏が他の棋士とどのような差異をもつのか、また、棋士の棋風を求めることによってどのような新しい評価関数を構築できるのかについて研究を進めていきたい。また今回は「羽生氏と他の棋士」という比較を行ったが、羽生氏自身に対する比較を今後行う予定である。時間経過によって羽生氏の棋風が変化したのか、変化したのであればそれはどのように変化したのかについて調べたい。

参考文献

- 1) 松原仁 編著, コンピュータ将棋の進歩5, 共立出版, (2005).
- 2) 鶴岡慶雅 横山大作 丸山孝志 近山隆, 局面の実現確率に基づくゲーム探索木アルゴリ

ズム, The 6th Game Programming Workshop(GPW2001) pp.17-24,(2001)

- 3) 松原仁 編著, コンピュータ将棋の進歩4, 共立出版, (2003).
- 4) 松原仁 竹内郁雄 編, ゲームプログラミング, 共立出版, (1998).
- 5) 伊藤毅志, コンピュータの思考とプロ棋士の思考, The 10th Game Programming Workshop in Japan 2005(GPW2005) pp.40-47,(2005).
- 6) 大槻知史, n-gram 統計からの「必然手」の抽出, The 10th Game Programming Workshop in Japan 2005(GPW2005) pp.89-96,(2005).