棋譜データの統計解析に基づく勝敗判別手順と感想戦解説方法の提案

A procedure of statistical discriminant analysis of winner of shogi games and a method for illustrating shogi games

華山　宣胤、野上　竜一、鷲津　昂大
HANAYAMA Nobutane, NOGAMI Ryuichi and WASHITSU Akihiro

[抄録]

棋譜は、特殊な質的変量からなる時系列データであるが、その表記方法複雑なもので、将棋の初心者が対局全体の流れを一目で把握することは難しい。そこで本報告では、棋譜データを統計的に分析するために、各局面における駒の配置を特徴づける指標として、歩の置かれている段の平均と分散、歪度、そして金銀桂が動いた数を提案し、勝敗判別分析の結果からこれらの指標が有用であることを示す。また、これからの将棋研究では、人間に勝つためのソフトウェア開発より、エンターテインメントの可能性を広げる研究が注目さえるようになると考えられる。そこで、判別分析から算出される判別得点を用いて、感想戦など既に終了した対局の各局面において、どちらの指し手が優勢であるか、勝負の分かれ目はどこにあったか、を解説する手法を、筆者らが開発したソフトウェアを用いて紹介する。

キーワード：

将棋勝者の判別分析、将棋局面の統計指標、将棋局面の特徴量

[Abstract]

Record of a game of shogi (Japanese chess) is special format of time series data. Because of its complexity, it is not easy for amateur or non-players of shogi to grasp or understand overall shogi games from it. In this study we suggest averages, variances and skewness of row numbers where pawns are put, and numbers of times gold and silver generals and nights moved as indices indicating features of phases of shogi game, and emphasize the usability of them by showing a result of discriminant analysis for winner of shogi game based on actual data for them. Meanwhile, in the recent study of shogi game, aspects of entertainment is received more attention rather than development of computer software to defeat human. So a method and a software for illustrating which player is advancing on every stage and decisive stage of games is suggested.

Keywords:

Discriminant analysis of winner of shogi game, feature of position of shogi game, statistical index of position of shogi game

はじめに

棋譜（将棋の対局者が行った手を順番に記入した記録）は、特殊な質的変量からなる時系列データである。そして、その表記方法は6つの項目（到達地点の筋、到達地点の段、駒の種類、駒の相対位置、駒の動作、成・不成・打）を使用する複雑なもので、対局全体の流れを一目で把握することは難しい。そこで本研究では、棋譜データに基づいて対局全体を統計的に分析するために、各局面における駒の配置（position）を特徴づける指標（feature）として、歩の置かれている段の平均と分散、歪度、そして金銀桂が動いた数を提案する。

将棋の対局分析は、チェスに関する研究と関係を持ちながら、人工知能とソフトウェア開発の分野において盛んに行われている。その中でも、ある局面（対局の途中の状態）でコンピュータが次の一手を決定するために、「評価関数」を用いて評価（指し手、またはコンピュータの優勢・劣勢を判断）する（Hoki and Kaneko、2014）[1)]。しかし、評価関数に用いられる feature は、material のように将棋の対局に関する知識が必要なものや、mobility のように相当な計算を必要とするものが多い。一方、歩の置かれている段の平均や分散、歪度、金銀桂の動いた数は棋譜データから簡単に計算することができる指標である。そこで本報告では、歩の対局中のある局面の position を把握するための feature として、歩の置かれている段の平均と分散、歪度および金銀桂の動いた数を用いることを提案し（第1節）、そして、提案指標を用いた勝敗判別分析の結果を示すことにより、提案した指標が有用であることを示す（第2節）。

これまで、人工知能とソフトウェア開発の分野での将棋研究では、人間（特に名人やプロ棋士）との対戦に勝利できるコンピュータソフトウェアを開発することが主な目的であった（松原　仁、2011）[2)]。しかし、情報処理学会が創立50周年を記念してトッププロ棋士に勝つコンピュータ将棋の実現を目指したプロジェクト（「コンピュータ将棋『あから』強化推進委員会」）（松原、2010）[2)] や Ponanza が目覚ましい成果を上げる（Takizawa、2013）[3)] など、人間に勝つためのソフトウェア開発研究は目標を達成したと考えられる[4)]。そこで今後は、感想戦を分かり易く解説する手法（梶原光輝、2016）[5)] など、将棋エンターテインメントの可能性を広げる研究が注目さえるようになると考えられる。そこで、本研究では、歩の置かれている段の平均と分散、歪度および金銀桂の動いた数を用いた勝敗の判別分析から算出される判別得点を用いて、感想戦など、既に終了した対局の各局面において、どちらの指し手が優勢であるか、勝負の分かれ目はどこにあったか、を棋譜データに基づいて解説する手法を、筆者らが開発したソフトウェアを用いて紹介する（第3節）。

歩の盤面における平均、分散、歪度

本説では、歩の置かれている段の平均と分散、歪度、そして金銀桂の動いた数（「打」を含む）を各局面の position の feature として用いることを提案する。まず、第 i 手（$i=1,2,\cdots$）において盤面に置かれている歩の数を $J_i^{(k)}$（$k=1$ は先手、$k=2$ は後手を表す）とし、第 i 手において盤面上に置かれている先手または後手の各歩の段を $h_{i,j}^{(k)}$（$j=1,2,\cdots,J_i^{(k)}$）とする。このとき、第 i 手における歩の段の平均、分散、歪度、それぞれ $\overline{h}_i^{(k)}$、$\tilde{h}_i^{(k)}$、$\ddot{h}_i^{(k)}$ は式で表される。

$$\overline{h}_i^{(k)} = \frac{1}{J_i^{(k)}} \sum_{j}^{J_i^{(k)}} h_{i,j}^{(k)} \text{、} \tag{1}$$

$$\tilde{h}_i^{(k)} = \frac{1}{J_i^{(k)} - 1} \sum_{j}^{J_i^{(k)}} (h_{i,j}^{(k)} - \overline{h}_i^{(k)})^2 \text{、} \tag{2}$$

$$\ddot{h}_i^{(k)} = \frac{\frac{1}{J_i^{(k)}} \sum_{j}^{J_i^{(k)}} (h_{i,j}^{(k)} - \overline{h}_i^{(k)})^2}{(\tilde{h}_i^{(k)})^{3/2}} \text{。} \tag{3}$$

図1は、歩の置かれている段の平均と分散、歪度を図示している。

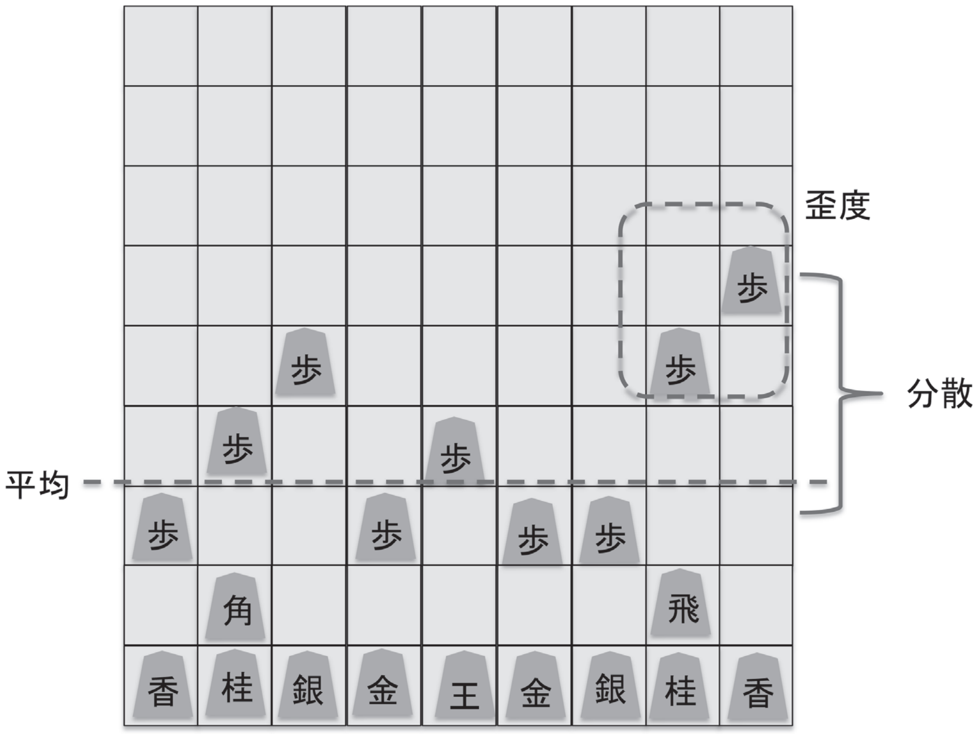

図1：歩の置かれている段の平均

上で定義した歩の段の平均：$\overline{h}_i^{(k)}$、分散：$\tilde{h}_i^{(k)}$、歪度：$\ddot{h}_i^{(k)}$ は各 i（手）について算出される指標であり、一手毎の対局の変化を表すものである、しかし、本研究では、一手々々の各局面におけるfeatureを考えるのではなく、対局全体の把握を目的としている。そこで、先手・後手それぞれ10手を一つの纏まり（フェーズ）と考える。また、第1手から投了までの手数は各対局で異なるため、第1手から第40手（先手後手合わせて80手）までから得られるfeatureを考える。つまり、第1～10手を第1フェーズ、第11～20手を第2フェーズ・・・第31～40手を第4フェーズとし、

$$\overline{p}_l^{(k)} = \frac{1}{10}\sum_{i=1+(l-1)\times 10}^{l\times 10} \overline{h}_i^{(k)} \tag{4}$$

$$\tilde{p}_l^{(k)} = \frac{1}{10}\sum_{i=1+(l-1)\times 10}^{l\times 10} \overline{h}_i^{(k)} \tag{5}$$

$$\ddot{p}_l^{(k)} = \frac{1}{10}\sum_{i=1+(l-1)\times 10}^{l\times 10} \ddot{h}_i^{(k)} \tag{6}$$

$(l=1,2,3,4)$ を一つの対局に対する feature として考える。これらに加え、下に示す「金銀桂」の動いた数（「打」を含む）を一つの対局に対する feature に加える。

$$g_l^{(k)} = count(< r_{1+(l-1)\times 10}^{(k)}, r_{1+(l-1)\times 10+1}^{(k)}, \cdots, r_{1+l\times 10}^{(k)} >, \text{"金"}) \tag{7}$$

$$s_l^{(k)} = count(< r_{1+(l-1)\times 10}^{(k)}, r_{1+(l-1)\times 10+1}^{(k)}, \cdots, r_{1+l\times 10}^{(k)} >, \text{"銀"}) \tag{8}$$

$$n_l^{(k)} = count(< r_{1+(l-1)\times 10}^{(k)}, r_{1+(l-1)\times 10+1}^{(k)}, \cdots, r_{1+l\times 10}^{(k)} >, \text{"桂"}) \tag{9}$$

ただし上式で $r_i^{(k)}$ は先手・後手（$k=1,2$）の棋譜の第 i 文字列、$< s_1, s_2, \cdots >$ は文字列 $s_1, s_2, \cdots$ の連結（concatenation）を、$count(< s_1, s_2, \cdots >, \text{"string"})$ は連結された文字列 $< s_1, s_2, \cdots >$内に含まれ文字列 "string" の数を表す。

提案指標を用いた勝敗判別分析の結果

本節では、第 14 回若獅子戦から 50 対局を選び出し（表 1)、その棋譜データから得られた $\overline{p}_{l,m}^{(k)}$、$\tilde{p}_{l,m}^{(k)}$、$\ddot{p}_{l,m}^{(k)}$、$g_{l,m}^{(k)}$、$s_{l,m}^{(k)}$、$n_{l,m}^{(k)}$（$m = 1, 2, \cdots, 50$）を説明変数として、先手勝利と後手勝利を予測する線形判別関数

$$\begin{aligned} z_m = \sum_{l=1}^{4}\sum_{k=1}^{2} (a_l^{(\overline{p})}\overline{p}_{l,m}^{(k)} + a_l^{(\tilde{p})}\tilde{p}_{l,m}^{(k)} + a_l^{(\ddot{p})}\ddot{p}_{l,m}^{(k)} \\ + a_l^{(g)} g_{l,m}^{(k)} + a_l^{(\mathrm{s})} s_{l,m}^{(k)} + a_l^{(\mathrm{n})} n_{l,m}^{(k)}) + c \end{aligned} \tag{10}$$

の推定を行う。ただし上式で、$a_{l,m}^{(\cdot)}$ は判別係数、 c は定数項、z_m は判別得点（$z_m \geq 0$ ならば先手の勝利、$z_m < 0$ ならば後手の勝利と判別）である。式（10）において説明変数の数は 4 × 6 × 2 = 28 で対局数の 3/5 近くある。そこで、実際の判別係数の推定では $F = 2$を規準とした変数選択を行った。

表 2 は、$F = 2$を規準とした変数選択による判別分析の結果概要を示している。表 2 に示すように、判別的中率は 84% であり、推定された判別関数がある程度の精度を持つことが分かった。

表 1：分析に用いた対局（第 14 回若獅子戦から 50 対局を抽出）

	対戦者	勝手と戦法			対戦者	勝手と戦法	
1	先崎学vs村山聖	先手勝ち	向飛車	26	先崎聖vs羽生善治	後手勝ち	横歩取り
2	森内俊之vs村山聖	先手勝ち	角変わり棒銀	27	村山聖vs小倉久史	先手勝ち	三間飛車
3	先崎学vs小倉久史	先手勝ち	その他の戦型	28	田中功vs羽生善治	後手勝ち	三間飛車
4	木下浩一vs森内俊之	後手勝ち	四間飛車	29	村山聖vs野田敬三	先手勝ち	中飛車
5	中川大輔vs小倉久史	後手勝ち	三間飛車	30	先崎学vs森下卓	先手勝ち	矢倉
6	佐藤康光vs先崎学	後手勝ち	相掛かり	31	小倉久史vs佐藤康光	先手勝ち	中飛車
7	畠山成幸vs村山聖	後手勝ち	矢倉	32	木下浩一vs先崎学	後手勝ち	横歩取り
8	森内俊之vs郷田真隆	先手勝ち	中飛車	33	森内俊之vs小倉久史	後手勝ち	三間飛車
9	木下浩一vs畠山鎮	先手勝ち	か相掛かり	34	阿部隆vs野田敬三	後手勝ち	矢倉
10	佐藤秀司vs先崎学	後手勝ち	三間飛車	35	中川大輔vs森下卓	後手勝ち	相掛かり
11	小倉久史vs丸山忠久	先手勝ち	中飛車	36	田中功vs屋敷伸之	先手勝ち	三間飛車
12	杉本昌隆vs畠山成幸	後手勝ち	四間飛車	37	佐藤康光vs田中大輔	後手勝ち	横歩取り
13	村山聖vs佐藤康光	先手勝ち	角交換その他	38	佐藤康光vs羽生善治	先手勝ち	矢倉
14	佐藤康光vs阿部隆	先手勝ち	矢倉	39	達正光vs中川大輔	後手勝ち	横歩取り
15	村山聖vs高田尚平	先手勝ち	角交換腰掛銀	40	森下卓vs佐藤康光	後手勝ち	矢倉
16	阿部隆vs先崎学	先手勝ち	三間飛車	41	神崎健二vs阿部隆	後手勝ち	矢倉
17	森内俊之vs高田尚平	後手勝ち	矢倉	42	日浦市郎vs羽生善治	後手勝ち	雁木
18	中川大輔vs佐藤康光	後手勝ち	角換わり棒銀	43	佐藤康光vs田中功	先手勝ち	三間飛車
19	畠山成幸vs村山聖2	後手勝ち	角交換腰掛銀	44	村山聖vs神崎健二	後手勝ち	三間飛車
20	阿部隆vs畠山鎮	先手勝ち	相掛かり	45	森内俊之vs中川大輔	後手勝ち	横歩取り
21	高田尚平vs屋敷伸之	先手勝ち	三間飛車	46	達正光vs櫛田陽一	先手勝ち	四間飛車
22	木下浩一vs中川大輔	後手勝ち	三間飛車	47	先崎学vs日浦市郎	後手勝ち	ひねり飛車
23	畠山成幸vs藤原直哉	先手勝ち	四間飛車	48	羽生善治vs井上慶太	先手勝ち	ひねり飛車
24	先崎学vs小倉久史	先手勝ち	矢倉	49	羽生善治vs石川陽生	先手勝ち	四間飛車
25	村山聖vs羽生善治	後手勝ち	矢倉	50	井上慶太vs田中宏樹	先手勝ち	矢倉

表 2：判別分析の概要

判別クロス表

n表

	1	2	全体
0以上	21	4	25
0未満	4	21	25
全体	25	25	50

横%表

	1	2	全体
0以上	84.0%	16.0%	100.0%
0未満	16.0%	84.0%	100.0%
全体	50.0%	50.0%	100.0%

的中度

1	84.0%
2	84.0%

縦%表

	1	2	全体
0以上	84.0%	16.0%	50.0%
0未満	16.0%	84.0%	50.0%
全体	100.0%	100.0%	100.0%

予測度

1	84.0%
2	84.0%

分析精度

判別的中率(%)	84.0%
誤判別の確率(%)	16.7%
マハラノビス平方距離	3.720349
相関比	0.492

また、表３は$F=2$を規準として選択された判別係数の推定値を表している。表中の判別係数の説明と（4）～（9）式との対応は下記のとおりである。

後手銀動いた数４：$g^{(2)}_{4,m}$、後手歩ゆがみ３：$\ddot{p}^{(2)}_{3,m}$、先手桂動いた数２：$n^{(1)}_{2,m}$

後手歩ばらつき２：$\tilde{p}^{(2)}_{2,m}$、先手歩ばらつき２：$\tilde{p}^{(1)}_{2,m}$、先手桂動いた数４：$n^{(1)}_{4,m}$。

これらの結果からは次のようなことが分かる。

- 後手銀の 31 ～ 40 手で動いた数、後手歩の 21 ～ 30 手でのゆがみ、後手歩の 11 ～ 20 手でのばらつき、先手歩の 11 ～ 20 手でのばらつきが大きいと先手が勝つ確率が高くなる。
- 先手桂の 11 ～ 20 手で動いた数、先手桂の 31 ～ 40 手で動いた数大きいと後手が勝つ確率が高くなる。

表３：推定された判別係数

判別式

	判別係数	標準判別	F値	p値	判定
後手銀動いた数4	0.958	0.041	15.60	0.000	[**]
後手歩ゆがみ3	1.538	0.291	3.60	0.064	[ ]
先手桂動いた数2	-20.036	-4.778	10.43	0.002	[**]
後手歩ばらつき2	6.112	1.772	9.64	0.003	[**]
先手歩ばらつき2	4.707	0.490	4.70	0.036	[*]
先手桂動いた数4	-0.539	-0.107	2.28	0.138	[ ]
定数項	-6.703				

将棋の対局解説方法の提案

本節では、判別分析から算出される判別得点を用いた「将棋の対局解説方法」を提案する。判別手得点は各対局に対して算出されるもので、対局中の優劣（「ある時点（手）での先手／後手の優勢／劣勢」を表すものではないが、本研究で判別分析に用いた説明変数は各フェーズ（$l=1,2,3,4$）について算出される値である。そこで、判別得点をフェーズ毎に分けて算出した値、すなわち

$$z_{l,m}=a^{(\bar{p})}_{l,m}\bar{p}^{(k)}_l+a^{(\tilde{p})}_{l,m}\tilde{p}^{(k)}_l+a^{(\ddot{p})}_{l,m}\ddot{p}^{(k)}_l+a^{(g)}_{l,m\ m}g^{(k)}_l+a^{(\mathrm{s})}_{l,m\ m}s^{(k)}_l+a^{(\mathrm{n})}_{l,m\ m}n^{(k)}_l+cl/4 \qquad (11)$$

を「各フェーズにおける先手／後手の優勢／劣勢」を表す指標として提案する。さらに、$z_{l,m}$を歩に関わる値

$$z^{(p)}_{l,m}=a^{(\bar{p})}_{l,m}\bar{p}^{(k)}_l+a^{(\tilde{p})}_{l,m}\tilde{p}^{(k)}_l+a^{(\ddot{p})}_{l,m}\ddot{p}^{(k)}_l \qquad (12)$$

と金銀桂に関わる値

$$z^{(gsn)}_{l,m}=a^{(g)}_{l,m\ m}g^{(k)}_l+a^{(\mathrm{s})}_{l,m\ m}s^{(k)}_l+a^{(\mathrm{n})}_{l,m\ m}n^{(k)}_l \qquad (13)$$

に分けて、これら３つを同時にグラフ化し対局を図２のように示すように解説する方法を提案する。

図２は、1991年10月22日；第14回若獅子戦；決勝；村山聖 vs 先崎学の対局（$m = 1$）を $z_{l,1}^{(p)}$、$z_{l,1}^{(gsn)}$ および $z_{l,1}$ を用いて、解説したものである。図に示されるように、判別得点から、勝敗を決める重要な局面を把握することができる。また、図３～５は提案した解説手法を用いた感想戦解説用ソフトウェアの表示画面を表している。提案した感想戦解説手法を用いて、対局の進行と伴に先手／後手の優勢／劣勢を定量的に示すことができる。

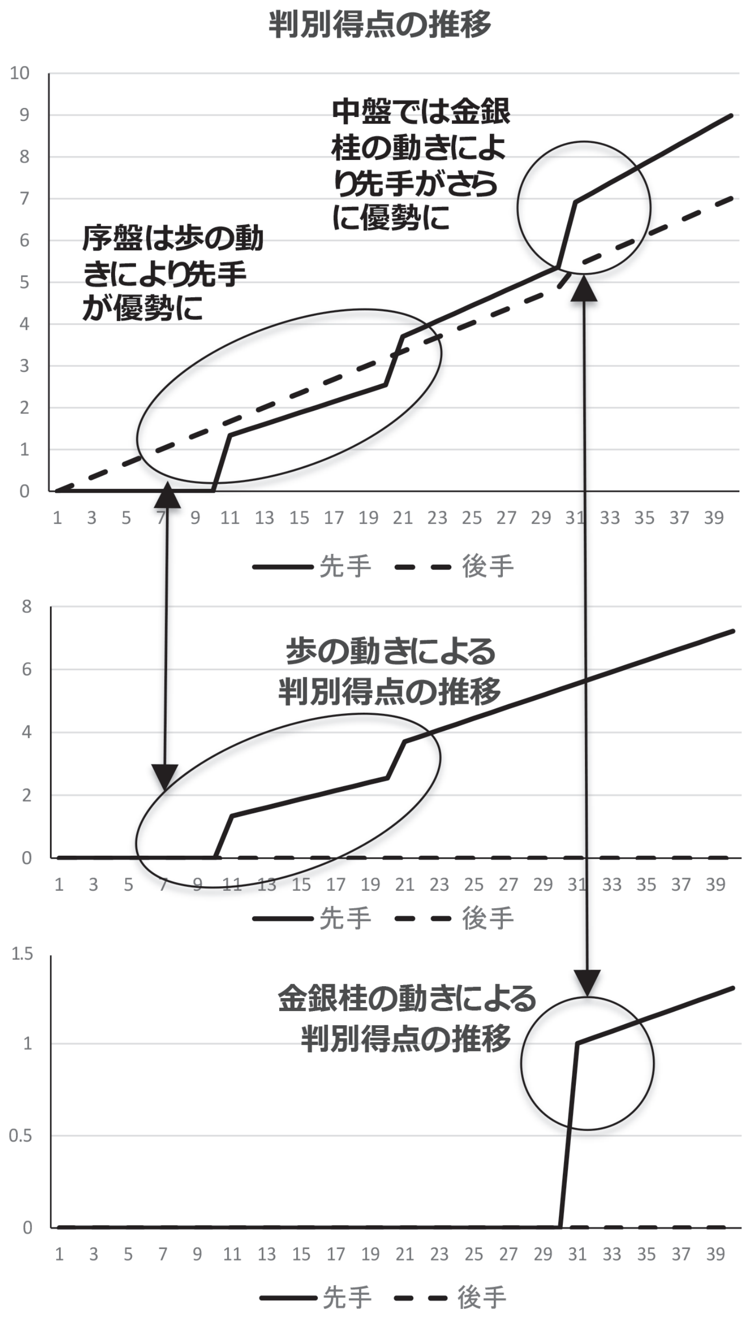

図２：判別得点を用いた対局解説方法

図3:対局の進行と伴に先手／後手の優勢／劣勢を定量的に示すソフトウェアの表示画面(初期画面)

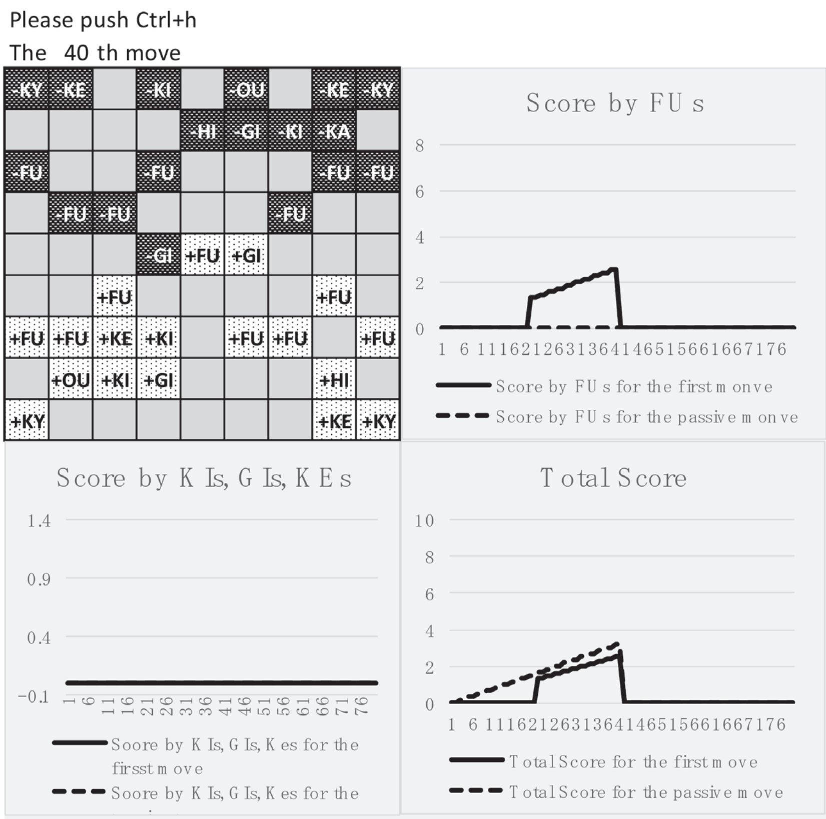

図4：対局の進行と伴に先手／後手の優勢／劣勢を定量的に示すソフトウェアの表示画面（第40手）

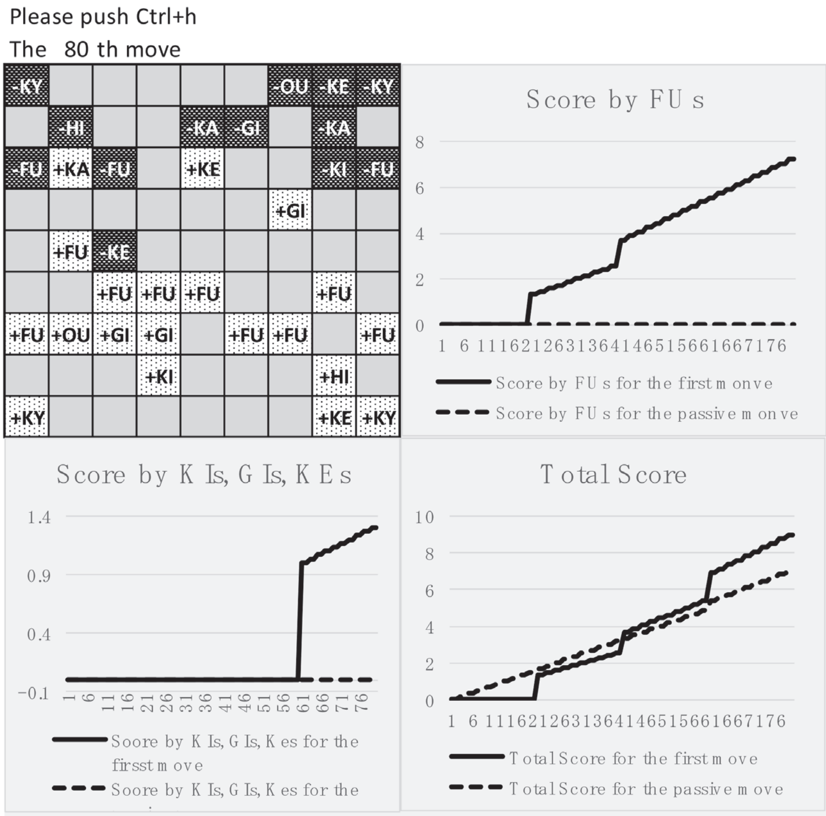

図5：対局の進行と伴に先手／後手の優勢／劣勢を定量的に示すソフトウェアの表示画面（第80手）

むすび

本報告では、「歩の置かれている段の平均と分散、歪度」と「金銀桂の動いた数（「打」を含む）」を対局の情勢を表す feature として提案し、それらの値に基づく勝敗の判別分析の結果から、提案指標が feature として有用であることを示した。また、判別分析から算出される判別得点を用いた「将棋の対局解説方法」を提案し、提案手法を用いて対局を解説するために筆者らが開発したソフトウェアを紹介した。

澤と伊藤（2011）[6] は、将棋において人間がどういった特徴要素に対してプレイスタイルを感じているかを棋譜から抽出した特徴要素を統計的分析によって明らかにしている。本研究ではプレイスタイルではなく、勝敗の判別に注目しているが、彼等が分析に用いた特徴要素には「金将、銀将、歩の使用頻度」が含まれており、この点は本研究の手法との共通点持つ。しかし、本研究では、10 手をフェーズとして金銀桂の動いた数（使用頻度）の時系列としての変化に注目している点が特徴となっている。

本研究の目的は、感想戦などを鑑賞する際に「勝負の分かれ目となった局面」などを分かり易く解説することにあるが、「歩の置かれている段の平均と分散、歪度」と「金銀桂の動いた数」など、単純な特徴で好手、妙手を表現する事は難しい（佐藤、高橋 2012）[7]。そこで、好手、妙手といった勝負の分かれ目を表現するための指標を検討し導入することが、本研究の今後の課題である。

引用文献

1）Kunihito Hoki and Tomoyuki Kaneko. “Large-Scale Optimization for Evaluation Functions with Minimax Search”. *Journal of Artificial Intelligence* 49, 2014, pp. 527-568.

2）松原　仁編. “ あから 2010 勝利への道特集 ”，情報処理，Vol.52，No.2，2011，p.152-190.

3）Takenobu Takizawa. “Contemporary Computer Shogi (May 2013)”. *IPSJ SIG Technical Reports,* Vol. 30, 2013, pp. 1 – 8.

4）小谷善行. “ 第 3 回将棋電王戦を振り返って：3. コンピュータ将棋の棋力の客観的分析 - 人間のトップに到達したか？ -”，情報処理，Vol.55，No.8，2014，pp.851-852.

5）梶原光輝　他. “ 将棋用語による棋譜からの局面検索 ”. エンタテインメントコンピューティングシンポジウム 2016 論文集.

6）澤 宣成，伊藤 毅志. “ 将棋における棋風を形成する要素に関する統計的分析 ”. 研究報告ゲーム情報学（GI）2011-GI-26(3) pp. 1-8.

7）佐藤佳州，佐藤佳州，高橋大介. “ 大規模な対局に基づいた教師データの重要度の学習 ”. 情報処理学会シンポジウム論文集，2012，pp. 22-29.