F-010

麻雀における行動戦略に関する研究
Research on The Strategy in Mahjong

高橋翔悟 †　　**六井淳** †
Shogo Takahashi　　Jun Rokui

1. はじめに

現在、完全情報ゲームや不完全情報ゲームについて様々な研究がなされている [1][2][6][7]。完全情報ゲームとは、対戦相手や自分が行うことのできる行動や行動の順番などすべての情報がそれぞれ分かっているゲームのことであり、将棋や囲碁がこれに当たる [3][7]。不完全情報ゲームとは、完全情報ゲームのように全ての情報が分かることはなく一部の情報が互いに知り得ることのできない未知の情報があるゲームのことで、麻雀やコンストラクタブリッジなどがこれに当たる [1][2][4]。

不完全情報ゲームにおいて、従来の完全情報ゲームで用いられてきた探索などの手法をそのまま同じように適用することは難しい。また麻雀においては行動順が一定でないためゲーム木を作ることも困難である。

麻雀は、不完全情報ゲームの中でも 1 対多不完全情報ゲームであり、1 対 1 不完全情報ゲームよりも対戦相手が多く考慮しなければならない要素が増えるため計算量も多くなる。我々の日常生活で扱う問題では 1 対 1 の関係性だけでなく複数人を相手にすることも多い。また相手の手の内が見えず不完全な情報から推察し問題を解決することもしばしばある。1 対多不完全情報ゲームを研究することは人間の思考過程や行動決定のメカニズムを解く糸口になると考えられる。

従来の麻雀における行動戦略として、統計を用いた研究やサポートベクタマシンによる研究があげられる [1][2]。前者は麻雀のゲームを行った際の行動を記録した牌譜を大量に用意し、ある局面における最適な行動を選択するものである [5]。後者は、麻雀の様々な特徴量を決め、同じように牌譜を用意しサポートベクタマシンを用いて学習を行い、最適な行動決定する [1]。

いずれの手法においても、良い成果は挙げられていない。本研究では、麻雀における行動戦略の新しい手法として遺伝的アルゴリズムと、モンテカルロ法を用いた手法を提案する。

2. 麻雀

2.1. 準備

麻雀では牌と呼ばれる 34 種 136 枚あるものを使用する。これらはマンズ・ピンズ・ソウズ・ジハイの 4 種類に分類される。マンズ・ピンズ・ソウズはそれぞれ 1 から 9 の数字が割り振られており、ジハイは 7 種類ある。これらの牌を組み合わせて様々な形を作っていく。組み合わせをそろえると点数を得ることが出来、最終的に最も点数の多い者が勝ちとなる。また役と呼ばれる決められた形があり、それらをそろえることでその難易度に応じたより多い点数を得ることが出来る。

† 島根大学総合理工学部総合理工学研究科数理・情報システム学専攻, Department of Mathematics and Computer Science Interdisciplinary Facutly Science and Engineering Shimane University

2.2. ゲームの流れ

初めに 4 人それぞれに 13 枚の牌が配られる。これを手牌という。そこから順番に 1 枚牌を取り、1 枚減らしていく。1 枚牌を取ることを自摸と言い、1 枚牌を減らすことを打牌という。自分が打牌する順番のことを打牌局面という。これを約 18 回ずつくり返し牌の組み合わせを作っていく。これを局と言う。この局を 4 回くり返すことを 1 ゲームとし、1 ゲーム行うことを対局という。対局が終わった際の各自の点数を比較し多い順番に 1 位から 4 位の順位が決定される。

2.3. 牌の組み合わせ

牌の組み合わせは 3 つの牌組み合わせが 4 つと 2 つの牌のペア 1 つを作る。3 つの牌の組み合わせはマンズ・ピンズ・ソウズの内、同じ種類で数字が順番に 3 つ並んでいるものか、同じ種類の牌で同じ数字のものを 3 つそろえるかのどちらかである、これを面子という。2 つのペアは同じ種類牌でかつ同じ数字のものを 2 つそろえたものである。

これらをそろえた状態が上がりの形になる。また上がりの形までの距離のことをシャンテン数と言い、上がっている状態を-1 シャンテン、あと 1 つの牌があれば上がりの形になるものは 0 シャンテンとなる。このときシャンテン数を減らすことのできる牌の枚数を有効牌という。

2.4. 和了と放銃

和了とは自分が他の対戦相手よりも早く上がりの形をそろえることをいう。自分の和了した回数を参加した全局数で割ったものが和了率となる。放銃とは自分よりも対戦相手のいずれかが和了し、かつ自分が打牌した牌によって和了を成立させた場合のことをいう。和了率と同じように、自分が放銃した回数を参加した全局数で割ったものを放銃率という。麻雀では、実力を評価する基準の一つとしてレートがある。レートは対局後の順位によりそれぞれ点数が振り分けられている。そこから対戦相手のレートと自分の現在のレートを比較して補正が掛る。

3. 麻雀における遺伝的アルゴリズムの適用

遺伝的アルゴリズムの麻雀への適応として、打牌選択局面における手牌をデータとして与える。そこから初期変異としてランダムに 1 つ牌を選択し他の牌へと突然変異させ初期集団を生成する。初期変異は集団に多様性を持たせるために行う。個体の評価方法としてシャンテン数を用いる。選択方法として評価値の高い上位 2 個体を選択する。交叉の方法として、選択された個体のうち面子を抜き出したものをランダムに複数個交叉させる。遺伝的アルゴリズムを用いて得られた個体集団に対して類似度の最も高い個体を最終的な

解とする。アルゴリズムは図 1 の流れになる。

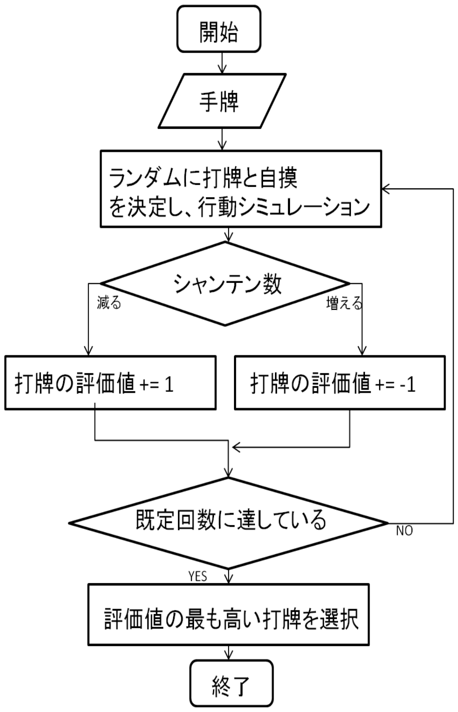

図 1: 遺伝的アルゴリズムのフローチャート

3.1. 類似度

類似度は、与えられた初期データと新たに作られた遺伝子がどれくらい似ているかを表したものである。類似度が良いほど元の手牌と似ていることになるため、類似度が悪い遺伝子よりもその遺伝子の形に変化しやすいということになる。遺伝的アルゴリズムの個体の評価方法としてこの式 (2) を用いる
ある特定の牌 i を自摸する確率 $T(i)$ が、

$$T(i) = \frac{1}{136} \tag{1}$$

で表される。ここからある特定の牌 i を自摸するために必要な順目 $J(i)$ を求める。ある牌 i の枚数を n とすると、

$$J(i) = \frac{1}{n * T(i)} \tag{2}$$

となる。同じ枚数の差でも、小さい時の方がより変化するようになっている。

4. 遺伝的アルゴリズムの検証実験

遺伝的アルゴリズムによる手法を用いて初期データとしてある手牌を与えて実験を行った。評価値として類似度を用い、類似度の値が小さい種が良い評価となる。
初めに初期データを与え遺伝的アルゴリズムを適応させ得られた解を新たに初期データとして与え、複数回遺伝的アルゴリズムを適応させる。複数回適応させる方法として、初期データから遺伝的アルゴリズムを適応させ得られた解を新たな初期データとして複数回適応させる広さ優先の探索と、得られた解を次々と初期データとして適応させていく深さ優先の探索を行う。
実験条件は図 1 となる。

表 1: 実験条件

初期遺伝子数	25 個
世代数	30
突然変異確率	5%
OS	Windows 7 32bit
CPU	Intel Core i3 3.33Ghz
言語	C 言語
コンパイラ	VC++

4.1. 手法による誤差の検証

図 2 は広さ優先探索を行った実験結果である。A の手牌が初期データである。B,C,D の手牌がそれぞれ 3,5,10 の広さでの最も類似度の良い手牌である。広さ優先探索では、探索の深さを 3 とした。
図 3 は深さ優先探索を行った実験結果である。広さ優先探索と同じように、A の手牌が初期データある。B,C,D の手牌がそれぞれの深さでの最も良い値の類似度の手牌である。深さ優先探索では、広さは 3 とした。

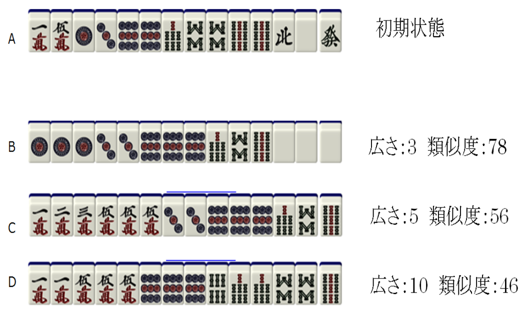

図 2: 広さ優先探索結果

どちらの探索においても広く深くすることで類似度が良くなっている。深さ優先探索では、深さ 10 程度で一度の計算に約 5 分ほど掛ってしまった。麻雀では一回の打牌に数秒程度しか掛けることができないため、実際の対戦において深さ優先の探索を行うことは難しい。遺伝的アルゴリズムは、処理時間の点で麻雀には適さない。

5. 麻雀におけるモンテカルロ法の適用

モンテカルロ法は、乱数を用いてシミュレーションや数値計算を行う手法の総称である [9]。解析的に解くことが出来ないような問題でも、十分な回数のシミュ

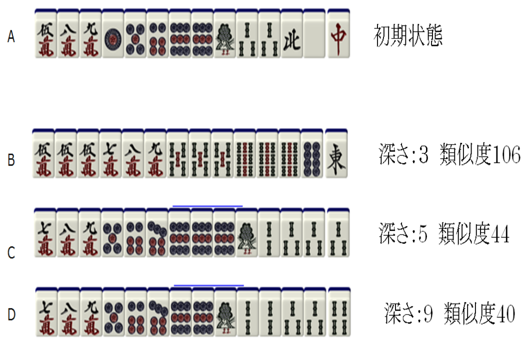

図 3: 深さ優先探索結果

レーションを繰り返すことにより近似解を求めることが出来る。モンテカルロ法は様々な問題に適応することが出来、不完全情報ゲームではポーカーやブリッジ、完全情報ゲームでは囲碁などで広く用いられている手法である。
モンテカルロ法の基本的なアルゴリズムは図 4 のようになる。

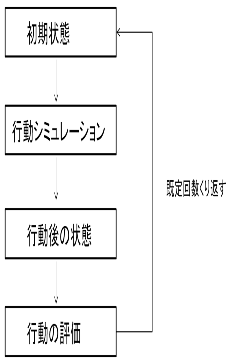

図 4: モンテカルロ法のアルゴリズム

5.1. 評価値増減法

評価値増減法では、打牌選択局面における自分の手牌を初期条件とし、その際に行える行動を乱数を用いてランダムに決定し行動シミュレーションを行う。行動として打牌と自摸を行う。初期状態と行動シミュレーション後の状態を比較し、その際の打牌に評価値を与える。評価の方法として、初期条件と行動シミュレーション後の結果のシャンテン数を比較する。これを既定回数くり返していき、それぞれ行動シミュレーションを行った際の行動の評価値の平均を求め、最も評価値の高い行動を最終的な行動として決定し、実際に行動を行う。
アルゴリズムの流れは図 5 のようになる。

5.2. 評価値変動法

シャンテン数を評価方法として用いる評価値増減法の改良案として、評価方法に更に受け入れ枚数を加えた評価値変動法を提案する。評価値増減法と同じように自分の手牌を初期条件として与え、乱数を用いて

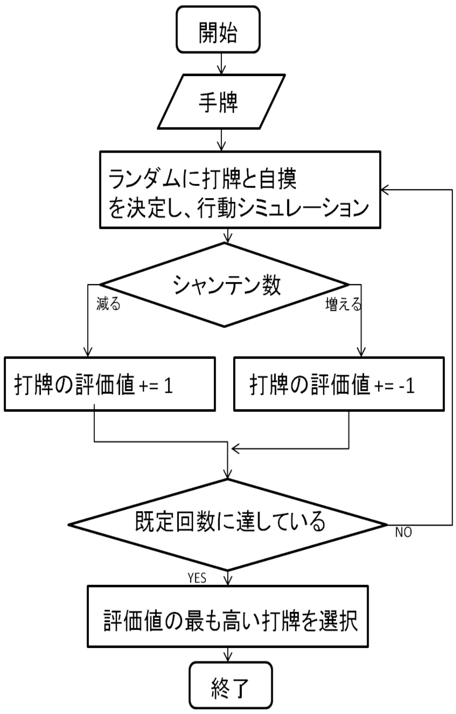

図 5: 評価値増減法フローチャート

行動を決定し行動シミュレーションを行う。この時、初期条件と行動シミュレーション後の状態を比較する際の評価方法として、シャンテン数と受け入れ枚数を用いる。シャンテン数に加えて受け入れ枚数を用いることにより、より価値の高い牌に対してはより評価値を高く与え、より価値の低い牌に対しては評価値を低く与えることが出来る。評価値増減法と同じように行動シミュレーションを行った状態に対して、さらにその状態からシャンテン数の下がる受け入れ枚数を数える。この時初めの行動シミュレーションの際にシャンテン数が下がっていれば打牌の評価値に受け入れ枚数分加える。下がっていた場合は打牌の評価値にその際の受け入れ枚数分を減らす。これを既定回数くり返し、それぞれの打牌の平均を求め最も評価値の高い打牌を最終的な行動とする。アルゴリズムの流れは図 6 のようになる。

6. モンテカルロ法の検証実験

モンテカルロ法を用いた評価値増減法と評価値変動法の検証実験を行った。実験条件は表 2 の通りである。ループ回数は 1 回の試行でのモンテカルロ法内のループ回数である。試行回数は、それぞれの手牌のパターンにおいて実験を行った回数である。手牌のパターンにはそれぞれあらかじめ正解の打牌が決められている。正解の打牌を選択出来た、正解率によりどちらの手法が優れているか比較する。

手牌のパターン別に行った実験結果が図 7、図 8、図 9 になる。上部に表示されている牌が手牌のパターンで、横軸が一度以上選択された牌、縦軸がその牌が

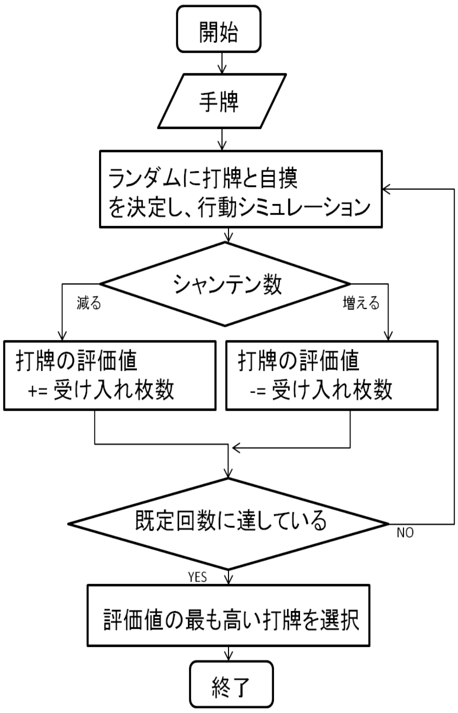

図 6: 評価値増減法フローチャート

表 2: 実験条件

手牌のパターン	3 パターン
試行回数	1000 回
ループ回数	20000 回
用いた手法	評価値増減法・評価値変動法

選択される確率である。丸の付いている牌がそれぞれのパターンにおける正解の打牌であり、この牌の選択される確率が高い方がより良い結果となる。

図 7 は 3 つの異なるパターンの手牌の 1 つ目の実験結果である。評価値増減法と評価値変動法を比較すると評価値変動法が約 30%正しい牌を選択できる確率が向上している。

図 8 は 2 つ目のパターンの手牌で、評価値増減法では選択された牌のほとんどが同じ確率で選択されているが、評価値変動法では、約 80%の確率で正しい牌が選択出来ており、選択される牌の種類も減っている。

図 9 は 3 つ目のパターンの手牌で、評価値増減法では選択されている牌 2 つともがほぼ同じ確率で選択されていたが、評価値変動法では正しい牌を選択する確率が約 30%向上している。

3 つのパターン全てにおいて、正しい牌を選択できる確率が向上している。とくに図 8 のパターン 2 においては、正しい牌を選択できる確率が約 60%向上している。これは、パターン 2 の場合、牌ごとの有効性の差が小さいのでシャンテン数のみの評価ではどれも同じように評価されているためである。評価値変動法のように受け入れ枚数を評価方法とした場合はこれをきちんと評価できており正しく牌を選択することが出来ている。いずれのパターンにおいても評価値増減法よりも評価値変動法の方が良い結果を出している。評価値増減法よりも評価値変動法の方が優れているといえる。

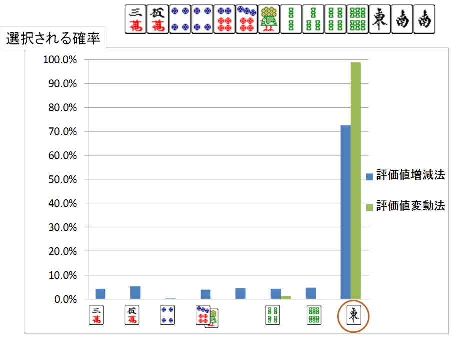

図 7: パターン 1

図 8: パターン 2

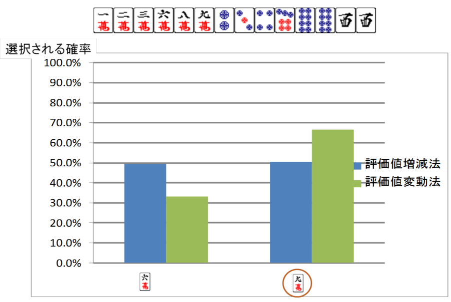

図 9: パターン 3

7. 東風荘を用いた対局実験

東風荘はインターネットを通じて他の人と麻雀の対戦が出来るサービスである [8]。また東風荘用入出力ライブラリを用いることで東風荘の画面から牌の情報を読み取ったり、東風荘上の機能を使用することが出来る。このライブラリを使用してプログラムに麻雀を

打たせる。その際の成績をプログラムの評価に用いる。用いる手法はモンテカルロ法による評価値増減法を用いて 483 試合、評価値変動法を用いて 2502 試合行った。図 10 は、実験を行った東風荘の実際の対局画面である。

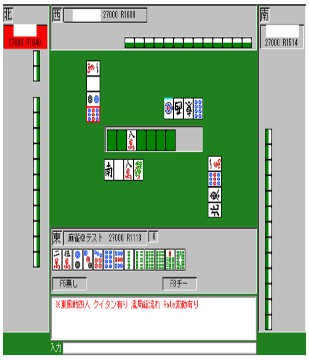

図 10: 実際の実験の様子

7.1. 必要な試合数

麻雀だけでなく他のゲームにおいても少ない試合数では実力が正しく反映されないことがある。また麻雀は見えない情報に左右されることが多く、最善手を選んだとしても必ず良い結果になるとは限らない。

i を麻雀の取り得る順位とし 1 位 2 位 3 位 4 位のいずれかの順位を取るとする。またそれぞれの順位を取る確率を $P(i)$ とすると平均順位 J と分散 A は以下の式になる。

$$J = \sum_{i=1}^{4} i * P(i) \tag{3}$$

$$A = \sum_{i=1}^{4} P(i) * (i - J)^2 \tag{4}$$

ここで平均順位と分散より試合数を n とすると、標準偏差 σ は式 (3)、式 (4) になる。

$$\sigma = \sqrt{\frac{A}{n}} \tag{5}$$

この時有意水準を 95%とすると、試合数 n は式 (6) を満たせばよい。

$$2\sigma < 0.05 \tag{6}$$

この時、1 位から 4 位までの順位をそれぞれ均等な確率で取るとすると分散 A は 1.25 となる。この分散 A を用いると σ は

$$2\sigma = 2\sqrt{\frac{A}{n}} < 0.05 \tag{7}$$

となり計算すると

$$\frac{5}{n} < 0.0025 \tag{8}$$

$$2000 < n \tag{9}$$

つまり試合数 n は最低 2000 試合必要となる。

7.2. 対局実験

対局実験の条件は表 3 になる。ワイワイ卓は誰でも参加できる場所である。その際の対戦相手の平均レートは 1483 出会った。レートは中央値が 1500 である正規分布に従っている。

表 3: 実験条件

実験場所	東風荘・ワイワイ卓
対戦相手平均レート	1483
用いた手法	評価値増減法 評価値変動法
対戦数	評価値増減法:483 試合 評価値変動法:2502 試合

図 11 は実験を行った結果の評価値増減法と評価値変動法それぞれの平均順位の遷移である。横軸が試合数、縦軸が平均順位である。評価値増減法では、483 試合行い最終的な平均順位が 2.767 となった。評価値変動法では、2502 試合行い最終的な平均順位が 2.674 となった。評価値変動法では、試合数が多くなるにつれてある一定の値に収束していることが分かる。
図 12 は実験を行った結果の評価値変動法の各順位 1 位から 4 位の順位分布の遷移である。最終的な結果は 1 位率が 19.8%、2 位率が 23.2%、3 位率が 27.1%、4 位率が 29.7%となった。図 11 の平均順位と同じように試合数が多くなるにつれてある一定の値に収束しようとしていることが分かる。
図 12 は実験を行った結果の評価値変動法の和了率と放銃率の遷移である。最終的な結果は和了率が 20.7%、放銃率が 18.2%となった。和了率と放銃率は 1 試合に 1 つしかとれない順位とは異なり、1 試合中で複数回和了や放銃の機会があるので平均順位よりも収束が早く、約 1000 試合ほどでほぼ収束している。

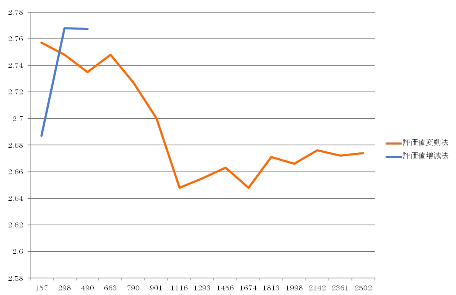

図 11: 平均順位遷移

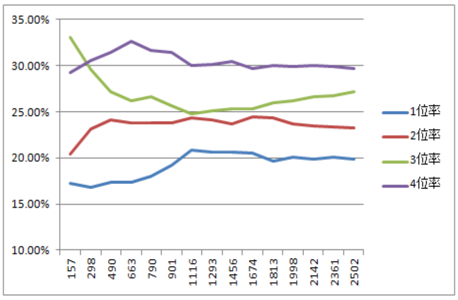

図 12: 評価値変動法：各順分布

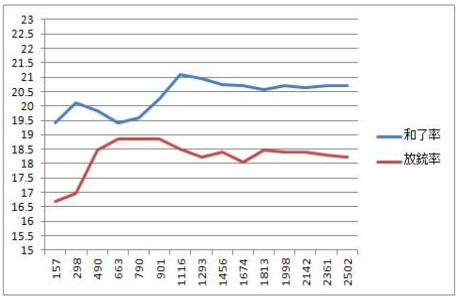

図 13: 評価値変動法：和了放銃率遷移

評価値増減法と評価値変動法を用いて東風荘にて対局実験を行ったが、評価値増減法では 483 試合で評価値変動法では 2502 試合と試合数に大きく差があり単純には比較することが出来ない。そこで式 (3)、式 (4)、式 (5) より実験結果より得られた平均順位を比較する。式 (3)、式 (4)、式 (5) より、それぞれの平均順位を求めると以下のようになる。

評価値増減法平均順位：2.716〜2.818
評価値変動法平均順位：2.651〜2.697

これより評価値増減法の平均順位の最も良い値 2.716 と評価値変動法の平均順位の最も悪い値 2.697 を比較すると評価値変動法の平均順位 2.697 の方が良い値となっている。評価値増減法よりも評価値変動法の方が優れているといえる。

8. まとめ

麻雀における行動戦略において遺伝的アルゴリズムを用いた手法とモンテカルロ法を用いた手法を提案した。モンテカルロ法を用いた手法ではシャンテン数を評価方法に用いた評価値増減法と受け入れ枚数を用いた評価値変動法を提案しそれぞれを比較した。またインターネット雀荘「東風荘」において対局実験を行った。実験を行った結果評価値増減法を用いた場合平均順位が 2.767、評価値変動法を用いて場合平均順位が 2.674 となった。これは東風荘全体でみると下位約 12%にあたる。行動の評価方法として自分の手牌のシャンテン数と受け入れ枚数のみを評価しており、相手の行動を評価していない。また麻雀は対局の終了時点で点数が多い人が勝ちである。よって対戦相手との点数の差を行動の評価に組み込むことは重要である。これらを考慮することにより、より良い行動が出来るようになり、平均順位の向上にも繋がると考えられる。
今後の課題として、点数の状況に応じた対応の変化や試合の序盤と終盤による行動の変化、相手の行動に対する対応なども考慮した行動戦略を考える必要がある。

参考文献

[1] 三木理斗, 近山隆, 三輪誠:木カーネルを用いた SVM による麻雀打ち手の順位学習 , 情報処理学会シンポジウム論文集. NC, Vol.2008 No.11 Page.60-66 (2008.10.31)

[2] ：北川竜平, 三輪誠, 近山隆:麻雀の牌譜からの打ち手評価関数の学習 , 情報処理学会シンポジウム論文集 Vol.2007 No.12 Page.76-83 (2007.11.09)

[3] 鬼沢武久, 風見覚, 高橋千晴:不完全情報ゲームプレイングシステムの構築 スタッドポーカーを例にして, 映像知能と情報 Vol.15 No.1 Page.127-141 (2003.02.15)

[4] 小田和友仁, 上原貴夫:コンピュータブリッジにおける他者のモデルを考慮したゲーム木探索の提案, 情報処理学会論文誌, D-II, 情報・システム, II-情報処理 Vol.47 No.11 Page.3005-3016 (2006.11.15)

[5] 伊藤毅志, とつげき東北:思考ゲーム 牌譜の解析による麻雀の分析 , 人工知能学会誌. Vol.24 No.3 Page.355-360 (2009.05.01)

[6] 池畑望 , 伊藤毅志, :Ms.Pac-Man におけるモンテカルロ木探索 , 情報処理学会論文誌ジャーナル. Vol.52 No.12 Page.3817-3827 (2011.12.15)

[7] 林伸也 , 浦晃, 近山隆 , 三輪誠 , 田浦健次朗 , :自己対戦棋譜を利用した半教師あり学習による将棋の評価関数の学習 , 情報処理学会シンポジウム論文集. Vol.2011 No.6 Page.143-149 (2011.10.28)

[8] インターネット雀荘「東風荘」 http://mj.giganet.net/

[9] 津田考夫 :モンテカルロ法とシミュレーション 電子計算機の確率論的応用, 中央出版 (1997.12.10)