

多人数性を分割した教師付き学習による 四人麻雀プログラムの実現

水上 直紀^{1,a)} 中張 遼太郎^{1,b)} 浦 晃^{†1,1,c)} 三輪 誠^{†2,2,d)} 鶴岡 慶雅^{1,e)} 近山 隆^{1,f)}

受付日 2011年11月4日, 採録日 2011年12月1日

概要: 本論文では, 牌譜を用いた教師付き学習による麻雀プログラムの作成法について述べる. まず, 上級者の牌譜を用いたパーセプトロン学習によって一人麻雀プレイヤーを作成し, それを拡張することによって四人麻雀への適用を行う. 拡張は, 一人麻雀プレイヤーに「降り」と「鳴き」の機能を教師付き学習によって導入することで行った. オンライン麻雀サイト「天鳳」で作成されたプログラムの実力を評価した結果, レーティングとして, 平均的な人間プレイヤーの実力を大きく上回る 1651 点が得られた.

キーワード: 多人数ゲーム, 麻雀, 教師あり学習

Realizing a Four-Player Computer Mahjong Program by Supervised Learning with Isolated Multi-Player Aspects

NAOKI MIZUKAMI^{1,a)} RYOTARO NAKAHARI^{1,b)} AKIRA URA^{†1,1,c)} MAKOTO MIWA^{†2,2,d)}
YOSHIMASA TSURUOKA^{1,e)} TAKASHI CHIKAYAMA^{1,f)}

Received: November 4, 2011, Accepted: December 1, 2011

Abstract: This paper describes a supervised machine learning approach for building a mahjong program. We start with building a one-player mahjong program by Perceptron learning with game records of expert human players, and adapt it to four-player mahjong. The adaptation is achieved by incorporating the “folding” and “calling” functionalities that are separately learned from game records. We have evaluated the playing strength of the resulting program on a large online mahjong site “Tenhou”. The program has achieved a rating of 1651, which is considerably higher than that of the average human player.

Keywords: Multi-player game, Mahjong, Supervised learning

1. はじめに

近年, 2 人零和確定完全情報ゲームであるチェス, オセ

ロ, 将棋といったゲームでは人間のトッププレイヤーと同等かそれ以上の実力を持つコンピュータプレイヤーが実現されている [1], [2], [3]. 一方, ポーカーや麻雀をはじめとする不完全情報多人数ゲームでは, 人間のトッププレイヤーを上回る実力のコンピュータプレイヤーが実現できている例は非常に少ない [4], [5], [6], [7].

代表的な不完全情報ゲームのひとつである Texas Hold'em (ポーカーの一種) では, オフラインで近似ナッシュ均衡戦略を計算しておくことで, 2 人ゲームであれば人間の世界チャンピオンを上回る実力のコンピュータプレイヤーが実現可能であることが示されている [8]. また, 2 人ゲームで得られたナッシュ均衡戦略を 3 人ゲーム用に拡張する方法 [4] やプレイヤーの行動を削減, 抽象化することでより少

¹ 東京大学

The University of Tokyo

² マンチェスター大学

The University of Manchester

^{†1} 現在, 富士通研究所

Presently with Fujitsu Laboratories Ltd

^{†2} 現在, 豊田工業大学

Presently with Toyota Technological Institute

a) mizukami@logos.t.u-tokyo.ac.jp

b) nakahari@logos.t.u-tokyo.ac.jp

c) ura@logos.t.u-tokyo.ac.jp

d) makoto-miwa@toyota-ti.ac.jp

e) tsuruoka@logos.t.u-tokyo.ac.jp

f) chikayama@logos.t.u-tokyo.ac.jp

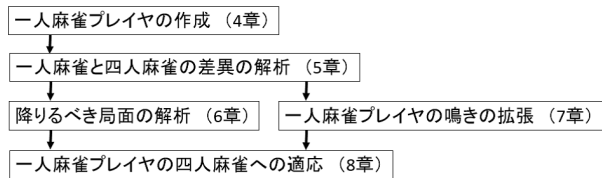


図 1 提案手法

Fig. 1 Proposed method

ない人数の少ないゲームを想定する方法 [5] などが提案されている。

本論文が対象とする「麻雀」は、4人零和不確定不完全情報ゲームである。コンピュータによる麻雀プレイヤを実現する手法として、これまでに、4人用の牌譜をそのまま学習したプレイヤ [6] やモンテカルロ木探索を用いてすべてのプレイヤの行動をシミュレートするプレイヤ [7] などが提案されているが、実力は人間の平均的プレイヤに大きく負け越すレベルである。

本論文で提案する麻雀プレイヤの実現法を図 1 に示す。まず、(四人) 麻雀から多人数性を排除した一人麻雀のためのプレイヤを、エキスパートの牌譜を用いた教師付き機械学習によって作成する。次に、作成された一人麻雀プレイヤの、四人麻雀における弱点を解消するため、「降り」と「鳴き」を実現する機構を導入する。最後に、四人麻雀用に拡張された一人麻雀プレイヤの実力を、オンライン麻雀サイトでの人間との対戦によって評価する。

本論文は以下の構成になっている。初めに 2 章で麻雀のルールと用語、3 章で関連研究を述べる。提案手法として、4 章で一人麻雀プレイヤの作成、5 章で一人麻雀と四人麻雀の差異の解析、6 章で降りるべき局面の解析、7 章で一人麻雀プレイヤの鳴きの拡張、8 章で一人麻雀プレイヤの四人麻雀への適応について述べる。最後に 9 章でまとめについて述べる。

2. 麻雀のルールと用語

麻雀は 4 人のプレイヤが互いの得点を奪い合うゲームである。

局の開始時には各プレイヤは配牌と呼ばれる 13 枚の牌を持ち、その後、ツモと呼ぶ 1 枚牌を引く行為と 1 枚牌を捨てる行為を各プレイヤが順番に行う。手持ちの 13 枚に、ツモで得た牌もしくは相手の捨てた牌を加えた 14 枚の牌の組み合わせに役 (特定の構成) がある時、プレイヤは、和了 (ホー) し、役に応じた点数を得ることができる。和了に関する用語としては、ツモして和了することもツモと呼び、相手の捨てた牌で和了することをロンと呼ぶ。またロンされることを放銃と呼び、得点をすべて放銃したプレイヤが払う。麻雀の 1 試合は、配牌から和了までを 1 局として、1 局を特定の回数 (典型的には 8 回) 行い最終的な点数の大きさを競う。

和了に必要な 14 枚の組み合わせを完成させるためには、メンツと呼ばれる特定の 3 枚の組み合わせが 4 組と、対子 (トイツ) と呼ばれる 2 枚の同一牌が 1 組必要である。プレイヤはメンツを作るために他のプレイヤの捨てた牌を利用することができる。これを鳴きまたは副露 (フーロ) と呼ぶ。ターツとはあと 1 枚でメンツになる組み合わせであり、鳴くことでメンツを一組作ることができる。

牌は全部で 34 種類あり、それぞれが 4 枚あるため合計は 136 枚となる。牌には数字の 1~9 のいずれかが書かれた数牌と文字の書かれた字牌がある。数牌は萬子 (マンズ)、筒子 (ピンズ)、索子 (ソーズ) の 3 種類に分けられる。この 3 種類に優劣はない。字牌は 7 種類あり、東、南、西、北の総称である風牌と白、発、中の総称である三元牌に分けられる。

また以下に一般的な麻雀用語を説明する。

リーチ 鳴きを一度も行わず、あと 1 枚で和了できる手牌になった時宣言出来る行為である。リーチ宣言の後は手牌を変更することはできない。すなわち、リーチ後にツモで入手した牌は、和了できる場合を除きすべてそのまま捨てなくてはならない。

七対子 (チートイツ) 同じ牌が 2 つという組み合わせを 7 組そろえることでできる役

国士無双 各色の 1, 9 牌と字牌をすべて 1 枚以上集める役

聴牌 (テンパイ) 和了に必要な牌の枚数が 1 枚の状態

向聴数 (シャンテンすう) 聴牌に必要な牌の枚数

役牌 3 つ集めると役になる牌

ドラ 1 局の始めにランダムに決定する牌。和了した時にこの牌があれば得点が増加する

山 同じ配牌、ツモになる牌の並び

降り 自分の和了を諦め、相手の和了牌を捨てないようにする行為

回し打ち 和了を目指した最善手を選ぶのではなく、放銃を避けつつ和了や聴牌を狙う行為

現物 相手がすでに捨てた牌。ルール上、現物をロンすることはない。

筋 現物の数字の 3 つ外側にある牌

壁 数字の連結がなくなった牌

染め 自分の手牌を同じ色で揃える役の総称

タンヤオ 各色の 1, 9 牌と字牌を一枚も使わない役

カン 鳴きの一種。同じ牌を 4 枚集めた時に行える行為で、これにより場のドラが 1 つ増える

チー 鳴きの一種。連続した 3 の数字の組み合わせであるメンツを作る

ポン 鳴きの一種。同じ牌 3 つ集めたメンツを作る

暗刻 同じ牌をポンすることなく集めたメンツ

3. 関連研究

ナッシュ均衡戦略を多人数ゲームに適用する研究として

はポーカーを用いた研究 [4], [5] がある。ポーカーは 2 プレイヤであれば近似ナッシュ均衡戦略を用いることで世界チャンピオンに勝利している [8] が、人数が増えると状態数が指数関数的に増大するためナッシュ均衡戦略の計算は困難になる。そこで 2 人ゲームで得られたナッシュ均衡戦略を 3 人ゲーム用に拡張する方法 [4] や、プレイヤの行動を削減、抽象化することでより少ない人数のゲームを想定する方法 [5] が提案されている。しかし、ナッシュ均衡戦略から外れる弱いプレイヤに対し搾取することができないことや、状態を抽象化しすぎて現実のゲームに適していないなどの問題がある。

麻雀に関しては、北川らが 3 層ニューラルネットワークを用いた麻雀プレイヤを提案している [6]。牌譜中の何を捨てるかという局面と鳴きの局面を自分の手番の合法手のみ探索し、3 層ニューラルネットワークで次の局面に対する手を学習させている。対人戦での挙動は、牌の組み合わせが適切に学習されていないため、和了する技術や、相手の和了を防ぐ技術が低いという問題がある。

三木らはモンテカルロ木探索を用いた麻雀プレイヤを提案している [7]。この手法では、麻雀のゲーム木の探索は探索の分岐数が膨大であり、正確に行うのは困難なため、相手の手牌や行動をランダムでシミュレートするモンテカルロ木探索 [9] を用いている。麻雀の知識をほとんど使わないにもかかわらず、和了に必要な牌の枚数を下げるように打つという単純なルールに基づいたプレイヤよりも成績が上回る結果となった。しかし、相手はシミュレート時にほとんど和了できないため、役を作ることが困難になる鳴きをするというように、人間が行うプレイとは大きく異なる挙動を示した。

上記のいずれの先行研究においても、作成されたプレイヤは東風荘 [10] の中で成績が下位 13% ほどの実力であり人間の平均レベルのプレイヤに達していない。

ゲームに機械学習の技術を適用した代表的な研究としては以下が挙げられる。オセロではディスクの形を特徴量とし棋譜の結果（勝敗）から特徴量の重みを計算する [1]。またバックギャモンでは TD 学習 [11] が用いられている。これは強化手法のひとつであり、結果がない局面でも親の局面の評価値を子の局面の評価値に近づけるように学習する手法である。しかし、麻雀は最善手を選択しても和了できないことがあるため、これらの手法を直接用いることは難しい。

4. 一人麻雀プレイヤの作成

本章では四人麻雀に拡張するための土台となる一人麻雀プレイヤを説明する。

4.1 一人麻雀

四人麻雀では成績の良いプレイヤほど、平均和了点が低

く和了率が高いことが統計で明らかになっている [12]。上手いプレイヤの平均和了点が低くなる最も大きな理由は、手役を無理に狙った打ち方をしないためである。麻雀の点数の特性上、ある一定以上の点数の役は難易度の割には点数が高くないため、比較的点数の低い役で多く和了することが多くの場合効率が良い。また、自分が和了することによって、相手の和了する機会を減らすという意味でも和了率の高さは重要である。

本研究では、平均和了点より和了率を重視したプレイヤ作成するため、一人麻雀の評価の際には和了したか否かのみで判断する。また以下のようなルールの一入麻雀を考える。

- 配牌とツモのみでプレイする
- 相手がいないため、相手の捨て牌はない
- 鳴きとリーチはない

4.2 平均化パーセプトロンによる一人麻雀プレイヤ

この節では一人麻雀プレイヤを作成方法について述べる。本研究では、平均化パーセプトロン [13] を用いた教師付き機械学習により一人麻雀プレイヤを作成した。パーセプトロン学習には、依存関係がある特徴量を大量に用いた学習を、比較的小さな計算コストで行うことができるという特長がある。

以下、 \mathbf{x} を局面の特徴ベクトル、 \mathbf{w} を重みベクトル、 n をベクトルの次元とする。 x_i は i 番目の特徴ベクトルの値、 w_i は i 番目の重みベクトルの値とすると、ある与えられた局面のスコアは式 (1) で計算される。

$$f(\mathbf{x}, \mathbf{w}) = \sum_{i=1}^n x_i w_i \quad (1)$$

14 枚の手牌から 1 枚切った 13 枚の手牌の特徴量を求め、それぞれについてスコアを求める。次に牌譜で実際に切られた手 t^* とスコアの最も高かった手 \hat{t} を比較する。それらが異なる場合、 t^* の特徴ベクトル \mathbf{x}_{t^*} と \hat{t} の特徴ベクトル $\mathbf{x}_{\hat{t}}$ を用いて、新しい重みベクトル \mathbf{w}' を式 (2) で計算する。

$$\mathbf{w}' = \mathbf{w} + \mathbf{x}_{t^*} - \mathbf{x}_{\hat{t}} \quad (2)$$

この更新を牌譜のすべての局面に対し繰り返す行い、毎局面の重みベクトルの平均値を最終的な重みとする。

特徴量としては北川らの論文 [6] を参考に設計した特徴量を用いた (表 1)。特徴ベクトルの次元は 37,514 となった。パーセプトロン学習における学習データ全体の走査回数は 3 回とした。

4.3 学習に用いる牌譜

教師データとしてはインターネット麻雀サイト天鳳 [14] の鳳凰卓の牌譜を用いた *1。鳳凰卓でプレイできるのは全

*1 2009 年 2 月 20 日から 2010 年 1 月 31 日までに行われた対局

特徴量	次元数
通常手, 七対子, 国士無双の向聴数	$15 + 7 + 14 = 36$
4 色のうち n 色 ($n=1\sim3$) の牌を 0 にしたときの n と向聴数	$3 \times 16 = 48$
複数の色 ($1\sim4$) と各色の向聴数 ($0\sim11$ まで) の組み合わせ	$12 + 78 + 364 + 1365 = 1819$
色の中で最も多い色の数	15
色の中で最も多い色の数+字牌の数	15
2 から 8 牌の数	15
メンツと 1,9 を含まない連続した 2 牌のターツの合計とメンツ候補の組み合わせ	$6 + 12 + 72 = 90$
各字牌の枚数と役牌かどうかとドラの組み合わせ	$5 \times 3 \times 2 = 60$
数牌の数字-5 の絶対値と枚数とドラの組み合わせ	$5 \times 5 \times 2 = 50$
各色の 1~9 の有無の組み合わせ	512
連続する n 種類の数牌の持っている枚数の組み合わせ ($n=2\sim6$)	$100 + 500 + 1860 + 8634 + 23760 = 34854$

表 1 一人麻雀プレイヤーの特徴量

Table 1 Features for the one-player Mahjong player

プレイヤーの中でも上位 0.1% 程度であり, 牌譜の質は高いと考えられる. ただし, この牌譜には, 四人麻雀に特有の局面と手が多く含まれており, このデータを一人麻雀の学習データとして直接用いるのは不適切である. そこで, 各局において初めてリーチをしたプレイヤーを確認して, そのプレイヤーの指された手のみを教師データとする. また局面は配牌からそのプレイヤーがリーチをかけるまでの局面とした. 最終的な教師データの数約 170 万局面となった.

4.4 一人麻雀の実力

学習が適切に行われているかを検証するため, 得られた一人麻雀プレイヤーの実力を評価した. 他のプレイヤーの影響をなくするため, テストはプレイヤー 1 人で行う. また運の要素はなるべく減らす必要があるため, 複数の一人麻雀プレイヤーの実力を同じ山を用いて比較した. テストセットとして異なる山を 100 個用意し, 各プレイヤーについて打った結果を評価した. プレイヤーは「上級者 (人間)」と「平均プレイヤー (人間)」を各 1 名と一人麻雀プレイヤー, そして先行研究 [7] で用いられた Plain UCT である. 平均プレイヤーは天鳳 [14] の中で成績が上位約 50% に位置するプレイヤーであり, 上級者は本論文の第一著者であり, 麻雀サイト天鳳 [14] において成績が上位 0.1% に位置するプレイヤーである.

結果を表 2 に示す. 一人麻雀プレイヤーは平均プレイヤー以上の和了率を達成している. 先行研究が平均プレイヤーにも大きく届かなかったことを考慮すると, 一人麻雀ではあるものの特徴量の構築と学習が適切に行われたため, 大きく性能が向上しているといえる.

この一人麻雀プレイヤーを小松ら [15] による一人麻雀プレイヤーと比較する. ただし小松らは, ドラを使用し, 結果を点数で評価しているため, 本研究とは設定がやや異なっている. 小松らの研究の提案手法は現在の手牌に和了できるまで牌を追加し, 和了点を牌に追加し必要でない牌を選択するという方法である.

プレイヤー	和了率 (%)	平均和了順目
上級者 (人間)	17	13.8
一人麻雀プレイヤー	13	12.7
平均プレイヤー (人間)	12	12.8
Plain UCT	2	15.0

表 2 一人麻雀の和了率

Table 2 Winning rates in one-player Mahjong

プレイヤー	和了率 (%)	1 局あたりの平均点数
一人麻雀プレイヤー	20.4	733
小松らの研究 [15]	14.0	850

表 3 1 万局の和了率と平均点数

Table 3 Winning rates and average scores in 10,000 games

山は異なるが 1 万局での和了率と 1 局あたりの平均点数の比較を行った. 結果を表 3 に示す. 1 局あたりの平均点数では小松らの研究に劣るものの和了率では上回っており, 我々の研究における一人麻雀においては牌譜を用いた学習が成功していることを示している.

5. 一人麻雀と四人麻雀の差異の解析

前章で, 一人麻雀に関しては, 牌譜を用いた機械学習によって人間の平均プレイヤーに近い実力を持つプレイヤーが実現できることを示した. 本章では, 一人麻雀と四人麻雀の差異を解析するため, 四人麻雀の牌譜をもとにプレイした手と牌譜の手の違いを調査する.

前章で実現した一人麻雀プレイヤーは「鳴く」ことができない. 実際の牌譜の和了した局面について 80 万局^{*2}を調査したところ, 鳴くことなく和了した割合は 52% であり, 鳴いて和了した割合は 48% であった. この結果から鳴きができないことの影響は極めて大きいと考えられるものの, 本章ではツモ時の局面についての一人麻雀と四人麻雀の差異を解析する.

具体的には以下のように行った. 最初に実際の牌譜の局

^{*2} 2010 年 1 月 1 日から 2010 年 12 月 31 日までに行われた対局

プレイヤー	Rank 1	Rank 2	Rank 3
上級者 (人間)	0.62	0.85	0.93
一人麻雀プレイヤー	0.53	0.77	0.85

表 4 牌譜との一致率

Table 4 Concordance rate

面に対して一人麻雀プレイヤーが切りたい牌の上位3つを選択する。牌譜で実際に切られた牌が、この上位3つの中に入らなかった局面を調査し、「降り」や「回し打ち」といったタグを人手で付与し、それらの回数を集計する。この集計結果から一人麻雀と四人麻雀の差異を解析する。

まず一人麻雀プレイヤーの四人麻雀における牌の選択精度を評価した。ここでプレイヤーが指定した第 n 番目までの候補に牌譜での打牌が含まれる割合を Rank n とする。比較として上級者も評価を行った。

テストに用いる局面は、天鳳 [14] で公開されている鳳凰卓の牌譜からランダムに抽出した 27 局から 1,342 局面を選択した。一人麻雀であるため鳴くことができる局面や鳴いた後の局面は 1,342 局面の中に含まれていない。一人麻雀では相手の捨て牌や副露した牌などはないが、テスト中ではそれらの牌は残りの枚数を数えるのに使用した。上級者 (人間) は四人麻雀を行うように手を選び、一人麻雀プレイヤーは一人麻雀としてプレイした。結果を表 4 に示す。

上級者の Rank 1 の正解率は 62% 程度で一人麻雀プレイヤーは 53% であった。テストセットは異なるものの Rank 1 の正解率として 56% の精度が報告されている。

次に一人麻雀プレイヤーの Rank 3 までに正解が含まれなかった局面、つまり残り 15% の 193 局面を調査し手動でタグ付け・分類を行った。その結果を図 2 に示す。外した局面のうち約 4 割が降りるべき局面である。また全体としては「役牌」、「七対子」、「染め」、「タンヤオ」など役が絡んだ局面が多い。その他には一人麻雀プレイヤー自体の手が悪手を打った局面もある。

タグについて説明する。役に絡んだタグは一人麻雀プレイヤーがその役を目指せなかった局面であり、「ドラ」は切る必要のないドラを切った局面である。「分類不可」はどれを切っても同じような牌が複数あり 3 つの候補のみでは選べない局面や牌譜自体が悪手を打った局面である。

この解析により、降りが一人麻雀を四人麻雀へと拡張するため一番大きな問題であることがわかり、降りを予測できれば、15% のうちの 4 割である 6% の牌譜との一致率の向上が予想される。

6. 降りるべき局面の認識

前章での解析により、降りる局面を正しく認識できれば牌譜一致率を上げ、実力が向上することを示した。そこでこの章では降りる局面を認識するための局面のタグ付についてまず説明し、次にその提案手法について説明する。

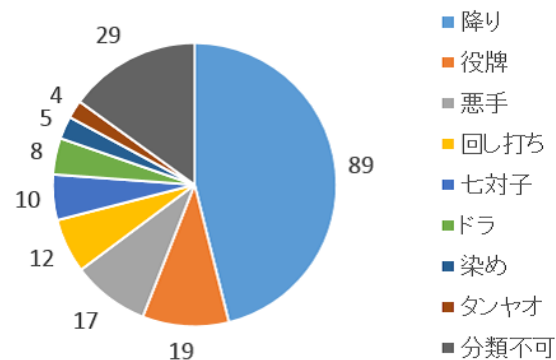


図 2 外した局面の分類

Fig. 2 Classification of errors



図 3 降りた局面

Fig. 3 A position in which the player folded

6.1 教師データの作成

教師あり学習として降りる局面を認識するにはその教師データが必要になる、しかし牌譜に「降りる」とは明示的に記されていない。したがって人間の手でタグ付けをする。タグとして降りかそうでないかなが必要になる。

具体例を図 3 に示す。図の局面はプレイヤー (画面下部) が 3 萬を切った局面である。いわゆるメンツの中抜きでありこの手の和了の可能性を諦めた手である。タグ付けにおいては一度降りるとその局では以降すべて降りたものとする。

この定義により牌譜から 509 局を通じ、5,716 局面に降りか否かのタグを第一著者 (A) が手動でつけた。すなわち牌譜の選ばれた手を見て、降りかどうかのタグをつけることになる。

このタグ付けの妥当性を検証するため第二著者 (B) が第一著者とは独立に、5,716 局面の一部である 1,053 局面にタグ付けを行った。第二著者は麻雀サイト天鳳 [14] において成績が上位 10% に位置するプレイヤーである。

*2 天鳳の開発者である角田氏に許可を頂いた。

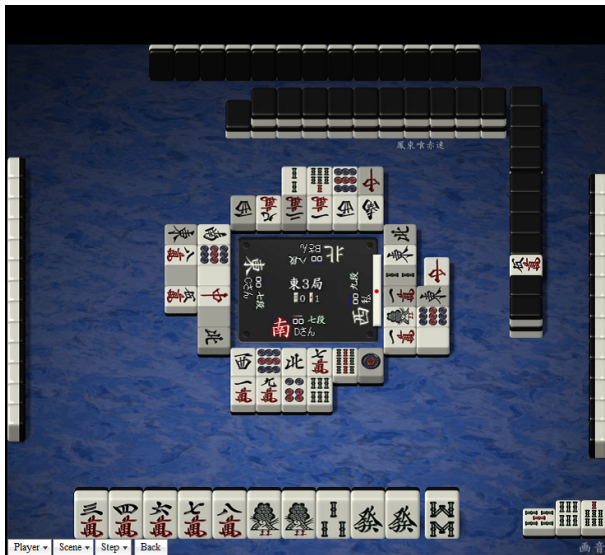


図 4 回し打ちした局面

Fig. 4 A position in which the player “Mawashi uti”

A/B	降り	そうでない
降り	83	23
そうでない	25	922

表 5 タグ付一致数

Table 5 Number of correct answers

両者のタグ付けの一致率を表 5 に示す。判断にはプレイヤーの主観的な考え方が大きく作用するため、困難な問題である。しかしながら一致率は一定の水準に達している。偶然によらない一致率の指標であるカッパ係数 [16] は 0.75 であり、おおむね降りの局面とそうでない局面を当てられているといえる。一致しなかった 48 の局面を見ると「回し打ち」の局面が 47 あり、残りの 1 つは牌譜が明らかな悪手を指した手であった。

回し打ちの具体的な例を図 4 に示す。他のプレイヤーのことを気にしなければ最善手は 8 索であるが、右のプレイヤーのリーチに対応して現物である 1 索を切ったというような局面である。

また最善手ではないと判断できても、それが降りなのか回し打ちなのか判断するのが困難な局面もあった。図 5 は二人の判断が分かれた局面である。ここで牌譜では 2 萬を切った。これをどう見るかであるが、メンツを崩しているという点で A は降りと判断した。しかし、見方を変えたと 2 萬や 6 萬は右のプレイヤーのリーチに現物であり、4 萬は筋で通りそうである。これらを切っている間に残りの形がメンツになることや危険な牌がリーチ者が切るかもしれないため、このプレイヤーはまだ加点することを諦めていないだろうという点で B は回し打ちと判断した。

このように手が和了や流局時の聴牌を狙えるかどうか判断するのはそのプレイヤーの主観的な考え方が大きく作用する。とくに七対子で和了できる可能性をどの程度見るかは



図 5 降りか回し打ちか判断に困る局面

Fig. 5 A position in which judging whether to fold or “Mawashi uti” is difficult

プレイヤーによってさらに幅が大きく困難である。

また一度降りと判断した場合、定義により次からすべて降りと判断するため、降りと回し打ちで判断が食い違った場合、その次のツモ牌が安全牌であるとき片方は降り、片方は降りてないと判断される。このように一度判断を間違えるとさらに間違いが引き起こされ一致率が低下する。

6.2 降りるべき局面の認識

本節では、降りる局面の認識を機械学習によって行う手法を説明する。まず局面からどのような特徴量を取り出したのかを説明し、認識精度を評価する。

本節の問題設定は 6.1 節での設定とは少し異なっている。6.1 節で行ったのは牌譜で何を選んだかを見てその手が降りかどうかを判断した。しかし実際には切る牌を選ぶ前に降りかどうかを判断して、その判断に基づいた手を選択する必要がある。したがって 6.1 節で行った問題とは異なり、局面についての降りかどうかを判断する分類問題とした。

降りかどうかを判断する局面の特徴量としては表 6 に示すものを取り出した。自分の状態と 3 人分の相手の状態と場の状況を考慮する必要があるため、特徴ベクトルの要素次元は 34 ($7 + 3 \times 8 + 3$) である。

この研究では問題を単純にするため、降り判断は以下のようなモデルで行われていると考える。人間は降りるべき局面かどうかを決定したときに周りを気にしないときの最善手を選ぶ。そしてこの牌が周りのプレイヤーに対して安全かどうかを考慮して、降りるかどうかを判断している。そのためこのモデルに必要な切りたい牌の周りのプレイヤーに対しての安全度を考える。

安全度はヒューリスティクスにより決定した。安全度の高い順に以下のようにになっている。

	項目	次元数
自分の状態	向聴数	1
	七対子に必要な牌の枚数	1
	リーチしているか	1
	親かどうか	1
	副露数	1
	副露により晒したドラの数	1
	持っているドラの数	1
相手の状態	リーチしているか	1
	副露数	1
	副露により見えたドラの数	1
	親かどうか	1
	切りたい牌のそのプレイヤーに対しての安全度	4
場の状況	山の残りの枚数	1
	カンの数	1
	捨て牌にあるドラの数	1

表 6 降り判断の特徴量

Table 6 Features for folding decisions

	降りる局面	そうでない局面
適合率	0.76	0.97
再現率	0.71	0.97
F 値	0.73	0.97

表 7 降りる局面の分類結果

Table 7 Result of classification of folding positions

- 現物
- 1 枚以上切れている字牌または筋の 1, 9 牌または壁
- 字牌または上記以外の筋
- 無筋

6.3 降りる局面認識の予備実験

6.2 節の特徴量の妥当性を検証するため、降りる局面認識の 2 クラス分類を行った。降り判断の学習に使用できる学習データは比較的少量である。そこで本研究では、マージン最大化により高い汎化性能を得られる、Gaussian カーネルによる Support Vector Machine (SVM) を分類器として用いた。

評価は 5 分割交差検定で行った。SVM の実装として LIBSVM [17] を用い、グリッドサーチによってガウシアンカーネルで用いるパラメータのコスト (c) とガンマ (g) を最適化した。グリッドサーチの範囲はそれぞれ 2^{-10} から 2^{10} まで値を 2 倍刻みで行った。

学習データ中の降りる局面数は 531, そうでない局面は 5,185 であった。結果を表 7 に示す。表中の適合率とはシステムが出力した予測が実際に正解であった割合である。表中の再現率はテストデータの正解をどれだけ予測してきたかという割合である。適合率と再現率の調和平均が F 値である。グリッドサーチの結果、 $c = 2^8$, $g = 2^{-7}$ の時が最も降りる局面での F 値が高い。

この結果より比較的高精度で降りを予測できることが分かった。

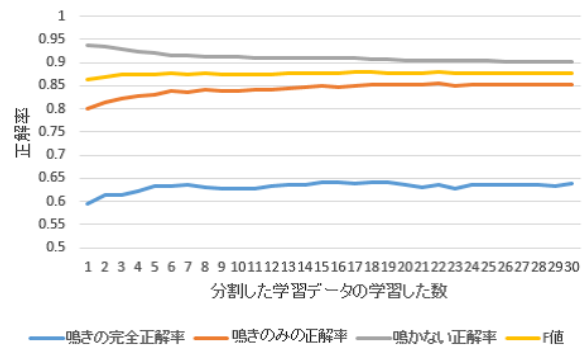


図 6 学習曲線

Fig. 6 Learning Curves

7. 一人麻雀プレイヤーの鳴きの拡張

本章では一人麻雀プレイヤーを鳴きができるように拡張する。またその鳴きの判断の正確さを牌譜をもとに確かめた。

7.1 平均化パーセプトロンによる鳴き局面の学習

鳴きについての学習も、4.1 節で行った平均化パーセプトロンを用いる。鳴いた後に切ることができる牌は、すでに完成しているメンツを鳴く行為（例えば 123 から 1 または 4 をチーして 1 を切る）を禁止するルールがあるため、ツモ局面と異なる。そのため鳴きとツモ局面の学習との相違点は、鳴ける局面の時に鳴かなかった手牌のスコアと鳴いて 1 枚切った手牌のスコアを比較するという点である。特徴量は 4.1 節に用いた特徴量に鳴きに関して以下の特徴量を追加する（表 8）。

特徴ベクトルの次元は 69,258 になった。パーセプトロン学習における学習データ全体の走査回数は 1 回とした。

7.2 鳴きの学習に用いる牌譜

4.1 節で使われたデータはリーチしたプレイヤーの局面を教師データとした。しかし、これでは鳴いた局面が含まれないため、鳴きの学習用の教師データを 4.1 節で用いたものから鳴きに対応できるように変更する。各局において初めてリーチをしたプレイヤーがリーチをかけるか、あるいは和了したプレイヤーが和了するまでの局面を教師データとした。最終的な教師データの数約 300 万局面となった。またテストデータ数は鳴きができる局面に限定し約 3 万局面となった。開発データとして 3,000 局面用意した。教師データ数を 30 分割し、ひとつ使用されるごとに開発データに対する正解率を観測した。

7.3 鳴きの正解率

鳴きの学習が適切に行われているかを検証するため、開発データに対する学習曲線とテストデータに対しての正解率を調査した。その結果を図 6 と表 9 に示す。

特徴量	次元数
副露数	5
リーチができる可能性があるか	2
各色の1~9の各数牌の数を0, 1, 2以上で変換したパターン	19472
副露した中から選んだn(1~2)個の副露した種類の組み合わせ	$136 + 9316 = 9452$
手牌の同じ牌が3枚以上の数と2枚以上の数の組み合わせ	$5 \times 7 = 35$
手牌の同じ牌が3枚以上の数+ボンの数と手牌の2枚以上のものとの組み合わせさらにチーをしたか	$5 \times 7 + 1 = 36$
手牌と副露の中のタンヤオ牌の数と向聴数	$16 + 16 = 32$
連続する3つの数字の各色の有無とその数字に1, 9が含まれとチーを含むかの組み合わせ	$512 \times 2 \times 2 = 2048$
チーを含む各色の1~9の数字の有無	512
色の中で最も多い色の数+同じ色または字牌の副露 $\times 3$	16
色の中で最も多い色の数+字牌の数+同じ色または字牌の副露 $\times 3$	16
特定の色と字牌以外を0枚としたときの向聴数	16
2つそろった役牌の数	8
3つそろった役牌の数	6
ドラの数(0, 1, 2, 3以上)または2~8牌のドラの数	$4 + 4 = 8$
副露数と向聴数の組み合わせ	$5 \times 15 = 75$
向聴数と役があるまたは条件付きの役または役がないかと順目の組み合わせ	$2 \times 3 \times 19 = 114$
役があるか条件付きの役かどうか役がない	3
1, 9牌を鳴いたかと1, 9牌を加えた時に向聴数が何枚減るか(0, 1, 2以上)	$2 \times 3 = 6$
手牌の1, 9牌の枚数を0にしたときに向聴数が何枚増えるか(0, 1, 2以上)	3
風牌4種類それぞれの副露を加えた手持ちの枚数(0, 1, 2, 3以上)の組み合わせ	$4 \times 4 \times 4 \times 4 = 256$
三元牌3種類それぞれの副露を加えた手持ちの枚数(0, 1, 2, 3以上)の組み合わせ	$4 \times 4 \times 4 = 64$
副露数と向聴数と自分の捨て牌を足して向聴数が減るか	$5 \times 4 \times 2 = 40$

表 8 一人麻雀プレイヤーの鳴きに関して追加する特徴量

Table 8 Additional features about calling for the one-player Mahjong player

牌譜の手	牌譜の数	完全一致した数	鳴きのみ正解	正解率
鳴く	6,230	3,755	1,435	84.2
鳴かない	22,164	19,392	N/A	90.7

表 9 鳴きの正解率

Table 9 Accuracy rates of calling

鳴くか鳴かないかの正解率が変動していないため、学習データの走査回数は1回でも十分だと考えられる。この鳴きの分類器は比較的、高精度で鳴きの局面に対して正解を出している。データが異なるため単純な比較はできないが、北川らの研究 [6] では鳴いた局面での正解率は21%である。鳴いた局面の正解率を大きく向上させた。鳴いた局面の我々の正解率が大きく上回った要因は学習した局面を全ての局面でなく攻めているであろう局面に限定したことにある。

8. 一人麻雀プレイヤーの四人麻雀への適応

前章までの結果から降りる局面の認識と鳴きを行う機能ができたため、この章では一人麻雀プレイヤーと組み合わせ、実際に四人麻雀に適用し、実力を評価する。

まず降り局面の認識と一人麻雀プレイヤーの組み合わせについてゲームでの挙動を交えて説明する。相手の手番では捨て牌を見てロン和了できるのならば、すべて行う。図 7 は自分の番での挙動のフローチャートである。まず和了で

きるかどうか確認して、和了できるときはすべてツモ和了する。和了できない時は一人麻雀として牌を選択する。その牌をもとに6.2節で説明した特徴量から降りかどうかを判断する。降りでなければ、最初に選んだ牌を以下のルールで切る。降りであれば降りに必要な牌を切る。今回は降りるかどうかが目目的であるため、回し打ちは考慮しない。降りる牌の選択は、以下の基準で決定する。

- それぞれのプレイヤーについて6.2節にある安全度の高い順位で牌を選ぶ。
- リーチしている人が複数人である場合、お互いに共通して安全度の高いものを選ぶ。
- リーチしたプレイヤーがいない場合、鳴いた数の多いプレイヤーに注目する。

図 8 は相手の番での挙動のフローチャートである。まず相手の牌で和了できる場合は和了する。次にその牌が鳴くことができ、なおかつ自分の番で一度も降りていない時に鳴きの判定を行う。この時に鳴く場合であれば、鳴く牌と切る牌を決定する。

8.1 評価設定

この一人麻雀プレイヤーの実力を測定するため、インターネット上の麻雀対戦サイトである天鳳 [14] で対戦を行った。ルールは東風戦、赤あり、持ち時間は一手5秒に考慮時間が10秒である。鳴きについても同様の持ち時間があ

プレイヤー	1 位率	2 位率	3 位率	4 位率	平均順位	試合数	レーティング
ベースライン+降り+鳴き	0.236	0.321	0.234	0.209	2.41	402	1651
ベースライン+降り	0.237	0.240	0.259	0.264	2.54	834	1507
ベースライン	0.181	0.216	0.252	0.351	2.77	504	1262

表 10 順位分布

Table 10 Rank distribution

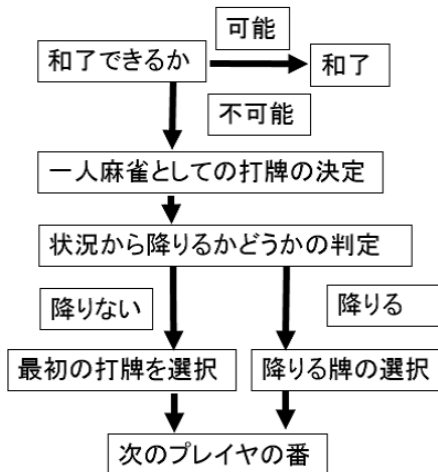


図 7 自分の番でのフローチャート

Fig. 7 Flowchart on the player's turn

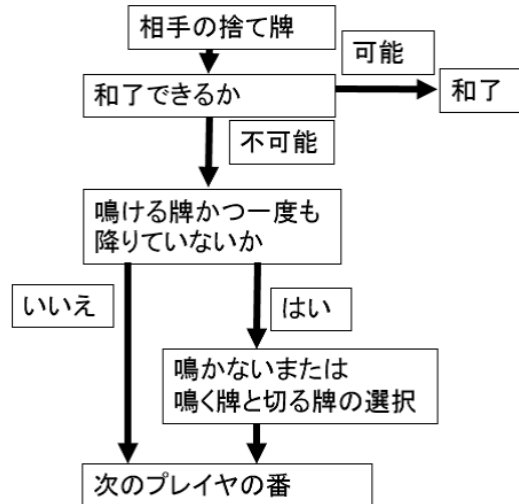


図 8 相手の番でのフローチャート

Fig. 8 Flowchart on opponents' turn

り、鳴ける局面では次のプレイヤーは勝手に牌を引くことができない。天鳳 [14] では成績に応じて対戦できる卓が異なる。卓は 4 種類あり、上から鳳凰卓、特上卓、上卓、一般卓がある。卓の種類は上の卓を選択可能な成績であっても一般卓のみに限定した。プログラムをサイト上で対戦させるための入出力インターフェースは自作した。評価としては平均順位（レーティング）を用いる。レーティング（ R ）とは平均順位と負の相関を持つ基準である。具体的には (3) で計算される。

$$R' = R + (50 - Rank \times 20 + \frac{AveR - R}{40}) \times 0.2 \quad (3)$$

$Rank$ は前回のゲームでの順位である。 $AveR$ は卓の平均の R である。初期 R は 1500 であり、およそ平均順位が 0.1 下がるごとにレーティングは 100 点ほど上昇する。

8.2 評価設定

4 章で作成した一人麻雀プレイヤーである「ベースラインプレイヤー」と「ベースラインプレイヤー」に降りを追加したプレイヤーとさらに鳴きを追加したプレイヤーとの 3 つを比較した。結果は表 10 に示す。ウェルチの t 検定により、2 つのプレイヤーの平均順位は有意水準 1% で差があることがわかった。

和了・放銃率の結果を表 11 に示す。和了率は 4.1 節で説明したとおり、上手いプレイヤーほど高い傾向にある。放銃とは相手の和了牌を切り和了されることである。放銃す

ると相手の得点をすべて自分が払うことになるため放銃率は原則として低いほうが良い。1 局の中で得られる得点は和了・放銃率に影響し、1 局の繰り返して最終的な順位が決まる。したがって和了・放銃率は順位分布と大きな関係がある [12]。

「ベースライン+降りプレイヤー」は「ベースラインプレイヤー」と比較すると和了率がわずかに低下したが放銃率は大きく低下した。「ベースライン+降りプレイヤー」は和了が困難な場合に他人の和了を察知して手を崩して放銃しないように降りることを目的としていたため、和了率は低下するが放銃率がより低下することが期待されこの結果は期待通りであったといえる。

「ベースライン+降り+鳴きプレイヤー」は「ベースライン+降りプレイヤー」と比較すると放銃率がわずかに上昇したが和了率は大きく上昇した。鳴きを行うと自分の手牌が少なくなるため、放銃率が上昇することが予想されるが、それ以上に鳴きによって和了率が良くなり、これも結果は期待通りであった。

「ベースライン+降り+鳴きプレイヤー」のレーティングは 1651 点になった。これはレーティングが 1500 点の平均プレイヤーと比較して強いといえる。具体的な役はあまり知らず、相手の点数状況なども一切、考慮しないにもかかわらず、これほどの実力を得たのは一人麻雀の必要性和降りと鳴きの重要性を示していると考えられる。

プレイヤー	和了率	放銃率
ベースライン+降り+鳴き	0.275	0.149
ベースライン+降り	0.181	0.144
ベースライン	0.188	0.190

表 11 和了・放銃率

Table 11 Winning and Hoju rates

8.3 考察

「ベースライン+降り+鳴きプレイヤー」の挙動を観察したところ他人の和了を察知して降りた局面についてはおおむね正しく判断していた。その局面は他人の和了が近そうで自分の和了が遠く、残りの順目もない場合である。

問題は第一著者が降りるべきと判断した局面で「ベースライン+降り+鳴きプレイヤー」が降りないことが多いことである。それは自分の手の向聴数は確かに小さいが受け入れ枚数が少ない等、あまり手牌がよくない場合である。原因は自分の手牌の良さを向聴数とドラの数でしか判断していないためである。向聴数は手牌の良し悪しを判断する基準としては重要であるが手牌のすべてを判断することはできない。この原因は降りに関して学習した局面が少なく、手牌の情報を入れても予備実験の結果があまり良くなかったため、向聴数程度しか判断基準を入れられなかったためである。

さらに降りの精度を高くするには相手の和了牌を読む技術が必要になる。今の手法では牌の安全度を4段階でしか判別していないが、今後は相手の捨て牌などから、それぞれの牌の安全度を予測することが考えられる。例としては、相手の副露から役を読み、安全牌を理解する技術である。染め手といった役は比較的読みやすい。また牌の安全度がルールで書いたそれとは大きく異なる。リーチに対しても一般的な上級者は無筋の牌の中に順位があり、その中で降りるか否かを判断しているため、ルールをより複雑にする方法が考えられる。

一人麻雀としては、平均的人間プレイヤーの和了率を上回っているものの、まだ悪手を選ぶこともある。例えば対子の価値である。対子は受け入れ枚数を2枚増やし、和了にも必ず必要であるため価値を判断するのは困難である。そのため対子を崩すという選択をあまりしない。その結果、対人戦での七対子の和了率が3.3%と平均の1.7% [12]の倍近い値になっている。まだ一人麻雀として上級者レベルにはなっていないため、特徴量を改良する必要がある。

鳴いた局面についてはおおむね正しかった。問題は教師データには和了した局面が多いため、自分の和了が困難な時に悪手となる鳴きを行うことである。教師データに和了をしていない局面を追加することで解決できる。

9. おわりに

本研究では、麻雀の多人数を削除した一人麻雀を考え、

それと四人麻雀の差を解析し、その差を埋めるように一人麻雀プレイヤーを拡張する手法を提案した。作成したプレイヤーを天鳳 [14] において 402 試合戦わせた結果、レーティングが 1651 点と平均的人間プレイヤーよりも強いことが確認された。平均的人間を上回る実力のコンピュータプレイヤーが報告されていないことを考慮すれば大きな進歩である。今後の課題としては一人麻雀プレイヤーの機能拡張による放銃率の改善などが挙げられる。

降りに関しては、完全ではないものの一定の成果は上げられた。次の課題は回し打ちについての機能拡張が必要であろう。今は回し打ちした局面を降りる必要はないとしか認識していない。回し打ちを認識することで、相手の攻めを認識しながらも自分の手牌を進めていくことができる。これにより相手の和了牌を切らないようにしつつ、自分の和了を目指すうち方ができ、放銃率の改善、成績の向上が見込める。

鳴きに関しても降りと同様に成果を上げられた。次は点数状況に合わせた手作りである。点数状況によっては、鳴いて安い手を和了することは悪手になることがある。今の一人麻雀の特徴量に点数状況を追加することでこれを避けることができると考えられる。

参考文献

- [1] Buro, M.: Logistello: A strong learning othello program, *19th Annual Conference Gesellschaft für Klassifikation eV*, pp. 1-3 (1995).
- [2] Campbell, M., Hoane Jr, A. J. and Hsu, F.-h.: Deep blue, *Artificial Intelligence*, Vol. 134, No. 1, pp. 57-83 (2002).
- [3] 田中哲朗, 金子知適: 4 大規模クラスタシステムでの実行: GPS 将棋の試み (<ミニ特集> コンピュータ将棋の不遜な挑戦), *情報処理*, Vol. 51, No. 8, pp. 1008-1015 (2010).
- [4] Risk, N. A. and Szafron, D.: Using counterfactual regret minimization to create competitive multiplayer poker agents, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, International Foundation for Autonomous Agents and Multiagent Systems, pp. 159-166 (2010).
- [5] 古居敬大, 浦 晃, 三輪 誠, 鶴岡慶雅, 近山 隆: 相手の抽象化による多人数ポーカーの戦略の決定, *Proceedings of the 17th Game Programming Workshop* (2013).
- [6] 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, *Proceedings of the 12th Game Programming Workshop* (2007).
- [7] 三木理斗, 近山 隆: 多人数不完全情報ゲームにおける最適行動決定に関する研究, 修士論文, 東京大学 (2010).
- [8] Bowling, M., Risk, N. A., Bard, N., Billings, D., Burch, N., Davidson, J., Hawkin, J., Holte, R., Johanson, M., Kan, M. et al.: A demonstration of the Polaris poker system, *Proceedings of 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2*, pp. 1391-1392 (2009).
- [9] Kocsis, L. and Szepesvári, C.: Bandit based monte-carlo planning, *Proceedings of the 17th European Conference*

- on *Machine Learning*, pp. 282-293 (2006).
- [10] mjman: 東風壮, <http://mj.giganet.net/>.
 - [11] Sutton, R. S.: Learning to predict by the methods of temporal differences, *Machine learning*, Vol. 3, No. 1, pp. 9-44 (1988).
 - [12] とつげき東北, 福地 誠: おしえて!科学する麻雀, 講談社現代新書 (2009).
 - [13] Collins, M.: Discriminative training methods for hidden markov models: Theory and experiments with perceptron algorithms, *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, Association for Computational Linguistics, pp. 1-8 (2002).
 - [14] 角田真吾: 天鳳, <http://tenhou.net/> (2014).
 - [15] 小松智希, 成澤和志, 篠原歩: 役を構成するゲームに対する効率的な行動決定アルゴリズムの提案, 情報処理学会研究報告. GI,[ゲーム情報学], Vol. 2012, No. 8, pp. 1-8 (2012).
 - [16] Series, S. C.: Measurement of observer agreement, *Radiology*, Vol. 228, pp. 303-308 (2003).
 - [17] Chang, C.-C. and Lin, C.-J.: Libsvm, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/> (2001).

水上 直紀 (学生会員)

2013 年金沢大学理工学域電子情報学類卒業。同年より、東京大学大学院工学系研究科修士課程に在学中。ゲーム AI に関する研究に従事。

中張 遼太郎 (学生会員)

2014 年東京大学工学部電気電子工学科卒業。同年より、同大学大学院工学系研究科修士課程に在学中。拡張現実感における触覚提示手法の研究に従事。

浦 晃 (正会員)

1985 年生。2009 年東京大学工学部電子情報工学科卒。2011 年同大学大学院工学系研究科電気系工学専攻修士課程修了。2014 年同博士課程修了。博士 (工学)。同年富士通研究所入社。機械学習と並列計算の研究に従事。

三輪 誠 (正会員)

2008 年東京大学大学院博士課程修了。博士 (科学)。同年東京大学大学院情報理工学系研究科特任研究員, 2011 年英国マンチェスター大学コンピュータ科学科リサーチアソシエイトを経て, 2014 年豊田工業大学知能数理研究室

准教授。ACL, 情報処理学会, 言語処理学会, 人工知能学会各会員。

鶴岡 慶雅 (正会員)

2002 年東京大学大学院博士課程修了。博士 (工学)。同年, 科学技術振興事業団研究員。2006 年, 英国マンチェスター大学研究員。2009 年, 北陸先端科学技術大学院大学准教授。2011 年より東京大学大学院工学系研究科准教授。

機械学習を用いた自然言語処理, ゲーム AI 等に関する研究に従事。

近山 隆 (正会員)

1982 年 3 月東京大学大学院工学系研究科情報工学専門課程博士課程修了。同年 4 月富士通株式会社入社。同年 6 月 (財) 新世代コンピュータ技術開発機構に出向。1995 年 4 月東京大学大学院工学系研究科助教授, 1996 年同

教授。