

麻雀の牌譜データ分析に基づく鳴き局面の頻出役の分析

2008MI114 窪田 真吾

2008MI135 宮田 良平

指導教員 河野 浩之

1 はじめに

既存のコンピュータ麻雀の研究としてインターネット麻雀ゲーム雀賢荘 [4] では過去の対局データから統計情報を分析し、麻雀が強くなるためにはどうすればよいのか、また強い人は何が違うのか、ということが研究されている。またとつげき東北 (<http://totutohoku.b23.coreserver.jp/hp/>) では過去の戦績から客観的な実力の計測を行うツールの開発や、東風戦と東南戦の違いによって、放銃率に対する和了率の重要性などの研究。また論文として麻雀において流れは存在しないことの証明、麻雀の成績比較の方法論も行っている。本研究では過去の対局データから統計データを調査し、ある局面において次に取れる手を分析し評価する。

2 コンピュータゲームについて

2.1 各コンピュータゲームの進具合

現段階においての各コンピュータゲームの進具合をまとめた。

表1 各ゲームの進具合

ゲーム名	勝った年	強さのレベル
チェス	1996	世界チャンピオン
リバーシ	1997	世界チャンピオン
ポーカー	未定	未定
麻雀	未定	未定
将棋	1990	プロ棋士レベル

表1 を見てみるとリバーシにおいては最強のリバーシソフトは、「MasterReversi」である。このソフトは、世界チャンピオンに勝る強さである [1]。チェスにおいても同様にコンピュータチェスのディーブ・ブルーが1996年に世界チャンピオンに勝利を収めた [3]。また将棋においてはチェスの思考アルゴリズムをさらに進化させ研究が進み、現在ではプロ棋士レベルぐらゐの実力をもっているであろうと言われている。麻雀においては公式にコンピュータとプロ雀士が対戦したという戦歴がない。

2.2 コンピュータゲームの思考

コンピュータゲームの思考法の根本的な原理は評価関数を用いて端末局面の有利・不利の度合いを数値化。そ

してミニマックス探索を使いお互いが自分にとって最も都合のいい手を選ぶと仮定し数手先をすべて読み逆算する。その数値から次の一手を決定する手法である。例え

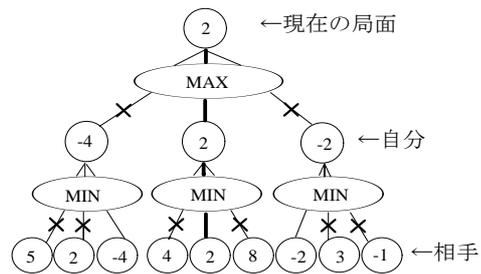


図1 ミニマックス法

ば図1 のようであれば自分の取れる手が3つありその最大の値を選択する。次に相手の取れる手がそれぞれに3つあり相手の取る手は自分にとって一番不利な手を取ってくると仮定する。よって2-2-2という手を取ると考える。この手法を使うことで相手にとって一番不利であり、自分にとって一番有利になる手を取ることができる。

2.3 現在のコンピュータ麻雀の分析

インターネット麻雀ゲーム雀賢荘 [4] というウェブサイトにおいて、様々な成績情報をルール毎に記録している。「雀賢荘」での麻雀対局データの統計情報を分析することで、麻雀が強くなるためにはどうすればいいのか、また強者とアマチュアは何が違うのかを研究している。

例えば強者のアガリ、スピード、守りについて分析しており、アガリのデータ分析では、アガリ率の高い人=強者というわけではないが、強者はアガリ率が高いと分析されている。しかし、平均アガリ点では、平均アガリ点の高い人=強者というわけではなく、まったく関係性がないと分析されている。次にスピードの分析においては、強者はどちらかといえば、平均アガリ巡目は早いという結果がでている。強者は、リーチを掛けることが多いのかという分析では、リーチ率と強さはまったく関係ないようで、それぞれのグループにリーチ派/ダメ派が混在するように見受けられる。但し、リーチ成功率では、強者は、総じてリーチ成功率は高い。守りの面の振込み率においては、やはり強者は総じて振込み率が低い。

3 コンピュータ麻雀の牌譜分析の提案

3.1 牌譜データからの鳴き局面の分析

コンピュータ麻雀における過去の牌譜「雀鬼百瀬」第一最上卓牌譜 (1000 試合超) から統計データを取り、雀賢荘の統計データと比較する。その抽出したデータから鳴きの局面においての取るべき手を分析をする。ここで計算を行う局面はある一定の条件を満たす局面であり、そこで行う判断が勝負の行方を左右するものとする。

3.2 研究の手順

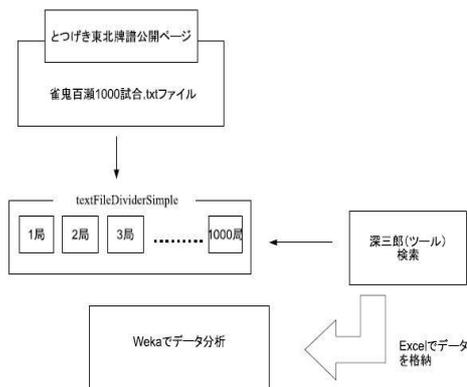


図2 研究の手順

今回行う研究の手順はまずとつげき東北過去の牌譜ページよりコンピュータ麻雀の過去の牌譜 txt をダウンロードし、そして textFileDividerSimple を使用し一局ごとの txt ファイルに分割し、新三郎を使用し鳴き局面のデータに絞り込みその結果を Excel に入力し、csv 形式で保存し Weka の J48 と SMO を用いて分析を行う。

3.3 Weka を用いた分析

集めたデータを分析するために Weka を使用する。Weka はニュージーランドのワイカト大学を中心に開発が進められるオープンソースのデータマイニングソフトである。データマイニングの問題を解決するための機械学習 (Machine Learning) アルゴリズムの一つである。これは、Java で書かれ、ほぼすべてのプラットフォーム上で動作する。アルゴリズムデータセットに直接適用するか、自身の Java のコードから呼び出すか、どちらでも可能である。成績表を Excel でデータ作成を行い、csv 形式で保存する。保存したデータを前処理で読み込ませる。読み込んだデータ決定木とクラスタリングにより分析を行う。

3.4 実装環境

本研究の実装環境は、以下のような環境、ツールを使用し行う。PC のスペック OS: Windows 7 Professional メ

モリ: 2GB CPU のスペック Intel「R」Core i3 CPU, 530 @ 2.93GHz 2.93GHz 使用するツール・Weka・深三郎・textFileDividerSimple

4 牌譜データ分析

4.1 戦績データ

統計的データをとるために実際に行われた牌譜公開ページ [2] のコンピュータ麻雀東風戦、約 1000 局からデータをとった。研究に用いた牌譜データは「雀鬼百瀬」第一最上卓牌譜 (1000 試合超) を用いた。実際に行われた 1000 局のうちロン、ツモなどで和了の形になったのは 844 局、また誰も和了の形にならなかった局、(チョンボ、九種九牌、四風連打などは除く) で流れた局は 156 局であった。図 3 を見ると鳴いてあがる場合は、1、2 回が

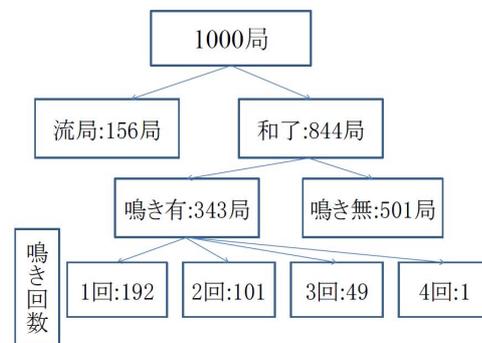


図3 戦績データ

極めて多い。これは 1 回だけ鳴くという形が多いことから鳴いたプレイヤーはテンパイの形になる確率が高いことを示している。全和了のうち約 3 回に 1 回鳴いて和了の形にしている。また鳴き無の和了ではリーチ有、無も数えており、リーチ有は 329 局、ダマでの形は 172 局であった。1000 局の対局データから平均アガリ巡をとり、

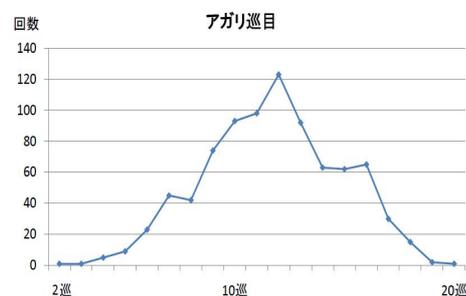


図4 アガリ巡グラフ

グラフ化したのが図 4 である。図 4 よりもっともアガリ順が多かったのは 12 巡目であり、12 巡目を頂点とした

凸グラフになった。このことから運要素が?なく、また手作りをしっかりとしていることがわかる。このグラフから「雀鬼百瀬」第一最上卓牌譜の麻雀プレイヤーはある程度の平均的強さを持ったプレイヤーであると考ええる。

4.2 データの比較

これらのデータをインターネット麻雀ゲーム [雀賢荘] の東風戦データと比較する。お互いのルールを確認する。

表 2 よりルールの違いとして延長戦、ダブルロン、ト

表 2 ルールの比較

	雀賢荘	雀鬼百瀬
喰いタン	あり	あり
後付け	あり	あり
延長戦	南家まで	北家まで 30000 点 を超えた場合即終了
ダブルロン	あり 点棒は上家取り	上家のみ和了
トリプルロン	あり 点棒は上家取り	流局
赤ドラ	あり	あり

トリプルロンにおいてのルールが違うが比較するにあたっては特に問題がないと考える。雀賢荘、雀鬼百瀬の戦績データを比較した。表 2 の比較データより、雀鬼百瀬のデータは比較対象であるデータに対してリーチ時の振り込みアガリ率が高いことからリーチがかかった後勝負をする人が多いと言える。また鳴き割合を見ると雀賢荘の方が鳴いて和了の形にするケースが多くみられる。平均アガリ巡を見ると、両方のデータとも局の中盤あたりになっている。これは麻雀の運要素が高くては役作りしなければいけないからというルールからこのような結果になったと考える。またその他のデータを比較した際に大きな差がでていないことから「雀鬼百瀬」第一最上卓牌譜のデータを用いても問題ないことが分かる。

表 3 戦績の比較

	雀賢荘	雀鬼百瀬
平均アガリ巡	11.6	11.73
(アガリ) ツモ割合 (%)	30.6	31.04
(アガリ) リーチ割合 (%)	37.6	38.98
(アガリ) 鳴き割合 (%)	44.2	35.78
(アガリ) ダマ割合 (%)	18.3	20.37
(振込み) リーチ割合 (%)	35.0	36.59
(振込み) 鳴き割合 (%)	47.2	44.15
(振込み) ダマ割合 (%)	17.8	19.24

5 実装結果

5.1 分析結果

まとめるデータは txt 番号, 鳴きの理由, 鳴き牌, 打牌, ボンとチー, 鳴き牌の種類, 打牌の種類, 鳴き巡だが, 鳴き牌, 打牌で分析を行ったところデータの種類の多すぎて法則性がとれなかったので鳴き牌を中張牌, 老頭牌, 字牌に分類し同様に打牌も分類を行った。本研究の提案に基づき Excel にまとめたデータを weka のアルゴリズムである J48 を使用し分析を行った。分析を行う上で学習させるカテゴリは鳴きの理由を選択した。J48 で分析を行った結果が図 5 である。このグラフの x 軸は minNumObj を示し, y 軸は ROC(Receiver Operating Characteristic) の平均である。ROC は 0~1 までの値をとり, 完全な分類が可能となるときは 1 で, ランダムな分類の場合は 0.5 となる。

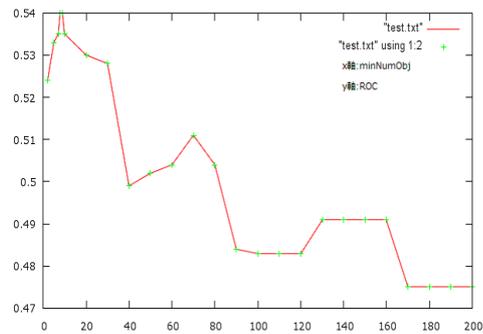


図 5 J48 の average

J48 の average のグラフは右肩下がりであり, minNumObj が増えるにつれて ROC の値は下がるという結果になった。J48 の average の minNumObj が 8.000 と 9.000 のときに ROC が最大値 0.5400 を示した。各鳴きの理由の ROC をまとめたものが表 4 である。

表 4 ROC(minNumObj=8)

minNumObj	8	minNumObj	8
発	0.404	一気通貫	0.750
西	0.529	純全帯	0.768
混一色	0.446	三色同順	0.660
東	0.581	北	0.532
対々和	0.557	全帯	0.503
中	0.503	混老頭	0.461
南	0.564	清一色	0.790
小三元	0.625	テンパイ	0.676
白	0.637		

表 4 を見ると一気通貫, 純全帯, 三色同順, 清一色の数値が高い. . . なので今回は ROC の高い清一色とデータを J48 を使用し分析を行った. また Weka の SMO でも分析を行ったが SMO は J48 に比べ分析処理に時間がかかる. 今回は?ないデータだったのでさほど誤差はないがデータが増えると差が大きくなるので今回は J48 の分析データを用いた. ROC の数値が高く信頼度の高い清

```

Classifier output
鳴き牌の種類 = 老頭牌
| 打牌の種類 = 字牌
| | 鳴き順 <= 9: 混一色 (11.0/7.0)
| | 鳴き順 > 9: 清一色 (5.0/3.0)
| 打牌の種類 = 中張牌
| | 鳴き順 <= 6
| | | 鳴き順 <= 3
| | | | 鳴き順 <= 1: 対々和 (3.0/1.0)
| | | | 鳴き順 > 1: 清一色 (4.0/2.0)
| | | | 鳴き順 > 3
| | | | | 鳴き順 <= 4: 発 (12.0/9.0)
| | | | | 鳴き順 > 4: 清一色 (9.0/5.0)
| | | | | 鳴き順 > 6
| | | | | | 鳴き順 <= 9
| | | | | | | 鳴き順 <= 7: 対々和 (6.0/4.0)
| | | | | | | 鳴き順 > 7: 混一色 (8.0/5.0)
| | | | | | | 鳴き順 > 9
| | | | | | | | 鳴き順 <= 11: 全帯 (11.0/8.0)
| | | | | | | | 鳴き順 > 11: 発 (7.0/4.0)
| | 打牌の種類 = 老頭牌: 純全帯 (4.0/2.0)

Number of Leaves : 37
Size of the tree : 69
Time taken to build model: 0.09 seconds

```

図 6 清一色のルール

一色のルール性は図 6 を見て 9 巡目以降に老頭牌を鳴き, 字牌を切り, 2~3, 5~6 巡目に老頭牌を鳴き, 中張牌を切る場合が清一色の役を狙っている可能性が高いということが分析結果から判明した. 図 7 は図 6 のデータを決定木として表示したものである. 図 8 は ROC の数値が最も高かった清一色の ROC のグラフである.

6 まとめ

本研究では, とつげき東北の牌譜公開ページより雀鬼百瀬の牌譜のみを利用し, その中の 1000 局からアガリ情報などのデータを取り Excel にまとめた. そしてそのデータを Weka の J48 などのアルゴリズムを使用し分析を行った. 結果として ROC を求めたがほとんどのデータが 0.7 未満であることを示した. SMO では平均値は最大で 0.57 であった. ROC の値は 1 に限りなく近いほど信頼度が高いデータであるといえる. 麻雀は将棋, 囲碁などの完全情報ゲームとは違い, 運要素を含むゲームであるため値が低いと考える. また Weka で他のアルゴリズムも使用すればより良い結果が発見できるであ

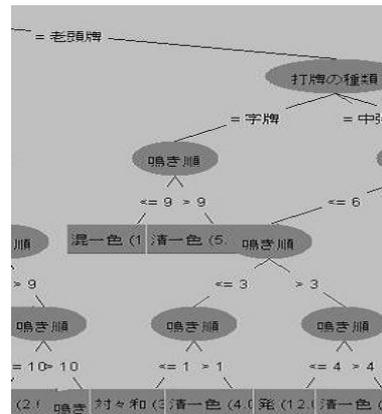


図 7 清一色の決定木

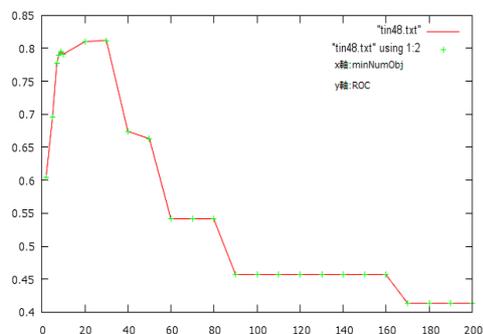


図 8 J48 清一色

うと考える. 今回私たちは鳴き局面に限って調査を行ったが, 打牌からの上がり役の調査や, リーチ局面における調査など他にも局面はたくさん存在する. それらの局面を調査し, 新たな発見があるだろうと考える.

参考文献

- [1] 石井隆, “最強のリパーシ (オセロ) ソフト”
<http://homepage2.nifty.com/tishii/mr/index.html>
 (accessed 2011, 9)
- [2] とつげき東北関係者, “牌譜公開ページ”
<http://totutohoku.b23.coreserver.jp/hp/HAIHU.htm>
 (accessed 2011, 8)
- [3] 村田浩, “ガスパロフ対ディープブルー”
<http://www.ne.jp/asahi/box/kuro/report/deepblue.htm>
 (accessed 2011, 9)
- [4] 有限会社アイ・エス・ジャパン, “インターネット麻雀ゲーム [雀賢荘]”
<http://janken.com/dataroom.asp> (accessed 2011, 9)