

ニューラルネットワークを用いた麻雀の打牌選択方法の提案

松井 一晃¹ 的場 隆一²

概要：本研究では，不完全情報ゲームの中でも特にルールが複雑である麻雀においてコンピュータプレイヤーに打牌選択させる方法を提案する．打牌選択の方法として，現在の局面の状態を入力することにより，各種の牌について打牌に適しているかを評価した値を出力する3層ニューラルネットワークを評価関数として使用している．評価関数の各パラメータの調整には，バックプロパゲーションを用いて教師データの打牌とコンピュータプレイヤーの打牌が一致するように調整している．教師データの打牌には，インターネット麻雀サーバである「東風荘」のレーティング2000以上のプレイヤーの牌譜を使用した．現在は，教師データの打牌とコンピュータプレイヤーの打牌の一致率は31.3%である．

キーワード：ニューラルネットワーク，バックプロパゲーション，強化学習

Selections of Discarding Mahjong Piece Using Neural Network

MATSUI KAZUAKI¹ MATOBA RYUICHI²

Abstract: Mahjong is one of games with imperfect information, and its rule is very complicated to construct mahjong AI. In this study, as a way of discarding mahjong piece, we employed three layer neural networks to calculate evaluation value of each pieces for discarding. Inputs of the neural networks are a current state and pieces on which a player holds, and outputs are evaluation values for deciding a discard piece. Each parameter of the evaluation function is adjusted by backpropagation. As learning data, we employed score sheets of players who have rating over 2000, from the Internet mahjong sever called "Tonpuso". As a result, our NN selects discarding pieces that correspond to learning data with 31.3% accuracy ratio.

Keywords: Neural Network, Backpropagation, Reinforcement Learning

1. はじめに

チェスや将棋といったゲームに関する情報が全て公開されているものは，二人零和確定完全情報ゲームと呼ばれる．これら二人零和確定完全情報ゲームのAI開発は，1950年代に初めてチェスを打つプログラムが開発されて以降盛んに研究されてきた．現在では，チェスの世界チャンピオンよりチェスのゲームAIの方が強いと言われている[1]．用いられる手法はモンテカルロ木探索やmini-max法の枝刈りを使ったものが一般的である．

一方で，麻雀やポーカーといった情報の一部が隠された状態で行われるゲームは，不完全情報ゲームと呼ばれる[2]．

不完全情報ゲームのAI開発は，世界的にみてポーカー（テキサスホールデム）が最も盛んに行われてきた．近年では，ゲーム理論を用いたポーカーAI[3]が人間と互角の勝負を出来るようになった．しかし，麻雀はルールが複雑なことからゲーム理論の適用は困難である．麻雀のルールに関しては2.1節にて記述する．さらに麻雀には「鳴き」と呼ばれるアクションがあり，ツモの順番が入れ替わる可能性がある．このため麻雀ではゲーム木の展開が難しく，従来の探索手法を用いるのは困難である．

麻雀の研究として有名なものには，とつげき東北が立ち上げたシステムティック麻雀研究所[4]の研究が挙げられる．システムティック麻雀研究所では，大量の牌譜を用いた統計的手法から麻雀の研究を行っている．とつげき東北の著書「科学する麻雀」[5]では，麻雀を打つ多くの人に古く

¹ 富山高等専門学校 制御情報システム工学専攻

² 富山高等専門学校 電子情報工学科

表 1 各牌の表記一覧

Table 1 Representation of each Mahjong Piece.

麻雀牌	表記方法
萬子 (1~9)	1m~9m
筒子 (1~9)	1p~9p
索子 (1~9)	1s~9s
東南西北	東南西北
白發中	白發中

から信じられてきた「流れ」などのオカルトが存在しないことや統計学的な「読み」や「ベタおり」に対する考えなどが述べられている。また、統計データと期待値計算に基づいて判断を下す AI として「まったく麻雀」[6] と呼ばれる AI がある。

その他、麻雀の牌譜を利用することで評価関数を学習する方法として、局面に対する評価値を求める手法が提案されている [7]。この研究では、麻雀の人間プレイヤーは局面にある情報を総合的に判断している点や、比較的ゲームの記録を入手しやすい点から 3 層ニューラルネットの教師あり学習を行うことにより、麻雀の局面に対する評価値を求めた。結果は、コンピュータプレイヤーの判断した手と教師データに使用した牌譜で選択された手との一致率はツモ局面において約 56% であった。この AI をインターネット麻雀サーバである東風荘 [8] で対戦させた結果はレーティング 1318 と低く、弱い AI となってしまった。東風荘におけるレーティングの定義は 2.2 節にて記述する。弱くなった原因として、相手のリーチなどに対して降りる場合であっても高い確率で面子を壊さないように打つことや、辺張や嵌張に関して上手く学習できなかったため牌効率の悪い打ち方になってしまったことが挙げられる。

本研究の目的は、ニューラルネットワークを利用することで麻雀の打牌選択を評価し、コンピュータに麻雀の上級者の打ち方を学習させることである。これにより、不完全情報ゲームの中でもルールが複雑なため難しいとされてきた麻雀 AI の開発を目指す。また、麻雀の上級者の打ち方を解明することにより、麻雀における人間プレイヤーへの学習教材の開発やアシスト機能の開発に貢献が期待できる。

2. 打牌選択の学習

2.1 麻雀のルール

本節では麻雀の基本的な知識と、本研究で使ったインターネット麻雀サーバである「東風荘」における麻雀のルールについて説明する。以降、麻雀牌の表記は表 1 にしたがって記す。例えば、萬子の 1 なら 1m、筒子の 5 なら 5p のように表記する。

麻雀は、4 人対戦型の不完全情報ゲームである。4 人で互いの点数を奪い合いゲーム終了時に最も多くの点数を持っていたプレイヤーの勝ちとなる。麻雀に使用する牌は、34 種

表 2 各順位に対する基本項

Table 2 Base point.

順位	基本項	順位	基本項
1 位	6	3 位	-2
2 位	2	4 位	-6

類で各種類それぞれ 4 枚ずつあり全部で 136 枚の牌を使用する。プレイヤーは親 1 人、子 3 人に分かれる。各プレイヤーは手牌として 13 枚の牌を持ち、山として 70 枚、王牌に 14 枚ある状態で親の手番から対局が始まる。3 枚で 1 組の形になっているもの (1m, 2m, 3m 東, 東, 東 など) を面子、同じ種類の牌が 2 牌あるもの (5p, 5p 西, 西 など) を雀頭と呼ぶ。和了の形は、一部の例外 (七対子や国士無双) を除き 4 面子 1 雀頭が基本となる。各プレイヤーは順に山から 1 枚牌を引き (ツモ) 手牌に加え、自分の手牌から 1 枚を川に捨てる (打牌)。これを繰り返し、山がなくなる (流局) か誰かがアガリ (和了) をすることでその局が終了する。親が和了した場合は、親はそのまま同じ局を始める。子が和了した場合は、親が変わり次の局が始まる。また、他のプレイヤーが捨てた牌で面子が出来る場合に鳴き (ポン、チー、カン) と呼ばれるアクションが行えることがある。この場合、ツモの順番が変わり鳴いた次のプレイヤーからとなる。

東風荘では、東場のみで東 4 局の終了時の点数で順位が決まる。ただし、東 4 局の終了時点でトップの点数が 30,000 点以上ない場合はトップの点数が 30,000 点を超えるまで次の局が行われる。その他のルールは、食いタンやオは有り後付け有りの「ありあり」ルールである。点数は 27,000 点持ちの 30,000 点返しである。二人同時の和了は無く、振り込んだプレイヤーからみてツモ順が近い人の和了となる。

2.2 東風荘のレーティングシステム

東風荘では新規のプレイヤーの場合、基準となるレーティング 1500 が与えられる。最初に与えられたレーティング 1500 が対局の結果によって変動することにより、そのプレイヤーの実力を評価する 1 つの指標として使われる。このレーティングの収束には約 400~500 試合は消化する必要があると言われている。レーティングの変化量は、基本項と補正項を求めることによって計算される。基本項の値は、対局の順位によって値が決定する。各順位に対する基本項の値を表 2 に示す。プレイヤーが獲得する基本項は、1 位の場合に 6、2 位の場合に 2、3 位の場合に -2、4 位の場合に -6 となる。

補正項の値は、対局相手のレーティングと自分のレーティングによって決まる。式 (1) に補正項の値を求める計算式を示す。レーティングを R と略して表記している。

$$(\text{補正項}) : \frac{\text{対局相手の平均 R} - \text{自分の R}}{300} \quad (1)$$

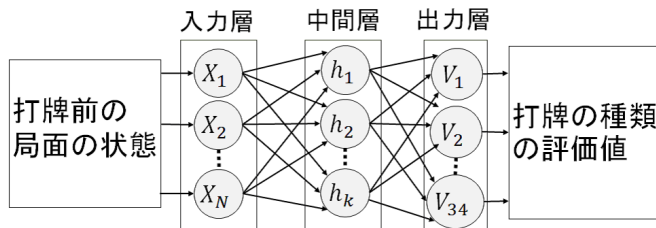


図 1 3層ニューラルネットワーク

Fig. 1 Three layer Neural Network

ただし、対局数が 400 に達していない場合のレーティングの変化量は、基本項と補正項に以下のような補正係数を掛けた値となる。式 (2) に補正係数の値を求める計算式を示す。

$$(\text{補正係数}) : 1 + \frac{400 - \text{対局数}}{100} \quad (2)$$

2.3 単純パーセプトロンによる 3 層ニューラルネットワーク

ニューラルネットワークの単純パーセプトロンとは、人間の脳のシナプスを模した構造をしている。3 層ニューラルネットワークは、入力層、中間層 (隠れ層)、出力層それぞれ 1 層ずつの構造からなる。

本研究で作成したニューラルネットワークの構造を図 1 に示す。入力層は、打牌前の局面の状態を表しており、ノード数は 820 となった。入力層に使用した評価要素と各評価要素を表すのに使用したノード数を表 3 に記す。入力データの評価要素はすべて BOOL 値で表した。評価要素として、リーチの有無、自分の手牌の情報、各プレイヤーの捨て牌の情報、残り牌の情報を使用した。本研究では、刻子や順子といった情報は自分の手牌の情報から得られると考えたため入力データとして用いなかった。出力層は、麻雀牌の種類数である 34 ノードとした。この各ノードがそれぞれの種類の打牌に対する評価値となる。最終的に打牌選択される牌は、手牌にある種類のうち最も打牌に対する評価の高いものとなる。中間層と出力層からの出力関数には、図 2 のようなシグモイド関数 $Sigmoid_a(x) = \frac{1}{1+e^{-ax}}$ を用いた。 a はシグモイド関数のゲインであり x の範囲に応じた値を決める必要がある。

入力層の i ノード目の値を X_i 、中間層の m ノード目の値を h_m として入力層 X_i と中間層 h_m 間の重みを w_{mi} とする。また出力層の o ノード目の値を V_o として中間層 h_m と出力層 V_o 間の重みを v_{om} とすると、中間層 h_m の値は式 (3) 出力層 V_o の値は式 (4) のようになる。 M は中間層の総ノード数である。

$$h_m = Sigmoid_a\left(\sum_{i=1}^{280} X_i w_{mi}\right) \quad (3)$$

$$V_o = Sigmoid_a\left(\sum_{m=1}^M h_m v_{om}\right) \quad (4)$$

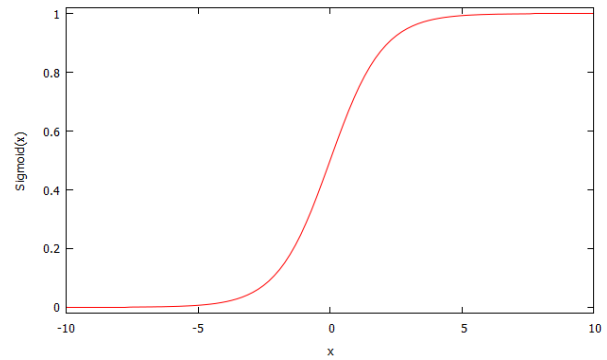


図 2 シグモイド関数

Fig. 2 Sigmoid function

表 3 入力データ

Table 3 Input Data.

評価要素	ノード数
リーチの有無	4
自分の手牌の情報	136
各プレイヤーの捨て牌の情報	544
残り牌の情報	136
計	820

2.4 誤差逆伝播法 (バックプロパゲーション) による学習

バックプロパゲーションは、ニューラルネットワークの各ノード間の重みを調整するための教師あり学習のアルゴリズムである。ニューラルネットワークの出力値と教師データの値の誤差を求め、この誤差が小さくなるように調整を行う。

出力層の値 V_o と教師データの値 T_o との誤差 E を求める式を式 (5) に記す。

$$E = \sum_{o=1}^{34} \frac{1}{2} (V_o - T_o)^2 \quad (5)$$

修正後の中間層 h_m と出力層 V_o 間の重み v'_{om} を求める式と、修正後の入力層 X_i と中間層 h_m 間の重み w'_{mi} を求める式をそれぞれ式 (6), (7) に記す。 η は学習率である。

$$v'_{om} = v_{om} - \eta \Delta v_{om} \quad (6)$$

$$w'_{mi} = w_{mi} - \eta \Delta w_{mi} \quad (7)$$

重みの修正量 Δv_{om} と Δw_{mi} を計算するとそれぞれ式 (8), (9) となる。

$$\begin{aligned} \Delta v_{om} &= \frac{\partial E}{\partial v_o} \\ &= -(V_o - T_o) V_o (1 - V_o) h_m \end{aligned} \quad (8)$$

$$\begin{aligned} \Delta w_{mi} &= \frac{\partial E}{\partial w_m} \\ &= -\left\{ \sum_{o=1}^{34} (V_o - T_o) V_o (1 - V_o) v_{om} \right\} h_m (1 - h_m) X_i \end{aligned} \quad (9)$$

3. 麻雀上級者の牌譜を用いた学習

3.1 打牌学習の実験方法

実験に使用したコンピュータは、プロセッサ：Intel Core i7 3.20GHz、実装メモリ：32.0GB のものを使用した。

教師データにはシステマティック麻雀研究所 [4] の公開牌譜・牌譜募集ページで公開されている牌譜の中でも特に強いプレイヤーとして、レーティング 2000 以上の牌譜を使用した。この牌譜から得られた打牌選択をする局面として 340,000 局面を用意した。用意した局面をプログラミング言語の Ruby を用いて解析を行い、入力データとして扱える形に変換を行った。教師データの出力層の値は、強者プレイヤーとして用意したレーティング 2000 以上のプレイヤーが選択した牌の評価値を 1、その他の種類の牌の評価値を 0 とした。

ニューラルネットワークの学習には、プログラミング言語の C++ を用いた。学習は用意した 340,000 局面すべてを入力したものと、教師データの打牌が各種類一定の回数になるよう前処理をしたものについて行った。学習に要した時間は、学習回数 340,000 において中間層のノード数 820、学習率 $\eta = 0.01$ の場合で約 24 時間となった。学習の評価として、教師データの打牌選択と学習後のニューラルネットワークによる打牌選択の一致率を計算した。

学習後のニューラルネットワークを用いて、インターネット麻雀サーバである「東風荘」にて実際に対戦を行った。これにより、作成した麻雀 AI がどのような打ち方をするのかを確認した。AI の実装には、とつげき東北、システマティック麻雀研究所で公開されている東風荘画面出力用ダイナミックリンクライブラリ (v0.92) を使用した。今回作成したニューラルネットワークでは打牌の選択のみを学習しているので、鳴きやリーチに関する行動には下記のようなルールを設けた。

- 鳴きは行わない。
- リーチ出来る状況になればすぐにリーチする。
- 和了れる場合は和了る。

3.2 打牌学習の実験結果

学習回数に対する一致率の推移のグラフを図 3 に示す。今回の実験では、学習回数 340,000 回、中間層のノード数 820、学習率 $\eta = 0.01$ の時に教師データとの一致率 31.3% が最も良い値となった。一方で、教師データの打牌が各種類一定の回数になるよう前処理をしたものに関しては、各種類 300 回ずつの学習で一致率が 5.95%、各種類 5000 回ずつの学習で一致率が 4.98% と非常に低い値となり学習回数を増やすほど一致率が下がってしまう結果となった。次に、中間層 h_m と出力層 V_o 間の重み v_{om} の推移と、入力層 X_i と中間層 h_m 間の重み w_{mi} の推移をそれぞれ図 4、図 5 に示す。重み v_{om} と w_{mi} 共に、学習により重みが収

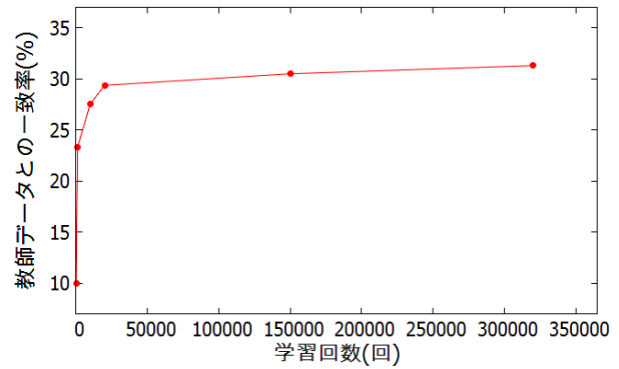


図 3 学習回数に対する一致率の推移

Fig. 3 Rate of concordance with learning frequency.

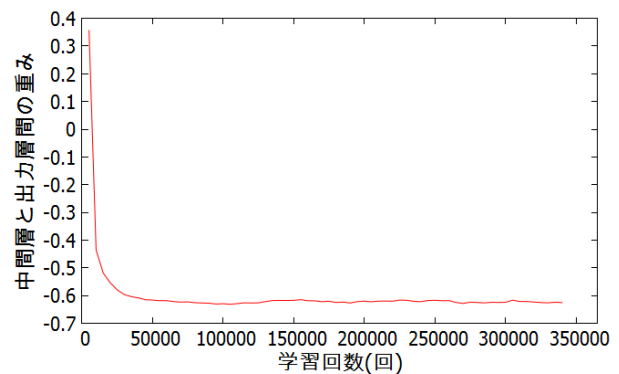


図 4 重み v_{om} の推移

Fig. 4 Transition of v_{om}

束していることが分かる。ただし、重み w_{mi} の修正量の式に入力層の値 X_i の項があるため、 X_i が多くの場合で 0 となるノードとの間の重みは変動が小さくなった。図 6 は、10,000 局面における麻雀牌の種類ごとの打牌選択された回数を表している。教師データの打牌選択ではすべての種類の牌を選択しているのに対し、作成したニューラルネットワークの打牌選択では一九字牌といったヤオチュウ牌の打牌選択が非常に多くなる結果となった。反対に、四五六といった中張牌の打牌選択は極端に少なくなった。これは、中間層のノード数が 30 の場合でも 820 の場合でも変わらなかった。

また、最も教師データとの一致率が高かった学習回数 340,000 回、中間層のノード数 820、学習率 $\eta = 0.01$ で学習後のニューラルネットワークを用いてインターネット麻雀サーバである「東風荘」にて実際に対局させた結果、図 6 の結果の通りに一九字牌を優先的に切ってしまうことが確認された。面子や雀頭については、学習ができなかった。

4. まとめ

本研究では、打牌の種類を評価するニューラルネットワークを作成した。また、インターネット麻雀サーバである「東風荘」にて実際に対局させてみたが、実際に人間相

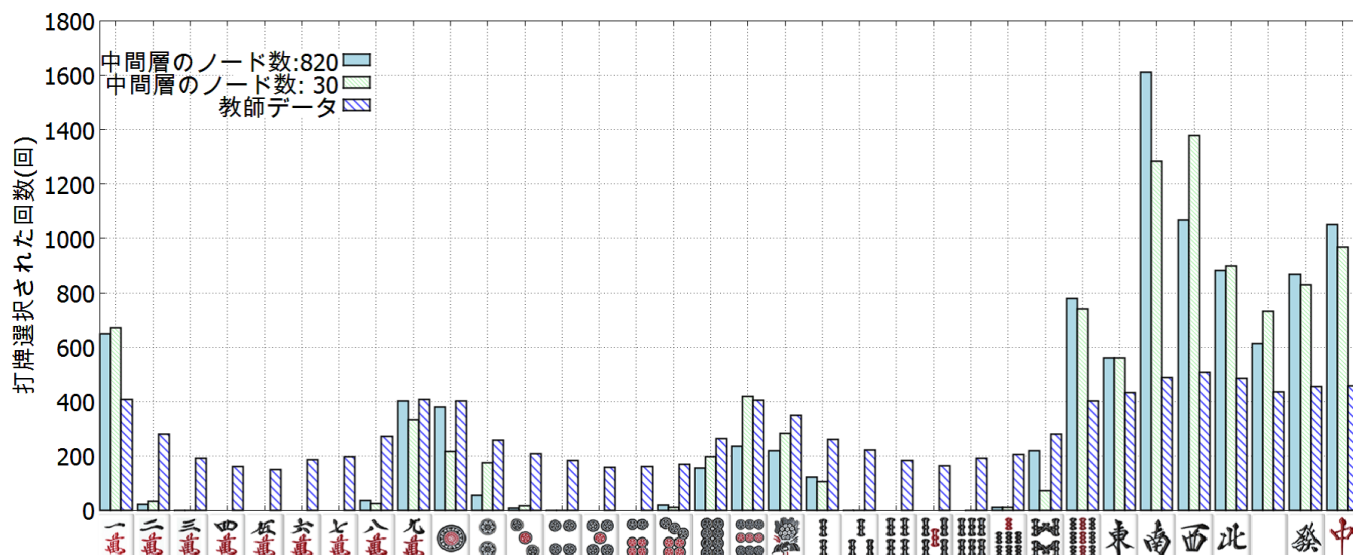


図 6 各牌の選択回数

Fig. 6 Number of saelection mahjong piece.

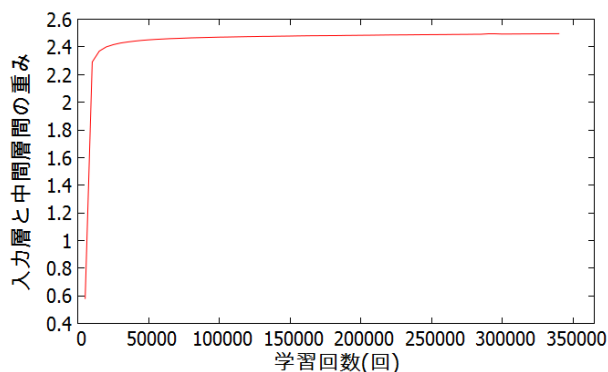


図 5 重み w_{mi} の推移

Fig. 5 Transition of w_{mi}

手に対局できるレベルには達しなかった。

今回作成したニューラルネットワークで教師データとの一致率が上がらなかった要因に、入力データの偏りが挙げられる。学習回数 340,000 回の場合、教師データの種類の打牌選択回数は最も少ない牌が 4m の 5,707 となり、最も多い牌が西で 16,711 と 3 倍以上もの差がある。このため、教師データにおいて多く選択された種類の牌の評価値が高くなるように学習されてしまった。これに対して、各種類の打牌が一定の回数となるものを入力データとしたものを試みたが、良い結果とはならなかった。これは、ニューラルネットワークのパラメータが局所解に陥ってしまった可能性が考えられる。

また、入力データの情報量が少なかったのが原因の 1 つに挙げられる。先行研究では面子や雀頭の情報を直接入力データとして使用していたのに対し、本研究では、手牌の情報から面子や雀頭をニューラルネットワークが勝手に学習してくれることを期待した。しかし、今回の手法では面子や雀頭の学習が出来ないことがわかった。

今後の課題として、パラメータが偏ることによって特定の種類の評価値が高くなってしまいう問題を解決する必要がある。その他に、面子や雀頭をニューラルネットワークに学習させるために、入力データとして新たに面子や雀頭の情報を加えることや、より面子や雀頭について学習し易いニューラルネットワークの構造を考える必要がある。これにより、学習後のニューラルネットワークと麻雀上級者の打牌一致率の向上が期待できる。

参考文献

- [1] 松原仁：完全情報ゲームから不完全情報ゲームへ (2012).
- [2] 作田誠：不完全情報ゲームの研究公益社団法人日本オペレーションズ・リサーチ学会 (2007).
- [3] Martin Zinkevich, Michael Johanson, Michael Bowling, Carmelo Piccione : *Regret Minimization in Games with Incomplete Information*, (2007).
- [4] とつげき東北：システムティック麻雀研究所 <http://totutohoku.b23.coreserver.jp/hp/>
- [5] とつげき東北：科学する麻雀，講談社現代新書 (2004).
- [6] まったり麻雀 <http://homepage2.nifty.com/kmo2/>
- [7] 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 (2004).
- [8] インターネット雀荘「東風荘」 <http://mj.giganet.net/>