

## ニューラルネットワークを用いた麻雀の捨て牌危険度推定

矢ノ口裕貴† 篠埜功‡

芝浦工業大学†‡

## 1. はじめに

麻雀においては、最終的な得点の増大を図るために、振り込みをしない、つまり他のプレイヤーの上がりとなる牌を捨てないことにより失点を抑えることは重要な戦略となる。本論文では、捨てようとしている牌の危険度を推定することを目的として、あるプレイヤーが他のプレイヤーに振り込んだ局面から評価要素を抽出し、ニューラルネットワークの教師あり学習により評価関数のパラメータの調整を行った。関連研究として、SVRを用いて牌の危険度を推定する我妻らの研究[1]、回帰モデルを用いて相手の上がり牌や得点などを予測する水上らの研究[2]などが挙げられる。これらの研究とは、学習方式や局面の評価要素が異なっている。

## 2. 捨て牌の危険度推定

本論文では捨て牌の危険度を、手牌の特定の牌が他のプレイヤーの上がりとなる牌である可能性とする。牌譜から他のプレイヤーに振り込んだ局面を構築し、そこから評価要素を抽出してニューラルネットワークへ入力することで、出力値が捨て牌の危険度になるような学習を行う。

## 2.1 学習データ

学習データとして、とつげき東北システムティック麻雀研究所[3]にて公開されている牌譜を用いる。この牌譜は、麻雀サイト東風荘[4]において上級者と考えられるレート 1800 以上のプレイヤーによるものである。牌譜とは、各プレイヤーの初期手牌、引いてきた牌、捨てた牌などの記録であり、それらの情報から各局面を構築して評価要素を抽出し、ニューラルネットワークの入力とする。

## 2.2 局面の評価関数

局面の評価方法については、局面から評価要素を抽出し、抽出した評価要素をニューラルネットワークへ入力することにより評価値を計算する。本論文では教師付き学習の代表的な手法である、多層パーセプトロンを用いる。

## 2.3 誤差修正

牌譜内の振り込み局面における合法手による次の局面の評価値と、牌譜内の手による次の局面の評価値との差を求め、誤差逆伝播法により調整することでノード間の重みを更新する。この方法は北川ら[5]の手法を参考にしたものである。

ニューラルネットワークの重みベクトル  $\omega$  の更新は以下のように求められる。

$$\omega_{new} = \omega_{old} - \eta \Delta \omega_{old} \quad \Delta \omega_{old} = \frac{\partial l(P, \omega_{old})}{\partial \omega_{old}}$$

$\omega_{old}$  は更新前の重みベクトル、 $\omega_{new}$  は更新後の重みベクトル、 $\eta$  は学習率である。 $l(P, \omega)$  は全合法手について、牌譜の手との評価値の違いの度合い(一致度)を足し合わせる関数であり、

$$l(P, \omega) = \sum_{m=1}^M T(V(p_m, \omega) - V(p_0, \omega))$$

で表される。 $T(x)$  は評価値の差を牌譜の手との一致度に変換する関数であり、本論文ではシグモイド関数を用いる。ここで  $p_0$  は実際に牌譜で取られた行動である。 $\omega_{n,i,j}$  の調整に使用される勾配  $\Delta \omega_{n,i,j}$  は以下の式で求められる。

$$\Delta \omega_{2,i,1} = \sum_{m=1}^M l(P, \omega) (1 - l(P, \omega)) (\varphi_{i,m} - \varphi_{i,0})$$

$$\Delta \omega_{1,i,j} = \sum_{m=1}^M l(P, \omega) (1 - l(P, \omega)) (\psi_{i,j,m} - \varphi_{i,j,0})$$

ここで  $\varphi_{i,m}$ 、 $\psi_{i,j,m}$  は以下の式で求められる。

$$\varphi_{i,m} = V(p_m, \omega) (1 - V(p_m, \omega)) f_{2,i,m}$$

$$\psi_{i,j,m} = V(p_m, \omega) (1 - V(p_m, \omega)) \omega_{2,j,1} f_{2,j,m} (1 - f_{2,j,m}) f_{1,j,m}$$

このような学習を繰り返すことによって得られたニューラルネットワークは、牌譜と似たような手による次の局面の評価値が高くなることが期待される。

## 2.4 学習方法

はじめに、牌譜から学習する局面を構築する。本論文では学習する局面を、いずれかのプレイヤーが他のプレイヤーに振り込んだ局面とした。

本論文では、以下の手順で学習を行った。

- (1) 各局面から牌譜で実際に移行した局面と、合法手で移行することの出来る局面を生成する。
- (2) 生成した局面から評価要素を抽出する。
- (3) 抽出した評価要素をニューラルネットワークに入力して評価値を計算し、牌譜で実際に移行した局

Estimating Risk Associated with Discarding Tiles in Mahjong by Neural Network

†Yanokuchi Yuki, Shibaura Institute of Technology

‡Sasano Isao, Shibaura Institute of Technology

面と合法手で移行出来る局面の評価要素の誤差を評価する。得られた誤差を元に勾配の計算を行い重みベクトルの更新をする。これを何度か繰り返す。

(1)~(3)の手順を全局面で行う。

### 2. 5 評価値の計算

評価値は以下の式で計算する。

$$V(p_m, \omega) = f_{3,1,m}$$

$$= \text{sigmoid} \left( \sum_{k=1}^{k_2} \omega_{2,k,1} f_{2,k,m} \right)$$

$$f_{2,i,m} = \text{sigmoid} \left( \sum_{k=1}^{k_1} \omega_{1,k,i} f_{1,k,m} \right)$$

ここで $p_m$ は抽出された局面  $P$  における合法手の数  $M$  個のうち、合法手  $m$  で進めた場合の局面である。 $m$  番目の合法手における第  $n$  層の  $i$  番目のユニットを  $f_{n,i,m}$ 、第  $n$  層のユニット数を  $K_n$ 、第  $n$  層の  $i$  番目のユニットから第  $n+1$  層の  $j$  番目のユニットへの重みを  $\omega_{n,i,j}$ 、各ユニットにはシグモイド関数を適用する。ここで、第 1 層のユニット  $f_{1,k,m}$  は局面  $p_m$  から抽出した評価要素である。

### 2. 6 評価要素の設定

牌譜から各局面を構築して表 1 に示すような評価要素を抽出し、ニューラルネットワークへの入力とする。この評価要素は、我妻らの研究[1]における特徴ベクトルを参考にして設定している。

表 1 評価要素

手牌の種類	相手の捨て牌の種類
リーチの有無	ドラ
牌が見えている枚数	関連する牌
面子作成の可否	現物かどうか

### 2. 7 学習結果の評価

牌譜内の各局面における合法手で進めた場合の局面に対するニューラルネットワークの評価値のうち、最上位のものが牌譜内の手と一致する割合を牌譜との完全一致率、上位 3 つのいずれかが一致する割合を牌譜との三位以内一致率とし、学習の精度を測る尺度とする。

### 3. 実験

Java 言語でプログラムを実装し、試合数 5000、ニューラルネットワークの中間層を 5, 10, 20, 50, 100, 200 個とし、局面ごとの学習回数を 1, 5, 10, 15, 20 回で行った。実験結果を以下に示す。

#### 3. 1 学習局面数

5000 試合分の牌譜内から学習する局面として抽出した、いずれかのプレイヤーが他のプレイヤーに振り込んだ局面は 3215 局面である。

### 3. 2 実験結果

局面ごとの学習回数と牌譜との一致率との関係について、中間層が 200 個のものについて図 1 に示す。多くの場合で、中間層の数を増やし、学習回数を多くするほど、より高い一致率を得ることが出来た。一致率が最大となったのは中間層 200, 学習回数 15 回の場合で、完全一致率が 93.7%, 三位以内一致率が 97.5%であった。

中間層の数にかかわらず局面ごとの学習回数が 1 回の場合に低い一致率となったが、これは学習する局面がそれぞれ違う試合であるため評価要素同士の関係が学習できなかったのではないと思われる。

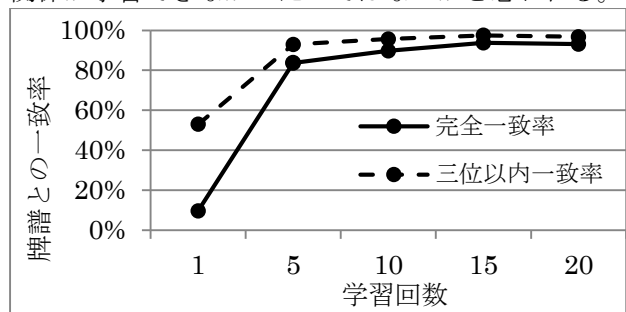


図 1 中間層 200 での完全一致率、三位以内一致率

#### 4. おわりに

本論文では、ニューラルネットワークを用いて捨て牌危険度を推定するシステムを制作した。ニューラルネットワークの評価値と牌譜内の手との一致率で評価を行ったところ、局面ごとの学習回数が 5 回以上の場合に平均 85.6%の一致率が得られた。

今後の課題としては、振り込んだ際の点数情報についても学習する必要がある。これは、どれだけの点数を失う可能性があるか推定することは、上がり優先するか振り込みを避けるか決定する指標となるためである。また、学習結果の評価として、学習していない局面においてニューラルネットワークの出力値が妥当なものになるか検証する必要がある。

#### 参考文献

- [1] 我妻敦, 原田将旗, 森田一, 古宮嘉那子, 小谷善行: SVR を用いた麻雀における酔手牌の危険度の推定, 情報処理学会研究報告ゲーム情報学, 2014-GI-31(12), pp. 1-3, 2014.
- [2] 水上直紀, 鶴岡慶雅: 牌譜を用いた対戦相手のモデル化とモンテカルロ法によるコンピュータ麻雀プレイヤーの構築, ゲームプログラミングワークショップ 2014 論文集, pp. 48-55, 2014.
- [3] とつげき東北システムテック麻雀研究所, <http://totutohoku.b23.coreserver.jp/hp/>.
- [4] 麻雀サイト東風荘, <http://mj.giganet.net/>.
- [5] 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, ゲームプログラミングワークショップ 2007 論文集, pp. 76-83, 2007.