

# カーネルPCAを用いた麻雀力考察

神野 莉子 (指導教員：吉田 裕亮)

## 1 はじめに

麻雀は基本的に4人(3人で行う方法もある)のプレイヤーが牌を用いて遊ぶボードゲームである。各プレイヤーが1枚牌をめくり1枚牌を捨てて役を作り、最終的に一番点数が高いプレイヤーの勝利となる。「和了率」「放銃率」「立直率」「副露率」など様々なデータがあり、プレイヤーによって様々な特徴が出る。牌を1枚めくり1枚捨てるというゲームのため、成績が実力だけではなく運に左右されることも多く、本来の実力を見極めることが難しい。今回は、カーネルPCAを用いて、いわゆる実力者の特徴を捉えることにより麻雀力に関する考察を行う。

## 2 先行研究

麻雀での研究は、何の牌を捨てるべきかをAIで判断する研究が盛んに行われている。麻雀AIはネット麻雀プレイヤーに浸透しており、麻雀AIを用いた学習もある。また、相手の待ち牌推測や点数予測といった先行研究も行われている。ここでは、上級者らのデータの特徴を非線形手法のひとつであるカーネルPCAを用いて、推定することを試みる。

## 3 カーネルPCA

### 3.1 PCA(主成分分析)

PCAとは、分散の大きい方向にデータを射影することで、多次元データの情報を、特性を保ちながらより低い次元に縮約させる方法である。しかし、スケールリング・フリーな識別ができる一方、線形データ解析手法のため、非線形なデータの構造が捉えにくいという欠点がある。

### 3.2 カーネル法

一般にカーネル法では、非線形変換を介して、データの様々な特徴量を取り出すことになる。

$\phi_1, \dots, \phi_d$  を特徴抽出するための非線形関数とし、特徴ベクトルを

$$\vec{\Phi}(x) = (\phi_1(x), \dots, \phi_d(x))^T$$

と書く。

このとき、カーネル関数  $k$  は特徴抽出の内積に基づき、以下のように定義される。

$$k(x, x') = \vec{\Phi}(x)^T \vec{\Phi}(x') = \sum_{m=1}^d \phi_m(x) \phi_m(x')$$

したがって、非線形に写像した空間での  $\vec{\phi}(x)$  と  $\vec{\phi}(x')$  の内積が、入力特徴  $x$  と  $x'$  のみで計算でき、 $k(x, x')$

から最適な非線形な非線形写像を構成することができる。このような関数  $k$  をカーネルと呼び、このように高次元に写像しながらカーネルの計算のみで最適な識別関数を構成することを、一般に、カーネルトリックという。本研究では、カーネル関数として最もよく用いられる Gauss カーネル

$$k(x, x') = \exp(-\beta \|x - x'\|^2)$$

を用いる。ここでは  $\beta > 0$  は学習により調整されるパラメータで、本研究ではカーネルパラメータと呼ぶ。

### 3.3 カーネルPCA

カーネルPCAとは、高次元の特徴空間内の変換してから、通常のPCAを行い、低次元の線形な構造部分空間を求めることにより、非線形な構造を取り出す多変量解析手法である。以下がその手順である。

- (i) 中心化されたデータ点の集合  $x^{(1)}, x^{(2)}, \dots, x^{(n)}$  から、カーネル(グラム)行列

$$K = \left( k(x^{(i)}, x^{(j)}) \right)_{n \times n}$$

を作る。

- (ii) 固有値問題  $K\alpha = \lambda\alpha$  を解く。

- (iii) 上から  $M$  個の固有値  $\lambda_1, \dots, \lambda_M$ , 固有値ベクトル  $\alpha_1, \dots, \alpha_M$  を用いて、 $M$  次元PCAプロットを行う。

本研究においては、 $M$  を2として、第2主成分まで用いて2次元PCAプロットを行った。

## 4 提案方法

本研究では、ネット麻雀に記録されているユーザーデータを用いて特徴ベクトルを取得し、それらに対してカーネルPCAを施しユーザのクラス分類を行う。

- I インターネットからネット麻雀高段位のデータと中段位のデータを採す。

- II 各ユーザーに対して18項目のデータを取得する。本研究で用いた18項目は、

和了率,	放銃率,	ダマ率,
副露率,	平均和了,	平均放銃,
立直率,	立直和了率,	立直収支,
立直収入,	立直支出,	立直巡目,
一発率,	放銃時立直率,	放銃時副露率,
副露後放銃率,	副露後和了率,	調整打点効率

である。

III 1人につき 18 次元の特徴ベクトルを用意してデータセットを構成し、それらの全体データに対してカーネル PCA を施す。

IV 第 1,2 主成分分析を用いた主成分プロットを行い、これらデータの識別の可能性を調べる。

V 学習に用いた特徴ベクトルにおいて、できる限り少ない変数によって IV で得られた PCA プロット図 (またはそれとほぼ等価な図) が維持するように変数の削減を行う。

## 5 実験概要

本研究では実データとして、60 人の各麻雀プレイヤーから 18 次元の特徴ベクトルを取得した。あらかじめ 60 人は麻雀の成績別に 2 つのグループに分けられている。全 60 人のデータに対して、PCA プロット図で出来る限り 2 グループに分かれるようにパラメータ  $\beta$  を学習し、その識別結果を考察する。

### ・目的

全データにカーネル PCA を施し、2 次元 PCA プロットにより識別し、その結果から識別の特徴を考察する。

### ・手法 I と結果

ネット麻雀での成績データを元にして、「(赤) ネット麻雀成績上位の麻雀プロ」と「(青) 成績中位の雀士」の 2 つのグループに分けた。あらかじめ分けられたデータが PCA プロット図において、できるだけ色ごとに識別されるように、パラメータ  $\beta$  を学習した結果が下図 1 である。18 変数から変数をできる限り減らしていき、識別に必要な変数も推定した。

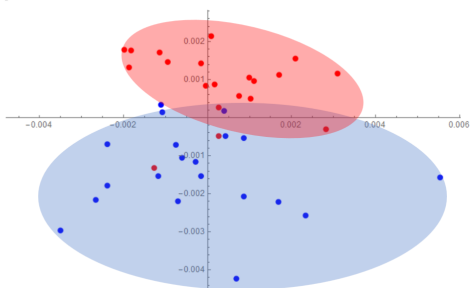


図 1: (赤): 成績上位のプロ, (青) 成績中位の雀士

### ・考察 I

青グループと赤グループにそれぞれ違う色のデータが含まれてしまい、誤判別が発生している。また、18 変数用意したが、4 変数「副露率」、「立直和了率」、「放銃時副露率」、「副露後放銃率」で十分に識別することができた。

### ・手法 II と結果

同様のカーネル PCA に「(緑) ネット麻雀成績上位の雀士 (麻雀プロではない)」を追加し、その結果から識別の特徴を考察する。結果が下図 2 である。

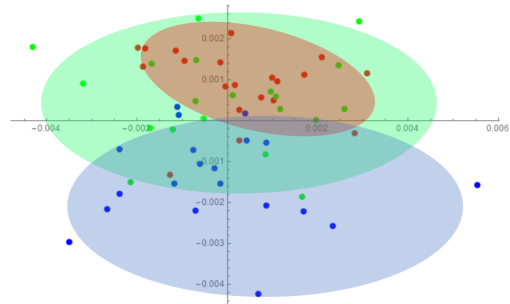


図 2: (赤) 成績上位のプロ, (青) 成績中位の雀士, (緑) 成績上位の雀士 (麻雀プロではない)

### 考察 II

3 グループに区別することができなかったが、「ネット麻雀成績上位」と「ネット麻雀成績中位」に識別することができた。また、(緑) ネット麻雀成績上位の雀士の中に (赤) ネット麻雀成績上位の麻雀プロがあるため、麻雀プロの識別の精度が高いと考えられる。

## 6 本研究のまとめと今後の課題

成績中位の雀士は、立直和了率、副露率、放銃時副露率、副露後放銃率に注目すると麻雀力向上に繋がると推測できる。カーネル PCA は、麻雀データを用いた麻雀力識別の手法の一つとして有効であると思われる。今回はサンプル数が 60 人で行ったが、サンプル数を増やして同様の実験を行い、同じような識別結果が得られるか、についても検証していく必要がある。

## 7 参考文献

[1] MajSoul Stats, <https://amae-koromo.sapk.ch/>