

# スキー選手の滑走ポジションと大会成績の関連性解析

田崎 ほの伽 (指導教員：吉田 裕亮)

## 1 はじめに

統計的機械学習法により大規模なデータセットや複雑な問題に対して、従来の手法では難しかった解析や予測が可能となっている。パターン認識とは、画像や音声などの膨大なデータから一定の特徴や規則性のパターンを識別して取り出す処理のことを指す。しかしながら、正則化された情報から特徴抽出を行う際に、どの情報が重要か分かりにくいという問題がある。そこで統計的に情報を低次元に圧縮する手法を用いて、認識に重要な情報を抽出することを可能とする。先行研究 [1] では、浮世絵役者の顔のパーツごとに、部位間の角度情報を用いて PCA(線形) を行っており、9 人の浮世絵師の識別を図っている。角度情報を用いることで、スケールフリーな識別を行うことができるという利点を応用している。本研究では、画像から得られたデータを用いて、識別の可能性を調べた。

## 2 カーネル PCA

本研究では、カーネル PCA と呼ばれる“非線形解析手法”を用いてデータを扱う。まず、カーネル法と PCA(主成分分析) について説明し、その後カーネル PCA について説明する。

### 2.1 カーネル法

「カーネル関数」を用いた計算方法で、高次元の特徴空間にデータを埋め込むことにより、非線形変換を介して与えられたデータ  $x$  のいろいろな特徴量を取り出す。

$$\phi_1, \dots, \phi_d$$

という非線形関数で特徴抽出された特徴ベクトルを書く。

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

カーネル関数は特徴抽出の内積に基づき、以下のよう  
に定義できる。

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

特徴空間での  $\vec{\phi}(x)$  と  $\vec{\phi}(x')$  の内積が、入力データ  $x$  と  $x'$  のみで計算でき、 $k(x, x')$  から最適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、ガウスカーネル

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

を用いる。パラメータは、 $\beta$  とし、本研究ではカーネルパラメータと呼ぶ。

### 2.2 PCA(主成分分析)

PCA とは、教師なし学習の一種で、多次元データを持つ情報をできるだけ損なうことなく、低次元空間に

情報を集約する方法である。具体的には、射影されたデータ点の分散を最大化しつつ、多次元空間上にあるデータ情報の特性を保ちながらさらに低次元空間上に縮約させる。しかし線形データ解析手法のため、非線形なデータに対しては構造が捉えにくいという欠点がある。そこで上述のカーネルトリックを用いて非線形への拡張を行うことを考える。

### 2.3 カーネル PCA

高次元の特徴ベクトルに変換してから、通常の線形 PCA の手法を行い、低次元の分類部分空間を求める非線形解析手法である。以下のアルゴリズムを用いる。

- (1) 中心化されたデータ点の集合  $x^{(1)}, x^{(2)}, \dots, x^{(n)}$  から、グラム行列  $K = (k(x^{(i)}, x^{(j)}))$  を作る。
- (2)  $K\alpha = \lambda\alpha$  という固有値問題を解く。
- (3) 上から  $M$  個の固有値  $\lambda_1, \lambda_2, \dots, \lambda_M$ , 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、第 2 主成分まで用いて 2 次元 PCA プロットを行った。

## 3 提案手法

本研究では、スキー選手の滑走中の画像から特徴ベクトルを取得し、それらに対してカーネル PCA を施し選手の識別を行う。

- (1) スキー選手の滑走中の同じポジション (左脚外足) 画像をスキャンする。(ただし、同じ種目かつ同じバーンでの滑走とし、3 または 4 ターン目の画像を用いる。)
- (2) 図 1 を例におき、12 次元の特徴ベクトルとして角度情報を取得する。

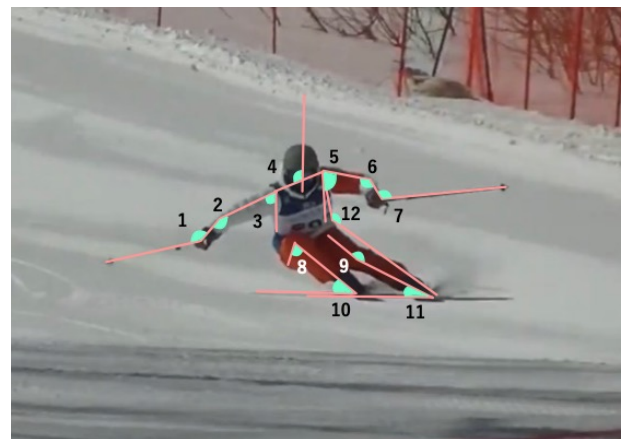


図 1: スキー選手の画像

- (3) 1人につき12次元の特徴ベクトルを複数人分まとめ、それら全体データに対してカーネルPCAを施す。
- (4) 第1, 2主成分分析を用いた主成分プロットを行い、全データの識別の可能性を調べる。

#### 4 実験概要

本研究では実データとして、33人のスキー選手から取得した12次元の特徴ベクトルを用いる。あらかじめ32人を大会成績順に2つのグループに分け、さらに成績上位グループには、プロ選手を1人追加した。全33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なパラメータ $\beta$ を調整する。また、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次元のPCAプロットによって識別を行い、その結果からデータの識別や特徴を考察する。

##### ● 手法Iと結果

2022年2月に行われた全国学生スキー大会での成績データを元にして「(赤)成績上位の選手+プロ選手1人」「(青)成績下位の選手」という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目視で限りなく色ごとに識別されるように、パラメータ $\beta$ を調整した。下図2の結果が得られた。

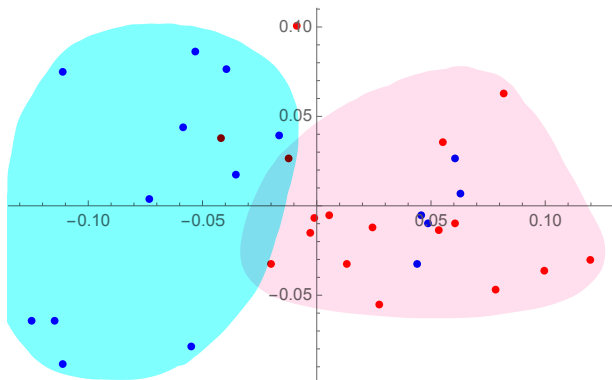


図2: (青): 成績下位選手, (赤): 成績上位選手

##### ● 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。滑り方が種目に適していなかったり、ポジションに癖がある選手もいるため、大会成績とだけでは、PCAプロット図において目視で色ごとに識別することが難しかった。成績には、スキャン画像から得られる角度情報のみならず、斜面状況への適応度(スピードと回転弧のコントロール)やリズム感、種目に適した滑り、エッジングなどの動的な評価も含まれていることが確認できた。そこでポジションの癖と成績の関係性に注目し、新たに識別することが可能であると示唆された。

##### ● 手法IIと結果

各々の選手の12次元の特徴ベクトルのうち、「腕」の角度に対する6次元の特徴ベクトルを削除した。プロの選手と比較して、腕ポジションの癖が十人十色だったため、腕に関する6次元の特徴ベクトルを削除した。さらに選手たちを成績順に「(青)成績上位の選手+プロ選手」と「(緑)成績下位の選手」「(赤)成績中位の選手」の3つのグループに分けた。これらのデータを目視でPCAプロット図ができる限り色ごとに3グループに分かれるように、パラメータ $\beta$ を調整した。図3の結果が得られた。PCAプロット図において、目視で3つのグループを色ごとに識別することが可能となった。

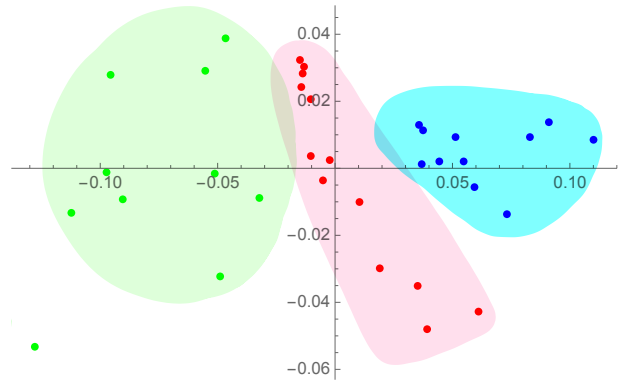


図3: (緑): 成績下位, (赤): 成績中位, (青): 成績上位

##### ● 考察II

結果から、今回のデータでは腕に関する情報がノイズであると判断できた。上の例以外にも、脚の角度に注目した識別を行う際には、それ以外の特徴量を減らすことで新たな識別の可能性が示唆される。

#### 5 まとめと今後の課題

本研究においてカーネルPCAは、画像から得られた特徴量を用いて、データを識別する手法の1つとして有効であった。結果として、ノイズと推測される特徴ベクトルを消して次元を下げることで、目視の範囲で識別が可能となった。今後の課題として、角度情報を減らすなどによってより特徴量を絞り込むこと、また、新たなデータを加えた際に識別の精度を上げていくことが重要であると考えられる。さらに、色々な分野のデータを用いての実験を試みたい。

#### 6 参考文献

- [1] 村上征勝, 浦部治一郎, 「浮世絵における役者の顔の描画法に関する数量分析」, 統計数理, 55(2), 223-233 (2007).
- [2] G. James, D. Witten, T. Hastie, and R. Tibshirani, An Introduction to Statistical Learning, Springer, New York, Heidelberg, Dordrecht, London, (2013).