

信号処理における 機械学習的クラスタリング手法の開発

富山県立大学情報システム工学専攻
1855001 麻生 到

指導教員：奥原浩之

1 はじめに

現代社会には様々な信号が存在しており、機械学習を用いてその信号の中の隠れた因子や成分を見つけ出し分析する技術は多くのことに応用することができる。また、測定する環境は毎回同じであるとは限らずその日の天候や測定する人の体調などにより毎回異なる仕様の信号が取得される可能性がある。

そのため、異なる環境にも対応できる信号処理の手法が重要となってくる。

従来では、音声に含まれる感情と音声の特徴との関係を見つけて、その音声から知覚される感情とを対応づける研究が行われてきた。この研究では、統計的な手法を用い、主観評価値を用いて、主成分分析によって声の高さや抑揚といった物理的基底と因子分析によって心理的基底をそれぞれ音声データから獲得している。

そして、獲得した基底を用いて重回帰分析を施すことにより、音声の物理量と感情との間の対応情報を得ている [1]。

また、医療分野において血小板 (PLT) の時系列データを用いて、B 型肝炎の患者と治療のためにインターフェロンを投与した C 型肝炎の患者、インターフェロンを投与していない C 型肝炎の患者を分類している。

この研究ではクラスタリングの手法としては決定木を用いており、用いる特徴量は時系列データに対する各特徴パターンとの相違度である [2]。

他には、動作を 3 次元時系列座標データをしてクラスタリングを行っている。ラジオ体操第一の 10 種類の動作をモーションキャプチャシステムで動作データとして記録し、データ間の類似度を求める DTW を拡張した連続 DP を用いてクラスタリングを行っている。

そして、それぞれの動作を複数の学習データセットから多数決によってアルゴリズムを統合する Bagging で再調整して最終的な分類結果を得ている [3]。

測定部位を頭頂部のみを使用して脳波データを用いてジャンケン動作のクラスタリングを行う研究がある。この研究では、周波数解析で 4Hz 帯と 8Hz 帯のパワースペクトルを特徴量としており、主成分分析で次元削減を行いマハラノビス距離関数でそれぞれのジャンケン動作を分類し、精度は 37.5 % といった結果となっている。[5]

従来の脳波データを用いたジャンケン動作分類の研究では、使用する測定部位の数が 1 つしかないことと周波数帯の範囲が狭いことが問題として挙げられる。また、運動準備電位を利用していることから加算平均が必要なためリアルタイム解析が不可能となり、実用性があまり高くないと考えられる。また、脳波計の着脱で測定部位のずれによるデータのずれや使用者の体調を考慮しておらず環境の違いに対応していない。

よって、本研究では 8 チャンネルの測定部位を使用し、周波数帯の範囲を広げて多くの情報から次元削減を行い有用な情報を取得する。また、次元削減では主成分分析と線形判別分析の組み合わせを検討することにより分類しやすい特徴量に変え機械学習であるランダムフォレストによる分類を行うことと分類精度の向上を目指す。

また、環境の変化に対応するために事前に複数の異なる環境の学習データを用意しておき学習データの選定により環境の変化にも対応できる処理を行う。

以上の処理を適用して使用者の脳波データからジャンケン動作を分類して、分類結果をもとに使用者が出したジャンケンの手に勝つシステムを作成する。実験では作成したシステムを用いて分類精度を求める。さらに、作成システムは背景脳波を使用していることから準リアルタイム処理が可能であることからすぐにジャンケン動作の脳波データを分析・分類をすることができ、環境の変化の対応も可能である。そのため、従来の研究より実用性が高いシステム作成する。

2 処理の流れ

2.1 環境の変化への対応

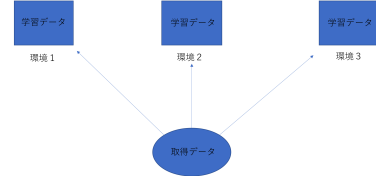


図 1: 環境の変化への対応

複数の異なる環境で取得した学習データをそれぞれ用意する。図 1 のように用意した複数の学習データと取得したデータとの相関を求めてどの環境が取得したデータの環境と近いかを選定する。

選定基準には相関係数を用いる。取得データと相関係数の最も高い学習データを用いてクラスタリングを行っていく。

2.2 前処理

従来の研究として、1 秒間の脳波のデータを 4 分割し、Hanning 窓をかけ FFT でパワースペクトルを求める。そして、4-32Hz 間の全体のパワースペクトルと 4Hz と 8Hz の周波数成分比を特徴量としている [4]。

しかし、この手法では 4Hz と 8Hz の周波数成分しか用いていないため効果的な情報を得られていない可能性がある。

よって、脳波の分類において知られている α 波、 β 波、 θ 波を特徴量成分として用いる。本論文では、まず同様に 1 秒間のデータに対して時間で 4 分割し窓関数をかける。ここでは、Hamming 窓を用いる。窓関数をかけた後、高速フーリエ変換を用いてパワースペクトル PS を求める。 PS は高速フーリエ変換によって得ることができる実部 Re と虚部 Im を用いて以下の式から求める。

$$PS = \sqrt{Re^2 + Im^2} \quad (1)$$

そして、 θ 波帯と呼ばれる 4-7Hz 間の平均パワースペクトルと α 波帯と呼ばれる 8-13Hz 間の平均パワースペクトル、14-32Hz 間の平均パワースペクトルをそれぞれ求めたパワースペクトルを 4-32Hz 間の合計のパワースペクトルで割った周波数成分比を特徴量として用いる。

また、表 1 に分析するデータの分析時間区分を示す。

表 1: 分析時間区分

T1	-1000ms ~ -750ms
T2	-750ms ~ -500ms
T3	-500ms ~ -250ms
T4	-250ms ~ 0ms

2.2 特徴抽出

使用する 1 つの特徴量については、8 チャンネル分のデータをそれぞれ 4 分割し、さらに 3 つの周波数帯の周波数成分比を用いているため $8 \times 4 \times 3$ 次元のデータつまり 96 次元のデータを用いる。

グー・チョキ・パーのデータ M 個の総データを用いたデータ行列 X の式を以下に示す。例としてはデータ行列 X の要素 x_{31} はデータ 3 個目の T1 における α 波帯の周波数成分比となる。

$$X = \begin{pmatrix} x_{11} \theta & x_{11} \alpha & \cdots & x_{14} \beta \\ x_{21} \theta & x_{21} \alpha & \cdots & x_{24} \beta \\ \vdots & \vdots & \ddots & \vdots \\ x_{M1} \theta & x_{M1} \alpha & \cdots & x_{M4} \beta \end{pmatrix}$$

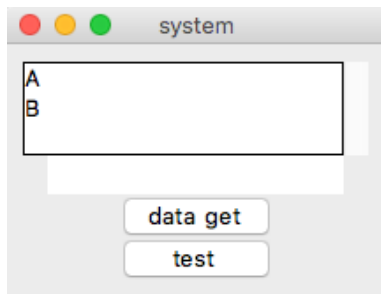


図 2: GUI



図 3: data get のスライドの流れ

作成したデータ行列 X に対して次元削減を行う。本論文では、次元削減を行う手法として主成分分析と線形判別分析を適用した。

主成分分析は、データを分散が最大となるような主成分軸上に順次に射影することによってデータの持つ情報の大部分を失うことなく元のデータより低い次元で表現することができる手法である。本論文では、一方、線形判別分析はクラス内分散・クラス間分散比を最大にする空間部分を求める手法であり、識別を考慮した次元削減法である。

2.3 クラスタリング

これまでに述べた前処理と特徴抽出の手法を用いて特徴量となるデータを作成する。クラスタリングを行う判別器としてランダムフォレストと SVM の手法を用いて比較する。

3 作成システム

脳波を利用してジャンケンで利用者に勝つシステムを作成する。作成するシステムの GUI を図 3 に示す。

3.1 data get

data get の項目では、学習データを取得する。利用者には図 2 のような流れで表示されるスライドを見てもらい 3・2・1 とカウントダウンされた後の黒い画像の表示時に任意のグー・チョキ・パーの手を出してもらう。分析には黒い画像の表示時の 1 秒前から 1 秒後のデータを用いる。

data get 項目では、図 2 の流れを 1 セットとすると 90 セットのデータを取得する。そして、その取得したデータを学習データとして使用する。

3.2 test

test 項目では、実際に data get 項目で取得した学習データを用いてジャンケンを行う。表示するスライドは図 2 の流れと同じスライドだが黒い画面の表示時の 1 秒前から 1 秒後のデータを未知のテストデータとして用いて黒い画面の表示後にその分析結果をグー・チョキ・パーの画像によって表示する。例として分析結果としてグーが表示された時のスライドを図 3 に示す。

4 評価実験

作成したシステムの精度を確認するために評価実験を行う。

4.1 実験方法

本実験に使用する脳波データは、OpenBCI 製のヘッドセットである ULTRACORTECH MARK 4 と Cyton Board を用いて取得する。この脳波計の主な特徴を以下に示す。

- ・サンプリング周波数 250Hz
- ・5-50Hz のバンドパス処理を行ったデータを送信
- ・1-8 チャンネルのデータ以外にも角速度とタイムスタンプも計測可能
- ・Wi-Fi と Bluetooth の 2 種類の無線通信方法が可能

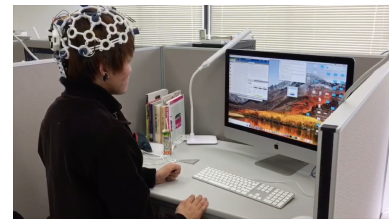


図 4: 実験の様子

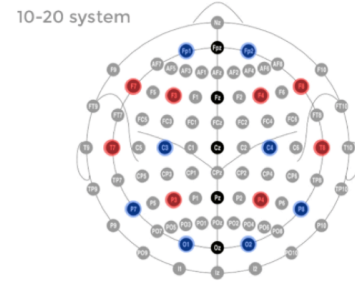


図 5: 測定部位

実際に脳波計を装着している様子を図 4 に示す。また、本実験で使用する 10-20 法に従った脳波の測定部位を図 5 に示す。

参考文献

- [1] 小松尚久, 内田薫, 池野修一, 坂野鋭, バイオメトリクスのおはなし, 日本規格協会, 2008.
- [2] 石川由羽, 吉田智奈美, 高田雅美, 城和貴, 脳波個人認証のための特徴抽出, 情報処理学会研究報告, 2014
- [3] 鈴木良次, 佐藤俊輔, 池田研二, 吉川昭, 生体信号—計測と解析の実践—, コロナ社出版
- [4] 波部齊 ランダムフォレスト, 情報処理学会, 2012
- [5] 山岡啓介 知っておきたいキーワードランダムフォレスト