

動的フラクタル次元を利用した EEG データの索引付け Indexing of EEG Data Using Dynamical Fractal Dimension

矢島 雄一 小西 修
Yuichi YAJIMA Osamu KONISHI

1. はじめに

近年の世の中を行き交う情報量は膨大なものになり、そのようなデータの中からまだ発見されていない有用となる知識や情報を発見するのがデータマイニングである。データマイニングに用いられるデータの一つとして時系列データがある。時系列データとは、時間軸に沿って観測された値の集合であり、生体データや自然現象データ、金融データなどがある。この時系列データをデータマイニングの対象として扱う際に問題となることが3つある。1つ目は、大規模なデータベースであるということ。2つ目は、時系列間の類似性が主観的であること。3つ目は、ノイズやエラー値など様々なデータが混在するということである。このような時系列データの問題を解決するために、時系列データに索引付けを行うが、時系列データの索引付けにも(1)類似性をどうするか、(2)いかに高速に検索するか、という問題が残る。これらの問題に対し、(1)には content based index によって、(2)には次元圧縮をすることで解決できる。

本研究では、実験データとして EEG (脳波) を用いる。EEG は頭皮上に複数の電極を配置し、神経細胞群が活動している部位の電気活動を電位変化として記録したものである。EEG データを扱うに当たって、EEG はノイズが多く含まれており、そのままデータとして扱うわけにはいかない。ノイズの影響を少なくするための前処理として、従来は加算平均を行っていた。しかし、加算平均処理では本来着目している波形を崩してしまい、情報を損失する可能性がある。そのため、前処理として ICA (主成分分析) を用いて、本来着目している波形を推定し、情報の損失を軽減することを目的としている研究があり、この研究では、さらにフラクタル次元を算出する際、時間を追って複雑さを定量化できるように動的フラクタル次元を用いている。

2. 目的

同一の被験者の EEG データを観測時の状態別に識別することを目的としている。使用するデータは、被験者に3つの状態をさせたとき記録した EEG を用い、被験者の3状態を高精度に識別できるか実験をする。3状態が高精度に識別出来れば、BCI に適用することができる。

3. 提案手法

本研究は、大量に蓄積されたデータに存在する共通の性質や特徴の抽出、識別を行うデータマイニングの研究として、不規則で複雑な時系列データである EEG を BCI に適用できるよう、高精度に識別することを目的とする使用する EEG データは、3つの状態、

- (1) 1つの絵を観ている状態
- (2) 2つの同じ絵を観ている状態
- (3) 2つの異なる絵を観ている状態

があり、被験者に対し、3状態それぞれの EEG をランダムに複数回記録している。以下に EEG データを処理する流れを示す。

(1). 電極の選択

頭皮上の 64 極の中からノイズの影響が少ないと思われる 11 電極のデータを選択する。

(2). ICA

選択したデータに対して ICA を行うが、その前処理として、PCA (独立成分分析) を行い、ICA に課せられる平均0分散1を満たすようにする。その後、ICA を行い、脳内からと思われる信号やノイズなど複数の信号分離する。

(3). 動的フラクタル次元

EEG を時間的に捉えるため、分離した複数の信号に対し動的フラクタル次元を用いる。

(4). 脳内からの信号を推定

複数の信号の中から、脳内からの信号を推定する。脳内からの信号は状態変化などのタスクが起きたとき、動的フラクタル次元が上昇するか、下降するかどちらかだと考えられる。そこで、動的フラクタル次元の変動に注目し、動的フラクタル次元を時間的に2つに分割し、次元の上昇値、または下降値を比較する。次元の変化が著しい信号を脳内からの信号と捉える。

(5). SOM (自己組織化マップ)

脳内からの信号を被験者の試行ごとに取りまとめ、得られた動的フラクタル次元をクラスタリングにかける。クラスタリングは3x3のSOMを用いる。各クラスタに状態ごとに識別されているか確認する。

4. 実験

4.1 実験目的

被験者に対して複数回記録された3状態の EEG データの複数個の電極を選択した場合に、高精度に識別することで、BCI への適用を目指す。状態ごとに識別が行えれば、BCI において EEG を観測した波形を入力として、識別したデータベース中のデータとマッチングを行うことで、入力に対応する動作などが出力できる。したがって、同じクラスタ内に複数の異なる状態が入ってしまったら、BCI に発展させることは難しいので、クラスタリングを高精度に行う必要がある。

4.2 実験データ

本研究の実験データは「UCI Knowledge Discovery in Databases Archive」で無償提供している大規模データ集合のオンラインアーカイブの EEG データを使用する。このデータは頭皮に置かれた 64 電極から測定された 1 秒間 256hz でサンプリングされたデータである。被験者には、

アルコール依存症者と健常者の2パターンがあり、本実験では健常者のデータを用いる。各被験者に対して1つの絵を観ている状態、2つの同じ絵を観ている状態、2つの異なる絵を観ている状態の3つの状態のEEGをそれぞれ複数回記録している。

4.3 実験方法

ここで実験の方法を説明する。

オンラインアーカイブからダウンロードしたデータ内にはアルコール依存症者36人と健常者36人のEEGデータがあり、各被験者は、3つの状態をそれぞれ複数回EEGを記録されている。EEGを測定する際の電極数は64極であるが、このうち前頭部や側頭部付近は眼球運動や呼吸などの生体信号がノイズとして含まれる可能性が高いため、ノイズが含まれにくいと思われる頭頂部付近の11電極のデータを実験に使用する。

次に、選択した11電極のデータに対してICAを行う。本実験では、統計的に独立した8つの成分に分解する。これにより脳からの信号とノイズとに分離ができる。

8つの独立成分に分離した後に、8つの信号それぞれに対して動的フラクタル次元を求める。ここでフラクタル次元を求める際の幅や、フラクタル次元の個数を複数設定し、動的フラクタル次元を算出する。

健常者の1人の1状態のEEGデータに対して、8つの動的フラクタル次元を求めた後に、8つのベクトルの動的フラクタル次元の上昇値に着目し脳内の活性部位からの信号をどれなのか推定する。ここで、8つの信号のうち脳内の活性部位からの信号以外を選択してしまうと、この後の実験が全て台無しになってしまう。

脳内の活性部位からの信号を推定したら、健常者の試行ごとに推定されたデータをまとめ、そのデータを入力としてSOMによるクラスタリングを行う。SOMのサイズは3×3の9クラスタとし、学習率は、近傍サイズは、学習回数 回とする。状態によって動的動的フラクタル次元が特徴付けられていれば、クラスタ内は同じ状態のデータが集まり、3状態を識別することができる。

4.4 実験結果

ここで健常者1人の3状態あるEEGをSOMでクラスタリングした結果を図1に示す。解析窓を160、解析窓のずらし幅を10pointとして脳内の活性部位からの信号を推定した場合の結果である。クラスタ数が9でないのはクラスタリングされないセルがあったためである。この図を見ると、

- ・3状態がそれぞれ別のクラスタに分類されている事がわかる。つまり動的フラクタル次元を100%で識別している。
- ・同じ状態でも、別のクラスタに分類されている場合がある。

事がわかる。状態別にクラスタリングを行えたことから状態別に高精度に識別する目的を達成している。

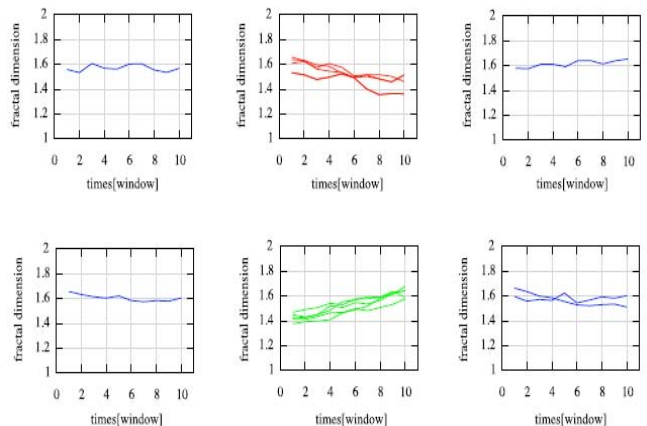


図1 クラスタリング結果

5. おわりに

本研究はノイズの入りやすいEEGデータに対し、電極を選択し、ICAを行った上で動的フラクタル次元にすることでデータの複雑さを定量的に捉えることができる。1人の3状態を記録されたデータをSOMでクラスタリングし、状態ごとに識別できた。今後は元のデータは3種類に分けられるため、2×2のSOMでクラスタを行うことで計算速度の向上をはかり、またどの被験者でも識別を行えるようにする。

参考文献

- [1]丸山 直樹, “EEGの動的フラクタルを索引とした類似検索”, 公立はこだて未来大学卒業論文, (2009).
- [2]澤谷 一志, “ICAを前処理としたEEGのフラクタル・クラスタリング”, 公立はこだて未来大学卒業論文, (2008).
- [3]伊藤 晶, “自己相似性に基づく時系列パターンクラスタリング”, 信学技報, DE2005-5, PRMU2005-26, (2005).
- [4]Monttri PHOTHISONTHAI 1, Masahiro NAKAGAWA 2, “EEG-Based Classification Motor Imagery Tasks Using Fractal Dimension and Neural Network for rain-Computer Interface”, (2006).
- [5]Monttri PHOTHISONTHAI 1, Masahiro NAKAGAWA 2, “EEG-Based Fractal Analysis of Different Motor Imagery Tasks using Critical Exponent Method”, International Journal of Biomedical Sciences, Vol.1, No.3 (2005).
- [6]Monttri PHOTHISONTHAI 1, Masahiro NAKAGAWA 2, “Fractal-Based EEG Data Analysis of Body Parts Movement Imagery Tasks”, J.Physiol.Sci., Vol.57, No.4 (2007).
- [7]Higuchi T., “Approach to an irregular time series on the basis of the fractal theory”, Physica D, Vol.31, (1998).
- [8]Babara D. 1, Chen P. 2, “Using Self-Similarity to Cluster Large Data Sets”, Data Mining and Knowledge Discovery, Vol.7, (2003).
- [9]門田 優1, 中川 匡弘2, “独立成分分析に基づく時系列パターンのクラスタリング”, 信学技報, NLP2004-61 (2005).
- [10]UCI Knowledge Discovery in Databases Archive, <http://kdd.ics.uci.edu/>.

†公立はこだて未来大学 システム情報科学部 複雑系科学科 Future University Hakodate, The School of Systems Information Science, Department of Complex Systems