

脳波を用いた運動想起型 Brain Computer Interface のための
End to End 学習

佐野 光

Sano Hikaru

2018 年 1 月 10 日

目次

第 1 章	序論	1
1.1	背景	1
1.2	BCI の概要	4
1.3	目的	7
1.3.1	モチベーション	7
1.3.2	達成目標	8
1.3.3	技術的貢献	9
第 2 章	BCI のための要素技術	11
2.1	周波数解析	11
2.1.1	4 種類のフーリエ変換	11
2.1.2	パワースペクトル密度	13
2.2	多変量解析	14
2.2.1	Principal Component Analysis(PCA)	14
2.2.2	Independent Component Analysis(ICA)	17
2.2.3	Blind Source Separation(BSS)	18
2.3	分類手法	20
2.3.1	Linear Discriminant Analysis(LDA)	20
2.3.2	Support Vector Machine(SVM)	23
2.3.3	Logistic Regression(LR)	27
第 3 章	従来手法 BCI	29
3.1	特徴量工学	29
3.1.1	頭皮領域と空間フィルタ	29
3.1.2	EEG と時間周波数解析	32

3.2	Common Spatial Pattern とその派生手法	33
3.2.1	脳波信号の定式化	33
3.2.2	Common Spatial Pattern(CSP)	33
3.2.3	Common Sparse Spatio Spectral Pattern(CSSSP)	35
3.2.4	Filter Bank Common Spatial Pattern(FBCSP)	37
第 4 章	提案手法	38
4.1	ニューラルネットワーク	38
4.1.1	ニューラルネットワークに用いられる層	38
4.1.2	活性化関数	40
4.1.3	ニューラルネットワークの設計と学習	42
4.2	提案手法のアイデア	43
第 5 章	実験結果と考察	44
5.1	提案手法の評価	44
5.1.1	EEG データ	44
5.1.2	評価方法	44
5.1.3	ニューラルネットワークの構成	46
5.1.4	結果	47
5.1.5	考察	48
5.2	被験者のデータを混合した BCI 構築	49
5.2.1	データの混合	49
5.2.2	混合データに対する結果	49
5.2.3	活性化関数に elu を用いた場合の結果	49
5.2.4	考察	50
5.3	新規被験者への適用検討	50
5.3.1	実験内容と結果	50
5.3.2	考察	51
第 6 章	結論	53
6.1	まとめ	53
6.2	今後の展望	53

第 1 章

序論

1.1 背景

約 1 世紀前、脳の電気活動の研究が最初に行われて以来、臨床、診断、およびリハビリのための脳活動の解析と解釈に大きな関心が寄せられている。いくつかの研究では、脳から放出される電気信号の特性は、各脳活動および個々人に特有であることが示されている。その結果、脳信号は以下のような領域で利用されている。

- 医療応用：脳信号は、認知症やてんかん発作のような様々な精神障害の診断、および治療のために広く活用されている。さらに、多くの精神障害の早期診断のために脳信号が使用できることが示されている。
- 生体認証：脳信号は偽造、盗聴が困難であることから、固体の識別のための普遍的な生体情報となりうる。商用応用には不適な可能性があるが、他の生体信号と組み合わせることで認証システムの信頼性を向上することができる。
- Brain Computer Interface(BCI)：脳信号はコンピュータや機械などの外界と、筋肉の動作無しに相互作用することができる。BCI は “direct neural interface” や “brain machine interface” とも表現され、基本的に脳と外部世界との間のインターフェースの役割を担う。脳の電氣的活動を外部装置への制御信号に変換することで動作する。初期の BCI は、麻痺患者や障害のある患者が、車いす、義手義足、音声合成装置などの生活補助装置を制御することに役立つように設計された。しかし、近年は高度な精神的タスクを実行する健常者の支援や仮想空間での入力装置としての商用 BCI が登場するに至っている。

BCI において脳信号を計測する方法は、大きく分けて以下の 3 つがある。

- 侵襲式：脳の灰白質へセンサを直接埋め込む方式
- 部分侵襲式：頭蓋骨の内部、脳の表面へセンサを埋め込む方式
- 非侵襲式：センサを頭蓋骨外部へ配置し、外科手術を必要としない方式

侵襲式と部分侵襲式ではセンサーを埋め込む外科手術を必要とし、センサの寿命は数年しか続かないため、約 2 年毎に取り換えの手術を必要とする。結果として侵襲式 BCI の利用範囲は臨床試験に限られている。対照的に、非侵襲式は外科手術の必要性がなく、商業及び医療的な用途の両方に用いられており、今後更に発展する可能性がある。

BCI で用いられる非侵襲式の計測に以下のようなものがある。

- functional Magnetic Resonance Imaging (fMRI) : MRI によって脳活動時の血流によって生ずる磁場を測定する。この測定は、脳の神経細胞が活動中により多くの酸素を含む血流を必要とするという事実に基づいている。
- functional Near-Infrared Spectroscopy (fNIR) : 近赤外線電磁波を用いて、脳皮質の異なる部分における酸素化および脱酸素化ヘモグロビンの濃度を測定する。これらの測定は、fMRI と同様に、皮質の活性部分を決定する。fNIR と fMRI との重要な違いは、fNIR 法は数センチメートルのオーダーの非常に限られた侵入深さでの測定、すなわち脳の表層上に限定されるが、fMRI 法は任意の深さで脳活動を測定できることである。
- Magnetoencephalography (MEG) : この方法は、高感度磁力計アレイを使用して、脳の神経活動によって生成される磁場を直接測定する。MEG は主に、接線電流源に由来する磁場を記録するが、これは通常皮質の漿膜壁に位置する。MEG を利用する大きな利点は、頭蓋骨や他の組織は磁場に対してほとんど透明であるため、MEG には減衰や歪みが生じないことが挙げられる。
- Electroencephalography (EEG) : この方法では頭皮上にいくつかの小さな電極を配置し、脳全体の神経アセンブリによって生成された電場が測定される。EEG は、通常、大脳皮質の表面上に位置する放射状の電流源によって生成される電場に敏感である。

これらの方法の中で、fMRI および MEG は、fNIR および EEG と比較して比較的高い空間分解能を提供する。しかし、fMRI 及び MEG は、非常に高価な装置である上に、その動作のために整備された環境を必要とする。さらに、fMRI(図 1.1) と MEG(図 1.2) は大型機器であるため、BCI アプリケーションで必要とされるような要件を満たすとは言い難い。fNIR はポータブルであるが、脳活動から数秒程度の遅れで測定が行われるという時間分解能の低さが問題となる。結果として、EEG は BCI において脳活動を測定するために最も広く使用されている方法であり、したがって、本論文の研究の焦点とする。

EEG には、BCI システムの設計で考慮すべき大きな制限事項が 2 つある。それは、限られた空間分解能と、限られた浸透深度である。これらの限界に対処するために、様々な記録方法を利用するマルチモーダル BCI システムが提案されている [1]。BCI アプリケーションでの移植性の重要性を考えると、マルチモーダル BCI システムの最良の候補は、EEG と fNIR 信号の組み合わせである。この論文の焦点は EEG の BCI システムに



图 1.1: fMRI



图 1.2: MEG



図 1.3: EEG キャップ (ミユキ技研)

あるが、研究の結果はマルチモーダル BCI システムの一部として利用可能である。

1.2 BCI の概要

過去 20 年間に、身体障害者を支援するために様々な EEG ベースの BCI システムが開発されてきた。しかし現存する BCI システムの大部分は、感覚刺激に応答して生成された EEG 信号である誘発電位に基づいて動作する BCI である。

誘発型 BCI

例として、ユーザが BCI システムを使用してマウスカーソルを動作させたいとする。この時、異なる周波数で点滅する複数の光源をユーザに提示することで、注視した光源に応じた誘発電位を生成させることができる。結果として、光源を見たユーザの脳信号を分析することでマウスカーソルの動作方向を決定することが可能である。このタイプの BCI は通常、SSVEP 型 BCI と呼ばれる。他にもオドボール課題と呼ばれるタスクを課した時に生じる EEG の呼称に因んだ P300 型 BCI など存在するが、外部刺激によって誘発されるタイプの BCI を本論文では誘発型 BCI とまとめて表記する。誘発電位を用いた BCI システムは非常に正確であるが、ユーザは常に刺激に直面するため、長期的な使用には向いていない。また、BCI システム自体が刺激装置などの外部機器を必要とする。

自発型 BCI

誘発型 BCI の問題を解決するために、近年は自発的な脳活動を用いた自発型 BCI の研究が盛んとなっている。自発型 BCI の中でも特に、特定の身体部位の動作を想像することで動作する運動想起型 BCI に注目する。運動想起型 BCI の簡単な例を示すために、ユーザが BCI システムを使用してマウスカーソルを動作させる例を見る。この時、マウスカーソルを左に動かしたい場合は左手の運動を想起し、マウスカーソルを右に動かす

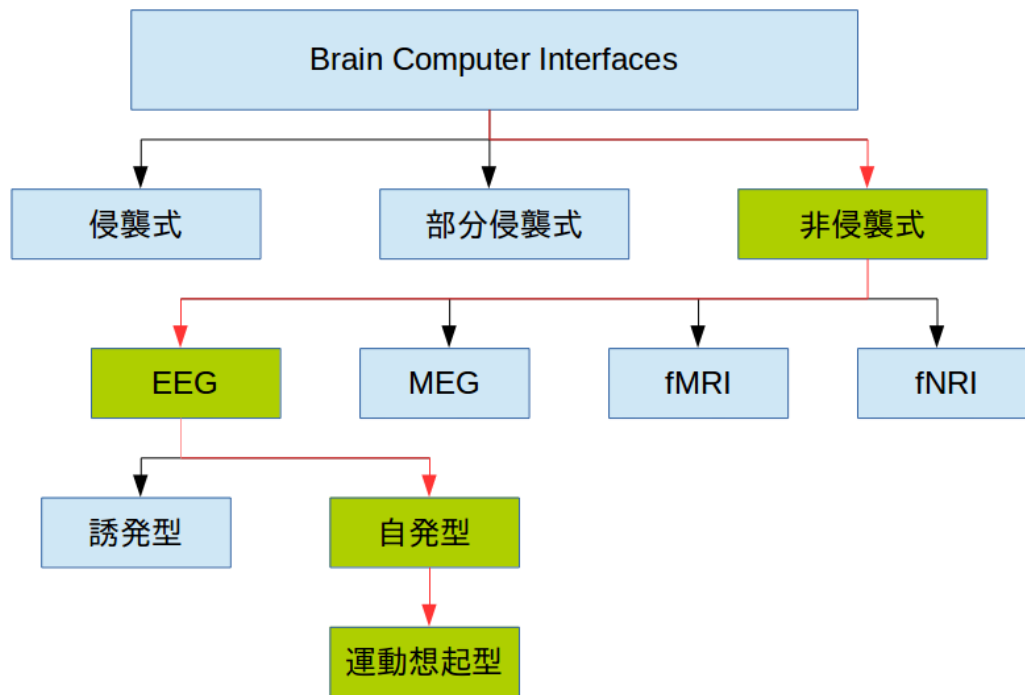


図 1.4: BCI の分類と本研究の焦点

たい場合は右手の運動を想起する。また、下に動かしたい場合は左足、上に動かしたい場合は右足、というように想起する身体部位に応じて外部機器への制御信号を対応させることが可能である。同様の応用方法が車いすなどにも適用できる。運動想起型 BCI を使う利点の 1 つは、運動想起によって生成された EEG 信号は、物体の想像、あるいは抽象的な概念を想像する他の精神的イメージのタスクと比較して一貫性がある点である。一般に、運動想起によって活性化される神経は、運動を実際に行う場合と同様であるとされる [2]。本研究では運動想起型 BCI に焦点を当てる。これまでの BCI の分類について図 1.4 に示す。

BCI の動作原理

BCI の動作は以下のスキームに従う。

1. 脳信号の獲得：センサによりアナログ信号を獲得しデジタル信号へ変換
2. 信号の前処理：データの成形及びアーチファクトの除去
3. 特徴量抽出：神経科学や統計に基づいた特徴量の選定
4. 分類：特徴量から閾値に基づいて意図を分類
5. 制御信号出力：分類結果に基づいて外部機器へ信号を出力

BCI の種類に関わらず動作原理の根本は同様であるが、スキームのどの段階に課題が生じるかは異なる。侵襲式の場合は脳信号の獲得自体が非常に困難であり、安全性やメンテナンス性に課題が生ずる。非侵襲式の場合は脳信号の獲得は気軽に実施できるが、空間分解能や時間分解能の問題、あるいはアーチファクトの存在に

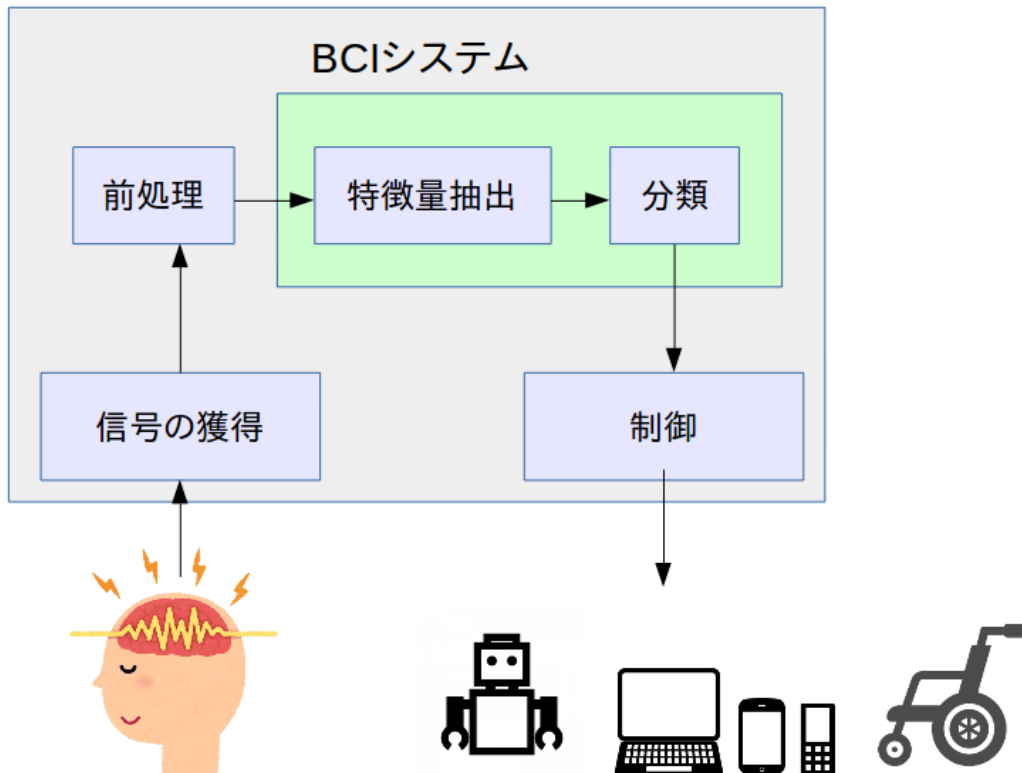


図 1.5: BCI の動作スキーム

よって意図を復元することは容易ではない。

EEG の場合は個々の脳細胞の活動時に発生した電位の総和を頭皮上で測定しており、深さの情報は失われている。これは侵襲式計測や MEG および fMRI に対する明確な欠点である。また、頭蓋骨や人体の細胞を電位がどのように伝搬しているかは定かではないため、頭皮上の位置と脳の表面のみを考えた場合でも位置情報は不明瞭となる。これは部分侵襲式に対する欠点となる。従って EEG を用いた運動想起型 BCI に焦点を当てた場合、主に特徴量抽出と分類の段階が課題となる (図 1.5)。

運動想起型 BCI 周辺の研究

運動想起時には運動野付近で特定の周波数帯域において活動電位が減少する事象関連脱同期 (Event Related Desynchronization: ERD) の存在が知られており、神経科学的な、あるいは BCI 応用のための研究が盛んに行われている [3-5]。従って特徴量として ERD が用いられる研究も数多くあり、EEG を用いた運動想起型 BCI でも基礎となっている [6-8]。ERD を発見することができた場合、EEG の振幅あるいはパワーに閾値を設けて分類を行うことができる。具体的にはスペクトル解析によって各周波数のパワーを推定し、パワーに閾値を設けることとなる。しかし、EEG からの ERD の検知は必ずしも容易ではない。ERD が生じる頭皮上の位置と周波数は個人差がある上、有効な周波数スペクトル解析に関しても未だ研究段階である [9]。

また、人の随意運動の約 0.5 秒から 1 秒前に現れる運動準備電位も特徴量になると考えられている。実動作以前に発生するという極めて特殊な現象であるため、その特性から BCI が人の意志に対して優れた反応速

度で動作することが期待される。BCI の研究としては実肢体動作の分類 [10] や実指動作の分類 [11] などがある。実運動を行わなくとも運動準備電位が特徴量として活用できるという研究 [12, 13] もある。

一方で特定の EEG の現象を直接獲得せずに運動想起型 BCI を構築する手法も提案されており、それらの多くが統計的信号処理や機械学習を活用したものである。特に Common Spatial Pattern(CSP) と呼ばれる統計的信号処理 [14, 15] は、EEG 解析のために発案されて以降、現在までに様々な派生手法を生み出している [16–19]。その中でも Filter Bank Common Spatial Pattern [20] と呼ばれる CSP の発展手法は、ERD に基づく BCI も含め他の手法と比べ高い性能を有している。

現在、EEG の現象として知られている既知の特徴量を検知する手法から、機械学習や統計的信号処理を活用した特徴量抽出手法と高度な分類器を組み合わせ高い性能を発揮する手法の模索が行われている [21]。

1.3 目的

1.3.1 モチベーション

特徴量抽出手法や分類器を拡張する方針とは異なり、より綿密に EEG の解析を行うことで性能の向上を達成しようとする試みも続いている [22–24]。EEG では BCI に限らず、頭皮領域と周波数帯域に関する研究が盛んに行われてきたため、特徴量抽出を行う場合にも電極と周波数に着目するが多い。特に ERD は特定の周波数における電位の減少であるため、スペクトル解析と時間周波数解析が有効利用でき、現在も運動想起型 BCI のための研究が行われている [9]。

一方で現在、運動想起型 BCI で高い性能を誇る手法は統計的信号処理や機械学習手法に基づいている。このことは不可解なことではない。綿密な EEG の解析を実施するか否かに関わらず、左手を動作させる場合と右手を動作させる場合とでは脳の活動は事実異なっているためである。タスク間での EEG の違いを可視化したい場合や、脳機能自体を解明したい場合には綿密な EEG の解析は有効であるが、その際には解析はあくまで人間が解釈する手段であり、解析のために処理された信号がそのまま特徴量として有望だとは限らない。なぜなら人間がデータを解釈、あるいは可視化できるような形に加工することで分類に有用な情報が失われている可能性もあるためである。

しかし、運動想起型 BCI を構築するための特徴量抽出手法、分類手法は数多く存在する。代表的なものを以下に記す。この中の幾つかは第 2 章にて解説する。

- Laplacian Filter(LF)
- Principal Component Analysis(PCA)
- Independent Component Analysis(ICA)

- Canonical Correspondence Analysis(CCA)
- Common Spatial Pattern(CSP)
- バンドパスフィルタバンク
- フーリエ変換
- ウェーブレット変換
- 自己回帰モデル
- Empirical Mode Decomposition(EMD)
- Linear Discriminant Analysis(LDA)
- Support Vector Machine(SVM)
- Logistic Regression(LR)

これらの手法は BCI の構成要素として組み込まれるが数多くある手法の中から個々人に応じて、あるいはタスクに応じて適切に組み合わせるのは容易ではない。その中、FBCSP はバンドパスフィルタバンクと CSP 及び LDA を組み合わせた数少ない成功例であり、現在でもバンドパスフィルタと CSP を基本とした派生手法が提案され続けている [25,26]。しかし、数多くの手法が提案されている中で、FBCSP 以降は運動想起型 BCI としての決定的なアーキテクチャは提案されていないと言える。

1.3.2 達成目標

本研究では音声認識や画像認識の分野で高い性能を発揮しているニューラルネットワークに着目する。ニューラルネットワークは学習時の計算量が膨大であることから、長い間研究が滞っていたが、近年の計算機の発展により応用研究に用いられるようになり、人間の設計した分類器を凌駕する性能の高さから注目が集まっている。

ニューラルネットワークの実体は巨大な合成関数であり、事実上単なるパラメトリックな数理モデルである。しかし、誤差逆伝播学習によって原理上極めて深い合成関数の形式であっても学習が可能であるため、問題に応じてネットワーク構造の設計を適切に行うことで、特徴量抽出と分類を同時に達成できる可能性がある。更にその学習のアルゴリズム自体は極めて単純であるが効果的に働き [27]、再学習可能であるため有用なニューラルネットワークの構造が発見された場合にその再利用性が高い。加えて、ニューラルネットワークの深層構造によって高い性能が引き出されること [28] や、モデルの複雑さと比較して最適化問題の目的関数は病的ではない、あるいは病的な形状を回避できることが示唆されるようになった [29]。

従って、運動想起型 BCI に適したニューラルネットワークの検討を行い、基礎的な構造を確立することは有益であると考え、本研究の目的とする。

1.3.3 技術的貢献

典型的な運動想起型 BCI は、運動想起に関連している EEG を取り出すための前処理 $\mathcal{H}(\cdot)$ によって EEG の生データ X から \hat{X} を獲得することが一般的である。

$$\hat{X} = \mathcal{H}(X) \quad (1.1)$$

次に、運動想起に関連している空間的な情報を抽出する処理を $f(\cdot)$ を適用し、特徴量 Z を取り出す。

$$Z = f(\hat{X}) \quad (1.2)$$

続いて Z に対して、運動想起部位 Y を出力する分類器 $g(\cdot)$ を準備することで、運動想起 BCI が構成されている。

$$Y = g(Z) \quad (1.3)$$

従って、BCI は EEG X を引数とした合成関数という形式を取る。

$$Y = (g \circ f \circ \mathcal{H})(X) \quad (1.4)$$

実際に合成関数としてどのようなものが選択されるか、EEG がどのように測定されるかはタスクに依存するが、典型的な CSP を用いた BCI では X を運動想起時の EEG、 \mathcal{H} をバンドパスフィルタ、 f を CSP、 g を LDA や SVM として各々個別に設計する。

一方で時間周波数解析に基づく BCI でも特徴抽出として何らかの時間周波数解析 $h(\cdot)$ が関数内に挿入され、

$$Y = (g \circ h \circ f \circ \mathcal{H})(X) \quad (1.5)$$

という形式で表せる。この時、 \mathcal{H} や f は ERD を検出するための神経科学的な知見に基づいた設計がなされる場合もあれば、機械学習の手法が用いられる場合もある。更に時間周波数解析によって得られるパワースペクトログラムに対して非負値行列分解などを用いて特徴量を抽出する試みもある [30, 31]。この場合も行列分解による変換を $a(\cdot)$ と置けば

$$Y = (g \circ a \circ h \circ f \circ \mathcal{H})(X) \quad (1.6)$$

と表され、形式上は合成関数である。それぞれの関数の役割を明示しなければ、BCI は単に以下の合成関数である。

$$Y = (f_K \circ \cdots \circ f_1)(X) \quad (1.7)$$

すなわち BCI は設計を終えた時には何らかの合成関数として構築されている。実際に設計を行う際には各々

の合成関数に対して関数族を仮定し、学習やヒューリスティクスによって関数を決定し、最後に合成する。しかしこのとき、BCI のデータの流れから明らかに関数 f_i を設計するためには f_{i-1} の設計が終了していなければならない。

ニューラルネットワークは万能な関数近似器である。ニューラルネットワークの構造を与えることで適切な関数族を仮定することができれば、誤差逆伝播学習によって End to End での BCI 設計が可能となる。従って以下の項目において貢献することができる。

- 各個人における EEG の解析の必要性を排除

また、ニューラルネットワークの応用研究が盛んな画像認識の分野では、既に模範的なニューラルネットワークの構造が発見されており、大量の画像によって事前学習を行い、達成したい認識対象を絞り込んだ後に転移学習することで手軽に高い性能の分類器が得られる。従って、BCI において模範的なニューラルネットワークの構築により以下の事項が達成できる可能性が示唆される。

- 人類に共通した一般的な BCI の基本モデル構築

第 2 章

BCI のための要素技術

2.1 周波数解析

スカラーの時間波形 $x(t)$ に対して、周波数スペクトルを算出する際に用いられるフーリエ変換について述べる。 t が連続量である時 $x(t)$ をアナログ信号と呼び、離散的であればデジタル信号と呼ぶ。このセクションではアナログ信号とデジタル信号が混在するため、アナログ信号に関しては $x(t)$ と表記し、デジタル信号を x_n と表記することで明確に区別する。

2.1.1 4 種類のフーリエ変換

時間を連続的に扱うか離散的に扱うかの違いだけでなく、周波数を連続的に扱うか否かによってもフーリエ変換の式は異なっており、 $2 \times 2 = 4$ つのフーリエ変換がある。通常はフーリエ変換について議論する場合は時間も周波数も連続であるとみなした式を使う場合が多いが、コンピュータ上で数値の処理を行う場合は原則離散的であるため、4 つのフーリエ変換について全て簡単に説明する。

フーリエ変換

スカラーの時間波形 $x(t)$ に対して、フーリエ変換は以下の (2.1) で表され、逆フーリエ変換の式は (2.2) で表される。

$$X(\Omega) = \int_{-\infty}^{\infty} x(t) \exp(-i\Omega t) dt \quad (2.1)$$

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\Omega) \exp(i\Omega t) d\Omega \quad (2.2)$$

ここで、 i は虚数単位であり $X(\Omega)$ は Ω を変数とするスカラー関数である。逆フーリエ変換はスカラーの時間波形 $x(t)$ をある周波数 Ω の波形 $\exp(i\Omega t)$ の線型結合によって表現した式である。このときの結合係数が $X(\Omega)$ である。工学的な立場では時間波形 $x(t)$ が与えられた時に、波形 $\exp(i\Omega t)$ による線型結合をすること

で、波形にどのような周波数の波形がどれくらいの割合で含まれているかを知りたいケースがある。それを可能にするのがフーリエ変換であり、 $x(t)$ から結合係数 $X(\Omega)$ を算出することが可能である。通常はこの時の結合係数 $X(\Omega)$ を周波数スペクトルと呼ぶ。逆フーリエ変換 (2.2) に登場する定数倍 $1/2\pi$ は、フーリエ変換 $\mathcal{F}(\cdot)$ と逆フーリエ変換 $\mathcal{F}^{-1}(\cdot)$ が、フーリエ変換可能な時間波形 $x(t)$ に対して $\mathcal{F}^{-1}(\mathcal{F}(x(t))) = x(t)$ となるように調整するための係数である。この係数は逆フーリエ変換の式ではなくフーリエ変換の式に付いていてもよく、あるいは両方の式に平方根の形で分配されていても構わない。またフーリエ変換可能な時間波形とは以下を満たす $x(t)$ である

$$\int_{-\infty}^{\infty} |x(t)| dt < \infty \quad (2.3)$$

離散周波数フーリエ変換

取りうる周波数を離散的にした場合のフーリエ変換について説明する。離散周波数フーリエ変換という命名は便宜的にこの論文内で行っているものであり一般的ではない。通常、ここで紹介するフーリエ変換は「フーリエ級数展開」として知られている。歴史的には微分方程式を解くために開発され、フーリエ変換よりも先に発見されている。離散周波数フーリエ変換と離散周波数逆フーリエ変換はそれぞれ (2.4) と (2.5) で表される。

$$X_k = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} x(t) \exp(-i\Omega_0 kt) dt \quad (2.4)$$

$$x(t) = \sum_{k=-\infty}^{\infty} X_k \exp(i\Omega_0 kt) \quad (2.5)$$

離散周波数逆フーリエ変換は、スカラーの時間波形 $x(t)$ を離散的な周波数 $\Omega_0 k$ の波形 $\exp(i\Omega_0 kt)$ の線型結合によって表現した式である。このときの結合係数が X_k である。ここに T_0 は時間波形の周期であり、 $\Omega_0 = 2\pi/T_0$ を基本周波数と呼ぶ。離散周波数フーリエ変換が、時間波形が与えられた時の周波数スペクトルを算出する役割を担うことは、フーリエ変換と同様である。フーリエ変換との最たる違いは、時間波形に対して周期性を仮定している点であり、その周期は既知でなければならない。時間波形が周期性を持つ場合には周波数は離散的な値を取る。

離散時間フーリエ変換

取りうる時間を離散的にした場合のフーリエ変換について説明する。離散時間フーリエ変換と離散時間逆フーリエ変換はそれぞれ (2.6) と (2.7) で表される。

$$X(\omega) = \sum_{n=-\infty}^{\infty} x_n \exp(-i\omega n) \quad (2.6)$$

$$x_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) \exp(i\omega n) d\omega \quad (2.7)$$

離散時間逆フーリエ変換が、時間波形 x_n を周波数スペクトル $X(\omega)$ を結合係数とした $\exp(i\omega n)$ の線型結合を表しているのはこれまでと同様である。しかし、離散時間フーリエ変換では周波数スペクトル $X(\omega)$ が周期的な関数となることは強調しておかねばならない。

離散フーリエ変換

離散フーリエ変換は、取りうる時間も周波数も離散的であるとした場合のフーリエ変換である。離散フーリエ変換と離散逆フーリエ変換はそれぞれ (2.8) と (2.9) で表される。

$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(-i\frac{2\pi}{N}kn\right) \quad (2.8)$$

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k \exp\left(i\frac{2\pi}{N}kn\right) \quad (2.9)$$

ここで N はサンプル時間点数の意味で周期である。時間波形に対する周期は N であるが、周波数スペクトルに対する周期も N となる。仮に 1 周期が 1000 点のサンプル点によって構成される時間波形に対し離散フーリエ変換を用いた場合は、得られる周波数スペクトルは 1000 点で周期を有する形式となる。離散フーリエ変換は有限の数列から有限の数列への変換であり、他のフーリエ変換と異なり無限大を扱う必要はないためコンピュータ上で計算を実行することが可能である。高速フーリエ変換と呼ばれる実応用で頻繁に用いられるアルゴリズムは、離散フーリエ変換 (2.8) を高速に実行する手続きのことである。

2.1.2 パワースペクトル密度

パワースペクトル密度 $PSD(\Omega)$ とは時間信号 $x(t)$ の周波数スペクトルを $X(\Omega)$ とした時、

$$PSD(\Omega) = |X(\Omega)|^2 \quad (2.10)$$

に相当する関数である。信号 $x(t)$ のエネルギーが周波数 Ω に関してどのように分布するかを示している。コンピュータでパワースペクトル密度を計算する場合は (2.8) で計算される X_k の二乗 $|X_k|^2$ を算出する。しかし、数学的な定式化を行う上では以下の定義が用いられ、 $|X_k|^2$ はピリオドグラムと呼び区別する。

$$PSD(\Omega) = \int_{-\infty}^{\infty} R(\tau) \exp(-i\Omega\tau) d\tau \quad (2.11)$$

ここに、 $R(\tau)$ は

$$R(\tau) = \mathbb{E}[x(t)x(t+\tau)] \quad (2.12)$$

であり、信号 $x(t)$ の自己相関関数である。(2.11) は、ウィーナー・ヒンチンの定理としても知られており、自己相関関数のフーリエ変換がパワースペクトル密度になることを示している。この定義から明らかなようにパ

ワースペクトル密度とは統計量であって、解析的に算出されるのではなく、推定されるものである。

ピリオドグラムはパワースペクトル密度の推定を行う手段として用いられる。一般的にピリオドグラムによる推定値は平均と標準偏差が同じ大きさを持ち、時系列の長さを長くしても推定誤差は改善されない。時系列を長くすることで周波数分解能を高くすることはできるが、個々の周波数でのスペクトルの相対誤差は変化しない。そこで、通常は何らかの平滑化を行って個々の周波数成分の推定値の誤差を減少させる方法が一般に採用される。主な方法として、時間領域で波形を分割し、複数のピリオドグラムの平均を算出する方法がある。時間領域で波形を分割する際には、波形が時間で互いに重なりを持つように分割する、Welch のオーバーラッピング法が用いられる。また最大エントロピー法を用いてパワースペクトル密度を推定する方法もあり、その場合には (2.11) を制約条件としたエントロピー最大化の変分問題を解くこととなる。このパワースペクトル密度推定方法はバーク法として知られている。スペクトル解析に関する記述は書籍 [32] に詳しい。

脳波は、運動想起時に特定の周波数領域でエネルギーが減少する事象関連脱同期が生じるとされているため、適切な電極選定を行い、脳波のパワースペクトル密度 $PSD(\Omega)$ を推定することができれば、とある Ω で著しくパワーが減少する様子が確認できる。

2.2 多変量解析

多変量解析とは多変量データを統計的に扱う手法である。扱うデータの科学的知識に基づいた特徴量抽出に加え、実データの統計的な性質を考慮した特徴量抽出も様々な分野で行われている。この章では脳波解析で用いられている多変量解析について紹介し、その有用性と限界について考察する。

2.2.1 Principal Component Analysis(PCA)

Principal Component Analysis(PCA) は特徴量抽出手法として幅広い分野で活用されている。簡単のため時間平均が 0 の多次元信号 $x(t) \in \mathbb{R}^D$ に対して PCA による特徴抽出を考える。PCA では変換行列 $W^T \in \mathbb{R}^{D \times D}$ を左から作用させ、特徴量

$$z(t) = W^T x(t) \in \mathbb{R}^D \quad (2.13)$$

を獲得するが、この際、 $z(t)$ の各成分が互いに無相関になるように W を決定する。 $z(t)$ が無相関となるためには、その分散共分散行列 $\Sigma_z = \mathbb{E}[z(t)z^T(t)]$ が対角行列になることが要請される。ここで $x(t)$ の分散共

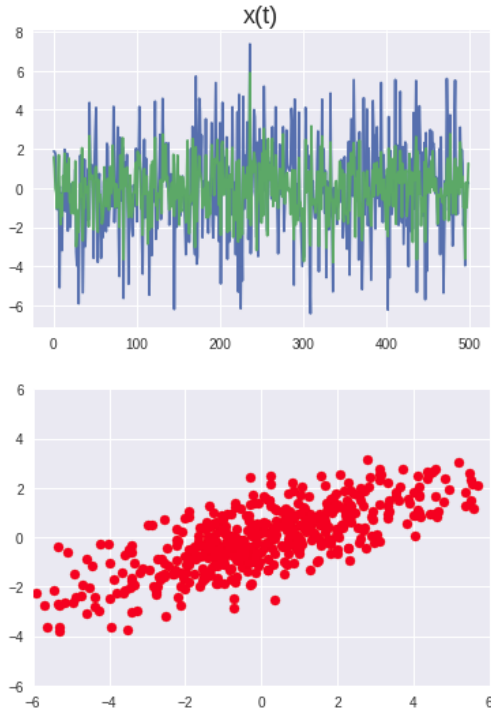


図 2.1: $x(t)$ の波形と散布図

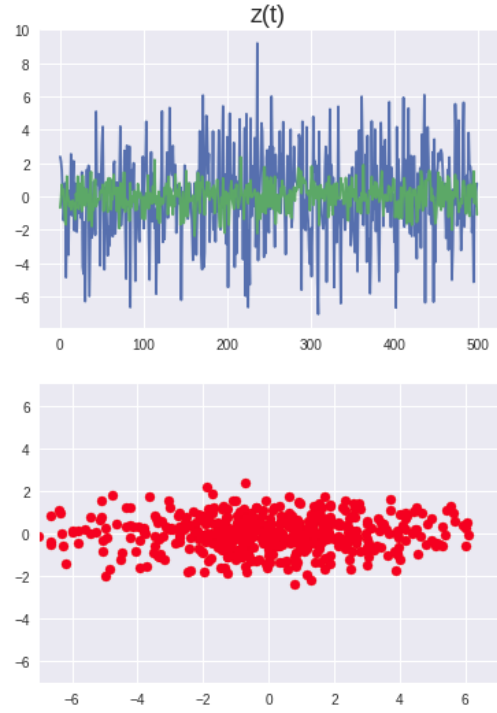


図 2.2: $z(t)$ の波形と散布図

散行列を $\Sigma_x = \mathbb{E}[x(t)x^T(t)]$ とすると、 $z(t)$ の分散共分散行列 Σ_z について

$$\begin{aligned}
 \Sigma_z &= \mathbb{E}[z(t)z^T(t)] \\
 &= \mathbb{E}[W^T x(t)x^T(t)W] \\
 &= W^T \mathbb{E}[x(t)x^T(t)]W \\
 &= W^T \Sigma_x W
 \end{aligned} \tag{2.14}$$

と表すことができる。(2.14) が対角行列になるような W は、 Σ_x の固有値分解によって求まる。今、 Σ_x の D 個の固有値を $\lambda_1 \geq \dots \geq \lambda_D$ とする。この固有値を対角成分に並べた行列を $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_D)$ とすると、 Σ_x の固有値分解は、ある U が存在して以下の形式となる。

$$\Sigma_x = U \Lambda U^{-1} \tag{2.15}$$

U^{-1} を左から、 U を右から掛けることで $U^{-1}\Sigma_x U = \Lambda$ が得られ、 $U^{-1} = W^T$ で $U = W$ とすれば (2.14) への要請を直ちに満たす。これは $W^T = W^{-1}$ という条件が満たされれば良く、実数信号の分散共分散行列 Σ_x は一般に正定値実対称行列となっており、直交行列によって固有値分解が可能であるため条件を満たす。また、このとき全ての i について固有値 λ_i は正の値となり、固有値 λ_i に属する固有ベクトルにデータを射影した際の分散を表す。これらの数学的な扱いやすさから PCA は非常に広く普及している。

PCA の働きを見るため $D = 2$ とした場合の線形変換 $z(t) = W^T x(t)$ を図で確認する。図 2.1 は人工的に

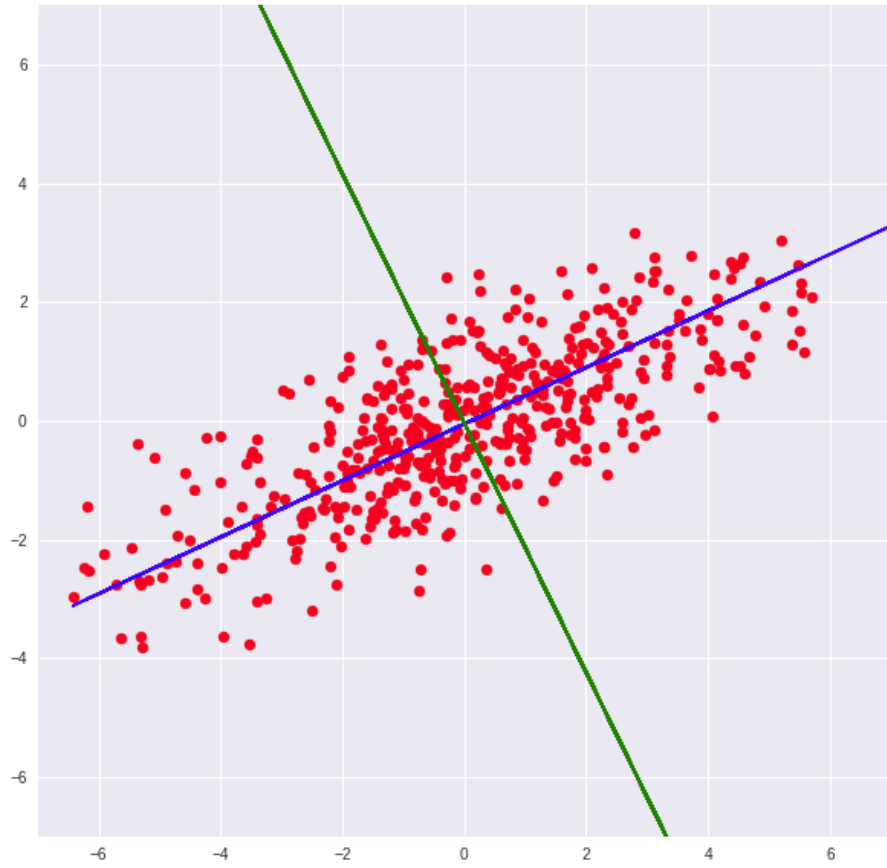


図 2.3: PCA によって得られる基底

作成した 2 次元波形 $x(t)$ と各成分の散布図である。この $x(t)$ に対して PCA を用いると、図 2.2 に示す $z(t)$ が得られ、散布図から $z(t)$ の各成分は無相関となっていることが確認できる。図 2.3 は $x(t)$ に対して PCA を施した場合に得られる新たな直交基底を表しており、 $z(t)$ は新たな直交基底に $x(t)$ を射影したものに他ならない。PCA では基底を取り直すことで各成分が無相関な信号を獲得でき、その結果、新たに得られた信号の各成分がどのような意味を持つのかを考察しやすくなる。

しかし、応用上は単に基底を取り直すことを目的とするケースは少ない。通常は固有値分解によって求めた固有ベクトル（新たな基底）を全て利用するのではなく、値の大きな固有値に属する固有ベクトル w を $d(< D)$ 個選び、 $W_d = (w_1, \dots, w_d)$ によって変換行列を構成することで、

$$z(t) = W_d^T x(t) \in \mathbb{R}^d \quad (2.16)$$

と次元削減を行う。値の大きな固有値に属する固有ベクトルによって基底を構成することは、 W_d による基底の元で、信号の分散（あるいは振幅）が最大化されることを要請することと等価である。また、射影先での分散最大化に伴い、(2.16) での変換行列 W_d は、元々の信号 $x(t)$ と、 $x(t)$ を \mathbb{R}^D の部分空間 \mathbb{R}^d へ射影した信号 $z(t)$ との二乗誤差を最小化する変換行列となっている。

以上から PCA は、元々の信号 $x(t)$ の情報損失を二乗誤差の意味で最小限に抑えながら、射影先で大きな変

動を有し、かつ各成分が無相関となる特徴量 $z(t)$ を抽出する。しかし、脳波への応用を考える上では PCA の性質は必ずしも有効には働かない。運動想起 BCI を考える上では、脳波に含まれる全ての情報の中から識別したい身体部位に関する情報のみを抽出する必要がある。この場合、脳波信号のごく一部のみが重要である可能性があり、射影先で大きな分散を持つような信号となっているかは定かではないためである。

2.2.2 Independent Component Analysis(ICA)

Independent Component Analysis(ICA) は PCA を発展させた比較的新しい信号解析手法である。ICA は PCA と同様に変換行列 $W^T \in \mathbb{R}^{D \times D}$ を左から作用させ、特徴量

$$z(t) = W^T x(t) \in \mathbb{R}^D \quad (2.17)$$

を獲得するが、この際、 $z(t)$ の各成分が互いに独立になるように W を決定する。独立性は無相関性の十分条件であり、PCA に比べて $z(t)$ により強い条件を要請する。PCA では無相関性が固有値分解という数学的によく知られた問題と関連していたが、独立性は簡単な問題への定式化は困難であるため、通常は独立性を測る目的関数を設定し、勾配法などの逐次最適化法を用いる。ICA の求解アルゴリズムを述べるのは本研究テーマから逸脱するため、主要なアルゴリズムに絞って簡単に説明する。最も普及している求解アルゴリズムは FastICA として知られており、独立性の必要条件である無相関性を要請した後、独立性を最大化することで解を得る。無相関性を満たすようにするためには前処理として PCA を用いることができる（正確には PCA を用いた中心化と白色化処理が行われる）。前処理後は独立性の最大化問題が、 $z(t)$ の各成分のエントロピー最小化問題に変換されるが、この最小化問題も容易ではない。従って更にネグエントロピーなる量を導入し、最大化問題に書き換える。ネグエントロピーは信号の非ガウス性を測る尺度であり、ガウス分布から遠いほど大きな値となる。FastICA ではネグエントロピーの最大化を各成分ごとに順次取り出していく。この時、非ガウス性が高い成分の順番に取り出されていく。独立成分分析の解説は次の文献が詳しい [33] [34]。

図 2.4 に PCA と ICA の振る舞いを Toy データを用いて比較を掲載する。図 2.4 の base vectors of ICA and PCA から分かるように ICA で得られる基底は直交するとは限らない。また、PCA の基底の大きさは等しく 1 であり正規直交基底を構築するが、ICA では正規性も持つとは限らない。

脳波への応用では、脳波計測時に混入した脳波以外の信号（筋電位、眼電位など）が脳波とは統計的に独立であると考え、脳波以外の信号成分を除去する目的で利用される。一方で運動想起 BCI において身体部位に関する脳波成分を直接抽出することは PCA 同様に難しい。筋電などの場合、脳波と独立であるという仮定は妥当であり、かつ振幅が目視可能なほど脳波に比べて大きくなる。従って、ICA によって分解された独立な成分から筋電などを見分けるのは比較的容易である。しかし、一方で身体部位に関する脳波成分がその他のあらゆる脳波と独立であるかは定かではなく、仮に独立であった場合にも分解された信号から目視によって特定す

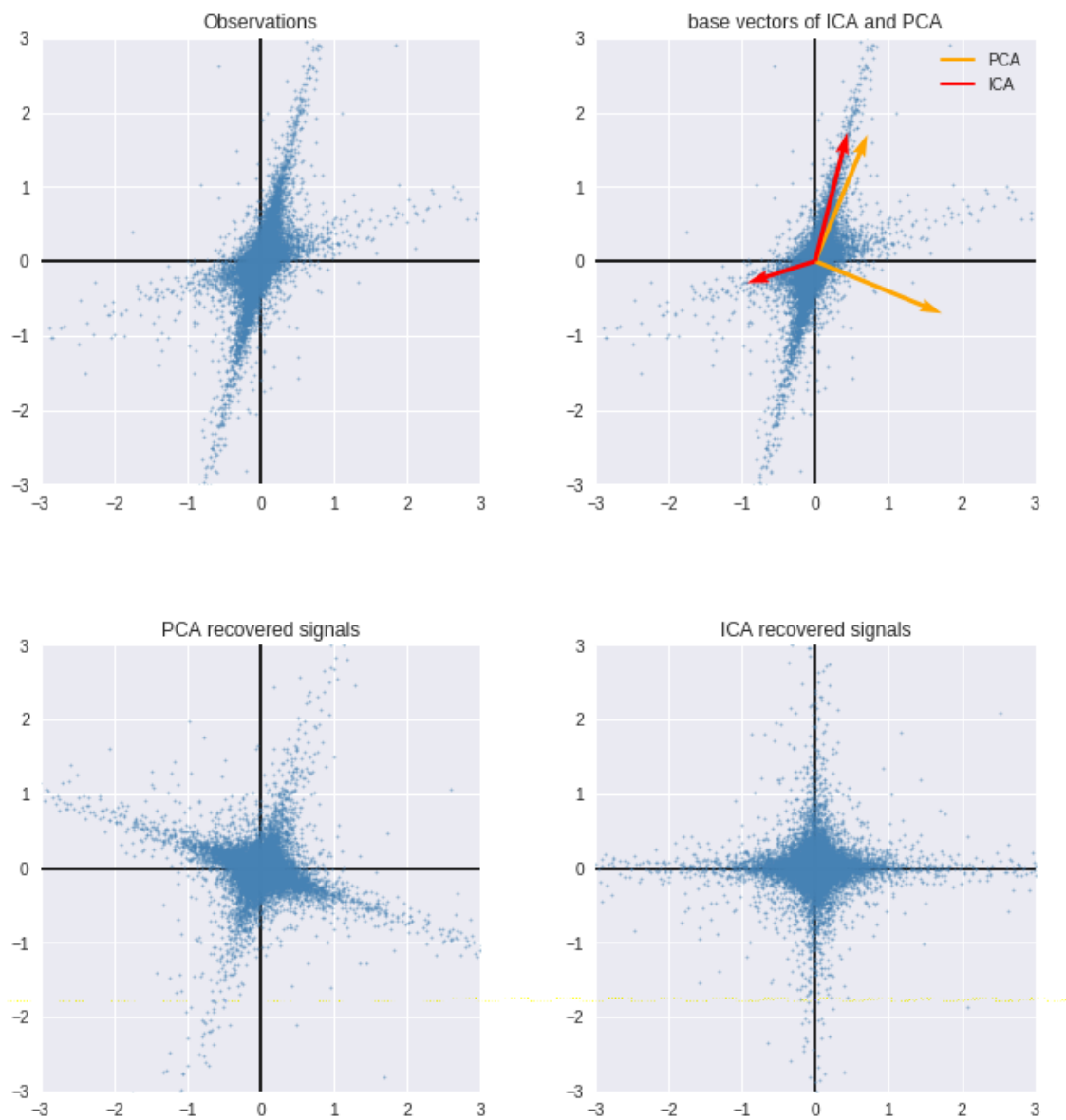


図 2.4: 無相関な基底を得る PCA と独立な基底を得る ICA の比較

ることは困難であると推察される。

2.2.3 Blind Source Separation(BSS)

上記では PCA と ICA が次元削減として用いられることを見た。一方でこれらの手法は Blind Source Separation(BSS) 問題の解法として解釈されることも多いため、ここで簡単に述べておく。まず信号源 $s(t) \in \mathbb{R}^d$ を直接観測できない場合に、 D 個のセンサで $x(t) \in \mathbb{R}^D$ という信号を観測したとする。このとき、観測信号 $x(t)$ のみから信号源 $s(t)$ を推定する問題が BSS 問題である。計測機器や環境に応じて、信号源 $s(t)$ は何らかの変換 $f(\cdot)$ を受けて観測されると考えられる。従って $x(t)$ は

$$x(t) = f(s(t)) \quad (2.18)$$

と表記できる。このときに観測に伴う変換 $f(\cdot)$ が線形変換 A であると仮定した場合、

$$x(t) = As(t) \quad (2.19)$$

と表記することができ、BSS 問題は $x(t)$ から A と $s(t)$ を同時に推定する問題であると見なせる。ここで仮に適当な線形変換によって、観測信号 $x(t)$ を

$$z(t) = Wx(t) \quad (2.20)$$

と変換することを考える。 W を上手く選ぶことに成功すれば、 $z(t) = Wx(t) \simeq s(t)$ となることが期待できる。ここで $D = d$ 、すなわち信号源の次元と観測信号の次元が一致している場合を考える。このとき (2.19) において、 A が正則であるとし、

$$s(t) = A^{-1}x(t) \quad (2.21)$$

と表すことが可能になる。従って、(2.20) の W を $W = A^{-1}$ とすることができれば、

$$z(t) = Wx(t) = A^{-1}x(t) = s(t) \quad (2.22)$$

と信号源を求めることが可能である。ただし、 $z(t) = s(t)$ となる W が存在するとしても、既知の $x(t)$ に対して未知の $W, s(t)$ を求めようとしている状況に変わりはなく、BSS 問題は基本的に不良設定問題である。また、実データでは信号源と観測信号の次元が一致しない場合が多く A は逆行列を持たないため、状況はより複雑である場合が多い。通常は BSS 問題を解くためには何らかの条件を追加するか、正則化の手法を導入する必要がある。PCA や ICA は信号源 $s(t)$ が各成分について無相関あるいは独立であると仮定することで条件式を追加し、観測信号 $x(t)$ から条件式を満たすような W と $z(t)$ を求め、 $z(t)$ が信号源 $s(t)$ の良い近似になっていると考える BSS 問題の解法の一つである。

PCA と ICA の BSS 問題への振る舞いを確認するために、トイデータによる実験結果を図 2.5 に示す。周波数の異なる 2 つの正弦波と 1 つのノコギリ波にそれぞれガウスノイズを加算した信号源 (図 2.5 の True Sources) を準備し、適当な線形変換を施して観測信号 (図 2.5 の Observations) とする。図 2.5 の ICA recovered signals が FastICA によって推定された信号源に適当なゲインを加えたものであり、図 2.5 の PCA recovered signals が PCA によって推定された信号源に適当なゲインを加えたものである。ICA では周波数の異なる正弦波とノコギリ波を明確に分解できており、信号源に近い波形が得られていることが確認できるが、PCA では信号源と異なる信号が得られている。PCA の振る舞いは各成分を無相関にしつつ、射影先で分散を最大化するような基底を求めるためノコギリ波と位相が一致している正弦波を 1 つの成分に集約してしまっている。

トイデータによる実験では ICA が BSS 問題に対して有効に働くことが確認できるが、本来信号源がどのよ

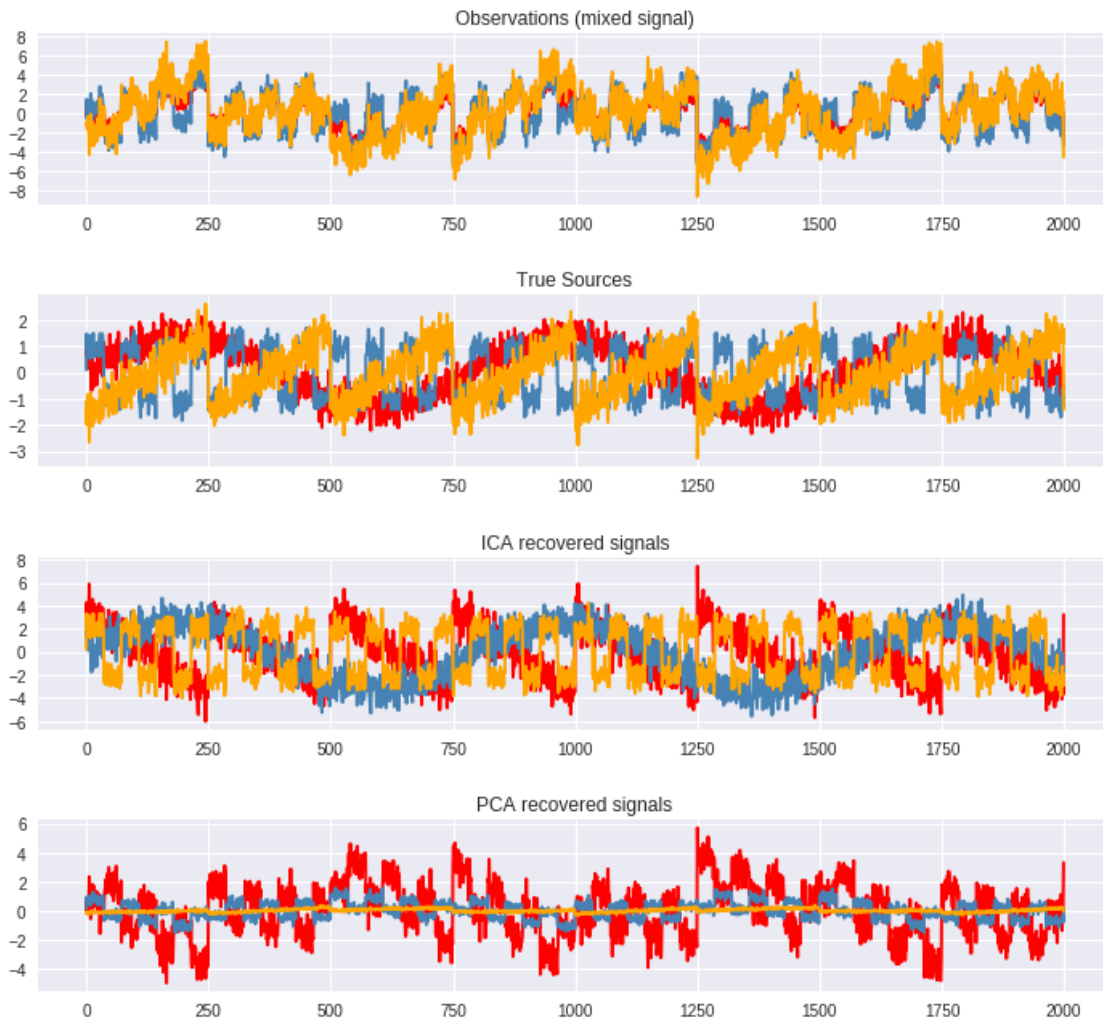


図 2.5: BSS 問題に対する PCA と ICA

うなものであるかは未知であるため、信号源推定が正しく行われたかを確認するのは実データでは困難である。運動想起 BCI を想定して脳波に BSS 問題を適用する場合には、観測された信号 $x(t)$ から脳波信号の根源である $s(t)$ を復元することを目的とするが、ICA によって推定された信号源のいずれの成分が運動想起と関連しているのかを判別するのは極めて難しい。ただし、脳波と筋電では波形が明らかに異なるため、筋電と脳波が独立であるという仮定を用いて ICA によって筋電成分を取り出すことは可能である。一般に、不良設定問題である BSS 問題に対しては何らかの仮定を置かざるを得なく、複雑な脳波に対して適切な仮定を設けることが重要な課題となる。

2.3 分類手法

2.3.1 Linear Discriminant Analysis(LDA)

Linear Discriminant Analysis(LDA) は統計分析において伝統的に用いられてきた歴史ある手法である。LDA では多次元データを部分空間で切り取り、切り取った空間で分類超平面を構築することでクラス分類を行う。

分類超平面を構築する手段を与えなければ、LDA は特徴量抽出手法としても機能する。まず、多次元データ $x \in \mathbb{R}^D$ を基底 $w \in \mathbb{R}^D$ へ射影すると、以下の式で表されるスカラー値を獲得できる。

$$z = w^T x \in \mathbb{R} \quad (2.23)$$

z に対してある閾値を設定し、 $z \geq -w_0$ の場合はクラス C_1 とし、そうでない場合はクラス C_2 であるとする
ことで分類器を獲得できる。多次元データを 1 次元空間へ射影した場合には多くの情報損失が生ずるが、 w の
取り方を上手く調整することによって、クラス分類を行いやすい射影を選択できる。まず以下のようにクラス
毎の平均ベクトル m_1, m_2 を定義する。

$$m_1 = \frac{1}{|C_1|} \sum_{x \in C_1} x \quad (2.24)$$

$$m_2 = \frac{1}{|C_2|} \sum_{x \in C_2} x \quad (2.25)$$

ここに、 $|C_i|$ はクラス C_i に属するデータの数である。クラス C_1 とクラス C_2 の平均間の距離が射影先で大きな値となれば、異なるクラスのデータは平均的に離れて配置され、クラス分類を行いやすい射影になっていると想定できる。従って、まずは以下の距離の最大化を考慮する。

$$d = |w^T(m_1 - m_2)| \quad (2.26)$$

しかし実際には (2.26) の最大化を考慮しただけでは分類が上手く行くとはいえない(図 2.6)。射影先での各クラスのデータの分散が大きい場合には、異なるクラスのデータが重なってしまう場合が生じるからである。この問題を解決するためにはデータの分散を考慮する必要がある。まず射影先での各クラスの分散は以下で表記できる。

$$\sigma_1^2 = \sum_{x \in C_1} \{w^T(x - m_1)\}^2 \quad (2.27)$$

$$\sigma_2^2 = \sum_{x \in C_2} \{w^T(x - m_2)\}^2 \quad (2.28)$$

ここで、全データのクラス毎の分散の和を総クラス内分散として以下で定義する。

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 \quad (2.29)$$

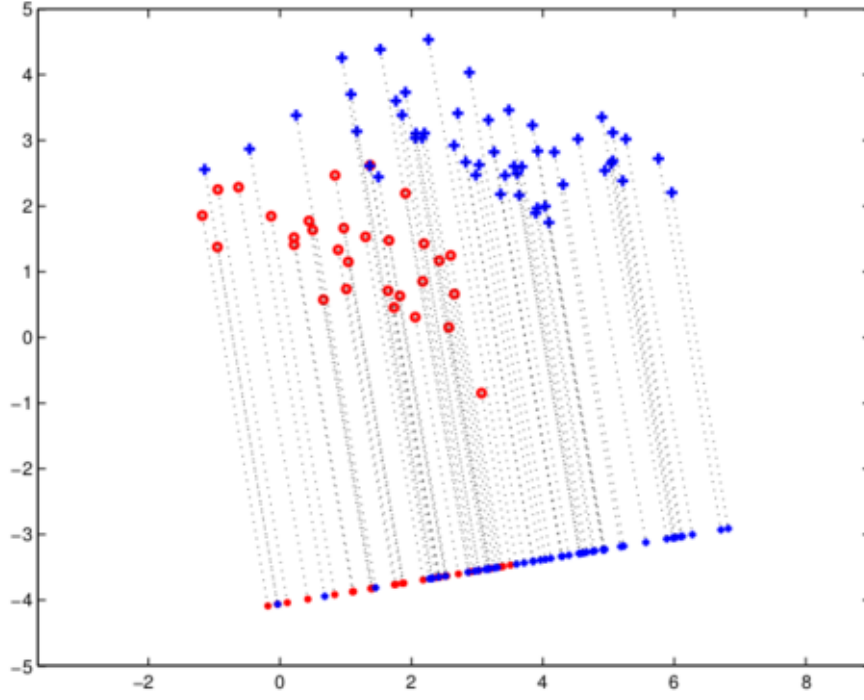


図 2.6: クラス毎の平均間の距離を射影先で最大化した判別分析

総クラス内分散 (2.29) を小さくしながらクラス間の平均の距離 (2.26) を大きくすることを考慮し LDA では以下の評価関数を用いる。

$$J(w) = \frac{d^2}{\sigma^2} = \frac{\{w^T(m_1 - m_2)\}^2}{\sum_{x \in C_1} \{w^T(x - m_1)\}^2 + \sum_{x \in C_2} \{w^T(x - m_2)\}^2} \quad (2.30)$$

このクラス間の平均とクラス内の分散を考慮した評価関数を用いることで、射影先でデータがクラス毎に小さくまとまり、かつ異なるクラスのデータがなるべく離れるようになる (図 2.7)。

1 次元の空間 (数直線上) でクラス毎にデータが上手く分離できた後、数直線上に閾値 $w_0 = (m_1 + m_2)/2$ を設けることで分離平面が得られる。これはクラス毎の平均値の平均値によって分離平面を設定したことに相当する。しかし、射影のされ方によってはこの閾値は適当ではない。より精密な分類を行うためには、 $z = w^T x$ がクラス毎に異なるガウス分布から生じる確率変数だと考え、クラス C_i の条件付き確率 $p(z|C_i)$ を算出した判別基準を設け、条件付き確率が大きいクラスへ分類するなどの方法を取る。

また LDA の評価関数 (2.30) は以下のように変形が可能である。

$$\begin{aligned} J(w) &= \frac{\{w^T(m_1 - m_2)\}^2}{\sum_{x \in C_1} \{w^T(x - m_1)\}^2 + \sum_{x \in C_2} \{w^T(x - m_2)\}^2} \\ &= \frac{w^T(m_1 - m_2)(m_1 - m_2)^T w}{w^T \left\{ \sum_{x \in C_1} (x - m_1)(x - m_1)^T + \sum_{x \in C_2} (x - m_2)(x - m_2)^T \right\} w} \\ &= \frac{w^T S_B w}{w^T S_W w} \end{aligned} \quad (2.31)$$

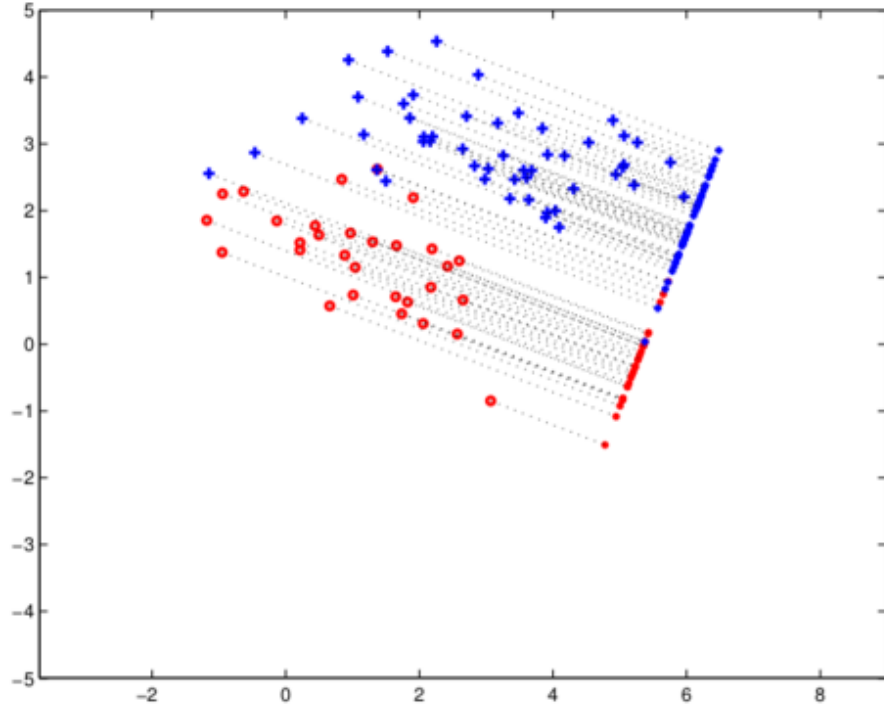


図 2.7: クラス毎の平均間の距離を射影先で最大化し、総クラス内分散を最小化した判別分析

ここで

$$S_B = (m_1 - m_2)(m_1 - m_2)^T \quad (2.32)$$

$$S_W = \sum_{x \in C_1} (x - m_1)(x - m_1)^T + \sum_{x \in C_2} (x - m_2)(x - m_2)^T \quad (2.33)$$

である。このとき S_B をクラス間共分散行列、 S_W をクラス内共分散行列と言う。(2.31) を w に関して微分して 0 と置くことで問題は解析的に解くことができる。また問題は一般化固有値問題となり、複数の固有ベクトルを用いて多次元の特徴量を獲得することも可能である。

2.3.2 Support Vector Machine(SVM)

脳波の分類では Support Vector Machine(SVM) の応用例もある。基本的に SVM はマージン最大化の考えによって汎化性能の向上に成功した 2 クラス分類のための線形分類器である。まずマージン最大化という概念について説明する。マージンとは端的に述べるとデータ点と分類超平面との距離のことを表す。学習データに対してマージンを最大化することで、学習データが空間上で僅かに移動した際にも誤分類を起こしづらくないと期待できる。SVM ではこのマージン最大化によって以下の分類超平面を定める。

$$y(x) = w^T x + w_0 \quad (2.34)$$

ここに x は D 次元のデータベクトルであり、 w は D 次元のパラメータベクトルである。 w_0 もスカラーパラメータであり閾値の役割を担う。分類面の役割により $y(x)$ は x がクラス C_1 に属する場合には正の値を、 C_2 に属する場合には負の値を取るように学習される。ここで、(2.34) の超平面と、あるデータ点 x_n との距離は以下で表される。

$$|r| = \frac{|y(x_n)|}{|w|} \quad (2.35)$$

ここで、 x_n がクラス C_1 に属する場合は $t_n = 1$ とし、クラス C_2 に属する場合には $t_n = -1$ と定めた t_n を導入する。さらに x_n には分類面から最も近いデータ点のみを考慮することとし、そのときの $|r|$ をマージンと呼び以下で表す。

$$|r|_{\text{margin}} = \min_{x_n} \frac{t_n y(x_n)}{|w|} \quad (2.36)$$

この (2.36) を最大化するようにパラメータを決定することでマージン最大化を実現することができる。従って、SVM のパラメータ決定は以下の最適化問題によって定式化される。

$$\arg \max_{w_0, w} \left(\min_{x_n} \frac{t_n y(x_n)}{|w|} \right) \quad (2.37)$$

しかしこの最適化問題において、 w, w_0 の大きさは本質的ではない。なぜなら w, w_0 を同時に kw, kw_0 と定数倍した場合にも (2.36) の値は変化しないためである。従って、 w, w_0 の大きさに関して制約を設ける必要がある。そこで分類面から最も近い x_n に関して $t_n(w^T x + w_0) = 1$ となるような制約を w, w_0 に対して要請する。この条件式に伴って、任意のデータ点において $t_n(w^T x + w_0) \geq 1$ という制約が与えられる。最終的にマージン最大化問題 (2.37) は以下で定式化される。

$$\begin{aligned} \arg \min_{w_0, w} \quad & \frac{|w|^2}{2} \\ \text{s.t.} \quad & t_n(w^T x + w_0) \geq 1 \end{aligned} \quad (2.38)$$

目的関数の分母は、単に勾配を計算する際に約分できるというテクニックによるものである。分子の二乗は、最適化問題の解を変更せずに勾配計算などを容易に行うための変形である。図 2.8 に分類面の定め方によりマージンが異なっている様子を見ることができる。左右いずれの図も学習データに対して正しく分類が行える分類面になっているが、新規のデータに対しての分類結果が異なってくる。SVM では右図の分類面の方が優れていると考える場合に用いる手法である。ここまで線形分類を行う場合の SVM を見てきたが、一般的に線形分類器はデータ点を非線形関数 $\phi(\cdot)$ によって別の特徴空間へ写像し、特徴空間上で線形分類を行う問題へ拡張することができる。SVM では分類超平面を以下の式によって構築することに相当する。

$$y(x) = w^T \phi(x) + w_0 \quad (2.39)$$

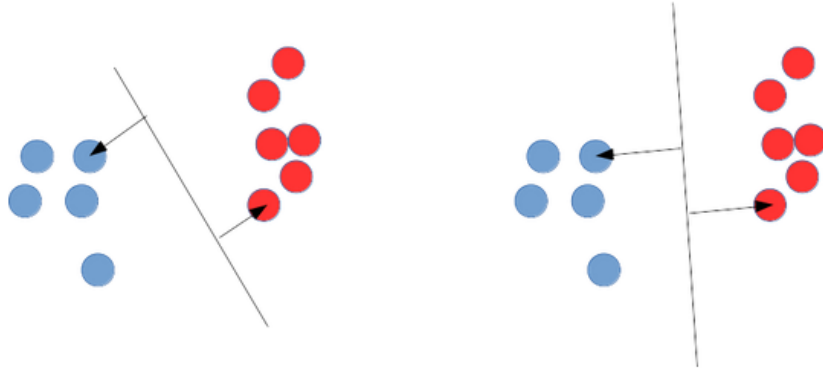


図 2.8: 分類面によってマージンの大きさが異なる様子

この場合においてもこれまでと同様の議論で最適化問題を以下のように定式化できる。

$$\begin{aligned} \arg \min_{w_0, w} \quad & \frac{|w|^2}{2} \\ \text{s.t.} \quad & t_n(w^T \phi(x) + w_0) \geq 1 \end{aligned} \quad (2.40)$$

実データは線形分離不可能な場合が多いため、通常 SVM を用いる場合は上記のような非線形に拡張されたものを用いる。また実データは異なるクラスのデータが重なって分布することも多々あるため、厳密な分類を行うことは不可能な場合が多い。そういった場合に対応したソフトマージンと呼ばれる考えがあり、学習データの誤分類に対して寛容になる指標を導入する。このソフトマージンの考え方は機械学習で過学習抑制に用いられる正則化の考えと本質的には変わらない。また、制約付き最適化問題をラグランジュ法によって変形することで双対問題を獲得することができる。双対問題においては非線形変換後の空間での内積のみが必要となり、具体的な非線形変換の計算をデータ x に対して実施する必要はない。すなわち、非線形変換 $\phi(x)$ による特徴空間への写像を具体的に考える代わりに、最終的に計算の必要性がある内積 $\phi(x)^T \phi(x')$ を定義することで SVM の非線形への拡張が可能である。このような方法はカーネル法として知られており、このときに用いられる内積計算をカーネル $k(x, x') = \phi(x)^T \phi(x')$ と呼ぶ。カーネルを定めることが特徴空間の設計を行うことに相当するが、BCI も含めた多くの応用では既に知られた優れた性質を持つカーネルを活用することがほとんどである。以下に代表的なカーネルについて記載する。

- ・線形カーネル：非線形変換を行わないことに対応。

$$k(x, x') = x^T x'$$

- ・多項式カーネル：多項式関数による非線形変換に対応。

$$k(x, x') = (x^T x' + c)^M$$

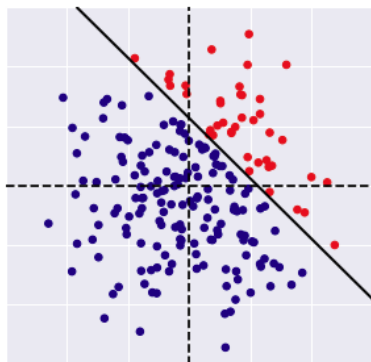


図 2.9: 線形 SVM (線形分離可能)

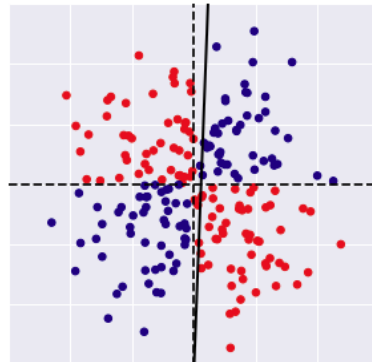


図 2.10: 線形 SVM (線形分離不可能)

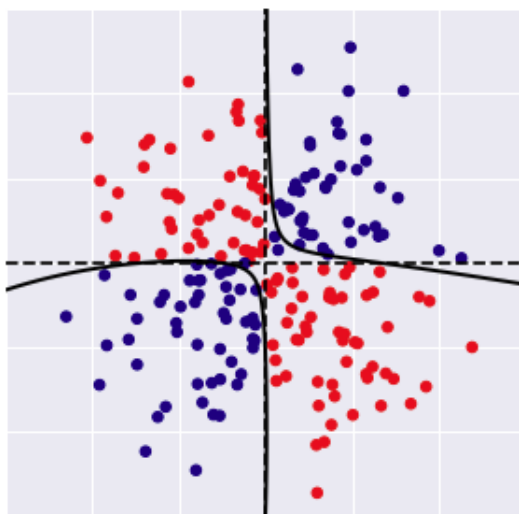


図 2.11: ガウシアンカーネル SVM (線形分離不可能)

- ・ ガウス基底カーネル：特徴空間が無限次元となる。

$$k(x, x') = \exp\left(-\frac{|x-x'|^2}{2\sigma^2}\right)$$

通常、カーネルを用いる場合にはハイパーパラメータが付随する。多項式カーネルの場合はスカラーの c, M 、ガウス基底カーネルの場合はスカラーの σ がハイパーパラメータとなり、これらの調整次第で得られる分類面は異なる。応用上はソフトマージンカーネル SVM を用いればよく、ソフトマージンのハイパーパラメータとカーネルの設計を変えることで通常の線形 SVM の働きをさせることも可能である。図 2.9-2.11 に 2 次元データに対する SVM の分類境界を示す。図 2.10 の通り、線形 SVM では線形分離不可能な問題に対して不適切な境界を設ける。

ソフトマージンカーネル SVM の識別関数は以下で表される。

$$y(x) = \sum_{n=1}^N a_n t_n k(x, x_n) + b \quad (2.41)$$

ここで学習データの数 N としている。また a_n はスカラーのパラメータであり、ラグランジュ法で双対問

題を考えた際のラグランジュ乗数である。また最適化問題は以下で定式化される。

$$\begin{aligned}
 & \arg \max_{a_1, \dots, a_N} \quad \sum_{n=1}^N a_n - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_i a_j t_i t_j k(x_i, x_j) \\
 & \text{s.t.} \quad 0 \leq a_n \leq C \\
 & \quad \sum_{n=1}^N a_n t_n = 0
 \end{aligned} \tag{2.42}$$

ここに C はソフトマージンのハイパーパラメータでありスカラーである。値が小さいほど a_n の取れる範囲が制限され、値が大きいくほど誤分類に寛容となる。最適化問題の解である a_1, \dots, a_N と任意の n を選んで以下の条件式に代入することで閾値 b も求まる。

$$t_n \left(\sum_{m=1}^N a_m t_m k(x_n, x_m) + b \right) = 1 \tag{2.43}$$

ただし、(2.43) は無駄な計算も含まれている。 $a_n = 0$ となるような x_n に対して $\sum_{m=1}^N a_m t_m k(x, x_m)$ の計算を実行する必要はない。 $a_n \neq 0$ となっている x_n のことをサポートベクトルと呼び、実際にはサポートベクトルのみ計算に考慮すれば良い。このことは新規のデータに対して (2.41) の計算を行うときも同様である。従って学習後に保持しておかねければならないデータはサポートベクトルのみに限定でき、実用上省メモリに貢献できる。

2.3.3 Logistic Regression(LR)

Logistic Regression(LR) は対数オッズ比を線形モデルで表現した分類手法である。クラス分類においてデータ x がクラス C_i に属する確率を $p(y = C_i | x)$ と表す。この時、事象 $y = C_i$ のオッズ比は以下で表される。

$$\text{odds}(y = C_i) = \frac{p(y = C_i | x)}{1 - p(y = C_i | x)} \tag{2.44}$$

2 クラス分類においては $\text{odds}(y = C_1) \geq 1$ であれば C_1 に分類するなどの規則を設けることで分類が可能になる。ここで、対数オッズは以下で表される。

$$\log \text{odds}(y = C_i) = \log \left\{ \frac{p(y = C_i | x)}{1 - p(y = C_i | x)} \right\} \tag{2.45}$$

(2.44) が 1 以上の値となる時、(2.45) は 0 以上の値となる。LR では多次元データ $x \in \mathbb{R}^D$ に対して $w \in \mathbb{R}^D$ として、線形モデル $z = w^T x \in \mathbb{R}$ によって対数オッズ比を出力するモデルを構築する。 $y = C_1$ の対数オッズ比を線形モデルで表現すると以下で表される。

$$\log \left\{ \frac{p(y = C_1 | x)}{1 - p(y = C_1 | x)} \right\} = w^T x \tag{2.46}$$

この時、以下の式変形によってよく知られたシグモイド関数 $\sigma(z) = 1/(1 + \exp(-z))$ が導出される。

$$\begin{aligned}
& \log \left\{ \frac{p(y = C_1 | x)}{1 - p(y = C_1 | x)} \right\} = w^T x \\
& \iff \frac{p(y = C_1 | x)}{1 - p(y = C_1 | x)} = \exp(w^T x) \\
& \iff \frac{1}{p(y = C_1 | x)} - 1 = \exp(-w^T x) \\
& \iff p(y = C_1 | x) = \frac{1}{1 + \exp(-w^T x)}
\end{aligned} \tag{2.47}$$

シグモイド関数に関しては、

$$\sigma(-z) = 1 - \sigma(z) \tag{2.48}$$

$$\sigma'(z) = \sigma(z)(1 - \sigma(z)) \tag{2.49}$$

という性質が知られている。ここで $y = C_1$ という事象を $y = 1$ に対応させ、 $y = C_2$ という事象を $y = -1$ に対応させると、(2.47) と (2.48) を用いて、

$$p(y | x, w) = \sigma(yw^T x) \tag{2.50}$$

とすることができる。i.i.d. を仮定して対数尤度の最尤推定を考えると以下のロジスティック損失の最小化問題に帰着される。ここで、 N は訓練データ数とした。

$$\begin{aligned}
\arg \max_w \sum_{i=1}^N \log \{ \sigma(y_n w^T x_n) \} &= \arg \min_w \sum_{i=1}^N -\log \{ \sigma(y_n w^T x_n) \} \\
&= \arg \min_w \sum_{n=1}^N \log \left\{ \frac{1}{\sigma(y_n w^T x_n)} \right\} \\
&= \arg \min_w \sum_{n=1}^N \log \{ 1 + \exp(-y_n w^T x_n) \}
\end{aligned} \tag{2.51}$$

以上から、LR の学習はオッズ比を対数線形モデルで表現し最尤推定を行うことに相当する。導出の過程から明らかなようにシグモイド関数は非線形変換であるが、スケールの変換を行うだけで線形分離不可能な問題に対応できるわけではなく、モデルとしては単層パーセプトロンと等価である。目的関数は凸であり、学習は勾配法、ニュートン法、準ニュートン法などが用いられる。

第 3 章

従来手法 BCI

3.1 特徴量工学

運動想起 BCI の標準的な役割は EEG 信号から運動想起部位を識別することであるが、そのためには EEG の生データに対して何らかの処理を施し、識別に有用な特徴を見出さねばならない。通常、EEG の生データを観測しただけでは運動想起部位を識別することは困難である。例として図 3.1 と図 3.2 を添付する。それぞれ足の運動想起と手の運動想起を行った際の EEG である。いずれも 0 秒以前はディスプレイを注視した状態であり、0 秒以降に運動想起を 4 秒間行っている。電極の個数は 64 個用いられており、全ての電極の波形が表示されている。足の運動想起が行われているのか手の運動想起が行われているのかを EEG の生データから識別するのは困難であることが分かる。そこで、何らかの変換 $f(\cdot)$ を EEG 信号 $x(t)$ に対して施すことで、特徴量 $z = f(x(t))$ の獲得を目指すのが特徴量工学である。この節ではまず運動想起に関する EEG の研究により得られた知見から特徴量を得る方法をいくつか紹介する。

3.1.1 頭皮領域と空間フィルタ

一般的に脳活動の計測では計測位置が極めて重要な要素であることは自明である。EEG のみならず非侵襲式測定を行った際には、センサ領域にどのような関連があるかを明らかにし、異なる領域へと変換することで新たな特徴量を得ることに関心がある場合も多く、この種の問題は信号源推定問題として知られている。

運動想起型 BCI についても初めに必要な電極についての選定やセンサ領域に関して考慮する必要がある。例えば足の動作に関する脳活動が EEG として観測される場合、頭皮上の全ての領域が対等に重要であるとは考えにくい。足の動作に関する脳の領域は頭頂部に存在するため、EEG として観測される場合、頭頂部に位置する Cz 電極が最も関係していると考えられる。図 3.4 は足の運動想起を行った際の Cz 電極によって記録された EEG である。しかし波形から足の運動想起を行っていることを判別する決定的な差異を見出すことは困難である。理由としては、脳の頭頂部での神経活動によって生じている電位が、必ずしも頭皮上の頭頂部に

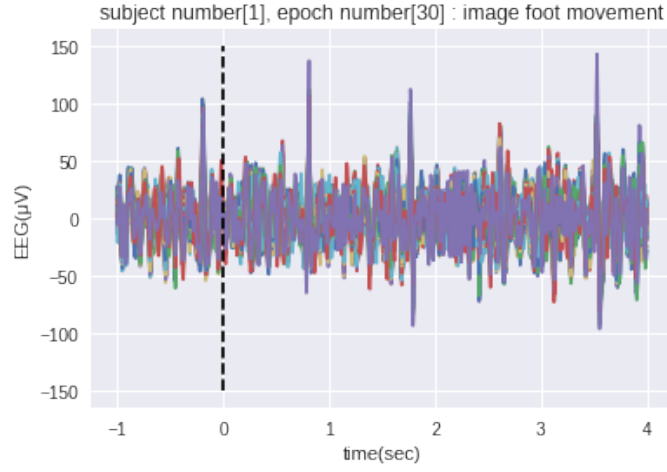


図 3.1: 0 秒 ~ 4 秒間に足の運動想起を行った際の EEG

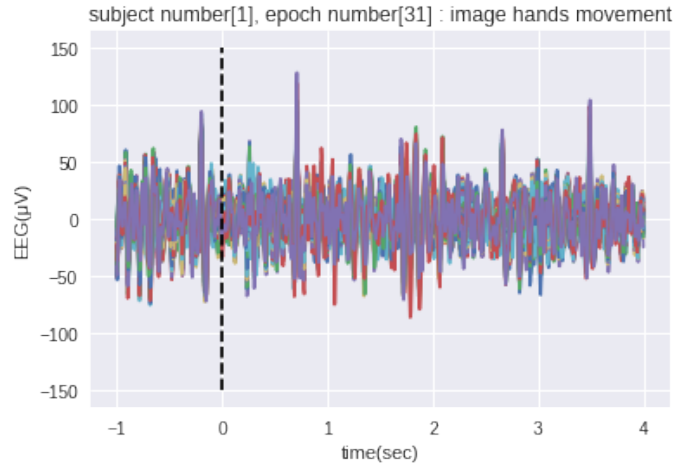


図 3.2: 0 秒 ~ 4 秒間に手の運動想起を行った際の EEG

のみ伝搬するとは限らないからである。電位を発生させている神経活動と電位を計測している電極との間には、頭蓋骨と皮膚があり、電位の伝搬がどのように行われているかを完全に把握することは難しい。ある程度空間的に広がりを持って電位が伝搬していると考えるのが自然である。そこで計測された EEG に対して、スモールラプラシアンフィルタと呼ばれる空間フィルタが用いられることがある。ラプラシアンフィルタは一般的に 2 次元配列 $X(i, j)$ に対して以下の計算によって二次差分を計算したものである。

$$X(i, j) = -4X(i, j) + X(i-1, j) + X(i+1, j) + X(i, j-1) + X(i, j+1) \quad (3.1)$$

二次差分フィルタは画像処理では輪郭検出のために用いられる処理であり、空間的に配置された値が際立って変化する位置で大きな値を持つ配列を返す。例として、A 電極によって計測された EEG を $x_A(t) \in \mathbb{R}$ と表記すると、Cz 電極に対するスモールラプラシアンフィルタが適用された $EEG_{x_{smallCz}}(t)$ は以下で表される。

$$x_{smallCz}(t) = x_{Cz}(t) - \frac{1}{4}(x_{C1}(t) + x_{C2}(t) + x_{FCz}(t) + x_{CPz}(t)) \quad (3.2)$$

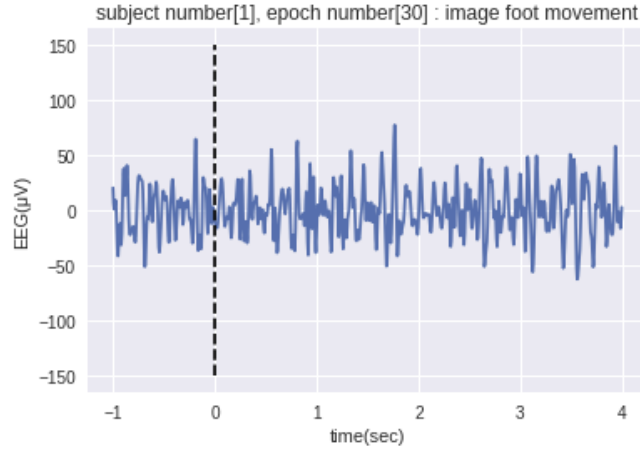


図 3.3: 0 秒 ~ 4 秒間に足の運動想起を行った際の Cz 電極の波形

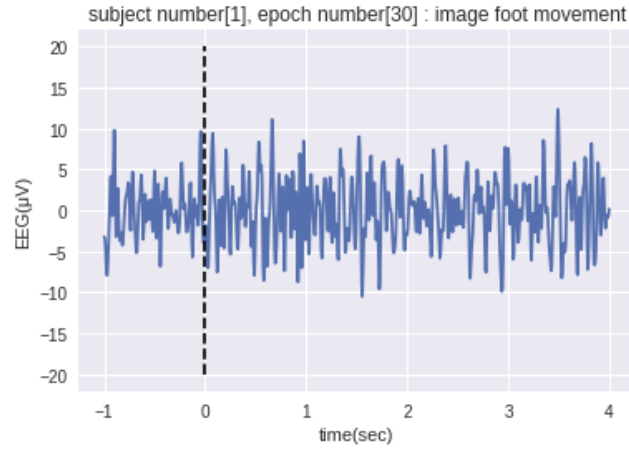


図 3.4: スモールラプラシアンフィルタを適用した際の Cz 電極の波形

信号に対する符号や絶対値については、多くの場合は着目した電極の係数の符号を正とした計算が行われる。図 3.4 は Cz 電極に対してスモールラプラシアンフィルタを用いた際の EEG である。スモールラプラシアンフィルタの処理によって必ずしも有効な特徴量が得られるとは限らないため、適切な空間フィルタを獲得するために統計的なフィルタ設計方法が用いる場合もある。ここに画像処理で空間フィルタという表現を使う場合、2 次元配列に於ける計算を考へるのが一般的であるが、EEG に対して空間フィルタという表現を用いる場合は、電極に対する重み付けを行う処理全般を指す。従って、電極を D 個用いて $\text{EEG}x(t) \in \mathbb{R}^D$ を計測した場合に、 $w \in \mathbb{R}^D$ によって

$$z(t) = w^T x(t) \in \mathbb{R} \quad (3.3)$$

を計算した場合は w を空間フィルタと呼ぶ。より一般的には行列 $W = (w_1, w_2, \dots, w_d) \in \mathbb{R}^{D \times d}$ によって、

$$z(t) = W^T x(t) \in \mathbb{R}^d \quad (3.4)$$

を計算する場合にも空間フィルタを用いたと言える。この場合は、 W の各列が空間フィルタ 1 つ 1 つ

表 3.1: EEG の周波数による分類

名称	読み	周波数帯域
波	デルタ波	0.5 ~ 4Hz
波	シータ波	4 ~ 8Hz
波	アルファ波	8 ~ 13Hz
波	ベータ波	13 ~ Hz

に対応し、 d 個の空間フィルタについて一挙に計算を行うことが可能である。特に Independent Component Analysis(ICA) による空間フィルタの設計は既に EEG 解析で一定の地位を築いている。また EEG 解析ソフトウェアの機能の 1 つとして実装されているケースが多く、書籍 [35] では、1 章を割いて ICA の応用方法について解説がなされている。

一方で東らの [36] では、グラフフーリエ領域と呼ばれる信号構造で定義される領域へ信号を変換する方法を提案している。この手法は、電極の配置における空間的距離と観測信号の類似性に基づいて、電極間の繋がりの強さを示すグラフ構造を獲得する。この手法を用いて空間的な平滑化を行った後に、従来の信号源推定手法である PCA を用いると、の信号源に近い信号の抽出が可能であると主張されている。

また、Common Spatial Pattern(CSP) は運動想起 BCI に対する非常に有効な空間フィルタ設計手法として考えられており、数多くの派生手法が存在するため後述する。

3.1.2 EEG と時間周波数解析

EEG には 波、 波、 波、 波などが存在するが、これらは周波数帯域で分類がなされており (表 3.1)、それぞれの波の位相や振幅が精神状態に応じて時間変化する。特にヒトが手や指、足等の運動およびそのイメージを行う際に生ずる、 律動や 律動での ERD および ERS は特徴量になると考えられ、ERD は運動およびそのイメージ中に、ERS は運動およびそのイメージ後に発生する [3]。従って EEG に対して時間周波数解析を行い、スペクトログラムを獲得することで ERD や ERS を直接的に検知し、運動想起 BCI を構築することが可能である [7]。ERD や ERS を検知するための時間周波数解析には、最も単純なものとして短時間フーリエ変換を用いることができる。しかし、フーリエ変換対に対する不確定性関係やスペクトル密度推定の精度の問題もありウェーブレット変換を用いる方法やバーク法に基づく方法も検討されている。また ERD 検知に有効な時間周波数解析の比較についても研究報告がある [9]。更に、非定常時系列に対する時間周波数解析として提案された Empirical Mode Decomposition(EMD) を用いる方法もいくつか検討がなされている [37, 38]。

3.2 Common Spatial Pattern とその派生手法

運動想起型 BCI において非常に活発に用いられている “Common Spatial Pattern(CSP)” は 1990 年に Zoltan [14] により提案されてから、数々の派生手法が生み出されてきた [15–20, 39]。このセクションでは CSP の基本的な概要と、BCI に適した形で発展してきた CSP の派生手法について述べる。

3.2.1 脳波信号の定式化

これまで多次元の信号を $x(t) \in \mathbb{R}^D$ と表記し、連続時間信号として扱ってきた。しかし、通常は計測された脳波はコンピュータで処理するために離散時間信号 $x_n \in \mathbb{R}^D$ に変換される。従って、以降、脳波信号を離散時間信号として取り扱う。また、運動想起時の脳波信号を計測する際には、被験者に対して定められたタイムスケジュールで運動想起を行うように指示がなされる。例として 64 個の電極を用い、サンプリング周波数 100Hz で 10 秒間の計測を行った場合、計測された運動想起 1 回分の脳波信号は $X \in \mathbb{R}^{64 \times 1000}$ と表すことができる。従って、以降統一のため、 M を電極の個数、 N を計測時間点数とした場合の脳波信号を以下で定義する

$$X = (x_1, \dots, x_N) \in \mathbb{R}^{M \times N} \quad (3.5)$$

運動想起を K 回行った場合には、 K 個の X が得られる。通常は運動想起時の脳波を数個から数十個集め、統計的な指標を元に有用な特徴量を抽出する。CSP は運動想起 BCI に対して極めて有効に働くとされている特徴量抽出手法である。

3.2.2 Common Spatial Pattern(CSP)

CSP は複数の電極で計測されることが一般的である脳波に対して、適切な電極の重み付けを行うことで、重要な頭皮領域の脳波を取り出す手法である。CSP を脳波に用いる際は、脳波信号 X を直接扱うのではなく、何らかの前処理を施した信号 $\hat{X} = \mathcal{H}(X)$ を用いる。通常、 \mathcal{H} には、運動想起に関連のある周波数帯域のみを通過させるバンドパスフィルタを用いる (図 3.5)。バンドパスフィルタ通過後の脳波信号を以下のように表記する。

$$\hat{X} = (\hat{x}_1, \dots, \hat{x}_N) \in \mathbb{R}^{M \times N} \quad (3.6)$$

$\hat{x}_i \in \mathbb{R}^M$ における添字 i はサンプル時刻の添え字である。CSP では、新たな基底 $w \in \mathbb{R}^M$ にバンドパスフィルタ通過後の脳波信号を射影し、スカラー時間信号である $z = w^T \hat{X} \in \mathbb{R}^N$ を抽出する。この時の w の決め方を以下に記す。

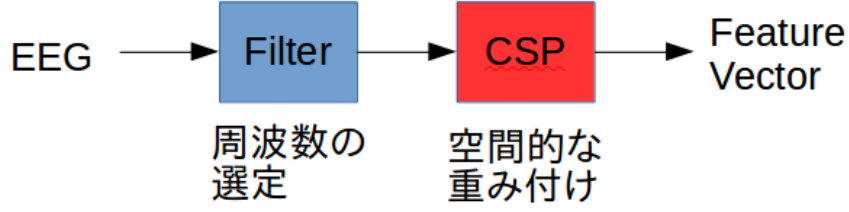


図 3.5: CSP の利用方法

まず \hat{X} を基底 w に射影した際の時間分散 $\sigma^2(\hat{X}, w)$ を以下で定義する。

$$\sigma^2(\hat{X}, w) = \frac{1}{N} \sum_{i=1}^N \left| w^T \left(\hat{x}_i - \frac{1}{N} \sum_{j=1}^N \hat{x}_j \right) \right|^2 \quad (3.7)$$

ここで計測された複数個の \hat{X} は必ず集合 C_1 か C_2 のいずれか一方に属するとし、 $C_1 \cap C_2 = \phi$ であるとする。CSP では、ベクトル $w \in \mathbb{R}^M$ を新たな基底とした電極空間において、一方のクラスに属する信号 $\hat{X} \in C_d$ についての時間分散 (3.7) が最大となるように、 w を決める。これは以下の最大化問題によって定式化される。

$$\begin{aligned} \max_w \quad & \mathbb{E}_{X \in C_1} [\sigma^2(\hat{X}, w)] \\ \text{s.t.} \quad & \sum_{d=1,2} \mathbb{E}_{X \in C_d} [\sigma^2(\hat{X}, w)] = 1 \end{aligned} \quad (3.8)$$

最大化問題 (3.8) を解いて得られる w_{csp} は、クラス C_1 に属する脳波の分散を最大化するような基底である。一方で、制約条件によって 2 つのクラスの分散の和が 1 であるとされているため、自動的にクラス C_2 に属する脳波の分散を最小化する基底ともなる。すなわち w_{csp} によって得られるスカラー信号は一方のクラスの信号のみを増幅させ、他方を減衰させる働きをする。(3.8) は更に以下で定式化することができる。

$$\begin{aligned} \max_w \quad & w^T \Sigma_1 w \\ \text{s.t.} \quad & w^T (\Sigma_1 + \Sigma_2) w = 1 \end{aligned} \quad (3.9)$$

ここで

$$\Sigma_i = \mathbb{E}_{X \in C_i} \left[\frac{\hat{X} \hat{X}^T}{\text{tr}(\hat{X} \hat{X}^T)} \right] \quad (3.10)$$

である。この問題の解はラグランジュ法によって解析的に求めることが可能である。簡単な計算によって (3.9) は以下の一般化固有値問題に帰着される [39]。

$$\Sigma_1 w = \lambda (\Sigma_1 + \Sigma_2) w \quad (3.11)$$

この一般化固有値問題は右辺の $\Sigma_1 + \Sigma_2$ が正則であれば、その逆行列を両辺左から掛けることで普通の固有値問題に変形できる。

一般化固有値問題が解けた時の M 個の固有値を $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$ とする。また、 λ_i に属する固有ベクトルを $w^{(i)}$ とする。このとき最適化問題の解に相当するのは $w^{(1)}$ であり、 C_1 に属する脳波を増幅し、 C_2 に属する脳波を減衰させたスカラー信号を獲得できる。一方で $w^{(M)}$ は C_1 に属する脳波を減衰し、 C_2 に属する脳波を増幅させたスカラー信号を獲得できる。従って通常は、 $w^{(1)}, w^{(M)}$ の固有ベクトルを対で用いる。同様に $w^{(2)}, w^{(M-1)}$ を対で取り出して 4 つの固有ベクトルを用いることもできる。一般に $2m (< M)$ 個の固有ベクトルを使うこととすれば、特徴量 $\sigma^2(\hat{X}, w^{(i)})$ を $2m$ 個準備することができる。ここで単に $z = w^{(i)} \hat{X}$ を用いないのは、CSP によって得られる特徴量は振幅に大きな違いがあるためである。通常は CSP によって獲得される特徴量 y は以下の形式となる。

$$y = (\sigma^2(\hat{X}, w^{(1)}), \dots, \sigma^2(\hat{X}, w^{(m)}), \dots, \sigma^2(\hat{X}, w^{(M-m+1)}), \dots, \sigma^2(\hat{X}, w^{(M)}))^T \quad (3.12)$$

あるいは $y = (y_1, \dots, y_{2m})$ に対して以下のような変換を行った特徴量 Y を使う (図 3.6)。

$$\begin{aligned} Y_i &= \log \left(\frac{y_i}{\sum_{i=1}^{2m} y_i} \right) \\ Y &= (f_1, \dots, f_{2m})^T \end{aligned} \quad (3.13)$$

また $m = 1, 2$ が多くの CSP の応用研究で使われており、 m を増やしても BCI の性能向上には繋がらないことが報告されている [19]。

3.2.3 Common Sparse Spatio Spectral Pattern(CSSSP)

CSP ではバンドパスフィルタ \mathcal{H} 通過後の信号を扱ったため、バンドパスフィルタの設計を終えてからでなければ CSP の問題を解くことができない。従って、あるバンドパスフィルタを通過した脳波に対して、最適化問題を解いているに過ぎなく、バンドパスフィルタ自体が適切であったかに関しては一切述べられない。このことによって生じる問題点に関しては、更に詳しく後述する。CSSSP の主なモチベーションはバンドパスフィルタの設計を CSP の問題の中に取り込むことである [16]。

まず脳波信号 $X = [x_1, \dots, x_N] \in \mathbb{R}^{M \times N}$ に対して、観測時間点を $\tau > 0$ だけずらした $X_\tau = (x_{1+\tau}, \dots, x_{N+\tau}) \in \mathbb{R}^{M \times N}$ を考える。ここで $\tau = 1, \dots, T$ である。更にクラス C_d に属する X に関して各 τ 毎に、

$$\Sigma_d^\tau = \mathbb{E}_{X \in C_d} [X X_\tau^T + X_\tau X^T] \quad (3.14)$$

を定義する。ただし $\tau = 0$ のとき $\Sigma_d^0 = \mathbb{E}_{X \in C_d} [X X^T]$ とする。更に、 T 個の要素を持つ係数ベクトル

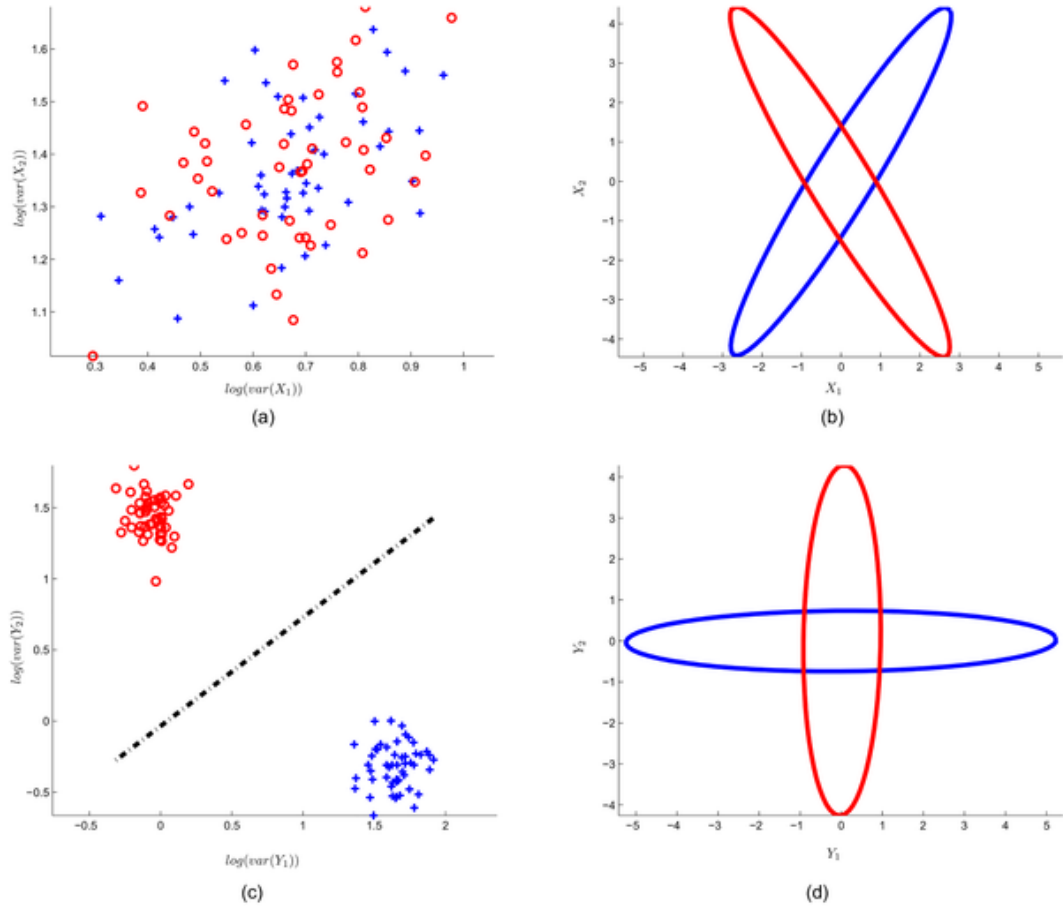


図 3.6: CSP の動作 (a) 観測点、(b) 観測時の座標、(c) 特徴量、(d) 変換された座標

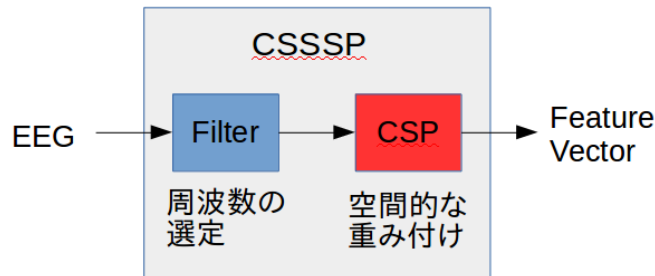


図 3.7: CSSSP の働き

$b = (b_1, b_2, \dots, b_T)^T$ とする。ここで CSSSP の最適化問題は以下で表現される。

$$\begin{aligned}
 & \max_{b_2, \dots, b_T} \max_w w^T \left\{ \sum_{\tau=0}^{T-1} \left(\sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right) \Sigma_1^T \right\} w - \frac{\alpha}{T} |b|_1 \\
 & \text{subject to} \quad w^T \left[\sum_{\tau=0}^{T-1} \left\{ \sum_{j=1}^{T-\tau} b_j b_{j+\tau} \right\} (\Sigma_1^T + \Sigma_2^T) \right] w = 1
 \end{aligned} \tag{3.15}$$

この最適化問題の目的関数第 1 項は b を FIR フィルタ f の係数ベクトルとした場合に、 f を通過後の脳波 X に対して CSP の最適化問題を適用していることに相当する。従って CSSSP では空間重みベクトル $w \in \mathbb{R}^M$ と次数が T の FIR フィルタのフィルタ係数 b を同時に評価することが可能である (図??)。目的関数における

第二項は正則化項であり、 α はハイパーパラメータである。この正則化によって FIR フィルタの系列ベクトルに対してスパース性が要請され、フィルタリングに用いられる時系列点数は必要最小限に抑えられる。このことにより FIR フィルタが必要以上に複雑になることが避けられる。解析的に解くことは困難であるため、勾配法あるいはニュートン法を用いた直線探索によって数値的に解を得る。

3.2.4 Filter Bank Common Spatial Pattern(FBCSP)

CSSSP は FIR フィルタの設計を CSP の最適化問題に含むことで、バンドパスフィルタと CSP の同時最適化の定式化に成功した。しかし、CSSSP によって求まる特徴量はフィルタ係数 b を持つ FIR フィルタを通過し、変換行列 W^T の CSP によって変換された脳波信号であり、ある 1 つの周波数領域のみしか考慮されていないこととなる。CSSSP を解いて得られるフィルタ係数 b は最適化問題の解であるが、必ずしもそのフィルタ係数のみを用いることが最良であるとは言い難い。なぜなら、運動想起を行う際にはその身体部位に応じて特徴的な変動を持つ周波数領域は異なっているためである。例えば左手の運動想起時と下肢の運動想起時には重要な周波数帯域が異なり、CSSSP の解がいずれかの情報を失っている可能性が考えられる。FBCSP はこのような問題の解決を図る手法である [20]。

まず、FBCSP ではバンドパスフィルタバンク $\{\mathcal{H}_1, \dots, \mathcal{H}_K\}$ を定義する。このバンドパスフィルタバンクは 1 つ 1 つのフィルタ \mathcal{H}_k が異なる周波数領域を通過させるように設計される。各バンドパスフィルタを通過した $\hat{X}_k = \mathcal{H}_k(X)$ に対して、それぞれ CSP の問題を解くことで各周波数領域における空間的な特徴量を獲得することが可能となる。 \hat{X}_k に対して CSP によって取り出した特徴量を Y_k とすれば、最終的な特徴ベクトル Y は

$$Y = (Y_1^T, \dots, Y_K^T)^T \quad (3.16)$$

であり、単に各周波数帯域の CSP による特徴量を連結したベクトルとなる。ただし CSSSP とは異なり、CSP とバンドパスフィルタの同時最適化を行っていない。従って、特徴ベクトル f のいずれの要素が重要であるか、すなわちいずれの周波数帯域が重要であるかについての最適化は行われておらず、あくまで複数の周波数帯域においてそれぞれの最適電極の重み付けを CSP によって獲得したにすぎない。そのため、FBCSP によって得られた特徴量から更に特徴量を抽出する手法が別途用いられる図 3.7。また、FBCSP 後の特徴量の選定を行う方法として相互情報量を用いた手法が提案されている [40]。

第 4 章

提案手法

4.1 ニューラルネットワーク

初めに提案手法の大部分を占めるニューラルネットワークについて説明する。ニューラルネットワークは近年、“深層学習”あるいは“ディープラーニング”の名の元に、再注目を浴びている学習モデルである。ニューラルネットワークは元々脳の神経活動を模倣したモデルとして提案されたが、現在では学習の力学特性や統計物理学の知見など数理的な解析が行われている [28, 29]。あるいはヒューリスティクスによる発展が著しい。従って、本論文では単にある特定の形式で表される数理モデルとしてニューラルネットワークを扱う。

4.1.1 ニューラルネットワークに用いられる層

Linear 層

ニューラルネットワークで最も基本的な層は Linear 層である。Linear 層は入力としてベクトル $x \in \mathbb{R}^D$ 、パラメータとして $W \in \mathbb{R}^{d \times D}$ と $b \in \mathbb{R}^d$ を有した関数 $L(x, W) = Wx + b$ である。具体的には入力を $x = (x_1, \dots, x_D)^T$ に対して出力 $y = (y_1, \dots, y_d)^T$ を以下で表すことができる。

$$y_j = \sum_{i=1}^D w_{j,i} x_i + b_j \quad (4.1)$$

ここに $w_{j,i}$ は行列 W の (j, i) 成分、 b_j はベクトル b の j 番目の成分である。

Bilinear 層

Bilinear 層は入力として $x^{(1)} \in \mathbb{R}^{D_1}$ と $x^{(2)} \in \mathbb{R}^{D_2}$ の 2 つを受け取り、1 つの出力 $y \in \mathbb{R}^d$ を返す。

$$y_k = \sum_{j=1}^{D_1} \sum_{i=1}^{D_2} w_{j,i,k} x_j^{(1)} x_i^{(2)} + b_k \quad (4.2)$$

ここで n 階のテンソル $A \in \mathbb{R}^{D_1, \dots, D_n}$ の (d_1, \dots, d_n) 成分を a_{d_1, \dots, d_n} と表記している。パラメータは $W \in \mathbb{R}^{D_1 \times D_2 \times K}$ と $b \in \mathbb{R}^K$ である。

Convolution 層

Convolution 層はパラメータとして $H \in \mathbb{R}^{P \times Q \times C \times L}$ と $B \in \mathbb{R}^{I \times J \times L}$ を有し、入力を $X \in \mathbb{R}^{I \times J \times C}$ として、出力 $Y \in \mathbb{R}^{I' \times J' \times L}$ は以下で表される。

$$y_{i,j,l} = \sum_{c=1}^C \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} x_{i+p,j+q,c} h_{p,q,c,l} + b_{i,j,l} \quad (4.3)$$

ここでも n 階のテンソル $A \in \mathbb{R}^{D_1, \dots, D_n}$ の (d_1, \dots, d_n) 成分を a_{d_1, \dots, d_n} と表記している。出力テンソルの大きさを決める I', J' は I, J と P, Q に依存し、それぞれ $I' = P - I + 1$ と $J' = Q - J + 1$ となる。しかし、Convolution 層は実応用では stride や padding、dilation などのテクニックによって更に複雑な動作をする。これらのテクニックはそれぞれハイパーパラメータを持ち、 I', J' はその全てに依存する。

Recurrent 層

Recurrent 層は入力 $X = (x_1, \dots, x_T) \in \mathbb{R}^{D \times T}$ 、出力 $Y = (y_1, \dots, y_T) \in \mathbb{R}^{d \times T}$ として以下で表される。

$$y_t = \tanh(W_{in}x_t + W_{out}y_{t-1} + b) \quad (4.4)$$

ここにパラメータは $W_{in} \in \mathbb{R}^{d \times D}$ と $W_{out} \in \mathbb{R}^{d \times d}$ ならびに、 $b \in \mathbb{R}^d$ である。元々系列データへの応用のために考案されたため、 D 次元のベクトル $x_t \in \mathbb{R}^D$ が時間変化していくような場合を想定して (4.4) と定式化される。しかし、実際には入力は画像のような静的なデータでも構わなく、画像を左から右に走査するような働きを担う。 $\tanh(\cdot)$ は双曲線正接関数であるが、ニューラルネットワークの活性化関数として広く持ちいられている。

LSTM 層

LSTM は Long Short-Term Memory の略であり、LSTM 層は Recurrent 層が長期的な系列データを上手く学習できない問題を解決した。LSTM 層は入力 $X = (x_1, \dots, x_T) \in \mathbb{R}^{D \times T}$ 、出力 $Y = (y_1, \dots, y_T) \in \mathbb{R}^{d \times T}$ と

して以下で表される。

$$i_t = \sigma(W_{ii}x_t + W_{hi}h_{t-1} + b_i) \quad (4.5)$$

$$f_t = \sigma(W_{if}x_t + W_{hf}h_{t-1} + b_f) \quad (4.6)$$

$$o_t = \sigma(W_{io}x_t + W_{ho}h_{t-1} + b_o) \quad (4.7)$$

$$g_t = \tanh(W_{ig}x_t + W_{hg}h_{t-1} + b_g) \quad (4.8)$$

$$c_t = f_t \odot c_{t-1} + i_t \odot g_t \quad (4.9)$$

$$y_t = o_t \odot \tanh(c_t) \quad (4.10)$$

ここに \odot は同じ次元のベクトルの要素ごとの積である。パラメータは $W_{ii}, W_{if}, W_{io}, W_{ig} \in \mathbb{R}^{d \times D}$ と、 $W_{hi}, W_{hf}, W_{ho}, W_{hg} \in \mathbb{R}^{d \times d}$ および、 $b_i, b_f, b_o, b_g \in \mathbb{R}^d$ である。Recurrent 層同様に、系列データを想定した定式化がなされているが、実際には入力画像でも構わない。 $\sigma(\cdot)$ はシグモイド関数であり (2.47) と同じものである。ここでのシグモイド関数は、値域が $(0, 1)$ であるために用いられている。

i_t, f_t, o_t, g_t は現在の入力 x_t と過去の出力 y_{t-1} を引数に取り、それぞれ個別のパラメータを用いて値が算出される。 c_t は Memory Cell と呼ばれ、過去の情報を層の内部に記憶する役割を担っている。 $f_t \in (0, 1)^d$ を乗じることで、過去の情報 c_{t-1} を一定の割合のみ保持し（あるいは忘却し）、 i_t に対して $g_t \in (-1, 1)^d$ を乗じた値を新たに加算して保持する。 y_t は内部情報である c_t を $(-1, 1)^d$ にスケーリングしたベクトルと o_t との要素積によって算出される。すなわち、 o_t の値が内部情報 c_t に応じて大きさと正負が調整されて LSTM 層の出力となる。

4.1.2 活性化関数

ここではパラメータを持たない変換について述べる。

シグモイド関数

シグモイド関数 $\sigma(x)$ は以下で定義される。

$$\sigma(z) = \frac{1}{1 + \exp(-z)} \quad (4.11)$$

ニューラルネットワークは当初、神経活動の数値モデルとしての一面を有しており、入力値がある閾値を超えると発火するというモデルを表現するためにステップ関数が利用されていた。シグモイド関数は、ステップ関数を微分可能な形で表現することを動機に考案された。

双曲線正接関数

双曲線正接関数 $\tanh(z)$ は以下で定義される。

$$\tanh(z) = \frac{\exp(z) - \exp(-z)}{\exp(z) + \exp(-z)} \quad (4.12)$$

先に述べたシグモイド関数とは $\sigma(z) = (\tanh(x/2) + 1)/2$ という関係にあり、負の値を取ることもできるため状況に応じてシグモイド関数の代わりに用いられる。

ランブ関数

ランブ関数はニューラルネットワークでは Rectified Linear Unit(ReLU) 関数と呼ばれ、以下で表される。

$$\text{ReLU}(z) = \max(z, 0) \quad (4.13)$$

入力が正で -1 の勾配を持ち、その他の場合は勾配は 0 となる。この性質がニューラルネットワークの勾配消失問題を大きく改善したとされる。

leaklyReLU 関数

ReLU 関数が入力が負の場合にも値を有するように提案された。

$$\text{leaklyReLU}(z) = \max(z, 0) + \alpha \min(z, 0) \quad (4.14)$$

入力が正で -1 の勾配を持ち、その他の場合は勾配は α でハイパーパラメータである。

elu 関数

ReLU 関数を全域で微分可能にするように調整された関数。

$$\text{elu}(z) = \max(z, 0) + \min(\alpha(\exp(z) - 1), 0) \quad (4.15)$$

図 4.1 はここまで述べてきた 5 つの活性化関数の入出力を表した図である。どのような場面でどの活性化関数があるか、理論的な根拠は確認されておらず、試行錯誤的に決定される。

ソフトマックス関数

ソフトマックス関数は一般的に、分類問題におけるニューラルネットワークの出力層に用いられる。ベクトルの入力を受け取りベクトルの出力を行うが、以下に要素毎の計算を示す。

$$\text{softmax}_i(z) = \frac{\exp(z_i)}{\sum_j \exp(z_j)} \quad (4.16)$$

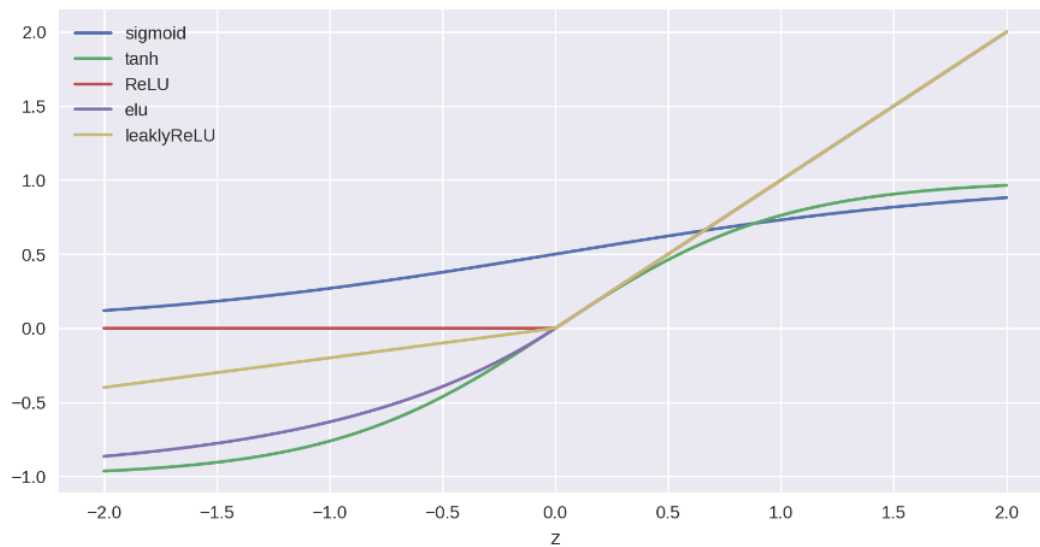


図 4.1: 5 つの活性化関数

ここで、 z_i はベクトル z の i 番目の要素である。定義から明らかに、値域は $(0, 1)$ に制限され、出力の要素の総和は 1 である。

4.1.3 ニューラルネットワークの設計と学習

ニューラルネットワークはパラメータを持つ関数とパラメータを持たない関数の合成関数を定義することにより設計される。パラメータを持つ層が入出力関係を柔軟に調整する役割を担い、活性化関数によって非線形性を与えることが可能となる。

設計されたニューラルネットワーク、すなわち合成関数を引数とする損失関数 (汎関数) を定義し、損失関数の最小化問題を解くことで学習を実行する。この際、損失関数はパラメータに関して微分可能なものが選択されなければならない。クラス分類においては多くの場合にソフトマックス交差エントロピーを用いる。

ニューラルネットワークは膨大なパラメータを有するため、学習させる入出力関係に対して冗長である場合がほとんどであり、過学習を抑制する手段を同時に検討しなければならない。機械学習で基本的な正則化に加え、ニューラルネットワークの過学習抑制方法としてはドロップアウトやバッチ正規化、ノイズの入力が用いられる。

4.2 提案手法のアイデア

Convolution 層による FIR フィルタバンク

EEG を $X \in \mathbb{R}^{M \times N \times 1}$ とする。ここに、 M は電極の数、 N はサンプル点数、1 は後に周波数を示すためのダミーインデックスである。 X に対する Convolution 層による演算は以下で表される。

$$u_{m,n,k} = \sum_{c=1}^C \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} x_{m+p,n+q,c} h_{p,q,c,k} + b_{m,n,k} \quad (4.17)$$

この演算に関して、 $P = 1, b_{m,n,k} = 0, C = 1$ でパラメータを与えることにより、

$$u_{m,n,k} = \sum_{q=0}^{Q-1} x_{m,n+q,1} h_{1,q,1,k} \quad (4.18)$$

という演算を行うことができる。ここでパラメータ $H \in \mathbb{R}^{1 \times D \times 1 \times K}$ と与えることで、フィルタ次数 $Q - 1$ の FIR フィルタ K 個からなるフィルタバンクとなる。

Convolution 層による空間フィルタ

続いて、第 2 層について $Q = 1$ でパラメータを与えることにより

$$v_{m,n,l} = \sum_{p=0}^{P-1} u_{m+p,n,k} g_{1,q,1,k} \quad (4.19)$$

という演算を行うことが可能であり、EEG に対して電極の重み付けを行っていることに相当する。ここで g がパラメータである。

以降ニューラルネットワークの演算を適宜定義することで、2 層目以降の層は、EEG を入力としたフィルタバンクからの出力を受け取る。FBCSP と異なる点は、FBCSP がフィルタバンクの設計とそれ以降の処理の設計が完全に分断されていることに対し、提案手法ではフィルタバンクのフィルタ係数と、その後の処理に用いられるパラメータを同時に最適化可能な点である。具体的なニューラルネットワークの構造は実験結果と共に記す。

第 5 章

実験結果と考察

5.1 提案手法の評価

5.1.1 EEG データ

実験データとして BCI2000 システム [41] を用いて記録された PhysioNet [42] が提供する運動想起データセットを用いた。電極の配置は標準的な 10-10 システム (図 5.1) であり、サンプリング周波数は 160Hz である。109 人の被験者が定められたタイムスケジュールに従い、左手、右手、両手および両足の運動想起を行っている。運動想起のタイムスケジュールを表す図を図 5.2 に示す。運動想起の開始時には被験者の前面に配置されたディスプレイから右手あるいは左手の運動想起の指示が出される。被験者は 4 秒間指示された手の指を開いたり閉じたりする運動想起試行を続け、その後 4.2 秒間の休息時間が与えられる。休息時間を終えると再びディスプレイから両手あるいは両足の運動想起の指示が出され、被験者は 4 秒間の支持された部位の指を開いたり閉じたりする運動想起試行を行う。その後 4.2 秒の休息時間が再び与えられる。計 16.4 秒を 1 サイクルとし、被験者 1 人につき 45 サイクルを繰り返し行う。従って被験者は左手、右手、両手、両足で計 90 回の運動想起を行っている。

PhysioNet が提供する運動想起時の EEG データベースは、世界最大規模の被験者数を有している。また、EEG の計測時に用いられた BCI2000 システムは運動想起型 BCI のみならず、多種多様な BCI システムを構築するために公開されたプログラム群であり、多くの脳研究者に利用されている。従って、手法の評価と検討を行うためのデータとして信頼性が高いと考えたため利用した。

5.1.2 評価方法

本研究では各運動想起時間である 4 秒間のデータを取り出して、2 クラス分類の問題を ${}_4C_2 = 6$ 種類準備した。ここで運動想起は一人あたり計 90 回行っており 4 部位あるため、1 つの部位につき 22 回あるいは 23 回の試行が行われている。従って準備した 2 クラス問題におけるデータ数は 44~46 となっている。また、以降

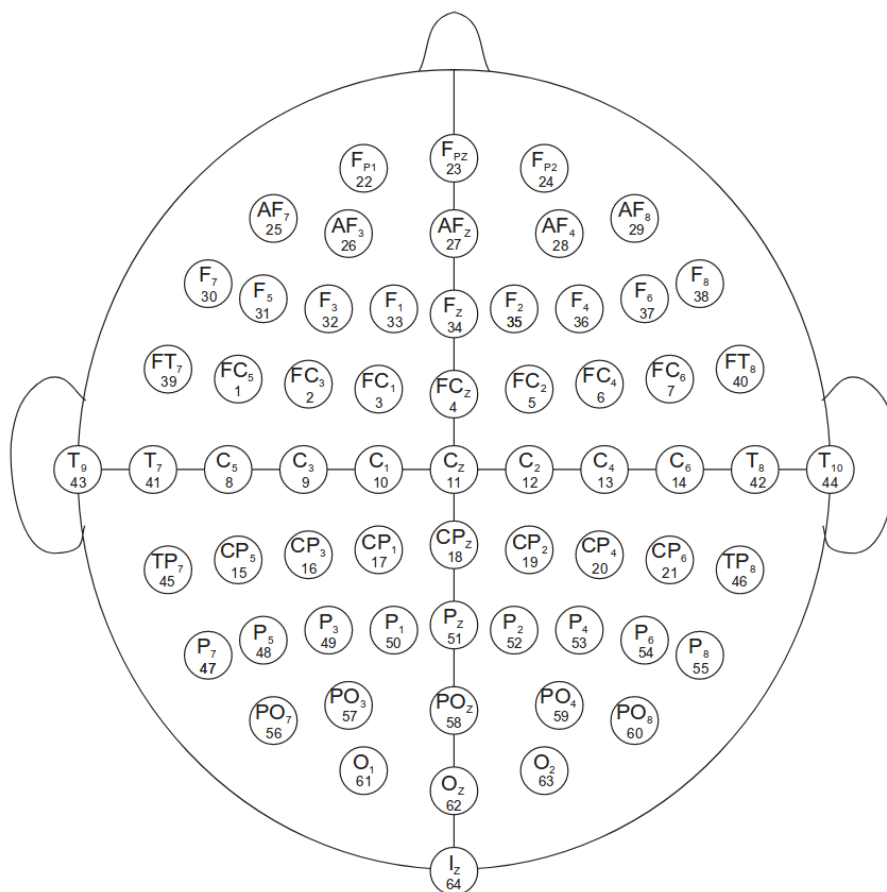


図 5.1: 10-10 システムの電極配置

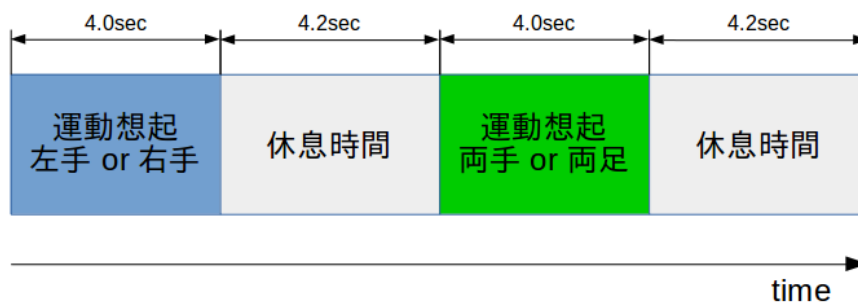


図 5.2: 運動想起 EEG 測定的时间スケジュール

2 クラス分類の問題は表 5.1 に示す通り表記する。問題 LR を例にする。ニューラルネットワークは 4 秒間の EEG データを受け取り、L か R かのいずれかを出力するように構成する。EEG データの与え方及びニューラルネットワークの構成を被験者や問題ごとに変更を行わないことで、単一のニューラルネットワークが被験者に関わらず、また問題に関わらず BCI を構成可能かを評価する。

また評価指標としては正解率を用いる。正解率はニューラルネットワークが EEG を元に出力する分類結果と、EEG 計測時に被験者が行った運動想起が一致している割合である。また、評価に用いられる EEG は学習には用いていないデータである。被験者は 99 人とし、それぞれ 6 つの分類問題が準備されているため計 594 の正解率が得られる。594 の正解率はそれぞれ 10-交差検証によって算出する。各問題について 99 人の被験

表 5.1: 分類問題の種類と論文内の表記

分類問題	論文内の表記
左手 or 右手	LR
両手 or 両足	HF
左手 or 両手	LH
左手 or 両足	LF
右手 or 両手	RH
右手 or 両足	RF

者の正解率の平均を取ることで、単一のニューラルネットワークが各問題に適応可能かを評価する。

5.1.3 ニューラルネットワークの構成

n 階のテンソル $A \in \mathbb{R}^{D_1, \dots, D_n}$ の (d_1, \dots, d_n) 成分を a_{d_1, \dots, d_n} と表記する。入力 $X \in \mathbb{R}^{M \times N \times C}$ に対して Convolution 層のパラメータを $H \in \mathbb{R}^{P \times Q \times C \times L}$ として

$$u_{m,n,k} = \sum_{c=1}^C \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} x_{m+p,n+q,c} h_{p,q,c,k} + b_{m,n,k} \quad (5.1)$$

と計算する。ここでテンソルの各次元について $X \in \mathbb{R}^{space \times time \times frequency}$ であると仮定する。提案手法のニューラルネットワークの構成は入力側から

1. 時間平均と時間分散による正規化
2. 畳み込み層 $H_1 \in \mathbb{R}^{1 \times 75 \times 1 \times 32}$
3. 畳み込み層 $H_2 \in \mathbb{R}^{64 \times 1 \times 32 \times 32}$
4. バッチ正規化
5. 平均プーリング 1 (プーリングサイズ:75 とストライド:25)
6. ドロップアウト
7. $H_3 \in \mathbb{R}^{1 \times 12 \times 32 \times 2}$
8. 平均プーリング 2 (出力サイズが $1 \times 1 \times 2$ となるように設定)
9. ソフトマックス

とした。損失関数としては交差エントロピー関数を用いた。最適化法としては比較的収束が速いとされる Adam を利用した。ドロップアウト率は 0.5 であり、 L_1 正則化や L_2 正則化などの手法は用いていない。平均プーリング 1 は time 方向のみに行った。平均プーリング 2 は出力の要素数が 2 になるように調整し、各要素が 2 クラス分類における logit を出力するようにした。ここでニューラルネットワークでは一般的には用いられない言葉である時間平均と時間分散による正規化について説明する。時間平均と時間分散による正規化は、

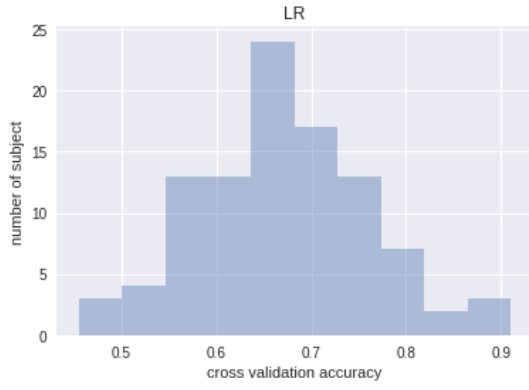


図 5.3: LR 問題の正解率ヒストグラム

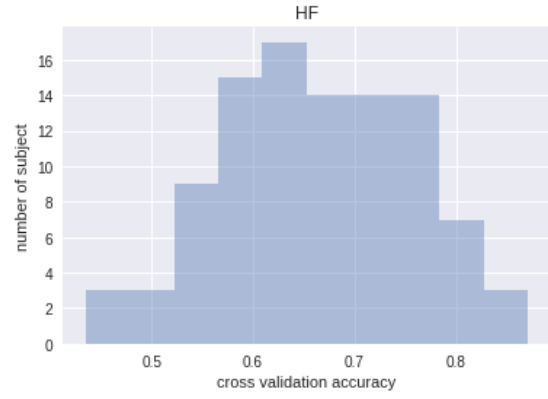


図 5.4: LR 問題の正解率ヒストグラム

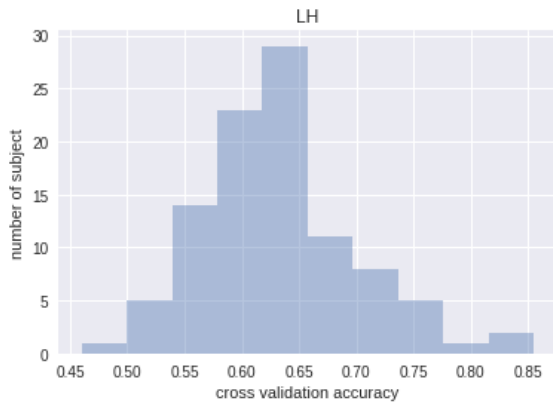


図 5.5: LH 問題の正解率ヒストグラム

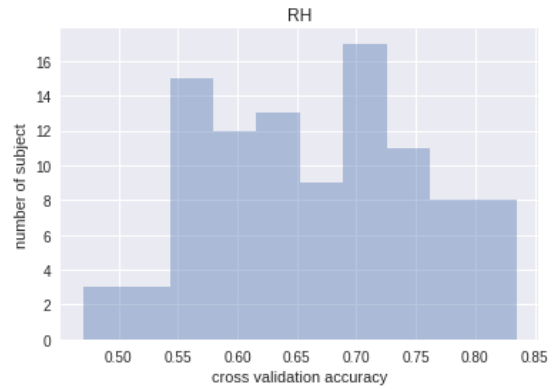


図 5.6: RH 問題の正解率ヒストグラム

$X \in \mathbb{R}^{M \times N \times C}$ に対して

$$\mu_{m,1,1} = \frac{1}{N} \sum_n x_{m,n,1} \quad (5.2)$$

$$\text{var}_{m,1,1} = \frac{1}{N} \sum_n (x_{m,n,1} - \mu_{m,1,1})^2 \quad (5.3)$$

$$x_{m,n,1} \mapsto \frac{x_{m,n,1} - \mu_{m,1,1}}{\sqrt{\text{var}_{m,1,1} + \epsilon}} \quad (5.4)$$

という処理を行う層である。

5.1.4 結果

各分類問題における正解率のヒストグラムを示す(図 5.3-5.8)。横軸が正解率で、縦軸が被験者の数である。どの分類問題においても 60~70% 付近の正解率となる被験者が多い。一方で、80% 以上の正解率となる被験者から 50% の正解率となる被験者までいる。

各分類問題における全被験者の平均を表 5.2 に示す。分類問題毎の平均の差異は小さく、タスクに依らず一定の性能を確保できることが分かる。しかし、この結果はパワースバンドパスフィルタと CSP と LDA を組

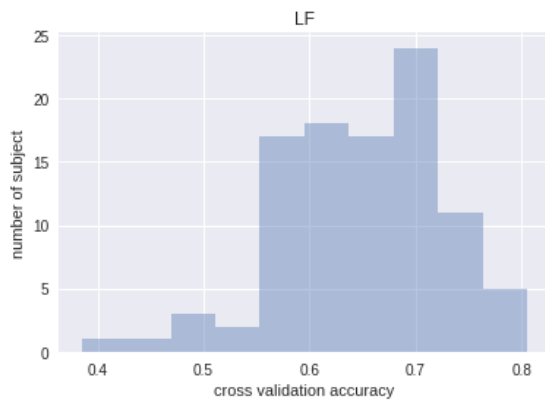


図 5.7: LF 問題の正解率ヒストグラム

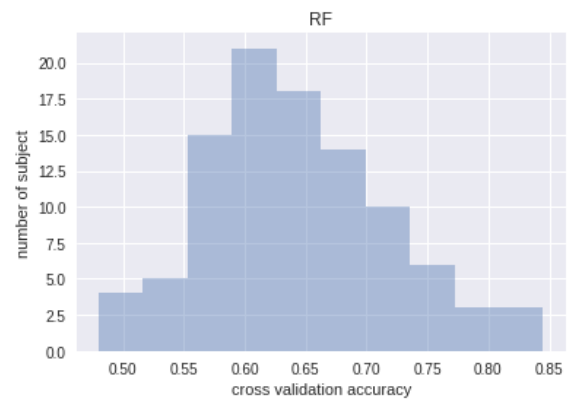


図 5.8: RF 問題の正解率ヒストグラム

表 5.2: 提案手法の問題毎の平均

分類問題	平均正解率
LR	0.6712
HF	0.6636
LH	0.6322
LF	0.6643
RH	0.6517
RF	0.6427

表 5.3: 従来手法の問題毎の平均

分類問題	平均正解率
LR	0.6124
HF	0.7719
LH	0.7760
LF	0.8093
RH	0.8196
RF	0.6701

み合わせた手法 [43] (表 5.3) に対して性能が高いとは言い難い。

5.1.5 考察

表 5.2 によると提案手法 1 のニューラルネットワークはいずれの問題に対しても平均的には同等の対応力を有するように捉えられる。しかし図 5.3-5.8 からは、問題ごとに異なったヒストグラムが得られており任意のタスクに対して同等に学習が可能であるとは言い難い。また、問題 LR 以外は従来手法よりも性能が劣っている。

このような結果となった理由として、実験に対して一回の学習に用いるデータ数が高々 40 個程であったことが問題であると考え。従来手法の最適化は第 2 章で述べた通り、解析的に実行が可能である。また凸最適化問題であるため、BCI の構成要素 1 つ 1 つは仮定した関数族の中で最適であることが保証されている。従って、与えられた分類問題に有効な入出力関係を包含する関数族となっていれば、極めて高い性能を保証することができる。一方でニューラルネットワークの学習は一般的に非凸の最適化問題であり、逐次的な最適化方法を用いなければならない。学習データが少ない場合は逐次最適化で得られる勾配のバリエーションが乏しくなり、探索範囲が極めて狭くなってしまうと推察できる。

一般にニューラルネットワークが表現できる関数の集合は非常に大きい。すなわち提案した手法は分類問題に有効な入出力関係を包含する関数族となっているはずである。しかし、学習データが少ない場合には大きな

表 5.4: 10 交差検証の内訳

検証 1	検証 2	検証 3	検証 4	検証 5	検証 6	検証 7	検証 8	検証 9	検証 10
0.8318	0.8136	0.7909	0.7955	0.7909	0.7500	0.8090	0.8318	0.8045	0.8136

関数の集合の中を十分に探索することができない。

従って特定のタスクに対する個人のための BCI を、ごく少数の学習データによって構築する場合にはニューラルネットワークを用いる方法は不適だと考えられる。

5.2 被験者のデータを混合した BCI 構築

5.2.1 データの混合

実験では LR 問題のみを扱った。LR 問題のデータを 50 人分集めすべてのデータを統合し、2200 個へと増加させた。すなわち左手の運動想起が 1100 回分と右手の運動想起が 1100 回分準備されたことになる。実際には質の悪い EEG に関しては排除したため、右手のデータを 1042 個、左手のデータを 1050 個用いることにし 10-交差検証によって正解率を算出した。

5.2.2 混合データに対する結果

ニューラルネットワークの構造やハイパーパラメータの調整は行っておらず、個人事に学習を行った際に使ったニューラルネットと全く同等のモデルを用いた。結果として交差検証の平均正解率は 0.8032 となった。

また表 5.4 に 10 回の検証のそれぞれの正解率を示した。各検証では 50 人の被験者のデータからランダムに 220 個のデータが検証データに用いられ、残りの 1980 個のデータが学習データとして用いられている。どのような組み合わせで学習が行われても 75% 以上の高い性能を発揮している。

一般にデータの多様性が上がることでニューラルネットワークの要求される表現力も高まるため、非線形活性化関数を用いて以下の実験を行った。

5.2.3 活性化関数に elu を用いた場合の結果

ニューラルネットワークのすべての畳み込み層の後に elu 関数を追加した。一般的に非線形関数を組み込むことでニューラルネットワークの表現力は向上する。結果としては交差検証の平均正解率は 0.8000 となった。活性化関数を用いない場合に比べ、性能は若干劣っている。しかし学習は極めて素早く収束したことが伺える(図 5.10)。ただし、学習を続けても検証データに対する性能は向上しなかった。一方で活性化関数を用いない場合には検証データに対して 80% を越える正解率にたどり着く代わりに収束は極めて遅い。また学習を続けていくと活性化関数を用いる場合と同様に過学習引き起こしている。

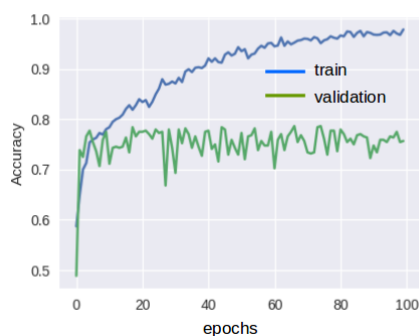


図 5.9: 過学習の様子：elu 活性化関数を用いた場合

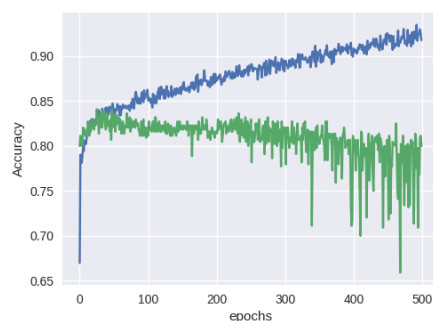


図 5.10: 過学習の様子：活性化関数を用いない場合

5.2.4 考察

結果は個々の人間に対して学習させた正解率の結果よりも向上している。すなわち、学習データが多い場合には、EEG に対してもニューラルネットワークの性能を活用することが可能であると言える。また、50 人すべての LR 問題に対応できるようにニューラルネットワークが学習を行っている点で、個人事に学習したニューラルネットワークよりも難しいタスクをこなしていると言える。

非線形活性化関数を用いない場合にはニューラルネットワークの表現力は著しく制限される。これは学習によって探索する範囲がもともと狭いためであり、その結果として過学習も引き起こりやすくなる。しかし、実際には学習を続けると次第に過学習が引き起こされているのが分かる。

一方で活性化関数を用いる方法では性能こそは劣っているものの極めて収束が速いのが分かる。これは学習によって探索する範囲が広く、勾配学習の際にパラメータ点が移動できる選択肢が多いため最適点を素早く見つけられたためと考えられる。結果として過学習が起こりやすくなるが、今回の場合は過学習した結果検証データに対して著しく性能が低くなることはなかった。活性化関数を用いない方法と正解率は僅差であるので、適切な正則化によって非線形の表現力を維持しながら過学習を抑えるという方向性が有効だと言える。

多数のデータを用いることでニューラルネットが学習可能になったが、多くの被験者の学習データを用いて構築されたニューラルネットワークが社会的に大きなメリットを有するためには、新規の被験者へ転用できることが望ましい。従って以下の追加実験を実施した。

5.3 新規被験者への適用検討

5.3.1 実験内容と結果

ニューラルネットワークの学習データに用いた 50 人の被験者を、以後トレーナーと記す。またトレーナー以外の被験者 49 人をテスターと記す。まずトレーナーの LR データを入力し、学習が確実に行われたのかを

確認した。トレーナーのデータは学習に使われた訓練データも学習に使われていない検証データもまとめてニューラルネットワークへ入力した。従ってニューラルネットワークの評価としては無意味な正解率を算出することとなるが、テスターとの比較を行うために用意した。次にニューラルネットワークに対してテスターの LR データを入力し、その正解率を算出した。すなわち、ニューラルネットワークにとっては全く未知の人間の EEG を受け取ることとなる。

テスターとトレーナーの正解率のヒストグラムを図 5.11-5.12 に示す。図の通りトレーナーとテスターではトレーナーの正解率のほうが比較的高い。これは学習に参加している被験者であるため当然の結果と言える。しかし、テスターのヒストグラム図 5.11 を図 5.13 とを比較した場合に、図 5.11 の方が比較的高い正解率を有する被験者が多いことが分かる。テスターのデータに対する正解率の平均値は 0.7696 であり、テスターに対して個別に学習を行った際の交差検証の正解率 0.6681 よりも高い。

5.3.2 考察

テスターのデータに対する正解率が個々の学習時よりも高かったことから、多数のデータを用いなければニューラルネットワークの性能を引き出せない一方で、多数のデータを用いて学習したニューラルネットワークは運動想起実行時に人間に共通した EEG の特徴を取り出すことができていると考えられる。

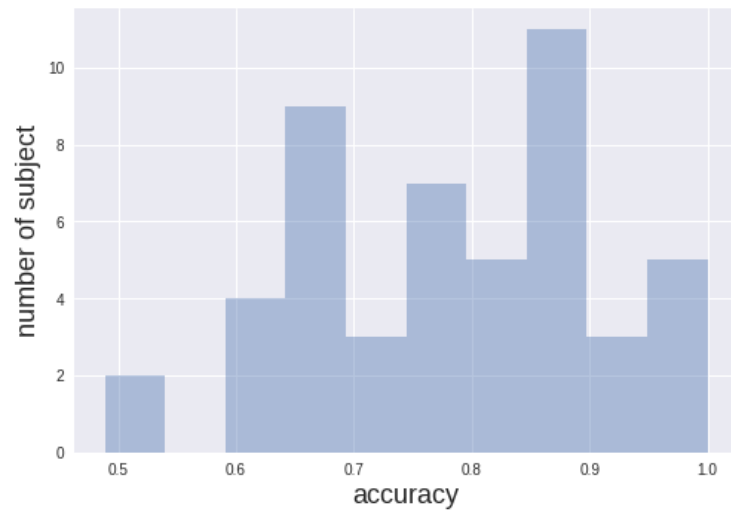


図 5.11: テスターの正解率ヒストグラム

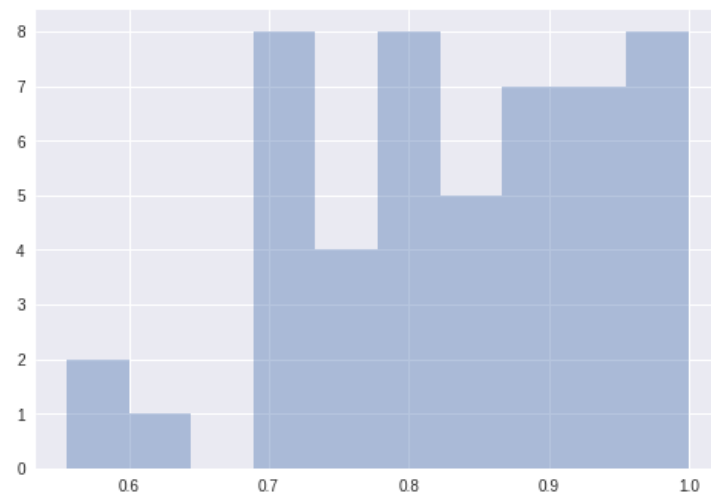


図 5.12: トレーナーの正解率ヒストグラム

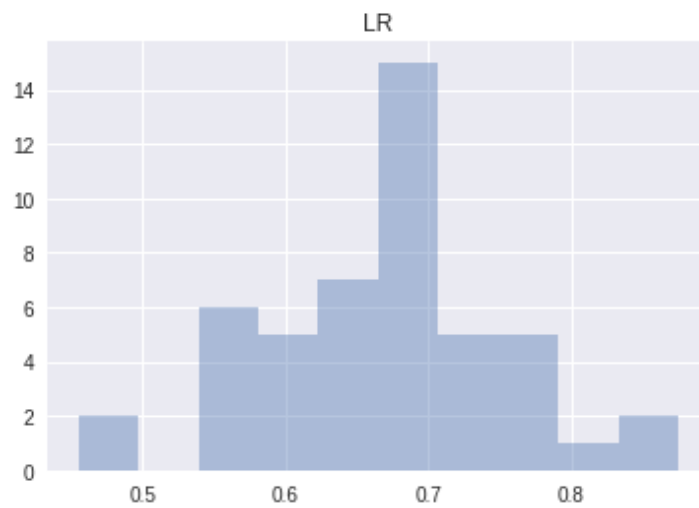


図 5.13: テスター毎に学習した際の交差検証正解率ヒストグラム

第 6 章

結論

6.1 まとめ

第 1 章で BCI に期待される社会的応用と、運動想起型 BCI に対する研究背景を述べた。現状多くの特徴量抽出手法や機械学習手法が試されている中で、FBCSP を除く決定的な BCI のアーキテクチャが生まれていないことについて触れた。本研究では BCI のアーキテクチャとしてニューラルネットワークを用いることを提案し、特に以下の 2 つの項目を強調した。

- 個人差における EEG の解析の必要性を排除
- 人類に共通した一般的な BCI の基本モデル構築

実験結果としては、ニューラルネットワークの構造を変更しなくとも複数の問題について学習によって高い性能を獲得可能かを調べたが、結果として満足のいくものではなかった。しかし、この結果はニューラルネットワークの構造上の問題ではなくデータ量が少ないことによる未学習であると考え、複数の被験者のデータを統合することを試みた。その結果、左手と右手の分類問題に関して学習が可能であることを示した。同時に、新規被験者のデータに対しても新たな学習なしに一定レベルの正解率を得られることから“個人差における EEG の解析の必要性を排除”に関してニューラルネットワークが大いに期待できることを示した。また、同様にして“人類に共通した一般的な BCI の基本モデル構築”の可能性を示唆した。

6.2 今後の展望

本研究では他分野のニューラルネットワーク応用に比べ、比較的簡易なニューラルネットワークの構造を採用した。理由としては End to End の学習が可能であるかを検討し、EEG への応用の裾野を広げたいと考えていたためである。従って、性能を追求した複雑かつ巨大なニューラルネットワークについては実験を行っていない。ある特定の人物と特定の問題に特化させる場合には更なる工夫が必要であると考えられる。しかし、本

研究が示唆する成果によれば、多様性のあるデータで学習した後に特定の問題に調整し直す転移学習が BCI でも可能であると推察される。従って特化型の BCI を考える場合においても、転移学習の可能性を模索していくことが重要だと思われる。

謝辭

参考文献

- [1] 笹山瑛由. ブレイン マシン・インタフェースに向けた多チャンネル脳波計測による脳内情報抽出法および識別法に関する研究. PhD thesis, 京都大学, 2012.
- [2] M. Jeannerod and V. Frak. Mental imaging of motor activity in humans. *Current opinion in neurobiology*, Vol. 9, No. 6, pp. 735–739, 1999.
- [3] G. Pfurtscheller and F. H. L. da Silva. Event-related eeg/meg synchronization and desyn-chronization: Basic principles. *Clinical Neurophysiology*, Vol. 110, No. 11, pp. 1842–1857, 1999.
- [4] 高橋光, 郷古学, 伊藤宏司. 機能的電気刺激 (fes) による感覚情報の事象関連非同期 (erd) への影響. 電子情報通信学会技術研究報告. NC, ニューロコンピューティング, Vol. 108, No. 130, pp. 7–12, jul 2008.
- [5] 高橋光, 郷古学, 伊藤宏司. 運動想起フィードバック訓練による事象関連脱同期 (erd) 出現の検証. システム制御情報学会論文誌, Vol. 22, No. 5, pp. 199–205, 2009.
- [6] D. Huang, K. Qian, S. Oxenham, D. Y. Fei, and O. Bai. Event-related desynchronization/ synchronization-based brain-computer interface towards volitional cursor control in a 2d center-out paradigm. In *2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB)*, pp. 1–8, April 2011.
- [7] G. Pfurtscheller and T. Solis-Escalante. Could the beta rebound in the eeg be suitable to realize a “ brain switch ” ? *Clinical Neurophysiology*, Vol. 120, No. 1, pp. 24–29, 2009.
- [8] Q. Xu, H. Zhou, Y. Wang, and J. Huang. Fuzzy support vector machine for classification of eeg signals using wavelet-based features. *Medical Engineering & Physics*, Vol. 31, No. 7, pp. 858–865, 2009.
- [9] Kenji Kato, Kensho Takahashi, Nobuaki Mizuguchi, and Junichi Ushiba. Online detection of amplitude modulation of motor-related eeg desynchronization using a lock-in amplifier: Comparison with a fast fourier transform, a continuous wavelet transform, and an autoregressive algorithm. *Journal of Neuroscience Methods*, Vol. 293, pp. 289–298, 2018.
- [10] 北山真平, 佐々木実, 伊藤聡, 安田晴信, 小嶋俊史. 運動準備電位による四肢動作推定. 日本 AEM 学会誌, Vol. 22, No. 2, 2014.

- [11] Yong Li, Xiaorong Gao, Hesheng Liu, and Shangkai Gao. Classification of single-trial electroencephalogram during finger movement. *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 6, pp. 1019–1025, June 2004.
- [12] 中村翔太郎, 吉川大弘, 古橋武. 脳波を用いた手足の運動想起判別への空間フィルタ適用に関する検討. 日本知能情報ファジィ学会 ファジィ システム シンポジウム 講演論文集, Vol. 25, pp. 19–19, 2009.
- [13] 翔太郎中村, 弘武高橋, 大弘吉川, 武古橋. 脳波を用いた手足の運動想起判別における準備電位の傾きを用いた特徴抽出法に関する検討. 知能と情報 : 日本知能情報ファジィ学会誌 : journal of Japan Society for Fuzzy Theory and Intelligent Informatics, Vol. 22, No. 4, pp. 427–433, aug 2010.
- [14] Zoltan J. Koles, Michael S. Lazar, and Steven Z. Zhou. Spatial patterns underlying population differences in the background eeg. *Brain Topography*, Vol. 2, pp. 275–284, 1990.
- [15] G. Pfurtscheller, C. Guger, and H. Ramoser. International work-conference on artificial neural networks. In *Engineering Applications of Bio-Inspired Artificial Neural Networks*, 1999.
- [16] Guido Dornhege, Benjamin Blankertz, Matthias Krauledat, Florian Losch, Gabriel Curio, and Klaus-Robert Muller. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. *IEEE TRANSACTIONS ON BIOMEDICAL ENGINEERING*, Vol. XX, No. Y, 2006.
- [17] F. Lotte and C. Guan. Regularizing common spatial patterns to improve bci designs: Unified theory and new algorithms. *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 2, pp. 355–362, 2011.
- [18] N. Behrooz, B. Reza, and J. M. Zolghadri. An efficient hybrid linear and kernel csp approach for eeg feature extraction. *Neurocomputing*, Vol. 73, pp. 432–437, 2009.
- [19] Jyoti Singh Kirara and R. K. Agrawala and. Optimal spatio-spectral variable size subbands filter formotor imagery brain computer interface. *Procedia Computer Science*, Vol. 84, pp. 14–21, 2016.
- [20] Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (fbcs) in brain-computer interface. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2390–2397, June 2008.
- [21] Bashashati H, Ward RK, Birch GE, and Bashashati A. Comparing different classifiers in sensory motor brain computer interfaces. *PLoS ONE*, Vol. 10, No. 6, 2015.
- [22] 板井陽俊, 船瀬親王. Bci システムの構築を目指した単一試行脳波の解析技術. 日本神経回路学会誌, Vol. 19, No. 3, pp. 118–125, 2012.
- [23] 吉岡将孝. 脳波分析に基づいた Brain-Machine Interface パワーアシストシステムの構築 -周期パワースペクトルによる運動に関する脳波の解析および関節トルクの推定-. PhD thesis, 前橋工科大学, 2016.
- [24] E. Thomas, J. Fruitet, and M. Clerc. Investigating brief motor imagery for an erd/ers based bci. In *2012*

- Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2929–2932, Aug 2012.
- [25] S. R. Sreeja, Joytirmoy Rabha, Debasis Samanta, Pabitra Mitra, and Monalisa Sarma. *Classification of Motor Imagery Based EEG Signals Using Sparsity Approach*, pp. 47–59. Springer International Publishing, Cham, 2017.
- [26] Rachel Rajan and Sunny Thekkan Devassy. *Improving Classification Performance by Combining Feature Vectors with a Boosting Approach for Brain Computer Interface (BCI)*, pp. 73–85. Springer International Publishing, Cham, 2017.
- [27] Henry W. Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, Vol. 168, pp. 1223–1247, 2017.
- [28] Hrushikesh Mhaskar and Tomaso A. Poggio. Deep vs. shallow networks : An approximation theory perspective. *Analysis and Applications*, Vol. 14, , November 2016.
- [29] Kenji Kawaguchi. Deep learning without poor local minima. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems* 29, pp. 586–594. Curran Associates, Inc., 2016.
- [30] M. Sakai. Kernel nonnegative matrix factorization with constraint increasing the discriminability of two classes for the eeg feature extraction. In *2013 International Conference on Signal-Image Technology Internet-Based Systems*, pp. 966–970, Dec 2013.
- [31] Hyekyoung Lee, Andrzej Cichocki, and Seungjin Choi. Kernel nonnegative matrix factorization for spectral EEG feature extraction. *Neurocomputing*, Vol. 72, No. 13-15, pp. 3182–3190, 2009.
- [32] 日野幹雄. スペクトル解析 (統計ライブラリー). 朝倉書店, 1977.
- [33] Hiromu Gotanda, Takaaki Ishibashi, Nobuo Iwasaki, and Katsuhiro Inoue. 独立成分分析の基礎と応用. Technical Report 1743, 数理解析研究所, 2011.
- [34] 村田昇. 入門 独立成分分析. 東京電機大学出版, 2004.
- [35] 開一夫, 金山範明河内山隆紀, 松本敦宮腰誠. 脳波解析入門:EEGLAB と SPM を使いこなす. 東京大学出版会, 2016.
- [36] 東広志, 田中聡久. 信号構造を利用する脳波処理. 計測と制御, Vol. 55, No. 11, pp. 960–965, 2016.
- [37] Guo Xiaojing, Wu Xiaopei, and Zhang Dexiang. Motor imagery eeg detection by empirical mode decomposition. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pp. 2619–2622, June 2008.
- [38] SachinTaran, VarunBajaj, DheerajSharma, SiulySiuly, and A.Sengur. Features based on analytic imf for

- classifying motor imagery eeg signals in bci applications. *Measurement*, Vol. 116, pp. 68–76, 2018.
- [39] B. Blankertz, M. Kawanabe, R. Tomioka, F. Hohlefeld, V. Nikulin, and K. R. M. In-variant common spatial patterns: Alleviating nonstationarities in brain-computer interfacing. *Advances in Neural Information Processing Systems*, Vol. 20, pp. 113–120, 2008.
- [40] Kai Keng Ang, Zheng Yang Chin, ChuanchuWang, , Cuntai Guan, and Haihong Zhang. Filter bank common spatial pattern algorithm on bci competition iv datasets 2a and 2b. *frontiers in NEUROSCIENCE*, Vol. 29, pp. 2390–2397, March 2012.
- [41] G. Schalk, D. J. McFarland, T. Hinterberger, N. Birbaumer, and J. R. Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on Biomedical Engineering*, Vol. 51, No. 6, pp. 1034–1043, June 2004.
- [42] Ary L. Goldberger, Luis A. N. Amaral, Leon Glass, Jeffrey M. Hausdorff, Plamen Ch. Ivanov, Roger G. Mark, Joseph E. Mietus, George B. Moody, Chung-Kang Peng, and H. Eugene Stanley. Physiobank, physiotoolkit, and physionet. *Circulation*, Vol. 101, No. 23, pp. e215–e220, 2000.
- [43] Hyun Seok Kim, Min Hye Chang, Hong Ji Lee, and Kwang Suk Park. A comparison of classification performance among the various combination of motor imagery tasks for brain-computer interface. In *6th IEEE Annual International EMBS Conference of Neural Engineering*, pp. 435–438, November 2013.