

マウスの脳波／筋電データを利用した睡眠ステージ判定

鈴木 悠太[†] 佐藤 牧人^{††} 塩川 浩昭^{†††} 柳沢 正史^{††} 北川 博之^{†††}

[†] 筑波大学 システム情報工学研究科 〒 305-8573 茨城県つくば市天王台 1-1-1

^{††} 筑波大学 国際統合睡眠医科学研究機構 (WPI-IIIS) 〒 305-8575 茨城県つくば市天王台 1-1-1

^{†††} 筑波大学 計算科学研究センター 〒 305-8577 茨城県つくば市天王台 1-1-1

E-mail: [†]kitamann.2312@kde.cs.tsukuba.ac.jp, ^{††}{sato.makito.ge,yanagisawa.masa.fu}@u.tsukuba.ac.jp,
^{†††}{shiokawa,kitagawa}@cs.tsukuba.ac.jp

あらまし マウスの脳波や筋電データから睡眠ステージを判定することは、睡眠解析において極めて重要な課題となっている。しかし、現在の睡眠ステージの判定は熟練の判定者の目視による判定が主流となっており、精度の高い自動睡眠ステージ判定手法は未だ存在していない。そのため、睡眠ステージの判定には膨大な人件費や時間が必要となり、睡眠研究および診断の大規模化の大きな障害となっている。精度の高い自動睡眠ステージ判定が実現できない原因として、脳波および筋電データから特徴抽出を行う際に全ての睡眠ステージの特徴を十分に抽出できていないこと、および、睡眠ステージの時系列性を考慮していないことの2点が考えられる。そこで本研究では、(1) 各睡眠ステージの特徴をよく表現できる特徴量の推定、(2) 睡眠ステージの時系列性を考慮した特徴量の作成、および(3) 分類器による判定において複数の睡眠ステージの特徴を持った確実性が低い判定に対する再判定により高精度な睡眠ステージ判定を行う。本研究では、実際のマウスの脳波および筋電データを用いた評価実験により提案手法の有効性を示した。

キーワード マウス, 脳波, 筋電データ, 睡眠ステージ判定, 時系列データ, 主成分分析, SVM, 機械学習

1. はじめに

臨床用、研究用を含めて、ノンレム睡眠やレム睡眠、覚醒といった睡眠ステージに関する情報は睡眠解析を行う上で基準となる情報である。しかし、睡眠ステージの自動判定は、現在市販されているソフトウェアも含めて、専門家の要求する判定精度95%以上を実現している手法は未だ存在していない。睡眠ステージを判定する場合、熟練の判定者の目視による手作業での睡眠ステージ判定が現在主流となっている。このため、睡眠研究および診断を行う上で手作業での睡眠ステージ判定では、人件費や時間といったコストが大きくなるため、睡眠研究の大規模化が困難となっている。ゆえに、睡眠ステージの高精度な自動判定技術の確立は、睡眠研究の大規模化において極めて重要な課題となっている。

この問題を解決するために、1960年頃から様々な動物を対象として睡眠ステージ判定手法が開発されてきた。これまでに、ニューラルネットワークを用いた手法 [1] や波形認識法と決定木学習を用いた手法 [2] などの様々な分類方法を利用した睡眠ステージ判定手法が開発されてきた。最近では、判定精度を向上させるために特徴量について解析・考察を行う手法が増えており、線形判別分析を用いた手法 [3] や単純ベイズ分類器を用いた手法 [4]、特徴量抽出と赤外線センサを用いた手法 [5]、密度ベースクラスタリングを用いた手法 FASTER [6] などが開発されてきた。今まで開発されてきたこれらの手法は、ノンレム睡眠や覚醒に対する判定精度は比較的専門家が要求する95%に近い値を実現するのに対し、レム睡眠に対する判定精度についてはまだ50~80%程度が限界という問題がある。精度の高い自

動睡眠ステージ判定が実現できない原因としては、次の2つのような問題が考えられる：(a) 脳波および筋電データから特徴抽出を行う際に全ての睡眠ステージの特徴、その中でも特にレム睡眠の特徴を十分に抽出することができていないこと、および(b) 人間が目視により判定する場合には前後のデータとの時系列的な繋がりを考慮しながら判定を行うのに対して、多くの既存手法では脳波および筋電データの波形の情報のみから睡眠ステージを判断しており、睡眠ステージの時系列性を考慮した判定を行っていないことの2点である。

そこで本研究では、上記の問題を解決するマウスの睡眠ステージ判定手法 MASC を提案する。本研究の提案手法 MASC は以下の3つのアプローチをとる：(1) 使用する特徴量の違いによる各睡眠ステージの判定精度の違いを調べることで、各睡眠ステージの特徴をよく表現できる特徴量を推定する、(2) 分類器に与えるパラメータとして睡眠ステージの時系列性を考慮した特徴量の作成を行う、(3) 分類を行った結果について、分類器による判定の確実性が低い判定に対する再判定を行う。分類器による判定の確実性が低い判定とは、複数の睡眠ステージの特徴を持つようなデータに対する判定のことを示す。確実性が低い判定の場合であっても、SVMなどの分類器では確率が一番高い睡眠ステージへ必ず分類されるが、確実性が低い場合正しくない判定をしている可能性が高い。

本研究では、上述のアプローチ(1)~(3)を適用することで高い睡眠ステージ判定精度を実現した睡眠ステージ判定手法の構築を目指す。本稿では、提案手法 MASC の詳細について述べるとともに、実際のマウスの脳波および筋電データを利用して、既存手法との比較を通じた提案手法 MASC の有効性について

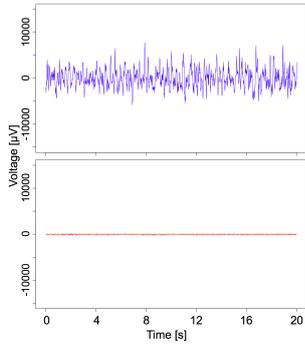


図1 典型的なマウスのノンレム睡眠の波形例 (上段：脳波，下段：筋電データ)

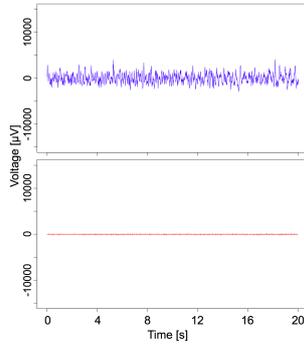


図2 典型的なマウスのレム睡眠の波形例 (上段：脳波，下段：筋電データ)

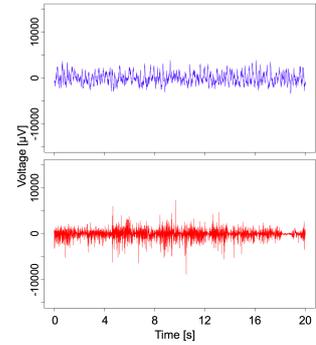


図3 典型的なマウスの覚醒の波形例 (上段：脳波，下段：筋電データ)

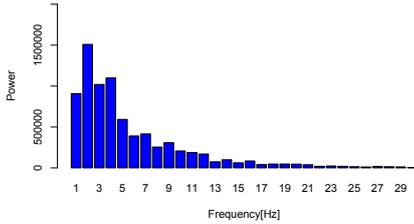


図4 典型的なマウスのノンレム睡眠における脳波の周波数スペクトル例

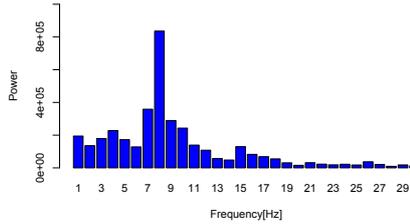


図5 典型的なマウスのレム睡眠における脳波の周波数スペクトル例

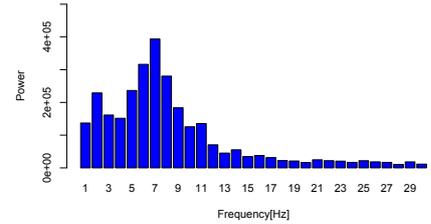


図6 典型的なマウスの覚醒における脳波の周波数スペクトル例

議論する。

本稿の構成は以下のとおりである。まず2節にて、マウスの睡眠ステージの概要と本研究で扱うマウスの睡眠ステージの判定方法について説明する。3節では、関連研究について説明する。4節では、前述の3つのアプローチを用いた提案手法であるMASCについて述べる。また、5節で提案手法の有効性を確かめるための評価実験について述べる。最後に6節にて、本研究のまとめと今後の課題について述べる。

2. 事前準備：マウスの睡眠ステージの判定方法

2.1 脳波/筋電データを用いたマウスの睡眠ステージ判定

現在利用されている睡眠ステージ判定方法は、マウスの脳波および筋電データと睡眠ステージとの関係性を用いて判定するものが主流である。脳波とはマウスの脳から生じる電気活動を脳表などに置いた電極で記録したものであり、筋電データとは脊髄から生じる電気活動を脊髄に置いた電極で記録したものである。睡眠ステージ判定では、まず脳波および筋電データについて2.2節で説明するエポックという単位にデータを分割する。その後、各エポックについて、2.3節で説明する脳波および筋電データの振幅や周波数特性といった睡眠ステージごとの特徴の違いを利用することで睡眠ステージを判定する。

2.2 エポック

睡眠ステージ判定を含む睡眠解析では、一般的に脳波および筋電データを一定区間ごとに区切って解析を行う。一定期間に区切った脳波および筋電データをエポックと呼ぶ。マウスの睡眠解析を行う場合、一般的にエポックの間隔は4秒から20秒が用いられており、本研究ではエポックの間隔は20秒間隔と

設定する。脳波および筋電データを20秒間隔のエポックに分割する例を図7に示す。一般的にマウスの睡眠ステージ判定では、各エポックに対してその脳波および筋電データの波形を比較することで睡眠ステージを判定する。

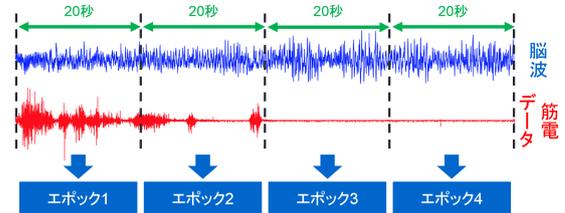


図7 エポックの分割方法

2.3 マウスの睡眠ステージ

マウスにおける睡眠ステージは、エポックをノンレム睡眠、レム睡眠、覚醒の3つのステージに分類することができる。

図1はノンレム睡眠における典型的な波形を示し、図4は図1における脳波の周波数スペクトルを示したものである。ノンレム睡眠では、脳波は1Hzから6Hzの低周波成分(デルタ波)を多く含んで振幅が大きくなるのに対し、筋電データの振幅は小さくなる特徴がある。

図2はレム睡眠における典型的な波形を示し、図5は図2における脳波の周波数スペクトルを示したものである。レム睡眠では、脳波は7Hzから11Hzの高周波成分(シータ波)を多く含むがノンレム睡眠と比べて振幅が小さくなる、また、筋電データの振幅は非常に小さいという特徴をもつ。

図3は覚醒における典型的な波形を示し、図6は図3における脳波の周波数スペクトルを示したものである。覚醒では、脳

波はレム睡眠時の脳波パターンと似た周波数特性を持つのに対し、筋電データの振幅は非常に大きいという特徴がある。

3. 関連研究

3.1 FASTER

密度ベースクラスタリングを用いた睡眠ステージ判定手法として、Sunagawa らの研究がある [6]。FASTER では、同じ睡眠ステージに分類されるエポックは脳波および筋電データを主成分分析することにより得られる特徴空間において分布が密集するという性質に基づいて睡眠ステージ判定を行う。FASTER ではまず、主成分分析 [10] により脳波および筋電データから特徴抽出を行う。次に、脳波および筋電データに関して、ノンパラメトリック密度推定クラスタリング [11] により特徴が似ているエポック同士を同じクラスタにまとめる。最後に、各クラスタの特徴から最も類似する睡眠ステージにエポックを分類する。

FASTER による睡眠ステージ判定では、ノンレム睡眠および覚醒に対しては非常に高い判定精度で分類することができるが、レム睡眠に対する判定精度は 80% であるという問題がある。

3.2 exFASTER

密度ベースクラスタリングと状態遷移モデルを用いた睡眠ステージ判定手法として我々の先行研究である exFASTER がある [7]。

Sunagawa らの睡眠ステージ判定手法 FASTER は、時系列性を考慮していない分類であるという特徴がある。しかし、時系列の情報には重要な情報が含まれており、人間が目視で睡眠ステージを判定する場合、前後数エポックについての睡眠ステージの流れを見て判定することが知られている。また、睡眠ステージは頻繁に遷移するものではないという特徴がある。したがって、コンピュータによる判定において、数秒から数十秒の間で異なる睡眠ステージに遷移している判定をした場合、その判定は誤っている可能性が高いと考えられる。

そこで我々は FASTER を改良し、各睡眠ステージを隠れマルコフモデル [13] に準じた状態としてモデル化した。これにより、FASTER による判定結果の中から確率的に不自然な睡眠ステージ遷移をしているエポックを検出し、より妥当と考えられる睡眠ステージへ再分類する手法 exFASTER を提案した。

exFASTER を用いた睡眠ステージ判定により、FASTER と比較してノンレム睡眠に対する判定精度を 4% 以上向上することができたが、FASTER において問題となっていたレム睡眠に対する判定精度は同等であった。

4. 提案手法 MASC

既存手法においてレム睡眠の Sensitivity が低い原因として、先行研究 [7] でも着目していた時系列性の問題に加えて、睡眠ステージの特徴抽出を行うときにレム睡眠の特徴が十分に抽出することができていない可能性があると考えた。そこで本研究では、各睡眠ステージをよく表現できる特徴量の抽出および睡眠ステージの時系列性の考慮した分類に着目した睡眠ステージ判定手法 MASC について提案する。

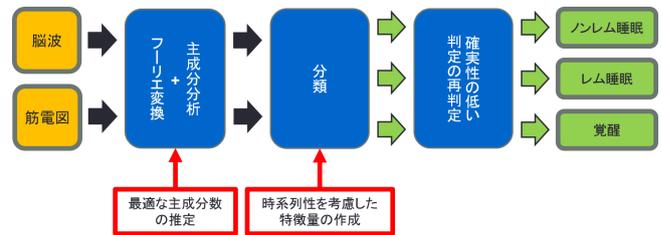


図 8 提案手法 MASC の概要

4.1 概要

本節で提案する手法 MASC の概要を図 8 に示す。

本手法では、まずエポックごとに分割された脳波および筋電データを FFT [12] により周波数スペクトルに変換し、それぞれの周波数スペクトルから主成分分析を用いて特徴量抽出 (4.2 節) を行う。その後、抽出した特徴量と睡眠ステージの時系列性を考慮した特徴量を入力パラメータとして与えた分類器を用いて睡眠ステージ判定を行う (4.3 節)。本手法では分類器として SVM (Support Vector Machine) [14] を使用して睡眠ステージ判定を行った。最後に、分類器を用いて睡眠ステージ判定を行った結果の中で、分類器による判定の確実性が低い判定に対する再判定について 4.4 節で説明する。

4.2 主成分分析を用いた特徴量抽出

脳波および筋電データから各睡眠ステージの特徴を抽出する方法として、本手法では主成分分析 [10] を使用する。特徴量として主成分を用いた分類では主成分数によって判定精度が大きく変動するため、妥当な主成分数を決める必要がある。そこで本研究では、脳波および筋電データの主成分数を変化することによる判定精度への影響について実験を行う。最適な主成分数の推定に関する評価実験について 5.3.1 節で説明する。

4.3 時系列性を考慮した特徴量を用いた分類

3.2 章で述べた通り、一般的に睡眠ステージは一定の規則にしたがって遷移することが知られており、睡眠ステージ判定において時系列性を考慮することは重要であるといえる。本研究では、分類器へ入力する特徴量として、脳波および筋電データの主成分に加えて、直前や直後の睡眠ステージ判定結果を加えることを考える。

本研究では、時系列性を考慮した特徴量として、(1) 睡眠ステージの判定結果を文字列シーケンスとする特徴量、(2) 直前または直後数エポックにおける各睡眠ステージの占有確率を重みなし確率とした特徴量、および (3) 各睡眠ステージの占有確率を重み付き確率とした特徴量の 3 種類の特徴量を提案する。これら 3 種類の特徴量について、判定対象のエポックの直前 k エポックからの時系列性のみを考慮した場合と、判定対象のエポックの前後 k エポックの時系列性を考慮した場合を組み合わせた 6 種類の特徴量を用いて分類を行う。

図 9 を例にして、考慮するエポックの数を $k = 3$ としたときの特徴量の詳細についていくつか説明する。図 9 は連続する 7 エポックについて、判定対象をエポック 4 としたときの前後 3 エポックの睡眠ステージ判定結果を示したものである。

MASC は、時系列性を考慮した特徴量として t^- および t^+

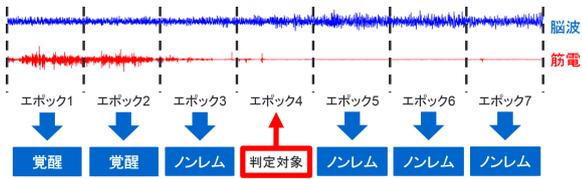


図9 時系列性を考慮した特徴量の作成

という特徴量を抽出する。\$t^-\$ は判定対象の直前 \$k\$ エポックの睡眠ステージ判定結果により得られる時系列特徴量を示し、\$t^+\$ は判定対象の直後 \$k\$ エポックの睡眠ステージ判定結果により得られる時系列特徴量を示したものである。

a) 直前 \$k\$ エポックの睡眠ステージ判定結果を文字列シーケンスとした特徴量

図9より、判定対象の直前の3エポックについての睡眠ステージ判定結果は、時系列が古い順に「覚醒」「覚醒」「ノンレム睡眠」であることがわかる。この特徴量では、睡眠ステージ判定結果を時系列が古い順に \$t^- = \{\text{覚醒}, \text{覚醒}, \text{ノンレム睡眠}\}\$ という3次元の文字シーケンスとして特徴量を保持する。

b) 前後 \$k\$ エポックにおける各睡眠ステージの占有確率を重みなし確率とした特徴量

図9より、判定対象の直前3エポックについてノンレム睡眠が占める割合は \$1/3 = 33\%\$、レム睡眠は \$0/3 = 0\%\$、覚醒は \$2/3 = 67\%\$ であり、直後3エポックについてノンレム睡眠が占める割合は \$3/3 = 100\%\$、レム睡眠は \$0/3 = 0\%\$、覚醒は \$0/3 = 0\%\$ である。この特徴量では、各睡眠ステージの占有確率を \$t^- = \{p_{NREM}^-, p_{REM}^-, p_{Wake}^-\} = \{0.33, 0.00, 0.67\}\$, \$t^+ = \{p_{NREM}^+, p_{REM}^+, p_{Wake}^+\} = \{1.00, 0.00, 0.00\}\$ という6次元の特徴量として保持する。ただし、\$p_{NREM}^-, p_{REM}^-, p_{Wake}^-\$ はそれぞれ直前 \$k\$ エポックにおける各睡眠ステージの占有確率を示し、\$p_{NREM}^+, p_{REM}^+, p_{Wake}^+\$ はそれぞれ直後 \$k\$ エポックにおける各睡眠ステージの占有確率を示す。

c) 前後 \$k\$ エポックにおける各睡眠ステージの占有確率を重み付き確率とした特徴量

図9のエポック4の睡眠ステージを判定するときに用いる時系列情報としては、エポック4から遠いエポック1やエポック7の睡眠ステージの情報よりもエポック4に近いエポック3やエポック5の睡眠ステージの情報のほうが重要度が高いといえる。そこで本特徴量では、判定対象であるエポックとの距離を示す位置関数および占有確率の重み関数を定義する。

[定義 4.1] (位置関数) 判定対象エポックを \$e_t\$、直前または直後のエポックを \$e\$ とするとき、エポック \$e\$ の位置を表す位置関数 \$P(e)\$ は下記の通りである。

$$P(e) = \begin{cases} -\frac{i}{k+1} & (e \text{ が } e_t \text{ より } i \text{ 個前のエポックである場合}) \\ 0 & (e = e_t) \\ \frac{i}{k+1} & (e \text{ が } e_t \text{ から } i \text{ 個後のエポックである場合}) \end{cases}$$

図9の例の場合では、\$P(\text{エポック } 3) = -\frac{1}{3+1} = -0.25\$、\$P(\text{エポック } 7) = \frac{3}{3+1} = 0.75\$ となる。

[定義 4.2] (重み関数) 睡眠ステージを \$s \in \{\text{レム睡眠}, \text{ノン}

レム睡眠, 覚醒\}, 位置 \$p\$ が表すエポックの睡眠ステージを \$S(p)\$ としたとき、重み関数 \$w_s\$ を以下に示す。

$$w_s = \frac{\sum_p g_s(p)}{\sum_s \sum_p g_s(p)}.$$

ただし、\$g_s(p)\$ はガウス関数であり、計算式を以下に示す。また、本研究では、\$\sigma = 0.8\$ を用いた。

$$g_s(p) = \begin{cases} \exp\left(-\frac{p^2}{\sigma^2}\right) & (S(p) = s) \\ 0 & (\text{Otherwise}) \end{cases}$$

\$k = 3\$ で直前の時系列性のみを考慮する場合、定義 4.2 より重みの値は判定対象に近いエポックから順に 0.82, 0.17, 0.01 となる。したがって、図9の場合、判定対象に近いエポック3とエポック5の重みが大きくなるため、この特徴量では、各睡眠ステージの重み付き占有確率を \$t^- = \{p_{NREM}^-, p_{REM}^-, p_{Wake}^-\} = \{0.82, 0.00, 0.18\}\$, \$t^+ = \{p_{NREM}^+, p_{REM}^+, p_{Wake}^+\} = \{1.00, 0.00, 0.00\}\$ という6次元の特徴量として保持する。

時系列性を考慮した特徴量を用いることにより判定精度への影響と考慮するエポックの数 \$k\$ の違いによる影響について 5.3.2 節で説明する。5.3.1 節で推定する脳波および筋電データの主成分と 5.3.2 節で作成する時系列特徴量により学習した SVM を用いて睡眠ステージを判定する。

4.4 確実性の低い判定に対する再判定

本実験における睡眠ステージの判定に用いた分類器は、One-Versus-Rest 形式の多クラス分類 SVM [15] を用いて分類を行う。One-Versus-Rest 形式の多クラス分類 SVM を用いた理由は、SVM は識別境界面から最近傍エポックまでの距離であるマージンが最大となるように識別境界面を設定することから、未知データに対する汎化性能が高いと考えられるためである。One-Versus-Rest 形式の多クラス分類 SVM では、「ノンレム睡眠とそれ以外の睡眠ステージを判定する \$SVM_N\$」、「レム睡眠とそれ以外を判定する \$SVM_R\$」、および「覚醒とそれ以外を判定する \$SVM_W\$」の3つの SVM を組み合わせて分類を行う。各 SVM では、特徴量空間において特定の睡眠ステージのエポックとそれ以外の睡眠ステージのエポックを識別境界面により二分する。特定の睡眠ステージのエポックが密集している方向を正方向、それ以外の睡眠ステージのエポックが密集している方向を負方向として識別境界面からの距離を求め、正方向に最も離れている SVM の睡眠ステージを判定結果とする。

3つの SVM による判定の中で、1つの睡眠ステージを分類する SVM でのみ正方向に大きく離れているような場合であれば、その判定は正しい判定である可能性が高い。しかし、複数の睡眠ステージの特徴を持つようなエポックの場合、2つ以上の睡眠ステージについて正方向に大きく離れておりそれぞれ識別境界面からの距離がほぼ同じような状況において、コンピュータは機械的に一番距離が離れている睡眠ステージを判定結果とするが、この判定は誤っている可能性が高いと考えられる。

あるエポック \$e\$ は、識別境界面からマージンまでの距離を 1 としたとき、各 SVM (\$SVM_N, SVM_R, SVM_W\$) の識別境界面から特徴点までの正方向の距離を \$d_N, d_R, d_W\$ とする。負方

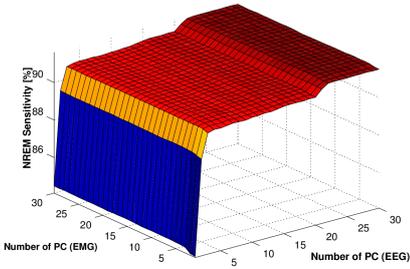


図 10 主成分とノンレム睡眠 Sensitivity の関係

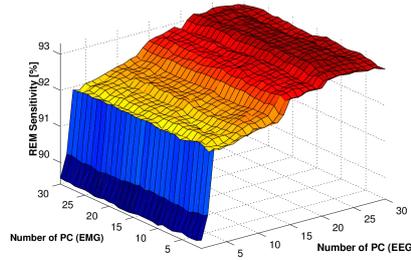


図 11 主成分とレム睡眠 Sensitivity の関係

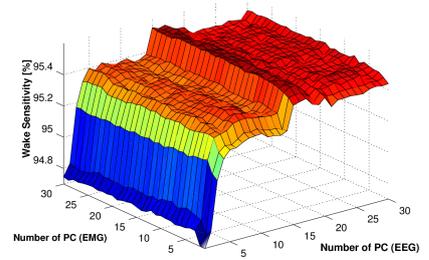


図 12 主成分と覚醒 Sensitivity の関係

向に特徴点が存在する場合、 d_s ($s \in \{N, R, W\}$) の値は負となる。正方向に最も離れた睡眠ステージと 2 番目に離れた睡眠ステージの距離 d_s の差 $D(e)$ は以下の式で表される。

$$D(e) = \max(d_s) - \text{next_max}(d_s)$$

MASC では、4.2 節および 4.3 節で抽出した特徴量を用いて分類した睡眠ステージ判定結果の中で、 $D(e)$ の値が閾値 θ_D よりも小さい判定を確実性が低い判定であるとする。そして確実性が低い判定に対して 5.3.2 節で作成する特徴量とは異なる時系列特徴量により学習した SVM を用いて再判定を行う。閾値 θ_D の推定および再判定に用いる分類器の選定について 5.3.3 節および 5.3.4 節で説明する。

5. 評価実験

本節では、提案手法 MASC のパラメータチューニングに関する予備実験、および MASC の有効性を示すために実施したノイズが少ないマウスおよびノイズが多いマウスを対象とした評価実験について述べる。本実験は 14-Fold Cross Validation で実験を行った。

5.1 データセット

本実験で用いるマウスのデータは、比較的ノイズが少ないマウス 14 匹に関する人手でラベル付けした脳波および筋電データを使用する。本実験で用いるマウスはすべて、薬物投与や遺伝子組み換えなどが行われていない健康体を表す WildType である。脳波および筋電データのサンプリング周波数は 250Hz、エポックサイズは 20 秒である。エポック数は 4 日分の 17,280 エポックであり、14 匹のマウスで計 241,920 エポックである。

5.2 評価指標

本実験では、睡眠ステージ判定の判定精度を計算するための評価指標として、Sensitivity [6] と Specificity [6], Accuracy [6] を用いる。Sensitivity と Specificity は、各睡眠ステージごとに計算することとする。Sensitivity は、表 1 を用いて以下の計算式で表される。

$$\text{Sensitivity} = \frac{TP}{TP + FN}$$

この式は、レム睡眠の Sensitivity の場合であれば、人間による判定でレム睡眠と判定されたエポックの中で、コンピュータによる判定でも正しくレム睡眠と判定されたものの割合を示す。

これに対して Specificity は、表 1 を用いて以下の計算式で表される。

$$\text{Specificity} = \frac{TN}{FP + TN}$$

この式は、レム睡眠の Specificity の場合であれば、人間による判定でレム睡眠ではないと判定されたエポックの中で、コンピュータによる判定でも正しくレム睡眠ではないと判定されたものの割合を示す。

Accuracy は、 TP_N , TP_R , TP_W をそれぞれノンレム睡眠、レム睡眠、覚醒の場合における TP の値とし、同様に FN_N , FN_R , FN_W をそれぞれノンレム睡眠、レム睡眠、覚醒の場合における FN の値とした時、以下の計算式で表される。

$$TP_{ALL} = TP_N + TP_R + TP_W$$

$$FN_{ALL} = FN_N + FN_R + FN_W$$

$$\text{Accuracy} = \frac{TP_{ALL}}{TP_{ALL} + FN_{ALL}}$$

この式は、人間による判定とコンピュータによる判定が一致した割合を示している。

本実験では、これら 7 つの評価指標を用いて、睡眠ステージ判定手法の判定精度を計算する。

表 1 評価指標の計算

		コンピュータによる判定	
		○	×
人間による判定	○	TP	FN
	×	FP	TN

5.3 予備実験：適切なパラメータの決定

5.3.1 最適な主成分数の推定

分類器の入力パラメータとして与える脳波および筋電データの周波数スペクトルの主成分を、それぞれ第 1 主成分から第 30 主成分まで変化させて分類を行った場合についてのノンレム睡眠の Sensitivity を図 10、レム睡眠の Sensitivity を図 11、覚醒の Sensitivity を図 12 に示す。

図 10 から図 12 より、脳波の主成分は主成分数が増加に伴い全ての睡眠ステージにおいて判定精度が段階的に向上する傾向にあり、第 20 主成分を超えたところで判定精度が一定になることがわかる。筋電データの主成分については、ノンレム睡眠およびレム睡眠の Sensitivity は主成分数に関係なくほぼ一定であるが、覚醒の Sensitivity は主成分数が 4 以上の時に値が

表 2 k の違いによるノンレム睡眠 Sensitivity の変化

method	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
直前のみ (シーケンス)	89.01 ± 3.54%	90.27 ± 2.67%	91.06 ± 2.55%	87.79 ± 4.70%	88.30 ± 4.09
直前のみ (重みなし確率)	93.22 ± 1.80%	93.93 ± 1.67%	94.16 ± 1.59%	94.22 ± 1.60%	94.21 ± 1.60
直前のみ (重み付き確率)	93.22 ± 1.80%	93.22 ± 1.80%	93.44 ± 1.77%	93.74 ± 1.68%	93.88 ± 1.67
前後考慮 (シーケンス)	91.13 ± 2.67%	92.38 ± 2.02%	92.58 ± 2.51%	88.30 ± 5.29%	89.97 ± 2.94
前後考慮 (重みなし確率)	93.35 ± 1.66%	93.94 ± 1.49%	94.09 ± 1.53%	94.19 ± 1.58%	94.20 ± 1.55
前後考慮 (重み付き確率)	03.35 ± 1.66%	93.36 ± 1.65%	93.40 ± 1.61%	93.65 ± 1.58%	93.80 ± 1.52
時系列性の考慮なし	90.82 ± 1.82%				

表 3 k の違いによるレム睡眠 Sensitivity の変化

method	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
直前のみ (シーケンス)	90.90 ± 3.55%	91.02 ± 3.57%	90.89 ± 3.40%	91.36 ± 3.39%	91.39 ± 3.52
直前のみ (重みなし確率)	94.23 ± 3.59%	93.78 ± 3.62%	93.51 ± 3.63%	93.27 ± 3.62%	93.17 ± 3.60
直前のみ (重み付き確率)	94.23 ± 3.59%	94.24 ± 3.60%	94.14 ± 3.59%	93.92 ± 3.59%	93.83 ± 3.63
前後考慮 (シーケンス)	92.93 ± 3.17%	92.72 ± 3.00%	93.09 ± 2.95%	94.04 ± 2.83%	94.54 ± 2.12
前後考慮 (重みなし確率)	96.03 ± 2.97%	95.30 ± 2.89%	94.75 ± 3.14%	94.31 ± 3.23%	94.05 ± 3.36
前後考慮 (重み付き確率)	96.03 ± 2.97%	96.01 ± 2.95%	95.89 ± 2.77%	95.50 ± 2.80%	95.25 ± 2.91
時系列性の考慮なし	92.29 ± 3.82%				

表 4 k の違いによる覚醒 Sensitivity の変化

method	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
直前のみ (シーケンス)	95.61 ± 2.40%	95.91 ± 1.97%	95.86 ± 1.95%	95.77 ± 2.20%	96.06 ± 1.84
直前のみ (重みなし確率)	95.24 ± 2.21%	95.59 ± 1.87%	95.88 ± 1.73%	95.99 ± 1.77%	96.07 ± 1.75
直前のみ (重み付き確率)	95.24 ± 2.21%	95.30 ± 2.08%	95.34 ± 2.07%	95.44 ± 1.97%	95.47 ± 1.92
前後考慮 (シーケンス)	95.36 ± 2.21%	95.34 ± 2.06%	95.28 ± 2.16%	95.48 ± 2.58%	94.57 ± 4.01
前後考慮 (重みなし確率)	95.32 ± 2.33%	95.59 ± 1.96%	95.84 ± 1.73%	95.97 ± 1.67%	96.08 ± 1.67
前後考慮 (重み付き確率)	95.32 ± 2.33%	95.36 ± 2.17%	95.40 ± 2.16%	95.51 ± 2.05%	95.53 ± 1.99
時系列性の考慮なし	95.43 ± 2.13%				

最大になり一定の値をとるようになることがわかる。以上の結果より、全ての判定精度について最大値をとる範囲は、脳波の主成分数が 20 以上で筋電データの主成分数が 4 以上であることがわかる。そこで本手法では、脳波の主成分数は 20、筋電図の主成分数は 4 を使用することとする。

5.3.2 睡眠ステージの時系列性を考慮した特徴量の選定

時系列性の考慮による判定精度への影響と、4.3 節で述べた 6 種類の特徴量の違いによる判定精度への影響について予備実験を行った。ノンレム睡眠の Sensitivity について判定精度の比較を行ったものを表 2、レム睡眠の Sensitivity については表 3、覚醒の Sensitivity については表 4 に示す。表に示した値は、(平均値 ± 標準偏差%) の形式で示した実験結果である。

表 2、表 3 より、ノンレム睡眠の Sensitivity、レム睡眠の Sensitivity は、直前または前後のエポックにおける各睡眠ステージの占有確率を特徴量とすることで、時系列性を考慮しない場合と比較して高い判定精度を示すことが明らかになった。一方で、文字列シーケンスを特徴量とする場合は時系列性を考慮しない場合と比べて同等かそれ以下の判定精度を示していることがわかる。したがって、本研究では、時系列性を考慮した特徴量として各睡眠ステージの占有確率を用いる。

また表 2、表 3 より、ノンレム睡眠の Sensitivity は直前の時系列性のみを考慮する場合と前後の時系列性を考慮する場合と比較して有意差はないが、レム睡眠の Sensitivity は前後の時系列性を考慮する場合の方が約 2%ほど向上していることがわかる。表 4 より、覚醒の Sensitivity は、時系列性を考慮することによる判定精度への影響はなく、考慮するエポックの数についても判定精度への影響はないことがわかる。したがって、時系列性を考慮した特徴量を作成する場合、前後の時系列性を

考慮するほうが高い判定精度を示すことがわかる。

ノンレム睡眠の Sensitivity は、各睡眠ステージの占有確率を特徴量とする分類については、考慮するエポックの数が大きいほど判定精度が向上し、重みなし確率の場合は k が 4 以上のときに一定である。これに対して、レム睡眠の Sensitivity は、重みなし確率の場合は考慮するエポックの数が大きいほど判定精度が低下し、重み付き確率の場合は k が 3 以下のときは一定で 4 以上のときは考慮するエポックの数が大きいほど判定精度が低下している。ノンレム睡眠の Sensitivity が最大となる k の範囲とレム睡眠の Sensitivity が最大となる k の範囲には重複がないため、本研究では、 k の値の変化による Sensitivity への影響が大きいレム睡眠を優先して、 $k = 3$ を使用する。

以上の結果より、本研究では、前後 3 エポックにおける各睡眠ステージの占有確率を重み付き確率とした特徴量を用いて分類を行うこととする。

5.3.3 再判定に用いる閾値 θ_D の推定

コンピュータによる判定においてノンレム睡眠と判定されたエポックについての $D(e)$ のヒストグラムを図 13 に示す。ただし、赤色で示したヒストグラムはコンピュータによりノンレム睡眠と判定された全てのエポックについてのヒストグラムを示す。また、青色のヒストグラムはノンレム睡眠と判定されたエポックの中で誤判定されたエポックのヒストグラムを示す。同様に、コンピュータによる判定においてレム睡眠と判定されたエポックについての $D(e)$ のヒストグラムを図 14、覚醒と判定されたエポックについての $D(e)$ のヒストグラムを図 15 に示す。図 13 から図 15 においてビンの幅は 0.2 を用いた。

図 13、図 15 より、コンピュータによりノンレム睡眠または覚醒と判定されたエポックの中で誤判定されたエポックの多く

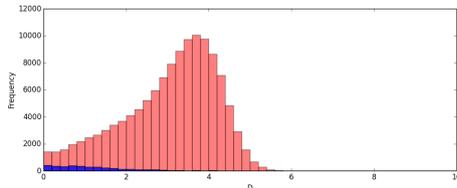


図 13 ノンレム睡眠と判定されたエポックの $D(e)$ のヒストグラム

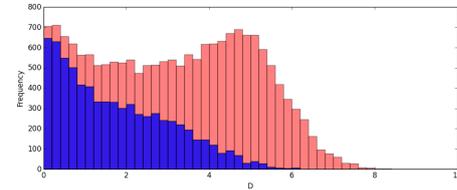


図 14 レム睡眠と判定されたエポックの $D(e)$ のヒストグラム

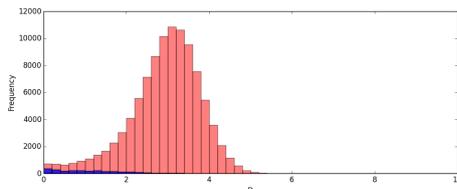


図 15 覚醒と判定されたエポックの $D(e)$ のヒストグラム

は D が 2 以下の非常に確実性が低いエポックであることがわかる。しかしながら、ノンレム睡眠または覚醒のエポック全体のごく一部であるため、これらのエポックに対する再判定による判定精度への影響は少ないと考えられる。これに対して図 14 より、コンピュータによりレム睡眠と判定されたエポックの中で誤判定されたエポックは全体の約 1/3 を占め、正しく判定されたエポックの分布は D が 3 から 6 付近に多いことがわかる。これに対して、誤判定されたエポックの D の分布は 0 から 4 付近に多いことがわかる。よって本研究では、確実性の低い判定を検出するための閾値として $\theta_D = 4$ を使用する。

5.3.4 再判定に用いる分類器の選定

5.3.3 節より、誤判定されたエポックの大半はコンピュータによりレム睡眠と判定されたエポックであることがわかった。また、コンピュータによりレム睡眠と判定されたエポックの中で誤判定されたエポックについて調査した結果、誤判定エポックの 90% 以上がノンレム睡眠のエポックであることがわかった。5.3.2 節の表 2 においてノンレム睡眠の Sensitivity が最も高い値を示した時系列特徴量は、前後 5 エポックにおける各睡眠ステージの占有確率を重みなし確率とした特徴量であることがわかる。そこで MASC では、コンピュータによりレム睡眠と判定されたエポックの中で $D(e)$ の値が閾値 $\theta_D = 4$ より小さいエポックについて、脳波の主成分 20 次元、筋電データの主成分 4 次元、および前後 5 エポックにおける各睡眠ステージの占有確率を重みなし確率とした特徴量 6 次元を用いて学習された SVM を用いて再判定を行う。

5.4 評価実験 1：ノイズが少ないマウス

5.4.1 比較手法

本実験では、提案手法 MASC, FASTER [6], exFASTER [7] の 3 つの手法について比較を行う。MASC の入力、脳波 (1Hz

～30Hz) の主成分 20 次元と筋電データ (30Hz～100Hz) の主成分 4 次元、判定対象の前後 3 エポックにおける各睡眠ステージの占有確率を重み付き確率とした特徴量 6 次元を与える。FASTER の入力、脳波 (1Hz～30Hz) の周波数スペクトル 30 次元と筋電データ (1Hz～30Hz) の周波数スペクトル 30 次元を合わせた 60 次元に対して主成分分析による特徴抽出を行うことで得られる主成分 4 次元を与える。exFASTER の入力、FASTER と同様の脳波と筋電データの周波数スペクトルの主成分 4 次元と FASTER による睡眠ステージ判定結果を与える。

5.4.2 実験結果と考察

ノイズが少ないマウスに対して各手法を適用した結果を表 5 に示す。表に示した値は、(平均値 ± 標準偏差%) の形式で示した実験結果である。

表 5 より、FASTER や exFASTER と比較して、レム睡眠の Specificity 以外の評価指標は判定精度が向上していることがわかる。特に、ノンレム睡眠の Sensitivity 以外の評価指標については、標準偏差の値から明らかに有意差があることがわかる。また、ノンレム睡眠の Specificity, レム睡眠の Specificity, 覚醒の Sensitivity と Specificity については、専門家が要求する精度である 95% を超えており、ノンレム睡眠の Sensitivity, レム睡眠の Sensitivity, Accuracy についても、95% に非常に近い値を示していることがわかる。

誤判定されたエポックについて調査した結果、誤判定の大半が睡眠ステージが遷移する前後のエポックであることがわかった。睡眠ステージが遷移する前後のエポックの判定は、専門家による判定であっても専門家によって意見が別れることが多く、異なる専門家により同じマウスの睡眠ステージ判定を行った場合の睡眠ステージ判定結果の一致率は約 95% と言われていることより、本手法によるノイズが少ないマウスを対象とした自動睡眠ステージ判定については、非常に専門家に近い判定精度で睡眠ステージ判定を行うことができていると考えられる。

5.5 評価実験 2：ノイズが多いマウス

5.5.1 データセット

本実験で用いる学習データは、5.1 節で説明したノイズが少ないマウスを使用し、評価に用いるデータはノイズが多いマウス 7 匹の脳波および筋電データを人手でラベル付けしたものである。ノイズが多いマウスの脳波および筋電データのサンプリング周波数は 250Hz, エポックサイズは 20 秒である。エポック数は 2 日分の 8,640 エポック、計 60,480 エポックである。

5.5.2 実験結果と考察

ノイズが多いマウスに各手法を適用した結果を表 6 に示す。

表 6 より、FASTER や exFASTER と比較して、ノンレム睡眠の Sensitivity, レム睡眠の Specificity 以外の評価指標は判定精度が向上していることがわかる。また、ノンレム睡眠の Sensitivity, レム睡眠の Specificity についても明らかに有意差がないことがわかる。特にノンレム睡眠の Specificity と覚醒の Sensitivity については値が大きく向上しており、標準偏差の値が大きく減少している。これは、FASTER および exFASTER を用いた睡眠ステージ判定では、ノンレム睡眠と覚醒の間で睡眠ステージが遷移する場合について、遷移する位置が正しく検

表5 ノイズが少ないマウスに対する睡眠ステージ判定精度の比較

method	ノンレム睡眠		レム睡眠		覚醒		Accuracy
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
MASC	94.37 ± 1.44%	96.63 ± 1.30%	94.74 ± 2.86%	97.37 ± 0.94%	95.42 ± 2.25%	98.09 ± 0.93%	94.76 ± 1.02%
FASTER	89.49 ± 0.41%	94.02 ± 0.73%	78.36 ± 1.00%	98.33 ± 0.06%	94.61 ± 0.72%	92.31 ± 0.36%	91.09 ± 0.21%
exFASTER	93.79 ± 0.21%	93.87 ± 0.54%	79.72 ± 0.59%	99.08 ± 0.05%	94.79 ± 0.59%	95.26 ± 0.20%	93.41 ± 0.21%

表6 ノイズが多いマウスに対する睡眠ステージ判定精度の比較

method	ノンレム睡眠		レム睡眠		覚醒		Accuracy
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity	
MASC	94.19 ± 5.47%	93.71 ± 2.07%	92.95 ± 7.11%	96.01 ± 2.60%	91.10 ± 4.98%	99.50 ± 0.34%	92.40 ± 3.31%
FASTER	91.26 ± 6.01%	79.04 ± 25.86%	88.95 ± 6.65%	97.24 ± 3.03%	75.55 ± 28.85%	94.79 ± 4.01%	82.32 ± 13.39%
exFASTER	95.02 ± 2.49%	78.65 ± 25.44%	88.08 ± 6.08%	97.89 ± 3.02%	75.40 ± 28.64%	97.14 ± 1.71%	83.93 ± 14.69%

出できず遷移付近の覚醒エポックまでノンレム睡眠と誤判定していることが原因であると考えられる。本手法では時系列性を考慮した特徴量を作成することで、ノンレム睡眠とレム睡眠の遷移を正しく検出でき、高い判定精度を示したと考えられる。

一方で、ノイズが少ないマウスの場合と比較して、レム睡眠の Sensitivity と覚醒の Sensitivity については値が減少しており、専門家が要求する判定精度である 95% をまだ満たしていないことがわかる。また、覚醒の Specificity 以外の評価指標において標準偏差が大きく増加しているため、マウスのごとに判定精度のばらつきが非常に大きく、安定していないことがわかる。

6. まとめと今後の課題

本研究では、マウスの睡眠ステージを自動判定するための手法を提案した。本研究では、既存手法の多くでは、睡眠ステージ判定において時系列性の考慮をしていないという問題および各睡眠ステージの特徴を正しく抽出することができていないという問題がある点に着眼した。また、分類器を用いた分類ではコンピュータはそれぞれの睡眠ステージへ分類される確率を求め、一番高い確率の睡眠ステージへ分類する。しかし、複数の睡眠ステージの特徴を持つデータの場合、それぞれの睡眠ステージへ分類される確率の差が小さいため、機械的に一番確率が高い睡眠ステージへ分類されるような確実性の低い判定が含まれることがあるという問題がある。そこで本研究では以下の3つのアプローチをとった手法 MASC を提案した：(1) 各睡眠ステージの特徴をよく表現できる特徴量の推定、(2) 分類器に与えるパラメータとして睡眠ステージの時系列性を考慮した特徴量の作成、(3) アプローチ (1) および (2) を用いて分類を行った結果について、分類器による判定の確実性が低い判定に対する再判定を行う。提案手法 MASC についてノイズが少ないマウスを用いて実験を行った結果、既存手法と比較して非常に判定精度が向上し、専門家が要求する判定精度 95% 以上という精度に非常に近い値を実現することができた。また、ノイズが多いマウスを用いて実験を行った結果、平均で 92.40% という高い判定精度を実現することができた。

今後は、脳波および筋電データに混入するノイズを除去することにより、ノイズが多いマウスに対しても高い判定精度を実現できる手法の提案、および睡眠障害のような健康体のマウスと異なる睡眠ステージ遷移の傾向があるマウスに対しても高い判定精度を実現できる手法の提案を予定である。[12, 14]

謝 辞

本研究の一部は、科学研究費補助金・新学術領域研究“宇宙に生きる”「超ストレス環境・宇宙を見据えた新規睡眠覚醒制御手法の開発」および文部科学省・地域イノベーション・エコシステム形成プログラム「つくばイノベーション・エコシステムの構築」による助成を受けたものである。

文 献

- [1] 横山 幸生, 島田 尊正, 竹村 淳, 椎名 毅, 斉藤 陽一, “睡眠脳波の主成分分析とニューラルネットワークによる特徴波検出”, 電子情報通信学会論文誌 A (1993) Vol.J76-A No.8 pp.1050-1058
- [2] 花岡 正明, 小林 正樹, 山崎 晴明, “波形認識法と決定木学習に基づく睡眠ステージの自動判定”, 電子情報通信学会論文誌 D (2001) Vol.J84-D2 No.12 pp.2672-2683
- [3] Jurij Brankačka, Valeriy I. Kukushka, Alexei L. Vyssotski, Andreas Draguhn, “EEG gamma frequency and sleep-wake scoring in mice: Comparing two types of supervised classifiers”, BRAIN RESEARCH 1322 (2010) 59-71
- [4] Michael J Rempe, William C Clegern, Jonathan P Wisor, “An automated sleep-state classification algorithm for quantifying sleep timing and sleep-dependent dynamics of electroencephalographic and cerebral metabolic parameters”, Nature and Science of Sleep 2015:7 85-99
- [5] Sayaka KOHTOH, Yujiro TAGUCHI, Naomi MATSUMOTO, Masashi WADA, Zhi-Li HUANG, Yoshihiro URADE, “Algorithm for sleep scoring in experimental animals based on fast Fourier transform power spectrum analysis of the electroencephalogram”, Sleep and Biological Rhythms 2008, 6, 163-171
- [6] Genshiro A. Sunagawa, Hiroyoshi Sei, Shigeki Shimba, Yoshihiro Urade, Hiroki R. Ueda, “FASTER: an unsupervised fully automated sleep staging method for mice”, Genes to Cells (2013) 18, pp.502-518
- [7] 鈴木悠太, 佐藤牧人, 塩川浩昭, 柳沢正史, 北川博之, “脳波/筋電図を利用したマウスの睡眠ステージ判定”, 第8回データ工学と情報マネジメントに関するフォーラム (DEIM2016), G4-3
- [8] 音成 龍司, 辻 貞俊, “よくわかる脳波判読”, 金原出版 (2014)
- [9] 市川 忠彦, “新版 脳波の旅への誘い”, 星和書店 (2006)
- [10] T. Jolliffe, “Principle Component Analysis”, Springer Series in Statistics (2002)
- [11] A. Azzalini, N. Torelli, “Clustering via nonparametric density estimation”, Stat Comput 17, pp.71-80 (2007)
- [12] 佐川 雅彦, 貴家 仁志, “高速フーリエ変換とその応用”, 昭晃堂 (1993)
- [13] Walter Zucchini, “Hidden Markov Models for Time Series: An Introduction Using R”, Chapman and Hall/CRC (2009)
- [14] C. Cortes, V. Vapnik, “Support-Vector Networks”, Machine Learning, Vol. 20, pp. 273-297 (1995)
- [15] Christopher M. Bishop, “Pattern Recognition and Machine Learning”, Springer (2006)