

大規模マウス睡眠ステージ解析のための深層学習モデルの開発

山部 剛士[†] 堀江 和正^{††} 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-IHS) 〒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: [†]{yamabe,horie}@kde.cs.tsukuba.ac.jp, ^{††}{shiokawa,kitagawa}@cs.tsukuba.ac.jp,

^{†††}yanagisawa.masa.fu@u.tsukuba.ac.jp

あらまし 睡眠解析・研究における重要な検査に、マウスの睡眠ステージ（睡眠状態）判定がある。現在、睡眠ステージ判定は、主に熟練者による目視で行われており、多くの時間と労力を必要としている。この問題を解決すべく、これまでも自動睡眠ステージ判定手法が提案されているが、専門家の代替として十分な精度は満たされていない。また、過去の睡眠研究と比べて小規模なデータによる検証されておらず、マウスの個体差やノイズに対する頑健性が十分考慮されていない点など、睡眠研究で実用できるか不明瞭であるといった問題があった。本研究では、波形検出等で優れた成果を示した深層学習（畳み込みニューラルネット、Long short-term memory）を用いることで、従来手法と比べて高い精度を持った睡眠ステージ判定手法を開発した。また、マウス 4200 匹からなる大規模なデータベースを用いることで、本手法がマウス個体差やノイズに対する高い頑健性を有することを示した。本手法は、一般的な睡眠研究で用いられるような幅広いマウスの睡眠ステージを高精度で判定することができることから、熟練者の代わりとして睡眠研究でも利用できると考えられる。

キーワード 脳波, 筋電位, 睡眠ステージ判定, マウス, 時系列データ, 深層学習

1 はじめに

睡眠に関する基礎研究を行う上で、マウスの睡眠のステージ（ノンレム・レム睡眠といった睡眠状態）を把握することは非常に重要である。これら睡眠ステージを脳波や筋電位といった生体信号から判定することを睡眠ステージ判定と呼ぶ。この判定は例えば、遺伝子操作や薬物投与による睡眠状態への影響を調査する場合などに必要とされる。

現在、マウスの睡眠ステージ判定は、主に熟練した専門家が脳波および筋電図を目視で確認、判定を行っている。しかし、この判定には多くの時間と労力が必要であり、睡眠研究の大規模化の妨げになっている。これらの問題を解決すべく、自動かつ高精度でマウスの睡眠ステージ判定を行うことができる手法が求められている。これまでに様々な提案手法が提案されているが、専門家の代替として十分な判定制度である 95% を達成していない。また、小規模なデータセットでの検証しか行われていないなど、実用化には課題が残っている [2] [3]。そこで本研究では波形検出等で優れた成果を示した深層学習（畳み込みニューラルネット、Long short-term memory）により、専門家の判定と高い一致率で判定可能な睡眠ステージ判定手法 MC-SleepNet を開発した。

また、本手法の判定精度を 4200 匹分の大規模なデータセットを用いて検証した。これによって、既存手法で用いられた数十匹程度のデータセットでは十分に検証できない、マウスの個体差やノイズの有無によるデータの分散に対する頑健性を評価し

た。本稿の構成は以下のとおりである。まず 2 節ではマウスの睡眠ステージ判定やニューラルネットワークなどの基本事項について説明する。3 節で既存の睡眠ステージ判定手法など関連研究について述べ、4 節で提案手法について説明する。5 節で実際のマウスのデータを用いた提案手法の評価実験について述べる。最後に 6 節でまとめと今後の課題について述べる。

2 前提知識

2.1 マウスの睡眠ステージとその判定

マウスの睡眠状態は覚醒、ノンレム睡眠、レム睡眠の 3 つの睡眠ステージに分けることができる。それぞれの睡眠ステージごとに脳波、筋電位には特有の性質が現れる。本節ではまず、各睡眠ステージの特性について説明する (表 1)。

(1) ノンレム睡眠

ノンレム睡眠は入眠後に現れる睡眠ステージである。このノンレム睡眠が睡眠時の状態の約 8 割を占めている。ノンレム睡眠時は脳が深い眠りについており、その脳波には 1Hz から 6Hz のデルタ波と呼ばれる比較的低周波な成分を多く含み、振幅は大きくなる。体を支える筋肉は働いており、寝返りをうつこともあるがその筋電データの振幅は小さい。

(2) レム睡眠

レム睡眠は基本的にノンレム睡眠後に現れる睡眠ステージである。レム睡眠時は脳が浅い眠りについておりため覚醒時と似た脳波を示し、7Hz から 11Hz のシータ波と呼ばれる比較的高周波な成分を多く含み、ノンレム睡眠と比較して振幅は小さい。

体を支える筋肉は弛緩しており寝返りをうつことは無いが、眼球が活発に動いておりこの眼球運動は急速眼球運動 (Rapid Eye Movement) と呼ばれ、この頭文字をとって REM 睡眠と呼ばれる。筋電データはノンレム睡眠同様振幅が小さくなる。また、レム睡眠の現れる時間は数十秒から数分程度と非常に短く睡眠中に一定の周期で現れることが知られている。

(3) 覚醒

覚醒は脳・筋肉ともに活発に働いている状態であり、レム睡眠と似た脳波パターンを示す一方で筋電データは睡眠時と比較して振幅が大きくなるという特徴がある。

各睡眠ステージの脳波、筋電位に現れるこれらの特性をもとに睡眠ステージ判定を行う。脳波/筋電位は図1のようにエポックと呼ばれる一定区間に区切って分割し、脳波/筋電データの周波数特性や振幅と各睡眠ステージの特性との関係性を考慮してエポック毎に判定を行う。睡眠ステージ判定において、専門家間の判定の一致率は95%程度とされており[12]、専門家の代替となるには、この程度の一致率が求められている。

表1 睡眠ステージの特性

	脳波のピーク周波数	筋電位の大きさ
覚醒	シータ波(7~11Hz)	大
ノンレム睡眠	デルタ波(1~6Hz)	小
レム睡眠	シータ波(7~11Hz)	小

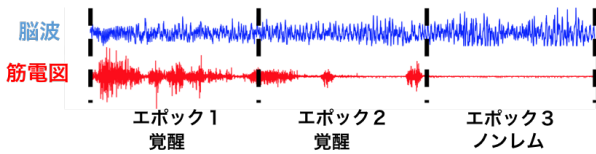


図1 エポックと睡眠ステージ判定の例

2.2 深層学習

深層学習とは、非常に多くのノードで構成され、高い表現力を持ったニューラルネットワークを用いた機械学習手法である。ニューラルネットワークとは学習能力を持ち回帰や分類を行うことができる数理モデルであり、現在までに様々なモデルが提案されている。本節では提案手法で用いる、Convolutional Neural Network(CNN)とLong Short-term Memory(LSTM)の2つのモデルについて述べる。

2.2.1 Convolutional Neural Network(CNN)

CNNは、画像や時系列信号の特徴を抽出するためのモデルである。一般的には最適化可能なフィルタを持つConvolution層と、次元圧縮や位置ずれに対する頑健性の向上を図るPooling層からなる。(図2)。

Convolution 層

Convolution層では入力データに対してフィルタによる畳込みを行い特徴マップを生成する。特徴マップを生成する式を以下に示す。

$$y_i = \sum_{s=0}^{n-1} w_s^k x_{i+s} + b^k$$

$\mathbf{x}$ と $\mathbf{y}$ はそれぞれ入力と出力を表している。 $\mathbf{w}^k$ と $\mathbf{b}^k$ は k 番目のフィルタの重みとバイアスである。 n はフィルタの幅である。このフィルタの幅によって抽出できる特徴が変化する。幅が広いほど、フィルタに含まれる信号の波長が長くなるため、低周波成分を中心に広い範囲の周波数成分を抽出することができる。また、幅が狭いほど信号に対する時間分解能が向上し、波形や振幅などの時間領域の特徴を抽出することができる。

Pooling 層

Pooling層ではConvolution層で生成された特徴マップの圧縮を行う。例えば、MaxPoolingと呼ばれる操作では、入力された時系列を一定の時間間隔で分割し、その分割範囲内の最大値をとることで、データ点数を減らしている。また、この操作を行うことで、パラメタ数の削減や入力データの位置的なずれ(脳波の位相のずれなど)にロバストであるといった効果が得られる。

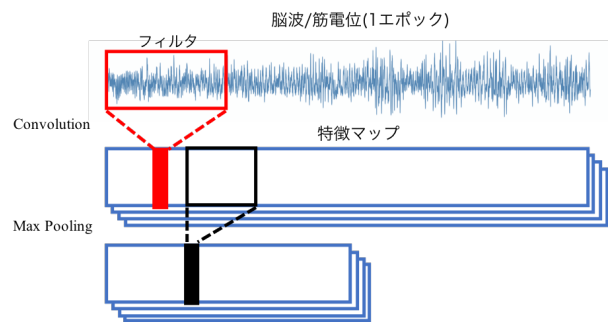


図2 脳波/筋電データに対するCNNによる特徴抽出

2.2.2 Long Short-term Memory(LSTM)

LSTMはLSTMブロックと呼ばれるノードを用いた再帰型ニューラルネットワークの一種であり、時系列データの解析に多く用いられている。LSTMブロックはノードの内部状態を保持するメモリセルと、その更新を制御する3種類のゲートからなる。これらのゲートは内部状態の更新量や忘却量、出力への反映の度合いを調整し、従来の再起型ニューラルネットワークよりも高い表現力と、より長い時系列情報の保持を可能にしている。時間 t での忘却ゲート、入力ゲート、出力ゲートを f_t, i_t, o_t とし、それぞれの計算を以下に示す。

$$f_t = \text{sigmoid}(\mathbf{W}_f \mathbf{a}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f)$$

$$i_t = \text{sigmoid}(\mathbf{W}_i \mathbf{a}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i)$$

$$o_t = \text{sigmoid}(\mathbf{W}_o \mathbf{a}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o)$$

$$g_t = \tanh(\mathbf{W}_g \mathbf{a}_t + \mathbf{U}_g \mathbf{h}_{t-1} + \mathbf{b}_g)$$

$$\mathbf{c}_t = i_t \otimes g_t + f_t \otimes \mathbf{c}_{t-1}$$

$$\mathbf{h}_t = o_t \otimes \tanh(\mathbf{c}_t)$$

$$\text{sigmoid}(\mathbf{x}) = \frac{1}{1 + e^{-\mathbf{x}}}, \tanh(\mathbf{x}) = \frac{e^{\mathbf{x}} - e^{-\mathbf{x}}}{e^{\mathbf{x}} + e^{-\mathbf{x}}}$$

a_t, h_t, c_t, g_t はそれぞれ時間 t での入力, 出力, 内部状態, 内部状態に加える候補値を示している. W_x, U_x, b_x はゲート X の重みとバイアスであり, $\otimes$ はベクトルの要素積を表す. このように各ゲートは, *sigmoid* 関数によって 0 1 の値をとることで, 更新量や忘却量, 出力への反映度合いを制御している. 例えば, 忘却ゲートは出力が 0 に近づくほど, 過去の内部状態を忘却させることができる.

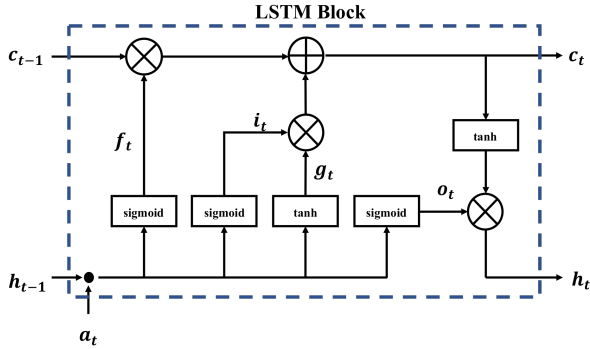


図3 LSTM ブロック

3 関連研究

本節では既存のマウスの睡眠ステージ判定手法である FASTER, MASC, LSTM モデルについて説明する. これらの手法は全て, 脳波/筋電位のデータを用いて, マウスの睡眠ステージ判定を行っている. これらの既存手法では, 数匹から数十匹の小規模なデータセットによる検証しか行われておらず, より多くのマウスを対象とする睡眠研究において, 実用可能であるかが不明瞭であるといった課題がある. また, 判定精度の点でも課題がある. 一般に異なる専門家が同一のデータを判定した時の検者間一致率は 95% 程度であり, 判定手法にも同程度の一致率が求められる. しかし, 既存手法で十分な一致率を達成した手法は未だない.

3.1 FASTER

FASTER(Sunagawa ら, 2013) [3] は密度ベースクラスタリング [4] に基づく睡眠ステージ判定手法である. エポックに分割された脳波/筋電位に対し, FFT と主成分分析によって周波数特性を特徴量として利用する. 各エポックは特徴量に基づき, ノンパラメトリック密度推定クラスタリングを用いてクラスタリングされる. それぞれのクラスタに対して, 最も特性に近い睡眠ステージをクラスタに属するエポックの睡眠ステージとする. このように, FASTER は周波数特性に注目して判定を行うが, 睡眠ステージの状態遷移を考慮していない. FASTER は覚醒, ノンレムについては高い精度で判定可能であるが, レムについては 80% 程度が限界である.

3.2 MASC

MASC(鈴木ら, 2017) [2] は各睡眠ステージの特徴をよく表現できる特徴量の推定, 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るマウスの睡眠ステージ判定手法である. まずエポック単位で FFT によって求めた脳波/筋電位の周波数成分に対して主成分分析による特徴量の推定を行う. それらの特徴量に加えて判定対象となるエポックの前後のエポックにおける睡眠ステージの占有確率を時系列的な特徴量として SVM [9] による判定を行う. さらに確実度の低い判定結果については再判定を行うことでその判定を修正することができる. MASC ではノイズの少ないデータについて 95% に近い一致率を達成しているが, ノイズを含むデータに対して判定精度が低下する.

3.3 LSTM モデル

これは我々が先行研究で脳波/筋電位の周波数成分に基づくディープラーニングによるマウスの睡眠ステージ判定手法である (2017) [7]. 睡眠ステージが時系列的な遷移をしていることに着目し, 時系列性を考慮できるように RNN モデルの 1 種である LSTM を用いて判定を行った. 事前に FFT(fast Fourier transform) [8] によって得た脳波/筋電データの周波数スペクトルを特徴量として用いている. この手法では覚醒・ノンレムの判定精度は 95% を達成したが, レムの判定精度が約 82% と実用可能な段階に達していない.

4 提案手法

提案手法 MC-SleepNet(図 4) は, 信号の前処理と, 特徴抽出部と識別部から構成される. それぞれについて詳細を説明する.

4.1 データの前処理

2.1 節で説明したように, マウスの筋電位はその振幅の大きさが重要である. そのため, 特徴抽出部に入力する筋電位のデータについて前処理として移動二乗平方根平均を求めた. 移動二乗平均平方根によってより信号の強さが分かりやすい形に変換される. 筋電位の信号の強さは覚醒時と睡眠時で大きく異なり, この前処理によってこれらの状態をモデルがより正確に判別できると考えられる.

4.2 特徴抽出部

特徴抽出部では CNN によって, エポック単位で脳波/筋電データからの特徴抽出を行う.

まず脳波については, 周波数成分および振幅がマウスの睡眠ステージを分ける重要な特徴であるため, 特徴抽出はフィルタサイズの異なる 2 種類の CNN によって行われる. フィルタサイズの小さな CNN では振幅などの波形的な特徴を行い, フィルタサイズの大きな CNN で低周波の波も含むような周波数成分の抽出を行う.

次に, 筋電位のデータは, 前処理として移動二乗平方根平均によって, 高周波成分が少ない波形へと変換が行われている. そのため, フィルタサイズの小さな CNN は用いず, フィルタサイズの大きい CNN のみを使用した.

4.3 識別部

識別部では双方向型 LSTM によって時系列的な特徴から睡眠ステージの判定を行う。入力に連続する 25 エポックに対して特徴抽出部から得られた特徴である。双方向型 LSTM によって、睡眠ステージの時系列的な遷移規則など、前後のエポックを考慮した判定が可能である。また、特徴抽出部から得られた特徴はそのままエポックの特徴量として利用するために全結合層にも入力される。双方向型 LSTM と全結合層の出力の要素和を Softmax 関数の入力とし各睡眠ステージへの判定は Softmax 関数によって得られた確率が最大であるステージに分類する。LSTM の各層の間には Dropout 層 [10] を挟むことで過学習を防止している。Dropout は前の層からの出力をランダムに一定数落とすことでモデルの表現力を落とし、過学習の状態を抑制する。また、アンサンブル学習のような効果も期待でき精度向上にもつながる。

4.4 事前学習

MC-SleepNet では、LSTM の学習に時間がかかるため、まず特徴抽出部のみ学習を行う。事前学習を行うときはモデルの識別部を Softmax 層と入れ替え、特徴抽出部のみで判定させている。これによってある程度良い初期値でモデル全体の学習を始めることができる。また、事前学習によって学習データに含まれる睡眠ステージの偏りの影響を軽減することができる。学習データの偏りは睡眠中にレム睡眠が現れる時間が非常に少ないために発生する。そこで、事前学習ではオーバーサンプリングやアンダーサンプリングを行い学習データに含まれるステージの数を最も多く出現するステージに合わせて学習を行っている。

5 実験

本研究では、小規模なデータセットを用いた既存手法との精度比較や、大規模データセットを用いた頑健性の評価を行っている。まずノイズの少ない小規模なデータセットによる、既存手法との比較実験について述べる。次に、大規模なデータセットによる、マウスの個体差などのデータのばらつきやノイズに対する頑健性を評価する実験について述べる。まず、データセットや評価指標などの実験の基本事項について説明する。

5.1 データセット

本実験では規模と性質の異なる 2 種類のデータセット (表 5.1) を用いた。まず 1 つは、ノイズの少ない小規模なデータセットである。このデータセットは従来手法の評価に用いられており、ノイズが少ないと思われるデータを収集したものである。マウスの脳波・筋電位のデータとして、筑波大学国際統合睡眠医科学研究機構 (IIS) で専門家が判定済みである 14 匹のマウスのデータを用いる。これらのマウスは薬物投与や遺伝子組み換えなどが行われていない野生型 (Wilde Type) のマウスである。2014 年 11 月 8 日から 2014 年 11 月 12 日の 4 日間 (17200 エポック) に渡って測定されたデータで、サンプリング周波数は 250Hz である。20 秒のエポックで分割され専門家に

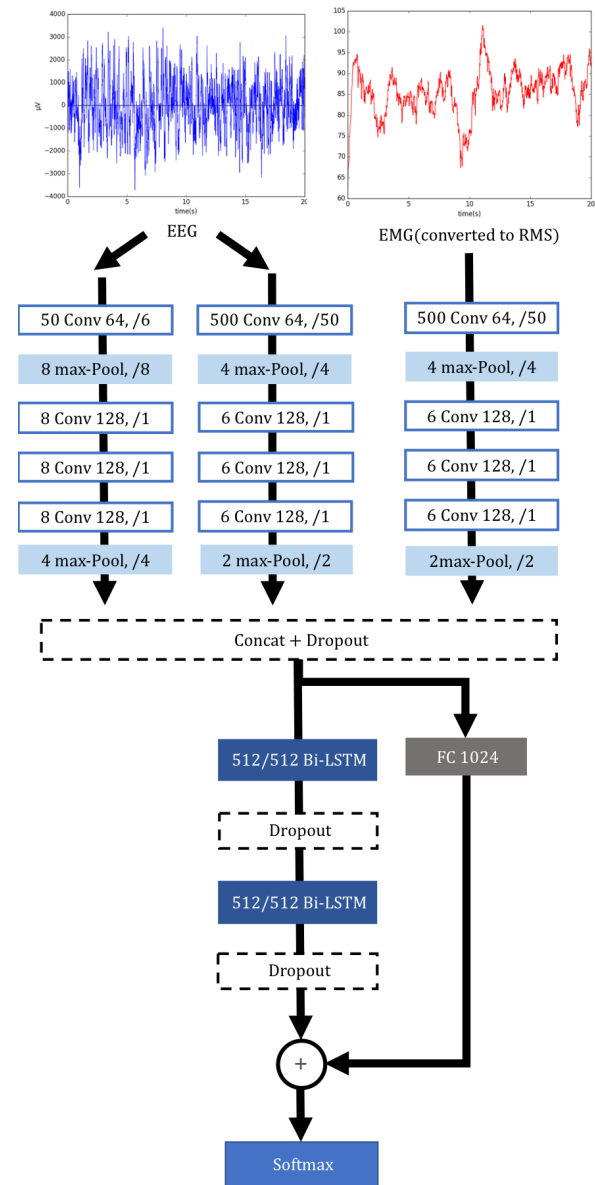


図 4 MC-SleepNet の構成

によって各エポックに覚醒/レム/ノンレムのラベルが付けられている。次に、大規模なデータセットである。このデータセットには 4200 匹分のマウスの脳波・筋電位のデータが含まれている。野生型のマウスではあるが、含まれているノイズの大小を問わず収集したデータであり、実際の睡眠研究とほぼ同等のバリエーションを持ったデータセットと考えられる。それぞれのデータは約 2 日間 (約 8500 エポック) に渡って測定され、小規模なデータセットと同様にラベルが付けられている。

表 2 データセットの概要		
マウスの数	エポック数	ノイズの有無
14	241920	無
4200	35700000	有

5.2 評価指標

本実験では、手法の精度評価に際し、Sensitivity と Speci-

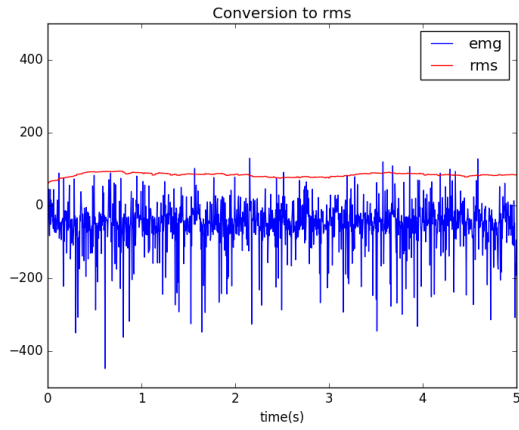


図 5 移動二乗平方根平均の前処理

ficity, Precision, Accuracy, Cohen の k 係数 [11] を使用した. Sensitivity と Specificity, Precision は各睡眠ステージに対して, 以下の式で計算する.

$$\text{Sensitivity} = \frac{TP_X}{TP_X + FN_X}$$

$$\text{Specificity} = \frac{TN_X}{TN_X + FP_X}$$

$$\text{Precision} = \frac{TP_X}{TP_X + FP_X}$$

TP_X , FN_X , TF_X , FP_X はそれぞれ, 睡眠ステージ X に対する判定が, true positive, false negative, false positive, true negative であったエポック数を表している. Sensitivity は専門家が X であると判定したエポックの中で判定器も X であると判定した割合, Specificity は専門家が X でないと判定したエポックの中で判定器も X でないと判定した割合, Precision は判定器が X であると判定したエポックの中で専門家も X であると判定した割合を示す.

Accuracy は専門家と判定器の判定が一致した割合を示し, 以下の式で計算する.

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

$$FN_{ALL} = FN_{NREM} + FN_{REM} + FN_{WAKE}$$

$$TP_{ALL} = TP_{NREM} + TP_{REM} + TP_{WAKE}$$

Cohen の k 係数は専門家と判定モデル間の判定結果の信頼性を示し, 一般に 0.8 以上であれば, 2 者の判定はほぼ完全に一致する結果であると言われている.

5.3 評価実験：ノイズの少ない小規模データに対する判定精度の評価

既存手法との精度比較を目的とし, ノイズの少ないデータセットを用いた評価実験を行った. 本実験では, 比較手法として既存手法である MASC, LSTM モデル, FASTER に加え, RandomForest を使用した. ここでは RandomForest の設定を, 決定木の個数を 20, 木の深さの最大値を 10 とした.

RandomForest に与える特徴量は, FASTER や MASC で検討された特徴量をもとに, それぞれのエポックについて脳波の振幅の総和, 脳波の周波数成分の総和, 脳波のデルタ波 (1 6Hz) とシータ波 (7 11Hz) それぞれの周波数成分の総和, 筋電位の周波数成分の総和, 筋電位の振幅の総和を特徴量として判定を行った. それぞれの手法について, 12 匹分のデータを学習データ, 2 匹分のデータをテストデータとした 7 分割交差検定による判定精度を比較する. ただし, FASTER の判定精度は MASC で同一のデータセットを判定した時の結果を引用する. それぞれの手法の判定精度について以下の表 3 に示す.

実験結果から, 覚醒, ノンレムは既存手法と比較しても同等以上の精度で判定可能である. これらのステージの Sensitivity は既存手法と比較して最も高く, これらのステージの他の指標についても 95% 以上の高い精度で判定できている. 一方でレムの Sensitivity と Precision はそれぞれ 92.1% と 73.0% であり, 95% には達していない. しかし, 既存手法と比較すると FASTER, LSTM モデルの Sensitivity が 78.4% と 82.9% であり, 提案手法の判定精度が上回っている. MASC と比較すると REM の Sensitivity が約 2% 低いが, Precision が約 7% 高い. また Accuracy は, 提案手法が最も高く, 95% を達成しており, k 係数も 0.92 で最も高い. これらの結果から, 提案手法は既存手法と同等以上の精度で判定可能であることを示した. また, LSTM モデルは FFT による周波数成分を特徴量としているのに対し, 提案手法は特徴量を CNN によって抽出している. どちらの手法も LSTM による分類を行っているが, 提案手法のほうが良い精度を示していることから, CNN が周波数成分に準じた特徴を抽出できていると考えられる.

5.4 評価実験：大規模なデータセットに対する判定精度の評価

この実験では大規模なデータセットを用いて判定精度の評価を行った. 実験を通して, 提案手法がマウスの個体差やノイズに頑健であるか確認する. 比較手法は, 小規模データセットに対して良い精度で判定を行った MASC と RandomForest である. 3360 匹分のデータを学習データ, 840 匹分のデータをテストデータとした 5 分割交差検定による判定精度を比較する. ただし, MASC については SVM を用いており, メモリを超えるサイズのデータの学習が困難である. そこで, MASC の評価用に 1340 匹分のデータを抽出し, 500 匹分のデータを学習データ, 840 匹分のデータをテストデータとして評価を行った. ここでは交差検定は行っていない.

4200 匹分のデータと 1340 匹分のデータに対する各手法の判定精度を以下の表 4 に示す.

まず, 1340 匹分のデータについては RandomForest が最も優れた精度を示した. 提案手法と MASC は, Accuracy がそれぞれ, 80.5% と 67.6% であり, RandomForest と比較すると, 精度は非常に低い.

一方, 4300 匹とサンプルを増やしたケースでは, 提案手法の精度が大きく向上し, RandomForest より優れた精度を示した. RandomForest については, データのバリエーションに対して非常に頑健であり, 比較的サンプルが少ない場合であっても適

表 3 ノイズの少ないデータに対する判定精度

method	覚醒			ノンレム			レム			Accuracy	k係数
	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision		
MC-SleepNet	96.4%	98.0%	97.3%	95.5%	96.7%	97.0%	92.1%	98.3%	73.0%	95.7%	0.92
FASTER	89.5%	94.0%	---	94.2%	93.7%	---	78.4%	98.3%	---	91.1%	---
MASC	95.6%	98.1%	97.3%	94.7%	96.6%	96.8%	94.4%	97.6%	66.3%	95.0%	0.91
LSTMモデル	96.2%	96.3%	94.9%	95.1%	95.2%	95.8%	82.9%	99.3%	85.5%	94.9%	0.90
RandomForest	96.0%	94.5%	93.4%	94.3%	95.1%	95.7%	74.7%	99.6%	89.5%	94.0%	0.89

表 4 大規模データセットに対する判定精度

method	覚醒			ノンレム			レム			Accuracy	k係数
	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision		
Dataset (1340 mice)											
MC-SleepNet	99.5%	64.2%	75.8%	60.7%	99.5%	98.6%	67.1%	98.6%	72.4%	80.5%	0.63
MASC	98.0%	45.9%	65.3%	36.7%	98.7%	95.8%	48.8%	95.6%	42.7%	67.6%	0.40
RandomForest	97.6%	93.5%	93.7%	92.4%	96.4%	95.7%	74.0%	99.3%	84.7%	94.1%	0.89
Dataset (4200 mice)											
MC-SleepNet	97.4%	98.2%	98.1%	97.5%	96.0%	95.5%	78.6%	99.6%	91.1%	96.6%	0.94
RandomForest	95.8%	95.1%	95.0%	94.0%	94.9%	94.4%	77.0%	99.2%	86.2%	94.1%	0.89

表 5 ノイズが多く含まれるデータに対する判定精度

Dataset	覚醒			ノンレム			レム			Accuracy	k係数
	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision	Sensitivity	Specificity	Precision		
ノイズが多いデータ	94.8%	98.6%	98.5%	97.5%	93.2%	92.5%	75.7%	99.3%	86.8%	95.0%	0.91
大規模データ	97.4%	98.2%	98.1%	97.5%	96.0%	95.5%	78.6%	99.6%	91.1%	96.6%	0.94

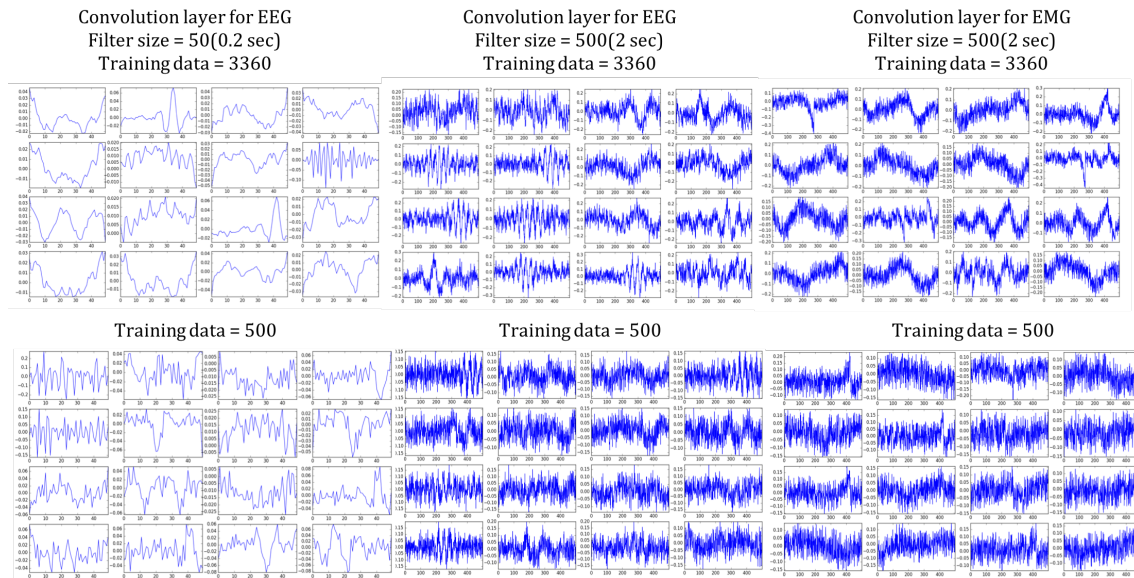


図 6 CNN フィルタの可視化

切な境界線を形成できている。ただし、95%の精度を達成できていない。また、RandomForest ではサンプル数を増やした場合であっても精度は改善されないことから、本手法を実用化するのは難しいと考えられる。

一方、深層学習を用いた提案手法では、サンプル数が少ない場合については最適な特徴抽出が学習できず、精度が低かったものの、サンプル数の追加に伴い精度が大きく改善した。95%という目安も達成できており、実用レベルに到達していると考えられる。サンプル数の増加によって、特徴抽出に用いているCNNのフィルタがどのように最適化されたのかを図6で示す。この図は、フィルタの重みを一次元データとして可視化したものである。この図から、学習サンプルの増加によって、よりはっ

きりとした特徴が抽出されていることが分かる。特に、フィルタサイズが大きい場合は、様々な周波数成分の波がはっきりと現れている。5.3節のノイズの少ない小規模なデータセットについては少ないサンプル数で学習することができていたが、本実験では500匹分の学習データでも十分な精度を達成できなかった。これは、データセットに含まれるノイズや個体差の影響が大きいのことを示していると考えられる。しかし、十分な学習データがあれば提案手法は十分これらのデータのバリエーションに対応が可能であることが確認できた。

5.5 評価実験：ノイズが多く含まれるデータに対する判定精度

この実験では、ノイズが多く含まれるマウスの脳波/筋電位

のデータに対して、5.4 節での実験で学習した MC-SleepNet を用いて判定を行った。判定対象のノイズが含まれるデータは、マウス 4 匹分の脳波/筋電位のデータである。サンプリング周波数などは 5.1 節で説明した条件と同じである。また、それぞれのデータの測定期間は約 2 日間である。判定結果を表 5 に示す。実験結果から、MC-SleepNet はノイズを含むデータに対しても十分な精度で判定することが出来ている。ノンレムや覚醒については 95% 程度を維持することができているが、レムの判定精度は 75.7% とやや低い精度になった。データに含まれるノイズには基線揺れなどによる低周波のノイズが多い。低周波成分が増えたことにより、脳波に低周波成分があまり含まれないレムと覚醒についてはやや精度が低下したと考えられる。また、筋電位にノイズが含まれることによって睡眠と覚醒状態の判別が困難なエポックも覚醒とレムの判定精度に影響を与えたと考えられる。しかし、特にノイズが多いデータであっても、数%の精度低下は見られるが、大きく悪化することはないことを確認した。これは、双方向型 LSTM によって、前後の睡眠ステージの状態遷移を判定に反映することで、判定対象エポックに含まれるノイズの影響を軽減できているためと考えられる。

5.6 考 察

ノイズの少ないデータセットを用いた実験から提案手法が他の既存手法と同等かそれ以上の精度で判定可能であることを確認した。特に Accuracy や k 係数は提案手法が最も高く、専門家の判定に非常に近い判定を行うことができることが分かった。また、大規模なデータセットを用いた実験によって提案手法がデータの個体差によるばらつきやノイズに対しても有効な判定を行えることを確認した。提案手法は学習データを増やすことによって性能が向上することが確認され、既存手法が利用している周波数成分や主成分の設計を行わずとも、CNN による特徴抽出が、データから自動的にこれらに準じる有効な特徴量を得られていると考えられる。特に、提案手法と同様に RNN モデルの一種である LSTM を用いた LSTM モデルでは、特徴量として FFT で得た周波数成分を用いているが、LSTM モデルより良い精度で提案手法は判定することが出来ている。これは、周波数領域の特徴を CNN によって抽出できているためだと考えられる。しかし、このように周波数成分に準じた特徴量を抽出できているため、主に低周波成分を多く含むノイズが混入しているデータに対しては、ノンレムへの誤判定が増えることが分かった。ただし、精度の低下は数%であることから、提案手法が周波数成分だけでなく、振幅や波形などの特徴抽出や LSTM による睡眠ステージの時系列的な状態遷移を有効に利用し判定できていると考えられる。また、提案手法は小規模なデータセット、大規模なデータセットそれぞれの実験結果からレムの Sensitivity が他の睡眠ステージの判定精度より低下することを確認した。これはレムが、覚醒に似た脳波の特性と、ノンレムに似た筋電位の特性をそれぞれ持っており、それによって覚醒、ノンレムにそれぞれ誤判定されるためと考えられる。

6 まとめと今後の課題

本稿では、深層学習を用いて生体信号の特徴抽出やステージ遷移のモデル化を行うことで、実用上十分な精度をもった睡眠ステージ自動判定手法、MC-SleepNet の開発を行った。また、4200 匹分の大規模な睡眠生体信号データを用い、MC-SleepNet の評価を行ったところ、本手法はマウスの個体差やノイズといった信号のばらつきに対して高い頑健性を有することが示された。専門家の代替として十分利用できる精度・頑健性を兼ね備えており、睡眠研究の能率向上に役立つと考えられる。また睡眠研究では、遺伝子操作や睡眠障害によって、特定のステージに遷移しないなど、睡眠の性質が異なるマウスを対象とする研究も行われている。本稿では、そのようなデータに対して、提案手法が有効であるか検証できていない。そのため、今後はそれらのデータに対して、提案手法を適用できるように検証と改善を行う必要がある。

謝 辞

本研究の一部は、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、ならびに科学研究費補助金・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」によるものである。

文 献

- [1] Elman, J. L., “Finding structure in time.”, Cognitive Science, 14(2), pp. 179–211, 1990.
- [2] 鈴木悠太, 佐藤牧人, 塩川浩昭, 柳沢正史, 北川博之, “マウスの脳波/筋電図を利用した睡眠ステージ判定”, 第 9 回データ工学と情報マネジメントに関するフォーラム (DEIM2017), E6-1.
- [3] 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 18, pp. 502-518, 2013.
- [4] Adelchi Azzalini, Nicola Torelli, “Clustering via nonparametric density estimation”, Stat Comput, vol. 17, pp. 71-80, 2007.
- [5] Akara Supratak, Hao Dong, Chao Wu, Yike Guo, “Deep-sleepnet: A model for automatic sleep stage scoring based on raw single-channel eeg”, IEEE transactions on Neural System and Rehabilitation Engineering, vol. 25, no. 11, 2017.
- [6] F.A. Gers, J. Schmidhuber, F. Cummins, “Learning to forget: continual prediction with hstm”, Neural Computation, vol. 12, pp. 2451-2471, 2000.
- [7] 山部剛士, 佐藤牧人, 塩川浩昭, 柳沢正史, 北川博之, “深層学習を用いたマウスの睡眠ステージ分析”, 情報処理学会第 79 回全国大会, 5K-03.
- [8] 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.
- [9] C. Cortes, V. Vapnik, “Support-Vector Networks”, Machine Learning, Vol. 20, pp. 273-297, 1995.
- [10] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”, Journal of Machine Learning Research 15, pp. 1929-1958, 2014.
- [11] J. Richard Landis, Gary G. Koch, “The Measurement of Observer Agreement for Categorical Data”, Biometrics, vol.

33, pp. 159-174, 1977.

- [12] Kirsi-Marja Rytönen, Jukka Zitting, Tarja Porkka-Heiskanen, "Automated sleep scoring in rats and mice using the naive Bayes classifier", *Journal of Neuroscience Methods*, vol. 202, pp. 60-64, 2011.