

睡眠段階ごとの生体振動特徴に着目したニューラルネットワークによる推定

○千住太希 中理怡恒 佐藤寛之 高玉圭樹 (電気通信大学)

Estimation by Neural Network Focusing on Biological Vibration Features in Each Stage of Sleep

* Taiki. Senju, I. Nakari, H. Sato, K. Takadama (The University of Electro-Communications)

Abstract—This paper focuses on the different features of the biological vibration in each sleep stage, and proposes the sleep stage estimation method that learns its features independently by the convolutional neural network (CNN) and selects the estimated sleep stage from the independently estimated ones. The experimental result derives the following implications: (1) the attention function in CNN contributes to classifying the two classes of wake and non-REM stage 1 but not classifying the two classes of REM and non-REM stages 2, 3 and 4; (2) the accuracy of the five classes of the sleep stage estimation in the proposed method is higher than in CNN with or without the attention function.

Key Words: Sleep stage prediction, bio signal, attention learning

1 序論

経済協力開発機構加盟国によると、日本の睡眠時間は最短である¹⁾。また、厚生労働省の国民調査によると、一般成人の5人に1人が何らかの睡眠の悩みを持ち、十分な睡眠を取れていないと推計される²⁾。睡眠不足が積み重なると、睡眠負債という状態になり、うつ病、認知症、生活習慣病などの病気の発症リスクを高める可能性がある。個人が健康であるためや、国の医療費を削減するためにも睡眠に関する問題は早期に改善されるべきである。

睡眠に関する問題を発見するためには、終夜睡眠ポリグラフ (Polysomnography: PSG) 検査によって取得された脳波計、筋電図、眼球運動などの生体データを用いて、WAKE, REM, Non-REM (NREM) 1~4 の6段階に分類される睡眠段階を導出して分析する必要がある。しかし、この PSG 検査には、頭部や身体に複数の電極を取り付ける必要があり、被験者に対する身体・精神的負担が非常に大きく、普段の睡眠のデータを取得できない問題がある。このような問題から、PSG 検査に代わるマットセンサの上に寝るだけで睡眠段階を推定できるような簡易的な検査方法に対する需要が高まっている。例えば、無拘束型の睡眠段階推定法として、マットセンサをマットレスの下に設置し、取得したセンサ値 (以下、生体振動データ) を分析することで睡眠の質を解析することや、アクチグラフ (腕時計型の加速度センサ) を被験者に取り付けて生体振動データを分析することがあげられる。それらの先行研究として、渡辺らはマットセンサから得られた生体振動データから心拍の中周波成分を抽出し、それと睡眠段階に相関性があることを示した³⁾。Xiao らは、アクチグラフから取得した心拍変動から時間領域と周波数領域の特徴量を設計し、Random Forest によって、WAKE, REM, NREM の3段階の分類をしている⁴⁾。しかし、これらの方法は推定精度に限界があり、実応用に展開できるレベルに至っていないため、より精度の高い推定精度の実現が求められる。特に、心拍や体動のみでは生体

振動データに含められる他の情報を活用できないため、マットセンサなどから取得できる全ての生体振動データ (呼吸、心拍、体動の振動などの複数の振動を混合した生体振動データ) を分析することが重要となる。

そこで、本研究では、複数の振動が混合した生体振動データを活用するアプローチをとる。具体的には、睡眠段階の基本単位であるエポック単位 (30 秒) で生体振動データを、フーリエ変換を用いて周波数領域に変換したパワースペクトルを算出し、それらを時系列順に並べたスペクトログラムを画像として畳み込みニューラルネットワーク (CNN) で学習する。しかし、各睡眠段階の特徴は全く異なり、1つの CNN で多クラス分類を学習させることは難しいため (ある CNN は WAKE の分類は得意だが、REM の分類は不得意であるなど)、本研究では各睡眠段階に対しての2クラス分類モデルを構築する。このとき、入力データのどこに着目して学習すべきかを指示するアテンション機構⁵⁾を用いることで、(CNN の学習した中身の解釈性の向上と) 分類精度の向上が期待できるが、各睡眠段階の特徴をアテンションとして捉えられるかどうか依存するので、アテンションとして特徴を捉えることができた睡眠段階に対してはアテンション機構を導入し、そうでない場合はアテンション機構を導入しない。以上より、本研究では、アテンション機構を各睡眠段階に応じて導入した CNN を提案し、マットセンサから得られる生体振動データのスペクトログラムから各睡眠段階の2クラス分類した後、5段階の睡眠段階 (WAKE, REM, NREM1, NREM2, NREM3&4) を推定するとともに、その有効性を検証することを目的とする。

以下、第2章で睡眠段階とその導出方法を述べた後、第3章で畳み込みニューラルネットワークを説明し、提案手法を第4章で説明する。提案手法の有効性を検証するために、第5章で実験をし、その結果の考察をする。最後に、本研究を結論および今後の課題について第6章でまとめる。

2 睡眠段階

睡眠段階とは、睡眠の深さを客観的に表す指標であり、PSG検査によって測定される、脳波、眼球運動図、筋電図などの生体データをもとに、Rechtschaffen and Kales法⁹⁾。よって推定され、浅い睡眠から深い睡眠にかけて、WAKE (浅い睡眠), REM, Non-REM (NREM) 1~4の6段階で表現される (NREM3とNREM4をNREM3&4と合わせることも多い)。Fig.1に一夜分の睡眠段階の例を示す。縦軸は睡眠段階、横軸は時間を表す。各睡眠段階に現れる一般的な生理学的特徴は次の通りである。WAKEは最も浅い睡眠段階であり、寝返りなどの体動が出現しやすい。REMは急速眼球運動を伴う睡眠段階であり、筋肉の緊張がなくなり、体動が生じにくい特徴と、心拍や呼吸などの自律神経機能が乱れやすい特徴がある。NREMは筋肉の緊張が保たれ、NREMの中で最も睡眠が浅いNREM1に体動が出現しやすく、NREM2~4では、体動が出現しにくくなる。また、睡眠が深くなるほど、心拍や呼吸などの自律神経機能が安定する特徴がある。

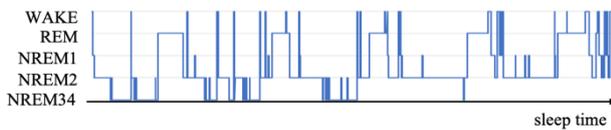


Fig. 1 Example of sleep stage.

3 畳み込みニューラルネットワーク (CNN)

3.1 CNN の学習

畳み込みニューラルネットワーク (CNN) は画像認識タスクに対して高い性能を示すニューラルネットワークである。このニューラルネットワークは全結合のニューラルネットワークと比較してスパースな結合を持っている。画像の位置 (i, j) の値を z_{ij} 、サイズ $H \times H$ の畳み込みフィルタの画像の位置 (p, q) の値を h_{pq} とすると畳み込み後の新しい入力は以下のように表せる。

$$u_{ij} = \sum_{p,q=0}^{H-1} z_{i+p,j+q} h_{pq} \quad (2.2.1)$$

モデルの活性化関数を f とすると次の層の入力 z' は以下のように表せる。

$$z'_{ij} = f(u_{ij}) = f\left(\sum_{p,q=0}^{H-1} z_{i+p,j+q} h_{pq}\right) \quad (2.2.2)$$

またプーリング層では畳み込み層で得られた局所的なパターンを平行移動などの位置ずれにロバストにするために用いられる。入力画像の (i, j) の位置を中心とした $H \times H$ の大きさの P_{ij} に対して、最大プーリングは入力 z'_{pq} に対して出力 u'_{ij} は以下のように表せる。

$$u'_{ij} = \max_{(p,q) \in P_{ij}} z'_{pq} \quad (2.2.3)$$

以上のような畳み込み層→活性化関数→プーリング層の処理を繰り返すことで畳み込みニューラルネットワ

ークは構築される。

3.2 インセプションモジュール (Inception module)

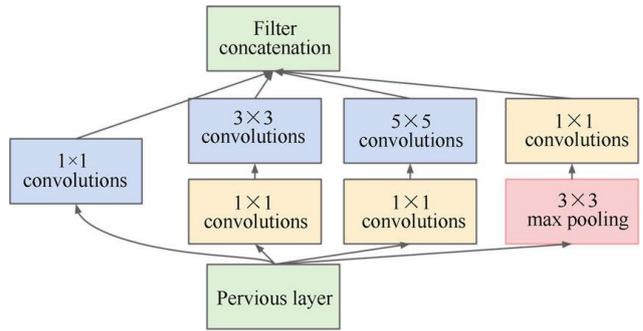


Fig. 2 Inception module

インセプションモジュールはGoogleNetで最初に導入されたネットワークの幅を活用したモデルである¹⁰⁾。Fig. 2ではPrevious layerからの入力に対して4箇所処理をする。このモジュール一つで様々なスケールでの畳み込み演算が同時に行うことができ、深層化せずとも高いパターン認識能力が実現できる。本研究ではインセプションモジュールを複数重ねたInception-v3と呼ばれるモデルの一部を使用した。Inception-v3とは深さが48層の畳み込みニューラルネットワークである。100万枚を超えるイメージで学習させた事前学習済みのネットワークを、ImageNetデータベースから読み込むことができる。この事前学習済みのネットワークは、イメージを1000個のオブジェクトカテゴリに分類できる。今回はInception-v3の中間層を取り出し全結合層 (FC) で結合し、睡眠段階を分類するモデルを作成する。

3.3 アテンション機構

アテンション機構はNNの可視化の際に用いることができるネットワーク構造である。隠れ層の中に以下の平行な処理を行い、アダマール積を取ることで注目する領域を学習することができる。アテンションを持つNNはFig. 3に示すような例が考えられる。Fig. 3は画像データを基に多クラス分類を行うNN内にアテンション機構を持つことで画像データのある領域に注目を当てて分類の根拠を学習することができる。

4 提案手法

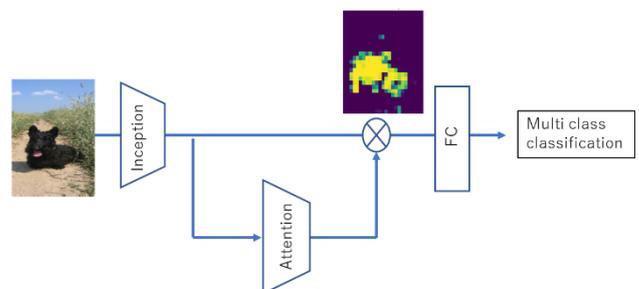


Fig. 3 Inception-v3 with Attention

本研究では3種類の分類手法を提案する。各提案手法では複数の分類器を組み合わせることで5クラス分類を再現する。

4.1 提案手法 1

Fig. 4 に提案手法 1 のフローを示す。図中の緑色の四角は生体振動データの正規化対数スペクトログラムの画像(5.2 節で説明), 緑色の四角の中の赤い線は5段階の睡眠段階を2段階に変換したものを表す。Inception はインセプションモジュール, FC は全結合層, WAKE class classification は WAKE か non-WAKE かの2クラスを出力する出力層, Multi class classification は5クラス分類機構である。なお, 2クラス分類のネットワーク構造は48層のインセプションモジュール, 全結合層, 出力層からなる。これから分かるように, 提案手法 1 ではアテンション無しのCNNを用いて各睡眠段階(WAKE, REM, NREM1, 2, 3&4)の2クラス分類をそれぞれ学習したのちに, それらの結果を基に Multi class classification で5クラスに分類する。ここで重要なことは, 2クラス分類の学習器の結果が競合した場合, その中から1つを選ぶ必要があるため, 5章での実験 1 により求めた各学習器の推定精度(F 値)に基づいて優先順位をつける。基本的な優先順位は NREM2, REM, NREM1, WAKE, NREM3&4

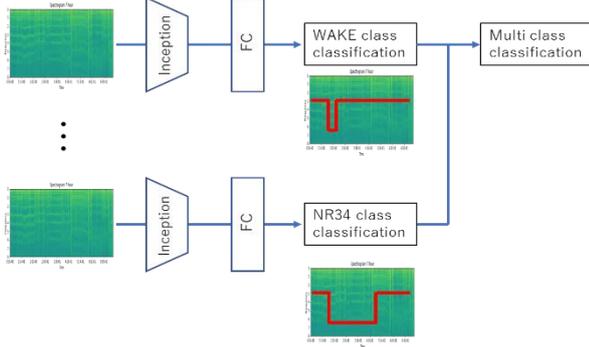


Fig. 4 Proposed method 1

となるが, 以下, それを基にした分類ルールを示す。

1. 5つのCNNの予測が1つのみ Positive であるとき, その予測を睡眠段階と推定する。
2. NREM2 と REM の CNN の予測が共に Positive, かつ, 予測が Positive である他の CNN が1つ以上存在するとき REM と推定する。
3. WAKE と NREM1 の CNN の予測のみが Positive のとき WAKE と推定する。
4. 上記のルールに当てはまらず, 複数の CNN が Positive と予測した場合, NREM2, REM, NREM1, WAKE, NREM3&4 の優先順位をつけて睡眠段階を推定する。
5. 全ての CNN の予測が Positive でなかった場合, NREM2 と推定する。

4.2 提案手法 2

Fig. 5 に提案手法 2 のフローを示す。同中の色やネットワークなどの意味はすべて Fig. 4 と同じである。異なるところは, インセプションモジュールの後にアテンション機構を導入した CNN になっているのみである。なお, アテンション機構のネットワーク構造は, 1層の畳み込み層からなる。

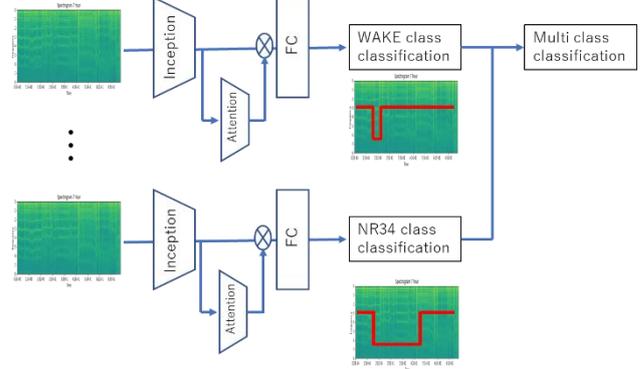


Fig. 5 Proposed method 2

4.3 提案手法 3

提案手法 3 は提案手法 1 と 2 を組み合わせたものである。具体的には, 各睡眠段階推定の2クラス分類ではアテンション機構が有効に働く場合と働かない場合があるため, 5章での実験 1 により求めた結果より, 有効に働く場合はアテンション有りのCNN, 働かない場合はアテンション無しのCNNを採用する。基本的な優先順位は提案手法 1 と 2 と同様となるが, NREM1 と WAKE はアテンション有りのCNN, NREM2 と REM と NREM3&4 はアテンション有りのCNNを用いる。以下, それを基にした分類ルールを示す。

1. アテンション有りの WAKE, NREM1 の CNN, アテンション無しの REM, NREM2, NREM3&4 の CNN の中で, 1つだけの CNN の予測が Positive であるとき, その予測を睡眠段階と推定する。
2. アテンション無しの NREM2 と REM の CNN の予測が共に Positive, かつ, 予測が Positive である他の CNN が1つ以上存在するとき REM と推定する。
3. アテンション有りの WAKE と NREM1 の CNN の予測のみが Positive であるとき WAKE と推定する。上記のルール当てはまらず, 複数の分類器が Positive と予測した場合提案手法 1 と同じ優先順位で推定する。
4. 上記のルールに当てはまらず, 複数の CNN が Positive と予測した場合, NREM2(アテンション無し), REM(アテンション無し), NREM1(アテンション有り), WAKE(アテンション有り), NREM3&4(アテンション無し)の優先順位をつけて睡眠段階を推定する(優先順位は提案手法 1 と同じだが, 採用する CNN は異なる)。
5. 全ての CNN の予測が Positive でなかった場合,

NREM2 と推定する。

5 実験

5.1 実験内容

本研究では提案手法の有効性を検証するために、次の実験を実施する。

- ・ 実験 1: 各睡眠段階におけるアテンションの効果
WAKE, REM, NREM1, NREM2, NREM3&4 のそれぞれの 2 クラス分類に対してアテンション機能無し CNN とアテンション機能有り CNN との比較
- ・ 実験 2: CNN と提案手法の比較
アテンション機能無し CNN, アテンション機能有り CNN, 提案手法 1, 2, 3 の 5 クラス分類の比較

5.2 実験環境とデータの前処理



Fig. 6 TANITA mat sensor

本研究では、Fig. 6 に示す TANITA のマットセンサ (サンプリング周波数は 16Hz) を用い、それをベッドの下に敷くことで体動、心拍、呼吸などの生体振動データを時系列データとして得る。同時に PSG 検査を実施し、各時刻の正解となる睡眠段階を取得することで、それらと各時刻の TANITA のデータから得られた生体振動データを紐づけることで教師あり学習のためのデータセットを作成する。具体的には、生体振動データに対して 64 秒間の短時間窓を取り、フーリエ変換を行うことで 8Hz の 512 次元のデータを得て、細かいスケール変化を捉えられるように対数に変換をする。このデータに対して正規化したものを 1 秒の対数スペクトログラムとし、この処理を 0.25 秒ずつずらし、32 秒間繰り返すことで、縦軸 512 個×横軸 128 (= 32×4) 個の正規化対数スペクトログラムの画像を得る。この正規化対数スペクトログラムの画像を入力、正解となる睡眠段階を出力として、CNN に学習させる。CNN の学習モデルとしては 4.1~4.3 節にある各睡眠段階に対する 2 クラス分類モデルをアテンション付きのものとアテンション付きでないものを作成する。この際睡眠段階のラベルの付け替えが必要になる。例えば WAKE の 2 クラス分類である場合 WAKE とそれ以外でクラスを分類するようなラベル付けにする。また 5 クラス分類のモデルを作成し 3 種類の提案手法と比較する。

5.3 評価基準

Table. 1 Details of subjects

Subject	age	WAKE	REM	NR1	NR2	NREM3&4
A	30s	95	173	33	357	0
B	40s	48	104	76	380	1
C	20s	49	106	45	345	1
D	40s	32	114	35	211	0
E	40s	70	61	44	390	2

F	60s	91	148	25	409	0
G	50s	32	234	51	488	0
H	40s	41	79	49	365	22
I	20s	41	165	37	395	155

実験 1 では、正確率(accuracy), 適合率(precision), 再現率(recall), F 値(f-measure)を評価し、実験 2 では正確性を評価する。なお、これらの値は、被験者 8 名に対して CNN で学習し、8 名以外の被験者に対して実験した結果に対して計算する。Table. 1 に各被験者の年齢、各睡眠段階の数を示す。

5.4 実験結果

5.4.1 実験 1:各睡眠段階におけるアテンションの効果

2 クラス分類の学習における適合性、再現率、F 値の 3 回の平均を Table 1 に示す。なお、NREM3&4 に関してはテストデータが 1 つしかないため算出していない。Table. 2 から各睡眠段階に対して以下のことがわかる。

Table. 2 Accuracy, Precision, Recall, F-measure

		Acc	Pre	Rec	F-m
WAKE	Attention	84.2%	32.0%	63.9%	41.2%
	No-attention	92.1%	32.6%	47.8%	36.1%
REM	Attention	66.7%	38.2%	64.2%	45.5%
	No-attention	85.3%	54.4%	79.6%	63.5%
NR1	Attention	86.4%	32.8%	65.2%	43.7%
	No-attention	90.8%	20.3%	60.3%	27.7%
NR2	Attention	68.7%	71.3%	87.1%	78.4%
	No-attention	68.5%	78.6%	82.6%	80.4%

- ・ WAKE と NREM 1 のときはアテンションがある方が高い F 値(41%)(43.7%)を示す
- ・ REM と NREM2 のときはアテンションがない方が高い F 値(68%)(80.4%)を示す
- ・ F 値は NREM2, REM, WAKE, NREM1 の順番順に低くなる

5.4.2 実験 2: CNN と提案手法の比較

5 クラス分類の CNN と 3 つの提案手法の正確性を Table. 3 に示す。この表から、一致率は提案手法 3 が一番高く、続いて提案手法 1, 提案手法 2, CNN(アテンション有り), CNN(アテンション無し)の順番となった。

Table. 3 Comparison of each methods

method	accuracy
CNN(Attention)	62.09%
CNN(No-attention)	52.88%
Proposed method 1	72.89%
Proposed method 2	66.12%
Proposed method 3	73.40%

5.5 考察

5.5.1 各睡眠段階におけるアテンションの効果

Table. 2 から、各睡眠段階によってアテンションの有無の効果は異なったが、この原因としては体動の影響が考えられる。Fig. 7 に WAKE のときの 13×13 のア

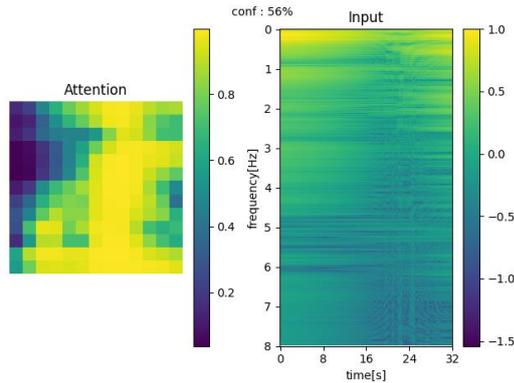


Fig. 7 Attention and input of WAKE

テンション(左)と 128×512 の正規化対数スペクトログラムの画像(右)を示す。なお、同図(左)の色は黄色/黒色であるほどその場所が分類にとって重要である/ないことを意味し、同図(右)の縦軸は周波数(0~8Hz)、横軸は時間(0~32 秒)、色は黄色/黒色であるほどその場所の周波数成分が多く存在する/存在しないことを意味する。同図(右)から、24s 付近に縦方向に伸びる線が確認でき、体動が生じていることを示しているが、同図(左)のアテンションからもこの領域を重視している。これは、アテンションが WAKE の特徴を適切に捉えていることを示す。同様の傾向が、NREM1 にも現れたが、これは WAKE と NREM1 は他の睡眠段階と比べて体動が起こりやすいからである。このように比較的体動の多い WAKE, NREM1 ではアテンションは分類に貢献する。

一方、REM, NREM2, NREM3&4 はどうであろうか？ Fig. 8 に REM のときの 13×13 のアテンション(左)と 128×512 の正規化対数スペクトログラムの画像(右)を示す。なお、同図の色、縦軸と横軸の意味は Fig. 7 と同じである。同図(右)から、0.2Hz 付近に横方向に伸びる黄色の線が確認でき、比較的安定した呼吸をしているように見えるが、同図(左)のアテンションからその領域以外だけでなく、他の領域も重視していることから、REM の特徴を適切に捉えられていない。生理学的にも REM は心拍と呼吸が乱れることが分かっており、それらの影響によるものであることが分かる。このような場合は、アテンション機構を導入しない方がよいことにあり、それが Table. 2 の結果として表れている。

また、NREM2 のアテンションと入力を Fig. 9 に示す。Fig. 9 は Fig. 7, Fig. 8 と同じようにアテンションとその時のスペクトログラムを示している。NREM2 と NREM3&4 は、REM のようにアテンションが睡眠段階の特徴を捉えることができているが、それ以外の高い周波数領域も重視しており(NREM2 と NREM3&4 は高周波成分がないという意味で重視していると捉えることはできる)、重視位置が分散していることから、

アテンションの効果はさほどないと言える。

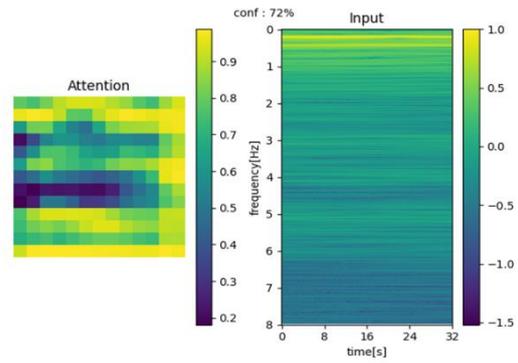


Fig. 8 Attention and input of REM

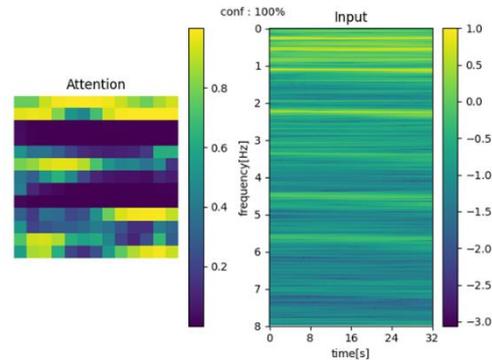


Fig. 9 Attention and input of NREM2

5.5.2 CNN と提案手法の比較

Table. 3 に示すように、実験 2 によって CNN よりも提案手法の方が一致率が改善された。これは本提案手法がアンサンブル学習の役割を果たしていたと考えられる。アンサンブル学習とは統計的機械学習の手法の一種であり、個々に学習した複数の学習器を組み合わせることで汎化性能を上げる学習アルゴリズムである。アンサンブル学習の枠組みでは個々の学習器を弱学習器 h 、学習器を H と表す。今回の提案手法 1 にアンサンブル学習の概念を当てはめると Fig. 10 のようになる。

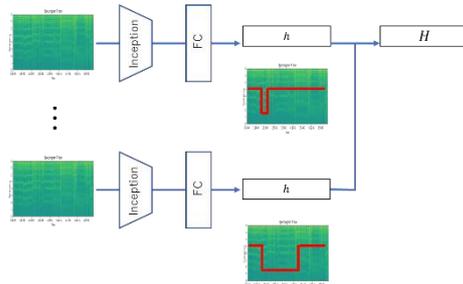


Fig. 10 Ensemble learning on proposed method 1

Fig. 10 は Fig. 4 で示した wake class classification の部分が弱学習器 h に変わり、multi class classification の部分が学習器 H 変更したものである。本研究では弱学習器 h から学習器 H を事前実験により分類ルールを作成することで、一度に分類する CNN

と比べて汎化性能を向上することができた。更なる拡張としては、アンサンブル学習の枠組みでバギングやブースティングを利用することで事前実験を必要としない分類ルールを生成できると考えられる。

結論

本研究では、睡眠段階推定精度の向上を目的とし、マットセンサから得られる生体振動データのスペクトログラムから各睡眠段階の2クラス分類した後、5段階の睡眠段階（WAKE, REM, NREM1, NREM2, NREM3&4）を推定する手法を提案した。健常者8名の睡眠データを基に学習し、それ以外の健常者で評価する被験者実験の結果、(1)提案手法、アテンション機構ありのCNN、アテンション機構なしのCNNの順で睡眠段階の精度が高いこと（つまり、多クラス分類として睡眠段階推定するよりも、2クラス分類を組み合わせた方が推定精度高いこと）、(2)体動の影響を受けやすいWAKEとNREM1は体動の特徴をアテンションとして捉えやすいが、REMはその特徴を捉えにくく、NREM2（とNREM3&4）は安定した生体振動であるため、アテンションがあってもなくても大きな効果はないこと（つまり、アテンション機構が有効である場合とそうでない場合の睡眠段階推定には存在すること）が明らかになった。今後は体動以外の特徴を捉えることができるアテンションの構造や、学習データを増やすために生体信号に基づくスペクトログラムに対する新しいデータ拡張方法が必要である。

参考文献

- 1) 経済協力開発機構, “GENDER EQUALITY,” Gender Data Portal 2019. <https://www.oecd.org/gender/data/>
- 2) 厚生労働省, 平成26年「国民健康・栄養調査」. <https://www.mhlw.go.jp/stf/houdou/0000106405.html>
- 3) 野田 明子: 睡眠ポリグラフ検査 (PSG) の測定方法と判定, 日本睡眠学会編, 睡眠学ハンドブック, pp.690-699, (2010)
- 4) T. Watanabe and K. Watanabe. Estimation of the Sleep Stages by the Non-Restrictive Air Mattress Sensor. *Transactions of the Society of Instrument and Control Engineers*, Vol. 37, No. 9, pp. 821-828, 2001.
- 5) Xiao, Meng, et al. "Sleep stages classification based on heart rate variability and random forest." *Biomedical Signal Processing and Control* 8.6 (2013): 624-633.
- 6) Yenigalla, Promod, et al. "Speech Emotion Recognition Using Spectrogram & Phoneme Embedding." *Interspeech*. 2018.
- 7) Huang, Jingshan, et al. "ECG arrhythmia classification using STFT-based spectrogram and convolutional neural network." *IEEE Access* 7 (2019): 92871-92880.
- 8) Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, Illia Polosukhin: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017).
- 9) A. Rechtschaffen and A. Kales, “A manual of Standardized terminology, Techniques and Scoring System for Sleep Stages of Human Subjects,” *Public Health Service US Gov. Printing Office*, Washington, DC, 1968.

- 11) Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, Andrew Rabinovich: Going deeper with convolutions