

特徴波の発生タイミングを考慮したモデルによる 睡眠ステージ判定精度の向上とエポックフリーなステージ判定

山田 空^{†,‡} 堀江 和正^{†,§} 北川 博之^{†,§}

[†]筑波大学 〒305-8574 茨城県つくば市天王台 1-1-1

E-mail: [‡]s2113516@s.tsukuba.ac.jp, [§]{horie, kitagawa}@cs.tsukuba.ac.jp

あらまし 睡眠診断の効率化を目指し、深層学習モデルを用いて生体信号から睡眠ステージ（レム睡眠など）を判定する手法が提案されている。中でも先行研究である Sleep-CAM は、判断根拠を説明できる実用性の高い手法であるが、判断根拠である「特徴波」の持続時間や発生タイミングを構造上把握できないといった問題があった。

本研究では、Sleep-CAM モデルをベースに LSTM を導入することで、説明可能性を損なわず上記の問題を解決できるモデルを目指した。現時点では判定精度の改善には至っていないが、従来よりも時間軸方向に細かいステージの判定を達成した。

キーワード 医療・ヘルスケア, 時系列データ処理, 睡眠ステージ判定

1. 序論

ヒトの睡眠は覚醒 (wake), レム睡眠 (REM, Rapid Eye Movement), ノンレム睡眠 (N1~N3 の3 段階に分類される) の3 つの睡眠ステージで構成される。睡眠時無呼吸症候群 (SAS) や不眠症等の睡眠障害の診断・治療には、睡眠ステージの分布や遷移といった睡眠の状態や質の把握が重要である。

生体信号を用いた睡眠ステージの判定は睡眠診断において不可欠であるが、手間がかかる。高い判定精度を達成している既存の自動スコアリング手法は、睡眠段階判定の根拠を説明する能力の欠如といった理由から臨床の場では広く適用されていない。この問題に対し、判定根拠の説明手法を備えた深層学習モデル, Sleep-CAM[1]が提案されている。判定ステージとは独立して提示する機能を有している。また、考慮された信号が、医師・技師がステージを判定する際の根拠である「特徴波」と重なることが示されている。

Sleep-CAMは判定精度の面でも高い水準を示している。しかし、判定を間違えやすいエポックにはある程度の共通点があり、判定精度を改善する余地があると思われる。特に

Sleep-CAM論文中では、モデルが特徴波の持続時間や発生タイミングなどの把握に適していないこと、またエポックに含まれる睡眠段階の割合を考慮できないといった問題が考察されている。これらの問題点を解決することで、Sleep-CAMの精度がさらに改善すると期待される。

本研究では、(1) Sleep-CAMの持つ説明可能性を損なうことなく、(2) 特徴波の持続時間や発生タイミングなどが把握でき、(3) 複数の睡眠段階が含まれるエポックも適切に

考慮できる自動スコアリング手法の提案を目標とする。

2. 睡眠ステージスコアリング

睡眠ステージの手動スコアリングは、以下の手順で行われる。

PSG の計測: 患者に測定機器を装着し、6 つの脳波 (EEG) , 2 つの眼電図 (EOG) , 1 つの脳波 (EOG) を測定する。

信号の前処理: 測定された PSG 信号のノイズ除去を行い、信号を 30 秒間隔 (1 エポックとする) に分割する。専門家は各エポックに1つの睡眠ステージラベルを割り当てる。

特徴波の検出: 専門家は睡眠ステージを判定する際に用いる根拠となる「特徴波」を検出する。

睡眠ステージの割り当て: 専門家は米国睡眠学会 (AASM) が制定するルール[2]に則り、睡眠ステージを割り当てる。

このルールは主に以下の原則に則る。

- 特にエポックの前半に特徴波が存在すれば、それと関連する睡眠ステージを割り当てる
- 特徴波がない場合は、先行エポックのステージと特徴波を考慮してステージを割り当てる
- 複数の睡眠段階を示す特徴波が含まれる場合、最も多くの割合を占めるステージを割り当てる

特徴波の検出・睡眠ステージの割り当ては、患者一人、一晚あたりに二時間近くもかかる作業であり、多大な労力と時間を必要とする。これらの作業を自動化することで、医師や判定技師の労力を軽減することができることから、多くの研究者が自動判定モデルの開発を行っている[1, 3-7]。

3. Sleep-CAM [1]

先行研究であるSleep-CAM[1]は、判断根拠の説明手法を備えた深層学習モデルである。Sleep-CAM のモデル構造を図1に示す。Sleep-CAM は大きく特徴抽出部とステージ割り当て部から構成され、それぞれ睡眠ステージスコアリングにおける特徴波の検出とステージの割り当てを再現する。

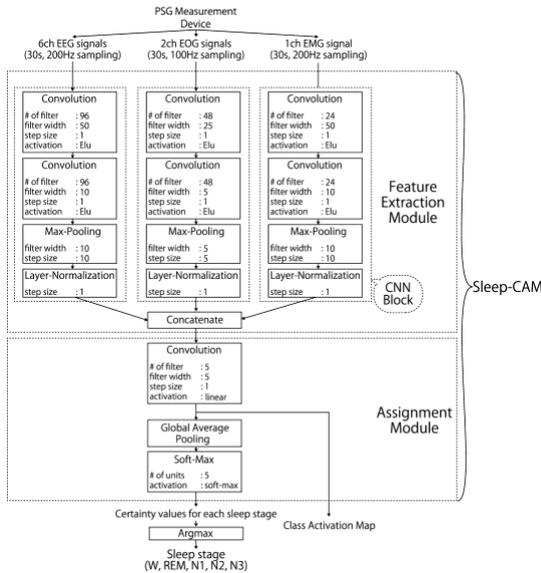


図 1 Sleep-CAM のモデル図

特徴抽出部は脳波・眼電位・筋電位それぞれの特徴を抽出する 3 つの CNN ブロックからなる。CNN を用いることにより定義が曖昧な特徴波を学習し、検出することが期待される。ステージ割り当て層では各時点での各睡眠ステージの尤度を示す Class Activation Map (CAM) が出力される。CAM は PSG 信号のどの部分に着目してステージを割り当てたかという判断根拠を提示する。実際の CAM を図 2 に示す。背景色のオレンジが濃い部分ほど判定時に注目されている。

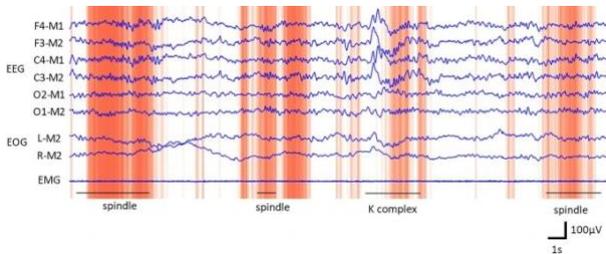


図 2 Sleep-CAM が提示する判断根拠

Sleep-CAM は技師判定との一致率 87.4%を達成している。両者の判定の一致度を示すカッパ係数は 0.8 を超え、技師判定と Sleep-CAM の判定がほとんど一致していると評価できる。異なる技師間一致率が 82.0%であることから、Sleep-CAM が実

際の臨床の場で利用することができるモデルであるといえる [8]。Sleep-CAM は異なる技師間の一致率を上回る性能を達成しているが、同一の技師が別の日に同じ信号を判定した場合の一致率は 90%を超えるため、Sleep-CAM にはまだ改善の余地があると考えられる。精度向上のためには Sleep-CAM と専門家の判定手順とのずれを修正することが必要だと考えられる。Sleep-CAM では特徴波を検出し、それをステージ判定に利用することは実現したが、特徴波の有無だけでなく、発生タイミングや持続時間を考慮するという専門家の判定手順は実現できていない。

4. 提案手法

「特徴波」の発生タイミングや持続時間といった「エポック内の時系列情報」を考慮するため、本研究では Sleep-CAM に LSTM を導入することを考えた。具体的には図 3 に示すように、ステージ割り当て部の Convolution レイヤと並列して LSTM を導入することで、Sleep-CAM の特徴波検出能力を損なうことなく、時系列情報を考慮できる。なお、Sleep-CAM と同一の部分（特徴抽出部と割り当て部の Convolution レイヤ）は、重みをコピー・固定し、LSTM 部分のみ学習している。

予備実験では、LSTM の出力は複数の特徴波の間、特定のステージを出力し続けることを確認している。これにより、エポックに複数のステージが含まれていた場合であっても、これらの割合を考慮することができそうだ。

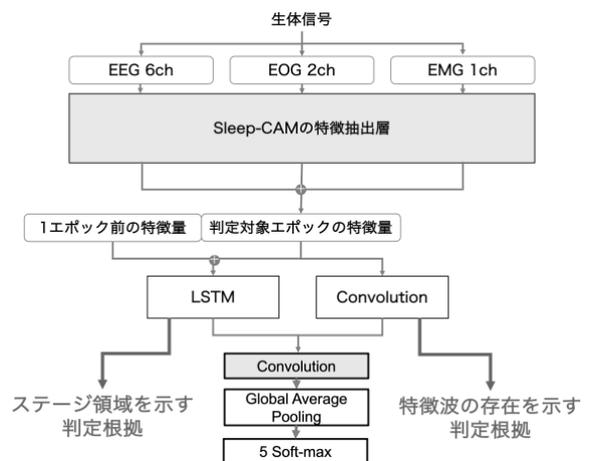


図 3 提案モデル図

提案モデルは、Convolution レイヤが出力する「特徴波」の有無を示す判定理由と LSTM が出力する「ステージ割合」を

示す判定根拠という 2 つの情報を兼ね備えた新しい判定根拠を提示することができる。この例では、被験者は深い眠り (N3) から寝返りを打つと同時に覚醒 (W), その後すぐに睡眠状態に戻っていく (N1) エポックを示している。ステージ割合を信号の下に帯状に表示することで、本来 N1 となるはずの区間が R に誤判定されていること、切り替わりのきっかけとなった部位に急速眼球運動に似た波形が存在することも一目で確認できる。Sleep-CAM の判定理由提示よりも詳細な検証が可能でダブルチェックもしやすい実用的な手法といえよう。提案モデルが提示する判定根拠を図 4 に示す。ステージ割合は信号の下に帯状に図示している。こちらの方が Sleep-CAM の判定理由提示よりも、情報量が多く、ダブルチェックのしやすい実用的な手法と言える。

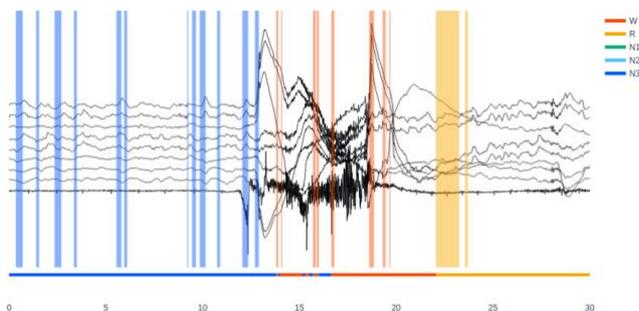


図 4 提案モデルが提示する判定根拠

5. 評価実験

20代と30代の異なる被験者の49夜分の生体信号（脳波・眼電位・筋電位）と判定技師により30秒ごとに割り当てられた睡眠ステージラベルをデータセットとする。この49件のデータを用い7分割交差検証としてSleep-CAMと提案モデルそれぞれで訓練と評価を行った。評価指標には判定技師との一致率をAccuracyとして用いた。結果はSleep-CAM がAccuracy = 87.4%, 提案モデルがAccuracy = 87.9%となり、若干ではあるが精度の改善が見られた。また技師判定に対する各ステージのRecallを示す混同行列を表1, 2 に示す。

表 1, Sleep-CAM の混同行列

		W	REM	N1	N2	N3
技師判定	W	95.0%	0.8%	3.5%	0.6%	0.1%
	REM	0.5%	88.6%	6.4%	4.6%	0.0%
	N1	8.9%	12.7%	47.7%	30.1%	0.6%
	N2	0.1%	1.5%	3.0%	92.3%	3.0%
	N3	0.4%	0.0%	0.0%	11.9%	87.7%

表 2, 提案手法の混同行列

		W	REM	N1	N2	N3
技師判定	W	95.3%	0.7%	3.4%	0.6%	0.1%
	REM	0.4%	89.6%	6.0%	4.0%	0.0%
	N1	8.2%	11.8%	51.7%	27.7%	0.6%
	N2	0.2%	1.4%	3.1%	91.7%	3.6%
	N3	0.4%	0.1%	0.0%	10.6%	88.9%

6. 考察

表1, 2 の混同行列から、各ステージの判定制度について見てみると、N1とN3の判定精度が大きく改善していることが見てとれる。N1の判定精度改善から、特徴波の発生・消失タイミングとステージの占有割合が考慮できるようになったと考えられ、N3の判定精度改善から特徴波の持続時間の考慮ができるようになり、K複合波とデルタ波の見分けができるようになったと考えられる。

7. LSTM 導入による副産物

研究の本来の目的とは異なるが、LSTM 層を導入したことにより副産物が得られた。LSTM が出力する「ステージ割合」を示す判定根拠を用いることで図 5 に示すようにエポックにとられない任意の時間間隔におけるステージ判定が得られると考える。これにより従来の睡眠ステージ判定よりも、詳細な睡眠遷移、睡眠ステージ割合を確認できるようになり、医学の新たな知見を得られるのではないかと考えられる。

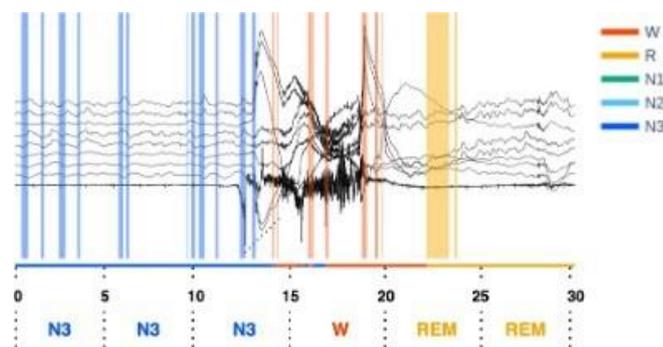


図 5 「エポックフリー」な判定

8. 今後の展望

本研究では、Sleep-CAM モデルをベースに LSTM を導入することで、説明可能性を損なわず「特徴波」の持続時間や発生タイミングを把握できないという問題を解決できるモデルを提案した。判定精度の改善は限定的であったが、Sleep-CAM よりも実用的な判定理由提示を実現した。

また、「エポックフリー」は判定を達成したが、その判定精度の評価は行えていない。現在、筑波大学国際統合睡眠医科学研究機構(IIIS)に数秒間隔での睡眠ステージ判定ラベルを依頼しており、それが完了次第「エポックフリー」な判定の精度評価を行うつもりである。

謝辞

本研究は、AMED (JP21zf0127005)、MEXT 地域イノベーション・エコシステム形成プログラム、MEXT WPI プログラム、JSPS 新学術領域研究 (15H05942Y)、JSPS 科研費 (17H06095、JP22K19802)、JSPS 最先端研究開発支援プログラム、JST 未来プログラム (JPMJMI19D8) 及び株式会社 S'UIMIN の支援の下行われています。

参考文献

- [1] Horie, K., Ota, L., Miyamoto, R., Abe, T., Suzuki, Y., Kawana, F., Kokubo, T., Yanagisawa, M., Kitagawa, H., Automated sleep stage scoring employing a reasoning mechanism and evaluation of its explainability. *Sci Rep* 12, 12799 (2022).
- [2] Berry, R. et al. The AASM manual for the scoring of sleep and associated events: Rules, terminology and technical specifications. Version 2.5. In *American Academy for Sleep Medicine* (2018).
- [3] Chambon, S. et al. A deep learning architecture for temporal sleep stage classification using multivariate and multimodal time series. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26**, 758-769. <https://doi.org/10.1109/TNSRE.2018.2813138> (2018).
- [4] Supratak, A., Dong, H., Wu, C. & Guo, Y. Deepsleepnet: A model for automatic sleep stage scoring based on raw single-channel EEG. *IEEE Trans. Neural Syst. Rehabil. Eng.* **25**(11), 1998-2008. <https://doi.org/10.1109/TNSRE.2017.2721116> (2017).
- [5] Zhang, L., Fabbri, D., Upender, R. & Kent, D. Automated sleep stage scoring of the Sleep Heart Health Study using deep neural networks. *Sleep*. <https://doi.org/10.1093/sleep/zz159> (2019).
- [6] Mousavi, S., Afghah, F. & Acharya, U. R. SleepEEGNet: Automated sleep stage scoring with sequence to sequence deep learning approach. *PLoSOne*. <https://doi.org/10.1371/journal.pone.0216456> (2019).
- [7] Perslev, M. et al. U-time: A fully convolutional network for time series segmentation applied to sleep staging. *Adv. Neural Inf. Process. Syst.* **32**, 4415-4426 (2019).
- [8] Danker-Hopfe H, Kunz D, Gruber G, Klösch G, Lorenzo JL, Himanen SL, Kemp B, Penzel T, Röschke J, Dorn H, Schlögl A, Trenker E, Dorffner G. Interrater reliability between scorers from eight European sleep laboratories in subjects with different sleep disorders. *J Sleep Res.* 2004 Mar;13(1):63-9. doi: 10.1046/j.1365-2869.2003.00375.x. PMID: 14996037.