Predicting Ordinal Level of Sedation from the Spectrogram of Electroencephalography

Haoqi Sun*, Sunil B. Nagaraj† and M. Brandon Westover*
*Department of Neurology, Massachusetts General Hospital, MA, USA
†Biomedical Signals and Systems Group, University of Twente, Enschede, Netherlands

***Abstract*—In Intensive Care Unit, the sedation level of patients is usually monitored by periodically assessing the behavioral response to stimuli. However, these clinical assessments are limited due to the disruption with patients' sleep and the noise of observing behaviors instead of the brain activity directly. Here we train a Gated Recurrent Unit using the spectrogram of electroencephalography (EEG) based on 166 mechanically ventilated patients to predict the Richmond Agitation-Sedation Score, scored as ordinal levels of -5, -4, ... up to 0. The model is able to predict 50% accurate with an error not larger than 1 level; and 80% accurate with an error not larger than 2 levels on hold-out testing patients. We show typical spectrograms in each sedation level and interpret the results based on the visualization of the gradient with respect to the spectrogram. Future improvements include utilizing the EEG waveforms since waveform patterns are clinically thought to be associated with sedation levels, as well as training patient-specific models.**

***Keywords*-EEG; sedation; machine learning; ordinal**

I. Introduction

Patients in Intensive Care Unit (ICU) are usually under continuous infusion of various sedatives and analgesia to reduce agitation and pain. Inappropriate levels of sedation could lead to longer ICU stay, delirium and increased morality [1]. To monitor the level of sedation, many clinical assessment protocols are proposed, such as the Ramsay scale [2], Richmond Agitation-Sedation Scale (RASS) [3] and Sedation Agitation Scales (SAS) [4]. For example, in RASS, the nurse or clinician applies incrementally intense stimuli to the patient and record the level on which the patient starts to respond. These assessments are clinically validated with relatively low inter-rater variance. On the other hand, since the assessment needs interaction with the patient periodically, it can disrupt patients' circadian rhythm and sleep. Further, the observational noise of assessing behaviors instead of the brain activity itself makes it harder to reflect the true brain status.

As an alternative, there are electroencephalography (EEG)-based anesthesia depth monitors commercially available, such as bispectral index (BIS) (Aspect Medical Systems, Norwood, MA, USA), Nacrotrend (Monitor Technik, Bad Bramstedt, Germany), and Patient State Index (PSI) (Hospira, Lake Forest, IL, USA), etc [5]. Clinically, since EEG is not yet commonly used in ICU, directly applying anesthesia depth monitors to ICU patients is not thoroughly validated [6]. Technically, these brain monitors use a two-step approach where a noisy prediction is first made for each EEG segment; then the noisy predictions are smoothed to get a relatively stable score. Such technique involves a trade-off between more/less smoothing and low/high time resolution, which is hard to find a theoretical optimum. We recently proposed EEG-based sedation level prediction using Gated Recurrent Unit (GRU) [7], which can output smooth prediction without ad hoc smoothing. The model was trained to classify binary sedation levels, i.e. deep vs. light, by combining several sedation levels. Here we extend our previous work [7] to allow predicting all sedation levels as an ordinal number.

The following sections are organized as follows. Section II provides the demographics about patients, EEG signal preprocessing and the model architecture. Section III presents the results including performance evaluation and visualization of typical spectrograms in each sedation level. Section IV describes the important limitations and future directions. Section V concludes the paper.

II. Methods

A. Richmond Agitation-Sedation Scale

The Richmond agitation-sedation scale (RASS) is used as the reference sedation level. It is a clinical sedation assessment applicable in ICU environment [3], as shown in Table I. It can be divided into the sedation and agitation parts, where we only use the sedation part. The sedation part has 6 levels from -5 to 0. The ICU nurses or researchers assess the patients about every 2 hours.

Table I
The sedation levels in Richmond Agitation-Sedation Scale

Score	Term	Description
0	alert and calm	spontaneously pays attention
-1	drowsy	sustained awake >10s, eye to voice
-2	light sedation	awake <10s, eye to voice
-3	moderate sedation	movement to voice, no eye contact
-4	deep sedation	no resp to voice, move to phys stimuli
-5	unarousable	no resp to voice or phys stimuli

978-1-5386-7315-7/18/$31.00 ©2018 IEEE
DOI 10.1109/CW.2018.00060

B. Patient

The Partners Institutional Review Board approved retrospective analysis of the dataset without requiring additional consent. All together there are 195 patients enrolled from 2014 to 2016. The inclusion criteria are (1) in the ICU; (2) on mechanical ventilator; and (3) no neurological disease such as stroke or dementia. We visually check the quality of EEG signals and remove the bad ones. We also remove EEG recordings without any RASS from -5 to 0 available. Finally, we are left with EEGs from 166 patients. Table II describes the patient demographics.

Table II
PATIENT DEMOGRAPHICS

Characteristic	Value
Number of patients	166
Age, yr, median (IQR)	60 (51 – 70)
Sex, n (%Male)	112 (67.5%)
BMI, kg/m^2, median (IQR)	29 (24 – 35)
Days in ICU, median (IQR)	12 (7 – 20)
CCI, median (IQR)	3 (2 – 5)
APACHE II, median (IQR)	22 (15 – 28)
Diagnosis at ICU admission, n (%)	
Acute respiratory failure	104 (63%)
Liver or renal failure	38 (23%)
Surgery	35 (21%)
Sepsis	27 (16%)
Cardiac diseases	11 (7%)

C. EEG Preprocessing and Artifact Removal

The EEG signals are obtained using Masimo Sedline brain function monitors (Masimo Corporation, Irvine, CA, USA) with sampling frequency of 250Hz. There are 4 EEG channels fixed to the frontal head at F7, F8, Fp1 and Fp2 based on the 10-20 international system. The EEG signals are re-referenced to obtain bipolar montage to reduce artifacts: Fp1-Fp2, F7-F8, Fp1-F7 and Fp2-F8. We bandpass the signals between 0.5Hz and 16Hz to avoid possible artifacts from various machines in the ICU.

For each RASS assessment, we take a 10min EEG segment before the assessment and compute spectrogram using the multitaper spectral estimation. The parameters are: K = 7 tapers, window size T = 4s and window step 2s. We convert the spectral power into decibels, and normalize the spectrogram for each patient to have zero mean.

The criteria to identify segments contaminated by artifacts are either (1) amplitude larger than 500uV; (2) spectrum is spuriously staircase-like, as defined convolving the detrended spectrum with step signals and being larger than a threshold. This is usually caused by nonphysiologic artifacts such as cooling blankets or pumps in ICU environment. About 10% of the data is identified as artifacts.

D. Training the Model

The EEG spectrograms are fed to Gated Recurrent Unit (GRU) [8], a type of recurrent neural network, which is in principle an autoregression with inputs and gates learned using neural networks. At each time step, the model consists of passing the information through input, hidden node, reset gate and input gate. We stack two layers of GRUs with hidden nodes in each layer. The output from the first layer of GRU is dropped out at rate 0.5 to avoid overfitting, and then fed to the second layer.

The output from the last layer of GRU is fed to an ordinal regression layer. The ordinal regression allows to train and predict ordinal values, which is commonly found in clinical assessments. There are many ways to model ordinal numbers with probabilistic interpretations, mostly by maximizing the likelihood of falling into the intervals defined by learnable thresholds. Here we use "ordistic regression" [9] which generalizes the logistic regression for categorical numbers to ordinal regression. When the number of ordinal levels is two, it reduces to logistic regression.

We validate the model using leave-one-patient-out cross validation, where we take a patient as the testing set in turn. There is no common patient in either the training or testing set to ensure its generalization ability to new patients. The performance is pooled over the testing patients in all rounds of the cross validation.

III. RESULTS

A. Prediction Performance

We compare the predicted RASS levels to the RASS levels assessed by nurses on the testing patients. As shown in Figure 1, the percentage of predictions that match exactly with the nurses is low, with median at around 15% of all the assessments. This can be due to (1) the EEG spectrograms does not contain enough information; (2) the relatively high patient heterogeneity in terms of response to sedatives; and (3) the observational noise in these assessments. To validate the last assumption, the accuracy of having prediction error not larger than 1 RASS level is about 50%, which is significantly larger than the exactly matched ones (Mann-Whitney U p-value <0.01). We leave the validation of the first and second assumptions to future studies. Allowing prediction error not larger than 2 RASS levels leads to 80% accuracy. Finally, allowing prediction error not larger than 3 RASS levels leads to close to 100% accuracy.

B. Typical EEG Spectrogram Examples

The GRU, similar to other neural networks, is hard to interpret what information it uses to make the prediction. On the other hand, interpretability is at the center of clinical applications. Here we show a typical EEG spectrogram example at each RASS level in Figure 2, Figure 3, Figure 4, Figure 5, Figure 6 and Figure 7 respectively. We define "typical" as having the prediction value before applying

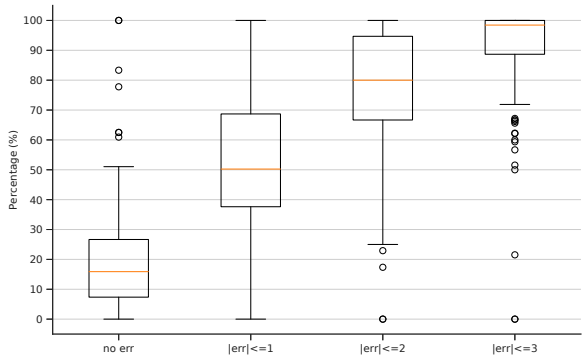

Figure 1. From left to right are the accuracies for each patient when allowing 0, 1, 2 and 3 levels difference in RASS prediction. These results are pooled over the testing set.

thresholds in ordinal regression at the final time step closest to the median value in all correctly predicted examples.

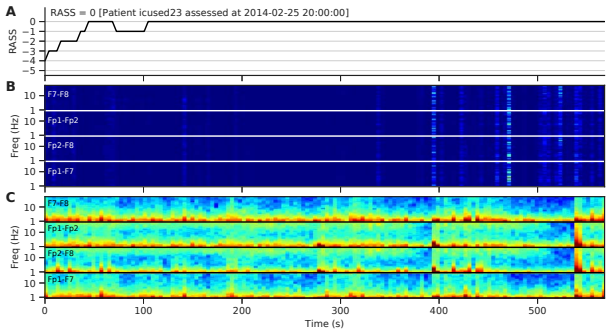

Figure 2. An example from an assessment with RASS at 0. (A) The predicted trace of RASS levels. (B) The absolute gradient of the final RASS with respect to the input spectrogram. For detailed explanations, see the main text. (C) The EEG spectrogram in decibel scale. Red means high power; blue means low power.

At the top panel of each figure, we show the predicted trace of RASS levels. The initial part is treated as "burn-in" period of GRU, which should not be used to interpret the results. This is because the GRU starts with initial zero hidden states, which are gradually washed out by the inputs. The burn-in period can be decided by giving a constant input and observe when the output reaches a plateau.

The middle panel shows the absolute value of the gradient of the final RASS prediction with respect to the spectrogram. In other words, the lighter parts indicate the important parts of the spectrogram, if we add noise to the important parts, the final RASS prediction would have a relatively large change compared to adding noise to dark parts. It turns out the import spectrogram parts are often associated with (1) changes in the spectrogram, such as Figure 2 around 400 and 480 seconds, Figure 3 around 100 seconds and Figure 6 around 250 seconds; (2) periodically reassuring the final RASS when the spectrogram keep similar, such as Figure 3 around 200 seconds, Figure 4 around 400 seconds, Figure 5 around 120 and 300 seconds and Figure 7 around 50, 120, 280, 380, 420 and 480 seconds.

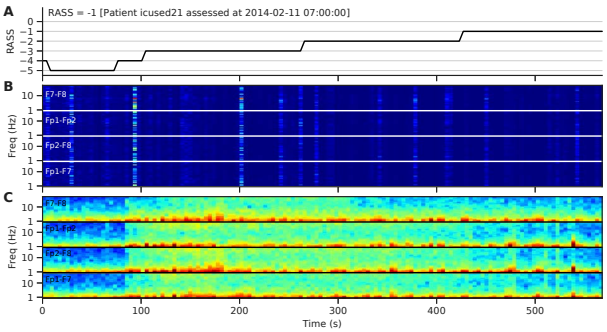

Figure 3. An example from an assessment with RASS at -1.

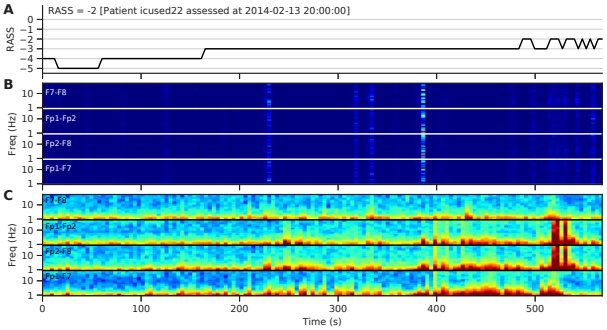

Figure 4. An example from an assessment with RASS at -2.

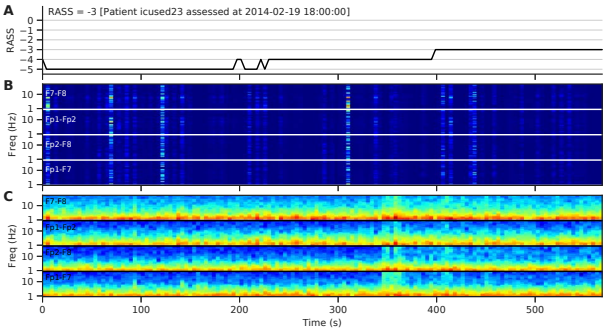

Figure 5. An example from an assessment with RASS at -3.

IV. Discussion

An EEG-based sedation monitoring system is developed based on 166 ICU patients. The use of ordinal regression enables the model to predict the RASS as an ordinal level. The use of recurrent neural network removes the need of ad hoc smoothing. The accuracy while allowing for prediction error not larger than 1 RASS level is about 50%. Allowing prediction error not larger than 2 RASS levels leads to 80% accuracy.

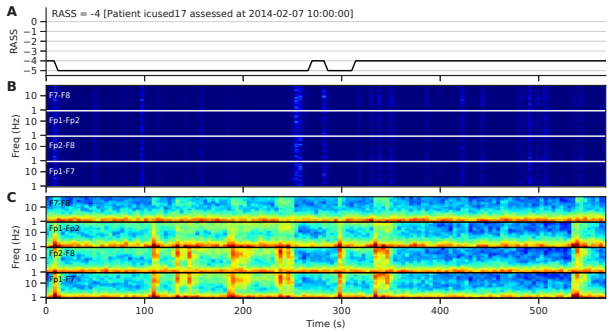

Figure 6. An example from an assessment with RASS at -4.

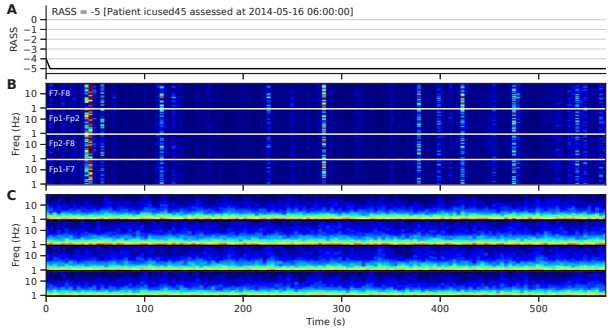

Figure 7. An example from an assessment with RASS at -5.

There are a few limitations. First, EEG spectrogram reflects the spectral information in EEG, but does not reflects intermittent waveform patterns, such as intermittent rhythmic delta slowing or burst suppressions. These waveform patterns are best captured by convolutional neural networks which can be highly specialized to these patterns as well as being invariant to its actual location in the EEG. Second, the wide range in the box plots in Figure 1 indicates a high amount of inter-patient heterogeneity, given various diagnoses for ICU admission. To alleviate the heterogeneity, training patient-specific models while still being able to benefit from the large amount of data from other patients, is a desirable property. Possible approaches include giving higher weight to the patient of interest, or using transfer learning to learn patient-specific models starting from an overall model. Third, standardized clinical validation protocol should be used to validate this model in a variety of clinical conditions as proposed in [10] in terms of clinical sign validation, pharmacological validation, clinical utility, cost-effectiveness, and ultimately outcome improvement.

V. Conclusion

As a real-time alternative to clinical sedation assessments, the ordinal sedation level can be predicted from EEG spectrogram using a combination of recurrent network and ordinal regression. It is expected that utilizing EEG waveform patterns and training patient-specific models can help improve the prediction performance.

Acknowledgment

MBW reports grants from NIH-NINDS (NIH-NINDS 1K23NS090900).

References

[1] C. G. Hughes, S. McGrane, and P. P. Pandharipande, "Sedation in the intensive care setting," *Clinical pharmacology: advances and applications*, vol. 4, p. 53, 2012.

[2] M. Ramsay, T. Savege, B. Simpson, and R. Goodwin, "Controlled sedation with alphaxalone-alphadolone," *British medical journal*, vol. 2, no. 5920, p. 656, 1974.

[3] C. N. Sessler, M. S. Gosnell, M. J. Grap, G. M. Brophy, P. V. O'neal, K. A. Keane, E. P. Tesoro, and R. Elswick, "The richmond agitation–sedation scale: validity and reliability in adult intensive care unit patients," *American journal of respiratory and critical care medicine*, vol. 166, no. 10, pp. 1338–1344, 2002.

[4] B. Vivien, S. Di Maria, A. Ouattara, O. Langeron, P. Coriat, and B. Riou, "Overestimation of bispectral index in sedated intensive care unit patients revealed by administration of muscle relaxant," *Anesthesiology: The Journal of the American Society of Anesthesiologists*, vol. 99, no. 1, pp. 9–17, 2003.

[5] B. Musizza and S. Ribaric, "Monitoring the depth of anaesthesia," *Sensors*, vol. 10, no. 12, pp. 10 896–10 935, 2010.

[6] C. J. Chisholm, J. Zurica, D. Mironov, R. R. Sciacca, E. Ornstein, and E. J. Heyer, "Comparison of electrophysiologic monitors with clinical assessment of level of sedation," in *Mayo Clinic Proceedings*, vol. 81, no. 1. Elsevier, 2006, pp. 46–52.

[7] H. Sun, S. B. Nagaraj, O. Akeju, P. L. Purdon, and M. B. Westover, "Brain monitoring of sedation in the intensive care unit using a recurrent neural network," in *Engineering in Medicine and Biology Society (EMBC), 2018 40th Annual International Conference of the IEEE*. IEEE, 2018 (Accepted).

[8] J. Chung, C. Gulcehre, K. Cho, and Y. Bengio, "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv preprint arXiv:1412.3555*, 2014.

[9] J. D. Rennie and N. Srebro, "Loss functions for preference levels: Regression with discrete ordered labels," in *Proceedings of the IJCAI multidisciplinary workshop on advances in preference handling*. Kluwer Norwell, MA, 2005, pp. 180–186.

[10] B. Heyse, B. Van Ooteghem, B. Wyler, M. Struys, L. Herregods, and H. Vereecke, "Comparison of contemporary eeg derived depth of anesthesia monitors with a 5 step validation process," *Acta Anaesthesiologica Belgica*, vol. 60, no. 1, p. 19, 2009.