34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Data-Driven Modeling of Sleep States from EEG

Alexander Van Esbroeck[1] and Brandon Westover[2]

Abstract— **Sleep analysis is critical for the diagnosis, treatment, and understanding of sleep disorders. However, the current standards for sleep analysis are widely considered oversimplified and problematic. The ability to automatically annotate different states during a night of sleep in a manner that is more descriptive than current standards, as well as the ability to train these models on a patient-by-patient basis, would provide a complementary approach for sleep analysis. We present a method that discovers latent structure in sleep EEG recordings, by extracting symbols from the continuous EEG signal and learning "topics" for a recording. These sleep topics are derived in a fully automatic and data-driven manner, and can represent the data with mixtures of states. The proposed method allows for identification of states in a patient-specific way, as opposed to the one-size-fits-all approach of the current standard. We demonstrate on a publicly available dataset of 15 sleep recordings that not only do the states discovered by this approach encompass the standard sleep stage structure, they provide additional information about sleep architecture with the potential to provide new insights into sleep disorders.**

I. INTRODUCTION

Sleep disorders, which include conditions such as sleep apnea and insomnia, have been estimated to affect over 50 million Americans [1]. These disorders are associated with higher rates of driving and occupational accidents, and an increased risk of cardiac disease [2,3].

Analysis of sleep is critical for the diagnosis, treatment, and scientific understanding of these disorders. Sleep analysis involves the identification of key properties in polysomnographic recordings, collections of a variety of physiological signals, throughout the course of a night of sleep. The most relevant of these signals for understanding the structure of sleep is the electroencephalogram (EEG). The EEG records the electrical activity of the brain, caused by the firing of millions of neurons, using electrodes placed on the scalp. EEG is the most commonly used signal in identifying the quality and progression of sleep, and often only a single electrode of EEG is necessary for sleep analysis.

Sleep is conventionally considered to be comprised of a number of stages. These stages consist of two main types, rapid eye movement (REM) and non-rapid eye movement (NREM). NREM can be divided further, into four stages (1, 2, 3, and 4) reflecting the continuum between drowsiness and deep sleep. This organization of sleep stages, designed by Rechtschaffen and Kales (R&K) [4], has been the sleep staging standard. Sleep staging is the process of annotating a recording of sleep with sleep stages, by evaluating all 30-second windows in the recording and assigning each data window to one of these conventional stages. This is primarily done using the EEG, although other signals such as the electrooculogram (EOG) provide useful supplementary information. The set of annotations for an individual's entire night of sleep is referred to as a hypnogram.

Normal sleep has a cyclic organization, in which individuals cycle from light to deep sleep, REM, and then return to light sleep. The transitions between states and the time spent in individual states carry information about the quality of sleep and insights into potential sleep disorders. The organization of sleep across the night, usually measured with the R&K stages, is referred to as sleep architecture, and is evaluated using the hypnogram.

Despite its essential role in sleep analysis, R&K staging is commonly considered to have many problems. These derive from both the subjectivity of the sleep staging process and the simplifications inherent in the stage definitions. We aim to address these problems with an automatic and unsupervised method for identifying latent states in sleep recordings. We explore an approach that extracts "words" from the EEG signal, and evaluate the EEG recordings as if they were natural language documents using a topic modeling approach. We hypothesize that this topic modeling approach can more expressively model structure across a night of sleep in an entirely data-driven and patient-specific manner. By tailoring states to a specific patient, and by allowing for mixtures of states, the resulting model improves in several ways over R&K staging and may provide new insights into sleep disorders.

II. BACKGROUND

There are a number of problems with the traditional approach to sleep staging. First, R&K is often regarded as an oversimplification of the actual structure of sleep. The hard distinctions between stages, such as between stage 3 and stage 4 which reflect different levels of deep sleep, impose unnecessary structure on the data to facilitate manual annotation when the true progression towards deeper sleep is likely a continuous process. Additionally, the stages may not represent the full variety of sleep activity well. For instance, drowsiness (stage 1 sleep) has been classified into as many as 9 different stages [5], a level of detail which the R&K system is unable to capture.

Another issue with R&K is that inter-rater reliability (how well the annotations of two different experts match) is low, meaning that hypnograms of the same night of sleep from two different annotators may differ significantly. This

[1] Computer Science and Engineering, University of Michigan, Ann Arbor, USA alexve@umich.edu

[2] Massachusetts General Hospital, Boston, MA

This material is based upon work supported by the National Science Foundation.

detracts from the value of the R&K stages as a standard, as annotations from different sleep labs are not directly comparable. The subjectivity of applying the R&K standard has motivated the design of a wide variety of automated sleep staging algorithms [6]. Unfortunately, due to the subjectivity of the R&K standard, evaluating the accuracy of these stagers is difficult, as prediction errors may actually reflect reasonable choices of annotation.

An unsupervised algorithm to identify sleep states could address both of these critical issues. An automatic method would give the consistency expected of a sleep staging standard, by avoiding the subjectivity of human annotators. By learning structure directly from the data, such an approach would also avoid the constrictive reliance on prior definitions of sleep stages, opening the door for richer descriptions and new ways to quantify sleep organization.

A final difficulty with the use of R&K is that the system was designed for young, healthy subjects. There are many individual differences in EEG and sleep organization [7], and applying a single set of staging definitions to a broad range of patients may pose difficulties in interpretation.

Prior work by Flexer et al. used a hidden Markov model (HMM) in an unsupervised approach to describe the structure of sleep recordings [8]. The method identified three population-wide states, corresponding to wake, sleep, and REM. The model used each window's posterior distribution over states as a continuous measure of sleep stages. A shortcoming of this approach is that although it results in a continuous mixture over states, the HMM assumes discrete states when estimating the model. This approach effectively simplifies R&K scoring even further, by reducing the set of states to three, limiting the model's ability to represent complex structure in the EEG.

We explore an alternative approach to unsupervised sleep analysis that allows for a more nuanced description of the sleep EEG. After extracting features from the EEG, we discretize the features and treat each value as a symbol. We treat these symbols as analogous to words, and apply Latent Dirichlet Allocation (LDA), a common approach to topic modeling in natural language documents, to the resulting symbols. Topic modeling identifies a set of themes in a collection of documents, which describe the latent structure behind the generation of the documents. In the context of sleep analysis, our goal is to identify latent states in the EEG recordings through "sleep topics" that can expand upon the information present in the R&K sleep staging. The benefit of using "sleep topics" lies in LDA's assumption that a given data instance (a document) can derive from multiple topics, as opposed to a single state. This allows for model flexibility in identifying sets of potentially concurrent time-varying states, as well as allowing for states that relate to only a subset of features.

We train models on individual patients, allowing each individual's sleep to be modeled separately. The development of patient-specific models avoids adverse effects inherent in fitting a single model to a population, where individual differences may be lost or misconstrued. While universally defined sleep stages have a critical role in sleep analysis, the development of more expressive patient-specific models can provide complementary information that expands the set of useful sleep analysis methods.

III. METHODS

For each patient, the single-channel EEG recording was divided into non-overlapping one-second segments. Each segment was analyzed to extract a variety of features. The features used were spectral power in four commonly used frequency ranges for EEG analysis (delta: $<$4 Hz, theta: 4-7 Hz, alpha: 8-13 Hz, and beta: 14-30 Hz).

After extracting features for each short window, we discretized each feature on a per-patient basis. The goal of the discretization is to convert the original continuous time series into a set of meaningful "words" that can be used to learn a topic model. The SAX approach to time-series symbolization was used [9], where for a given feature (e.g. delta energy), the full range of values for the patient were divided into 5 equiprobable bins, with boundaries at each quintile. Each bin was assigned a different symbol, corresponding to low through high values of that feature. After symbolization, each one second window was represented by four symbols, one for each spectral power feature in the data.

We divided each recording into non-overlapping 30-second segments, corresponding to "documents". The length of 30 seconds was chosen to correspond with the time scale used by R&K sleep staging, so that the results between the two methods would be comparable. In LDA, these documents are considered as "bags of symbols", where a document is represented as a vector of symbol frequencies, without considering order. We consider the set of symbols generated by SAX analogously to words in a natural language document. As the features were generated from one second segments, each document consisted of 120 symbol instances. The Latent Dirichlet Allocation model was applied to learn topics from the collection of documents (all 30-second windows in the recording).

LDA is a generative model for a collection of documents that assumes that the collection was derived from an underlying set of "topics" [10]. A topic is defined as a set of related symbols, for example symbols related to the waking state (high alpha energy, low delta energy). Specifically, a topic is a distribution over all symbols in the data. Each document is assumed to have been generated by a combination of topics, and has its own multinomial distribution over the set of all topics (e.g., p(light sleep) = 0.5 and p(deep sleep) = 0.5). The model assumes that each symbol in the document was obtained by first sampling a topic from the document's distribution over topics, and then sampling a symbol from that topic's distribution over symbols.

More precisely, LDA defines an underlying set of K topics, where each topic k can be defined by a distribution over all of the symbols in the vocabulary. Each document d is itself generated by a distribution over topics, where the generative process assumes a two-part process for each symbol w_{di} in the document. First, a topic z_{di} is chosen by

sampling from that document's topic distribution θ_d. Then the symbol w_{di} is sampled from that topic's distribution over symbols β_k. More explicitly, the model defines for a document d the prior distributions over observed symbols w_{di}, their latent topics z_{di}, and the document distributions over topics θ_d as:

$$\theta_d \sim Dir(\alpha)$$

$$z_{di} \sim Multi(\theta_d)$$

$$w_{di} \sim Multi(\beta_{z_{di}})$$

where α parameterizes a symmetric Dirichlet prior over topic distributions. In training the model, the parameters of interest are the θ_d and β_k values, which characterize the topics of semantically related symbols and the proportions of these topics in each document. The parameters of the model (the probabilities of these multinomial distributions) can be inferred from a set of unlabeled data using a variety of methods (Markov Chain Monte Carlo, Variational Inference).

A benefit of applying LDA to sleep analysis when compared with the commonly used HMM is that the topic model does not assume discrete states. Under the LDA generative model there can be multiple processes responsible for the generation of single window, whereas the HMM assumes that each window was generated by a single state. While it is possible to interpret the HMM's posterior distribution over states as a continuous mixture, the fact that the model optimization makes the discrete state assumption greatly affects the parameter estimates. Factorial HMMs remove this assumption, but like HMMs each state models the joint distribution of all features, so that each state must account for all features in the data. By considering windows as a bag of symbols from different features, each topic has the option of modeling only a subset of features, resulting in a more flexible representation of the data.

We trained a model for each patient using all of the documents generated from their recording. Parameter estimation was done with variational inference. The number of topics for each patient was chosen in a data-driven manner using the Akaike Information Criterion. The estimates for the posterior distribution of each window over the topics was used as a representation of the sleep recording, analogous to the R&K stages.

We evaluated the generated models to confirm two properties. First, we assessed whether the unsupervised model captured the same structure as the R&K standard. The ability of the unsupervised model to encompass well-established properties of sleep structure acts as validation of the quality of the results. This was done using two approaches: first, visualization was used to establish a correspondence between the derived topics and the R&K stages. Second, we trained a support vector machine (SVM) classifier that used the topic mixtures as features in predicting the R&K stages. High accuracy for a sleep stager built on topic mixtures would confirm the retention of the R&K relevant information. This also illustrates a method to establish correspondence between the unsupervised model's structure and the currently accepted gold standard of sleep staging. For training and evaluating the SVM, the data was split into equally-sized training and test sets. Due to the patient-specificity of the topic models, predictors were trained on a per-patient basis.

As a second point of evaluation, we assessed whether the topic models provided more information about the structure of sleep than the R&K stages. Due to the difficulty of experimentally validating such qualities, we evaluate the novelty of the results by visualizing the resulting topic mixtures and qualitatively comparing with the R&K standard.

We conducted all evaluations on the publicly available MIT-BIH polysomnographic database from Physionet, comprised of multimodal recordings from 14 patients with sleep apnea. Each recording contained a single channel of EEG, with recording durations ranging from 3 hours and 40 minutes to 6 and a half hours.

IV. EVALUATION

The derived topics demonstrate a clear visual concordance with the R&K standards. Figure 1 compares the generated model with the R&K annotations for several patient recordings. The top panels depict the model estimates for each window's distribution over topics, with each vertical strip corresponding to a 30 second time span. Each color in the mixture diagram represents a different topic. The size of a color band for a window indicates that topic's contribution to that window. The bottom panels show the R&K annotations using the same 30 second boundaries. The models for each patient were trained separately, and there is no correspondence between the topic coloration between the two patients.

Figure 1 indicates that the inferred model contains similar structure to the R&K stages. For the left patient, the black topic (bottom) corresponds to stage 2, with increases as the patient enters stage 2 sleep and decreases as they begin entering stages 3 and 4. The white topic (top) reflects deeper sleep (stages 3 and 4), dominating the recording from epochs 180 to 330. The second patient reflects similar structure, with a topic corresponding to the awake state, another reflecting light sleep, and a third reflecting deeper sleep.

The accuracies in Table I indicate that the topic mixtures can predict R&K stages within the range of inter-rater reliability, which has been estimated between 70 to 90% [11], achieving a mean accuracy of 70.1% and a median accuracy of 71.2%. Scores in this range by an automatic stager can be considered good performance as they achieve the same correspondence with the reference as another human annotator might. This demonstrates that despite the proposed method's discretization and unsupervised learning of structure, it preserves the information relevant to the current standard of sleep staging, and has the ability to convert from the derived topics back to the R&K standard.

Visual inspection reveals several properties that indicate greater expressive power than the sleep stages. First, the sleep topic model shows a continuous shift between states, reflected by gradual rather than abrupt transitions between

TABLE I

ACCURACY OF SVM MODELS TRAINED WITH TOPIC MIXTURE FEATURES WHEN PREDICTING R&K STAGES ON THE MIT-BIH DATASET

Recording ID	**01a**	**01b**	**02a**	**02b**	**03**	**04**	**14**	**16**	**32**	**37**	**41**	**45**	**48**	**59**	**60**	**61**	**66**
Accuracy	73.1	78.9	82.4	81.3	58.6	76.4	58.4	71.2	85.0	85.4	53.1	64.7	62.8	61.1	71.5	68.1	59.8

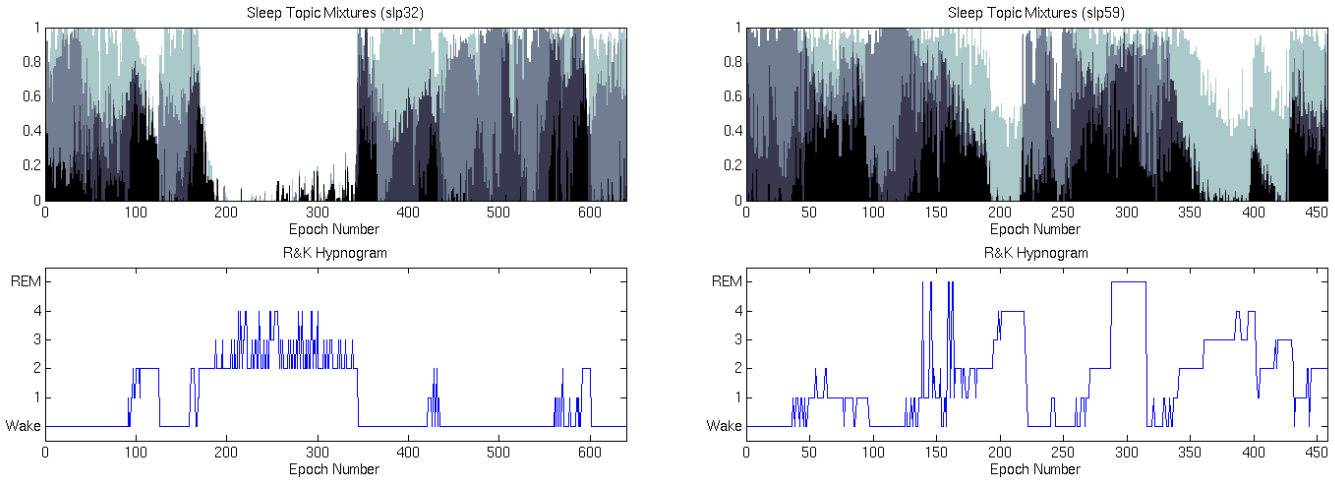

Fig. 1. Topic mixture diagrams (top) and R&K hypnograms (bottom) for patients slp32 (left) and slp59 (right).

different depths of sleep. This allows for better assessment of the relative depth of sleep, for example, where a light sleep topic increases during waking epochs. The relative smoothness of transitions is a property of the data as opposed to the model, as the exchangeability of documents in the LDA model means that there is no explicit relationship between adjacent windows as in an HMM.

An observation with regard to the second patient shows potential utility for sleep topics in going beyond the capabilities of R&K scoring. In the second patient's diagram, there are three long periods of stage 2 sleep, in epochs 160 to 200, 270 to 290, and 340 to 360. In the first and third instances, the patient progresses into deeper sleep (stages 3 and 4), while in the second instance they progress into REM. The topics capture this distinction well before the state transition, where the deeper sleep topic (top) is present and increasing in the first and third cases well before the transition begins, yet absent in the second. This indicates a difference in stage 2 sleep preceding the transitions, verifying the ability of the model to detect variations within a single R&K stage.

V. CONCLUSION

We present an unsupervised approach to identifying structure in sleep EEG recordings. The sleep topic model improves on the current standard for sleep analysis by providing an automated, data-driven algorithm for learning patient-specific sleep states. The approach relaxes the traditional assumption of discrete states, allowing for a more expressive model. The use of patient-specific models allows for better modeling of individual differences, and complements the current use of universal sleep staging systems. The resulting models are capable of representing standard sleep stages, while including additional information.

A clinical validation of the method by investigating the derived topics or the relationships between the derived models and sleep disorders is an area for future work, and currently under investigation. A possible extension to the presented methods could incorporate additional features, from other time-scales or additional physiological signals, to fully leverage the benefits of LDA.

REFERENCES

[1] J. Hossain and C. Shapiro, "The prevalence, cost implications, and management of sleep disorders: an overview," *Sleep and Breathing*, vol. 6, no. 2, pp. 85–102, 2002.

[2] A. Vgontzas, D. Liao, E. Bixler, G. Chrousos, and A. Vela-Bueno, "Insomnia with objective short sleep duration is associated with a high risk for hypertension," *Sleep*, vol. 32, no. 4, p. 491, 2009.

[3] M. Daley, C. Morin, M. LeBlanc, J. Gregoire, J. Savard, and L. Baillargeon, "Insomnia and its relationship to health-care utilization, work absenteeism, productivity and accidents," *Sleep Medicine*, vol. 10, no. 4, pp. 427–438, 2009.

[4] A. Rechtschaffen and A. Kales, "A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects," *US Department of Health, Education and Welfare, National Institute of Health Publ*, no. 204, 1968.

[5] J. Santamaria and K. Chiappa, *Electroencephalography of drowsiness*. Demos Medical Publishing, 1987.

[6] T. Penzel, K. Stephan, S. Kubicki, and W. Herrmann, "Integrated sleep analysis, with emphasis on automatic methods." *Epilepsy research. Supplement*, vol. 2, p. 177, 1991.

[7] J. Buckelmüller, H. Landolt, H. Stassen, and P. Achermann, "Trait-like individual differences in the human sleep electroencephalogram," *Neuroscience*, vol. 138, no. 1, pp. 351–356, 2006.

[8] A. Flexer, G. Gruber, and G. Dorffner, "A reliable probabilistic sleep stager based on a single eeg signal," *Artificial intelligence in Medicine*, vol. 33, no. 3, pp. 199–207, 2005.

[9] J. Lin, E. Keogh, S. Lonardi, and B. Chiu, "A symbolic representation of time series, with implications for streaming algorithms," in *proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery*. ACM, 2003, pp. 2–11.

[10] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *The Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003.

[11] Y. Kim, M. Kurachi, M. Horita, K. Matsuura, and Y. Kamikawa, "Agreement in visual scoring of sleep stages among laboratories in japan," *Journal of Sleep Research*, vol. 1, no. 1, pp. 58–60, 1992.