Dynamic Bayesian Networks for the Classification of Sleep Stages

Jiří Vomlel[1,2] and **Václav Kratochvíl**[1,2]
vomlel@utia.cas.cz, velorex@utia.cas.cz

[1]Institute of Information Theory and Automation
Czech Academy of Sciences

[2]Faculty of Management
University of Economics, Prague

Abstract

Human sleep is traditionally classified into five (or six) stages. The manual classification is time consuming since it requires knowledge of an extensive set of rules from manuals and experienced experts. Therefore automatic classification methods appear useful for this task. In this paper we extend the approach based on Hidden Markov Models by relating certain features not only to the current time slice but also to the previous one. Dynamic Bayesian Networks that results from this generalization are thus capable of modeling features related to state transitions. Experiments on real data revealed that in this way we are able to increase the prediction accuracy.

1 Introduction

Human sleep occurs in cycles. Each cycle lasts approximately 90 minutes. Typically, there are four or five cycles per night. Traditionally, sleep stages are scored into the following five stages of sleep: (W) Wakefulness, (REM) the stage named after "rapid eye movement" that occurs during this stage, and three Non-REM stages named (N1), (N2), and (N3). In Figure 1 we present an example of a real hypnogram from the dataset we used in our experiments.

Traditionally, a set of rules from a manual is used for sleep scoring. The rules for sleep scoring were standardized by [5]. An up-to-date manual is provided by the American Academy of Sleep Medicine, see [3]. However, it should be beneficent if the scoring system is based on probabilistic principles rather than on a predefined set of rules. Hidden Markov models (HMMs) can play this role.

In [4] the authors verify on synthetic data that Gaussian Observation HMMs (GOHMMs) can detect the state transitions and are thus a model well-suited for

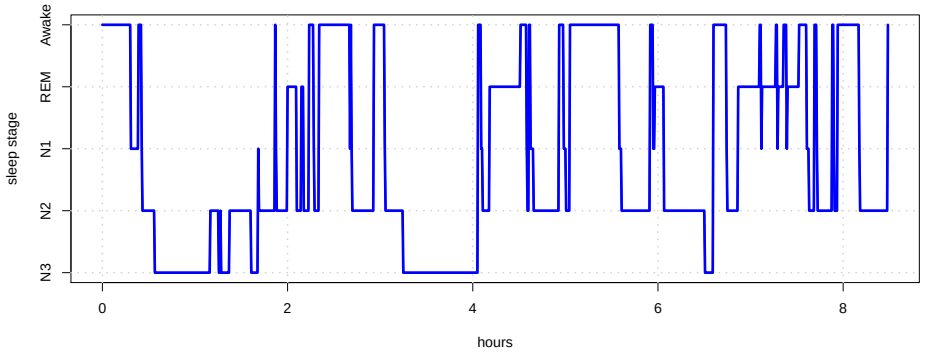

Figure 1: An example of a hypnogram.

the EEG analysis.In [2] a probabilistic continuous sleep stager based on Hidden Markov models is developed using only a single EEG signal. In the work of [1] an automatic diagnosis system based on HMMs is proposed to help clinicians in the diagnosis of sleep apnea syndrome.

In our paper we build on these works and use a generalization of HMMs – Dynamic Bayesian networks. This allows us to properly include features related to hidden states of two consecutive stages into the model. We used one such feature from our dataset – Arousal, which appears typically when a transition between sleep stages occurs. We will see that already treating a single feature properly (in a DBN model) helps us to significantly increase the prediction accuracy.

In this paper, first, we discuss probabilistic models suitable for the sleep classification task. We start with two versions of the Naive Bayes (NB) model in Section 2. In Section 3 we extend the NB model in two steps. First, the time factor is included into the model by considering the transition probabilities between sleep stages – in this way we get Hidden Markov Models (HMMs). Second, the HMMs are generalized by relaxing the Markov property. HMMs are generalized to Dynamic Bayesian Networks (DBNs) by including features related to transitions between two consecutive stages. We conclude the paper by experimental evaluation of the considered probabilistic models (in Section 4), by a summarizing our results, and by a discussion of possible future work (in Section 5).

2 Naive Bayes models

Probabilistic models offer an advantage of being objective by not relying on human scorers and being based on solid probabilistic principles rather than a predefined set of rules. A Naive Bayes (NB) model is an example of a simple probabilistic

model. It is a Bayesian network model that builds on the assumption of conditional independence of observed features given the state (class) variable. In the context of sleep analysis, it means that the observed attributes are assumed to be independent given the sleep stage. The model is static in the sense that features observed in time t are relevant for the classification of the sleep at time t only.

In Figure 2 we present an example of a Naive Bayes model. Variable Y represents the sleep stage at a given time t and variables $X_1, \ldots, X_5$ are features. Examples of features used in the sleep analysis are:

- the spectral density at given frequency ranges of the Electro Encephalogram (EEG) signal,
- Electro Oculogram (EOG), which identifies eye movements,
- Electro Myogram (EMG), which identifies muscle activity,
- Central Sleep Apnea,
- Snoring, etc.

Most of these features are continuous variables (in figures displayed as double circles) and some of them are discrete – often binary (in figures displayed as single circles).

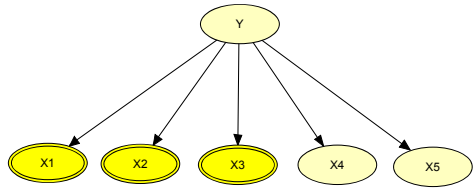

Figure 2: The structure of the Naive Bayes Model for clock time t.

For continuous features X a natural model for the conditional density $P(X = x|Y = y)$ seems to be the Conditional Gaussian distribution $\mathcal{N}(\mu_y, \sigma_y)$ where μ_y and σ_y are the mean and the standard deviation of feature X given the value y of parent variable Y. At the top of Figure 3 we can see the Gaussian density estimates for the delta power spectral density learned from real data.

In the experiments (reported in Section 4) it appeared that the CGDs did not lead to a good prediction accuracy. A better option appeared to be the Kernel Density Functions (KDFs) that can better fit the actual shape of the conditional density functions. At the bottom of Figure 3 we can see the KDFs for the delta power spectral density learned from real data. The density shape for the sleep class N3 seems surprising complex (two maxima) but KDFs lead to a significant improvement of accuracy on testing data (i.e. it does not seem to be an overfitting effect).

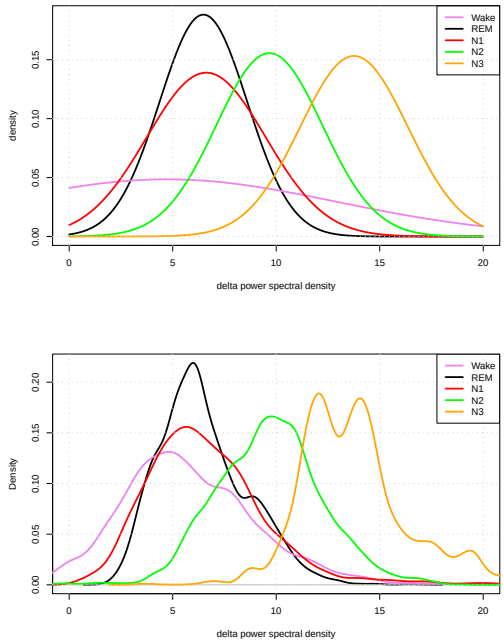

Figure 3: Gaussian density estimates (top) and Kernel Density Functions (bottom) for the delta power spectral density.

3 From Hidden Markov Models to Dynamic Bayesian Networks

At each clock time t, a Hidden Markov Model (HMM) consists of:

- an unobserved state variable Y_t taking a finite number of states. In the sleep analysis the unobserved variable will be the sleep stage and it will take states $\mathcal{Y} = \{Wake, N1, N2, N3, REM\}$ and
- a set of observed variables $X_{i,t}, i = 1, \ldots, k$ (e.g., the spectral density at a given frequency range of the EEG signal).

A new state y_{t+1} is entered based upon a transition probability distribution $P(Y_{t+1} = y'|Y_t = y)$ for $y, y' \in \mathcal{Y}$ which depends on the previous state Y_t (this is called the Markovian property). This allows to exploit the probabilistic dependence of successive sleep stages. Transitions between some of the sleep stages are much more likely than between others. For example, a transition from stage "wake" (W) directly to stage "deep sleep" (N3) is quite unlikely. After each transition is made,

an observation x is produced according to a conditional probability distribution $P(X_{i,t+1} = x|Y_{t+1} = y)$ which depends on the current state y of Y_{t+1} only. In Figure 4 we present and example of a structure of the two consecutive stages of a Hidden Markov Model.

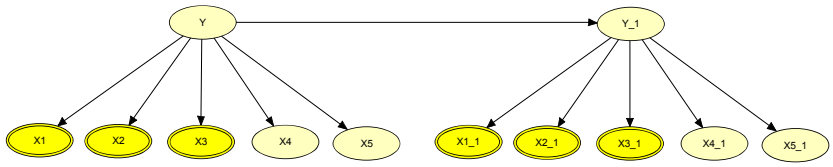

Figure 4: The structure of two consecutive stages of a Hidden Markov Model.

The conditional probability distributions $P(X_{t+1} = x|Y_{t+1} = y)$ are assumed to be stationary (i.e., they do not depend on time t) and can be defined (and learned) in the same way as the conditional probability distributions of the NB model as discussed in Section 2. We used the CDFs due to their better performance in the experiments on real data. The transition probability distribution $P(Y_{t+1} = y'|Y_t = y)$ is discrete and it is also assumed to be stationary. It is easily estimated from training data by normalizing the corresponding contingency table.

As we will see in the Section 4 the HMMs perform better than NB models. However, we conjectured that there is still a room for additional improvement since some of the features should not be treated as dependent on the current state only but also on the previous state. This is because some features are witnesses of state transitions. Therefore we include the model binary variables $Z_{j,t+1}$ that depend on state variables Y_{t+1} and Y_t. These distributions are again assumed to be stationary. They are estimated from training data by normalizing the corresponding contingency table.

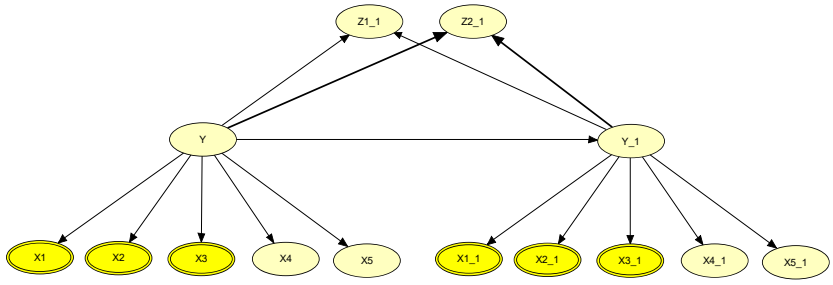

Figure 5: The structure of two consecutive stages of a Dynamic Bayesian Network.

In Figure 5 we present an example of a structure of the two consecutive stages of a Dynamic Bayesian Network. There are two state transition witness features –

denoted Z_1 and Z_2 in the figure. We will refer to them as transition features. Note that the Markovian property does not hold any more since the values of features $X_{i,t+1}$ are not independent of the past given the current state of Y_{t+1}. We can see that if a transition feature $Z_{j,t+1}$ is known a path from the past gets open.

The probabilistic inference in these DBN is very similar to the standard inference in the HMMs, where the Viterbi algorithm [6] is used. The difference is that in each time step we modify the transition matrix representing transition probability distribution $P(Y_{t+1}|Y_t)$ by multiplying it by a probability potential $P(Z_{j,t+1} = z_{j,t+1}|Y_{t+1}, Y_t)$ where $z_{j,t+1}$ is the state of the transition feature observed at time $t+1$. We do this multiplication for each transition feature. The computational process in time slice t is completed (as it is also done in the standard Viterbi algorithm) by selecting the most probable state y_t for each state y_{t+1} and by normalizing the distribution over the states of Y_{t+1}. Then a backward pass is performed to find a most probable configuration of state variables. See Algorithm 1.

Input: A DBN defined by the conditional probability distributions $P(Y_{t+1}|Y_t)$, $P(Z_{j,t+1}|Y_{t+1}, Y_t)$, $P(X_{i,t+1}|Y_{t+1})$, and $P(Y_1)$, feature evidence:
- $x_{i,t}$ for features $X_{i,t}, i = 1, \ldots, n_X, t = 1, \ldots, N$, and
- $z_{j,t}$ for transition features $Z_{j,t}, j = 1, \ldots, n_Z, t = 2, \ldots, N$.

Output: The most probable state values y_t for $t = 1, \ldots, N$

$S_1 \leftarrow P(Y_1) \cdot \prod_i P(X_{i,1} = x_{i,1}|Y_1)$;
for $t \leftarrow 2$ ***to*** N **do**
  for $y \in \mathcal{Y}$ **do**
    $R \leftarrow S_{t-1} \cdot P(Z_{j,t} = z_{j,t}|Y_t = y, Y_{t-1}) \cdot P(Y_t = y|Y_{t-1})$
     $\cdot P(X_{i,t} = x_{i,t}|Y_t = y)$;
    $T_t(y) \leftarrow \arg\max_{y' \in \mathcal{Y}} R(y')$;
    $S_t(y) \leftarrow \max_{y' \in \mathcal{Y}} R(y')$;
  end
  $S_t \leftarrow \frac{S_t}{\sum_{y'} S_t(y')}$;
end
$y_N \leftarrow \arg\max_{y \in \mathcal{Y}} S_N(y)$;
for $t \leftarrow N$ ***to*** 2 **do**
  $y_{t-1} \leftarrow T_t(y_t)$;
end

Algorithm 1: The inference algorithm for solving the sleep analysis DBN

4 Experiments

We learned our models on a training dataset that consisted of 37 hypnograms. The models were tested on a (testing) dataset that also consisted of 37 (different) hypnograms. Altogether we used 46 features. In the DBN model we used the

Arousal feature as the transition feature while in all other models it was treated as a standard feature related to the current state only. The results of experiments are summarized in Table 1 where the methods are presented in the ascending order given by their accuracy.

Table 1: The average accuracy of the tested methods on testing data.

Method	Accuracy
No Information Rate	41.57%
Naive Bayes with CGDs	49.13%
Philips Respironics	57.43%
Naive Bayes with KDFs	65.06%
Hidden Markov Model	65.71%
Dynamic Bayesian Network	67.02%

In Figure 6 we compare the accuracy of the proposed methods on the testing dataset. Each point in the plot corresponds to one hypnogram from the testing dataset. To see whether the methods differ significantly we performed the Wilcoxon signed rank test for the pairs of methods results presented in Figure 6. We can conclude that:

- The NB model with Kernel Density Functions achieved significantly better accuracy than the NB model with Conditional Gaussian Distributions with the p-value = 5.093e-10
- Hidden Markov Models achieved significantly better accuracy than the NB model with Kernel Density Functions with the p-value = 7.577e-05.
- DBNs achieved significantly better accuracy than HMMs with the p-value = 1.533e-05.
- DBNs achieved significantly better accuracy than Philips Respironics with the p-value = 0.001416

See Figure 7 for the comparisons of the hypnograms predicted by tested methods with the expert for a selected hypnogram of one person. Though, the predicted hypnograms may look very similar, at a closer look, we can see differences that imply different accuracy on the testing dataset. For example, if we have a look at the first two hours of the patient's sleep we can see that the HMM and DBN leads to less oscillations than NB. The DBN further improves the fit by widening the third period of the N2 sleep stage. The DBN also further reduces some oscillations in the latter periods of the sleep. The accuracy for this patient hypnogram was: 77.13% (for NB), 77.53% (for HMM), and 79.98% (for DBN). The Philips Respitronics has the accuracy of 52.50% for this patient.

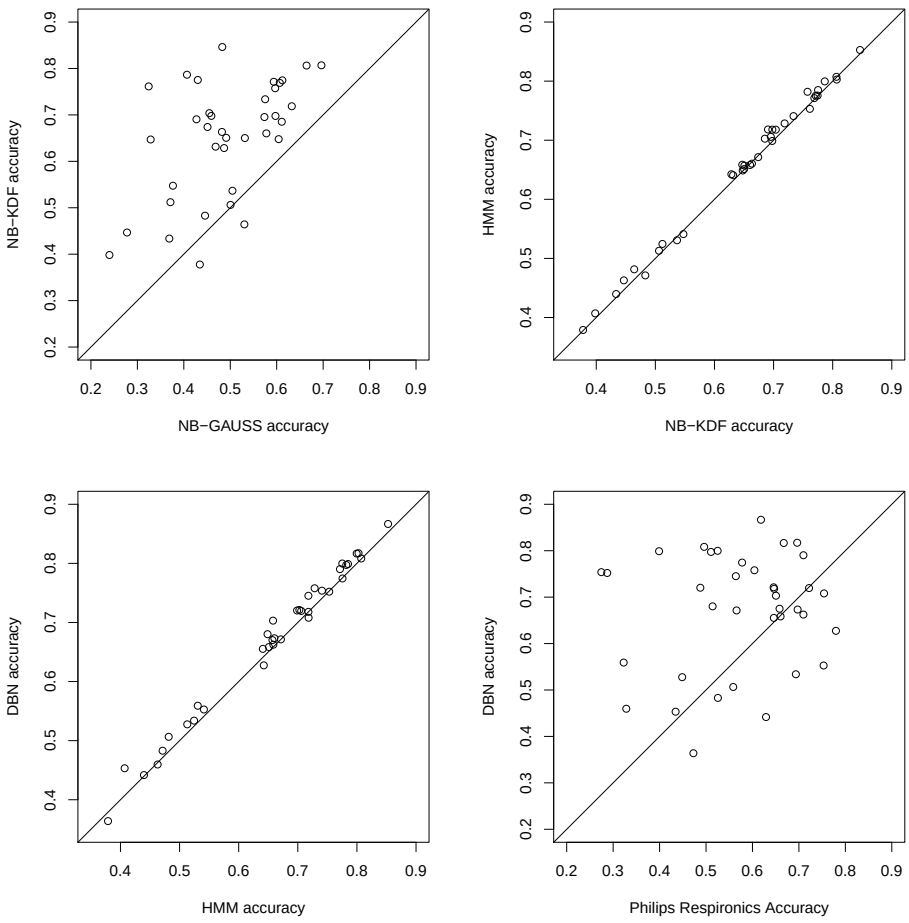

Figure 6: Comparisons of the methods' accuracy on real data.

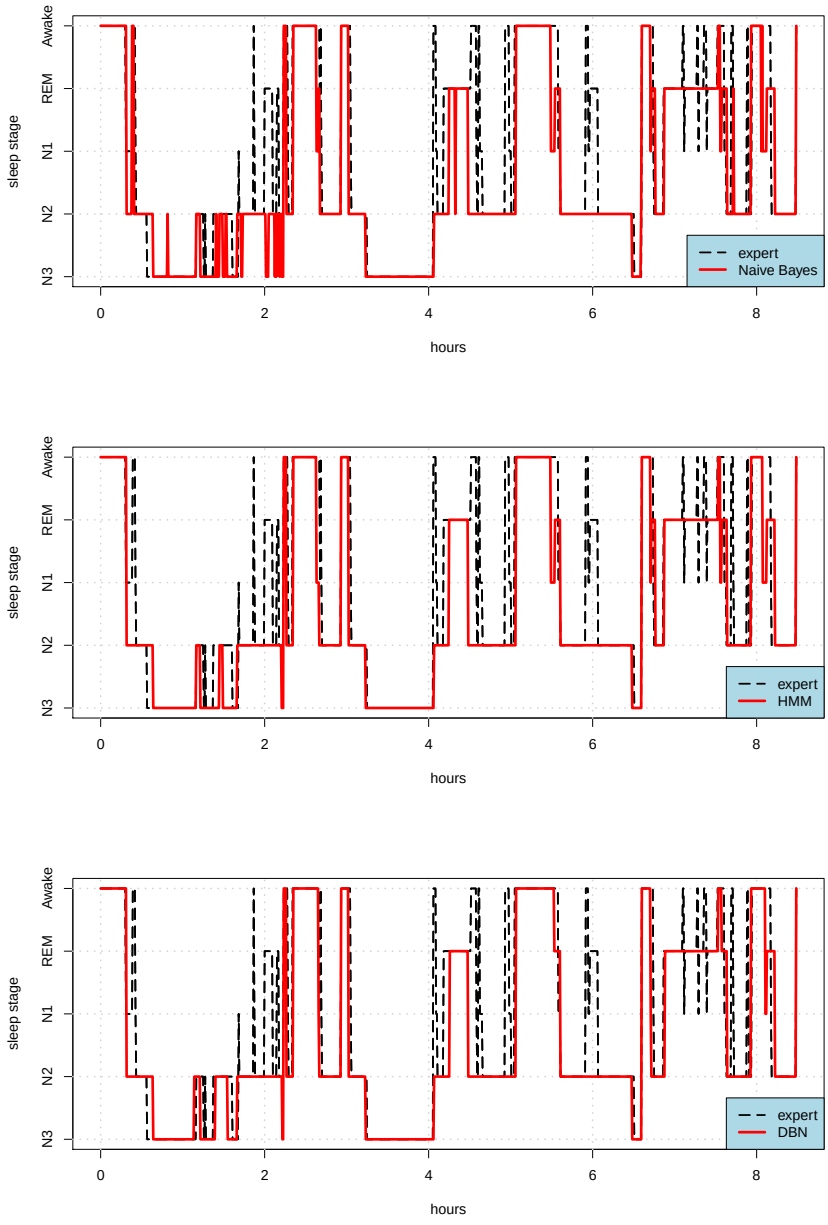

Figure 7: Comparisons of the predicted hypnograms with the expert hypnogram for one selected patient.

5 Conclusions and Future Work

In the problem of the classification of sleep stages Hidden Markov models and Dynamic Bayesian networks have achieved a better accuracy. They also better filter oscillations in hypnograms than Naive Bayes models. It is because they take into their consideration the sleep stage in the previous time step which is closely related to the sleep stage in the current time step. Using the Arousal feature we have demonstrated that the implementation of transition-related features in a Dynamic Bayesian Network helps to further improve the predictions accuracy. Other candidate features of this type are spindles, K-complexes, etc. These were not available in our dataset.

A model we did not use in our experiments but with which we would also like to compare in our future work is a Recurrent Neural Network. We also may consider to include time factors into our models - as are the time spent in a sleep stage and the total time spent in a sleep.

Another future goal is to use the proposed methods for automatic detection of sleep related disorders, which is of great interest of medical doctors. Yet another interesting research direction might be opened by lifting the assumption of five predefined sleep stages and by considering the variable representing the sleep stages as a truly hidden variable with an unknown number of states.

Acknowledgements

This work was supported by the Czech Science Foundation (projects 16-12010S and 17-08182S). This work benefited from a personal communication with Hynek Kružík and Nela Grimová.

References

[1] T. Al-Ani, Y. Hamam, R. Fodil, F. Lofaso, and D. Isabey. Using hidden Markov models for sleep disordered breathing identification. *Simulation Modelling Practice and Theory*, 12:117–128, 2004.

[2] A. Flexer, G. Gruber, and G. Dorffner. A reliable probabilistic sleep stager based on a single EEG signal. *Artificial Intelligence in Medicine*, 33:199–207, 2005.

[3] C. Iber, S. Ancoli-Israel, A. L. Chesson, and S. F. Quan. *The AASM Manual for the Scoring of Sleep and Associated Events.* American Academy of Sleep Medicine, Westchester, Illinois, 2007.

[4] W. D. Penny and S. J. Roberts. Gaussian observation Hidden Markov models for EEG analysis. Technical Report Technical Report, TR-98-12, Imperial College, London, 1998.

[5] A. Rechtschaffen and A. Kales, editors. *A manual of standardized terminology, techniques and scoring system for sleep stages of human subjects.* Public Health Service, U.S. Government Printing Office, Washington, D.C., 1968.

[6] A. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory*, 13(2):260–269, 1967.