Heterogeneous Hidden Markov Models for Sleep Activity Recognition from Multi-Source Passively Sensed Data

Fernando Moreno-Pino
María Martínez-García
Pablo M. Olmos
Antonio Artés-Rodríguez

Signal Processing and Learning Group, Dept. of Signal Theory and Communications
Universidad Carlos III de Madrid, Spain

Abstract

Psychiatric patients’ passive activity monitoring is crucial to detect behavioural shifts in real-time, comprising a tool that helps clinicians supervise patients’ evolution over time and enhance the associated treatments’ outcomes. Frequently, sleep disturbances and mental health deterioration are closely related, as mental health condition worsening regularly entails shifts in the patients’ circadian rhythms. Therefore, Sleep Activity Recognition constitutes a behavioural marker to portray patients’ activity cycles and to detect behavioural changes among them. Moreover, mobile passively sensed data captured from smartphones, thanks to these devices’ ubiquity, constitute an excellent alternative to profile patients’ biorhythm. In this work, we aim to identify major sleep episodes based on passively sensed data. To do so, a Heterogeneous Hidden Markov Model is proposed to model a discrete latent variable process associated with the Sleep Activity Recognition task in a self-supervised way. We validate our results against sleep metrics reported by clinically tested wearables, proving the effectiveness of the proposed approach.

Keywords: Unsupervised learning, Semisupervised learning, Human Activity Recognition, Behavioral markers

1. Introduction

Mental illnesses affect as much as the 2% of world’s population (WHO, 2019) and they are estimated to be the cause of about 14% of global diseases (Prince et al., 2007). Mental disorders lead to quality of life deterioration along with social disabilities that burden patients’ professional and social lives (Rapaport et al., 2005). Therefore, monitoring patients that suffer from mental illnesses is vital to detect trend changes that may indicate behavioural shifts. Historically, psychiatric patients’ mental state monitoring has been assessed via questionnaires, patients’ reports, and periodic appointments, which entail long periods before detecting any possible risk. Recent technological advances allow real-time patients monitoring with tools like the smartphone-based active Ecological Momentary Assessment (EMA) (Porras-Segovia et al., 2020). Nevertheless, collected data via electronic EMA suffer from patients’ subjective insight and presumes patients’ ability and interest to fill in the required information (Fischer and Kleen, 2021). Objective assessment methods can be achieved through passive data collected from patients’ smartphones, using information as the visited locations via GPS, smartphone and social media interaction, etc. The ubiquity of smartphone devices allows Human Acti-
vity Recognition (HAR) methods (Jobanputra et al., 2019; Rios-Munoz et al., 2020) to track patients in a non-invasive way, creating a personal profile for each user. To do so, the variety of smartphones' available sensors must be merged into relevant indicators of patients’ daily life. These indicators constitute biomarkers that represent patients' state in real-time and can be used to profile their evolution, allowing real-time monitoring and detection of behavioural shifts. Recently, several studies have identified relations between mental health condition worsening and sleep disruption in psychiatric patients (Tseng et al., 2019; Littlewood et al., 2019; Freeman et al., 2020). Considering the available data through patients’ smartphones, sleep estimation arises as a natural biomarker to integrate smartphones’ data into a single behaviour metric that characterizes each individual’s circadian rhythm over time. Furthermore, a wide range of studies have reported that wearable devices show high reliability while detecting asleep-awake states in different populations, consistently reporting good estimators of the sleep activity (de Zambotti et al., 2018; Chinoy et al., 2021), which allows them to be used for validating predictions performed by Sleep Activity Recognition (SAR) models, as they can assess sleep quality during long time windows.

2. Sleep Measurement Techniques

This work aims to identify major sleep episodes based on passively sensed behavioural data, obtaining objective behaviour markers that could be used to recognise major disturbances in psychiatric patients. Among traditional sleep assessment methods, Polysomnography (PSG) stands as the most reliable technique. Nonetheless, PSG requires medical assistance for its application and sleep studies need to be recorded in an ambulatory setting (Smith et al., 2020). Besides, it is an expensive procedure that does not provide a continuous evaluation of patients’ sleep as it is required to assess their behaviour evolution. Sleep questionnaires and diaries arise as a natural alternative when trying to monitor sleep evolution constantly. Regardless, we discard them as they do not express an objective sleep evaluation (Hennig et al., 2017). In these conditions, wearable devices appear as an excellent alternative: they are not expensive and they collect daily information in a non-intrusive way. Several studies have shown that, even when they are not sufficiently accurate to be used in clinical settings, mainly due to poor specificity and low accuracy, they can be relied on for detecting asleep-awake states (Moreno-Pino et al., 2019). Nonetheless, the percentage of users who possess these devices is minimal, implying that we cannot rely on wearable devices to assess patients’ sleep continuously. Therefore, smartphones constitute the most reliable source of data: most patients have one, they do not require an additional investment, and most people carry them along most of the day; the ubiquity of these devices makes them the perfect passive assessment. However, mobile health data that can be collected this way usually entails several problems: missing data is quite common, signals are typically corrupted by noise, the different number of sensors that smartphones use imply data heterogeneity, and main sleep trends must be detached from false positives while making the predictions. Therefore, the proposed model must be capable of dealing with these problems. Several frameworks are available in the literature to obtain sleep estimates from smartphone data (Gautam et al., 2015; Eichi et al., 2021), as well as certain HMM implementations that handle mixed discrete and continuous observations (Epaillard and Bouguila, 2015; Missaoui et al., 2013). Anyhow, none of these models can simultaneously deal with all the
complex requirements previously mentioned, making them not suitable for creating behavioural markers in the considered scenario. Furthermore, the literature lacks models validated with clinical patients and tested in large samples (Aledavood et al., 2019). For these reasons, we propose our own method for the SAR problem, detailed in Section 4.

3. Passively Sensing Mobile Data

In this section, we aim to specify the details of the dataset used. As previously stated, smartphones’ data constitute a tool for patient monitoring due to their omnipresence, reporting objective measures of patients’ circadian rhythm that can be used as indicators for detecting behavioural shifts. A data gathering system (eB2, 2019) was used to collect passively sensed mobile data from patients’ smartphones under clinicians’ supervision. The available population consisted of a set of 57 patients who suffered from different mental health conditions. For 10 of those patients, 386 days of data with associated wearables’ sleep metrics were available. Data from this subset of patients was exclusively used for validation purposes and did not take part during the training process. Demographic information for this subset was not available. The rest 47 patients, 38 females and 9 males, had an average age of $43 \pm 15$ years. For this second subset of patients, 647 days of data were collected. Thus, considering both sets, 1033 days of data were collected. Thus, we should clarify that train/test split was done solely depending on the availability of wearable devices’ data: those patients who possessed wearable devices were not used during the training phase. Instead, their corresponding data was kept for testing purposes, as wearable devices’ reported sleep metrics allow us to assess the quality of the model’s predictions. This is possible as the proposed model operates in an unsupervised way, which entails that no labelled data is required during training, allowing us to use all wearable devices’ data for evaluation purposes. Data acquisition was conducted periodically, collecting data from different sensors, i.e. actigraphy, light, smartphone usage (a binarized feature that indicates if the patient is using or not the smartphone), and pre-computed metrics, i.e. the daily amount of steps, which we binarize for each time-slot. Data’s granularity was 10 minutes; hence 144 observations per day were available in absence of missing data. Table 1 summarizes the smartphones’ used sensors and Figure 1 shows the collected data’s structure for one of the patients. The models developed to extract relevant sleep trends from this real-world data (incomplete, noisy, and heterogeneous) need to consider its structure, as it encloses different challenges while building optimal methods capable of dealing with these problems.

4. Sleep Activity Recognition

Passively sensing mobile data’s characteristics, enumerated in Sections 2 and 3, demarcates the features that we demand from the proposed model. Firstly, it needs to be able to manage data heterogeneity, handling gaussian and categorical observations at the same time. Secondly, it needs to perform inference on missing data, critical while dealing with healthcare datasets. Finally, it needs to allow semi-supervised training, granting users the possibility to fix certain observation probabilities, therefore manually design-
E.g., fixing to zero the probability of observing smartphone usage for those hidden states that we aim to interpret as sleeping activity. Fulfilling these requirements, we present a model easy to train and with a low number of parameters, which allows it to be implemented in a production setting and to be used by clinicians as a tool to monitor patients’ circadian rhythm, unveiling a powerful mechanism to detect behavioural changes.

To merge the heterogeneous input features that define the collected data into a SAR indicator, capable of detecting different activity levels along the day and identifying sleep episodes among them, we require the proposed model to be capable of projecting the high dimensional input features into a lower dimensional space. To do so, latent variable models arise as a natural alternative (Bishop, 1998). Considering the time dependence of our data, Hidden Markov Models (HMMs) (Rabiner, 1989) constitute a natural way of modelling a discrete latent process associated with the SAR task. HMMs are generative models that assume that the phenomena being modelled is a Markov process (Ethier and Kurtz, 2009) with associated hidden states. HMMs’ objective is to learn about these hidden states sequence, denoted $S = \{s_1, s_2, \ldots, s_T : s_t \in \{1, \ldots, I\}\}$, with $t = \{1, 2, \ldots, T\} \in \mathbb{N}$, by means of the generated observations, $Y = \{y_1, y_2, \ldots, y_T : y_t \in \mathbb{R}^M\}$. To do so, they use an observation model $p(y_t|s_t)$. Therefore, each observation $y_t$ depends exclusively on its associated state $s_t$. Here, we aim to use this hidden states sequence $S$ as the low dimensional representation of the input features. Therefore, each possible state of the HMM is associated with different activity levels, being sleep episodes one of them.

Nevertheless, standard HMM implementations use Multinomial and Gaussian observation models, depending on the probability distribution chosen to model the emission probabilities $p(y_t|s_t)$. Due to the heterogeneity of the collected data, neither of these options would fit our requirements. While some of the used features, like the actigraphy and light sensors, can be naturally modelled via Gaussian distributions, others, as the app usage and steps count (binarized and converted into an movement indicator), are better modelled with a Bernoulli distribution. To manage the heterogeneity of our input sources, we developed a variation of ordinary HMMs which we refer to as Heterogeneous Hidden Markov Model (HHMM).
The proposed HHMM can simultaneously manage categorical and continuous data thanks to its observation model, visible in Fig. 2, which is responsible for jointly modelling the continuous and discrete observations, $y_t$ and $l_t$. The joint distribution of the proposed model is stated in Eq. 1, where $N \in \mathbb{Z}$ represents the number of sequences. More information about the HHMM, its parameters inference procedure, and how it performs inference on missing data can be found in the Appendixes A and B.

$$p(S, Y, L) = \prod_{n=1}^{N} \left( \prod_{t=2}^{T_n} p(s^n_t | s^n_{t-1}) \right) \left( \prod_{l=1}^{T_n} p(y^n_l | s^n_l) \right) \left( \prod_{t=1}^{T_n} p(l^n_t | s^n_t) \right)$$ (1)

5. Experiments

We will now train the proposed model and evaluate its performance in comparison with different baseline methods. We should remark that because of the dataset used, where just a relatively small portion of data has the corresponding labels indicating the sleep activity of the patients, as stated in Section 3, it is not feasible to train supervised methods due to this lack of labelled data.

HHMM’s parameters are trained via the Baum-Welch algorithm using a subset the 47 patients with 647 sequences of daily data. The remainder 10 patients are retained for model evaluation purposes. The optimal number of hidden states is selected using two different model selection methods, the Bayesian Information Criteria (BIC) and the Akaike Information Criterion (AIC). For further details regarding the selection criteria for the optimal number of hidden states, as well as its projection into a binary output (awake or asleep), we refer the reader to Appendix C. Once the optimal model’s parameters are obtained, we proceed to evaluate the accuracy of the model in the task of detecting sleep episodes. To do so, we use the subset of 10 patients with 386 days of data for whom associated wearables’ sleep metrics were available, reporting the sleep state of the patients as awake, light sleep, deep sleep or REM phase. We binarize this output into two categories: awake or asleep, so we can directly compare them with HHMM’s final binary predictions.

In order to illustrate HHMM performance, we include two baseline models for comparison purposes. The first of them is a K-means algorithm that cluster each datapoint as asleep/awake. Two variations of this K-means model are considered, each of them uses different missing data imputation techniques in order to cluster each datapoint. The first imputation method (Imputation Method 1 in Table 2) uses the mean to infer the missing values on the continuous features of the dataset, and it sets to zero the missing observations on the binary features. The second imputation method (Imputation Method 2 in Table 2) still infers continuous missing values using the mean, but it substitutes missing observations on the binary features with the most common values (per sequence). Besides the K-means algorithm, the second baseline considered is a Gaussian Mixture Model (GMM), which also uses the two previously explained imputation techniques to deal with missing values.
| Method                               | Accuracy      | Specificity    | Sensitivity    |
|-------------------------------------|---------------|----------------|----------------|
| Random Classifier (Most Frequent)   | 0.6798 ± 0.043 | 0.9973 ± 0.0510 | 0.0026 ± 0.0511 |
| Random Classifier (Uniform)        | 0.4967 ± 0.0426 | 0.5011 ± 0.0502 | 0.4983 ± 0.0762 |
| K-means                             |               |                |                |
| Imputing Method 1                   | 0.7393 ± 0.1156 | 0.6338 ± 0.1651 | 0.9641 ± 0.0324 |
| Imputing Method 2                   | 0.7386 ± 0.1167 | 0.6356 ± 0.1672 | 0.9587 ± 0.0767 |
| GMM                                 |               |                |                |
| Imputing Method 1                   | 0.5126 ± 0.1327 | 0.2968 ± 0.1926 | 0.9717 ± 0.0416 |
| Imputing Method 2                   | 0.7386 ± 0.1167 | 0.6356 ± 0.1672 | 0.9587 ± 0.0767 |
| HHMM                                | 0.8780 ± 0.0804 | 0.9052 ± 0.0779 | 0.8207 ± 0.2382 |

Table 2: Experiments results.

Moreover, besides these two baselines, two dummy classifiers are tested on the labelled dataset. The first of these dummy classifiers follows a “most frequent” approach, always returning the most frequent class label. The second one is “uniform”, as it generates predictions uniformly at random from the list of unique classes observed. The inclusion of these two dummy classifiers provides a quantification of the label distribution of the dataset, which helps to understand class imbalances, as the portion of time the patients are asleep is considerable smaller than the time they are awake.

As done with the HHMM, we train the unsupervised baseline methods in the 47 patients within the training set. The trained methods are evaluated then in the cohort of 10 patients with associated labelled wearables’ data. The results reported by these baselines, together with HHMM’s performance, are collected in Table 2. As these results show, no baseline methods report better performance than the proposed algorithm in terms of accuracy. Regarding specificity and sensitivity, some of the utilised baselines report better results if we consider them separately, due to class imbalances, but the HHMM report the best trade-off between both metrics. For illustrative examples of the HHMM’s predictions, we refer the reader to Figures 4 and 5 of the Appendix C, which show two examples of the HHMM’s reported predictions for different patients.

6. Conclusion

We introduce a new approach that performs Sleep Activity Recognition from passively sensed data, providing an unsupervised method that can operate in realistic scenarios and produces sleep activity indicators. Accordingly, the model comprises a tool that can be used by clinicians to monitor patients’ circadian rhythm.

A discrete latent variable model with temporal dependences is used to monitor patients’ activity cycles, integrating smartphones’ data into a single behaviour metric. The proposed model can deal with passively sensed data associated problems: corrupted signals, heterogeneity, and missing values. We validated our results against wearables’ reported sleep metrics, proving the effectiveness of the proposed approach. Therefore, Heterogeneous Hidden Markov Models comprise a powerful tool to supervise patient’s evolution in real-time, providing a behavioural marker to portray patients’ activity cycles and to detect behavioural changes.

References

Talayeh Aledavood, John Torous, Ana Maria Triana Hoyos, John A Nashlund, Jukka-Pekka Onnela, and Matcheri Kesha-
HHMMs for Sleep Activity Recognition from Multi-Source Passively Sensed Data

van. Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Current psychiatry reports, 21(7): 1–9, 2019.

Christopher M Bishop. Latent variable models. In Learning in graphical models, pages 371–403. Springer, 1998.

Evan D Chinoy, Joseph A Cuellar, Kirbie E Huwa, Jason T Jameson, Catherine H Watson, Sara C Bessman, Dale A Hirsch, Adam D Cooper, Sean PA Drummond, and Rachel R Markwald. Performance of seven consumer sleep-tracking devices compared with polysomnography. Sleep, 44(5):zsaa291, 2021.

Massimiliano de Zambotti, Aimee Goldstone, Stephanie Claudatos, Ian M Colrain, and Fiona C Baker. A validation study of fitbit charge 2™ compared with polysomnography in adults. Chronobiology international, 35(4):465–476, 2018.

eB2. Evidence-based behavior, 2019. URL "https://eb2.tech".

Habiballah Rahimi Eichi, Garth Coombs, Constanza M Vidal Bustamante, Jukka-Pekka Onnela, Justin T Baker, and Randy L Buckner. Dpsleep: Open-source longitudinal sleep analysis from accelerometer data. bioRxiv, 2021.

Elise Epaillard and Nizar Bouguila. Hybrid hidden markov model for mixed continuous/continuous and discrete/continuous data modeling. In 2015 IEEE 17th International Workshop on Multimedia Signal Processing (MMSP), pages 1–6. IEEE, 2015.

Stewart N Ethier and Thomas G Kurtz. Markov processes: characterization and convergence, volume 282. John Wiley & Sons, 2009.

Florian Fischer and Sina Kleen. Possibilities, problems, and perspectives of data collection by mobile apps in longitudinal epidemiological studies: scoping review. Journal of medical Internet research, 23(1):e17691, 2021.

Daniel Freeman, Bryony Sheaves, Felicity Waite, Allison G Harvey, and Paul J Harrison. Sleep disturbance and psychiatric disorders. The Lancet Psychiatry, 7(7):628–637, 2020.

Alvika Gautam, Vinayak S Naik, Archie Gupta, SK Sharma, and K Sriram. An smartphone-based algorithm to measure and model quantity of sleep. In 2015 7th International Conference on Communication Systems and Networks (COMSNETS), pages 1–6. IEEE, 2015.

Timo Hennig, Katarina Krkovic, and Tania M Lincoln. What predicts inattention in adolescents? an experience-sampling study comparing chronotype, subjective, and objective sleep parameters. Sleep medicine, 38:58–63, 2017.

Charmi Jobanputra, Jatna Bavishi, and Nishant Doshi. Human activity recognition: A survey. Procedia Computer Science, 155:698–703, 2019.

Donna L Littlewood, Simon D Kyle, Lesley-Anne Carter, Sarah Peters, Daniel Pratt, and Patricia Gooding. Short sleep duration and poor sleep quality predict next-day suicidal ideation: an ecological momentary assessment study. Psychological medicine, 49(3):403–411, 2019.

Oualid Missaoui, Hichem Frigui, and Paul Gader. Multi-stream continuous hidden markov models with application to landmine detection. EURASIP Journal on Advances in Signal Processing, 2013(1):1–23, 2013.
Fernando Moreno-Pino, Alejandro Porras-Segovia, Pilar López-Esteban, Antonio Artés, and Enrique Baca-García. Validation of fitbit charge 2 and fitbit alta hr against polysomnography for assessing sleep in adults with obstructive sleep apnea. *Journal of Clinical Sleep Medicine*, 15 (11):1645–1653, 2019.

Fernando Moreno-Pino, Emese Sőkei, Pablo M Olmos, and Antonio Artés-Rodríguez. Pyhmm: A python library for heterogeneous hidden markov models. *arXiv preprint arXiv:2201.06968*, 2022.

Alejandro Porras-Segovia, Rosa María Molina-Madueño, Sofian Berrouiguet, Jorge Lopez-Castroman, María Luisa Barrigón, María Sandra Pérez-Rodríguez, José Heliodoro Marco, Isaac Díaz-Olíván, Santiago de León, Philippe Courtet, et al. Smartphone-based ecological momentary assessment (ema) in psychiatric patients and student controls: A real-world feasibility study. *Journal of affective disorders*, 274:733–741, 2020.

Martin Prince, Vikram Patel, Shekhar Saxena, Mario Maj, Joanna Maselko, Michael R Phillips, and Atif Rahman. No health without mental health. *The lancet*, 370(9590):859–877, 2007.

Lawrence R Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

Mark Hyman Rapaport, Cathryn Clary, Rana Fayyad, and Jean Endicott. Quality-of-life impairment in depressive and anxiety disorders. *American Journal of Psychiatry*, 162(6):1171–1178, 2005.

Gonzalo R Ríos-Muñoz, Fernando Moreno-Pino, Nina Soto, Pablo M Olmos, Antonio Artés-Rodríguez, Francisco Ferández-Avilés, and Angel Arenal. Hidden markov models for activity detection in atrial fibrillation electrograms. In *2020 Computing in Cardiology*, pages 1–4. IEEE, 2020.

Claire Smith, Barbara Galland, Rachael Taylor, and Kim Meredith-Jones. Actigraph gt3x+ and actical wrist and hip worn accelerometers for sleep and wake indices in young children using an automated algorithm: validation with polysomnography. *Frontiers in psychiatry*, 10:958, 2020.

Wen-Che Tseng, Yun-Chieh Liang, Mei-Hsin Su, Yun-Ling Chen, Hao-Jan Yang, and Po-Hsiu Kuo. Sleep apnea may be associated with suicidal ideation in adolescents. *European child & adolescent psychiatry*, 28 (5):635–643, 2019.

WHO. Mental disorders key facts, 2019. URL "https://www.nrel.gov/grid/solar-power-data.html".

### Appendix A. Parameters Inference

Heterogeneous Hidden Markov Models, as classic HMMs, present three inference problems that must be solved to deliver an useful application.

The first of these problems is related with obtaining the probability of the observed variables, $Y$ and $L$, given the model parameters, $\theta = \{A, B, D, \pi\}$, for the considered time-steps $t = \{1, 2, \ldots, T\}$, i.e., to calculate the probability $p(Y, L|\theta)$.

The second problem consists of determining the optimal hidden states path that better explains the observed data. This can be achieved by using the Forward-Backward algorithm, calculating $p(s_t|y_1, l_t)$ each timestep, or through the Viterbi algorithm, which maximizes the probability of the hidden states sequences by considering all time-steps $t = \{1, 2, \ldots, T\}$, i.e., calculating $p(S|y_t, l_t)$.

Finally, the third problem consists of determining the optimal parameters $\theta$
that maximize the conditional probability $p(Y, L | \theta)$. The Baum-Welch algorithm, a special case of the Expectation-Maximization (EM) algorithm, can be used for obtaining this optimal parametrization. The joint distribution required for this third task is expressed in Eq. 2. For further information regarding these three problems and their solution, we refer the readers to Rabiner (1989).

$$p(S, Y, L) = \prod_{n=1}^{N} \left( \prod_{t=2}^{T_n} p(s^n_t | s^n_{t-1}) \right) \left( \prod_{t=1}^{T_n} p(y^n_t | s^n_t) \right) \left( \prod_{t=1}^{T_n} p(l^n_t | s^n_t) \right)$$ (2)

Appendix B. Missing Data Inference

As Table 1 proves, missing data is relatively common on passively sensing data. These missing observations can be either full missing, where none of the sensors hold a value, or partially missing, where specific sensors exhibit missing values. The model differentiates these two different scenarios and handles both of them through modifications in the Baum-Welch algorithm.

The first scenario, full missing data, where all the features are non observable at a specific time instant, is straightforward solved by the Baum-Welch algorithm, which substitutes the missing values of each feature with its average per each hidden state.

In order to illustrate the second scenario, where partial missing data is present among the observed features, we assume for simplicity’s sake two jointly gaussian features at each time instant, $x = (x_1, x_2)$, with means:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$ (3)

and covariance matrix:

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix},$$ (4)

The inverse of this covariance matrix can be defined as follows:

$$\Lambda = \Sigma^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix},$$ (5)

We can obtain the marginals of both features as:

$$p(x_1) = N(x_1 | \mu_1, \Sigma_{11}),$$
$$p(x_2) = N(x_2 | \mu_2, \Sigma_{22}),$$ (6)

which finally allows us to compute the posterior of the missing values conditioning on the observed ones:

$$p(x_1 | x_2) = N(x_1 | \mu_{1|2}, \Sigma_{1|2}),$$ (7)

where

$$\mu_{1|2} = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (x_2 - \mu_2)$$
$$= \mu_1 - \Lambda_{11}^{-1} \Lambda_{12} (x_2 - \mu_2)$$ (8)

and

$$\Sigma_{1|2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} = \Lambda_{11}^{-1}.$$(9)

Appendix C. Experiments

Figure 3 shows BIC and AIC evolution for different numbers of hidden states. $N = 6$ was selected as the optimal value, with complexity increasing for a higher number of states. The mapping from the six hidden states to the binary classification (asleep/awake) is done by taking advantage of the capability of our proposed model to train in a semi-supervised way. Semi-supervised training within our framework allows users to manually design certain aspects of the model’s behaviour, which is especially important when training models from smartphones’ data. This allows users to fix certain discrete probability distributions. In our case, we fix the probability of sleeping to zero when observing that the patient is using the smartphone.
for those hidden states we aim to interpret as sleeping (while letting the model infer the observation probabilities for the rest of the features). This makes the mapping from the original six hidden states to the binary classification straightforward.

Figures 4 and 5 show two examples of the HHMM’s reported predictions for different patients. The first sequence suffered from scarce missing data, gathered in the continuous features, while the second sequence exhibited significant amounts of missing values for most of the features. HHMM’s prediction for the first sequence correctly integrated the available sources and the model could detect awake periods that the wearable device missed, as the app usage feature proves. For the second sequence, the model failed predicting as sleep activity some periods distant from the primary sleep trend due to the lack of available data. This behavior could be corrected by applying post-processing techniques to the model’s predictions.