Mixture of Hidden Markov Models for Pattern-Recognition of Accelerometer data

Jonathan Y. Bernard [1,2,3] & Marie Du Roy de Chaumaray [4] & Matthieu Marbac [4] & Fabien Navarro [4]

[1] *Inserm, Centre de Recherche en Épidémiologie et StatistiqueS (CRESS), Équipe de recherche sur les déterminants précoces de la santé (EAROH), Villejuif, France jonathan.bernard@inserm.fr*
[2] *Université Paris Descartes, Villejuif, France*
[3] *Singapore Institute for Clinical Sciences (SICS), Agency for Science, Technology and Research (A*STAR), Singapore, Singapore*
[4] *CREST, ENSAI, Campus de Ker-Lann, Rue Blaise Pascal, BP 37203, 35172 Bruz*
{*marie.du-roy-de-chaumaray, matthieu.marbac-lourdelle2, fabien.navarro*}*@ensai.fr*

Résumé. L'analyse de données d'accelerométrie consiste à extraire des informations sur les temps passés à différents niveaux d'activité. Ces informations sont généralement utilisées ensuite dans un modèle prédictif. Nous proposons une modélisation de ce type de données utilisant un mélange de chaînes de Markov cachées, afin de pouvoir automatiquement détecter le nombre de niveaux d'activités ainsi que leurs caractéristiques. Pour tenir compte de la spécificité des données d'accéléromètre, les données sont modélisées par une distribution de *Zero-inflated Gamma* dont les paramètres sont spécifiques à l'état caché. La modélisation par un mélange permet de tenir compte de l'hétérogénéité de la population. Les propriétés de cette modélisation (identifiabilité, gestion de valeurs manquantes, probabilité de détecter la vraie partition) sont discutées.
Mots-clés. Chaînes de Markov cachées, données d'accéléromètre, modèles de mélanges.

Abstract. Accelerometer data analysis consists of extracting information on time spent at different levels of activity. Then, this information is generally used in a predictive model. We propose to model these data with a mixture of hidden Markov model, to automatically detect the amount of activity levels and their characteristics. To take into account the specificities of the accelerometer data, observations are modeled with a Zero-inflated Gamma distribution whose parameters depends of the hidden state. The use of mixtures allows to consider the heterogeneity of the population. Properties of this model (identifiability, handling missing values, probability of detection of the true partition) are discussed.
Keywords. Accelerometer data, mixture models, hidden Markov model.

1 Introduction

Inadequate sleep and physical inactivity affect physical and mental well-being while often exacerbating health problems. They are currently considered major risk factors for several health conditions. Therefore, appropriate assessment of activity and sleep periods is essential in disciplines such as medicine and epidemiology. The use of accelerometers to evaluate physical activity—by measuring the acceleration of the part of the body to which they are attached—is a classic method that has become widespread in public health research. Indeed, the analysis of actigraphy data has been the subject of extensive studies over the past three decades. Since the pioneering work of Cole et al. (1992), the objective of automating the classification of activity periods directly from an accelerometer signal has been pursued and continues to attract the interest of researchers, medical experts and the industrial community. Many studies have focused on the classification of sleep and wake-up periods (see Wallace et al. (2018)) while others focus on the classification of different levels of activity (see, Yang and Hsu (2010); Huang et al. (2018)). In epidemiology studies, the times spent by activity levels are often used as covariate in predictive models (*e.g.,* as covariates for obesity prediction). However, the definition of the different activity levels (*e.g.,* number of levels, characterization of the levels) and their detection is often made by *ad hoc* methods.

In this work, we develop a mixture of Hidden Markov Models (HMM mixture) for modeling the accelerometer data. The hidden states define the different activity levels. The distribution of the observed data, at any time, is modeled by a zero-inflated gamma (ZIG) distribution whose parameters depend on the hidden state. To the best of our knowledge, the first HMM-based methodology in this context was used in discrete time by Pober et al. (2006), and then in Witowski et al. (2014); Huang et al. (2018) and in continuous time by Xu et al. (2018). HMM mixture permits to estimate the mean time spent by one individual into each activity level. Moreover, it gives estimators of the activity level, for one individual, at any time. Inference by maximum likelihood is performed by an EM algorithm which generalizes the Baum–Welch algorithm (Cappé et al. (2005)) to the case of HMM mixture. The model identifiability and the probability of estimating the true partition are discussed. Missing data occur when the accelerometer is not worn, but they are managed using Markovian properties.

This paper is organized as follows. Section 2 presents the general model and its properties. Section 3 presents a specific version of this model for accelerometer data. Section 4 shows preliminary results on real data. Section 5 discusses some future developments.

2 Mixture of hidden Markov models

The model The observed data $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \dots, \boldsymbol{y}_n^\top)$ are composed of n i.i.d sequences $\boldsymbol{y}_i = (y_{i(1)}, \dots, y_{i(T)})$ measured by an accelerometer at discrete times $t \in \{1, \dots, T\}$. Heterogeneity of the population is modeled by a K-component mixture model providing a partition $\boldsymbol{z} = (\boldsymbol{z}_1, \dots, \boldsymbol{z}_n)$ among the observations $\boldsymbol{y}$. The probability distribution function (pdf) of an observation $\boldsymbol{y}_i$ is

$$p(\boldsymbol{y}_i; \boldsymbol{\theta}) = \sum_{k=1}^{K} \delta_k \, p(\boldsymbol{y}_i; \boldsymbol{\pi}_k, \boldsymbol{A}_k, \boldsymbol{\lambda}), \tag{1}$$

where $\boldsymbol{\theta}$ groups the model parameters, δ_k is the proportion of component k with $\delta_k > 0$ and $\sum_{k=1}^{K} \delta_k = 1$, and $p(\cdot; \boldsymbol{\pi}_k, \boldsymbol{A}_k, \boldsymbol{\lambda})$ being the pdf of component k. Under component k, $\boldsymbol{y}_i$ follows a HMM where the hidden state sequence $\boldsymbol{x}_i = (\boldsymbol{x}_{i(1)}, \dots, \boldsymbol{x}_{i(T)}) \in \mathcal{X}$ takes M values for each observation $\boldsymbol{x}_{i(t)} = (x_{i(t)1}, \dots, x_{i(t)M})$ where $x_{i(t)h} = 1$ if observation i is at state h at time t and $x_{i(t)h} = 0$ otherwise. Its pdf is

$$p(\boldsymbol{y}_i; \boldsymbol{\pi}_k, \boldsymbol{A}_k, \boldsymbol{\lambda}) = \sum_{\boldsymbol{x}_i \in \mathcal{X}} p(\boldsymbol{x}_i; \boldsymbol{\pi}_k, \boldsymbol{A}_k) p(\boldsymbol{y}_i \mid \boldsymbol{x}_i; \boldsymbol{\lambda}),$$

where $\boldsymbol{\pi}_k = (\pi_{k1}, \dots, \pi_{kM})$ defines the initial probabilities so that $\pi_{kh} := \mathbb{P}(X_{i(1)} = h \mid Z_i = k)$ and $\boldsymbol{A}_k$ is the transition matrix so that $\boldsymbol{A}_k[h, \ell] := \mathbb{P}(X_{i(t)} = \ell \mid X_{i(t-1)} = h)$.

$$p(\boldsymbol{x}_i; \boldsymbol{\pi}_k, \boldsymbol{A}_k) = \prod_{h=1}^{\ell} \pi_{kh}^{x_{i(1)h}} \prod_{t=2}^{T} \prod_{h=1}^{M} \prod_{\ell=1}^{M} (\boldsymbol{A}_k[h, \ell])^{x_{i(t-1)h} x_{i(t)\ell}}, \tag{2}$$

The goal is to obtain summary statistics about the time spent at different activity levels for the people having worn the accelerometer, and then to use these statistics in the same predictive model for the whole population. The activity levels are defined by the distribution of $y_{i(t)}$ given $x_{i(t)}$ therefore, it is important that these distributions are equal among mixture component. Thus, clusters are only defined by the transition probabilities $(\boldsymbol{\pi}_k, \boldsymbol{A}_k)$. Therefore, we have

$$p(\boldsymbol{y}_i \mid \boldsymbol{x}_i; \boldsymbol{\lambda}) = \prod_{t=1}^{T} \prod_{h} f(y_{i(t)}; \boldsymbol{\lambda}_h)^{x_{i(t)h}}. \tag{3}$$

The choice of the distribution $f(\cdot; \boldsymbol{\lambda}_h)$ is discussed in the next section.

Model properties The next two propositions present the model identifiability and the probability of misclassifying an observation when the model parameter is known. Mild conditions are required for the proposition: T must be large enough according to K and the distributions of the latent states must be sufficiently different among components.

Proposition 1 *Under mild assumptions, the model defined by eqs.* (1) *to* (3) *is identifiable.*

Proposition 2 *Let θ_0 be the true model parameter and $\mathbb{P}_0 := \mathbb{P}(\cdot | Z_{ik_0} = 1, \theta_0)$ denote the true conditional distribution (labels and parameters are known). Under mild assumptions,* *if the density functions $g(\cdot;\lambda_h)$ are defined on disjoint spaces for any h. Then, for every $a > 0$ and every $k \neq k_0$,*

$$\mathbb{P}_0 \left[\frac{\mathbb{P}(Z_{ik} = 1|\boldsymbol{y}_i)}{\mathbb{P}(Z_{ik_0} = 1|\boldsymbol{y}_i)} > a \right] = \mathcal{O}(e^{-cT}),$$

where $c > 0$ is a positive constant that depends on a and θ_0.

Therefore, by considering $a = 1$, Proposition 2 shows that the probability of misclassifying an observation $\boldsymbol{y}_i$, using the MAP rule, tends to zero when T increases, if the model parameters are known. This proposition can be extended in the case where the density functions are defined on overlapping spaces. We investigate an extension of Proposition 2 that considers an estimator of θ_0 (*e.g.,* the maximum likelihood estimate).

3 Mixture of HMMs of ZIG distributions

Specific distributions conditionally on the state Pre-processed accelerometer data are given every second by combining the vector amplitude of all the movements recorded in the previous second. They are therefore composed of positive values with a large number of zeros. For this reason, we will consider that $y_{i(t)}$ is drawn from a ZIG distribution whose parameters are defined by the state $x_{i(t)}$. Thus,

$$f(y_{i(t)}; \boldsymbol{\lambda}_h) = (1 - \varepsilon_h) g(y_{is(t)}; a_h, b_h) + \varepsilon_h \mathbf{1}_{\{y_{i(t)}=0\}},$$

where $g(u;\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\beta u} \mathbf{1}_{\{u \geq 0\}}$ is the pdf of a Gamma distribution with shape α and rate β, $\boldsymbol{\lambda}_h = (\varepsilon_h, a_h, b_h)$ and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_1, \ldots, \boldsymbol{\lambda}_M)$.

Dealing with missing values Missing values appear when the accelerometer is not worn. Thus, we will not observe isolated missing values but rather wide ranges of missing values. Our idea is that after as many iterations as the number of missing values, the transition matrix can be considered sufficiently close to stationarity, which has been chosen as initial distribution (*e.g.*, let d be the number of successive missing values, $A_k^d[h,\ell] \simeq \pi_\ell$; for any (h,ℓ)). Therefore, an observation $\boldsymbol{y}_i$ with r observed sequences split with missing value sequences of size at least d are modeled as a product of r observed sequences. Once the parameter estimation is completed, we ensure that this assumption was justified by verifying that the width of the smallest range of missing values is large enough to be greater than the mixing time of the transition matrix obtained. To do so, we use an upper bound for the mixing time given by Theorem 12.4 of (Levin et al. 2006).

Maximum likelihood inference All parameters $\delta_k, \boldsymbol{\pi}_k, \boldsymbol{A}_k$ and $\boldsymbol{\lambda}$ are estimated simultaneously using the well-known iterative Expectation-Maximization algorithm for mixtures together with the forward/backward formulas specific to HMM.

4 Analyze of accelerometer data

This section illustrates our proposition on real data. The aim is to extract, from accelerometer data, information on the activity behaviour of individuals and then to plug-it a predictive model to prevent obesity. The application on this dataset is in progress, so we present here only the results obtained on 4 individuals where the accelerometer measures activity every 5 minutes for 5 days. Thus, we only use the HMM of ZIG (*i.e.*, the model we presented with $K = 1$). It should be noted that subject 4 did not wear the accelerometer for a day and a half. Table 1 describes the states with their parameters and gives the mean time per state spent by each individual. Figure 1 shows the accelerometer data of observation 4 and its probability of each state at each time.

	Model parameters				Mean time per state per individual			
State name	ε_h	a_h	b_h	mean	obs. 1	obs. 2	obs. 3	obs. 4
intensive-level	0.00	5.65	0.07	78.23	0.21	0.11	0.15	0.16
moderated-level	0.00	3.91	0.13	30.09	0.24	0.20	0.25	0.18
low-level	0.00	2.08	0.19	11.09	0.19	0.31	0.28	0.15
sleeping 1	0.34	0.76	0.18	2.82	0.28	0.23	0.22	0.38
sleeping 1	0.95	36.02	0.13	14.05	0.07	0.15	0.11	0.13

Table 1: Parameters per state and mean time per states for the four individuals.

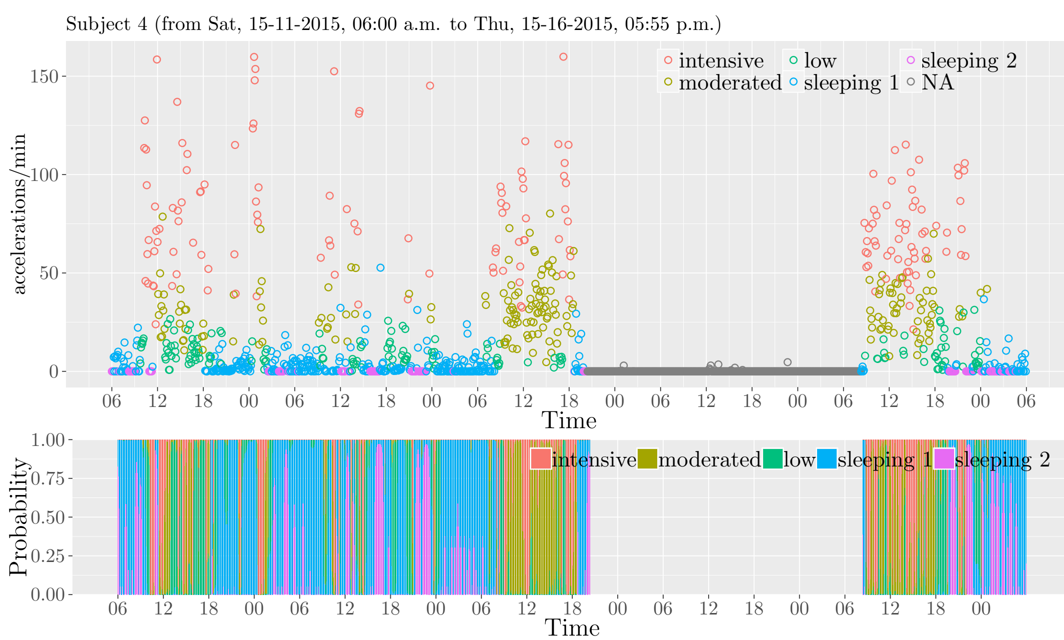

Figure 1: State estimation of individual 4 (with a wear-free period): (a) accelerometer data where color indicates the most likely state; (b) probability of each state at each time.

5 Conclusion

We proposed a new mixture of HMM with ZIG distributions in each state which represent the different activity levels. Parameter inference was performed through a generalization of the Baum-Welch algorithm to this context of mixture and applied to accelerometer data. Future work will focus on developing predictive models for the whole considered population based on the time obtained at each activity level for people who have worn the accelerometer.

Bibliography

Cappé, O., E. Moulines, and T. Rydén (2005). *Inference in hidden Markov models*. Springer Series in Statistics. Springer, New York.

Cole, R. J., D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin (1992). Automatic sleep/wake identification from wrist activity. *Sleep 15*(5), 461–469.

Huang, L., J. Bai, A. Ivanescu, T. Harris, M. Maurer, P. Green, and V. Zipunnikov (2018). Multilevel matrix-variate analysis and its application to accelerometry-measured physical activity in clinical populations. *J. Am. Stat. Assoc*, 1–12.

Huang, Q., D. Cohen, S. Komarzynski, X.-M. Li, P. Innominato, F. Lévi, and B. Finkenstädt (2018). Hidden markov models for monitoring circadian rhythmicity in telemetric activity data. *Journal of The Royal Society Interface 15*(139), 20170885.

Levin, D. A., Y. Peres, and E. L. Wilmer (2006). *Markov chains and mixing times*. American Mathematical Society.

Pober, D. M., J. Staudenmayer, C. Raphael, P. S. Freedson, et al. (2006). Development of novel techniques to classify physical activity mode using accelerometers. *Medicine and science in sports and exercise 38*(9), 1626.

Wallace, M. L., D. J. Buysse, A. Germain, M. H. Hall, and S. Iyengar (2018). Variable selection for skewed model-based clustering: application to the identification of novel sleep phenotypes. *J. Am. Stat. Assoc 113*(521), 95–110.

Witowski, V., R. Foraita, Y. Pitsiladis, I. Pigeot, and N. Wirsik (2014). Using hidden markov models to improve quantifying physical activity in accelerometer data–a simulation study. *PloS one 9*(12), e114089.

Xu, Z., E. B. Laber, and A.-M. Staicu (2018). Hierarchical continuous time hidden markov model, with application in zero-inflated accelerometer data. *arXiv preprint arXiv:1812.01162*.

Yang, C.-C. and Y.-L. Hsu (2010). A review of accelerometry-based wearable motion detectors for physical activity monitoring. *Sensors 10*(8), 7772–7788.