Fusion of estimations from two modalities using the Viterbi's algorithm: application to fetal heart rate monitoring

Rémi Souriau^{1,2}, Julie Fontecave-Jallon¹ and Bertrand Rivet²

1- Univ. Grenoble Alpes, CNRS, UMR 5525, VetAgro Sup, Grenoble INP, TIMC, 38000 Grenoble, France

2- Univ. Grenoble Alpes, CNRS, UMR 5216, Grenoble INP, GIPSA-lab, 38000 Grenoble, France

Abstract. The Viterbi's algorithm allows to estimate latent time series according to observations in a hidden Markov model. This algorithm can be used to merge estimations from different modalities as proposed in this paper. Such a multi-modal estimation is more efficient than mono-modal estimations when the modalities are subject to independent noises. In this paper, this improvement is evaluated in function of noise level of modalities. Experiences on toy data and actual signals to estimate the fetal heart rate show that merging modalities will provide better estimations on average than using the modalities separately.

1 Introduction

Fetal well-being monitoring relies on the analysis of the fetal heart rate (fHR) [1]. The estimation of the fHR can be based on different cardiac signals: e.g., the heart electrical activities through the electrocardiogram (ECG) [2] or the mechanical heart information through the phonocardiogram (PCG) [3]. Their measures are both subject to different noise sources since these two modalities have different physiological origins. Based on this, the merging of both modalities for the estimation of the fHR seems to be an interesting path to gain in performance [4].

In this paper, we are interested in estimating an hidden sequence from multimodal observations. More precisely, we are investigating under which condition adding a second modality can improve the estimation of the hidden sequence compared to using only a single set of observations. The proposed algorithm to merge the two observations is the *Viterbi's algorithm* [5, chap.13] which allows to estimate a latent time series according to observations. Recent studies already used the Viterbi's algorithm for denoising data [6] or for merging different modalities for classification problem [7].

In the next section, the hidden Markov model (HMM) and the Viterbi's algorithm are presented. The third section is dedicated to numerical experiments. A discussion and a conclusion are finally given in the last section.

2 Multimodal Viterbi's algorithm

In this section, the multimodal HMM is first presented (Section 2.1) and then the proposed multimodal Viterbi's algorithm is detailed (Section 2.2).

2.1 Model's description

Let $Y_k^{(1)}$ and $Y_k^{(2)}$ be two noisy estimations of X_k at time k. The evolution of the state X_k and the observations $Y_k^{(1)}$ and $Y_k^{(2)}$ are modelled with a HMM given in figure (1a). The matrix A refers to the transition matrix to X_k given X_{k-1} . $B^{(1)}$ and $B^{(2)}$ are emission matrices of both observations, respectively, $Y_k^{(1)}$ and $Y_k^{(2)}$, given the state X_k . The definition of the model parameters are provided by

$$\forall (i,j) \in [\![1,N]\!]^2, \quad \begin{cases} a_{ij} = \Pr(X_k = j | X_{k-1} = i) \\ b_{ij}^{(1)} = \Pr(Y_k^{(1)} = j | X_k = i) \\ b_{ij}^{(2)} = \Pr(Y_k^{(2)} = j | X_k = i) \end{cases}$$
(1)

where N is the number of states which are directly written with index i and j for simplification. Emission noises of $Y_k^{(1)}$ and $Y_k^{(2)}$ are assumed to be independent, so that the joint probability of observations given the state satisfies $\forall (i, j, m) \in [\![1, N]\!]^3$, $\Pr(Y_k^{(1)} = j, Y_k^{(2)} = m | X_k = i) = b_{ij}^{(1)} b_{im}^{(2)}$. (2)

 $\forall (i, j, m) \in [\![1, N]\!]^3, \quad \Pr(Y_k^{(1)} = j, Y_k^{(2)} = m | X_k = i) = b_{ij}^{(1)} b_{im}^{(2)}.$ (2) Matrices A, B⁽¹⁾ and B⁽²⁾ are fixed for the application. Figures (1b) and (1c) give the *i*th row plot of the transition matrix and emission matrices (before normalization), respectively. First, the transition matrix A is built according to the following idea: large jumps in the variation of the desired state are impossible. This idea is reflected by the introduction of a threshold q for which $\Pr(X_k = j | X_{k-1} = i) = \epsilon$ if |i - j| > q, with ϵ very small. Values of the transition matrix for $|i - j| \leq q$ are given by a triangular function centered in i = j. The sum of coefficients for each row of A is normalized to one. Second, $\forall (i, j) \in [\![1, N]\!]^2, \Pr(Y_k^{(1)} = j | X_k = i)$ and $\Pr(Y_k^{(2)} = j | X_k = i)$ are given by :

$$\begin{cases} \Pr(Y_k^{(1)} = j | X_k = i) = \alpha_i^{(1)} \left(\pi^{(1)} \mathcal{N}(j; i, (\sigma_1^{(1)})^2) + (1 - \pi^{(1)}) \mathcal{N}(j; i, (\sigma_2^{(1)})^2) \right) \\ \Pr(Y_k^{(2)} = j | X_k = i) = \alpha_i^{(2)} \left(\pi^{(2)} \mathcal{N}(j; i, (\sigma_1^{(2)})^2) + (1 - \pi^{(2)}) \mathcal{N}(j; i, (\sigma_2^{(2)})^2) \right) \end{cases} (3)$$

where $\mathcal{N}(\cdot; \mu; \sigma)$ is the value of Gaussian distribution of mean μ and standard deviation σ evaluated at \cdot . The first Gaussian has a small standard deviation $(\sigma_1^{(1)} \text{ and } \sigma_1^{(2)})$ to model instants where observations are close to the desired state. The instants where noise is predominant are modelled by the second Gaussian with a large standard deviation $(\sigma_2^{(1)} \text{ and } \sigma_2^{(2)})$. $\pi^{(1)}$ and $\pi^{(2)}$ give the proportion of the first Gaussian in comparison of the second Gaussian. $\pi^{(1)}$ and $\pi^{(2)}$ can be interpreted as a measure of quality of the modalities. $\alpha_i^{(1)}$ and $\alpha_i^{(2)}$ are normalization coefficients to ensure the sum of coefficients per row is equal to one.

This model setting allows to build toy data similar with the observed fHR estimations in real conditions.



(a) HMM. X_k is the latent state and $Y_k^{(1)}$ and $Y_k^{(2)}$ the two observations.



(c) Emission matrices (i-th row). See equation (3) for the parameters' definition.

Fig. 1: Hidden Markov Model to model the fetal heart rate (X_k) according to two estimations $(Y_k^{(1)} \text{ and } Y_k^{(2)})$. (b) and (c) are, respectively, the *i*-th row construction of the transition matrix $a_{:j}$ and the emission matrices $b_{:j}^{(1)}$ and $b_{:j}^{(2)}$.

2.2 Viterbi's algorithm

The aim of the Viterbi's algorithm is to find the state path $X_{0:T} = \{X_k\}_{k \in [0,T]}$ according to the observation paths $Y_{0:T}^{(1)}$ and $Y_{0:T}^{(2)}$:

$$\widehat{X}_{0:T} = \arg\max_{X_{0:T}} \Pr(X_{0:T} | Y_{0:T}^{(1)}, Y_{0:T}^{(2)}).$$
(4)

To solve this problem, the following recursive formula is used:

$$\begin{aligned}
\mu(X_k) &= \max_{X_{0:k-1}} \Pr(X_{0:k}, Y_{0:k}^{(1)}, Y_{0:k}^{(2)}) \\
&= \begin{cases} \Pr(X_0) \Pr(Y_0^{(1)} | X_0) \Pr(Y_0^{(2)} | X_0), & \text{for } k = 0 \\ \max_{X_{k-1}} \mu(X_{k-1}) \Pr(X_k | X_{k-1}) \Pr(Y_k^{(1)} | X_k) \Pr(Y_k^{(2)} | X_k), & \text{for } k > 0. \end{cases}
\end{aligned} \tag{5}$$

 $\mu(X_k)$ can be interpreted as, for all possible states (X_k) , the maximum probability of transition among all X_{k-1} given $Y_k^{(1)}$, $Y_k^{(2)}$ and the previous value $\mu(X_{k-1})$. Once $\mu(X_k)$ is computed for all k, the path $\hat{X}_{0:T}$ can be deduced by selecting the index of the maximum $\mu(X_k)$ for each k.

3 Experiments

Numerical tests have first been performed on toy data to evaluate how using a HMM with two modalities gives better results than a HMM with only a single modality. Then, we illustrate the result of the proposed multimodal Viterbi fusion on actual fHR signals. In this section, the influence of $\pi^{(1)}$ and $\pi^{(2)}$ defined in (3) is studied. For all experiments, the set of fixed parameters are summarized in Table 1.

Once the state and observations are sampled, the Viterbi's algorithm is used

to predict the state. Three models are compared: two mono-modal models, which rely on a single observation $Y_k^{(1)}$ or $Y_k^{(2)}$, and a multi-modal, which use the two observations $Y_k^{(1)}$ and $Y_k^{(2)}$. Let $\hat{X}^{(1)}$, $\hat{X}^{(2)}$ and $\hat{X}^{(1,2)}$ be the estimated state according to the HMM with the modality $Y^{(1)}$, the modality $Y^{(2)}$ and the two modalities $Y^{(1)}$ and $Y^{(2)}$, respectively. For each HMM, the root mean square error (RMSE) is computed to measure the similarity between the true state X and the predicted state $\hat{X}^{(i)}$: $rmse(\hat{X}^{(i)}, X) = \sqrt{\sum_{k=0}^{T} (\hat{X}_k^{(i)} - X_k)^2/(T+1)}$. The following ratio between RMSEs is then computed:

 $r^{(1)} = \text{rmse}(\widehat{X}^{(1)}, X) / \text{rmse}(\widehat{X}^{(1,2)}, X), \tag{6}$

which quantifies the improvement of RMSE between the multi-modal estimation, $X^{(1,2)}$, and the mono-modal one, $\hat{X}^{(1)}$. Note that we consider arbitrarily that the first modality $Y^{(1)}$ provides in average the best mono-modal estimation, $X^{(1)}$, of X, *i.e.* $\pi^{(1)} \geq \pi^{(2)}$. A ratio $r^{(1)}$ larger than one means that the multi-modal HMM provides a better result than the mono-modal HMM. $r^{(1)}$ provides a view of the gain of performance if a second modality, $Y^{(2)}$, is added compared to only use $Y^{(1)}$ to estimate X. The gain is computed for many sets of $\pi^{(1)}$ and $\pi^{(2)}/\pi^{(1)}$. For a fixed $\pi^{(1)}$, the ratio $\pi^{(2)}/\pi^{(1)}$ varying between 0.1 and 1 gives how much the second modality is degraded in comparison with the first one. $\pi^{(2)}/\pi^{(1)} = 1$ means the two modalities have the same quality. Each configuration $\{\pi^{(1)}, \pi^{(2)}/\pi^{(1)}\}$ is tested fifty times.

Table 2 gives mean values of $r^{(1)}$ for all configurations and Fig. (2) gives box plots of $r^{(1)}$ for cases $\pi^{(1)} = \{0.2; 0.4; 0.6\}$ as illustration. As shown in Table 2, the ratio $r^{(1)}$ is always greater than one, meaning that adding a second modality always leads to an improvement in average. As expected, for each value of $\pi^{(1)}$, the gain $r^{(1)}$ increases with the ratio $\pi^{(2)}/\pi^{(1)}$. This phenomenon can also be observed in figure (2) for $\pi^{(1)} = 0.4$ or 0.6. However, gain values decrease and get closer to 1 as $\pi^{(1)}$ increases. In this latter case, this means that adding a second modality will not improve the estimation if the first modality is already quite good (*i.e.* $\pi^{(1)}$ close to 1). On the contrary, when the first modality is quite noisy, the ratio $r^{(1)}$ is larger (e.g., it varies from 6.95 to 8.89 when $\pi^{(1)} = 0.1$). However, in such a situation, the RMSE using only the modality $Y^{(1)}$ is very high (*rmse* = 68.85 for $\pi^{(1)} = 0.1$) and the gain does not compensate the low performance with a single modality in comparison with the case of $\pi^{(1)}$ tends to 1. Note that, when the additional modality is too noisy (*i.e.* $\pi^{(2)}/\pi^{(1)}$ is low), even if the average gain is larger than 1, this is not always the case for each individual configuration: see Fig. 2 $\pi^{(1)} = .4$ and $\pi^{(2)}/\pi^{(1)} = .1$, the smallest value is less than 1. In other words, in some few cases, adding a too noisy modality can deteriorate the estimation compared to using only one modality.

Set of all	Number of	Transition matrix	Emission matrices	Initial state	
possible states	time sample	parameters	parameters	law	
$[\![30, 250]\!]$	T = 5000	q=3,	$\sigma_1^{(1)} = \sigma_1^{(2)} = 5,$	$\Pr(X_0) =$	
		$\epsilon = 10^{-10}$	$\sigma_2^{(1)} = \sigma_2^{(2)} = 80$	N(120, 10)	

Table 1: Parameters values for numerical tests.



Fig. 2: Gain of performance $r^{(1)}$ in function of $\pi^{(2)}/\pi^{(1)}$ for three values of $\pi^{(1)}$. Box plots give median and quantile values of $r^{(1)}$ over the fifty tests for the different configuration.

An illustration on real data for fHR estimation is given in Fig. (3). Data preprocessing and fHR estimations are detailed in [2, 3]. Figure (3a) gives the noisy input observations $Y^{(1)}$ and $Y^{(2)}$ of fHR, X_k , estimated from ECG and PCG, respectively. Output estimations, $\hat{X}^{(i)}$, are plotted in Figure (3b). The fHR estimation using mono-modal HMMs sometime fails (e.g., ECG between 690s and 720s or PCG between 640s and 660s). The multi-modal HMM provides better results and shows the interest of using both modalities for the fHR estimation. The relative improvement compared to the ECG modality alone is equal to $r^{(ecg)} = 1.37$ with a ratio $\pi^{(pcg)}/\pi^{(ecg)} = 0.97$. Despite the difference with the standard deviations between the real data and the numerical tests, the gain of performance with real data problem is in the same order of magnitude.

4 Discussion

In this paper, the Viterbi's algorithm is used to merge two estimations of the same process. These estimations are jointly modelled using a multi-modal HMM and the Viterbi's algorithm allows to build a new estimation according to the two previous ones used as inputs of the multi-modal HMM. The multi-modal model is compared to the same model using a single modality. As expected, the Viterbi's algorithm succeeds to merge the two estimations into a better one. This study shows, with numerical tests, how the improvement varies in function of the noise level of each modality. The illustration on real data provides a coherent result with the numerical tests. Experiences also show that merging a highly noisy signal can worsen results. Next studies will be focus on modifications of the Viterbi's algorithm to be more robust against poor modelling of the emission (from the state to the observation) and to help the model to reject modalities without information.

Mean values		$\pi^{(1)}$									
		0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$\frac{\pi^{(2)}}{\pi^{(1)}}$	0.2	6.95	1.43	1.03	1.03	1.03	1.03	1.03	1.03	1.03	1.03
	0.4	7.16	1.45	1.08	1.08	1.08	1.08	1.08	1.08	1.07	1.07
	0.6	7.79	1.77	1.16	1.15	1.14	1.13	1.13	1.13	1.12	1.12
	0.8	7.99	1.79	1.21	1.22	1.19	1.19	1.19	1.19	1.18	1.17
	1.0	8.89	1.96	1.29	1.26	1.25	1.25	1.24	1.23	1.23	1.23
$RMSE^{(a)}$		68.85	9.37	5.00	4.24	3.77	3.44	3.19	3.00	2.83	2.67

^(a) Mean RMSE over 500 tests of the Viterbi's algorithm with the modality $Y^{(1)}$ only. Table 2: Gain of performance $r^{(1)}$ in function of $\pi^{(1)}$ and $\pi^{(2)}/\pi^{(1)}$.



Fig. 3: fHR estimation. (a) gives the fHR estimation with the two modalities separately. The three plot in (b) are the comparison between the reference state and estimated states using three HMMs: two HMMs with a single modality: ECG or PCG (the left and the middle plots) and a HMM with the two modalities: ECG and PCG (the right plot). RMSEs are computed over the full recording session (~880 seconds). The initial probability distribution is given by $Pr(X_0) = \mathcal{N}(135, 50)$ to be coherent with the range of the fHR variation (110 to 160 bpm).

References

- D. Ayres-de-Campos, C.Y. Spong and E. Chandraharan, FIGO consensus guidelines on intrapartum fetal monitoring: Cardiotocography. *International Journal of Gynecology & Obstetrics*, vol. 131, no. 1, pages 13-24, 2015
- [2] N. Dia, J. Fontecave-Jallon, M. Resendiz, M.C. Faisant, V. Equy, D. Riethmuller, P. Y. Guméry and B. Rivet, Fetal heart rate estimation by non-invasive single abdominal electrocardiography in real clinical conditions, *Submitted in Biomedical Signal Processing and Control*, 2021.
- [3] N. Dia, J. Fontecave-Jallon, P. Y. Guméry and B. Rivet, Fetal heart rate estimation from a single phonocardiogram signal using non-negative matrix factorization. *Proc. Int. Conf. IEEE Eng. Med. and Bio. Society*, pages 5983-5986, Berlin (Germany), 2019.
- [4] S. Gobillot, J. Fontecave-Jallon, V. Equy, B. Rivet, P.Y. Gumery and P.Hoffmann, Noninvasive fetal monitoring using electrocardiography and phonocardiography: A preliminary study. *Journal of Gynecology Obstetrics and Human Reproduction*, vol. 49, no. 9, pages 455-459, 2018.
- [5] C.M. Bishop. Pattern recognition and machine learning, springer, New York, 2006.
- [6] R.N. Vargas and A.C.P. Veiga, Empirical Mode Decomposition, Viterbi and Wavelets Applied to Electrocardiogram Noise Removal. *Circuits, Systems, and Signal Processing*, vol.40, pages 691-718, 2021.
- [7] A. Pieropan, G. Salvi, K. Pauwels and H. Kjellström, Audio-visual classification and detection of human manipulation actions. Proc. Int. Conf. Intelligent Robots and Systems, pages 3045-3052, 2014.