Unsupervised fine-tuning of Markov-based Neural Networks for Heart Sound Segmentation

Miguel L. Martins
miguel.l.martins@inesctec.pt
Miguel T. Coimbra
mcoimbra@fc.up.pt
Francesco renna
francesco.renna@fc.up.pt

INESC-TEC and FCUP
University of Porto
Porto, Portugal

Abstract

We present a novel hybrid framework that allows for joint learning of a Hidden Markov Chain and Artificial Neural Network in the context of fundamental heart sound segmentation. The Markovian nature of the model allows for unsupervised end-to-end training and our experiments reveal improvement of up to 3.90% Positive Predictive Value over a pre-trained baseline using the PhysioNet 2016 and 2022 datasets.

1 Introduction

Cardiovascular diseases are currently the primary cause of mortality worldwide. Due to its low cost, simplicity, and broad range of diagnostic capabilities, cardiac auscultation is a particularly attractive diagnostic tool to mitigate the burden in such challenging scenarios. With the newfound progress in machine learning, automatic solutions can now extract meaningful clinical information from *Phonocardiogram* (PCG) recordings during the screening phase. These valuable clinical insights depend on key events in the PCG recording, such as the two basic sounds present in each heart cycle: the *first sound* or S1, generated by the mitral and tricuspid valve vibrations from the systolic onset, and the *second sound* or S2, which results from the aortic and pulmonary valve closure at the diastolic onset. For these purposes, we put forward an automatic fundamental heart sound segmentation method using an end-to-end framework that couples Artificial Neural Networks (ANNs) with Hidden Markov Models (HMMs) we named *Markov-based Neural Networks* (MNNs)[1]. Specifically, we show that these models can learn to fit unseen datum sampled from dissimilar distributions in an unsupervised way using a novel gradient descent-based approach. We select two golden standard PCG datasets for this effect and measure very substantial improvements over a pre-trained baseline.

2 Markov-based Neural Networks

Suppose you have a dataset $D = \{(\mathbf{o}^{(i)}, \mathbf{s}^{(i)})\}_{i=1}^{N}$ such that each observation sequence $\mathbf{o}^{(i)} = [\mathbf{o}_1, \mathbf{o}_2, \ldots, \mathbf{o}_{T_i}]$ is a function of a state sequence $\mathbf{s}^{(i)} = [s_1, s_2, \ldots, s_{T_i}]$ that was generated by some (latent) homogeneous first-order Markov chain with discrete states $\mathbf{s}_t \in \mathcal{S}, \mathcal{S} = \{0, 1, \ldots, L-1\}$. The joint distribution of a pair of *emissions* $\mathbf{o}^{(i)}$ and *states* $\mathbf{s}^{(i)}$ is given by:

$$P(\mathbf{o}^{(i)}, \mathbf{s}^{(i)}) = P(s_1) \prod_{t=2}^{T_i} P(s_{t-1}|s_t) \prod_{t=1}^{T_i} P(\mathbf{o}_t|s_t). \tag{1}$$

Presume ignorance of the class of distributions to which $P(\mathbf{o}_t|s_t)$ pertains. Consider instead access to a highly discriminant ANN such that $\text{ANN}(\mathbf{o}_t) \sim P(s_t|\mathbf{o}_t)$. One can approximate (1) by using Bayes' rule to estimate the emission distribution given this approximated posterior estimation, so that $P(\mathbf{o}_t|s_t) = \frac{\text{ANN}(\mathbf{o}_t)P(\mathbf{o}_t)}{P(s_t)}$. Thus, the likelihood of $\mathbf{o}^{(i)}$ follows as:

$$P(\mathbf{o}^{(i)}) = \sum_{\mathbf{s} \in \mathcal{S}^{T_i}} P(s_1) \prod_{t=2}^{T_i} P(s_{t-1}|s_t) \prod_{t=1}^{T_i} \frac{\text{ANN}(\mathbf{o}_t)P(\mathbf{o}_t)}{P(s_t)}, \tag{2}$$

which can be computed efficiently without overflow errors using a scaled forward-backward algorithm. These equations bind the HMM and ANN into a single, unified framework since they depend on the parameters of both models. A *Markov-based Neural Network* (MNN) is thus a HMM that shares the parameter space with an ANN, which models its emissions. This also allows us to use the Viterbi decoder to find the most likely sequences during inference.

2.1 Training

Let $\Psi = \{\lambda, \Theta\}$ be the set of all parameters of an MNN, where Θ denotes the parameters of the ANN, and $\lambda = \{\pi, \Gamma\}$ collects the HMM's parameters, *i.e.*, initial state probabilities $\pi \in \mathbb{R}^L$ and transition matrix $\Gamma \in \mathbb{R}^{L \times L}$.

Equations (1) and (2) are amenable to be adapted to loss functions in an optimization context. Firstly, for the supervised case, the *complete log-likelihood loss* simply follows as:

$$\mathcal{L}_{\text{CL}}(D; \Psi) = -\sum_{i=1}^{N} \log P(\mathbf{o}^{(i)}, \mathbf{s}^{(i)}). \tag{3}$$

Secondly, we adapt (2) to an *unsupervised fine-tuning loss* ($\mathcal{L}_{\text{FT}}$), which enables Ψ to be adaptive given co-variate shifts in previously unseen (and unlabelled) data:

$$\mathcal{L}_{\text{FT}}(D; \Psi) = -\sum_{i=1}^{N} \log P(\mathbf{o}^{(i)}). \tag{4}$$

We use Glorot's normalized initialization for the parameters Θ of the ANN. In a supervised scenario, we find the maximum likelihood estimate of Γ by calculating the expected number of transitions from $\{s^{(i)}\}_{i=1}^{N}$. Afterwards, we find π so that it describes the *steady state distribution* by solving $\pi\Gamma = \pi$, so that $\pi \geq \mathbf{0}$ and $||\pi||_1 = 1$.

2.1.1 Gradient update projection

We use a gradient descent approach to train the ANN and HMM jointly. Note that the rows of Γ and π are probabilistic, and thus lie in the canonical simplex of $\mathbb{R}^L$:

$$\mathcal{K}^L = \left\{ \mathbf{x} \in \mathbb{R}^L : x_i \geq 0, i = 0, \ldots, L-1, \sum_{i=0}^{L-1} x_i = 1 \right\}. \tag{5}$$

After each gradient descent update over Γ we use Michelot's finite projection algorithm [2] to project its rows to $\mathcal{K}^L$. We then set π to be the steady state distribution characterized by the new Γ.

3 Experiments

We use two datasets in our experiments: the 2016 PhysioNet Challenge [1] (PhysioNet'16) dataset and the 2022 PhysioNet Challenge CirCor DigiScope dataset [3] (CirCor'22). The former spans 2435 recordings from 1297 healthy or pathological patients. These recordings were originally re-sampled at 2000 Hz with anti-aliasing [1]. We use only the 792 heart sounds (181 healthy, 611 pathological from a total of 135 patients) that have an associated ECG recording[2]. The fundamental heart sound sequence was estimated through analysis of the synchronous ECG recordings following [5]. We discarded 39 samples, accounting for the cases where the signal lasted less than 1 second or had noisy labels (*i.e.*, illegal state sequences, such as those that allow transitions from state S1 directly to state S2). Secondly, we use CirCor'22 which is currently the largest publicly available pediatric heart sound dataset. It spans 5282 PCG recordings with expert manual annotations of the fundamental heart sounds from 1568 subjects. The signals were sampled at 4000 Hz with a 16-bit resolution. Of the 5282 recordings, we used the 3279 samples publicly available in the training set for the 2022 PhysioNet Challenge[3].

[1] Our implementation is available at `https://github.com/miguelmartins/mnn`

[2] The dataset is available at `https://PhysioNet.org/physiotools/hss/`

[3] The dataset is available at `https://physionet.org/content/circor-heart-sound/1.0.3/`

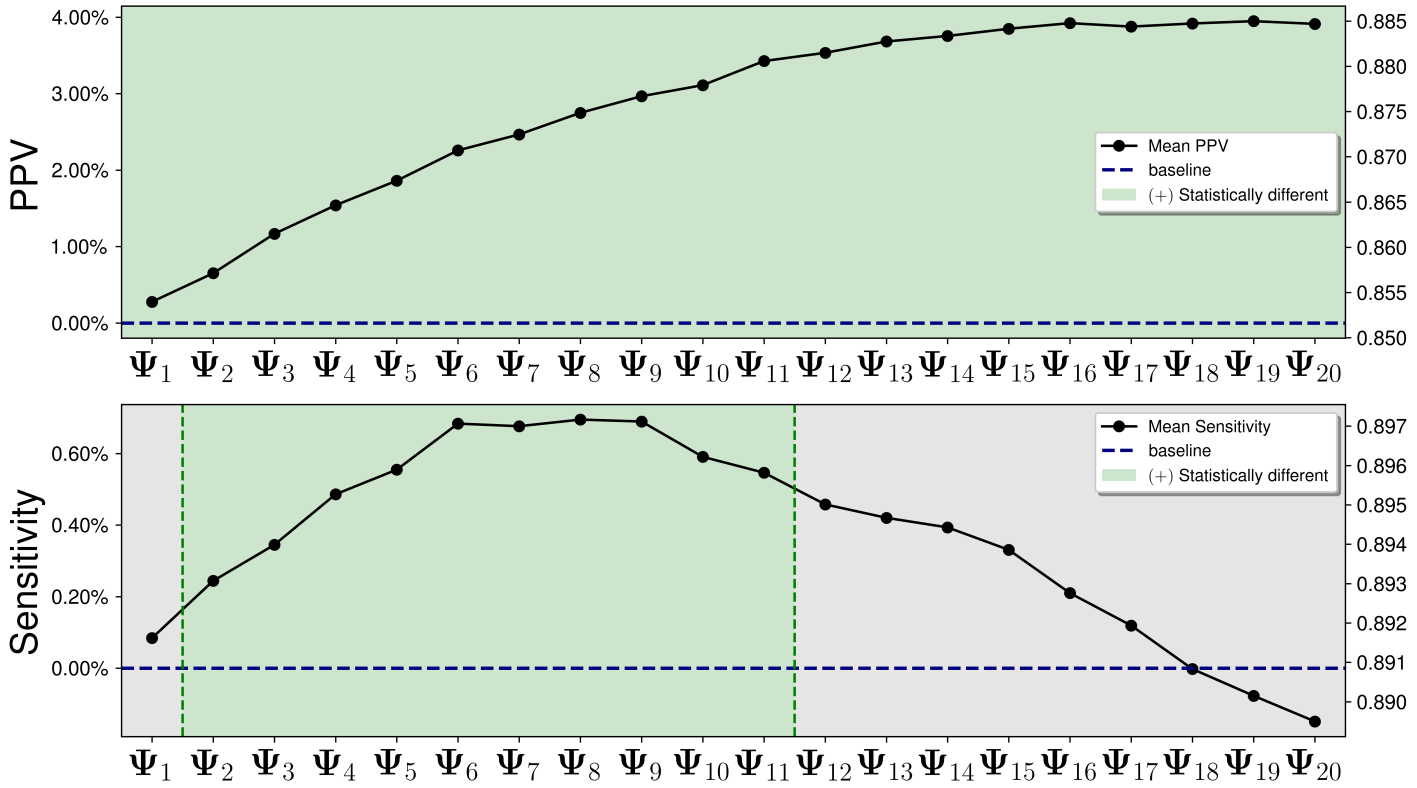

Figure 1: Fine-tuning average performance of $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$ compared with $\text{MNN}^{\Psi_i}_{\mathcal{L}_{\text{FT}}}$, for $1 \leq i \leq 20$. The baseline is pre-trained on CirCor'22 and fine-tuned throughout different number of epochs to the PhysioNet'16 dataset. (Top) PPV. (Bottom) Sensitivity.

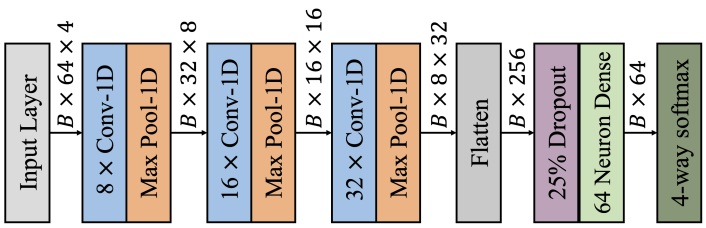

Figure 2: The architecture of the convolutional ANN discriminator.

3.1 Experimental methods

We follow [4] to compute the Positive Predictive Value (PPV) and Sensitivity. We consider a prediction a true positive if the centre of an S1 (S2) prediction is closer than 60 ms to the next S1 (S2) sound in the true sequence. All other predicted S1 or S2 segments are considered false positives. A set of four envelograms is extracted for each sound in order to serve our models after downsampling the signals to 50 Hz following [5]. We adopt a patient exclusive 10-fold cross-validation setup throughout our experiments, where the folds have a fixed length of roughly 10% of the length of the dataset in its entirety. We set aside 10% of the out-of-fold data for early stopping.

We selected the ANN depicted in Figure 2 as the descriminator. A left-to-right HMM is implemented that enforces state transitions of the type: …→S1→Systole→S2→Diastole→S1→…. The procedure presented in Section 2.1.2 is applied solely on the two possible non-zero components of each row while adding/subtracting a small perturbation ε in order to avoid absorbing states.

3.2 Results

We pre-train the MNN in the CirCor'22 dataset using $\mathcal{L}_{\text{CL}}$ with a batch size of 1 in an 80/10/10 random holdout split with early-stopping at the best loss value using the *Adam* algorithm. The parameters of the model attained at this pre-training stage are denoted as Ψ_0 and the associated model as $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$. Then, as described in Section 2.1, we fine-tuned $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$ to each observation of the PhysioNet'16 dataset seperately and recorded the average PPV and Sensitivity of each additional round of fine-tuning until $k = 20$ epochs in the entire dataset, which is denoted by $\text{MNN}^{\Psi_i}_{\mathcal{L}_{\text{FT}}}$. Pairwise t-tests between $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$ and $\text{MNN}^{\Psi_i}_{\mathcal{L}_{\text{FT}}}$, $1 \leq i \leq k$, with significance $\alpha = 0.05$ were performed to grasp wether there was a statistical significant improvement of the fine-tuned models over the baseline $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$, as depicted in Fig. 1. A steady increase in PPV is observed throughout all i with a positive or negligible impact in Sensitivity. In fact, baseline $\text{MNN}^{\Psi_0}_{\mathcal{L}_{\text{CL}}}$ scored a mean PPV of 0.847 and mean Sensitivity of 0.891, while the model $\text{MNN}^{\Psi_{20}}_{\mathcal{L}_{\text{FT}}}$ scores the best average PPV at 0.886 (3.90% statistically significant increase) with mean Sensitivity of 0.889 (0.20% statistically non-significant decrease), which is a substantial improvement in performance.

4 Conclusion

We presented Markov-based Neural Networks as a unifying framework between ANNs and HMMs in the context of heart sound segmentation. The expression that characterizes the likelihood of the model was used as the optimization objective for the goal of unsupervised training in unseen datum. Concretely, we observed that a pre-trained model in the CiCor'22 dataset could be enhanced through fine-tuning with an improvement of up to 3.90% PPV over its baseline performance in the PhysioNet'16 dataset.

References

[1] Chengyu Liu et al. An open access database for the evaluation of heart sound algorithms. *Physiological Measurement*, 37(12):2181–2213, 2016.

[2] Christian Michelot. A finite algorithm for finding the projection of a point onto the canonical simplex of α^n. *Journal of Optimization Theory and Applications*, 50(1):195–200, 1986.

[3] Jorge Oliveira et al. The CirCor DigiScope dataset: from murmur detection to murmur classification. *IEEE journal of biomedical and health informatics*, 26(6):2524–2535, 2021.

[4] SE Schmidt et al. Segmentation of heart sound recordings by a duration-dependent hidden Markov model. *Physiological measurement*, 31(4):513–529, 2010.

[5] David B Springer, Lionel Tarassenko, and Gari D Clifford. Logistic regression-hsmm-based heart sound segmentation. *IEEE transactions on biomedical engineering*, 63(4):822–832, 2015.