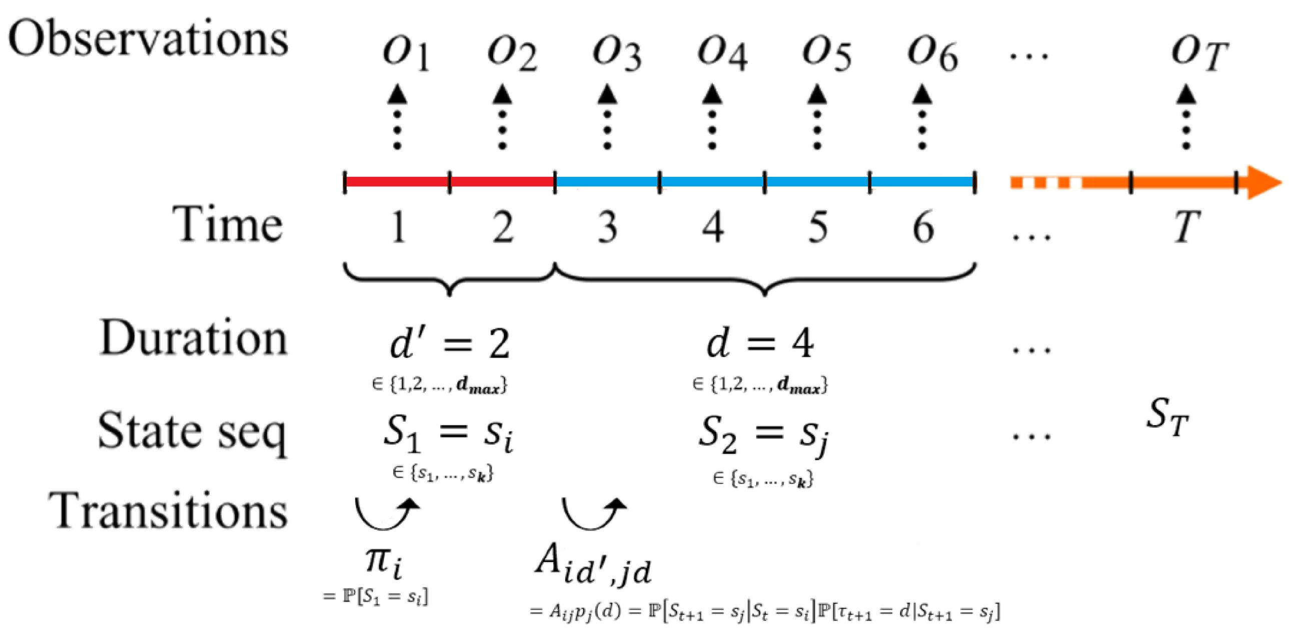
“High accuracy and low computational requirement define a novel heuristic method for **selecting the optimal order** in **hidden semi-Markov models**”

Selection of the number of states in Hidden semi-Markov models

RENTE, Filipa[1]; MARINHO, Zita[2]; FIGUEIREDO, Mário[3]

1 Instituto Superior Técnico (IST), Universidade de Lisboa (ULisboa), Portugal
2 Instituto de Sistemas e Robótica, IST & Priberam Labs, Lisboa, Portugal
3 Instituto de Telecomunicações, IST, ULisboa, Portugal

HIDDEN SEMI-MARKOV MODEL (HSMM)



INTRODUCTION

- HSMMs are powerful probabilistic models used in several fields, e.g., speech, health and genetics.
- More expressive extension of hidden Markov models (HMMs) with explicit state duration.
- Order selection problem: find the optimal hyperparameters **k** (nº states of the model) and $\mathbf{d_{max}}$ (maximum allowed state duration).

SOLUTION

- Single HSMM selection criterion for both the number of states and the maximum allowed state duration.

METHODS

- Empiric estimation of $\mathbf{d_{max}}$:

1. Find θ_j (parameter of the state duration distribution) that corresponds to the highest duration amongst all states s_j – θ_j^*;
2. $\mathbf{d_{max}}$ is the (first) integer duration for which the cumulative state duration probability, with parameter θ_j^*, is smaller than $\varepsilon = 0{,}01$.

- Estimation of **k**:

Sequential pruning strategy + Mixture minimum description length (MMDL) criterion	**VS**	Standard BIC criterion

CRITERIA ADAPTATION FOR HSMMs

BIC criterion for HSMMs

$$\mathbf{BIC_{HSMM}}(\boldsymbol{k}) = \mathbf{LL} - \frac{\boldsymbol{k^2 + k}}{\mathbf{2}}\boldsymbol{log(n)}$$

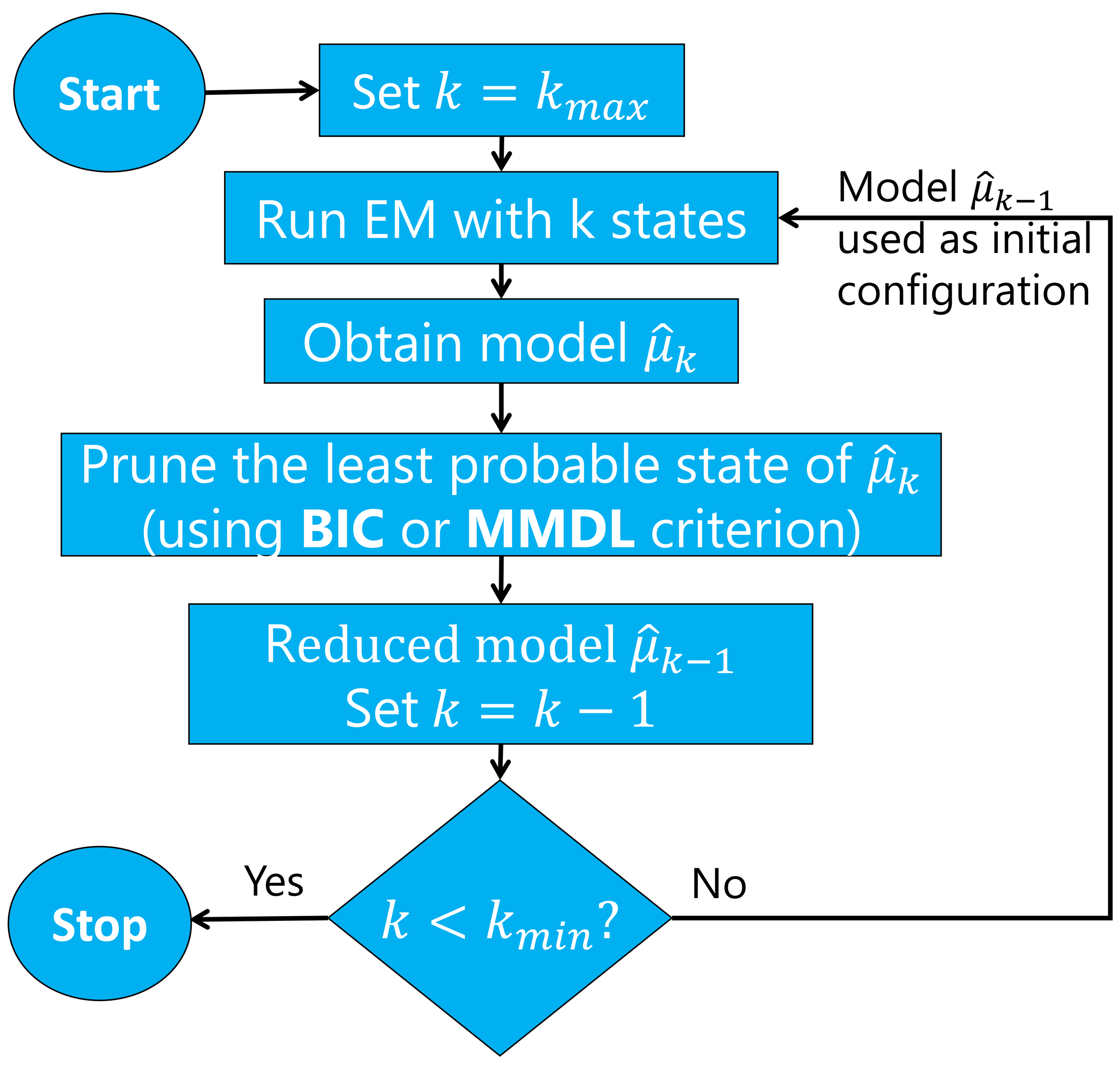
$\boldsymbol{k}$: number of states.
$\boldsymbol{n}$: total number of samples.
$\mathbf{LL}$: log-likelihood of the observations.

MMDL criterion for HSMMs

$$\mathbf{MMDL_{HSMM}}(\boldsymbol{k}) = \mathbf{LL} - \frac{\boldsymbol{k^2 - k}}{\mathbf{2}}\boldsymbol{log(n)} - \sum_{\boldsymbol{m=1}}^{\boldsymbol{k}}\boldsymbol{log(np_\infty(m))}$$

$\boldsymbol{p_\infty(m)}$: stationary probability distribution for HSMMs.

Sequential pruning strategy

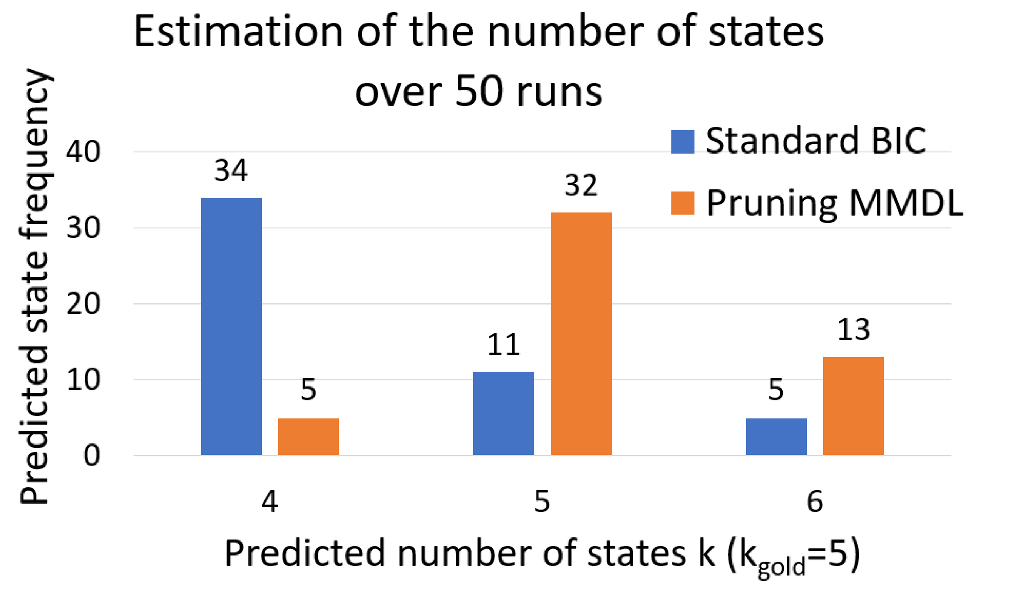


$\hat{\mu}_k = (\boldsymbol{s}, A, \pi, B, D)$: predicted model parameters.
$\boldsymbol{s} = \{s_1, \dots, s_k\}$: set of k states.
$A \in \mathbb{R}^{k\times k}$: transition matrix.
$\pi \in \mathbb{R}^k$: initial state probability distribution.
$B \in \mathbb{R}^{k\times m}$: emission matrix, for m different observation symbols.
$D \in \mathbb{R}^{k\times d_{max}}$: duration probability matrix.

RESULTS with synthetic data

Pruning MMDL vs. BIC

Method	Nº states	Nº iter.
BIC	4.4 ± 0.7	38.6 ± 34.2
MMDL	5.2 ± 0.6	5.8 ± 2.9



- **Accuracy** in the estimation of the number of states (*Nºstates* field; true value is 5);
- **Computational requirement** reflected in the number of required iterations in the EM algorithm (*Nºiter* field).

DISCUSSION

Comparison between the pruning MMDL for HSMMs and the standard BIC criterion:

- **Higher accuracy** in the selection of the optimal number of states;
- Less demanding **computational requirement**;
- The sequential pruning strategy guarantees a **lower sensitivity** to the initialization of the EM algorithm.

FUTURE WORK

- Compare with other standard methods;
- Demonstrate the effectiveness in a more substantial application study using real data;
- Design feature conditioned state transitions or add latent variables to the HSMM model so that external information can be captured.

TÉCNICO LISBOA

Interested in the full paper? SCAN ME