Bayesian nonparametric methods for dynamics identification and segmentation for powered prosthesis control

Neil Dhir
Machine Learning Research Group [Osborne], Applied AI Laboratory [Posner] and (and Probabilistic Programming Group [Wood])

Viva voce

January 25, 2018

Overview

1. Motivation
   - High-level
   - Mid-level
   - Appropriate state-space modelling strategy for sequence modelling
2. Understanding animal behaviour from observations
   - What is the problem and why do we care?
   - Appropriate state-space modelling strategy for sequence modelling
3. "Automatic inference" $\rightarrow$ probabilistic programming
   - A primer
   - Probabilistic programming languages
4. State-space models
5. Bayesian nonparametrics + state-space models
6. Synthetic experiments
7. Labelling lion observations
8. Conclusion

Motivation: high-level

- Multiple-sclerosis: severely affects and reduces mobility
- Diabetes: can give rise to complications e.g. heart disease, kidney disease retinopathy (a complication which damages retina) and neuropathy.
- *Diabetes is the leading cause of amputation*
- Quality of living; activities of daily living such as locomotion

(a) Parents

Motivation: mid-level

- Multiple-sclerosis: severely affects and reduces mobility
- Diabetes: can give rise to complications e.g. heart disease, kidney disease retinopathy (a complication which damages retina) and neuropathy.
- *Diabetes is the leading cause of amputation*
- Quality of living; activities of daily living such as locomotion

(b) Parents

Sequence modelling

- Time-series labelling is laborious and subjective $\rightarrow$ use semi/un-supervised learning to support labelling exercise
- Want to discover new behaviours $\rightarrow$ Bayesian nonparametrics might help
- Model structure still far from clear $\rightarrow$ probably need to iterate over models
- Great opportunity for *probabilistic programming*

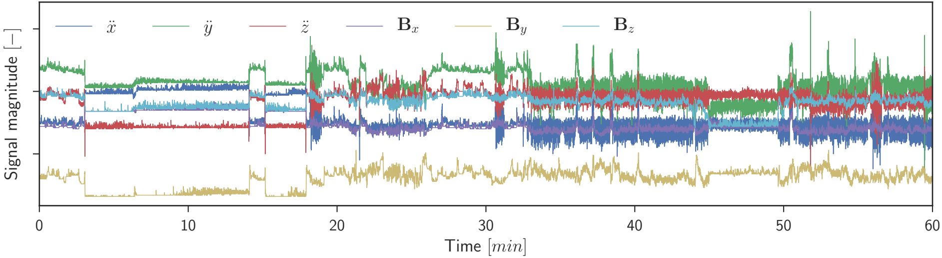
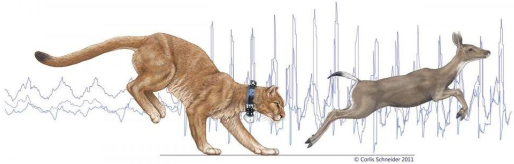


Figure: one hour of raw data captured at 32Hz, containing a total of 115,200 multivariate readings, wherein tri-axial accelerometry $(\ddot{x}, \ddot{y}, \ddot{z})$ and magnetometer $(\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z)$ observations are shown.

Understanding animal behaviour from observations

- Oxford's zoologists have been tracking prides of lions for years
- Famous members include Cecil and Xanda (killed by trophy hunters in July, two years after Cecil)
- Observations $(y \in \mathbb{R}^d, d \gg 1)$ often sampled at years at a time, sometimes at very high frequencies
- Use of especially accelerometry is widespread within biotelemetry as a means of measuring an animals activity quantitatively
- Biotelemetry is used as a means of classifying behaviour i.e. to *understand* their ecology



(a) Puma. Not lion. Still has collar, so we're ok.

(b) Collared lions

Sequence modelling

- Time-series labelling is laborious and subjective $\rightarrow$ use semi/un-supervised learning to support labelling exercise
- Want to discover new behaviours $\rightarrow$ Bayesian nonparametrics might help
- Model structure still far from clear $\rightarrow$ probably need to iterate over models
- Great opportunity for *probabilistic programming*

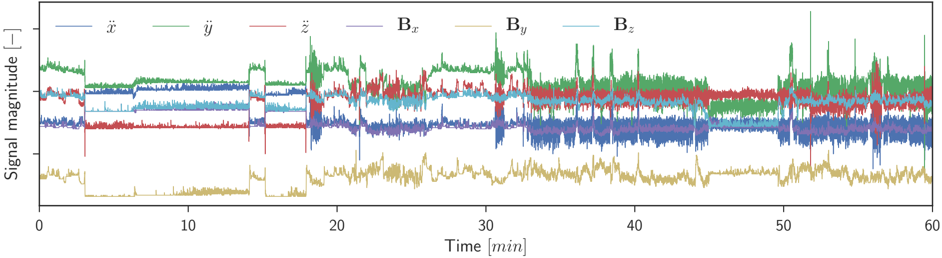


Figure: one hour of raw data captured at 32Hz, containing a total of 115,200 multivariate readings, wherein tri-axial accelerometry $(\ddot{x}, \ddot{y}, \ddot{z})$ and magnetometer $(\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z)$ observations are shown.

Probabilistic programming

What *is* it?

- Languages for probabilistic modelling and inference
- Separate modelling and inference
- Use general purpose inference (i.e. 'black-box' that can just be applied on the fly)

Why care?

- Make complex statistical modelling/ML available to non-experts
- Think about the *what* rather than *how*
- Computing power will increase but we will not get (much) smarter $\rightarrow$ scientist's time more important than computing time, hence generic inference worth it
- Mix statistics with classical computer science e.g. data structures and higher order functions (map, reduce)

Probabilistic programming languages

Many languages, with different focus:

- Stan (HMC for continuous variables)
- PyMC3
- BUGS (pioneer in MCMC for graphical models)
- Infer.NET (graphical models, variational inference, EP)
- Church/**Anglican**/Venture (functional programming, data structures, SMC and MCMC, BNP, experimental)
- Edward and many others...

```
(defquery hmm
 [observations initial-dist
    transition-dists observation-dists]
 (reduce
  (fn [states observation]
   (let [state (sample (get
       transition-dists (peek states)))]
    (observe (get observation-dists
      state) observation)
    (conj states state)))
  [(sample initial-dist)]
  observations))
```

Figure: example Anglican code for conditioning HMM on observations.

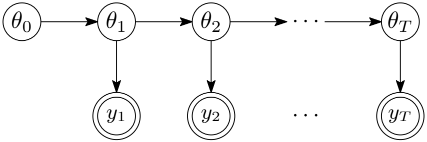
State space models

Problems with HMMs

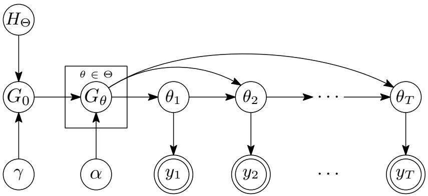
1. Geometric state duration: $\mathbb{P}(d) = a^{d-1}(1-a)$ where d denotes the duration of a given state and a denotes the Markov transition probability of a self-transition
2. Number of latent states must be set a priori

Solutions

1. Employ explicit state duration HMMs e.g. EDHMM or HSMM
2. Use BNP to place an unbounded prior on the latent state cardinality



(a) Vanilla HMM



(b) BNP HMM (i.e. HDP-HMM)

Figure: From vanilla to infinity.

Bayesian nonparametrics + state-space models

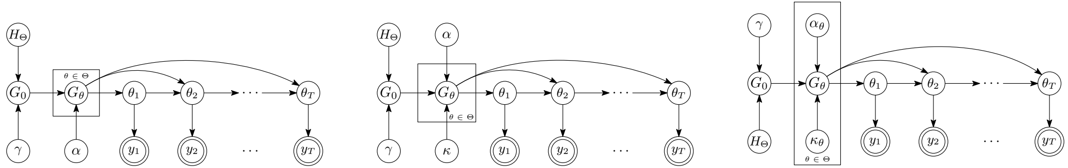
Hierarchical Dirichlet process hidden Markov model (HDP-HMM)

$$
\begin{array}{lll}
G_0 \mid \gamma, H \sim \mathcal{DP}(\gamma, H) & & \text{Sample random base measure } G_0 \\
G_\theta \mid \alpha, G_0 \sim \mathcal{DP}(\alpha, G_0) & \theta \in \Theta & \text{Sample transition distribution } G(\cdot) \\
\theta_i \mid \theta_{i-1} \sim G_{\theta_{i-1}} & i = 1, 2, \ldots & \text{Sample state from transition distribution} \\
y_i \mid \theta_i \sim F_{\theta_i} & i = 1, 2, \ldots & \text{Sample from emission distribution } F(\cdot)
\end{array}
$$

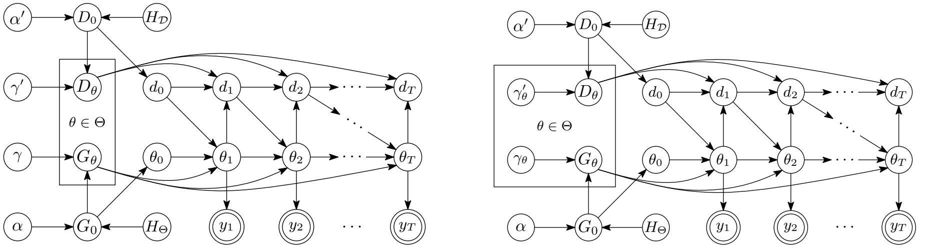
We add:

- Infinite duration hidden Markov model (IDHMM) – *nonparametric state durations*
- Stateful IDHMM – *stateful nonparametric state and duration statistics*

A smörgåsbord of models (and that's the whole point)



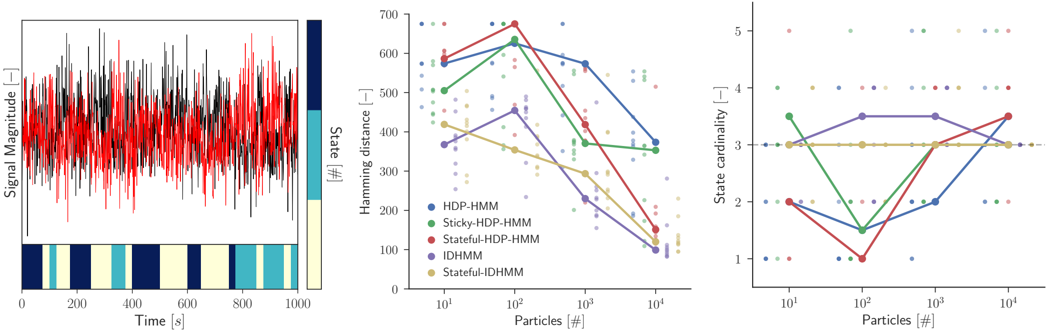
(a) HDP-HMM (b) Sticky HDP-HMM (c) Stateful HDP-HMM



(d) IDHMM (e) Stateful IDHMM

Figure: BNP *discrete* SSMs used in this work.

Synthetic experiments

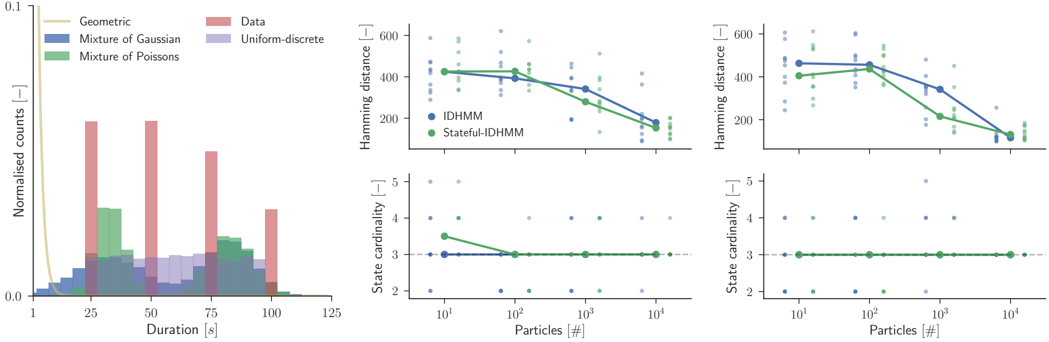


(a) Observations w. GT (b) Hamming distance (c) State cardinality

Figure: results from experiments on multivariate synthetic Gaussian observations with sequential Monte-Carlo inference. Connected bullets are median scores.

*Hamming distance: the number of positions at which the corresponding symbols are different, for two sequences of equal length – i.e. measuring the *edit distance*

Synthetic experiments



(a) Duration priors (b) Gaussian mixture duration (c) Poisson mixture duration

Figure: results from experiments with different duration priors. Connected bullets show median scores.

Labelling lion observations: fuzzy ground-truth

Labelling lion observations

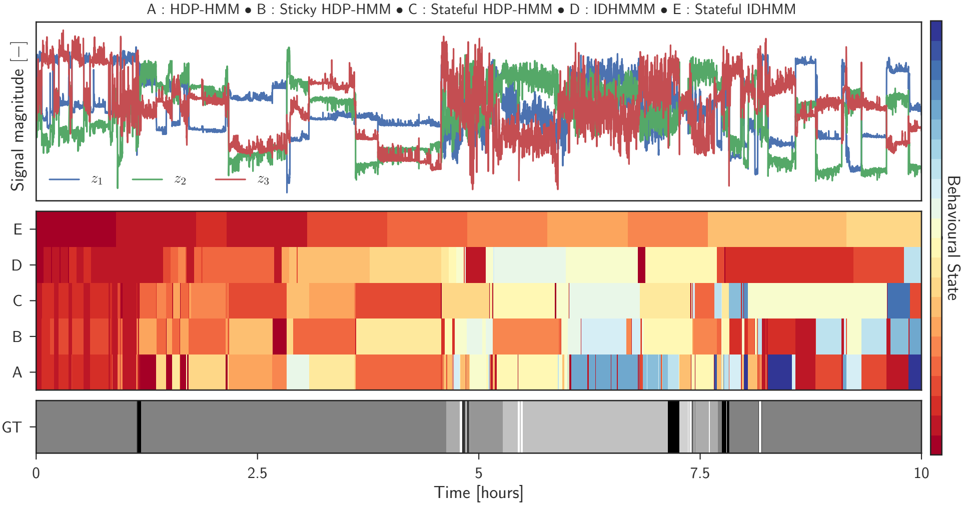


Figure: **top** – signal, **middle** – state sequences inferred by models through unsupervised learning and **bottom** – manually labelled fuzzy ground truth state sequence

Detailed analysis: a hunt

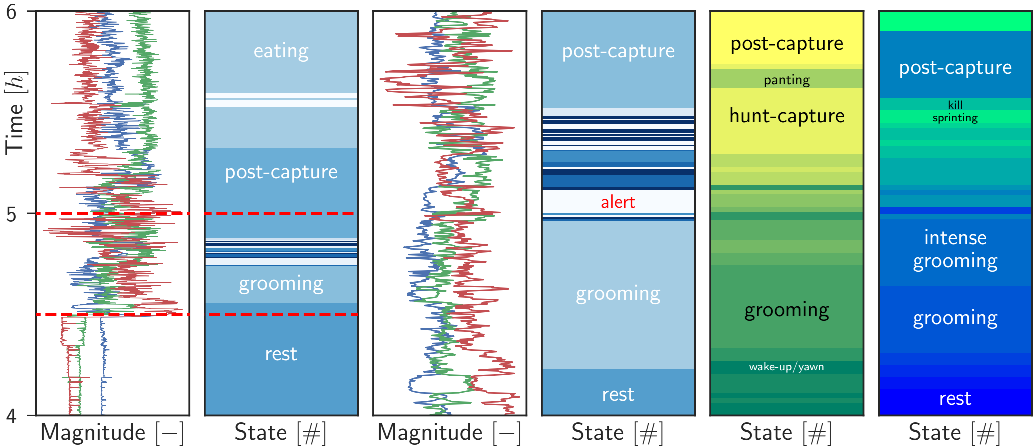
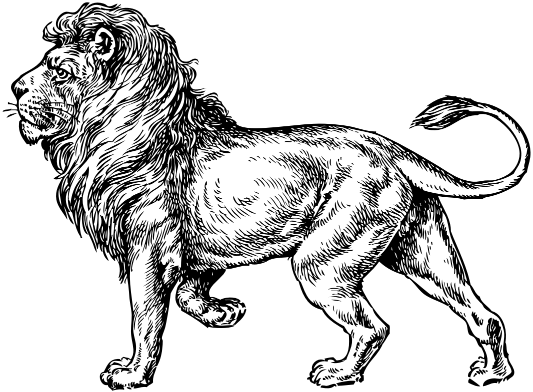


Figure: **first two panels** – signal and ground truth; **next two panels** – zoomed in; **final two panels** – assigned detailed labelling by IDHMM and stateful IDHMM as established by listening to audio, "concluding" that models learned *meaningful* new behaviours.

Conclusion and future work

- Bayesian nonparametric state-space models show promise in this difficult domain
- For sequence modelling and novelty detection, these models could make animal ecology more interpretable
- Purely as a first pass through the data, this approach allows the zoologist to identify regions of interest
- By using PPS we can quickly iterate over state-space models which are
  - Non I.I.D. and which have non-geometric durations
  - Unsupervised
  - Nonparametric
- **Future work**: incorporate domain knowledge through priors, semi-supervised learning, use audio as observations, transfer learning between members of the pride and like everyone else we too are exploring sequence modelling via *deep learning*



Questions?