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

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

Viva voce

January 25, 2018

Overview

Motivation: high-level

- Multiple-sclerosis: severely affects and reduces mobility
- Diabetes: can give rise to complications e.g. heart disease, kidney disease retinopathy and neuropathy.
- *Diabetes is the leading cause of amputation*
- Quality of living through ability to perform activities of daily living such as locomotion

Figure: Parents

Motivation: mid-level

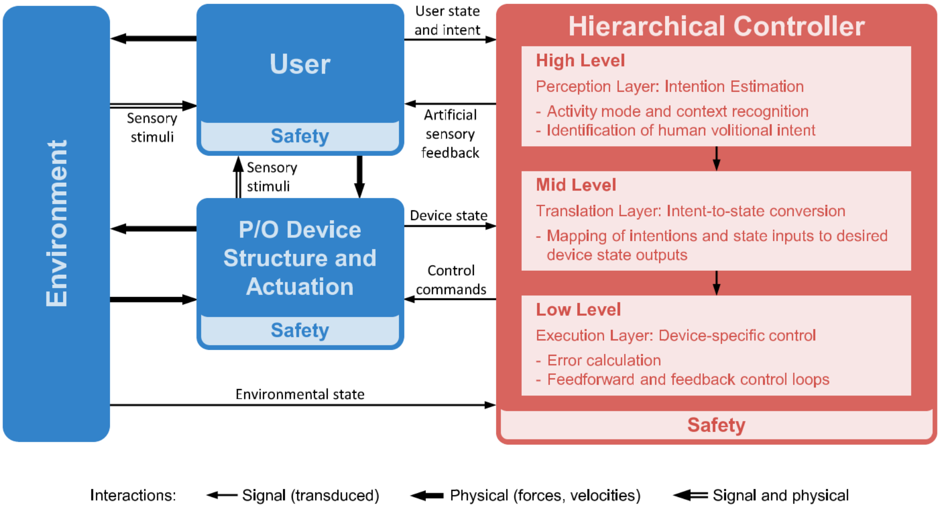


Figure: Generalised control framework for active prostheses and orthoses (Tucker et al., 2015).

Motivation: low-level

- Perception layer
  - **Chapter III:** Incidence detection
  - **Chapter IV:** Dynamics identification via time-series segmentation
- Translation/Execution layer
  - **Chapter V:** Gaussian process regression for prosthesis control

Three broad themes were addressed which fit into the control stratification proposed by Tucker et al. (2015).

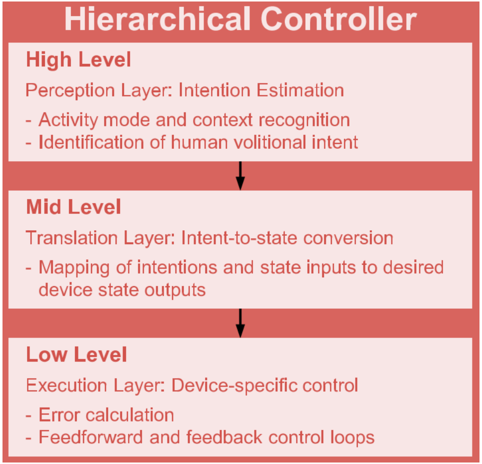


Figure: Parts of a hierarchical controller (Tucker et al., 2015).

Chapter III: Incidence detection [Perception layer]

The purpose of the high-level control is to perceive the locomotive intent of the user through activity mode detection.

- Use standard discriminative and generative approaches for classification
- Demonstrate that popular original study by Luštrek and Kaluža (2009) could be improved
- Missing data was in-painted using a generative model (Kalman smoother)
- Dimensionality reduction operated upon complete data
- Increased performance was demonstrated with this simple approach

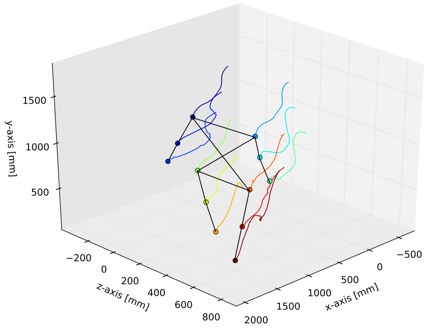


Figure: Example of ‘walking’ incidence (Dhir and Wood, 2014).

Chapter III: Example observations

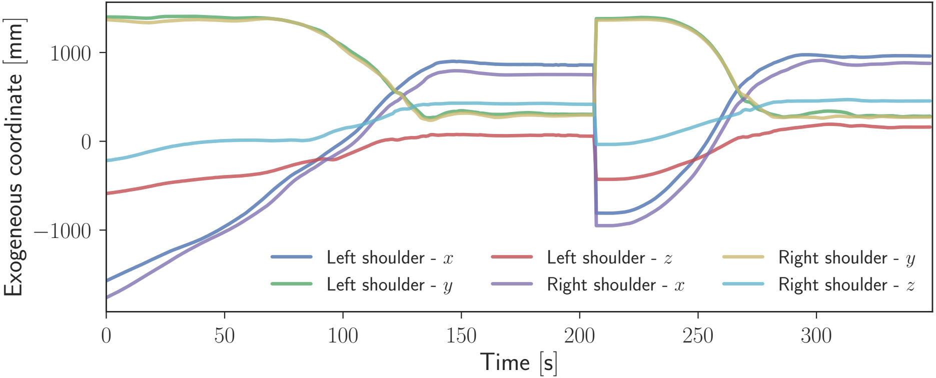


Figure: Recordings made from an infrared motion capture system. Window shows marker trajectories of the coordinates of two (out of 12) markers attached to the bodies of three volunteers. The scenario depicted includes three activities, enacted in the following order: walking → falling → lying.

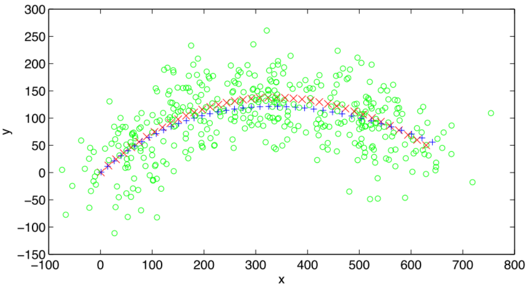


Figure: Kalman smoothing to estimate trajectory; with observations, ground truth and estimates (Barber, 2012).

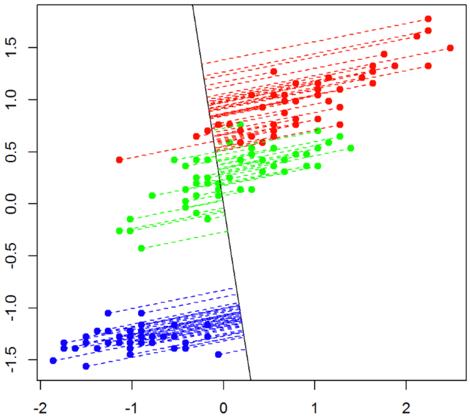


Figure: Multiclass linear discriminant analysis of a two-dimensional demonstration space (Bishop, 2006).

Chapter III: Results

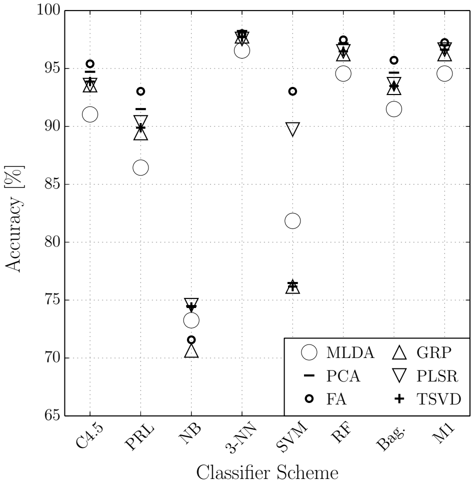


Figure: Methods applied to raw data.

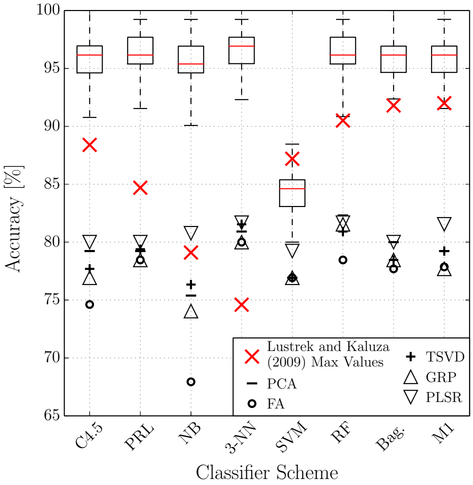


Figure: Methods applied to feature vectors.

Chapter III: Conclusions

- Able to significantly improve **supervised** classification performance using very simple, off-the-shelf, tools
- In-painting proved valuable to complete the dataset
- Though complete, the dataset also has lots of redundancy, removed through dimensionality reduction (DR)
- DR was the primary driver of improved performance
- Additional experiments investigated information content of individual tags (see RHS figure)

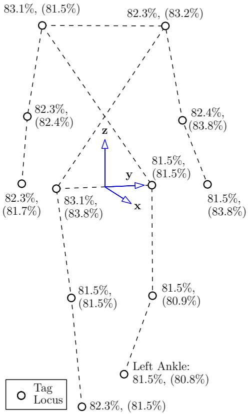


Figure: Illustration of activity, with best individual tag classification accuracy quoted with each tag (Dhir and Wood, 2014).

Reminder: Where are we now in the control hierarchy?

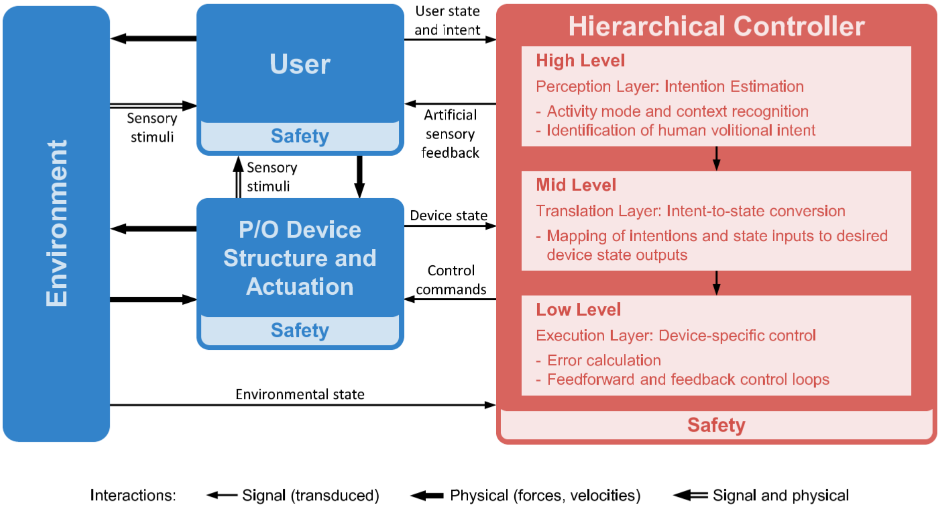


Figure: Generalised control framework for active prostheses and orthoses (Tucker et al., 2015).

Chapter IV: Dynamics identification via time-series segmentation [Perception layer]

The purpose of the high-level control is to perceive the locomotive intent of the user through activity mode detection.

- Traditional setting: supervised learning (Ch. III) using discriminative models
- Must consider unsupervised setting using generative modelling
- Behaviours and activities are likely to grow with time, hence it is unsatisfactory to bound the state-space
- Incidence detection and labelling must be **automatic**

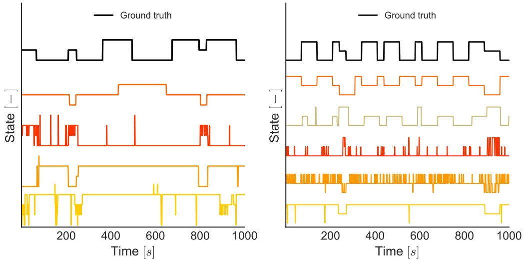


Figure: Unsupervised learning of slow and fast dynamics.

Chapter IV: Generative 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 in many situations (new and old) $\rightarrow$ probably need to iterate over models
- Great opportunity for *probabilistic programming*

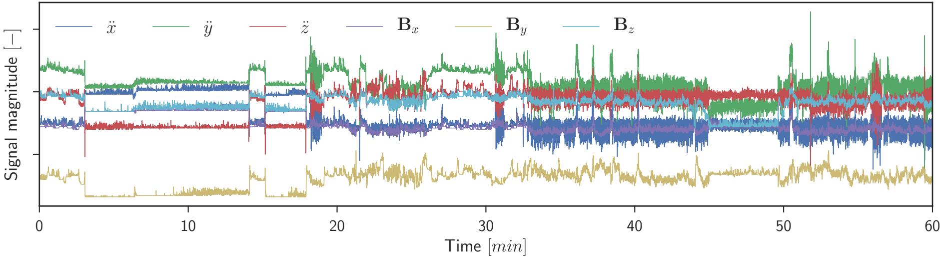


Figure: Raw data captured at 32Hz, containing a total of 115,200 multivariate readings, recorded with tri-axial accelerometry $(\ddot{x}, \ddot{y}, \ddot{z})$ and magnetometers $(\mathbf{B}_x, \mathbf{B}_y, \mathbf{B}_z)$.

Chapter IV: 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 (e.g. `map`, `reduce`)

Chapter IV: Probabilistic programming languages

- Functional PPL: Church/**Anglican**/Venture (data structures, SMC and MCMC, BNP primitives, experimental)
- Stan (big following with a focus on HMC, thus limited support for models with discrete latent spaces)
- PyMC3/Edward (make use of auto-diff packages such as Theano and TF to calculate gradients in MCMC and VI settings)
- BUGS (pioneer in MCMC for graphical models)
- Infer.NET (graphical models, variational inference, EP)

```
(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.

Chapter IV: Example of a probabilistic program

The following example appeared in (Tolpin et al., 2015).

Statistical model:

$$
\begin{aligned}
x_1 &\sim \mathcal{N}(2, 1) \\
x_2 &\sim \mathcal{N}(x_1, 1) \\
y_1 \mid x_1 &\sim \mathcal{N}(x_1, 0.1^2) \\
y_2 \mid x_2 &\sim \mathcal{N}(x_2, 0.1^2)
\end{aligned}
$$

```
(query
  (let
    [unknown-mean-t1 (sample (normal 2 1))
     unknown-mean-t2 (sample (normal unknown-mean-t1 1))
     noise 0.1]
    (observe (normal unknown-mean-t1 noise) 3)
    (observe (normal unknown-mean-t2 noise) 3.1)
    (predict unknown-mean-t1)
    (predict unknown-mean-t2)))
```

Figure: A probabilistic model with Gaussian emissions with unknown means and a known standard deviation (Perov, 2016).

As noted by Perov (2016), an example of the execution trace for the program is $(x_1 = 2.7, x_2 = 2.3, y_1 = 2.0, y_2 = 2.1)$. For this program, each list of four random variables constitutes a valid execution trace: $(x_1 \in \mathbb{R}^1, x_2 \in \mathbb{R}^1, y_1 \in \mathbb{R}^1, y_1 \in \mathbb{R}^1)$.

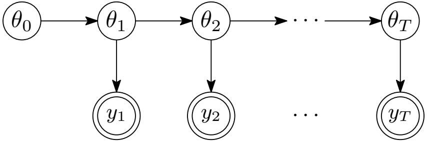
State space models

Problems with HMMs

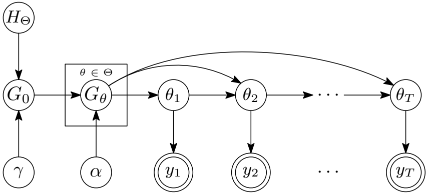
1. Geometric state duration: $\mathbb{P}(d) = a^{d-1}(1-a)$ where d denotes the duration of a given state and a denotes the Markov transition probability of a self-transition
2. Number of latent states must be set a priori

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

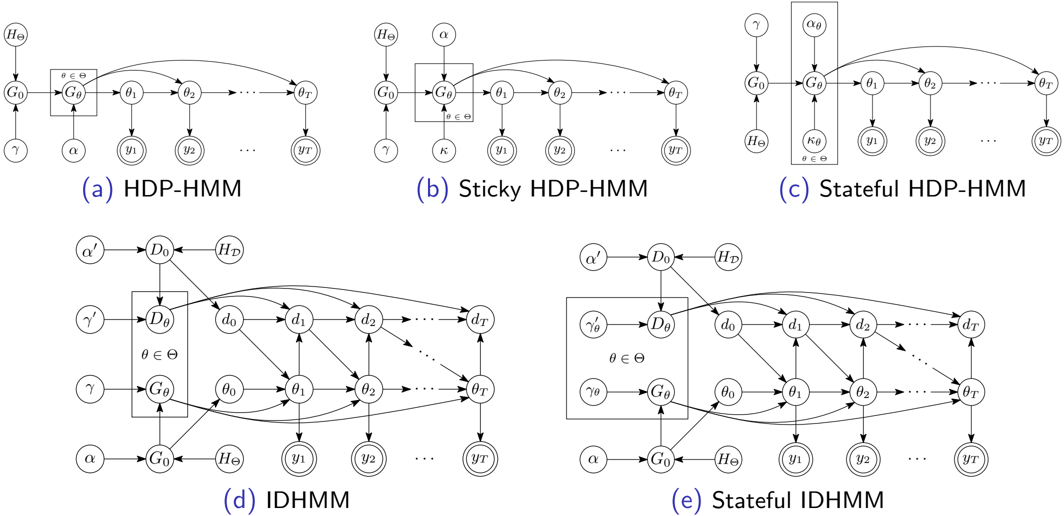
Hierarchical Dirichlet process hidden Markov model (HDP-HMM)

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

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)

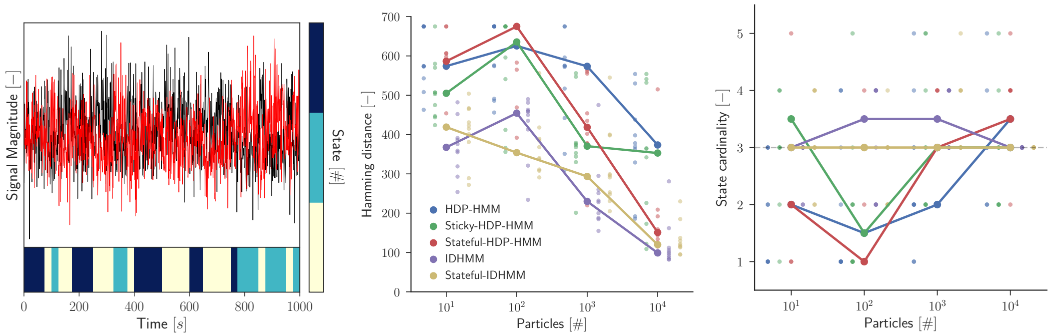


(a) HDP-HMM (b) Sticky HDP-HMM (c) Stateful HDP-HMM

(d) IDHMM (e) Stateful IDHMM

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

Synthetic experiments

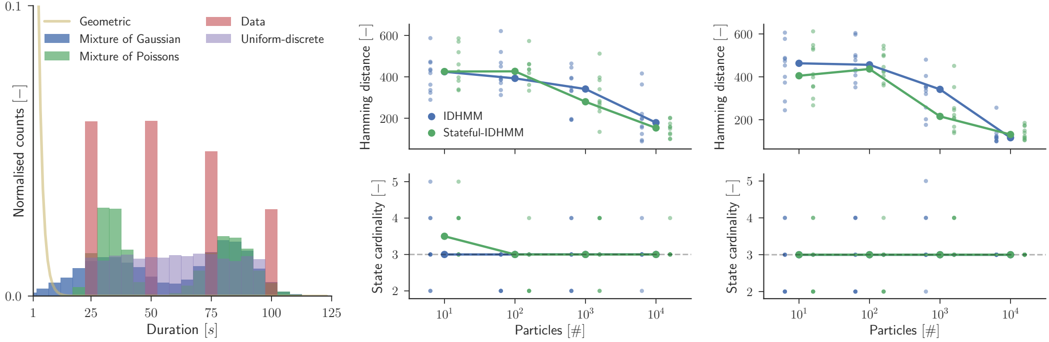


(a) Observations w. GT (b) Hamming distance (c) State cardinality

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

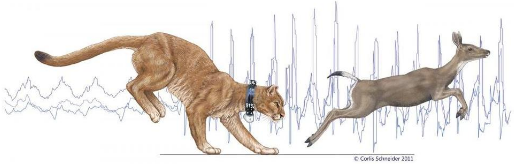


(a) Duration priors (b) Gaussian mixture duration (c) Poisson mixture duration

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

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

Labelling lion observations: fuzzy ground-truth

Labelling lion observations

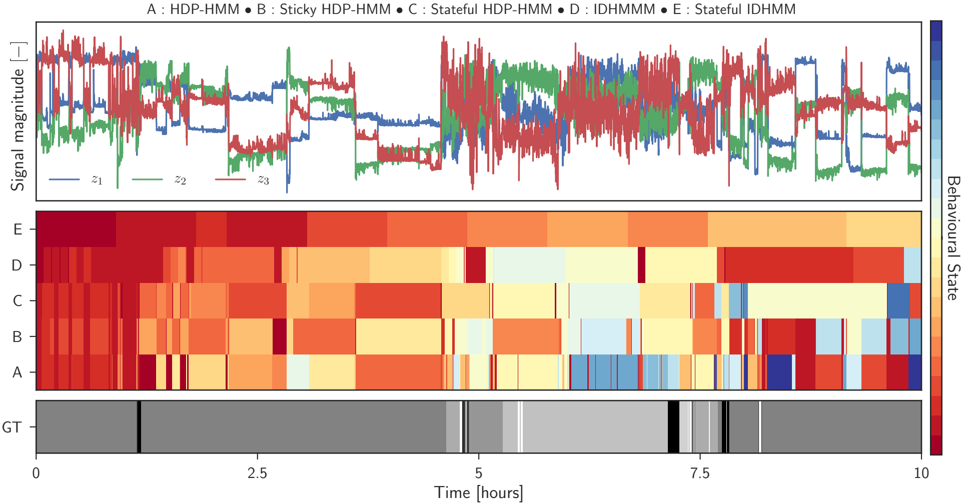


Figure: **top** – signal, **middle** – state sequences inferred by models through unsupervised learning and **bottom** – manually labelled fuzzy ground truth state sequence

Detailed analysis: a hunt

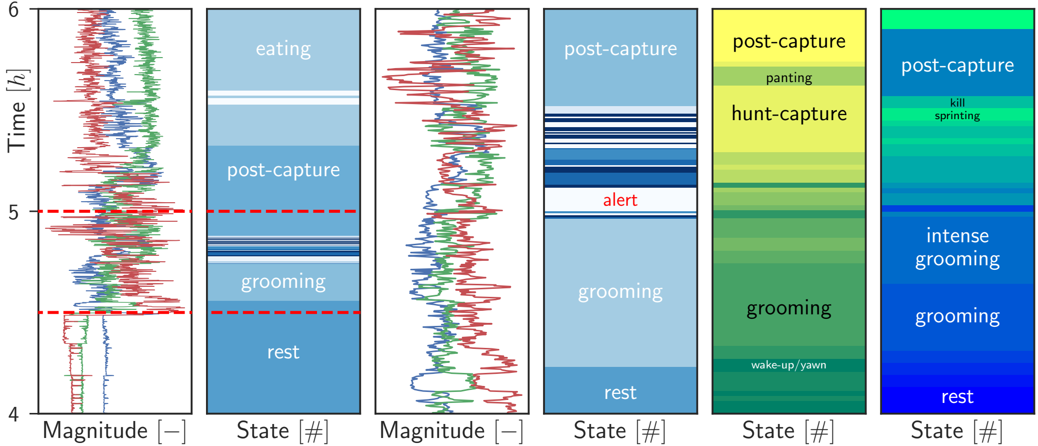
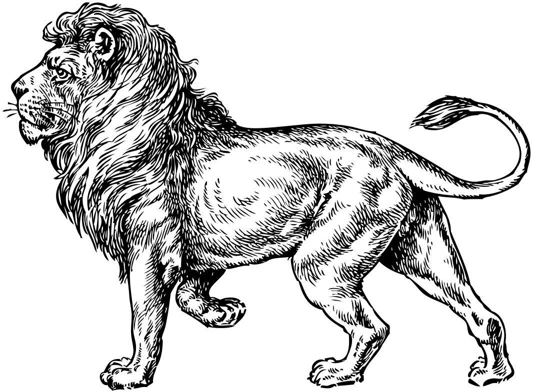


Figure: **first two panels** – signal and ground truth; **next two panels** – zoomed in; **final two panels** – assigned detailed labelling by IDHMM and stateful IDHMM as established by listening to audio, "concluding" that models learned *meaningful* new behaviours.

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?

References

David J Aldous. *Exchangeability and related topics*. Springer, 1985.

David Barber. *Bayesian reasoning and machine learning*. Cambridge University Press, 2012.

Christopher M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006. ISBN 0387310738.

Neil Dhir and Frank Wood. Improved activity recognition via kalman smoothing and multiclass linear discriminant analysis. In *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, pages 582–585. IEEE, 2014.

Neil Dhir, Frank Wood, Matthijs Vákár, Andrew Markham, Matthew Wijers, Paul Trethowan, Byron Du Preez, Andrew Loveridge, and David MacDonald. Interpreting lion behaviour with nonparametric probabilistic programs. UAI, 2017.

Thomas S Ferguson. A bayesian analysis of some nonparametric problems. *The annals of statistics*, pages 209–230, 1973.

Mitja Luštrek and Boštjan Kaluža. Fall detection and activity recognition with machine learning. *Informatica*, 33(2), 2009.

Peter Orbanz. Lecture notes on bayesian nonparametrics. *Journal of Mathematical Psychology*, 56: 1–12, 2012.

Yura N Perov. Applications of probabilistic programming (master's thesis, 2015). *arXiv preprint arXiv:1606.00075*, 2016.

Yee Whye Teh and Michael I Jordan. Hierarchical bayesian nonparametric models with applications. *Bayesian nonparametrics*, 1, 2010.