Learning transition times in event sequences: the Event-Based Hidden Markov Model

Peter A. Wijeratne,
Daniel C. Alexander

Email: p.wijeratne@ucl.ac.uk
Twitter: @WijDr

Connecting ideas from event-based and hidden Markov modelling to derive a new interpretable model of disease progression

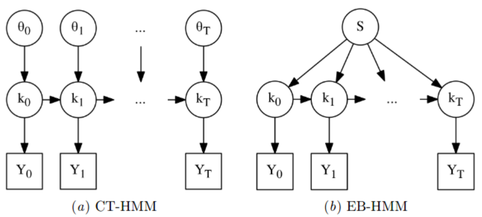
Progressive diseases such as Alzheimer's disease (AD) are characterised by monotonic deterioration in functional, cognitive and physical abilities over a period of years to decades. Data-driven models of disease progression can be used to learn hidden information, such as individual-level stage, from observed data. Here we address the problem of how to learn transition times in event sequences of disease progression, by introducing a new generative Event-Based Hidden Markov Model (EB-HMM) of disease progression. The main novelties of our work are as follows.

- We generalise a formerly cross-sectional model (the EBM: event-based model Fonteijn et al. 2012), allowing it to account for longitudinal data.
- We define a Bayesian 'event-based' framework to inject prior information into structured inference from longitudinal data.
- We use our model to learn a new clinically interpretable sequence and timing of events in AD and to predict individual-level trajectories.

Methods

The Event-Based Hidden Markov Model

To formulate EB-HMM, we make three assumptions, namely i) monotonic feature changes; ii) a consistent event sequence, S, across the whole sample; and iii) Markov (memoryless) stage transitions.



(a) CT-HMM (b) EB-HMM

Assuming independence between observed features i=1,...,I, if a patient j=1,...,J is at latent state $k_{j,t}$= 0,...,N at time t=1,...,T_j in the progression model, the likelihood of their data $Y_{j,t}$ is given by:

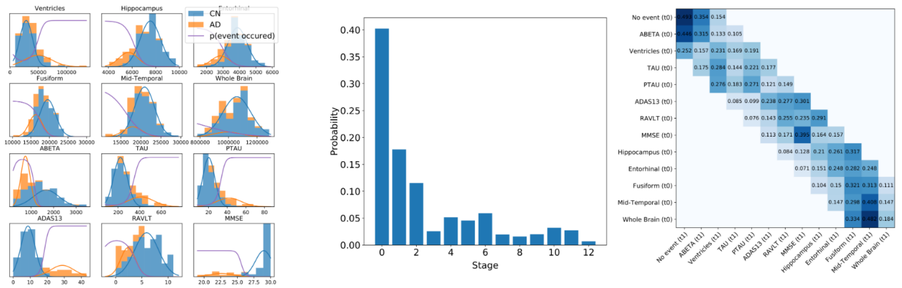
$$P(Y|\theta,S)=\prod_{j=1}^{J}\left[\sum_{k=0}^{N}P(k_{j,t=0})\prod_{t=1}^{T_j}P(k_{j,t}|k_{j,t-1})\prod_{t=0}^{T_j}\prod_{i=1}^{k_{j,t}}P(Y_{i,j,t}|k_{j,t},\theta_i^p,S)\prod_{i=k_{j,t}+1}^{I}P(Y_{i,j,t}|k_{j,t},\theta_i^c,S)\right]$$

Here θ^p_i= [μ^p_i,σ^p_i,w^p_i] and θ^c_i=[μ^c_i,σ^c_i,w^c_i] are the mean, μ, standard deviation, σ, and mixture weights, w, for the patient and control mixture model distributions, which define the event-based model. For a full derivation see Wijeratne & Alexander, 2020.

Results

Inferring the timeline of feature changes in Alzheimer's disease

We use data from the ADNI study, a longitudinal multi-centre observational study of AD. We select 468 participants (119 CN: cognitively normal; 297 MCI: mild cognitive impairment; 29 AD: manifest AD; 23 NA: not available), and three time-points per participant (baseline and follow-ups at 12 and 24 months). Individuals were allowed to have missing data at any time-point.



(a) Mixture models (b) $\overline{\pi}$ (c) $\overline{Q}$

Figure 1: EB-HMM parameters inferred from ADNI. (a) Gaussian mixture model fits to distributions of CN and manifest AD groups. (b) Initial probability density, π, inferred by EB-HMM. (c) Event-based transition matrix, Q, inferred by EB-HMM. Events are ordered by the maximum likelihood sequence, S.

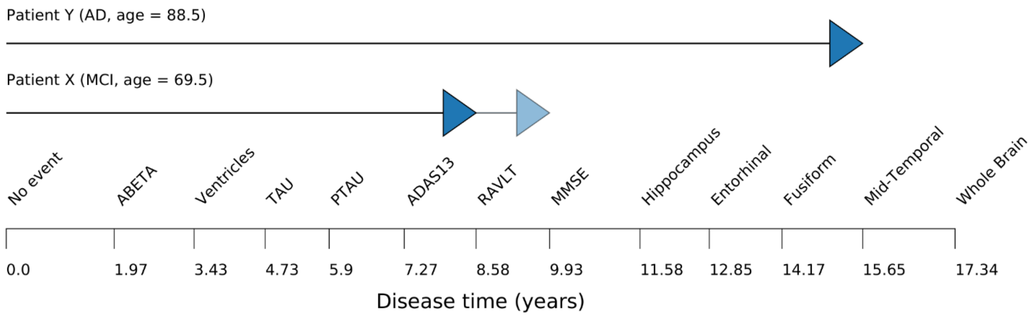


Figure 2: AD timeline inferred by EB-HMM. The order of events on the horizontal axis is given by the maximum likelihood sequence S, and the time between events is calculated from Q. Baseline stage (solid arrow) and predicted next stage (shaded arrow) estimated by EB-HMM for two example patients are shown, chosen from the MCI and AD sub-groups.

Model	AU-ROC
EB-HMM (full)	$\mathbf{0.804 \pm 0.07}$
EB-HMM (subset)	0.737 ± 0.09
CT-HMM (subset)	0.579 ± 0.12

Table 1: EB-HMM improves predictive utility over a standard continuous time hidden Markov model (CT-HMM). Performance for the task of predicting conversion, using either the full data (including individuals with missing data) or subset data (only individuals with complete data).

% missing	AU-ROC
25%	0.722 ± 0.09
50%	0.719 ± 0.13
75%	0.669 ± 0.15

Table 2: EB-HMM maintains performance with missing data. Performance for the task of predicting conversion with % missing data.

Discussion

A new interpretable model of disease progression

A key corollary benefit of EB-HMM's formulation is that it can infer probabilistic estimates of group- and individual-level progression from datasets with missing data, both in terms of observed features and time-points. This gives EB-HMM high utility in clinical applications where resources are scarce and/or it is too expensive to observe a patient multiple times, making EB-HMM an ideal tool for advancing on the objective of accessible healthcare. Future work with EB-HMM will be focused on relaxing its assumptions, in particular allowing for non-monotonic trajectories and multiple event sequences (subtypes).

References

Fonteijn et al. (2012). "An event-based model for disease progression and its application in Alzheimer's disease and Huntington's disease." In: NeuroImage 60(3). pp. 1880-9.
Wijeratne & Alexander. (2020). "Learning transition times in event sequences: the Event-Based Hidden Markov Model of disease progression." In: ML4H Extended Abstract arXiv Index. https://arxiv.org/pdf/2011.01023.pdf