A Bayesian Linear Mixed Model for Bi-Level Feature Selection

Applications to an Alzheimer's Disease Progression Study

Daniel R. Baer[1], Andrew B. Lawson[2], Yeonhee Park[3], Sharon X. Xie[1], Andreana Benitez[4], and for The Alzheimer's Disease Neuroimaging Initiative

[1]University of Pennsylvania, Department of Biostatistics, Epidemiology and Informatics, Philadelphia, PA, [2]Medical University of South Carolina, Department of Public Health Sciences, Charleston, SC, [3]University of Wisconsin–Madison, Department of Biostatistics Medical Informatics, Madison, WI, [4]Medical University of South Carolina, Department of Neurology, Charleston, SC

Background

- Alzheimer's disease (AD) is a neurological disorder with an increasing rate of prevalence in the United States.
- There is currently no cure for AD, and there is therefore interest in characterizing AD in order to develop novel AD therapies.
- Characterization of AD can be facilitated by Bayesian feature selection models, which allow us to identify patient feature data that are associated with longitudinal AD outcome data.
- However, current Bayesian feature selection models are limited by the complexities of data arising in the longitudinal study of AD.
- In particular, there are no Bayesian feature selection models that can simultaneously:
  - account for irregularly-spaced longitudinal outcome data,
  - account for feature data group structure,
  - and specify time-varying feature parameters.
- We believe that accounting for these data complexities can lead to a feature selection model with superior performance.

Methods

- We therefore propose a Bayesian linear mixed model for bi-level feature selection which addresses these complexities:

$$\begin{bmatrix} \mathbf{y}_i|\mathbf{x}_i, \mathbf{B}, \sigma^2, \mathbf{Z}_i, \mathbf{b}_i \\ \mathbf{b}_i|\mathbf{G} \end{bmatrix} \overset{ind.}{\sim} N_{J+q}\left(\begin{bmatrix} \mathbf{B}^T\mathbf{x}_i + \mathbf{Z}_i\mathbf{b}_i \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \sigma^2\mathbf{I}_J & \mathbf{0} \\ \mathbf{0} & \mathbf{G} \end{bmatrix} \right), \quad (1)$$

$$\begin{aligned} &vec\left(\tilde{\mathbf{B}}_k^T\right)|\boldsymbol{\Sigma}, \pi_{0k} \overset{ind.}{\sim} \\ &\pi_{0k} N_{p_k J}\left(\mathbf{0}, \mathbf{I}_{p_k} \otimes \boldsymbol{\Sigma}\right) + \left(1 - \pi_{0k}\right)\delta_0\left(vec\left(\tilde{\mathbf{B}}_k^T\right)\right), \end{aligned} \quad (2)$$

$$\tau_l^2|\pi_{0l}, s^2 \overset{ind.}{\sim} \pi_{0l} N_1^+\left(0, s^2\right) + \left(1 - \pi_{0l}\right)\delta_0\left(\tau_l^2\right). \quad (3)$$

- The algorithm used to fit our model is available in R via our GitHub repository:

$$devtools :: install\_github\left('danielrbaer/BayesianLMMFS'\right). \quad (4)$$

- By the inclusion of random effect parameters, our model can also identify subjects with the greatest propensity for AD progression.

Results

- We evaluated our model via a simulation study and an analysis of longitudinal AD outcome data from the AD Neuroimaging Initiative (ADNI).
- Our model was compared to a bi-level feature selection model which does not account for irregularly-spaced longitudinal outcome data:

$$\mathbf{y}_i|\mathbf{x}_i, \mathbf{B}, \boldsymbol{\Sigma} \overset{ind.}{\sim} N_J\left(\mathbf{B}^T\mathbf{x}_i, \boldsymbol{\Sigma}\right). \quad (5)$$

- Relative to the competing model, our model provided markedly narrower posterior credible intervals (CIs) for the selected feature parameters
  - These CIs were upwards of 20% narrower (e.g. figure 2).
- Advantageously, our model also allowed us to identify subjects with the greatest propensity for AD progression (e.g. figure 3).

Conclusions

- Our model identifies a parsimonious subset of patient feature data associated with AD risk.
- Moreover, by accounting for the aforementioned data complexities, our model provides improved precision of feature parameter estimates.
- Therefore our model is well suited to perform feature selection given data complexities arising in the longitudinal study of AD.

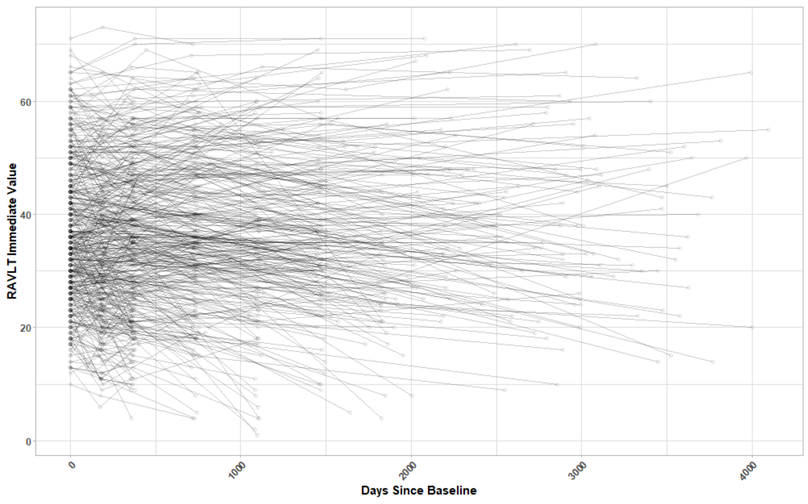

Figure 1:The Rey's Auditory Verbal Learning Test (RAVLT) Immediate longitudinal outcome data consisting of $n = 451$ and $J = 3$ from the ADNI. Note that lower values of RAVLT Immediate indicate worse patient memory.

Feature (l)	Feature Group (k)	p_k	Feature Description
ABETA	CSF biomarkers	3	Amyloid beta peptide
TAU	CSF biomarkers	3	Total tau protein
Age	Demographics	5	Age
Education	Demographics	5	Education level (in decades)
Sex	Demographics	5	Sex (female vs. male)
Race	Demographics	5	Race (white vs. other)
Marriage	Demographics	5	Marital status (married vs. other)
MCI/Dementia	Neurological diagnoses	1	MCI or Dementia vs. cognitively normal diagnosis
APOE4 1/2	Genetic markers	1	1 or 2 vs. 0 APOE-ϵ4 alleles
Hippocampus	MRI measurements	7	Volumetric quantification of the hippocampus
Ventricles	MRI measurements	7	Volumetric quantification of the ventricles
Whole Brain	MRI measurements	7	Volumetric quantification of the whole brain
Entorhinal	MRI measurements	7	Volumetric quantification of the entorhinal cortex
Fusiform	MRI measurements	7	Volumetric quantification of the fusiform gyrus
MidTemp	MRI measurements	7	Volumetric quantification of the middle temporal gyrus
ICV	MRI measurements	7	Intracerebral volume
FDG	PET measurements	2	Average FDG PET of the angular, temporal, and posterior cingulate
AV45	PET measurements	2	Average Florbetapir F 18 PET of THE whole cerebellum

Table 1:Baseline feature data from the ADNI; we consider a total of $p = 18$ features partitioned into $K = 6$ feature groups. Abbreviations: MCI: mild cognitive impairment, APOE: Apolipoprotein E, and FDG: Fluorodeoxyglucose.

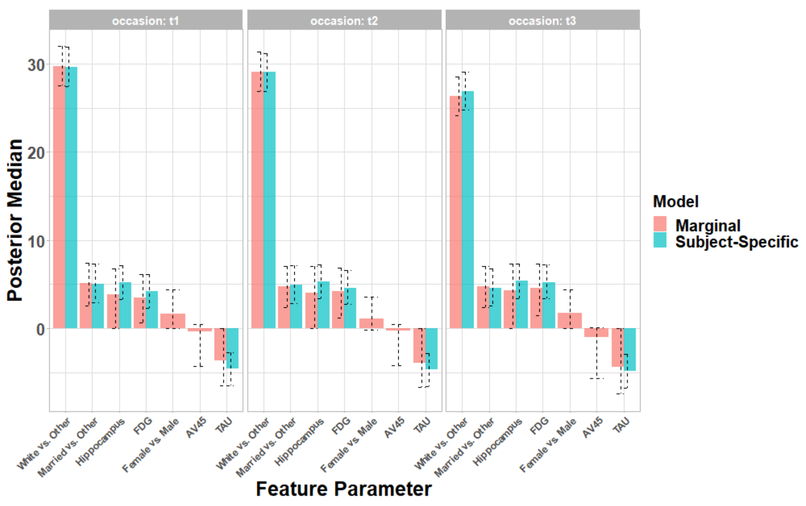

Figure 2:The posterior medians of the selected features from our analysis of the RAVLT Immediate longitudinal outcome data over the $J = 3$ measurement occasions. The errorbars here represent the 95% equal-tailed credible intervals. We observe that our model provides more precise estimates of important feature parameters associated with AD risk.

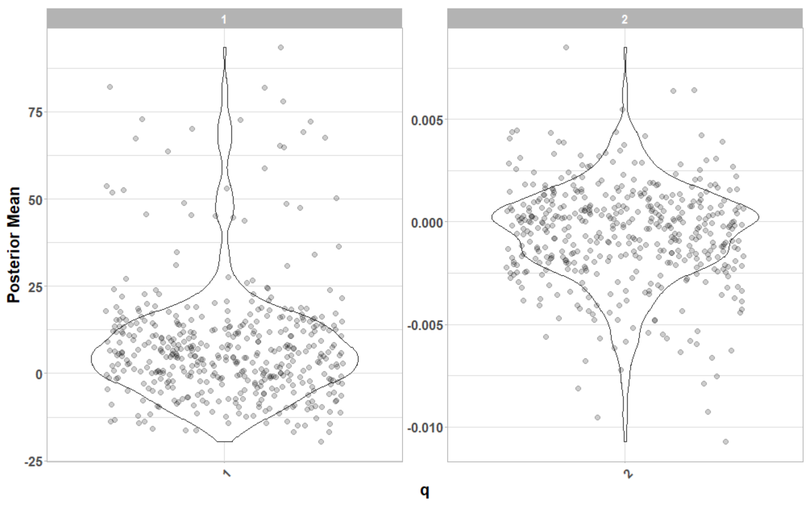

Figure 3:The posterior means of the random effect parameters, $\mathbf{b}_i$, from our ADNI analysis; $q = 1$ denotes the random intercept parameters, and $q = 2$ denotes the random slope parameters. Subjects with negative random slope parameters ($q = 2$) are estimated to have the greatest propensity for AD risk over time.