An Analysis of Activity Changes in MS Patients: A Case Study in the Use of Bayesian Nonparametrics

Finale Doshi-Velez
MIT
Cambridge MA 02139
finale@mit.edu

Nicholas Roy
MIT
Cambridge MA 02139
nickroy@mit.edu

Introduction and Research Objectives. In the fall of 2009, the Boston Home Team at MIT attached wifi-enabled tablets to the wheelchairs of twenty residents at The Boston Home (TBH), an eldercare facility specializing in multiple sclerosis. Since then, all access points visible to each tablet and their signal strengths, have been logged every 30 seconds between 6am and 11pm each day. During the last 18 months of operation, the care for several of these residents has changed. Our system has experienced several outages, and some access points at the facility have been replaced.

Our goal was to automatically recover these resident and system changes from these longitudinal data, as well as characterize the nature of the change. For example, if we discover that a resident is spending more or less time in their room, it may reflect a change in their health. Bayesian nonparametric methods seem to be an appropriate fit for this kind of analysis because they can both characterize changes at an "appropriate" level of granularity and provide quantifications of uncertainty for subtle changes. In the following, we focus on analyzing the data in an offline setting; being able to flag potential changes in near realtime is subject for future work.

Models and Methods. We focused on five residents. For each resident r, we aggregate the data into a $T \times N$ matrix X_r, where $T = 82$ weeks of data collection between December 2009 and July 2011 (we summed over days of the week to avoid weekday/weekend effects). The data dimension $N = 6 \cdot 52$, where 52 is the number of access points at TBH, summed over 6 2-hour time blocks between 9am-9pm (we considered 2-hour blocks to avoid effects due to variation in daily habits).

We compared five methods to find and characterize changes in these data. The first, hierarchical clustering (HC), combined adjacent clusters if their means differed by less than certain threshold. Bayesian clustering (BC) used a Dirichlet process to do the clustering (inference: 1000 iterations of collapsed MCMC). The next three methods found low-rank factorizations of the data matrix X_r as the product $Z_r A_r$, a $T \times K$ weight matrix Z_r multiplied by a $K \times N$ activity matrix A_r. We compared a Bayesian nonparametric approach (BNP) to two very simple matrix factorization models, principle components analysis (PCA) and non-negative matrix factorization (MM) set to 5 activities [1], that used a simple difference-of-differences threshold for detecting changes.[1]

Our Bayesian nonparametric model is an extension of the Indian Buffet Process Compound Dirichlet Process [2]. The generative process placed an independent exponential prior on the activity matrix A_r and an independent Gaussian noise ϵ on the factorization $X_r = Z_r A_r + \epsilon_r$. Independently for each resident, the 'customer' representing the first time-step, samples a set of IBP sticks π_k representing the probability of activity k being present and a set of gamma-distribution parameters α_k, β_k representing the distribution of weights assigned to an activity if it is present. The customer then samples values $z_{1k} = \texttt{Be}(\pi_k)\texttt{Gamma}(\alpha_k, \beta_k)$ for each activity k. These values are normalized across the row z_1 before being multiplied by the activity matrix A_r. The later customers sample their values of z_n using the parameters π_k, α_k and β_k. At each time-step, with probability p_r, a new leader enters the particular resident's buffet and samples new values for π_k, α_k, and β_k (representing

[1] We also investigated the use of more standard permutation tests and parametric tests with multiple hypothesis (BH) corrections, but 82 weeks was too short a sequence to use these to identify multiple changes.

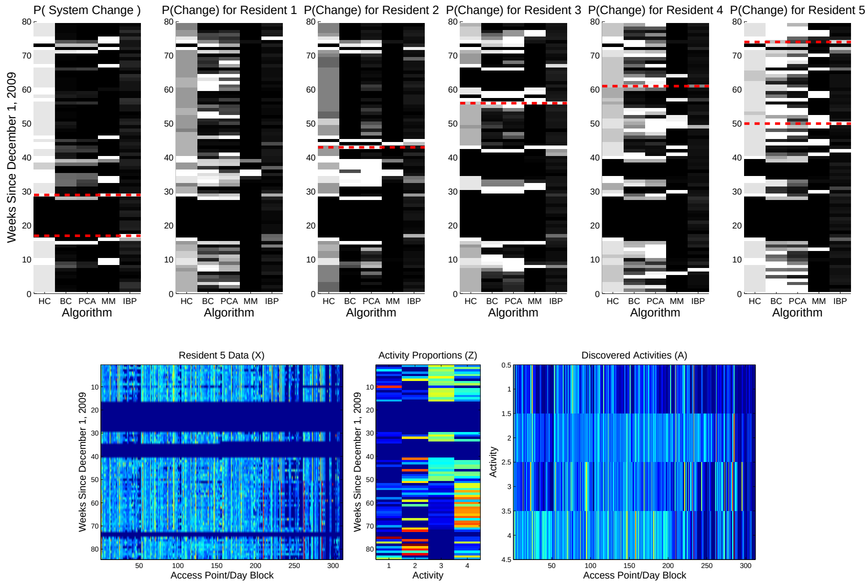

Figure 1: Change-points found by the five approaches (above), example BNP factorization (below).

a resident-level change). With probability p_s, a new leader enters all the restaurants (a system-level change). Inference was performed with a combination of Gibbs sampling and Metropolis-Hastings on a weak-limit approximation with all of the hyper-parameters integrated out (500 iterations).

Results and Discussion Running in Matlab, with fairly optimized code, HC took 9.3 seconds; BC, 262 seconds; PCA, 4.1 seconds; MM 5.6 seconds; and the BNP approach 4900 seconds. Figure shows the change-points detected by the different approaches. Lighter means more likely. For HC, more "likely" was based on different cut-points on the inferred tree. For PCA and MM, more "likely" was derived from the magnitude of the difference-of-differences. The red dashed lines show known system and resident changes from discussions with a subset of the TBH staff.

All of the approaches have some noise, but the sparsity encoded in our BNP model helped it find known change-points with few false-positives, even in these relatively short sequences of only 82 weeks (we are now investigating whether the additional change-points found by BNP correspond to real events). In both MM and BNP, Z_r is a positive weight matrix; looking for changes in the low-rank space of activity weights Z_r, instead of directly on the data matrix X_r, also seemed to help identify changes more accurately. The BNP posterior found 3-8 activities per resident.

Figure shows one factorization. The data (left) was factored into proportions (center) and activities (right). Initially the resident has mixture of activities 3 and 4. The sparser entries in activity 3 correspond to the resident spending more time in one place (validated using additional scan data). The much less sparse activities 2 and 4 correspond to spending more time at certain activities during the early part of the day, activity 1 corresponds to a different lunch location. The increasing proportion of activity 4 and the appearance of activities 1 and 2 in the resident's week correspond to known changes in socialization following a successful care plan.

Overall, we observed both the claimed advantages and drawbacks of Bayesian nonparametric approaches in this case study. The BNP approach matched well with known change-points, resulted in interpretable activities, and its internal measure of uncertainty proved to be more accurate than the heuristics applied to the other approaches. However, the BNP hyper-hyper parameters required some (non-intuitive) tuning before the model "automatically" discovered the appropriate structure. The simpler factorizations also found similar gross structure much faster, emphasizing the gap between inference used in Bayesian nonparametrics and more "engineering" approaches.

References

[1] D. Lee and H. Seung. Algorithms for non-negative matrix factorization. In *NIPS*, 2001.

[2] S. Williamson, K. Heller, C. Wang, and D. Blei. The IBP compound Dirichlet process and its application to focused topic modeling. In *ICML*, 2010.