# Latent Features and Manifolds

## Non-Parametric Bayesian

In this article we can observe in the findings that the several key aspects of existing marital interaction models are inadequate. We review how this unsupervised learning technique generates plausible dyadic sequences that are sensitive to relationship quality and provide a natural mechanism for computational models of behavioral and affective micro-social processes.

Decades of marital interaction research demonstrate an unequivocal relationship between moment-to-moment behavioral exchanges seen in a conversing couple and the quality of their relationship: each gesture and word is cradled in affect; each nuanced pause conveys a semi-private message understood only to the inhabitants of the relationship and the entire process is uniquely sensitive to the situation. The language is public but the conversation is private. And despite almost forty years of investigators dissecting sequential verbal and non-verbal data, the micro-social processes that critically determine and reflect marital quality remain opaque.

Over the last few decades most of the significant insights into couple dynamics arose not from better data or measurement methods but improved data analytic strategies. The area moved steadily from summary statistics describing individual behavior during an interaction (e.g., mean number of positive statements) to probability estimates of shifting dyadic structures (e.g., sequential analysis) to state transition models (e.g., Hidden Markov Models).

Despite these analytic advances, it’s commonly acknowledged that the algorithms that generate the evolving contingency structures embedded in sequential data are not well understood. Even with the abundance of real-time couple interaction data from multiple labs throughout the world there are no published generative models of dyad dynamics and certainty none that discriminate by relationship satisfaction level. To construct realistic and informative generative models of dyadic interactions three analytic problems need addressing:

(1) how to articulate the state space of the dyad.

(2) how to generate a tenable state transition matrix.

(3) how to incorporate duration expectancy into states and transitions.

To advance our understanding of intimate dyadic affective and behavioral structures we need generative models of sequential latent processes, i.e., profiles of sequential movement across latent states with estimated probabilistic structures.

Recently HMM (Hidden Markov Models) methodologies began incorporating the capacity of nonparametric Bayesian approaches to define prior distributions on transition matrices over countably infinite state spaces; adopting this technique allows a greater range of use with real, somewhat messy, data. Contemporaneously, Fox and colleagues began developing algorithms that parameterize the likelihood of state self-transitions; this, combined with the use of a Hierarchical Dirichlet process to generate priors while leaving unspecified the expected number of states, permits the modeling of duration sensitive latent state transitions (i.e., semi-Markovian processes). Such models are more realistic of natural dynamic stochastic processes.

Borrowing heavily from the innovations of Fox and colleagues along with the recent work by Johnson and Willsky [, we use a explicit-duration HDP-HSMM of couple affect dynamics to build generative models illustrating differential affect patterns in marital couples classified by self-report satisfaction. This paper illustrates how this methodology can capture pertinent sequential dyadic state dynamics and accurately discriminate affective processes associated with relationship satisfaction.

Fortunately, with increased investigations of real-world complex datasets, the standard HMM has been transformed over the last decade—computer scientists created numerous sequential analytic techniques that are sensitive to the nuances of evolving latent structures (e.g., infinite HMMs) akin to the type seen in micro-social dynamics. Among these is the HDP-HMM. The hierarchical Dirichlet process (HDP)–a nonparametric technique using the Dirichlet process (i.e., a distribution over distributions)–models the dependence among groups through sharing the same set of discrete parameters. Yet its assumption of exchangeability makes it inappropriate for sequential data; fortunately, this restrictive assumption spurred new models that are appropriate for time sensitive data.

The goal of extracting patterns from sequential data is conceptually and quantitatively similar to their search for structure in real and synthetic data. Whereas they used multiple time-series of household appliance data, we inserted husband and wife sequence data, and have added the additional dimension of relationship satisfaction as a classification problem.

For this study we show that integrating temporal and event data with modern machine learning—techniques provides a novel way of capturing the nuances of sequential dyadic data that differentiate couples by levels of satisfaction and provide the foundation for building robust computational models in both areas.

## Materials and Methods

To use the HDP-HSMM methodology required that we partition the couples into homogeneous sub-groups. To do this, we moved away from the traditional reliance on self-report data and developed a dyad feature set which allowed us to use hierarchical clustering techniques to construct the sub-groups.

From a list of typical areas of conflict in a marriage, each marital partner selected and ranked three topics that he or she thought was most problematic in their relationship. With the help of a lab assistant the couples then negotiated on the top three areas from their joint list. Prior to beginning an interaction based on these topics the couple moved to separate section of the laboratory and was shown how to use the affect rating software. After becoming familiar with the software, they returned to their chairs and the lab assistant, prior to leaving the room, instructed the couple to attempt to resolve the topics while engaging in a 12-minute discussion.

Spouses were separated immediately after the conversation and taken to another section of the lab where each individual simultaneously rated his or her own affect during the interaction while viewing a split-screen playback.

In this method of metric retrieval, each affect has a subjective reference that is unique to the rater, within the context of the interaction, given the dyad's history. For each individual there is only an internal template referencing their affective state; an internal state that is pleasant to one individual may be only neutral to another. Moreover, because it is self-report, it arguable that such a recall procedure provides a good proxy of the true affect state, and requires less inference than other, outsider perspective data collection procedures. This method of assessing affect effectively discriminates, by sex, the propensity to exit negative states, and Griffin, using Hidden Markov Models, found that affect ratings and their durations successfully discriminated distressed and nondistressed couples.

For analyzing dyadic interaction, either, e.g., mother-child or spousal interactions, the capabilities of the explicit-duration HDP-HSMM extends the traditional Hidden Markov Model in two fundamental ways: (1) by incorporating varying state durations, micro-social event dynamics are not constrained to a geometric form—acknowledging that time in state makes a difference in sequential behavior; and (2) by allowing a countably infinite number of states the model incorporates dyadic histories with ideographic state spaces—each dyad has a unique number of states that best capture their behavioral propensities.

Models were selected by minimizing the difference between the realized and simulated data. Specially, model selection was done by comparing and selecting the best fitting model derived from 50,000 Gibbs samples over a parameter sweep of values that initial analyzes suggested were most sensitive to model fit.

Model selection was determined by assessing sequence dissimilarity between the model simulated sequence and the raw data for each couple within a cluster; best fit was derived by averaging the 5 waves of each parameter configuration and the combination with the smallest average difference was considered the best model. We used the States values for comparison.

Each feature (*MAT Score*, *States mean*, *States standard deviation*, *Shannon entropy*, *Dynamic Time Warping*) was initially normalized (0,1) using the MinMax method. After normalizing the data, a series of pair-wise euclidean distance estimates were taken (row 1 vs row 2, etc). This normalized distance matrix was then used to construct clusters.

We illustrate how contemporary Bayesian extensions of traditional HMMs can generate realistic multidimensional (i.e., dyadic) interactions. These new methods add realism by incorporating state durations along with the ability to model a countably infinite number of states. These HDP-

HSMMs permit investigators to model complex social processes that can vary substantially as a function of the relationship quality of the interactants. Our current study had two objectives: First, we sought to classify dyad dynamics, using self-report affect, into homogeneous groups beyond what had been obtained in the existing literature in this area; Second, we sought to build generative models of these dynamics using contemporary Bayesian machine learning techniques.

Next, we used realized data extracted from each cluster to build a HDP-HSMM. These generative models provided a simulation of the captured data as well as a generic model of the dyadic interactions observed in the cluster. Not surprisingly, clusters having greater entropy fit less well than lower entropy clusters, as measured by their Hamming distance. This suggests that affect variability associated with specific satisfaction types influences model fit.

These results have a clear implication for subsequent work in Affective Computing—among some actors in an intimate relationship, even those reporting high levels of attraction or satisfaction, sometimes generate elevated and variable levels of negativity. Prior research in the area of couple dynamics report, or assumed, that highly satisfied couples were generally homogeneous and characterized by consistent and high levels of positive behavior and affect. This assumption is inconsistent with the clusters derived from the current data. The HDP-HSMM effectively reproduced the realized data across all cluster types, even those with greater affect variability (e.g., Clusters 1, 2, 3). These generative models demonstrated it is possible to take multi-dimensional data reflecting social processes and reasonably recreate complex interactions.

Not surprisingly, our results suggest that well-behaved data are easier to model, yet even the poorer fitting models did an acceptable job of capturing duration sensitive state transition dynamics.