Week 4

Github URL: https://github.com/nevilshah444/CMPE-257/tree/main/Assignment_4
Colab Link:
https://colab.research.google.com/drive/1vUKmYkVTpOeahiCZCMe5dDVJ2kCnJy85#scrollTo=bmXUjmWQcqoK

Bayesian Nonparametrics

In fact, and predictably, these findings show that marriage quality impacts the affective producing process, which then manifests the observed dyadic behaviors.
However, no attempt has been made to articulate a catholic analytic solution allowing investigators to reanalyze existing data in the hopes of uncovering additional crucial tempo-behavioral facets linked to self-report satisfaction in intimate dyads; nor has there been any attempt to articulate a catholic analytic solution allowing investigators to reanalyze extant data.

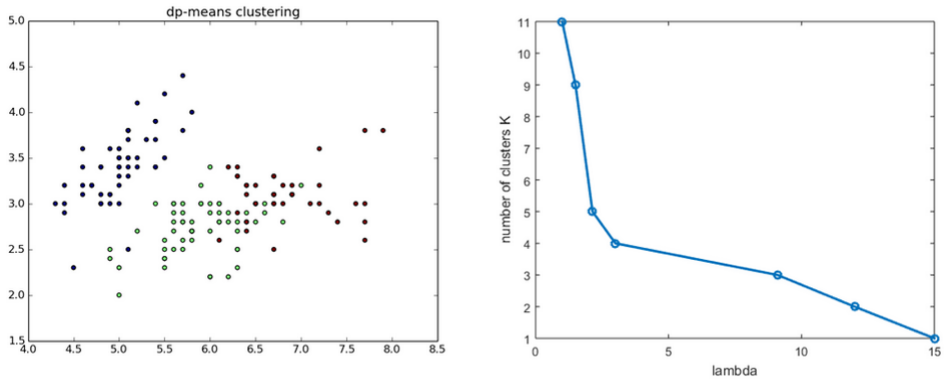
Three analytic problems must be addressed in order to construct realistic and informative generative models of dyadic interactions:

- how to articulate the dyad's state space.
- how to generate a tenable state transition matrix.
- how to incorporate duration expectancy into states and transitions.

Bayesian Non - parametric models are those in which the number of parameters expands as the data grows. Non-parametric K-means clustering [1] is a simple example. Rather than deciding on a fixed number of clusters, we let the data decide on the optimal number of clusters. We can better characterize the data and produce new data given our model by allowing the number of model parameters (cluster means and covariance matrices) to expand with the data.

Of course, we penalize the number of clusters K via a regularization parameter that governs the pace at which new clusters are produced to avoid over-fitting. As a result, our new K-means goal is:

$$\min \sum_{c=1}^{K} \sum_{x \in l_c} ||x - \mu_c||^2 + \lambda K$$



Fortunately, understanding latent generating processes has been a cornerstone of modern machine learning theory during the last decade. The emphasis switched toward identifying methods that capture temporal grouping and feature extraction in sequential data when faster computers and theoretical improvements in Bayesian approaches became available. The non - parametric Bayesian approaches are at the forefront of this field. For example, investigators can use Hierarchical Dirichlet Process approaches to divide data into themes and sequence sensitive feature states Similarly, recent attention in the classic Hidden Markov Model (HMM) has led in the development of many multi-state, hierarchical models that may describe time-sensitive latent state transition processes.

In Bayesian non - parametric models, the Dirichlet process (DP) is a stochastic process. A discrete distribution is each drawn from a Dirichlet process. The finite dimensional marginal distributions of a random distribution G must be Dirichlet distributed in order for it to be distributed according to a DP. Let H denote a theta distribution and alpha a positive real integer.

Dir is a Dirichlet distribution, which is defined as:

$$p(x_1, ..., x_k) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_{k=1}^{K} x_k^{\alpha_k - 1}$$

Observations:
A continuous Gaussian distribution was used to model observations (i.e., impact ratings).
We employed the normal-inverse-Wishart distribution, a multivariate four-parameter (= prior mean, = scale matrix, = prior observations, v = degrees of freedom) family of continuous probability distributions, to produce the Gaussian distributions.

Two of the four hyper - parameters are set at initialization and have only minor influence on the model:
The other two, and (covariance), were formed from realized data for each cluster, where denotes the number of dimensions (i.e., male and female affect ratings) * 2; the other two, and (covariance), were developed from realized data for each cluster.

Durations:
State duration were modeled as Poisson distributed with parameter ; suitable values for a Gamma conjugate prior were determined using hyperparameters (= shape, = scale). The gamma parameters for each cluster were computed using L-moments [48] from realized data utilizing the durations of each States state.

States:
The multinomial distribution of states was created using a single parameter Dirichlet Process conjugate prior with an initial concentration parameter.

Transitions:
Across couples, state transition distributions were represented using a Hierarchical Dirichlet Process (HDP).
The initial generation of the HDP model was controlled by two hyperparameters: and, where was the hyperparameter employed in the stick breaking distribution (denoted as GEM) to create the random base measure [49].
The transition matrix was constructed as a Dirichlet Process at each pair level, and they were all linked together via the HDP via, where GEM().

Model Selection

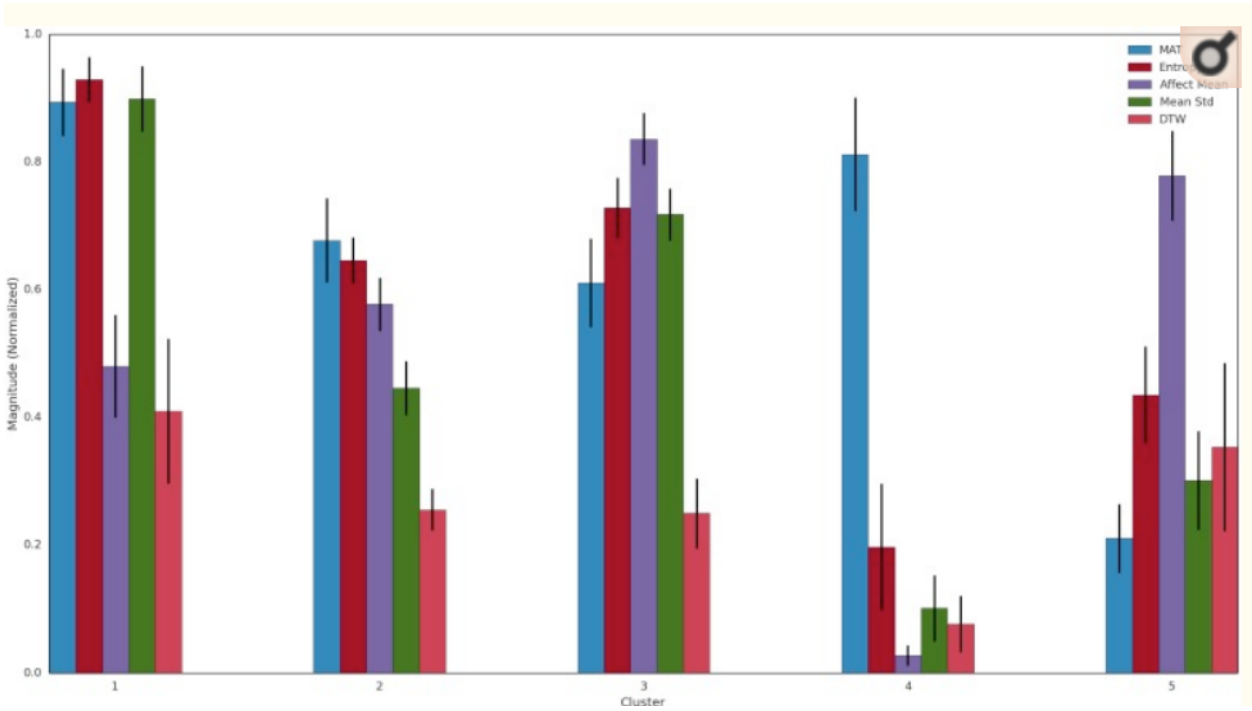
The best fit was obtained by averaging the 5 waves of each parameter configuration, with the combination with the least average difference being deemed the best model.
For comparison, we utilized the values from the United States.
TraMineR [50], an R-based program for estimating differences across sequential categorical data, was used to do the comparisons.
We employed Hamming distance and Dynamic Hamming Distance (DHD), which measures distance by discovering contemporaneous similarities and producing substitution costs as a function of transition density.

Clusters

The MinMax approach was used to normalize (0,1) each characteristic (MAT Score, States mean, States standard deviation, Shannon entropy, Dynamic Time Warping).
A set of pair-wise euclidean distance estimations were taken after normalizing the data (row 1 vs row 2, etc).
Clusters were then built using the normalized distance matrix.

We used hierarchical clustering approaches to build clusters, alternating between average linkage and Ward's method [52]; the final number of clusters ranged from 4 to 5, and both strategies produced comparable, but not identical clusters.
The average linkage approach with 5 clusters was slightly different from Ward's method for a single pair; it resulted in a singleton, whereas Ward's method resulted in the identical dyad being put into an existing cluster.
Using Ward's technique, we eventually landed on the 5-cluster model.

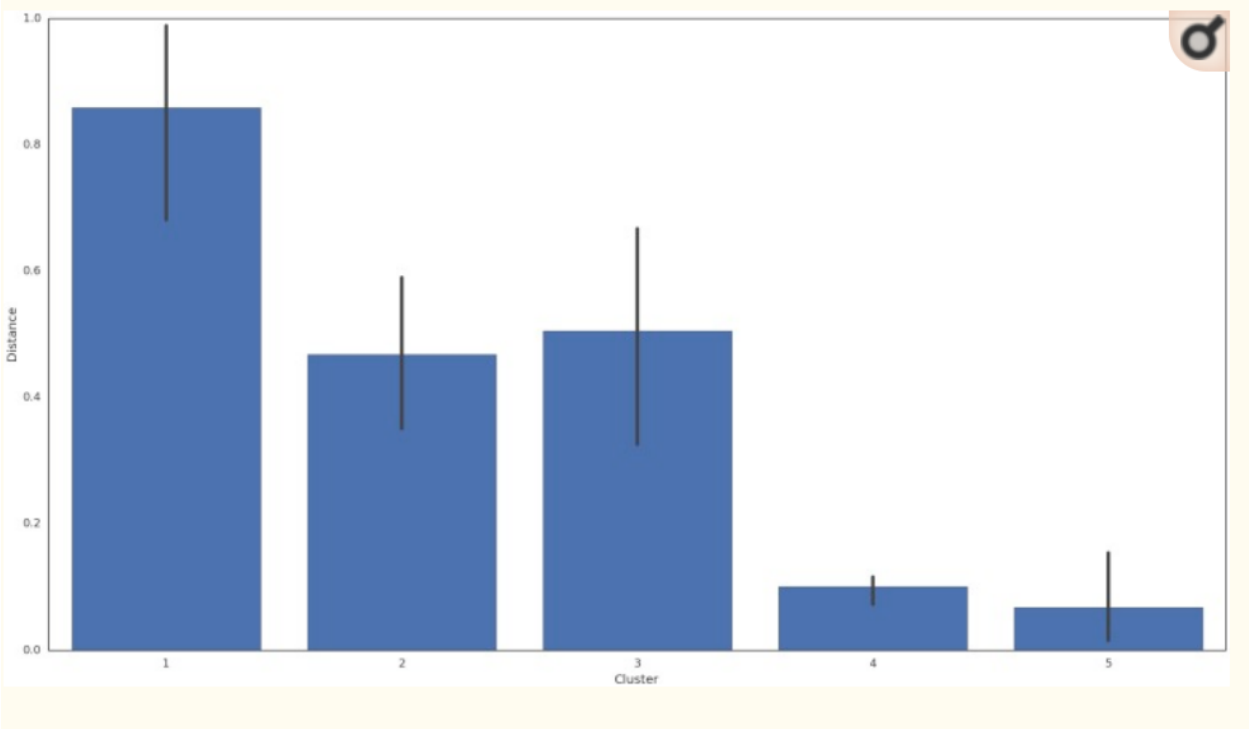
The final clusters had sample sizes ranging from three couples (Cluster 4) to ten (Cluster 2), and in terms of marital satisfaction, there was one low (Cluster 5), two medium (Clusters 2,3), and two high satisfaction groups (Clusters 1,4).
The validity of the clusters was checked using a variety of approaches, both intuitive and statistical.



HDP-HSMM Compatibility:

We only give the Hamming distance because a review of the findings revealed little variations between the Hamming and Dynamic Hamming distances.
All of the best fit combinations, with the exception of one, were with 20 States.
Aside from the maximum number of States, no other single parameter (e.g., or in the transition distribution) predominated using the best fitting model criterion, either at the couple level or within Cluster.

For the Distance estimations, one apparent trend emerged: Cluster characteristics affect the amount of the average distance between realized and simulated data. Cluster 5's low satisfaction couples and Cluster 4's conventional high pleased couples had the best simulation estimates, as shown in Fig 7, whereas Cluster 1's highly satisfied yet negativity laced couples had a significantly lower match.
Clusters 2 and 3, which had moderate levels of satisfaction, had estimations that were in the middle of the range.
Cluster 1 distance estimates were nearly 7 times greater than Cluster 4 and 9.15 times greater than Cluster 5. Despite the fact that Cluster 1 and Cluster 4 both contain highly satisfied couples, Cluster 1 distance estimates were nearly 7 times greater than Cluster 4 and 9.15 times greater than Cluster 5.



PCA: Principal component analysis is an unsupervised linear dimensionality reduction technique. PCA is used to identify patterns in the data based on correlation between features. This technique finds the maximum variance from the given high-dimensional data and then it is represented into lower dimensions.

Decision Tree:
One of the simplest and most straightforward classification techniques is the Decision Tree. It is a type of supervised learning method that can be used for classification and regression. The fundamental goal is to learn simple decision rules from the training data and then predict the class of the test data using those rules.

The decision tree is divided into two types based on the type of target variable (continuous/categorical):

1. A decision tree with a continuous variable.
2. A decision tree based on categorical variables.

The root node of the decision tree is where class prediction begins.

Its value is compared to the value of the record.
We jump to a branch that corresponds to that value and then update the node based on the comparison.

Random Forest:
It's also a regression and classification supervised learning algorithm.
It can handle datasets with continuous variables for regression and categorical data for classification. It's a lot slower than the Decision Tree. This algorithm selects observations at random, then creates a decision tree and averages the results.
If all trees are appropriate and diverse, it will perform better than a decision tree.

References:
https://medium.com/cube-dev/bayesian-nonparametrics-9f2ce7074b97
https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4871360/
https://arxiv.org/abs/1203.1365