

CSE 291D Project Proposal

Names of group members: Wangmuge Qin, Shuang Song and Aaron Trefler

Research Goals: Within the realm of human decision making, not many decisions are as important than one's choice of a marriage partner. "In contemporary Western societies, this decision usually follows a long learning period during which people engage in more informal ... relationships, i.e., dating" (Fisman, et al. 2008). Therefore, whom people choose to date during this "learning period" will likely end up affecting future marriage partner decisions. Our research aims to use latent variable modeling to help better understand how people choose whom to date.

Using Speed Dating Data: In speed dating, potential partners go on several quick (i.e., several minutes) speed dates with other participants before individually deciding whom they would like to go on a real date with in the future. How does the complex relationship between two people's attributes, attitudes, interests, and judgements affect the likelihood of a brief encounter (i.e., a speed date) eventually leading to a real date? It is not hard to see why shedding light on this answer would be of interest to the millions of single men and women currently dating.

Data: The dataset we will use is from Kaggle.com, and named "Speed Dating Experiment"[1]. Originally, the dataset is gathered and used in a research about gender difference in mate selection [2]. The data was gathered from 2002 to 2004 and includes data related to demographics, dating habits, lifestyle information, participant's ratings in terms of attractiveness, sincerity, intelligence, fun, ambition, and shared interests for others during the 4 minute speed dating interactions, and whom each participant wanted to see again for a real date. The speed dating was conducted over 14 separate sessions, involving ~200 males and females, which resulted in ~40,000 unique dates. For this dataset, dates were only conducted between males and females.

Models & Algorithms: We believe that each person's preferences and attributes can respectively be represented by a vector, which acts as a latent variable. The rating one would give to another is then somehow related to the similarity between one's preferences and another's attributes, which is commonly measured as the inner product of two vectors. Our intention is to construct a model to infer the unobserved vectors from peoples' observed features, for which there are a lots of models to try, from logistic regression to a neural network.

As is mentioned in the original research [2] on this dataset, there is a significant gender difference in dating partner selection, we thus would like to build model for female-selecting-male and male-selecting-female separately. We also consider training the classifiers in the Bayesian way.

References

[1] <https://www.kaggle.com/annavictoria/speed-dating-experiment>

[2] Fisman, Raymond, et al. "Gender differences in mate selection: Evidence from a speed dating experiment." *The Quarterly Journal of Economics* 121.2 (2006): 673-697.