CSE291D Final Report

Aaron Trefler
trefler.aaron@gmail.com

Shuang Song
shs037@eng.ucsd.edu

Wangmuge Qin
qwangmug@eng.ucsd.edu

Abstract

In this project, we study people's decision making in mate selection through a speed dating dataset. We build a model that combines logistic regression and collaborative filtering together to find patters of people' decision making in mate selection. We present some experimental results on a speed dating dataset from Kaggle.com.

1 Introduction

Within the realm of human decision making, not many decisions are as important than ones 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., 2006]. 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.

The data we aim at exploring is from speed dating. 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 peoples attributes, attitudes, interests, and judgments 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.

In this project, we analyzed a speed dating dataset [Fisman and Iyengar,], and model people's decision on whether they wanted to see their partner again for a traditional date using Bayesian logistic regression and collaborative filtering.

We present some related work in Section 2, details about the dataset in Section 3, models, training algorithms and experimental results in Section 7, **??** and 8 respectively.

2 Related Work

There exists a long line of work on factors that influence human mate selection. Previously, researchers have aimed to discover the factors that influence people's mate selection, the relative importance of these factors, and how these factors interact with individual differences such as gender and race.

Differing mate selection preferences based on individual differences has been studied for quite some time. For example, [Buss, 1989], [Wiederman and Allgeier, 1992], [Kenrick et al., 1993], [Buston and Emlen, 2003], [Tadinac and Hromatko, 2007] and [Garcia et al., 2012] studied the differences in the preference of mates across genders, the relation between people's self-perceived qualities and their expectation for potential mates, as well as the difference between mate selection preferences based on whether people were seeking long-term or short-term relationships. Common to their discoveries was the following "exchange" between the two genders – on average men place the most importance on a women's physical attractiveness, whereas women consider men's income levels as the most important factor. Additionally, men tend to prefer short-term relationships while women prefer long-term relationships. However, the data from all these studies were collected from questionnaires or advertisements in newspapers' mate seeking sections, and the analysis methods used in most of the papers are basic statistical tests.

There are also some works studying real mate selection data, i.e., data that includes a meeting in person and decision making and/or date evaluation process. Some of the work are more qualitative or on experimental design; and there are also many quantitative works [Todd et al., 2007], [Houser et al., 2008], [Luo and Zhang, 2009], [Fisman et al., 2006], [Tidwell et al., 2013] that studied the relation between different factors and mate selection decision and gender difference and similarities. Generally, they found similar trend as those previous on questionnaire data, that is, men respond more to physical attractiveness while women consider intelligence and race as more important. They either studied data from speed-dating event with small number (at most 80 of each gender) of participants or used simple statistics to measure the importance of difference factors. The only work that used data from a larger scale speed-dating experiment is [Fisman et al., 2006], which we note is where our data comes from. The only work that used machine learning algorithm is [Houser et al., 2008]. They used logistic regression to predict the data decision and discover the influence of multiple factors. There is another work [Wade and Gisler, 2015] which used logistic regression to model people's mate selection based on their mate intelligence, self-esteem and self-perceived attractiveness; however, mate selection in their experiments is done through choosing one from two pictures of the opposite sex downloaded from the Internet.

In this project, apart from logistic regression which is a fairly standard method for predicting binary variables and is being used in previous research, we also combine it with collaborative filtering to further find latent patterns in participants decision making. Specially, we use collaborative filtering with matrix factorization. Matrix factorization for collaborative filtering is widely used in recommender systems, where there are a bunch of ratings given by a set of users to a set of items, and the goal is to predict unseens ratings. The intuition of using matrix factorization in recommender systems is that there exists some latent features of the users and items that determines the rating, and the number of latent features is much smaller than the number of users or items. For example, in a movie recommendation system, each latent feature might represents one genre. If two users give similar ratings to a bunch of movies, they might share similar appetite for a genre; and if a movie is an action movie and a user has strong preference on action movie, then he/she might give high rating on it. In a Bayesian view, there is a generative process that draws the latent features of each user and each item according to certain distribution, and then draw the ratings according to some other distribution characterized by the inner products of users and items. Given the user-item matrix, the goal is to learn the posterior distribution of user and item features. There are numerous work on matrix factorization and collaborative filtering, for example, [Mnih and Salakhutdinov, 2008], [Salakhutdinov and Mnih, 2008], [Ma et al., 2008], [Koren et al., 2009]. We note that the method used in our project is the biased matrix factorization, which also takes into consideration the bias of users' ratings and the ratings received by items. The bias terms intuitively captures the average rating a user gives, and the average rating an item would receive. It has been shown in many existing work that adding the bias terms can significantly improve prediction accuracy.

3 Dataset

The dataset is the "Speed Dating Experiment" from Kaggle.com [Fisman and Iyengar,]. Originally, the dataset was gathered and used in a research study about gender differences in mate selection [Fisman et al., 2006].

Data was gathered from participants in several experimental speed dating events from 2002-2004. It consists of two different sets of information, one gathered prior to dating and one after. Before the actual speed dating, the participants were required to submit general information such as age, race, demographics and educational background. They were also asked about self perception, expectations for potential partners, belief about others' perceptions regarding the six attributes (i.e., *attractiveness*, *sincerity*, *intelligence*, *fun*, *ambition*, and *shared interests*).During the event, participants had a 4 minute "first date" with every other participant of the opposite sex. At the end of their 4 minutes, participants were asked to rate their date on the 6 attributes as above, and whether they would like to see their date again.

The speed dating was conducted over 14 separate sessions, involving more than 200 males and females, which resulted in around 40,000 unique dates. Dates were only conducted between males and females.

4 Model

Our final goal is to construct a model instructive for dating recommendation, for which the one-sided decision is of great importance and the logistic regression is in our consideration. Instead of predicting these decisions directly from observed features (like ages), we choose to utilize as intermediate variables the ratings one gave to another, which are in principle unobservable, but are available in the dataset.

We believe these ratings demonstrate people's preferences in a straightforward way and should be, thus, useful for the prediction task.

As for the ratings themselves, we consider predicting them using collaborative filtering, which seeks for a latent vector representation for each rater and each ratee. The reason why we do not use the observed features is twofold. For one thing, the personal features available in the dataset are intuitively too weak and incomplete - for example, it is hard to evaluate one's physical attractiveness without his/her photo. For another, all our attempts on this, from regular linear regression to a two-layer neural network, failed, proving such a prediction task is indeed intractable.

The general structure of our model is illustrated in Figure 1.

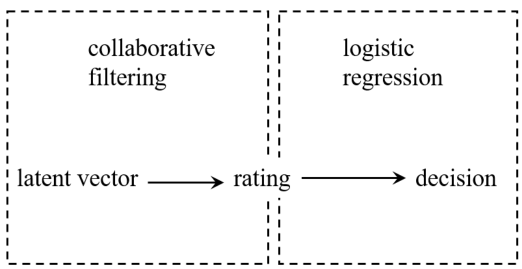

Figure 1: General Structure

4.1 Bayesian Logistic Regression

A natural choice for the likelihood model for the one-sided decision is Bernoulli distribution, as its output is binary. We model the mean of the distribution as a linear combination of the ratings one gave to another in various aspects, passed through a sigmoid function, which is also known as the logistic regression.

To specialize our model, we assign to each rater a unique weight vector, which can be interpreted as his/her preferences over all the aspects.

More specifically, let i, j be two people, $d_{ij} \in \{0, 1\}$ the binary one-sided decision of i regarding j and r_{ij} a column vector denoting the ratings i gave to j. The probability i would like to continue the dating with j is modeled as

$$\Pr(d_{ij} = 1 | r_{ij}, w_i) = \text{sigm}(w_i^\intercal r_{ij}) \tag{1}$$

where w_i is i's personal weight vector and $\text{sigm}(x) = 1/(1 + e^{-x})$ is the sigmoid function.

Moreover, we make our model Bayesian by assuming the weight vectors are draw from some prior. Intuitively, these weights should be non-negative. But for mathematical convenience, we choose to use Gaussian distribution as the prior, that is

$$\Pr(w_i|\mu_w, \sigma_w^2) \propto \exp[-\frac{1}{2\sigma_w^2}(w_i - \mu_w)^\intercal (w_i - \mu_w)]$$

4.2 Collaborative Filtering

Collaborative filtering attempts to find a latent vector representation for each person. Specifically, the rating $r_{ij}^{(k)}$, which is the k-th aspect of r_{ij}, is modeled with a Gaussian distribution:

$$\log \Pr(r_{ij}^{(k)} = r|\theta) = -\frac{1}{2}(\alpha^{(k)} + \beta_i^{(k)} + \beta_j^{(k)} + \gamma_i^{(k)\intercal}\gamma_j^{(k)} - r)^2 + \text{const} \tag{2}$$

where $\theta = \left\{\alpha^{(k)}, \beta_i^{(k)}, \beta_j^{(k)}, \gamma_i^{(k)}, \gamma_j^{(k)}\right\}$ is the parameters, among which $\alpha^{(k)}, \beta_i^{(k)}$ and $\beta_j^{(k)}$ are scalars and $\gamma_i^{(k)}, \gamma_j^{(k)}$ are column vectors. Notice that the variance is assumed to be 1 without loss of generality.

From an intuitive point of view, α is the overall average of that rating; β_i, β_j explain the average rating i gives or j receives above/below the overall average α; γ_i and γ_j represent the personal preferences of i and the personal attributes of j respectively and the inner product, $\gamma_i^\intercal \gamma_j$, measures their similarity.

We further assign a Gaussian prior to each of the β and γ for regularization purposes. The mean is assumed to be zero for simplification.

$$\log \Pr(\beta_i^{(k)}|\sigma^{(k)2}) = -\frac{1}{2\sigma^{(k)2}}\beta_i^{(k)2} + \text{const}$$
$$\log \Pr(\gamma_i^{(k)}|\tau^{(k)2}) = -\frac{1}{2\tau^{(k)2}}\|\gamma_{ij}^{(k)}\|^2 + \text{const}$$

To summarize, the generating process of our model is shown in Figure 2, where i indexes the raters, j indexes the ratees and k the aspects of ratings. The hyper-parameters are omitted.

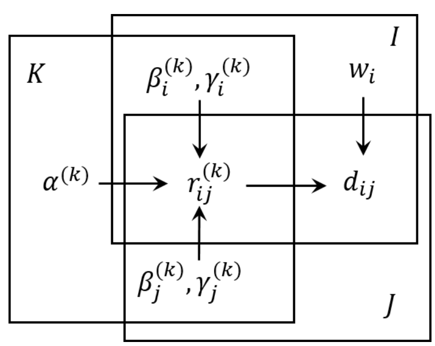

Figure 2: Graphical Model Digram

5 Inference Algorithm

Now that our model consists of two parts, namely collaborative filtering and Bayesian logistic regression, we decide to conduct the inference separately.

5.1 Hard-EM algorithm

EM algorithm is a natural choice for the inference of models with latent variable and we would like to use it for our logistic regression model. However, deriving the update equation for a standard EM algorithm involves computing the expected logarithm likelihood of the joint probability of data and latent variables, which is mathematically difficult in our case. The hard-EM algorithm is hence considered.

The hard-EM algorithm alternatively optimizes the latent variables (i.e. the personal weight w_i) and the hyper-parameters (denoted as θ). In particular, we

1. Initialize w_i and θ to random values.
2. Optimize w_i to maximize the joint probability $\Pr(d_{i\cdot}, w_i) = \prod_j \Pr(d_{ij}|w_i, r_{ij})\Pr(w_i|\theta)$ for each i.
3. Optimize θ to maximize $\prod_i \Pr(w_i|\theta)$.
4. Repeat 2-3 until convergence.

Step 3 is exactly to conduct the maximum likelihood estimation (MLE) for a Gaussian prior, which makes it relatively easy. The update equations are given as follow.

$$\begin{aligned} \mu_w &\leftarrow \frac{1}{N}\sum_i w_i \\ \sigma^2 &\leftarrow \frac{1}{N}\sum_i (w_i - \mu_w)^\intercal (w_i - \mu_w) \end{aligned} \tag{3}$$

In practice, we make σ_w^2 fixed and only update μ_w to avoid the potential overflow issue, which makes σ_w^2 a hyper-parameter that needs to be predefined.

Step 2, however, is far more difficult, for which approximate algorithms like gradient ascent can apply.

5.2 Gradient Ascent with Adaptive Learning Rate

We use gradient ascent algorithm to cope with the optimization in step 2 and the update is as follows.

$$w_i \leftarrow w_i + \lambda \frac{\partial \log \Pr(d_{i\cdot}, w_i)}{\partial w_i} \tag{4}$$

where λ is the learning rate. Aiming at a better performance, we also make the learning rate adaptive. In particular, at each epoch, new weights are calculated using the current learning rate and new log-probability are then calculated. If the new log-probability is less than the old one, the updates are discarded and the learning rate λ is decreased by multiplying by a predefined factor.

5.3 Alternating Least Squares

The inference of the collaborative filtering model involves maximizing

$$\log \Pr(r, \alpha, \beta, \gamma) = \sum_{i,j,k} \log \Pr(r_{ij}^k|\theta) + \sum_{i,k} \log \Pr(\beta_i^{(k)}|\sigma^{(k)2}) + \sum_{i,k} \log \Pr(\gamma_i^{(k)}|\tau^{(k)2}) \tag{5}$$

which turns out to be some kind of least square regression. The alternating least squares algorithm applies here. Properly speaking, we

1. Initialize α, β, γ to random values.
2. Alternatively optimize $\alpha, \beta_i, \beta_j, \gamma_i$ and γ_j with other variables fixed.
3. Repeat step 2 until convergence.

The updates in step 2 are derived by simply letting the partial gradients be zero and are given as below.

$$\alpha^{(k)} \leftarrow \frac{\sum_{i,j}(r_{ij}^{(k)} - \beta_i^{(k)} - \beta_j^{(k)} - \gamma_i^{(k)\intercal}\gamma_j^{(k)})}{N}$$
$$\beta_i^{(k)} \leftarrow \frac{\sum_j(r_{ij}^{(k)} - \alpha - \beta_j^{(k)} - \gamma_i^{(k)\intercal}\gamma_j^{(k)})}{1/\sigma^{(k)2} + |J|}$$
$$\gamma_{il}^{(k)} \leftarrow \frac{\sum_j \gamma_{jl}^{(k)}(r_{ij}^{(k)} - \alpha - \beta_i^{(k)} - \beta_j^{(k)} - \gamma_i^{(k)\intercal}\gamma_j^{(k)} + \gamma_{il}^{(k)}\gamma_{jl}^{(k)})}{1/\tau^{(k)2} + \sum_j \gamma_{jl}^{(k)2}}$$

6 Evaluation & Analysis

6.1 Experiment Setting

Considering the huge difference between female and male, men-to-women and women-to-men decisions are analyzed separately. Incomplete dating data is eliminated in advance as preprocessing, which leaves us with 3,584 pieces of men-to-women data and 3,111 pieces of women-to-men data.

In both cases, 90% of the data is then randomly chosen as training set. The rest of the data is used for validation.

6.2 Bayesian Logistic Regression

6.2.1 Should Weights be Non-Negative?

Intuitively, a weight vector should have all of it's components be non-negative, in the sense that a higher rating in a specific aspect should not decrease the probability of "continuing to date". This can be achieved by clipping the gradient in Equation 4 to force the weights to remain non-negative.

Control experiment is conducted to justify the gradient clipping - we train our model using the men-to-women data twice with gradient clipping on and off respectively. In both cases, we set the initial learning rate $\lambda = 0.01$, variance $\sigma_w^2 = 0.01$, and initialize the weights to random values draw from a uniform distribution between 0 and 1. The results are illustrated in Figure 3.

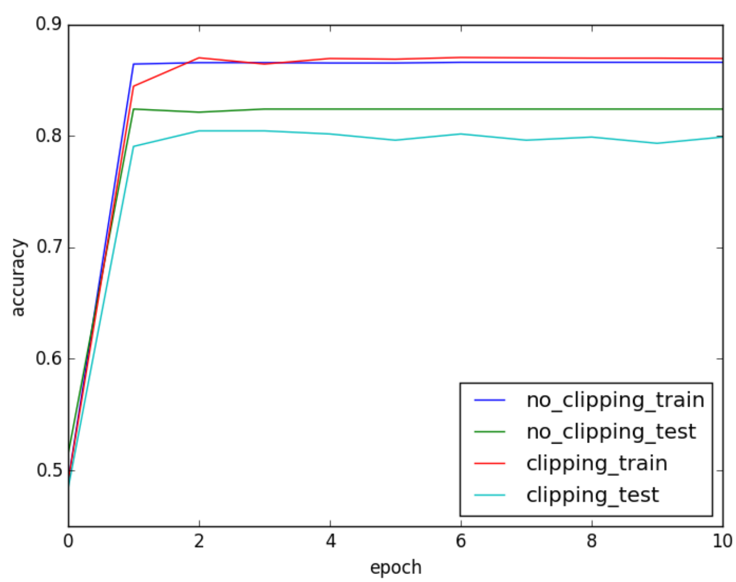

Figure 3: With vs. Without Gradient Clipping

Table 1: Average Weights of Men and Women

Gender	Attractiveness	Sincerity	Fun	Ambition	Intelligence	Shared Interest
Men	0.66973065	-0.14845873	0.27078013	-0.16384133	-0.06353842	0.26389261
Women	0.41922765	-0.09260041	0.26893471	-0.16203004	0.08500969	0.27362003

It can be seen that the model without gradient clipping uniformly outperforms the model with gradient clipping. Moreover, it is found that about 38% of the weights trained with gradient clipping are zeros, which is to say that these values should be negative.

This result is rather surprising as it implies that a high score in certain aspects actually reduces people's willingness to continue dating. We therefore choose the model without gradient clipping, where the weights can be any real value, to use in the following sections.

6.2.2 Tuning Hyper-Parameters

The hyper-parameters to tune in our model are the initial learning rate λ and the variance σ_w^2. The corresponding accuracies of various λ and $sigma_w^2$ are respectively shown in Figure 4 and Figure 5.

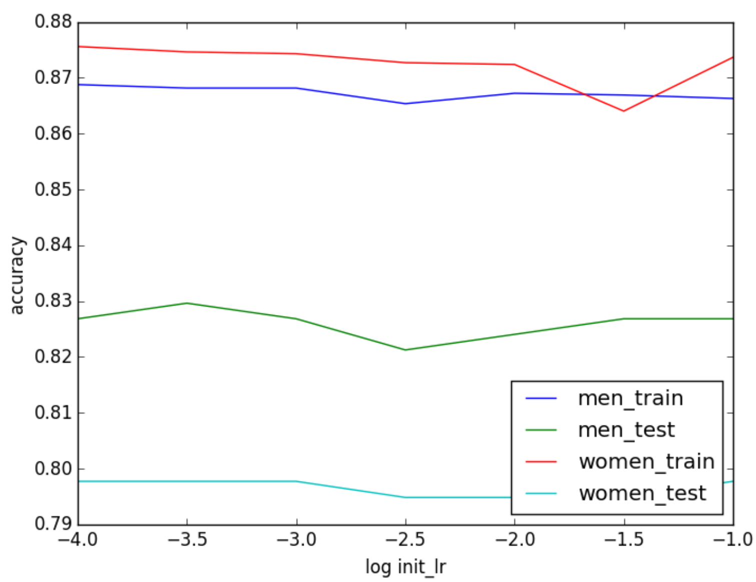

Figure 4: Accuracy vs. Initial Learning Rate

We see that the initial learning rate makes very little difference to the model, which is what we would expect as the learning rate is adaptive. However, the predefined variance σ_w^2 is of great significance. The peak of test accuracy for both men-to-women and women-to-men data is approximately marked by $\sigma_w^2 = 10^{-2.5}$.

6.2.3 Analysis

We then fix $\lambda = 0.01$ and $\sigma_w^2 = 10^{-2.5}$ and compute the personal weight vectors for each participant. The average weights are shown in Table 1.

We see that both men and women put the greatest weight on physical attractiveness, which makes intuitive sense to the authors. However - in line with previous work - men place a greater weight on

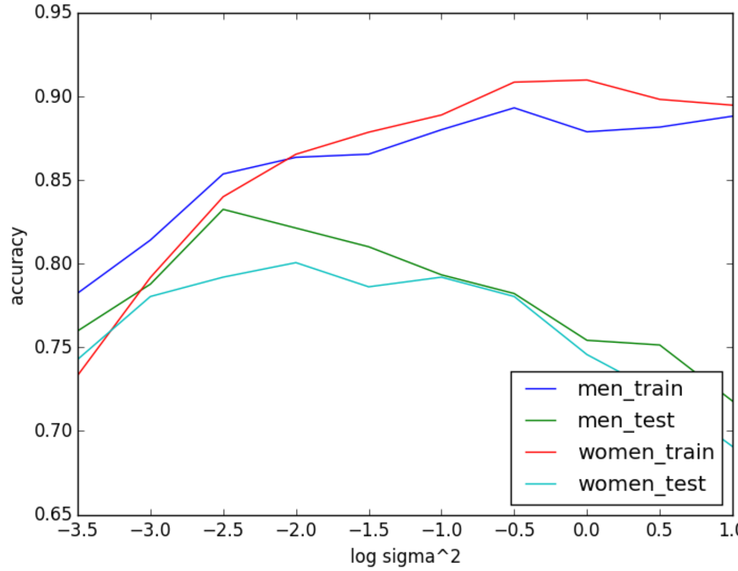

Figure 5: Accuracy vs. σ_w^2

Table 2: First Two Eigenvectors of Men and Women

Eigenvector	Attractiveness	Sincerity	Fun	Ambition	Intelligence	Shared Interest
Men_1	0.39554214	0.43415339	0.41054783	0.4104145	0.44963592	0.34043384
Women_1	0.32314752	0.45571797	0.40996021	0.43557307	0.47172536	0.327995
Men_2	0.65497223	0.03298053	-0.19666785	0.00278033	0.11405796	-0.71988088
Women_2	-0.38452353	0.61919907	-0.30230595	-0.01143151	0.30797667	-0.53138025

attractiveness than women, as shown by the fact that the weight of men (0.67) for this attribute is much greater than that of women (0.42).

For both men's and women's decision making, fun and shared interest play similar and positive roles. On the other hand though, the effect of sincerity is actually negative, which is indeed a surprising finding. Perhaps participants who were candid during dating got high sincerity ratings for their honesty, but poor ratings overall for sharing too much. Another difference between men and women is the weight of intelligence, which is positive for women while negative for men. This too is a surprising finding.

It is also found that weights of different people are actually highly correlated. Principle component analysis shows that over 70% of the variance for the weights is explained by the first principle eigenvector component, for both men and women (see Figure 6).

The first two eigenvectors of men and women are illustrated in Table 2. We see that the first eigenvector, of both men and women, uniformly increases the weight in each aspect. This can be interpreted as people's general willingness above/below the average for future dating.

The second eigenvector of men significantly increases the weight on attractiveness and, meanwhile, decreases that of shared interest, leaving weights in other aspects almost unchanged. On the contrary, the second eigenvector of women corresponds to greater weight on sincerity and intelligence and less on both attractiveness and shared interest.

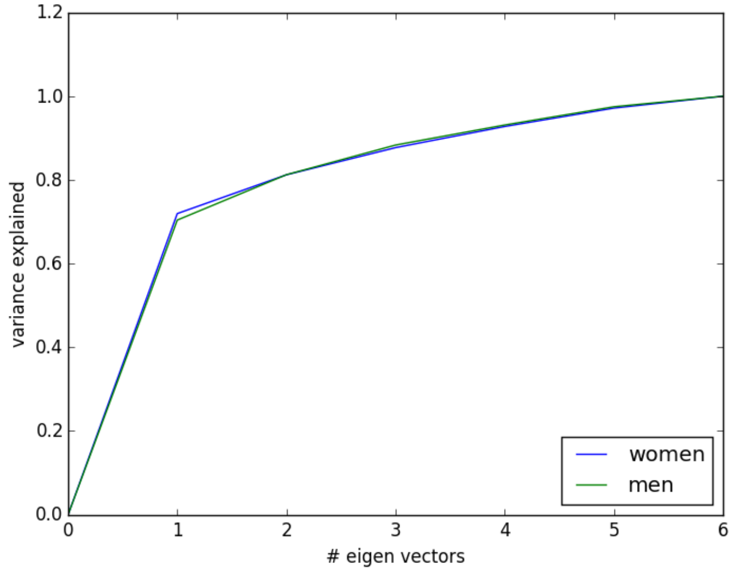

Figure 6: Explained Variance Ratio

7 Task II

Another goal of this project was to model participant's one-sided decisions using only the one-sided decision matrix itself. Each entry in the one-sided decision matrix consists of either a positive or negative rating given by a rater to a ratee. Unlike the first model, here we do not assume we have access to the ratings participants' gave their ratees over the six attributes. We also do not assume we have significant information about the participants themselves (e.g., dating preferences, income, etc.) This situation is analogous to the growing world of online dating (e.g., Tinder), where companies have to recommend potential mates to users with limited knowledge of their dating preferences or characteristics outside of who they have previously indicated they would like to match with or not.

We attempt to perform this training two models: a cosine-similarity model and a latent factor model.

7.1 Cosine Similarity Model

For this model, the likelihood of the one-sided decision is modeled as a Bernoulli distribution. The mean of this distribution is the average rating given to that ratee by all other raters who have rated them, weighted by the cosine similarity between those rater's decisions and decisions of the rater we are trying to predict.

More specifically, let i, j be two people, $d_{ij} \in \{-1, 1\}$ the one-sided decision of i regarding j. The probability i would like to continue dating j is modeled as

$$\Pr(d_{ij} = 1 | d_{-i,j}, \theta) = \text{sigm}(d_{-i,j}^{\intercal} \theta) \tag{6}$$

where θ is a vector of cosine similarities between i and all other users who have rated j. The cosine similarity is defined between to vectors $\mathbf{A}$ and $\mathbf{B}$ as follows

$$\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A_2}\| \|\mathbf{B_2}\|} \tag{7}$$

7.2 Latent Factor Model

This latent factor model follows the same general structure indicated in Figure 1. The model attempts to find the best 2 dimensional representations for how each person is as a rater and ratee. Specifically if p is a 2-D vector representation of each participant as a rater and q is a 2-D vector representation of each participant as a ratee, then the model selects p and q such that the following objective function is minimized

$$error^2 = (r_{ij} - p_i^\intercal q_j)^2 + \frac{\beta}{2} \sum_{n=1}^{N} (||p_k||^2 + ||q_k||^2) \tag{8}$$

where β is a regularizing hyperparameter which we set to be 0.02.

We optimize this equation using the alternating least squares procedure described in 5.3. The updates are derived by setting the partial gradients to zero and are given below

$$p'_{ik} = p_{ik} + \alpha \frac{\partial}{\partial p_{ik}} e_{ij}^2 = p_{ik} + 2\alpha e_{ij} q_{kj}$$
$$q'_{kj} = q_{kj} + \alpha \frac{\partial}{\partial q_{kj}} e_{ij}^2 = q_{kj} + 2\alpha e_{ij} p_{ik}$$

where α is the gradient step size and k is the kth dimensional component of the latent vector representations.

7.3 Evaluation & Analysis

8 Conclusion

References

[Buss, 1989] Buss, D. M. (1989). Sex differences in human mate preferences: Evolutionary hypotheses tested in 37 cultures. *Behavioral and brain sciences*, 12(01):1–14.

[Buston and Emlen, 2003] Buston, P. M. and Emlen, S. T. (2003). Cognitive processes underlying human mate choice: The relationship between self-perception and mate preference in western society. *Proceedings of the National Academy of Sciences*, 100(15):8805–8810.

[Fisman and Iyengar,] Fisman, R. and Iyengar, S. S. Speed Dating Dataset. `https://www.kaggle.com/annavictoria/speed-dating-experiment/`.

[Fisman et al., 2006] Fisman, R., Iyengar, S. S., Kamenica, E., and Simonson, I. (2006). Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*, 121(2):673–697.

[Garcia et al., 2012] Garcia, J. R., Reiber, C., Massey, S. G., and Merriwether, A. M. (2012). Sexual hookup culture: a review. *Review of General Psychology*, 16(2):161.

[Houser et al., 2008] Houser, M. L., Horan, S. M., and Furler, L. A. (2008). Dating in the fast lane: How communication predicts speed-dating success. *Journal of Social and Personal Relationships*, 25(5):749–768.

[Kenrick et al., 1993] Kenrick, D. T., Groth, G. E., Trost, M. R., and Sadalla, E. K. (1993). Integrating evolutionary and social exchange perspectives on relationships: Effects of gender, self-appraisal, and involvement level on mate selection criteria. *Journal of Personality and Social Psychology*, 64(6):951.

[Koren et al., 2009] Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8).

[Luo and Zhang, 2009] Luo, S. and Zhang, G. (2009). What leads to romantic attraction: Similarity, reciprocity, security, or beauty? evidence from a speed-dating study. *Journal of personality*, 77(4):933–964.

[Ma et al., 2008] Ma, H., Yang, H., Lyu, M. R., and King, I. (2008). Sorec: social recommendation using probabilistic matrix factorization. In *Proceedings of the 17th ACM conference on Information and knowledge management*, pages 931–940. ACM.

[Mnih and Salakhutdinov, 2008] Mnih, A. and Salakhutdinov, R. R. (2008). Probabilistic matrix factorization. In *Advances in neural information processing systems*, pages 1257–1264.

[Salakhutdinov and Mnih, 2008] Salakhutdinov, R. and Mnih, A. (2008). Bayesian probabilistic matrix factorization using markov chain monte carlo. In *Proceedings of the 25th international conference on Machine learning*, pages 880–887. ACM.

[Tadinac and Hromatko, 2007] Tadinac, M. and Hromatko, I. (2007). Own mate value and relative importance of a potential mate's qualities. *Studia Psychologica*, 49(3):251.

[Tidwell et al., 2013] Tidwell, N. D., Eastwick, P. W., and Finkel, E. J. (2013). Perceived, not actual, similarity predicts initial attraction in a live romantic context: Evidence from the speed-dating paradigm. *Personal Relationships*, 20(2):199–215.

[Todd et al., 2007] Todd, P. M., Penke, L., Fasolo, B., and Lenton, A. P. (2007). Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences*, 104(38):15011–15016.

[Wade and Gisler, 2015] Wade, T. J. and Gisler, S. (2015). The role of intelligence in mating: an investigation of how mating intelligence relates to mate selection and mating-relevant constructs. *Human Ethology Bulletin*, page 8.

[Wiederman and Allgeier, 1992] Wiederman, M. W. and Allgeier, E. R. (1992). Gender differences in mate selection criteria: Sociobiological or socioeconomic explanation? *Ethology and Sociobiology*, 13(2):115–124.

9 Midterm report (for reference)

9.1 Model & Results

Thus far we have implemented and tested several techniques to model aspects of the Kaggle speed dating dataset. Attempts have been made on the following models.

9.1.1 Collaborative Filtering

Our first try is collaborative filtering. For each rater i we have a constant β_i and a vector $\vec{\gamma_i}$ representing his/her preferences. Similarly we have β_j and $\vec{\gamma_j}$ for each ratee j. Then the score rater gives ratee is modeled as

$$r_{i,j} \simeq \alpha + \beta_i + \beta_j + \vec{\gamma_i} \cdot \vec{\gamma_j}$$

We then try to find a model f for the relationship between the latent attributes $\vec{Z_i} = (\beta_i; \vec{\gamma_i})$ and the observed attributes $\vec{X_i}$ (like habits) of each person i.

$$\vec{Z_i} = f(\vec{X_i})$$

Collaborative filtering itself works well, an MSE $\sim$ 1.0 is achieved in predicting the scores (10-scaled) on test set. But after trying standard linear regression, gradient boosting regression and a 2-layer neural network, we still failed to find a suitable model f. We conclude that the observed attributes available in the dataset just cannot be used to predict $\vec{Z_i}$.

9.1.2 Bayesian-Style Logistic Regression

We realize that the previous task is somewhat too complicated and, therefore, try to predict the people's decision of whether to continue the dating instead. The scores are used as features now. We believe that these scores demonstrate people's will in a more direct way.

Specifically, we assume that the preferences of person i is represented by some latent vector $\vec{w_i}$, and let the scores he/she gives be $\vec{R_{i,j}}$ (in terms of attractiveness, sincerity, fun and so on). The binary decision is modeled following standard logistic regression:

$$P(d = 1|\vec{R_{i,j}}, \vec{w_i}) = \text{sigmoid}(\vec{R_{i,j}} \cdot \vec{w_i})$$

Moreover, we assume people's preferences $\vec{w_i}$ have some prior distribution, which makes our model a Bayesian one.

We consider using the hard-EM algorithm, which alternatively optimizes the hyper-parameters and the latent variables. We utilize gradient descent in the M step to optimize the latent variables.

Our first attempt for the prior is a gamma distribution, in order to constrain the weights $\vec{W_i}$ to the domain of positive real. But we find that our optimizer tends to make some of the weights negative. This is an interesting fact that some aspect of the scores actually reduces people's willingness for the dating. So a Gaussian distribution is tried instead.

This model works fine - an accuracy $\sim$ 0.75 is achieved on the test set. But we find it difficult to make any further improvement.

9.1.3 Hidden Topic Model

We also make an attempt to implement a latent variable model to automatically assign each date to one of two latent classes in an unsupervised fashion. To get the model off the ground we utilize a single variable, x, the age difference between the two participants (i.e., man_age - women_age). This generative model works as follows:

- we draw a latent class variable, z, from a multinomial distribution parameterized by $\vec{\pi}$, which itself is drawn from a Dirichlet distribution;
- we draw x from a Gaussian distribution, parameterized by μ and σ^2, chosen based on the latent z variable selected.

Making use of conjugate priors we implement a Gibbs sampling procedure to learn the distributions over all model parameters. Looking at the MAP estimates for our model parameters, we see that our first latent class corresponds to a Gaussian (μ=0.5, σ^2=4.2) that is selected 93% of the time; and the second latent class corresponds to a much wider Gaussian (μ=-0.01, σ^2=9.6) that is selected only 7% of the time.

9.2 Future Plan

Our future plan is two-fold. On the one hand, we decide to continue improving the Bayesian-style logistic regression model. As we find it difficult to further increase the accuracy on test set (due to overfitting), we consider trying some simpler priors like discrete distribution. We will also try to plug in the observed attributes (like habits) to make better prediction.

On the other hand, we believe it would be much more interesting to apply the hidden topic model with (1) multiple features, (2) using data only from matches, and (3) more latent classes. This will allow us to potentially uncover whether there is a simple underlying latent-structure behind a successful speed-date.

10 Proposal (for reference)

Research Goals: Within the realm of human decision making, not many decisions are as important than ones 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 peoples 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, participants 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 persons 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.