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 studied people's decision making in mate selection using a relatively large speed dating dataset from Kaggle.com. We built several models relevant to this task, which are detailed throughout the report. Our most successful model combined logistic regression and collaborative filtering to uncover hidden structure laying behind people's mate selection decisions. In the following report we detail model construction, experimental results, as well as model evaluation.

1 Introduction

Within the realm of human decision making, not many decisions are more 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" [5]. 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.

We aim to do this through 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 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 taken from Kaggle.com [4], and modeled people's decisions on whether they wanted to see their partner again for a traditional date. The models we chose to use include Bayesian logistic regression, collaborative filtering, and cosine similarity.

The structure of the report is the following. We present some related work in Section 2; details about the dataset in Section 3; models, training algorithms and experimental results in Section 4, 5 and 6 respectively; a second modeling task in 7; and we finally conclude our work and discuses future directions in Section 8.

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, [1], [19], [8], [2], [15] and [6] 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 high 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, and the analysis methods used in most of the papers were simple statistical tests (e.g., t-tests).

There is also work studying real mate selection data (i.e., data that includes a meeting in person as well as a decision making and/or date evaluation process). These works are sometimes quantitative [17], [7], [10], [5], [16]. Generally, these works find similar trends as those previous discussed, that is, men respond more to physical attractiveness while women consider other factors like intelligence and race as more important. Some of these works studied data from a speed-dating, however, in these cases the data set was small (i.e., at most 80 participants of each gender).

The only work that used data from a larger scale speed-dating experiment is [5], which we note is where our data originated from. The only two works we found that used machine learning algorithms were [7] and [18]. [7] used logistic regression to predict the decision to date and discovered the relative influence of multiple factors. [18] used logistic regression to model people's mate selections based on their mate's intelligence, self-esteem and self-perceived attractiveness. However, in both their experiments mate selection is done by choosing one from two pictures of the opposite sex downloaded from the Internet.

In this project, apart from logistic regression, we also combine it with collaborative filtering to further find latent patterns in participants decision making. Specially, we use collaborative filtering with matrix factorization in our primary model. 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, where the number of latent features is much smaller than the number of users or items. In a Bayesian view, there is a generative process that draws the latent features of each user and each item according to a certain distribution, and then draws 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 works on matrix factorization and collaborative filtering, for example, [13], [14], [11], [9]. Among them, there are also many works on incorporating auxiliary user or item information into the collaborative filtering process [3] and [12]. We note that the method used in our primary model 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 both the average rating a user gives and the average rating an item would receive. It has been shown in many existing works that adding the bias terms can significantly improve prediction accuracy.

3 Dataset

The dataset is the "Speed Dating Experiment" from Kaggle.com [4]. Originally, the dataset was gathered and used in a research study about gender differences in mate selection [5].

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 six key attributes (i.e., *attractiveness*, *sincerity*, *intelligence*, *fun*, *ambition*, and *shared interests*).During the event, participants had a 4 minute "speed 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 stated 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.

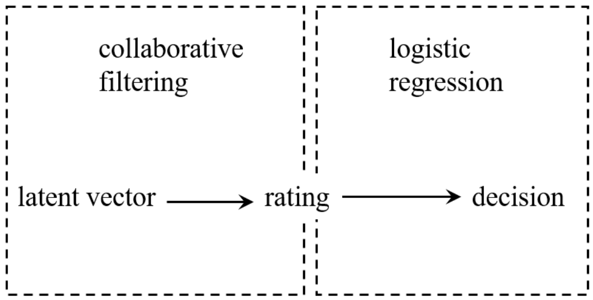


Figure 1: General Structure

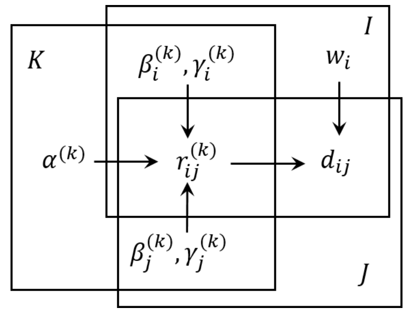


Figure 2: Graphical Model Digram

4 Model

Our final goal is to construct a model instructive for dating recommendation, where importance is placed on predicting the one-sided decisions of the speed dating participants. Instead of predicting these decisions directly from observed features (e.g., age), we choose to utilize as intermediate variables the ratings raters gave to ratees after the speed date, 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. We also believe analyzing trained model parameters based on these features will be quite informative in terms of shedding light on participants' mate selection decision processes.

As for the ratings themselves, we consider predicting them using collaborative filtering, which seeks for a latent vector representation for each rater and 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 to do 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.

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 drawn 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.

A variant with user-user similarities Apart from the collaborative filtering model described above, we also investigate on a slight variant of it with user-user similarity information. Suppose there exists background information from which we can compute the pairwise similarities between users, and each user's top K neighbors. To incorporate the background information to the latent features learned through matrix factorization, we can consider the true latent feature as a weighted sum among its own latent features and the latent features from its neighbors. Specifically, instead of $\gamma_i^{(k)}$, we consider $\sum_{j \in \text{neighbor}_i} \lambda_j \gamma_j^{(k)}$ to be the latent features of user i, where λ_j is the weight of its neighbor (including itself), e.g., we can consider $\lambda_i = \lambda$ and $\lambda_j = (1-\lambda)/K$ for $j \neq i$. In this way, we could address the user similarity encoded in the background information. If it is known from background information that two users are similar, they would become closer in the latent feature space. In this project, we consider the answers to the questionnaire prior to the actual speed dating as the background information used to calculate similarities. Similarity is measured with cosine similarity.

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 variables 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 weights 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 steps 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.

$$\mu_w \leftarrow \frac{1}{N}\sum_i w_i \tag{3}$$

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 is 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.

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

6 Evaluation & Analysis

6.1 Experiment Setting

Considering the huge difference between female and male decision patterns, men-to-women and women-to-men decisions are analyzed separately. Incomplete dating data is eliminated in advance

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

Table 1: Average Weights of Men and Women

during 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.

A control experiment is conducted to justify 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 drawn from a uniform distribution between 0 and 1. The results are illustrated in Figure 3a.

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 σ_w^2 are respectively shown in Figure 3b and Figure 3c.

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 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.

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

Table 2: First Two Eigenvectors of Men and Women

MSE/Variance	Attractiveness	Sincerity	Fun	Ambition	Intelligence	Shared Interest
Men-to-women	0.34863	0.39834	0.38802	0.49230	0.43599	0.42747
Women-to-men	0.31104	0.41695	0.44842	0.48082	0.40411	0.51629

Table 3: Performance of Collaborative Filtering on Different Tasks

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 3d).

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.

6.3 Collaborative Filtering

6.3.1 Tuning Hyper-Parameters

For collaborative filtering, the only hyper-parameters to tune are $\sigma^{(k)2}$ and $\tau^{(k)2}$, which act as regularization for $\beta^{(k)}$ and $\gamma^{(k)}$ respectively, and the length of latent vector γ (denoted as L). To find the optimal parameters .In practice, we found that $\sigma^{(k)2}$ and $\tau^{(k)2}$ have very limited influence on the model's performance **on test set**, as long as they are large enough to avoid overflow. We therefore set $\sigma^{(k)2} = 1/2$ and $\tau^{(k)2} = 1/6$ for all k.

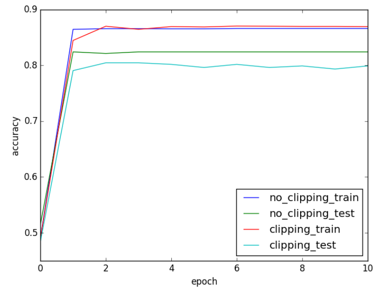
The MSE's corresponding to various lengths of latent vector in predicting men-to-women attractiveness ratings are shown in Figure 3e. We see that $L = 3$ is a good compromise between performance and time complexity.

6.3.2 Evaluation

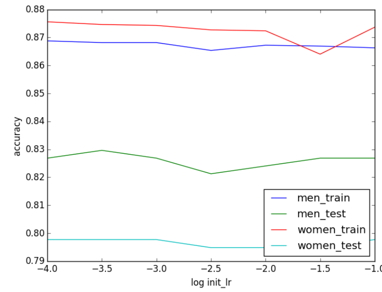
The optimal performance we achieved, measured in the ratio of MSE to variance, on test set for ratings in each aspect, are shown in Table 3. The performance is not as good as we would expect, but considering the small size of training data (about 3000 pieces), it is also acceptable to us.

6.3.3 Using Background Similarity Information

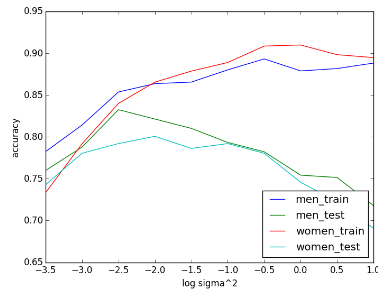
Table 4 shows the MAE for collaborative filtering with and without background similarity information. We picked the top 20 nearest neighbors of each user, and set the number of latent features as 5. We can see an improvement in MSE for all 6 fields. However, in this model we used the original matrix factorization without the bias term, and the MSE obtained is worse than that of biased matrix factorization. We chose to use the latent factor learned from the original collaborative filtering as input features for logistic regression.



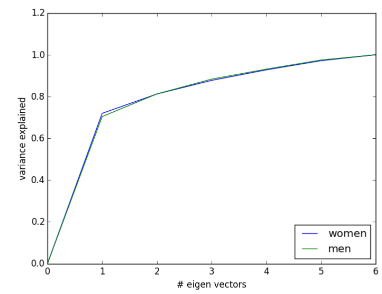
(a) With vs. Without Gradient Clipping



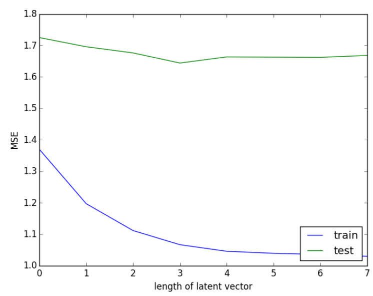
(b) Accuracy vs. Initial Learning Rate



(c) Accuracy vs. σ_w^2



(d) Explained Variance Ratio



(e) MSE vs. Length of Latent Vector

Figure 3: Figures for experiments

	Attractiveness	**Sincerity**	**Fun**	**Ambition**	**Intelligence**	**Shared Interest**
w/o similarity	1.98748	1.97504	1.99122	2.00535	1.9831	1.99435
w/ similarity	1.9131	1.91548	1.91507	1.91435	1.91322	1.9197

Table 4: MSE w/ and w/o similarity information (averaged over 10 repeated runs).

7 Task II

7.1 Collaborative Filtering on Decision Matrix

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 is either positive or negative decision 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 key attributes. We also do not assume we have information about the participants themselves (e.g., dating preferences, age, etc.) This situation is analogous to the growing world of online dating, where companies (e.g., Tinder) 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 using two models: a cosine-similarity model and a latent factor model.

7.2 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, weighted by the cosine similarity between those raters' decisions and the decisions of the rater whom 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 two vectors **A** and **B** as $\frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A_2}\| \|\mathbf{B_2}\|}$.

7.3 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 representation 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

$$error^2 = (d_{ij} - p_i^{\intercal} q_j)^2 + \frac{\beta}{2} \sum_{n=1}^{N} (||p_k||^2 + ||q_k||^2) \tag{7}$$

where β is a regularizing hyper-parameter that we set to be 0.02, and $d_{ij} \in \{0, 1\}$.

We optimize this equation using an alternating least squares procedure similar to that 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 set to 0.0002 and k is the k-th dimensional component of the latent vector representations.

7.4 Evaluation & Analysis

The complete dataset was randomly split into training and test sets. 90% of the data (i.e., 7,531 dates) was used of the training set and the remaining 10% for the testing set (i.e., 837 dates).

For a baseline reference, always predicting d_{ij} to be negative produced a test accuracy of 59%. Though both men and women more often reject future dating than accept it, women were much more selective. On average women rejected future dating 64% of the time compared with men who rejected much less often at 51%.

The Cosine Similarity model produced a test accuracy of approximately 70%. The Latent Factor model produced a similar test accuracy of approximately 71% (train set accuracy $= 87\%$), which corresponded to 8 more correct classifications.

The fact that the Latent Factor model achieved comparable results to the Cosine Similarity model is intriguing, as the Latent Factor model was able to significantly reduce the dimensions needed to model participant's dating preferences. This is an indication that the latent factors produced by the model may be indicative of a simple hidden low dimensional structure that dictated the participants date selection decisions. Thus, future work my aim to analyze these latent factors further, and see if they can be mapped onto human interpretable labels.

8 Conclusion and Future Work

In this project, we studied factors that influence mate selection through a speed dating dataset. We combined logistic regression and collaborative filtering to predict ratings and decisions between participants, and found similarity as well as differences in the two genders, which agrees with findings from previous work in the mate selection literature.

Additionally in Task II, we also showed that participants' date selection decisions can be modeled adequately using a low dimensional latent structure.

As for future directions, we think it might be interesting to further interpret the latent features produced by our investigation, as well as incorporate more background information into our models (e.g., incorporating background similarity into our biased-matrix factorization model).

References

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

[2] Peter M Buston and Stephen T Emlen. 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, 2003.

[3] Yi Fang and Luo Si. Matrix co-factorization for recommendation with rich side information and implicit feedback. In *Proceedings of the 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, pages 65–69. ACM, 2011.

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

[5] Raymond Fisman, Sheena S Iyengar, Emir Kamenica, and Itamar Simonson. Gender differences in mate selection: Evidence from a speed dating experiment. *The Quarterly Journal of Economics*.

[6] Justin R Garcia, Chris Reiber, Sean G Massey, and Ann M Merriwether. Sexual hookup culture: a review. *Review of General Psychology*, 16(2):161, 2012.

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

[8] Douglas T Kenrick, Gary E Groth, Melanie R Trost, and Edward K Sadalla. 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, 1993.

[9] Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8), 2009.

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

[11] Hao Ma, Haixuan Yang, Michael R Lyu, and Irwin King. Sorec: social recommendation using probabilistic matrix factorization. In *CIKM*, pages 931–940. ACM, 2008.

[12] Hao Ma, Dengyong Zhou, Chao Liu, Michael R Lyu, and Irwin King. Recommender systems with social regularization. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 287–296. ACM, 2011.

[13] Andriy Mnih and Ruslan R Salakhutdinov. Probabilistic matrix factorization. In *NIPS*, 2008.

[14] Ruslan Salakhutdinov and Andriy Mnih. Bayesian probabilistic matrix factorization using markov chain monte carlo. In *ICML*, pages 880–887. ACM, 2008.

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

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

[17] Peter M Todd, Lars Penke, Barbara Fasolo, and Alison P Lenton. Different cognitive processes underlie human mate choices and mate preferences. *Proceedings of the National Academy of Sciences*, 104(38):15011–15016, 2007.

[18] T Joel Wade and Stefanie Gisler. The role of intelligence in mating: an investigation of how mating intelligence relates to mate selection and mating-relevant constructs. *Human Ethology Bulletin*.

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