**Dating recommendation system**

Table of Contents

[**1. Introduction** 2](#_Toc498879580)

[**2. Data** 2](#_Toc498879581)

[**3. Problems to be Solved** 2](#_Toc498879582)

[**4. Data Processing** 2](#_Toc498879583)

[**5. Methods and Process** 2](#_Toc498879584)

[**6. Evaluations and Results** 2](#_Toc498879585)

[6.1. Evaluation Methods 2](#_Toc498879586)

[6.2. Results and Findings 2](#_Toc498879587)

[**7. Conclusions and Future Work** 2](#_Toc498879588)

[7.1. Conclusions 2](#_Toc498879589)

[7.2. Limitations 2](#_Toc498879590)

[7.3. Potential Improvements or Future Work 2](#_Toc498879591)

# **1. Introduction**

Recommender system is now a day popular in many fields such ecommerce sites for recommending items which we wish. Netflix, recommending movies based on the user’s taste/likes. In this project we have used speed dating dataset to recommend the perfect partner to a user based on his/her preferences.

Speed dating concept is being popular since last decade. The main idea behind this speed dating is to date the partners in an event & rate them personally based on individual interest match. This event is usually conducted among few people mostly 30-40, where both males & females are given pamphlets with couple of attributes in it to rate their partner. In the beginning, a host lets both the males and females to date for 4-5 mins where they rate, after the completion of each date participants are shuffled so that they could date a different person, by the end of the event the host would collect pamphlets and notify the participants about their mutual date interests.

# **2. Data**

Data was gathered from participants in experimental speed dating events from 2002-2004. During the events, the attendees would have a four minute "first date" with every other participant of the opposite sex. At the end of their four minutes, participants were asked if they would like to see their date again. They were also asked to rate their date on six attributes: Attractiveness, Sincerity, Intelligence, Fun, Ambition, and Shared Interests. The dataset also includes questionnaire data gathered from participants at different points in the process. These fields include: demographics, dating habits, self-perception across key attributes, beliefs on what others find valuable in a mate, and lifestyle information.

* Source of the dataset: [https://www.kaggle.com/annavictoria/speed-dating-experiment](https://www.kaggle.com/annavictoria/speed-dating-experiment%20)
* The data is composed of 8,378 observations and 195 variables.

Attributes:

UserID: User ID

PartnerID: Partner ID

Attr: Attractive

Sinc: Sincere

Int: Intelligent

Fun: Fun

Amb: Ambitious

Shar: Shared Interests

# **3. Problems to be Solved**

To make a better and reliable decision in any situation, its necessary to possess a certain sufficient amount of information. So, the major problem here would be reliability, recommending a partner to a user just based on their overall rating wouldn’t be an efficient. Every user has their own preferences, it’s not that an item having highest rating would be recommended to everyone. If we take a deeper dive into the dataset we get to know few insights like what’s the user’s preferences and using this as a key element we can provide better recommendations.

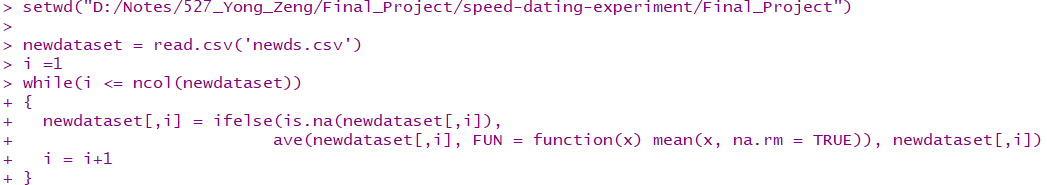
Here in this speed dating dataset we have various attributes like attractiveness, sincerity, intelligence etc. In addition to the overall rating, we have utilized all these attributes using Multi-criteria recommender system to achieve better recommendation.

# **4. Data Processing**

The following data processing is done to the given dataset.

Data cleaning – Unwanted special characters like dots, colons etc. from the obtained dataset.

Missing values – write a function to find the missing values or NULL values in the dataset. If any missing or NULL values were found, the function should find the mean in the class and fill in the data in the dataset.



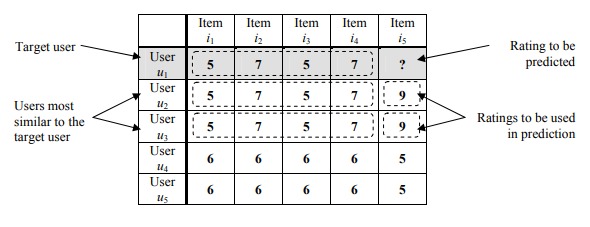


# **5. Methods and Process**

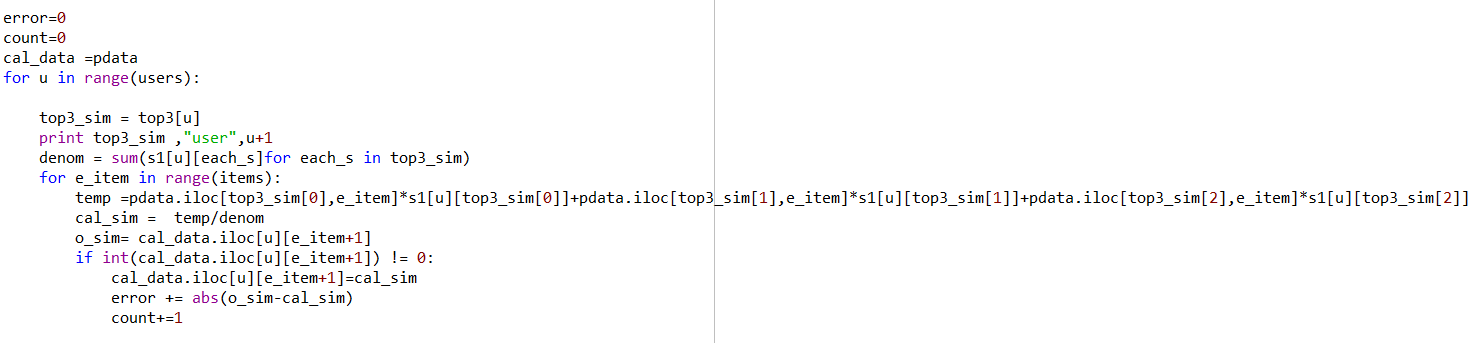
Traditional Collaborative Filtering:

Works by using single-criterion rating which specifies how much the user rated for a certain attribute that implicitly shows users preference. Here we are using user based collaborative filtering where the similarity between the users are computed and the overall rating is predicted and once the ratings are predicted, the profiles with highest rating is recommended to the user.





# 



Multi-criteria Recommendation system:

Even though single rating traditional recommender system is successful in many applications, multi criteria recommendation system is being widely used considering the fact that multiple attributes could be considered and each attribute would be given specific weight and using the specific weight the overall rating is being predicted. The profiles with highest rating based on user preferences would be recommended to the user.



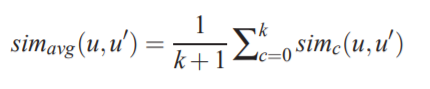
A picture containing text

Description generated with high confidence

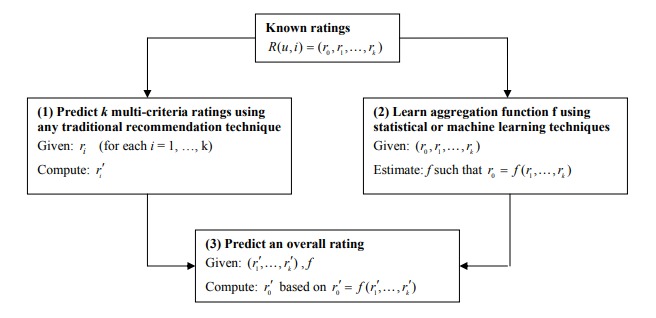
Heuristic Approach (Average Similarity)

Extending the traditional heuristic-based collaborative filtering technique to reflect multi-criteria rating.

Here, the first step of the prediction process is to choose the similarity computation method to find a set of neighbors for each user. There are several similarity measures to be used, we have incorporated cosine similarity to find the distance between two users.



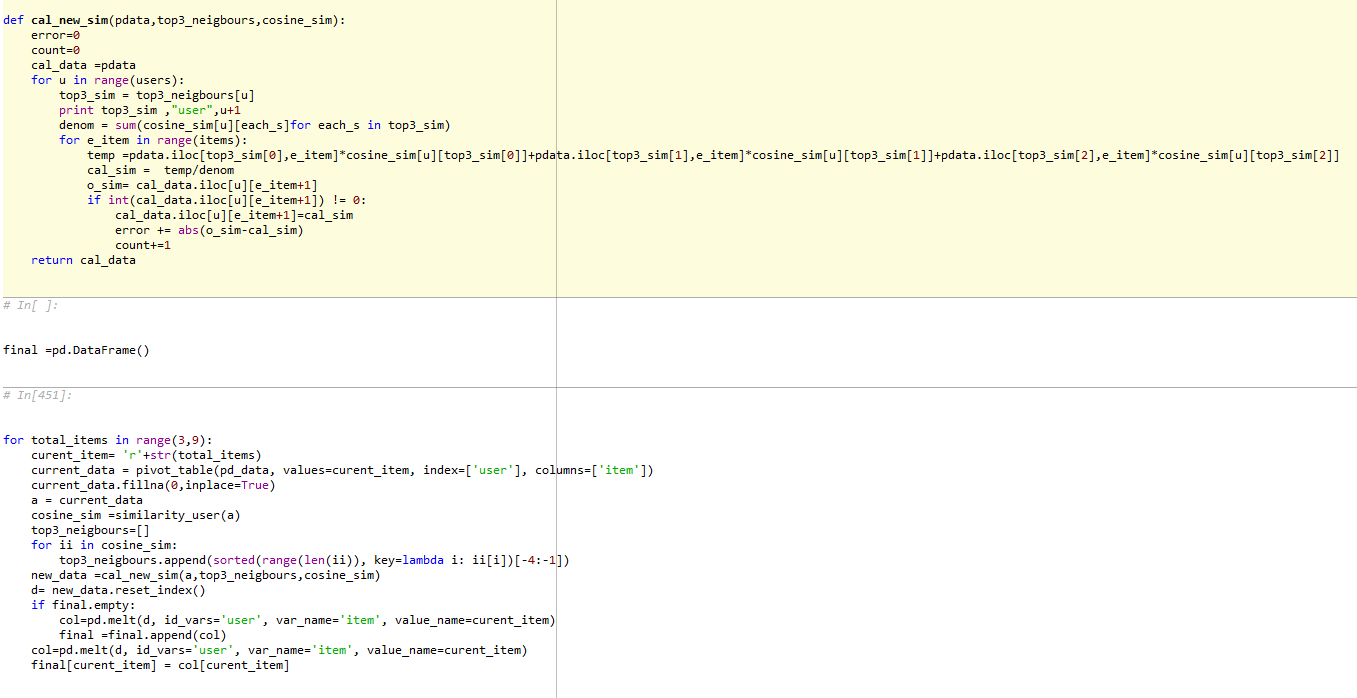
Model-Based Approach:



This approach constructs a predictive model to estimate unknown ratings by learning from the observed data.

We have incorporated the Aggregate function approach, consists of 3 steps.

1. Estimating the k individual ratings using a recommendation technique.
2. aggregation function f is chosen using domain expertise, statistical techniques, or machine learning techniques.
3. Lastly, overall rating of each unrated item is computed based on the k predicted individual criteria ratings and the chosen aggregation function f.



# **6. Evaluations and Results**

## 6.1. Evaluation Methods

We have calculated MAE and precision to compare the models build with that of the traditional collaborative filtering.

MAE: Calculating the difference between actual overall rating with that of the predicted overall rating.

Precision: Calculating the fraction of relevant instances among the retrieved instances.

## 6.2. Results and Findings

|  |  |
| --- | --- |
|  | MAE |
| Traditional CF | 0.85 |
| Average similarity | 0.75 |
| Aggregate Function | 0.72 |

A screenshot of a cell phone

Description generated with high confidence

We predicted the rating for individual attributes using traditional collaborative filtering and two multi-criteria recsys approaches Heuristic (Average similarity) and Model based approach (Aggregate function). And predicted the overall rating using multiple linear regression and calculated the MAE comparing the predicted overall rating with the actual overall rating. Its clear that the other two approaches average similarity and aggregate function performs better than traditional CF.

|  |  |
| --- | --- |
|  | [Precision@10](mailto:Precision@10) |
| Traditional CF | 0.35 |
| Average similarity | 0.47 |
| Aggregate Function | 0.5 |

A screenshot of a cell phone

Description generated with very high confidence

Also calculated the precision@10 using traditional collaborative, average similarity and aggregate function, hence we could see a significant difference in these precisions.

# **7. Conclusions and Future Work**

## 7.1. Conclusions

In this paper, we are motivated by the assumption that only some of the given selection criteria dominate the user’s decision and propose a clustering method which positions users in a preference lattice dependent on which criteria they prefer. Based on the clustering, we then present three different methods for predictions which to instantiate the overall clustering approach.

The three methods are:

a. Predicting for overall rating based on the aggregation function of criteria ratings. Within this method, two directions are presented, by using all the criteria as the independent variables, or by using only significant criteria for the users to aggregate overall rating;

b. Predicting overall rating based on the overall ratings of the same item from other users within the same or close clusters;

## 7.2. Limitations

1. We cannot rely much on these results as the data is collected from specific school. As there is very less diversity in dataset, the predictions we made can’t be considered to represent a large group.
2. As the participants could meet their partners for only 4 mins, Attraction attribute played a major role in dating decision.

## 7.3. Potential Improvements or Future Work

We have multi criteria recommendation systems for recommending potential matches to the users, but when it comes dating there is a specific problem, where person “A” might like person “B”, but that doesn’t mean that person “B” likes person “A”. Therefore, future work could be extended to incorporate reciprocal matching algorithms to suggest recommendation for each user accordingly.

Furthermore, the future work for the proposed methods could be extended to handle qualitative and fuzzy metrics for computing users experience with recommender system.