

Roll No: CS18B013

Name: PRASHANTH

- Dear Student, You may have tried or thought of trying different methods for your data contest. Choose one of the methods that was not taught in class, and submit a writeup of this "new method" in the template provided below. This is an individual submission - i.e., while you would've done your kaggle submission as a team of two members or while you may've discussed this method with your teammate, **you will have to write about the new method in your own words independently and submit it individually.**
- **Template:** Fill in whatever fields are applicable for your algorithm (overall 1-2 page writeup; since some fields may not be applicable for certain methods, we haven't shown points below).

1. (points) [Name of Method, and its ML Problem and Paradigm: problem could be regression/classification/clustering/etc., and paradigm could be supervised/unsupervised/..., generative/discriminative/direct, linear/non-linear models, etc.)]:

Solution: Method Name - Collaborative Filtering, This technique is used for generating recommendation systems. It is based on the users predictions (ratings) made on the particular product/products. customers prefers the product/products in the past, so that they may like in future. Collaborative Filtering is an unsupervised learning. Collaborative Filtering depends on the users similarities to recommend the products. Nearest Neighborhood algorithm is great to use in collaborative filtering technique. Collaborative filtering is regression problem.

2. (points) [Brief introduction/motivation: One paragraph to describe briefly the new method (its name, what it does, its main application, etc.)]

Solution: Collaborative Filtering, This method uses similarities between users and products. Collaborative filtering techniques will recommend a product to user X which is purely based on the interests of a similar user Y. As we see in the day to day life recommendation engines are playing crucial part in the major companies like Netflix, spotify, Amazon, Youtube etc. Recommendation engine uses one of the machine learning methods called collaborative filtering. Collaborative filtering technique for recommender systems are methods that are based on past interactions recorded between customers and products. There are two types of ratings this can be implicit ratings or explicit ratings. Moreover there are two types of algorithms for collaborative filtering are Non-probabilistic algorithms and probabilistic algorithms ,this also includes

model based approach and memory based approach are there are some techniques such as item based, user based, hybrid based, Bayesian-network models, EM algorithm. The main application of collaborative filtering is music recommendation, movie recommendation, books recommendation etc. If we consider movie recommendation When a user visits the netflix site, the recommending engine should recommend movies based on both similarity to movies the user has liked in the past movies that similar users liked. The main collaborative filtering applications are. This technique is used in various fields such as e-commerce, marketing , e-learning , social networking sites and this method is also applied in many different kinds of data including sensing and monitoring data, financial data etc. Mathematical analysis of collaborative filtering method.

$$r_{xi} = \frac{1}{k} \sum_{y \in N} r_{yi}$$

r_x is the vector of costumer x 's ratings, and N be the set of k customers most similat to x and also rated for product i . For user to user collaborative filtering the mathematical analysis is.

$$r_{xi} = \frac{\sum_{j \in N(i,x)} s_{ij} \cdot r_{xj}}{\sum_{j \in N(i,x)} s_{ij}}$$

- 1) s_{ij} =Similarity of products i and j
- 2) r_{xj} =Rating of user x on item j
- 3) $N(i, x)$ =items rated by x similar to i

Some pros/cons for collaborative filtering are, cold start, first rater, popularity bias, sparsity.

3. (points) [Closely related method seen in class, and relation of your selected new method to method you eventually used for the data contest]:

Solution: The concept taught in class that is similar is finding K nearest neighbours since here also we are looking for 'neighbour's with same 'taste' to group them together. This method is the method we used in data contest. we used item-item filtering.

4. (points) [Training Input and Output: (e.g., $\{x_i, y_i\}_{i=1 \dots N}$, $x_i \in \mathbb{R}^d$, $y_i \in \{-1, +1\}$, etc.):

Solution: Training input is the ratings which the customers have given for the songs. training output is a matrix such that it contains the predicted score for every customer-user. Training input is the ratings which the customers have given for the songs. training output is a matrix such that it contains the predicted score for every customer-song.

5. (points) [Training Objective function (e.g., loss function that is to be optimized) or probabilistic model (over which MLE or Bayesian inference done):]

Solution: we need to find a set of items which share the same features as one other. finding this set is our objective function. actually we have no objective function, but just to fill you can write this.

6. (points) [Training Algorithm: Brief description of key aspects of the algorithm]

Solution: Training Algorithm :- Is An algorithm that can obtain sensitivity from a provided set of training data. The brief description of key aspects of the algorithm, which is as follows.

Collaborative Filtering :- In Collaborative Filtering, we tend to find similar costumers and recommend what similar customers like.

1) Measuring Similarity :- Here, we recommend items to costumers/users. W.r.t to the users of similar interest domain, they have liked.

2) Cosine Distance :- In this method the cosine distance between the users to find out the users with similar interests is known as cosine distance.

3) Rounding the Data :- In collaborative filtering we round off the data because, comparison becomes more easy.

4) Normalizing Rating :- we consider Average user rating and subtract all the given ratings from it. So, this results are either positive or negative values as rating. We can also classify this further into similar groups.

7. (points) [Testing Input and Output: (e.g., $x \in \mathbb{R}^d$, $y \in \{-1, +1\}$)]

Solution: Testing input is the customer id and song id whose predicted score we have to predict. testing output is the corresponding score.

8. (points) [Testing Algorithm: Brief description of key aspects of the algorithm]

Solution: Not applicable

9. (points) [Critique of the method: (1-2 paragraphs discussing its strengths and weaknesses in your own words)]

Solution: The strengths and weaknesses of Collaborative filtering are: Strengths of this particular methods are, As we discussed this method helps users to pick up their products easily by recommendation engines. the advantage of this method includes there is no requirement of domain knowledge because the embeddings are already learned automatically. This model also helps costumers to explore new interests. Where as this system recommend the model to the costumer if he/she may not interested in that particular recommended product, this is called Serendipity. In this technique includes based on the recommendation technique, predict target costumers preferences for each user product. This technique Sort the customers products according to the prediction probability and recommended to them. So, there are some weaknesses for this particular method, they are feature representation of the products are hand-engineered to some extent, for this technique it requires lots of domain knowledge. This model can only make recommendations based on existing interests of the customers. The model has limited ability to expand on the customers existing interests. This method cannot handle new items/products. It's difficult to add side features for item.