SOEN 691 - Big Data Analytics

Winter 2020

Project Presentation
Recommendation System for Anime Dataset

Presented By :
Arsalaan Javed - 40085994
Gunvansh bhatia - 40082036
Manpreet Singh - 40083737

Index

- Data Set
- Content Based Recommender System
  - Item Profile
  - User Profile
  - Evaluation Metrics
- Collaborative Filtering Recommender System
  - Utility Matrix
  - User User based recommendation
  - Common Practices - regularization
  - Evaluation

Dataset

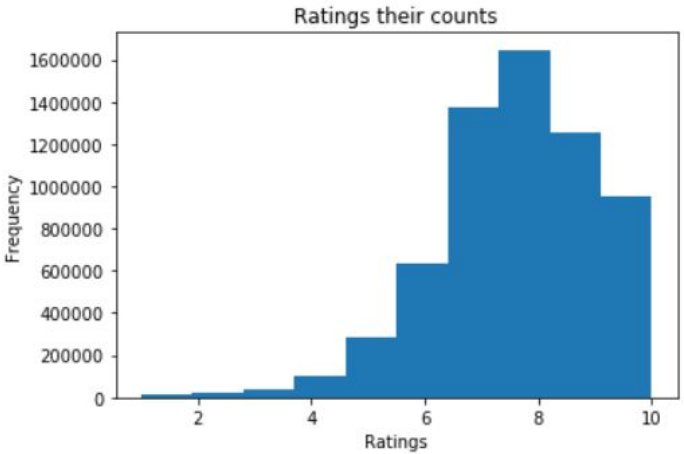
```
df.isnull().sum()
```

```
anime_id      0
name          0
genre        62
type         25
episodes      0
rating      230
members       0
dtype: int64
```

Anime dataset from kaggle having two csv files:

- anime.csv with 7 columns : Anime_id, Name, Genre, Type, Episodes, Rating, members
  - Total 12294 rows or we can say 12294 unique animes and their details
  - Missing values in columns Genre, Type and Rating

- Rating.csv with 3 columns : User_id, Anime_id, rating
  - Total entries : 7813737
  - Total unique users : 73516
  - Missing values for rating column (value of -1)
  - Ratings are in range from 1 - 10
  - with global avg rating of 7.80



Recommendation System

In this project we have implemented two types of recommender systems as per the requirement for Implementing two algorithms covered in class:

- Content Based Recommender System
- Collaborative Filtering Recommender System

Content Based Recommender System

Main idea: Recommend items to customer ***x*** similar to previous items rated highly by ***x.*** To do so we need to create Item and User Profile.

Item Profile: We have taken columns **Genre** and **Type** of each anime and the final vector that we will get for each item would be of 1 x 51 dimensions. To calculate the weightage of each content word like “action” we implemented **TFIDF** score wrt each document.

	anime_id	name	genre	type
1854	332	Dokidoki Densetsu: Mahoujin Guru Guru	adventure, comedy, fantasy, magic, shounen	TV

Item vector looks something like this:

supernatural	drama	Movie	school	romance	military	magic	adventure	TV	action	...
0.494556	0.361599	0.331109	0.462052	0.425588	0.000000	0.000000	0.000000	0.000000	0.000000	...

Content Based Recommender System

User Profile: Will also be a vector of dimension 1 x 51 with the same set of content words.

It is calculated by taking weighted sum of each anime item vector that the user has watched and rated. The weight for each anime is calculated by taking the mean rating of the user and subtracting it with the rating for each anime. This will shift the rating values about zero. So, the animes with a positive rating now are the ones that user has liked while vice versa for the not liked animes.

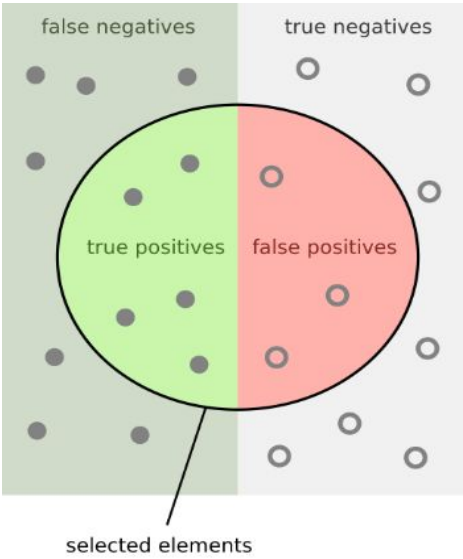
Best Matched: A **cosine similarity** of user profile with all the item profile will provide the best matched anime list and this can be our recommendation list.

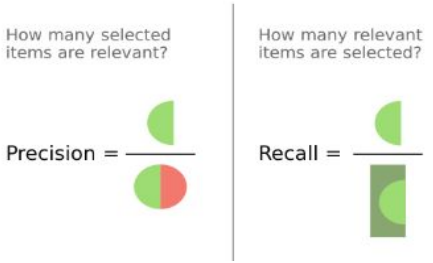
Evaluation Metric:

We are not predicting any ratings for the anime so, to calculate RMSE score was not possible.

Instead, after researching and discussing about it, we created

The evaluation metric with the animes which were rated by the user and comparing with predictions made by the model.





This was one of the challenge for us in this project. To be frank, we are still not so sure if this is the right approach for evaluating this model.

Sample Result for CBR:

```
Recommending movies for user 59325
```

	anime_id	name	genre	type
335	12189	Hyouka	mystery, school, slice of life	TV
3076	4202	Mokke Specials	mystery, slice of life, supernatural	Special
4392	31636	Dagashi Kashi	comedy, shounen, slice of life	TV
1707	32648	Handa-kun	comedy, shounen, slice of life	TV
9813	25139	Oh! My Konbu	comedy, shounen, slice of life	TV
1239	10378	Shinryaku!? Ika Musume	comedy, shounen, slice of life	TV
1297	8557	Shinryaku! Ika Musume	comedy, shounen, slice of life	TV
1431	13267	Shinryaku!! Ika Musume	comedy, shounen, slice of life	OVA
1664	27969	Hana to Alice: Satsujin Jiken	drama, mystery, slice of life	Movie
2684	2931	Mokke	mystery, slice of life, supernatural	TV

```
evaluation_metrics(cosine_sim_u
```

```
true_pos : 247
true_neg : 145
false_pos : 65
false_neg : 217

Precision : 79.16666666666666
Recall : 53.23275862068966
Accuracy : 58.160237388724035
```

Collaborative Filtering Recommender System

Main Idea : Find a set N of other users whose ratings are “similar” to user x’s ratings. Then estimate user x’s ratings based on ratings of users in *N*.

For this model we have used rating.csv file for preparing data i.e. columns : user_id, anime_id and rating.

Utility Matrix : {user_id : {anime_id : rating}}

Task 1: To calculate most similar users and for this we have used Pearson Correlation Coefficient between user x with rest of all users.

$$sim(x,y) = \frac{\sum_{s \in S_{xy}} (r_{xs} - \overline{r_x})(r_{ys} - \overline{r_y})}{\sqrt{\sum_{s \in S_{xy}} (r_{xs} - \overline{r_x})^2} \sqrt{\sum_{s \in S_{xy}} (r_{ys} - \overline{r_y})^2}}$$

Collaborative Filtering Recommender System

Task 2: After getting a list of similar users as user x we can predict the ratings of the animes that this user has not watched but similar users have watched.

$$r_{xi} = \frac{\sum_{y \in N} s_{xy} \cdot r_{yi}}{\sum_{y \in N} s_{xy}}$$

Task 3 : Applying common practises i.e. **regularization** in predicting the rating for th user x for improving the prediction result.

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

Evaluation

Taking test set i.e. 50 randomly selected users and 30% of randomly selected animes. Then we removed the test set from the data.

Evaluation : RMSE

1: Before applying regularization term:

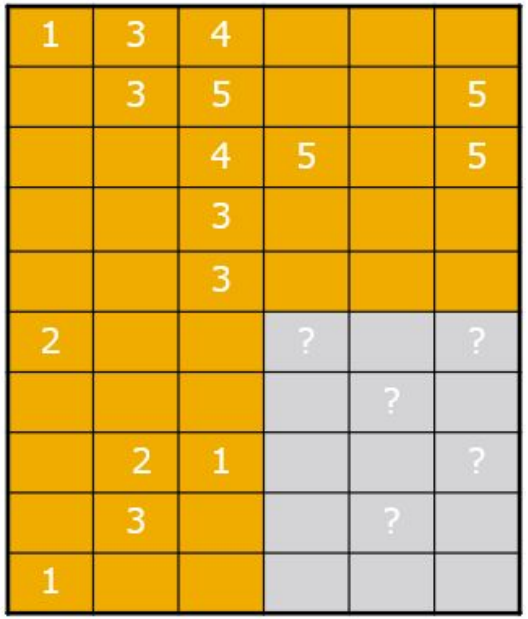
```
RMSE(utility, predicted1)
```

```
2.0913927814953595
```

2: With Regularization:

```
RMSE(utility, predicted)
```

```
1.5373710265944023
```



Future Possible Ideas!

Content Based Recommendation System:

- Creating user profile with gradient descent, as the more appropriate a user profile is the better would be the result.

Collaborative Filtering:

- Modeling Item Item based CF

Thank You!

Any Questions/ Comments/ Suggestions