Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Collaborative Recommendation System in Users of Anime Films

To cite this article: A S Girsang *et al* 2020 *J. Phys.: Conf. Ser.* **1566** 012057

Collaborative Recommendation System in Users of Anime Films

A S Girsang[1*], B Al Faruq[1*], H R Herlianto[1*] and S Simbolon[1*]

[1] Computer Science Department, BINUS Graduate Program-Master of Computer Science, Bina Nusantara University, Jakarta, 11480, Indonesia

*Email: agirsang@binus.edu, barra.faruq@binus.ac.id, hemdani.herlianto@binus.ac.id, sihar.simbolon@binus.ac.id

Abstract. The recommendation system is one method to know the preference consumer by showing the potential object. This recommendation also helps the consumer gets the preference object. One of the popular objects in the recommendation is anime film. In this case, we conduct research to recommend anime films based on ratings of previously watched films. Collaborative filtering is a technique that consists of calculating similarities, predictions, and recommendations. This study is taken from dataset Kaggle which consists of 73,516 users and 12,294 anime. A user's history will be matched with the whole user's history with alternating least squares (ALS) method. The anime will be recommended based on that results. This method is expected to help millions of users find the desired anime.

Keywords: Recommendation technique, collaborative filtering, prediction, anime

1. Introduction

With the development of technology, all forms of activities can be done digitally. Some information in cyberspace more and more with the growth of the internet in real-time which records every information received. As a consequence, much information is uncertain. It makes consumers or users are hard to find their preference because the information can be irrelevant to them. The recommendation system aims to resolve this problem. with a recommendation system, one can quickly access relevant information without searching manually. Many websites today benefit from a recommendation system for promoting and selling their products such as movies, hotels, restaurants, food and so forth. The recommendation can be built based on user profiles social media or track history in online stores which knows the user's behavior.

The recommendation system helps users to determine items that might be used. In terms of anime and manga, anime and manga versions are increasingly appearing with an interesting story and series quality. this will make it difficult for users to determine which movie to watch first. Collaborative filtering is a technique that is often used because it only relies on ranking data from previous users to be able to predict recommended films [1]

In this paper, we propose the use of a recommendation system using collaborative filtering (CF) to recommend anime. CF [2], which is one of the most popular techniques used wildly for recommendation systems. A simple recommendation system is used to measure the similarity between shows, users and helps to predict whether users will enjoy certain anime.

2. Related work

Many types of research about recommendations have been conducted using collaborative, content-based and hybrid both of them. Dev, in his research, uses a big data platform with MapReduce to develop a collaborative reconciliation system [3]. This system is able to reduce costs by removing the redundant computational process. Kumar [4] proposes a film recommendation system using collaborative filters that focus on the ratings given by users to provide recommendations. This allows the user to choose his choice from a given set of attributes and then recommends him to list the films based on the cumulative weights of the different attributes use algorithms k-means to recommend films based on the highest order. Shahjalal et al [5] implement a reconciliation system that combines both collaborative and user-based filtering approaches based on items. Researchers use the KUNN algorithm, a new algorithm for collaborative screening of one class, a setting that includes many applications.

3. Proposed method

3.1. Problem formulation

Table 1. Movie ratings by users.

Movies	User A	User B	User C	User D
P	5	5	1	1
Q	5	?	?	1
R	?	4	1	?
S	1	1	5	4
T	1	1	5	?

To know this problem clearly, table 1 is given as an example problem. It consists of five movies (P, Q, R, S, and T) and four users (A, B, C, and D). User A gives a rating for movies P, Q, S, and T with 5, 5, 1 and 1. User A likes movie P and Q very much with giving the highest score, 5. Contrarily, He/she doesn't like movies S and T with the lowest score 1. As a short, it shows that the more user like, the score will be more with 5 as the highest score. User A does not have a rating for movie R. It means user A has not watched this movie yet. With the same logic, Table 1 also shows the rating for the other users B, C and D for movies P, Q, R, S and T. The system is built to recommend for a user which movie should be watched by using the history data. It can be said the system will offer to each user which movies should be watched by sorting from the highest movie rating [6]. Figure 1 shows the process to build this recommendation system.

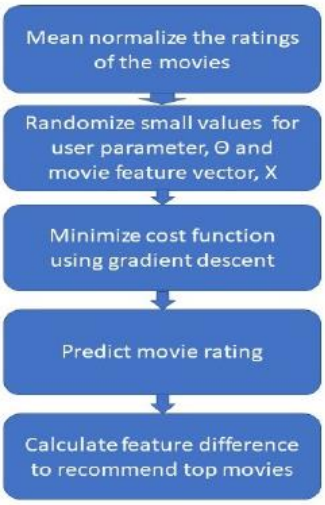

Figure 1. Recommendation system algorithm.

Journal of Physics: Conference Series **1566** (2020) 012057 doi:10.1088/1742-6596/1566/1/012057

3.2. Mean normalization

For new users who are registered will not have preferences related to the film because it has not been recorded, because of the supported activities. So, it is quite difficult to use the required film. This is the initial part of the debate which is known as a cold start problem. To overcome this, it is calculated from the average value of each film calculated from the film rating. So that the film rating will be normalized with films that have an average of zero. this will increase efficiency much better.

3.3. Similarity matrices

Interactions on users and movies are expressed in a matrix ranking that can be transformed. Where the node corresponds to users and items, and an edge that connects users and items. so that structural steps can be used in the bipartite graph to identify the environment and then determine the nearest neighbor of the target user. SimRank is a well-known tool and proven effective in reducing sparsity problems. In this paper, we include SimRank for calculating the similarity of users in the collaborative filtering model.

3.4.Recommend top movies

The results of minimizing the X matrix and user ratings by using the parameter θ and films with the X feature will get the formula θ TX. The approach is to calculate the difference in features between films. so for small differences, it will be recommended for users. For example, small differences | Xi - Xj | between films i and j mean they are similar. In this study, Top Anime will be recommended based on the friendliness of the anime from the largest to the biggest difference.

4. Result and analysis

4.1. Dataset

The film dataset used in our experiment was obtained from the anime recommendation database "Kaggle". Data provide two files namely anime data and rating on anime. This data consists of a relation between 73,516 users and 12,294 anime. Each user is able to add anime to their completed list and give it a rating and this data set is a compilation of those ratings.

```
In [5]: #anime_data$episodes <- as.numeric(anime_data$episodes)
        summary(anime_data)

    anime_id                               name                             genre
 Min.   :    1   Saru Kani Gassen           :    2   Hentai                 :  823
 1st Qu.: 3484   Shi Wan Ge Leng Xiaohua:    2   Comedy                 :  523
 Median :10260   _Summer                    :    1   Music                  :  301
 Mean   :14058   _Summer Specials           :    1   Kids                   :  199
 3rd Qu.:24794   .hack//G.U. Returner       :    1   Comedy, Slice of Life:  179
 Max.   :34527   .hack//G.U. Trilogy        :    1   Dementia               :  137
                 (Other)                    :12286   (Other)                :10132
      type          episodes          rating           members
        :  25   1      :5677   Min.   : 1.670   Min.   :      5
 Movie  :2348   2      :1076   1st Qu.: 5.880   1st Qu.:    225
 Music  : 488   12     : 816   Median : 6.570   Median :   1550
 ONA    : 659   13     : 572   Mean   : 6.474   Mean   :  18071
 OVA    :3311   26     : 514   3rd Qu.: 7.180   3rd Qu.:   9437
 Special:1676   3      : 505   Max.   :10.000   Max.   :1013917
 TV     :3787   (Other):3134   NA's   :230
```

Figure 2. Summary of anime data.

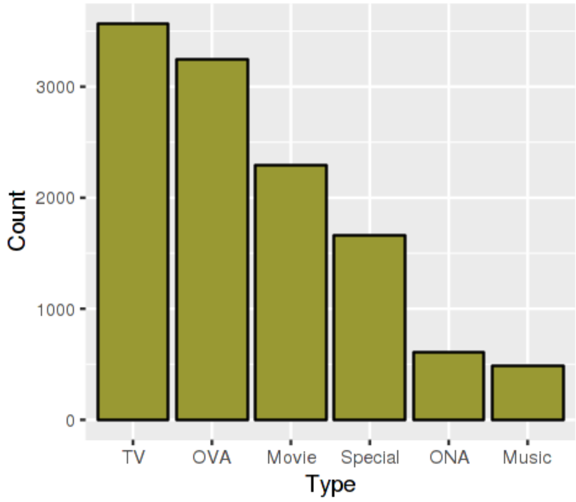

Figure 3. Distribution of types of shows.

4.2. Model evaluation

The collaborative filtering method is a method that guests the rating of the user's data based on the similarity history of the other user preference. Collaborative filtering is implemented by using matrix factorization [7]. Figure 5, Figure 6, Figure 7, and Figure 8 are the process to show the rating recommendations. The recommendation system is built by Alternating Least Squares (ALS) on the training data as shown in Figure 9.

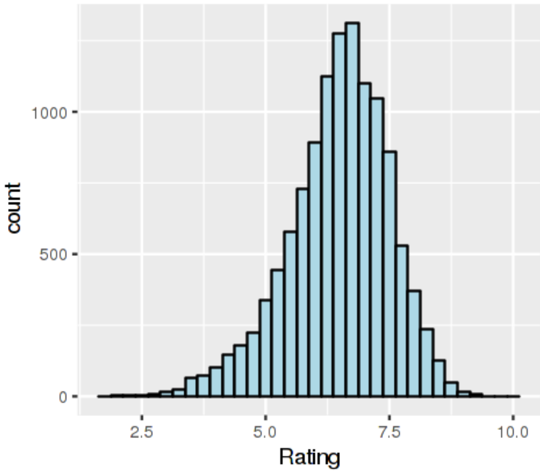

Figure 4. Histogram of the ratings.

```
In [29]: top_animes('Slam Dunk')

Similar shows to Slam Dunk include:

No. 1: Hajime no Ippo: New Challenger
No. 2: Beet the Vandel Buster Excellion
No. 3: Major S5
No. 4: Major S1
No. 5: Whistle!
No. 6: Gintama&#039;: Enchousen
No. 7: Aoki Densetsu Shoot!
No. 8: Major S4
No. 9: Major S2
No. 10: Major S6
```

Figure 5. Top animes recommended.

```
In [30]: top_users(3)

Most Similar Users:

User #934, Similarity value: 0.41
User #298, Similarity value: 0.27
User #335, Similarity value: 0.26
User #552, Similarity value: 0.22
User #969, Similarity value: 0.22
User #534, Similarity value: 0.21
User #977, Similarity value: 0.20
User #442, Similarity value: 0.19
User #312, Similarity value: 0.18
User #504, Similarity value: 0.17
```

Figure 6. Calculate the similarity value.

```
In [31]: similar_user_recs(3)

Out[31]: [('Death Note', 6),
          ('Fullmetal Alchemist: Brotherhood', 6),
          ('Shingeki no Kyojin', 5),
          ('Clannad', 2),
          ('Clannad: After Story', 2)]
```

Figure7. Recommendation similar user.

```
In [32]: predicted_rating('Slam Dunk', 3)

Out[32]: 8.245414581525795
```

Figure 8. Prediction rating for anime.

```
In [8]: (training, test) = ratings.randomSplit([0.7, 0.3])
```

```
In [9]: # Build the recommendation model using ALS on the training data
        # Note we set cold start strategy to 'drop' to ensure we don't get NaN evaluation metrics
        als = ALS(rank=20, #10 was by default
                  maxIter=2, regParam=0.01,
                  userCol="user_id", itemCol="anime_id", ratingCol="rating",
                  coldStartStrategy="drop",
                  implicitPrefs=False)
        model = als.fit(training)

        # Evaluate the model by computing the RMSE on the test data
        predictions = model.transform(test)
        evaluator = RegressionEvaluator(metricName="rmse", labelCol="rating",
                                        predictionCol="prediction")

        rmse = evaluator.evaluate(predictions)
        print("Root-mean-square error = " + str(rmse))

        Root-mean-square error = 2.5375315343383655
```

Figure 9. Build the recommendation model using ALS.

5. Conclusion

A better anime recommendation system based only on user watch history. A simple recommendation system can measure the similarity between performances, users and help predict whether users will enjoy certain anime. However, due to limitations calculating very large data sets, 100k data sets selected. Using a much more powerful machine, if a 1 million data set could be experimenting with. This paper shows a very simple but efficient recommendation system.

Reference

[1] Vie J J, Yger F, Lahfa R, Clement B, Cocchi K, Chalumeau T and Kashima H 2017 Using posters to recommend anime and mangas in a cold-start scenario IEEE 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) Vol 3 pp 21-26.

[2] Wang X, Xu Z, Xia X, Mao C 2017 Computing user similarity by combining simrank++ and cosine similarities to improve collaborative filtering IEEE 14th Web Information Systems and Applications Conference (WISA) pp 205-210

[3] Dev AV, Mohan A 2016 Recommendation system for big data applications based on set similarity of user preferences IEEE International Conference on Next Generation Intelligent Systems (ICNGIS) pp 1-6

[4] Kumar M, Yadav DK, Singh A, Gupta VK 2015 A movie recommender system: Movrec International Journal of Computer Applications Jan 1 124(3).

[5] Verstrepen K, Goethals B 2014 Unifying nearest neighbors collaborative filtering Proceedings of the 8th ACM Conference on Recommender systems pp 177-184

[6] Shahjalal MA, Ahmad Z, Arefin MS, Hossain MR 2017 A user rating-based collaborative filtering approach to predict movie preferences IEEE 3rd International Conference on Electrical Information and Communication Technology (EICT) pp 1-5

[7] C. M. Wu, D. Garg, and U. Bhandary 2018 A Collaborative Filtering Recommender System with Randomized Learning Rate and Regularized Parameter IEEE 9th Int. Conf. Softw. Eng. Serv. Sci. pp 11–15