I−018

Latent Factor Model for User Preferences and Prediction of TV Program in IPTV Environment

Quoc Hung DINH † Pao SRIPRASERTSUK† Wataru KAMEYAMA †

1. INTRODUCTION

Over the past decade, recommendation system has become a popular tool for supporting consumers who are deluging with their choices. Especially, many movie and TV program recommendation systems are available both in markets and research institutes. Nevertheless, those systems are not exploited deeply enough to find the user preferences together with user biases to predict effectively user evaluation of the recommended movies or TV programs. Thus, our work in this paper aims to explore the way to integrate user history of viewing, user evaluation and user bias in TV programs to build the model for recommendation system in IPTV environment. We propose a novel recommendation model which can generate list of programs that user will potentially want them with the highest possibly assessment. The proposed method is implemented by the approach of the latent factor model.

2. BACKGROUND

In general, recommendation system suggests those most potentially consumed items to users based on their explicit and implicit preferences. Consequently, recommendation system can be done by two different major techniques which are content filtering and collaboration filtering [1].

The content filtering approach such as [2] is quite simple, and it requires users to define their own preferences to characterize user's nature preferences. It can be used widely in Web based services.

The collaboration filtering approach is more sophisticated and there are so many primary algorithms used to accomplish this method. According to the study in [3], memory-based algorithm, vector similarity, or using a model-based method is an effective method.

In the work, we employ the collaboration filtering approach using model-based method for TV programs and users. In addition, the model uses latent factor techniques to predict user's future evaluation, thus it is also called latent factor model.

3. PROPOSED MODEL

3.1 Introduction of Latent Factor Model

To be assumed that user has a number *f* of different preferences for TV programs. Additionally, a TV program also has a number *f* of different characteristics to describe itself. Preferences can be sport TV program, romantic TV program, horror TV program and so on. Likewise, characteristics can also be sport, romance, horror and so on. Hence, each user's preferences and each TV program's characteristics can be denoted in *f* factors in f dimensional vector space as $\overrightarrow{p_u}$ and $\overrightarrow{q_i}$, respectively. In other words, in the same *f* factors in *f* dimensional space, more identical $\overrightarrow{p_u}$ and $\overrightarrow{q_i}$ means there are more potential that user will like TV program. In mathematical, the dot product of $\overrightarrow{p_u}$ and $\overrightarrow{q_i}$ is equal to the evaluation from user toward a TV program. Furthermore, the evaluation can be inferred from TV program rating patterns because of the logic that the higher rating means user enjoy the TV program.

The model based on the latent factor model that contains *N* users, *M* TV programs and user's evaluation in which each user's evaluations can be explained by the dot product of *f* factors of user vector and *f* factors of TV program. Therefore, the rating r_{ui} indicates the preference by user *u* of TV program *j*, where higher value means higher preference, denoted by the equation in [4].

Therefore, the major challenge for the model is how to compute the mapping process of each user and each TV program through the process of creating factor vector p_u and q_i.

After completing the mapping process, we can easily predict user's future evaluation by using equation (1):

$$r_{ui} = \overrightarrow{q_i}^T \overrightarrow{p_u} \quad (1)$$

3.2 Adjustment for Latent Factor Model

There are two adjustments for the model that are regularization, and adding user and movies biases.

The equation (1) is too simple, and it is easy to prone model overfitting. In order to avoid this overfitting, we need to regularize the model using Tikhonov regularization [4]. Moreover, the system minimizes the regularized squared error on the set of known behaviors in order to learn the factor vectors q_i and p_u. The algorithm to perform minimization is either stochastic gradient descent [4] or alternating least squares.

With the regard to the fact that each user has different preferences as well as bias to contents, the ratings from users are so varied. Hence, user and movies biases model is included in to the system in order to make it be more realistic and effective. Movie bias and user bias are denoted by mb_i and ub_u, respectively. Thus, equation (1) can be extended to equation (2):

$$r'_{ui} = r_{ui} + mb_i + ub_u \quad (2)$$

3.3 Proposed Model

Fig.1 shows the flow chart for the proposed IPTV recommendation system model

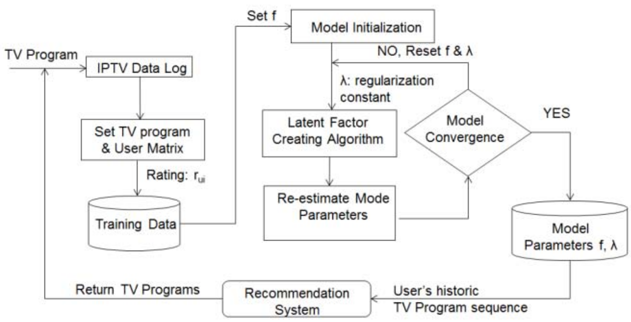

Fig1. The Proposed Model Flow Chart

† GITS, Waseda University

4. PRELIMINARY EXPERIMENT

We have built a prototype to evaluate the latent factor model by analyzing prediction results of user's future evaluation. The prototype accounts to two model adjustments which are regularization and adding biases.

4.1 Data Description

To test the model, we use dataset (10 movies) collected from IMDB (the Internet Movie Database) website, where each movie is rated from 1 to 10 by a large number of IMDB users. IMDB has total 17 demographics breakdown and their rating for each movie. Consequently, we have a matrix 10 x 17 and 170 rating data.

Subsequently, we randomly delete some rating data from the matrix to make empty spaces, and use our prototype to fill in those empty spaces in order to evaluate the accuracy of the proposed model.

4.2 Evaluation Methodology

We use root mean squared error (RMSE) to evaluate the prototype. The reason to use RMSE to evaluate the system is RMSE focuses more on large errors comparing with the alternative of mean absolute error. The equation of RMSE is show in equation (2) below:

$$RMSE = \sqrt{\frac{\sum_{(i,u)\in TestSet}(predict\_evaluation_{(i,u)} - real\_evaluating_{(i,u)})^2}{|TestSet|}} \quad (2)$$

4.3 Evaluation Result

In the first experiment, we use the test with 40 empty spaces out of 170, and then use two prototypes to predict the ratings in empty spaces and calculate the RMSE for each prediction, and the result is shown in Table 1 below:

Evaluation	Regularization	Adding Biases
RMSE	0.87	0.78

Table 1: The Evaluation Result

The meaning of 0.78 or 0.87 in Table 1 is clarified as following:

- If the real rating is 5
- Prototype averagely predicts: (5 - 0.78 = 4.22) to (5 + 0.78 = 5.78)

Fig.2, Fig3 and Fig.4 shows the original dataset, the regularization model result, and the biased model results.

7.5	7.9	8.1	8	8.5	7.6	7.6
7.5	8.4	6.4	5.7	9	7.9	7.7
7.7	8.3	8.5	8.5	9	7.9	7.8
5.5	6.2	6.5	6.4	7.2	5.8	5.7
7.2	7.9	7.9	7.7	8.3	7.4	7.3
3.4	6	5.4	5.3	6.2	3.7	3.2
5	6.9	7.3	7	8.4	5.7	5
6.6	7.9	7.6	6.6	8.7	7.1	6.6
2.2	5.5	3.1	2.6	6.4	2.3	2
5.8	8.1	6.8	5.9	8.4	6.2	5.7

Fig2. Original Data Set.

7.3911	8.3068	8.1	7.6143	8.5	7.6452	7.6
7.5	7.8879	6.4	5.7	9	7.9	7.7
7.3115	8.3	8.0044	7.5042	9	7.9	7.3269
5.5	6.2	6.5	6.4	6.4149	5.8	5.7
7.2	7.9	7.9	7.7	8.3	7.3352	7.3
3.4	4.5125	5.4	5.3	6.2	3.7	3.2
5	6.9	7.3	7	6.7667	5.7	5
6.6	7.7613	7.6	6.6	8.7	7.1194	6.6
2.2	5.5	3.6359	2.6	6.4	2.3	2
6.434	8.1	6.8	5.9	7.6894	6.2	6.4245

Fig3. Result of the Regularization Model.

7.5327	8.4443	8.1	7.7525	8.5	7.7796	7.6
7.5	8.0086	6.4	5.7	9	7.9	7.7
7.3947	8.3	8.0143	7.5868	9	7.9	7.4047
5.5	6.2	6.5	6.4	6.6944	5.8	5.7
7.2	7.9	7.9	7.7	8.3	7.4619	7.3
3.4	4.9365	5.4	5.3	6.2	3.7	3.2
5	6.9	7.3	7	6.9867	5.7	5
6.6	7.9069	7.6	6.6	8.7	7.2362	6.6
2.2	5.5	3.86	2.6	6.4	2.3	2
6.5053	8.1	6.8	5.9	7.7993	6.2	6.4682

Fig4. Result of the Model with User and Movie Biases.

In Fig3 and Fig4, those numbers with for 4 digits in decimal expansion are the predicted ratings.

4.4 Comparing with Mean Rating

Mean rating (MR) value is calculated by equation (3):

$$\text{MR for movie} = \frac{\text{Sum movie Rating Point}}{\text{Number of movie Rating Point}} \quad (3)$$

We fill the MR for those movies to all the empty spaces in each movie row of the matrix, because this is the simplest way to predict future evaluation from users.

Finally, RMSE for Mean Rating value is 1.32. Compared with RMSEs derived from latent factor model, it is shown that the model gives much better accuracy.

5. CONCLUSION & FUTURE WORKS

In this paper, we have studied the latent factor model for collaborative filtering in IPTV environment. In addition, user's movie evaluations can be transformed into the magnitudes of user's preferences and movie's characteristics. We have built a prototype to evaluate the proposed model. Using the dataset from IMDB in our experiment, the results show that the proposed model gives much better accuracy than the mean rating.

There are a lot of future works needed to be tackled. For instance, we need to find or create better algorithm to enhance the accuracy of model (reduce RMSE). Moreover, we need to consider other relevant information from user such as user's demographics (age, gender ...).

REFERENCES

[1] R. Burke, "*Hybrid recommender systems: survey and experiment*", user modeling and user-adapted interaction, 2002, vol. 12, No. 4, pp. 331-370.

[2] M.J. Pazzani, D. Billsus, "*Content-based recommendation systems*", in: P. Brusilovsky, A. Kobsa, W. Nejdl (Eds.), The Adaptive Web, Lecture Notes in Computer Science, vol. 4321, Springer-Verlag, 2007, pp. 325–341.

[3] J. S. Breese, D. Heckerman and C. Kadie, "*Empirical analysis of predictive algorithms for collaborative filtering*", in: Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, 1999, pp. 43–52.

[4] Y. Koren, R. Bell and C. Volinsky, "*Matrix Factorization Techniques for Recommender Systems*", Computer, vol.42, No.8, 2009, pp. 30-37.