Movie Recommendation based on User Similarity of Consumption Pattern Change

Minjae Kim
Dept. of Data Science,
Graduate School, Dankook University
Yongin-si, Korea
mbf3279@naver.com

SungHwan Jeon
Dept. of Data Science,
Graduate School, Dankook University
Yongin-si, Korea
72160261@dankook.ac.kr

Heeseong Shin
Dept. of Applied Computer Eng.,
Dankook University
Seoul, Korea
dk.hsshin@gmail.com

Wonseok Choi
Dept. of Computer Science & Eng.
Graduate School, Dankook University
Yongin-si, Korea
one8291@gmail.com

Haejin Chung
Software-Centric University Project Office
Dankook University
Yongin-si, Korea
haejini.chung@gmail.com

Yunmook Nah
Dept. of Applied Computer Eng.,
Dankook University
Yongin-si, Korea
ymnah@dankook.ac.kr

***Abstract*—The recurrent neural network(RNN) deep learning algorithm, which mainly learns and predicts sequence data and time series data, is mainly used in language modeling, stock price prediction, and chat bot. In this paper, we propose a method of predicting and recommending a movie by considering movie consumption patterns of users. We measure the similarity between users based on movie rating data, classify users with similar movie preferences, and learn the consumption pattern of each similar user group to improve the prediction accuracy by considering the change of preference over time. In order to show the effectiveness of the proposed method, we apply the collaborative filtering algorithm, the simple RNN and our modified RNN and compare their prediction accuracies.**

Keywords—movie recommendation, user similarity, consumption pattern, sequence data, Recurrent Neural Network

I. Introduction

Recently, many studies are being conducted to predict and recommend products to be purchased in the near future through customized analysis of individual users, and applications such as Netflix recommending movies and Amazon recommending products are increasing. In order to predict future consumption, user-based or item-based collaborative filtering algorithms are usually used. However, these methods are usually based on rating data provided by users, and this means that it can't be predicted without rating data. Therefore, a product without rating data can't be included in the recommendation list, which causes a sparsity problem. Also, since these conventional methods do not consider the time changes, it does not reflect the consumption pattern changes of users.

In this paper, we propose a method of predicting and recommending a movie by considering movie consumption patterns of users. We measure the similarity between users based on movie rating data and classify users with similar movie preferences. We then apply RNN to learn movie consumption pattern of similar user groups and later to predict or recommend movies. To show the effectiveness of the proposed method, we apply the collaborative filtering algorithm, the simple RNN and our modified RNN for the MovieLens Latest Dataset and compare their prediction accuracies.

This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program(IITP-2019-2015-0-00363) and under the National Program for Excellence in SW(2017-0-00091) supervised by the IITP(Institute for Information & communications Technology Promotion).

The remainder of this paper is organized as follows. Section 2 proposes a recommendation model based on RNN and user similarity. In Section 3, we compare the prediction accuracies of the collaborative filtering algorithm, the simple RNN and our modified RNN, to show the effectiveness of the proposed method. Finally, Section 4 concludes the paper.

II. Movie Recommendation Method

In recommender systems, recommendations can be based on demographics of the users, overall top selling items, or past buying habit of users[1]. Collaborative Filtering(CF) is the most successful recommendation technique to date. The basic idea of collaborative filtering-based algorithms is to provide item recommendations or predictions based on the opinion of other like-minded users[2]. The recurrent neural network(RNN) uses sequence data as input data, unlike existing deep artificial neural networks, in which each input/output is independent of each other[3,4,5,6,7,8].

In order to solve the sparsity problem arising from the recommendation of the collaborative filtering method and to consider the consumption pattern changes of the users, we use the Pearson correlation coefficient to classify like-minded users and apply recurrent neural networks to similar user groups.

We calculated the Pearson correlation coefficient, which ranges from 1 to -1, for the movie rating data to measure the similarity of users having similar ratings for the same movie. Generally, if the similarity is more than 0.3, it is interpreted that there is a clear positive correlation. Therefore, users with similarity of 0.3 or more to a specific user are classified into similar user groups.

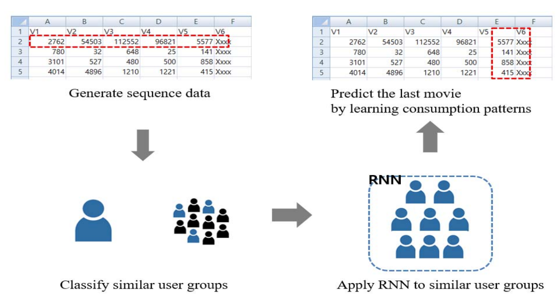


Fig. 1. Proposed method

978-1-7281-1488-0/19/$31.00 ©2019 IEEE
DOI 10.1109/AIKE.2019.00064

In our method, the input layer of the RNN is the sequence data which lists the last movies watched by each user in chronological order. The hidden layer learns and memorizes the movie consumption pattern of each user group by time order. The output layer uses the movie consumption pattern learned in the hidden layer, and predicts the movie to be recommended to users. Figure 1 shows the summary of our proposed method.

III. EXPERIMENTS

In our experiment, we used the MovieLens Latest Dataset (http://movielens.org) which includes the rating and the time of the movies consumed by users from 1995 to 2017. From this dataset, we used data consisting of 1,048,576 records, 10,656 users, and 45,000 movies.

A. Recommendation by Collaborative Filtering

We constructed the user-item rating matrix, as shown in Figure 2, which shows the rating data given by each user for each movie.

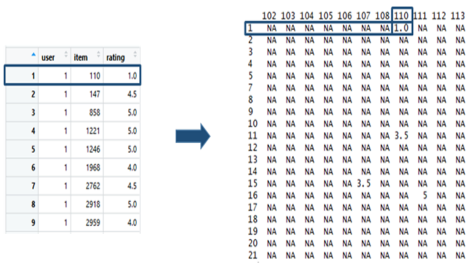


Fig. 2. User-item rating matrix

Figure 3 shows the recommendation result by collaborative filtering. The left column is user id and the right column is recommended movie id. For this method, the root mean square error(RMSE) is 1.03, the mean square error(MSE) is 1.07 and the mean absolute error(MAE) is 0.80.

```
$ 1   : chr "904"
$ 2   : chr "1196"
$ 3   : chr "296"
$ 4   : chr "858"
$ 5   : chr "356"
$ 7   : chr "593"
$ 8   : chr "356"
$ 9   : chr "24"
$ 10  : chr "1265"
$ 11  : chr "467"
$ 12  : chr "665"
$ 13  : chr "1201"
$ 14  : chr "79132'
$ 15  : chr "593"
```

Fig. 3. Recommendation result by collaborative filtering

B. Recommendation by RNN

In RNN, each data in a sequence becomes a feature, thus longer sequence has more features than shorter sequence. This leads to poor prediction accuracy when the given sequence is relatively long. We measured prediction accuracy, while increasing the length of sequence from 5 to 20. In our experiment, the sequence with length 5 showed best prediction accuracy, as shown in Figure 4. Therefore, we selected the last 5 movies seen by 10,656 users, as shown in Figure 5.

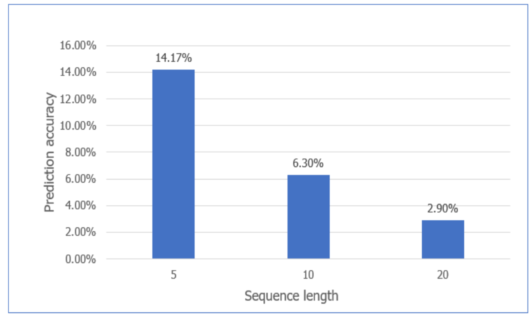
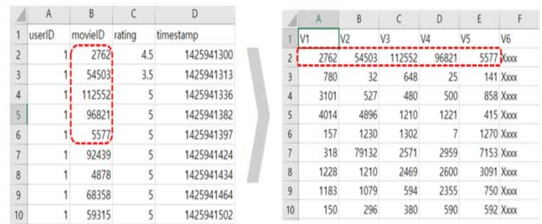


Fig. 4. Prediction accuracy according to the sequence length



userID	movieID	rating	timestamp
1	2762	4.5	1425941300
1	54503	3.5	1425941313
1	112552	5	1425941336
1	96821	5	1425941382
1	5577	5	1425941397
1	92439	5	1425941424
1	4878	5	1425941434
1	68358	5	1425941464
1	59315	5	1425941502

V1	V2	V3	V4	V5	V6
2762	54503	112552	96821	5577	Xxxx
780	32	648	25	141	Xxxx
3101	527	480	500	858	Xxxx
4014	4896	1210	1221	415	Xxxx
157	1230	1302	7	1270	Xxxx
318	79132	2571	2959	7153	Xxxx
1228	1210	2469	2600	3091	Xxxx
1183	1079	594	2355	750	Xxxx
150	296	380	590	592	Xxxx

Fig. 5. Generation of sequence data

We applied the RNN model by using 80% of user group data as training data set and 20% of user group data as testing data set. The training data set is used to learn and model the movie consumption patterns of similar user groups, and this model is tested by using the test data set to predict the movies that each user last consumed in the test data set. Experiments were conducted to measure prediction accuracies, while changing the number of hidden layer nodes, the number of learning iterations, and the learning rate.

We set up the number of hidden layer nodes as 16, 32, 64, 96 and 128, and the best prediction accuracy was obtained when the number of hidden layer nodes was 96, as shown in Figure 6. We also set up the number of iterations as 1500, 2000 and 2500, and the best prediction accuracy was obtained when the number of iterations was 2000, as shown in Figure 7. Experiments were conducted with learning rates 0.01, 0.02, and 0.03, and the highest prediction accuracy was obtained when the learning rate was 0.02, as shown in Figure 8.

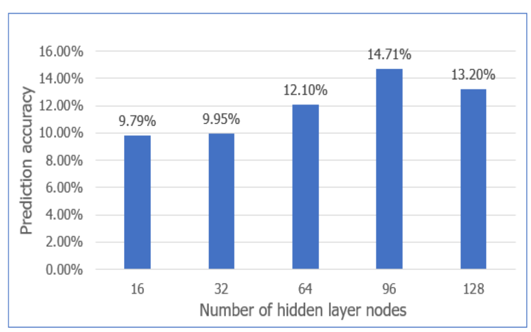


Fig. 6. Prediction accuracy according to the number of hidden layer nodes

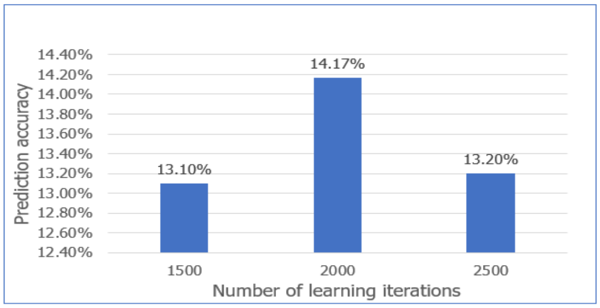


Fig. 7. Prediction accuracy according to the number of iterations

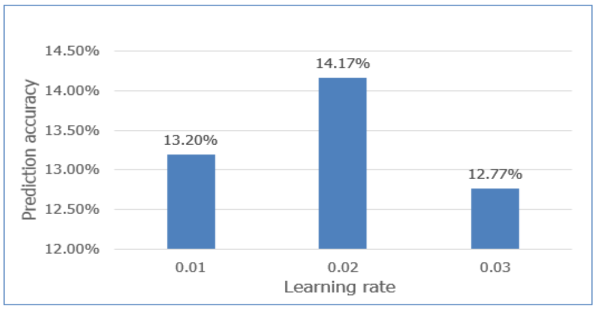


Fig. 8. Prediction accuracy according to the learning rate

C. Comparison of Results

We compared the prediction accuracies of the collaborative filtering, the simple RNN for all users and the modified RNN for similar user groups. The prediction accuracy was calculated by dividing the number of correct prediction of last movie by the number of records in the test data set. Experimental results show that the accuracy of the cooperative filtering is 4.8%, the accuracy of the simple RNN is 11.5% and the accuracy of the modified RNN for similar user groups is 14.17%, as shown in Figure 9. The modified RNN for similar user groups shows 2.95 times better prediction accuracy than the collaborative filtering and 1.23 times better accuracy than the simple RNN for all users.

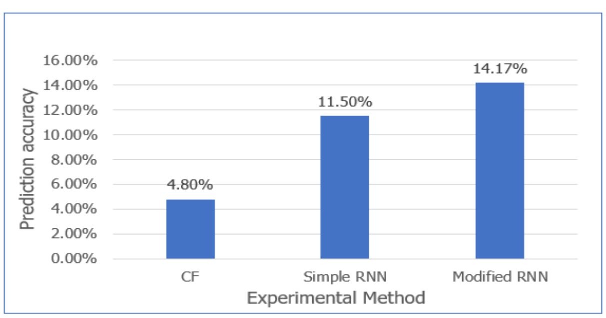


Fig. 9. Comparison of prediction accuracies

IV. CONCLUSION

In this paper, we classify similar user groups with similar taste preferences through movie rating data set for 'movie' items, and apply RNN learning method to them. In this way, the movie consumption pattern of the similar user group is learned, and a model which recommends movies to users is proposed. This model overcomes the sparsity problem which is the biggest problem in the current recommendation systems based on the existing rating data. Also, it can recommend movies by considering the dynamically changing consumption patterns over time.

In addition to the 'movie' item, the recommendation model can also be used as a recommendation system in areas such as books and clothing that have individual taste and are likely to change with the passage of time.

In this paper, we predicted a single movie which is likely to be consumed by the given user from 45,000 movies. For more practical applications, multiple recommendation methods that recommend multiple similar movies at the same time, considering genre, actor, director, etc., will be more desirable.

REFERENCES

[1] Son, J., Kim, S.B., Kim, H. and Cho, S., "Review and Analysis of Recommender Systems," J. of the Korean Institute of Industrial Engineers, Vol.41, No.2, pp.185-208, 2015.

[2] Sarwar, B., Karypis, G., Konstan, J., and Riedl, J. "Item-Based Collaborative Filtering Recommendation Algorithms," in Proc. WWW, pp.285-295, 2001.

[3] Mikolov, T., Karafiat, M., Burget, L. Cernocky J. and Khudanpur, S., "Recurrent neural network based language model," in Proc. ISCA(International Speech Communication Association), pp.1045-1048, 2010.

[4] Sutskever, I., Vinyals, O. and Le, Q.V., "Sequence to Sequence Learning with Neural Networks," in Proc. NIPS(Advances in Neural Information Processing Systems), pp.3104-3112, 2014.

[5] Zaremba, W., Sutskever, I. and Vinyals, O., "Recurrent Neural Network Regularization," CoRR abs/1409.2329, 2014.

[6] Sundermeyer, M., Schluter, R. and Ney, H., "LSTM Neural Networks for Language Modeling," in Proc. ISCA(International Speech Communication Association), pp.194-197, 2012.

[7] Yu, F., Liu, Q., Wu, S., Wang, L. and Tan, T. "A Dynamic Recurrent Model for Next Basket Recommendation," in Proc. SIGIR, pp.729-732, 2015.

[8] Liu, D.Z. and Singh, G. "A Recurrent Neural Network Based Recommendation System," 2015.