Tourism Site Recommender System Using Item-Based Collaborative Filtering Approach

R.A. Nugroho[1,*], A.M. Polina[1], Y. D. Mahendra[1]

[1]*Department of Informatics, Faculty of Science and Technology,*
Sanata Dharma University, Yogyakarta, Indonesia
*Corresponding Author: robertus.adi@usd.ac.id

Abstract

Many people like traveling. However, often it is difficult for them to find a tourism site that they like much. Too many information about tourism is the problem. To overcome this problem, we need to filter the information. Recommender System could filter the information. By considering the advantages, the system used item-based collaborative filtering approach to give recommendation. Some tourism sites around Daerah Istimewa Yogyakarta province were used in this research. The system is able to give recommendation to users. The accuracy of the rating prediction is 0,6293 and the average time consumption is 1693,33 milliseconds.

Keywords: recommender system, tourism, collaborative filtering

1 Introduction

Traveling has become a lifestyle for Indonesian people today. Various tourist destinations opened and managed to attract tourists. One of the provinces that is always attractive as a tourist destination is Yogyakarta. Many tourists visit this area. According to tourism statistics, from 2014 to 2018 there has been an increase in the number of tourists visiting Yogyakarta [1]. Various types of tourist destinations, including natural

tourism, artificial tourism, and cultural tourism can be found here. Sometimes these attractions are not detected by tourists. It is not that there is no information, but too much information about tourism sites is the cause. Too much information make tourists have difficulty finding interesting objects for them. To overcome this problem, we need to reduce the information given to the user. Only relevant information is provided to its users (tourists).

The recommender system is a system that is able to suggest items that users like [2]. This system is able to provide information according to the user's preferences. Various fields have implemented recommender systems as a solution in filtering the information that will be provided to users. In the field of tourism, a recommender system is needed, especially in reducing the number of information of tourism sites that will be provided to users [3].

In the recommender system, there are two approaches that are commonly used, namely collaborative filtering and content-based filtering [4]. Collaborative Filtering tries to predict what users like by comparing user profiles with one another. In Collaborative Filtering, information about user's preferences are very important. If there is too little information about user's preferences, the system will have a cold-start problem so it cannot predict well. Meanwhile, Content-based Filtering predicts what users like now by looking at what users liked in the past. The system will look for similarities between the content of objects and the user profile. The more similar the object is, the more recommended it is to users.

Collaborative filtering consists of user-based and item-based approaches [4]. Both require information about the preferences of users. The differences are the user-based looks at the relationship between users, while item-based looks at the relationship between items [5]. Relationships between items are considered more static (do not change much) than relationships between users. This results in less computational burden in providing recommendations.

Therefore, in this research, we proposed a system for recommending tourism sites using item-based collaborative filtering approach.

2 Research Methodology

This research used 10 tourism sites in Yogyakarta, Indonesia. The chosen tourism sites were popular or widely visited tourism sites in Yogyakarta according to the Yogyakarta Tourism Statistics [1]. Table 1 shows the tourism sites used in this research.

Table 1. Tourism Site List

No	Tourism Site	Abbreviation
1	Museum TNI AU Dirgantara Mandala	MTAU
2	Monumen Jogja Kembali	MJK
3	Tebing Breksi	TB
4	Kraton Ratu Boko	KRB
5	Museum Benteng Vredeburg	MBV
6	Taman sari	TS
7	Kraton Yogyakarta	KY
8	De Mata Art Museum	DMA
9	Taman Pintar	TP
10	Candi Prambanan	CP

Cold-start problem is a condition when we do not have enough ratings related to items [6]. To avoid cold-start problem, we first collected some ratings from tourists for these sites through a survey. The survey involved five respondents. The rating range given by tourists were 1 to 5 (1; 1,5; 2; 2,5; 3; 3,5; 4; 4,5; 5).

Table 2. User-Item Matrix

	MTAU	MJK	TB	KRB	MBV	TS	KY	DMA	TP	CP
User 1	5,00	5,00	4,00	4,00	5,00	5,00	5,00	4,00		
User 2		3,50	4,00			3,50	4,00		3,00	4,50
User 3	1,00		2,50	2,50	2,50	3,50	1,00	2,00		3,50
User 4		4,00		4,00	4,00	5,00	3,00	5,00	5,00	5,00
User 5	3,50	4,00	4,00	4,00	3,00	3,00	4,50	3,50	4,00	5,00

Higher value of ratings indicates that the tourist is more interested with the tourism site. After the rating was obtained, a user-item matrix was formed. This matrix shows the rating given by tourists (users) to certain tourism sites (items). In the user - item matrix, some cells appear empty (see Table 2). This means that the tourist did not give a rating for a tourism site. Not giving a rating because tourists had never visited these tourism sites.

After the user - item matrix was formed, the process was continued by looking for similarity between item which the rating will be predicted to all items in the matrix.

Only co-rated cases (the users rated both item i and j) was used in the calculation (see Figure 1).

$$S(i,j) \rightarrow \text{Similarity item } i \text{ and item } j$$

	I$_1$	I$_2$	...	I$_i$	I$_j$	...	...	I$_{n-1}$	I$_n$
U$_1$				R	R				
U$_2$				-	R				
U$_3$				R	-				
U$_4$				R	R				
U$_5$				R	R				

Figure 1. Finding Similarity between Items

To calculate the similarity, this research used Pearson Corellation as follows

$$PC(i,j) = \frac{\sum_{u \in U_{ij}}(r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{ij}}(r_{ui} - \bar{r}_i)^2 \cdot \sum_{u \in U_{ij}}(r_{uj} - \bar{r}_j)^2}} \quad . \tag{1}$$

From equation (1), it is denoted $PC(i,j)$ is a similarity value between item i and item j; $r_{u,i}$ dan $r_{u,j}$ are ratings given by user i and j to item i; while, $\bar{r}_i$ and $\bar{r}_j$ are the average ratings of item i and j.

By considering the similarity value between items, top-n-neighbors were chosen. These neighbors were used to predict the rating that the active user would give to an item. The predictive rating is calculated by

$$P(a,j) = \frac{\sum_{i \in I}\left(R_{a,i} \cdot S_{i,j}\right)}{\sum_{i \in I}\left| S_{i,j} \right|} \quad . \tag{2}$$

From equation (2), it is denoted $P(a, j)$ is predicted rating given by user a to item j; $R_{a,i}$ is a rating given by user a to item i; while, $S_{i,j}$ is similarity value between item i and item j.

Based on this predictive rating, the system considered to recommend it to user or not. The higher the predictive rating leads to the greater the chance that the item will be recommended to a user. In general the recommendation process that use item-based collaborative filtering approach can be depicted in Figure 2 [7]. This research followed this process.

Figure 2. Recommendation Process

At the end, the quality of the system will be evaluated by measuring the accuracy of the predictive rating and measuring the time consumption of the predicting process.

3   Results and Discussions

The system is evaluated by measuring the magnitude of the error rate in predicting the rating given by the user for a tourist site. To measure the level of error prediction, we use MAE (Mean Absolute Error). System evaluation is carried out with several scenarios. The scenario is to use several numbers of the nearest neighbors (top N neighbor) in predicting the ratings. The number of nearest neighbors that we used are top 4 neighbors, top 6 neighbors, and top 8 neighbors. We choosed those number of nearest neighbors because we only involved 10 tourism sites.

Table 3. Evaluation Results

Top N Neighbors	MAE	Time Consumption (miliseconds)
4	0,6334	1654
6	0,6254	1797
8	0,6291	1629

From the results of the evaluation (see Table 3), it can be seen that the smallest error rate occurs when using the top 6 neighbors, which is 0,6254 (see Figure 3). However, the differences in error rates in the three scenarios are not significant.
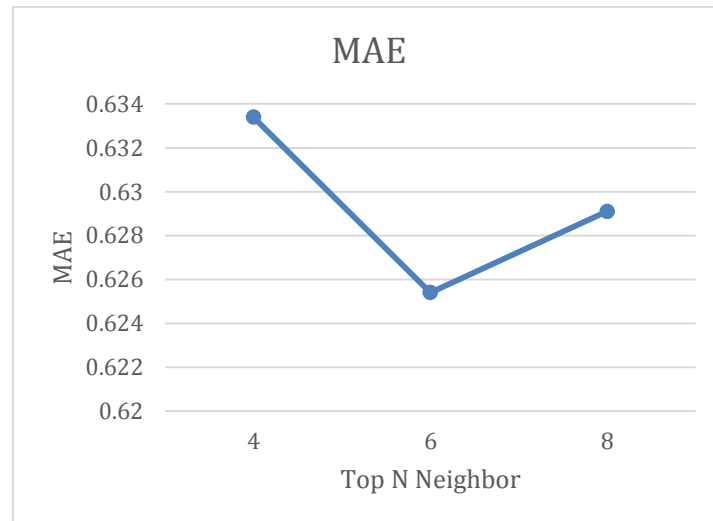
Figure 3. Mean Absolute Error

In addition, we also measure the time consumption required by the system to complete the recommendation process. From the results of test, it can be seen that the top 6 neighbors require the highest time consumption (see Figure 4), which is 1797 milliseconds. The differences in time consumption between one scenario and another is not significant.
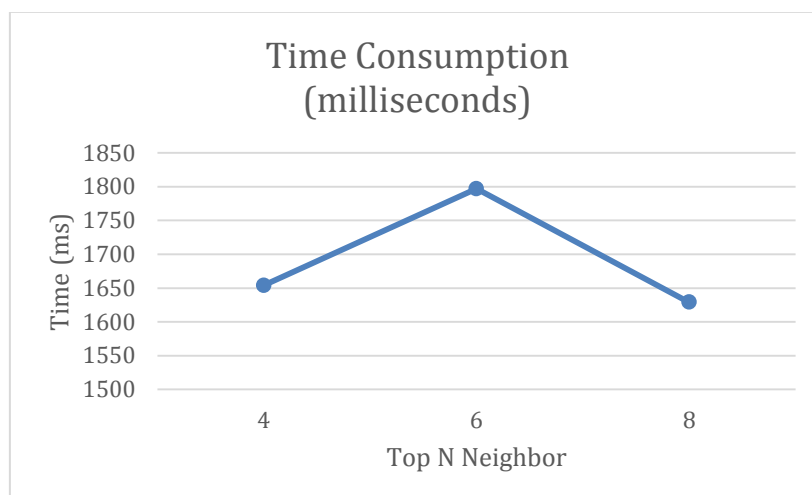
Figure 4. Time Consumption

4 Conclusions

From the experimental results, it can be concluded that the tourism site recommendation system is able to provide recommendations to users quite well. The item-based collaborative filtering approach is able to predict the rating that given by users with an average MAE of 0,6293 and an average time consumption of 1693,33 milliseconds. The weakness of this research is the small number of users and tourism sites involved. In future, it is necessary to involve more users and tourism sites so that the scalability of the system can be measured properly, especially regarding the computational load. To improve accuracy, it is necessary to implement another similarity function.

References

[1] Statistik Kepariwisataan. Dinas Pariwisata Daerah Istimewa Yogyakarta, 2018.

[2] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook." in *Recommender Systems Handbook,* F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor (Eds.) Boston, Springer, 1–35, 2011.

[3] D. Gavalas, C. Konstantopoulos, K. Mastakas, and G. Pantziou, "Mobile recommender systems in tourism." *Journal of Network and Computer Applications,* **39**, 319–333, 2014.

[4] C. Desrosiers and G. Karypis, "A Comprehensive Survey of Neighborhood-based Recommendation Methods." in *Recommender Systems Handbook,* F. Ricci, L. Rokach, B. Shapira, and P. B. Kantor (Eds.) Boston, Springer, 107–144, 2011.

[5] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-Based Collaborative Filtering Recommendation Algorithms." in *Proceedings of the 10th International Conference on World Wide Web,* New York, 285–295, 2001.

[6] S. Jain, A. Grover, P. S. Thakur, and S. K. Choudhary, "Trends, problems and solutions of recommender system." in *International Conference on Computing, Communication Automation,* 955–958, 2015.

[7] K. Falk, *Practical Recommender Systems,* Manning Publications, 2019.

This page intentionally left blank