NON-NEGATIVE MATRIX FACTORIZATION OF STORY WATCHING TIME OF TOURISTS FOR BEST SIGHTSEEING SPOT AND PREFERENCE

Motoki Seguchi[1], Fumiko Harada[2] and Hiromitsu Shimakawa[1]

[1]College of Information Science and Engineering,
Ritsumeikan University, Kusatsu, Shiga, Japan
[2]Connect Dot Ltd., Kyoto, Japan

ABSTRACT

In this research, we propose a method of recommending the best sightseeing spot through watching stories of sightseeing spots. It predicts the rating for each sightseeing spot of a target tourist based on Non-negative Matrix Factorization on the story watching times and ratings of tourists. We also propose to estimate the degree of the target tourist's preference for a sightseeing spot. Tourists visit a sightseeing spot for a certain purpose of tourism. The preferences of tourists appear prominently in their purposes of tourism. In addition, the degree of the tourists' preferences for sightseeing spots differs depending on the sightseeing spot. If we can estimate the degree of preference of a tourist, it will be possible to recommend a sightseeing spot that can achieve his purpose of tourism.

KEYWORDS

Sightseeing, Recommendation, Interest Estimation, Story Watching, Preference.

1. INTRODUCTION

In recent years, the development of the global tourism industry has been remarkable [1]. One of the backgrounds for the development of the tourism industry is the enhancement of tourist information. Tourist information is transmitted not only by organizations such as companies and mass media but also by various people through social networking service (SNS). The amount of tourist information is increasing every day. With the spread of smartphones, anyone can easily obtain tourist information. However, it is difficult to find the best sightseeing spot for oneself from the huge amount of tourist information.

To solve such the problem, various research on recommendation methods of the best sightseeing spots have been developed [2], [3], [4], [5], [10], [13], [14]. However, these researches aim to improve the accuracy of recommendation, and they did not discuss the effort tourists spend to select a sightseeing spot. There are also many tourist guide apps that help tourists select sightseeing spots. However, they also provide a wide variety of information, which complicates tourists' select of sightseeing spots. Tourists want a method that allows them to select a sightseeing spot that suits their wishes with little effort.

Dhinaharan Nagamalai et al. (Eds): CSEIT, WiMoNe, NCS, CIoT, CMLA, DMSE, NLPD - 2020
pp. 177-192, 2020. CS & IT - CSCP 2020 DOI: 10.5121/csit.2020.101116

Tourists visit sightseeing spots for a certain purpose. The purpose of tourism of a tourist is always explicit, but variable. Therefore, it is necessary to recommend sightseeing spots that suit the purpose of tourism at that time. The preferences of tourists appear prominently in the purposes of tourism. In addition, the degree of the tourists' preferences for sightseeing spots differs depending on the sightseeing spot. If we can estimate the degree of preference of a tourist, we will be able to recommend sightseeing spots that can achieve the tourist's purpose of tourism.

It is possible to use storytelling marketing for recommending sightseeing spots [5]. The storytelling marketing can arouse tourists' empathy by using stories from a third-party perspective on sightseeing spots. Storytelling marketing is useful because it allows tourists to receive detailed information about attractions of sightseeing spots [7]. In addition, reviews of various sightseeing spots can be used as stories [11]. Therefore, many posts about sightseeing spots written on SNS can be used as stories.

In this research, we propose a method of recommending the best sightseeing spot by using non-negative matrix factorization on the story watching times of tourists. This method estimates the best sightseeing spot from the similarity of the behavior regarding the watching of the story between the target tourist who will visit a sightseeing spot and the tourists who have visited the sightseeing spot in the past. In this study, the only task to be given to the target tourists to recommend sightseeing spots is to watch the stories. Tourists can use this method as if they were enjoying surfing the internet on SNSs, and the load on the tourists is small. In this study, we propose the methods not only to recommend the best sightseeing spot, but also to estimate the degree of the tourists' preferences for each sightseeing spot. If we can estimate the degree of a tourists' preference for a sightseeing spot, we will be able to recommend the other sightseeing spots which brings him the attractiveness same to that of the sightseeing spot according to his preference.

In this paper, Section 2 describes the current state of tourism. Section 3 describes related work. Section 4 proposes the method of estimating the best sightseeing spot and degree of tourist preference for sightseeing spots. Section 5 describes the experiments that evaluate the proposed method and the results. Section 6 describes the evaluation of experimental results and their consideration.

2. CURRENT STATE OF TOURISM

2.1. Purpose of Tourism

Tourists visit sightseeing spots for a purpose of tourism, such as "I want to see a beautiful scenery" and "I want to heal fatigue." Van Harssel's research [6] classified tourism into the following 10 types, based on the purposes of tourism that tourists mainly aim at.

1) Nature trip
2) Cultural trip
3) Social trip
4) Activity trip
5) Recreational trip
6) Sports trip
7) Special trip
8) Religious trip
9) Health trip
10) Ethnic trip

For example, the nature trip applies to the case that a tourist wants to see a beautiful scenery. The Health trip applies to the case that he wants to heal fatigue. Naturally, tourists want to achieve their purposes of tourism. Therefore, tourists should select the tourism type according to the purpose of tourism. In addition, any sightseeing spot has a suitable tourism type. Therefore, tourists cannot achieve the purpose of tourism unless they select a sightseeing spot that matches the tourism type corresponding to the purpose of tourism.

On the other hand, people have a preference that is a desire to be satisfied in daily life. One person's preference is "eating" and another person's preference is "seeing." When preferences are not satisfied, people induce extraordinary behavior to satisfy them. It is considered that one of the behaviors is that people are going to travel [10]. Therefore, tourists' preferences are prominent for purposes of tourism. For example, the tourism purpose of "healing fatigue" would strongly include the preference of "healing." If we can grasp the potential preference and its degree of each tourist that he does not want to say, we can consider that the preference is not satisfied in recent daily life. Therefore, if we can recommend the tourist a sightseeing spot where he will be able to achieve satisfaction of the preference as the purpose of tourism, he can be pleased much with the tourism.

2.2. Attractions Associated with Sightseeing Spots

Sightseeing spots have various attractions. "Nature," "environment," "facility," and "events" are often cited as attractive factors of sightseeing spots. Mill [9] roughly divided the attractions of sightseeing spots into the following 8 categories.

1) Sun, sea and resort
2) Landscape
3) Animal
4) Hot springs and health resorts
5) Urban attractive conditions
6) Local attractive conditions
7) Sports event
8) Systematically developed attractive conditions

Hudman and Hawkins [8] divided the attractive factors in the 8 kinds of attractions into the following twelve categories.

1) Buildings and their environment
2) Cultural activities
3) Religion
4) Politics
5) Science
6) Nature
7) Climate
8) Scenery
9) Outdoor life
10) Outdoor recreation and sports
11) Entertainment
12) Health and hot springs

There is a correspondence between the preferences of tourists and the attractions associated with sightseeing spots. For example, tourists who have a strong preference to "seeing" strongly like sightseeing spots that have a strong attraction of "scenery."

Sightseeing spots provide various attractions in various degrees. On the other hand, how to feel the attraction is different for each person. The preferences that tourists have for sightseeing spots differ depending on how they feel the attraction. For example, a person who feels that a national park is strongly associated with the attraction of "scenery" has a strong preference of "seeing." A person who feels that the park is strongly associated with the attraction of "entertainment" has a strong preference of "playing." It is possible to estimate the attractions of a sightseeing spot by estimating the tourists' preferences for the sightseeing spot because the attractions of sightseeing spots and the preferences of persons correspond. If we can estimate the attractions associated with the sightseeing spot for a person, we can recommend him another sightseeing spot that he will feel the same attractions as that sightseeing spot.

3. RELATED WORK

In recent years, the tourism industry has grown [1]. It can be said that the enrichment of tourist information has contributed significantly among many factors for growth. However, as tourist information continues to increase, it has become difficult for tourists to obtain suitable tourist information. Tourists have to select the most suitable one for themselves from the vast amount of tourist information. Therefore, it is difficult for tourists to find the best sightseeing spot for themselves. Therefore, in recent years, research that recommends the best sightseeing spot to tourists has become popular.

There are methods of recommending sightseeing spots using location information obtained by GPS [13], [14]. These methods recommend the best tourist plan to tourists by learning the movement history of the tourists in the sightseeing spots. However, radio waves from positioning satellites cannot be captured everywhere and lack stability. For example, there is a large error in the location information in undergrounds or in forests where radio waves are hard to connect. In some cases, there are even things that cannot be supplemented. Also, these studies do not consider tourists' preferences.

There is also sightseeing spots recommendation methods that use personal data of tourists such as time spent for sightseeing [2], [3], [4]. Because these methods allow users to input personal data such as budget into a device before sightseeing, personalized tourist information can be provided. However, tourists must decide in advance the time and budget to spend on sightseeing. These methods cannot be used when tourist time and budget are undecided, because personal data of tourists are insufficient. Moreover, some tourists may find it annoying to input their personal data into the device. Furthermore, these methods do not take into account the preferences of tourists.
No matter how wonderful the attraction of a sightseeing spot is, if the attraction of the sightseeing spot does not match the preference of a tourist, the sightseeing is worthless for him. It is necessary to consider a tourist's preference and propose a method of recommending a sightseeing spot that has an attraction that matches the preference.

4. A RECOMMENDATION METHOD OF SIGHTSEEING SPOTS THROUGH WATCHING STORIES

4.1. Recommendation of Sightseeing Spots and Estimation of Preferences Using Story Watching Time

In this research, it is assumed that the user watches stories of sightseeing spots of interest. Therefore, it is considered that the user's interest in sightseeing can be estimated by the length of the story watching time. In this research, we propose a method of recommending the best sightseeing spot to the target user by analyzing the target user's story watching time with the past user's data. In this research, the user who is going to visit sightseeing spots from now on is called the target user. The user who has visited a sightseeing spot in the past is called the past user. In this research, we consider recommending a sightseeing spot to the target user based on the similarity of watching behavior between past users and the target user. In this research, we consider only that the accuracy of recommendation to the sightseeing spots itself is improved, but also whether the degree of the user's preference for the sightseeing spots can be estimated. If the latter can be achieved, it is possible not only to recommend a target user a suitable sightseeing spot visited by past users, but to recommend another sightseeing spot having the same attraction as that of the suitable sightseeing spot according to the preference of the target user. The overall diagram of the proposed method is shown in Figure 1.

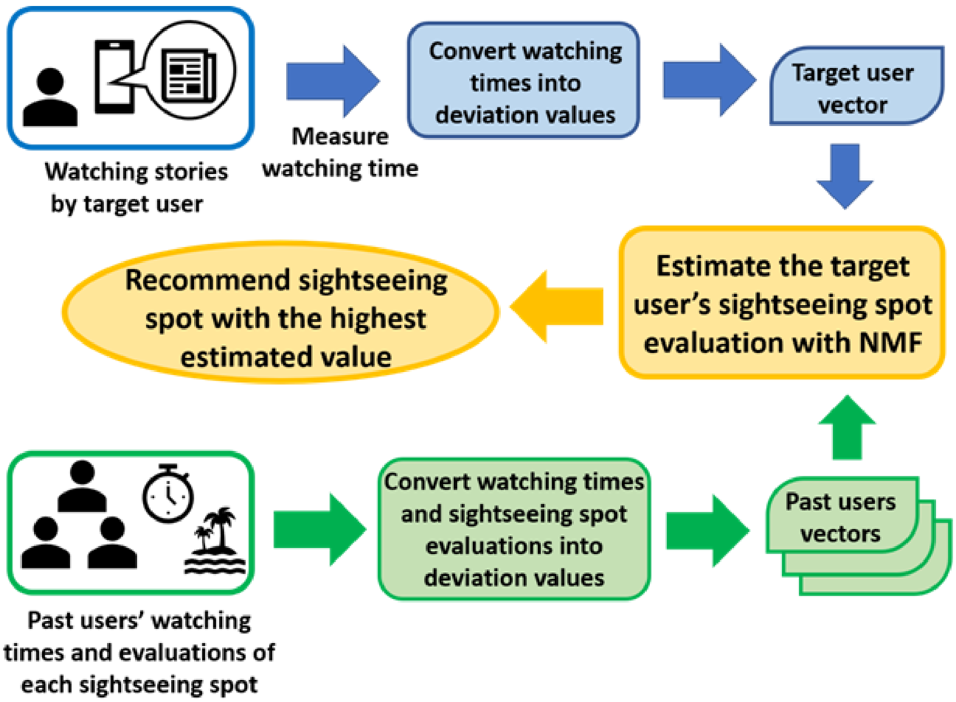


Figure 1. Method outline

In order to provide the target user with basic data for recommending sightseeing spot, past users' data is firstly generated as follows. Past users freely select the interesting stories from the many prepared stories of the sightseeing spots and watch them for a short time. After that, the past users actually visit the sightseeing spots and evaluate each sightseeing spot. The story watching times and evaluations are recorded as the past users' data. On recommendation to the target user, the target user selects the interesting stories from the prepared stories and also watches them freely for a short time. The watching time of each story watched is recorded. In the proposed method, non-negative matrix factorization (NMF) [12] is used to estimate the evaluations of the target user for each sightseeing spot from the past user's data and the target user's story watching times. The sightseeing spot with the highest estimated evaluation value is recommended as the best sightseeing spot for the target user. By using NMF, we can recommend the best sightseeing spot only by imposing watching the stories on the target user. In this research, we also consider whether NMF can be used to estimate the degree of the user's preference for sightseeing spots.

With the proposed method, the only task for target users to recommend sightseeing spot is to watch interesting stories. Target user do not wear special sensors or answer questionnaires. Therefore, by using this method, it is possible to recommend a best sightseeing spot without imposing a heavy load on the target user.

4.2. Story Watching by Target User

In the proposed method, the target user watches the stories in order to enjoy recommendation of the best sightseeing spot. In this research, the stories about each sightseeing spot are collected from the posts about the sightseeing spot in SNSs such as Instagram. Each story consists of text-format experiences in the sightseeing spot and photographs of the sightseeing spot written by various people. Because there are many people who posted stories on SNSs, there are many different stories about a same sightseeing spot. Since there are many different stories, it is possible to grasp sightseeing spots from various viewpoints. Therefore, it is possible to prevent a biased view.

In this research, we provide the target user with many different stories. The target user can freely watch the stories on the smartphone. Figure 2 shows the screenshot in the story watching on a Smartphone.

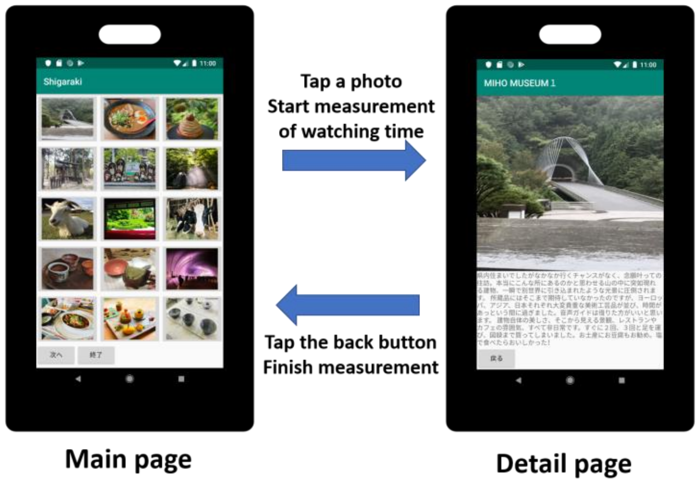


Figure 2. The screenshot at watching stories on a smartphone

A story is made up of a pair of text-format of experience and photograph. First, only multiple photos are displayed on the screen. Each photo is associated with a particular story. This state is the main page (see the left part in Figure 2). The target user can tap a photo of interest on the main page. On tap a specific photo on the main page, the screen moves to the story page associated with the photo. This page is the detail page (see the right part in Figure 2). On the detail page, both the enlarged photo of the tapped photo and the experience are displayed. On tap the back button on the detail page, the screen returns to the main page again.

The time from tapping a photo on the main page to tapping the back button on the corresponding detail page is the story watching time. On the main page, the photos of each story are displayed in small size and the target user cannot watch the experiences in each story, so it is not in a watching state.

The target user does not have to watch all stories. The target user can select and tap favorite ones from many stories. If he/she tapped once a story but it was a story he was not interested in, he can tap the back button immediately and select another story again on the main page. Therefore, the target user watches the stories that he/she is interested in for a long time, and immediately stops or does not watch the stories that he/she is not interested in. The target user can intuitively watch as many stories as they like. Therefore, this system can be used with the same feeling as if one normally enjoys surfing the internet on SNSs, and the load on the user is small.

4.3. Acquisition of Past User's Data

In this research, past users evaluate each sightseeing spot. This method uses the target user's story watching time and past user's data. A past user's data includes the scores of each sightseeing spot in addition to the story watching time. The past user scores two points, which are the evaluation of each sightseeing spot and the degree of each preference for the sightseeing spot. The evaluation of a sightseeing spot is expressed by a real number from 0 to 100. The closer to 0, the lower the evaluation, and the closer to 100, the higher the evaluation.

In this research, it is assumed that the target user feels seven preferences of "eating," "making," "playing," "seeing," "healing," "history," and "nature" for any sightseeing spot at individual degrees. The degree of each preference is evaluated on a scale of 5 from 1 to 5, with a degree closer to 1 being lower and a degree closer to 5 being higher. The stories given to the target user and the past users are the same. The past users select favorite stories from a large number of given stories and watch them.

4.4. Estimating Sightseeing Spot Evaluation by NMF

The proposed method uses the NMF to estimate the evaluation of each sightseeing spot from the target user's story watching time and past user's data.

NMF is an algorithm that decomposes a non-negative matrix X into two non-negative matrices W and H. At this time, the product of the decomposed matrices W and H is an approximation of the matrix X. NMF allows that some elements of X are missing (unknown values). The missing elements are replaced with 0 in advance. The unknown values are estimated when the matrix X is approximated by the product of the matrices W and H. NMF approximates matrix X with the product of matrices W and H. That is, the (i, j)-element of X is represented by the inner product of the i-th row vector of W and the j-th column vector of H. Each row vector of W and each column vector of H are used to compute multiple elements of X. NMF attempts to approximate all elements of X except the missing elements by the inner product of the row vectors of W and the column vectors of H. When the number of missing elements is small enough and the approximation of all non-missing elements is achieved, the inner product of the row vector of W and the column vector of H corresponding to each missing element can be calculated. NMF considers this inner product value to be an estimation of the missing value. That is, in NMF, missing values can be estimated by decomposing the matrix.

In the method, we consider a vector that summarizes the story watching times and the evaluations of sightseeing spots. The vector for the target user and the vector for each past user will be called the target user vector and the past user vector, respectively. The proposed method applies NMF to the matrix that combines the target user vector and the past user vector. An example of this matrix is shown in Figure 3.

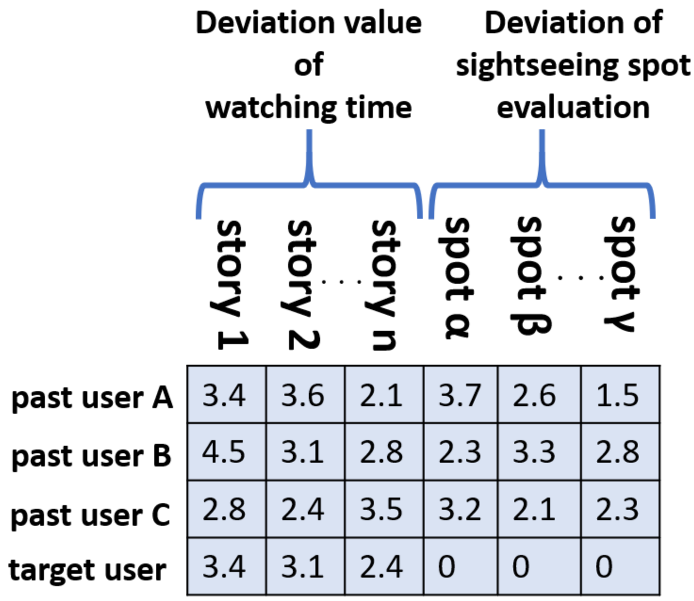


Figure 3. Example of matrix for estimating evaluation

The target user has never visited the sightseeing spots and his evaluation of the spots are missing. Therefore, the proposed method estimates these missing evaluations through NMF.

The matrix for NMF is generated from the target and past vectors based on the following idea. The row vectors of the matrix are the past and target user vectors. The past vector of each past user is a (n + m) row vector for n stories and m sightseeing spots. The watching time of i-th story is given as the i-th element of the past user vector. The evaluation of j-th sightseeing spot is given as the (n+j)-th element of the past user vector. The target user vector is generated in the similar manner.

However, NMF does not recommend treating values with different units such as watching time and evaluation as element values in one matrix. Moreover, in NMF, the accuracy of estimation increases when the variations in the values are similar. Evaluation of sightseeing spots is subjective. Among some sightseeing spots, some users evaluate them with a large variation while some other users evaluate them with a small variation. In order to improve the estimation accuracy, it is better to make the size of the evaluation variations uniform. The variation among watching times should be uniformed. Therefore, the variations in the watching times and in the evaluation of sightseeing spots are adjusted based on the deviation values within the corresponding user. In the proposed method, standardization is performed so that the evaluation value takes a value from 0 to 5 by taking the deviation value with the average being 2.5. The watching times are replaced in the similar manner.

In the matrix generated in the above manner, the elements corresponding to the evaluations of sightseeing spots in the target user vector is missing. Through NMF on this matrix, these missing elements can be estimated as a value represented by the deviation value.

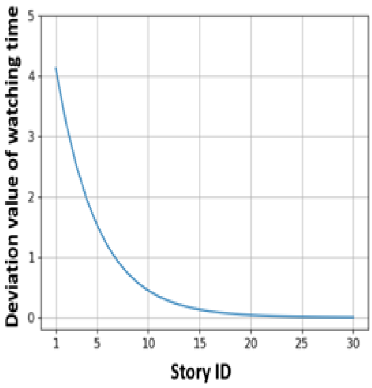
4.5. Watching Time for Unwatched Stories

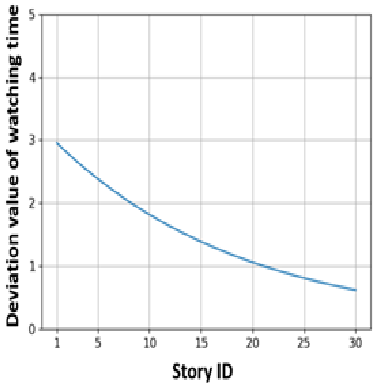
NMF can estimates missing values in the matrix. In the proposed method, the matrix for NMF includes missing elements as the missing data. A past or the target user do not watch all stories, where the corresponding watching times become missing. There are the cases where the users do not watch a story because they are not interested in it, and where they cannot watch it because of spending time for other stories even though they are interested in it. We consider, if they avoided watching a story because they are not interested, to set the corresponding watching times close to 0. We also consider, if they are interested but cannot watch, to set the corresponding watching times as a positive-valued watching time.

If NMF is used, it is possible to estimate these deviation values by setting the story watching time that the target user and the past users have not watched as a missing value. However, the purpose of the proposed method is to estimate the evaluation of the sightseeing spots scored by the target user. Estimating the deviation values of the watching times of stories that has not been watched does not fit for the purpose. Therefore, we calculate the deviation values of the story watching times that the target user and the past users did not watch are found by regressing within each of the user in advance. The deviation value of the watching time of a story that is not watched by a user is calculated from the decay curve expressed by the following equation.

$$y=S(1/a)^{n-1} \tag{1}$$

y is the deviation value of the watching time and is the objective variable of the regression. n is the explanatory variable of the regression equation, and it is the story ID given based on the length of watching time. In other words, IDs are assigned in the descending order of the length of the watching time. Obviously, the ID of the story that watched in the longest is 1, and the ID of the story that watched the second longest is 2. S is the deviation value of the story watching time that was watched in the longest. a is a regression coefficient and represents the degree of attenuation. The degree of attenuation a here is the degree to which the story watching time is reduced. The degree of attenuation varies from person to person. Examples of the decay curves are shown in Figure 4.





[a]User with largely different watching times [b] User with almost consistent watching times

Figure 4. Difference in decay curves depending on user

Figure 4 [a] is an example of a user with significant variations in the story watching times. Since the watching time varies significantly among the stories, the degree of attenuation large. On the other hand, Figure 4 [b] is an example for a user who does not have much variation in watching times for stories. Since the watching time does not vary so much, the degree of attenuation is small. The degree of attenuation of the watching time of each user is found by regressing a for each user. By substituting the total number of stories for n, the deviation value of the virtual watching time of the story that is not watched is obtained. Substituting the total number of stories for n gives the watching time as a very small positive value. If this value is set in the matrix shown in Figure 3 and applied to NMF, this is not considered as a missing value.

4.6. Estimating User's Preference for Sightseeing Spots

The proposed method also estimates the degree of each preference that the target user has for each sightseeing spot. As mentioned in Subsection 2.2, there are roughly 10 types of tourism types according to the purpose of tourism. The target user's purpose of tourism is always explicit, but variable. Therefore, it is necessary to recommend sightseeing spots according to the occasion. The preferences that are emphasized differ depending on the purpose of tourism. If we can grasp the degree of each preference that the target user has for sightseeing spots, we can recommend a sightseeing spot that suits the target user's purpose of tourism.

In this research, it is assumed that the target user has seven preferences of "eating," "making," "playing," "seeing," "healing," "history," and "nature" for any sightseeing spot. There are individual differences in the degree of each preference that target user has for sightseeing spots. Therefore, unlike Subsection 4.4, we consider the target user vector that combines the watching times and the degrees of preference for sightseeing spots. Then, the matrix that combines the target user vector and the past user vectors is decomposed by NMF. The degree of each preference that the target user has for the sightseeing spot is estimated. Figure 5 shows an example of the matrix that combines target user vector and past user vectors for estimating preferences.

	Deviation value of watching time			Preferences of sightseeing spot		
	story 1	story 2 · · ·	story n	eating	making · · ·	nature
past user A	3.4	3.6	2.1	4	5	3
past user B	4.5	3.1	2.8	4	3	2
past user C	2.8	2.4	3.5	5	3	2
target user	3.4	3.1	2.4	0	0	0

Figure 5. Example of matrix for estimating preferences

In this study, the story watching times of the target user and the past users are replaced by the deviation values while the degree of each preference are not replaced by a deviation value. This is because the past users have evaluated the degree of each preference they have for the sightseeing spot they visited in five levels. The evaluation of each sightseeing spot is scored on a 100-point scale while the degree of preference is evaluated on a scale of five. Thus, the variations in preference degree values are uniform. By using NMF, we can estimate the degree of each preference that the target user has for the sightseeing spot simply by watching the story. If we can estimate the degree of each preference that the target user has for the sightseeing spot, we can recommend the sightseeing spot that suits the target user's purpose of tourism.

5. EXPERIMENT

5.1. Outline of Experiment

We conducted an experiment to verify the usefulness of the proposed method. The following two were verified by the experiment.

- The accuracy of recommending sightseeing spot to target users
- The accuracy of estimating target users' preferences for sightseeing spots

In this experiment, we used three actual sightseeing spots: a museum, a restaurant and a pottery hall in Shigaraki, Shiga Prefecture in Japan. The subjects were nine men and two women in their twenties. The total number is 11. The 11 subjects have never visited the three sightseeing spots. We prepared 30 stories about Shigaraki. We got the stories from Instagram and personal blogs. Each story was chosen to have the same amount of text. This is to prevent a difference in watching time due to a difference in the amount of text.

The experiment was conducted according to the following procedure.

1. Each of the 11 subjects selected interesting stories from the prepared stories within the 3-minute time limit and freely watched the stories on his smartphone. Each story watching time was automatically recorded by the smartphone.
2. The 11 subjects visited the three sightseeing spots. Each of them gave an evaluation to each of the sightseeing spot out of 100. He/she also gave the degree of each of the seven preferences of each of the sightseeing spots in a 5-point scale.
3. We calculated the estimation accuracy of the proposed method through the leave-one-out cross-validation by using one of the subjects as the target user and the remaining 10 subjects as the past users. By applying NMF, we estimated the evaluations of the sightseeing spots scored by the target user. Only the story watching times were used as the target user data. The evaluations for the sightseeing spots were treated as missing values.
4. We calculated the correlation coefficient between the estimated evaluation values of the sightseeing spots and the actual evaluations of the sightseeing spot actually given by each of the target user.
5. We similarly estimated the degrees of the 7 preferences of the target user and calculated the correlation coefficient between the estimated degrees and the actual degrees given by the target user, as Steps 3 and 4.

5.2. Sightseeing Spot Recommendation Accuracy

In order to verify the recommendation accuracy of sightseeing spot using the proposed method, we investigated the correlation between the sightseeing spots evaluations actually scored by the subjects and the estimated sightseeing spots evaluations. In this study, the evaluation of each sightseeing spot actually scored by the subject is called the actual evaluation value, and the evaluation of the sightseeing spot estimated using NMF is called the estimated evaluation value. In this experiment, the 11 subjects evaluated the 3 sightseeing spots. There are 11 × 3 = 33 actual evaluation values and estimated evaluation values. We calculated the correlation coefficient and the rank correlation coefficient of these 33 actual evaluation values and estimated evaluation values. The rank correlation coefficient is the correlation coefficient obtained by converting each variable into ranks. In this study, we calculated the correlation coefficient by converting the actual evaluation values and estimated evaluation values of each user into ranks. In addition, a p-value was also obtained by performing a test for no correlation with the significance level 1%. The null hypothesis here is "no correlation." Table 1 shows the correlation coefficient and rank correlation coefficient between the actual evaluation value and the estimated evaluation value. Figure 6 shows the correlation diagram between the actual evaluation value and the estimated evaluation value. The horizontal axis of the figure is the actual evaluation value. The vertical axis is the estimated evaluation value. and the straight line is the regression line.

Table 1. Correlation coefficient and rank correlation coefficient between actual evaluation value and estimated evaluation value and their p-values

correlation coefficient	0.647
p-value of correlation coefficient	4.64×10^{-5}
rank correlation coefficient	0.820
p-value of rank correlation coefficient	5.07×10^{-9}

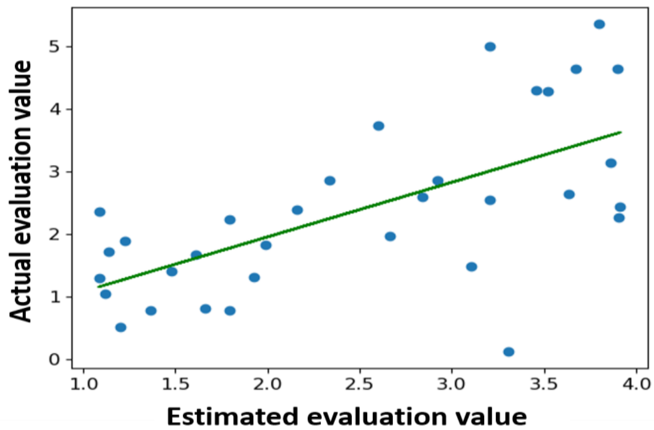


Figure 6. Correlation diagram of actual evaluation value and estimated evaluation value

From Table 1, the correlation coefficient was 0.647 and the rank correlation coefficient was 0.820, both of which show a significant strong positive correlation. The p-value is smaller than the significance level in all cases. In addition, we succeeded in estimating the evaluation ranking of 8 out of 11 subjects. It can be said that the recommendation accuracy is high because a strong positive correlation was found in the correlation between the actual evaluation value and the estimated evaluation value.

5.3. Accuracy of Estimating Tourists' Preferences for Sightseeing Spots

We verified the estimation accuracy of each preference that the target user has for each sightseeing spot. We investigated the correlation between the degree of each preference that the target user actually scored to each of the sightseeing spot and the degree of each estimated preference. In this study, the degree of each preference that the target user actually has for the sightseeing spot is called the actual preference value, and the degree of each preference for the sightseeing spot estimated using NMF is called the estimated preference value.

In this experiment, since the 11 subjects gave 7 preference values for each of the 3 sightseeing spots, there are $11 \times 7 \times 3 = 231$ actual preference values and estimated preference values. As described in Subsection 5.2, we calculated the correlation coefficient and rank correlation coefficient among these 231 actual preference values and estimated preference values. We confirmed the relationship between them. In addition, the p-value was also obtained by performing a test for no correlation with the significance level 1%. The null hypothesis is "no correlation." Table 2 shows the correlation coefficient and rank correlation coefficient between the actual preference values and the estimated preference values. Figure 7 shows the correlation diagram between the actual preference values and the estimated preference values. The horizontal axis of the figure is the actual preference value. The vertical axis is the estimated preference value. The straight line is the regression line.

Table 2. The correlation coefficient and rank correlation coefficient between the actual preference values and estimated preference values and their p-values

correlation coefficient	0.485
p-value of correlation coefficient	5.06×10^{-15}
rank correlation coefficient	0.576
p-value of rank correlation coefficient	8.99×10^{-22}

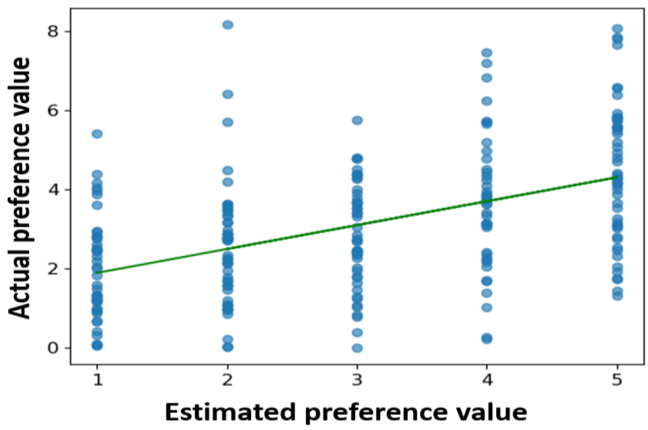


Figure 7. The correlation between actual preference values and estimated preference values

From Table 2, the correlation coefficient was 0.485 and the rank correlation coefficient was 0.576, both of which show a significant weak positive correlation. The p-value is smaller than the significance level in all cases. It was found that there is a generally positive correlation between the actual preference values and the estimated preference values. However, it cannot be said that there is a strong correlation in either result. Therefore, it can be said that the preference can be estimated to some extent, but the accuracy is lower than that of estimating the evaluation of sightseeing spots.

6. DISCUSSION

In this research, we proposed a method of recommending the best sightseeing spot to the target user by using the story watching time of the target user and the data of past users. In this study, we did not only recommend the best sightseeing spot but also estimated the degree of the tourists' preference for each sightseeing spot. Here, we consider these estimating accuracies.

6.1. Usefulness of Story Watching Time in Recommending Sightseeing Spots

Regarding the recommendation of the best sightseeing spot, a significant strong positive correlation was found between the actual evaluation value and the estimated evaluation value. In addition, the accuracy of recommending sightseeing spots can be said to be high because we succeeded in estimating the evaluation ranking of 8 out of 11 subjects. In the future, the following point can be considered to further improve the recommendation accuracy.

In the experiments in this paper, the time during which the subjects can watch the story was fixed at 3 minutes. Some users may find this 3 minutes long, while others may find it short. A subject who found 3 minutes long may have watched all the stories that he/she was interested in and then have watched the stories that he/she was not interested in until the 3 minutes have passed. A subject who found the 3 minutes short may have spent 3 minutes before watching all the stories of interest. Thus, it cannot be said that the story watching time reflects the user's interest entirely. It was necessary to allow the subjects to finish watching the stories at any time without limiting the watching time of the stories to 3 minutes. By not limiting the watching time, it is considered that the user's interest appears significantly in the watching time of each story.

We used only three sightseeing spots in the experiment: a museum, a restaurant, and a pottery hall. The correlation coefficient and rank correlation coefficient may take higher values due to the small number of sightseeing spots. In the future, it is necessary to increase the number of sightseeing spots and confirm the accuracy.

6.2. Estimating Preferences of Tourists for Sightseeing Spots

Regarding the estimation of the preference of the subjects to sightseeing spots, a weak positive correlation was found between the actual preference values and the estimated preference values. However, the estimation accuracy was lower than the recommendation of the best sightseeing spot. This is because the number of the sightseeing spots was three while it was necessary to estimate the seven values in terms of preference. The greater the number of values to be estimated, the harder it is to estimate all of them because the known information available for estimating is limited.

In the proposed method, NMF is used to estimate the user's preference for sightseeing spots. In NMF, the accuracy of estimation increases as the number of data increases. Therefore, it is considered that the accuracy of preference estimation becomes higher by increasing the number of past users' data.

In this research, assuming that the recommended sightseeing spot is Shigaraki, the seven preferences that target users have for each sightseeing spot are "eating," "making," "playing," "seeing," "healing," "history," and "nature." This is based on the website of the Shigaraki Tourism Association [15] and is considered to be specialized in sightseeing spots of Shigaraki. In the future, it is necessary to verify whether these seven preferences are appropriate for the target user for each sightseeing spot in Shigaraki.

Also, all p-values were quite small. This is probably because the number of subjects was 11. The number of subjects should be increased in the future. In addition, the ages of the subjects in this study were all in their 20s, which was quite biased. In the future, subjects of various ages should be recruited.

7. CONCLUSION

In this research, we proposed a method of recommending the best sightseeing spot through watching videos of stories of sightseeing spots. This method recommends best sightseeing spot based on the similarity of the behavior regarding the watching of stories of tourist who are going to visit the sightseeing spot and tourists who have visited the sightseeing spot in the past. Moreover, this method does not only recommend a sightseeing spot but also estimates the degree of tourists' preferences for sightseeing spots.

As the experimental result of verifying the usefulness of this method, it was suggested that the story watching time is useful for recommending the best sightseeing spot. It was also suggested that it is possible to estimate the preferences to some extent, although it is inferior to the recommendation of the best sightseeing spot.

In the future, we will increase the number of recommended sightseeing spots to confirm the recommendation accuracy. We will also increase the amount of data of past users to improve the accuracy of preference estimation.

REFERENCES

[1] International Tourism Highlights, 2019 Edition. World Tourism Organization, Madrid, Spain, 2019.

[2] Ricardo Anacleto, Lino Figueiredo, Ana Almeida, and Paulo Novais Mobile application to provide personalized sightseeing tours. Journal of Network and Computer Applications, Vol. 41, pp. 56–64, 2014.

[3] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Where shall we go today? planning touristic tours with tripbuilder. In Proceedings of the 22nd ACM international conference on Information and Knowledge Management, pp.757–762, 2013.

[4] Igo Brilhante, Jose Antonio Macedo, Franco Maria Nardini, Raffaele Perego, and Chiara Renso. Tripbuilder: A tool for recommending sightseeing tours. In European Conference on Information Retrieval, pp. 771–774. Springer, 2014.

[5] Fabio Clarizia, Saverio Lemma, Marco Lombardi, and Francesco Pascale. An ontological digital storytelling to enrich tourist destinations and attractions with a mobile tailored story. In International Conference on Green, Pervasive, and Cloud Computing, pp. 567–581. Springer, 2017.

[6] J van Harssel. Tourism: an exploration. No. Ed. 2. National Publishers of the Black Hills, Inc., 1986.

[7] Kuo-Lun Hsiao, Hsi-Peng Lu, and Wan-Chin Lan. The influence of the components of storytelling blogs on readers' travel intentions. Internet Research, 2013.

[8] Lloyd E Hudman, Donald E Hawkins. Tourism in contemporary society: an introductory text. Prentice-Hall, Inc., 1989.

[9] Mill, Robert Christie. Tourism: The international business. Prentice Hall Englewood Cliffs, NJ, 1990.

[10] Hiromitsu Shimakawa, Momoko Kato. Recommendation of tour route from tourist motivation improving serendipity occurrence. International Journal of Latest Research in Engineering and Technology, Vol. 3, No. 2, pp. 26–36, 2017.

[11] Rebecca Pera. Empowering the new traveller: storytelling as a cocreative behaviour in tourism. it Current Issues in Tourism, Vol. 20, No. 4, pp. 331–338, 2017.

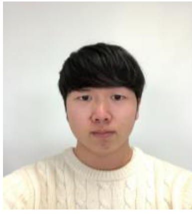
[12] Toby Segaran. Programming collective intelligence: building smart web 2.0 applications. "O'Reilly Media, Inc.", 2007.

[13] Hyoseok Yoon, Yu Zheng, Xing Xie, and Woontack Woo. Smart itinerary recommendation based on user-generated gps trajectories. In International Conference on Ubiquitous Intelligence and Computing, pp. 19–34. Springer, 2010.

[14] Hyoseok Yoon, Yu Zheng, Xing Xie, and Woontack Woo. Social itinerary recommendation from user-generated digital trails. Personal and Ubiquitous Computing, Vol. 16, No. 5, pp.469–484, 2012.

[15] Shigaraki Tourism Association, Hottosuru Shigaraki, https://www.e-shigaraki.org/

AUTHORS



Motoki Seguchi received B.E degrees from Ritsumeikan University in 2020. He is currently pursuing the M.E degree with Ritsumeikan University. He is currently studying data engineering.

Dr. Fumiko Harada received B.E., M.E, and Ph.D degrees from Osaka University in 2003, 2004, and 2007, respectively. She joined Ritsumeikan University as an assistant professor in Ritsumeikan University in 2007, and is currently a counselor of Connectdot co.ltd. She engages in the research on real-time systems and data engineering. She is a member of IEEE.

Prof. Hiromitsu Shimakawa received Ph.D degree from Kyoto Univ. in 1999. Since 2002, He has worked in Ritsumeikan Univ. as a professor. His research interests include data engineering, usability, and integration of psychology with IT. He is a member of IEEE and ACM.

© 2020 By AIRCC Publishing Corporation. This article is published under the Creative Commons Attribution (CC BY) license