A Japanese Tourism Recommender System with Automatic Generation of Seasonal Feature Vectors

Guan-Shen Fang
Department of Information Engineering
Hiroshima University
Hiroshima, Japan

Sayaka Kamei
Department of Information Engineering
Hiroshima University
Hiroshima, Japan

Satoshi Fujita
Department of Information Engineering
Hiroshima University
Hiroshima, Japan

Abstract—Tourism recommender systems have been widely used in our daily life to recommend tourist spots to users meeting their preference. In this paper, we propose a content-based tourism recommender system considering travel season of users. In order to characterize seasonal variable features of spots, the proposed system generates seasonal feature vectors in three steps: 1) to identify the vocabulary concerned through Wikipedia; 2) to identify the trend over all spots through Twitter for each season; and 3) to highlight the weight of words contained in each identified trend. In the decision of recommendation, it does not only match the user profile with features of spots but also takes user's travel season into account. The effectiveness of the proposed system is evaluated by a series of experiments, *i.e.* computer simulation and questionnaire evaluation. The result indicates that: 1) those vectors certainly reflect the similarity of spots for designated time period, and 2) with using such vectors of spots, the system successfully realized a tourism seasonal recommendation.

Keywords—Tourism recommender system; seasonal feature vector; Wikipedia; Twitter

I. Introduction

In recent years, tourism recommender system is widely used to support users' choices of tourist spots. Unlike commodity such as books or movies, many of the features of Japanese spots change following season. A typical example is the famous tourist spot Niseko in Hokkaido, Japan, whose red leaves attract visitors in autumn. In winter, the snow makes its scene totally different so that visitors may enjoy skiing. Therefore, the decision of spots made for the user must not only fit his interest, but also realize the seasonal fashion.

In this paper, we focus on content-based seasonal recommendation. Although content-based recommendation has been applied successfully in the tourism domain [2], [3], [4], [5], they seldom produce recommendation considering season.

We propose a content-based seasonal tourism recommender system which fits the designated season of travel. For example, for a user who likes Matsushima island because of cherry-blossom viewing and wishes to travel in spring, the system can recommend other spots with the attraction of cherry-blossom viewing. However, such spots may not be the recommendations of another user who wishes to travel in autumn even if is fond of Matsushima's sea and coast. In order to characterize the

Fig. 1. A part of the Wikipedia article about Ritsurin Park. A detailed description of the main features is included.

dynamically changed seasonal features, the proposed system generates seasonal feature vector for each spot for a given season. Its generation consists three steps. Firstly, it identifies the vocabulary concerned through Wikipedia[1] document about the spot. The reason for using Wikipedia is that each tourist spot has its unique document which introduces its detail features (see Fig. 1 which shows the description about the spot "Ritsurin Garden" in Kagawa prefecture including seasonal features for each season). Secondly, it identifies the trend (*i.e.*, seasonal variance of features for a designated period) over all spots in Japan through Twitter[2]. Finally, for the words (*i.e.*, the features) in the vocabulary, it highlights the ones contained in the identified trend which corresponds to the given season. With these vectors of spots, the proposed system match them with user's profile which presents his/her preference to decide the recommendations for the travel period of user.

As the implementation, the proposed system gathered 6,057 Wikipedia documents to cover almost all sightseeing spots in Japan. On the other hand, it also collected more than 500 thousands tweets that are published during 7 months. In the experiments we conducted a series of experiments including both a computer simulation and a questionnaire evaluation.

The contribution of current paper is as follows:

- For a designated season, it provides a method to generate seasonal feature vector for sightseeing spot as the characterization of its features. The experimental results demonstrate that the property of those vectors certainly reflect the similarity of spots.

- The proposed recommender system provides the user a

---

[1]http://ja.wikipedia.org
[2]https://www.twitter.com

recommendation of spots with awareness of the travel season. The experiment of questionnaire shows that the seasonal recommendations have higher precision of user's actual choices than the one without applying seasonal feature vectors.

The remainder of this paper is organized as: Section II overviews related works. Section III describes the detail of the proposed system including the generation of seasonal feature vectors and recommendation process as well. Section IV represents the implementation of the prototype system. Section V represents the method of evaluation and shows its results. Finally, a conclusion is given and future works are discussed in Section VI.

II. RELATED WORK

A. Content-based Tourism Recommendation

Content-based tourism recommendation systems try to recommend spots similar to those users have liked in the past (*i.e.*, history) [6]. From the history, user's profile is built to represent his preference. On the other hand, the features of spots are characterized in order to match with user's profile to decide the recommendation. Some of existing researches aim to provide the user an appropriate tour plan to meet his/her constraints, such as time or cost [7], [8], [3]. The features of spots are given from experts of tourism, which simply includes available time, normal visiting time and geographical information, etc.. Therefore, the recommendation of the tour plan turns to an integer programming problem or traveling salesman problem to approximate a combination of spots with minimization of the travel path or time wasted in movement. Győrödi *et al.* [9] proposes a spot recommendation with a mobile application. In order to determine user's interest and features of spot, they use tags such as food, music etc., which can be established by users and assigned to a specific spot. The recommendation is produced by matching such tags of the given user and spots.

In existing content-based researches much efforts are made to provide tourist spots or plan to meet user's needs. However, the travel season which is an essential factor in decision of spots is seldom taken into account.

By utilizing the proposed method not only in content-based recommendations but also in some hybrid approaches [10], [11], [12], a seasonal recommendation can be easily realized.

B. Tourism Recommendation using Wikipedia

In many recent researches of tourism recommendation, Wikipedia is integrated as an external source in identification of spots. It is effective at reducing the cost of manually construction or maintenance of spots' information. A common idea is to take advantage of geographic information included in Wikipedia documents about spots to filter users' geo-tagged photos (e.g., photos in Flickr) and extract their visiting trajectories of spots [13], [14], [15]. Techniques such as T-pattern tree [16] are exploited to mine the traveling patterns potentially contained in extracted trajectories. Additionally, in Wikipedia since categories are assigned to each tourist spot, some of researches further transform user's trajectories into sequences of categories to represent user's preference [14], [15].
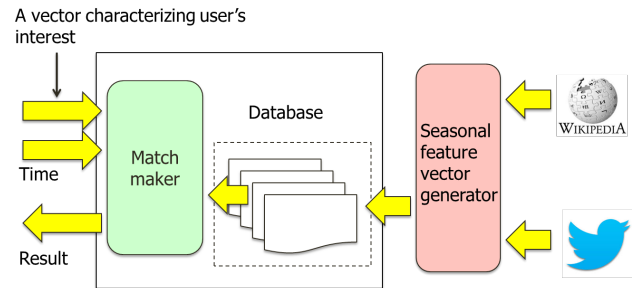
Fig. 2. Components in proposed tourism recommender system.

Although in many existed researches Wikipedia is used to combine with SNS to improve the performance of recommendations, few researches take advantage of the content of textual article in Wikipedia, even the detailed description of both permanent and seasonal features is contained.

C. Identification and Analysis of Tweets for Tourism

Recently, Twitter has been paid much attention as a source of data mining and characterization of spots for tourism informatics. In order to detect tourism related tweets which are posted at specific spots, Shimada *et al.* [17] applies a Support Vector Machine(SVM) to their gathered tweets. Their idea is that the target tweets are similar on their textual content. With the aid of geo-tag, Oku *et al.* [18] proposes another SVM-based method of detection of tweets relevant to tourism, and extract temporal features of spots from them. They regard tweets issued within a week as a single document and obtain a temporal feature vector for each week by calculating the TF-IDF weight of keywords contained in such documents. However, it does not generate vectors to cover a sufficient number of spots because it is solely based on tweets. Similarly, Menchavez *et al.* [19] focus on the identification of tourism related tweets and a Naive Bayes based sentiment analysis to mine the opinions when tourists visit spots in Philippines. Furthermore, such mined opinions are classified into positive and negative polarity and presented at the geographical map as references for tourists. Similar study is done by Claster *et al.* [20] for Thailand, in which the sentiment analysis of tweets is applied in time-series. Although the problem definitions of previously mentioned researches are different with the current paper, the objectives are similar which aim to discover useful information of tourism via Twitter.

In such researches, many efforts have made to reduce the noise in tweets. However, due to the irregularity of tweets, many meaningless words like punctuation or prefix are always extracted and significantly influence the accuracy of the analysis. In addition, those techniques based on machine learning are sometimes difficult to conduct for minor spots having few related tweets. Since in this paper the proposed system uses Wikipedia as the corpus combined with Twitter, it is free of the influence from such noise. Furthermore, for the minor spots which are seldom tweeted Wikipedia can cover their features and avoid the failure of their recommendation.

III. SEASONAL RECOMMENDATION OF TOURIST SPOT

In this section, we represent the proposed recommendation system in detail. Fig. 2 shows the architecture, which consists of two processes: 1) to generate seasonal feature vectors for each spot; 2) to identify user's preference as profile, and match it with such vectors of spots to produce recommendation. Following subsections detail each part.

A. Generation of Seasonal Feature Vectors

Firstly, the time axis is assumed to be separated into several ranges so that the features of spots are regarded to be invariant, as in the year end season, the season of cherry-blossom viewing or the bathing season. Each range is called a **season**. The proposed system generates one seasonal feature vector (SFV, for short) of each spot for each season. SFV is calculated by extending the basic feature vector (BFV, for short), in such a way that it reflects the trend of words in each season. More concretely, BFV is a vector of TF-IDF weights (defined bellow) and SFV is its extension.

Let O be the set of spots and d_i be the Wikipedia document about spot $o_i \in O$. Generally d_i is a summarization of the entire information of o_i. Therefore, the reader should note that d_i is the union of statements on spot o_i relevant to various seasons. In other words, in order to generate SFV for each season, the system needs to distinguish word sets relevant to each season in document d_i. Let W_i be the set of words included in document d_i and $W = \bigcup_i W_i$ (*i.e.*, W is the set of words included in Wikipedia documents about O). Then, the term frequency (TF, for short) weight of word w_j in document d_i is defined as

$$TF_{i,w_j} = \frac{n_{i,w_j}}{\sum_{w \in W} n_{i,w}}$$

and the inverse document frequency (IDF, for short) weight of word w_j over $|O|$ documents is defined as

$$IDF_{w_j} = \log\left(\frac{|O|}{m_j}\right)$$

Where, n_{i,w_j} is the number of occurrences of w_j in d_i and $m_j(\leq |O|)$ is the number of documents containing w_j. With these notions, the BFV $\vec{v}_i^b$ of spot o_i is defined as:

$$\vec{v}_i^b = \{(w_j, TF_{i,w_j} \times IDF_{w_j}) \mid w_j \in W_i\}. \quad (1)$$

The words which are frequently mentioned in d_i and seldom contained in other documents would have high weights in BFV.

For a given season, the key idea is to extend the definition of the TF weight in (1) by considering the trend of words. Let t_k be a collection of tweets issued in season s_k. By considering t_k as a single document, the TF weight of word w_j in season s_k is defined as follows:

$$TF'_{k,w_j} = \frac{n'_{k,w_j}}{\sum_{w \in W} n'_{k,w}} \quad (2)$$

Where, n'_{k,w_j} is the number of occurrences of word w_j in t_k. Because W is the set of words contained in Wikipedia documents about O, *the proposed system omits words in tweets which do not appear in any Wikipedia document*. With the above notions, SFV $\vec{v}_{i,k}^s$ of spot o_i for season s_k is defined as

$$
\begin{aligned}
\vec{v}_{i,k}^s = \ & \{(w_j, ((1-\alpha)TF_{i,w_j} + \alpha TF'_{k,w_j}) \times IDF_{w_j}) \\
& \mid w_j \in W_i\}
\end{aligned}
\quad (3)
$$

Where, $0 \leq \alpha \leq 1$ is an appropriate parameter. Note that for o_i, only for the word $w_j \in W_i$ it has TF_{i,w_j} to be non-zero. If word $w_j \in W_i$ has not tweeted in specific season s_k, $TF'_{k,w_j} = 0$; otherwise $TF'_{k,w_j} > 0$, then we say that w_j is highlighted in s_k.

B. Identification of User's Preference and Recommendation Process

Although an analysis of user's history of tweets would help us to extract his/her preference on the features of sightseeing spots, it may fail for the users who even do not have Twitter accounts or seldom tweet about travel. In order to fit such users, the proposed system extracts user's preference in an explicit way that it directly asks the user for a history of travel. In other words, the user answers two easy questions when he/she begins to use the system: 1) the season that he/she wishes to travel; 2) the most favorite spot that he/she has visited during assigned season until now. Assume that user u chooses tourist spot $o_{i'}$ and the period of season $s_{k'}$. His/her profile which presents preference is defined by the SFV $\vec{v}_{i',k'}^s$ as $\vec{U}_{k'}$. Although the user profiling is simple, it effectively characterizes user's seasonal preference and is with various benefits: first, it does not suffer the cold start problem; second, such questions are easy to answer and time-saving.

With the constructed user profile, the proposed system matches it with SFVs of spots to decide recommendations for season $s_{k'}$. To quantify the correspondence of $o_{i'}$ and a given spot $o_l(o_l \neq o_{i'})$ in $s_{k'}$, the system calculates their cosine similarity as follows:

$$Sim_{i',l} = \frac{\vec{U}_{k'} \cdot \vec{v}_{l,k'}^s}{|\vec{U}_{k'}||\vec{v}_{l,k'}^s|}, \quad for \ each \ o_l \neq o_{i'}, o_l \in O.$$

The spots with Top-t similarities are the recommendations to user u in $s_{k'}$ as a ranking. Note that the recommendations vary for different seasons designated.

IV. IMPLEMENTATION OF A PROTOTYPE SYSTEM

A. Datasets Description

Since the objective of recommendation is entire tourist spots in Japan, for this prototype system, we focus on 6,057 spots given in the category of "tourist spots in Japan" in Wikipedia, and download the Japanese document for each of them from the Wikipedia server. The prototype system uses only nouns as words in each document d_i. The set of words W_i in d_i is obtained by conducting the morphological analysis using MeCab [3] with the default IPA dictionary. From all collected Wikipedia documents, 608,390 words are extracted overall, in average 100.5 words for a document. The relationship of the count of spots and the size of words which are extracted from
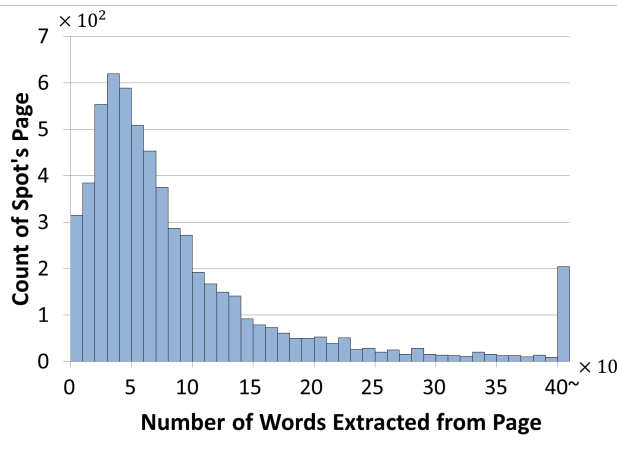
---

[3]http://mecab.sourceforge.net/

Fig. 3. The counts of spots with various sizes of words extracted from their documents.

Fig. 4. Diameter of spot in the vector space.

TABLE I. THE NUMBER OF MINOR SFVS FOR EACH s_k WHEN $\ell = 3$.

$\alpha < 1.0$	$\alpha = 1.0$						
	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
61	220	199	205	175	274	601	191

their documents is shown in Fig. 3. It represents that most spots are introduced in detail in their Wikipedia documents.

A set of tweets relevant to tourism is gained from Twitter using Twitter Streaming API. More concretely, 50 million Japanese tweets issued from September 2013 to March 2014 are acquired. For each of the tweet, its textual content is matched with the names of collected spots. As a result, about 500 thousands tweets containing at least one name of 6,057 spots are regarded as tweets relevant to the tourism and extracted as a part of dataset. Although it may contain tweets which are not relevant to tourism and may miss tweets relevant to the tourism, we didn't evaluate the precision of such a naive extraction since it is the out of scope of this paper. Let T be the resulting set of tweets.

B. Parameter Assignments

Considering seasons always last more than one month with different periods of time, assume that one year is divided into 12 disjoint seasons of (almost) equal length in the way that the first season is from January 1st to January 31st, the second season is from February 1st to February 28th (or 29th), and so on. More precisely, seven documents, which represent the trend of each season, is derived from T because collected tweets are for seven months. For a given season and its corresponding document in T, the prototype system generates one SFV for each spot. As a result, in all spots' SFVs 170,978 words are highlighted by Tweets, 28.3 words for one spot's SFVs in average.

Finally, another task is to identify appropriate value for α in (3). The value α is assumed more than 0 without loss of generality. Recall that W_i is the word set contained in the Wikipedia document about spot o_i. Relatively, let W'_k be the word set contained in tweets in season s_k. When $\alpha < 1$, the L^0-norm of SFV $\vec{v}^s_{i,k}$ coincides with $|W_i|$ which is independent
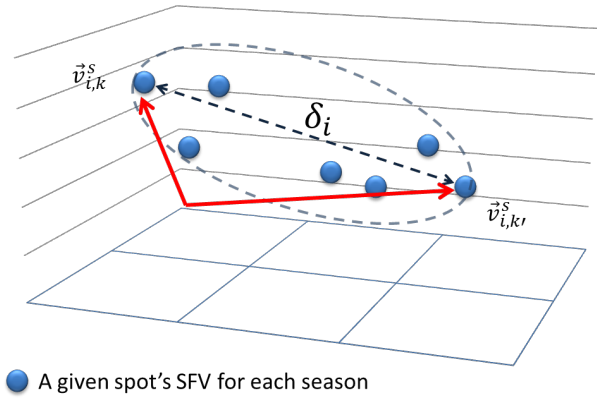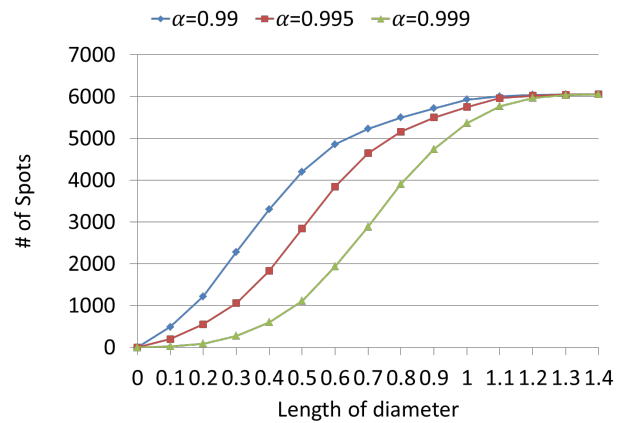
Fig. 5. Distributions of the diameters of spots with $\alpha = 0.99$, 0.995 and 0.999.

of α and k, but when $\alpha = 1$, it coincides with $|W_i \cap W'_k|$ which varies depending on k. It may cause failure of recommendation for some spots which are lack of vocabulary in their Wikipedia documents and have been seldom tweeted. Such spots are defined as **minor** if the L^0-norm of its SFV is smaller than or equal to ℓ. Table I shows the number of minor spots for given α and k. Although the number of minor ones is 61 when $\alpha < 1$, it exceeds 170 by increasing α to 1. Thus, in the following, the prototype system will restrict our attention to the case of $\alpha < 1$.

Next, consider the variance of SFVs in various α to decide its assignment. Let Ω be the vector space spanned by all SFVs (of all spots). In the following, each vector is normalized by the length in the L^2-norm to have an unit length in Ω. Therefore, each spot o_i is mapped to a point by SFV $\vec{v}^s_{i,k}$ in Ω for each season s_k. This implies that the "intensity" of the variance of SFVs is characterized by

$$\delta_i = \max_{k \neq k'}\{|\vec{v}^s_{i,k} - \vec{v}^s_{i,k'}|\} \tag{4}$$

Where, $|\cdot|$ denotes the L^2-norm. See Fig. 4 for the illustration. δ_i is affected by α and called the **diameter** of o_i hereafter.

Fig. 5 illustrates the cumulative distributions of δ for $\alpha = 0.99$, 0.995 and 0.999, where the horizontal axis is the length of the diameter and the vertical axis is the accumulative size of the spots having diameter less or equal to a specific value. It indicates that the diameter follow Gaussian distribution and its mean and variance of diameter certainly increase as α increases. In general, a large δ implies that for corresponding spot its seasonal features are well highlighted in SFVs.

Additionally, for each of the word sets $W_i \setminus W_k'$ and $W_i \cap W_k'$, we observe its average weight of words in SFVs. When $\alpha = 0.995$, a comparison is made that the two averages are both at range of 2.3×10^{-4}. As α increased to 0.999 the words in $W_i \setminus W_k'$ are weighted as one-fifth as the ones in $W_i \cap W_k'$ overall. It implies that although in the latter case the seasonal features are well highlighted, static features which do not relate with the seasons are weakened significantly and with failures of characterization. On the other hand, it is observed that the average distance of a spot's BFV to its nearest neighbor is almost 1.2, which is nearly twice of the average of all spots' δ in the case of $\alpha = 0.995$(almost 0.55). Therefore, α is fixed to 0.995.

V. EVALUATION

In this section, the effectiveness of the proposed system is evaluated with respect to the following two aspects: 1) whether the proposed SFV certainly extracts and characterizes seasonal features from Wikipedia and Twitter; and 2) whether the proposed system effectively provides seasonal tourist spot's recommendation.

A. Variance of SFVs

1) Evaluation Methodology: In this section, rather than a direct observation of the difference of SFVs for a given spot, the evaluation of time transition of the similarity of spots is conducted. They are obtained by applying the K-means method [21] to SFVs of all spots. More concretely, if the spots contained in a cluster in season s_k are separated into several clusters in other seasons, those spots are given similar SFVs for s_k and the set of words characterizing the cluster should represent the feature of those spots for s_k. Considering the size of spots is over 6,000, the value of K is setted to 70 in the process of clustering. This evaluation examines the mean of each resulting clusters and focus on several typical ones for the convenience of presentation.

2) Result: From the resulting clusters, four typical clusters, say C_r, C_i, C_s and C_c, are identified. Their details are summarized in Table II. Note that each of these four clusters is defined only for a specific season. Since *red leaves* and *cherry-blossom* have higher popularities than *illumination* and *snow* in Japan, the corresponding ones also have larger sizes than the others.

Results on the time transition of the similarity of spots are summarized in Fig. 6. The left-most figure of the first line in Fig. 6 shows the result of cluster C_r and the other three figures concern with clusters C_i, C_s and C_c, respectively. In November, all spots in C_r form a distinct cluster, but in other seasons, they separate into different ones. It indicates that the common features of those spots are *highlighted* in November, although they have various features in other seasons. Similar

TABLE II. DETAILS OF FOUR TYPICAL CLUSTERS

	relevant features	season	# of spots
C_r	red leaves, waterfall	November	109
C_i	illumination,cafe	January	36
C_s	snow,event	January	35
C_c	cherry-blossom,park	March	61

TABLE III. THE VALUE OF $\xi(30, k)$ WITH $k \in \{$ SEP, OCT, NOV, DEC, JAN, FEB, MAR $\}$.

	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
C_r	5	5	**5**	5	5	4	4
C_i	−8	4	4	4	**4**	4	4
C_s	6	4	4	5	**8**	3	3
C_c	11	10	11	11	12	13	**13**

phenomenon can also be found in other clusters. On the other hand, several spots in C_c are also confirmed to remain in the same cluster through all seasons. It indicates that SFVs of those spots are close with each other in vector space Ω regardless of the transition of seasons.

B. Impact on Recommendation

1) Evaluation Methodology: In this subsection, the performance of the proposed seasonal tourism recommender system is evaluated with simulated users. Although a questionnaire evaluation is also conducted in the next subsection, here we aim to compare and observe the difference between recommendations which are generated with considering season (*i.e.*, SFV) and without season (*i.e.*, BFV) in detail. The evaluation focuses on the aforementioned clusters C_r, C_i, C_s and C_c, and regards the mean of SFVs contained in each cluster as the preference of users[4]. In other words, there are four users who are fond of *red leaves*, *illumination*, *snow* and *cherry-blossom*, with the spots in clusters C_r, C_i, C_s and C_c as the answers respectively. The performance as the proposed system is evaluated by analyzing the Top-t spots' recommendation to the designated points for each season k. Such a subset of spots is denoted as Q_t^k hereafter. As comparison, according to the cosine similarity of the corresponding BFVs to the designated points, the Top-t spots (denoted as P_t) are also calculated.

For the mean of a given cluster C, the goodness of a subset X concerned is measured by $|C \cap X|$. Thus, the advantage of using SFV instead of ordinary BFV can be measured by calculating

$$\xi(t, k) \stackrel{\text{def}}{=} |Q_t^k \cap C| - |P_t \cap C|, \quad (5)$$

which depends on the value of parameter t and the selection of season k.

2) Result: Table III summarizes the results for $t = 30$, where the emphasized numbers designate the seasons in which the corresponding clusters are defined (e.g., cluster C_r is defined for November). The result implies that by using SFVs, the proposed system can recommend more spots to fit simulated users' preferences and the effect is maximized when the designated season coincides with the one defining the cluster.

---

[4]A typical scenario assumed in this paper is that the user designates an interested spot as the preference with a hoped travel season and the system returns a set of spots as recommendation.
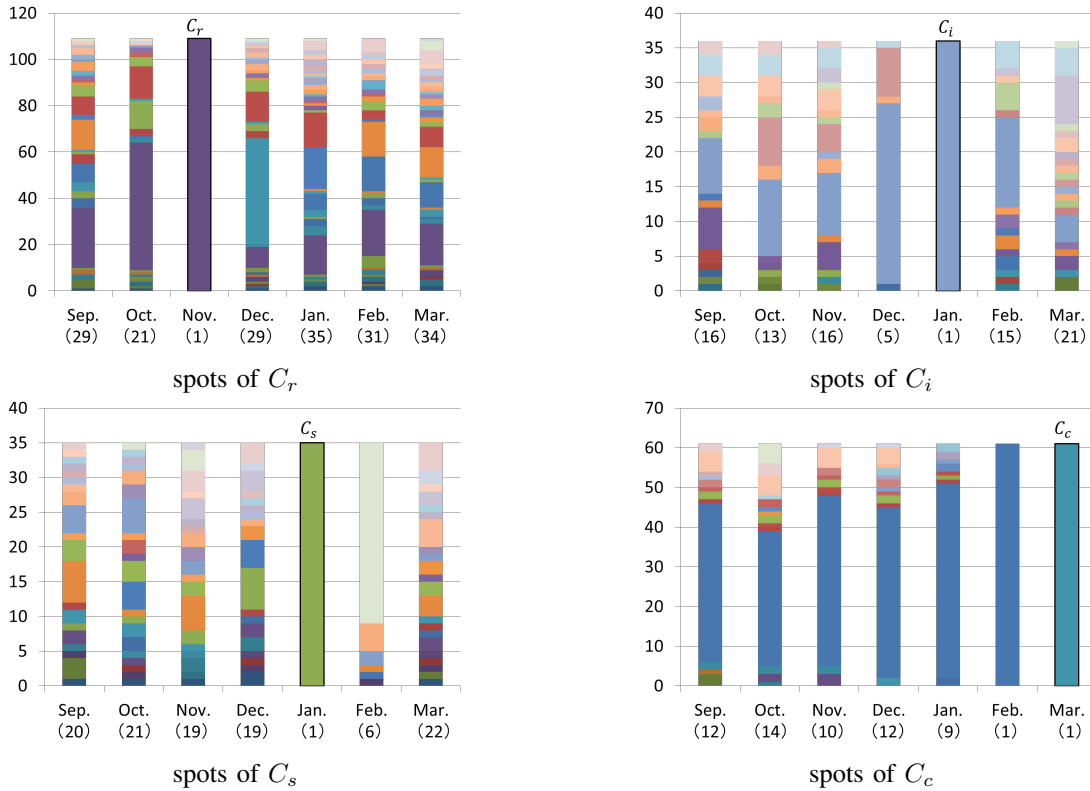
Fig. 6. The distributions of clusters for the spots that are included in C_r, C_i, C_s and C_c. The bars with various colors represent different clusters and their lengths depend on the sizes of focused spots. For a given cluster and a season, the number of clusters that focused spots separate into is also given in parentheses at the bottom.

TABLE IV. THE VALUE OF $\xi(t,k)$, WHERE THE VALUE OF t IS FIXED TO BE EQUAL TO THE CORRESPONDING CLUSTER SIZE

	Sep.	Oct.	Nov.	Dec.	Jan.	Feb.	Mar.
C_r	6	8	**7**	8	9	−4	2
C_i	−15	−5	0	0	**0**	−3	0
C_s	7	6	4	5	**9**	4	4
C_c	8	8	11	10	8	17	**17**

Fig. 7. Precision and recall of Q'_t and P'_t following t.

Recall that the value of $\xi(t,k)$ depends on parameter t. Table IV summarizes the results for each cluster, where the value of t is fixed to be equal to the cluster's size, e.g., let $t = 109$ for cluster C_r. Comparing with Table III, in each row a larger gap of $\xi(t,k)$ is observed for each season. It indicated that there are various $Q^k_{|C|}$ for given cluster C. In other words, if the designated season is not relevant with given C, fewer spots that contained in C will be recommended.

C. Questionnaire Evaluation

1) Evaluation Methodology: Finally, a two-steps' seasonal tourism questionnaire is conducted to evaluate whether the proposed system can provide a seasonal recommendation of spots in an actual case. Recall that the proposed system extracts user's preference from his visited favorite tourist spot in Section III-B. Therefore, as the first step of the questionnaire, participant selects the spot and the season (*i.e.* month) $s_{k'}$ that he/she wishes to travel. According to his/her selections, system generates a list of recommendations for $s_{k'}$ denoted as $Q'^{k'}_t$. For comparison, another recommendation list P'_t using BFVs
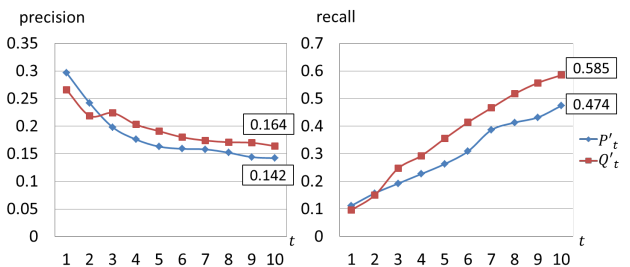
instead of SFVs of spots are generated. Here t denotes the number of spots that are included in the recommendations, *i.e.* the lengths of $Q'^{k'}_t$ and P'_t. $Q'^{k'}_t$ and P'_t are randomly combined into one list of recommendations, as $Q'^{k'}_t \cup P'_t$. In the second step, from $Q'^{k'}_t \cup P'_t$ the participant chooses at most 5 spots that he/she wishes to visit in $s_{k'}$. Since in following we focus on entire participants and their experimental results, the superscript k' in $Q'^{k'}_t$ is omitted for convenience.

As quantification, this evaluation calculates average precision and recall of all participants' choices in Q'_t and P'_t as follows:

$$precision = E(\frac{h_t}{t})$$

$$recall = E(\frac{h_t}{H})$$

TABLE V.   THE NUMBERS OF TRIALS WITH EACH $s_{k'}$ HAVING BEEN CHOSEN. FOR EXAMPLE, 10 QUESTIONNAIRES ARE SUBMITTED WITH $s_{k'} = Dec.$.

month	Jan.	Feb.	Mar.	Sep.	Oct.	Nov.	Dec.
# of trials	6	6	12	11	14	5	10

TABLE VI.   THE SIZES OF SPOTS CHOSEN FROM P'_t AND Q'_t BY ALL PARTICIPANTS

	P'_t	Q'_t	$Q'_t \cup P'_t$	$Q'_t \cap P'_t$
$t = 5$	52	61	107	6
$t = 10$	91	105	185	11

Where, h_t is the size of spots having been chosen from $Q'_t \cup P'_t$ by a participant, and H is such size of spots with $t = 10$.

In this evaluation, 55 participants' cooperation is received, including 17 college students major in information engineering and 38 second-year high school students, and 64 spots are chosen as their favorite spots overall, *i.e.* 64 trials by 55 participants. Table V summarizes the favorite spots having been chosen by the participants with various $s_{k'}$.

2) Result: Table VI shows the detail of participants' choices from recommendations $Q'_t \cup P'_t$. In either case of t, the size of spots having been chosen in Q'_t is higher than P'_t. It represents that participants prefer the recommended spots in Q'_t than P'_t. Also note that from $Q'_t \cup P'_t$, 107 and 185 spots are chosen when $t = 5$ and $t = 10$ respectively, which contain duplicated spots in Q'_t and P'_t. More concretely, in all trials 6 spots are chosen from $Q'_t \cap P'_t$ when $t = 5$ by participants, and 11 spots with $t = 10$ respectively. In average, almost 2.89 spots are chosen from $Q'_t \cup P'_t$ in one trial.

The results of precision and recall are given in Fig. 7. It represents that when $t \leq 2$, although the precision of Q'_t is worse than P'_t, the recall of Q'_t and P'_t are almost at the same level. This phenomenon represents that the users who wish to visit the spots in the top of Q'_t are with little interest to the spots having been included in P'_t. Furthermore, such users tend to choose fewer spots overall than the ones who have not chosen the spots in the top of Q'_t. On the other hand, in the case of $t > 2$, proposed seasonal recommendation outperforms the ordinary recommendation only utilizing BFV. Also considering the fact that in most of the recommender systems the list of recommendations often includes more than 3 spots, the spots recommendations provided by the proposed system more fit user's demand than ordinary ones without considering travel season.

VI.   CONCLUDING REMARKS

This paper proposes a seasonal tourism recommender system using Wikipedia and Twitter to provide a list of tourist spots as seasonal recommendation. The effectiveness of the proposed system is experimentally evaluated by detailed observation of seasonal feature vector of spot and questionnaires of users' actual choices of spots. The results of evaluations indicate that SFVs certainly characterize the variable seasonal features of the spots. More concretely, the variance of SFVs follows Gaussian distribution and the similarity of SFVs reflects the similarity of the features of the corresponding

spots in a designated season. Further more, the result of questionnaire verifies that in most of the case the proposed system successfully provides seasonal spots recommendations to fit user's demand in tourism.

A future work is to extend the proposed recommender system to extract and characterize *spatial-temporal* features of the spot. Another issue is to integrate user modeling techniques into proposed recommender system, in order to improve the accuracy of recommendations. On the other hand, we also consider that in some hybrid recommender systems like [11], [22], our proposed method can be used as a component to improve them to achieve a seasonal recommendations. In future, we wish to combine the proposed method with such approaches and evaluate the performance of recommendations.

REFERENCES

[1] G.-S. Fang, S. Kamei, and S. Fujita., "Automatic generation of temporal feature vectors with application to tourism recommender systems," in *Proceedings of the 4th International Symposium on Computing and Networking*, 2016.

[2] G. Büyüközkan and B. Ergün, "Intelligent system applications in electronic tourism," *Expert systems with applications*, vol. 38, no. 6, pp. 6586–6598, 2011.

[3] Y. Kurata, Y. Shinagawa, and T. Hara, "Ct-planner5: a computer-aided tour planning service which profits both tourists and destinations," in *Workshop on Tourism Recommender Systems, RecSys*, vol. 15. ACM, 2015, pp. 35–42.

[4] Ángel García-Crespo, J. L. López-Cuadrado, R. Colomo-Palacios, I. González-Carrasco, and B. Ruiz-Mezcua, "Sem-fit: A semantic based expert system to provide recommendations in the tourism domain," *Expert Systems with Applications: An International Journal*, vol. 38, no. 10, pp. 13 310–13 319, September 2011.

[5] M. Zanker, M. Jessenitschnig, and W. Schmid, "Preference reasoning with soft constraints in constraint-based recommender systems," *Constraints*, vol. 15, no. 4, pp. 574–595, 2010.

[6] P. Lops, M. De Gemmis, and G. Semeraro, "Content-based recommender systems: State of the art and trends," in *Recommender systems handbook*. Springer, 2011, pp. 73–105.

[7] D. Herzog and W. Wörndl, "A travel recommender system for combining multiple travel regions to a composite trip," in *Workshop on New Trends in Content-Based Recommender Systems*. Citeseer, 2014, pp. 42–48.

[8] R. A. Abbaspour and F. Samadzadegan, "Time-dependent personal tour planning and scheduling in metropolises," *Expert Systems with Applications*, vol. 38, no. 10, pp. 12 439–12 452, 2011.

[9] R. Győrödi, C. Győrödi, and M. Deridan, "An extended recommendation system using data mining implemented for smart phones," *International Journal of Computers and Technology*, vol. 11, no. 3, pp. 2360–2372, 2013.

[10] A. Garca-Crespo, J. Chamizo, I. Rivera, M. Mencke, R. Colomo-Palacios, and J. M. Gmez-Berbs, "Speta: Social pervasive e-tourism advisor," *Telematics and Informatics*, vol. 26, no. 3, pp. 306 – 315, 2009.

[11] G. Fenza, E. Fischetti, D. Furno, and V. Loia, "A hybrid context aware system for tourist guidance based on collaborative filtering," in *Fuzzy Systems (FUZZ), 2011 IEEE International Conference on*. IEEE, 2011, pp. 131–138.

[12] J. M. Noguera, M. J. Barranco, R. J. Segura, and L. MartíNez, "A mobile 3d-gis hybrid recommender system for tourism," *Information Sciences: an International Journal*, vol. 215, pp. 37–52, December 2012.

[13] R. Baraglia, C. Frattari, C. I. Muntean, F. M. Nardini, and F. Silvestri, "A trajectory-based recommender system for tourism," in *International Conference on Active Media Technology*. Springer, 2012, pp. 196–205.

[14] C. Lucchese, R. Perego, F. Silvestri, H. Vahabi, and R. Venturini, "How random walks can help tourism," in *European Conference on Information Retrieval*. Springer, 2012, pp. 195–206.

[15] K. H. Lim, "Recommending tours and places-of-interest based on user interests from geo-tagged photos," in *Proceedings of the 2015 ACM SIGMOD on PhD Symposium*. ACM, 2015, pp. 33–38.

[16] F. Giannotti, M. Nanni, F. Pinelli, and D. Pedreschi, "Trajectory pattern mining," in *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2007, pp. 330–339.

[17] K. Shimada, S. Inoue, and T. Endo, "On-site likelihood identification of tweets for tourism information analysis," in *Proceedings of the IIAI International Conference on Advanced Applied Informatics*, 2012, pp. 117–122.

[18] K. Oku, K. Ueno, and F. Hattori, "Mapping geotagged tweets to tourist spots for recommender systems," in *Proceedings of the IIAI 3rd International Conference on Advanced Applied Informatics*, 2014, pp. 789 – 794.

[19] J. C. L. Menchavez and K. J. P. Espinosa, "Fun in the philippines: Automatic identification and sentiment analysis of tourism-related tweets," in *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, Nov 2015, pp. 660–667.

[20] W. B. Claster, M. Cooper, and P. Sallis, "Thailand–tourism and conflict: Modeling sentiment from twitter tweets using naïve bayes and unsupervised artificial neural nets," in *2010 Second International Conference on Computational Intelligence, Modelling and Simulation*, 2010, pp. 89–94.

[21] C. Elkan, "Using the triangle inequality to accelerate k-means." in *ICML*, T. Fawcett and N. Mishra, Eds. AAAI Press, 2003, pp. 147–153.

[22] C. Tan, Q. Liu, E. Chen, H. Xiong, and X. Wu, "Object-oriented travel package recommendation," *ACM Trans. Intell. Syst. Technol.*, vol. 5, no. 3, pp. 43:1–43:26, Sept. 2014.