
Spot extraction and analysis using an automatic detection method of tourist spots using SNS

Munenori Takahashi*

Polytechnic University,
Kodaira-shi, Tokyo, Japan
Email: m20305@uitech.ac.jp
*Corresponding author

Masaki Endo and Shigeyoshi Ohno

Division of Core Manufacturing,
Polytechnic University,
Kodaira-shi, Tokyo, Japan
Email: endou@uitech.ac.jp
Email: ohno@uitech.ac.jp

Masaharu Hirota

Faculty of Informatics,
Okayama University of Science,
Okayama-shi, Okayama, Japan
Email: hirota@mis.ous.ac.jp

Hiroshi Ishikawa

Graduate School of System Design,
Tokyo Metropolitan University,
Hino-shi Tokyo, Japan
Email: ishikawa-hirosi@tmu.ac.jp

Abstract: Tourism information collection using the social network services (SNSs) has become popular in recent years. Geotagged tweets are useful as a social sensor for estimating and acquiring local tourist information in real time because the information can reflect real-world situations. Earlier studies of methods of estimating cherry blossom viewing times have typically relied on the assumption that one knows a tourist destination: it is impossible to estimate cherry blossom tourist spots that a system user does not know. In its early stages, it can use tweets to find spots already featured in magazines and on the internet. As described herein, spots were detected automatically using a geotagged tweet by visualisation with a heat map and by setting conditions. The proposed method achieved it in about 80% of cases. We also used geotagged Tweets to assess observations of cherry blossom ‘front lines’ of viewing.

Keywords: social network service; SNS; mining; sightseeing; spot detection.

Reference to this paper should be made as follows: Takahashi, M., Endo, M., Ohno, S., Hirota, M. and Ishikawa, H. (2022) ‘Spot extraction and analysis using an automatic detection method of tourist spots using SNS’, *Int. J. Intelligent Information and Database Systems*, Vol. 15, No. 1, pp.6–27.

Biographical notes: Munenori Takahashi is currently enrolled in a Master’s course at a Polytechnic University. His research areas include big data and web mining. In particular, there is interest in the prediction using machine learning from the big data. He is also a member of NPO Society for Tourism Informatics.

Masaki Endo earned a BE degree from Polytechnic University, Tokyo and graduated from the course of Electrical Engineering and Computer Science, Graduate School of Engineering Polytechnic University. He received an ME degree from NIAD-UE, Tokyo. He earned a PhD in Engineering from Tokyo Metropolitan University in 2016. He is currently an Associate Professor of Polytechnic University, Tokyo. His research interests include web services and web mining. He is also a member of DBSJ, NPO STI, IPSJ, and IEICE.

Shigeyoshi Ohno earned an MSc and Dr. Sci. degrees from Kanazawa University and Dr. Eng. degree from Tokyo Metropolitan University. He is currently a full Professor of Polytechnic University, Tokyo. His research interests include big data and web mining. He is a member of DBSJ, IPSJ, IEICE and JPS.

Masaharu Hirota received his Doctor of Informatics degree in 2014 from Shizuoka University. After working for the National Institute of Technology, Oita College, he has been working as an Associate Professor in the Faculty of Informatics, Okayama University of Science since April, 2017. His research interests include photograph, GIS, multimedia, and visualisation. He is a member of ACM, DBSJ, and IPSJ.

Hiroshi Ishikawa earned BS and PhD degrees in Information Science from The University of Tokyo. After working for Fujitsu Laboratories and becoming a full Professor at Shizuoka University, he became a full Professor at Tokyo Metropolitan University in April, 2013. He is an Emeritus Professor there. His research interests include databases, data mining, and social big data. He has published actively in international refereed journals and conferences such as ACM TODS, IEEE TKDE, VLDB, IEEE ICDE, and ACM SIGSPATIAL. He has authored several books: *Social Big Data Mining* (CRC Press). He is a Fellow of IPSJ and IEICE and is a member of ACM and IEEE.

This paper is a revised and expanded version of a paper entitled ‘Automatic tourist spot detection method using SNS’ presented at Proceedings of the 10th International Conference on Web Intelligence, Mining and Semantics (WIMS’20). ACM, Biarritz, France, 30 June to 3 July 2020.

1 Introduction

In recent years, opportunities for tourists to obtain tourism information using the web are increasing. Particularly because of the spread of social networking services (SNSs), information of diverse kinds is distributed and accumulated on the web. Some SNS can use location information. One of them is Twitter. We are currently estimating the best time to view cherry blossoms using information related to Twitter’s geotagged tweets. A

low-cost method (Endo et al., 2016) was proposed to use this information for estimating optimal times for observing phenology using a moving average. The proposed method can estimate the best time to view cherry blossoms in prefectures and municipalities where a certain number of tweets with geotags are visible. In addition, the geotagged tweets used for this method are useful as a social sensor to grasp real-world situations. Therefore, this effective method of estimating viewing times can provide local tourism information in real time. By contrast, existing methods require users to have a known location of the cherry blossom viewing spot they want to visit. The spot must be translated correctly. Then the user must find related information using the internet. Therefore, when using an existing system, a person is unable to obtain information related to the optimal time because the correct information necessary to see cherry blossom highlights is insufficient. For users to be comfortable using the system, it must be independent of the amount of knowledge the user already has. We also considered whether verifying the possibility of observing the cherry blossom front line would improve the accuracy of the best time to estimate. Therefore, this study investigates a method for automatic spot detection that is useful when applied in combination with existing methods to estimate the optimal timing. This method establishes a comprehensive extraction method that is particularly useful for tourism by people making phenological observations. The method can detect spots in addition to providing cherry blossom viewing estimates (<https://www.data.jma.go.jp>). Therefore, as described herein, we conducted an experiment to assess automatic detection of cherry blossom sites from tweets. This report also describes that the automatic detection of spots can engender the discovery of unknown spots. Results indicate that a user can observe a calculated cherry blossom viewing ‘front line’.

2 Related research

Large amounts of diverse information accumulate continuously in the archives of SNSs. Examples include location information, images, and character strings. Hereinafter, we describe research that has examined information extraction from SNSs.

Chen et al. (2019) studied space–time similar area mining using SNSs to elucidate the atmosphere of a destination in advance. For the present study, we mine similar areas only at a specific time with clustering by time information in addition to position information.

Isokawa et al. (2020) detected changes in local regional characteristics based on patterns from prior statistical congestion data. They analysed changes while particularly addressing regional characteristics. By statistical analyses for which a spatial mesh and change points were set in two counties, they confirmed regional changes quantitatively when a specific new, large facility was created.

Rocío et al. (2020) performed sentiment analysis by finding correlation in public reactions to government tweets. After they demonstrated identification of people’s emotions in response to government tweets, they visualised them.

Cheng (2008) described that the core content improves a category-based collaborative filtering algorithm based on the user authority. The newly proposed algorithms overcome shortcomings of traditional collaborative filtering, i.e., the problem of an extreme sparsity matrix.

Maenaka et al. (2015) proposed the Sakura Sensor, a participatory sensing system that extracts landscape route information automatically from videos taken using an in-vehicle smartphone. It then shares data among users nearly in near real time. Using the method described by Maenaka et al., we were able to confirm cherry blossoms in a flowering state with accuracy of about 74% and a recall rate of 84%. In addition, the k-stage sensing method achieved the same point of interest detection rate in half the sensing time as that shown by the conventional method.

Finally, although research on SNS and regional changes has been conducted as described above, no study of cherry blossom spot analysis has been reported. Therefore, for this study, we detect cherry blossom sightseeing spots automatically and estimate the best time to view them.

3 Proposed method

For this research, we propose a method that can estimate an optimal viewing time irrespective of the amount of available knowledge. Earlier studies demonstrated that if a person can see the number of geotagged tweets in a prefecture or city, then that person can estimate the best time to visit. Taking the estimate of the best time in Tokyo, then comparison of the cherry blossom bloom date to the full bloom day (JMA – Meteorological Agency Information on Phenology, 2020) observed by the Japan Meteorological Agency indicates a recall rate of 100% and a precision rate of 53.3% for the method in 2018. It is noteworthy that the full-bloom day is a state in which about 80% or more buds of the sample tree are open. Therefore, the best time estimate continues even after the full-bloom day. Using this conventional method, one can estimate the best time for viewing after the day of full bloom for a prefecture or municipality unit. Furthermore, specifying the prefecture name and municipality at the time of estimation is necessary. Using the proposed method, latitude and longitude information attached to the tweet is used for application to the map. The number of tweets was visualised using a heat map. Subsequently, the conditions were set using the heat map. The tourist spots were detected. Using the existing method, we conducted an experiment to ascertain whether the automatically detected spots can be best-time estimated.

4 Experiment method

This section presents descriptions of preprocessing and the data used. Using the Streaming API (<https://dev.twitter.com/>), we collected geotagged tweets with location information including latitude and longitude in Japan. Then we analysed tweets with biological names. For this experiment conducted during February 17, 2015 through July 29, 2019, the transition of tweets related to cherry blossoms was confirmed with analysed organism names as ‘さくら’, ‘サクラ’, and ‘桜’. In addition, from latitude and longitude information of the geotagged tweet including location information, the simple reverse geocoding service (Agricultural Research Institute, 2020) of the National Research Institute for Agricultural Sciences and Food Industry was used for latitude and longitude information, and for the prefecture / municipality from which each tweet was sent. Analyses were conducted using general towns and streets within the named city planning area.

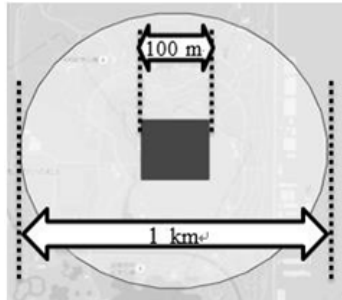
4.1 Automatic spot detection method

This section presents a description of the correct data used for automatic detection of cherry spots and the existing method used in this experiment. First, from Walker + 2019 ver. 1,000 views, we set 64 correct spots corresponding to cherry blossoms in Tokyo and 40 correct spots corresponding to cherry blossoms in Shizuoka and 91 correct spots corresponding to cherry blossoms in Kyoto. Spots were determined when generating a 3D map in Excel and using heat map shading to meet the conditions. Figure 1 presents the heat map shading and condition range. A small tweet quantity is displayed in black. Large quantities are shown in white: range (1) is 85%–100%; range (2) is 60–85% (0% is black, 100% is white).

Figure 1 Light and shade of the heat map and condition range



Figure 2 Range of correct answers



The proposed method is a system for automatically detecting sightseeing spots: it is independent of the amount of related user knowledge. At the initial stage, we investigate whether the proposed method can achieve automatic detection of existing spots. Therefore, we set a square with a 100 m side and a circle with 1 km diameter centred on the latitude and longitude indicated by Walker + as the correct answer spot. Then, three conditions were set as conditions for automatic detection:

Condition 1 – the square contains tweet volume (1) [range (1) in Figure 1]

Condition 2 – the circle contains tweet volume (1) [range (1) in Figure 1]

Condition 3 – the circle contains tweet volume (2) [range of (2) presented in Figure 1].

Using this condition, we can confirm whether the correct answer spot can be detected automatically from the tweets. We show how many correct answer spots were detected by automatic spot detection. Then we calculate the spot automatic detection rate. The method obtains the value by dividing the spots detected automatically by the number of correct answer spots. Figure 2 portrays the squares and circles set within the correct range.

4.2 Judgment of cherry blossom viewing time

As the best time estimation method, we use the existing method (Takahashi et al., 2019), which uses a weighted moving average. This section presents a description of existing methods.

4.2.1 Weighted moving average

The weighted moving average used for the existing method is a moving average with each value assigned a weight. By adding weights, the recall and accuracy of the best time estimation are improved. With the existing method, the median was set to 1. The values ± 0.5 from the median were taken respectively as the minimum and maximum values. In addition, except for the median, weights from the lowest value to the highest value were assigned linearly. The third decimal place was rounded. Taking the 5-day weighted moving average used in the existing method as an example, the following equation (1) was obtained. Here,

Havg5 represents the weighted moving average for 5 days; x_y is the number of tweets x before y days prior.

$$Havg5 = (x_5 \times 0.5 + x_4 \times 0.75 + x_3 \times 1 + x_2 \times 1.25 + x_1 \times 1.5) / 5 \quad (1)$$

4.2.2 Best time estimation method

Using a simple moving average and a weighted moving average, the following estimation method was set for the frequency of appearance of each geotagged tweet containing the target word. Results were analysed by date to estimate the best viewing period.

- 1 We used a one-year simple moving average to ascertain the period during which tweets about cherry blossoms increased.
- 2 Because the number of tweets tends to be higher on Saturdays and Sundays, the seven-day weighted moving average was used on a weekly basis.
- 3 A five-day weighted moving average was used based on the average number of days from flowering of cherry blossoms to full bloom, five days.

The best time to see each tourist spot was inferred using these best time estimation criteria.

5 Experiment results

This chapter presents results of automatic spot detection and its best time estimation.

5.1 Automatic spot detection

One can visualise the frequency of tweets by application of the latitude and longitude information of geotagged tweets to the Excel heat map. An experiment was conducted to

assess automatic spot detection using the conditions shown in 4.1 and the range of correct answer spots. As an example, Figure 3 shows spots (Jindai Botanical Park) that meet condition 1. Figure 4 shows spots (Ukima Park) that meet condition 2. Table 1 presents the same decision results for all 64 cases in Tokyo. Table 2 shows spot automatic detection rates under conditions 1–3 obtained based on results of all 64 judgments in Tokyo. Table 3 presents the same decision results for all 40 cases in Shizuoka. Table 4 shows the spot automatic detection rates under conditions 1–3 obtained based on results of all 40 judgments in Shizuoka. Table 5 shows the same decision for 60 cases randomly selected in Kyoto to fit the space. Table 6 shows the automatic spot detection rate under conditions 1–3 based on the results of all 91 judgments made in Kyoto.

Figure 3 Correct answer spots that meet condition 1

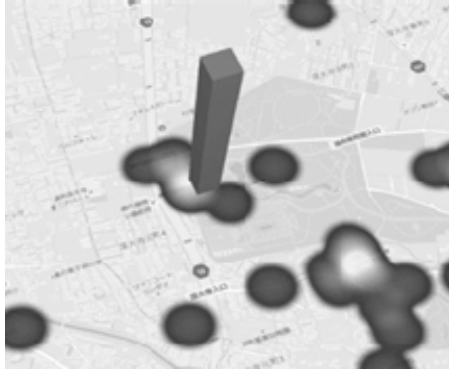


Figure 4 Correct answer spot that meets condition 2

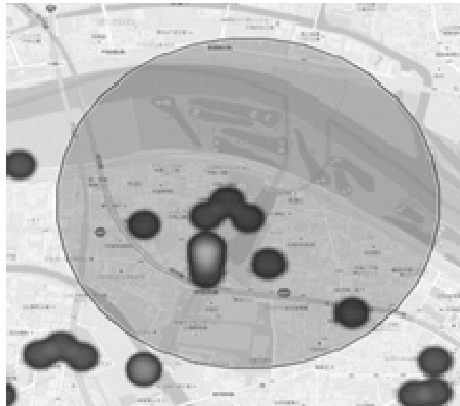


Table 1 Judgment results of all 64 correct answer spots in Tokyo

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>		<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Ark Hills	0	0	0	Chidorigafuchi Park	1	1	1
Asukayama Park	1	1	1	Central Park	0	0	0
Senzoku Pond Park	0	0	1	Tokyo Midtown	1	1	1
Ikegami Honmonji Temple	1	1	1	Toshimaen Sakura Festival	0	0	1
Inokasira Gift Park	1	1	1	Toneri Park	0	0	1
Ueno Gift Park	1	1	1	Toyama Park	1	1	1
Ukima Park	0	1	1	Metropolitan Sakuragaoka Park	0	0	1
Baigan Temple	1	1	1	Metropolitan Sayama Park	0	0	1
Ota-ku Tamagawadai Park	0	1	1	Metropolitan Jindai Botanical Garden	1	1	1
Lake Okutama	0	0	0	Metropolitan Hikarigaoka Park	0	1	1
Otonashishinsui Park	1	1	1	Metropolitan Yoyogi Park	1	1	1
One Green Road	0	1	1	Nishiarai Park	0	0	1
Kamanofuchi Park	0	0	0	Hibiya Park	1	1	1
Former iwasaki Garden	0	1	1	Hamarikyu Gift Garden	0	1	1
Former Shiba Rikyu Garden	1	1	1	Hamura Weir	1	1	1
Former Furukawa Garden	0	1	1	Fujimori Park	1	1	1
Kiyosumi Garden	0	1	1	Homyouji Temple Approach	1	1	1
Koishikawa Korakuen	0	1	1	Mizumoto Park	0	0	0
Koganei Park	0	1	1	Myoujinshita Park	0	0	0
National Showa Memorial Park	0	0	1	Mukojima Hundred Flower Gaeden	0	0	0
Kotta river	0	0	0	Musashino Park	0	0	1
Komazawa Olympic Park	0	1	1	Meiji jingu Gaien	0	1	1
Sunshine City	1	1	1	Meguro River	1	1	1
Shiotakochitsutsumi	0	0	0	Mohri Garden, Roppongi	1	1	1
Shiba Park	1	1	1	Yaesu Sakura Street	1	1	1
Shakujii Park	0	0	0	Yomiuri Land	0	0	0
Shinjuku Gyoen	1	1	1	Rikugien	1	1	1
Sumida Park	1	1	1	Gaien Park	1	1	1
Sendaihorikawa Park	0	0	0	Kinuta Paek	0	0	1
Zenpukuji River Margin	0	0	0	Tatsuminomori Green Park	0	0	1
Takiyama Park	0	0	1	Harimazaka	1	1	1
Tama river Embankment	0	0	1	Sakura at Yasukuni Shrine	1	1	1

Table 2 Spot automatic detection rates in Tokyo

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Hit	27	39	51
Answer spot	64	64	64
Answer spot automatic detection rate	42.2%	60.9%	79.7%

Table 3 Judgment results of all 40 correct answer spots in Shizuoka

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>		<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Shizuoka sengen shrine	0	0	1	Mt. ie	1	1	1
Sumpu castle park	1	1	1	Nomori pond	0	1	1
Nihondaira	0	1	1	Kiya river bank	0	0	0
Shimiz funakoshitsutsumi park	0	0	1	Kurigatake	0	0	0
Mt. goten	0	0	1	Tsumagoi resort saino sato	0	1	1
Hamamatsu castle park	1	1	1	Renge temple pond park	0	0	1
Sanaru lake park	0	0	0	Mt. Konpira green park	1	1	1
Okuyama park	0	0	1	Fuji peace park	0	0	1
Akiba dam	0	0	0	Chichibunomiya sakura park	0	0	1
Mt. toba park	0	0	0	Kano river sakura park	0	1	1
Chitose river	0	0	0	Katsumata river	0	0	0
Ito river promenade	1	1	1	Kawazu river	1	1	1
Atami castle	0	1	1	Shimogamo hot spring	1	1	1
Mishima taisha	1	1	1	Matsuzaki town	1	1	1
Kariyado no Gebazakura	0	0	1	Koganezaki park	0	0	0
Taiseki temple	0	0	0	Surugadaira natural park	0	0	1
Izu kougen	0	0	0	Fuji cemetery	0	0	0
Sakura no sato	1	1	1	Ookuni shrine	0	0	0
Ito hot spring	0	1	1	ota river sakura dank	0	0	0
Kawaramachi old embankment	0	0	0	Sengoku yumekaido hiking course	0	0	0

Table 4 Spot automatic detection rate in Shizuoka

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Hit	10	16	25
Answer spot	40	40	40
Answer spot automatic detection rate	25.0%	40.0%	62.5%

Table 5 Judgment result of 60 randomly selected correct answer spots in Kyoto to fit the space

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>		<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Hirano Shrine	1	1	1	Maruyama Park	1	1	1
Kamigamo Shrine	0	1	1	Kodaiji Temple	1	1	1
Elegant Rendaiji Temple	1	1	1	Chion-in	0	1	1
Joshoji Temple	0	0	0	Kiyomizu Temple	1	1	1
Myorenji Temple	0	1	1	Rokusonno Shrine	0	1	1
Ryuhonji Temple	1	1	1	Omuro (Ninnaji Temple)	1	1	1
Myokenji Temple	1	1	1	Arashiyama	1	1	1
Honpo-ji Temple	0	1	1	Tenryu temple	1	1	1
Honmanji Temple	1	1	1	Jojakkoji Temple	0	0	1
Shinen	1	1	1	Nison-in	0	0	1
Philosophy road	0	1	1	Daikakuji Temple	0	0	1
Jukkokubune tour	1	1	1	Joshokoji Temple	0	0	0
Kyoto Botanical Garden	0	1	1	Houkongouin	1	1	1
Yase beauty	0	0	0	Umenomiya Shrine	0	1	1
Sanzenin Monzeki	1	1	1	Kaohsiung / Mio area	0	0	0
Kyoto International Conference Center	0	1	1	Ryoanji Temple	1	1	1
Kempon Hokke Sect	0	0	0	Taizo-in (Myoshinji Yamauchi)	0	1	1
Kitazono Pond	0	0	1	Okochi Sansoen	0	1	1
Konkai Komyoji Temple	1	1	1	Daigoji Temple	1	1	1
Hyakumanben Chionji Temple	0	1	1	Jonangu Shrine Rakusuien	0	0	0
Nanzenji Temple	0	1	1	Yodo Waterway	1	1	1
Iwakura Jisso-in Gate Ruins	0	0	0	Gokonomiya Shrine	0	0	1
Shogoin Gate Ruins	0	1	1	Fushimi Momoyama Castle	0	0	1
Incline	1	1	1	Sumizome Temple	0	0	1
Anrakuji	0	1	1	Sakura police	1	1	1
Kurama Temple	0	0	1	Bishamon-do	0	1	1
Kamogawa Park	1	1	1	Kajuji Temple	0	0	0
Cherry blossom garden	0	1	1	Lake Biwa Canal	0	0	0
Clear stream garden	0	1	1	Oishi Shrine	1	1	1
Shinsen-en	1	1	1	Iwayaji Temple	0	1	1

As presented in Table 2, 27 correct answer spots were found for which the range of tweet volume (1) was included in the 100 m square of condition 1; the spot automatic detection rate was 42.2%. In condition 2, the number of correct answer spots in which the tweet volume (1) range was included in a circle of 1 km diameter was 39; the spot automatic detection rate was 60.9%. Under condition 3, the number of correct answer spots which included the range of tweet volume (2) within a circle of 1 km diameter was 51; the spot automatic detection rate was 79.7%.

As presented in Table 4, 10 correct answer spots were found for which the range of tweet volume (1) was included in the 100 m square of condition 1; the spot automatic detection rate was 25.0%. In condition 2, the number of correct answer spots in which the tweet volume (1) range was included in a circle of 1 km diameter was 16; the spot automatic detection rate was 40.0%. Under condition 3, the number of correct answer spots which included the range of tweet volume (2) within a circle of 1 km diameter was 25; the spot automatic detection rate was 62.5%.

Table 6 Spot automatic detection rates in Kyoto

	<i>Condition 1</i>	<i>Condition 2</i>	<i>Condition 3</i>
Hit	30	52	65
Answer spot	91	91	91
Answer spot automatic detection rate	33.0%	57.1%	71.4%

As presented in Table 6, 30 correct answer spots were found for which the range of tweet volume (1) was included in the 100 m square of condition 1; the spot automatic detection rate was 33.0%. In condition 2, the number of correct answer spots in which the tweet volume (1) range was included in a circle of 1 km diameter was 52; the spot automatic detection rate was 57.1%. Under condition 3, the number of correct answer spots which included the range of tweet volume (2) within a circle of 1 km diameter was 65; the spot automatic detection rate was 71.4%.

The low spot automatic detection rate obtained under condition 1 is explainable by the fact that the range of the 100 m square was narrow. The reason for this low rate is that many spots are famous for cherry blossoms. They are all correct answers. They include parks and gardens, which cover large areas, but a 100 m square cannot encompass such a large park. In Condition 2, the correct answer spot that was leaked in Condition 1 was also determined. However, a new difficulty is that tweets are scattered to that extent in large parks and gardens. Because the tweet points of origin are not concentrated in one place, one can infer that the number of tweets did not increase to the threshold range of tweet volume (1) when regarding the tweet volume. A large tourist spot such as a park or garden cannot be estimated correctly if the range of conditions is narrow.

We can explain why the automatic spot detection rate for Kyoto is lower than that for Tokyo. Some correct spots in Kyoto are in the mountains. There were a few more correct spots in the mountains than in Tokyo. As a result, an increased number of tweets were made from the flat ground where one can see the famous cherry blossoms in the mountains. However, fewer tweets were made near the mountain spots because few tourists visited the mountain cherry blossom spots.

Differences between results obtained in Tokyo and Shizuoka are explainable: an important difference is the number of tweets which can be extracted. About 145,000 tweets were extracted in Tokyo, but only about 14,000 were extracted in Shizuoka. The

difference in the extracted tweets is therefore about 10 times: the number of spots that can be detected decreased automatically. However, that decrease might be improved by increasing the number of data using other SNS. In Shizuoka Prefecture, correct answer spots are located at the mountains and at the foot of the mountains. Some of them are beyond the reach of radio waves. Some can only accommodate four buses a day. Therefore, moving by car is necessary. Perhaps tweeting is impossible because a person is driving. The tweet might not be detected because parking spaces are sparse. Figure 5 to Figure 7 present the respective communication areas, marked as grey, of three telephone services in Japan: au™ (KDDI Corp.), NTT Docomo Inc., and Softbank Group Corp.

Figure 5 au™ (KDDI Corp.) communication areas

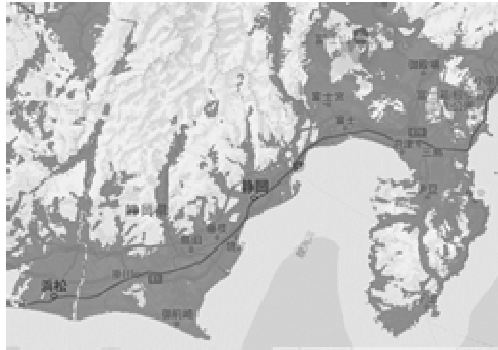


Figure 6 NTT Docomo Inc. communication areas

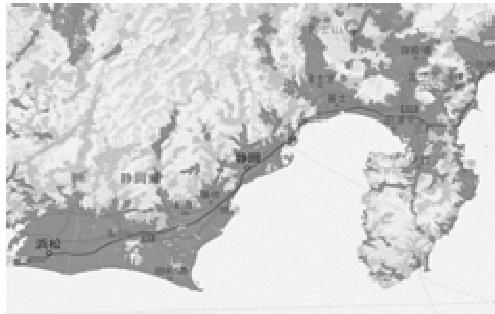
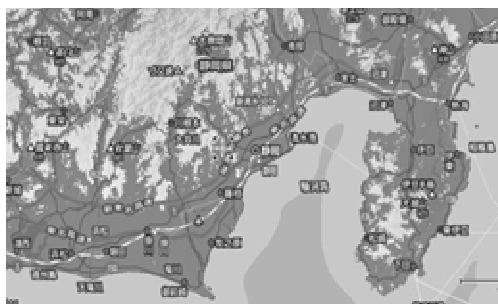


Figure 7 Softbank Group Corp. communication areas



The possibility of detecting an unknown spot is described next. For this experiment, we investigated the condition settings for automatic detection of existing spots. However, the tweet amounts during the experiment met conditions in some places other than those of existing spots. For example, spots such as Notsuda Park and Onda River in Tokyo, and Kakegawa Castle Park and Fujikawa Rakuza in Shizuoka, Kinmei Park and Hirosawa Pond in Kyoto were also extracted. Survey results confirmed that, other than spots with correct answers, these are also cherry blossom tourism spots. Further examination of extraction conditions must be done in future studies, but the proposed method can support the discovery of unknown spots.

5.2 Best time estimation

The automatic spot detection method described in 5.1 identified 51 cherry blossom tourist spots. We used the existing method to infer the best time to view cherry blossoms at the detected spots. As an example, we address the famous Meguro River cherry blossoms because the total amount of tweets is large. Using the existing method, we did not conduct an experiment when the tweets were few. Therefore, we chose a spot with a large total amount of tweets that is being tested using the existing method. Figures 8 to 10 present results of the best time to view cherry blossoms on the Meguro River during 2017–2019.

Figure 8 Meguro River in 2017

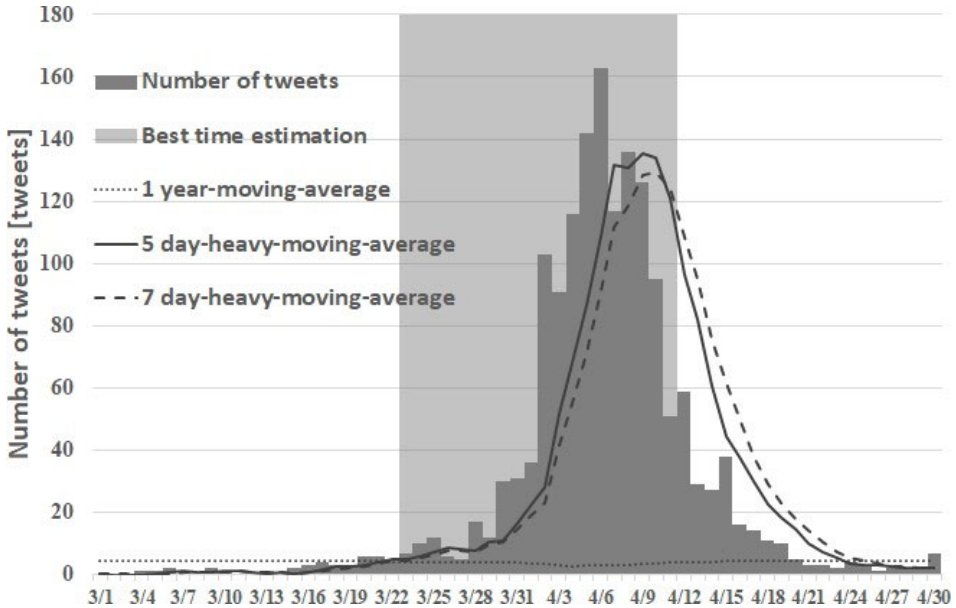


Figure 9 Meguro River in 2018

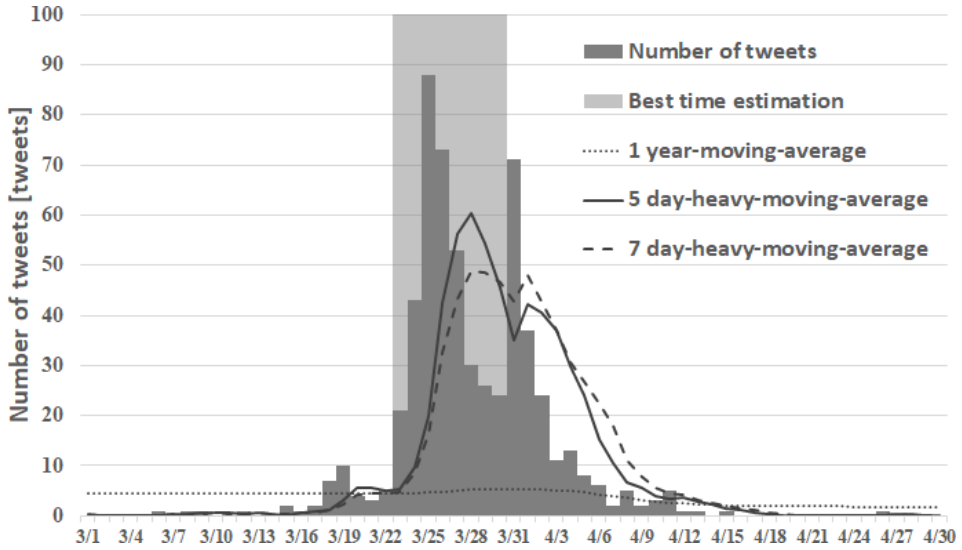
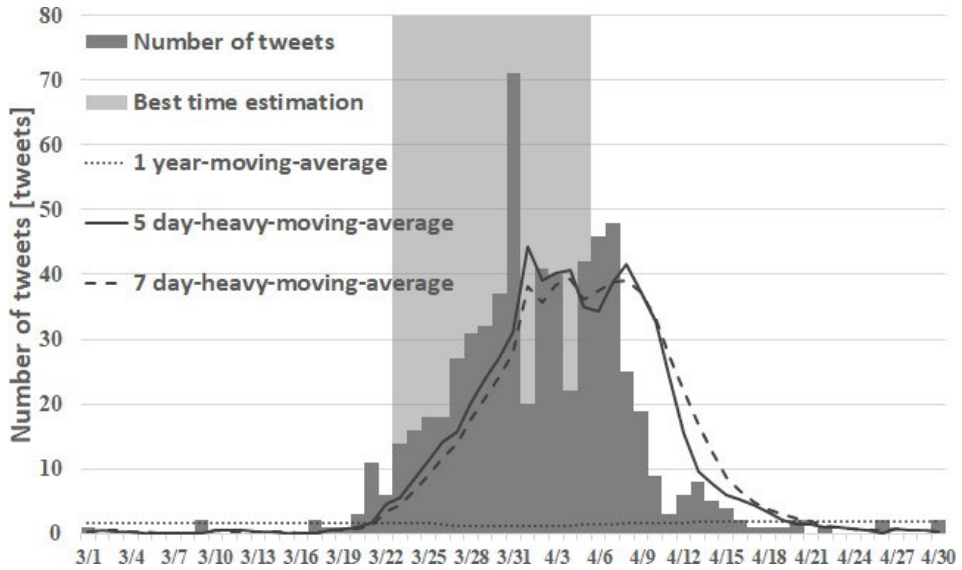


Figure 10 Meguro River in 2019



Next, Table 7 presents a comparison between the period from the blooming of cherry blossoms in Tokyo to full bloom observed by the Japan Meteorological Agency and the best time to estimate the cherry blossoms in the Meguro River. The period from flowering to full bloom of the Japan Meteorological Agency is indicated by a black arrow. The grey part represents the best estimated time to see the blossoms. Table 8 presents the recall and precision for each year when the Meguro River cherry blossoms are estimated.

Table 7 Comparison of cherry blossom viewing times

<i>2017</i>	<i>Estimated</i>	<i>2018</i>	<i>Estimated</i>	<i>2019</i>	<i>Estimated</i>	
3/1		3/1		3/1		
3/2		3/2		3/2		
3/3		3/3		3/3		
3/4		3/4		3/4		
3/5		3/5		3/5		
3/6		3/6		3/6		
3/7		3/7		3/7		
3/8		3/8		3/8		
3/9		3/9		3/9		
3/10		3/10		3/10		
3/11		3/11		3/11		
3/12		3/12		3/12		
3/13		3/13		3/13		
3/14		3/14		3/14		
3/15		3/15		3/15		
3/16		3/16		3/16		
3/17		3/17	↑	3/17		
3/18		3/18	↓	3/18		
3/19		3/19		3/19		
3/20		3/20		3/20		
3/21	↑	3/21		3/21		
3/22		3/22		3/22		
3/23		3/23	↓	3/23	↑	
3/24		3/24		3/24		
3/25		3/25		3/25		
3/26		3/26		3/26		
3/27		3/27		3/27		
3/28		3/28		3/28		
3/29		3/29		3/29		
3/30		3/30		3/30		
3/31		3/31		3/31		
4/1			4/1			4/1
4/2		4/2		4/2		
4/3		4/3		4/3		
4/4		4/4		4/4		
4/5		4/5		4/5		
4/6		4/6		4/6		
4/7		4/7		4/7		
4/8		4/8		4/8		
4/9		4/9		4/9		
4/10		4/10		4/10		
4/11		4/11		4/11		
4/12		4/12		4/12		

Table 8 Recall rate and precision rate for cherry blossom viewing

	<i>Recall</i>	<i>Precision</i>
2017	100%	59.1%
2018	25.0%	25.0%
2019	71.4%	35.7%

As shown in Tables 7 and 8, we obtained good results for best time estimates in 2017 because the period from flowering to full bloom, as indicated by the Japan Meteorological Agency, can be estimated as in full bloom. In 2019 as well, the recall rate did not reach 100%, but sufficient results were obtained. From start to completion of best time estimation, an estimate can be done without interruption. However, in 2018, less than half of the period from flowering to full bloom indicated by the Japan Meteorological Agency was estimated. Probably, estimation from the start to the end of the full bloom was made to some degree. Based on the discussion presented above, estimation results of sufficient cherry blossom viewing times were obtained, except for 2018. The method is somewhat applicable, even when used as it is. Of the last three years, two have been successful: in 2018, the recall and precision decreased considerably. Depending on the cause in 2018, it might be difficult to apply the existing method as it is. The next section presents an explanation of the causes of decreased accuracy in 2018.

5.2.1 Causes of reduced recall in 2018

The decreased recall which occurred in 2018 was investigated. Results indicate weather as a factor reducing the recall rate. Weather information for March 2018 was obtained from the goo weather (2020). Table 9 presents weather information for March 2018. ● are sunny, ▲ are cloudy, × are rain, and ■ are snow.

Table 9 Weather information for March 2018

<i>Day</i>		<i>11</i>	<i>12</i>	<i>13</i>	<i>14</i>	<i>15</i>	<i>16</i>	<i>17</i>
Temperature (°C)	Max	16.1	15.9	17.5	21.5	22.1	16.9	12.1
	Min	5.5	5.2	5.3	7.2	9	15.7	3.1
Time	9 A.M.	▲	●	●	●	▲	×	●
	12 P.M.	●	●	●	●	▲	×	●
	15 P.M.	●	●	●	●	▲	×	●
<i>Day</i>		<i>18</i>	<i>19</i>	<i>20</i>	<i>21</i>	<i>22</i>	<i>23</i>	<i>24</i>
Temperature (°C)	Max	15.8	18	8.8	3.8	14.5	18.8	16.8
	Min	5.4	8.7	7.7	3.7	3.7	9.4	7.1
Time	9 A.M.	●	▲	×	×	×	▲	●
	12 P.M.	▲	▲	×	■	×	●	▲
	15 P.M.	▲	▲	▲	■	●	●	▲

Table 9 shows good weather for March 17, which the meteorological agency indicated as the best time for viewing. Because of bad weather, the number of outings probably decreased on March 15 and 16. It is possible that the temperature was low on March 17 because of the bad weather the day before. It is also possible that the number of people going out decreased. Temperatures rose on March 18 and 19, but the number of tweets did not increase, probably because the weather was cloudy. Then, during March 20–22, bad weather continued with clouds, rain, and snow. As a result, the number of outings apparently decreased for two reasons: bad weather and low temperatures. These effects probably reduced the estimation results of cherry blossom viewing time considerably because the existing method is intended for best time estimation without consideration of weather information because it was possible to estimate the best time for viewing for each prefecture in the survey using the existing method. However, results of this experiment demonstrated that the influence of weather on the tweet becomes strong when one examines one tourist spot unit, as in this case. In other words, one can not simply apply the existing method to estimate the best time to see cherry blossoms in large prefectures. These results demonstrate that noise greatly disturbs the estimation result, such as the estimation result in 2018, because the weather effect becomes large when the granularity is fine. Each tourist spot becomes a unit.

6 Cherry blossom front examination

This chapter presents consideration of the possibility of observing the cherry blossom front using tweets. This time, setting Kagoshima Prefecture as the start of the cherry blossom front, we tentatively investigated whether one can follow the cherry blossom front as it moves northward with warmer weather. Some points were set as intermediate points. They are Okayama, Kyoto, Aichi, Tokyo, Niigata, and Aomori. Tables 10–12 present results of optimal time estimation at all stations and shows the expected cherry blossom front line. The expected cherry blossom front line is indicated by a black arrow. The grey part represents the best estimated time for viewing the blossoms.

First, prediction of the cherry blossom front is explained. To explore the possibilities, the black arrow indicates the transition of the estimation, which shows a tendency as a cherry blossom front. The reason is that it is difficult to distinguish the varieties entirely from the tweet information. Distinguishing the varieties is difficult. It is not possible to distinguish early bloomers from late bloomers. At this stage, the black arrows indicate the transition of signs of the cherry blossom front. Next, regarding Table 10, Kagoshima appears to be in full bloom early, but one can roughly observe the cherry blossom front line. Next, regarding Table 11, the cherry blossom front line is visible, except for Niigata and Aomori. Similarly, in Table 12, the cherry blossom front line can be roughly observed, except for Niigata and Aomori. The cherry blossom front cannot be observed well in Aomori and Niigata for the following reasons. The Sea of Japan side of Japan and the Pacific side of Japan respectively have very different climates. The Sea of Japan side has much snow each year. Moreover, the hours of daylight are few in many areas because of the proximity of mountains (JMA – Information on the daylight hours of the Japan Meteorological Agency, 2020). The midpoint was therefore set without considering the Pacific Ocean side or the Sea of Japan side. Tables 10–12 suggest that Niigata and

Aomori prefectures, which are adjacent to Tokyo, generally lag in terms of flower development. Results show that the cherry blossom front might not be continuous at the prefectural border, where large differences exist in temperature and sunshine amounts. Furthermore, this discontinuity might apply to snowy areas, such as the Tohoku region, that are not on the Sea of Japan side.

Table 10 2017 estimated cherry blossom front line

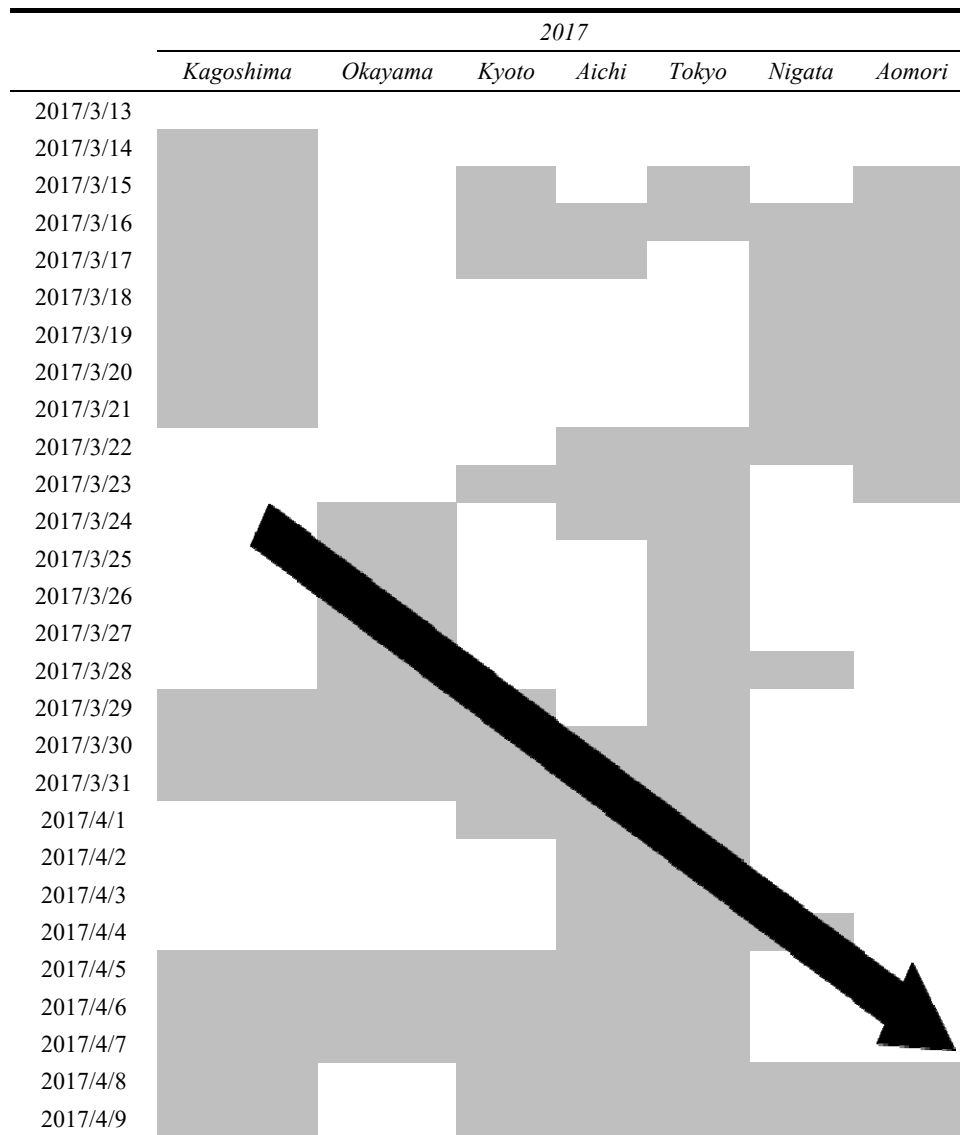


Table 11 2018 estimated cherry blossom front line

	2018						
	<i>Kagoshima</i>	<i>Okayama</i>	<i>Kyoto</i>	<i>Aichi</i>	<i>Tokyo</i>	<i>Nigata</i>	<i>Aomori</i>
2017/3/10							
2017/3/11		■					
2017/3/12		■					
2017/3/13		■		■			
2017/3/14		■		■			
2017/3/15		■		■			
2017/3/16	■	■		■	■		
2017/3/17	■	■		■	■		
2017/3/18		■	■	■	■		
2017/3/19		■	■	■	■		
2017/3/20		■	■	■	■		
2017/3/21		■	■	■	■		
2017/3/22		■	■	■	■	■	
2017/3/23		■	■	■	■	■	
2017/3/24		■	■	■	■	■	
2017/3/25		■	■	■	■	■	
2017/3/26		■	■	■	■	■	
2017/3/27		■	■	■	■	■	
2017/3/28	■	■	■	■	■	■	■
2017/3/29	■	■	■	■	■	■	■
2017/3/30	■	■	■	■	■	■	■
2017/3/31	■	■	■	■	■	■	■
2017/4/1		■	■	■	■		
2017/4/2		■	■	■	■		
2017/4/3		■	■	■	■		
2017/4/4		■	■	■	■	■	
2017/4/5		■	■	■	■	■	■
2017/4/6		■	■	■	■	■	■
2017/4/7		■	■	■	■	■	■

Table 12 2019 estimated cherry blossom front line

	2019						
	<i>Kagoshima</i>	<i>Okayama</i>	<i>Kyoto</i>	<i>Aichi</i>	<i>Tokyo</i>	<i>Nigata</i>	<i>Aomori</i>
2017/3/10							
2017/3/11							
2017/3/12		■					
2017/3/13		■		■			
2017/3/14		■	■	■			
2017/3/15			■	■			
2017/3/16							
2017/3/17	■						
2017/3/18	■						
2017/3/19							
2017/3/20	■				■		
2017/3/21					■		
2017/3/22		■			■		
2017/3/23		■	■		■		
2017/3/24		■	■		■		
2017/3/25			■		■		
2017/3/26					■		
2017/3/27					■		
2017/3/28	■				■		
2017/3/29					■		
2017/3/30	■				■		
2017/3/31					■		
2017/4/1		■			■		■
2017/4/2		■			■		■
2017/4/3		■			■		■
2017/4/4		■			■		■
2017/4/5		■			■		■
2017/4/6		■			■		■
2017/4/7							

7 Conclusions

This study was conducted to establish a tourist spot detection system requiring no prior knowledge of a tourist area. Earlier research examining this system detected spots that had been featured in magazines and in tweets from the internet. Next, we checked whether the optimal time can be estimated for automatically detected spots. We applied it to a heat map using latitude and longitude information of the geotagged tweet. By applying the conditions there, cherry blossom tourist spots were detected automatically. From experiments conducted for Tokyo, Shizuoka and Kyoto, spot detection rates were found to be 42.2%, 25.0% and 30.0% under condition 1, 60.9%, 40.0% and 57.1% under condition 2, and 79.7%, 62.5% and 71.4% under condition 3. The experiment detected many existing spots, confirming the possibility of automatic tourist site detection. Moreover, the discovery of some spots other than the correct answer spot presented the possibility of discovering unknown spots. Next, we estimated the best time to view cherry blossoms along the Meguro River. The recall rates were 100% in 2017, 25.0% in 2018, and 71.4% in 2019. Except for 2018, satisfactory results were obtained. In 2018, the model accuracy was reduced greatly because of the influence of weather. The problem persisted. Nevertheless, the system presents opportunities for development because it requires no prior knowledge of cherry blossom sightseeing spots to estimate the best time to view cherry blossoms.

In addition, even at spots from which vastly numerous tweets are issued, the weather exerts a strong influence. Therefore, results demonstrated that the existing method using the weighted moving average cannot be applied as it is. The proposed method demonstrated the possibility of automatic detection of cherry blossom tourist spots depending on the setting of conditions. Nevertheless, the existing method using a weighted moving average was not directly applicable to the detected tourist spots. Future studies must be conducted to improve the spot detection rate of the automatic detection method of tourist spots and improve the spot detection rate to assess spot situations. Furthermore, optimal time estimation method for tourist spots can be improved by considering weather information and sunshine amounts. The cherry blossom front can probably be estimated from tweets. However, climatic differences are not considered. Therefore, it is necessary to estimate the best time to view cherry blossoms in all prefectures by considering regional characteristics and then re-examining cherry blossom front estimation. More detailed cherry blossom front predictions will also be attempted in future research.

References

- Agricultural Research Institute (2020) *Simple Reverse Geocoding Service* [online] <https://www.finds.jp/rgeocode/index.html.ja> (accessed 26 April 2020).
- Chen, L., Kawanishi, Y., Ide, I., Hirayama, T., Doman, K., Deguchi, D. and Murase, H. (2019) 'SNS toukou shashin no visual concept ni motodoku jikukan ruiji chiiki mining, (Spatiotemporal similar area mining based on visual concept of photographs posted on SNS)', *Proceedings of the 11th Annual Conference on the 11th Forum on Data Engineering and Information Management (DIEM'19)*, Nagasaki, Japan, p.6 (in Japanese).
- Cheng, G. (2008) 'Interest mining and recommender system for local activities', *WI-IAT '08*, pp.90–93.

- Endo, M., Mitomi, K., Saeki, K., Ehara, Y., Hirota, M., Ohno, S. and Ishikawa, H. (2016) ‘Study of information provided by the best time to see estimation method of phenological observations using tweets’, *Proceedings of the 12th Annual Conference on Japan Society for Tourism Informatics*, Shizuoka, Japan, pp.47–60 (in Japanese).
- goo weather (2020) [online] <https://weather.goo.ne.jp> (accessed 23 May 2020).
- Isokawa, H., Toyota, M. and Kitsuregawa, Y. (2020) ‘Keitaidenwa jinkou toukei data wo mochiita shinki shisetsu ni kakawaru toshi doutai no henka kaiseki, (Analysis of changes in urban dynamics related to new facilities using mobile phone demographic data)’, *Proceedings of the 12th Annual Conference on the 12th Forum on Data Engineering and Information Management (DIEM'20)*, pp.75–82 (in Japanese).
- JMA – Information on the daylight hours of the Japan Meteorological Agency (2020) [online] <https://www.data.jma.go.jp/obd/stats/data/mdrr/tenkou/indexTenkouSun10dhi.html> (accessed 20 May 2020).
- JMA – Meteorological Agency Information on Phenology (2020) [online] <https://www.data.jma.go.jp/sakura/data/index.html> (accessed 20 May 2020).
- Maenaka, S., Morishita, S., Nagata, D., Tamai, M., Yasumoto, K., Fukukura, T. and Sato, K. (2015) ‘SakuraSensor: a system for realtime cherry-lined roads detection by in-vehicle smartphones’, *Proceedings of the 15th Annual Conference on the 15th Forum on International Symposium on Wearable Computers (ISWC'15)*, pp.345–348.
- Rocío, B.H., Estevez, E., Maguitman, A. and Janowski, T. (2020) ‘Analyzing and visualizing government–citizen interactions on twitter to support public policy-making’, *Digit. Gov.: Res. Pract.*, April, Vol. 1, No. 2, Article 15, p.20.
- Takahashi, M., Endo, M., Ohno, S., Hirota, M. and Ishikawa, H. (2019) ‘SNS de-ta wo riyou shita kankou supotto no riaru taimu na joukyou tyuusyutu syudan no kentou’, *Proceedings of the 16th Annual Conference of Japan Society for Tourism Information*, Tottori, Japan, pp.33–34 (in Japanese).
- Twitter [online] <https://twitter.com/> (accessed 30 May 2020).
- Twitter Developers (2015) <https://dev.twitter.com/> (accessed 10 February 2015).
- Walker + 1000 Views of Cherry Blossoms Nationwide (2019) [online] <https://hanami.walkerplus.com/> (accessed 21 September 2019).