Examination of Best-time Estimation using Interpolation for Geotagged Tweets

Masaki Endo, Shigeyoshi Ohno
Division of Core Manufacturing
Polytechnic University
Tokyo, Japan
e-mail: endou@uitec.ac.jp, ohno@uitec.ac.jp

Masaharu Hirota
Department of Information Engineering
National Institute of Technology, Oita College
Oita, Japan
e-mail: m-hirota@oita-ct.ac.jp

Yoshiyuki Shoji, Hiroshi Ishikawa
Graduate School of System Design
Tokyo Metropolitan University
Tokyo, Japan
e-mail: y_shoji@tmu.ac.jp, ishikawa-hiroshi@tmu.ac.jp

***Abstract*—Various studies have been conducted to analyze social media data in real time and to extract events in the real world. A benefit of analysis using data with position information is that it can accurately extract the event from a target area to be analyzed. However, because the proportion of data with position information in social media data is small, the amount to analyze is insufficient in almost areas. In other words, the problem indicates that we cannot fully extract most events. Therefore, efficient analytical methods are necessary for accurate extraction of events with position information, even in areas with few data. For this research, we conducted an experiment using information interpolation to estimate the times for biological season observation using tweets with Twitter location information. Then we evaluated its effectiveness. Herein, we explain results obtained using interpolated information and analysis of cherry blossoms in Japan in 2016.**

Keywords–trend estimation; phenological observation; Twitter.

I. INTRODUCTION

In recent years, because of wide dissemination and rapid performance improvement of various devices such as smart phones and tablets, diverse and vast data are generated on the web. Particularly, social networking services (SNSs) have become popular because users can post data and various messages easily. Twitter [1], an SNS that provides a micro-blogging service, is used as a real-time communication tool. Numerous tweets have been posted daily by vast numbers of users. Twitter is therefore a useful medium to obtain, from a large amount of information posted by many users, real-time information corresponding to the real world.

Here, we describe the provision of information to tourists using the web. Such information is useful for tourists, but providing timely and topical travel information entails high costs for information providers because they must update the information continually. Today, providing reliable information related to local travel is not only strongly demanded by tourists, but also by local governments, tourism organizations, and travel companies, which bear high costs of providing such information.

Therefore, providing current, useful, real-world information for travelers by ascertaining the change of information in accordance with the season and time zone of the tourism region is important for the travel industry. As described herein, we define "now" information as information that travelers require for tourism and disaster prevention, such as best flower-viewing times, festivals, and local heavy rains. As one might expect, the period estimated for disaster prevention information would be an estimate of the “worst” time instead of the best time.

We propose a method to estimate best-time viewing for phenological observations for tourists, such as cherry blossoms and autumn leaves, in each region by particularly addressing phenology observations that must be made “now” in the real world. Tourist information for the best time requires a peak period: the best time is not a period after or before the falling of flowers, but a period when one can view blooming flowers. Furthermore, such best times differ among regions and locations. Therefore, it is necessary to estimate a best time of phenological observation for each region and location. To estimate best-time viewing, one must collect large amounts of information with real-time properties. For this study, we use Twitter data obtained from many users throughout Japan.

However, to analyze the information of each region from Twitter data, it is necessary to specify the location from tweet information. Because geotagged tweets can identify places, they are effective for analysis, but because the proportion of geo-tagged tweets accounts for a very small proportion of the total information content of tweets, it is not possible to analyze all regions. Therefore, we propose an information interpolation method using geo-tagged tweets. We conducted experiments to estimate the position around areas not identified by location information.

The remainder of the paper is organized as follows. Section II presents earlier research related to this topic. In Section III, we propose a method for estimating the best time for phenotypic observations using information interpolation.

 ISBN: 978-1-61208-548-7

Section IV describes experimentally obtained results for our proposed method and a discussion of the results. Section V presents a summary of the contributions and future work.

II. RELATED WORK

The amounts of digital data are expected to increase greatly in the future because of the spread of SNSs. Reports describing studies of the effective use of these large amounts of digital data are numerous. Some studies use microblogs to conduct real-world understanding and prediction by analyzing information transmitted from microblogs. Kleinberg [2] detected a "burst" of keywords signaling a rapid increase in time-series data. Sakaki et al. [3] proposed a method to detect events such as earthquakes and typhoons based on a study estimating real-time events from Twitter. Kaneko et al. [4] proposed a method of detecting an event using geotagged non-photo tweets and non-geotagged photo tweets, as well as geotagged photo tweets. Yamagata et al. [5] proposed a real-time urban climate monitoring method using geographically tagged tweets, demonstrating the effectiveness of tweets for urban risk management. Consequently, various methods for extracting event and location information are discussed. However, event detection has been done in earlier studies, but discussion of the validity period of the event has not been reported. As described in this paper, we propose a method for extracting such information. Then we estimate “now” in relation to tourism information, such as the full bloom period of phonological observations. Additionally, we treat a crucially important difficulty related to analysis of geotagged contents: what amount of data is effective for an analyzed area.

III. OUR PROPOSED METHOD

This section presents a description of an analytical method for target data collection and presents our best-time estimation to obtain a guide for phenological change from Twitter in Japan. Our proposal is portrayed in Figure 1.

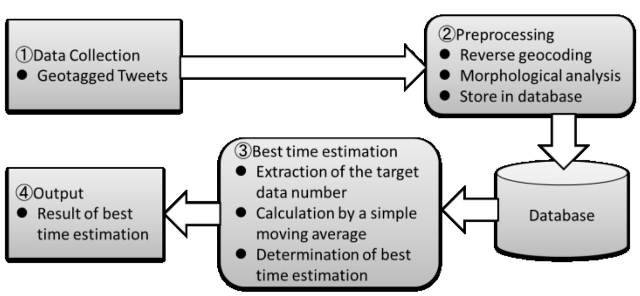

Figure 1. Our proposal.

A. Data collection

This section presents a description of the Method of (1) data collection presented in Figure 1. Geotagged tweets sent from Twitter are a collection target. The range of geo-tagged tweets includes the Japanese archipelago (120.0°E ≤ longitude ≤ 154.0°E and 20.0°N ≤ latitude ≤ 47.0°N) as the collection target. Collection of these data was done using a streaming API [6] provided by Twitter Inc.

Next, we describe the number of collected data. According to a report described by Hashimoto et al. [7], among all tweets originating in Japan, about 0.18% are geotagged tweets: they are rare among all data. However, the collected geo-tagged tweets, shown as an example in Table I, number about 70,000, even on weekdays. On some days during weekends, more than 100,000 such messages are posted. We use about 30 million geo-tagged tweets from 2015/2/17 through 2016/4/30. For each day of collection, the number during the period covered was about 72,000. We calculated the best time for flower viewing, as estimated by processing the following sections using these data.

TABLE I. TRANSITION OF GEOTAGGED TWEETS (2015/5/9 – 6/3)

Date (Day of the week)	Volume [tweet]	Date (Day of the week)	Volume [tweet]
5/9 (Sat)	117,253	5/22 (Fri)	92,237
5/10 (Sun)	128,654	5/23 (Sat)	55,590
5/11 (Mon)	91,795	5/24 (Sun)	72,243
5/12 (Tue)	87,354	5/25 (Mon)	82,375
5/13 (Wed)	67,016	5/26 (Tue)	83,851
5/14 (Thu)	88,994	5/27 (Wed)	83,825
5/15 (Fri)	89,210	5/28 (Thu)	85,024
5/16 (Sat)	116,600	5/29 (Fri)	121,582
5/17 (Sun)	126,705	5/30 (Sat)	119,387
5/18 (Mon)	89,342	5/31 (Sun)	81,431
5/19 (Tue)	83,695	6/1 (Mon)	76,364
5/20 (Wed)	87,927	6/2 (Tue)	76,699
5/21 (Thu)	86,164	6/3 (Wed)	78,329

B. Preprocessing

This section presents a description of the method of (2) preprocessing presented in Figure 1. Preprocessing includes reverse geocoding and morphological analysis, with database storage for data collected using the process explained in Section III.A.

Reverse geocoding identified prefectures and municipalities by town name using latitude and longitude information from the individually collected tweets. We use a simple reverse geocoding service [8] available from the National Agriculture and Food Research Organization in this process: e.g., (latitude, longitude) = (35.7384446N, 139.460910W) by reverse geocoding becomes (Tokyo, Kodaira City, Ogawanishi-cho 2-chome). Furthermore, based on latitude and longitude information of collected tweets, data are accumulated for each division of land using tertiary mesh data provided by the Land Numerical Information download service of the Ministry of Land, Infrastructure, and Transport [9]. The tertiary mesh is a section of about 1 km square.

Morphological analysis divides the collected geo-tagged tweet morphemes. We use the “Mecab” morphological analyzer [10]. As an example, “桜は美しいです” (“Cherry blossoms are beautiful.” in English) is divisible into "(桜 / noun), (は / particle), (美しい / adjective), (です / auxiliary verb), (。 / symbol)".

Preprocessing performs the necessary data storage for the best-time viewing, as estimated based on results of the processing of data collection, reverse geocoding, and morphological analysis. Data used for this study were the tweet ID, tweet post time, tweet text, morphological analysis result, latitude, and longitude.

C. *Estimating the best-time viewing*

This section explains the method of (3) best-time estimation presented in Figure 1. In our method of estimating best-time viewing, we first process the target number of extracted data and then calculate a simple moving average, yielding an inference of the best flower-viewing time. The method defines a word related to best-time viewing, estimated as the target word. The target word includes Chinese characters, hiragana, and katakana, which represents an organism name and seasonal change, as shown in Table II.

TABLE II. TARGET WORD EXAMPLES

Items	Target Words	In English
さくら	桜, さくら, サクラ	Cherry blossoms
かえで	楓, かえで, カエデ	Maple
いちょう	銀杏, いちょう, イチョウ	Ginkgo
こうよう	紅葉, 黄葉, こうよう, もみじ, コウヨウ, モミジ	Autumn leaves

Next, the granularity for estimation is shown. For an estimate for Japan as a whole, prefecture units are assumed and acquired by reverse geocoding. However, when conducting more detailed analyses, a difficulty arises: it is impossible to estimate the number of geotagged tweets for each city or town or village or tourist spot. Therefore, we attempted estimation through information interpolation using data aggregated for each section of tertiary mesh data. For interpolation, we used Kriging [11], an estimation method used for estimating values for points where information was not acquired, to ascertain the distribution of the information in the whole space in geostatistics. The estimated value of the target data at a certain point S_0 is represented in formula (1) as a weighted average of the measured values $Z(S_i)$ (i = 1, 2..., N) at N points S_i existing around point S_0. As described in this paper, we experimentally assigned a +1 weight for 'full bloom' and 'beautiful', and assigned -1 on 'still' or 'falling'. Then we assigned value Z to tweets including the target word and Z. N denotes the 30 nearby targeted tweets. λ has adopted a spherical model that decreases the influence as the distance increases.

$$\hat{z}(S_0) = \sum_{i=1}^{N} \lambda_i \, Z(S_i) \tag{1}$$

$Z(S_i)$: Measurement value at i-th position
λ_i: Unknown weighting of measured value at i-th position
S_0 : Predicted position
N : Number of measurements

Next, we describe the simple moving average calculation, which uses a moving average of the standard of the best-time viewing judgment. It calculates a simple moving average using aggregate data on a daily basis by the target number of data extraction described above. Figure 2 shows an overview of the simple moving average of the number of days.

We calculate the simple moving average in formula (2) using the number of data going back to the past from the day before the estimated date of the best-time viewing.

Standard lengths of time we used for the simple moving averages were seven days and one year. A 7-day moving

$$X(Y) = \frac{P_1 + P_2 + \cdots + P_Y}{Y} \tag{2}$$

$X(Y)$:Y day moving average
P_n:Number of data of n days ago
Y: Calculation target period

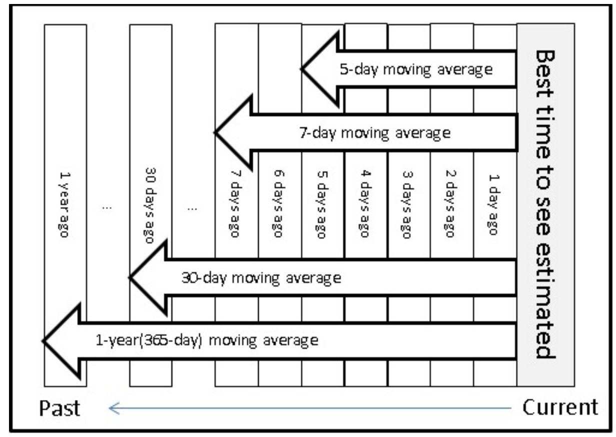

Figure 2. Number of days simple moving average.

average has one week as the criterion of the estimated period of full bloom because, as shown in Table I, a tendency exists for a transition of geotagged tweets of the increases on weekends compared to weekdays. In addition, phenological observations are based on the moving average of best-time viewing estimated in prior years because many such “viewing” events occur every year: cherry blossom viewing, autumn leaf viewing, and even moon viewing.

A simple moving average of the number of days is described for each type of event to compare the 7-day moving average and the one-year moving average. In this study, the period of the best-time viewing depends on the specified type of event, the individual event, and the number of days during the biological period related to the event.

As an example, we describe cherry blossoms. The Japan Meteorological Agency [12] carries out phenological observations of "Sakura," which yields two output items of the flowering date and the full bloom date observation target. "Sakura flowering date" [13] is the first day of blooming 5–6 or more wheels of flowers of a specimen tree. "Sakura in full bloom date" is the first day of a state in which about 80% or more of the buds are open in the specimen tree. In addition, "Sakura" is the number of days from general flowering until full bloom: about 5 days. Therefore, "Sakura" in this study uses a standard 5-day moving average.

Next, we describe an estimated judgment of best-time viewing, which was calculated using the simple moving average (7-day moving average, 1-year moving average, and another biological moving average). It specifies the two conditions as a condition of an estimated decision for best-time viewing. Condition 1 is the number of data one day before expression. Formula 3 is a simple moving average greater than that of the estimated best-time viewing date.

 ISBN: 978-1-61208-548-7

Condition 2 is a case that follows formulas 4 ((A) / (2)) or more.

$$P_1 \geqq X(365) \tag{3}$$
$$X(A) \geqq X(B) \tag{4}$$

Finally, an estimate is produced using conditions 1 and 2. By the proposed method, a day that satisfies both condition 1 and condition 2 is estimated as best-time viewing.

D. Output

This section presents a description of the method of (4) output presented in Figure 1. Output can be visualized using a best-time viewing result, as estimated by processing explained in the previous section. A time-series graph presents the results inferred for best-time viewing. The graph presents the number of data and the date, respectively, on the vertical axis to the horizontal axis. We are striving to develop useful visualization techniques for travelers.

IV. EXPERIMENTS

In this section, we describe an estimation experiment of best-time viewing for cherry blossoms using the method proposed in Section III.

A. Dataset

Datasets used for this experiment were collected using streaming API, as described for data collection in Section 3.1. Data are geo-tagged tweets from Japan during 2015/2/17 – 2016/4/30. The data include about 27 million items. We are using these datasets for experiments to infer the best time for cherry blossom viewing in 2016.

B. Estimation experiment for best-time viewing of cherry blossoms

The estimation experiment to ascertain the best-time viewing of cherry blossoms uses the target word in Table II: "Sakura". The target word is “cherry blossom,” which is “桜” and “さくら” and “サクラ” in Japanese. The subject of the experiment was set as tourist spots in Tokyo. In this report, we describe “Takao mountain,” “Showa Memorial Park," “Shinjuku gyoen,” and “Rikugien.”

Figure 3 presents the target area location. A, B, C and D in the figure respectively denote “Takao mountain,” “Showa Memorial Park,” “Rikugien,” and “Shinjuku Gyoen.” A and B are separated by about 16 km straight line distance. B and C are about 32 km apart. C and D are about 6 km apart.

The following two experiments were conducted. The first is an experiment using the number of tweets including the target word and the sightseeing spot name without information interpolation. This is Experiment 1 described in this paper. The second is an experiment using information interpolation for tertiary mesh including sightseeing spots. In the second experiment, the numerical value obtained by summing the result of the information interpolation and the first experiment result was used for observation estimation. This is Experiment 2 in this paper.

Figure 3. Position of target area.

C. Experimental result

We can present experimentally obtained results from tweets including a target word and a tourist spot name. Figure 4 shows those results for the estimated best-time viewing in 2016 using the target word ‘cherry blossoms’ in the target tourist spots of Figure 3. The dark gray bar in the figure represents the number of tweets. The light gray part represents best-time viewing as determined using the proposed method. In addition, the solid line shows a 5-day moving average. The dashed line shows a 7-day moving average. The dotted line shows a 1-year moving average.

At tourist spots targeted for the experiment in 2016, as portrayed in Figure 4, much data was obtained in C and D. The maximum number of tweets per day was about 30. For this reason, it was confirmed that some estimation can be done by near-sight estimation method without interpolation. However, best-time viewing cannot be done in A and B because of the very small number of tweets.

Next, Figure 5 portrays an experimentally obtained result from interpolation results for a tertiary mesh including tourist spots we examined. The notation is the same as that in Figure 4.

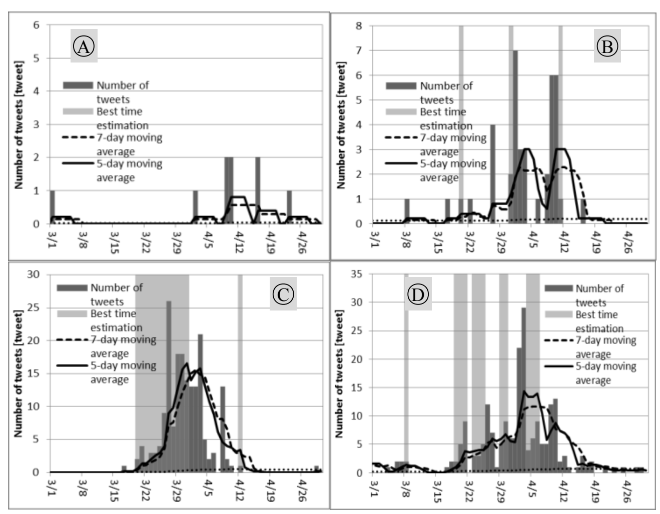

Figure 4. Experimental results obtained using tweets including the target word and the tourist spot name.

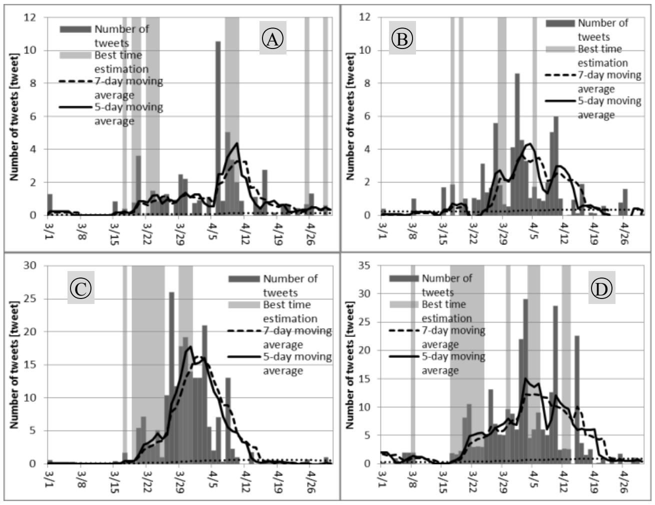

Figure 5. Experimental results obtained using interpolation.

Apparently, A and B were able to produce an estimate using the proposed method by increasing the number of tweets using information interpolation with surrounding tweets. In C and D, there are days when it can be determined more accurately by interpolating the number of tweets. However, because there were tweets of negative judgments such as "still" or "scattered" among surrounding tweets, in some cases, interpolation excluded the day determined as the best time in Experiment 1. Therefore, the judgment condition of the tweet is subject to further study.

These results confirmed the possibility of estimating the peak period, even in an area without tweets, using data interpolation and overall tweet number interpolation.

Table III presents results of the optimal time for viewing in 2016, as estimated using the proposed method. Experiment 1 used co-occurring words in tweets including the sightseeing spot name coexisting with the target word "Sakura." Experiment 2 used information interpolation on a tertiary mesh including sightseeing spots. The numerical values in the table are the numbers of tweets including the target word and co-occurrence word in Experiment 1. Experiment 2 uses the sum of the number of tweets in Experiment 1 plus numerical values by interpolation. The light gray area indicates the date when the fullness prediction was made using the proposed method.

Confirming the flowering day and full bloom period of each sightseeing spot using JMA data is difficult, but this experiment to evaluate SNS data for flowering is valid also for weather forecasting companies [14] and public service organizations [15] to evaluate optimum times for viewing based on services and blogs that are used. Arrows indicating the flowering time can be checked manually at tourist sites.

Experimental results confirmed the tendency by which the relevance ratio and the recall rate become higher for tourist spots with few tweets. Therefore, we presented the possibility of estimating sightseeing sites with few tweets using information interpolation. However, because the interpolation information amount is insufficient in the current method, it is necessary to improve the information interpolation method further. In addition, sightseeing spots with many tweets are affected by tweets of minus judgments in the surroundings, so accuracy is lower than in Experiment 1. However, one might be able to estimate more details, such as the start time, using interpolation.

TABLE III. ESTIMATION RESULTS AT RESPECTIVE SIGHTSEEING SPOTS

	Takao mountain		Showa Memorial park		Rikugien		Shinjuku gyoen	
	Exp.1	Exp.2	Exp.1	Exp.2	Exp.1	Exp.2	Exp.1	Exp.2
3/1	1	1.27	0	0.40	0	0.58	0	0.28
3/2	0	0.00	0	0.00	0	0.00	0	0.00
3/3	0	0.00	0	0.00	0	0.00	0	0.00
3/4	0	0.00	0	0.00	0	0.00	1	1.00
3/5	0	0.00	0	0.00	0	0.00	0	0.00
3/6	0	0.00	0	0.00	0	0.00	2	2.00
3/7	0	0.00	0	0.00	0	0.00	2	2.00
3/8	0	0.00	1	1.00	0	0.00	2	2
3/9	0	0.00	0	0.00	0	0.00	0	0.00
3/10	0	0.00	0	0.00	0	0.00	0	0.00
3/11	0	0.00	0	0.00	0	0.00	0	0.00
3/12	0	0.00	0	0.00	0	0.00	0	0.00
3/13	0	0.00	0	0.00	0	0.00	0	0.00
3/14	0	0.00	0	0.00	0	0.00	0	0.00
3/15	0	0.84	0	1.68	0	0.41	0	0.24
3/16	0	0.00	0	0.00	0	0.00	0	0.00
3/17	0	0.40	1	1.85	1	1.67	1	1.85
3/18	0	−0.51	0	−0.43	0	0.02	2	1.64
3/19	0	0.78	0	0.49	0	0.22	2	2.30
3/20	0	3.61	1	1.00	2	5.42	5	8.18
3/21	0	0.00	0	−4.81	4	7.16	9	10.49
3/22	0	0.14	1	1.06	1	2.93	0	2.97
3/23	0	1.51	0	0.93	3	2.70	3	3.06
3/24	0	0.84	0	3.13	3	5.11	3	3.00
3/25	0	1.06	0	1.40	4	1.00	5	5.00
3/26	0	1.27	0	1.67	9	10.41	12	13.14
3/27	0	−0.08	4	5.57	26	26.00	7	7.00
3/28	0	0.94	0	1.85	7	11.74	1	5.19
3/29	0	2.50	0	0.66	18	17.83	5	5.00
3/30	0	2.21	0	0.52	18	19.18	9	9.62
3/31	0	0.00	2	4.13	14	14.00	6	8.82
4/1	0	0.74	7	8.60	13	13.00	6	6.00
4/2	1	0.91	3	4.56	13	13.00	22	22.00
4/3	0	0.00	3	3.30	21	21.00	29	29.00
4/4	0	1.12	0	1.05	5	5.62	4	4.55
4/5	0	0.00	0	1.73	2	2.00	6	6.00
4/6	0	10.52	1	1.00	3	7.05	9	9.00
4/7	0	0.89	0	0.88	0	0.33	5	6.06
4/8	0	5.05	2	2.00	13	13.00	5	5.00
4/9	2	3.37	6	5.05	2	2.29	12	12.62
4/10	2	2.00	6	6.00	1	1.00	13	27.88
4/11	0	0.88	1	1.00	0	0.00	2	2.47
4/12	0	0.00	0	0.00	1	−0.24	3	2.61
4/13	0	−0.50	0	0.02	0	1.79	1	2.55
4/14	0	1.11	0	0.95	0	−0.67	0	−0.37
4/15	0	0.51	0	0.12	0	0.00	1	22.54
4/16	2	2.75	1	1.91	0	0.60	3	3.63
4/17	0	0.00	0	0.00	0	0.42	1	1.54
4/18	0	0.14	0	0.17	0	0.36	2	2.51
4/19	0	−0.34	0	0.13	0	0.30	0	0.07
4/20	0	0.66	0	−0.43	0	−0.20	0	0.21
4/21	0	0.58	0	0.58	0	0.00	1	1.00
4/22	0	0.00	0	0.00	0	0.00	0	0.00
4/23	1	0.43	0	−4.08	0	0.00	1	1.19
4/24	0	0.00	0	0.00	0	0.00	0	0.00
4/25	0	0.68	0	0.76	0	0.64	1	1.75
4/26	0	1.33	0	1.61	0	−1.03	0	−0.26
4/27	0	0.00	0	0.00	0	0.00	0	0.00
4/28	0	−0.25	0	−0.34	0	0.19	1	1.11
4/29	0	0.60	0	0.00	1	1.00	1	1.00
4/30	0	0.00	0	0.09	0	0.00	0	0.00
Precision	0.26	0.38	0.72	0.75	0.89	0.89	0.84	0.75
Recall	0.00	0.20	0.11	0.22	0.67	0.61	0.56	0.50

V. CONCLUSION

As described herein, to improve best-time estimation accuracy and thereby enhance tourist information related to phenologic observation, we proposed an information interpolation method. For the proposed method, information was interpolated using neighbor-weighted tweets on a tertiary mesh including sightseeing spots, thereby indicating the optimum time to view flowers at sightseeing spots.

The results of cherry blossom experiments at sightseeing spots in Tokyo in 2016 confirm the tendency for improvement of accuracy of estimation by information interpolation. The proposed method using information interpolation for tweets related to organism names might improve the accuracy of estimating the best time in the real world. We confirmed the possibility of applying this proposed method to estimation of the viewpoint and line of sight in areas and sightseeing spots with few tweets and little location information. However, because the experimental case was related only to cherry blossoms, it is necessary to verify other cases as well.

Future research with manual experimental weighting and geotagged tweets will facilitate further improvements to overcome insufficiencies in measured values used for interpolation. Additionally, we expect to reconsider the viewing angle estimation conditions. Eventually, this system might be extended to a system by which travelers can obtain travel-destination-related event information and disaster information in real time.

ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI Grant Nos. 16K00157 and 16K16158, and by a Tokyo Metropolitan University Grant-in-Aid for Research on Priority Areas “Research on Social Big Data.”

REFERENCES

[1] Twitter. *It's what's happening*. [Online]. Available from: https://Twitter.com/ [retrieved: 2, 2015]

[2] J. Kleinberg, “Bursty and hierarchical structure in stream,” In Proc. of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.1–25, 2002.

[3] T. Sakaki, M. Okazaki, and Y. Matsuo, “Earthquake shakes Twitter users: real-time event detection by social sensors,” WW W 2010, pp.851–860, 2010.

[4] T. Kaneko and K. Yanai, “Visual Event Mining from the Twitter Stream,” WWW '16 Companion Proceedings of the 25th International Conference Companion on World Wide Web, pp.51–52, 2016.

[5] Y. Yamagata, D. Murakami, G. W. Peters, and T. Matsui, “A spatiotemporal analysis of participatory sensing data “tweets” and extreme climate events toward real-time urban risk management,” arXiv preprint arXiv:1505.06188, pp.1–34, 2015.

[6] Twitter Developers. *Twitter Developer official site*. [Online] Available from: https://dev.twitter.com/ [retrieved: 2, 2015]

[7] Y. Hashimoto and M. Oka, “Statistics of Geo-Tagged Tweets in Urban Areas (<Special Issue>Synthesis and Analysis of Massive Data Flow),” JSAI, vol. 27, No. 4, pp.424–431, 2012 (in Japanese).

[8] National Agriculture and Food Research Organization. *Simple reverse geocoding service*. [Online]. Available from: http://www.finds.jp/wsdocs/rgeocode/index.html.ja [retrieved: 4, 2015]

[9] Ministry of Land, Infrastructure and Transport. *Land Numerical Information download service*. [Online]. Available from: http://nlftp.mlit.go.jp/ksj-e/index.html [retrieved: 4, 2015]

[10] MeCab. *Yet Another Part-of-Speech and Morphological Analyzer*. [Online]. Available from: http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html [retrieved: 4, 2015]

[11] M. A. Oliver, “Kriging: A Method of Interpolation for Geographical Information Systems.,” International Journal of Geographic Information Systems 4, pp.313–332, 1990.

[12] Japan Meteorological Agency. *Disaster prevention information XML format providing information page*. [Online]. Available from: http://xml.kishou.go.jp/ [retrieved: 4, 2015]

[13] Japan Meteorological Agency. *Observation of Sakura*. [Online]. Available from: http://www.data.jma.go.jp/sakura/data/sakura2012.pdf [retrieved: 4, 2015]

[14] Weathernews Inc. *Sakura information*. [Online]. Available from: http://weathernews.jp/sakura [retrieved: 4, 2015]

[15] Japan Travel and Tourism Association. *Whole country cherry trees*. [Online]. Available from: http://sakura.nihon-kankou.or.jp [retrieved: 4, 2015]