Co-occurring Word Determination Used for Estimating Best Times for Viewing Cherry Blossoms

Yusuke Takamori*, Kenji Terada, Masaki Endo, Shigeyoshi Ohno

Polytechnic University

Hiroshi Ishikawa

Tokyo Metropolitan University

Outline

1. Introduction
2. Proposed Method
3. Conclusion
4. References

1. Introduction

1. Introduction

- **Biological seasonal observations**
  - Indicators of seasonal changes
  - Used also in the tourism industry
  - Decreased numbers of observation targets

Alternative indicators for bioseasonal observations are needed.

Introduction

Earlier research

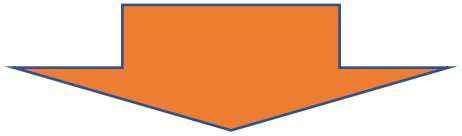
- Proposed method using Twitter for estimating the best time to view cherry blossoms [1]

- Estimation of cherry blossom viewing period for a certain period of time in the future [2]

Introduction

Prediction with **small amounts of data**

Decreased forecast accuracy occurs



Proposed method to increase data volume with **co-occurring words**

Introduction

Prediction with **small amounts of data**

Decreased forecast accuracy occurs

Proposed method to increase data volume with **co-occurring words**

- **Improved accuracy for estimating the best time for viewing**
- **Estimation of the viewing period for a certain period of time in the future**

2. Proposed Method

2. Proposed Method

1. Extract candidate co-occurring words
2. Determine co-occurring words
3. Collect tweets using co-occurring words
4. Use the collected tweets to estimate the best viewing times
5. Combine with results obtained using prior methods

Proposed Method

Extract candidate co-occurring words

Original data: Tweets containing the keyword "Sakura" in the text

Collection period: February 1, 2015 -- May 31, 2022 Extraction

Method: Morphological analysis using MeCab

Conditions: Must be in the top 1% of frequently appearing keywords;
Parts of speech are nouns, shape verbs or verbs

Proposed Method

Determine Co-occurring words

Period used: January 1, 2018 -- December 31, 2018

Co-occurrence criteria

- $K > 2$
- $(Ss - 1) \leq S \leq (Ss + 1)$

However, the skewness kurtosis values of the words are denoted respectively by *S* and *K*; the cherry skewness is denoted by *Ss*.

Proposed Method

Collect tweets using co-occurring words

Collection period: February 1, 2015 -- June 30, 2022

Collection criteria: one determined co-occurrence word is included in the text

Proposed Method

Use the collected tweets to estimate best viewing times

Estimated best viewing period: March 1, 2022 -- April 30, 2022

Criteria for judging the best viewing period

- $x_i > \text{Avg } 365$
- $\text{Avg } 10 < \text{Avg } 20$

However, x_i denotes the number of tweets on day *i* is for finding the best time for viewing, Avg *Y* stands for the *Y*-day simple moving average.

Proposed Method

Combine with results obtained using prior methods

- Results obtained using the prior method and the time period estimated here differ.

- The period when either of the two methods is estimated as the best time to visit is regarded as the best time for viewing.

3. Results

3. Results

Example of a co-occurrence judgment result

cherry blossom, Sakura, Ueno Park,

Yasukuni [shrine name], Someiyoshi [Sakura variety name]

Results

Example of a co-occurrence judgment result

cherry blossom, Sakura, Ueno Park,
Yasukuni [shrine name], Someiyoshi [Sakura variety name]

Words indicating cherry blossoms and the names of their places of interest can be extracted.

Results

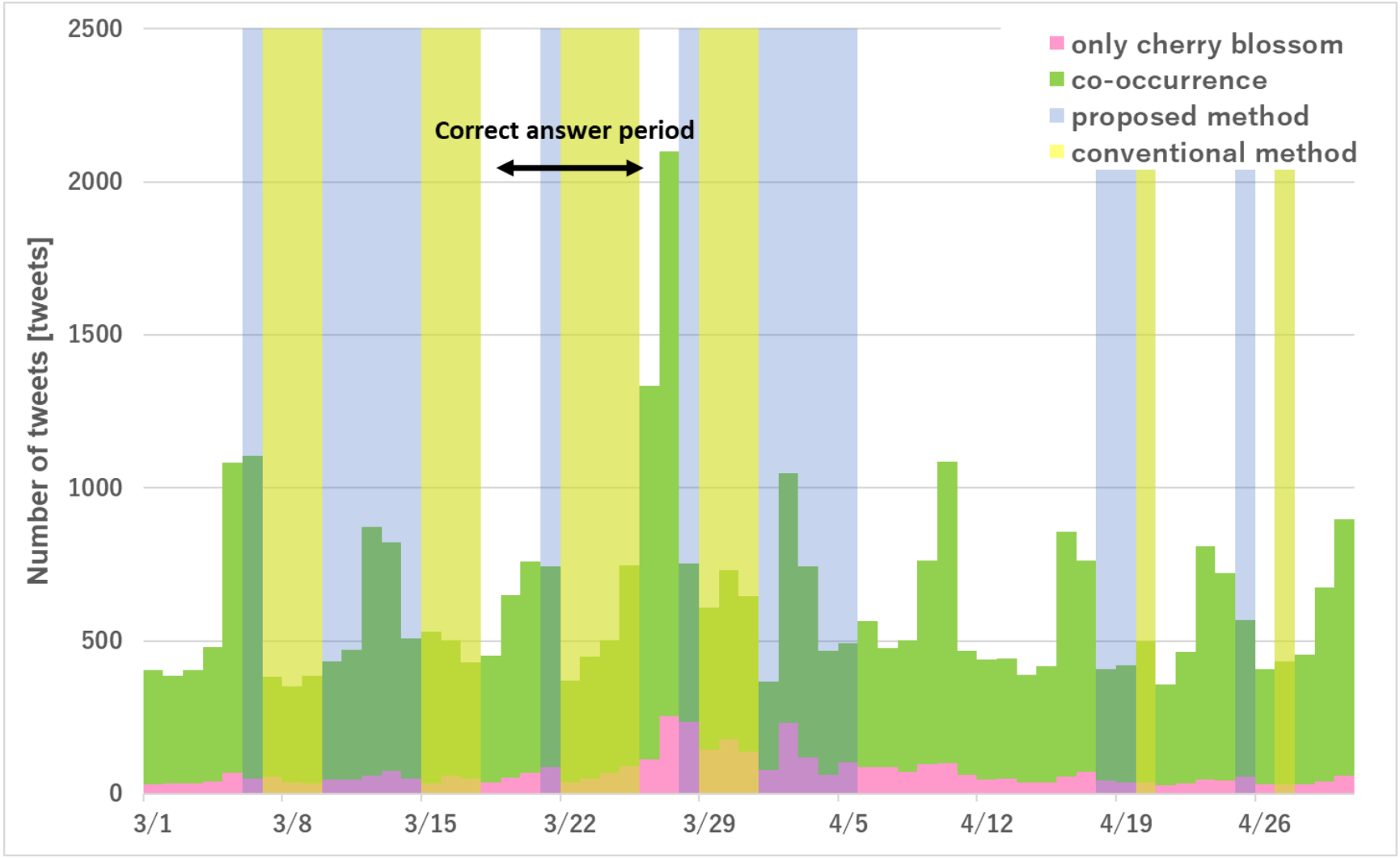


Figure 1 Estimated results of best time for viewing (Tokyo).

Results

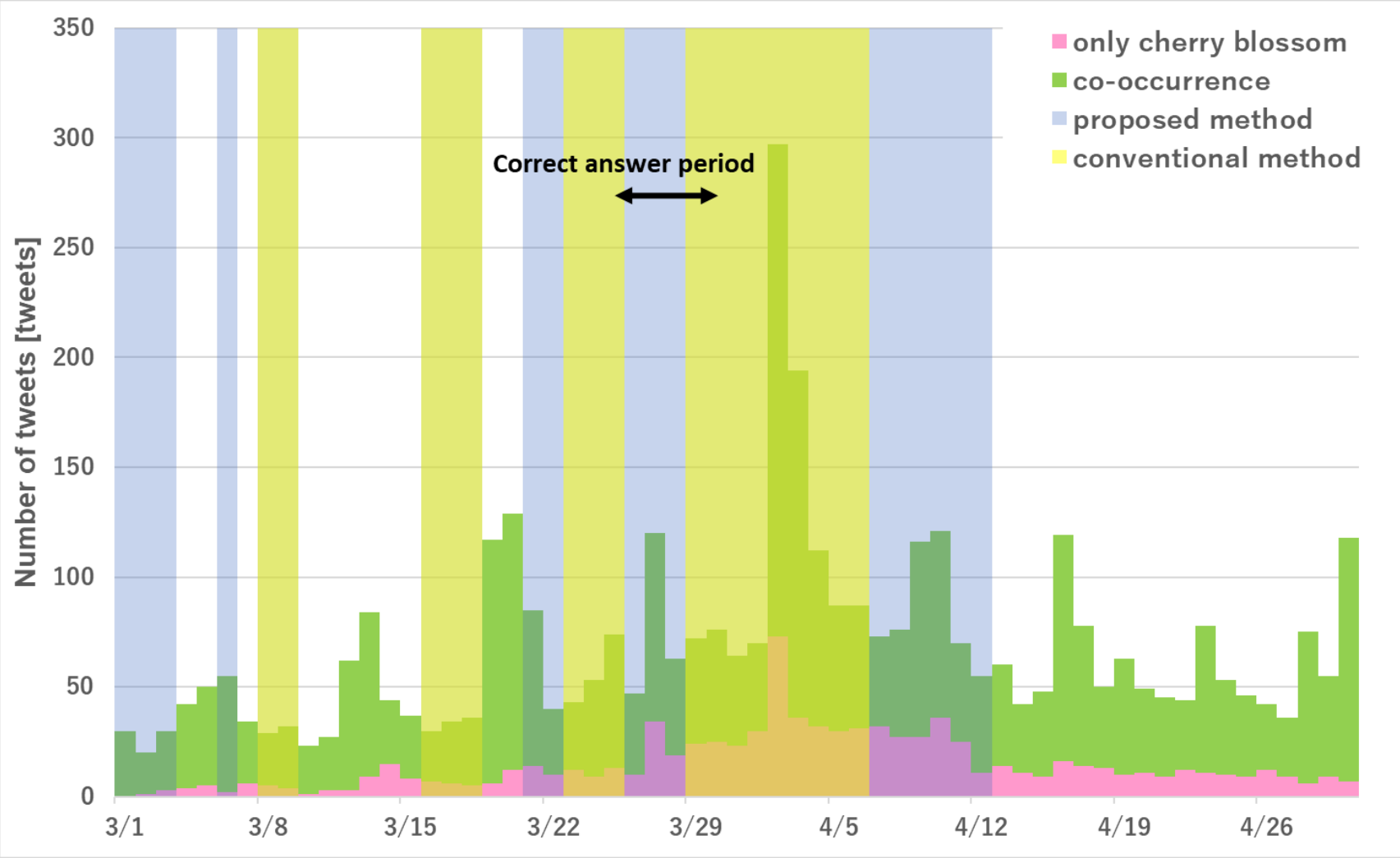


Figure 2 Estimated results of best time for viewing (Kyoto).

Results

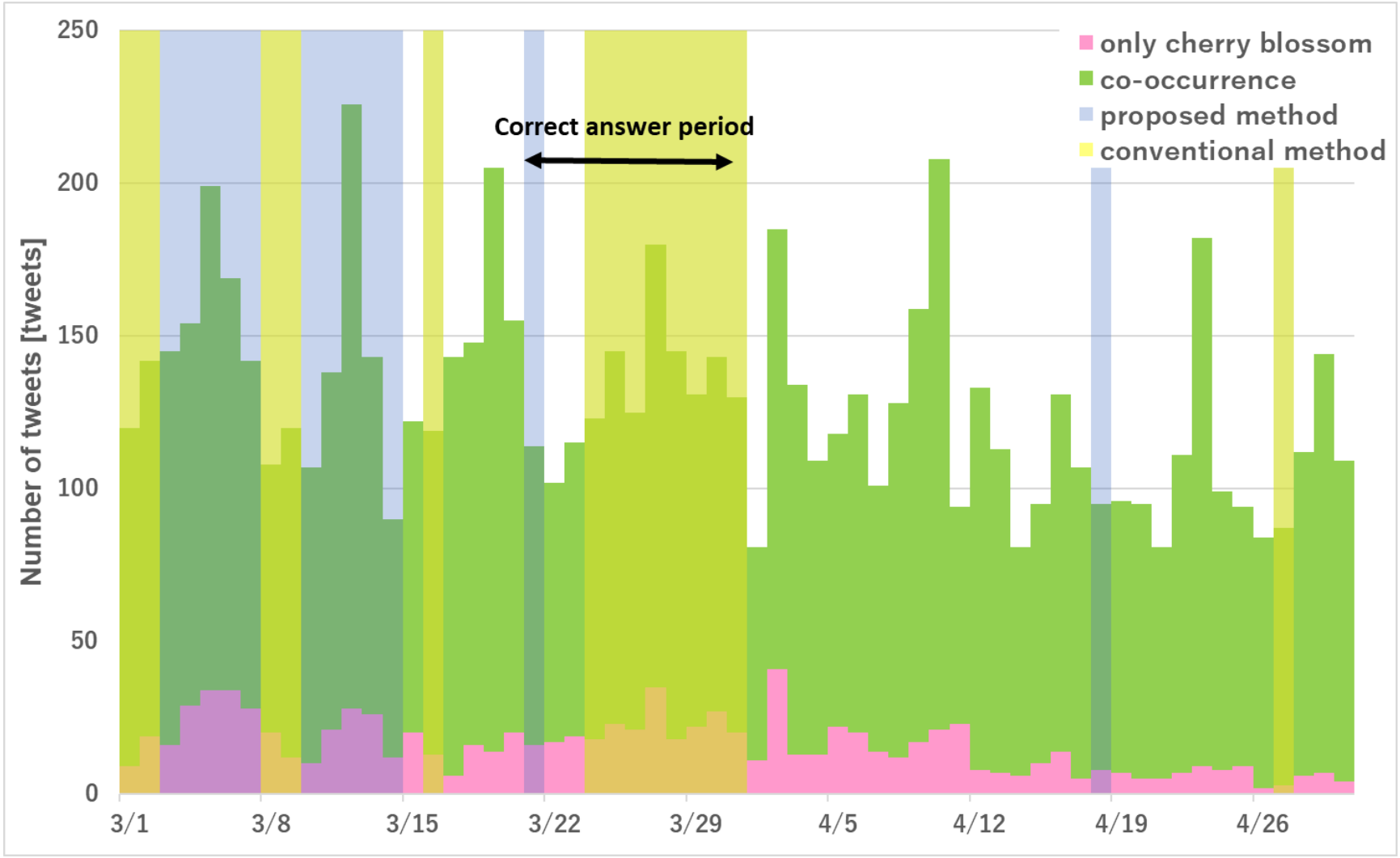


Figure 3 Estimated results of best time for viewing (Shizuoka).

Results

Table 1　Results of evaluation of the estimation of the best time for viewing

Method	Prefecture	Recall (%)	Precision (%)
Conventional	Tokyo	50.0	26.7
Proposed		52.5	16.1
Conventional	Kyoto	57.1	23.5
Proposed		100.0	21.9
Conventional	Shizuoka	70.0	50.0
Proposed		80.0	30.8

Results

Table1　Results of evaluation of the estimation of the best time for viewing

Method	Prefecture	Recall (%)	Precision (%)
Conventional	Tokyo	50.0	26.7
Proposed		52.5	16.1
Conventional	Kyoto	57.1	23.5
Proposed		100.0	21.9
Conventional	Shizuoka	70.0	50.0
Proposed		80.0	30.8

Results

Table1　Results of evaluation of the estimation of the best time for viewing

Method	Prefecture	Recall (%)	Precision (%)
Conventional	Tokyo	50.0	26.7
Proposed		52.5	16.1
Conventional	Kyoto	57.1	23.5
Proposed		100.0	21.9
Conventional	Shizuoka	70.0	50.0
Proposed		80.0	30.8

4. Conclusion

4. Conclusion

- Problem: small amounts of data in earlier studies
- Proposal: Tweet collection method using co-occurrence word determination
- Result: Recall rate increased by 18.5% on average
  However, the average precision rate dropped by 10.5%
- Conclusion: Increased amount of data, improved reproducibility

Conclusion

Future work

- Consideration of a method to estimate the best time to view cherry blossoms without lowering the precision rate

- Further improvement of accuracy by eliminating tweets that are unrelated to cherry blossoms

5. References

1. M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, “Automatic detection of tourist spots and best-time estimation using social network services,” International Workshop on Informatics 2020, pp.65-72, 2020.
2. T. Horikawa, M. Takahashi, M. Endo, S. Ohno, M. Hirota, and H. Ishikawa, “Estimating the best time to see cherry blossoms using SNS and time-series forecasting of tweet numbers using machine learning,” International Workshop on Informatics 2021, pp.37-44, 2021.