Journal of Information Processing Vol.26 386–395 (May 2018)

[DOI: 10.2197/ipsjjip.26.386]

Regular Paper

Modeling Weather Context Dependent Food Choice Process

Taku Ito[1] Yusuke Fukazawa[1,a)] Dandan Zhu[2] Jun Ota[1]

Received: August 17, 2017, Accepted: February 1, 2018

Abstract: In this paper, we investigate the impact of weather context on the process of choosing foods. We mine social media on the web to create datasets on food choice, and associate food selections with weather context gathered from governmental meteorological information. From the dataset, we find that not only weather but also food events or social events on special days significantly impacts food choice. Accordingly, we propose a topic model that include the event class to represent the relationship between weather context and food choice. We quantitatively evaluated the model by perplexity, and discovered that considering both weather and event context improves prediction performance. Perplexity of the proposed model (weather and event context-aware topic model with separate topics) is (4663.0), which beats the benchmark model (4943.4). An analysis shows that combining contexts in the topic generation process yields better results that combining contexts in the word generation process. We also conduct a qualitative evaluation on the learned topic and associated foods.

Keywords: meteorological context, topic models, food analysis, Twitter

1. Introduction

The process of determining what foods to eat has been widely investigated [1], [2], [3], [4], [5], [6], [7], [8], [9], [10]. A number of factors are thought to influence people's dietary choices, including health, cost, convenience and taste. Falk et al. clustered these factors into five types: ideals, personal factors, resources, social factors, and context [1]. **Figure 1** show the food choice process model proposed in Ref. [1], [2]. Ideals represent normative forms of what and how one should eat. Ideals are culturally learned from families and other institutions, and reflect the plans and expectations for food and eating. Personal factors include physiological factors (genetic, etc.) and psychological characteristics (preferences, moods, etc.) Resources are assets available to people for making food choices such as money, equipment, transportation and space. Social factors are the relationships that influence food choices such as roles, families, organizations that provide opportunities and obligations for constructing eating relationships. Contexts include physical surroundings and behavior settings, social institutions, and seasonal and temporal climate.

Most of the factors have been widely investigated in the field of psychology. However, the impact of weather context has not been deeply investigated. This is because the amount of effort needed to acquire an adequate dataset of people's food choices under a variety of climate conditions from many people is thought to be excessive. In this paper, we focus on Twitter posts to gather the datasets needed. Twitter is one of the most popular services and allows users to post their status using short sentences. As most user posts focus on daily activities, the tweets include eating activity. In addition, we can collect meteorological information from weather services. By associating tweet contents that include eating activities with weather data, we can obtain rich datasets from which the relationship between weather and food choice can be extracted.

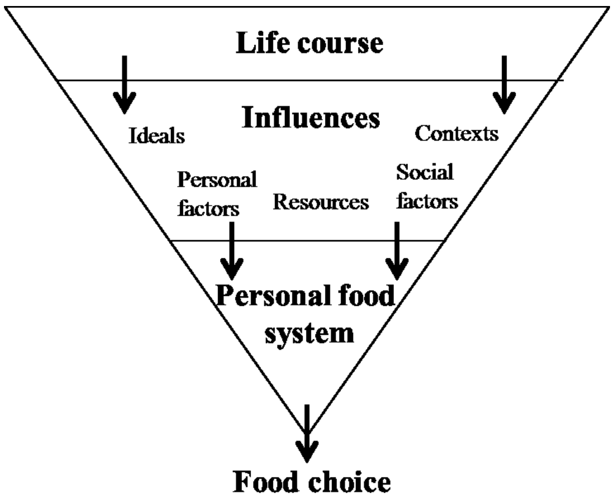


Fig. 1 A food choice process model.

In order to investigate the relationship between documents and contexts, many topic models have been proposed. Nowadays, context-aware topic models are being proposed such as location, time, and companions. Unfortunately, weather-aware topic models have yet to be proposed. Hence, the purpose of this paper is to reveal the relationships among weather context, topics, and foods posted on Twitter.

Contributions of our paper are the following three points:

1 Research into Artifacts, Center for Engineering (RACE), The University of Tokyo, Kashiwa, Chiba 277–8568, Japan
2 China University of Petroleum, Shangdong, China
a) yusuke.fukazawa@gmail.com

- We mine social media on the web to create datasets on food choice, and associate food selections with weather context gathered from govermental meteorological information.
- We show that food events or customs on a special day have a significant impact on the characteristics of the food choice on the datasets mined from social media.
- We propose a topic model that includes event class to represent the relationship between weather context and food choice. We quantitatively evaluated the model, and discuss the learned relationship between food choice and weather context.

2. Related Works

2.1 Food Choice Process

The team of Bisogni and Sobal is active in developing the food choice process. Falk, L.W. et al. conducted in-depth interviews of 16 individuals to learn about how they chose foods. The results showed that social structure played an important role in the food choices of the interviewees [1]. T. Furst extended the model through the incorporation of value negotiations and behavioral strategies [3]. They interviewed 29 adults to examine the food choice process who were primarily individuals making grocery store food choice decisions. These people were asked about how they chose foods when shopping and in other settings, and what influenced their choices. Connors, M.M. et al. incorporated a personal food system in the food choice process [4]. They conducted in-depth qualitative interviews of subjects drawn from a diverse population of urban adults living in upstate New York. Bisogni, C.A. et al. investigated the influence of food management skills on food choice by the interview [5]. Shepherd, R. published a book that summarizes the findings about food choice in the field of psychology [2]. Sobal, J. et al. developed a unified food decision making process that incorporates existing multiple findings [6], [7].

Other than the team of Bisogni and Sobal, there are many interesting studies on food choice process. Oliver, G. et al. investigated experimentally whether acute stress alters food choice during a meal [8]. Sixty-eight healthy men and women volunteered for a study. They found that in a laboratory setting, emotional eaters under stress increased their consumption of sweet fatty foods. Holsten, J.E. investigated children's food choices in the home with particular attention to environmental influences [9]. They interviewed 11 to 14 year old children (n = 47) from one middle school. They found that children evaluated potential food options based on their hunger level, food preferences, time pressure and activity prioritization, food preparation effort and skills, and expected physical consequences of food. Huneke, T. et al. examined the effect of service employees' appearance on consumer food choices using an experimental study involving video manipulation and eye-tracking [10]. They found that exposure to an overweight employee did not stimulate greater (i.e., earlier or longer) attention to unhealthy meal alternatives, whereas exposure to the employee who displayed an unhealthy lifestyle did.

Most of the above research was conducted in the field of psychology. They constructed a model by analyzing the results of interviews in the laboratory; indeed, most psychological research has been done in the laboratory. Accordingly, the relationship between weather and food choice has not been deeply investigated.

2.2 Twitter and Weather

In this section, we discuss studies about the relationship between weather and Twitter. Chen et al. proposed an approach for predicting the time and location at which a specific type of crime would be most likely to occur by joint analysis of Twitter and weather data [11]. Demirbas et al. proposed a collaboration of Twitter and crowd-sourced sensing and designed a collaboration system for Twitter with crowd-sourced weather radar [12]. These studies showed that weather strongly affects the content of a Twitter post.

2.3 Topic Modeling for Relationship Analysis

In this section, we discuss studies on topic models. The most basic topic model is Latent Dirichlet Allocation (LDA) [13]. Because LDA introduces the prior probability distribution, it can overcome the problem of over-fitting to the learning data and can be applied to new documents. Several researchers have studied the application of LDA. Andrzejewski et al. proposed a method of incorporating domain knowledge into LDA to guide the recovery of latent topics [14]. AlSumait et al. proposed online LDA (OLDA), which can automatically capture the thematic patterns and identify emerging topics in text streams and their changes over time [15]. Ahmed et al. proposed topic models that can capture temporal streams and the distributions of time-evolving topics [16]. Krestel et al. proposed LDA-based tag recommendation systems for users searching for multimedia content [17]. Lau et al. proposed a method for automatically labeling topics learned via LDA and generated a label candidate set from top-ranking topic terms and Wikipedia titles [18]. Ding et al. used LDA to propose a topical translation model for microblog hashtag suggestion [19]. Wang et al. proposed an extension of LDA that considers the word order and phrases to capturing the meaning of text in many texts [20]. Chen et al. designed and implemented a solution to behavioral targeting using the Hadoop MapReduce framework and built more than 450 behavioral targeting category models from all of Yahoo's users [21]. Ahmed et al. proposed a method for capturing users' profile changes for improved prediction and recommendation performance [22]. These studies show that LDA is suitable for analyzing the background elements of the document generation process in social network services. However, LDA does not consider context in any form.

2.4 Context-aware Topic Modeling

Some researchers have proposed context-aware topic models that include location, time, and companions. Eisenstein et al. proposed a location context-aware topic model that can estimate spatially distributed latent topic classes [23]. The model first generates global topics and then generates local topics from global topics. This model can extract words posted in specific geographical regions. Yin et al. proposed latent geographical topic analysis (LGTA) and compared three models: a location-driven model, a text-driven model, and LGTA [24]. They found that adding a geographical distribution can help to model topics, and that topics provide important cues for grouping different geographical regions. Several researchers have proposed topic models that analyze trending topics as time-context-aware topic models. Blei et al. proposed a dynamic topic model (DTM) [25]. They sliced time and treated topics as being dependent on the sliced time. Kawamae proposed a time analysis model (TAM) that extracts trend words to estimate time-dependent topics [26]. Tsolmon et al. proposed an event extraction model to extract events by using an event extraction method that combines user reliability and a time-line analysis from Twitter [28]. Fukazawa et al. proposed a companion-aware topic model that introduces the companion class; they use switch variables and so can extract words related to companions [29].

Many studies have proposed context-aware topic models, but no research has proposed weather-context-aware topic models to analyze the relationship between weather and Twitter posts, even though weather is known to strongly affect the content of a Twitter post.

3. Dataset Construction

In order to investigate the relationship between weather context and Twitter content, it is necessary to link posted tweet data with the weather. This section presents the tweet database and how we link tweet data with weather data.

3.1 Tweet Content

We collected tweets posted with geographical data from the Web by using the Twitter API from May to December in 2011. To eliminate bots, we used the Levenshtein distance, which represents the closeness between two documents or words. We eliminated any tweet from the dataset if the Levenshtein distances between the tweet and the past 1,000 tweets were less than 30. We extracted tweets posted in Tokyo prefecture based on the geographical information of each tweet. This yielded a database consisting of 928,051 Japanese tweets.

3.2 Food list

First, from the above 928,051 tweets, we extracted 7,000 tweets that included the lexico-syntactic pattern "eat NOUN". We treated the extracted NOUN as a food that the tweet poster was interested in at that time and at that location. This yielded 2,533 unique food nouns. **Table 1** shows some examples. There is the possibility of extracting multiple nouns from one tweet. There are some words that do not explicitly identify a food such as "what," "thing." To exclude those unrelated words, we introduce switch variables to the topic model.

3.3 Weather Data

We used the weather data published by the Japan Meteorological Agency [30]. They publish the daily weather data captured by each meteorological station. **Table 2** summarizes the weather elements used in this research. There are other kinds of weather data such as precipitation, wind velocity, sunlight. Our previous work [31] showed that the weather elements that most affect tweet content are temperature and humidity. Therefore this study adopts temperature and humidity as weather context. We obtained weather data from the Tokyo meteorological station in the Tokyo prefecture.

Table 1 Example list of food name extracted by lexico-syntactic pattern from tweets.

rice	curry	lunch	dish	ice
supper	bread	gyoza	udon	buckwheat
eel	steak	burger	shaved ice	oden
rice ball	breakfast	takoyaki	oyster	chicken
noodles	soft cream	tan	pizza	chocolate
cake	ramen	cutlet	hormone	sweets

Table 2 Weather data published by Japan Meteorological Agency.

attribute	type	explanation
date	date	date
temperature	double	temperature (degrees)
humidity	double	humidity (%)

Table 3 Dataset structure of tweets data linked with weather data.

attribute	type	explanation
id	int	control number of tweets
date	int	Date when tweets were posted in Tokyo. It starts 0 at Jan. 1 and ends at Dec. 31
food	text	food name
temperature	double	temperature (degrees)
humidity	double	humidity (%)

3.4 Linking Tweet Content with Meteorology Information

We used the date to link tweet data with weather data. This yielded a database of 7,000 tweets with the structure shown in **Table 3**.

4. Topic Model Proposal

4.1 Weather Context

Here, we explain why we need the weather class (m_d). m_d represents the weather class associated with tweet d. In real life, users tend to determine or change their leisure or free time daily activity in response to the weather. Outside eating activity is strongly related to the weather. If it is rainy (humidity is high), the possibility of going outside for lunch is low, but the possibility of lunch box will be high. On the other hand, if the weather is comfortable (medium temperature and low humidity), the possibility of going outside for lunch or dinner is high. In addition, the food type is related to temperature. If the temperature is high, the possibility of eating ice cream will be high. Therefore, temperature (a_{1d}) and humidity (a_{2d}) are taken to be inputs of topic (z_{di}). z_{di} represents the weather topic associatred with the ith food in tweet d. There are some topic models that include links from known variables to unknown variables. In this case, however, the known variables are discrete variables, such as the author topic model [32]. It is difficult to create a model that includes links from continuous variables to variables. Therefore, temperature (a_{1d}) and humidity (a_{2d}) cannot be inputs to topics (z_{di}). Instead, we create discrete variable m_d, a link between m_d and z_{di}, and a link between m_d and a_{1d}/a_{2d}.

4.2 Event context

Here, we explain why we need event class (t_d). In **Fig. 2**, we show the occurrence frequency of the word "cake" in Twitter from the viewpoint of temperature and date. The upper figure

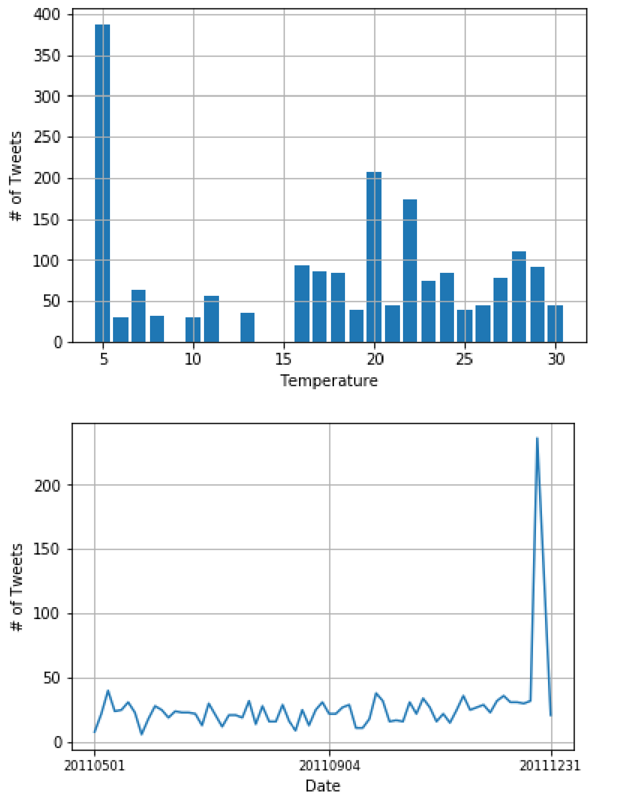


Fig. 2 Frequency of occurrence of the word “cake” from the viewpoint of temperature/date.

shows the change in the number of tweets against change of temperature. The bottom figure shows the change in the number of tweets against date transition. These data were extracted from the dataset used in this research. The upper graph shows that many users post the word “cake” when the temperature is around 5 degrees. On the other hand, when we look at the bottom graph, many users post the word “cake” on or around 24th December. Obviously, the reason is that the 24th of December is Christmas Eve, and many people eat cake on that occasion. The word “cake” is not related to weather, but related to a specific event. If we do not consider the event context, we might come to the wrong conclusion. Therefore, it is important to know the relationship between Twitter posts and both weather and events.

4.3 Switch Variables

We introduce switch variable (s_{di}) to the model in order to classify common words that do not represent food in each topic such as “what,” “thing” into background topics.

(1) Weather context specific words ($s_{di} = 0$)
(2) Event context specific words ($s_{di} = 1$)
(3) Background topic words ($s_{di} = 2$)

The model can learn the switch variables automatically. When $s_{di} = 0$, the food w_{di} is chosen from weather context topic k. When $s_{di} = 1$, the food w_{di} is chosen from event context topic z. When $s_{di} = 2$, the topic of the word (w_{di}) is chosen from the background topic o.

4.4 Topic Model

Figure 3 shows a graphic representation of this model. This

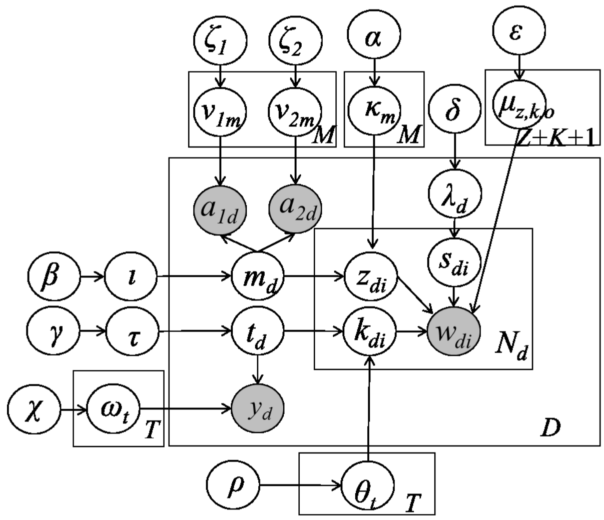


Fig. 3 Graphical representation of weather and event context-aware topic model with separate topics (WETMS).

model represents how people choose food based on the input of temperature, humidity and date information. First, weather class (m_d) is determined by the input of temperature (a_{1d}) and humidity (a_{2d}), and event class (t_d) is also determined by the input of date (y_d). Then, users choose a weather topic (z_{di}) according to the weather class (m_d) and choose an event topic (k_{di}) according to the event class (t_d). Then, users determine which topic they tweet, e.g., a weather topic (z_{di}) or event topic (k_{di}), and finally choose a food (w_{di}) according to the weather topic (z_{di}) or event topic (k_{di}). In summary, the input of the model is temperature, humidity and date information, and the output of the model is the food. We call this model, a weather and event context-aware topic model with separate topics (WETMS).

We list the notation used in this paper in **Table 4**. This notation is common to all topic models.

4.4.1 WETMS Inference

We use Collapsed Gibbs Sampling (CGS) [33] as the inference engine of our proposed model. CGS is widely adopted in LDA research. First, we describe the document generation process of WETMS.

(1) Draw multinomial ι from Dirichlet prior β
(2) Draw multinomial τ from Dirichlet prior γ
(3) Draw M multinomials κ_m from Dirichlet prior α, one for each weather class m
(4) Draw T multinomials θ_t from Dirichlet prior ρ, one for each weather class t
(5) Draw M normal distributions ν_{1m} and ν_{2m} from normal distributions ζ_1 and ζ_2, one for each weather class m
(6) Draw T normal distributions ω_t from normal distribution χ, one for each weather class t
(7) Draw Z + K + 1 multinomials $\mu_{z\ or\ k\ or\ o}$ from Dirichlet prior ϵ, one for each topic z, k or background topic o
(8) For each document d
  (a) Draw weather class m_d from multinomial ι
  (b) Draw event class t_d from multinomial τ
  (c) Draw temperature a_{1d} from normal distribution ν_{1m}
  (d) Draw humidity a_{2d} from normal distribution ν_{2m}

Table 4 Notation used in this paper.

SYMBOL	DESCRIPTION
M, T	number of weather/event classes
Z, K	number of weather/event topics
D	number of tweets
V	number of unique foods
N_d	number of foods in tweet d
m_d, t_d	weather/event class associated with tweet d
z_{di}, k_{di}	weather/event topic associated with the ith food in tweet d
s_{di}	switch variable associated with the ith food in tweet d
w_{di}	ith food in tweet d
ι, τ	multinomial distribution of the weather/event classes
ν_{1m}, ν_{2m}	normal distribution of the temperature/humidity specific to the weather class m
a_{1d}, a_{2d}	temperature/humidity associated with tweet d
ω_t	normal distribution of the date specific to the event class t
y_d	date associated with tweet d
$\mu_{z\ or\ o}$	multinomial distribution of foods specific to topic z or background topic o
λ_d	multinomial distribution of the switch variables specific to document d
κ_m	multinomial distribution of the topics specific to weather class m
θ_t	multinomial distribution of the topics specific to event class t
$\alpha, \beta, \gamma, \delta, \epsilon, \rho$	fixed parameters of the Dirichlet priors
χ, ζ_1, ζ_2	fixed parameters of the normal distribution

(e) Draw date y_d from normal distribution ω_t
(f) Draw multinomial λ_d from Dirichlet prior δ
(g) For each food i in document d
  (i) Draw switch variable s_{di} from he multinomial λ_d
    if $s_{di} = 0$
    a) Draw topic z_{di} from multinomial κ_{m_d}
    b) Draw food w_{di} from multinomial $\mu_{z_{di}}$
    if $s_{di} = 1$
    a) Draw topic k_{di} from multinomial θ_{t_d}
    b) Draw food w_{di} from multinomial $\mu_{k_{di}}$
    if $s_{di} = 2$
    a) Draw food w_{di} from multinomial μ_o

The total probability of the entire document can be written as follows:

$$
\begin{aligned}
&p(\mathbf{m},\mathbf{t},\mathbf{a_1},\mathbf{a_2},\mathbf{y},\mathbf{s},\mathbf{z},\mathbf{k},\mathbf{w},\iota,\tau,\nu_1,\nu_2,\omega,\lambda,\kappa,\theta,\mu;\beta,\gamma,\zeta_1,\zeta_2,\chi,\delta,\alpha,\epsilon)\\
&= p(\iota|\beta) \times p(\tau|\gamma) \times \prod_m^M p(\nu_{1m}|\zeta_1) \times \prod_m^M p(\nu_{2m}|\zeta_2)\\
&\times \prod_t^T p(\omega_t|\chi) \times \prod_d^D p(\lambda_d|\delta) \times \prod_m^M p(\kappa_m|\alpha)\\
&\times \prod_t^T p(\theta_t|\rho) \times \prod_z^{Z+K+1} p(\mu_z|\epsilon) \times \prod_d^D p(m_d|\iota)\\
&\times \prod_d^D p(t_d|\tau) \times \prod_d^D p(a_{1d}|\nu_{1m_d}) \times \prod_d^D p(a_{2d}|\nu_{2m_d})\\
&\times p(y_d|\omega_{t_d}) \times \prod_d^D \prod_i^{N_d} p(s_{di}|\lambda_d) \times \prod_d^D \prod_i^{N_d} p(z_{di}|s_{di}, \kappa_{m_d})\\
&\times \prod_d^D \prod_i^{N_d} p(k_{di}|s_{di}, \theta_{t_d}) \times \prod_d^D \prod_i^{N_d} p(w_{di}|s_{di}, \mu_{z_{di}}).
\end{aligned}
$$

By using CGS, we can integrate out parameters such as $\iota, \nu_1, \nu_2, \lambda, \kappa$ and μ since we cannot obtain them analytically. For each document, we can obtain the conditional distribution $p(m_d = m|\mathbf{m}\backslash d, \mathbf{z}; \alpha, \beta, \epsilon)$ is as follows.

$$
\begin{aligned}
&p(m_d = m|\mathbf{m}\backslash d, \mathbf{z}; \alpha, \beta, \epsilon)\\
&\quad \propto (n^{m,-d} + \beta_m) \times \frac{n_m^{z,-d} + \alpha_z}{\sum_{z'}(n_m^{z',-d} + \alpha_{z'})} \times f(a_{1d}, m) \times f(a_{2d}, m).
\end{aligned}
$$

where $\mathbf{m}\backslash d$ represents the vector of weather classes associated with each document, except d, and $\mathbf{z}$ represents the vector of topics associated with all words. $n^{m,-d}$ represents the number of documents assigned to weather class m, except d, and $n_m^{z,-d}$ represents the number of documents assigned to weather class m and topic z, except d. Since topics are associated with words, not documents, we define the most major topic in a document as the topic of the document. $f(a_{1d}, m)$ represents the normal distribution function of temperature in document d, where the mean and standard deviation of the normal distribution are those of documents assigned to weather class m, and $f(a_{2d}, m)$ represents the normal distribution function of humidity in document d, where the mean and standard deviation of the normal distribution are those of documents assigned to weather class m.

For each document, we obtain the conditional distribution $p(t_d = t|\mathbf{t}\backslash d, \mathbf{k}; \rho, \gamma, \epsilon)$ as follows.

$$
\begin{aligned}
&p(t_d = t|\mathbf{t}\backslash d, \mathbf{k}; \rho, \gamma, \epsilon)\\
&\quad \propto (n^{t,-d} + \gamma_t) \times \frac{n_t^{k,-d} + \rho_k}{\sum_{k'}(n_t^{k',-d} + \alpha_{k'})} \times f(y_d, t).
\end{aligned}
$$

where $\mathbf{t}\backslash d$ represents the vector of event classes associated with each document, except d, and $\mathbf{k}$ represents the vector of topics associated with all words. $n^{t,-d}$ represents the number of documents assigned to event class t, except d, and $n_t^{k,-d}$ represents the number of documents assigned to event class t and topic k, except d. $f(y_d, t)$ represents the normal distribution function of date in document d, where the mean and standard deviation of the normal distribution are those of documents assigned to event class t.

The conditional distribution $p(z_{di} = z|\mathbf{z}\backslash di, m_d, w_{di}; \alpha, \epsilon)$ is as follows:

$$
\begin{aligned}
&p(z_{di} = z|\mathbf{z}\backslash di, m_d, w_{di}; \alpha, \epsilon)\\
&\quad \propto (n_{m,(.)}^{z,-(d,i)} + \alpha_z) \times \frac{n_{(.),w}^{z,-(d,i)} + \epsilon_w}{\sum_v^V (n_{(.),v}^{z,-(d,i)} + \epsilon_v)}.
\end{aligned}
$$

where $\mathbf{z}\backslash di$ represents the vector of topics associated with each word, except ith word in document d. $n_{m,(.)}^{z,-(d,i)}$ represents the number of words assigned to topic z in documents assigned to weather class m, except the ith word in document d, and $n_{(.),v}^{z,-(d,i)}$ represents the number of words v assigned to topic z, except the ith word in document d.

The conditional distribution $p(k_{di} = k|\mathbf{k}\backslash di, t_d, w_{di}; \rho, \epsilon)$ is as follows:

$$
\begin{aligned}
&p(k_{di} = k|\mathbf{k}\backslash di, t_d, w_{di}; \rho, \epsilon)\\
&\quad \propto (n_{t,(.)}^{k,-(d,i)} + \rho_k) \times \frac{n_{(.),w}^{k,-(d,i)} + \epsilon_w}{\sum_v^V (n_{(.),v}^{k,-(d,i)} + \epsilon_v)}.
\end{aligned}
$$

where $\mathbf{k}\backslash di$ represents the vector of topics associated with each word, except the ith word in document d. $n_{t,(.)}^{k,-(d,i)}$ represents the number of words assigned to topic k in documents assigned to event class t, except the ith word in document d, and $n_{(.),v}^{k,-(d,i)}$ represent the number of words v assigned to topic k, except the ith word in document d.

The conditional distribution $p(s_{di} = 0|\mathbf{s}\backslash di, w_{di}; \alpha, \epsilon)$ is as follows:

$$p(s_{di} = 0|\mathbf{s}\backslash di, w_{di}; \alpha, \epsilon)$$
$$\propto (n_{d,(.)}^{0,-(d,i)} + \delta_0) \times \sum_z \left(\frac{n_0^{z,-(d,i)} + \alpha_z}{\sum_{z'}^{Z} (n_0^{z',-(d,i)} + \alpha_{z'})} \times \frac{n_{(.),w}^{z,-(d,i)} + \epsilon_w}{\sum_{v'}^{V} (n_{(.),v}^{z',-(d,i)} + \epsilon_{v'})} \right),$$

where $n_{d,(.)}^{0,-(d,i)}$ represents the number of words assigned to switch variable 0 in document d, except the ith word in document d. $n_0^{z,-(d,i)}$ represents the number of words assigned to topic z and switch variable 0, except the ith word in document d.

The conditional distribution $p(s_{di} = 1|\mathbf{s}\backslash di, w_{di}; \rho, \epsilon)$ is as follows:

$$p(s_{di} = 1|\mathbf{s}\backslash di, w_{di}; \rho, \epsilon)$$
$$\propto (n_{d,(.)}^{1,-(d,i)} + \delta_1) \times \sum_k \left(\frac{n_1^{k,-(d,i)} + \rho_k}{\sum_{k'}^{K} (n_1^{k',-(d,i)} + \rho_{k'})} \times \frac{n_{(.),w}^{k,-(d,i)} + \epsilon_w}{\sum_{v'}^{V} (n_{(.),v}^{k',-(d,i)} + \epsilon_{v'})} \right),$$

where $n_{d,(.)}^{1,-(d,i)}$ represents the number of words assigned to switch variable 1 in document d, except the ith word in document d. $n_1^{k,-(d,i)}$ represents the number of words assigned to topic k and switch variable 1, except the ith word in document d.

The conditional distribution $p(s_{di} = 2|\mathbf{s}\backslash di, w_{di}; \epsilon)$ is as follows:

$$p(s_{di} = 2|\mathbf{s}\backslash di, w_{di}; \rho, \epsilon)$$
$$\propto (n_{d,(.)}^{2,-(d,i)} + \delta_2) \times \frac{n_{(.),w}^{o,-(d,i)} + \epsilon_w}{\sum_{v'}^{V} (n_{(.),v}^{o',-(d,i)} + \epsilon_{v'})},$$

where $n_{d,(.)}^{2,-(d,i)}$ represents the number of words assigned to switch variable 2 in document d, except the ith word in document d.

5. Evaluation

In this chapter, we conduct a quantitative evaluation to compare existing topic models. Then we analyze the relationship between food choice and weather/event context based on the learned model.

5.1 Compared Methods

In order to determine the effects of a combination of two contexts, we compare the following three models. Taking the same approach as the context aware topic model proposed in Ref. [29], we introduce a single context class as the compared methods 1 and 2. On the other hand, we introduce multiple context class to allow comparison with method 3, but the integration of multiple context class is rather simple.

(1) **Weather context-aware topic model (WTM)**
This model, shown in **Fig. 4**, considers only weather context. This model can identify the food choice based on the topics as influenced by temperature and humidity.

(2) **Event context-aware topic model (ETM)**
This model, shown in **Fig. 5**, considers only event context. This model can identify the food choice based on the topics as influenced by the event factor.

(3) **Weather and event context-aware topic model with integrated topics (WETMI)**
This model, shown in **Fig. 6**, considers both weather and event context. The difference between WETMI and WETMS is the way in which the weather and event contexts are combined. WETMI combines the two contexts during the topic choice process, whereas WETMS combines them in the word choice process. In other words, users determine foods according to some combination of weather/event contexts in WETMI, whereas users determine foods according to either of weather context or event context in WETMS.

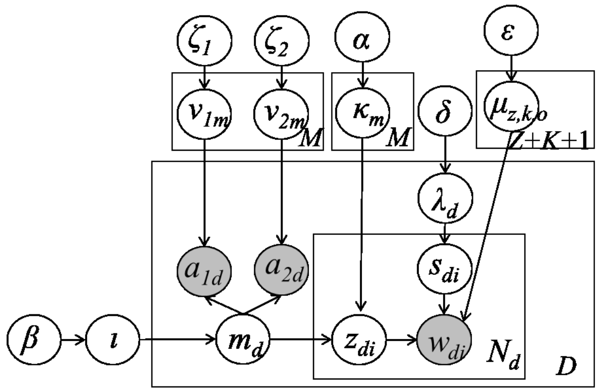


Fig. 4 Graphical representation of weather context-aware topic model (WTM).

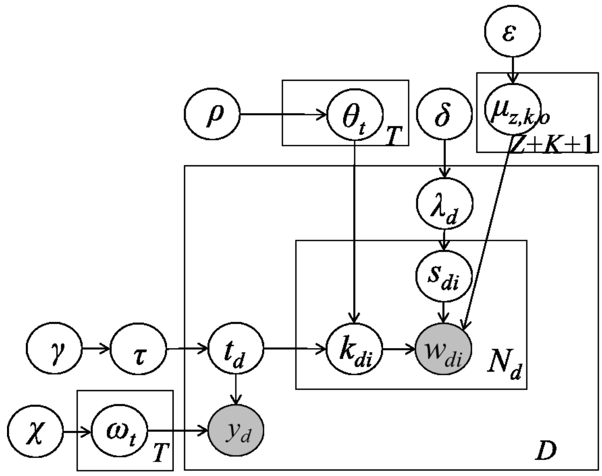


Fig. 5 Graphical representation of event context-aware topic model (ETM).

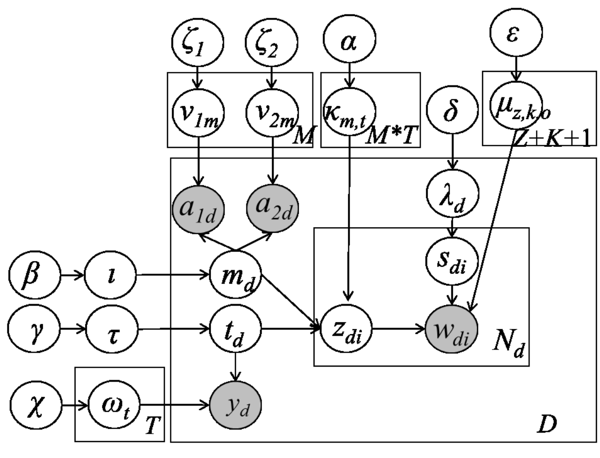


Fig. 6 Graphical representation of weather and event context-aware topic model with integrated topics (WETMI).

Table 5 The relationship between perplexity and the number of topics.

parameters	models	perplexity	best models
Z, K = 20	WTM	5843.7	
	ETM	5863.6	
	WETMI	5655.2	
	WETMS	5603.1	*
Z, K = 40	WTM	5282.6	
	ETM	5350.2	
	WETMI	5242.1	
	WETMS	5022.7	*
Z, K = 60	WTM	5128.8	
	ETM	5065.2	
	WETMI	5052.8	
	WETMS	4783.0	*
Z, K = 80	WTM	4952.9	
	ETM	4965.3	
	WETMI	4797.8	
	WETMS	4663.0	*

Table 6 The relationship between perplexity and the number of topics.

parameters	models	perplexity	best models
M, T = 5	WTM	4943.4	
	ETM	4962.9	
	WETMI	4803.1	
	WETMS	4738.6	*
M, T = 10	WTM	4952.9	
	ETM	4965.3	
	WETMI	4797.8	
	WETMS	4663.0	*

5.2 Quantitative Evaluation

To measure the ability of the topic models as document generation models, we computed the perplexity and compared the resulting values. The perplexity is equivalent to the inverse of the word likelihood [26]. A lower perplexity means that the words in a document are not surprising to the topic models and therefore a lower perplexity is better. The definition of perplexity is as follows:

$$Perplexity = \exp\left(-\frac{1}{\sum_{d=1}^{D} N_d} * \sum_{d=1}^{D}\sum_{i=1}^{N_d} \log(p(w_{di}))\right), \qquad (1)$$

where $p(w_{di})$ means the likelihood of the ith word in document d. We randomly took 20% of each tweet as the test part and the remainder as the learning part.

5.2.1 Comparison of Topic Models

We describe how to set parameters. There are two kinds of parameters; hyperparameters $\alpha, \beta, \gamma, \delta, \epsilon, \rho, \chi, \zeta_1, \zeta_2$ and a number of topics or classes (Z, K, M, T). As for hyperparameters, we often use fixed values in LDA and its extensions [33]. The hyperparameters of Dirichlet priors are usually set at 1/|number of topics or classes| [26], [27]. Here the parameters of Dirichlet priors $\alpha, \beta, \gamma, \delta, \epsilon, \rho$ were set at 1/Z, 1/M, 1/T, 1/3, 1/V and 1/K respectively. The hyperparameters of normal distribution χ, ζ_1, ζ_2 were set at mean and variance of y_d, a_{1d} and a_{2d} for all d respectively.

We tuned the parameters Z, K, M, T for each model by varying the parameters one by one and choosing the value yielding the lowest perplexity. In **Table 5** and **Table 6**, we show the perplexity when we change the parameters (Z, K, M, T), and we show the comparison of the perplexity with the best parameter value in **Table 7**. In Table 5, M, T is fixed at 10, and in Table 6, Z, K is fixed at 80. The differences between WTM, ETM, WETMI, and WETMS are significant (p of t-test are $p < 0.01$) according to one-tailed t-tests.

Table 7 Comparison of the perplexity.

WTM	ETM	WETMI	WETMS
4943.4	4962.9	4797.8	4663.0

Table 8 Results of the distributions of words associated with each weather class by WETMI. Among foods marked with [1], the result includes food that is not related to weather such as "dish." Among foods marked with [2], "soba"(buckwheat noodles) and "cake" are not related to coldness but special events occurring at the Christmas Eve and end of the year.

Weather class		high temp.	low temp.
Avg. temperature (°C)		31.28±1.29	9.29±0.61
Avg. humidity (%)		55.28±3.64	20.70±4.34
Foods of each topic	1	curry	soba[2]
		burger	sushi
		riceball	croquette
	2	dish[1]	cake[2]
		Thai	parfait
		crepe	short rib
	3	udon	New Year's Eve
		cheese	ponzu sauce
		fried cutlet	alamode

From these results, we we draw on two observations. One is that WETMI and WETMS are better than WTM and ETM from the viewpoint of perplexity. This is because the first two models reduce the number of possible words in a specific tweet by considering two contexts. The other one is that WETMS is superior to WETMI. This is because the words related to weather and event are clearly different. The hypothesis that users independently choose words according to weather or event is more accurate than the hypothesis that users choose words according to the combination of weather and event. These two observations suggest the importance of properly introducing weather and event contexts into context-aware topic models.

5.3 Qualitative Evaluation

In this section, we discuss the food choices related to temperature, which is found from the topics learned by the proposed method. In **Table 8** and **Table 9**, we selected two topics z assigned to the weather class m that have the highest and the lowest temperatures. Table 8 and Table 9 show the result learned by the WETMI and WETMS respectively. The tables also show the list of foods assigned to the selected topics.

5.3.1 Food Likely to Be Eaten When Temperature Is Low

First, we discuss the food choices related to low temperature. In the case of compared model WETMI (Table 8), the highest frequency words in the weather class of low temperature contain "soba" (buckwheat noodles) and "cake." The foods mentioned above are not related to coldness but events such as the end of the year or Christmas. As proof, in **Fig. 7**, we show the frequency for "soba" from the viewpoints of temperature and date. The upper figure shows the change in the number of tweets against change of temperature. The bottom figure shows the change in the number of tweets against date transition. The frequency peaks in the word "soba" is 5 degrees and 20 degrees, and in the date graph, the peak is 31st December. This is because most Japanese eat

Table 9 Results of the distributions of foods associated with weather class of high and low temperature by WETMS. As for foods marked with [1], it is natural for those foods to be appeared in the list as they are specially eaten in hot/cold temperature. As for foods marked with [2], it is new discovery for those foods to be appeared in the list. In the case of "natto," we investigated the twitter post and found that "natto" is good for preventing summer heat fatigue, and some people tend to eat natto in hot temperature.

weather class		high temp.	low temp.
mean temperature (°C)		33.69±1.11	9.09±1.21
mean humidity		46.63±5.21	44.98±9.84
foods of each topic	1	ice cream[1]	curry[1]
		natto [2]	udon[1]
		tea[1]	crepe [2]
	2	curry	snacks [2]
		udon	manju [2]
		crepe	stall
	3	shaved ice[1]	okonomiyaki[1]
		soba [2]	Chinese noodles[1]
		water melon[1]	mont blanc [2]

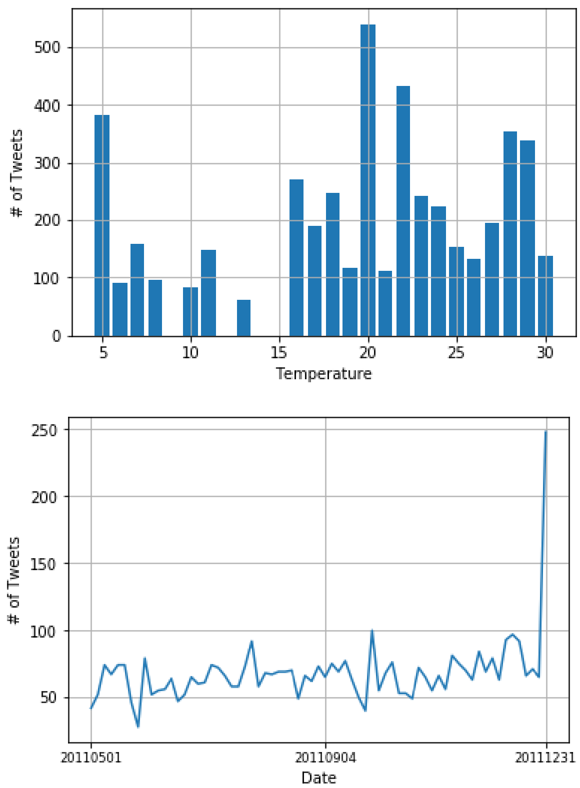


Fig. 7 Frequency of occurrence of the word "soba" from the viewpoint of temperature/date.

buckwheat noodles on New Year's Eve as an old custom. Therefore, the choice of "soba" is related to a specific event, and not to low temperatures.. Therefore, such choices are event context related and not weather context related.

On the other hand, in the case of proposed model WETMS (Table 9), the highest frequency words in the weather class of low temperature include "curry," "udon," "okonomiyaki" and "Chinese noodles" which are rather hot foods. Note that "curry" and "udon" appear as independent foods, but actually they represent "curry udon" which is an original Japanese dish that mixes udon with curry soup. In addition, we found that sweets such as snacks, manju and Mont Blanc (dessert) are eaten in the low temperature

Table 10 Part of the results of the distributions of foods associated with event class by WETMS. As for food marked with [1], the result includes food that is not related to weather such as "dish."

Event Class	Avg. date	Foods of the topic assigned to event class
Class3	212.3±21.7	lunch, rice, lunch meal
Class10	361.8±2.24	soba, gyoza, chocolate

season. In **Table 10**, we selected the event topic k to which food "soba" is assigned by WETMS. As can be seen from Class 10 in the table, we can extract the "soba" eating event which is held at the end of the year (see 31st December in Fig. 7). From the above, by dividing the topics into weather topics and event topics, the model can extract foods that are more strongly related to weather.

5.3.2 Foods Likely to Be Eaten When Temperature Is High

First, we discuss the food choices related to high temperature. In the case of compared model WETMI (Table 8), the highest frequency words in the weather class for high temperature include "curry," "burger," "dish" and "udon." Those foods are not related to events, however foods not related to the weather such as "dish" are included. On the other hand, in the case of proposed model WETMS (Table 9), the highest frequency words in weather class of high temperature contain "ice cream," "shaved ice," "tea," "watermelon" which are appropriate for high temperatures. In addition, we found that "soba" is a hot temperature choice. Japanese eat "soba" in the hot season as zaru-soba (cold soba noodles served on a woven bamboo tray). By excluding soba popularity at the end of the year event as event class, we can find the relationship between temperature and food more correctly. We found "natto" as a choice here which is surprising since it is not usually considered a hot season dish. Upon investigating the twitter post, we found that "natto" is good for preventing summer heat fatigue, and so some people tend to eat natto in hot temperature.

Here we explain unexpected but useful effects from introducing the event class. In Table 10, we selected event topic k that has non-weather related foods such as lunch or rice as assigned to Class 3 by WETMS. Considering the deviation of the class is almost six weeks (i.e., 21.7 days = 3 weeks), the event class acts to extract words which are used on a daily basis and thus not related to weather. On the other hand, as in Table 8, there are non-weather-related daily words for food such as "dish" at the top of the assignment list in the case of WETMI (Table 8). From the above, by dividing the topics into weather topics and event topics, the model can extract words that are more strongly related to weather by assigning non-weather-related daily words of food as event class.

6. Conclusion

In this research, we investigate the relationship between the food choice process and weather context. This paper provides several contributions. The first contribution is that it mines social media on the web to create datasets on food choice, and associate food selections with weather context gathered from govermental meteorological information. A second contribution is that it shows that food events or customs on special day have a significant impact on the characteristics of the food choice process. A third contribution is that it creates a topic model that excludes the effects of events to better represent the abstract relationship between food choice and weather. We quantitatively evaluated several models by perplexity, and discovered that prediction performance was improved by considering the contexts of weather and events. Perplexity of proposed method (weather and event context-aware topic model with separated topics) was (4663.0), which is an improvement over the compared model (4943.4). In addition, we qualitatively evaluated the models. Through the example of "soba," we showed that the proposed model can learn the correct relationship between low temperature and "soba" by eliminating the effect of the soba eating event at the end of the year. In the future, we intend to consider the impact of other factors such as user preference on the food choice process.

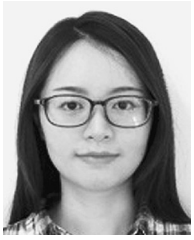
Acknowledgments A part of the results of this research were developed through collaborative research with Inc. GLOBALPRENEURS and The University of Tokyo.

References

[1] Winter Falk, L., Bisogni, C.A. and Sobal, J.: Food choice processes of older adults: A qualitative investigation, *Journal of Nutrition Education*, Vol.28, pp.257–265 (1996).
[2] Shepherd, R. and Raats, M. (Eds.): The psychology of food choice, *Frontiers in Nutritional Science*, Vol.3 (2006).
[3] Furst, T., Connors, M., Bisogni, C.A., Sobal, J. and Falk, L.W.: Food choice: A conceptual model of the process, *Appetite*, Vol.26, No.3, pp.247–266 (1996).
[4] Connors, M.M., Bisogni, C.A., Sobal, J. and Devine, C.: Managing values in personal food systems, *Appetite*, Vol.36, pp.189–200 (2001).
[5] Bisogni, C.A., Jastran, M., Shen, L. and Devine, C.M.: A biographical study of food choice capacity, Standards, circumstances, and food management skills, *Journal of Nutrition Education and Behavior*, Vol.37, No.6, pp.284–291 (2005).
[6] Sobal, J. and Bisogni, C.A.: Constructing food decision making, *Ann Behav Med.*, Vol.38, pp.37–46 (2009).
[7] Sobal, J., Bisogni, C.A. and Jastran, M.: Food choice is multifaceted, contextual, dynamic, multilevel, integrated, and diverse, *Mind, Brain, and Education*, Vol.8, No.1, pp.6–12 (2014).
[8] Oliver, G., Wardle, J. and Gibson, E.L.: Stress and food choice: A laboratory study, *Psychosomatic medicine*, Vol.62, No.6, pp.853–865 (2000).
[9] Holsten, J.E., Deatrick, J.A., Kumanyika, S., Pinto-Martin, J. and Compher, C.W.: Children's food choice process in the home environment: A qualitative descriptive study, *Appetite*, Vol.58, No.1, pp.64–73 (2012).
[10] Huneke, T., Benoit, S., Shams, P. and Gustafsson, A.: Does service employees' appearance affect the healthiness of food choice?, *Psychology & Marketing*, Vol.32, No.1, pp.94–106 (2015).
[11] Chen, X., Cho, Y. and Jang, S.: Crime prediction using twitter sentiment and weather, *Systems and Information Engineering Design Symposium*, pp.63–68 (2015).
[12] Demirbas, M., Bayir, A., Akcora, G., Yilmaz, S. and Ferhatosmanoglu, H.: Crowd-sourced sensing and collaboration using twitter, *World of Wireless Mobile and Multimedia Networks* (*WoWMoM*), pp.1–9 (2010).
[13] Blei, D., Ng, A., Jordan, M. and Lafferty, J.: Latent dirichlet allocation, *JMLR*, Vol.3, pp.993–1022 (2003).
[14] Andrzejewski, D., Zhu, X. and Craven, M.: Incorporating domain knowledge into topic modeling via dirichlet forest priors, *ICML*, pp.25–32 (2009).
[15] AlSumait, L., Barbara, D. and Domeniconi, C.: Online lda: Adaptive topic models for mining text streams with applications to topic detection and tracking, *ICML*, pp.3–12 (2008).
[16] Ahmed, A. and Xing, E.P.: Timeline: A dynamic hierarchical dirichlet process model for recovering birth/death and evolution of topics in text stream, *ICML*, pp.2–29 (2010).
[17] Krestel, R., Fankhauser, P. and Nejdl, W.: Latent dirichlet allocation for tag recommendation, *RecSys*, pp.61–68 (2009).
[18] Lau, J.H., Grieser, K., Newman, D. and Baldwin, T.: Automatic labeling of topic models, *ACL*, pp.1536–1545 (2011).
[19] Ding, Z., Qiu, X., Zhang, Q. and Huang, X.: Learning topical translation model for microblog hashtag suggestion, *IJCAI*, pp.2078–2084 (2013).
[20] Wang, X., McCallum, A. and Wei, X.: Topical n-grams: Phase and topic discovery, with an application to information retrieval, *ICDM*, pp.697–702 (2007).
[21] Chen, Y., Pavlov, D. and Canny, J.F.: Large-scale behavioral targeting, *ICDM*, pp.209–218 (2009).
[22] Ahmed, A., Low, Y., Aly, M., Josifovski, V. and Smola, A.J.: Scalable distributed inference of dynamic user interests for behavioral targeting, *KDD*, pp.114–122 (2011).
[23] Eisenstein, J., O'Connor, B., Smith, N. and Xing, E.: A latent variable model for geographic lexical variation, *EMNLP*, pp.1277–1287, Association for Computational Linguistics (2010).
[24] Yin, Z., Cao, L., Han, J., Zhai, C. and Huang, T.: Geographical topic discovery and comparison, *WWW*, pp.247–256, ACM (2011).
[25] Blei, D. and Lafferty, J.: Dynamic topic models, *WWW*, pp.113–120, ACM (2006).
[26] Kawamae, N.: Trend analysis model: Trend consists of temporal words, topics, and timestamps, *WSDM*, pp.317–326, ACM (2011).
[27] Kawamae, N.: Predicting Future Reviews: Sentiment Analysis Models for Collaborative Filtering, *WSDM*, pp.605–614, ACM (2011).
[28] Tsolmon, B. and Lee, K.: An event extraction model based on timeline and user analysis in latent dirichlet allocation, *WWW*, pp.1187–1190, ACM (2014).
[29] Fukazawa, Y. and Ota, J.: Companion context dependent topic modeling, *International Conference on Mobile Computing and Ubiquitous Networking*, pp.105–110 (2014).
[30] Japan meteorological agency, http://www.jma.go.jp/jma/
[31] Itoh, T., Fukazawa, Y., Zhu, D. and Ota, J.: Climate condition that mostly affects the change of tweet content, *ICMU*, pp.72–73 (2015).
[32] RosenZvi, M., Griffiths, T., Steyvers, M. and Smyth, P.: The author-topic model for authors and documents, *UAI*, pp.487–494, AUAI Press (2004).
[33] Griffiths, T. and Steyvers, M.: Finding scientific topics, *PNAS*, No.1, pp.5228–5235 (2004).

Taku Ito received his B.Eng. and M.Eng. degrees from The University of Tokyo in 2015 and 2017, respectively. He joined NTTDOCOMO, Inc. in 2017. He is a member of IPSJ.

Yusuke Fukazawa received his B.Eng. and M.Eng. degrees from The University of Tokyo in 2002 and 2004, respectively. He joined NTTDOCOMO, Inc. in 2004. He received his Ph.D. at The University of Tokyo in 2011. He has also joined RACE (Research into Artifacts, Center for Engineering), The University of Tokyo as a collaborative researcher in 2011–2016 and has been a visiting researcher since 2016. His research interests include human behavior understanding and content recommendation. He is a member of IEEE, JSAI and IPSJ.



Dandan Zhu received her B.E. degree in Automation from Central South University in 2008, M.E. degree in Air- craft Design from Beihang University in 2011, and Ph.D. degree in Precision Engineering from The University of Tokyo in 2014. She subsequently worked in Research into Artifacts Center for Engineering (RACE), The University of Tokyo, as a project researcher. In 2015, she joined China University of Petroleum, Beijing (CUPB) as a lecturer, and from July, she became an associate professor in Computer Science Dept. in College of Geophysics and Information Engineering in CUPB. Her current research fields include machine learning and artificial intelligent in industrial production fields.

Jun Ota He received his B.E., M.E. and Ph.D. degrees from Faculty of Engineering, The University of Tokyo in 1987, 1989 and 1994, respectively. From 1989 to 1991, he joined Nippon Steel Cooperation. In 1991, he was a research associate of The University of Tokyo. In 1994, he became a Lecturer. In 1996, he became an associate professor. From April 2009, he became a professor at Graduate School of Engineering, The University of Tokyo. From June 2009, he became a professor at Research into Artifacts, Center for Engineering (RACE), The University of Tokyo. From 2015, he is a guest professor of South China University of Technology. From 1996 to 1997, he was a visiting scholar at Stanford University. He was accepted as an RSJ (the Robotics Society of Japan) Fellow in 2016. His research interests include multi-agent robot systems, embodied-brain systems science, design support for large-scale production/material handling systems, human behavior analysis and support.