Electronic Preprint for Journal of Information Processing Vol.25

Regular Paper

# Event Identification for Explicit and Implicit References on Microblogs

Jun-Li Lu[1,a) ] Makoto P. Kato[1,b)] Takehiro Yamamoto[1,c)] Katsumi Tanaka[1,d)]

Received: December 20, 2016, Accepted: April 10, 2017

***Abstract:*** We address the problem of event identification on microblogs with special attention to implicit reference cases in which events are not referred to by event's information. Most studies identify events referred to by event's information, while there are many implicitly referred events by microblogs, which are difficult to identify for short text such as microblogs. We therefore tackled implicit reference cases by analyzing links from microblogs. The links are able to connect opinions or feeling to their referred events. The analysis of links is particularly important for certain types of implicit references. In addition, we predict reference type of a microblog for accurately ranking referred events. The experimental results suggest that our method was effective for implicit references and predicting reference type was essential for identifying implicitly or explicitly referred events together.

***Keywords:*** event identification, implicit reference, reference prediction

## 1. Introduction

People write microblogs (posts) to share their opinions about information in the real world and refer to events, such as news, in various ways. Event identification, a problem of identifying events referred to by documents, has been addressed extensively in the literature [1], [2], [3]. Most related studies addressed the problem of identifying events referred to by event's information, such as event's subjects and their actions. However, there are many events referred to by microblogs without using event's info., which are often difficult to identify, especially for short text such as microblogs. E.g., given a microblog "United Kingdom's decision made people unbelievable!," it is hard to know *which event* made people unbelievable if one does not know the fact that "United Kingdom's withdrawal from the European Union was far away from people's expectations." In this paper, we address the event identification problem with special attention to *implicit reference* cases in which events are not referred to by event's information. Especially, if people refer to events by using their opinions or feeling, it is difficult to identify referred events because opinions/feeling are usually irrelevant to event's info.

We therefore advocate to utilize *links* from microblogs, on social media platforms, because links are able to connect opinions or feeling, which are irrelevant to event's info., to their referred events. E.g., given a microblog $m$ "It's a sad news!" linking to an event $e$ "Two injured in gun shooting," if a microblog $m_2$ "It's a broken-hearted news!" is semantically similar to $m$, we may infer that microblog $m_2$ could be used to refer to $e$. We furthermore find similar links among a set of links from microblogs because similar links generate more evidence between microblogs and their referred events. We then propose a clustering-based algorithm for accurately finding similar links.

We also predict reference type of a microblog because referred events can be different according to reference type, E.g., given a microblog "I am a Mysterious fan". referring to event $e$ "A movie, The Mysterious Men, released.", if we think that this microblog is an implicit reference because of the feeling of mysterious, we may infer the referred events that make people feel mysterious and are possibly far from $e$. We therefore predict reference type of a microblog for accurately ranking referred events. Experiments were carried out with 15,996 tweets referring to total 231 events. The experimental results suggested that (1) our method was effective for implicit references, (2) predicting reference type was essential for identifying implicitly or explicitly referred events together, and (3) finding similar links from microblogs was useful for composing implicitly referred events. We summarize the contributions of this work:

- We address the problem of identifying implicitly referred events from microblogs and argue that there is a considerable amount of implicit references in real microblog data.
- We propose a method for identifying implicitly referred events and predict reference type of a microblog for accurate event identification.
- We demonstrate that our method was effective for implicit references and predicting reference type was essential for identifying implicitly and explicitly referred events together.

The rest of this paper is organized as follows. We discuss the related work on event identification in Section 2. We explain how an event is referred through an explicit or implicit reference and the event identification problem in the presence of explicit and implicit references in Section 3. We predict reference type of a

1 Department of Social Informatics, Graduate School of Informatics, Kyoto University, Kyoto 606–8501, Japan
a) jllu@dl.kuis.kyoto-u.ac.jp
b) kato@dl.kuis.kyoto-u.ac.jp
c) tyamamot@dl.kuis.kyoto-u.ac.jp
d) tanaka@dl.kuis.kyoto-u.ac.jp

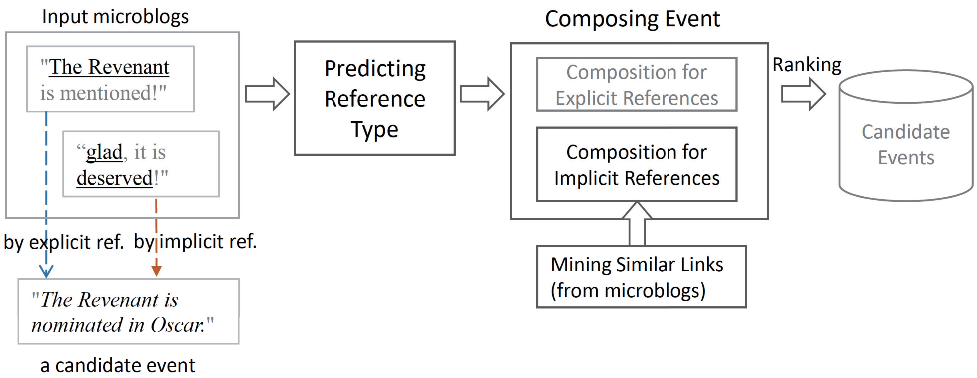



**Fig. 1** Event identification on microblogs in the presence of explicit (event info.) and implicit (non-event-info.) references and our proposed system.

microblog in Section 4, compose possible event for implicit references, and use the predicted reference type to compose event for implicit and explicit references in Section 5. We conduct a survey and experiments on real-wold microblog data, discuss the experimental results on event identification and composition for implicit references, give a case study in Section 6, and conclude the paper in Section 7.

## 2. Related Work

Event identification or detection has been an area of active research and attracted many researchers. Given a document or documents, one task is to detect whether documents are referred to events or not, and another task to identify which events are referred to by documents [1], [2], [4], [5], [6]. We categorize studies related to event identification or detection as follows. Some studies extracted real-time events from microblogs (e.g., tweets) of a social media platform (e.g., Twitter [*1]) for providing up-to-date information about the real world [7], [8]. Some studies analyzed the detected events from historical microblogs across a time period for understanding behaviors of people mentioning events of past or expectation to future [9], [10]. Some studies searched the first-story of events from microblogs for knowing the reason or origin of events [3], [11], [12]. Some studies summarized an amount of microblogs across time periods and geographical locations for providing an overview or visualization of events with respect to time and locations [10], [13], [14], [15], [16], [17], [18]. Some studies acquired the information related to events by extracting geospatial location mentions or people's opinions, sentiments, or actions from microblogs [4], [19], [20], [21]. Some past studies detected events from news articles [22] and some studies compared behaviors of references of events between a social media platform and a traditional news platform [23], [24]. For the above past studies, the generation of referred events of documents was usually according to event's information. However, in the real world, people can refer to events from documents by using various ways, such as opinions or feeling of people. In this paper, we shed light into implicit references, where events are not referred to by event's information, and propose a method to address the event identification problem in the presence of explicit and implicit references.

To evaluate a referred event, especially from microblogs, the features used in previous studies are listed as follows. Microblog features [8], [25]: some studies used the content, symbols (e.g., hashtag of Twitter and emoticons of writer's expression), hyperlinks, locations, post time of a microblog, and replied microblogs. Social media platform features [7], [8], [24], [26], [27], [28]: some studies used trending or key terms, bursty (suddenly frequently appeared) terms, frequently queried terms on a social network platform, and modeled topics and spatial or temporal aspects of a set of various microblogs. Knowledge base features [8], [26], [29]: some studies used additional knowledge such as information of events and entities, which could be extracted from Wikipedia, characteristics of language. NLP (natural language processing) features [2], [3], [11]: some studies used the parsed results of documents by NLP techniques such as proper nouns (e.g., "Kyoto"), semantics or synonyms of terms, detected sentiments, part-of-speech (nouns, verbs, etc.), or functional terms (e.g., question words). As we discussed in our experiments, the keyword-based features from microblogs were not effective for implicit references. The difficulty is that implicitly referred events are usually mentioned by opinions or feeling of people. We therefore propose to use links from microblogs that are able to connect opinions or feeling, which are irrelevant to event's information, to their referred events. We also predict reference type of a microblog and use the predicted reference type to compose possible event for implicit and explicit references.

## 3. Problem: Event Identification on Microblogs

We introduce several terms to define the problem addressed in this paper as shown in **Fig. 1**. *Microblog*: a microblog is a sequence of words posted by a user on a social media platform. E.g., a tweet posted by a user on Twitter is a kind of microblog. *Event*: an event consists of subjects and the action/status of these subjects. E.g., an event $e$ consists of subject "The Revenant" (a movie) and status "is nominated in Oscar." We then discuss how an event is referred to by a microblog.

**Explicit (event info.) reference**: previous studies [1], [4] generated referred events by using event's information as follows. If microblog $m$ is used to refer to event $e$ and $m$ contains event's info. of $e$, then the reference to $e$ by $m$ is a explicit reference.

[*1] https://twitter.com/.

E.g., $m$ = "The Revenant is mentioned!" refers to event $e$ = "The Revenant is nominated in Oscar". by using the subject of $e$. Or, $m$ = "The movie is nominated!" refers to $e$ by the subject's status.

Note that $m$ is used to refer to $e$ if the writer post $m$ to reply $e$, where we can check the reply by evidence (e.g., *hyperlink* or *reply* of Twitter). However, people can refer to events without using event's information. From real-world microblog data, we found that many people refer to events by posting their opinions, feeling, or related events.

**Implicit (non-event-info.) reference**: if microblog $m$ is used to refer to event $e$ and $m$ does not contain event's info. of $e$, then the reference to $e$ by $m$ is an implicit reference as the following cases. E.g., $m$ = "It deserved!" refers to event $e$ by using opinion or $m_2$ = "Congrats, glad to know this!" refers to $e$ by using feeling. Or, $m_3$ contains another event "This is directed by Alejandro," which is used to refer to $e$. We therefore define the problem as follows.

**Definition 1** (Event Identification on microblogs in the presence of Explicit/Implicit References, *EI-EIR*): given a microblog $m$ from a user $u$ and a set of events $E$, the *EI-EIR* problem is to identify the event $e \in E$, which is referred to by $m$, through an explicit or implicit reference.

## 4. Predicting Reference Type

Given a microblog, we predict reference type of the microblog because the referred event can be different according to the reference type. E.g., given a microblog "I am a Mysterious fan" referring to event $e$ "A movie, The Mysterious Men, released," if we infer that the microblog is a implicit reference by using mysterious feeling, we may regard the referred event that makes people feel mysterious, which is possibly far from $e$.

We predict the reference type of a microblog as follows. Because subjects of an event are usually terms of proper nouns, we assume that proper nouns in a microblog imply an explicit reference. We also assume that feeling terms in a microblog imply an implicit reference because people implicitly refer to events by opinions, which may not be event's information and often includes feeling terms. Given a microblog $m$, we therefore compute the probability of reference type $t \in$ {"explicit ref","implicit ref"} of $m$ by

$$p(t|m) = \frac{1}{z}\exp\left(\mathbf{w}^T\mathbf{f}(t,m)\right) \tag{1}$$

where $\mathbf{f}(t,m)$ is a vector of feature values $f_1, f_2, f_3, f_4$ (described as follows), $\mathbf{w}$ is a vector of weights of $\mathbf{f}(t,m)$, and $z$ is the normalizing constant. Suppose the number of proper nouns in $m$ is $n_p$ and the number of feeling terms in $m$ is $n_f$, we use $f_1 = n_p$ and $f_2 = n_p - n_f$ for $t$ = "explicit ref" and use $f_3 = n_f$ and $f_4 = n_f - n_p$ for $t$ = "implicit ref." Note that we train $\mathbf{w}$ as described in Appendix A.1.

## 5. Composing Event

We propose a method to identify referred events from explicit or implicit references. We especially focus on implicit references and propose a method to compose possible events for implicit references. We then use the predicted reference type to compose event for both implicit and explicit references.

### 5.1 Composition for Implicit References

The difficulty of evaluating an implicit reference between a microblog $m$ and a referred event $e$ is that $m$ does not contain event's info. of $e$, which makes it difficult to identify $e$. We therefore utilize *link* because links are able to connect opinions or feeling, which are irrelevant to event's info., to their referred events. A link $r = (s, o)$ on a social media platform is a directed mapping from a source unit of microblog $s$ to a mapped unit $o$, where $o$ is one of the following two instances. $o$ is another microblog if $s$ links to $o$. E.g., a user can post $s$ to reply to $o$ in Twitter. Or, $o_2$ is an outside resource if the content of $s$ contains a link to $o_2$. E.g., $s$ contains a hyperlink of a web page $o_2$.

We argue that if two different microblogs with similar content of text (e.g., opinions) and they link to similar mapped units (e.g., events), we may infer that each microblog can be used to refer to another mapped unit. We therefore find similar links for connecting microblogs to their referred events.

**A problem of finding similar links**. It would be difficult to find similar links due to that source units (e.g., opinions) and mapped units (e.g., fact of events) of links come from different domains (types of text). E.g., considering the following equation used for connecting two links $(s, o)$, $(s_2, o_2)$,

$$a \cdot sim(s, s_2) + b \cdot sim(o, o_2) \geq \theta,$$

where $s, s_2$ are from domain of source units $\mathcal{S}$, $o, o_2$ are from domain of mapped units $O$, $sim(s, s_2)$ is the cosine similarity between the bag-of-words model of $s$ using tf-idf weighting [30] and that model of $s_2$, $0 \leq a, b \leq 1$ are weights, $\theta$ is the threshold. It would be hard to compute weights $a$ and $b$ for finding similar links because we do not know the correlation between domain $\mathcal{S}$ and domain $O$. To avoid the above problem, we utilize link instances for building the correlation between domain $\mathcal{S}$ and domain $O$.

**Finding Similar Links by Clustering Sequentially**. Given a set of links $R$, we find similar links by considering similarities of source units (of links), similarities of mapped units, and each link instance. We use clustering for time-efficiently [*2] finding similar source units (or mapped units) of links according to $sim(s, s')$ (or $sim(o, o')$). The process of extracting similar links from $R$ is as follows.

(1) We extract similar mapped units of links by clustering on each $o$ of $(s, o) \in R$ according to $sim(o, o')$ using *k-means clustering* [31] and acquire each cluster as $C^o_i = \{o_{i,1}, o_{i,2}, ...\}$.
(2) For each cluster of mapped units $C^o_i$, we acquire a set of source units $S_i$ that are linked by each $o \in C^o_i$. I.e., $S_i = \{s|(s, o) \in R, o \in C^o_i\}$.
(3) For each linked source units $S_i$, we then extract similar source units by clustering on each $s \in S_i$ according to $sim(s, s')$ using *k-means clustering* and acquire each cluster as $C^s_{i,j} = \{s_{i,j,1}, s_{i,j,2}, ...\}$.
(4) According to each cluster of source units $C^o_i$ and each cluster

[*2] The complexity of computing the similarity between each pair of links $r_1$ and $r_2$, $\forall r_1, r_2 \in R$, is in quadratic, $|R|^2$.

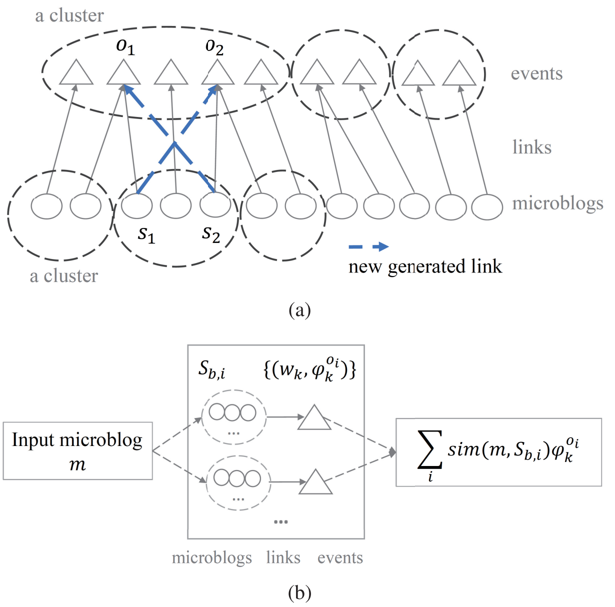



**Fig. 2** (a) Finding similar links by clustering sequentially. (b) Composition for implicit references.

of mapped units $C^s_{i,j}$, we can extract similar links and generate more linkable source and mapped units of links. I.e.,

$$R^* = \{(s, o')|(s, o), (s', o') \in R, s, s' \in C^s_{i,j}, \quad o, o' \in C^o_i, \forall i, j\} \cup R. \tag{2}$$

**Composed event for implicit reference**. Given a microblog $m$, we find the referred event of $m$ by using a set of links $R^*$. However, $R^*$ may not contain the referred event instance. We therefore compose the *possibly referred event* of $m$ as

$$\{(w_k, \varphi^\alpha_k)\},$$

where $\varphi^\alpha_k$ is the weight of word $w_k$, as follows.

(1) We summarize a set of source units $S_i$ that are linked by each mapped unit $o_i$ in $R^*$. *I.e.*, $S_i = \{s|(s, o) \in R^*, o = o_i\}$. We then build a bag-of-words model $S_{b,i}$ from the documents of $S_i$, where $S_{b,i}$ is the summarization of the bag-of-words model of each document in $S_i$ using tf-idf weighting.

(2) For each $S_{b,i}$ of source units, we compute the similarity between microblog $m$ and $S_{b,i}$, which is $sim(m, S_{b,i})$.

(3) We sum up the weight of words from each mapped unit $o_i$ using $sim(m, S_{b,i})$ as follows. Note that we build a bag-of-words model for each $o_i$ as $\{(w_k, \varphi^{o_i}_k)\}$. Therefore, the weight of each word $w_k$ is

$$\varphi^\alpha_k = \sum_i sim(m, S_{b,i}) \cdot \varphi^{o_i}_k. \tag{3}$$

We then rank candidates of referred events of $m$ by the composed event by

$$e' = \arg\max_{e \in E} sim(e, \{(w_k, \varphi^\alpha_k)\}),$$

where $E$ is a set of candidate events referred to by $m$.

### 5.2 Composition for Implicit and Explicit References

Since referred events can be different according to reference type, given a microblog $m$, we compute the probability of reference types of $m$ by using Eq. (1) and compose the referred event of $m$ by merging the composition for implicit reference and that for explicit reference based on probabilities of reference types, which is

$$\{(w_k, \varphi_k)\},$$

as follows.

(1) We build the composition $\{(w_k, \varphi^\alpha_k)\}$ for implicit references by using Eq. (3) and build the composition $\{(w_k, \varphi^\beta_k)\}$ for explicit references by using Eq. (6), which is a baseline method of microblog similarity and was shown to be effective for explicit references in experiments (as shown in **Fig. 3** (c)).

(2) We combine the two compositions $\{(w_k, \varphi^\alpha_k)\}$ and $\{(w_k, \varphi^\beta_k)\}$ with the consideration of probabilities of reference types as follows. We normalize the weight of words in each composition by using max weight and acquire $\{(w_k, \varphi'^\alpha_k)\}$ and $\{(w_k, \varphi'^\beta_k)\}$, which are normalized from $\{(w_k, \varphi^\alpha_k)\}$ and $\{(w_k, \varphi^\beta_k)\}$, respectively. We then sum up the weight of each word $w_k$ by

$$\varphi_k = p(t\text{=“}implicit\ ref\text{''}|m)\varphi'^\alpha_k + p(t\text{=“}explicit\ ref\text{''}|m)\varphi'^\beta_k, \tag{4}$$

where $p(t|m)$ is the probability of reference type $t$ given $m$ from Eq. (1). We then rank candidates of referred events of $m$ by

$$e' = \arg\max_{e \in E} sim(e, \{(w_k, \varphi_k)\}). \tag{5}$$

## 6. Experimental Results

We investigated implicit and explicit reference cases in real-world microblog data. We described baseline methods for event identification. We discussed the experimental results of event identification and composition of implicit references and gave a case study. **Table 1** shows the summary of experimental data.

### 6.1 Microblog Annotation

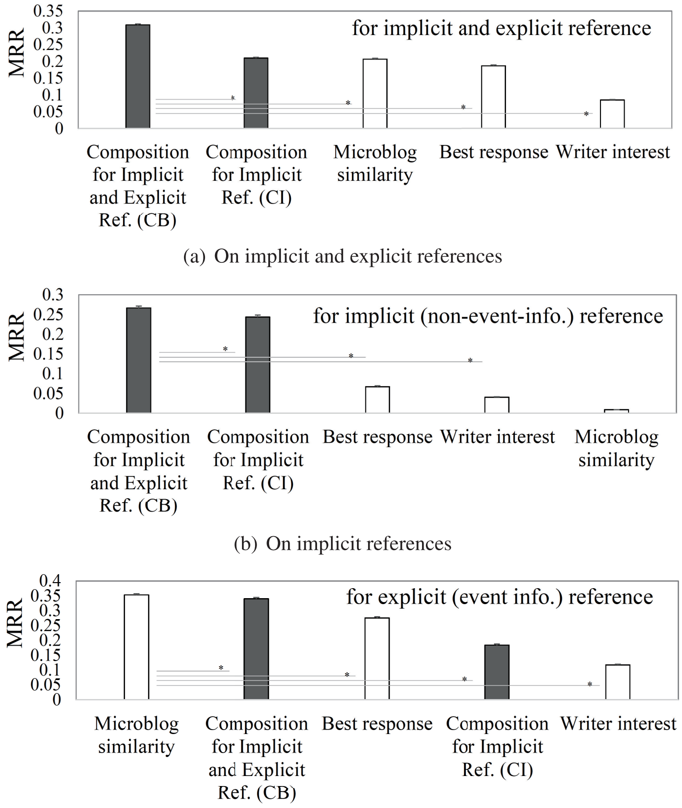
We specifically collected 15,996 tweets (microblogs) from Twitter, each of which referred to one of 231 events, during July 19-August 23, 2016, as follows.

- We collected events by extracting news-like tweets posted by a news-feeder account (e.g., “@cnnbrk”) from Twitter. We made sure that each event $e$ was from a news article, which was linked by some news-like tweet $m_e$ via hyperlink. Note that we assume a one-to-many relationship between an event and news articles (e.g., news articles from CNN [*3]); thus, an event must correspond to at least one news article.
- We then collected tweets (microblogs) that referred to (linked to) each event $e$ by collecting tweets, which were posted (by Twitter-users) for replying the news-like tweet $m_e$ of $e$.

We judged the reference type of tweet (microblog) $m$ referring to $e$ by using the following rules.

- The reference type is explicit (event info.) reference if $m$ contained the information of $e$.
- The reference type is implicit (non-event info.) reference if $m$ did not contain the information of $e$.

[*3] http://edition.cnn.com/.




(a) On implicit and explicit references

(b) On implicit references

(c) On explicit references

**Fig. 3** Comparison of performance of methods (+SEM).

**Table 1** Data summary.

| Data | # |
|---|---|
| Tweets (microblogs) | 15,996 |
| Events | 231 |
| Implicit (non-event-info.) references | 6,810 |
| Explicit (event info.) references | 9,186 |
| Writers | 10,523 |
| Writers' past microblogs | 506,904 |
| Feeling terms | 2,628 |

**Table 2** Annotation results on 15,996 tweets referring to 231 events.

| | Mean per event | Std. error |
|---|---|---|
| Tweets referring to | 69.25 | 6.31 |
| Implicit references | 29.48 | 3.65 |
| Explicit references | 39.77 | 4.26 |

As shown in **Table 2**, the annotation results showed that the number of implicit references was not small (29.48 tweets per event and the ratio of implicit and explicit references was 0.74:1). The result implied that predicting reference type of a microblog and using the predicted type to identify the referred event were required because the possibility of appearance of each type were not small.

### 6.2 Baseline Method

We used the following baseline method for ranking referred events from previous studies.

**Microblog similarity** (Ms) [1]: given microblog $m$, the best referred event $e'$ is ranked according to the similarity between the content of $m$ and event's content. I.e.,

**Table 3** Parameters used in experiments.

| Parameter | Value |
|---|---|
| # of candidate events | 231 |
| # of top-k words (of a bag-of-words model) | 100 |
| # of clusters for mapped units | 15 |
| # of clusters for the corresponding source units | 4 |
| # of iterations of running k-means clustering | 1,000 |
| ratio of # of top-k links | 0.1 |

$$e' = \arg\max_{e \in E} sim(e, m), \tag{6}$$

where $sim(e, m)$ is the cosine similarity between the bag-of-words model trained by $e$'s content using tf-idf weighting and that model trained by $m$ and $E$ is a set of candidate events.

**Best response** (Br) [2]: given $m$, the best $e'$ is ranked according to the top-k links $R'$ from a set of links $R$. I.e.,

$$e' = \arg\max_{e \in E} sim(e, R'), \tag{7}$$

where $R'$ is a bag-of-words model trained by $\{o|(s, o)$ is a top-$k_1$ link according to $sim(m, s)$, $o$ is a document$\}$ and we set $k_1 = 0.1|R|$ in experiments.

**Writer interest** (Wi) [4]: given $m$, the best $e'$ is ranked according to writer's interest $U$, where $U$ is a bag-of-words model trained by the profile $p$ of the writer of $m$ and a sequence of past $k_2$ microblogs $\{m_1, m_2, ...\}$ of the writer and we set $k_2$=50 in experiments. I.e.,

$$e' = \arg\max_{e \in E} sim(e, U). \tag{8}$$

**Experiment setting**. We list some parameters used in experiments in **Table 3** and experiment settings as follows. (1) We mea-

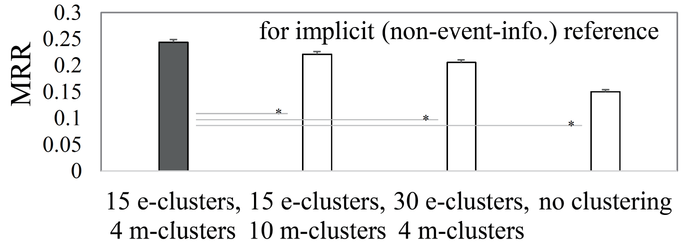



**Fig. 4** Performance of composition for implicit references with varying cluster sizes (+SEM).

sured the performance of event identification by using the mean reciprocal rank (*MRR*), which is

$$MRR = \frac{1}{|D|}\sum_{i=1}^{|D|} 1/rank_i,$$

where $D$ is a set of tests and $rank_i$ is the ranked position of the ground-truth referred event compared with other candidates for a microblog by score at test $i$. Note that the ground-truth event is one of candidate referred events. (2) We ran 4-fold cross-validation for the given data and we learned the composition for both implicit and explicit references (denoted as $CB$) in Section 5.2 and the composition for implicit reference (denoted as $CI$) in Section 5.1 by training data. (3) The cluster size for mapped units and for source units was set 15 and 4, respectively. The feeling terms we used in Eq. (1) were from this vocabulary list [*4].

### 6.3 Performance of Event Identification

We evaluated each method on event identification as follows. **Overall performance**. As shown in Fig. 3 (a), the performance of MRR of our proposed $CB$ was 0.309 and that of the top-one baseline method of microblog similarity was 0.207. I.e., $CB$ achieved the performance of MRR 1.494 times compared with baseline method [*5]. The result suggested that composing event for both implicit and explicit references based on the predicted reference type was important. We then investigated the performance on implicit or explicit references in detail.

**Performance on implicit references**. For implicit references, both the proposed $CB$ and only $CI$ were effective than baseline methods as shown in Fig. 3 (b). E.g., $CB$ achieved the performance of MRR 3.981 times compared with the top-one baseline method of best response. We also observed that the performance of $CI$ was very close to that of $CB$. The result suggested that composing event for implicit references was essential and we modeled that composition well. We also found that using writer's past microblogs were not useful for implicit references.

**Performance on explicit references**. For explicit references, our $CB$ was slightly worse than the baseline of microblog similarity, *cf.* the MRR difference was only 0.0135, as shown in Fig. 3 (c). We also observed that $CI$ did not performed well for explicit references. The reason might be that $CI$ preferred to link terms, which were other than event's information. However, the terms (e.g., names of subjects) used in explicit reference might

*4 https://www.vocabulary.com/lists/12827.
*5 We showed that the experimental results were trustable by statistical hypothesis testing as shown in Appendix A.2.

**Table 4** Performance of prediction of reference type.

| Reference type | Prediction accuracy |
|---|---|
| Explicit reference | 66.7 % |
| Implicit reference | 58.3 % |

not be intended to link others. The result suggested that referred events for explicit and implicit references were different and we need the composition for explicit reference and that composition could be modeled by using keywords of the posted microblog (e.g., proper nouns of subjects)

We also showed the accuracy of prediction of reference type in **Table 4**. The result showed that our method might infer the reference type of a microblog on some extent. Therefore, we could compose an event with balanced weight of words for the implicit and explicit references. Note that, for event identification, we used the probability of reference type computed by Eq. (1). The probability value was not a binary-value indicating either implicit or explicit reference type as used in experiments of Table 4.

### 6.4 Performance of Composition for Implicit References

As shown in **Fig. 4**, we tuned the performance of the composition for implicit reference ($CI$) by examining the effect of sizes of clusters of mapped units of links (denoted as e-clusters) and sizes of clusters of source units of links (denoted as m-clusters) when finding similar links. Intuitively, we regarded a cluster of mapped units as a type of events (e.g., "gun abuse") and regarded a cluster of source units as a type of opinions or feeling (e.g., "mysterious"). For implicit references, we observed that modeling $CI$ by setting 15 e-clusters and setting 4 m-clusters for each e-cluster was better than other assignments of more e-clusters or more m-clusters. The result suggested that modeling the composition for implicit references with reasonable parameters might be useful. One reason might be that people reply to a type of events using limited types of responses. E.g., modeling $CI$ with limited types of source units of links (m-clusters was set 4) was better, when e-clusters was 15.

### 6.5 Discussion and Limitation

According to the experiments in Sections 6.3 and 6.4, we summarized the findings as follows. (1) Our method was effective for implicit references. (2) Predicting reference type was essential for identifying implicitly or explicitly referred events together. The reason might be that referred events were different according to reference type. (3) Composing event for implicit references by using similar links was effective and finding similar links with reasonable parameters might be useful. One reason

**Table 5** Case study on implicit (non-event-info.) references.

| **Event** *e* | | | *e*'s info. "Florida investigating non-travel related case of Zika..." |
|---|---|---|---|
| **Microblogs of implicit reference to** *e* | **Method** *CB* | **baseline** | **Difficulty** |
| "@cnnbrk global warming is going to..." | **.083** | .004 (*MRR*) | Opinion |
| "@cnnbrk leave it down there" | .01 | .004 | Opinion |
| "@cnnbrk this also outbroke in Brazil" | **.2** | .004 | Related event |
| "@cnnbrk frightening, but this was..." | **.067** | .004 | Feeling (emotion) |

might be that people reply a type of events using limited types of responses. We also listed some implicit reference cases for an event in our dataset and the difficulty in **Table 5**. (4) We found that our method was helpful for identifying opinions referring to events, compared with identifying emotions. This implied that we should increase the accuracy by considering more on personalization in our method.

The limitation of our method is listed as follows. (1) Given a link between a microblog $m$ and an mapped unit $e$, we did not check which part of $e$'s content the microblog $m$ actually refers to and we assumed $m$ refers to the whole part of $e$'s content. We may infer the referred part of $e$'s content by mining about the conversations (from past microblogs) of the writer of $m$ on those mapped units similar to $e$. (2) Word mismatch: people may use general terms (e.g., the man) to refer subjects of an event. Word mismatch may reduce the accuracy of the prediction of reference type and we may improve word mismatch problem by applying the methods of entity (subject) identification from documents [32], [33]. (3) We currently used the static cluster sizes for clustering links of source units and mapped units and we should use dynamical cluster sizes for clustering links because the posted microblogs of people on events is diverse.

## 7. Conclusion and Future Work

We addressed the problem of event identification on microblogs and focused on implicit references instances, where events are not referred to by event's information and are accordingly difficult to identify. We therefore propose to use links from microblogs that are able to connect opinions or feeling, which are irrelevant to event's information, to their referred events. We also predict reference type of a microblog, compose possible event for implicit references, and use the predicted reference type to compose event for implicit and explicit references. We surveyed 15,996 tweets referring events and found that the number of implicit references was not small. The experimental results suggested that our method was effective for implicit references, predicting reference type was essential for identifying implicitly or explicitly referred events together, and composing event for implicit references by using similar links was useful. In future work, the systematic analysis for links from microblogs, e.g., personalization of links, is required for accurately resolving implicit references. The on-line application may need to speed up the system, where there is huge amounts of microblog data from social network platforms.

**Acknowledgments** This work was supported in part by JSPS KAKENHI Grant Numbers 15H01718, 26700009, 16H02906, and 16K16156.

## References


[1] Stilo, G. and Velardi, P.: Efficient temporal mining of micro-blog texts and its application to event discovery, *Data Min. Knowl. Discov.*, Vol.30, No.2, pp.372–402 (2016).

[2] Paltoglou, G.: Sentiment-based event detection in Twitter, *JASIST*, Vol.67, No.7, pp.1576–1587 (2016).

[3] Moran, S., McCreadie, R., Macdonald, C. and Ounis, I.: Enhancing First Story Detection using Word Embeddings, *SIGIR 2016*, pp.821–824 (2016).

[4] Ritter, A., Mausam, Etzioni, O. and Clark, S.: Open domain event extraction from twitter, *KDD 2012*, pp.1104–1112 (2012).

[5] Weng, J. and Lee, B.: Event Detection in Twitter, *ICWSM 2011* (2011).

[6] Becker, H., Naaman, M. and Gravano, L.: Beyond Trending Topics: Real-World Event Identification on Twitter, *ICWSM 2011* (2011).

[7] Abdelhaq, H., Sengstock, C. and Gertz, M.: EvenTweet: Online Localized Event Detection from Twitter, *PVLDB*, Vol.6, No.12, pp.1326–1329 (2013).

[8] Sakaki, T., Okazaki, M. and Matsuo, Y.: Earthquake shakes Twitter users: Real-time event detection by social sensors, *WWW 2010*, pp.851–860 (2010).

[9] Wang, Q., She, J., Song, T., Tong, Y., Chen, L. and Xu, K.: Adjustable Time-Window-Based Event Detection on Twitter, *WAIM 2016*, pp.265–278 (2016).

[10] Jatowt, A., Antoine, É., Kawai, Y. and Akiyama, T.: Mapping Temporal Horizons: Analysis of Collective Future and Past related Attention in Twitter, *WWW 2015*, pp.484–494 (2015).

[11] Petrovic, S., Osborne, M. and Lavrenko, V.: Using paraphrases for improving first story detection in news and Twitter, *NAACL HLT 2012*, pp.338–346 (2012).

[12] Petrovic, S., Osborne, M. and Lavrenko, V.: Streaming First Story Detection with application to Twitter, *NAACL HLT 2010*, pp.181–189 (2010).

[13] Antoine, É., Jatowt, A., Wakamiya, S., Kawai, Y. and Akiyama, T.: Portraying Collective Spatial Attention in Twitter, *KDD 2015*, pp.39–48 (2015).

[14] Benhardus, J. and Kalita, J.: Streaming trend detection in Twitter, *IJWBC*, Vol.9, No.1, pp.122–139 (2013).

[15] Aiello, L.M., Petkos, G., Martín, C.J., Corney, D., Papadopoulos, S., Skraba, R., Göker, A., Kompatsiaris, I. and Jaimes, A.: Sensing Trending Topics in Twitter, *IEEE Trans. Multimedia*, Vol.15, No.6, pp.1268–1282 (2013).

[16] Nichols, J., Mahmud, J. and Drews, C.: Summarizing sporting events using twitter, *IUI 2012*, pp.189–198 (2012).

[17] Marcus, A., Bernstein, M.S., Badar, O., Karger, D.R., Madden, S. and Miller, R.C.: Twitinfo: aggregating and visualizing microblogs for event exploration, *CHI 2011*, pp.227–236 (2011).

[18] Chakrabarti, D. and Punera, K.: Event Summarization Using Tweets, *ICWSM 2011* (2011).

[19] Stefanidis, A., Crooks, A. and Radzikowski, J.: Harvesting ambient geospatial information from social media feeds, *GeoJournal*, Vol.78, No.2, pp.319–338 (online), DOI: 10.1007/s10708-011-9438-2 (2013).

[20] Ghiassi, M., Skinner, J. and Zimbra, D.: Twitter brand sentiment analysis: A hybrid system using n-gram analysis and dynamic artificial neural network, *Expert Syst. Appl.*, Vol.40, No.16, pp.6266–6282 (2013).

[21] Conover, M., Gonçalves, B., Ratkiewicz, J., Flammini, A. and Menczer, F.: Predicting the Political Alignment of Twitter Users, *SocialCom 2011*, pp.192–199 (2011).

[22] Allan, J.(ed.): *Topic Detection and Tracking: Event-based Information Organization*, Kluwer Academic Publishers, Norwell, MA, USA (2002).

[23] Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., Ounis,

I. and Shrimpton, L.: Can Twitter Replace Newswire for Breaking News?, *ICWSM 2013* (2013).

[24] Zhao, W.X., Jiang, J., Weng, J., He, J., Lim, E., Yan, H. and Li, X.: Comparing Twitter and Traditional Media Using Topic Models, *ECIR 2011*, pp.338–349 (2011).

[25] Watanabe, K., Ochi, M., Okabe, M. and Onai, R.: Jasmine: A real-time local-event detection system based on geolocation information propagated to microblogs, *CIKM 2011*, pp.2541–2544 (2011).

[26] Hua, T., Chen, F., Zhao, L., Lu, C. and Ramakrishnan, N.: Automatic targeted-domain spatiotemporal event detection in twitter, *GeoInformatica*, Vol.20, No.4, pp.765–795 (2016).

[27] Zhang, Y. and Qu, Z.: A novel method for online bursty event detection on Twitter, *ICSESS 2015*, pp.284–288 (online), DOI: 10.1109/ICSESS.2015.7339056 (2015).

[28] Zhou, X. and Chen, L.: Event detection over twitter social media streams, *VLDB J.*, Vol.23, No.3, pp.381–400 (2014).

[29] Sankaranarayanan, J., Samet, H., Teitler, B.E., Lieberman, M.D. and Sperling, J.: TwitterStand: news in tweets, *SIGSPATIAL 2009*, pp.42–51 (2009).

[30] Domeniconi, G., Moro, G., Pasolini, R. and Sartori, C.: A Study on Term Weighting for Text Categorization: A Novel Supervised Variant of tf.idf, *DATA 2015*, pp.26–37 (2015).

[31] Jain, A.K.: Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, Vol.31, No.8, pp.651–666 (2010).

[32] Shen, W., Wang, J. and Han, J.: Entity Linking with a Knowledge Base: Issues, Techniques, and Solutions, *IEEE Trans. Knowl. Data Eng.*, Vol.27, No.2, pp.443–460 (2015).

[33] Lu, J., Kato, M.P., Yamamoto, T. and Tanaka, K.: Entity Identification on Microblogs by CRF Model with Adaptive Dependency, *IEICE Transactions*, Vol.99-D, No.9, pp.2295–2305 (2016).

[34] Joachims, T., Finley, T. and Yu, C.J.: Cutting-plane training of structural SVMs, *Machine Learning*, pp.27–59 (2009).

[35] Vedaldi, A.: A MATLAB wrapper of SVM$^{\text{struct}}$, available from ⟨http://www.robots.ox.ac.uk/~vedaldi//svmstruct.html⟩ (2011).

# Appendix

## A.1 Training Prediction of Reference Type

Given a model of prediction of reference type $(X, Y)$, we learned the optimal weight $\mathbf{w}' \in \mathbb{R}^n$ of features in Eq. (1) based on the objective function of Hinge loss [34], i.e.,

$$\mathbf{w}' = \arg\min_{\mathbf{w}\in\mathbb{R}^n}\left(\epsilon\|\mathbf{w}\| + \sum_{i=1}^{\nu}\max_{y\in Y}\left(l(y) + \mathbf{w}^T\phi(x^i, y)\right) - \mathbf{w}^T\phi(x^i, y^i)\right),$$

where $n$ is the number of features, $\nu$ is the number of training instances of microblogs, $y \in Y = \{$"explicit ref","implicit ref"$\}$ is a candidate reference type, $x^i$ is the $i$-th microblog, $y^i \in Y$ is the ground-truth reference type for $x^i$, $l(y)$ is the difference between $y$ and $y^i$, which is $|\{1|y \neq y^i\}|$, and $\phi(x^i, y)$ is a vector of feature values of $y$ given $x^i$. We used the tool [35] to obtain the optimal weight $\mathbf{w}'$.

## A.2 Statistical Hypothesis Testing

For the overall performance of Fig. 3 (a), the Analysis of variance (Anova) result ($F(4, 79975) = 854.3$, $p < 2e^{-16}$) showed that at least the effects of two of five methods ($CB$, $CI$, $Ms$, $Br$, and $Wi$, which are denoted by abbreviation) were significantly different. We then ran the Tukey's HSD test, a post-hoc test, and the results ($CB$-$CI$ $p = 0.0$, $CB$-$Ms$ $p = 0.0$, $CB$-$Br$ $p = 0.0$, and $CB$-$Wi$ $p = 0.0$) showed that our method $CB$ was significantly different with other methods.

For the performance on implicit references of Fig. 3 (b), the Anova result ($F(4, 34045) = 1229$, $p < 2e^{-16}$) showed that at least the effects of two of five methods were significantly different. Then, the results of the Tukey's HSD test ($CB$-$CI$ $p = 2.86e^{-5}$, $CB$-$Ms$ $p = 0.0889$, $CB$-$Br$ $p = 0.0$, and $CB$-$Wi$ $p = 0.0$) showed that our method $CB$ was significantly different with other methods except $Ms$.

For the performance on explicit references of Fig. 3 (c), the Anova result ($F(4, 45925) = 714.9$, $p < 2e^{-16}$) showed that at least the effects of two of five methods were significantly different. Then, the results of the Tukey's HSD test ($CB$-$CI$ $p = 0.0$, $CB$-$Ms$ $p = 0.0$, $CB$-$Br$ $p = 0.0$, and $CB$-$Wi$ $p = 0.0$) showed that our method $CB$ was significantly different with other methods.

For the performance of Fig. 4, the Anova result ($F(3, 27236) = 73.31$, $p < 2e^{-16}$) showed that at least the effects of two of four clustering settings, (15 e-clusters,4 m-clusters), (15,10), (30,4), and *no* clustering, were significantly different. Then, the results of the Tukey's HSD test ($(15, 4)$-$(15, 10)$ $p = 0.0033$, $(15, 4)$-$(30, 4)$ $p = 0.0$, and $(15, 4)$-*no* $p = 0.0$) showed that our used setting $(15, 4)$ was significantly different with other settings.

**Jun-Li Lu** received his M.S. degree in Department of Electrical Engineering from National Taiwan University, Taiwan, in 2010. From 2013, he became a Ph.D. student in Graduate School of Informatics, Kyoto University, Japan. His research interests include text mining and social-network mining.

**Makoto P. Kato** received his B.S., M.S., and Ph.D. degrees from Kyoto University, Japan, in 2008, 2009, and 2012, respectively. He is currently an assistant professor at Kyoto University. He serves as a program co-chair of NTCIR from 2015. He is also a program committee member of the SIGIR and WSDM conferences. His research interests include interactive information retrieval, user behavioral analysis in search, and search intent detection.

**Takehiro Yamamoto** received his M.S. and Ph.D. degrees in Informatics from Kyoto University, Japan, in 2008 and 2011, respectively. He has worked as a research fellow at Kyoto University from 2012 to 2013. Since 2014 he has been an assistant professor at Kyoto University. His research interests include information retrieval, human-computer interaction, and web mining.

**Katsumi Tanaka** received his B.S., M.S., and Ph.D. degrees in Information Science from Kyoto University, Japan, in 1974, 1976 and 1981, respectively. In 1986, he joined Department of Instrumentation Engineering, Faculty of Engineering at Kobe University, Japan, as an associate professor. In 1994, he became a full professor in Faculty of Engineering, Kobe University. Since 2001, he has been a professor of Graduate School of Informatics, Kyoto University. His research interests include database theory and systems, web information retrieval, and multimedia retrieval.

(Editor in Charge: *Takayuki Tamura*)