

Early Discovery of Disappearing Entities in Microblogs

Satoshi Akasaki*
Yahoo Japan Corporation
sakasaki@yahoo-corp.jp

Naoki Yoshinaga
Institute of Industrial Science,
The University of Tokyo
ynaga@iis.u-tokyo.ac.jp

Masashi Toyoda
Institute of Industrial Science,
The University of Tokyo
toyoda@iis.u-tokyo.ac.jp

Abstract

We make decisions by reacting to changes in the real world, particularly the emergence and disappearance of impermanent entities such as restaurants, services, and events. Because we want to avoid missing out on opportunities or making fruitless actions after those entities have disappeared, it is important to know when entities disappear as early as possible. We thus tackle the task of detecting disappearing entities from microblogs where various information is shared timely. The major challenge is detecting uncertain contexts of disappearing entities from noisy microblog posts. To collect such disappearing contexts, we design time-sensitive distant supervision, which utilizes entities from the knowledge base and time-series posts. Using this method, we actually build large-scale Twitter datasets of disappearing entities¹ for English and Japanese. To ensure robust detection in noisy environments, we refine pretrained word embeddings for the detection model on microblog streams in a timely manner. Experimental results on the Twitter datasets confirmed the effectiveness of the collected labeled data and refined word embeddings; the proposed method outperformed a baseline in terms of accuracy, and more than 70% of the detected disappearing entities in Wikipedia are discovered earlier than the update on Wikipedia, with the average lead-time is over one month.

1 Introduction

Our daily actions depend on the state of the real world and its changes, especially changes in the entities available to us. Among the various changes, the beginning of entities, or emerging entities (Akasaki et al., 2019) such as new songs, movies, and events, are useful for understanding

*Part of this work was carried out during the period the author was at the University of Tokyo.

¹We release the datasets (tweet IDs) used in the experiments to promote reproducibility. URL

Westfield State University Athletics - **Dave Laing**
Named Little East Coach of the Year URL

I'm so *sad to hear* that **Dave Laing** has *died*. Dave was a very accomplished music industry journalist.

Don't forget you can follow our posts and leave feedback about recent visits on our **Google+** page!

Can't believe **Google+** is being *shut down*. It's like when they just pulled Google Friends Connect...

PHOTOS!! **Martin Tower**'s mosaic artwork goes on public display URL URL

Here's your *Demolition* Day Planner for **Martin Tower**. A brief, stray shower can't be ruled out...

Pilot lineup for 2018 **Red Bull Air Race World Championship**: Britain is back! URL

Red Bull Air Race World Championship *will not continue* after 2019. URL

Happy new year everyone! I hope this year will be better for all of us and **Pristin** will achieve a lot!

Pristin to *disband* after 2 years promoting as a group K-Netz express how *raged* they are...

Figure 1: Tweets of **entities** with non-disappearing and disappearing contexts.

the trends in interests. At the same time, we want to be made aware of the end or disappearance of entities, such as closing stores or discontinuing services (Figure 1), as soon as possible so that we can avoid missing out on opportunities or taking fruitless actions after the entities are no longer available. Portal sites also need to address this issue by quickly removing unavailable companies, locations, and events from their services. Moreover, it is important to collect these entities to maintain knowledge bases (KBs) in which information about the entities is continuously accumulated.

Whereas disappearing entities remain to be discovered, studies on discovering out-of-KB or

emerging entities (Lin et al., 2012; Färber et al., 2016; Wu et al., 2016; Akasaki et al., 2019) have been successful to an extent, as most of these entities have distinct names and can be characterized by mentions to their unseen names. In contrast, we do not have such clues to detect disappearing entities, since most of them continue to be mentioned with the same names even after they actually disappear.

Given such a situation, we take on the new task of discovering disappearing entities from microblogs where news and personal experiences are widely shared. To detect the entities' disappearance, we exploit the specific expressions that people use when mentioning disappearing entities in microblogs (Figure 1) (§ 2). By capturing these contexts, we can discover a variety of entities in the early stage even before they disappear. To develop a dataset of disappearing entities and contexts, we use time-sensitive distant supervision (Akasaki et al., 2019), which collects specific contexts of entities by utilizing KB entities and timestamps of microblogs. Because this method requires the timing of the desired contexts, we extract the year of disappearance for each entity described in Wikipedia and incorporate it into the distant supervision (§ 4.1).

We then train a named entity recognition (NER) model on the collected entities and contexts to discover and type disappearing entities. However, the NER model performs poorly for microblogs (Derzynski et al., 2017) where posts are short and noisy, and the training data collected by the distant supervision also contains noise, making it difficult to train a reliable model. We address this issue by considering that multiple posts are likely to mention the target entity when it disappears. Concretely, we utilize these posts to refine pretrained word embeddings and incorporate them into the NER model (§ 4.2). This enables the model to consider the tokens that frequently appear among multiple posts and to recognize disappearing entities robustly.

To evaluate our method, we built large-scale English and Japanese datasets from Twitter using the proposed time-sensitive distant supervision method (§ 5.1). The experimental results (§ 5.4) demonstrated that our method outperformed the baseline, which simply collected the latest burst of posts about the disappearing entities as the disappearing contexts using the original version of time-sensitive distant supervision and used them to train the NER model. In addition, the evaluation of relative recall indicated that our method successfully found

more than 70% of the target disappearing entities in Wikipedia. Except for entities whose disappearances are sudden and receive media exposure (e.g., persons), our method detected entities such as services, facilities, and events on average more than 100 days earlier than the update of the disappearance in Wikipedia.

2 Definition of Disappearing Entity

In this section, we define the meaning of the term *disappearing entity* in this study. We consider the entities' disappearance to be the disappearance of its existence from the real world or official announcement of its discontinuation, with reference to the list of ending entities in Wikipedia.²

To reach a solid definition of disappearing entities, we refer to the definition of emerging entities (Akasaki et al., 2019). They reported that emerging entities have a specific process from their first appearance to when they become known to the public; during this process, they appear with specific expressions, *i.e.*, emerging contexts. We define disappearing entities and contexts by focusing on the fact that specific expressions that indicate plans and signs of disappearance appear in contexts not only at the time of disappearance but also in the process leading up to that time as follows:

Disappearing contexts. *Contexts in which the writers assumed the readers do not know the disappearance of the entities.*

Non-disappearing contexts. *Contexts other than disappearing contexts.*

Disappearing entities. *Entities still being observed in disappearing contexts.*

Figure 1 lists examples of disappearing entities and their disappearing and non-disappearing contexts. By properly identifying disappearing contexts, we can detect mentioned disappearing entities in the early stages even before they actually disappear. This is especially important for entities such as stores that are closing or services and events that are ending so that we can take action before they are gone.³ We later confirm the solidness of these definitions by evaluating the inter-rater agreement of disappearing entities acquired from microblogs and by demonstrating that the disap-

²https://en.wikipedia.org/w/index.php?title=Category:Endings_by_year&from=2000

³Note that how early the disappearance is detected depends on the entity type.

pearing entities can be detected before the update of the disappearance in Wikipedia (§ 5.4).

3 Related Work

There are no studies that detect disappearing entities in a timely manner. We briefly review the existing studies related to our task.

3.1 Entity Extraction

Because existing entity-related tasks such as NER (Nadeau and Sekine, 2007; Lample et al., 2016; Akbik et al., 2018, 2019; Yu et al., 2020; Cui et al., 2021) and entity linking (Shen et al., 2014; Kolitsas et al., 2018; Martins et al., 2019; Oba et al., 2022) do not take into consideration whether the entities have already disappeared from the real world, it is difficult to detect disappearing entities by using these techniques simply.

As a complementary task to detecting disappearing entities, Akasaki et al. (2019) attempted to find emerging entities.⁴ They exploited the fact that people use expressions that suggest novelty when mentioning emerging entities, and defined the emerging entities based on these expressions (contexts). They proposed a variant of the distant supervision method (Mintz et al., 2009) called time-sensitive distant supervision to collect emerging contexts efficiently by utilizing KB entities and microblog timestamps. They then developed an NER model using the obtained data to detect emerging entities.

Although this study has some similarities to ours, finding emerging entities is a more easy task since the initial occurrences of entities could be easily regarded as emerging contexts. We cannot apply the analogous assumption to disappearing entities because entities will be mentioned even after their disappearance. We thus propose another method of time-sensitive distant supervision for disappearing entities (§ 4.1) to properly capture disappearing contexts. We compare our method to Akasaki et al. (2019) in the experiments and demonstrate its effectiveness. Also, unlike them, we experimented with multiple languages (English and Japanese) and even jointly performed entity typing rather than the cascaded pipeline (Akasaki et al., 2021) for broader applications.

⁴Although there are other studies that detect ‘emerging’ entities like Hoffart et al. (2014) and Wu et al. (2016), they actually target out-of-KB entities.

3.2 Temporal Event Extraction

As part of information extraction, researchers have tackled the task of extracting an event (*e.g.*, birth of a person) and its predefined attributes and arguments (such as time and places) from the text (Ritter et al., 2012; Nguyen and Grishman, 2015; Liu et al., 2020). Although it is possible to use these techniques to detect disappearing entities, they rely on manual annotations (Doddington et al., 2004; Aguilar et al., 2014) and only a few entity types such as ‘person’ and ‘military conflict’ are chosen for the target entity types; it is thus unrealistic to cover a wide range of entity types.

KBP2011 (Ji et al., 2010; McClosky and Manning, 2012) introduced the temporal slot filling task, in which the duration of an event is identified from the given text, entities (*e.g.*, Steve Jobs), and their events (*e.g.*, become CEO). The events are attributes defined in Freebase, and they include the disappearance of certain entities such as when a person dies. However, the types of entities handled in this task are also only a few (‘person’ and ‘organization’). Moreover, although target entities are given in advance for this task, we have to detect mentions of the entities in the settings of microblogs.

4 Proposed Method

To accurately collect contexts of disappearing entities (§ 2), we adopt the framework of time-sensitive distant supervision (Akasaki et al., 2019), which efficiently collects specific contexts of given entities utilizing microblog timestamps. Here, we utilize the timings of the entities’ disappearance described in Wikipedia to refine the original time-sensitive distant supervision since the original method is designed for emerging entities and it is difficult to apply the method to disappearing entities directly.

To ensure that a detection model can make robust predictions using the noisy dataset constructed using distant supervision and for noisy microblog posts, we refine pretrained word embeddings to acquire features from multiple occurrences of disappearing entities and feed them into the detection model. As a microblog, we target Twitter, where various sources including news articles and personal posts, are extensively and instantly shared.

4.1 Time-sensitive Distant Supervision

The original time-sensitive distant supervision involves two main steps (Akasaki et al., 2019); 1) Collecting candidates of emerging entities by as-

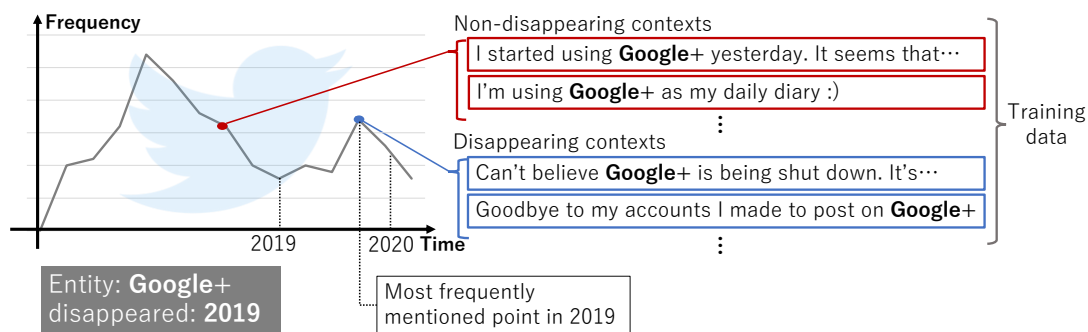


Figure 2: Time-sensitive distant supervision; given an entity, its non-disappearing and disappearing contexts are collected from microblogs by utilizing the year of the disappearance. We then train a detection model using the obtained contexts.

sociating the time-stamps of registration with titles of articles in Wikipedia, and 2) collecting contexts of emerging entities by retrieving early-stage microblog posts posted before the time-stamps of registration. To prevent overfitting, non-emerging contexts for the same entities with enough time elapsed are collected as negative examples.

While this method was able to gather contexts of emerging entities by collecting the posts when they first appeared, it is difficult to directly apply this implementation of time-sensitive distant supervision to disappearing entities because mentions of entities are made even after they actually disappear, and thus the timing of disappearance is not clear. Therefore, we explicitly feed the timing of the entities’ disappearance extracted from Wikipedia to time-sensitive distant supervision to collect disappearing entities and their contexts more accurately (Figure 2). The specific procedure is as follows:

Step 1: Collecting disappearing entities We first collect candidates of disappearing entities while excluding noise. To collect only entities that have actually disappeared, we refer to the list of ending entities in Wikipedia¹ and gather the titles of articles, categories, and their year of disappearance. We excluded entities whose year of their first appearance on Twitter was the same as the year of their disappearance since they could be emerging entities and contaminate the contexts. We also remove entities that have the ambiguity page so that the contexts are not contaminated by homographic entities which share the same namings with other entities (e.g., “Go” can refer to a programming language, a board game, or a verb). To acquire entity types, we manually map the category of the article to the coarse-grained type; for example, the entity ‘Daft Punk’ is mapped to the type ‘GROUP.’ We

adopted this method instead of Akasaki et al. (2019, 2021), which utilized the DBpedia mappings because of few mappings of disappearing entities.

Step 2: Collecting disappearing contexts In contrast to emerging entities, where the first appearance of the entity is often the emerging context, the latest posts of disappearing entities contain noisy unrelated contexts because they continue to be mentioned in microblogs even after they have disappeared. Therefore, for each collected entity, we utilize the year of disappearance and frequency of appearance on Twitter to gather their disappearing contexts. Specifically, we randomly collect k posts of the day with the highest number of occurrences in the given year, assuming that the timing that received the most attention in the year of the disappearance includes the disappearing contexts.

For each collected entity, Akasaki et al. (2019) extracted contexts that differed from the positive examples as negative examples to avoid overfitting the detection model when recognizing mentions of positive examples. Thus, we similarly collected random k non-disappearing contexts as negative examples from posts prior to the year in which we collected positive examples for each entity. This also enables the model to discriminate between disappearing contexts and other contexts.

We adopt the BILOU scheme (Ratinov and Roth, 2009) for the NER tags; we label the disappearing entities in the positive examples with BILU and their entity types and label the rest with O.

4.2 Finding Disappearing Entities

We train an NER model for recognizing and typing disappearing entities using the collected data. Because we target short and noisy microblog posts and build the dataset using distant supervision, it is

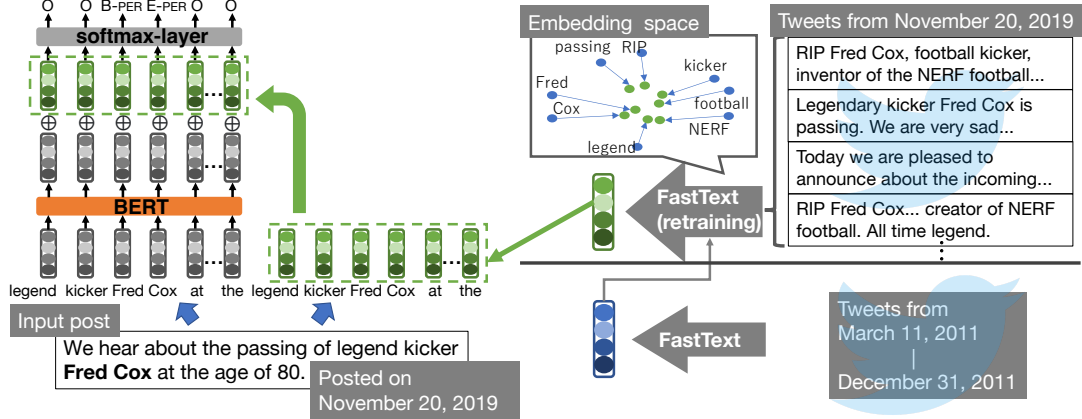


Figure 3: Sequence labeling with refined word embeddings; we refine pretrained word embeddings using the Twitter stream on the day of the input post, and feed them into the NER model for robust training and prediction.

difficult to train the model stably. Thus, we focus on the fact that multiple posts mentioning the disappearance of the entities often appear when entities disappear from microblogs. By obtaining features from these posts, we can enhance the training and prediction of the model even with noisy microblog posts. To do this, we propose an unsupervised method of refining pretrained word embeddings as features from multiple posts of a Twitter stream. We also devise a sequence labeling model for finding disappearing entities using the refined word embeddings as additional input (Figure 3).

4.2.1 Refining Pretrained Word Embeddings

Our objective is to extract features from multiple posts in microblogs. However, because the surface of the target entity is unknown at the prediction time of NER, it is difficult to collect only relevant posts of that entity from the massive Twitter stream. To address this issue, we use, as the additional embedding layers of our NER model, pretrained word embeddings further fine-tuned on the posts on the day of detecting disappearing entities. This enables the refined word embeddings to reflect the tokens and their co-occurrences in the Twitter stream of the target day without the need for post selection. The specific procedure is as follows:

First, we train the base word embeddings v_{base} using the posts prior to the period in which we collected the data in § 4.1. We use fastText (Bojanowski et al., 2017), a method of constructing word embeddings, to deal with unknown words.

Next, we use the Twitter stream from each date d of the posts collected in § 4.1 to retrain the word vector v_{base} for obtaining v_d . The resulting v_d can be interpreted as capturing the temporary semantic

change of v_{base} on date d , and it can be treated as an auxiliary input of models for various tasks. Refining takes about 5 minutes per 2M tweets and can be done in parallel without using GPU.

4.2.2 NER with Refined Word Embeddings

We adopt BERT (Devlin et al., 2019) model with softmax layer as an NER model. Since the pretrained BERT models available on the web may have learned text with future timestamps than the training data (§ 4.1) and suffer from data leaks, we pretrain BERT from scratch using tweets from before the time at which we construct the training and test data.

Based on this model, we input each post of the constructed data into BERT and obtain the sequence of hidden states. We then concatenate those hidden states with the refined word embeddings v_d corresponding to the date of the input post. We finally feed them into the softmax layer and predict labels. This enables the model to consider the global information of the given Twitter stream other than the input post.

5 Experiments

We performed our task of discovering disappearing entities using datasets built from Twitter.

5.1 Data

We built the Twitter datasets by using our time-sensitive distant supervision. We targeted English and Japanese, which are the top two languages used on Twitter (Alshaabi et al., 2021), and use our archive of Twitter posts that were retrieved⁵

⁵Timelines of 26 popular Japanese users starting from 2011 have been continuously collected using user_timeline

TYPE	# entities	# posts	examples of disappearing context (truncated)
Wikipedia categories			
PERSON	780	24689	
Deaths	780	24689	Roger Ailes died of complications of a subdural hematoma after he fell at home, hit his head. ...to ski one day with Olympic Legend Stein Eriksen at Deer Valley. He passed yesterday at... RIP Dave Rosenfield : @USER executive, International League schedule maker, one-time...URL
CREATIVE WORK	975	24222	
American_television_series	381	12413	RT @USER: See what's coming up in the Diggnation Finale airing next week...
British_television_series	139	3650	Wait...is this episode supposed to be the Metalocalypse finale or the start...
Web_series	57	1679	Final Series of McLevy this week on @USER The Scotland office are missing cast & crew...
Others (36 types)	398	6480	Crying while watching the last episode of Packed to the Rafters :(such a great tv show
LOCATION	240	3545	
Buildings_and_structures	54	949	@USER: After 67 years Clemson House is gone in seconds. (This version is sped up) URL
Educational_institutions	36	647	Coleman University closing its doors after loss of accreditation URL
Restaurants	24	286	...food news of HASH Everyone loves. Lynn's Paradise Cafe abruptly serves its final meal...
Others (20 types)	126	1663	Sad news, crime sleuths. The National Museum of Crime and Punishment in D.C. will close at...
GROUP	1187	30384	
Musical_groups	453	11559	...Flyleaf anymore. Adam's not in Three Days Grace anymore. My Chemical Romance broke up...
Retail_companies	79	3187	The convenience store chain, My Local , is to be placed in administration,9 months after it was sold...
Airlines	60	1535	The second hand car giant Carcraft has gone bust, with the loss of around 500 jobs across the UK...
Others (68 types)	595	14103	The Foreign Policy Initiative , a right-leaning foreign-policy think tank, will cease operations...
EVENT	186	4940	
Sporting_events	111	2492	The Adidas Grand Prix in NY, replaced by Rabat meet in Diamond League series, is set to ...
Events	33	1462	Live music blow. MT @TomTilley: C3 confirms Big Day Out will NOT go ahead in 2015.
Sports_leagues	16	137	The cancellation of Women's Professional Soccer League in US is bad news for England & Team...
Others (2 types)	26	849	The Foreign Policy Initiative , a right-leaning foreign-policy think tank, will cease operations...
SERVICE&PRODUCT	709	17574	
Magazines	187	2998	was what inspired me to write for mags. RT @USER: Future closes Nintendo Gamer magazine
Internet_properties	131	5916	Google is going to shut down Orkut on September 30...Haven't seen that site since 2007 L...
Products_and_services	119	3450	The PlayStation 3 has ended production according to the official PlayStation Japan website.
Others (10 types)	272	5210	Just heard that Yorkshire Radio is no more. Great shame as @USER was pretty much the voice of...
TOTAL	3213	81850	

Table 1: Statistics of English disappearing entities and disappearing contexts in training data obtained from our Twitter archive by time-sensitive distant supervision.

TYPE	#ent.	#posts	TYPE	#ent.	#posts
PERSON	147	422	PERSON	73	220
CREATIVE WORK	10	23	LOCATION	42	114
LOCATION	46	113	GROUP	64	173
GROUP	103	270	EVENT	9	25
EVENT	8	19	SERVICE&PRODUCT	47	131
SERVICE&PRODUCT	43	114	TOTAL	235	663
TOTAL	357	961			

(a) English (b) Japanese

Table 2: Statistics of disappearing entities and disappearing contexts in test data.

using APIs. The archive contains more than 50B posts, 32% in English and 20% in Japanese, a trend similar to the actual data (Alshaabi et al., 2021). We tokenized posts using spaCy (ver. 2.0.12)⁶ with the en_core_web_sm model for English and MeCab (ver. 0.996)⁷ with ipadic (ver. 2.7.0) for Japanese. We removed URLs and usernames from the text.

API, while the user set has been iteratively expanded to those who were mentioned or retweeted by existing users.

⁶<https://spacy.io>

⁷<https://taku910.github.io/mecab>

In Step 1 of § 4.1, we collected article titles of ending entities in Wikipedia from 2012 to 2019 using the Wikipedia dump from June 20th, 2021. We undersampled the types PERSON and CREATIVE WORK to 1,000 entities as they are much larger than the other types. We then excluded entities as described and carried out Step 2 by setting k to 100, as in (Akasaki et al., 2019).

We split the collected data into training data (2012-2018) and test data (2019).⁸ For the training data, we obtained a total of 163,700 English and 150,204 Japanese tweets, including the same number of disappearing and other contexts for 3,213 English entities and 1,906 Japanese entities, respectively. For model selection, we used 10% of the training data as the development data.

For the test data, from the collected positive examples of disappearing entities, we randomly selected three posts for each entity and asked three annotators (the first author and two graduate stu-

⁸Note that there is no overlap of target disappearing entities between the training data and test data. Because there are no duplicate entities in the year of disappearance in Wikipedia.

dents who are Japanese native speakers and fluent in English) to judge whether each context is accompanied by a disappearing context. We adopt the positive contexts with the answers agreed upon by two or more annotators. Then, for each entity of the resulting data, we asked the annotators to determine the non-disappearing contexts using the collected negative examples of the entity and selected the same number of posts as the positive examples. We obtained an inter-rater agreement of 0.722 for English and 0.786 for Japanese by Fleiss’s Kappa (Fleiss and Cohen, 1973), both of which show substantial agreement. As a result, we obtained a total of 1,922 English and 1,326 Japanese tweets for 357 English entities and 235 Japanese entities, respectively, as the test data.

Tables 1 and 2 show the resulting disappearing entities and disappearing contexts of training data (English) and test data (English and Japanese), respectively. The entity types that are manually categorized into PERSON and GROUP account for a large proportion. These types of entities, occupied by persons, musical groups, and companies, are more likely to disappear. We see that disappearing contexts could be diverse according to the type of entity they include. We thus have to capture those contexts properly. Because we use Wikipedia as the list of disappearing entities, there is a bias in the composed categories, and some categories are absent in certain languages. However, we can still detect entities of the missing categories because if the coarse type of entities is the same, their contexts tend to be somewhat similar regardless of their category (*e.g.*, LOCATION types tend to be mentioned with the term *close*).

Note that the test data for EVENT and CREATIVE WORK are scarce. This is due to the absence of disappearing contexts in the data collected by distant supervision. In the case of EVENT, there are many entities that appear and then disappear within a year, and many of them are filtered out (§ 4.1). Regarding CREATIVE WORK, the disappearance of these entities is actually uncertain for the nature of the type. For example, even when the final episode of a television drama airs on TV, it remains in various other media. It is thus difficult to determine their disappearing contexts.

5.2 Models

We implemented the three models for comparison:

Proposed (TDS + Emb): As the proposed method,

we implemented BERT (Devlin et al., 2019) as an NER model and refined word embeddings (Emb) using the training data constructed by the proposed time-sensitive distant supervision (TDS). We refined pretrained fastText embeddings for each input post using tweets on the day of the input post (about 1M to 2M tweets for each day from 2012 to 2019) and additionally fed them into the BERT model at the time of fine-tuning and the test time.

Proposed (TDS): To verify the effectiveness of refined word embeddings, we removed the part related to Emb (the refined word embeddings; the green part in Figure 3) from Proposed (TDS + Emb) and used with the same training data, optimization, and parameters. Hence, this method does not consider the multiple posts of the target day when recognizing entities.

Baseline: To verify the effectiveness of the constructed training data, we collected the latest posts of disappearing entities using the original version of time-sensitive distant supervision (Akasaki et al., 2019), which does not consider the timing of the entities’ disappearance. Specifically, for each ending entity in Wikipedia from 2012 to 2018, we collected up to 100 retweets of the last day (through 2018) in which the entity appeared more than ten times as positive examples. For negative examples, we obtained the same number of posts from more than one year before the date when we collected the positive examples. By using the collected 25,920 posts for 2867 entities in English and 16,777 posts for 1733 entities in Japanese, we fine-tuned BERT using the same optimization and parameters as Proposed (TDS). Since this method does not consider the timing of the entities’ disappearance, many noisy contexts may be collected.

5.3 Settings

Implementation and model hyperparameters

We use TensorFlow 2 (ver. 2.6.0)⁹ for implementing the models and ran on NVIDIA RTX A6000 GPU with AMD EPYC 7313 @ 3.0GHz CPU. For the BERT model, we use *bert-base-cased* without next sentence prediction¹⁰ and pretrained from scratch using 2B English tweets for English and 800M Japanese tweets, respectively, both posted from Mar. 11th, 2011 to Dec. 31st, 2011. Using the same tweets, we pretrained 300-dimensional

⁹<https://www.tensorflow.org/>

¹⁰It is reported that there is little or no effect of next sentence prediction (Liu et al., 2019).

Parameter	Value
Epoch	5
Sentence length	128
Batch size	32
Dropout rate	0.1
Learning rate	2e-5

Table 3: Hyperparameters of fine-tuning BERT.

word embeddings using fastText¹¹ with default parameters. We refine them and feed to the BERT model at the time of fine-tuning. We fine-tuned all the models using AdamW (Loshchilov and Hutter, 2019) and chose the model in the epoch with the highest F₁-score on the development data. The hyperparameters of fine-tuning are listed in Table 3.

Evaluation methods To evaluate the accuracy, we apply the models to each post in the test data built in § 5.1. We utilize CoNLL-2003 (Tjong Kim Sang and De Meulder, 2003) schema, which measures precision, recall, and F₁-score. We performed training and test five times and reported the average performances.

To evaluate the relative recall and the detection immediacy of our method, we follow the experiments designed for evaluating emerging entities (Akasaki et al., 2019). Specifically, for each ending entity in Wikipedia that disappeared in 2019 (2608 for English and 763 for Japanese), we applied our method to all the posts in 2019 in which each entity appeared (437,816 for English and 202,666 for Japanese). For each entity, we consider an entity to have been discovered if the model was able to recognize the target string at least once. Then we determined how many target entities could be discovered from the posts and how much earlier those entities could be detected before the corresponding Wikipedia articles were categorized as ending using edit histories.

5.4 Results and Analysis

Overall accuracy of models Table 4 shows performances for all models. Both the proposed methods outperformed the baseline, which collected training data without considering the timing of the entities’ disappearance. The performance of the baseline is low because it was trained with data including many noisy contexts. This shows that our time-sensitive distant supervision successfully collected the disappearing contexts. Our Pro-

¹¹<https://fasttext.cc/>

	Precision	Recall	F ₁
Proposed (TDS + Emb)	0.721	0.666	0.692
Proposed (TDS)	0.747	0.589	0.658
Baseline (TDS)	0.561	0.263	0.358

(a) English

	Precision	Recall	F ₁
Proposed (TDS + Emb)	0.831	0.580	0.683
Proposed (TDS)	0.822	0.513	0.631
Baseline (TDS)	0.767	0.291	0.421

(b) Japanese

Table 4: Overall performance of each method for English and Japanese.

	Precision	Recall	F ₁
PERSON	0.858	0.887	0.872
CREATIVE WORK	0.446	0.505	0.473
LOCATION	0.711	0.519	0.600
GROUP	0.562	0.540	0.550
SERVICE&PRODUCT	0.549	0.397	0.460
EVENT	0.810	0.431	0.562

(a) English

	Precision	Recall	F ₁
PERSON	0.939	0.835	0.883
LOCATION	0.821	0.554	0.661
GROUP	0.801	0.439	0.567
SERVICE&PRODUCT	0.789	0.529	0.633
EVENT	0.291	0.114	0.163

(b) Japanese

Table 5: Performance of Proposed (TDS + Emb) for each coarse type.

posed (TDS + Emb) detected the entities with the highest accuracy, which means that the refined word embeddings worked effectively. In particular, the recall was improved in both Japanese and English, indicating that entities that could not be recognized by only using the features of a single post can be detected by utilizing multiple posts.

Detailed accuracy of optimal model Table 5 shows the performances for each type to analyze the behavior of Proposed (TDS + Emb). The accuracy of PERSON type entities is high in both English and Japanese. This is likely because numerous entities of this type exist in the training data and the person’s names themselves are easy to recognize from the surface. The CREATIVE WORK type is not present in Japanese, and the accuracy for that type is low in English because the disappearance of these entities is uncertain for the nature of

TYPE	# entities	# found (%)	lead-days mean (median)	
PERSON	1838	1622 (88.24%)	23	(0)
CREATIVE WORK	351	119 (33.90%)	175	(66)
LOCATION	73	45 (61.64%)	141	(45)
GROUP	163	102 (62.57%)	190	(129)
SERVICE&PRODUCT	93	44 (47.31%)	182	(95)
EVENT	25	7 (28.00%)	83	(33)
UNMAPPED	65	32 (49.23%)	240	(170)
TOTAL	2608	1971 (75.57%)	53	(0)
TOTAL (w/o PERSON)	770	349 (45.32%)	143	(78)

(a) English

TYPE	# entities	# found (%)	lead-days mean (median)	
PERSON	515	423 (82.13%)	32	(0)
LOCATION	48	29 (60.41%)	155	(120)
GROUP	121	63 (52.06%)	174	(138)
SERVICE&PRODUCT	63	41 (65.07%)	156	(98)
EVENT	16	7 (43.75%)	173	(162)
TOTAL	763	563 (73.78%)	69	(0)
TOTAL (w/o PERSON)	248	140 (56.45%)	128	(98)

(b) Japanese

Table 6: Relative recall and time advantage over entity types of English and Japanese disappearing entities detected with Proposed (TDS + Emb).

the type as described in § 5.1. This causes the training data to be contaminated with diverse contexts, resulting in low accuracy. The accuracy for EVENT type entities is the lowest in Japanese. This is likely because the number of training data is small and thus the model could not be sufficiently trained.

Relative recall and detection immediacy Table 6 shows the distribution of the entity types, detection ratio, and lead time against the Wikipedia update time for both languages. Overall, Proposed (TDS + Emb) detected 1971 (75.57%) English and 563 (73.78%) Japanese disappearing entities. More than 80% of PERSON entities were detected in both languages, while the other types were found 45.32% for English and 56.45% for Japanese. Note that some of the target entities are low frequency in our Twitter archive and do not appear in disappearing contexts, which may affect the performance. Also, it is difficult to discover entities that are mentioned without disappearing contexts because our method utilizes such disappearance signals.

For detection immediacy, we confirmed that 76.71% of the discovered English entities (1,512 out of 1,971) and 83.33% of the discovered Japanese entities (485 out of 582) were detected earlier than their update in Wikipedia. Remaining

entities were often either mentioned infrequently with disappearing contexts or their updates were unusually fast. The mean (and median) lead days of the first day when our method detected each entity against their update date were 53 (and 0.15) for English and 69 (and 0.47) days for Japanese. In particular, for other types of entities other than person, the lead days were 143 days (and 78) for English and 128 days (and 98) for Japanese. This demonstrates the detection immediacy of our method.

Our method could discover entities such as *Durgin-Park* (restaurant), *Rolling Acres Mall* (shopping mall), and *CiteULike* (Internet property), which are the types of entities whose disappearance is important to know in advance. This should prevent us from making fruitless actions or missing out on opportunities because the average lead time for these types is more than 100 days. Interestingly, the updates for PERSON type entities in Wikipedia are faster for both languages. Since this kind of uncontrolled disappearance (death) occurs suddenly and often receives media exposure, it is difficult to detect these entities before they actually disappear. It also shows that only certain types of entities are updated faster in Wikipedia. Therefore, detecting disappearing entities not only helps avoid missing opportunities but also serves as a means to quickly update the knowledge base.

6 Conclusion

In this paper, we have introduced the task of discovering disappearing entities in microblogs (§ 1, § 2, and § 3). To deal with the uncertainty of entity disappearance, we considered the year of disappearance in Wikipedia and fed it into time-sensitive distant supervision (§ 4.1). To perform the detection from noisy microblog posts, we proposed the method of refining pretrained word embeddings using the Twitter stream (§ 4.2). We actually constructed a multi-lingual dataset in Japanese and English using the proposed time-sensitive distant supervision (§ 5.1). Experimental results (§ 5.4) demonstrated that our method outperformed the baseline method and successfully found more than 70% of the target disappearing entities in Wikipedia and they were detected more than a month earlier than the update of the disappearance in Wikipedia. We release the dataset used in our experiments.¹

We plan to extract information such as the specific time of disappearance to make the discovered entities more useful for various applications.

7 Limitations

Our method cannot discover entities that are mentioned without disappearing contexts since we utilize such disappearance signals.

Although our experiments focused solely on Wikipedia entities, they did not sufficiently cover certain entities, such as stores and food products, for which disappearance is important for us. Therefore, it is necessary to collect and conduct experiments specifically for these entities.

8 Ethical Considerations

The dataset was collected using Twitter’s official API¹² and in compliance with Twitter’s terms of use. Only the tweet IDs of the tweets used in the experiments will be made public, and we ensure that their redistribution is in compliance with Twitter’s developer policy.¹³ Researchers cannot collect deleted tweets or tweets of private users, thus protecting user privacy.

We paid the annotators \$0.03 per post they checked (approximately 3,000 to 5,000 posts).

Although the first author is affiliated with Yahoo Japan Corporation, all the data used in the experiments belong to the affiliation of the second and third authors.

Acknowledgments

This work was partially supported by JSPS KAKENHI Grant Number JP21H03494, JP21H03445 and by JST, CREST Grant Number JPMJCR19A, Japan.

References

Jacqueline Aguilar, Charley Beller, Paul McNamee, Benjamin Van Durme, Stephanie Strassel, Zhiyi Song, and Joe Ellis. 2014. [A comparison of the events and relations across ACE, ERE, TAC-KBP, and FrameNet annotation standards](#). In *Proceedings of the Second Workshop on EVENTS: Definition, Detection, Coreference, and Representation*, pages 45–53, Baltimore, Maryland, USA. Association for Computational Linguistics.

Satoshi Akasaki, Naoki Yoshinaga, and Masashi Toyoda. 2019. [Early discovery of emerging entities in microblogs](#). In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 4882–4889.

¹²<https://developer.twitter.com/en/docs/twitter-api>

¹³<https://developer.twitter.com/en/developer-terms/agreement-and-policy>

Satoshi Akasaki, Naoki Yoshinaga, and Masashi Toyoda. 2021. [Fine-grained typing of emerging entities in microblogs](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4667–4679, Punta Cana, Dominican Republic. Association for Computational Linguistics.

Alan Akbik, Tanja Bergmann, and Roland Vollgraf. 2019. [Pooled contextualized embeddings for named entity recognition](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 724–728, Minneapolis, Minnesota. Association for Computational Linguistics.

Alan Akbik, Duncan Blythe, and Roland Vollgraf. 2018. [Contextual string embeddings for sequence labeling](#). In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1638–1649, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Thayer Alshaabi, David Rushing Dewhurst, Joshua R. Minot, Michael V. Arnold, Jane L. Adams, Christopher M. Danforth, and Peter Sheridan Dodds. 2021. [The growing amplification of social media: measuring temporal and social contagion dynamics for over 150 languages on twitter for 2009–2020](#). *EPJ Data Science*, 10(1).

Piotr Bojanowski, Edouard Grave, Armand Joulin, and Tomas Mikolov. 2017. [Enriching word vectors with subword information](#). *Transactions of the Association for Computational Linguistics*, 5:135–146.

Leyang Cui, Yu Wu, Jian Liu, Sen Yang, and Yue Zhang. 2021. [Template-based named entity recognition using bart](#). In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1835–1845.

Leon Derczynski, Eric Nichols, Marieke van Erp, and Nut Limsopatham. 2017. [Results of the WNUT2017 shared task on novel and emerging entity recognition](#). In *Proceedings of the 3rd Workshop on Noisy User-generated Text*, pages 140–147, Copenhagen, Denmark. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. [BERT: Pre-training of deep bidirectional transformers for language understanding](#). In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

George Doddington, Alexis Mitchell, Mark Przybocki, Lance Ramshaw, Stephanie Strassel, and Ralph Weischedel. 2004. [The automatic content extraction \(ACE\) program – tasks, data, and evaluation](#). In *Proceedings of the Fourth International Conference*

- on *Language Resources and Evaluation (LREC'04)*, Lisbon, Portugal. European Language Resources Association (ELRA).
- Michael Färber, Achim Rettinger, and Boulos Asmar. 2016. On emerging entity detection. In *Proceedings of the 20th International Conference on Knowledge Engineering and Knowledge Management (EKAW)*, pages 223–238.
- Joseph L Fleiss and Jacob Cohen. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement*, 33(3):613–619.
- Johannes Hoffart, Yasemin Altun, and Gerhard Weikum. 2014. [Discovering emerging entities with ambiguous names](#). In *Proceedings of the 23rd International Conference on World Wide Web (WWW)*, pages 385–396.
- Heng Ji, Ralph Grishman, Hoa Trang Dang, Kira Griffitt, and Joe Ellis. 2010. Overview of the TAC 2010 knowledge base population track. In *Third text analysis conference (TAC 2010)*, volume 3, pages 3–3.
- Nikolaos Kolitsas, Octavian-Eugen Ganea, and Thomas Hofmann. 2018. [End-to-end neural entity linking](#). In *Proceedings of the 22nd Conference on Computational Natural Language Learning*, pages 519–529, Brussels, Belgium. Association for Computational Linguistics.
- Guillaume Lample, Miguel Ballesteros, Sandeep Subramanian, Kazuya Kawakami, and Chris Dyer. 2016. [Neural architectures for named entity recognition](#). In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 260–270, San Diego, California. Association for Computational Linguistics.
- Thomas Lin, Mausam, and Oren Etzioni. 2012. [No noun phrase left behind: Detecting and typing un-linkable entities](#). In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 893–903, Jeju Island, Korea. Association for Computational Linguistics.
- Jian Liu, Yubo Chen, Kang Liu, Wei Bi, and Xiaojiang Liu. 2020. [Event extraction as machine reading comprehension](#). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1641–1651, Online. Association for Computational Linguistics.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. RoBERTa: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Ilya Loshchilov and Frank Hutter. 2019. [Decoupled weight decay regularization](#). In *International Conference on Learning Representations (ICLR)*.
- Pedro Henrique Martins, Zita Marinho, and André F. T. Martins. 2019. [Joint learning of named entity recognition and entity linking](#). In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 190–196, Florence, Italy. Association for Computational Linguistics.
- David McClosky and Christopher D. Manning. 2012. [Learning constraints for consistent timeline extraction](#). In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 873–882, Jeju Island, Korea. Association for Computational Linguistics.
- Mike Mintz, Steven Bills, Rion Snow, and Dan Jurafsky. 2009. Distant supervision for relation extraction without labeled data. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing (ACL-IJCNLP)*, pages 1003–1011.
- David Nadeau and Satoshi Sekine. 2007. [A survey of named entity recognition and classification](#). *Linguisticae Investigationes*, 30(1):3–26.
- Thien Huu Nguyen and Ralph Grishman. 2015. [Event detection and domain adaptation with convolutional neural networks](#). In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 365–371, Beijing, China. Association for Computational Linguistics.
- Daisuke Oba, Ikuya Yamada, Naoki Yoshinaga, and Masashi Toyoda. 2022. Entity embedding completion for wide-coverage entity disambiguation. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 6333–6344.
- Lev Ratinov and Dan Roth. 2009. [Design challenges and misconceptions in named entity recognition](#). In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning (CoNLL-2009)*, pages 147–155, Boulder, Colorado. Association for Computational Linguistics.
- Alan Ritter, Oren Etzioni, and Sam Clark. 2012. [Open domain event extraction from twitter](#). In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, pages 1104–1112.
- Wei Shen, Jianyong Wang, and Jiawei Han. 2014. [Entity linking with a knowledge base: Issues, techniques, and solutions](#). *IEEE Transactions on Knowledge and Data Engineering*, 27(2):443–460.
- Erik F. Tjong Kim Sang and Fien De Meulder. 2003. [Introduction to the CoNLL-2003 shared task: Language-independent named entity recognition](#). In *Proceedings of the Seventh Conference on Natural*

Language Learning at HLT-NAACL 2003, pages 142–147.

Zhaohui Wu, Yang Song, and C Lee Giles. 2016. [Exploring multiple feature spaces for novel entity discovery](#). In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 3073–3079.

Juntao Yu, Bernd Bohnet, and Massimo Poesio. 2020. Named entity recognition as dependency parsing. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 6470–6476.

ACL 2023 Responsible NLP Checklist

A For every submission:

A1. Did you describe the limitations of your work?

7

A2. Did you discuss any potential risks of your work?

8

A3. Do the abstract and introduction summarize the paper's main claims?

1

A4. Have you used AI writing assistants when working on this paper?

Left blank.

B Did you use or create scientific artifacts?

8

B1. Did you cite the creators of artifacts you used?

8

B2. Did you discuss the license or terms for use and / or distribution of any artifacts?

8

B3. Did you discuss if your use of existing artifact(s) was consistent with their intended use, provided that it was specified? For the artifacts you create, do you specify intended use and whether that is compatible with the original access conditions (in particular, derivatives of data accessed for research purposes should not be used outside of research contexts)?

8

B4. Did you discuss the steps taken to check whether the data that was collected / used contains any information that names or uniquely identifies individual people or offensive content, and the steps taken to protect / anonymize it?

8

B5. Did you provide documentation of the artifacts, e.g., coverage of domains, languages, and linguistic phenomena, demographic groups represented, etc.?

5

B6. Did you report relevant statistics like the number of examples, details of train / test / dev splits, etc. for the data that you used / created? Even for commonly-used benchmark datasets, include the number of examples in train / validation / test splits, as these provide necessary context for a reader to understand experimental results. For example, small differences in accuracy on large test sets may be significant, while on small test sets they may not be.

5

C Did you run computational experiments?

5

C1. Did you report the number of parameters in the models used, the total computational budget (e.g., GPU hours), and computing infrastructure used?

5

The Responsible NLP Checklist used at ACL 2023 is adopted from NAACL 2022, with the addition of a question on AI writing assistance.

- C2. Did you discuss the experimental setup, including hyperparameter search and best-found hyperparameter values?
5
- C3. Did you report descriptive statistics about your results (e.g., error bars around results, summary statistics from sets of experiments), and is it transparent whether you are reporting the max, mean, etc. or just a single run?
5
- C4. If you used existing packages (e.g., for preprocessing, for normalization, or for evaluation), did you report the implementation, model, and parameter settings used (e.g., NLTK, Spacy, ROUGE, etc.)?
5
- D** **Did you use human annotators (e.g., crowdworkers) or research with human participants?**
5
- D1. Did you report the full text of instructions given to participants, including e.g., screenshots, disclaimers of any risks to participants or annotators, etc.?
5
- D2. Did you report information about how you recruited (e.g., crowdsourcing platform, students) and paid participants, and discuss if such payment is adequate given the participants' demographic (e.g., country of residence)?
5
- D3. Did you discuss whether and how consent was obtained from people whose data you're using/curating? For example, if you collected data via crowdsourcing, did your instructions to crowdworkers explain how the data would be used?
5
- D4. Was the data collection protocol approved (or determined exempt) by an ethics review board?
5
- D5. Did you report the basic demographic and geographic characteristics of the annotator population that is the source of the data?
5