Generating Spoofing Tweets considering Points of Interest of Target User

Jeongwoo Lim*, Naoko Nitta*, Kazuaki Nakamura*, and Noboru Babaguchi*,
* Osaka University, Osaka, Japan
E-mail: jeongwoolim@nanase.comm.eng.osaka-u.ac.jp, {naoko, k-nakamura, babaguchi}@comm.eng.osaka-u.ac.jp,
Tel: +81-6-6879-7746

***Abstract*—Personal information of legitimate users shared on social networking services (SNS) can be used for identity spoofing. The simplest approach is to clone the profile information of the target user. Recent deep learning techniques have enabled us to even automatically generate spoofing messages by imitating the past messages of the target user; however, such message generators can only be trained for target users who have posted sufficient number of messages to train the generator. Further, since the legitimate users actually exist in the real world, their messages are often related to the situations in the real world. Such relations to the real world have not been considered in generating the spoofing messages, which can be the cues for detecting the identity spoofing. This paper further examines the possibility of identity spoofing even for target users who have posted only a limited number of messages based on the assumptions that messages about semantically related points of interest (PoIs) in the real world can be similar regardless of users. Our proposed method firstly collects messages about various PoIs posted by arbitrary users and estimates the semantic topic of each PoI based on the content of its messages. A topic-based message generator trained on the collected messages can be commonly used to generate spoofing messages about PoIs in the real world according to the interest of each target user.**

I. Introduction

Social network services (SNS) such as Twitter, Facebook, and Instagram have enabled people all over the world to easily communicate with each other. Besides the legitimate users who are actual humans in the real world, there exist many bot accounts controlled by software. For example, the ratio of the bot accounts on Twitter is estimated between 9 and 15% [1]. Some of them impersonate humans to deceive other legitimate users. Especially, those which pretend to be specific users are considered *identity spoofing*, and they can be created automatically by using the personal information of legitimate users shared on SNSs. For example, Bilge et al. [2] created bot accounts on various SNSs by simply cloning the profiles of legitimate users and have verified that more than 50% friend requests sent by the bot accounts were accepted by the friends of the legitimate users.

Due to the simplicity of the short text messages posted to Twitter, even spoofing messages are often automatically generated on Twitter. A typical approach is to use deep neural networks such as the recurrent neural network (RNN) for a specific target user who has published sufficient number of messages for training [3], which makes users who have posted rather limited number of messages seem to be safe from such identity spoofing. In fact, since the networks require rather a large corpus of texts for training, they are more often trained on a corpus of texts generated by arbitrary users to impersonate humans, rather than specific users. Since the networks are more effective when trained on a domain-specific corpus of shorter texts, it is often trained separately to generate specific short-text content such as online restaurant reviews of the same score [4], phishing emails [5], etc. A network can also be trained to generate texts according to the given input. For example, a network can be trained on pairs of images and their captions so that it can generate suitable caption for a given image [6]. Similarly, a network can be trained on pairs of phishing message and its topic to generate a phishing message according to the topic which would interest the target user [7].

Based on such ideas, we also propose to train a network which can be commonly used to generate spoofing messages for arbitrary users. As a common patterns of legitimate users who exist in the real world, we focus on the fact that they often post observations about their points of interests (PoIs), each of which is a specific point location that each user finds interesting. Further, messages about the semantically similar PoIs can be similar regardless of users. For example, messages about different restaurants tend to be about the food served at each restaurant, and messages about different baseball parks tend to be about the baseball game being played at each park. Thus, by estimating the semantics of each PoI based on its past messages from arbitrary users, a network can be trained on the messages about various PoIs posted by arbitrary users to generate messages according to the semantics of a given PoI. In addition, a PoI which would interest each target user can be selected based on his/her past messages. By feeding the semantics of the selected PoI to the trained network, a spoofing message which seems to be posted by the target user can be generated.

II. Proposed method

Targeting on the identity spoofing on Twitter users, our goal is to generate a spoofing tweet (text message posted to Twitter) $s^{U_{spoof}}$ of a specific target user U at the location x^U, who has posted only a limited number of tweets $S^U = \{s_n^U | n = 1, \cdots, N_U\}$. Our assumptions are that 1) real users who exist in the real world often post about their PoIs in the real world such as their experiences at the places or events they have visited and 2) tweets about semantically similar PoIs can be similar among different users. Based on these assumptions, our

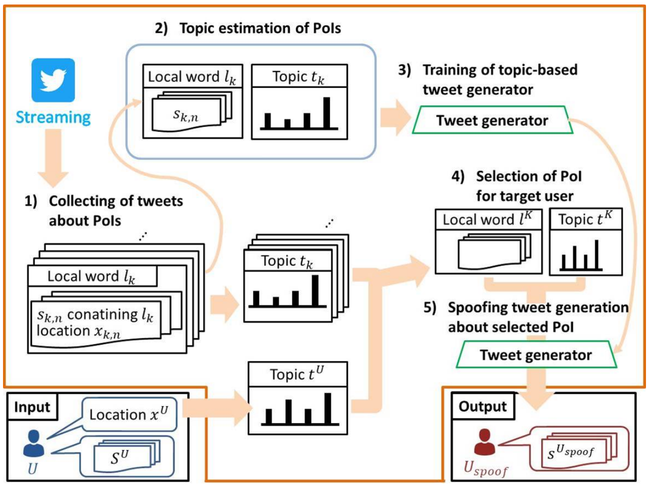

Fig. 1. Overview of proposed method

proposed method collects tweets from arbitrary users about their PoIs in the real world, estimates the topics of each PoI based on its tweets, and trains a topic-based tweet generator by using the collected tweets. Then, when given a target user with a sufficient number of past tweets to estimate his/her interest, a PoI is selected according to his/her estimated interest, and a spoofing tweet about the selected PoI is generated based on the estimated topic of the PoI. Fig. 1 shows the overview of the proposed method composed of the following 5 steps.

1) Collection of tweets about PoIs
   The tweets about PoIs in the real world are collected from geotagged tweets (tagged with geographic coordinates) posted by arbitrary users by firstly collecting *local words*, which are considered to represent each PoI such as its name [8]. For each PoI represented by the local word $l_k (k = 1, 2, \cdots)$, the tweets $S_k = \{s_{k,n} | n = 1, \cdots, N_k\}$, each of which contains the word l_k, are collected with their posted geographic coordinates $X_k = \{x_{k,n} | n = 1, \cdots, N_k\}$.
2) Topic estimation of PoIs
   The semantics of each PoI represented by the local word l_k can be characterized by the words used in S_k. After selecting the representative words from S_k, the topic of PoI t_k is estimated based on the latent semantic representations of the selected words.
3) Training of topic-based tweet generator
   Given the collected pairs of the tweet $s_{k,n}$ and the estimated topic t_k, the topic-based tweet generator, $s_{k,n} = G(t_k)$, is trained.
4) Selection of PoI for target user
   The interest of the target user U is estimated in the same way as Step 2) based on the latent semantic representations of the words selected from a set of his past tweets S^U. Then, from the PoIs around the current location of the target user x^U, a PoI whose topic t^K best matches the estimated interest of the target user is selected.
5) Spoofing tweet generation about selected PoI
   The estimated topic t^K of the selected PoI is given to the tweet generator G to generate a spoofing tweet by $s^{U_{spoof}} = G(t^K)$.

Steps 1), 4), and 5) are real-time processes while Steps 2) and 3) are processed in a batch mode by using the data collected in Step 1) during a certain time duration. The details of each process are described in the following subsections.

A. Collection of tweets about PoIs

Based on the assumption that tweets about the same PoIs posted by different users are likely to contain common words such as the names of places or events, we try to collect tweets about PoIs by using such words. While place names can be collected from existing geographical dictionaries such as GeoNames [9], in order to handle the diversity of words used by Twitter users, we firstly collect such commonly used words to represent PoIs. Since these words are likely to be used by several users only at the locations of the corresponding PoIs, we call these words *local words*, and collect them from geotagged tweets based on the spatial locality of words.

In order to continuously extract as many real-time tweets about PoIs as possible, we use a method proposed in [8] for collecting local words from streaming geotagged tweets. Their method continuously collects the word usage history within a specific time window which is adaptively determined for each word, so that the spatial locality can be examined at the proper timing according to the usage pattern of each word. Such approach enables us to collect diverse types of local words including the temporary ones representing temporary events and regardless of the popularity of PoIs. [8] should be referred for more details. For each extracted local word l_k, tweets $S_k = \{s_{k,n} | n = 1, 2, ..., N_k\}$, each of which contains l_k, are collected with their posted geographic coordinates $X_k = \{x_{k,n} | n = 1, 2, ..., N_k\}$.

B. Topic estimation of PoIs

Since users would post about their experiences at each PoI, the content of the tweets posted from the same PoI would be related to the topic of the PoI. Thus, we estimate the topic of each PoI represented by l_k based on the words in its collected tweets S_k. Word2vec embedding [10] is often used as the latent semantic representation of words, so that words used in similar context in a collection of documents would be represented by feature vectors which are close in the latent semantic space. In order to learn the semantic representation of words based on how the words are used in the tweets about PoIs, word2vec embedding is learned from the collected tweets $S = \{S_k | k = 1, 2, \cdots\}$ by considering S_k as a document. Note that words which are expected to be unnecessary to represent the semantic topics such as local words l_k, stop words, URLs starting with *http(s)://*$\cdots$, and numbers are removed from $s_{k,n}$ as a preprocess. As a result, each word w_m used in S is represented by a feature vector v_m.

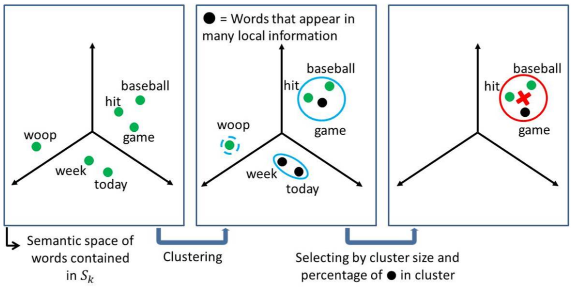

Fig. 2. Word selection for topic estimation of PoI

TABLE I
EXAMPLE OF TWEETS BEFORE/AFTER PREPROCESS

Before	So good to be back to football!!! Go Broncos!!!! #gobroncos ... URL
After	<START> so good to be back to football go <SLW> <HASHTAG> gobroncos </HASHTAG> <END>

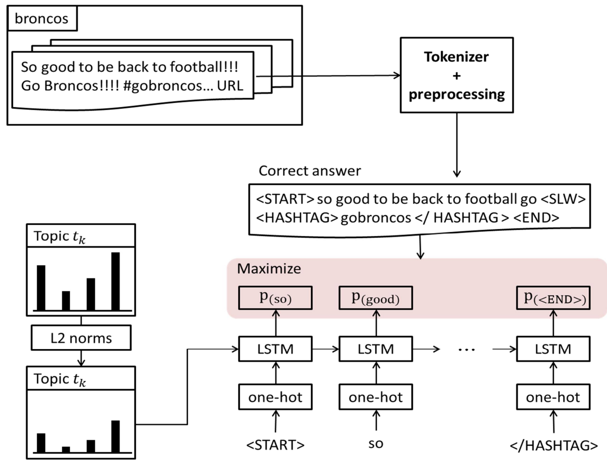

Fig. 3. Training of topic-based tweet generator

Then, the topic of each PoI represented by l_k can be estimated based on the semantic representations of words in S_k. However, for each word in S_k, its strength of relations to the topic of PoI should differ. Given a corpus of documents, term frequency inverse document frequency (tfidf) is often used as a measure of the word importance in each document. However, tfidf which is based on the word frequency in each document and the number of documents containing each word is not proper measure for S, since each tweet posted by different user is very short and the topic of the same PoI is often expressed with different words by each user.

Thus, for each PoI, we consider the semantic similarity among words to select the important words which represent its topic. The words $w_m \in S_k$ are firstly clustered based on the cosine similarity of their semantic representations v_m in the latent semantic space as shown in Fig. 2. The hierarchical clustering is applied to find the clusters of words with high similarity. The cluster relevant to the topic of the PoI should contain words used by many users uniquely for the corresponding PoI. Thus, we consider words commonly used for many PoIs as stop words. Then, the cluster containing the words used by the largest number of users, few of which are stop words, is selected. The mean of v_m of the words in the selected cluster is used as the topic of the PoI.

C. Training of topic-based tweet generator

Our assumption is that the tweets at semantically similar PoIs are similar. Thus, the patterns of tweets can be learned based on the topics of PoIs. Long Short-Term Memory network (LSTM) [11], which is a variant of RNN to learn long-term dependencies, is often used to generate a sentence of good quality [6], [7].

Here, we train an LSTM on pairs of ($s_{k,n}$, t_k) so that, when t_k is fed to the LSTM as the input, each word in the tweet $s_{k,n}$ is accurately predicted sequentially. As the pre-processes, the semantic representations of PoIs are normalized by their L2 norms so that semantically similar PoIs would be close by the Euclidean distance. Additionally, each sentence is normalized by using Tokenizer [12]. In order to further simplify the normalized sentence, specially identified words represented with <*> are removed except <HASHTAG> and </HASHTAG> each of which represents the start and end of hashtag. In order to learn the common patterns for different PoIs, the local word l_k indicating the selected PoI is replaced with <SLW> and other local words are replaced with <LW>. Finally, the words used fewer than 5 times in S are replaced with <UNK> to handle them as unknown words, and the start and end of each tweet are represented with <START> and <END>. An example of the tweets before and after the pre-process are shown in Table I.

Afterwards, each word in the tweet $s_{k,n}$ is represented as a one-hot vector of dimension equal to the number of unique words used in S and LSTM outputs a vector of the same dimension which represents the likelihood of each word at each step in a sentence as shown in Fig.3. LSTM is trained to maximize the log likelihood of the correct tweets in the training dataset.

D. Selection of PoI for target user

The interest of the target user t^U is estimated using the his/her past tweets S^U. The cluster of important words is selected as described in Section II-B. Here, instead of the number of users, the cluster is selected based on the maximum frequency of words in the cluster. Since users can be interested in several topics, clusters are ranked based on the maximum frequency of words in the cluster and the top B clusters $t_b^U (b = 1, 2, ..., B)$ are selected. Then, PoIs that have been collected are ranked based on the similarity of t_k to t_b^U to select a PoI represented with the local word l^K, which is relevant to the user's interest.

E. Spoofing message generation about selected PoI

Finally, using the model G trained in Section II-C and the estimated topic t^K of the seleted PoI, the spoofing message

$s^{U_{spoof}}$ is generated by $s^{U_{spoof}} = G(t^K)$ Then, if $s^{U_{spoof}}$ contains <SLW>, it is replaced with l^K and <LW> is replaced with one of the local words l_k used in S^K except l^K. They are randomly selected according to their frequency distributions in S^K. A sequence of words between <HASHTAG> and </HASHTAG> are concatenated and added $\sharp$ in front to make them a hashtag, while <HASHTAG> and </HASHTAG> are removed.

III. Experiments

The experiments are conducted by using 6,655,763 geotagged tweets posted during 30 days from September 2016 to October 2016 from the United States defined with the latitude and longitude ranges of [24, 49] and [125, 66], respectively. The geotagged tweets about PoIs are firstly collected iteratively in Step 1). Since the local words which should represent PoIs are considered mainly nouns such as the name of places, a part of speech tagging is applied to each geotagged tweet to collect nouns [13]. Additionally, compound nouns such as *Boston Museum* are extracted as nouns using TermExtract [15], since they represent more specific areas than their component words such as *Boston* or *Museum*. Then, the spatial locality of each noun is examined to firstly collect local words. As a result, 61,058 local words with 770,896 tweets in total were collected after 30 days. The collected tweets are used in the following experiments.

A. Evaluation of PoI Selection

In order to estimate the topics of PoIs and the interest of the target user, we firstly learn word2vec embedding using the 610,463 geotagged tweets for 56,696 local words collected during the first 26 days as a corpus. Table II shows the examples of the geotagged tweets in S_k collected for each PoI represented by the local word l_k and the words in the cluster selected in Step 2). 'URL' in the tweets represents any URL in the form of *http(s)://*$\cdots$. For example, for the PoIs related to football and music, relevant words were properly selected from the collected geotagged tweets.

In order to evaluate the validity of the estimated topics, we examined if the PoIs actually mentioned by the target users can be selected by the process described in Section II-D. Among the users who posted tweets about any PoI on the 28th day, we randomly selected 500 users and collected their past tweets up to the corresponding tweets posted on the 28th day. The number of the past tweets collected for each user varies from 500 to 3,000. The interest of each user is estimated from the collected tweets and the 500 PoIs mentioned by the users were ranked according to their similarity to the estimated interest of each user.

Fig. 4 shows the number of users whose corresponding PoIs were ranked within the top R ranks. For comparison, we also show the results when estimating the topics of PoIs and the interest of the target user with different approaches. The baseline is when PoIs are randomly ranked. As other approaches, LDA is when the topics of S_k or S^U are estimated with Latent Dirichlet Allocation (LDA) model [14], w2v+mean is when taking the average of v_m for all words in S_k or S^U, and w2v+tfidf is when taking the weighted average of v_m for all words with the tfidf weights. The LDA model is widely used for estimating a topic for each document in a given corpus. Here, the same corpus which is used to learn the word2vec embedding is firstly used to estimate word distribution for each topic. A set of tweets containing the same local word is treated as a document. The estimated word distributions for topics are then used to infer the topic distribution of S_k or S^U. For the proposed method, B was set to $B = 1, 2, 5$.

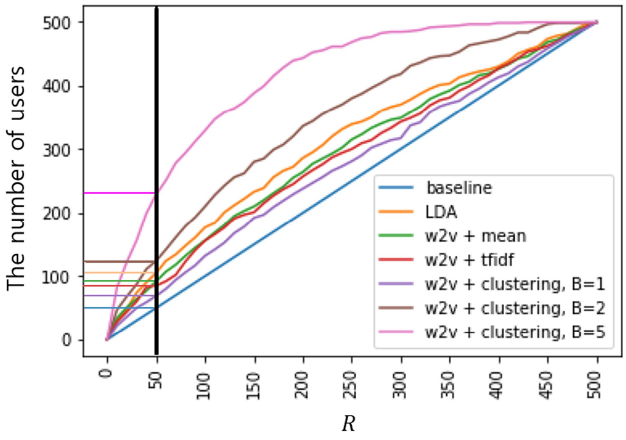

Fig. 4. Number of users whose corresponding PoI is ranked in top R ranks

All approaches performed better than the baseline. The proposed method performed worse than w2v+tfidf, w2v+mean, and LDA when $B = 1$; however, the performance improved largely when $B = 2$. This is because each user can have multiple interests and the PoI mentioned by each user does not necessarily match his/her top interest area. As an example, Table III and Table IV show the past tweets of a target user who posted about PoI represented by *believeland*, the tweets of other users who posted about the same PoI, and the ranks of the PoI *believeland* for the target user obtained by each approach. From the tweets, the target user can be inferred to be interested in food and sports, while the PoI represented by *believeland* can be inferred to be related to sports. Since w2v+tfidf, w2v+mean, and LDA all consider all words in S^U, the interest of user is inferred as a mixture of multiple topics. On the other hand, the proposed method can separately consider words relevant to each topic. Thus, the proposed method can rank the PoI more accurately for each target user especially when considering his/her multiple interests.

B. Evaluation of spoofing tweet generation

We trained the tweet generator G on the geotagged tweeets about PoIs collected during the first 26 days. Although we have collected geotagged tweets containing the local words, not all of them are observations about the PoIs represented by the local words. As described in Section II-B and shown in Table II, we have selected words which are considered as relevant to the topic of each PoI. Conversely, these words are

TABLE II
EXAMPLES OF TWEETS ABOUT POIS REPRESENTED BY LOCAL WORDS AND RELEVANT WORDS IN SELECTED CLUSTER

Local words	Tweets	Relevant Words
asu	ASU vs. Cal! Even though I'm an alum, this is actually my first time at a college football game, ... URL	football, game, beat, watch, win, score
	Game day vs asu @ 4pm? #gameon #gameday #totamsterritory #tusksup #csufwomenssoccer...URL	
fedexfield	Getting our game faces on! Gofing to see wvu vs byu at #redskins home @ FedExField URL	game, football, win, games, watch, beat
	Gearing up for the season opener @steeler vs @redskins #GoSteelers @ FedExField URL	
boone picken stadium	Game days are too fun with you girly gal. <;3 @ Boone Pickens Stadium URL	game, win, football, lose, cheer, college
	Hook 'em am I right?! Just kidding GO POKES!!!!?? @ Boone Pickens Stadium URL	
adele	adele my favorite Song of the night !!! #dontyouremember #adele #adelelive #ilovethissong #tears URL	music, singer, singing songwriter, song, sing
	Very fitting song to end an amazing concert #Adele #adelelive2016 #adeleconcert #nyc URL	
atl	My favorite part of music midtown #twentyonepilots #musicmidtown #atl #ride twentyonepilots URL	music, mic, indie songwriter, musician
	Listening to the best of #mUptop at the hands of @djshaolin at pyramidslounge in #mATL	
kaaboo	Group love playing Beast Boys, Sabotage ! #kaaboo2016 #sandiego #friends @ KAABOO URL	music, stage, set, band guitar, playing, rockin
	@JimmyBuffett Maybe play Captain America tonight. John Lovell on kazoo at Kaaboo.	

TABLE III
EXAMPLES OF TWEETS (OF A TARGET USER WHO POSTED ABOUT POI *believeland* AND ABOUT THE SAME POI POSTED BY DIFFERENT USERS)

Past Tweets of Target User	Brew and lunch while biking around #StanleyPark with the w... (Stanley Park 1897 Amber Ale) URL #photo
	This beer is much better than the play of the #Cleveland #Browns #... (Scarlett Red Rye) URL #photo
	Trophy is coming east #believeland #nbanals - Drinking a Go West! IPA @ Bridgewater Clubhouse - URL #photo
	Not much wrong with the Ale - Drinking a Smoked Spliff Ale by @divingdogbrew @ Bacon and Beer Classic URL #photo
	Can't believe how he was able to stick with curry at times late in the game. Effort goes a long way. URL
	Watching some of this #DemDebate during college football commercial breaks and feel that @MartinOMalley is making a great impression
Tweets about *believeland*	2nd last home game of the year. Great seats! #Believeland #Windians @ Progressive Field - URL
	ARE YOU READY!!!???? #mlbplayoffs #2016 #Rallytogether #Believeland #YourTurn #GoTribe @ URL
	Nice and moist out here this morning ?? #golfislife #golfife #golf #believeland @ Ridge Top URL

TABLE IV
RANK OF *believeland* FOR EACH APPROACH

	w2v + mean	w2v + tfidf	LDA	w2v + clustering ($B = 1$)	w2v + clustering ($B = 2$)
Ranks	238	265	94	125	10

TABLE V
MEAN VALUES OF $BLEU_n$ FOR EACH METHOD

	$BLEU_1$	$BLEU_2$	$BLEU_3$	$BLEU_4$
[16]	0.1556	0.0352	0.0131	0.0064
Proposed	0.4255	0.2889	0.2032	0.1398

used to select relevant tweets as the observations at each PoI. Since the number of tweets vary largely among PoIs, for each PoI, we have randomly selected at most 5 tweets containing the selected relevant words which were used by more than 2 different users, resulting in total of 69,216 tweets. LSTM model is trained by using 80% of them as the training data and the remaining 20% for validation. In order to generate spoofing tweets with more variations, we selected 50 candidate words as the first word according to the predictions of the model, then for each of the first word, selected 2 candidate words as the second word, and for each of the second word, generated a sentence which maximizes the log likelihood. This process generates 100 tweets in total.

As an existing method, [16] also generates tweets according to a topic, although the target differs. Their target is to impersonate humans, instead of specific users, and to generate tweets to get into a community group. Their method collects the tweets exchanged within a community group, and considers that the frequent words represent their topic of interest. Then, tweets which contain the frequent words are randomly duplicated from other users' tweets. We use this method for comparison. Here, we randomly select 100 tweets which contain the words frequently used in S^K from recent tweets.

We randomly selected 50 local words with S^K consisting of the tweets collected after the 27th day, and for each S^K, generated 100 tweets according to its topic by the trained G and [16] respectively. The generated tweets are evaluated with BLEU [17], which evaluates the quality of a sentence based on n-gram, by using the tweets in S^K as the ground truth. As shown in Table V, $BLUE_n(n = 1, \cdots, 4)$ was higher for any n for the proposed method.

C. Examples of generated spoofing tweets

As case studies, we selected target users who are inferred to be most interested in different topics: music, sports, food, national park, beer, or politics. Tables VI– XI show the past

TABLE VI
EXAMPLES OF GENERATED TWEETS (TOPIC: MUSIC)

Selected PoIs	1	madison square	2	insustrybar	3	uadnyc	4	rockwoodmusichall	5	bowery electric
Examples of past tweets of user	And congratulations to a true artist and musician Bradley Cooper, nominated alongside me. I couldn't be more proud URL									
	Im humbled & grateful that my album Joanne was nominated & also my song Million Reasons. Thank u so much Monst URL									
	This was the 1st time I ever played Sonja the song I wrote about and FOR her #GrigioGirls This was her last birthday URL									
Examples of generated spoofing tweets (local word: madison square)	just posted a photo @ madison square									
	we are #hiring ! click to apply : <UNK> <UNK> - #<UNK> #throwback									
	at the madison square . #sweden #madisonsquare #thank #music #<UNK>									
	my favorite song of the #adele #madisonsquare #<UNK> #iloveilluminatemsg									

TABLE VII
EXAMPLES OF GENERATED TWEETS (TOPIC: SPORTS)

Selected PoIs	1	autzenstadium	2	autzen stadium home of the	3	university of oregon	4	uoregon	5	cuthbert amphitheater
Examples of past tweets of user	I like how Demaryius Thomas tried to sneak the football over to the 1st down marker during that flight. #aintcheatinainttryin									
	Behind the Scenes — Chris Betts — Student Sports Baseball — Sweat Machine URL									
	Checking out "Garvey & Gretsky: Big Sports Names Chasing Big League Dreams" on 4 uoregon The Baseball Network: URL									
Examples of generated spoofing tweets (local word: autzenstadium)	just posted a photo @ autzenstadium									
	can you recommend anyone for this #job ? autzenstadium - #autzen #sunset #autzenstadium , plw									
	first game of the season ! #football #goduck #autzenstadium #<UNK>									
	best game ever ! #autzenstadium #goduck #autzenstadium #goautzen #sunset									

TABLE VIII
EXAMPLES OF GENERATED TWEETS (TOPIC: FOOD)

Selected PoIs	1	uws	2	158th	3	amyruths	4	seasoned vegan	5	fishtag
Examples of past tweets of user	Lunch!! (@ Triumph Brewing - @triumphnewhope in New Hope, PA) URL									
	Wow it seems to be a little warm out. I guess I'll start the grill to cook dinner.									
	I just ousted @sarahtomko as the mayor of Yummy Sushi on @foursquare! URL									
Examples of generated spoofing tweets (local word: amyruths)	just posted a photo @ amyruths									
	(@ amyruths in harlem , amy ruth)									
	if you are looking for work in #amyruths , newyork , check out this #job :									
	best friend ! #nyc #amyruths #<UNK> #foodporn									

tweets of each user, the top 5 candidate PoIs, and examples of the generated spoofing tweets for one of the PoIs. It can be seen that spoofing tweets can be generated according to the interest of each user such as music, sports, food, national park, and beer. For the target user interested in politics, the generated tweets sometimes contained words unrelated to politics such as *#music*. It might be due to the closness between *music* and selected PoI *white house* in the semantic space of words. Furthermore, tweets often posted commonly for any PoI such as *just posted a photo* and tweets for job recruiting were contained in the training data, and as a result, such tweets tend to be generated for any PoI. Such bias and noise in the training data need to be removed to further improve the quality of the generated tweets. In addition, the generated tweets often contain <UNK>. They also need to be replaced with certain words by considering the context of the generated sentence.

IV. CONCLUSIONS

This paper proposed a method for generating a spoofing tweet without training separate tweet generator for each target user. By focusing on the fact that legitimate users often tweet about their PoI in the real world and tweets about semantically similar PoIs tend to be similar, our proposed method trains a topic-based tweet generator by using the tweets about PoIs posted by arbitrary users. Experiments using the geotagged tweets posted during a month have verified that the topics of PoIs and the interest of users can be accurately estimated from their past tweets, and PoIs can be selected according to the interest of the target user. Further, the tweets can also be generated according to the topic of each PoI with a common tweet generator. Handling the bias in the training data and quantitative evaluations of the generated spoofing tweets are our future work.

ACKNOWLEDGMENT

This research was partly funded by Japan Society for the Promotion of Science KAKENHI: Grant Number 16H06302 and 19K12019.

TABLE IX
EXAMPLES OF GENERATED TWEETS (TOPIC: NATIONAL PARK)

Selected PoIs	1	glaciernationalpark	2	niagara	3	grand canyon	4	griffithpark	5 catskill
Examples of past tweets of user	Someone didn't want to get their feet wet. @ Neptune State Park URL								
	Scenes like this make me want to try fly-fishing. @ North Umpqua Trail URL								
	A lesser-known trail in Crater Lake National Park to celebrate the beauty of all the parks. URL								
Examples of generated spoofing tweets (local word: glaciernationalpark)	just posted a photo @ glaciernationalpark								
	if you are looking for work in #glaciernationalpark , montana , check out this #job :								
	so glad i am at glaciernationalpark in nps100 , montana								
	ready for the <UNK> ! #avalanchelake #<UNK> #nofilter #glaciernationalpark								

TABLE X
EXAMPLES OF GENERATED TWEETS (TOPIC: BEER)

Selected PoIs	1	cocoa	2	cascade	3	world of beer	4	gr	5 beermenus
Examples of past tweets of user	Enjoying a brew and some #za after setting a personal record walking a 5k thi... (Mischief) URL #photo								
	When #househunting #forfun in #NorCal, drinking a few strong brews ... (Super Saint Thomas) URL #photo								
	Starting off the fantasy draft with a #tangerine #ipa #fantasybaseball ... (Citradelic IPA) URL #photo								
Examples of generated spoofing tweets (local word: world of beer)	just posted a photo @ world of beer								
	if you are looking for work in #worldofbeer , south tampa , check out this #job :								
	one of the best of the best <UNK> . #worldofbeer #<UNK> #westchase								
	a little bit of the <UNK> . #<UNK> #worldofbeer #5thanniversaryipa #love								

TABLE XI
EXAMPLES OF GENERATED TWEETS (TOPIC: POLITICS)

Selected PoIs	1	white house	2	georhia brown	3	black women s agenda	4	timkaine	5 john lewis
Examples of past tweets of user	The most important way to stop gangs, drugs, human trafficking and massive crime is at our Southern Border. We need URL								
	It should not be the job of America to replace regimes around the world. This is what President Trump recognized i URL								
	I just had a long and productive call with President @RTErdogan of Turkey. We discussed ISIS, our mutual involveme URL								
Examples of generated spoofing tweets (local word: white house)	just posted a photo @ white house								
	(can you recommend anyone for this #job ? white house - #whitehouse #un , obama)								
	had a great time at the white house ! #<UNK> #sxsl #music								
	so excited to be a <UNK> <UNK> . #whitehouse #obama #champion								

REFERENCES

[1] O. Varol, E. Ferrara, C. A. Davis, F. Menczer, and A. Flammini, "Online Human-Bot Interactions: Detection, Estimation, and Characterization," *Proc. Int. AAAI Conf. on Web and Social Media (ICWSM)*, pp. 280–289, May 2017.

[2] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda, "All Your Contacts are Belong to Us: Automated Identity Theft Attacks on Social Networks," *Proc. the 18th int. Conf. on World wide web (WWW)*, pp. 551–560, April 2009.

[3] @DeepDrumpf, https://twitter.com/deepdrumpf.

[4] Y. Yao, B. Viswanath, J. Cryan, H. Zheng, and B. Y. Zhao, "Automated Crowdturfing Attacks and Defenses in Online Review Systems," *Proc. the 2017 ACM SIGSAC Conf. on Computer and Communications Security (ACM CCS)*, pp. 1143–1158, November 2017.

[5] A. Das and R. Verma, "Automated Email Generation for Targeted Attacks using Natural Language," *Text Analytics for Cybersecurity and Online Safety (TA-COS)*, 8 pages, 2018.

[6] O. Vinyals, A. Toshev, S. Bengio, and, D. Erhan, "Show and Tell: A Neural Image Caption Generator," *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 3156–3164, June 2015.

[7] J. Seymour and P. Tully, "Weaponizing Data Science for Social Engineering: Automated E2E Spear Phishing on Twitter," *Black Hat USA*, 2016.

[8] J. Lim, N. Nitta, K. Nakamura, and N. Babaguchi, "Constructing Geographic Dictionary from Streaming Geotagged Tweets," *ISPRS Int. J. Geo-Inf.*, 8(5), 216, 24 pages, 2019.

[9] GeoNames, Available online: http://www.geonames.org

[10] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient Estimation of Word Representations in Vector Space," *Proc. Int. Conf. on Learning Representations (ICLR)*, May 2013

[11] S. Hochreiter, J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, 9(8), pp. 1735-1780, 1997.

[12] C. Baziotis, N. Pelekis, and C. Doulеridis, "Datastories at Semeval-2017 Task 4: Deep LSTM with Attention for Message-Level and Topic-Based Sentiment Analysis," *Proc. the 11th Int. Workshop on Semantic Evaluation (SemEval-2017)*, pp. 747–754, August 2017.

[13] E. Brill, "A Simple Rule-based Part of Speech Tagger," *Proc. the 3rd Conf. on Applied natural language processing (ANLC)*, pp. 152–155, April 1992.

[14] D. Blei, A. Ng, and M. Jordan, "Latent dirichlet allocation," *Journal of Machine Learning Research*, vol.3, pp. 993-1022, July 2003

[15] "Term Extract," http://gensen.dl.itc.u-tokyo.ac.jp.

[16] Z. Coburn and G. Marra, "Realboy: Believable TwitterBots," Available online: http://ca.olin.edu/2008/realboy/index.html, 2011.

[17] K. Papinei, S. Roukos, T. Ward, and W. Zhu, "BLUE: a Method for Automatic Evalueation of Machine Translation," *Proc. the 40th Annual Meeting of Association for Computational Linguistics (ACL)*, pp. 311-318, July 2002.