Detection of Tweets Where Birthdays are Revealed to Other People

Yasuhiko Watanabe, Naohiro Miyagi, Kenji Yasuda, Ryo Nishimura, and Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, Japan
Email: watanabe@rins.ryukoku.ac.jp, t120499@mail.ryukoku.ac.jp, t130522@mail.ryukoku.ac.jp,
r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

***Abstract*—These days, many people use a social networking service (SNS). When we use SNSs, we carefully protect the privacy of personal information: name, age, gender, address, birthday, etc. However, we often reveal birthdays on SNS, not only ours but also of others. Birthday information can threaten our privacy and security when combined with other personal information. In this study, we investigated tweets where birthdays were revealed to other people. We collected 1,000 Japanese tweets including word "*tanjyobi* (birthday)" and found about 30% of them were tweets where birthdays were revealed to other people. Furthermore, 70% of tweets where birthdays were revealed to other people were ones where receivers' birthdays were revealed. We obtained 87% accuracy when we applied support vector machine (SVM) machine learning techniques to classify tweets including word "*tanjyobi* (birthday)" into ones revealing birthdays of senders, receivers, and others. However, the recall rate of tweets where senders' birthdays were revealed was only 20%.**

Keywords–birthday; personal information; Twitter; SNS; privacy risk.

I. INTRODUCTION

These days, many people use a social networking service (SNS). These users, especially young users, tend to disclose personal information on their SNS profiles seemingly without much concern for the potential privacy risks. They seem to believe the benefits of disclosing personal information in order to use SNSs as greater than the potential privacy risks. Furthermore, they often reveal personal information on SNSs, not only theirs but also of others. For example, (exp 1) is a comment on a Facebook user profile.

(exp 1) I hope you had an amazing birthdayyy!

This comment was time-stamped. As a result, anyone, including unwanted audiences, could understand this user's birthday even if the user did not disclose his/her birthday on the profile. Also, we often find tweets where we can understand someone's birthday.

(exp 2) *Atashi no tanjyobi ha 8 gatu youka yo, Risshu tte itte 1 nen de mottomo atsui hi rashii wane–. Koyomi no ue deha dayo?*
(My birthday is August 8th, that is, the beginning day of autumn, and seems to be the hottest day of the year. Well, it is according to the calendar, you know?)

(exp 3) *@kahuhi kahuhi san tanjyobi omedetou gozaimasu!!*
(@kahuhi Mr. kahuhi, happy birthday!!)

Both (exp 2) and (exp 3) are tweets on Twitter. The sender of (exp 2) disclosed her birthday by herself. On the other hand, the sender of (exp 3) revealed his/her friend's birthday. In this paper, we focus on birthday information because we treat it differently than other personal information. For example, if someone revealed our name, address, age, gender, telephone number, or social security number on a SNS, we would get upset him/her for doing it. On the other hand, interestingly, if someone revealed our birthday in his/her birthday message on a SNS, like (exp 3), most of us would appreciate what he/she does, like (exp 4) and (exp 5).

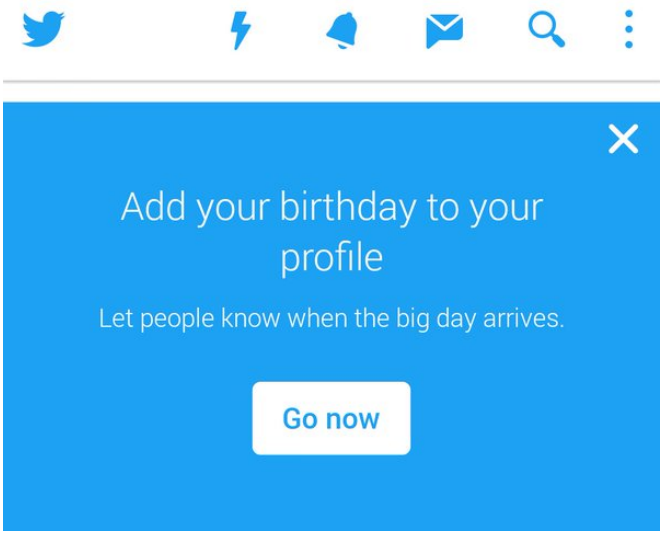

Figure 1. Twitter recommends us to add our birthday to our profiles.

(exp 4) *message kureta minna arigatou. yoi tanjyobi ni narimashita* – (＊^^＊)
(Thank you for birthday messages. I have a nice birthday – (＊^^＊))

(exp 5) *@taguma6 reina no mama no tanjyobi oboete kurete runyane, arigatou, sasuga*
(@taguma6 I'm glad to hear that you remember my mother's birthday. Thank you. Amazing.)

Birthday messages often give us opportunities to start new communications. As a result, as shown in Fig. 1, Twitter recommends us to add our birthday to our profiles. It is likely that these kinds of recommendations let SNS users discount the potential risks related to disclosing personal information. However, birthday information can be linkable to a specific individual when it is combined with other information. In order to deal with the privacy risks, it is important to investigate how we disclose or reveal personal information on SNSs, not only ours but of others. Birthday information especially should

 ISBN: 978-1-61208-516-6

be investigated carefully because we treat it differently than other personal information. Furthermore, it is important to investigate whether unwanted audiences can collect revealed personal information automatically. To solve these problems, in this paper, we investigate tweets where birthdays are revealed to other users and show how we communicate with each other about our birthdays. Furthermore, we discuss whether unwanted audiences can collect revealed birthday information by using machine learning techniques.

The rest of this paper is organized as follows: In Section II, we survey the related works. In Section III, we report how we disclose or reveal birthday information on Twitter. In Section IV, we discuss whether unwanted audiences can collect revealed birthday information by using machine learning techniques. Finally, in Section V, we present our conclusions.

II. Related works

Personally identifiable information is defined as information which can be used to distinguish or trace an individual's identity such as social security number, biometric records, etc. alone, or when combined with other information that is linkable to a specific individual, such as date and place of birth, mother's maiden name, etc. [1] [2]. Internet users are generally concerned about unwanted audiences obtaining personal information. Fox et al. reported that 86% of Internet users are concerned that unwanted audiences will obtain information about them or their families [3]. Also, Acquisti and Gross reported that students expressed high levels of concern for general privacy issues on Facebook, such as a stranger finding out where they live and the location and schedule of their classes, and a stranger learning their sexual orientation, name of their current partner, and their political affiliations [4]. However, Internet users, especially young users, tend to disclose personal information on their profiles, for example, real full name, gender, hometown and full date of birth, which can potentially be used to identify details of their real life, such as their social security numbers. In order to discuss this phenomenon, many researchers investigated how much and which type of information are revealed in SNSs, especially, in Facebook. Stutzman investigated Facebook profiles of University of North Carolina at Chapel Hill freshmen and found that 96.2% of them published their birthdays on their Facebook profiles, 74.7% their political views and 83.2% their sexual orientation [5]. Gross and Acquisti investigated Facebook profiles of Carnegie Mellon University students and found that 87.8% of them reveal their birth date on their profiles, 39.9% list their phone number, and 50.8% list their current residence [6]. Taraszow et al. observed Facebook profiles of 131 young people (68 females and 63 males, ages ranged from 14 to 29 years) and found that all participants disclosed their birthdays and 54.2% list their hometowns on their Facebook profiles [7]. Taraszow et al. also observed Cypriot Facebook users and found that they were willing to share personal information: All of them published their real names, 97% revealed their gender, 97% published a facial profile picture of themselves, 97% published their facial profile pictures, 51% indicated their hometowns and 88% published their birth date [8]. Huffaker and Calvert studied 70 teenage bloggers and found that 70% of them published their first names, 20% list their full names, 67% list their ages, and 39% list their birthdays [9]. Based on these results, researchers discussed the reasons why users willingly disclose personal information on their SNS profiles. Dwyer concluded in her research that privacy is often not expected or undefined in SNSs [10]. Barnes argues that Internet users, especially teenagers, are not aware of the nature of the Internet and SNSs [11]. Hirai reported that many users had troubles in SNSs because they did not mind that strangers observed their communication with their friends [12]. Viseu et al. reported that many online users believe the benefits of disclosing personal information in order to use an Internet site as greater than the potential privacy risks [13]. On the other hand, Acquisti and Gross explain this phenomenon as a disconnection between the users' desire to protect their privacy and their actual behavior [4]. Also, Livingstone points out that teenagers' conception of privacy does not match the privacy settings of most SNSs [14]. Joinson et al. reported that trust and perceived privacy had a strong affect on individuals' willingness to disclose personal information to a website [15]. Also, Tufekci found that concern about unwanted audiences had an impact on whether or not students revealed their real names and religious affiliation on MySpace and Facebook [16].

Next, we survey studies that focus on the issue of potential privacy risks of disclosing personal information. Birthday information alone cannot threaten the privacy and security of users. However, it can expose users' identities and threaten their privacy when combined with other personal information disclosed in their profiles. Sweeney reported 87% of Americans can be uniquely identified from a birth date, five-digit zip code, and gender [17]. Acquisti and Gross reported the existence of a potential ability to reconstruct users' social security numbers utilizing a combination of information often found in profiles, such as their full name, date of birth and hometown [4]. Many banks and credit-card companies recommend their customers to select a personal identification number (PIN) that cannot be easily guessed, for example, birth date [18] [19]. Bonneau et al. investigated 805 participants and found that 23% of them chose their PINs representing dates [20]. Furthermore, Bonneau et al. asked users about the significance of the dates in their PINs: 29% of them used their own birthday, 26% the birthday of a partner or family member, and 25% an important life event like an anniversary or graduation. As a result, we should be aware of the potential privacy risks on SNSs and manage our personal information carefully. SNSs do not force users to reveal personal information. However, we think, they actually recommend and encourage them to do so. As shown in Fig. 1, Twitter recommended users to add their birthdays on their Twitter profiles. On the other hand, Twitter enables each user to set the visibility preferences for his/her birthday on the profile from options [21] [22]:

- public,
- limited audience, or
- closed.

Fig. 2 shows a Twitter profile where a user sets the visibility preferences for his/her birthday. However, even if a user set it closed, his/her birthday would be revealed to others when the following kind of tweets was submitted.

(exp 6) *@446xx110rn tanjyobi omedetou!!*
(@446xx110rn Happy birthday!!)

We found many tweets where someone's birthdays were revealed and linked to specific Twitter accounts. We may say that

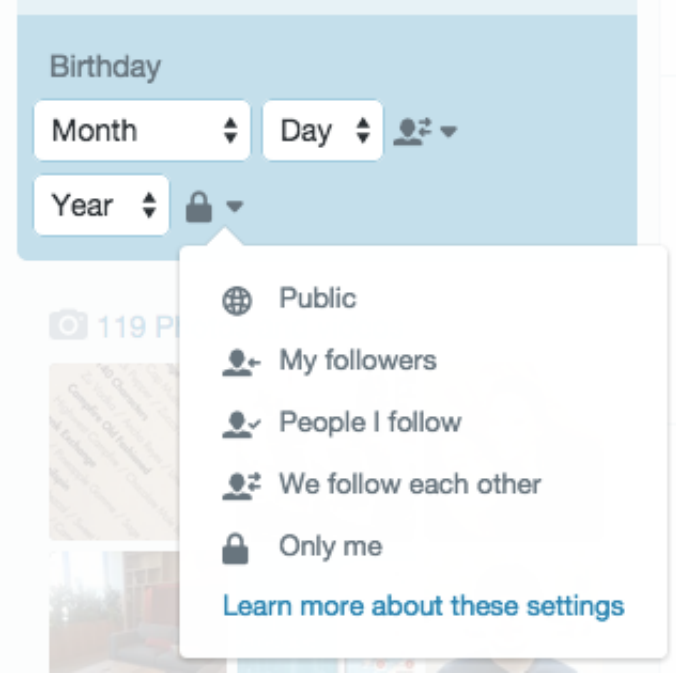

Figure 2. A Twitter user can set the visibility preferences for his/her birthday on the profile.

Fig. 1 and Fig. 2 show a disconnection between the Twitter's desire to protect their users' privacy and their actual behavior.

III. INVESTIGATION OF TWEETS WHERE BIRTHDAYS ARE REVEALED TO OTHER PEOPLE

In this section, we show how we disclose or reveal birthday information on Twitter.

A. The investigation object

We collected 1,000 Japanese tweets including word "*tanjyobi* (birthday)" in December 2015. We used these 1,000 tweets for investigating tweets where birthdays were revealed to other people.

Tweets can be classified into three types [23]:

- reply
  A reply is submitted to a particular person. It contains "@username" in the body of the tweet. For example, (exp 3), (exp 5), and (exp 6) are replies.
- retweet
  A retweet is a reply to a tweet that includes the original tweet.
- normal tweet
  A normal tweet is neither reply nor retweet. For example, (exp 2) and (exp 4) are normal tweets. Normal tweets are generally submitted to general public.

Table I shows the numbers and percentages of normal tweets, replies, and retweets in the 1,000 tweets. As shown in Table I, there were no retweets in the 1,000 tweets. On the other hand, Table II shows the numbers and percentages of normal tweets, replies, and retweets in the 7,085,267 Japanese tweets obtained in November and December 2012 by using the streaming API [24]. The comparison of Table I with Table II shows that

TABLE I. THE NUMBERS AND PERCENTAGES OF NORMAL TWEETS, REPLIES, AND RETWEETS IN THE 1,000 JAPANESE TWEETS INCLUDING "*tanjyobi* (BIRTHDAY)" (IN DECEMBER 2015).

	number	(percentage)
normal tweet	560	(56.0 %)
reply	440	(44.0 %)
retweet	0	(0.0 %)
total	1,000	(100.0 %)

TABLE II. THE NUMBERS AND PERCENTAGES OF NORMAL TWEETS, REPLIES, AND RETWEETS IN THE 7,085,267 JAPANESE TWEETS (IN NOVEMBER AND DECEMBER 2012).

	number	(percentage)
normal tweet	3,813,164	(53.8 %)
reply	2,528,642	(35.7 %)
retweet	743,461	(10.5 %)
total	7,085,267	(100.0 %)

TABLE III. THE CLASSIFICATION RESULT OF THE 1,000 TWEETS OBTAINED IN DECEMBER 2015 (BY HUMAN EXPERTS).

TYPE	whose birthday is revealed	normal tweet	reply	total
TYPE S	sender	51	32	83
TYPE R	receiver	0	211	211
TYPE N	no one	509	197	706
	total	560	440	1,000

word "*tanjyobi* (birthday)" was used more frequently in replies than normal tweets. We classified these 1,000 tweets into three types:

TYPE S tweets where sender's birthdays were disclosed by themselves,
TYPE R tweets where receiver's birthdays were revealed by senders, and
TYPE N tweets where no one's birthdays were revealed.

Table III shows the classification result. As shown in Table III, there were 294 tweets revealing senders' or receivers' birthdays. Furthermore, the number of tweets revealing receivers' birthdays (211 tweets) was more than twice the number of tweets revealing senders' birthdays (83 tweets). In this study, a tweet where someone's birthday was revealed but could not be linked to a specific Twitter account was classified into TYPE N: tweets where no one's birthdays were revealed. For example, the birthdays of *oniichan* (brother) in (exp 7) and *Chihiro Iwasaki* in (exp 8) were revealed but could not be linked to their Twitter accounts. As a result, in this study, these tweets were classified into TYPE N.

(exp 7) *kyou ha jikkei no tanjyobi! oniichan tanjyobi omedetou – ! 18 kin kaikin toka otona yana...*
(Today is my elder brother's birthday! Happy birthday, brother. Now, you can watch movies for adults only...)

(exp 8) *Iwasaki Chihiro san no tanjyobi nanoka*
(Today is the birthday of Chihiro Iwasaki.)

Chihiro Iwasaki was a famous Japanese artist.

 ISBN: 978-1-61208-516-6

B. Tweets where birthdays are revealed

1) Tweets where sender's birthdays are revealed (TYPE S): In order to start new communications on Twitter, many users submitted tweets where their birthdays were disclosed by themselves. The point is that senders disclosed their birthdays not only in normal tweets but replies. Both (exp 9) and (exp 10) were normal tweets where senders' birthdays were disclosed by themselves.

(exp 9) *kyou tanjyobi nanode dareka nonde kudasai!!!!*
(Today is my birthday. Does anyone keen to go drinking with me!!!!)

(exp 10) *shi-a-wa—se suggoi tanoshii tanjyobi deshita– !!! minasan no okagedesu. arigatou gozaimasu. toriaezu ashi itasugiru. hayo ie tsukan ka na-n*
(H-A-P-P-Y I had a very happy birthday!!! I do appreciate you. Thank you. Just say my foot hurts. I want to go home soon.)

On the other hand, (exp 11) was a reply where sender's birthday was disclosed by himself/herself.

(exp 11) @*takutwu_w takuto kun– kyou tanjyobi nanda oiwai rep hoshii na*
(@takutwu_w Takuto kun–, today is my birthday. Give me your birthday message, please.)

As shown in Table III, sender's birthdays were disclosed in normal tweets more frequently than replies. (exp 9) and (exp 10) were normal tweets and the senders of them wanted to communicate with anyone. On the other hand, (exp 11) was a reply and the sender of it wanted to communicate with a particular person (@takutwu_w). However, all of (exp 9), (exp 10), and (exp 11) were submitted for starting new communications on Twitter. On the other hand, (exp 12) was a reply where the sender disclosed her birthday not because she wanted to start a new communication but because she was asked when her birthday was.

(exp 12) @*kmns6_n teru-chan kon* (*´∇`*) *sou nano– kinou tanjyobi deshita. arigatoune– ♡ mata hitotsu toshi wo totte shimatta wa zutto nannimo itte kurenai kara akirame tetanda kedo, ureshii ♪*
(@kmns6_n Teru-chan hello (*´∇`*) Yes. Yesterday was my birthday. Thank you ♡ I got another year older again. I have got your birthday message out of my mind because you said nothing for a long time. I am happy ♪)

All of (exp 9), (exp 10), (exp 11), and (exp 12) were submitted within one day of senders' birthdays. On the other hand, (exp 13) and (exp 14) were not. The senders of (exp 13) and (exp 14) disclosed their birthdays by showing the dates.

(exp 13) *boku no tanjyobi ha, 2007 nen 9 gatsu 20 nichi goro da nya– (^^)*
(My date of birth is September 20, 2007 –(^^))

(exp 14) @*alex_hayate shigusa...uwame dukai toka? a, tanjyobi ha 8 gatsu nanoka desu*
(@alex_hayate gesture... up-from-under look? Oh, my birthday is August 7.)

The sender of (exp 15) disclosed her birthday by showing not the dates but whom she shared a birthday with.

(exp 15) *masaka no furukawa yuuki kun to onaji tanjyobi ww majime ni ureshii desu*
(Oh, I share a birthday with Yuuki Furukawa kun ww Very happy.)

Yuuki Furukawa in (exp 15) was an actor and his birthday might be published. However, we did not understand his birthday with just (exp 15). As a result, we determined that sender's birthday of (exp 15) was unclear. In this study, tweets where birthdays were revealed unclearly, such as (exp 15), were classified into TYPE N.

2) Tweets where receiver's birthdays are revealed (TYPE R): As shown in Table III, tweets where receivers' birthdays were revealed by senders were all replies. Furthermore, the number of replies where receivers' birthdays were revealed was almost half of the number of replies including word "*tanjyobi* (birthday)".

(exp 16) @*nami_1215_ nami tanjyobi omedetou!!!*
(@nami_1215_ Nami happy birthday!!!)

Tweets revealing receivers' birthdays were almost birthday messages to them, such as (exp 16).

3) Tweets revealing no one's birthdays (TYPE N): Tweets where birthdays could not be linked to specific Twitter accounts, such as (exp 17), (exp 18), and (exp 19), were classified into TYPE N: tweets where no one's birthdays were revealed.

(exp 17) *ke-taman tanjyobi omedetou –*
(ke-taman happy birthday –)

(exp 18) *kyou ha daisuki na aya chan no tanjyobi!!!*
(Today is my favorite Aya's birthday!!!)

(exp 19) @*hokoa_a Valentine Day- yade w Jingu no tanjyobi ww tsuraa www watashi ha iroiro dashi sugite tsurai ww*
(@hokoa_a Valentine's Day w Jingu's birthday ww hard www I had a hard time of it ww)

Just like (exp 15), we did not understand chiipopo's birthday with just (exp 20). As a result, (exp 20) was classified into TYPE N.

(exp 20) *watashi chiipopo to tanjyobi onaji yawa*
(I share a birthday with chiipopo.)

The senders of (exp 21) and (exp 22) showed what had happened or would happen on their birthdays. However, they did not show when their birthdays were. As a result, (exp 21) and (exp 22) were classified into TYPE N.

(exp 21) *22 sai no tanjyobi ni −20 °C no yukiyama de fuhatsudan shori shiteta.*
(On my 22th birthday, I did bomb disposal work in a snowy mountain, minus 20 degrees.)

(exp 22) *tanjyobi ni intern kakutei shita shini tai*
(I have to work on an internship program on my birthday. I'd rather die.)

The sender of (exp 23) asked the receiver when her birthday was. We could not understand her birthday with just (exp 23). As a result, (exp 23) was classified into TYPE N.

(exp 23) *iku chan kyou tanjyobi jya nakatta?*
(Iku chan. Is today your birthday?)

Tweets dealing with topics related to "birthday", but not someone's birthday, such as (exp 24) and (exp 25), were classified into TYPE N.

TABLE IV. FEATURES USED IN SVM METHOD FOR DATA TRAINING AND CLASSIFYING TWEETS INCLUDING WORD "*tanjyobi* (BIRTHDAY)".

$s1$	word unigrams of the tweet
$s2$	word bigrams of the tweet
$s3$	the number of words in the tweet
$s4$	word unigrams of the first sentence of the tweet
$s5$	word bigrams of the first sentence of the tweet
$s6$	the number of words in the first sentence of the tweet
$s7$	the last word of the first sentence of the tweet
$s8$	character unigrams of the tweet
$s9$	character bigrams of the tweet
$s10$	character 3-grams of the tweet
$s11$	the length of the tweet
$s12$	character unigrams of the first sentence of the tweet
$s13$	character bigrams of the first sentence of the tweet
$s14$	character 3-grams of the first sentence of the tweet
$s15$	the length of the first sentence of the tweet
$s16$	whether the tweet is a reply

(exp 24) *jissai, 2/29 umare no hito tte inno?? koseki ni 2/29 tte touroku shitara 4 nen ni 1 kai shika tanjyobi konai yona.*
(Actually, are there people born on Feb.29?? If the birthdays were registered correctly, they would have their birthday every four years.)

(exp 25) *@BBCNNHK douse nara suihanki to nanige nai kaiwa shite tanjyobi oboete kureru tekina yatsu ga eena*
(@BBCNNHK I might as well buy a rice cooker that deduces my birthday from a daily chat.)

IV. DETECTION OF TWEETS WHERE BIRTHDAYS ARE REVEALED TO OTHER PEOPLE

If we detect tweets revealing someone's birthdays automatically, we can give warnings to users before they submit their tweets where someone's birthdays are revealed. In this section, we discuss whether we can automatically detect tweets where someone's birthdays are revealed by using machine learning techniques.

In this study, we used the support vector machine (SVM) for data training and classifying. Table IV shows feature $s1 \sim s16$ used in machine learning on experimental data. $s1 \sim s7$ were obtained by using the results of morphological analysis on experimental data. In the experiments, we used a Japanese morphological analyzer, JUMAN for word segmentation of tweets [25]. $s8 \sim s10$ and $s12 \sim s14$ were obtained by extracting character N-gram from experimental data. Odaka et al. reported that character 3-gram is good for Japanese processing [26]. $s4 \sim s7$ and $s12 \sim s15$ were obtained from first sentences of tweets. This is because, we thought, clue expressions of birthday messages are often found at first sentences of tweets.

In this study, we used the 1,000 tweets investigated in Section III for the experimental data. We conducted this experiment using TinySVM [27]. Table V shows the experimental result. The experimental result was obtained with 10-fold cross-validation. As shown in Table III, the experimental data consisted of 560 normal tweets and 440 replies. We divided the experimental result (Table V) into those of 560 normal tweets (Table VI) and 440 replies (Table VII).

TABLE V. THE SVM CLASSIFICATION RESULT OF THE 1,000 TWEETS INCLUDING WORD "*tanjyobi* (BIRTHDAY)".

whose birthday is revealed	SVM result: sender	receiver	no one	recall
sender	17	5	61	0.20
receiver	0	185	26	0.88
no one	7	36	663	0.94
precision	0.71	0.82	0.88	

TABLE VI. THE SVM CLASSIFICATION RESULT OF THE 560 NORMAL TWEETS INCLUDING WORD "*tanjyobi* (BIRTHDAY)".

whose birthday is revealed	SVM result: sender	receiver	no one	recall
sender	9	0	42	0.18
receiver	0	0	0	—
no one	5	3	501	0.98
precision	0.64	0.00	0.92	

TABLE VII. THE SVM CLASSIFICATION RESULT OF THE 440 REPLIES INCLUDING WORD "*tanjyobi* (BIRTHDAY)".

whose birthday is revealed	SVM result: sender	receiver	no one	recall
sender	8	5	19	0.25
receiver	0	185	26	0.88
no one	2	33	162	0.82
precision	0.80	0.83	0.78	

As shown in Table V, 865 tweets were classified correctly and 135 tweets incorrectly in this experiment. 66 tweets out of 135 incorrectly classified tweets were ones where sender's birthdays were revealed. As shown in Table V, the recall of tweets revealing senders' birthdays were 20%. As shown in Table VI and Table VII, many tweets revealing senders' birthdays were classified incorrectly into tweets revealing no one's birthdays. As a result, it is difficult to detect tweets revealing senders' birthdays and give warnings to senders before they submit tweets revealing their birthdays. On the other hand, as shown in Table V, the precision of tweets revealing senders' and receivers' birthdays were 71% and 82%, respectively. Our method is useful for collecting tweets revealing birthdays precisely. As a result, it is easy for attackers to collect birthday information related to specific Twitter accounts by using our method.

V. CONCLUSION

Many people willingly disclose their birthdays on their SNS profiles and reveal others' birthdays on their SNS messages. They seem unaware of the potential risks of doing it. Birthday information alone cannot threaten their privacy and security. However, it can expose users' identities and threaten their privacy when combined with other personal information disclosed in their profiles. Interestingly, we treat birthday information differently than other personal information. For

 ISBN: 978-1-61208-516-6

example, if someone revealed our personal information except birthday on a SNS, we would get upset him/her for doing it. On the other hand, if someone revealed our birthday in his/her birthday message on a SNS, most of us would feel happy and appreciate what he/she does. However, we have not sufficiently investigated how we reveal birthday information on SNSs. As a result, the authors investigated how we reveal birthday information on SNSs, not only ours but of others.

In this study, we investigated tweets where someone's birthdays were revealed to other people. We collected 1,000 Japanese tweets including word "*tanjyobi* (birthday)" and found that about 30% of them were tweets where someone's birthdays were revealed to other people. Furthermore, about 70% of tweets revealing someone's birthdays were ones where receivers' birthdays were revealed by senders. In this study, we proposed a method of detecting tweets revealing someone's birthday by using machine learning techniques. The experimental result showed that our method was able to classify tweets including word "*tanjyobi* (birthday)" with accuracy of 87%. However, the recall of tweets revealing senders' birthday was only 20%. As a result, in our method, it is difficult to detect tweets revealing senders' birthdays and give warnings to senders before they submit them. On the other hand, the precision of tweets revealing senders' and receivers' birthdays were 71% and 82%, respectively. As a result, in our method, it is not difficult to collect tweets revealing birthdays precisely. We recommend that birthday messages should not be sent via SNSs. This is because unwanted audiences can read and collect them. We are now investigating other language tweets where birthdays are disclosed or revealed to other people.

REFERENCES

[1] C. Johnson III, Safeguarding against and responding to the breach of personally identifiable information, Office of Management and Budget Memorandum, 2007. [Online]. Available: http://www.whitehouse.gov/omb/memoranda/fy2007/m07-16.pdf [accessed: 2016-10-4]

[2] B. Krishnamurthy and C. E. Wills, "On the leakage of personally identifiable information via online social networks," in Proceedings of the 2Nd ACM Workshop on Online Social Networks, ser. WOSN '09. New York, NY, USA: ACM, 2009, pp. 7–12. [Online]. Available: http://doi.acm.org/10.1145/1592665.1592668 [accessed: 2016-10-4]

[3] S. Fox et al., Trust and Privacy Online: Why Americans Want to Rewrite the Rules, The Pew Internet & American Life Project, 2000. [Online]. Available: http://www.pewinternet.org/˜/media//Files/Reports/2000/PIP_Trust_Privacy_Report.pdf.pdf [accessed: 2016-10-4]

[4] A. Acquisti and R. Gross, Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 36–58.

[5] F. Stutzman, Student life on the Facebook, 2006. [Online]. Available: http://www.ibiblio.org/fred/facebook/stutzman_fbook.pdf [accessed: 2016-10-4]

[6] R. Gross and A. Acquisti, "Information revelation and privacy in online social networks," in Proceedings of the 2005 ACM Workshop on Privacy in the Electronic Society, ser. WPES '05. New York, NY, USA: ACM, 2005, pp. 71–80.

[7] T. Taraszow, E. Aristodemou, G. Shitta, Y. Laouris, and A. Arsoy, "Disclosure of personal and contact information by young people in social networking sites: An analysis using Facebook profiles as an example," International Journal of Media and Cultural Politics, vol. 6, no. 1, 2010, pp. 81–101.

[8] T. Taraszow, A. Arsoy, G. Shitta, and Y. Laouris, "How much personal and sensitive information do cypriot teenagers reveal in facebook?" in Proceedings of the 7th European Conference on E-Learning, 2008, pp. 606–611.

[9] D. A. Huffaker and S. L. Calvert, "Gender, identity, and language use in teenage blogs." Journal of Computer-Mediated Communication, vol. 10, no. 2, 2005. [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1111/j.1083-6101.2005.tb00238.x/full [accessed: 2016-10-4]

[10] C. Dwyer, "Digital relationships in the "myspace" generation: Results from a qualitative study," in Proceedings of the 40th Annual Hawaii International Conference on System Sciences, ser. HICSS '07. Washington, DC, USA: IEEE Computer Society, 2007, p. 19.

[11] S. B. Barnes, "A privacy paradox: Social networking in the united states." First Monday, vol. 11, no. 9, 2006. [Online]. Available: http://firstmonday.org/article/view/1394/1312 [accessed: 2016-10-4]

[12] T. Hirai, "Why does "Enjyo" happen on the Web? : An Examination based on Japanese Web Culture," Journal of Information and Communication Research, vol. 29, no. 4, mar 2012, pp. 61–71. [Online]. Available: http://doi.org/10.11430/jsicr.29.4_61 [accessed: 2016-10-4]

[13] A. Viseu, A. Clement, and J. Aspinall, "Situating privacy online: Complex perception and everyday practices," Information, Communication & Society, 2004, pp. 92–114.

[14] S. Livingstone, "Taking risky opportunities in youthful content creation: teenagers' use of social networking sites for intimacy, privacy and self-expression." New Media & Society, vol. 10, no. 3, 2008, pp. 393–411.

[15] A. N. Joinson, U.-D. Reips, T. Buchanan, and C. B. P. Schofield, "Privacy, trust, and self-disclosure online." Human-Computer Interaction, vol. 25, no. 1, 2010, pp. 1–24. [Online]. Available: www.joinson.com/home/pubs/HCI_journal.pdf [accessed: 2016-10-4]

[16] Z. Tufekci, "Can You See Me Now? Audience and Disclosure Regulation in Online Social Network Sites," Bulletin of Science, Technology & Society, vol. 28, no. 1, 2008, pp. 20–36.

[17] L. Sweeney, "Uniqueness of Simple Demographics in the U.S. Population," LIDAP-WP4 Carnegie Mellon University, Laboratory for International Data Privacy, Pittsburgh, Pennsylvania, 2000. [Online]. Available: http://dataprivacylab.org/projects/ identifiability/index.html [accessed: 2016-10-4]

[18] VISA, "Issuer PIN Security Guidelines," http://usa.visa.com/dam/VCOM/download/merchants/visa-issuer-pin-security-guideline.pdf [accessed: 2016-10-4], 2010.

[19] HSBC, "New service for HSBC cards — PIN (personal identification number) change via HSBC ATMs," https://www.hsbc.am/1/2/am/en/new-service-for-hsbc-cards [accessed: 2016-10-4], 2016.

[20] J. Bonneau, S. Preibusch, and R. Anderson, "A birthday present every eleven wallets? the security of customer-chosen banking pins," in The 16 th International Conference on Financial Cryptography and Data Security, 2012, pp. 25–40.

[21] Twitter, "Customizing your profile," https://support.twitter.com/articles/127871 [accessed: 2016-10-4].

[22] ——, "Profile visibility settings," https://support.twitter.com/articles/20172733 [accessed: 2016-10-4].

[23] Y. Watanabe, K. Nakajima, H. Morimoto, R. Nishimura, and Y. Okada, "An investigation of a factor that affects the usage of unsounded code strings at the end of japanese and english tweets," in Proceedings of the Seventh International Conference on Evolving Internet (INTERNET 2015), Oct 2015, pp. 50–55. [Online]. Available: https://www.thinkmind.org/index.php?view=article&articleid=internet_2015_2_40_40038 [accessed: 2016-10-4]

[24] Twitter, Inc. The Streaming APIs. [Online]. Available: https://dev.twitter.com/streaming/overview [accessed: 2016-10-4]

[25] S. Kurohashi and D. Kawahara, JUMAN Manual version 5.1 (in Japanese). Kyoto University, 2005.

[26] T. Odaka et al., "A proposal on student report scoring system using n-gram text analysis method," The transactions of the Institute of Electronics, Information and Communication Engineers. D-I, vol. 86, no. 9, sep 2003, pp. 702–705. [Online]. Available: http://ci.nii.ac.jp/naid/110003171273/en/ [accessed: 2016-10-4]

[27] Taku Kudoh. TinySVM: Support Vector Machines. [Online]. Available: http://chasen.org/taku/software/TinySVM/index.html [accessed: 2016-10-4]

 ISBN: 978-1-61208-516-6