2014 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)
July 6-11, 2014, Beijing, China

Behavior Extraction from Tweets using Character N-gram Models

Yuji Yano, Tomonori Hashiyama, Junko Ichino and Shun'ichi Tano

Abstract— **Human daily activities are stored in various kinds of data representations using ICT devices nowadays, named lifelogs. It is highly requested to retrieve useful information from lifelogs because these raw data are hard to handle. Extracting human activities from these logs is promising to enrich our life. Context-awareness services can be provided depending on user activities extracted from these logs. Recently, a lot of people post a message called tweet within Twitter to show what they are doing, thinking, feeling, and so on. Tweets have potential to record human activities, because many people post tweets so frequently every day. This paper focused on the tweets to retrieve human behavior from them. The length of tweets are limited within short sentence, so this causes some difficulties. The users will use domain-specific terms and will post grammatically incorrect sentences to fit with the constraints. These make us hard to analyze tweets with grammatical manner or with dictionaries. To tackle them, we are applying character n-gram tokenization and naïve Bayes classifier to extract appropriate behavioral information from tweets. Using n-gram tokenizer, domain-specific words can be identified and incorrect grammar can be handled. Our approach is examined using real tweets in Japanese. The index of precision, recall and F-measure shows the promising results. Some experiments have been carried out to show the feasibility of our approach. At this point, our system applied to Japanese tweets but it is applicable to any other languages.**

I. Introduction

Context-awareness services [1] which provide appropriate information depending on users' situation are getting quite popular recent years. The context is estimated based on users' current behavioral information what they are doing, the environmental information by which they are surrounded and locational information where they are, and so on. It is getting easier to obtain these data with information technologies, especially using smartphones, some of them are called lifelogs. It is also very important to know users' preferences for better context recognition. To obtain these contextual information, there are researches using extra sensors [2] [3], user operation logs on PC or smartphone [4], text data [5], food photos captured by themselves [6] and so on. With these kinds of daily data of lifelogs, context-awareness services are able to provide suitable information for the users. In this paper, we focus on Twitter for better estimation of users' behavior. The message posted on Twitter is named "tweet". Tweet can represent users' behavior in almost real-time responses with much rich information than sensor data themselves. For instance, "I'm going to sleep" directly represents that he/she is going to bed.

Yuji Yano, Tomonori Hashiyama, Junko Ichino and Shun'ichi Tano are with the Department of Human Media Systems, Graduate School of Information Systems, The University of Electro-Communications, Chofu, Tokyo, Japan (email: yano@media.is.uec.ac.jp, {hashiyama, ichino, tano)}@is.uec.ac.jp.

Twitter is a kind of social networking services, which spread rapidly these years. A tweet is limited within a short text message. With this limitation, users post tweets very frequently while blogs and other web services are updated once a day or so. Sometimes, their tweets represent what they are doing and thinking. The word "now" is usually attached to each tweet in Japanese. Therefore, we try to extract users' behaviors from tweets. Typical examples are that tweets show the behaviors themselves like using verbs. There are many other tweets which indirectly represents users' behavior. For example, in Japanese greeting word "ittekimasu" means that they are going out and mimetic word "mukuri" means that they woke up. There are two approaches [7] [8] to extract the users' behaviors from tweets. One is based on the dictionary on co-occurrence frequency of words used to indicate behavior. The accuracy of the approach depends on whether appropriate words can be prepared in dictionaries to refer which words represents users' behaviors. Another is to use grammatical model which define the attributes such as actor, action, object, time and location. It is well known that tweets have their domain-specific words which are not in the dictionaries, and are not written in correct grammer. These grammatical approaches for ill-formed texts are not so suitable for tweets.

In this paper, we are investigating in Japanese tweets. Another problem arise using Japanese. In Japanese sentences, words are not usually separated by spaces. To find appropriate division point is another challenge to extract the behavioral information from ill-formed texts.

In this paper we will propose a novel approach to the challenges for extracting behavioral information from ill-formed texts in Japanese tweets. Character n-gram tokenization is applied to handle both domain-specific words and incorrect grammars. Naïve Bayes classifier is applied to classify tweets into behavior or not. The decisions whether they represent behaviors or not are provided manually by the users in the training phase. With this approach, frequently used domain-specific words can be identified. It is also applicable for the new expression which are not used correctly, for example "mukuri" as mentioned above. In addition, classified tweets are presented to users in a single unit such as one day. Users correct only misclassified results through simple user interface, and these tweets are added to training data. By repeating this, the classifier is able to classify with higher degree of accuracy, depending on each user.

In the remainder of this paper, related researches are roughly investigated in section II. Our novel approach based on character n-gram tokenization is described in section III. Section IV shows the experimental conditions. Section V verifies our approach empirically from the F-measure index

978-1-4799-2072-3/14/$31.00 ©2014 IEEE

and discuss in detail one by one for each user. Finally, we will draw some conclusions and future scopes.

II. Related Works

In this section, we will mention about previous researches to extract users' behavioral information from several kinds of daily logs.

A. Supporting Manual Input for Lifelogs

Lifelog Systems are getting popular recently. In the lifelog systems, it is possible to investigate human activities and behaviors. To retrieve human behavior from lifelogs, naïve approach is to record their activity manually by the users themselves. But it is easy to imagine that without any support for input procedure, users will immediately give up the recording, because it will be a burden for them. There are some support systems to reduce those users' input efforts. In [3], using user's location obtained from GPS or Wi-Fi connections, the system will show the only possible behaviors in that place. For example, user can not select "deskwork" on the road. Users will select appropriate activity within the limited choices. In addition, the system set the tag that the user most chosen on same place in the past as a default value. Another approach is handling meal and analyze the nutrients in the food using meal photo taken by the user[6]. In this research, the system classify all the taken photos into meal or not. Secondly, the system estimate the nutrients in the dish by analyzing the photos. Next, the system visualize the estimated nutrients and typical icon which represent the dish. Finally it will suggest desirable menu for next meal.

In these researches, these systems use manually inputted data for extraction of behavioral information. The error in analysis for the correct behavior or not is small. On the other hand, to input data manually is a burden for a user and several users surrender in the half way of using these systems. Therefore, we have to tackle the problem on user's burden and continuation of use.

B. Extracting Behavior Using Sensory Data

There are approaches to extract useful information from sensory data attached to human bodies or to the surrounding objects. In [2], the system obtains user's postural information using three accelerometers that the user worn. Secondly, the system recognize a object with a RFID tags attached to the object such as cup, toothbrush, iron, and so on. RFID reader is also attached to user's hand. Using these information, the system estimates his/her behavior and the environmental information using by decision tree classification. Another research is trying to extract behavioral information from knowledge workers using sensory data [4]. In this research, the sensors on the smartphone and the histories of PC operations are used. PC operation histories are the logs of the application activations and mouse movements.

In these researches, these systems use sensory data when extracting users' behavior. They automatically estimate and extract behavior with specific algorithm. However, in these methods, to extract behavior with small motion is difficult. Handling noisy sensory data also cause implementation difficulties.

C. Behavior Extraction from Text Data

There are researches to extract behavioral information from Twitter like ours are performed [7] [8]. In [7], the system extract behavioral information based on co-occurrence frequency of the words that indicate behavior and a category or time in a tweet. In this research, the list of words that indicate behavior, a category and time are given a for the system a priori. The system estimate user's interest of action. For example, by finding words that indicate behavior or time that co-occur with words that indicate a category in a tweet. In [8], the system extract the attributes such as actor, action, object, time and location from a structure of a sentence in a tweet with Japanese. The system is composed of the self-supervised learner and the behavior extractor. In the self-supervised learner, by extracting the attributes of behavior using morphological analysis and syntactic analysis, the learner learn a set of feature function by a template file and conditional random field using these result. The behavior extractor extract the attributes using learned results of the self-supervised learner by applying morphological analysis to a tweet.

These approach are based on conventional natural language processing. It is hard to apply these approach to tweets because there are many sentences with incorrect grammars and used domain-specific words which are not listed on dictionary. To extend the dictionary, we proposed to construct behavioral dictionary extracting from tweets[9]. In [9], human subjects manually choose keywords which seems to indicate behaviors. These keywords are stored into the behavioral dictionary. This behavioral dictionary approach is considerably useful, it has limitation to apply for a large scale of tweets. It is required to reduce human manual operation for registering to the dictionary. Based on this motivation, character n-gram tokenization and naïve Bayes classifier are introduced in this paper

III. Proposed System

A. Overview

The investigation described in the last section leads us to the novel system to extract the behavioral information from tweets. As mentioned, tweets include rich information about users' behavior, because it can represent the users intention other than analyzing raw sensory data. The analyzing problem in tweets mainly lies on two problems. One is that some words which indicate user activities are domain-specific. This will raise a problem in using ordinal dictionaries. Few appropriate words can be drawn from ordinal dictionaries. Another is caused by the limitation of maximum number of characters which can be used in tweets. Because of this limitation, users sometimes do not follow the grammar. To overcome these problems, we will propose a novel approach based on character n-gram tokenization and naïve Bayes

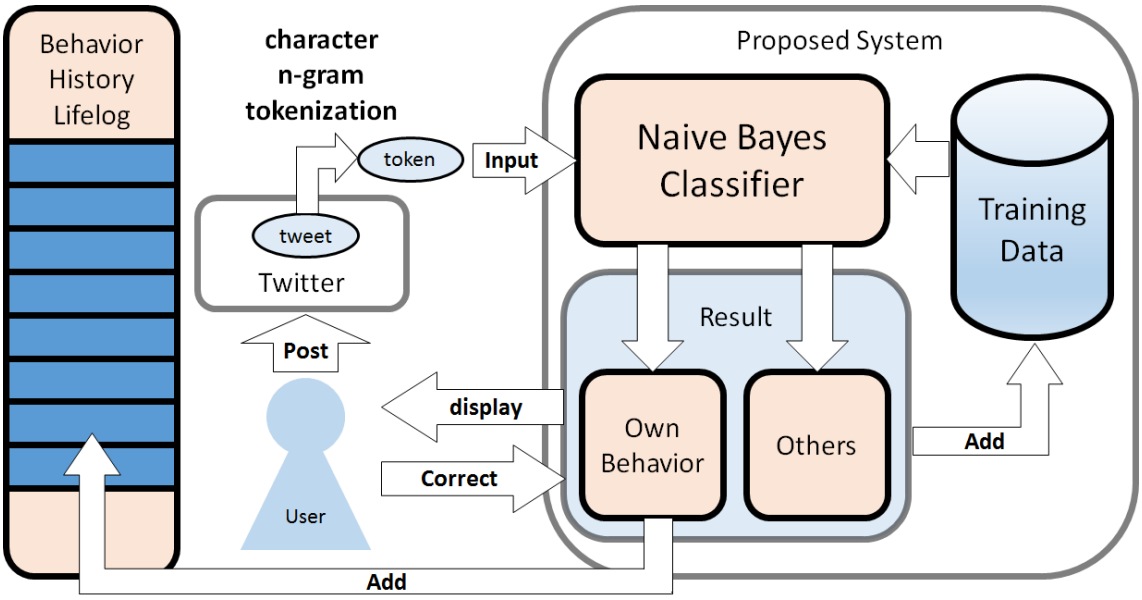

Fig. 1. Outline of Proposed System

classifier. Fig. 1 shows the outline of proposed system. When user posts his/her tweet, the system classified the tweet whether it indicates behavioral information or not. The classification procedures are as follows; The tweet posted by the users are passed to the character n-gram tokenizer and divided into tokens. These tokens are used as inputs for the classifier. The classifier is constructed based on naïve Bayes classifier. Classification results are shown to the user when they represent the behaviors. If and only if user find misclassification, user will re-classify it into appropriate category. This data is stored into the training database which are used to train the classifier. Repeating these procedure, as the training database are updated with users' preferences, classification error should be smaller and smaller. Suppose that training data are shared with other users, the system is able to classify a tweet using other user's tweet even when the users use the system for first time.

B. Character N-gram Tokenization

Sentences are decomposed into arbitrary n-characters by the character n-gram tokenizer. The tokens obtained from this process are used as features of the sentences. These feature can overlap with other features. When the length of a sentence is m, we obtain features of $m+n+1$. Character n-gram tokenization is a kind of bag-of-words model. It deconstructs the sentence into n-character units and does not consider the order of words and structure of the sentence. This characteristics enables us to handle sentence with incorrect grammars and with new domain-specific words in tweets. On the other hand, there is a problem that computational complexity increases with a large amount of data. We will take care of this problem by applying feature selection that trim unnecessary ones. Feature selection is carried out using pointwise mutual information and using relative entropy [10]. In this paper, we classify behavior sentences using frequency of feature appearance that are obtained by character n-gram tokenization.

C. Naïve Bayes Classifier

Naïve Bayes classifier is a classifier using Bayes' theorem. In document classification, given the feature vector $\mathbf{w}$ of a document, the probability $p(c_j|\mathbf{w})$ that the document belongs to the class c_j is given by eq. (1).

$$p(c_j|\mathbf{w}) = \frac{p(\mathbf{w}|c_j)p(c_j)}{p(\mathbf{w})} \tag{1}$$

In eq. (1), we remove probability $p(\mathbf{w})$ that are independent from the class c_j. We assume that each feature is generated independently, we obtain the value proportional to probability that the document belongs the class c_j. The system calculate the value for each class, and classify the document into appropriate class referring maximum value. Unclassified document composed of feature vector $\mathbf{w}$ is classified to c_{new} is given by eq. (2).

$$c_{new} = \arg\max_j \Big((\prod_{i=1}^{n} p(w_i|c_j)) p(c_j) \Big) \tag{2}$$

Naïve Bayes classifier assumes that each feature is conditional independent. Even when each feature is not conditional independent, the probability of misclassification is practically very low as shown in [11]. Now, we can suppose that this classifier is also effective to features that is not conditional independent from the feature divided from character n-gram tokenization. In naïve Bayes classifier, it suffered from

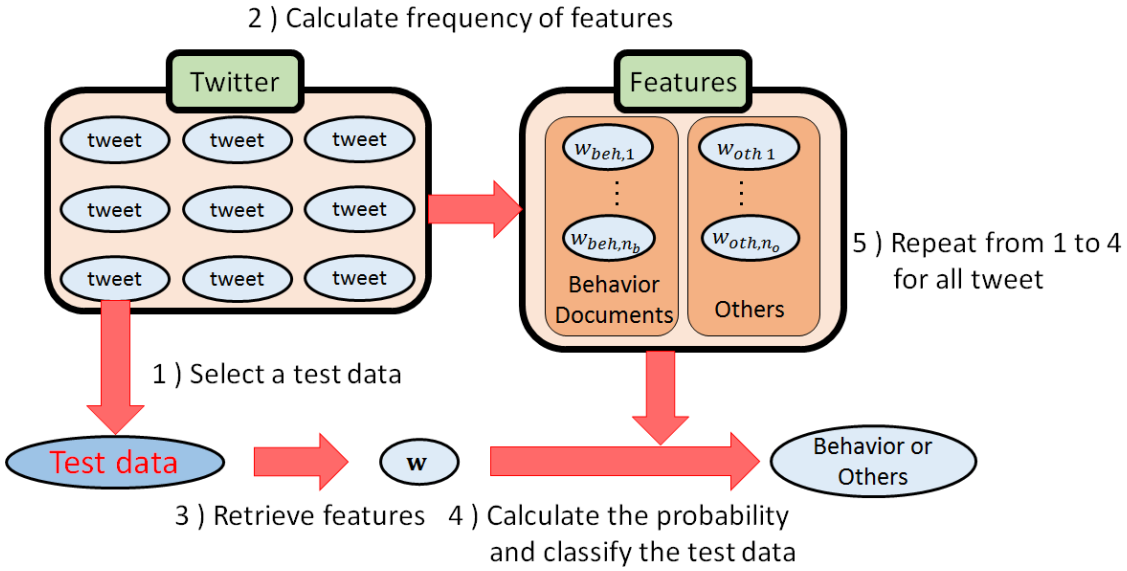

Fig. 2. Experimental Flow

problem named zero frequency problem that conditional probability $p(w_i|c_j)$ goes to 0 in the case where the feature w_i never appeared in the class c_j in the training data. This causes problem when this classifier runs with new data. This problem often occurs when the number of training data is small. We handle this by using a method that allow a very small conditional probability to a feature when conditional probability of the feature is 0. This modification is called smoothing.

IV. Experiments

Experiments are designed to show the effectiveness of our approach using real Japanese tweets. We use Twitter API provided by Twitter Inc. to retrieve each tweet. The value n of character n-gram tokenization is set to 1, 2 and 3.

A. Data Set

Japanese tweets of 10 users were chosen at random in our experiments. They are not our acquaintances. We don't know their ages, sexes and occupations, etc. These are posted on January 21 until January 31, 2012. Reply and retweet are excluded because they do not indicate individual behavior. The number of tweet posted in 11 days is 445.8 in average per person and the standard deviation is 218.0. The average and the standard deviation show that we could choose various users with different frequency of posting tweets. Referring our experiments with human decision, 5 collaborators from our lab students are asked to classify tweets whether it represents behavior or not. When the decision is split into yes and no, it follows majority vote.

B. Evaluation Index

We evaluate proposed method using precision, recall and F-measure shown in eqs. (3) to (5).

$$precision = \frac{R}{N} \tag{3}$$

$$recall = \frac{R}{C} \tag{4}$$

$$F - measure = \frac{R}{\frac{1}{2}(N+C)} \tag{5}$$

Where R represents number of the correct data among all the extracted data reffering to the decision number subjects. N shows the number of tweets extracted as behavior with proposed method. C is tweet number of all the correct data assigned in IV.A. We calculate the value of precision, recall and F-measure of each 10 user.

C. Behavior Extraction using All the Data

The system is trained using all tweets in data set. The experiments are carried out by classifying the tweets using these trained results. This assumes the situation when a user continues to use the proposed system as shown in Fig. 1. Cross-validation with the following procedure are examined. Fig. 2 shows the flow of the experiments.

1) Selecting one tweet from set of tweets.
2) Rest of tweets are used as training data, the system calculate each frequency of features in a document that indicate behavior or not.
3) Applying character n-gram tokenization to the tweet, and retrieve the features of this tweet.
4) Using frequency of features in training data, the system classify the tweet by calculating each probability for each class.

TABLE I
THE NUMBER OF CLASSIFICATION FOR EACH METHOD

	Tweet	Manual		Proposed Method		Previous Method	
		Behavior	Behavior Ratio	Estimated	True Positive	Estimated	True Positive
user1	962	172	0.179	194	138	190	119
user2	384	60	0.156	50	36	60	39
user3	290	25	0.086	41	19	33	13
user4	675	58	0.086	59	30	67	31
user5	529	65	0.123	66	35	59	26
user6	338	34	0.101	25	15	48	21
user7	347	61	0.176	47	44	64	46
user8	352	52	0.148	42	30	86	33
user9	235	36	0.153	32	27	46	28
user10	446	81	0.182	64	53	72	54
average	455.8	64.4	0.144	62.0	42.7	72.5	41.0

5) The system repeats the procedure from 1 to 4, we evaluate the result using precision, recall and F-measure of the classification results.

To show the improvement of the approach, these results are compared with our previous method based on behavior dictionary[9]. We compare proposed method with previous method by calculating the value of precision, recall, and F-measure. In previous method, there are two problem. First, the method require additional work to choose the features manually that is registered for *behavior dictionary*. Second, the method must always update *behavior dictionary* as time goes.

To examine effective features in the classifier constructed by proposed method, we calculate information gain $IG(w_i)$ shown in eq. (6) for each feature w_i to obtain the feature at high information gain.

$$IG(w_i) = H(c) - H(c|w_i) \quad (6)$$

Where $H(c)$ is entropy of the class c, $H(c|w_i)$ is conditional entropy that conditioned by the feature w_i. Because high information gain of the feature w_i show that the feature w_i decrease fuzziness of classification, we use information gain to evaluate effectiveness of features.

To examine the sensitivity to the number of training data, we carried out the experiments with changing the number of training data from 100 to 4000. Training data are selected randomly from all tweets except test data. These experiments are carried out 100 times for each number of training data.

In addition, the system are trained by using tweets that are posted users except the user who post test data, and classify the tweet using this training result. This assumes the situation when the user uses the proposed system as shown in Fig. 1 for first time. The results are cross-validated.

TABLE II
F-MEASURE FOR EACH METHOD

	Proposed	Previous
user1	0.754	0.657
user2	0.655	0.650
user3	0.576	0.448
user4	0.513	0.496
user5	0.534	0.419
user6	0.508	0.512
user7	0.815	0.736
user8	0.638	0.478
user9	0.794	0.683
user10	0.731	0.706
average	0.652	0.579

V. EXPERIMENTAL RESULTS AND DISCUSSIONS

A. Behavior Extraction using All Data

TABLE I shows the experimental results both by proposed method and previous method [9]. Tweet means total number of tweet, Behavior represents the number of tweets classified as indicated behavior manually, which are regarded as correct data in these experiments. Behavior Ratio is the ratio of the correct data to total number of tweet. In average, Behavior Ratio 0.144 means that about 14% of tweets correspond to the user behavior. Estimate is the number of extracted data as behavior by each method. True Positive is the number of tweet that matched to the correct data. The number that extracted as indicating behaviors by proposed method is less than the number by previous method, 62.0 and 72.5 in average, respectively. On the other hand, the number of True Positive in the proposed method is slightly larger than that by previous method, 42.7 and 41.0 in average, respectively. Proposed method appropriately extracts tweets that indicate behavior with smaller number than previous method. TABLE II shows F-measure of behavior extraction both by proposed method and previous method. Results of F-measure shows the advantage of our proposed method.

Fig. 3 shows the difference of the evaluation result for each users both by proposed method and previous method. In

TABLE III
FEATURE AND INFORMATION GAIN

Features	Meaning	Information Gain
o ha yo	A part of the word that indicates action of waking up	0.076977
ya su mi	A part of the word that indicates going to bed	0.076971
ta da i	A part of the word that indicates returning home	0.076803
ki ta ku	The word that indicates returning home	0.076777
na u	Twitter-specific word that indicates behavior “now”	0.076728
mu ku ri	Twitter-specific word that indicates action of waking up	0.076693
go ha n	The word that indicates meal	0.076689
o wa ri	The word that indicates end of behavior	0.076687
ne ru	The word that indicates going to sleep	0.076685
o hu ro	The word that indicates bathing	0.076671

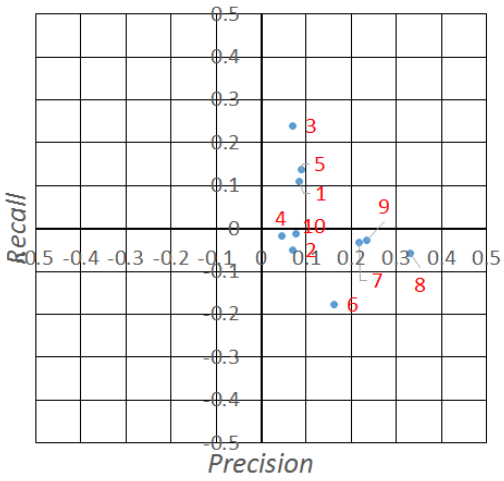

Fig. 3. Difference of Precision and Recall for each Method

Fig. 3, horizontal axis is the difference of precision, vertical axis is the difference of recall and the number indicate the user ID. From Fig. 3, we can see that the precisions for all the users are better for proposed method because all the data are plotted in the positive domain in horizontal axis. On the other hand, the number of the users that recall by proposed method is more than recall by previous method is 3. However, in user 1, user 3 and user 5 that recall by proposed method is more than recall by previous method, the range that increase recall is more than 0.1 in all users. In users that recall proposed method is less than by previous method, the user who range that decrease recall more than 0.1 is only user 6. Therefore, the average of recall by proposed method is almost identical to the average of recall by previous method. F-measure is increasing in all users except user 6. Proposed method is effective in almost users because precision and recall by proposed method is increasing.

In Fig. 3, precision and recall by proposed method is better than those by previous method in almost users. Precision by proposed method is especially improving a lot. In behavior extraction, we consider that users expect high recall to extract behavior. On the other hand, we do not expect to extract behavior by mistake. The indices for proposed method is higher than previous method shows our approach effective. In addition, proposed method does not need troublesome task to prepare behavioral dictionary. The classifier in proposed method is updated to add new correct data as training data with small operation when it misclassified. There still exist a problem that does not attach the label of behavior in proposed method unlike previous method.

TABLE III shows ten examples of extracted features with highest information gain. In the table, the meaning of the features are also described, because all the original tweets are written in Japanese. All these ten words or part of words indicate behavior. For instance, “o ha yo” in the top of TABLE III means the part of greeting word “o ha yo u”, which is a morning greeting words when we woke up. Almost all the extracted features in the table corresponds our daily life, such as waking up, going to sleep, going back home, eating and bathing. It should be noticed in detail about “na u” which corresponds to the English word “now”. We have Japanese words which represent ”now”, but using “na u” in tweets shows specific meaning of doing something. “go ha n na u” means “I am having breakfast or lunch now”. These kinds of omission or abbreviation is often used in tweets. Our proposed method based on n-gram approach can handle this situation.

Fig. 4 shows the average of 100 trials of precision, recall and F-measure of the 10 users when changes the numbers of training data from 100 to 4000. Fig. 4 shows that the evaluation result is increasing with increasing number of training data. The evaluation results under about 1000 are increasing rapidly. On the other hand, the improvements are gradual with over 1,000 training data. We consider that at least about 1000 tweets are needed for learning.

Suppose that the user is going to use this system from scratch, which means the user has no training data. Table I shows that 64.4 behavioral tweets are posted during 11 days in average. To collect 1,000 behavioral tweets, users needs to use this system more than 150 days or half a year. Although Fig 3 shows that performance is getting better and better during the usage, it should be concerned to use this system

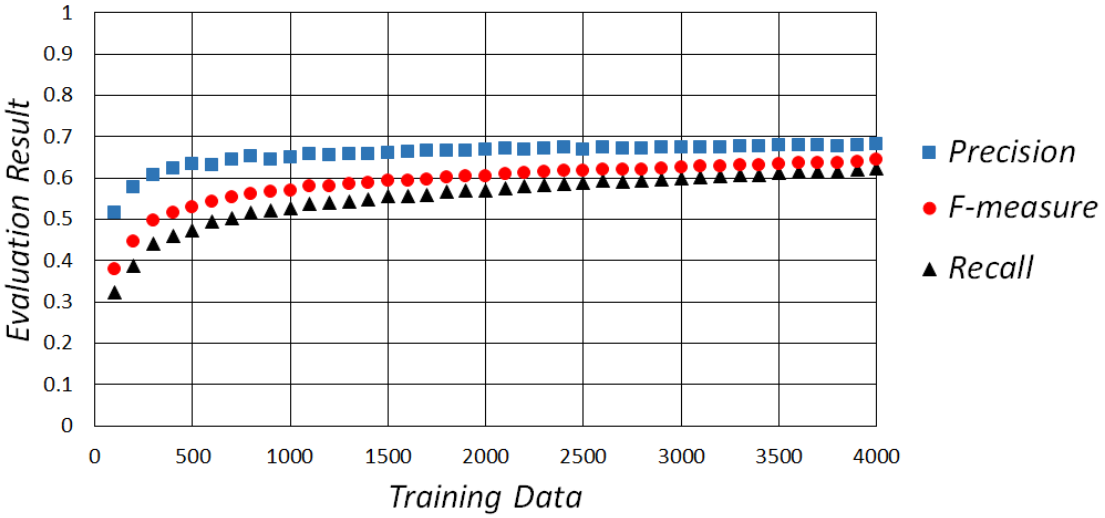

Fig. 4. Evaluation Result when Changes Number of Training Data

TABLE IV
F-MEASURE FOR EACH NUMBER OF TRAINING DATA

	All Data	Other's Data
user1	0.754	0.692
user2	0.655	0.515
user3	0.576	0.493
user4	0.513	0.491
user5	0.534	0.410
user6	0.508	0.311
user7	0.815	0.652
user8	0.638	0.598
user9	0.794	0.762
user10	0.731	0.541
average	0.652	0.547

with a high performance from the initial state. It is possible to use the tweets by others as training data. The feasibility of using others' tweets are examined in the next experiments.

B. Behavior Extraction using Other's Data

TABLE IV shows F-measure of behavior extraction by proposed method when using all tweets and other users' tweets as training data. In TABLE IV, F-measures with that of other users' tweets are less than all tweets as training data.

Fig. 5 plotted the difference that the result of previous experiment minus this result of the evaluation result for each user. As seen in Fig. 5, the data are distributed positive and negative domain for horizontal axis which represents the precision. With or without user's own data, precision is getting better and worse depending on the user. In average, precision is almost the same with these conditions. As for the vertical axes which represent recall, all the index degraded without user's own data. Misclassification depends on precision, when the users start to use this system, they are not bothered for correction of misclassification. This is a good characteristic considering continuous use of the system. This means the index on recall is getting better.

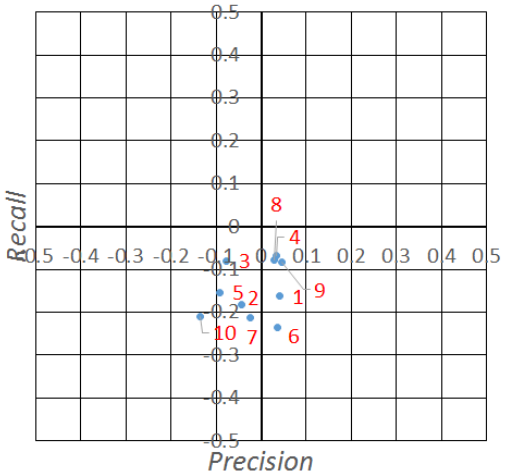

Fig. 5. Difference of Precision and Recall for each User Data

VI. CONCLUSIONS

We propose a novel system to extract behavioral information from tweets for context-awareness services. Words that indicate behavior are extracted from real Japanese tweets. The experiments are evaluated with precision, recall and F-measure. As a result, proposed method shows higher precision and recall than those of previous method. In addition, because proposed method does not need to prepare dictionary both in advance and during usage, it is able to reduce much burdens for humans. Experimental results of behavior extraction with or without the user's own tweets show the promising results using the system continuously. The performance indices are getting better while training data is increasing. Proposed method is able to handle not only common words but also Twitter-specific words as effective features in the classifier when the system extract behavioral

information using character n-gram tokenization. The words extracted in our experiments are quite trivial ones in this paper. It is considerable to use our method with morphological analysis. In the experiments, we assume that greeting words represent the behavior, “o ha yo”(good morning) for example. To ensure it is really behavioral information or not, other sensory data accompanied by the tweets may help us, such as location, time, body motion and so on. To assign the label that indicate concrete behavior for extracted tweets is planned for our future work. This will make us easy to use extracted behavioral information for various context-awareness services. In addition, we must investigate the relation between the number of training data and evaluation indices with the number of training data. This is important for estimating behavior of the system when the limited system in practical use.

REFERENCES

[1] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles, “Towards a Better Understanding of Context and Context-Awareness,” *Proceedings of the 1st International Symposium on Handheld and Ubiquitous Computing*, pp. 304-307, 1999.

[2] Ig-Jae Kim, Sang Chul Ahn, Heedong Ko, and Hyoung-Gon Kim, “Automatic Lifelog Media Annotation based on Heterogeneous Sensor Fusion,” *Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp. 703-708, 2008.

[3] Masanobu Abe, Daisuke Fujioka, and Hisashi Handa, “A Life Log Collecting System Supported by Smartphone to Model Higher-Level Human Behaviors,” *Proceedings of 6th International Conference on Complex, Intelligent, and Software Intensive Systems*, pp. 665-670, 2012.

[4] Masayuki Okamoto, Nayuko Watanabe, ShinichiNagano, and Kenta Cho, “Annotating Knowledge Work Lifelog: Term Extraction from Sensor and Operation History,” *Proceedings of 20th Conference on Information and Knowledge Management*, pp. 2581-2584, 2011.

[5] Ansheng Ge, Wenji Mao, Daniel Zeng, and Lei Wang, “Action Knowledge Extraction from Web Text,” *Proceedings of 11th International Conference on Intelligence and Security Informatics*, pp. 368-370, 2013.

[6] Kiyoharu Aizawa, Gamhewage C. de Silva, Makoto Ogawa, and Yohei Sato, “Food log by Snapping and Processing Images,” *Proceedings of 16th International Conference on Virtual Systems and Multimedia*, pp. 71-74, 2010.

[7] Nilanjan Banerjee, Dipanjan Chakraborty, Koustuv Dasgupta, Anupam Joshi, Sumit Mittal, Seema Nagar, Angshu Rai, and Sameer Madan, “User Interests in Social Media Sites: An Exploration with Micro-blogs,” *Proceedings of 18th International Conference on Information and Knowledge Management*, pp. 1823-1826, 2009.

[8] The-minh Nguyen, Takahiro Kawamura, Yasuyuki Tahara, and Akihiko OhsugaC “Self-Supervised Capturing of Users’ Activities from Weblogs,” *International Journal of Intelligent Information and Database Systems*Cvol. 6CNo. 1, pp. 61-76, 2012.

[9] Yuji Yano, Takeru Yokoi, and Tomonori Hashiyama, “Behavior Extraction from Behavioral Words on Twitter (in Japanese),” *Proceedings of 12th Forum on Information Technology*Cvol. 4, pp. 157-164, 2013.

[10] Andreas Stolcke, “Entropy-based Pruning of Backoff Language Models,” *Proceedings DARPA Broadcast News Transcription and Understanding Workshop*, pp. 270-274, 1998.

[11] Pedro Domingos, and Michael Pazzani, “On the Optimality of the Simple Bayesian Classifier under Zero-One Loss,” *Journal of Machine Learning*Cvol. 29, pp. 103-130, 1997.