Acquisition of Knowledge with Time Information from Twitter

Kohei Yamamoto
Department of Artificial Intelligence
Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502 Japan
Email: k_yamamoto@pluto.ai.kyutech.ac.jp

Kazutaka Shimada
Department of Artificial Intelligence
Kyushu Institute of Technology
680-4 Kawazu Iizuka Fukuoka 820-8502 Japan
Email: shimada@pluto.ai.kyutech.ac.jp

***Abstract*—In this paper, we propose a knowledge acquisition method for non-task-oriented dialogue systems. Such dialogue systems need a wide variety of knowledge for generating appropriate and sophisticated responses. However, constructing such knowledge is costly. To solve this problem, we focus on a relation about each tweet and the posted time. First, we extract event words, such as verbs, from tweets. Second, we generate frequency distribution for five different time divisions: e.g., a monthly basis. Then, we remove burst words on the basis of variance for obtaining refined distributions. We checked high ranked words in each time division. As a result, we obtained not only common sense things such as "sleep" in night but also interesting activities such as "recruit" in April and May (April is the beginning of the recruitment process for the new year in Japan.) and "raise the spirits/plow into" around 9 AM for inspiring oneself at the beginning of his/her work of the day. In addition, the knowledge that our method extracts probably contributes to not only dialogue systems but also text mining and behavior analysis of data on social media and so on.**

***Keywords*-Knowledge acquisition; Text mining; Hourly things; Daily things; Weekly things; Monthly things;**

I. Introduction

Recently, dialogue systems, especially non-task-oriented dialogue systems, have been increasingly important and popular, such as Rinna[1] (Microsoft) and Xiaoice[2] (Microsoft). In addition, task-oriented dialogue systems, such as QA systems, often contain a chatting function for increasing user satisfaction: e.g., Siri[3] (Apple) and Pepper[4] (Softbank). The success of neural networks or reinforcement learning approaches improves performance dramatically. However, these models tend to generate simple responses and are not sufficient to satisfy users. The main problem is caused by a lack of knowledge that such dialogue systems retain. When we talk to someone about something, we often utter not only direct responses and answers to what he/she talks but also implicit information that we share, namely a matter of common sense. To realize a human-like conversation by dialogue systems, we need to acquire much knowledge related to experiences and events in real life. Narisawa et al. [1] have proposed an automatic acquisition method about numerical common sense from the Web. Young et al. [2] have proposed a method for generating utterances by using common sense knowledge. Many researchers have studied knowledge acquisition methods from Web news [3], asynchronous dialogues from Ubuntu forums [4], [5] and Reddit [6], [7], and Wikipedia [8]. Some researchers have also handled Twitter as the resource for dialogue systems [9], [10], [11]. Such knowledge and approaches lead to the improvement of user satisfaction for dialogue systems.

In this paper, we propose a method for extracting knowledge about daily occurrences from social media. We focus on Twitter as the social media service. It is one of the most famous microblogging services and text-based posts of up to 140 characters. The posted sentences are described as "tweet". In microblogging services such as Twitter, users tend to post tweets in real-time. It denotes that tweets often contain significant information about daily occurrences, as lifelog data.

The purpose of our study in this paper is to automatically extract daily, weekly, and monthly activities with time information from Twitter. In other words, our purpose is to link an event and time information, e.g., "snowfall" appears in winter (around December to February in Japan) and "going to bed" frequently occurs in approx. 10 p.m. to 12 p.m. Here we explain the importance of the knowledge for dialogue systems with some examples.

- If a system retains knowledge, "eating lunch" = around 12 a.m., the system can generate the following utterance by using the current time.
  Ex1 (2/22 <u>12:30</u>)
  System: Have you eaten lunch?
  User: Yes, I have.
- If a system retains knowledge that "swimming" is an event in summer, the system can generate the following utterance by using the current month.
  Ex2 (<u>2</u>/22 15:00)
  User: I went swimming in the sea yesterday.
  System: Oh! It is uncommon to swim in February, isn't it?
- If a system retains knowledge about the peak of cherry blossoms, the system can generate the following utter-

[1]https://www.rinna.jp/
[2]https://www.msxiaobing.com/
[3]https://www.apple.com/siri/
[4]https://www.softbank.jp/robot/pepper/

ance by using the current date.

Ex3 (3/15 12:00)

User: Spring is coming soon.

System: I would like to go on a cherry viewing picnic.

These utterances can be realized by the knowledge about relations between an event and the most relevant time.

One simple approach to realize the system is to construct rules by handwork. However, the construction is costly and often possesses a problem of the coverage of rules. Moreover, Higashinaka et al. [12] have reported that manual modification and extension of the rules do not always tend to lead the improvement of the system's performance. On the other hand, the contribution of this paper is that our method links a word/phrase with time information automatically. In addition, the knowledge that our method extracts probably contributes to not only dialogue systems but also text mining and behavior analysis of data on social media and so on. Furthermore, our method does not depend on specific languages although we handle Japanese tweets in this paper because the method is just based on rankings by frequency distribution and burst detection. To the best of our knowledge, this is the first research about knowledge acquisition with time information from Twitter

II. Related work

As a study about handling time information for dialogue systems, Sato et al. [13] have proposed a neural chatbot system with knowledge about season type information. In other words, they proposed a method handling a variety of situations that affect the system outputs. However, they used only seasons, namely spring (Mar. - May.), summer (Jun. - Aug.), autumn (Sep. - Nov.), and winter (Dec. - Feb.) as the time information. On the other hand, we apply various types of time information, such as daily and monthly, to our knowledge acquisition task.

For knowledge acquisition, many researchers have reported their approaches [1], [14], [15], [16]. As mentioned above, Narisawa et al. [1] proposed an automatic acquisition method about numerical common sense from the Web. It must be useful knowledge for dialogue systems. Mitsuda et al. [14] have collected and clustered information that humans perceive from each utterance by handwork. Machida et al. [15] and Otani et al. [16] have proposed knowledge acquisition methods via dialogue systems. On the other hand, we use tweets on Twitter for knowledge acquisition.

The purpose of our study is to extract knowledge about hourly, daily, weekly, monthly occurrences. Ge et al. [17] have constructed a resource, EventWiki, from Wikipedia. Their target is major events, such as earthquakes and Olympic events. It is different from the target in our task. As studies focusing on relations between human activity and time information, Tandon et al. [18] have proposed an acquisition method of knowledge about activities from narratives, such as movies. Yao et al. [19] have proposed a method for acquiring rich temporal "before/after" event knowledge across sentences in narrative stories. In this paper, we propose an acquisition method for a wide range of knowledge from Twitter as compared with the related work.

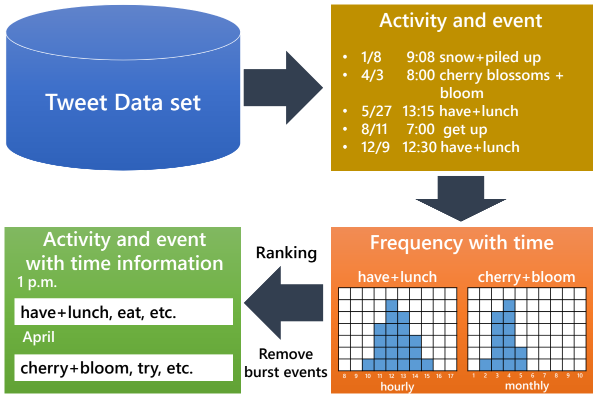

Figure 1. Overview of our method.

III. Method

In this section, we describe our acquisition method. Figure 1 shows the outline of the method. First, we extract activity and event word/phrases from a tweet data set (Section III-A.) Next, we generate the frequency distribution of each activity and event word/phrase on the basis of the timestamp of each tweet (Section III-B.) Then, we generate the rankings of each activity and event word/phrase (Section III-C.) Finally, we remove burst words by using a burst detection process (Section III-D.) As a result, we obtain activities and events that appear in a specific time range: e.g., "have lunch" at 1 p.m. in Figure 1.

A. ActEvn word extraction

Tweets tend to contain mention about users' real-time activities or events that users attend. In general, verbs have the most important role in such mention. In addition, the object of each verb also has an important role in recognizing the content of activities and events. Therefore, we extract verbs and verb-object pairs from tweets (hereinafter this is called "ActEvn words".)

We use a morphological analyzer, MeCab[5], for the ActEvn word extraction. Tweets tend to contain new words and proper nouns that are not included in normal dictionaries. Therefore, we also use the NEologd dictionary[6] to solve this problem.

B. Frequency distribution

Then, we generate frequency distribution tables of ActEvn words. In this paper, we focus on the appearance of ActEvn

[5]http://taku910.github.io/mecab/

[6]https://github.com/neologd/mecab-ipadic-neologd

Table I
TIME PERIODS.

Type	Unit for counting
Month	Monthly (Jan., Feb. ,..., Dec.)
Week	Weekday (Mon, to Fri.) and Weekend (Sat. and Sun.)
Day-of-Week	Each day of the week (Sun. to Sat.)
Day	Morning (3 to 8 a.m.), Daytime (9 a.m. to 5 p.m.), Night (6 p.m. to 2 a.m.)
Hour	Hourly (1 a.m., 2 a.m., ..., 12 p.m.)

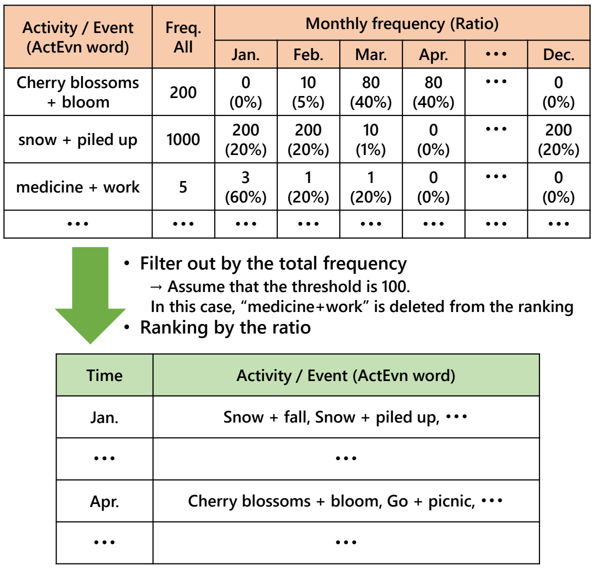

Figure 2. Ranking by the ratio that computed from the total frequency.

words in several time periods for the knowledge acquisition with temporal information. For the purpose, we need a suitable definition of time periods. We utilize the time unit defined by the Japan meteorological agency.

Table I shows the definition and the time ranges: units that our method sums up by. We handle five types of time ranges. We compute the frequency of each ActEvn word for each type of time range. The upper part of Figure 2 shows an example of frequency distribution tables about the Month type, namely a monthly distribution.

C. Ranking

Typical activities and events closely related to a specific time lead to a massive amount of posts, namely tweets, about the activities and events. For example, the number of tweets that contain the ActEvn word "have+lunch" increases around noon. In a similar way, tweets with "Cherry blossom+bloom" frequently occur in March and April in Japan. Therefore, we generate rankings about each time range type from the distribution tables obtained in Section III-B.

Figure 2 shows an example of a ranking process. The ranking process is based on the ratio of each time range. Our method sorts each ActEnv word in descending order as

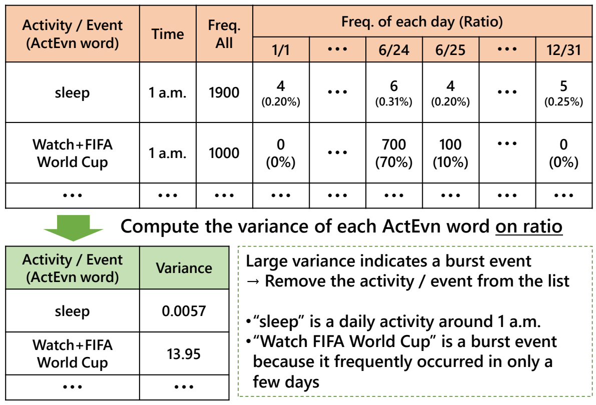

Figure 3. Deletion of burst activities and events by using the variance value.

the ranking of each type: Year, Week, Day-of-Week, Day, and Hour explained in Table I. In the process, we use a filter based on the total frequency of each ActEvn word because low-frequency words usually should be ignored for the real ActEnv word detection: e.g., "medicine+work" in Figure 2. For example, the threshold in Figure 2 is 100. The threshold value of the filter is determined experimentally.

D. Burst detection and removal

In Twitter, a phenomenon in which many users post tweets occurs at the same time. It is called "burst". In other words, the burst is a point in which the number of tweets suddenly increases. It usually occurs in out-of-ordinary events: e.g., FIFA World Cup and Olympic games. Such burst situations in Twitter should be ignored in our knowledge acquisition process because our purpose is to extract common activities of people and ordinary events in the world. Therefore, we need to remove ActEvn words related to the burst situation.

In this paper, we focus on distribution uniformity of each ActEvn word. If an ActEvn word is not a burst word, the distribution of the word frequency becomes uniform. In other words, the variance of ratios based on the frequency becomes small. On the other hand, if an ActEvn word is a burst word, that becomes non-uniform. In other words, the variance of ratios based on the frequency becomes large.

Figure 3 shows an example of burst and non-burst ActEvn words. The word "sleep" is an ActEvn word and a daily activity. Hence, the variance is sufficiently small. On the other hand, "Watch+FIFA World Cup" is not an ActEvn word because it is a temporary event at 1 a.m. However, monthly time range is excluded from the burst detection process due to a lack of enough yearly data because we need tweets of several years to compute the statistically unwarranted variance.

Table II
MONTH RANKING WITHOUT BURST DETECTION. THE THRESHOLD VALUES OF THE FREQUENCY IS 1000.

Month	1st	2nd	3rd	4th
Jan	積もる (snow cover)	雪+降る (snowfalls)	お願い+いたす (please)	お願い+致す (please)
Feb	積もる (snow cover)	雪+降る (snowfalls)	渡す (give)	受かる (pass)
Mar	卒業+する (graduate)	受かる (pass)	狙う (target)	染める (dye)
Apr	募集+する (recruit)	うつ (depression)	病む (worry about)	フォロー+する (follow)
May	黙る (be silent)	募集+する (recruit)	掘る (dig)	つづく (continue)
Jun	揺れる (shake)	勝つ (win)	追いつく (catch)	攻める (attack)
Jul	溶ける (melt)	浴びる (shower)	雨+降る (rain falls)	刺す (bite)
Aug	鳴る (sound/thunder)	掘る (dig)	雨+降る (rain falls)	刺す (bite)
Sep	発表+する (announce)	合わせる (fit)	刺す (bite)	晴れる (fine weather)
Oct	当選+する (get elected)	風邪+ひく (catch cold)	晴れる (fine weather)	風邪+引く (catch cold)
Nov	晒す (expose)	風邪+ひく (catch cold)	風邪+引く (catch cold)	冷える (get cold)
Dec	雪+降る (snowfalls)	掃除+する (clean up)	実家+帰る (homecoming)	迎える (to start the new year)

IV. DISCUSSION

We analyzed the outputs from our method. First, we discuss the longest and shortest time ranges in our setting: monthly and hourly. Table II shows the top 4th about monthly ranking. We can see intuitive seasonal events from the table, such as “snowfalls” in the winter season. In addition, we obtained many interesting ActEvn words for the monthly ranking. In Japan, February and March are an entrance exam period for high schools and universities. As a result, the ActEvn word “pass” frequently appeared in these months. April is the beginning of a new fiscal year and the beginning of the recruitment process for the next year in Japan. Therefore the word “recruit” became the 1st rank in April. The words “depression” and “worry about” probably express the sentiments and feelings of new employees. The word “bite” indicates that July, August, and September are mosquito-infested months. The words “clean up”, “homecoming”, and “to start the new year” in December express the Japanese culture: the whole house cleaning and staying in parents’ home at the end of the year. On the other hand, some unsuitable ActEvn words appeared on the list. The appearance of “get elected” in October as the 1st ActEvn word is caused by the national election that was held in this month of the year that we collected tweets from Twitter. It is not a habitual activity that we want to acquire. One reason is that we were not able to apply the burst detection process, namely removal by variance, due to a lack of yearly tweets. To improve the correctness and validity of the ranking list, we need to capture tweets with a long-term plan.

Table III shows the top 4th about hourly ranking with burst detection. The threshold values of frequency and burst are 100 and 1, respectively. We can observe typical hourly actions from the table, such as “sleep” during midnight time, “eating lunch” around noon, and “eating dinner” in the evening. In other words, we obtained a good ActEvn word list that related to common-sense. In addition, we also obtained some interesting ActEven words. One example is the 1st ActEvn word at 9 AM, “raise the spirits/plow into”. It probably indicates that he/she posts the tweet for inspiring oneself at the beginning of his/her work of the day. The word “do the best today” at 7 and 8 AM also contains a similar meaning.

Here we imagine an application of the knowledge that was acquired by our method in a similar way to Section I, namely a chat dialogue system with the knowledge.

- If the system retains knowledge, “bite:mosquito-infested” = Jul. to Sep., the system can generate “Be careful about mosquitoes when you go out.”
- If a system retains knowledge that “raise the spirits/plow into” is around 9 AM and the current time is 8 to 9 AM, the system can generate “Plow into your work! Do your best and good luck.”

Thus, we might well be able to develop an extremely expressive dialogue system in the future.

Table IV, Table V, and Table VI show the results of the weekly ranking, the day-of-week ranking, and the daily ranking, respectively. We obtained similar tendencies between the week ranking and the day-of-week ranking. For example, “go to school/hospital/work” on weekdays and “go out” on weekends. The result of the daily ranking was also similar to the hour ranking shown in Table III.

Here we focus the effectiveness of the burst detection and the removal. Table VII shows a part of a result by our method without the burst detection for the hourly ranking. The table contained many noise words as compared with that using the burst detection process (see Table III.) In addition, the method without the bust detection acquired non-habitual activities, such as “watch fireworks” at 8 PM as the 4th. These results show the effectiveness of burst detection.

Words in each ranking often accompany auxiliary verbs about report and conjecture markers, such as “らしい (It is said that)” and “みたい (It seems that).” Some researchers have studied fact analysis methods [20], [21]. Incorporating the methods into our method is interesting future work.

In this paper, we evaluated our result qualitatively. To provide a quantitative evaluation by test subjects is the most important future work.

V. CONCLUSIONS

In this paper, we proposed a method for extracting knowledge about daily occurrences from social media. We focused on Twitter as the social media service. Users in Twitter tend to post tweets in real-time. We handled tweets as lifelog data by using this characteristic. First, we extracted event words (ActEvn words) from tweets. Then, we generated

Table III

HOUR RANKING WITH BURST DETECTION. THE THRESHOLD VALUES OF THE FREQUENCY AND BURST ARE 100 AND 1, RESPECTIVELY.

Hour	1st	2nd	3rd	4th
0 AM	お誕生+ござる (be birthday)	夢+見れる (be able to have a dream)	明日+起きる (wake up tomorrow)	5 時+起きる (wake up at five)
1 AM	ねむれる (be able to sleep)	仕事+寝る (sleep in working)	ねれる (be able to sleep)	ため+寝る (sleep for)
2 AM	これ+寝る (this sleep)	の+眠れる (be able to sleep)	ねれる (be able to sleep)	目+冴える (be wakeful)
3 AM	目+冴える (be wakeful)	ねれる (be able to sleep)	時間+起きる (wake up at)	目+さめる (wake up)
4 AM	時間+起きる (wake up at)	就寝+する (go to bed)	2 時間+寝る (sleep two hours)	の+眠れる (be able to sleep)
5 AM	表す (express)	お過ごし+くださる (stay)	指す (point)	4 時+起きる (wake up at 4)
6 AM	使用+する (use)	予測+する (predict)	お過ごし+くださる (stay)	結局+寝る (finally sleep)
7 AM	日+がんばる (do the best today)	今日+頑張る (do the best today)	朝+迎える (in the morning)	今週+頑張る (do the best this week)
8 AM	日+がんばる (do the best today)	今週+頑張る (do the best this week)	つづく (continue)	遅延+する (delay)
9 AM	はりきる (raise the spirits/ plow into)	遅延+する (delay)	病院+来る (go to hospital)	元気+過ごす (keep well)
10 AM	瞬+殺る (in a flash)	洗濯+終わる (finish washing)	電話+くる (get a phone call)	朝ごはん+食べる (have breakfast)
11 AM	病院+来る (go to hospital)	昨日+行う (do yesterday)	マック+食べる (eat McDonald's)	昼ごはん+食べる (have lunch)
12 PM	昼ご飯+食べる (have lunch)	弁当+食べる (have lunch)	飯+食える (eat lunch)	瞬+殺る (in a flash)
1 PM	昼ごはん+食べる (have lunch)	昼飯+食う (have lunch)	昼ご飯+食べる (have lunch)	お昼+食べる (have lunch)
2 PM	昼ごはん+食べる (have lunch)	用事+済ませる (finish a job)	先+ある (be ahead)	買い+来る (go buy)
3 PM	モード+なる (go into work mode)	仕事+戻る (go back to work)	昼飯+食べる (have lunch)	バイト+休む (get off a part-time job)
4 PM	バイト+行く (go to a part-time job)	昼+食べる (have lunch)	夕方+なる (in the evening)	電車+座る (get a seat on a train)
5 PM	夕方+なる (in the evening)	定時+帰れる (leave work on time)	冊+買う (buy books)	駆る (punch)
6 PM	選択+する (select)	夕飯+作る (make dinner)	夕飯+食べる (have dinner)	定時+上がる (leave work on time)
7 PM	閉じる (close)	ご飯+炊く (cooking rice)	自炊+する (cook own meal)	お仕事+終わる (finish the work)
8 PM	与える (give)	晩御飯+食べる (have dinner)	チケット+届く (get a ticket)	点+入る (get a score)
9 PM	明日+届く (get tomorrow)	収束+する (settle down)	明日+楽しむ (enjoy tomorrow)	ドラム+叩く (beat a drum)
10 PM	ふむ (step)	心+響く (touch a heart)	湯船+浸かる (get in a bath)	いらっしゃる (coming)
11 PM	明日+寝る (sleep for tomorrow)	人+飲む (dirnk with)	どっか+見る (watch something)	時間+食べる (eat late at night)

Table IV

WEEK RANKING (WEEKDAY AND WEEKEND) WITH BURST DETECTION. THE THRESHOLD VALUES OF FREQUENCY AND BURST ARE 1000 AND 1, RESPECTIVELY.

Week	1st	2nd	3rd	4th
weekday	学校+行く (go to school)	サボる (wag off)	病院+行く (go to hospital)	今日+頑張る (do my best today)
weekend	出かける (go out)	並ぶ (get in line)	参加+する (join)	向かう (go to)

frequency distributions for five different time divisions, e.g., a monthly basis. We introduced burst word detection on the basis of variance for obtaining refined distributions. As a result, we obtained not only common sense things such as in night but also interesting activities such as "raise the spirits/plow into" around 9 AM for inspiring oneself at the beginning of his/her work of the day. Although we handled Japanese tweets in this paper, our method does not depend on specific languages. This is also the effectiveness of our method. The result often contained noise words in

Table V

DAY-OF-WEEK RANKING WITH BURST DETECTION. THE THRESHOLD VALUES OF FREQUENCY AND BURST ARE 1000 AND 1, RESPECTIVELY.

DoW	1st	2nd	3rd	4th
Mon	学校+行く (go to school)	書ける (write)	揺れる (swing)	今日+寝る (sleep today)
Tue	使用+する (use)	掛かる (hang)	つづく (continue)	黙る (shut)
Wed	つづく (continue)	晒す (expose)	曇る (cloudy)	つぶやく (tweet)
Thu	相談+する (have a talk)	腹立つ (be angry)	サボる (wag off)	今日+寝る (sleep today)
Fri	いらっしゃる (come)	使用+する (use)	今日+頑張る (do my best today)	当選+する (get elected)
Sat	出かける (go out)	並ぶ (get in line)	売り切れる (sold out)	呑む (drink)
Sun	充実+する (fulfilling)	浸る (soak)	参加+する (join)	出掛 ける (go out)

Table VI
DAY RANKING (MORNING, DAYTIME, NIGHT) WITH BURST DETECTION. THE THRESHOLD VALUES OF FREQUENCY AND BURST ARE 1000 AND 1, RESPECTIVELY.

Day	1st	2nd	3rd	4th
3 to 8	今日+頑張る (do my best today)	寝坊+する (oversleep)	仕事+行く (go to work)	早起き+する (early rising)
9 to 17	昼寝+する (napping)	売り切れる (sold out)	出かける (go out)	混む (jam-up)
18 to 2	今日+寝る (sleep)	風呂+入る (take a bath)	眠れる (sleep)	酒+飲む (drink)

Table VII
HOUR-RANKING WITHOUT BURST DETECTION. THE THRESHOLD VALUES OF FREQUENCY IS 100.

Hour	1st	2nd	3rd	4th
4 PM	%+見る (watch)	白猫+遊ぶ (play a game)	それ+合わせる (fit)	抽選+当たる (win a lottery)
8 PM	今日+閉じる (close)	逢う (meet)	声+変わる (voice-change)	花火+見る (watch fireworks)

the list, namely non-habitual activities. The improvement of the method using another burst detection approach and fact analysis is important future work.

The current method only focused on five-time ranges. We need to combine the time ranges, such as 24 hours in a month, for extracting much rich knowledge. In addition, user attributions are interesting features for the knowledge acquisition process: e.g., the difference between 24 hours of young men and old men. It is our important future work. Applying our extracted knowledge to dialogue systems and text mining systems is also our future work.

REFERENCES

[1] K. Narisawa, Y. Watanabe, J. Mizuno, N. Okazaki, and K. Inui, "Is a 204 cm man tall or small? acquisition of numerical common sense from the web," in *Proceedings of ACL*, vol. 1, 2013, pp. 382–391.

[2] T. Young, E. Cambria, I. Chaturvedi, H. Zhou, S. Biswas, and M. Huang, "Augmenting end-to-end dialogue systems with commonsense knowledge," in *AAAI Conference on Artificial Intelligence*, 2018.

[3] K. Yoshino and T. Kawahara, "Conversational system for information navigation based on pomdp with user focus tracking," *Computer Speech & Language*, vol. 34, no. 1, pp. 275–291, 2015.

[4] R. Lowe, N. Pow, I. Serban, and J. Pineau, "The ubuntu dialogue corpus: A large dataset for research in unstructured multi-turn dialogue systems," in *Annual Meeting of SIGDIAL*, 2015, pp. 285–294.

[5] R. T. Lowe, N. Pow, I. V. Serban, L. Charlin, C.-W. Liu, and J. Pineau, "Training end-to-end dialogue systems with the ubuntu dialogue corpus," *Dialogue & Discourse*, vol. 8, no. 1, pp. 31–65, 2017.

[6] P.-E. Mazare, S. Humeau, M. Raison, and A. Bordes, "Training millions of personalized dialogue agents," in *Proceedings of EMNLP*, 2018, pp. 2775–2779.

[7] A. C. Curry, I. Papaioannou, A. Suglia, S. Agarwal, I. Shalyminov, X. Xu, A. Eshghi, I. Konstas, V. Rieser *et al.*, "Alana v2: Entertaining and informative open-domain social dialogue using ontologies and entity linking," *Proc. Alexa Prize*, 2018.

[8] K. Niina and K. Shimada, "Trivia score and ranking estimation using support vector regression and ranknet," *Proceedings of PACLIC*, 2018.

[9] J. Li, M. Galley, C. Brockett, G. Spithourakis, J. Gao, and B. Dolan, "A persona-based neural conversation model," in *Annual Meeting of the ACL*, vol. 1, 2016, pp. 994–1003.

[10] Y. Luan, C. Brockett, B. Dolan, J. Gao, and M. Galley, "Multi-task learning for speaker-role adaptation in neural conversation models," in *Eighth International Joint Conference on Natural Language Processing*, vol. 1, 2017, pp. 605–614.

[11] I. V. Serban, A. Sordoni, R. Lowe, L. Charlin, J. Pineau, A. Courville, and Y. Bengio, "A hierarchical latent variable encoder-decoder model for generating dialogues," in *AAAI Conference on Artificial Intelligence*, 2017.

[12] R. Higashinaka, T. Meguro, H. Sugiyama, T. Makino, and Y. Matsuo, "On the difficulty of improving hand-crafted rules in chat-oriented dialogue systems," in *APSIPA*. IEEE, 2015, pp. 1014–1018.

[13] S. Sato, N. Yoshinaga, M. Toyoda, and M. Kitsuregawa, "Modeling situations in neural chat bots," in *Proceedings of ACL Student Research Workshop*, 2017, pp. 120–127.

[14] K. Mitsuda, R. Higashinaka, and Y. Matsuo, "What information should a dialogue system understand?: Collection and analysis of perceived information in chat-oriented dialogue," in *Advanced Social Interaction with Agents*. Springer, 2019, pp. 27–36.

[15] Y. Machida, D. Kawahara, S. Kurohashi, and M. Sassano, "Design of word association games using dialog systems for acquisition of word association knowledge," in *Proceedings of AKBC*, 2016, pp. 86–91.

[16] N. Otani, D. Kawahara, S. Kurohashi, N. Kaji, and M. Sassano, "Large-scale acquisition of commonsense knowledge via a quiz game on a dialogue system," in *Proceedings of OKBQA*, 2016, pp. 11–20.

[17] T. Ge, L. Cui, B. Chang, Z. Sui, F. Wei, and M. Zhou, "Eventwiki: a knowledge base of major events," in *Proceedings of LREC*, 2018.

[18] N. Tandon, G. De Melo, A. De, and G. Weikum, "Knowlywood: Mining activity knowledge from hollywood narratives," in *Proceedings of ACM CIKM*, 2015, pp. 223–232.

[19] W. Yao and R. Huang, "Temporal event knowledge acquisition via identifying narratives," in *Proceedings of ACL*, 2018, pp. 537–547.

[20] R. Sauri and J. Pustejovsky, "Factbank: a corpus annotated with event factuality," *Language resources and evaluation*, vol. 43, no. 3, p. 227, 2009.

[21] M.-C. De Marneffe, C. D. Manning, and C. Potts, "Did it happen? the pragmatic complexity of veridicality assessment," *Computational linguistics*, vol. 38, no. 2, pp. 301–333, 2012.