International Journal of Smart Computing and Artificial Intelligence
International Institute of Applied Informatics
2022, Vol. 6, No. 2, IJSCAI700

Estimating Passive Twitter-User's Interests from Followed Users' Tweets by Technique Transfer

Tessai Hayama *, Qi Zhang *

Abstract

User modeling based on the contents of social network services has been developed to recommend information related to the preference of each user. Most of the previous studies have analyzed active users' tweets and estimated their interests. Meanwhile, although there are more than a certain number of passive users, who do not tweet but only gather information, little research has been conducted on interest estimation for them due to the lack of clues for estimating their interests. In this study, we developed an interest estimation method for passive Twitter users from the tweets of followed users by applying an interest topic extraction method for active users. In our evaluation, we confirmed the effectiveness of the proposed method by comparison with simple topic extraction methods based on data with interest topic evaluation of 12 users.

Keywords: user interests, SNS data analysis, topic extraction, user modeling

1 Introduction

Social network services (SNS) have been rapidly gaining popularity since their emergence in 2000. Currently, many users browse and collect various types of information (medical information, news, etc.) on SNS such as Twitter [1]. The behavior of users on SNS is influenced by their interests. As such, it is likely to play a paramount role in providing recommendations for content and information that match the users' interests. Thus, many studies have been conducted to estimate the interest profile of users from the contents posted by the target users on Twitter. Most of these studies have focused on user modeling of active users who actively post content on Twitter. However, the number of passive users on SNS who only browse and collect information without posting content has been increasing, accounting for a large proportion of users - 44% of users who have never posted a tweet [2].

Several studies have been conducted to estimate the interests of passive Twitter users. For example, Besel et al. [3] developed a method to extract the users' interests by linking the names of users they follow to Wikipedia entities and applying the category structure of Wikipedia entities. However, although entity-based methods can improve the accuracy of extracting user's in-

* Nagaoka University of Technology, Niigata, Japan

terests from the names of users included in Wikipedia entities, the method can only handle Twitter accounts of celebrities, and thus cannot use the information of most followed users. Piao et al. [2] developed a method to extract a user's interests using the biography of Twitter accounts the user follows. The biography of each user on Twitter contains direct descriptions of occupations and interests with a 160-character limit, which can be useful cues for creating interest profiles of each user. However, it is difficult to extract phrases that indicate the interests of each user from the biography of the followed users due to the lack of updates and few descriptions. Meanwhile, it is possible to obtain the interests of passive Twitter users using the tweets of followed users. Generally, tweets are frequently updated and contain the interests of the user who posted the tweets. However, for passive users, there are few clues to identify the users' interests from the tweets that the users posted; it is difficult to extract the appropriate interests of passive users because of the numerous topics, including tweets of followed users other than interests of the user.

Therefore, in this study, we developed a method for estimating the interests of a passive Twitter-user from the tweets of their followed users by technique transfer. To develop the method, we first developed an interest topic extraction for an active user from the tweets of his/her followed users, which supports extracting common topics of some followed users, setting the appropriate number of topics to be extracted for each user and removing tweets irrelevant to interest topics. Then, the developed method for an active user is applied to the interest topic estimation of a passive user. In the evaluation experiment, the effectiveness of the interest topic extraction method for active users by comparison with simple topic extraction methods based on the dataset was first confirmed. Next, by comparing with simple topic extraction methods based on survey data on 12 users' interest topic ratings, the effectiveness of the developed method for passive users was confirmed.

2 Previous Studies

Most of the studies on estimating the interest profile of active Twitter-users who actively tweet have analyzed tweets posted by users [4]. For example, Siehndel et al. [5] developed a profile generation method of user's interest based on Wikipedia entities from tweets that the user posted. Kapanipathi et al. [6] generated a weighted hierarchical interest graph of a user using Wikipedia entities and their categories and extracted categories containing many Wikipedia entities that appear in the user's tweets as activation nodes. Piao et al. [7] and Orlandi et al. [8] used DBpedia instead of Wikipedia categories to generate profiles of user's interests. Piao et al. [9] used structurally different categories and related entities in DBpedia to enhance the quality of user modeling for web-page recommendation in Twitter. These methods cannot be applied directly to our targeted passive users since they do not tweet.

Moreover, some researchers have studied methods that employ follow-relationships for active users. For example, Faralli et al. [10] developed a method for estimating profiles of users' interests for user recommendation by using their followed users' names linked to Wikipedia entities. They showed that this method could construct more accurate user interest profiles than analyzing tweets of followed users by utilizing the followed users' profiles. However, only 12.7% on average of the followed users could be linked to Wikipedia entities. For passive users' interest estimation, although follow relationship can be employed, it is difficult to use only the names of their followed users.

The interest estimation methods for Twitter passive users have also employed the following relationship in the same way as active users. For example, Piao et al. [11] used the entities in the biographies of the users they follow, and Piao et al. [12] used the entities in the titles and descriptions of the lists containing famous users they follow. The previous estimation methods of interests of Twitter passive users have been developed based on the assumption that passive users have the same behavior tendencies except posting as active users. However, the differences in features on SNS between the passive user and active user have yet to be clarified. Hayama [13] have analyzed the Twitter data with the user features used in the previous studies by using statistical methods to clarify the clue for extracting the interest of the passive user. The results showed that there are some passive user types with many celebrity followers or celebrities and it is appropriate to develop an interest estimation method for each such passive type. This study introduces a method for estimating the interests of different user types.

The relationships between passive SNS browsing and mental health problems in university students and adults have been studied in psychology [14, 15]. Passive users were positively correlated with online social comparison and fear of missing out (FoMO), implying that they were linked to depressive symptoms and self-perception. However, the studies conducted by these researchers were primarily based on psychological questionnaire surveys and did not address the characteristics of passive user behavior.

In this study, we analyze the tweets of followed users of passive users and extract their interest topics to improve the quality of the user interest profile. This technique is distinct from those developed in previous studies.

3 Approach

In this section, we show the settings, considerations, and design policies for developing a method to estimate the interests of a passive user based on their followed users' tweets.

3.1 Settings

The topic model of this study used a set of words with the probability of occurrence (Bag-of-Words) to represent a list of interest topics as user modeling. Thus, the tweets of all followed users of the target user are input, and a list of word sets representing the topics that the target user is interested in is output. The procedure is as follows:

Step 1: Input Twitter ID of a target user,
Step 2: Collect Twitter IDs of the target user's followed users and their last three-month tweets,
Step 3: Perform topic analysis of the collected tweets and extract interest topics from the collected topics,
Step 4: Output a list of the interest topics for the target user.

In Step 3, the target user's followed users' tweets are analyzed, and a list of topics is generated as the result. The list of topics generally includes topics irrelevant to the target user, so it is necessary to extract only the topics relevant to the target user. However, for passive users, since there are few clues to estimate their interests, it is difficult to extract the interest topics of the users accurately. In this study, a method for extracting interest topics from the tweets of active users is first developed, and then a method for estimating the interest topics of passive users by employing the interest topic extraction method of active users is developed.

3.2 Topic Extraction Method for Active Users

In this section, how the interest topics of active users can be extracted with high accuracy from the topics included in the tweets of users who the active users follow is considered.

More frequent topics that are included in tweets of the users who target users follow are probably of interest to the target user. Therefore, it is a useful method to analyze the tweets of users that the active users follow and then select more frequent topics to extract interest topics of active users. However, it is difficult to identify topics that are community-specific or special with simply the frequency of topic occurrence since such topics may be included only in the tweets of some followed users.

It is necessary to adjust the number of topics extracted from the tweets of the followed users based on the interests of the target user. The more tweets the target user posts, the more topics the tweets include. Meanwhile, if the number of tweets of the followed users is large, more topics are extracted other than the target user interest topic. Thus, it is inappropriate to set the same number of extracted interest topics for all target users.

In addition, tweets containing many decorative characters or indirectly containing a topic are disruptive to topic analysis. Meaningless sentences to decorate tweets and sentences not specific to a topic, such as ASCII art and simple response, interfere with topic analysis, which handles the co-occurrence of words, and irrelevant words are probably included in the topic models. Thus, such tweets are unnecessary for the interest topic extraction.

3.3 Design Policies

The followings are design policies to solve the issues described in Section 3.2.

- Enable to extract frequent topics from tweets of followed users based on the number of the users' tweets.
- Enable to set the appropriate number of topics to be extracted, depending on the amount of data for topic analysis.
- Exclude tweets with a small amount of content from topic analysis.

4 Implementation of Interest Estimation Method of Passive Users

Based on the design policies described in Section 3.3, we implemented a method to estimate the interest topics of a passive user from the tweets of his/her followed users. Figure 1 shows the procedure of the proposed method.

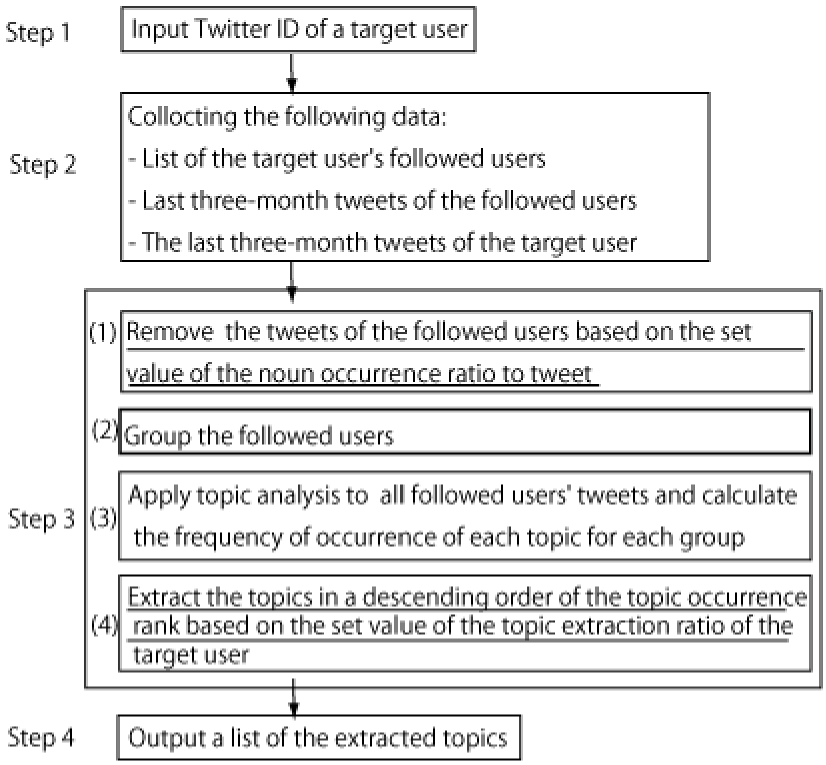

Figure 1: Procedure of the proposed method for estimating the interest topics of a passive user. The underlined parts are the processing modules that employ the set value of the interest topic extraction of an active user.

In Step 2, using Twitter ID input in Step 1 and the Twitter API†, the data of "the list of the target user's followed users," "last three-month tweets of the followed users," and "the last three-month tweets of the target user" are collected from the Twitter server and registered in the database of the local server. In Step 3, the following processes are performed: (1) remove the tweets of the followed users based on the set value of the noun occurrence ratio to tweet, (2) group the followed users, (3) apply topic analysis to all followed users' tweets and calculate the frequency of occurrence of each topic for each group, and (4) extract the topics in a descending order of the topic occurrence rank based on the set value of the topic extraction ratio of the target user. In Step 4, a list of the extracted topics is outputted.

The proposed method is applied with g means [16], which is an unsupervised classification method that automatically determines the number of groups based on the number of followed users and average number of tweets. For topic analysis to tweets, HDP-LDA [17] is used for topic modeling of the proposed method. The HDP-LDA is an unsupervised learning algorithm that is LDA [18] to automatically estimate the number of topics by the hierarchical Dirichlet process (HDP).

In Step 3, we employ the analysis of the dataset of active users (described in Section V) to determine the optimal values of the topic extraction ratio and noun occurrence ratio in Step 3-(1) and (4). In Step 3-(1), tweets with the number of nouns less than the set value are removed. To

† https://developer.twitter.com/ja/docs

decide the set value, the accuracy of extracting interest topics by varying the ratio of noun occurrence to tweet is investigated, and the ratio of nouns with the highest accuracy is applied to the proposed method. For the topic extraction ratio of a target user in Step 3-(4), a different optimal set value is assigned to each target user. To decide the set value, we group the target users using g means with the number of followed users of each target user, and the set value of the topic extraction ratio with the highest accuracy is used for each group in the topic extraction.

The features of the proposed method are summarized as follows.

- The appropriate number of extracted topics for each target user is determined by assigning the optimal topic extraction ratio according to the number of tweets of the followed users.
- Selecting the topics with high-frequency occurrence for each group of the followed users with the same number of tweets, the frequent topics among tweets of the followed users who have a small number of tweets are probably extracted.
- An appropriate topic model is constructed, excluding the co-occurrence of words irrelevant to the interest topic, by eliminating tweets that contain many words other than nouns from topic analysis.
- By determining the optimal set values of the proposed method by the datasets, the accuracy of interest topic extraction of active users is improved.
- It is possible to apply the interest topic extraction of an active user to interest topic estimation of a passive user because of not directly using information about a target user.

5 Dataset of Active Users

To determine the set values for the proposed method, a Twitter dataset of active users was created by avoiding bias in users' interests and behaviors. The dataset includes the Twitter ID of a user, the user's tweets, and the number of followers and followings of the user, the Twitter IDs of the followed users, and the tweets of the followed users.

To create this dataset, Twitter IDs of about 1.5 million users from the followers of a news site account (@YahooNewsTopics) were first collected using the Twitter API, and then 2000 users were randomly selected. From these users, 541 users who restrict their access permissions, 54 users who have not tweeted in the last three months, and 154 users whose followed users' tweets were not sufficiently collected were excluded, thus 1251 users were selected. Then, for the selected 1251 users, tweets in the last three months, the number of the followings, Twitter IDs of the followed users, and tweets of the followed users in the last three months were collected. Table 1 shows the number of target users, mean and standard deviation of the number of tweets by the target users, mean and standard deviation of the number of followers, and mean and standard deviation of the number of followings in the dataset. Figures 2 and 3 shows histograms of the followers of the users and tweets of the users, respectively.

Table 1: Twitter dataset created from 1251 active users

	Mean	S.D.
# of tweets	724.33	753.75
# of followers	265.56	316.14
# of followed users	392.86	357.74
# of tweets of followed users	181.89	77.41

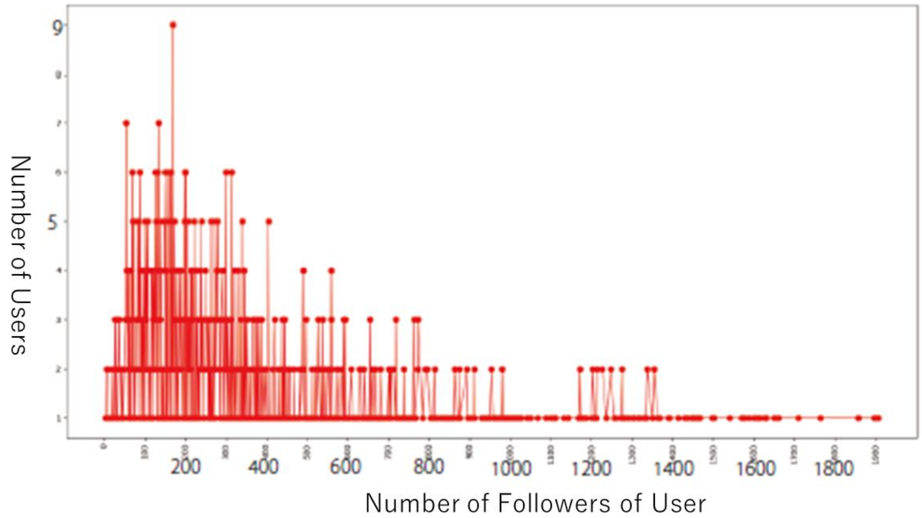

Figure 2: Histogram of the number of followers of the users

6 Evaluation

6.1 Overview

To confirm the effectiveness of the proposed method for estimating the interest topics of a passive user, we created the evaluation data of topics with interest ratings and compared the proposed method with simple topic extraction methods using the evaluation data.

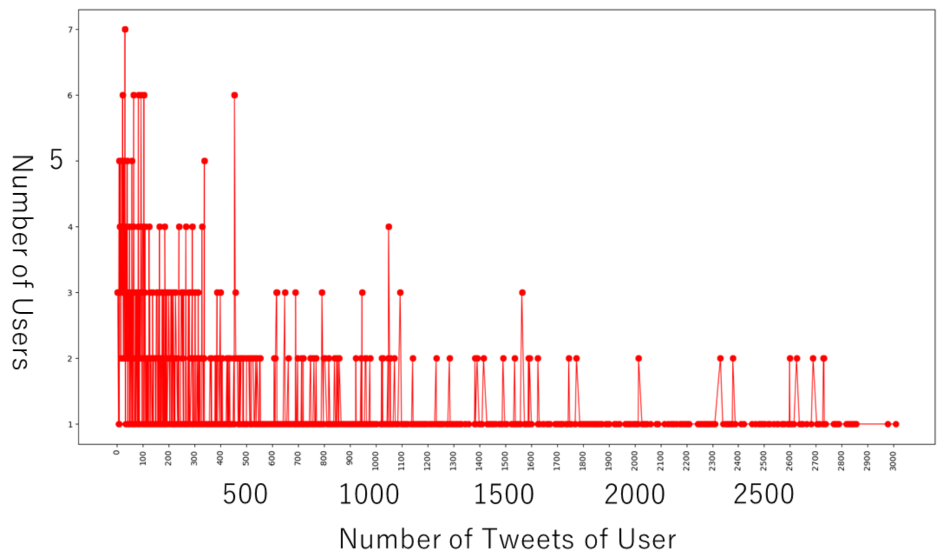

Figure 3: Histogram of the number of tweets of the users

To create the evaluation data of topics with interest ratings, 12 Twitter users were recruited, and a topic list for each user was extracted from the results of the topic analysis by inputting their Twitter IDs. Then, for the topic list of each user, systematic sampling, which extracted 20 topics in 20 segmented positions from the topic list, was conducted, and an interest survey questionnaire, with a word cloud and a 5-Lickard scale for each topic, was designed like Figure 4. Each user answered the questionnaire generated by inputting their Twitter ID, and the topics with high-interest ratings were as the correct items of the evaluation data. Table 2 shows statistics of the Twitter data of the users who participated in this experiment.

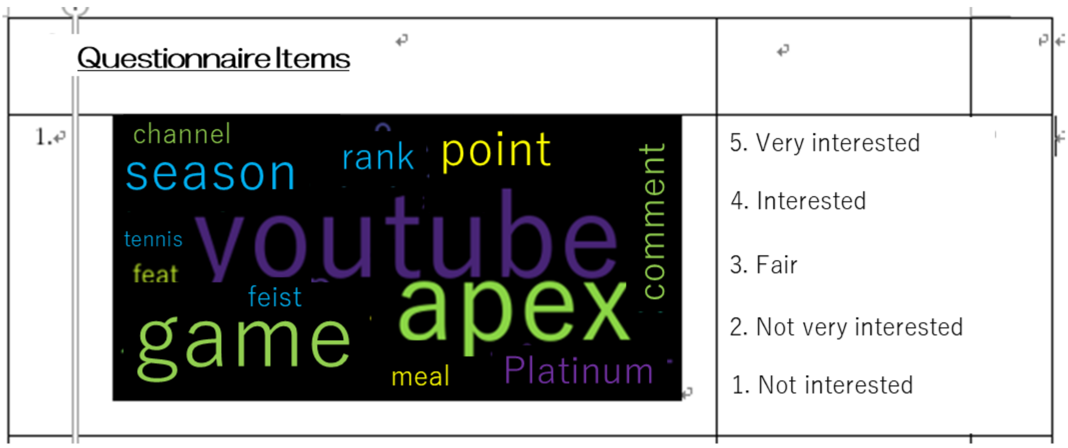

Figure 4: Example of Questionnaire Used in User Interest Survey.

Table 2: Statistics of Twitter data of the 12 users who participated in the experiment.

Topic extraction ratio				Ave. # of followed users				Ave. # of tweets of followed users			
Mean	S.D.	Max.	Min.	Mean	S.D.	Max.	Min.	Mean	S.D.	Max.	Min.
0.47	0.10	0.95	0.20	183.58	38.91	491.00	13.00	232.60	24.46	420.45	77.21

The precision rate, recall rate, and F-value are calculated using equations (1)-(3) to determine the set value of the proposed method and evaluate the proposed and comparison methods.

$$Precision = \frac{Num(Ts(user) \cap ExTs(followees))}{Num(Ts(user))} \quad (1)$$

$$Recall = \frac{Num(Ts(user) \cap ExTs(followees))}{Num(ExTs(followees))} \quad (2)$$

$$F - value = \frac{2 * Presicion * Recall}{Precision + Recall} \quad (3)$$

where *user*, *followees*, *Ts(A)*, *ExTs(A)*, and *Num(Ts(A))* represent a target user, followed users, topics included in the tweets posted by *A*, topics extracted by the proposed/comparison method from the tweets posted by *A*, and the number of topics included in the tweets posted by *A*, respectively. The precision rate indicates the ratio of the number of topics included in the tweets of a target user to the number of topics extracted by the proposed/comparison method. The recall rate indicates the ratio of the number of correct topics extracted by the proposed/comparison method to the number of topics extracted by the proposed/comparison method. The F-value indicates an evaluation value that considers both the precision and recall rates.

In the comparison method, the topics extracted by HDP-LDA [17] from the tweets of followed users are sorted in descending order of frequency, and some of them are selected by topic extraction ratio. In the experiment, two kinds of topic extraction ratios were used: the topic extraction ratio of 0.3 with the highest F-value and topic extraction ratio of 1.0.

6.2 Two kinds of set values of the proposed method

“topic extraction ratio for each target user" and “noun occurrence ratio for tweet removal" were determined with the accuracies of the proposed method by using the dataset described in Section 5. The results of the two kinds of set values are described in Sections 6.2.1 and 6.2.2, respectively. In addition, the effectiveness of the proposed method for active users by utilizing the optimal set values is described in Section 6.2.3.

6.2.1 Topic extraction ratio for each target user

The 1251 users in the dataset were categorized into 53 groups. The topic extraction ratios of the 53 groups were investigated with the accuracy in extracting interest topics of interest in the range of 0.05-1.00. The result of the classification of the 1251 active users and the accuracies of the interest topic extraction of each group for the topic extraction ratio are shown in Figures 5 and 6, respectively.

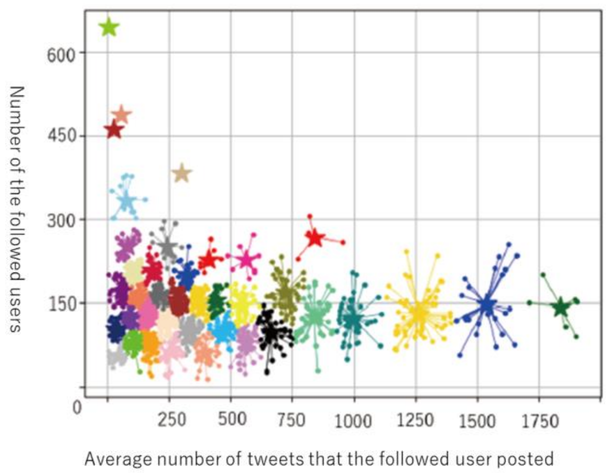

Figure 5: Result of Classification of 1251 Active Twitter Users in the Experiment.

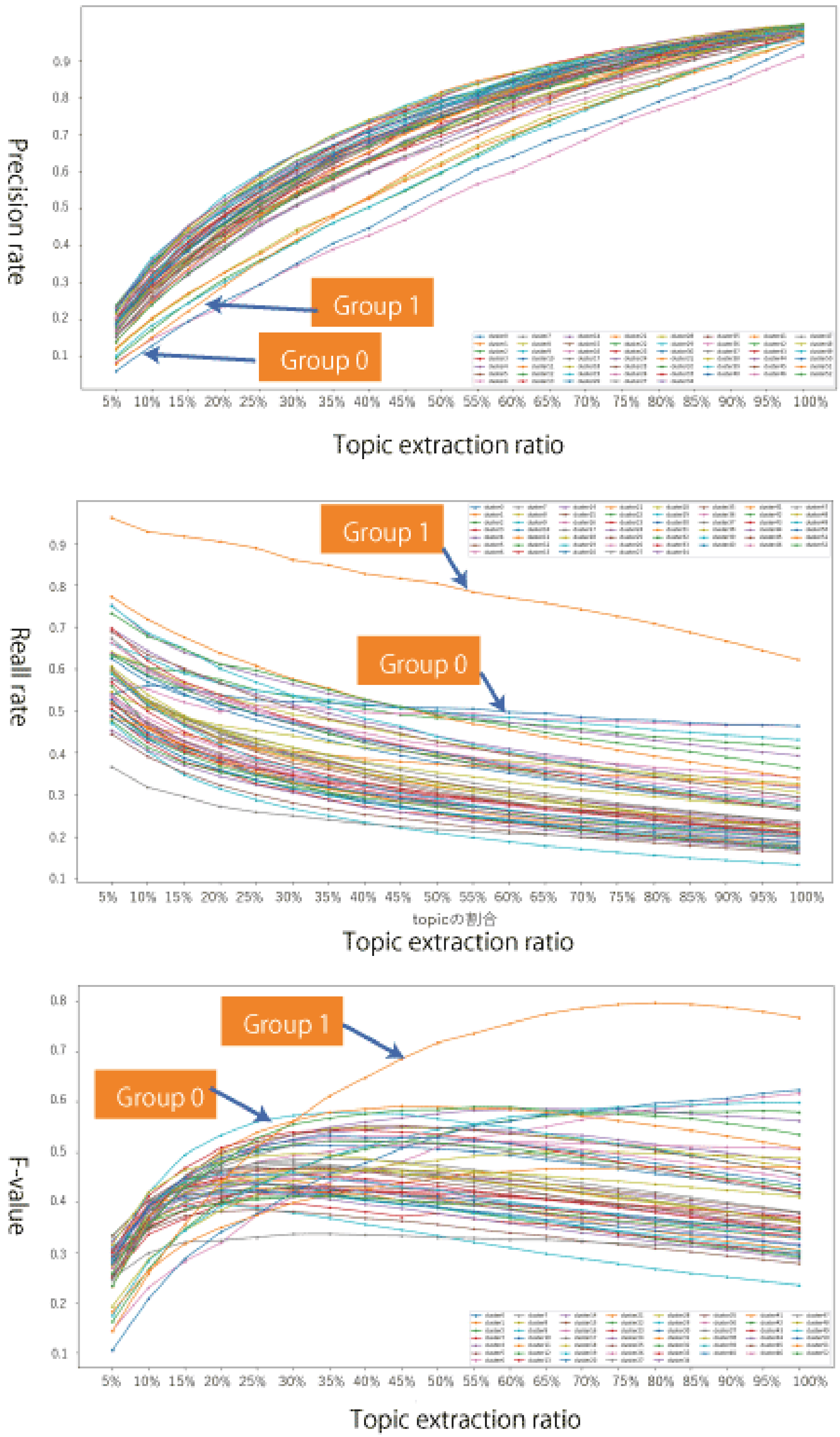

Figure 6: Accuracies of interest topic extraction of each the group for topic extraction ratio.

As shown in Figure 6, the accuracies of interest topic extraction of each group were different in terms of the precision rate, recall rate, and F-value of the same topic extraction ratio. When the F-value was considered the criterion for the evaluation, for the group with the largest topic extraction ratio, the topic extraction ratio was set to 0.95 and its F-value was 0.54. Moreover, for the group with the smallest topic extraction ratio, the topic extraction ratio was set to 0.15 and its F-value was 0.43. For the group with the highest F-value of 0.84, the topic extraction ratio was 0.70, and for the group with the lowest F-value of 0.34, the topic extraction ratio was 0.20. Thus, the optimal set value of the topic extraction ratio is assigned with different values for each user group. Table 3 shows the statistics of Twitter data of the groups in the dataset.

Table 3: Statistics of Twitter data of the 53 groups of the active user in the dataset

Topic extraction ratio				Ave. # of followed users				Ave. # of tweets of followed users			
Mean	S.D.	Max.	Min.	Mean	S.D.	Max.	Min.	Mean	S.D.	Max.	Min.
0.31	0.02	0.95	0.15	311.30	293.72	1163.00	13.00	184.83	30.83	279.99	135.63

6.2.2 Noun occurrence ratio for tweet removal

The accuracies of extracting the interest topics of users with noun occurrence ratios of 0.0, 0.1, 0.2, 0.3, and 0.4 were investigated. The accuracies of interest topic extraction for the noun occurrence rate for tweet removal are shown in Figure 7.

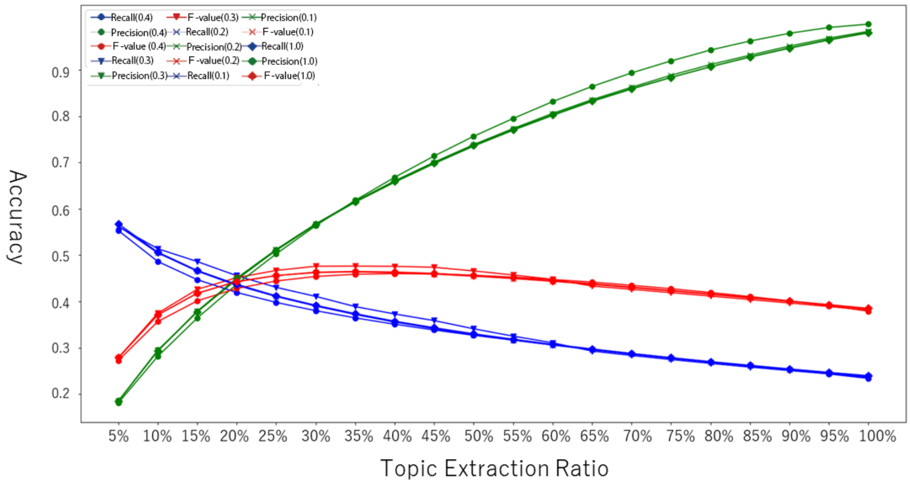

Figure 7: Accuracies of the interest topic extraction for the noun occurrence rate for tweet removal.

By considering the F-value as the criterion, the highest topic extraction accuracy was obtained with the noun occurrence ratio of 0.3 in the range of the topic extraction ratio of 0.05-0.60. The second highest topic extraction accuracies were obtained with the noun occurrence ratios of 0.1, 0.2, and no setting in the range of the topic extraction ratio of 0.05-0.60. Thus, it was found that the optimal set value of the noun occurrence ratio for tweet removal was 0.3 to extract the interest topics of the users with the highest accuracy.

6.2.3 Effectiveness of the proposed method for active users

To confirm the effectiveness of the proposed method for active users, the accuracies of the proposed method with the optimal set values and comparison methods were investigated. The result is shown in Figure 8.

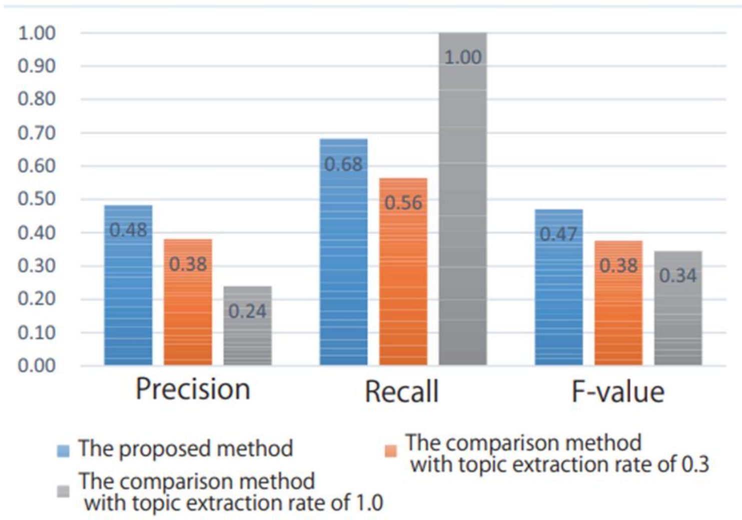

Figure 8: Accuracies of the interest topic extraction for active users.

The accuracies of the proposed method were 0.48, 0.68, and 0.47 for the precision rate, recall rate, and F-value, respectively. Moreover, the accuracies of the comparison method with the topic extraction ratio of 0.3 were 0.38, 0.56, and 0.38 for the precision rate, recall rate, and F-value, respectively, and the accuracies of the comparison method with the topic extraction ratio of 1.0 were 0.24, 1.00, and 0.34 for the precision rate, recall rate, and F-value, respectively. Comparing the F-values and precision rates of these methods, the proposed method obtained the highest accuracies, which are statistically significantly different from the comparison methods with the topic extraction ratios of 0.3 and 1.0 [$F(1, 1250) < 883.98$, $p < .01$ and $F(1, 1250) < 2898.02$, $p < .01$]. Comparing the recall rates of the proposed method and comparison method with the topic extraction ratio of 0.3, the proposed method obtained higher accuracy, which is statistically significantly different from another one [$F(1, 1250) < 727.81$, $p < .01$]. Thus, the proposed method is effective in extracting interest topics for active users in terms of the precision rate, recall rate, and F-value.

6.3 Effectiveness of interest topic estimation for passive user

The accuracies of the interest topic estimation for passive users and the number of topics per interest rating assigned by the Twitter users are shown in Figures 9 and 10, respectively. The number of interest topics in the evaluation data is 84 of 240.

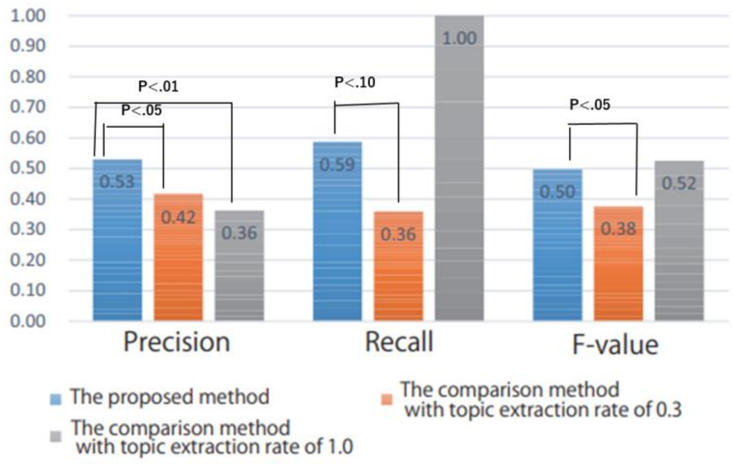

Figure 9: Accuracies of the interest topic estimation for passive user.

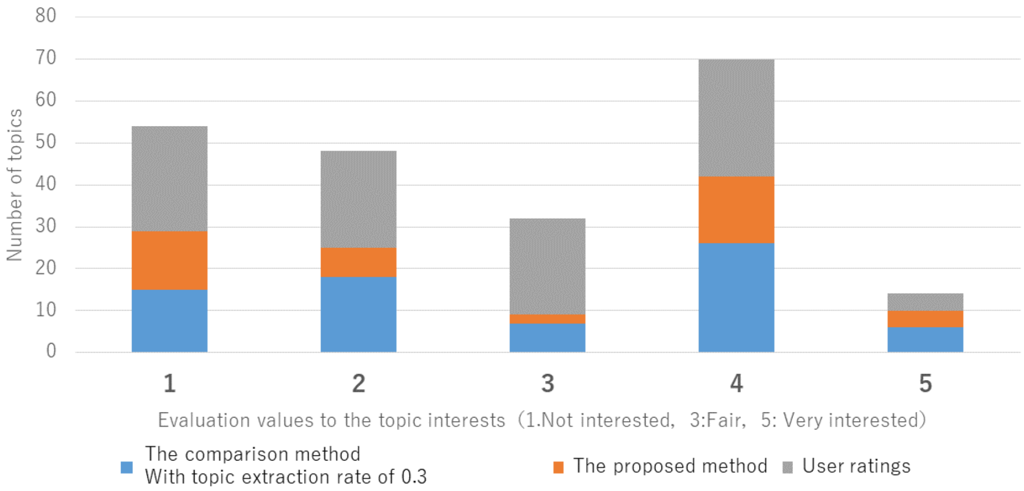

Figure 10: Number of the topics per interest rating assigned by the Twitter Users.

The accuracies of the interest topic extraction of the proposed method were 0.53, 0.59, and 0.50 for the precision rate, recall rate, and F-value, respectively. Moreover, the accuracies of the interest topic extraction of the comparison method with the topic extraction ratio of 0.3 were 0.42, 0.36, and 0.38 for the precision rate, recall rate, and F-value, respectively, and the accuracies of the interest topic extraction of the comparison method with the topic extraction ratio of 1.0 were 0.36, 1.00, and 0.50 for the precision rate, recall rate, and F-value, respectively. Comparing the precision rates of these methods, the proposed method obtained the highest accuracy, which is statistically significantly different from the comparison methods with the topic extraction ratios of 0.3 and 1.0 [$F\ (1,\ 11) < 7,\ 08,\ p < .05$, and $F(1,\ 11) < 13.68,\ p < .01$]. Comparing the recall rates of the proposed method and comparison method with the topic extraction ratio of 0.3, the proposed method obtained higher accuracy, which is statistically significantly different [$F(1,\ 11) < 4.59,\ p < .10$]. In the F-value, comparing the proposed method with the comparison method with the topic extraction ratio of 0.3, the proposed method obtained higher accuracy, which is statistically significantly different [$F(1, 11) < 7.72, p < .05$]. Moreover, comparing the proposed method with the comparison method with the topic extraction ratio of 1.0, the proposed method obtained lower accuracy, but there is no statistically significant difference between them [$F(1, 11) < 0.68, p > .10$]. The proposed method is based on the approach of transferring the interest topic estimation technique for passive users from the interest topic extraction technique for active users. The accuracies of the proposed method for active and passive users were similar in terms of the precision rate and F-value, with the difference being 0.05 and 0.02, respectively.

The proposed method was also found to be effective for passive users. Therefore, the proposed method is effective in estimating the interest topics of passive users by adapting the interest topic extraction of active users.

To further improve the proposed method, we conducted an interview survey on the reasons for the topics with low ratings in this experiment. The results showed that if a topic contained a few words unknown to a user, the user was likely to judge the topic as uninteresting, such as interesting topic rate of 1 and 2 as shown in Figure 10. To solve this problem, it is necessary to consider words unknown to the user in estimating the user's interest, but this requires a new method to be considered, which is beyond the scope of this research.

7 Conclusion

Although there are passive users who only collect information without tweeting, there are no clues to estimate the interests of passive users, and thus research on interest estimation of passive users is inadequate. In this study, we developed a method to estimate the interest topics of passive users from the tweets of their followed users by employing the interest topic extraction method of active users. In the evaluation, we confirmed the effectiveness of the proposed method by comparison with simple topic extraction methods based on the data of 12 users' interest topic evaluation.

In our future work, we will improve the topic analysis of the proposed method and evaluate it on an applied system. To improve the topic analysis, it is necessary to make words in input tweets more general, which can reduce the number of unknown words for the user because the topics are more acceptable to the user. We will also verify the effectiveness of the proposed method by incorporating it into an applied system. In addition, we will develop a topic extraction method adapted to each type of passive user. According to the literature [13], passive twitter users tend to have more celebrities among their followers than active twitter users. Therefore, it would be useful to use external information such as Wikipedia to estimate the interests of passive users by utilizing the information of followed celebrities.

References

[1] Sheth, A., and Kapanipathi, P., "Semantic Filtering for Social Data," IEEE Internet Computing 20(4), 2016, pp.74--78.

[2] Piao G., and Breslin J.G., "Inferring User Interests for Passive Users on Twitter by Leveraging Followee Biographies," In: Jose J. et al. (eds) Advances in Information Retrieval. In Proceedings of 39th European Conference on Information Retrieval; Lecture Notes in Computer Science, vol 10193, Springer, 2017, pp.122--133.

[3] Besel, C., Schl¨otterer, J., and Granitzer, M., "Inferring Semantic Interest Profiles from Twitter Followees: Does Twitter Know Better Than Your Friends?," In Proceedings of the 31st Annual ACM Symposium on Applied Computing. 2016, pp.1152--1157.

[4] Abel, F., Gao, Q., Houben, G.J., and Tao, K., ``Analyzing user modeling on twitter for personalized news recommendations," In Proceedings of the 19th international conference on User modeling, adaption, and personalization (UMAP'11), 2011, pp.1--12.

[5] Siehndel, P., and Kawase, R., "TwikiMe!: user profiles that make sense," In Proceedings of the 2012th International Conference on Posters \& Demonstrations Track-Volume 914. 2012, pp.61--64.

[6] Kapanipathi, P., Jain, P., Venkataramani, C., and Sheth, A. "User interests identification on twitter using a hierarchical knowledge base," In: Presutti, V., d'Amato, C., Gandon, F., d'Aquin, M., Staab, S., Tordai, A. (eds.), In Proceedings of 11th International Semantic Web Conference (ESWC'14), LNCS, vol. 8465, 2014, pp. 99--113.

[7] Piao, G., and Breslin, J.G., "Exploring Dynamics and Semantics of User Interests for User Modeling on Twitter for Link Recommendations," In Proceedings of 12th International Conference on Semantic Systems, 2016, pp.81--88.

[8] Orlandi, F., Breslin, J., Passant, A., "Aggregated, interoperable and multi-domain user profiles for the social web," In: Proceedings of the 8th International Conference on Semantic Systems, 2012, pp. 41--48.

[9] Piao, G., "User Modeling on Twitter with WordNet Synsets and DBpedia Concepts for Personalized Recommendations," In Proceedings of 25th ACM International Conference on Information and Knowledge Management, 2016, pp.2057--2060.

[10] Faralli, S., Stilo, G., and Velardi, P., "Recommendation of microblog users based on hierarchical interest profiles," Social Network Analysis and Mining 5(1), 2015, pp.1--23.

[11] Piao G., and Breslin J.G., "Inferring User Interests for Passive Users on Twitter by Leveraging Followee Biographies," In: Jose J. et al. (eds) Advances in Information Retrieval. In Proceedings of 39th European Conference on Information Retrieval; Lecture Notes in Computer Science, vol 10193, Springer, pp.122–133, 2017.

[12] Piao G., Breslin J.G., "Leveraging Followee List Memberships for Inferring User Interests for Passive Users on Twitter." In Proceedings of the 28th ACM Conference on Hypertext and Social Media (HT '17), pp.155–164, 2017.

[13] Hayama, T. ""Analyzing Features of Passive Twitter's. Users to Estimate Passive Twitter-User's. Interests". The 20th IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT'21) , 2021.

[14] Burnell, K., George, M. J., Vollet, J. W., Ehrenreich, S. E., and Underwood, M. K. "Passive social networking site use and well-being: The mediating roles of social comparison and the fear of missing out," Cyberpsychology: Journal of Psychosocial Research on Cyberspace, Vol.13(3), Article 5, pp.1–14, 2019.

[15] Cesar G. Escobar-Viera, Ariel Shensa, Nicholas D. Bowman, Jaime E. Sidani, Jennifer Knight, A. Everette James, and Brian A. Primack.“Passive and Active Social Media Use and Depressive Symptoms Among United States Adults," Cyberpsychology, Behavior, and Social Networking, Vol.21(7), pp.437–443, 2018.

[16] Hamerly, G., and Elkan, C., “Learning the k in k-means," In Proceedings of 16th International Conference on Neural Information Processing Systems (NIPS'03), 2003, pp.281--288.

[17] Teh, Y. W. and Jordan, M. I., “Hierarchical bayesian nonparametric models with applications," Bayesian Nonparametrics (eds. N. L. Hjort, C. Holmes, P. Muller and S. G. Walker), 2010, pp.158--207.

[18] Blei, D. M., NG, A. Y., and Jordan, M. I., “Latent dirichlet allocation," the Journal of machine Learning research, Vol.3, 2003, pp. 993--1022.