Investigation of Inadequate Multiple Account Users in a Q&A Site by Considering Deviations of Answer Submission Order

Kenji Umemoto, Naoki Ishikawa, Yasuhiko Watanabe, Ryo Nishimura, Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga, 520-2194, Japan
Email: t11m074@mail.ryukoku.ac.jp, t10m096@mail.ryukoku.ac.jp,
watanabe@rins.ryukoku.ac.jp, r_nishimura@afc.ryukoku.ac.jp, okada@rins.ryukoku.ac.jp

Abstract—Some users in a question and answer (Q&A) site use multiple user accounts and attempt to manipulate communications in the site. In order to detect these inadequate multiple account users precisely, it is important to investigate them from various points of view. In this paper, we investigate suspicious users from the viewpoint of deviations of answer submission order and discuss the reasons why and how the deviations occurred. The results of this study will give us a chance to investigate purposes and behaviors of inadequate multiple account users in a Q&A site.

Keywords- *multiple account; Q&A site; deviation; submission order; credibility.*

I. INTRODUCTION

In these days, many people use question and answer (Q&A) sites, where users share their information and knowledge. Q&A sites offer greater opportunities to users than search engines in the following points:

1) Users can submit questions in natural and expressive sentences, not keywords.
2) Users can submit ambiguous questions because other users give some supports to them.
3) Communications in Q&A sites are interactive. Users have chances to not only submit questions but give answers and, especially, join discussions.

As a result, Q&A sites are promising media. One of the essential factors in Q&A sites is anonymous submission. In most Q&A sites, user registrations are required for those who want to join the Q&A sites. However, registered users generally need not reveal their real names to submit messages (questions, problems, answers, comments, etc.). It is important to submit messages anonymously to a Q&A site. This is because anonymity gives users chances to submit messages without regard to shame and reputation. However, some users abuse the anonymity and attempt to manipulate communications in a Q&A site. For example, some users use multiple user accounts and submit messages to a Q&A site inadequately. Manipulated communications discourage other submitters, keep users from retrieving good communication records, and decrease the credibility of the Q&A site. As a result, it is important to detect users suspected of using multiple user accounts and manipulating communications in

a Q&A site. In this case, identity tracing based on user accounts is not effective because inadequate users are likely to hide their true identity to avoid detection. A possible solution is authorship identification based on analyzing stylistic features of messages. In recent years, a large number of studies have been made on authorship identification [1] [2] [3] [4] [5], however, few researchers addressed the identification issues of authors suspected of using multiple user accounts and manipulating communications in a Q&A site. To solve this problem, we proposed methods of detecting two kinds of inadequate multiple account users:

- Multiple account users suspected of submitting questions and their answers repeatedly [6].
- Multiple account users suspected of submitting many answers to the same question repeatedly [7].

However, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users from various points of view. One example is whether these inadequate users use multiple user accounts in different ways. Suppose that one user intends to advocate or justify his/her submitted answer and uses multiple user accounts as follows:

- A main account.
- Secondary accounts for advocating or justifying his/her answer submitted by the main account.

In this case, the deviation of answer submission order is likely to occur. As a result, we investigate user pairs who had large deviations of answer submission order and discuss the reasons why and how the deviations occurred.

By the way, we should notice that it is difficult to verify the credibility of our investigation. This is because there is no reliable information about users who used multiple user accounts and manipulated communications in Q&A sites. In order to discuss the credibility of our investigation, we show the results of our investigation in detail. The results of this study will give us a chance to investigate purposes and behaviors of users who use multiple user accounts and intend to manipulate communications in a Q&A site.

category	N_{qst}	U_{qst}	N_{ans}	U_{ans}	NP_{qst}	UP_{qst}	NP_{ans}	UP_{ans}	$UP_{userpair}$	N_{mfe}
PC	171848	43493	474687	27420	124210	36771	427049	26634	463438	67846
social issues	78777	13259	403306	25766	70886	12238	395415	25552	828812	74781
all 286 categories	3116009	165064	13477785	183242	2576718	150835	12938494	179773	23053308	–

N_{qst} and N_{ans} are the numbers of questions and answers, respectively. U_{qst} and U_{ans} are the numbers of users who submitted questions and answers, respectively. NP_{qst} is the number of questions which had two or more answers, and NP_{ans} is the number of answers submitted to questions which had two or more answers. UP_{qst} is the number of questioners who submitted questions which had two or more answers, and UP_{ans} is the number of answerers who submitted answers submitted to questions which had two or more answers. $UP_{userpair}$ is the number of user pairs who submitted answers to one or more of the same questions. N_{mfe} is the total number of each user's answers which were submitted with his/her most frequently encountered user in the category.

This paper is organized as follows. Section II describes some related works. Section III explains Yahoo! chiebukuro, the data of which we used for observation and examinations. Section IV describes submissions by using multiple user accounts in Q&A sites and deviation of answer submission order. Section V proposes a detection method of too large deviations of answer submission order. Section VI shows the experimental results and discussions. Section VII concludes this study.

II. RELATED WORKS

One of the essential factors in the Internet is anonymity. Joinson discussed the anonymity in the Internet from various points of view [8]. In these days, many users abuse the anonymity: they use multiple user accounts inadequately and submit inadequate messages, such as, deceptive opinion spams. In recent years, a large number of studies have been made on authorship identification [1] [2] [3] [4] [5], however, few researchers addressed the identification issues of authors suspected of using multiple user accounts and manipulating communications in the Internet. One of the difficulties of this problem is that we did not have sufficient number of examples of inadequate multiple account users. To solve this problem, some researchers tried to extract inadequate submissions by using heuristic methods based on text similarities and ranking results [9] [10]. On the other hand, Ott et al. pointed that these heuristic methods were insufficient to detect inadequate submissions precisely, and showed they could detect inadequate submissions precisely when they used large number of examples of inadequate submissions [11]. However, Ott et al. obtained examples of inadequate submissions by using Amazon Mechanical Turk. The examples of inadequate submissions created by workers in Amazon Mechanical Turk have the following problems.

- Little is known about the purposes and methods of inadequate submissions. As a result, it is possible that their instructions to workers in Amazon Mechanical Turk were insufficient.
- There are unreliable workers in Amazon Mechanical Turk [12].

As a result, it is important to obtain inadequate submissions from the Internet. To solve this problem, we proposed methods of detecting inadequate multiple account users and their submissions [6] [7]. However, as mentioned, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users and their inadequate submissions from various points of view.

III. YAHOO! CHIEBUKURO

In this study, we used the data of Yahoo! chiebukuro for observation, data training, and examination. The data of Yahoo! chiebukuro was published by Yahoo! JAPAN via National Institute of Informatics in 2007 [13]. This data consists of about 3.11 million questions and 13.47 million answers which were posted on Yahoo! chiebukuro from April/2004 to October/2005. In the data, each question has at least one answer because questions with no answers were removed. In order to avoid identifying individuals, user accounts were replaced with unique ID numbers. By using these ID numbers, we can trace any user's questions and answers in the data. Table I shows the numbers of users and their submitted messages (questions and answers) to PC category, social issues category, and all 286 categories in the data. Many users have other users who submitted answers to one or more of the same questions with them. We will use the term *most frequently encountered user* of a certain user to refer to a user who submitted answers to the same questions most frequently with the user.

[most frequently encountered user] Suppose $\mathbb{U}$ is a set of users who submitted answers to the same questions with user i. The most frequently encountered user of user i, that is, $mfe(i)$ is defined as follows:

$$mfe(i) = \arg \max_{j \in \mathbb{U}} N_{ans\_together}(i, j)$$

where $N_{ans\_together}(i, j)$ is the number of questions to which user i and j submitted answers together.
N_{mfe} in Table I is the total number of each user's answers which were submitted with his/her most frequently encountered user. As a result, it is expected that, when a user

submitted 100 answers to social issues category, the user and his/her most frequently encountered user submitted

$$\frac{N_{mfe}}{N_{ans}} \times 100 = \frac{74781}{403306} \times 100 = 18.5$$

answers together to the same questions.

Furthermore, the following kinds of information are described in the data.

- Submission time of question.
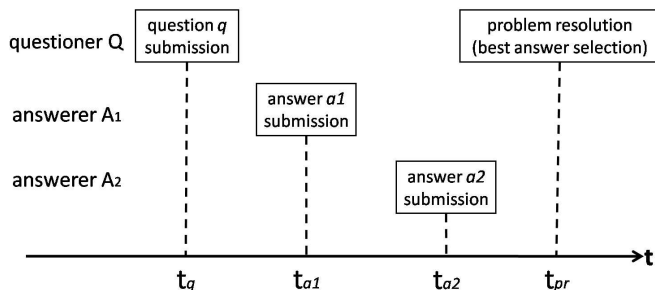- Submission time of answer.
- Problem resolution time.

Figure 1 shows an example of a series of events that occur after a questioner submits his/her question to Yahoo! chiebukuro. In Figure 1, the submission time of question q is t_q. Also, the submission time of answer a_1 and a_2 are t_{a1} and t_{a2}, respectively. Finally, the problem resolution time of question q is t_{pr}. At the problem resolution time, questioner Q stopped accepting answers and determined which answer was the best answer. By using these kinds of time information, we measured two kinds of submission time lags:

- Submission time lags between questions and their answers (e.g., $t_{a1} - t_q$ and $t_{a2} - t_q$ in Figure 1).
- Submission time lags between answers submitted to the same question (e.g., $t_{a2} - t_{a1}$ in Figure 1).

Figure 2 shows the cumulative relative frequency of the submission time lags between questions and their answers in the data of Yahoo! chiebukuro. Also, Figure 3 shows the cumulative relative frequency of the submission time lags between answers submitted to the same question. As shown in Figure 3, the median of the submission time lags between answers submitted to the same question in social issues category was greater than those of PC category and all 286 categories. In social issues category, there were many answers criticizing or against previous answers. As a result, many answerers in this category made and submitted answers after they read other answers to the same question. We think this is one of the reasons why the median of the submission time lags between answers submitted to the same question in social issues category was greater than those of PC category and all 286 categories.

IV. SUBMISSIONS BY USING MULTIPLE USER ACCOUNTS

There are many reasons why users in a Q&A site use multiple user accounts. First, we discuss a proper reason. In Yahoo! chiebukuro, users need not reveal their real names to submit their questions and answers. However, their submissions are traceable because their user accounts are attached to them. Because of this traceability, we can collect any user's submissions and some of them include clues of identifying individuals. As a result, to avoid identifying individuals, it is reasonable and proper that users change their user accounts or use multiple user accounts. However,

Questioner Q submitted question q at t_q. Also, answerer A_1 and A_2 submitted their answers at t_{a1} and t_{a2}, respectively. Finally, questioner Q stopped accepting answers and determined which answer was the best answer at t_{pr}.

Figure 1. An example of a series of events that occur after a questioner submits his/her question to Yahoo! chiebukuro.
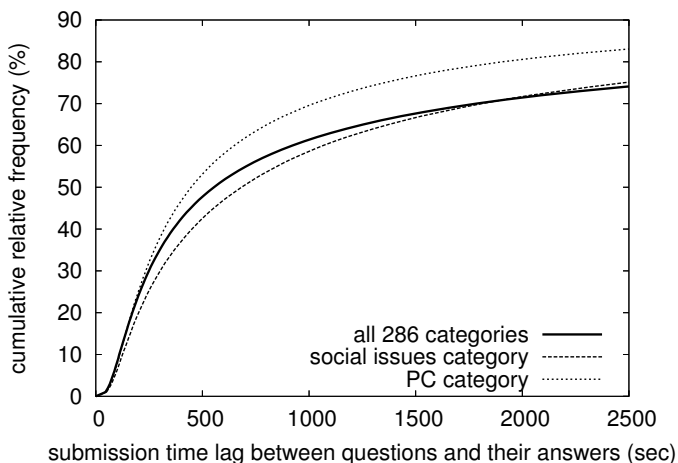

Figure 2. The cumulative relative frequency of the submission time lags between questions and their answers in social issues category, PC category, and all 286 categories of the data of Yahoo! chiebukuro.
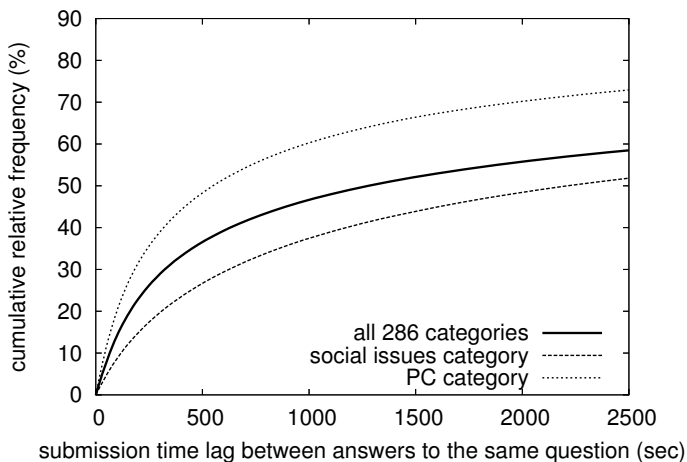

Figure 3. The cumulative relative frequency of the submission time lags between answers submitted to the same question in social issues category, PC category, and all 286 categories of the data of Yahoo! chiebukuro.
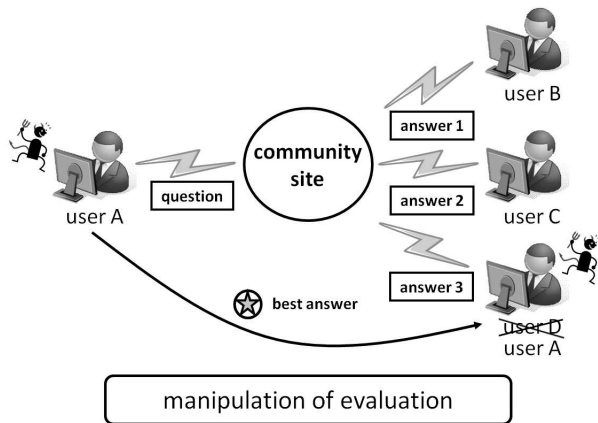
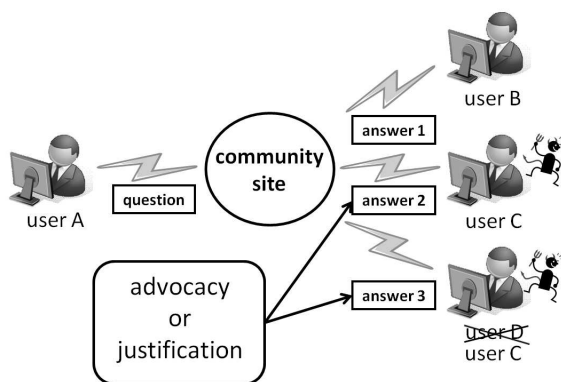Figure 4.   An example of TYPE QA submissions.

Figure 5.   An example of TYPE AA submissions.

the following types of submissions by using multiple user accounts are neither reasonable nor proper.

TYPE QA One user submits a question and its answer by using multiple user accounts, as shown in Figure 4. In Figure 4, user A submits a question and its answer by using two user accounts.
We think that the user intended to manipulate the submission evaluation. For example, in Yahoo! chiebukuro, each questioner is requested to determine which answer is best and give a *best answer* label to it. These evaluations encourage answerers to submit new answers and increase the credibility of the Q&A site. We think, the user repeated this type of submissions because he/she wanted to get many best answer labels and be seen as a good answerer.

TYPE AA One user submits two or more answers to the same question by using multiple user accounts, as shown in Figure 5. In Figure 5, user C submits two answers by using two user accounts. We think that the user intended to dominate or

disrupt communications in the Q&A site. To be more precise, the user intended to control communications by advocating or justifying his/her opinions, or disrupt communications by submitting two or more inappropriate messages.

These two types are not all types of inadequate submissions. However, these kinds of submissions seriously disrupt communications in a Q&A site, discourage other submitters, keep users from retrieving good communication records, and decrease the credibility of the Q&A site. As a result, it is important to detect these kinds of inadequate submissions. To solve this problem, we proposed methods of detecting multiple account users suspected of repeating TYPE QA and TYPE AA submissions [6] [7]. However, little is known about the purposes and methods of inadequate multiple account users. As a result, it is important to investigate these inadequate multiple account users from various points of view. In this study, we investigate the purposes and methods of inadequate multiple account users who use multiple user accounts in different ways.

Inadequate users repeating TYPE QA submissions are likely to use multiple user accounts as follows:

- Main accounts.
- Secondary accounts for submitting questions and manipulating evaluations of main accounts.

However, little is known whether inadequate users repeating TYPE AA submissions use multiple user accounts somehow. To solve this problem, it is important to detect inadequate multiple account users who used multiple user accounts in different ways and repeated TYPE AA submissions, and investigate the purposes and methods of them.

If one user uses multiple user accounts in different ways, some deviations are likely to occur. Suppose that one user intends to advocate or justify his/her submitted answer and uses multiple user accounts as follows:

- A main account.
- Secondary accounts for advocating or justifying his/her answer submitted by the main account.

In this case, the user is likely to submit first answers from his/her main account and other answers from their secondary accounts. In order to detect this kind of inadequate users, we introduce *deviation of answer submission order*.

[deviation of answer submission order]   Suppose user i and user j submitted their answers to the same N questions, and, user i submitted N_i answers earlier than user j and user j submitted N_j answers earlier than user i. The deviations of answer submission order of this user pair is $N_i - N_j$.

As a result, in this study, we investigate user pairs who had large deviations of answer submission order and discuss the reasons why and how the deviations occurred.

In Yahoo! chiebukuro, there were many questions the purpose of which was to collect opinions. For example,

(Q)   What do you think about Prime Minister Koizumi? He has maintained high approval ratings and does well in his work.

This kind of question often had many answers. Some of them were criticizing or against previous answers. Because of such criticizing submissions, some users were likely to use multiple user accounts and submit new answers for advocating or justifying their previous answers. We think some users used multiple user accounts as follows:

- Main accounts.
- Secondary accounts for advocating or justifying their answer submitted by the main accounts.

This is because it is easy to manage multiple user accounts. When multiple user accounts were used as above, it is easy to avoid submitting new answers which were inconsistent with the previous answers. Inconsistent answers often gave suspicious impressions to others. However, if multiple user accounts were used in this way, the deviation of answer submission order is likely to occur. As a result, in this study, we investigate user pairs who had large deviations of answer submission order and discuss the reasons why and how the deviations occurred.

V.   DETECTION OF TOO LARGE DEVIATIONS OF ANSWER SUBMISSION ORDER

In order to detect users who were suspected of repeating TYPE AA submissions by using multiple user accounts in different ways, we introduce two ideas. If one user repeated TYPE AA submissions too many times by using two user accounts, user i and user j, it is expected that

(idea 1) user i and user j submit too many answers to the same questions together.

Furthermore, if the user used these two user accounts in different ways, it is expected that

(idea 2) there are too large deviations of answer submission order between user i and user j.

Based on these two ideas, we determine whether users repeated TYPE AA submissions by using multiple user accounts in different ways.

A. *Detection of user pairs who submitted too many answers to the same questions*

As mentioned, if one person used two user accounts, user i and user j, and repeated TYPE AA submissions in a Q&A site too many times, it is expected that we observe abnormal submissions:

user i submitted abnormally too many answers to the same questions responded by j.

To detect these abnormal submissions, we test one hypothesis: Hypothesis AA.

[Hypothesis AA]   If user i did not submit abnormally too many answers to the same questions with user j, we would expect that user i submitted at most $N_{AA}(i)$ answers to the same questions with user j.

$$N_{AA}(i) = \frac{N_{mfe}}{N_{ans}} \times ans(i)$$

where $ans(i)$ is the total number of answers submitted by user i. As shown in Table I, N_{ans} is the total number of answers submitted to the category, and N_{mfe} is the total number of each user's answers which were submitted with his/her most frequently encountered user.

If this hypothesis is rejected by an one-sided binomial test, we determine that user i submitted abnormally too many answers to the same questions with user j.

B. *Detection of user pairs who had too large deviations of answer submission order*

If one user repeated TYPE AA submissions by using two user accounts, user i and user j, in different ways, it is expected that we observe

too large deviations of answer submission order between user i and user j.

To detect too large deviations of answer submission order between user i and user j, we test one hypothesis: Hypothesis AASO.

[Hypothesis AASO]   Suppose that there are $N_{AA}(i, j)$ cases where user i and user j submitted their answers to the same question. If one of these users did not submit answers too many times before the other did, we would expect that there are at most $N_{AASO}(i, j)$ cases where one user submitted his/her answer before the other did.

$$N_{AASO}(i, j) = P_{AASO}(i, j) \times N_{AA}(i, j)$$

where $P_{AASO}(i, j)$ is the probability that one user submitted an answer before the other did. In this study, $P_{AASO}(i, j)$ was set to 0.5. In other words, user i and user j have equal probability that one user submitted an answer before the other did.

If this hypothesis is rejected by a two-sided binomial test, we determine that one of these users, user i or user j, submitted answers abnormally too many times before the other did.

VI.   RESULT OF THE INVESTIGATION

In order to detect too large deviations of answer submission order, we test Hypothesis AA and AASO. In this study, the target user pairs are 828812 user pairs each of whom submitted answers to at least one same question in social issues category of Yahoo! chiebukuro. This is because there were many discussions between answerers in this category. As a result, it seems more likely that some multiple account users intended to advocate or justify their answers and repeated TYPE AA submissions in this category.

In this experiment, the significance level for Hypothesis AA was extremely low: 0.000005. This is because we intend

Table II
THE RESULT OF THE INVESTIGATION ON 7 USER PAIRS WHO HAD TOO LARGE DEVIATIONS OF ANSWER SUBMISSION ORDER.

A_1	A_2	$N_{AA}(A_1, A_2)$	$NE_{AA}(A_1, A_2)$	$T_{QA}(A_1, A_2)$	$T_{AA}(A_1, A_2)$	decision
691911	802184	47	43	5.0 min.	83 sec.	same user
267614	76731	62	44	22 min.	22 min.	same user
458523	518681	86	61	9.0 min.	26 min.	different users
414445	733881	20	18	4.0 min.	2.3 hrs.	different users
649164	622996	40	30	6.6 hrs.	30 hrs.	same user
471690	471692	12	11	16 hrs.	50 hrs.	same user
622996	471692	12	11	18 hrs.	74 hrs.	different users

User $A1$ more often submitted his/her answers before user $A2$ did. $N_{AA}(A_1, A_2)$ is the number of questions to which both user $A1$ and $A2$ submitted answers. $NE_{AA}(A_1, A_2)$ is the number of questions to which user $A1$ submitted answers before user $A2$ did. $T_{QA}(A_1, A_2)$ is the median of submission time lags between questions and the earlier of their answers of $A1$ or $A2$. $T_{AA}(A_1, A_2)$ is the median of submission time lags between answers of $A1$ and $A2$ submitted to the same question. Decision shows our judgements. By considering the similarity of writing styles and opinions, we determined whether each user pair is one and the same user or not.

to detect extreme abnormal submissions. On the other hand, the significance level for Hypothesis AASO was 0.01.

In this experiment, we first applied Hypothesis AA on 828812 user pairs in social issues category, and detected 20 user pairs who repeated submitting answers to the same question too many times. Then, we applied Hypothesis AASO on these 20 user pairs and detected 7 user pairs who had too large deviations of answer submission order. Table II shows the result of the investigation on these 7 user pairs. In Table II, user $A1$ mainly submitted answers before user $A2$ did. $N_{AA}(A_1, A_2)$ is the number of questions to which both user $A1$ and user $A2$ submitted answers. $NE_{AA}(A_1, A_2)$ is the number of questions to which user $A1$ submitted answers before user $A2$ did. $T_{QA}(A_1, A_2)$ is the median of submission time lags between questions and the earlier of their answers of $A1$ or $A2$. $T_{AA}(A_1, A_2)$ is the median of submission time lags between answers of $A1$ and $A2$ submitted to the same question. Figure 2 showed the cumulative relative frequency of submission time lags between questions and their answers. Also, Figure 3 showed the cumulative relative frequency of submission time lags between answers submitted to the same question. By considering the similarity of writing styles and opinions, we determined whether each user pair is one and the same user or not. Decision shows our judgements. We discuss the following points in detail below.

- Whether each of these seven user pairs is one and the same user or not.
- The purposes of inadequate multiple account users.
- The reasons why and how the deviations of answer submission order occurred.

User pair (267614, 76731) submitted many answers to the questions about foreign residents in Japan. We determined that user 267614 and 76731 were one and the same user. This is because their writing styles and opinions were quite similar and their answers often included special words, for example, personal HP and comic artists, which other users did not cover in this category. These accounts were likely

to be used for repeating the same words. For example,

[Q: 654871] I found this exhibitor in the auction [URL].
   I think it is against the rule.
[A: 76731] It is scratchbuild. Let it go. You are a snitch.
[A: 267614] You are like a snitch in North Korea. Or a
   hound.

We thought there were this kind of inadequate users in Yahoo! chiebukuro. However, we did not think we found them by detecting too large deviations of answer submission order. This is because we did not think of any reasons why this kind of users used their multiple user accounts in this way. We are searching more examples of this kind of inadequate users and intend to find the reasons.

Also, in case of user pair (691911, 802184), we determined these users were one and the same user. This is because the median of submission time lags between their answers was only 83 seconds although user 691911 submitted answers at different times of a day. Furthermore, when user 691911 submitted questions, user 691911 selected user 802184's answers as best answers in too many times in various categories. Like the case of user pair (267614, 76731), these accounts were likely to be used for repeating the same words.

In contrast, in cases of user pair (458523, 518681) and (414445, 733881), we determined that the users of each pair were different users. This is because we found many opinion conflict between the users of each pair. Each pair used Yahoo! chiebukuro almost at the same time of each day. For example, user 458523 and 518681 mainly used Yahoo! chiebukuro from 8:00 am to 5:00 pm. Also, user 414445 and 733881 mainly used Yahoo! chiebukuro from 8:00 pm to 1:00 am. As a result, the users of each pair read questions almost at the same time. On the other hand, the median of submission time lags from questions to user 458523's answers and user 518681's answers were 9.9 minutes and 28 minutes, respectively. Also, the median of submission time lags from questions to user 414445's answers and user 733881's answers were 7.4 minutes and 66 minutes,

respectively. We think, these time lags gave the deviations of answer submission order between the users of each user pair.

Both user pair (649164, 622996) and (471690, 471692) submitted answers repeatedly to questions about a certain religious group. We determined that the users of each pair were one and the same users. This is because they had similar writing styles and opinions respectively. Especially, there was only one opinion conflict between user 649164 and 622996 just after they were pointed out that they were one and the same user. These accounts were likely to be used for criticizing other users' answers, or advocating or justifying their previous answers. In these cases, user 622996 and 471692 mainly criticized other user's answers, and advocated or justified their previous answers. As shown in Table II, user 622996 and 471692 mainly submitted their answers after user 649164 and 471690 did, respectively. In both cases, two user accounts were used in different ways as follows:

- Main accounts (user 649164 and 471690).
- Secondary accounts (user 622996 and 471692) for criticizing other users' answers, or advocating or justifying answers submitted by the main accounts.

Especially, user 471692 often criticized user 622996's answers. As a result, user pair (622996, 471692) was detected although the users of this pair were different users and had different opinions.

VII. CONCLUSION

In this study, we investigated the user pairs who had large deviations of answer submission order and discussed the reasons why and how the deviations occurred. In social issues category of Yahoo! chiebukuro, we found four user pairs suspected of being one and the same users and submitting many answers to the same questions repeatedly by using multiple user accounts in different ways. The purposes of these users seemed to be

- To repeat the same words.
- To criticize other users' answers which were against their answers.
- To advocate or justify their previous answers.

We intend to use the results of this study for further investigation of purposes and behaviors of inadequate multiple account users in Q&A sites. Especially, we think, opinion similarity is a promising clue to the detection of inadequate users and the investigation of their purposes and behaviors.

REFERENCES

[1] O. de Vel, A. Anderson, M. Corney, and G. Mohay, "Mining e-mail content for author identification forensics," ACM SIGMOD Record, Vol.30 No.4, 2001, pp. 55–64.

[2] M. Koppel, S. Argamon, and A. R. Shimoni, "Automatically Categorizing Written Text by Author Gender," Literary Linguistic and Computing, Vol.17 No.4, 2002, pp. 401–412.

[3] M. Corney, O. de Vel, A. Anderson, and G. Mohay, "Gender-Preferential Text Mining of E-mail Discourse," Proc. 18th Annual Computer Security Applications Conference (ACSAC 2002), 2002, pp. 21–27.

[4] S. Argamon, M. Saric, and S. S. Stein, "Style mining of electronic messages for multiple authorship discrimination: first results," Proc. the ninth ACM SIGKDD international conference on Knowledge discovery and data mining, 2003, pp. 475-480.

[5] R. Zheng, J. Li, H. Chen, and Z. Huang, "A Framework of Authorship Identification for Online Messages: Writing Style Features and Classification Techniques," Journal of the American Society for Information Science and Technology, Vol.57 No.3, 2006, pp. 378-393.

[6] N. Ishikawa, Y. Watanabe, R. Nishimura, K. Umemoto, Y. Okada, and M. Murata, "Detection of users suspected of using multiple user accounts and manipulating evaluations in a community site," Proc. the 6th IEEE International Conference on Natural Language Processing and Knowledge Engineering (NLPKE 2010), 2010, pp. 600–607.

[7] N. Ishikawa, K. Umemoto, R. Nishimura, Y. Watanabe, and Y. Okada, "Detection of users in a Q&A site who suspected of submitting multiple answers to a question by using multiple user accounts," Proc. the Fourth International Conference on Internet Technologies and Applications (ITA 2011), 2011, pp. 236–244.

[8] A. N. Joinson, "Understanding the Psychology of Internet Behaviour: Virtual Worlds, Real Lives," Palgrave Macmillan, 2003.

[9] N. Jindal and B. Liu, "Opinion spam and analysis," Proc. First ACM International Conference on Web Search and Data Mining (WSDM 2008), 2008, pp. 219–230.

[10] G. Wu, D. Greene, B. Smyth, and P. Cunningham, "Distortion as a validation criterion in the identification of suspicious reviews," Technical report UCD-CSI-2010-4, University College Dublin, 2010.

[11] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, "Finding Deceptive Opinion Spam by Any Stretch of the Imagination," Proc. the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011), 2011, pp. 309–319.

[12] C. Akkaya, A. Conrad, J. Wiebe, and R. Mihalcea, "Amazon Mechanical Turk for Subjectivity Word Sense Disambiguation," Proc. NAACL-HLT 2010 Workshop on Creating Speech and Language Data With Amazon's Mechanical Turk, 2010, pp. 195–203.

[13] http://research.nii.ac.jp/tdc/chiebukuro.html, [retrieved: June, 2009].