Who Stays Longer in community QA Media?
- Analysis of User Participant Behavior in cQA -

Yoshiyuki Shoji[1], Sumio Fujita[2], Akira Tajima[2], and Katsumi Tanaka[1]

[1] Department of Social Informatics
Graduate School of Informatics, Kyoto University
Kyoto, Japan
{shoji, tanaka}@dl.kuis.kyoto-u.ac.jp
[2] Yahoo Japan Corporation
Tokyo, Japan
{sufujita, atajima}@yahoo-corp.jp

Abstract. Macro and micro analyses of why and when users stop asking and/or answering questions on a community question answering (cQA) site were done for a ten years' worth of questions and answers posted on Yahoo! Chiebukuro (Japanese Yahoo! Answers), the biggest cQA site in Japan. The macro analysis focused on how long participants were active in the QA community from the viewpoints of several user characteristics. In turn, the micro analysis focused on how the participants ' behaviors changes. The behaviors of both askers and answerers were found to change over the time of their active participation: the askers tended to expand the range of categories for which they asked questions while the answerers tended to contract the range of categories for which they answered questions.

1 Introduction

The popularity of community question answering (cQA) media seems to have declined. More than 15 years have passed since online cQA sites appeared. Such sites quickly became popular and one of the most successful types of consumer generated media, serving as a social knowledge community. Their decline in popularly is attributed to the rapid growth of other social network services (e.g., Twitter and Facebook) and the appearance of specific knowledge services (e.g., Stack Overflow). For instance, Yahoo! Chiebukuro (Japanese Yahoo! Answers)[3], the biggest cQA site in Japan, continues to lose users. As shown in Figure 1, the number of posts and new users (askers + answerers) started decreasing around the beginning of 2012. The decreasing number of new answerers is particularly important as it could indicate the beginning of a negative spiral; no one wants to ask in a community without answerers.

Nevertheless, cQA sites continue to be considered useful information resources: cQA contents frequently appear at the top of search results, and many

[3] http://chiebukuro.yahoo.co.jp/

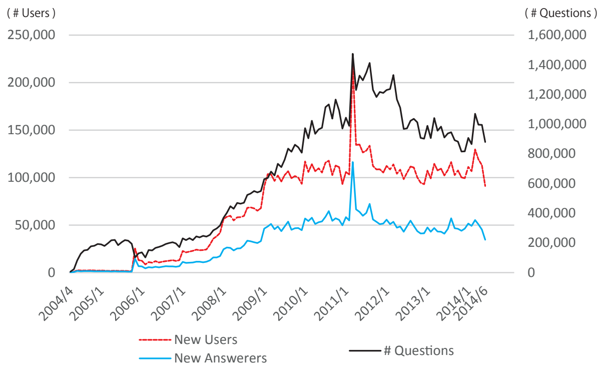

Fig. 1. Number of First Time Users and Posts

Staying Time (month)	
Average	9.98
Mode	1
Median	20

Table 1. User Staying Time

	%user
Asking only	52.8%
(10 times +)	2.2%
Answering only	14.9%
(10 times +)	1.9%

Table 2. Only Ask/Answer

people read them. For instance, Yahoo! Japan reports that there were 714,000,000 Yahoo! Chiebukuro page views in October 2014 [4]. This number is much larger than the number of posts in the same month, 524,000. This indicates that cQA contents are still of interest to read-only viewers. Additionally, cQA is a useful information source for researchers. The contents of cQA sites are a good corpus for many research areas, such as question answering systems, information retrieval, and community analysis. For cQA to continue as a useful resource, the decline in the number of users must be reversed.

We analyzed data on cQA users to clarify their reasons for leaving the cQA community. Such understanding should be helpful in stopping them from leaving. We used a dataset obtain from Yahoo! Chiebukuro that contains ten years' worth of data on users, questions, and answers. We conducted basic statistical analysis at the macro level and sequential user behavior analysis at the micro level. The first step was to statistically compare the data on users who soon left with those for users who stayed longer from several aspects. The next step was to model the sequence of actions for users from when they started using the site to when they stopped using the site. To clarify our analysis, we distinguished between long-stay users and short-stay users. Moreover, we compared the user's monthly behavior from joining to leaving the community.

The rest of this paper is organized as follows. Section 2 describes related work. Section 3 explains our method and presents the results of basic statistical analysis. Section 4 explains the comparison of early-phase behavior and late-phase behavior. Section 5 explains our findings and how they might be utilized. We conclude with a summary of the key points in Section 6.

[4] http://i.yimg.jp/images/marketing/portal/paper/media_sheet_open.pdf

2 Related Work

This section introduces research on cQA service analysis in Section 2.1, research on cQA user motivation on Section 2.2, and research on sequential behavior analysis and phase modeling in Section 2.3.

2.1 cQA Services Analysis

Community question answering is a hot research topic in many areas [20][6]. How to find experts on cQA sites is a typical topic in knowledge community research [21][22][15]. Another typical topic is how to support cQA services. For instance, a question recommendation function helps keep answerers connected to a cQA site [13]. Yang et al. [23] used user lifespan as a tool for comparing cQA communities. User lifespan is tightly related to when and why a user leaves the community. Data on cQA is also used for research in other areas, such as corpus. Several Web information retrieval conferences have included question-answering tracks. For instance, the NTCIR (NII Test Collection for IR Systems) project circulated a cQA dataset as a dataset for a factoid-based retrieval system [5].

There has also been research on the relationship between cQA sites and other social media sites. Morris et al. [19] analyzed data on users who asked questions on an SNS site rather than on a cQA one. One of their findings was that some questions are better suited for a cQA site (e.g., excessively private questions and political questions) while some are better suited for an SNS site (e.g., social invitations and context-dependent questions).

2.2 cQA User Motivation

One of the biggest research topics related to the analysis of cQA services is classification of question types and estimation of the askers' intents. Many researchers have proposed rules for classifying questions and asker intents. Kim et al. [11] classified questions as information, suggestion, opinion, or other for use in estimating the best answer automatically. Harper et al. [8] classified questions into factual, opinion, and advice for use in finding expert answerers on cQA sites. Both classification methods cover opinion. This is because askers on cQA sites often want not only factual information, but also the opinions of others. Liu et al. [17] classified cQA asker intent into navigational, informational, transactional, and social. This classification method supports the conventional Web search query classification [12]. It also highlights a characteristic of cQA sites; that is, the difference between web search users and cQA users is whether they have social intent. Chen et al. [4] also focused on the social intent of the asker, such as subjective, objective, and social. These research efforts generally focused on investigating the reasons that cQA users ask questions. Our work aims to broaden this target; focusing not only on the askers but on all users, including the answerers.

Harper et al. [7] observed that users of cQA sites have two general types of intent. One is an **informational** intent, and the other is a **conversational** intent. Users who have an informational intent tend to ask specific questions that have a clear answer. Those who have a conversational intent want to communicate through cQA sites, and they tend to ask questions that have no clear answer. Harper et al. asserted that users can be classified by using category, text, and user information. They also asserted that conversational questions have low archival quality. Aikawa et al. [2] proposed a method for sorting subjective questions in the same way. Liu et al. [14] discovered the same direction for askers who switch from web search to cQA asking; that is, they ask informational questions more than sticky askers, i.e., those who ask question on a cQA site over a long period of time.

2.3 Sequential Behavior Analysis and Phase Modeling

There has been much research on the analysis of user behavior by using sequential action modeling techniques, not only for cQA but also for other services. One typical example is user modeling for electronic commerce sites. Some researchers have split the total user time into sessions and then created a sequential model for use in estimating whether the user will buy something.

They considered many users actions to be features, such as page transitions, click-throughs, and search queries [10]. Several modeling methods have been used for this kind of situation. For instance, the hidden Markov model has been widely used for modeling the behaviors of web searchers and web service users [18]. Hassan et al. used relational Markov models [3] to model Web user behavior as a means to enhance Web navigation [9].

3 Macro Analysis of cQA User Behaviors

In this section, we explain our macro analysis of users who stop using a cQA service. We compared two types of users: those who use the service for a relatively long time and those who use it for a relatively short time. To do this, we set the following research question and hypotheses:

RQ1: How do users who leave soon differ from those who stay longer?
H1.1: They make different types of contributions.
H1.2: The range of categories in which they participate differs.
H1.3: Their satisfactions and rewards differ.

To answer this research question and to test these hypotheses, we used a massive dataset to analyze user behavior, such as the characteristics of users who leave the QA community. Section 3.1 describes the dataset, and section 3.2 explains our analysis.

3.1 Dataset

The Yahoo! Chiebukuro (Japanese Yahoo! Answers) is the biggest online question answering community in Japan and is freely available to all Yahoo! Japan users. It began operation in 2004. The dataset contained slightly more than ten years' worth of data (March 2004 to December 2014).

- Questions: 84,123,965 questions that received one or more answers
- Answers: 224,969,887 answers to the questions
- Users: 10,391,194 anonymized users who had asked and/or answered one or more questions

Questions that do not receive an answer are deleted from the system, so the dataset contained only questions that received one or more answers. An Asker has the option to specify the best answer (BA). If the asker does not specify a BA, other users can specify one by voting. Yahoo! Chiebukuro has three levels of categories. The top 16 categories are shown in Table 4.

Since our research focused on users leaving the cQA community, we neglected users still using the site. It has been reported that a certain percentage of users who leave for a while never come back [23]. We assumed that the threshold is six months; the dataset contained user data from March 2004 to December 2014, so we eliminated users who asked or answered one or more times after June 2014. In this dataset, 87.85% of the users had already stopped using the site. The average staying time was 9.98 months, as shown in Table 1.

3.2 Basic Statistics

To tackle research question 1, to clarify the difference between shot-stay users and long-stay users, we verify three hypotheses.
H1.1: Difference in Contributions

First of all, we set a simple hypothesis: a user ' s contribution, such as frequency and ratio of asking and answering, reflects his or her length of stay. A breakdown of the user contributions is shown in Table 3. One simple finding is that many Yahoo! Chiebukuro users were light users; they used it once and then stopped using the service. A third or more of the users left the community after asking or answering once. Some were askers who asked only once and left the service. Apparently, they had a problem and they used the service to find a solution. Once they had a solution, they stopped using the service. There were some heavy users as well. They used the service for a long time and frequently post both questions and answers.

Users of cQA sites can contribute as an asker or answerer or as both. The relationship between staying time and the ratio of asking and answering is shown in Figure 6. The horizontal axis represents the number of months that the user used the service (binned with month.). The vertical axis represents the average ratio of questions to answers that users posted. The ratio of asking correlates significantly with staying time; the rank correlation coefficient was -0.88. Users who stayed longer tended to answer more than to ask. Moreover, they posted both questions and answers, as opposed to inactive short-stay users who posted either a few questions or a few answers. Few users only asked or answered more than ten times (see Table 2). Another phenomenon that can also be seen is that users in cQA differentiate into two types: asker and answerer. It is unusual to find users who post both questions and answers in similar numbers. (see Figure 5)

The number of posts per month shows the characteristics of askers and answerers. Figure 4 shows the relationships between question and answer posting frequencies and staying time. It shows that it is generally difficult to continue posting very frequently for long time. While users who stay longer post at their own pace, the answerers tend to be more frequent posters later in the staying time. The number of heavy answerers who posted more than 100 times per month and stayed 3 years or more was 2011, and the number of users who stayed 5 years or more was 861. The corresponding numbers for askers were much smaller: 82 and 37, respectively. Thus, users who use cQA sites for a longer period apparently like answering more than asking.

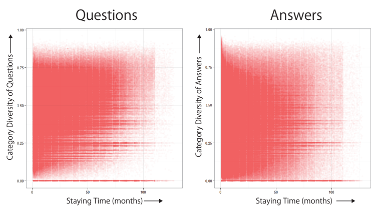

Fig. 2. Category Diversity vs Staying Time

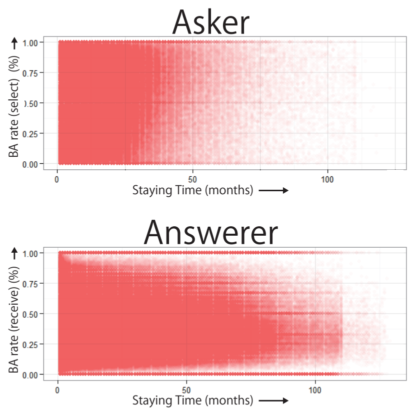

Fig. 3. Ratio of Giving or Receiving Best Answer vs Staying Time

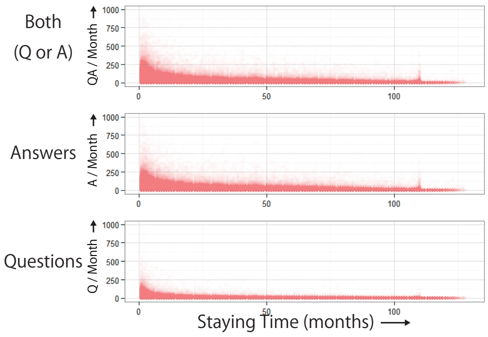

Fig. 4. Number of Posts Per Month vs Staying Time

	Asking	Answering	Both
Maximum	29,757	353,126	362,692
Median	8	733	741
Average	7.96	20.36	28.32
Mode	1	0	1
only 1 time	35.7%	14.9%	37.0%
10 times or less	87.8%	86.9%	78.0%
50 times or less	97.6%	95.3%	92.9%
100 times or more	1.1%	2.8%	4.0%

Table 3. Breakdown of User Contributions

H1.2: Difference in Category Participation

The categories on a cQA site directly reflect the users' interests. Moreover, the interests may affect user behavior. To test this, we defined a category for each user on the basis of how many times the user posted answers and/or questions in each category. The defined category was the one with the most posts. If a user posted a similar number of questions and/or answers to two or more categories, we randomly set the user's category to one of them.

Table 4 shows the details for each category. The staying times ranged from 5.3 to 11.6 months. The users who were defined with a daily life category tended to stay longer: the top three categories by staying time were "Entertainment and Hobbies," "Life, Love, and Human Relationship," and "Sports, Outdoor, and Cars." Conversely, the users who were defined with more serious categories, such as "Business, Economics, and Money," "Manner and Ceremony," and "Computer Technology," tended to stay for a shorter time.

Another factor related to staying time was the activeness of the community: the correlation coefficient between staying time and the number of posts was 0.91, and the correlation between staying time and the number of users was 0.88. Each category had a different distribution ratio of long-stay and short-stay users. The columns "< 0.5 (0.5 year or less)," "$0.5 - 1$ ($0.5 - 1$ years)," "1 - 3 ($1 - 3$ years)," and "> 3 (more than 3 years)" shows the component ratios. The categories with many short-stay users seem better suited for askers with sudden problems. The top three categories in this regards were "Computer Technology," "Yahoo! Japan," and "Manner and Ceremony." Each of them contained many questions that needed a quick answer, like "My computer has broken down; what should I do?" and "A relative has died; how much money should I give as a condolence gift."

Moreover, the average length of posts in each category was inversely correlated with staying time. Most of the long-stay users tended to participate in short and frequent communication rather than lengthy discussions. Several researchers have observed that cQA users generally have two types of intent: informational and conversational. This is highly related to the categories. Harper et al. [7] pointed that intent can be classified into informational and conversational only by category with an accuracy of 70 %. For analysis, we classified 100 randomly selected QA pairs for each category by hand. The ratio of informational questions ("% Info" in Table 4) did not correlate with staying time ($R = -0.00$). Although the component percentages of users in the community correlated with staying time, the ratio of long-stay users was low in the informational categories. The percentage of users who stayed three or more years had a weak inverse correlation with % info ($R = -0.28$). As another perspective, we focused on category diversity. Adamic et al. pointed that the diversity of categories in which a user posts questions and/or answers indicates whether a user is a specialist or not [1]. That is, the diversity of questions and answers represents the range of the user's interests and knowledge. We calculated the entropy of each user on the basis of

Category	Staying	% info	# users	# board	<0.5	0.5–1	1–3	>3	Length
Entertainment and Hobbies	11.6	66%	2,328,386	17,571,681	43%	35%	10%	13%	68.6
Life, Love, and Human Relationship	10.7	0%	1,750,675	8,255,538	46%	34%	9%	11%	181.9
Sports, Outdoor, and Cars	10.3	48%	1,094,271	6,805,630	48%	38%	6%	8%	75.3
Living	9.1	77%	1,092,231	7,303,245	44%	43%	5%	8%	108.9
Kids and Schools	8.9	25%	1,015,797	6,193,570	47%	36%	7%	11%	85.3
Health, Beauty, and Fashion	8.4	68%	843,838	7,185,539	54%	29%	7%	10%	111.7
Internet, PC, and Electronics	7.9	82%	1,204,661	8,348,942	37%	53%	4%	6%	107.9
Local, Travel, and Trip	7.9	56%	707,311	3,887,467	60%	27%	6%	6%	80.7
Knowledge, Education, and Science	7.7	80%	794,644	3,848,875	48%	39%	6%	7%	155.2
News, Politics, and Global	6.8	27%	621,021	2,984,093	62%	25%	6%	8%	96.9
Jobs and Career	6.6	41%	721,055	2,360,837	49%	38%	5%	7%	146.8
Others	6.3	9%	541,983	3,498,145	64%	23%	6%	7%	44.3
Business, Economics, and Money	6.1	51%	495,974	2,065,734	69%	19%	6%	6%	121.7
Manner and Ceremony	6.0	28%	463,642	1,283,267	72%	15%	6%	7%	142.3
Yahoo! Japan	5.7	31%	430,842	2,110,650	75%	11%	7%	8%	85.6
Computer Technology	5.3	92%	358,264	420,537	81%	6%	6%	7%	255.7

Table 4. Category Detail

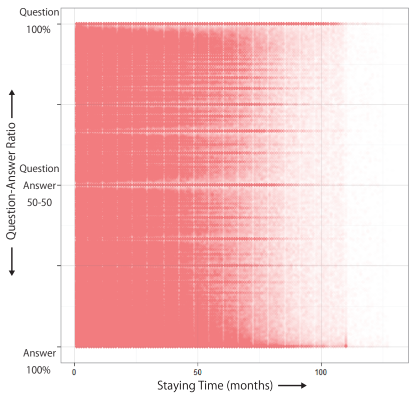

Fig. 5. Asking Answering Ratio vs. Staying Time

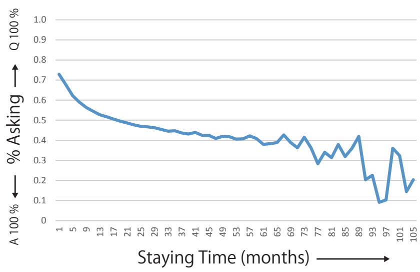

Fig. 6. Asking Answering Ratio vs Staying Time

the categories in which he or she posted questions and answers:

$$etp(u) = -\sum_{c \in C} \frac{|p \in P(u)|cat(p) = c|}{|P(u)|} log_{|C|} \frac{|p \in P(u)|cat(p) = c|}{|P(u)|}, \tag{1}$$

where $P(u) = p_1, p_2, p_3 \cdot$ is the set of posts by user u, $cat(p)$ is the category of post p, and $C = c_1, c_2, c_3 \ldots$ is the top level category. It drops between zero to one by normalizing by taking logarithm $|C|$. The overall tendency is that a user who stays longer posts in a broader range of categories. We calculated category diversity for both questions and answers. Spearman's rank correlation coefficient for staying time and category diversity for both was 0.99. While both were correlated with staying time, they were markedly different in terms of distribution and detail. As shown in Figure 2, there was a sharp contrast between askers and answerers. The most notable feature in this figure is the sparseness at the bottom-right corner of the "Questions" scattergram. Askers who stay longer tend to post questions in a broader range of categories, so it is quite possible that, if an asker continues asking questions in a specific category, he or she will eventually run out of questions. In contrast, answerers who stay longer tend not to post answers to such a broad range of categories. The distribution does not appear to converge or follow a normal distribution. The range of categories in which a user posts answers is assumed to reflect the answerers' field of expertise. Since no one is truly omniscient, no one can continue to answer questions in a wide range of categories. Therefore, answerers who consistently post answers in several specialized categories tend to stay longer.

H1.3: Difference in satisfaction and reward

The satisfaction of askers and the reward to answerers have also been used to estimate the motivation of users to continue using a cQA service. The best answer (BA) rate is related to satisfaction and reward [16].

For askers, indicating the BA is a way to indicate satisfaction. Askers typically select the BA when they are satisfied and may not select a BA if all the answers are unsatisfactory. (As mentioned above, other users can then vote for the BA.) Similarly, having his or her answer selected as the BA is rewarding to the answerer. Answerers wanting to get more BA selections tend to answer questions more frequently, and politely.

The BA ratios of askers and answerers are plotted against staying time in Figure 3. The vertical axis represents the ratio of selecting or receiving a BA, and the horizontal axis shows the staying time. For askers, there is no correlation between the rate of selecting the BA and staying time ($R = 0.12$). This is because there are many short stay users, and they tend to select the BA randomly. The graph does show, however, that users who select fewer BAs tend to stay a shorter length of time. The average probability of a user selecting the BA is 0.46. For answerers, the correlation is not much better ($R = 0.17$), indicating that they do not seem particularly concerned about the BA rate.

In our dataset, the average number of answers for a question was 2.7. It is thus natural for the BA rate to converge to 0.37. In other words, people who answer questions over a long period of time do not seem eager for their answers to be selected as the BA.

4 Micro Analysis of cQA User Behaviors

To analyze user behavior from the micro view, we compared the users' actions during the user action phase, from the month in which the user joined the service to the month in which the user left the service. Again we set a research question and three hypotheses.

RQ2: Do users change their actions during the period from when they start using the site to when they stop using the site?

- H2.1: The contribution type changes.
- H2.2: The category in which they participate changes.
- H2.3: The satisfaction and reward change.

To test these hypotheses, we compared users' monthly behaviors from when they started using the site to when they stopped using it.

4.1 User Phase Analysis

Before analyzing the user action phase, we removed the data on users who posted less than five questions and/or answers. The number of users who posted 5 times or more was 3,256,402, i.e., 31.24% of all users. We compared the monthly activity of users from two viewpoints: staying time and period of time from start using or stopped using. For instance, we focused on what users who stayed m years did in the n-th month after starting to use the site or on the last action of long-stay users before they left.

H2.1: Difference in Contributions

How users commit to the site may vary depending on the their period of use. We focused on changes of their contribution, such as the frequency of posts and ratio of asking and answering.

Figure 9 shows the ralation between their period of use and frequency of posts. The upper graph shows the number of post per month, and the lower graph shows the cumulative number of posts. The horizontal axis means how many months have passed after users start using the site. It shows the behavior of ten user groups divided by their staying time, from a year to 10 years. The graph shows the overall trend, that is, users start using the cQA site in the active state. They frequently post questions and answers. After that, they decrease the frequency of posts, and finally they leave. For most users, the period when they post with high frequency is right after joining. Both short-stay users and long-stay users post about ten times in a month when they start using the service. Their participation decreases as time goes on, although shor-stay users are easily excited but quick to cool down again (see Table 5). The long-stay users tend to continuously use cQA even after their participation decreases, in contrast to the short-stay users who post more frequently and leave abruptly.

Asking and answering also depends on the period of use. Figure 7 shows the ratio of asking and answering in n-th month from joining the service. The vertical axis shows the users' asking ratio up to that time. As explained above, most users start using cQA as askers, and increase their percentage of answering during their using. Except the remarkable long-stay users (those who stay nine or more years), users start using the cQA site by posting 60 % questions and 40% answers, and leave the site if the ratio of asking becomes 55%. It is possible that they exhaust the questions and lose interest in the site. For the users who leave after 1 year to 9 years, the speed of decreasing of asking ratio is slower. On the other hand, users who use the cQA site for nine or more years are more heterogeneous than the others. What we can say is that the users who use Yahoo! Chiebukuro for 9 years or more are Early Adopters; because the service started in 2004. They answered more frequently than most of the people joining later, especially in the early period of the service. The amount of users were smaller at that time, therefore they may retained communication by posting many answers for rare questions.

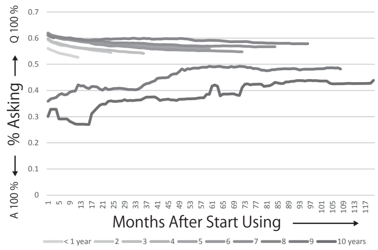

Fig. 7. QA ratio vs n-th month

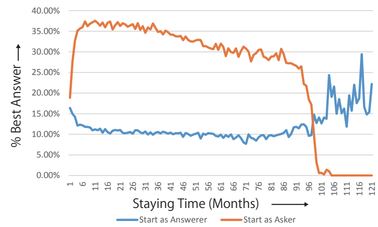

Fig. 8. BA Rates on First Posts

We found that 67.76% of the users who posted five times or more, started by asking questions. As mentioned in section 3.2, many users stop using a cQA site after they post a few questions. Others apparently started to like the site and become regular visitors, and some of the regular visitors become long-staying users. There was only a small difference between the first action of short-stay users (stay less than a year) and that of long-stay users (stay 5 or more years). The percentage of short-stay users whose first post was a question was 62.48 % and that for long-stay users was 70.10%. That is, more long-stay users start by asking a question. The last posts of most users before they leave the site are also asking, as 67.76% of users stop using the service after they post a question. These facts suggest two factors: a great multitude of users post answer in their middle period, and some particular users who are enthusiastic answerers post a large amount of answers. It is possible that users who leave by asking a question are losing their interest in the site and they just do not want to search a question relevant to their interest. They may be tired to be commited to the cQA site.

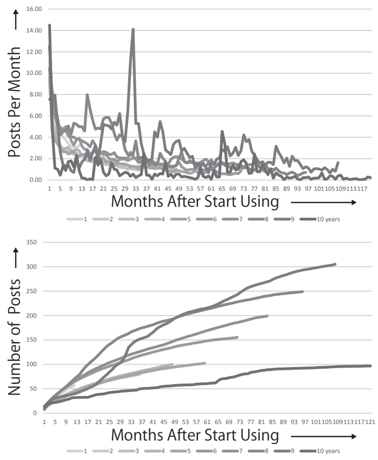

Fig. 9. Posts frequency vs n-th month

Staying years	Period of use starting	1 year	3 years	5 years
1	12.53	3.25	-	-
2	10.98	2.36	-	-
3	10.21	2.31	1.15	-
4	9.23	2.75	1.31	-
5	9.63	2.19	1.27	0.68
6	12.39	3.92	1.74	1.10
7	12.56	3.82	1.80	1.70
8	10.47	4.98	2.19	1.60
9	7.55	1.09	2.40	0.71
10	14.50	0.89	2.50	0.39

Table 5. Frequency of Post and Time Prsiod

H2.2: Difference in Categories

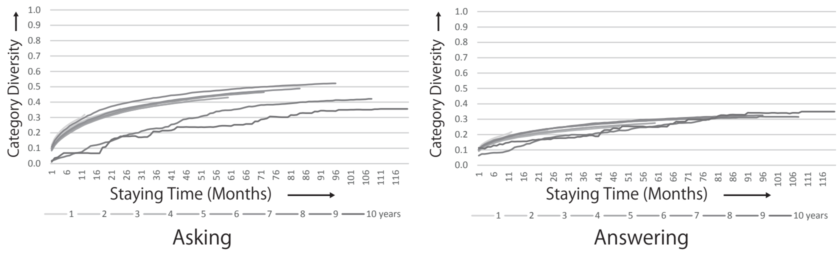

Fig. 10. Category-diversity vs Time Period

	First Post						Last Post					
	Asking			Answering			Asking			Answering		
Category / Staying Years	<1	1–5	5–10	<1	1–5	5–10	<1	1–5	5–10	<1	1–5	5–10
Entertainment and Hobbies	12.6%	13.1%	8.3%	8.9%	7.7%	4.6%	10.4%	10.3%	7.7%	10.7%	7.8%	5.1%
Internet, PC, and Electronics	6.7%	9.3%	11.0%	2.0%	1.6%	2.0%	7.5%	11.4%	13.5%	2.4%	2.0%	2.2%
Life, Love, and Human Relationship	7.1%	4.9%	2.8%	5.7%	4.4%	3.1%	5.3%	4.4%	3.5%	7.0%	5.7%	5.7%
Living	5.0%	6.8%	8.4%	2.6%	2.3%	3.0%	4.8%	7.1%	9.3%	3.3%	2.6%	2.9%
Health, Beauty, and Fashion	6.2%	6.6%	6.5%	3.0%	2.4%	2.9%	5.7%	6.3%	5.5%	3.5%	2.3%	2.1%
Sports, Outdoor, and Cars	4.0%	5.5%	6.2%	3.0%	2.8%	3.4%	3.5%	5.1%	6.0%	3.5%	3.0%	3.2%
Knowledge, Education, and Science	3.9%	4.1%	4.0%	1.8%	1.5%	1.8%	3.5%	4.2%	4.0%	2.1%	1.5%	1.4%
Local, Travel, and Trip	2.4%	3.8%	4.1%	1.0%	1.0%	1.3%	2.5%	4.8%	6.4%	1.2%	1.2%	1.5%
Kids and Schools	3.8%	3.8%	3.6%	2.4%	2.3%	2.6%	3.0%	3.4%	2.6%	2.9%	2.5%	2.4%
Others	2.2%	1.2%	0.9%	1.8%	0.7%	0.6%	1.8%	1.0%	0.7%	2.2%	0.9%	0.6%
News, Politics, and Global	1.0%	0.7%	0.7%	1.3%	0.9%	0.9%	1.0%	0.8%	0.7%	1.6%	1.2%	1.2%
Jobs and Career	2.2%	2.6%	2.5%	0.9%	0.8%	0.9%	1.8%	2.6%	2.5%	1.0%	0.9%	0.9%
Yahoo! JAPAN	2.4%	2.6%	4.6%	1.6%	0.9%	1.6%	1.9%	1.0%	0.8%	1.8%	0.5%	0.3%
Business, Economics, and Money	1.7%	2.5%	3.4%	0.7%	0.6%	0.8%	1.6%	2.6%	3.6%	0.8%	0.6%	0.7%
Manner and Ceremony	0.9%	1.3%	1.9%	0.7%	0.6%	0.8%	0.8%	1.1%	1.5%	0.8%	0.6%	0.7%
Computer Technology	0.4%	0.6%	0.8%	0.1%	0.1%	0.1%	0.4%	0.6%	0.8%	0.1%	0.1%	0.1%

Table 6. first Post and Last Post

The users must change the category which they commit in response to current time period. First, we focused on the category which the user commited at the first posting. Table 6 shows the category and the type of first posts. As explained above, user tend to start using the cQA site as an asker. Regardless of users' staying time, a lot of user start using cQA in entertainment and advice categories. These categories are good appeal to get new users. They are originally popular categories. Let us compare the last category they commit before leave. The ratio of answering increases especially in counseling categories and Entertainment categories. This may reflects that users who stays longer in cQA tend to love communication but not serious arguments.

5 Discussion

Category diversity which users ask or answer also change with their phase. Diversity generally increases when a user posts in more categories in a balanced manner. Therefore, it has correlation to the length of using. Users post questions and answers in new categories for a while after joining the service. On some level, increasing of the diversity stops because they stop penetrating new categories or they posting their own specific categories disproportionately. **H2.3:** Difference in satisfaction and reward Users' bahavior on getting or giving BA also depends on their phase and staying time. The simplest hypothesis that the asker who got good answer must stay, and answerer who got BA must stay. To test it, we focus on their first actions. Figure 8 shows the relation between first action and staying time by the BA. Surprisingly, in the case of the user who post a question first, the ratio of giving BA has inverse correlation to the staying time ($R = -0.82$). Through the whole data, probability of giving BA at first asking time is 33.54%. One of long-stay users is lower, especially in the case of staying months is longer than 90 months. One reason is that they are early adopters and stay the site since it was immaturity. In that time, the quarity of answer might be lower, and the rule of giving BA was not enough spreaded.

In the case of users who join the cQA as answerers, user who got BA at first post tend to stay linger($R=0.45$). To get BA on first post is not easy; the probability if 10.88%. Such answerers might imbibe a taste for getting BA as an incentive.

We identified several tendencies of long-stay and short-stay users of cQA sites from our analysis that should be useful in keeping users active on cQA sites for a longer period of time. First of all, many users use cQA only once as an asker and leave. Keeping such users is an important issue. Related to this, we have that answerers tend to stay longer than askers. It may be possible to recommend a catchy question for askers that would cause them to unconsciously switch from being an asker to being an answerer. Most users start using cQA as askers; then, as they stay longer, they differentiate into askers and answerers over time. For user who regularly uses cQA for a reasonable amount of time, navigation or recommendation should be personalized to their stance.

Our findings related to categories may also be useful for keeping users. Askers who post questions in a wide range of categories tend to stay longer, while answerers who post in a narrow range of categories tend to stay longer. For askers, personalized navigation may help them to expand the range of categories in which they ask questions. For answerers, it is better to let them focus on their field of expertise. cQA sites are also used to get information about various types of amusement. Users who mainly post in entertainment-related categories tend to stay longer. The categories in which the posts are short tend to be more popular and attract more long-stay users. Some desultory exchanges of words in short sentence are not useful for improving the information quality of cQA sites, but it is useful to sustain the community in active state.

Askers who tend to select the BA tend to stay longer. In other words, askers who feel that they did not receive a good answer tend to leave sooner. It is safe to assume that good answers make askers more satisfied, and this encourages them to stay longer. In contrast, answerers who stay longer are apparently less concerned about their BA rate.

Our phase analysis revealed that most users of a cQA site are not always active. While users who stay longer are also not always active, they tend to come back sooner to answer or to ask. One simple way to keep users active is to send them reminders or advertisements.

6 Conclusion

We have conducted both macro and micro analysis to clarify why and when users stop asking or answering questions on cQA sites. We used a dataset obtained from the Yahoo! Chiebukuro site and analyzed it both statistically and sequentially. For the macro analysis, we set three simple hypothesis related to contribution, category, and reward and satisfaction and analyzed how long users participate in the QA community. For the micro analysis, we analyzed how cQA users behave by comparing users' actions in early-phase and late-phase. We found that askers who stay longer in cQA sites tend to ask questions in a wider range of categories while answerers who stay longer tend to answer questions in a narrower range of categories.

Acknowledgments

This work was supported in part by the following projects: Grants-in-Aid for Scientific Research (Nos. 24240013) from MEXT of Japan.

References

1. L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman. Knowledge sharing and yahoo answers: everyone knows something. In *Proceedings of the 17th international conference on World Wide Web*, pages 665–674. ACM, 2008.
2. N. Aikawa, T. Sakai, and H. Yamana. Community qa question classification: Is the asker looking for subjective answers or not? *IPSJ Online Transactions*, 4:160–168, 2011.
3. C. R. Anderson, P. Domingos, and D. S. Weld. Relational markov models and their application to adaptive web navigation. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 143–152. ACM, 2002.
4. L. Chen, D. Zhang, and L. Mark. Understanding user intent in community question answering. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 823–828. ACM, 2012.
5. T. S. Daisuke Ishikawa and N. Kando. Overview of the ntcir-8 community qa pilot task (part i): The test collection and the task. In *Proceedings of NTCIR-8 Workshop Meeting*, pages 421–432, 2010.
6. D. Dearman and K. N. Truong. Why users of yahoo!: answers do not answer questions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 329–332. ACM, 2010.

7. F. M. Harper, D. Moy, and J. A. Konstan. Facts or friends?: distinguishing informational and conversational questions in social q&a sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 759–768. ACM, 2009.
8. F. M. Harper, D. Raban, S. Rafaeli, and J. A. Konstan. Predictors of answer quality in online q&a sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 865–874. ACM, 2008.
9. A. Hassan, R. Jones, and K. L. Klinkner. Beyond dcg: user behavior as a predictor of a successful search. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 221–230. ACM, 2010.
10. T. Iwata, S. Watanabe, T. Yamada, and N. Ueda. Topic tracking model for analyzing consumer purchase behavior. In *Proceedings of the 21st international jont conference on Artifical intelligence*, pages 1427–1432. Morgan Kaufmann Publishers Inc., 2009.
11. S. Kim, J. S. Oh, and S. Oh. Best-answer selection criteria in a social q&a site from the user-oriented relevance perspective. *Proceedings of the American Society for Information Science and Technology*, 44(1):1–15, 2007.
12. U. Lee, Z. Liu, and J. Cho. Automatic identification of user goals in web search. In *Proceedings of the 14th international conference on World Wide Web*, pages 391–400. ACM, 2005.
13. B. Li and I. King. Routing questions to appropriate answerers in community question answering services. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1585–1588. ACM, 2010.
14. Q. Liu, E. Agichtein, G. Dror, Y. Maarek, and I. Szpektor. When web search fails, searchers become askers: understanding the transition. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 801–810. ACM, 2012.
15. X. Liu, W. B. Croft, and M. Koll. Finding experts in community-based question-answering services. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 315–316. ACM, 2005.
16. Y. Liu, J. Bian, and E. Agichtein. Predicting information seeker satisfaction in community question answering. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 483–490. ACM, 2008.
17. Y. Liu, S. Li, Y. Cao, C.-Y. Lin, D. Han, and Y. Yu. Understanding and summarizing answers in community-based question answering services. In *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, pages 497–504. Association for Computational Linguistics, 2008.
18. D. R. H. Miller, T. Leek, and R. M. Schwartz. A hidden markov model information retrieval system. In *Proceedings of the 22Nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '99, pages 214–221, New York, NY, USA, 1999. ACM.
19. M. R. Morris, J. Teevan, and K. Panovich. What do people ask their social networks, and why?: a survey study of status message q&a behavior. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 1739–1748. ACM, 2010.
20. K. K. Nam, M. S. Ackerman, and L. A. Adamic. Questions in, knowledge in?: a study of naver's question answering community. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 779–788. ACM, 2009.

21. A. Pal and J. A. Konstan. Expert identification in community question answering: exploring question selection bias. In *Proceedings of the 19th ACM international conference on Information and knowledge management*, pages 1505–1508. ACM, 2010.
22. F. Riahi, Z. Zolaktaf, M. Shafiei, and E. Milios. Finding expert users in community question answering. In *Proceedings of the 21st international conference companion on World Wide Web*, pages 791–798. ACM, 2012.
23. J. Yang, X. Wei, M. S. Ackerman, and L. A. Adamic. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. In *ICWSM*, 2010.