

An Analysis of Questions in a Q&A Site Resubmitted Based on Indications of Unclear Points of Original Questions

Masahiro Kojima and Yasuhiko Watanabe and Yoshihiro Okada

Ryukoku University, Seta, Otsu, Shiga, 520-2194, Japan

t10m101@mail.ryukoku.ac.jp, {watanabe, okada}@rins.ryukoku.ac.jp

Topics language processing, pragmatic dimensions

Abstract

In this study, we analyzed how answerers indicated unclear points in questions, and how questioners modified and resubmitted their questions based on indications of unclear points.

1 Introduction

In these days, many of us use question and answer (Q&A) sites where we share our problems and get solutions of them. For example, about 3.11 million questions were submitted to Yahoo! chiebukuro¹ from April/2004 to October/2005. Because of this large numbers of questions, questioners had better submit questions which give enough information to answerers. However, it is difficult to make good questions. For example, in Yahoo! chiebukuro, we often found unclear questions (e.g. Q1 in Figure 1) and their answers where answerers indicated unclear points of the questions (e.g. A1 in Figure 1).

(Q 1) I cannot access a web page which I could read yesterday. What should I do?

(A 1) Show URL.

In (A 1), the answerer pointed out that the questioner did not describe important information to answer the question: URL. Unclear questions may decrease chances of getting good answers. As a result, it is important to investigate supporting methods of making clear questions. One idea is to indicate unclear points of questions, as the questioner of (A 1) did. In order to obtain helpful knowledge and develop a help system for making clear questions, it is important to analyze

- how answerers indicated unclear points in questions, and

¹Yahoo Answers in Japan. <http://chiebukuro.yahoo.co.jp>

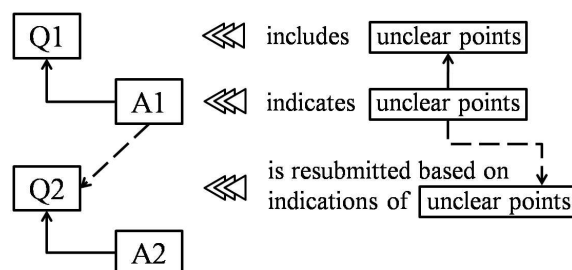


Figure 1: An example of a resubmitted question based on the indication of unclear points in the original question.

- how questioners modified and resubmitted their questions based on indications of unclear points in their original questions.

Our approach differs from previous analyses on Yahoo! Answers (Su et al., 2007) (Adamic et al., 2008). In this study, we used the data of Yahoo! chiebukuro for observation and examination. The data of Yahoo! chiebukuro was published by Yahoo! JAPAN via National Institute of Informatics in 2007². 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.

2 Types of indication of unclear points in questions and modification of questions

2.1 Types of indication of unclear points in questions

We observed answers submitted to PC category of Yahoo! chiebukuro and found the following five types of indication of unclear points in questions (Table 1).

TYPE (A1-1) Answerers wanted detailed accounts of what questioners did.

(Q 2) I got a warning from Symantec. How do I extend the software license.

²<http://research.nii.ac.jp/tdc/chiebukuro.html>

Table 1: Types of indication of unclear points in questions

TYPE	indication of unclear points in questions
A1-1	detailed accounts of what questioners did
A1-2	detailed accounts of what happened
A1-3	detailed accounts of conditions
A1-4	information other than (A1-1), (A1-2), and (A1-3)
A1-5	unhelpful solution

(A 2) Did you buy an extension key?

TYPE (A1-2) Answerers wanted detailed accounts of what happened.

(Q 3) When I try to maximize my IE window, it is positioned about 2cm below from the top of the screen! What should I do?

(A 3) What's there? blank?

TYPE (A1-3) Answerers wanted detailed accounts of conditions. For example, the indication of (A 1) is classified into this type.

TYPE (A1-4) Answerers wanted detailed accounts of information which were not asked in answers of TYPE (A1-1), (A1-2), and (A1-3).

(Q 4) I don't know the connection type. What should I do?

(A 4) Which connection type do you want to know?

TYPE (A1-5) Answerers submitted solutions, however, they were not helpful to solve questioners' problems. In these cases, answerers did not indicate unclear points of questions. However, we think these unhelpful solutions are one type of indication of unclear points of questions. This is because these unhelpful solutions often made questioners aware of unclear points of their questions. For example, the answerer of (A 5) showed one solution with detailed instruction. The questioner of (Q 5) tried to solve his/her problem according to the instruction and found the solution was unhelpful.

(Q 5) Windows XP crashed. How do I boot my computer?

(A 5) Just put a recovery disc into CD/DVD drive. And restart your PC.

(Q 6) Windows XP crashed. I put a recovery disc into CD drive, but my PC didn't work. How do I boot my computer?

Table 2: Types of modification of questions based on indication of unclear points

TYPE	modification of questions
Q2-1	added explanation based on the indication
Q2-2	added explanation based on other information
Q2-3	described the solution was unhelpful
Q2-4	asked about unknowns in the indication
Q2-5	resubmitted in disregard of the indication

2.2 Types of modification of questions

We observed resubmitted questions in PC category of Yahoo! chiebukuro and found the following five types of modification of questions (Table 2).

TYPE (Q2-1) Questioners added explanations based on the indications of unclear points to their questions and resubmitted them. In this type, it is likely that answerers asked about what questioners knew or could find out easily.

(Q 7) How do I reset my iMac to default settings, except IE and Outlook settings.

(A 7) Show the versions of OS, IE, and Outlook Express.

(Q 8) How do I reset my iMac to default settings, except IE and Outlook settings. My mac OS is version 9 and both IE and Outlook are version 5.

TYPE (Q2-2) Questioners added explanations based on information other than the indications to their questions and resubmitted them. In this type, it is also likely that answerers asked about what questioners knew or could find out easily.

(Q 9) Can I boot my XP PC with Windows 98 HDD?

(A 9) Did you install 98 first? You cannot boot your PC by using Windows 98 if your primary OS is XP.

(Q 10) Can I boot my XP PC with Windows 98 HDD? I don't want to set up a dual boot system. I want to know my PC gets in trouble when I boot it with Win 98 HDD.

TYPE (Q2-3) In resubmitted questions, questioners described that solutions received from answerers were unhelpful to solve their questions. For example, in (Q 6), the questioner described the solution received from the answerer of (A 5) was unhelpful to solve his/her problem. In this type, it

is likely that answerers showed one solution which questioners did not know.

TYPE (Q2-4) In resubmitted questions, questioners asked about unknown points in the indication received from answerers. For example, the questioner of (Q 11) received one solution and got the key to solve his/her problem: module deletion. However, he/she did not know it and submitted (Q 12) for requesting detailed information about it.

(Q 11) I removed all macros from my excel file. Then, whenever I open the file, excel asks me if I want to enable/disable macros. How do I stop it?

(A 11) Open visual basic editor and delete the module.

(Q 12) I removed all macros from my excel file. Then, whenever I open the file, excel asks me if I want to enable/disable macros. I want to stop it. The module should be deleted by using visual basic editor. But, how do I do it?

TYPE (Q2-5) Questioners resubmitted almost the same questions as they had. They did not mention any kinds of information received from answerers. For example, in (Q 14), the questioner did not mention any kinds of information described in (A 14) although he/she selected (A 14) as a best answer.

(Q 13) My optical mouse is faulty. The cursor sometimes freezes. Is it end of life?

(A 13) Look the back side and remove dust gathered around the red light.

(Q 14) My optical mouse is faulty. The cursor sometimes freezes. Is it end of life?

3 Extraction of original and resubmitted questions and their answers from Yahoo! chiebukuro

We intended to extract

- original questions which included unclear points (e.g. Q1 in Figure 1),
- answers which indicated unclear points in the original questions (e.g. A1 in Figure 1),
- resubmitted questions based on the indications of unclear points in the original questions (e.g. Q2 in Figure 1), and
- answers to the resubmitted questions

from PC category of Yahoo! chiebukuro in the next way.

step 1 extract an answer which indicated unclear points in a question (e.g. A1 in Figure 1). This kind of answer can be extracted by using a method based on machine learning techniques (Isogai et al., 2009).

step 2 extract the question which had the answer extracted in step 1 (e.g. Q1 in Figure 1). This question is regarded as an original question.

step 3 extract the first question submitted by the questioner after he/she received the answer extracted in step 1.

step 4 examine whether the questions extracted in step 1 and step 3 met one of the following conditions:

- they shared more than 10 content words when both of them consisted of more than 20 content words, or
- they shared more than 5 content words.

When one of the conditions was satisfied, the question extracted in step 3 is regarded as a resubmitted question (e.g. Q2 in Figure 1).

step 5 extract the answers to the resubmitted question extracted in step 4 (e.g. A2 in Figure 1).

4 Experimental results

We applied our method described in section 3 to 171848 questions and 474687 answers which were submitted to PC category of Yahoo! chiebukuro from April/2004 to October/2005, and extracted 4271 cases of questions and their answers. Among them, we selected 200 cases randomly and found 133 cases of them where

- an original question (e.g. Q1 in Figure 1),
- the answer which indicated unclear points in the original questions (e.g. A1 in Figure 1),
- the resubmitted questions based on the indication of unclear points in the original questions (e.g. Q2 in Figure 1), and
- the answers to the resubmitted questions (e.g. A2 in Figure 1)

were extracted adequately. We observed these 133 cases and show

Table 3: The results of the analyses: (1) how answerers indicated unclear points in original questions (A1-1, 2, 3, 4, and 5), (2) how questioners utilized indications in answers when they resubmitted their questions (Q2-1, 2, 3, 4, and 5). The numbers in parentheses are the numbers of best answers.

(a) 122 cases where questioners obtained good answers.

	A1-1	A1-2	A1-3	A1-4	A1-5	total
Q2-1	5 (5)	15 (14)	21 (16)	4 (4)	10 (10)	55 (49)
Q2-2	0 (0)	1 (1)	1 (1)	7 (6)	6 (4)	15 (12)
Q2-3	1 (1)	0 (0)	0 (0)	0 (0)	18 (14)	19 (15)
Q2-4	1 (1)	4 (3)	0 (0)	0 (0)	12 (10)	17 (14)
Q2-5	0 (0)	4 (4)	1 (1)	1 (1)	10 (8)	16 (14)
total	7 (7)	24 (22)	23 (18)	12 (11)	56 (46)	122 (104)

(b) 11 cases where questioners did not obtain good answers.

	A1-1	A1-2	A1-3	A1-4	A1-5	total
Q2-1	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Q2-2	0 (0)	0 (0)	0 (0)	1 (1)	0 (0)	1 (1)
Q2-3	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Q2-4	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Q2-5	0 (0)	2 (1)	2 (1)	1 (1)	5 (4)	10 (7)
total	0 (0)	2 (1)	2 (1)	2 (2)	5 (4)	11 (8)

1. whether questioners obtained good solutions or useful clues by resubmitting their questions, especially, when they added explanations based on indications to their questions.
2. how answerers indicated unclear points in questions.
3. whether questioners gave good evaluations to answerers who indicated unclear points of their questions.
4. how questioners utilized indications of unclear points of their original questions when they resubmitted their questions.

Table 3 shows the results of these analyses.

First, we examined whether questioners obtained good solutions or useful clues by resubmitting their questions. As shown in Table 3, there were 26 cases where questioners resubmitted TYPE (Q2-5) questions (questions resubmitted in disregard of indications), and then, 16 of these 26 cases where questioners obtained good solutions or useful clues. On the other hand, there were 107 cases where questioners resubmitted TYPE (Q2-1), (Q2-2), (Q2-3), and (Q2-4) questions, in other words, they accepted indications from answerers and modified their questions. Then, there were 106 of these 107 cases where questioners obtained good solutions or useful clues. As a result, questioners can increase their chances to obtain good solutions or useful clues when they accepted indications from answerers and modified their questions.

Secondly, we examined how answerers indicated unclear points in questions. As shown in Table 3, TYPE (A1-5) answer (unhelpful solution) was the most common answer. As a result, questioners had chances to recognize unclear points in their questions even if they received unhelpful solutions.

Thirdly, we examined whether questioners gave good evaluations to answerers who indicated unclear points of their questions. As shown in Table 3, there were 112 cases (84%) where questioners gave good evaluations to answerers who indicated unclear points of their questions. On the other hand, in PC category of Yahoo! chiebukuro, 474687 answers were submitted and 171848 of them (36%) were received good evaluations. As a result, many questioners gave good evaluations to answerers who indicated unclear points of their questions.

Finally, we examined how questioners utilized indications of unclear points of their original questions when they resubmitted their questions. In this experiment, there were 107 cases where questioners resubmitted TYPE (Q2-1), (Q2-2), (Q2-3), and (Q2-4) questions, in other words, they accepted indications from answerers and modified their questions. Then, there were 17 of these 107 cases where questioners resubmitted TYPE (Q2-4) questions, in other words, they had unknown points in indications received from answerers and needed to ask about what they were. In other 90 cases, questioners resubmitted TYPE (Q2-1), (Q2-2), and (Q2-3) questions, in other words, they knew or could find out easily what answerers indicated. As a result, just to indicate unclear points in questions is useful to make good questions.

References

- Su,Q., Pavlov,D., Chow,J., and Baker,W.: Internet-scale collection of human-reviewed data, WWW2007, (2007).
- Adamic,L., Zhang,J., Bakshy,E., and Ackerman,M.: Knowledge Sharing and Yahoo Answers: Everyone Knows Something, WWW2008, (2008).
- Isogai,N., Nishimura,R., Watanabe,Y., and Okada,Y.: Information Extraction for Supporting a Learner's Efforts to Recognize What the Learner did not Understand, CSEDU 2009, (2009).