

Collecting and Classifying Examples of Consumer Troubles on “Contract and Cancellation” in a Question-Answer Site

Shota Arai[†] Tian Nie[†] Takehito Utsuro[‡]

Yasuhide Kawada*

Noriko Kando**

[†]Graduate School of Systems and Information Engineering, University of Tsukuba

[‡]Faculty of Engineering, Information and Systems, University of Tsukuba

*Logworks Co., Ltd.

**National Institute of Informatics

Abstract

In Japan, there exist plenty of consumer troubles caused by crimes such as fraud and illegal business operation these days. The National Consumer Affairs Center of Japan (NCAC) collects typical examples of consumer trouble cases and makes their detailed information publicly available via the Internet, where the detailed information includes questions from the consumers, how to solve the case, and the actual result of solving the case. However, on the NCAC Web site, the number of consumer trouble examples publicly available is quite limited. This paper, on the other hand, focuses on question-answer sites as another information source for collecting plenty of examples of such consumer troubles. We propose how to collect questions and answers on consumer troubles regarding cancelling certain contracts from a question-answer site. We actually collect more than 2,000 candidates of questions and answers on consumer troubles, and then manually analyze and classify them into four cases including real consumer troubles and the cases where it is difficult to legally judge that the service / product provider is liable for the consumer's claim. This paper reports detailed statistics of examining those question-answer examples on consumer troubles.

Keywords: Question-Answer Site, Consumer Troubles, Question-Answer Examples,

1 Introduction

In Japan, there exist plenty of consumer troubles caused by crimes such as fraud and illegal business operation these days. The National Consumer Affairs Center of Japan (NCAC)¹ collects typical examples of consumer trouble cases and

makes their detailed information publicly available via the Internet, where the detailed information includes questions from the consumers, how to solve the case, and the actual result of solving the case. However, on the NCAC Web site, the number of consumer trouble examples publicly available is quite limited.

This paper, on the other hand, focuses on question-answer sites such as Yahoo! Japan Chiebukuro² and Oshiete! goo³ as another information source for collecting plenty of examples of such consumer troubles. In question-answer sites, questions regarding such consumer troubles are increasingly posted, and one can also find how to solve those troubles as answers to those questions. Considering such situations, this paper aims at assisting those who are engaged in any consumer trouble to solve their cases by automatically discriminating questions that are closely related to real consumer troubles and those that are not related to any real consumer troubles.

More specifically, this paper first analyzes 219 typical examples of consumer troubles that are available from the Web site of NCAC. We then classify those examples into the following two types: (1) the cases where the consumer hopes to cancel a certain contract, while the service / product provider does not accept the cancellation, and (2) other cases. This paper concentrates on consumer troubles of the first type mainly because more than half of those 219 typical examples are of this type. Another reason why we concentrate on this type is that it is relatively easy to collect text data of question-answer examples of this type, since question-answer examples written in Japanese tend to include both of the query terms “契約” (contract) and “解約”

¹http://www.kokusen.go.jp/ncac_index.html

²<http://chiebukuro.yahoo.co.jp/>

³<http://oshiete.goo.ne.jp/>

(cancellation) in their question part. By considering this fact, we then propose how to collect questions and answers on consumer troubles regarding cancelling certain contracts from the Yahoo! Japan Chiebukuro question-answer site. Based on the categories in Yahoo! Japan Chiebukuro, we construct our own list of categories for classifying consumer troubles. We further collect more than 2,000 candidates of questions and answers on consumer troubles regarding cancelling certain contracts from Yahoo! Japan Chiebukuro, and then manually analyze and classify them into the following four cases: (i) real consumer troubles where the service / product provider is liable for the consumer's claim, (ii) the cases where the consumer claims that he/she faces with a consumer trouble, although it is difficult to legally judge that the service / product provider is liable for the consumer's claim, (iii) an issue other than consumer troubles, and (iv) not a trouble but just a simple question. This paper reports detailed statistics of examining those question-answer examples on consumer troubles.

2 Reports on Consumer Trouble Examples from Consumer Centers

The National Consumer Affairs Center of Japan (NCAC) is an independent administrative agency of Japan and works as a core consumer advocate organization in accordance with the Consumer Affairs Fundamental Act of Japan. NCAC handles consumer issues in collaboration with the government and local consumer centers located throughout Japan. NCAC collects typical examples of consumer trouble cases and makes their detailed information publicly available via the Internet, where the detailed information includes questions from the consumers, how to solve the case, and the actual result of solving the case.

We collect those typical examples from the Web site of NCAC, which amount to 219 examples. We first manually analyze those 219 examples and then found that they can be roughly classified into the following two types:

- Typical consumer troubles where the consumer hopes to *cancel a certain contract*, while the service / product provider does not accept the cancellation (henceforth referred to as “contract and cancellation”). The number of consumer trouble examples

of this type is 124.

- Consumer troubles other than the cases where the consumer hopes to cancel a certain contract (henceforth referred to as other than “contract and cancellation”). The number of consumer trouble examples of this type is 95.

Next, we further classify those 219 examples according to the domain of the consumer troubles into the 26 categories⁴ listed in Table 1. As we introduced in the previous section, we focus on consumer troubles of the “contract and cancellation” type and collect question answer examples from Yahoo! Japan Chiebukuro. Here, out of the 20 categories in Table 1 for which at least one example of the consumer troubles of “contract and cancellation” type is listed in the NCAC Web site, we ignore three categories (ID = 18, 19, 20). This is simply because, for each of the three categories, Yahoo! Japan Chiebukuro includes just a small number of question-answer examples having both of the query terms “契約” (contract) and “解約” (cancellation) in their question part.

3 Collecting and Manually Analyzing Examples of Consumer Troubles on “Contract and Cancellation” in a Question-Answer Site

This section describes the procedure and the result of collecting and manually analyzing examples of consumer troubles on “contract and cancellation” from Yahoo! Japan Chiebukuro.

3.1 Question-Answer Examples collected in Yahoo! Japan Chiebukuro

Yahoo! Japan Chiebukuro is known as one of the most popular Japanese question-answer sites. As of August 12th, 2013, the total number of questions recorded in Yahoo! Japan Chiebukuro is 111,692,915, while that of answers is 257,851,479. The text data of questions and answers of Yahoo! Japan Chiebukuro is available for academic purposes. In this paper, we use one of those questions and answers text data sets that are published from Yahoo! Japan Chiebukuro for academic purposes. More specifically, we use the text data of

⁴In the Web site of NCAC, they categorize those 219 examples into their own 13 categories, while we refine those 13 categories into finer grained 26 categories.

Table 1. Classification of Consumer Trouble reported in the National Consumer Affairs Center of Japan

ID	Category (proposed in this paper)	# of troubles on “contract and cancellation”	# of troubles other than “contract and cancellation”	Total
1	at-home sales, catch sales	30	6	36
2	sales of learning materials	21	1	22
3	esthetic clinic	15	3	18
4	CATV/Internet and home utilities	13	9	22
5	fraud of part-time job contract	10	1	11
6	charged site	7	0	7
7	pyramid scheme	6	1	7
8	buying / selling real estate	4	7	11
9	investment	3	3	6
10	insurance	3	2	5
11	real-estate rental	2	0	2
12	lease (car, telephone, etc.)	2	0	2
13	credit card	1	4	5
14	cell phone	1	3	4
15	loan	1	2	3
16	newspaper subscription	1	0	1
17	sales of PC	1	0	1
18	travel agency and accommodation	1	7	8
19	insertion into a who’s who	1	0	1
20	gamble	1	0	1
21	defective product	0	26	26
22	individual information	0	12	12
23	auction	0	5	5
24	moving	0	1	1
25	elevator in a department store	0	1	1
26	college / university tuition	0	1	1
—	Total	124	95	219

questions and answers for five years from April 1st, 2004 to April 7th, 2009, which amounts to 16,257,413 questions and 50,053,894 answers. Those question-answer text data are classified according to categories constituting a tree structure hierarchy. The tree structure hierarchy of categories has three levels, and it has 16 categories at the topmost level, 108 at the second level, and 348 at the bottom level, which amount to 472 categories in total. When a person posts a question to the Web site of Yahoo! Japan Chiebukuro, he/she chooses one of those 472 categories, then the posted question as well as answers to the question are allocated to the chosen category.

3.2 The Procedure of Collecting Examples of Consumer Troubles on “Contract and Cancellation”

Table 2 describes the current status of our work on collecting question answer examples of consumer troubles from Yahoo! Japan Chiebukuro. The structure of Table 2 is as follows:

- We roughly classify consumer troubles into (a) “contract and cancellation” type, and (b) the rest, and we focus on the type (a).
- We roughly classify question answer examples of Yahoo! Japan Chiebukuro into (c) those including both of the query terms “契約” (contract) and “解約” (cancellation) in their question part, and (d) the rest, and we collect questions-answers examples of the type (c) in this paper.

Table 2. Collecting Examples of Consumer Troubles from Yahoo! Japan Chiebukuro: Current Status of This Paper

classification of question examples	current status	classification of consumer troubles	
		consumer troubles on “contract and cancellation” (type (a))	consumer troubles other than “contract and cancellation” (type (b))
questions including “契約” (contract) and “解約” (cancellation) (type (c))	out of the 10,459 examples in total, 2,547 examples corresponding to 17 categories (proposed in this paper) in Table 1 (ID=1~17) are randomly selected from 6,000 candidates and are manually analyzed.	for the 17 categories (proposed in this paper) in Table 1 (ID=1~17), estimated numbers of consumer troubles (service provider is liable / service provider is NOT liable) is 485 / 1,181 (details to be found in Table 5)	for the 17 categories (proposed in this paper) in Table 1 (ID=1~17), estimated numbers (service provider is liable + service provider is NOT liable + an issue other than consumer troubles + simple question) is 2,427 (details to be found in Table 5), while their further classification is not manually annotated yet
other examples (type (d))	not examined yet in this paper	Future Work: Step 1. list up the categories of Yahoo! Japan Chiebukuro from which candidates of consumer trouble examples are to be collected. Step 2. randomly select samples of question examples from each category. Step 3. manually analyze those selected samples.	

First, the number of type (c) questions in Yahoo! Japan Chiebukuro is 10,459 in total. Out of those 10,459 question examples, we focus on those corresponding to 17 categories (proposed in this paper) in Table 1 (ID=1~17). To do this, we manually examine 348 categories at the bottom level of Yahoo! Japan Chiebukuro and allocate each of the 348 Yahoo! Japan Chiebukuro categories to one or more of the 17 categories (proposed in this paper) in Table 1 (ID=1~17). For each of our 17 categories, Table 3 lists allocated Yahoo! Japan Chiebukuro categories, which amount to 16 categories in total (as listed in Table 4). Out of the 16 categories, the bold-faced three categories *consumer affairs*, *legal consultation*, and *real estate* have characteristics which differ from the other 13, in that each of the three categories is allocated to more than half of our 17 categories⁵. This is mainly be-

⁵As shown in the column of “# of corresponding categories (proposed in this paper) in Table 3” of Table 4, 15 for each of *consumer affairs* and *legal consultation* and 9 for *real estate*.

cause, in Yahoo! Japan Chiebukuro, categories are manually assigned to each question by the person who posts the question and the criterion on how to assign a category to a question is not controlled at all by the administrator of Yahoo! Japan Chiebukuro.

Next, in Table 4, for each of the 16 Yahoo! Japan Chiebukuro categories, we list the number of type (c) questions including both of the query terms “契約” (contract) and “解約” (cancellation) in their question part, which amount to 6,000 examples in total. Among those 6,000 examples, those with the three categories *consumer affairs*, *legal consultation*, and *real estate* consist of the mixture of the question examples of at most 15 categories. Thus, we manually analyze all of those question examples according to the criterion to be presented in the next section. For each of the remaining 13 categories, on the other hand, we randomly select at most 20 examples, and then manually analyze them. Finally, the number of question examples we manually

Table 3. Allocating Categories in Yahoo! Japan Chiebukuro to Our Categories

ID	Category (proposed in this paper)	Categories of Yahoo! Japan Chiebukuro
1	at-home sales, catch sales	consumer affairs / legal consultation
2	sales of learning materials	consumer affairs / legal consultation
3	esthetic clinic	consumer affairs / legal consultation / esthetic and hair removal clinic
4	CATV/Internet and home utilities	consumer affairs / legal consultation / real estate / new construction of residence / PC
5	fraud of part-time job contract	consumer affairs / legal consultation
6	charged site	consumer affairs / legal consultation
7	pyramid scheme	consumer affairs / legal consultation / real estate
8	buying / selling real estate	consumer affairs / legal consultation / real estate / new construction of residence
9	investment	stock / investment trust fund / consumer affairs / legal consultation
10	insurance	insurance / consumer affairs / legal consultation / real estate / car / new construction of residence
11	real-estate rental	consumer affairs / legal consultation / real estate
12	lease (car, telephone, etc.)	consumer affairs / legal consultation / real estate / car
13	credit card	credit card / consumer affairs / legal consultation / real estate
14	cell phone	cell phone (au / NTT docomo / Softbank) / consumer affairs / legal consultation / real estate
15	loan	loan / new construction of residence
16	newspaper subscription	consumer affairs / legal consultation / real estate / savings
17	sales of PC	PC

analyze in this paper amounts to 2,547 in total as shown in in Table 4.

3.3 Manually Analyzing and Estimating the Number of Consumer Troubles on “Contract and Cancellation”

We manually analyze the 2,547 question examples according to the following criterion:

- We classify all of the question examples into (a) “contract and cancellation” type, and (b) the rest⁶.

The result of this manual classification is shown in Table 5, where the boldfaced numbers are not estimated but real numbers counted after manually analyzing all the examples, while other numbers without boldfaced fonts are those estimated after manually analyzing samples randomly selected from the whole examples. Here, given c

⁶For each of the two types (a) and (b), the question examples include all of the four types “service provider is liable”, “service provider is NOT liable”, “an issue other than consumer troubles”, and “simple question” to be described in the following.

as one of our 17 categories and $z (\in \{a, b\})$ as one of the two types (a) and (b), the number of consumer troubles is estimated according to the following formula as $\bar{N}(c, z)$:

$$\bar{N}(c, z) = \sum_y \left(N(y) \times \frac{n(y, c)}{\sum_{c'} n(y, c')} \right) \times \frac{n(y, c, z)}{\sum_{z' \in \{a, b\}} n(y, c, z')}$$

where y denotes one of the 16 categories of Yahoo! Japan Chiebukuro, c' denotes one of our 17 categories, $N(y)$ denotes, for the category y , the number of questions including both of the query terms “契約” (contract) and “解約” (cancellation) in their question part as listed in Table 4, $n(y, c')$ denotes the number of questions which belong to our category c' out of the manually analyzed samples in the category y of Yahoo! Japan Chiebukuro, $z' (\in \{a, b\})$ denotes one of the two types (a) and (b), and $n(y, c, z')$ denotes the number of questions which belong to the type

Table 4. # of Questions for Each Category in Yahoo! Japan Chiebukuro

ID	Category of Yahoo! Japan Chiebukuro	# of corresponding categories (proposed in this paper) in Table 3	Total # of Questions	# of Questions including “契約” (contract) and “解約” (cancellation)	# of Manually Analyzed Questions
1	insurance	1	60,025	1,068	20
2	real estate	9	88,659	866	866
3	cell phone (Softbank)	1	45,827	854	20
4	legal consultation	15	170,929	747	747
5	cell phone (NTT docomo)	1	63,701	717	20
6	consumer affairs	15	32,861	687	687
7	cell phone (au)	1	31,565	478	20
8	loan	1	34,133	167	20
9	new construction of residence	4	27,511	103	20
10	credit card	1	16,542	70	20
11	car	2	225,035	63	20
12	esthetic and hair removal clinic	1	11,387	57	20
13	savings	1	28,393	51	20
14	PC	2	292,657	42	20
15	investment trust fund	1	8,909	23	20
16	stock	1	59,357	7	7
—	Total	—	—	6,000	2,547

z' out of the $n(y, c)$ questions. As summarized in Table 5, estimated numbers of consumer troubles are 2,964 for the type (a) and 2,427 for the type (b).

Furthermore, the question examples of type (a) can be further classified into the following four types, where the result of this manual classification is also shown in Table 5:

“service provider is liable” This type is a true consumer trouble⁷. For example, if the service / product provider does not follow the procedure clearly written in the contract, then it is judged as “service provider is liable”. Or, as in the case where the service / product provider is coercive so that the consumer, against his/her own intent, makes the contract, if the service / product provider’s conduct or statement is unethical, then it is also judged as “service provider is liable”.

⁷Exact judgment whether each case is a true consumer trouble where the service / product provider is liable for the consumer’s claim or not is mainly based on the judgments given in the answers to the question.

“service provider is NOT liable” In the case of the questions of this type, the consumer claims that he/she faces with a consumer trouble, although it is difficult to legally judge that the service / product provider is liable for the consumer’s claim. For example, if the service / product provider exactly follows the procedure clearly written in the contract, and it is said in the contract that the consumer has to pay penalty charges or that he/she has to carry out the contract, then it is judged as “service provider is NOT liable”.

“an issue other than consumer troubles” If the focus of the consumer’s question is not on consumer troubles, but it is on other troubles, then it is judged as “an issue other than consumer troubles”. Examples of this type include a case of “cell phone” category, where the question is on the troubles in manipulating cell phones.

“simple question” If the focus of the consumer’s question is not on consumer troubles, but it is just a simple question, then it

Table 5. Analysis of Consumer Troubles on “Contract and Cancellation”

ID	Category (proposed in this paper)	# of troubles: “contract and cancellation” (type (a)) / other than “contract and cancellation” (type (b))	Rates (%) among troubles on “contract and cancellation”: (type (a)) service provider is liable / service provider is NOT liable / an issue other than consumer troubles / simple question
1	at-home sales, catch sales	66 / 15	33.3 / 66.6 / 0 / 0
2	sales of learning materials	29 / 12	31.0 / 69.0 / 0 / 0
3	esthetic clinic	116 / 20	21.6 / 73.5 / 0 / 4.9
4	CATV/Internet and home utilities	91 / 73	13.2 / 86.8 / 0 / 0
5	fraud of part-time job contract	11 / 39	27.3 / 72.7 / 0 / 0
6	charged site	21 / 30	90.5 / 9.5 / 0 / 0
7	pyramid scheme	5 / 2	80.0 / 20.0 / 0 / 0
8	buying / selling real estate	196 / 62	22.8 / 66.7 / 7.9 / 2.6
9	investment	15 / 15	20.0 / 40.0 / 13.3 / 26.7
10	insurance	465 / 702	24.2 / 29.2 / 0.7 / 45.9
11	real-estate rental	477 / 411	11.1 / 88.9 / 0 / 0
12	lease (car, telephone, etc.)	22 / 7	22.6 / 77.4 / 0 / 0
13	credit card	51 / 130	15.7 / 76.5 / 1.9 / 5.9
14	cell phone	1,176 / 842	7.8 / 4.2 / 17.4 / 70.6
15	loan	13 / 17	62.0 / 22.8 / 0 / 15.2
16	newspaper subscription	199 / 22	31.7 / 68.3 / 0 / 0
17	sales of PC	11 / 28	20.0 / 0 / 0 / 80.0
—	Total	2,964 / 2,427	16.4 / 39.8 / 7.6 / 36.2

Boldfaced numbers and ratios are not estimated but real numbers counted after manually analyzing all the examples, while other numbers and ratios without boldfaced fonts are those estimated after manually analyzing samples randomly selected from the whole examples.

is judged as “simple question”. Examples of this type include a case of “insurance” category, where the question is just a simple question on an insurance product.

As summarized in Table 2, their estimated numbers are 485 for “service provider is liable” and 1,181 for “service provider is NOT liable”.

Finally, from the Yahoo! Japan Chiebukuro categories *consumer affairs*, *legal consultation*, and *real estate*, we collect 199 examples classified into “service provider is liable”, i.e., true consumer troubles, and then classify them into the three types shown in Table 6. As can be seen from this result, about 70% of true consumer trouble cases are those where the service / product provider violates a certain contract. The re-

maining cases are roughly classified into frauds and unethical acts.

4 Related Work

Before studying the issue of consumer troubles, we previously studied the domain of crime, where we proposed a framework of extracting people’s concerns and reports on crimes in their own blogs [1]. Among various domains of crime, we especially studied “*fraud*” and “*Internet crime*”. We focused on experts in crime domain and address the issue of extracting concerns on crimes such as tips for preventing being victimized. We also focused on victims of criminal acts, and address the issue of extracting reports on being victimized. We proposed

Table 6. Classification of Examples of “service provider is liable” in Consumer Troubles on “Contract and Cancellation”

classification	# of examples (rate (%))
<u>Type (1)</u> The service provider, who violates the contract, is in violation of the Civil Code or of the Commercial Code.	143 (71.9)
<u>Type (2)</u> The service provider is a fraud or a scam.	31 (15.6)
<u>Type (3)</u> Other than types (1) and (2), but the service provider’s conduct or statement is unethical and inappropriate under normal social conventions, such as those being coercive so that the consumer, against his/her own intent, made the contract.	25 (12.5)
total	199 (100)

a technique which is based on detecting linguistic expressions representing experiences of being a victim of certain fraud, such as “*being a victim of*” and “*deceived*”. We incorporated dependency relations for identifying victims’ blog posts and showed its effectiveness. We are planning to develop similar techniques in our future task of identifying question examples of consumer troubles on “contract and cancellation”.

De Saeger et al. [2] and Torisawa et al. [3] studied how to collect linguistic expressions which represent trouble situation, where Web documents including writer’s own trouble experiences such as blogs are used for evaluation. Their technique itself as well as the collected expressions representing trouble situation can be applicable to our future task of identifying question examples of consumer troubles on “contract and cancellation”. Yangarber et al. [4] studied how to combine reports on epidemic threats from over 1,000 portals in 32 languages. Bautin et al. [5] studied how to analyze sentiment distribution in news articles across 9 languages.

5 Conclusion

This paper focused on question-answer sites as an information source for collecting plenty of examples of consumer troubles. We actually collected more than 2,000 candidates of questions and answers on consumer troubles, and then manually analyzed and classified them into four cases including real consumer troubles and the cases where it is difficult to legally judge that the service / product provider is liable for the consumer’s claim.

As we described in Table 2, our future work roughly consists of collecting and manually analyzing consumer troubles of the following types: (i) consumer troubles other than “contract and cancellation” (type (b)), and (ii) question-answer examples other than those including both of the query terms “契約” (contract) and “解約” (cancellation) in their question part (type (d)). Another future work is to develop a framework for assisting those who are engaged in a consumer trouble to solve their cases by automatically collecting questions and answers that are closely related to the users’ consumer trouble cases.

References

- [1] Y. Abe, T. Utsuro, Y. Kawada, T. Fukuhara, N. Kando, M. Yoshioka, H. Nakagawa, Y. Kiyota, and M. Tsuchiya. Extracting concerns and reports on crimes in blogs. In *6th AMT*, volume 6335 of *LNCS*, pages 498–509. Springer, 2010.
- [2] S. De Saeger, K. Torisawa, and J. Kazama. Looking for trouble. In *Proc. 22nd COLING*, pages 185–192, 2008.
- [3] K. Torisawa, S. De Saeger, Y. Kakizawa, J. Kazama, M. Murata, D. Noguchi, and A. Sumida. TORISHIKI-KAI, an Autogenerated Web Search Directory. In *Proc. 2nd ISUC*, pages 179–186, 2008.
- [4] R. Yangarber, C. Best, P. von Etter, F. Fuart, D. Horby, and R. Steinberger. Combining Information about Epidemic Threats from Multiple Sources. In *Proc. Workshop: Multi-source, Multilingual Information Extraction and Summarization*, pages 41–48, 2007.
- [5] M. Bautin, L. Vijayarenu, and S. Skiena. International Sentiment Analysis for News and Blogs. In *Proc. ICWSM*, pages 19–26, 2008.