

A Question Answer System Based on Confirmed Knowledge Developed by Using Mails Posted to a Mailing List

Ryo Nishimura
Ryukoku University
Seta, Otsu, Shiga,
520-2194, Japan
t030461a@ryukoku-u.jp

Yasuhiko Watanabe
Ryukoku University
Seta, Otsu, Shiga,
520-2194, Japan
watanabe@rins.ryukoku.ac.jp

Yoshihiro Okada
Ryukoku University
Seta, Otsu, Shiga,
520-2194, Japan
okada@rins.ryukoku.ac.jp

Abstract

In this paper, we report a QA system which can answer how type questions based on the confirmed knowledge base which was developed by using mails posted to a mailing list. We first discuss a problem of developing a knowledge base by using natural language documents: wrong information in natural language documents. Then, we describe a method of detecting wrong information in mails posted to a mailing list and developing a knowledge base by using these mails. Finally, we show that question and answer mails posted to a mailing list can be used as a knowledge base for a QA system.

1 Introduction

Because of the improvement of NLP, research activities which utilize natural language documents as a knowledge base become popular, such as QA track on TREC (TREC) and NTCIR (NTCIR). However, there are a few QA systems which assumed the user model where the user asks how type question, in other words, how to do something and how to cope with some problem (Kuro 00) (Kiyota 02) (Mihara 05). This is because we have several difficulties in developing a QA system which answers how type questions. We focus attention to two problems below.

First problem is the difficulty of extracting evidential sentences. It is difficult to extract evidential sentences for answering how type questions only by using linguistic clues, such as, common content

words and phrases. To solve this problem, (Kuro 00) and (Kiyota 02) proposed methods of collecting knowledge for answering questions from FAQ documents and technical manuals by using the document structure, such as, a dictionary-like structure and if-then format description. However, these kinds of documents requires the considerable cost of developing and maintenance. As a result, it is important to investigate a method of extracting evidential sentences for answering how type questions from natural language documents at low cost. To solve this problem, (Watanabe 04) proposed a method of developing a knowledge base by using mails posted to a mailing list (ML). We have the following advantages when we develop knowledge base by using mails posted to a mailing list.

- it is easy to collect question and answer mails in a specific domain, and
- there is some expectation that information is updated by participants

Next problem is wrong information. It is almost inevitable that natural language documents, especially web documents, contain wrong information. For example, (DA1-1) is opposed by (QR1-1-1).

(Q1) How I set up my wheel mouse for the netscape navigator?

(DA1-1) You can find a setup guide in the Dec. issue of SD magazine.

(QR1-1-1) I can not use it although I modified /usr/lib/netscape/ja/Netscape according to the guide.

Wrong information is a central problem of developing a knowledge base by using natural language

documents. As a result, it is important to investigate a method of detecting and correcting wrong information in natural language documents. (Watanabe 05) reported a method of detecting wrong information in question and answer mails posted to a mailing list. In (Watanabe 05), wrong information in the mails can be detected by using mails which ML participants submitted for correcting wrong information in the previous mails. Then, the system gives one of the following confirmation labels to each set of question and their answer mails:

positive label shows the information described in a set of a question and its answer mail is confirmed by the following mails,

negative label shows the information is opposed by the following mails, and

other label shows the information is not yet confirmed.

Our knowledge base, on which our QA system bases, is composed of these labeled sets of a question and its answer mail. Finally, we describe a QA system: It finds question mails which are similar to user's question and shows the results to the user. The similarity between user's question and a question mail is calculated by matching of user's question and the significant sentence extracted from the question mail. A user can easily choose and access information for solving problems by using the significant sentences and these confirmation labels.

2 Confirmed knowledge base developed by using mails posted to a mailing list

There are mailing lists to which question and answer mails are posted frequently. For example, in Vine Users ML, several kinds of question and answer mails are posted by participants who are interested in Vine Linux ¹. We reported that mails posted to these kinds of mailing lists have the following features.

1. Answer mails can be classified into three types: (1) direct answer (DA) mail, (2) questioner's reply (QR) mail, and (3) the others. Direct answer mails are direct answers to the original question. Questioner's reply mails

¹Vine Linux is a linux distribution with a customized Japanese environment.

are questioner's answers to the direct answer mails.

2. Question and answer mails do not have a firm structure because questions and their answers are described in various ways. Because of no firm structure, it is difficult to extract precise information from mails posted to a mailing list in the same way as (Kuro 00) and (Kiyota 02) did.
3. A mail posted to ML generally has a significant sentence. For example, a significant sentence of a question mail has the following features:
 - (a) it often includes nouns and unregistered words which are used in the mail subject.
 - (b) it is often quoted in the answer mails.
 - (c) it often includes the typical expressions, such as,

(ga / shikashi (but / however)) + ... + mashita / masen / shouka / imasu (can / cannot / whether / current situation is) + .

(ex) *Bluefish de nihongo font ga hyouji deki masen.* (I cannot see Japanese fonts on Bluefish.)
 - (d) it often occurs near the beginning.

Taking account of these features, (Watanabe 05) proposed a method of extracting significant sentences from question mails, their DA mails, and QR mails by using surface clues. Furthermore, (Watanabe 05) proposed a method of detecting wrong information in a set of a question mail and its DA mail by using the QR mail.

For evaluating our method, (Watanabe 05) selected 100 examples of question mails in Vine Users ML. They have 121 DA mails. Each set of the question and their DA mails has one QR mail.

First, we examined whether the results of determining the confirmation labels were good or not. The results are shown in Table 1. Table 2 shows the type and number of incorrect confirmation. The reasons of the failures were as follows:

- there were many significant sentences which did not include the clue expressions.
- there were many sentences which were not significant sentences but included the clue expressions.

Table 1: Results of determining confirmation labels

| type | correct | incorrect | total |
|----------|---------|-----------|-------|
| positive | 35 | 18 | 53 |
| negative | 10 | 4 | 14 |
| other | 48 | 6 | 54 |

Table 2: Type and number of incorrect confirmation

| incorrect confirmation | type and number of correct answers | | | |
|------------------------|------------------------------------|----------|-------|-------|
| | positive | negative | other | total |
| positive | – | 4 | 14 | 18 |
| negative | 2 | – | 2 | 4 |
| other | 4 | 2 | – | 6 |

Table 3: Results of determining confirmation labels to the proper sets of a question and its DA mail

| labeling result | positive | negative | other | total |
|-----------------|----------|----------|-------|-------|
| correct | 29 | 8 | 27 | 64 |
| failure | 4 | 4 | 15 | 23 |

- some question mails were submitted not for asking questions, but for giving some news, notices, and reports to the participants. In these cases, there were no answer in the DA mail and no sentence in the QR mail for confirming the previous mails.
- questioner’s answer was described in several sentences and one of them was extracted, and
- misspelling.

Next, we examined whether these significant sentences and the confirmation labels were helpful in choosing and accessing information for solving problems. In other words, we examined whether

- there was good connection between the significant sentences or not, and
- the confirmation label was proper or not.

For example, (Q2) and (DA2–1) in Figure 1 have the same topic, however, (DA2–2) has a different topic. In this case, (DA2–1) is a good answer to question (Q2). A user can access the document from which (DA2–1) was extracted and obtain more detailed information. As a result, the set of (Q2) and (DA2–1) was determined as correct. On the contrary, the set of (Q2) and (DA2–2) was a failure. In this experiment, 87 sets of a question and its DA mail were determined as correct and 34 sets were failures. The reasons of the failures were as follows:

- (Q2) *vedit ha, sonzai shinai file wo hirakou to suru to core wo haki masuka.* (Does vedit terminate when we open a new file?)
 (DA2–1) *hai, core dump shimasu.* (Yes, it terminates.)
 (DA2–2) *shourai, GNOME ha install go sugu tsukaeru no desu ka?* (In near future, can I use GNOME just after the installation?)
- (Q3) *sound no settei de komatte imasu.* (I have much trouble in setting sound configuration.)
 (DA3–1) *mazuha, sndconfig wo jikkou shitemitekudasai.* (First, please try ‘sndconfig’.)
 (QR3–1–1) *kore de umaku ikimashita.* (I did well.)
 (DA3–2) *sndconfig de, shiawase ni narimashita.* (I tried ‘sndconfig’ and became happy.)
- (Q4) *ES1868 no sound card wo tsukatte imasu ga, oto ga ookisugite komatte imasu.* (My trouble is that sound card ES1868 makes a too loud noise.)
 (DA4–1) *xmixer wo tsukatte kudasai.* (Please use xmixer.)
 (QR4–1–1) *xmixer mo xplaycd mo tsukaemasen.* (I cannot use xmixer and xplaycd, too.)

Figure 1: Examples of the significant sentence extraction

- wrong significant sentences extracted from question mails, and
- wrong significant sentences extracted from DA mails.

Failures which were caused by wrong significant sentences extracted from question mails were not serious. This is because there is not much likelihood of matching user’s question and wrong significant sentence extracted from question mails. On the other hand, failures which were caused by wrong significant sentences extracted from DA mails were serious. In these cases, significant sentences in the question mails were successfully extracted and there is likelihood of matching user’s question and the significant sentence extracted from question mails. Therefore, the precision of the significant sentence extraction was emphasized in this task.

Next, we examined whether proper confirmation labels were given to these 87 good sets of a question and its DA mail or not, and then, we found that proper confirmation labels were given to 64 sets in them. The result was shown in Table 3.

We discuss some example sets of significant sentences in detail. Question (Q3) in Figure 1 has two answers, (DA3–1) and (DA3–2). (DA3–1) is



Figure 3: A QA example which was generated by our system

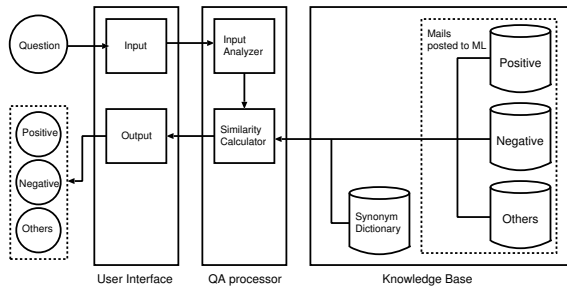


Figure 2: System overview

a suggestion to the questioner of (Q3) and (DA3–2) explains answerer’s experience. The point to be noticed is (QR3–1–1). (QR3–1–1) contains a clue expression, “*umaku ikimashita* (did well)“, which gives a positive label to the set of (Q3) and (DA3–1). It guarantees the information quality of (DA3–1) and let the user choose and access the answer mail from which (DA3–1) was extracted.

(DA4–1) in Figure 1 which was extracted from a DA mail has wrong information. Then, the questioner of (Q4) confirmed whether the given information was helpful or not, and then, posted (QR4–

1–1) in order to point out and correct the wrong information in (DA4–1). In this experiment, we found 16 cases where the questioners posted reply mails in order to correct the wrong information, and the system found 10 cases in them and gave negative labels to the sets of the question and its DA mail.

3 QA system using mails posted to a mailing list

3.1 Outline of the QA system

Figure 2 shows the overview of our system. A user can ask a question to the system in a natural language. Then, the system retrieves similar questions from mails posted to a mailing list, and shows the user the significant sentences which were extracted the similar question and their answer mails. According to the confirmation labels, the sets of the similar question and their answer mails were classified into three groups, positive, negative, and other, and shown in three tabs (Figure 3). A user can easily choose and access information for solving problems by using the significant sentences and the con-

firmation labels. The system consists of the following modules:

Knowledge Base It consists of

- question and answer mails (50846 mails),
- significant sentences (26334 sentences: 8964, 13094, and 4276 sentences were extracted from question, DA, and QR mails, respectively),
- confirmation labels (4276 labels were given to 3613 sets of a question and its DA mail), and
- synonym dictionary (519 words).

QA processor It consists of input analyzer and similarity calculator.

Input analyzer transforms user's question into a dependency structure by using JUMAN(Kuro 98) and KNP(Kuro 94).

Similarity calculator calculates the similarity between user's question and a significant sentence in a question mail posted to a mailing list by comparing their common content words and dependency trees in the next way:

The weight of a common content word t which occurs in user's question and significant sentence S in the mails M ($i = 1 \dots N$) is:

$$w_{WORD}(t, M) = t(t, S) \frac{N}{d(t)}$$

where $t(t, S)$ denotes the number of times content word t occurs in significant sentence S , N denotes the number of significant sentences, and $d(t)$ denotes the number of significant sentences in which content word t occurs. Next, the weight of a common modifier-head relation in user's question and significant sentence S in question mail M is:

$$w_{LINK}(l, M) = w_{WORD}(dier(l), M) + w_{WORD}(head(l), M)$$

where $dier(l)$ and $head(l)$ denote a modifier and a head of modifier-head relation l , respectively.

Therefore, the similarity score between user's question and significant sentence S of question mail M , $SCO E(M)$, is set to the total weight of common content words and modifier-head relations which occur user's question and

significant sentence S of question mail M , that is,

$$SCO E(M) = SCO E_{WORD}(M) + SCO E_{LINK}(M)$$

where the elements of set T and set L are common content words and modifier-head relations in user's question and significant sentence S in question mail M , respectively.

When the number of common content words which occur in user's question and significant sentence S in question mail M is more than one, the similarity calculator calculates the similarity score and sends it to the user interface.

User Interface Users can access to the system via a WWW browser by using CGI based HTML forms. User interface put the answers in order of the similarity scores.

3.2 Evaluation

For evaluating our method, we gave 32 questions in Figure 4 to the system. These questions were based on question mails posted to Linux Users ML. The result of our method was compared with the result of full text retrieval

Test 1 by examined first answer

Test 2 by examined first three answers

Test 3 by examined first five answers

Table 4 (a) shows the number of questions which were given the proper answer. Table 4 (b) shows the number of proper answers. Table 4 (c) shows the number and type of confirmation labels which were given to proper answers.

In Test 1, our system answered question 2, 6, 7, 8, 13, 14, 15, 19, and 24. In contrast, the full text retrieval system answered question 2, 5, 7, 19, and 32. Both system answered question 2, 7 and 19, however, the answers were different. This is because several solutions of a problem are often sent to a mailing list and the systems found different but proper answers. In all the tests, the results of our method were better than those of full text retrieval. Our system answered more questions and found more proper answers than the full text retrieval system did. Furthermore, it is much easier to choose and access information for solving problems by using the answers of our QA system than

- (1) I cannot get IP address again from DHCP server.
- (2) I cannot make a sound on Linux.
- (3) I have a problem when I start up X Window System.
- (4) Tell me how to restore HDD partition to its normal condition.
- (5) Where is the configuration file for giving SSI permission to Apache ?
- (6) I cannot login into proftpd.
- (7) I cannot input kanji characters.
- (8) Please tell me how to build a Linux router with two NIC cards.
- (9) CGI cannot be executed on Apache 1.39.
- (10) The timer gets out of order after the restart.
- (11) Please tell me how to show error messages in English.
- (12) NFS server does not go.
- (13) Please tell me how to use MO drive.
- (14) Do you know how to monitor traffic load on networks.
- (15) Please tell me how to specify kanji code on Emacs.
- (16) I cannot input \ on X Window System.
- (17) Please tell me how to extract characters from PDF files.
- (18) It takes me a lot of time to login.
- (19) I cannot use lpr to print files.
- (20) Please tell me how to stop making a backup file on Emacs.
- (21) Please tell me how to acquire a screen shot on X window.
- (22) Can I boot linux without a rescue disk?
- (23) Pcmcia drivers are loaded, but, a network card is not recognized.
- (24) I cannot execute PPxP.
- (25) I am looking for FTP server in which I can use chmod command.
- (26) I do not know how to create a Makefile.
- (27) Please tell me how to refuse the specific user login.
- (28) When I tried to start Webmin on Vine Linux 2.5, the connection to localhost:10000 was denied.
- (29) I have installed a video capture card in my DIY machine, but, I cannot watch TV programs by using xawtv.
- (30) I want to convert a Latex document to a Microsoft Word document.
- (31) Can you recommend me an application for monitoring resources?
- (32) I cannot mount a CD-ROM drive.

Figure 4: 32 questions which were given to the system for the evaluation

by using the answers of the full text retrieval system.

Both systems could not answer question 4, “Tell me how to restore HDD partition to its normal condition”. However, the systems found an answer in which the way of saving files on a broken HDD partition was mentioned. Interestingly, this answer may satisfy a questioner because, in such cases, our desire is to save files on the broken HDD partition. In this way, it often happens that there are gaps between what a questioner wants to know and the answer, in several aspects, such as concreteness, expression and assumption. To overcome the gaps, it is important to investigate a dialogue system which

Table 4: Results of finding a similar question by matching of user’s question and a significant sentence

| | Test 1 | Test 2 | Test 3 |
|---------------------|--------|--------|--------|
| our method | 9 | 15 | 17 |
| full text retrieval | 5 | 5 | 8 |

(a) the number of questions which is given the proper answer

| | Test 1 | Test 2 | Test 3 |
|---------------------|--------|--------|--------|
| our method | 9 | 25 | 42 |
| full text retrieval | 5 | 9 | 15 |

(b) the number of proper answers

| | positive | negative | other | positive & negative |
|--------|----------|----------|-------|---------------------|
| Test 1 | 2 | 2 | 5 | 0 |
| Test 2 | 9 | 4 | 12 | 0 |
| Test 3 | 10 | 5 | 25 | 2 |

(c) the number and type of labels given to proper answers

can communicate with the questioner.

References

- TREC (Text REtrieval Conference) : <http://trec.nist.gov/>
- NTCIR (NII-NACSIS Test Collection for IR Systems) project: <http://research.nii.ac.jp/ntcir/index-en.html>
- Kurohashi and Higasa: Dialogue Helpsystem based on Flexible Matching of User Query with Natural Language Knowledge Base, 1st ACL SIGdial Workshop on Discourse and Dialogue, pp.141-149, (2000).
- Kiyota, Kurohashi, and Kido: “Dialog Navigator” A Question Answering System based on Large Text Knowledge Base, COLING02, pp.460-466, (2002).
- Kurohashi and Nagao: A syntactic analysis method of long Japanese sentences based on the detection of conjunctive structures, Computational Linguistics, 20(4),pp.507-534, (1994).
- Kurohashi and Nagao: JUMAN Manual version 3.6 (in Japanese), Nagao Lab., Kyoto University, (1998).
- Mihara, fujii, and Ishikawa: Helpdesk-oriented Question Answering Focusing on Actions (in Japanese), 11th Convention of NLP, pp. 1096–1099, (2005).
- Watanabe, Sono, Yokomizo, and Okada: A Question Answer System Using Mails Posted to a Mailing List, ACM DocEng 2004, pp.67-73, (2004).
- Watanabe, Nishimura, and Okada: Confirmed Knowledge Acquisition Using Mails Posted to a Mailing List, IJCNLP05, (2005).