ASQA: Agent Supported Question Answering System

Takashi Ishikawa, Thanachai Wongvibulsin, Yu Dongming
Nippon Institute of Technology
tisikawa@nit.ac.jp

Keywords: Question Answering, Intelligent Agent, Text Mining, Collaborative Learning

Abstract The paper describes Agent Supported Question Answering system (ASQA) that assists collaborative learning in the classroom using networked computers. When a student sends any question to the system, the supporting agent searches FAQ stored in the system or forwards the question to another student selected, then the student can get an answer for the question. The agent in the system utilizes text mining technology and vector space model to select FAQ or students with respect to the question autonomously. The system assists not only giving an answer to the students but also giving chance of learning to students by answering the question of other students.

1. Introduction

In the classroom of programming practice or system development practice, students are hard to communicate with each other because they are devoting to each work for almost of the time. When a student has any question, the student raises hand and tells the question to the teacher of TA. However one teacher or a few TA cannot provide answer simultaneously in the classroom. The motivation of the research arose from the circumstance in the classroom using computers.

The purpose of the research is to provide any facility for question answering in a classroom using computers. The goal of the research is to develop the online communication system that can provide an answer to the question as soon as possible and encourages the collaborative learning in such a classroom. To achieve this goal we have developed Agent Supported Question Answering system (ASQA) applying intelligent agent technology, distributed computing technology and text mining technology. The experimental result to evaluate the feasibility of the system exhibits that the system is almost feasible for the purpose.

2. Online Question Answering

The progress of information technology has provided various online communication tools such as e-mail system, online chat system, electronic bulletin board and online messaging system. These tools are applicable to question answering in the classroom using networked computers. Some of them have real time feature, some provide one to one communication, and some are suitable for communication in a community. In the classroom using networked computers, students are also able to communicate with each other by oral communication. However in lessons such as programming practice or system development practice, the students are difficult to communicate with other students because they are busy to accomplish each work. They need to use online communication tools for question answering without disturbing other students.

The needed facility for online question answering in the classroom has requirements as the following.
(1) Students can send a question and receive the answer in any time.
(2) Students should receive any answer as soon as possible and can send any other question for the answerer if needed.
(3) Students should receive questions that are interesting to answerer for their learning.
(4) Students should receive questions of equal number for fairness.

The first requirement means that the needed communication tool must have real time feature in such as online chat system or online messaging system. The second requirement means that the question should be sent quickly to someone who is able to answer the question and communicate with the questioner till the question will be solved. This requirement may be satisfied by e-mail system when the question is sent to selected person who is able to answer. The third requirement means that the answerer should be selected whether the student has interest in the topic of the question so that answering the question promotes self learning of the student. The fourth requirement means that the answerer should be selected also by the number of questions that the student has solved. The third and fourth requirement arise from that online question answering should encourage collaborative learning in the classroom.
3. Supporting Agent

Intelligent agent performs some decision making tasks instead of person. In the online question answering system, a supporting agent performs selection of answerers instead of the questioner. In order to perform the task, the agent must decide who is suitable for the question from the information stored in the question answering system. Such information includes preference of the users, characteristics of the users, question answering log of the users. From the standpoint of autonomy needed in intelligent agent, it is preferable that the agent learns how to select answerers by doing the task itself. This autonomy of the supporting agent is a issue of the research that applies intelligent agent to the online question answering system.

The approach to the autonomy of the supporting agent is to use question answering log for the learning of the agent. The reasons of adopting this approach are the following.

(1) It is hard to provide the preference of characteristics of the students as prior knowledge of the agent.
(2) The agent can acquire question answering log as the data to perform its task.
(3) The agent can decide whether it performs the task well or not from the question answering log.

The first reason still exists if we could provide some initial data that are some approximation. The second reason is the most important point of the research applying text mining technology to solve the problem as described in the next section. The third reason is the possibility of applying machine learning techniques to realize the autonomy of the agent.

The supporting agent for the online question answering performs the task to select suitable answerers for the question that a questioner sent. In this decision making, the agent must account the following requirements to encourage collaborative learning in the classroom.

(1) The answerer receives only new interesting question as possible.
(2) The questioner receives any answer as soon as possible.

To satisfy the first requirement, Agent Supported Question Answering system (ASQA) described in the paper uses FAQ (Frequent Asked Question) search to provide an answer without disturbing other student when the similar question and its answer exists in the question answering log (see Section 5). When the similar question does not exist or the questioner does not satisfy with the searched FAQ, the agent selects a suitable answerer using similarity between feature vectors of the question and the candidate answerer, where the feature vectors are constructed by applying text mining technology (see Section 6).

4. Text Mining

The supporting agent for the question answering utilizes text mining technology to construct feature vectors for a question, an FAQ entry, and a student. Text mining is a technology that extracts some information or knowledge from given text data. The process of text mining to construct feature vectors involves word extraction, word weighting, word counting, and vector construction (Figure 1).

![Figure 1. Process of constructing a feature vector](image)

Word extraction is performed with natural language processing using a package program such as Chasen [4] for Japanese. The program outputs words and their part of speech in the input text data. From the purpose of using feature vectors, the words corresponding to each dimension of the feature vector are limited to noun that express topics of the text.

Word weighting is performed based on statistics of words appearing in the text data. First we define a message as a group of words that can be separated by position in a text, time of creation, or any structural characteristics. Then we assume that a message expresses at least one topic to communicate with others and the topic is expressed by some words. Hence we can construct feature vectors for a question, an FAQ entry, and messages of a student using the frequencies of words in the text data. In usual some words are important in the communication and the others are less important. Therefore we assign a weight to each word.
to express the importance of the word. Finally a feature vector is expressed as an array of word frequency multiplied by word weight.

A word weight can be calculated with the following expression [3].

$$\text{word weight} = \text{TF} \times \text{IDF}$$

where TF stands for *term frequency* in a text and IDF stands for *inverse document frequency*. In this paper a term corresponds to a word and a document corresponds to a message. It is known that some modifications based on statistical distribution of word frequency improve the accuracy of TF or IDF [2]. So we have implemented slightly modified TF and IDF. It is needed that the words used in the computation of decision making should be limited with word weight because of trade off between accuracy and efficiency.

The final step of constructing a feature vector is word counting in a given text for the limited number of words. Each component of a feature vector is calculated as word frequency multiplied by word weight with respect to the word associated with its dimension. In this definition of a feature vector, we assume that word frequency in a text expresses the local importance of a word in the text. On the contrary we assume that word weight in a set of texts expresses the global importance of a word in the set.

5. FAQ Search

In usual FAQ entries in Web pages are convenient for solving a question about something by ourselves. Such FAQ entries are collected by hand in a mailing list or an electronic bulletin board. An FAQ entry is composed of a question and a corresponding answer(s) as text data. Applying text mining technology to FAQ text data, it is possible to find any answer for a given question [1]. In the online question answering system, it is preferable that the answer for a given question is selected first in question answering log as a FAQ database. So we adopt FAQ search facility in ASQA.

The process of FAQ search is composed of FAQ indexing and similarity based retrieval. FAQ indexing is performed when a question is solved with a satisfied answer to add a new feature vector for the question and its answer in the FAQ database. The feature vector is an array of word counts without multiplied by word weight. The reason is that word weight will change its value according to the growth of the question answering log. Also similarity based retrieval is performed when a question is sent to the system. First the feature vector for the question is constructed by word counting as same as FAQ indexing then the similarity between the feature vectors for the question and an FAQ index is calculated by the following expression. This similarity is called cosine similarity.

$$\text{similarity} = \frac{(\mathbf{q}, \mathbf{i})}{|\mathbf{q}| |\mathbf{i}|}$$

where \(\mathbf{q}\) stands for the feature vector for the question, \(\mathbf{i}\) stands for the feature vector for the FAQ index, which are multiplied by word weight for each dimension. The numerator represents an inner product of two feature vectors and the denominator represents a product of norms of the two vectors. In a preliminary experiment, the cosine similarity can discriminate the relevant FAQ from the others. The result is shown in Figure 2, where horizontal axis corresponds to similarity, the vertical axis corresponds to frequency of FAQ entries, and the dark bars represent relevant FAQ entries.

![Figure 2. Result of FAQ search](image)

6. Answerer Selection

When the questioner does not satisfy with the FAQ entry that the supporting agent searched, the agent forward the question to a student selected by the answer selection process. The process is composed of constructing feature vectors for answerers and selecting an answerer by the similarity between the two feature vectors.

Constructing feature vectors for answerers is performed when a new FAQ entry (a question and its answer accepted) is added. First the feature vector for the FAQ entry is constructed by word counting for the FAQ text, then the feature vector is added to the existing feature vector for the answerer of the FAQ entry. The initial feature vector is a zero vector. Here the feature vectors are composed of word counts without multiplied by word weights.
Selecting an answerer is performed in the following way. First similarities between the feature vector for the unsolved question and all of the feature vectors for answerers. The similarity is calculated by the following expression.

$$\text{similarity} = \frac{(q \cdot a)}{|q|^2}$$

where $a$ stands for the feature vector for an answerer. The reason why the denominator does not include the norm of $a$ is that the dimensions where components are zero in $q$ do not have relevance with the question. The expression of similarity is slightly modified version of Bin Yu similarity [5]. In a preliminary experiment, Bin Yu similarity can discriminate the relevant answerer from the others. The result is shown in Figure 3, where horizontal axis corresponds to similarity, the vertical axis corresponds to number of answerers, and the dark bars represent relevant answerers. In the beginning of experiment the feature vectors for answerers are not obtained well, it is the reason of similarities for some relevant answerers are low.

Finally the agent selects a candidate answerer in the order of similarities between $q$ and $a$ in general. But if the number of FAQ entries that the next candidate has solved is less than that of the first candidate, then the agent selects the next candidate in order to equalize the number of questions that students will answer. Furthermore, if the selected answerer rejects to answer the forwarded question, then the agent selects the other promising answerer in the order of similarities between $q$ and $a$.

7. Experiment

In order to evaluate the feasibility of Agent Supported Question Answering system (ASQA), we have implemented the system with Java distributed computing technology and natural language processing techniques. The system comprises client application and Web server. The users download the client application from the Web server, and send (or receive) question or answer using it. The Web server includes supporting agents for each user that communicate with the client application and each other. Figure 4 shows an overview of the experimental system and Figure 5 shows an example display of client application.

Figure 3. Result of answerer selection

Figure 4. An overview of the system

Figure 5. Example of client display

The evaluation experiment has done in the classroom of system development practice for the third grade students in the Department of Computer Engineering at Nippon Institute of Technology. The number of students using the system is twenty, the number of questions processed is a hundred, and the number of FAQ prepared before the experiment is about five hundreds. The topic of questions diverge usage of programming language, cause of program errors, and so on. After the experiment, 90 % of the users satisfied with the system in the evaluation questionnaires. They have got some answer for each question finally. The numbers of answers for each user shown in Figure 6 are nearly equal for the sake of the answerer selection by the supporting agents.
8. Conclusions

From the result of evaluation experiment for Agent Supported Question Answering system (ASQA), the system almost properly selects candidates of FAQ or answerer with respect to the question and fairly assists collaborative learning in the classroom. Students in the classroom may not answer to the similar question already exists in FAQ database and are provided with new questions which are of interest. This feature encourages question answering using ASQA.

References


