Interactive Question Answering Based on FAQ

Song Liu[1,2], Yi-Xin Zhong[1] and Fu-Ji Ren[2]

[1]School of Computer Science, Beijing University of Posts and Telecommunications, 10th Xitucheng Road,Beijing, 100876, China.

[2]Faculty of Engineering, The University of Tokushima, 2-1 Minamijosanjima,Tokushima 770-8506, Japan.

Songliu84@gmail.com, yxzhong@ieee.org, ren@is.tokushima-u.ac.jp

Abstract. A question answering system receives the user's question in nature language, and answers it in a concise and accurate way. An interactive question answering (IQA) provides a natural way for users to express their information requirement. There are two key points for IQA. The first is how to answer a user's question in a continuous question answering process. The second is the way that the question answering system interacts with the user. In this work the answers are from FAQ knowledge base which is extracted from community question answering web portals. The syntactic, semantic and pragmatic features between question and candidate answers and context information are used to construct models by ranking learning method to extract the answers. And the question answering system requests user to feedback of the answer. It is a naive and effective interactive method. The results of experiments show that our method is effective for interactive question answering.

Keywords: context, FAQ, Interactive question answering, ranking learning.

1 Introduction

A question answering system receives the user's question in nature language, and answers it in a concise, accurate and natural way. Question answering system analyzes the question to acquire the information requirement of user. And then based on the knowledge question answering system convert the information requirement into constraint condition while searching answers in knowledge space. The interaction between user and question answering system is introduced in the interactive question answering. There are two differences between interactive question answering and single round question answering. The first difference is that interactive question answering is a continuous question answering. How to use the context information is one key point. The second difference is that the question answering system interacts with users besides answering questions. It is the second key point of interactive question answering.

In this work, the answers are from the FAQ. FAQ provides information from the question-answer pair. The community question answering in web portals brings out abundant FAQs. In this work the FAQ is from the Baiduzhidao. For the first key point of interactive question answering, the ranking learning method is used to train a statistic model to predict answers. The features that describe the training and testing instances are from the question, related FAQ and question answering context. And the syntactic, semantic and pragmatic information is concerned to extract the features.

For the second key point, the interaction between the question answering system and user is designed as that after answering a user's question the system will request the user to feedback whether he is satisfied with the answer. This interactive mode has two advantages. First is that the white or black feedback is easy to be caught and fully used by QA system. Secondly this kind of feedback matches FAQ. If the feedback is positive the question and answer will be added into the FQA base. If the feedback is negative, QA system will provide another answer based on the question and repudiated answer.

The rest of this paper is organized as follows: Section 2 introduces related work. Section 3 is about the model and syntactic, semantic and pragmatic features in context question answering. Section 4 describes the interaction between user and QA system. Section 5 is the experiments and results. Section 6 is conclusions.

2 Related work

There are two key points for interactive question answering. The first is how to get answers in continuous question answering. The second is the interaction between QA system and users. To the first point, since 2004 in TREC QA task [1], a group of questions were around one topic. Hence a question was related to the question answering context. In real interaction the topic transformation is possible. The questions may not be around one topic. Yang estimated whether the topic is transferred through anaphora and ellipsis by decision tree [2]. Sun used center theory to deal with the anaphora in context question answering [3]. Sun [4] and Chai [5] used discourse theory in context question answering. Kirschner adopted logistic regression with contextual information to find answers in context QA [6]. There are three methods to obtain the data of context question answering. The first was getting questions from the QA evaluation such as TREC QA task [3]. However there was a gap between this kind of data and the realistic context question answering. The second method is wizard of OZ [7] which simulated the interaction between user and QA system to get the question answering data. The third method was collecting the question answering data while using the question answering system. This kind of data is the most realistic data. However, this kind of data is limited by the ability of the QA system.

The second key point is the method of interaction between user and QA system. The TREC QA task also explored the interaction between QA system and user. Interactive QA was first introduced in TREC 2006[9]. And in TREC 2007 QA task[10] reviewers interact with the QA system online. As reported, for most systems the answers are improved after interaction. However the improvements were not significant. Hickl employed CRF model to construct question-answer pairs. And then it showed the question answering pair to user to impact the following question that users will ask in the next round [11]. Misu used reinforcement learning to learn the interacting strategy in the interactive process [12]. Similarly, adaptive learning was also used in an interactive question answering [13].

The FAQ is a kind of knowledge source that is easy to use by QA system. In 1997 Burke researched the FAQ question answering in semantic level [14]. Kolkata discussed how to get answers when there are no related questions in question-answer pairs [15]. Cong used sequence pattern based classification method to extract FAQ pairs form forum and a community question answering [16]. In this work ranking learning method is used to select answers [17]. Ranking learning methods are widely used in information retrieval [18] and question answering [19].

3 Context question answering in Interactive question answering

This section introduces the first key point for interactive question answering, how to extract answers in context question answering. The answer extraction is converted to a ranking learning process. It is supervised learning that uses labeled instances to train the statistic model. In the following part of this section the support vector based ranking learning methods and the syntactic, semantic and pragmatic features in context question answering are introduced.

3.1 Support vector based ranking learning method

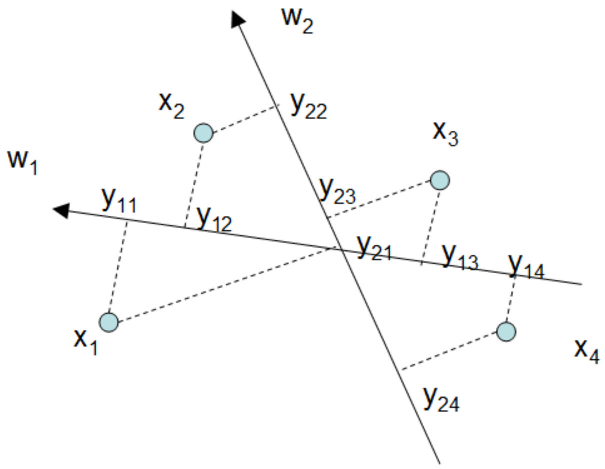
Two support vector based ranking learning methods Ranking SVM[20] and SVM-MAP[21] are used to rank the candidates question-answering pairs. The idea of support vector was first exploited in support vector machine (SVM) [22]. The core of SVM is finding the plane (support vector) $\hat{\gamma} = y_i(w \cdot x_i + b)$, which makes the training data correctly classified and the geometric intervals maximize. The optimization of SVM is described in the following equations. ξ_i is the slack variable for linearly inseparable conditions. The parameter C is the penalty parameter.

$$\min_{w,b,\xi} \quad \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \tag{1}$$

$$s.t. \quad y_i(w \cdot x_i + b) \geq 1 - \xi_i, \quad i = 1,2,...,N \tag{2}$$

$$\xi_i \geq 0, \quad i = 1,2,...,N \tag{3}$$

Support vector based ranking learning has some differences from SVM classification. The mainly difference is that output Space of ranking learning is an ordered sequence space. The optimization target of support vector based ranking learning is that the order of the output sequence is correct and maximizes the interval of distances between the elements which are mapped on the plane.



A group of instances (x1, x2, x3, x4) in the input space X. The correct ordering is <y1, y2, y3, y4>. After mapping the instances onto plane w1 the ordering of them is <y11, y12, y13, y14>. And After mapping the instances onto plane w2 the ordering of them is <y22, y23, y21, y24>. Hence the plane w1 can order the instances correctly.

Fig. 1. the input instances are mapped onto different plane.

Two support vector based ranking learning methods will be introduced. The first is Ranking SVM which is a pair-wise approach. The second is SVM-MAP which is a list-wise approach. The mainly difference between the two approaches is the difference between their lose/risk functions.

In Ranking SVN the risk function is described by Kendall's τ [23]. Kendall's τ is a metric to measure the consistency of two finite strict orderings. For two finite strict orderings ra and rb which are in the same space, $r_a \subset D \times D$ and $r_b \subset D \times D$, the Kendall's τ is defined as:

$$\tau(r_a, r_b) = \frac{P-Q}{P+Q} = 1 - \frac{2Q}{\binom{m}{2}} \tag{4}$$

In the two orderings. If a pair $d_i \neq d_j$ has the same order in ra and rb, the pair is concordant. Otherwise the pair is discordant. P is the number of the concordant pairs and Q is the number of the discordant pairs. And m is the number of the elements in ordering. The summation of P and Q is $\binom{m}{2}$. The higher of Kendall's τ, the more consistency the two orderings are.

SVM-MAP is a support vector based list wise ranking learning approach. The risk function of SVM-MAP is described by mean average precision (MAP) [24]. MAP is also a metric which measure the consistency of two finite strict orderings. It is widely used in the evaluation of information retrieval and question answering. Average precision is the basis of MAP, which is calculated as equation 9. In the equation p(i) is the precision of the elements from the first to the ith. And Δr(i) is the variation of recall in the ith position. It is the difference between r(i-1) and r(i). And r(i) is the precision of the elements from the first to the ith. On the right of equation 9 rel(i) represents whether the ith element is a correct answer for the question. Average precision describes the consistency fo the two orderings from precision and recall. MAP is the mean of APs for several groups of orderings.

$$AP = \sum_{i=1}^{n} p(i)\Delta r(i) = \frac{\sum_{i=1}^{n}(p(i)\times rel(i))}{element \quad number} \tag{5}$$

$$MAP = \frac{1}{qnum}\sum_{k=1}^{qnum} AP = \frac{1}{qnum}\sum_{k=1}^{qnum}\left(\sum_{i=1}^{n} p(i)\Delta r(i)\right) \tag{6}$$

The two support vector based ranking learning methods, Ranking SVM and SVM-MAP has been introduced. The mainly differences between them are the definitions of the risk function. And the two ranking learning approaches will tested in the experiments.

3.2 The features for context question answering

The features to describe the instances of context question answering from syntactic, semantic and pragmatic level. Here the syntactic, semantic and pragmatic from question Q and candidate question-answer(Q', A') pairs are first introduced.

3.2.1 Syntactic feature

The syntactic feature describes the similarity between question and candidate question-answer pair in grammatical form. The overlap of words are used to calculate the syntactic similarity[19]. The overlap of words is the proportion of the concurrence words in the sentence. The overlap of words between question Q and the question Q' and answer A' of candidate question-answer pair(Q', A') are calculated separately. Firstly the question Q and QA pair (Q',A') are segmented and tagged the POS and the entities. Then the verbs, nouns, adjectives and entities are retained to calculate the overlap of words between question Q and question Q' and answer A' in candidate QA pair, as the equation 7 and 8. C is the number of the overlap words in sentence. And n is the number of words of a sentence.

$$overlap(Q,Q') = \frac{C_{QQ'} + C_{Q'Q}}{n_Q + n_{Q'}} \tag{7}$$

$$overlap(Q,A') = \frac{C_{QA'} + C_{A'Q}}{n_Q + n_{A'}} \tag{8}$$

3.2.2 semantic features

The semantic features indicate the similarity between question Q and candidate QA pairs (Q', A') in content. The semantic similarity between two sentences is calculated based on the word semantic similarity. Here the word semantic similarity is calculated based on Howent [25]. We refer Liu's method[26] to calculate the word semantic similarity between words based on sememes. The similarity between two sememes is calculated based on the distance of the two sememes in the sememes tree(equation 9).

$$Sims(p_1, p_2) = 1 - \frac{dis(p_1, p_2)}{d_1 + d_2} \tag{9}$$

And the similarity between concepts is calculated based on the sememes similarity. There are four kinds of sememes describing a concept. They are first sememe, basic sememe, relational symbol and relational sememe. The similarities of the four kinds of sememes between two concepts are calculated separately. And then the semantic similarity is calculated as equation 10, where $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 1$, $\beta_1 \geq \beta_2 \geq \beta_3 \geq \beta_4$.

$$Simw(S_1, S_2) = \sum_{i=1}^{4} \beta_i \prod_{j=1}^{i} Sims_j(S_1, S_2) \tag{10}$$

The semantic similarity between sentences is calculated based on the word semantic similarity. Equation 20 shows how to calculate the sentence semantic similarity. The sentence similarities between question Q and question Q and answer A in the candidate QA pair are calculated as equation 12 and 13.

$$simsem(q,s)=\frac{\sum_{i=1}^{n}\max_{j=1,\dots,m}(sim(qnoun_i,snoun_j))+\sum_{i=1}^{p}\max_{j=1,\dots,q}(sim(qverb_i,sverb_j))}{num(nouns)+num(verbs)} \quad (11)$$

$$simsem(Q,Q')=\frac{\sum_{i=1}^{n}\max_{j=1,\dots,m}(sim(Qnoun_i,Q'noun_j))+\sum_{i=1}^{p}\max_{j=1,\dots,q}(sim(Qverb_i,Q'verb_j))}{num(nouns)+num(verbs)} \quad (12)$$

$$simsem(Q,A')=\frac{\sum_{i=1}^{n}\max_{j=1,\dots,m}(sim(Qnoun_i,A'noun_j))+\sum_{i=1}^{p}\max_{j=1,\dots,q}(sim(Qverb_i,A'verb_j))}{num(nouns)+num(verbs)} \quad (13)$$

3.2.3 Pragmatic features

Pragmatic feature describes the effect of information for the goal of subject. In question answering system the subject is the user. And the goal of the user is obtaining the answer that satisfies the user' information requirements. In question answering the pragmatic information is the information that indicates the whether the answer can meet the user's information requirement. In this work, the pragmatic information is from the question Q and Q' in candidate QA pair. The pragmatic feature of question Q describes the expected speech act of answers. The user has expectation on the speech act of answers. For example, when a user asks "Why France rejected the EU constitution treaty?", the expected speech act of answer is "explanation" . From the expected act of answer, questions are classified in five categories: statement, instruct, explanation, verifying and opining.

Table 1. Expected answering acts.

Expected act	Description	Question
statement	Information, claim, or announcement	When will the next train arrive? What are greenhouse gases? Who is Justin Bieber?
instruct	An idea or a manner that is suggested	How can I drive from Beijing to Shanghai?
explanation	An explanation	Why the American invasion of Iraq?
Verifying	An affirmation or negation	Were you born in 1990?
Opining	An opinion, evaluation, or attitude	How about my new longuette?

Table 1. common Chinese interrogatives

types	interrogatives
interrogative pronoun	什么(what)，谁(who)，哪里(where)，哪儿(where)，何(what)，孰(which)，哪个(which)，哪(where)，啥(what)，

	哪些(which)
interrogative adverbs	为何(why)，怎样(how)，怎么办(how to)，为什么(why)，咋(how, why)，怎(how)，多少(how much)，多高(how tall)，多久(how long)，多长(how long)，多重(how heavy)，何如(how)，怎么(how)，为啥(why)，怎么样(how)，如何(how)，何以(why)，缘何(why)
Confirmation verb	*不* (* not *)，*没* (* not *)，*否 (* not)，*(了)没 (* not)
modal particle	吗(ma)，呢(ne)，么(me)[1]

A maximum entropy [27] method is used to classify the expected speech acts of answers. The classification features include n-gram, interrogative, the words modified by interrogative and syntactic structure. We collect 3124 questions by a search engine based on Chinese interrogatives and label the expecedt speech act of answer. And 70% data is used to train the model and 30% data is used to test it. The results are shown in table 3. And the average F score is 91.3%

Table 2. Evaluation of classification for expected questioning acts.

	Statement	Instruct	Explanation	Opining	Verifying
Precision	0.933	0.852	0.961	0.766	0.769
Recall	0.976	0.784	0.805	0.855	0.682
F-score	0.954	0.790	0.876	0.808	0.723

And for the question Q' in the candidate QA pair, the expected speech act is also classified as another pragmatic feature. And if the two expected speech acts are matched, there is more possibility that the candidate QA pair is the answer for the question.

3.3 Context features

In context question answering, the previous questions and answers construct the context. The context features include whether the topic of QA is continuous and the syntactic and semantic similarities between the candidate QA pair and context.

The transferring of topic in continuous question answering is common. If the QA topic transfers, the question has no relations with the context, and the context information cannot assist to find answers. Hence whether the QA topic is continuous is an important feature when using the context information.

Here whether the topic continuity is converted to a dichotomy problem. SVM is an effective classification method in dichotomy problem. Here the tool libsvm[28] is

[1] They are the common Chinese modal particles. We labeled them with pinyin.

used. RBF kernel is chosen and the parameter c and r are confirmed by the tool grip. And the features for classification include the features from the current question and the features from context. Firstly the features from the question are introduced. The anaphora is an important feature. If the question contains anaphora, it is high possibility the question has relation with context. And some conjunctions and adverbs such as “既然”(since) and“那么”(then) also show the continuous relation. Ellipsis is also an important feature. We use the dependency of the question to judge the ellipsis in question. If the subject is lacking in the question, the ellipsis are confirmed. The second kind of features is from the question and context. The syntactic and semantic similarities between question and question answer pair in the previous round are introduced as features.

The training data have two sources. First is the IT question answering in CSDN community. Second is the Confucius and analects of Confucius question answering in Baiduzhidao. We collect 400 group context question answering and tag the continuity manually. And 5 times cross validation is used. Table 4 shows the result.

Table 3. the result of context continuity

	precision	recall	F score
continuous	0.714	0.781	0.746
discontinuous	0.829	0.773	0.8

The features of the second kind are the context syntactic and semantic similarity features. These features are: (1) the syntactic and semantic similarities between the question Q’ in candidate QA pairs and the last user question Qp, (2) the syntactic and semantic similarities between the question Q’ in candidate QA pairs and the last answer Ap, (3) the syntactic and semantic similarities between the answer A’ in candidate QA pairs and the last user question Qp, (4) the syntactic and semantic similarities between the answer A’ in candidate QA pairs and the last answer Ap.

4 The interaction between QA system and user

The second key point of interactive question answering is the interaction mode. As presented by the TREC CIQA task, complex interaction did not assist to get answers signally. In this work a naive and effective interaction mode is adopted. The interactive mode is that after providing answers, the question answering system will ask user for feedback whether he is satisfied with the answer. It is the direct representation of the answer effect. In this interaction mode the form of user's feedback is restricted. Hence the feedback information is easy to be obtained and used.

There are positive and negative feedbacks from the user. For the positive feedbacks, user's question and system's answer are combined together as the QA pair and is stored in the FAQ knowledge base. It makes the knowledge of QA system increases in the process that the user uses a question answering system. For the negative feedbacks, QA system should supply a new answer for the user. When finding the new answer, the information from question and previous answer is used. The syntactic and semantic similarity between question Q and new answer are calculated. And the semantic similarity negative answer and new answer are also calculated. Then the score for new answer is calculated as equation 14. It follows the hypothesis that the more related the new with the question and the less related to the negative answer, the more possibility the new answer is correct.

$$Score(Q, A^{-}, A) = \frac{syn(Q, A) + sem(Q, A)}{syn(A, A^{-}) + sem(A, A^{-})} \tag{14}$$

5 Experiment

First the method to acquire the data in question answering is introduced. In this paper the QA system is FAQ based QA system about the Analects of Confucius. The FAQ is from Baiduzhidao. Baiduzhidao is a portals community of question answering. In community question answering the answers are also from users which contain the domain experts. And users can vote, commit the answers to filter the best answers. Hence the community question answering is an effective source to get the FAQs. Here we crawled more than 26000 questions about the the Analects of Confucius. And about 7100 QA pairs that are labeled "best answer" are restored in the FAQ knowledge base.

Baiduizhiao is also used to acquire the training and testing data. Volunteers are required to supply 10 questions about the Analects of Confucius.volunteers can get the answers from Baiduzhidao, and consider the next question based on the answer. Form 10 volunteers 100 questions are collected. And after filtering the improper questions 10 groups 90 questions are retained. This method simulates the interaction process between users and QA system, so that it is a Wizard of oz method.

The 10 groups questions are divided into 5 parts. And 5 cross validation is used to evaluate the results. The evaluation metrics are MAP and p@1. MAP has introduced in section 3.1. P@1 is the precision of the first answer. In this experiment the effect of context features is evaluated. The results with and without context features are compared.

The results show that the results of SVM MAP are better than Rank SVM. And it is consistent with the related works [21]. And after adding the context features the MAP and p@1 using SVM MAP are both improved. And for Ranking SVM the MAP decreases slightly. These prove that the context features are important for continuous question answering. And after feature selection the MAP and p@1 are both increase for SVM MAP and Ranking SVM. And the results are also comparable with recent related work [15].

Table 5. The results for continuous question answering

	Ranking SVM		SVM MAP	
feature	map	p@1	map	p@1
Without context features	0.635	0.533	0.737	0.644
With context features	0.631	0.522	0.769	0.677

The second part of the experiments is about the interaction method between QA system and user to verify the effect of the interaction method. Based on the previous experiments, there are 62 answers are correct for 90 questions. For the correct answers users give positive feedbacks. These question answer pairs are stored in the FAQ knowledge base. And for the rest 28 answers with negative feedback, the method in section 4 is used to find a new answer for the user. And 15 new answers are correct. The precision of the second round answers is 0.536. And combining the first round answers and second round answers the total precision is 0.856. It proves that the interaction between QA system and user is effective and helpful to find the collect answer.

6 Conclusions

By interactive question answering user can obtain information conveniently. The interactive answering faces two major problems. The first is how to answer the user question in the process of continuous answering, which is mainly manifested as how to use context information. The Second is the interaction mode between QA system and the user. For the continuous question answering, this paper adopts the frequently answers to questions (FAQ) as the source of the answer and uses ranking learning methods based on support vector to build model. The features of describing training and test instances mainly come from two sources: one is the syntax, semantic and pragmatic features of the candidates question answer pairs, the other is the continuity of the question answering and the syntax and semantic features from context. As to the interaction mode between QA system and the user, the QA system asks users whether they are satisfied with the answer. This interaction is simple and effective, because the information obtained is easily understood and used by the QA system. It can provide the basis for a QA system to get new answers or adding correct answers to the questions to the knowledge base. Experiments show that it is effective to answer user questions by using the ranking learning method and multiple features. And the interaction between QA system and users further significantly improve the accuracy of the QA system to answer user questions. Future studies need to address the following questions. Firstly more features describing the relations between questions and answers are needed to understand the questions and answers and to improve the performance of QA system. The second is the interaction way between QA system and the user. A naive interactive way is used in this paper, by this way, the QA system is easy to understand and use the information. However, the expression of the user's information is limited, so in future studies it should also be concerned about the way that enables users to more freely express their intention.

Reference

1. Voorhees E M. Overview of the TREC 2004 robust retrieval track[C] Proceedings of TREC 2004. 2004.
2. Yang F， et a1． A data driven approach to relevancy recognition for contextual question answering[A]．Proc of the IQA Workshop at HLT-NAACL 2006[C]．New York City，USA：Association for Computational Linguistics．2006．33—4
3. Son M， Chai J． Discourse processing for context question answering based oil linguistic knowledge knowledge-based systerns lJJ． Special Issues on Intelligent User Interfaces，2007，20(6)：511．526．
4. Sun M, Chai J Y. Discourse processing for context question answering based on linguistic knowledge [J]. Knowledge-Based Systems, 2007, 20(6): 511-526.

5. Chai J Y, Jin R. Discourse structure for context question answering[C] Proceedings of the Workshop on Pragmatics of Question Answering at HLT-NAACL 2004. 2004: 23-30.

6. Kirschner M, Bernardi R, Baroni M. Analyzing interactive QA dialogues using logistic regression models[M] AI* IA 2009: Emergent Perspectives in Artificial Intelligence. Springer Berlin Heidelberg, 2009: 334-344.

7. Rieser V, Lemon O. Learning effective multimodal dialogue strategies from Wizard-of-Oz data: Bootstrapping and evaluation[C] Proceedings of ACL. 2008: 638-646.

8. Van Schooten B W, Op den Akker R, Rosset S, et al. Follow-up question handling in the IMIX and Ritel systems: A comparative study[J]. Natural Language Engineering, 2009, 15(01): 97-118.

9. Kelly D, Lin J. Overview of the TREC 2006 ciQA task[C] ACM SIGIR Forum. ACM, 2007, 41(1): 107-116.

10. Dang H T, Lin J, Kelly D. Overview of the TREC 2007 question answering track[C]//Proceedings of TREC. 2007, 2007(5.3): 3.

11. Hickl A，Harabagiu S． Enhanced interactive question-answering with conditional random fields[A]．Proc of the IQA Workshop at HLT-NAACL[CJ. New York City，USA：Associationfor Co mputationalLinguistics． 2006． 25—32．

12. Misu T, Georgila K, Leuski A, et al. Reinforcement learning of question-answering dialogue policies for virtual museum guides[C]. Proceedings of the 13th Annual Meeting of the Special Interest Group on Discourse and Dialogue. Association for Computational Linguistics, 2012: 84-93.

13. Rzeniewicz J, Szymański J, Duch W. Adaptive Algorithm for Interactive Question-Based Search[M]. Intelligent Information Processing VI. Springer Berlin Heidelberg, 2012: 186-195.

14. Burke R D, Hammond K J, Kulyukin V, et al. Question answering from frequently asked question files: Experiences with the faq finder system[J]. AI magazine, 1997, 18(2): 57.

15. Pal S, Bhattacharya S, Datta I, et al. A Framework For Automatic Generation Of Answers To Conceptual Questions In Frequently Asked Question (FAQ) Based Question Answering System [J]. INTERNATIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE, 2012, 1(2).

16. Cong G, Wang L, Lin C Y, et al. Finding question-answer pairs from online forums[C]//Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2008: 467-474.

17. Cao Z, Qin T, Liu T Y, et al. Learning to rank: from pairwise approach to listwise approach[C]//Proceedings of the 24th international conference on Machine learning. ACM, 2007: 129-136.

18. Liu T Y. Learning to rank for information retrieval[J]. Foundations and Trends in Information Retrieval, 2009, 3(3): 225-331.

19. Verberne S. In search of the why: Developing a system for answering why-questions[M]. [Sl: sn], 2010.

20. Joachims T. Optimizing search engines using clickthrough data[C] Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2002: 133-142.

21. Yue Y, Finley T, Radlinski F, et al. A support vector method for optimizing average precision[C] Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2007: 271-278.

22. Furey T S, Cristianini N, Duffy N, et al. Support vector machine classification and validation of cancer tissue samples using microarray expression data[J]. Bioinformatics, 2000, 16(10): 906-914.

23. A. Mood, F. Graybill, and D. Boes. Introduction to the Theory of Statistics. McGraw-Hill, 3 edition, 1974.

24. Zhu M. Recall, precision and average precision[J]. Department of Statistics and Actuarial Science, University of Waterloo, Waterloo, 2004.

25. Dong Z, Dong Q. HowNet[J]. 2000.

26. Q. Liu and S. J. Li, Word semantic similarity computation based on HowNet, Proc. 3rd Chinese Word Semantic Conference (2002).

27. Blunsom P, Kocik K, Curran J R. Question classification with log-linear models[C]//Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2006: 615-616.

28. Chang C C, Lin C J. LIBSVM: a library for support vector machines[J]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3): 27.