A Hierarchical Relevance Feedback Algorithm for Improving the Precision of Virtual Tutoring Assistant Systems

JI-WEI WU, JUDY C.R. TSENG
Department of Computer Science and Information Engineering
Chung Hua University
Hsinchu, Taiwan 300, ROC
{m09502050, judycrt}@cc.chu.edu.tw

Abstract: In recent years, several virtual tutoring assistant systems had been proposed based on question-answering systems. These virtual tutoring assistant systems are very helpful for the students to get instant helps when the teachers are not available. Pedagogic scholars think that when a student is stuck on a certain problem while learning, an instant tutoring assistant is very helpful to promote his/her study. Therefore, the qualities (precision) of the answers of virtual tutoring assistant systems are very important. It determines the effectiveness of these systems. In practices, some students are suffered from not being able to properly express their need; some may not even know exactly what information they need. As a result, some problems do not match any existing solutions, even if they do contain some clues for finding solutions. That makes the effectiveness of the virtual tutoring assistant systems not being acceptable. In the literatures, researchers found that relevance feedback information are quite useful for information retrieval systems to improve their effectiveness. Among them, the Rocchio's Relevance Feedback (RRF) algorithm is the most well-known and have been employed in several information retrieval systems. In this paper, we proposed a novel pseudo relevance feedback algorithm, called Hierarchical Relevance Feedback (HRF) algorithm. The HRF algorithm is used for improving the precision of virtual tutoring assistant systems. After a query is submitted by some student, the new virtual tutoring assistant system will automatically modify the student's query according to the HRF algorithm and re-submit it to the system. Experimental results showed that the effectiveness of the system is improved by automatically modifying user's query using the HRF algorithm. Moreover, the HRF algorithm is also outperformed the famous RRF algorithm.

Keywords: E-learning, Tutoring assistant, Information Retrieval, Question-answering, Relevance feedback

1 Introduction

The development of network not only facilitates the acquirement of information, but also highlights the value and feasibility of e-learning. Most of the existing e-learning platforms focus on providing contents and performing assessments, however, tutoring assistance is less to be considered. Pedagogic scholars think that when a student is stuck on a certain problem while learning, an instant tutoring assistant is very helpful to promote his/her study. Nevertheless, the teacher is not always available to answer students' question immediately. Moreover, it would be too time-consuming for a teacher to answer students' questions one by one, especially for a popular e-course where there are thousands of students and most of the students' questions are same-alike.

Several virtual tutoring assistant systems are proposed and developed [1, 7, 8, 12]. These systems provide instant assistants for students' learning problems even when the teachers are not available. However, the qualities of the answers determine the effectiveness of these systems. In practices, some students are suffered from not being able to properly express their need; some may not know exactly what information they need. Sometimes, even though the students properly express exactly what they need, the information returned by the systems are not relevant to the students' need. That makes the effectiveness of the virtual tutoring assistant systems not being acceptable.

In the literatures, researchers found that relevance feedback information are quite useful for information retrieval systems to improve their effectiveness. Among them, the Rocchio's Relevance Feedback (RRF) algorithm is the most well-known and have been employed in several information retrieval systems.

To improve the effectiveness of a virtual tutoring assistant system, a novel pseudo relevance feedback algorithm, called Hierarchical Relevance Feedback

(HRF) algorithm, is proposed in this paper. After a query is submitted by some student, the new virtual tutoring assistant system will automatically modify the student's query according to the HRF algorithm and re-submit it to the system. Experimental results showed that the effectiveness of the system is improved by automatically modifying user's query using the HRF algorithm. Moreover, the HRF algorithm is also outperformed the famous RRF algorithm.

2 Background and Related Works

Recent progress of computer and network technologies has encouraged the development of web-based learning environments [3, 4, 5, 6, 7, 8, 9]. However, as most of the existing web-based learning systems simply provide subject materials for browsing, the students are likely to be stuck while encountering problems during the learning process without instant aid, and hence their learning performances could be significantly affected [10].

Some researchers attempted to employ on-line discussion groups to cope with the problem. Nevertheless, most of the answers obtained from the discussion groups could be incorrect or incomplete; therefore, the most desirable approach is to obtain the answers directly from the teacher. Unfortunately, for a popular on-line course with thousands of students, it is almost impossible for the teacher to answer every question submitted by the students, not to mention the provision of instant aids to them. Rau et al. [11] indicated that, without face-to-face interaction, it is important to provide immediate help and interactions while proceeding online instruction.

Several virtual tutoring assistant systems [7, 8, 12] had been proposed based on question-answering systems [13]. However, most of the services provided are not ubiquitous, which are unable to provide instant problem-solving services for mobile learners. Furthermore, the solutions rely heavily on teachers to add onto the systems manually, which consume a lot of labor hours and thus reduce the efficiency of problem-solving.

To cope with these problems, Judy et al. [14] proposed a course content restructuring approach to reduce the high communication cost required in a mobile learning environment. Wu et al. [1] propose and develop a Ubiquitous Virtual Tutoring Assistant System (UVTAS) which incorporates a supplement-material base as well as a solution extraction module to search, analyze, and extract automatically solutions from the supplement-material base, which relief the burden of solving problems manually by the teachers. Furthermore, no matter where a student is, while he/she encounters or thinks about a learning problem, he/she can obtain immediate solutions via his/her mobile device, such as a cell-phone or a PDA. UVTAS not only provides a web-interface for mobile learners, it also incorporates a SMS-interface of question-answering to support mobile learners while Internet access is not available.

Although UVTAS can automatically search, analyze, and extract solutions from the supplement-material base to solve student's problems, sometimes the students are not fully satisfied by the given solutions. We know that the effectiveness of a virtual tutoring assistant system heavily depends on the quality of solution it provides. Besides of expanding the solution base as UVTAS does, the precision of the information retrieval algorithm used plays another important role.

In the literatures, researchers found that relevance feedback information are quite useful for information retrieval systems to improve their effectiveness. Among them, the Rocchio's Relevance Feedback (RRF) algorithm [2] is the most well-known and will be discussed in the following section.

3 The Process of Relevance Feedback

The process of information retrieval with relevance feedback [2, 15, 16] is shown in Fig. 1. First, each of the solutions in the solution-base is represented as a Characteristic Vector (CV) by the vector space model [17] as follows:

$$\mathrm{Q}_i = \{(K_1, W_{i,1}), (K_2, W_{i,2}), \ldots, (K_j, W_{i,j}), \ldots, (K_n, W_{i,n})\} \quad (1)$$

where K_j represents *j-th* keyword and W_{ij} is the weighting of K_j in the *i*-th solution.

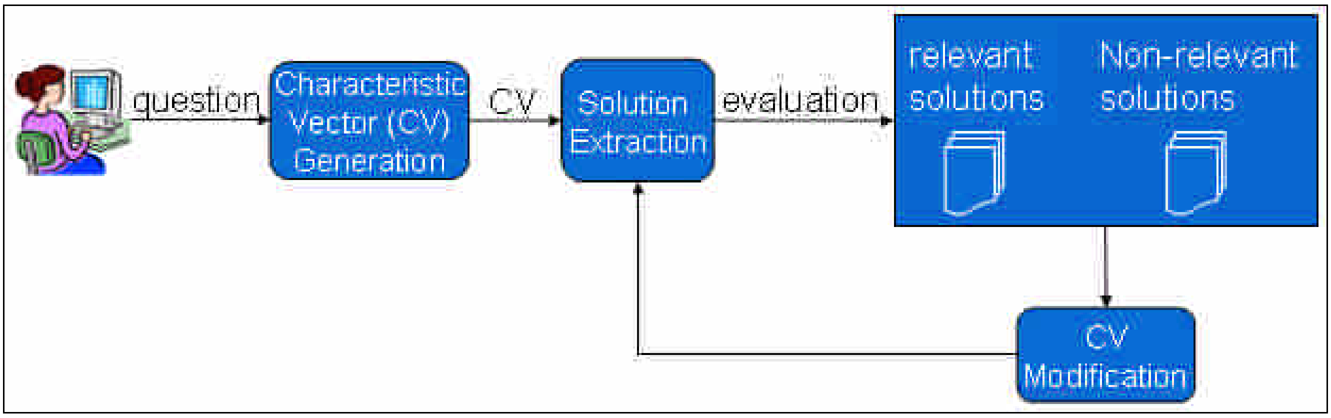

Fig. 1 The process of the relevance feedback algorithm

To construct the characteristic vector, the question is transformed into an indexed form by some basic steps, such as tokenization, stop word removal, stemming and term weighting [18]. An example is shown in Fig. 2.

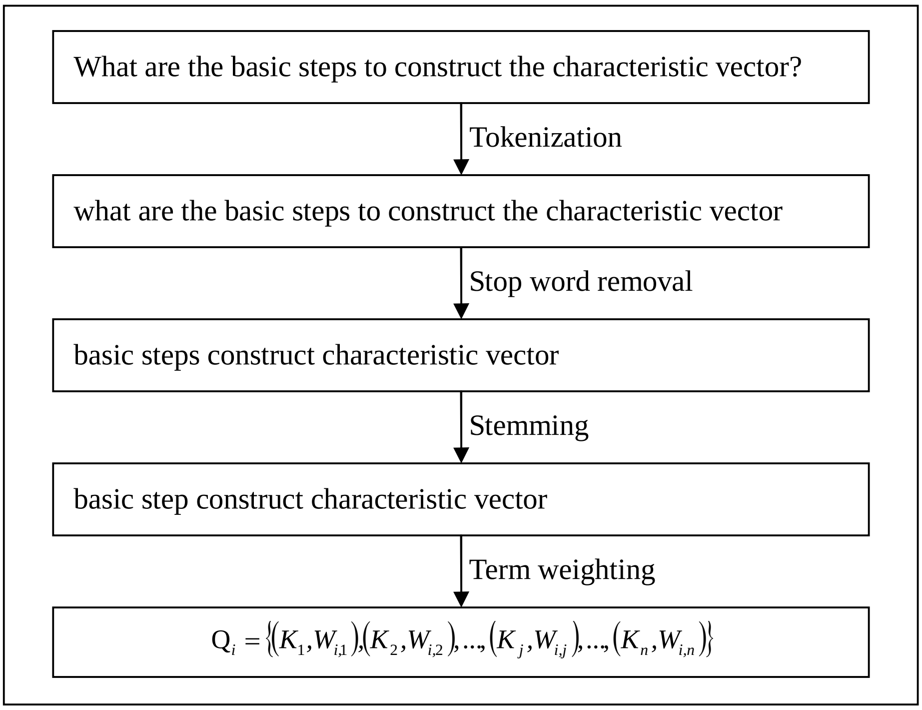

Fig. 2 An example of the basic steps to construct the CV

Tokenization

Convert the problem into a stream of terms, including converting all the terms into lower case and removing punctuation characters. For example, a sentence "What is Tokenization?" can be converted into "what is tokenization".

Stop word removal

Remove all the common terms (e.g. a, and, the, not etc.) which will not being used in document retrieval.

Stemming

Reduce terms to their root variant to avoid the user having to instantiate every possible variation of each query term (e.g. stemming → stem).

Term weighting

Evaluating the importance of each term to the problem it is contained. One of the best known schemes is the TF×IDF (term frequency × inverse document frequency) [17]. It is a statistical measure for evaluating the importance of a term. The importance increases proportionally to the frequency of the term (Term Frequency, TF) in the problem, but is offset by the frequency of the term in the whole solutions.

When a question is submitted by a user, the characteristic vector of the question is also constructed by vector space model as follows:

$$Q = \{(K_1, W_1), (K_2, W_2), \ldots, (K_j, W_j), \ldots, (K_n, W_n)\} \quad (2)$$

Q is then compared to each Q_i to find out the similarity between the user's question and each of the solutions. Inner product is usually used in this step as shown in Eq. (3).

$$Sim(Q, Q_i) = \sum_{k=1}^{n} (W_k \times W_{ik}) \quad (3)$$

In a question-answering system without relevance feedback, the question with the largest similarity measure is judged as the most feasible solution to the user's query. Nevertheless, relevance feedback information can be used to modify the original query to obtain more precise information. In the literatures, relevance feedback information come from two folds: one from user designation, which is more accurate but consume a lot of user's efforts; another from system designation, called *pseudo relevance feedback* or *blind relevance feedback*, which releases the user's burden while the promise of improving precision has been shown [2,19, 20].

4 The Rocchio's Relevance Feedback (RRF) Algorithm

Rocchio's RF algorithm is a commonly used algorithm in researches of relevance feedback. It is a query modification method which consists of adding or removing terms and term re-weighting using relevance judgments. After ranking the feasible solutions by their similarity measurement to the user's query, the question-answering system based on the RRF algorithm will select the first n_1 solutions as relevant solutions, while selecting the last n_2 solutions as irrelevant solutions. The characteristic vector of the user's question Q is then modified by the following formula:

$$Q' = \alpha \times Q + \frac{\beta}{n_1} \times \sum_{i=1}^{n_1} R_i - \frac{\gamma}{n_2} \times \sum_{i=1}^{n_2} S_i \quad (4)$$

where Q' is the new query vector, Q is the original query vector, n_1 is the number of relevant solutions, n_2 is the number of the irrelevant solutions, R_i is the *i-th* vector of relevant solutions, S_i is the *i-th* vector of irrelevant solutions. α, β and γ are constants, specify the degree of effect of each component in the RRF algorithm. Typically, α=1 and $\beta + \gamma = 1$ [19, 20].

The new query vector is then used by the system to retrieve feasible solutions again. Previous works found that user's are more satisfied by the new solutions.

5 The Hierarchical Relevance Feedback (HRF) Algorithm

Although the RRF algorithm improves the retrieval performance by reformulating the original query, we found that the degrees of relevance and irrelevance are not discriminated well in Rocchio's algorithm and its varieties [19, 20]. To cope with this problem, the HRF algorithm is proposed. In order to reduce the user's burden, we adopt the pseudo RF as our basic scheme to reformulation the original query vector as Rocchio's algorithm. However, the rankings of each solutions considered, either relevant or irrelevant, are used to represent the degree of relevance/irrelevance.

The HRF algorithm is given as follow:

$$Q' = \alpha \times Q + \frac{\beta}{n_1} \times \sum_{i=1}^{n_1} (WR_i \times R_i) - \frac{\gamma}{n_2} \times \sum_{i=1}^{n_2} (WS_i \times S_i) \quad (5)$$

where WR_i represent the degree of relevance for each relevant solutions and WS_i represent the degree of irrelevance for each irrelevant solutions. We define the WR_i and WS_i as Eq. (6) and Eq. (7) according to the ranks of relevant (or irrelevant) solutions.

$$WR_i = (n_1 - i) + 1 \quad (6)$$

$$WS_i = (n_2 - i) + 1 \quad (7)$$

In Eq. (6) and Eq. (7), the inverse ranked values are given as the degree of relevant (or irrelevant). For example, assume that the first 5 ranked solutions are considered to be relevant, then the degree of relevance, WR_1, of the top one relevant solution is (5-1+1) = 5, the degree of relevance, WR_2, of the second relevant solution is (5-2+1) =4, and so on.

In the next section, experimental results will show the superiority of the HRF algorithm to the RRF algorithm.

6 Experiments and Evaluation

6.1 Test Collection

We need a test collection that contains: (1) a set of documents (2) a set of queries (3) a list of judged relevant documents for each query. We used the test collections Medlars to evaluate the performance of our proposed algorithm. Medlars is publicly available at ftp://ftp.cs.cornell.edu/pub/smart/med/.

Medlars is based on MEDLINE reference collections from 1964 to 1966. There are 1033 documents and 30 queries in the collection. An example of query is shown as Table 1, and an example of documents is shown as Table 2. For each query, there is a list of documents associated with it for relevant judgments. The relevant judgments are provided by human experts.

Table 1 Query example of Medlars

.I 1
.W
the crystalline lens in vertebrates, including humans

Table 2 Documents example of Medlars

.I 1
.W
correlation between maternal and fetal plasma levels of glucose and free fatty acids . correlation coefficients have been determined between the levels of glucose and ffa in maternal and fetal plasma collected at delivery . significant correlations were obtained between the maternal and fetal glucose levels and the maternal and fetal ffa levels . from the size of the correlation coefficients and the slopes of regression lines it appears that the fetal plasma glucose level at delivery is very strongly dependent upon the maternal level whereas the fetal ffa level at delivery is only slightly dependent upon the maternal level .

6.2 Effectiveness Measurement

To measure the performance of solution retrieval, we use Mean Average Precision (MAP) [21] and the well-known precision and recall. Precision is portion of the retrieved solutions which are relevant. It is used to measure the ability to retrieve top-ranked solutions that are mostly relevant and defined as:

$$precision = \frac{A}{A+B}$$

Recall is portion of the relevant solutions which had been retrieval. It is used to measure the ability of the search to find all of the relevant solutions and defined as:

$$recall = \frac{A}{A+C}$$

Under those measurements, the parameters A, B and C are defined as Table 3. It can also present as Fig. 3 and Fig. 4.

Table 3 Definition of the parameter A, B and C

Parameter	Definition
A	Number of relevant solutions retrieved
B	Number of irrelevant solutions retrieved
C	Number of relevant solutions not retrieved

MAP has been shown to have especially good discrimination and stability. It is defined as:

$$MAP(Q) = \frac{1}{|Q|}\sum_{j=1}^{|Q|}\frac{1}{m_j}\sum_{k=1}^{m_j} Precision(R_{jk}) \qquad (8)$$

where Q is presented as query set. For each $q_j \in Q$, there is a set of relevant solutions for it, $\{d_1, . . . d_{mj}\}$. And R_{jk} is a set of solutions from the top solution until get to solution d_k. For example, there is a user's query and four solutions relevant to it. Assume the four relevant solutions are ranked as rank 1, 2, 4 and 7 in ranked solutions respectively. The MAP is computed as:

$$MAP = \frac{\frac{1}{1}+\frac{2}{2}+\frac{3}{4}+\frac{4}{7}}{4} = 0.83$$

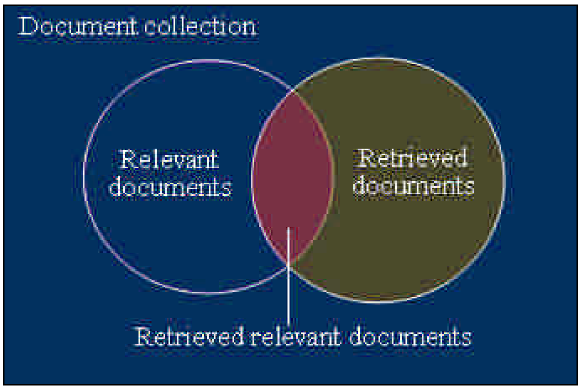

Figure 3. The precision figure

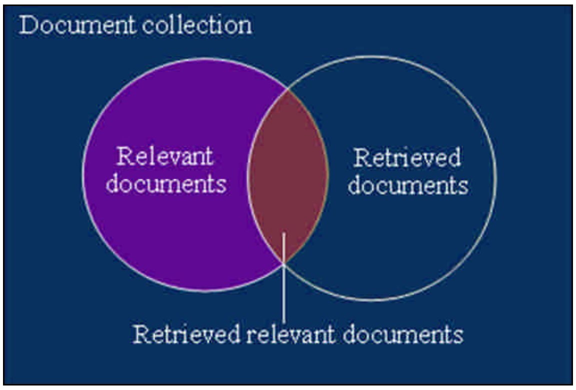

Figure 4. The recall figure

For the effectiveness of the virtual tutoring assistant systems, the higher average precision, in other words, obtain more relevant solutions in a fixed number of retrieved documents. Nevertheless, the higher MAP, in other words, make the relevant documents ranked in the ranking list more front.

6.3 Experimental Results and Analysis

Firstly, we evaluate the performance of RRF algorithm by four representative parameter sets (as shown in Table 4) to find a best parameter set for RRF. The parameter set 1, α = 1, β = 1 and γ = 0, means only relevant documents are used as the feedback information. The parameter set 2, α = 1, β = 0.75 and γ = 0.25, means relevant documents are the higher portion of the feedback information in RRF algorithm. The parameter set 3, α = 1, β = 0.5 and γ = 0.5, means relevant documents are equal to irrelevant documents in RRF algorithm. The parameter set 4, α = 1, β = 0.25 and γ = 0.75, means relevant documents are the lower portion of the feedback information in RRF algorithm.

Table 4 Four type parameter set for RRF algorithm

Parameter set	Parameter value
Set 1	α = 1, β = 1 and γ = 0
Set 2	α = 1, β = 0.75 and γ = 0.25
Set 3	α = 1, β = 0.5 and γ = 0.5
Set 4	α = 1, β = 0.25 and γ = 0.75

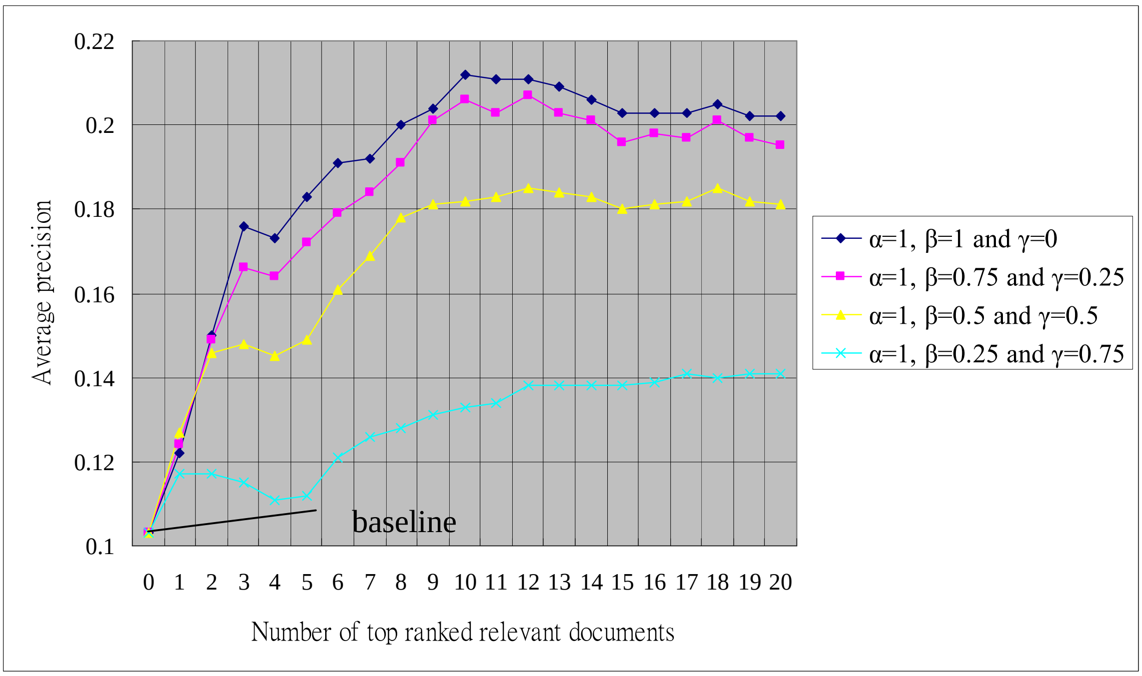

Fig. 5 Precision comparison of four parameter set for RRF algorithm when recall = 1

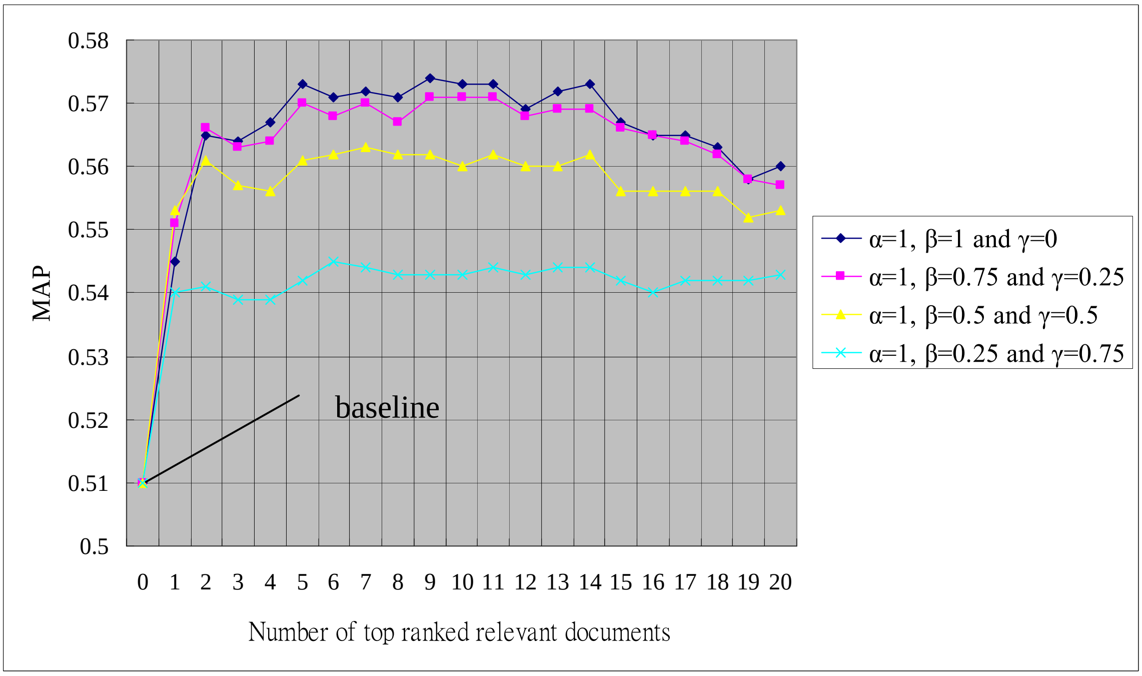

Fig. 6 MAP comparison of four parameter set for RRF algorithm when recall = 1

Fig. 5 and Fig. 6 are the results of evaluate the performance of RRF algorithm by four representative parameter sets. Fig. 5 represents the precision of the performance of RRF algorithm by four parameter sets. Fig. 6 represents the MAP of the performance of RRF algorithm by four parameter sets.

Each of the precision and MAP in Fig. 5 and Fig. 6 are compared when the number of top ranked relevant documents considered is ranged from 1 to 20 under the same recall (set to 1 in our experiments). When the number of the top ranked relevant documents is equal to 0, no relevant information will be used. That is, no query modification will be occurred. This case is called "baseline", which will be used to evaluate the effect of relevance feedback.

Our experiments show that the parameter set 1, $\alpha = 1$, $\beta = 1$ and $\gamma = 0$, outperformed parameter sets 2, 3 and 4. We also found that the parameter set 2 outperformed parameter set 3 and 4 and the parameter set 3 outperformed parameter set 4. According to the results, we conclude that the relevance feedback information with higher portion of relevant documents in RRF algorithm performs better than that with lower portion of relevant documents. That is, it is more advantageous not to consider irrelevant feedback information. Hence, only parameter set 1 is used in the following experiments.

The precision and MAP of the two algorithms, RRF and HRF, are compared when the number of top ranked relevant documents considered is ranged from 1 to 20 under the same recall (also set to 1 in these experiments). The experimental results are show in Fig. 7 and Fig. 8. From the experimental results, both RRF and HRF outperform the baseline in precision, as well as in MAP.

In Fig. 7 and Fig. 8, the precision and MAP of RRF and HRF algorithm are almost the same when the number of top ranked relevant documents is small (about 1 to 5). It is because of the relevance feedback information is insufficient. However, when the number of top ranked relevant documents is increased, the HRF algorithm outperforms the RRF algorithm. Moreover, when the number of top ranked relevant documents is bigger than 10, the performance of RRF algorithm is decreased, while the performance of HRF algorithms is increased. It proved that the degrees of relevance and irrelevance are better discriminated in the HRF algorithm.

When the number of top ranked relevant documents is bigger than 3, HRF outperforms RRF algorithm at about 1.7~14.9% in average precision and 0.4~4.3% in MAP. It means that using the HRF algorithm to modify user queries will obtain more relevant solutions.

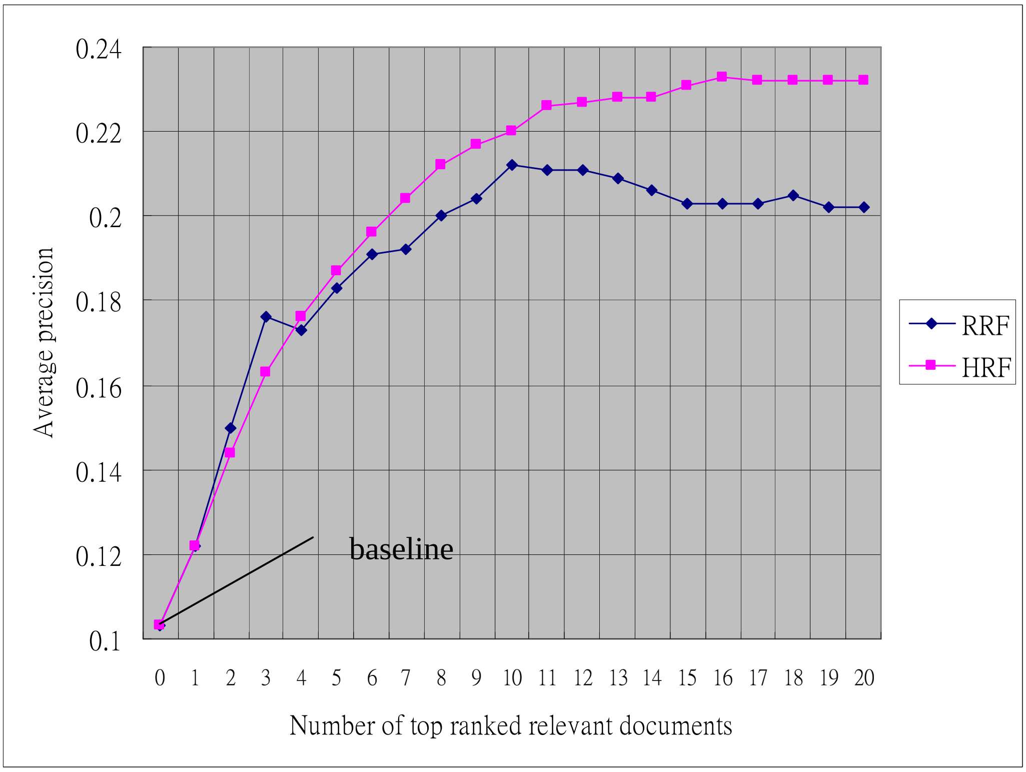

Fig. 7 Precision comparisons when Recall = 1

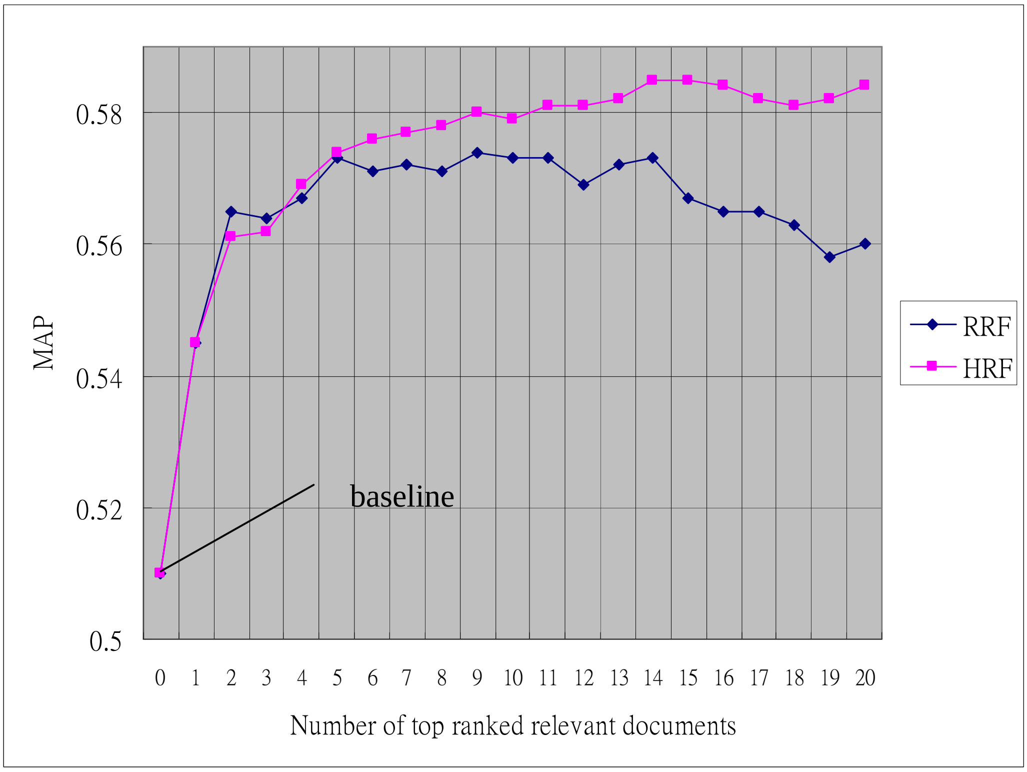

Fig. 8 MAP comparisons when Recall = 1

7 Conclusions and Future Works

In this paper, A Hierarchical Relevance Feedback (HRF) algorithm is proposed and employed in designing a virtual tutoring assistant system. In our novel approach, the degrees of relevance and/or irrelevance are considered, which improve the effectiveness of a virtual tutoring assistant system. Through our experiments, no matter precision or MAP is evaluated, the HRF algorithm outperforms the famous RRF algorithm and the baseline (no relevance feedback at all). From the experimental results, HRF outperforms RRF algorithm at about 1.7~14.9% in average precision and 0.4~4.3% in MAP.

In the near future, variations of the HRF algorithm, which incorporate various measurements to determine the degrees of relevance/irrelevance, will be proposed and compared to the original design. Extensive experiments with larger test collections will be done to further verify the superiority of the HRF.

Although relevance feedback will clearly improve the quality of virtual tutoring assistant systems, the results of query modification are not kept in the system for future use. Furthermore, the user's profile or preferences are not considered in the information retrieval process [22, 23, 24, 25, 26]. To make the virtual tutoring assistant systems more effective, in our future works, we will focus on adaptiveness to make the virtual tutoring assistant system more intelligent.

Reference:

[1] Wu, J.J., Tsou, Y.H., Chiou, C.K. and Tseng, J.C.R (2007). Development of a Ubiquitous Virtual Tutoring Assistant System, *Proceedings of World Conference on Educational Multimedia, Hypermedia and Telecommunications 2007* (pp. 1577-1585).

[2] Rocchio, J.J. (1971), Relevance feedback in information retrieval, In: G. Salton (ed.), *The Smart retrieval system: experiments in automatic document processing*, Prentice Hall, pp. 313-323.

[3] Hwang, G.J. (2003), A Concept Map Model for Developing Intelligent Tutoring Systems, *Computers & Education*, 40 (3), 217-235.

[4] Hwang, G.J., Hsiao, J.L., and Tseng, J.C.R. (2003), A Computer-Assisted Approach for Diagnosing Student Learning Problems in Science Courses, *Journal of Information Science and Engineering*, 19 (2), 229-248.

[5] Hwang, G.J., Lin, B.M.T. and Lin, T.L. (2006), An Effective Approach for Test-Sheet Composition from Large-Scale Item Banks, *Computers & Education*, 46 (2), 122-139.

[6] Tseng, J.C.R. and Hwang G.J. (2004), A Novel Approach to Diagnosing Student Learning Problems in E-Learning Environments, *WSEAS Transactions on Information Science and Applications*, 1(5), 1295-1300.

[7] Wang, C.C., Hung, J.C., and Yang, C.Y. and Shih, T.K. (2006), An Application of Question Answering System for Collaborative Learning, *Proceedings of the 26th IEEE International Conference on Distributed Computing Systems Workshops(ICDCSW'06)*.

[8] Wang, C.C., Hung, J.C., Shih, T.K. and Lin, H.W. (2006), A Repository-Based Question Answering System for Collaborative E-learning, *Journal of Computers*, 17 (3), October 2006.

[9] Hwang, G.J., Yin, P.Y., Tseng, J.C.R. and Hwang, G.H. (2007), An Enhanced Genetic Approach to Optimizing Auto-reply Accuracy of An E-learning System, *Computers & Education*.

[10] Jonassen, D.H. (2000), *Computers as mind tools for schools*, Merrill, Upper Saddle River: NJ.

[11] Rau, P.L. Patrick, Chen, S.H. and Chin, Y.T. (2004), Developing web annotation tools for learners and instructors, *Interacting with Computers*, 16, 163-181.

[12] Wei, C.C. and Tseng, J.C.R. (2006), The Development of an On-line Virtual Assistance System Based on the Automatic Document Classification Mechanism, *Proceedings of the 2006 Taiwan E-Learning Forum (TWELF 2006)*, Tainan, Taiwan.

[13] Tseng, J.C.R., and Hwang, G.J. (2006), Development of an Automatic Customer Service System on the Internet, *Electronic Commerce Research and Applications*, 6, 18-29.

[14] Tseng, J.C.R, Hwang, G.J. and Chan, Y. (2005), An Efficient Approach to Restructuring Subject Materials in Mobile Learning Environments, *WSEAS Transactions on Advances in Engineering Education*, 302-308.

[15] Harman, D. (1992), Relevance feedback revisited, *In Proceedings of the Fifth International SIGIR Conference on Research and Development in Information Retrieval*, pages 1-10.

[16] Salton, G. and Buckley, C. (1990), Improving retrieval performance by relevance feedback, *Journal of the American Society of Information Science*, 41:288-297.

[17] Salton, G. and McGill, M.J. (1983), *Introduction to Modern Information Retrieval*, New York N.Y.: McGraw-Hill.

[18] Ruthven, I. and Lalmas, M. (2003), A survey on the use of relevance feedback for information access systems, *Knowledge Engineering Review*, 18 (2). pp. 95-145.

[19] Jordan, C. and Watters, C. (2004), Extending

the Rocchio Relevance Feedback Algorithm to Provide Contextual Retrieval, *In the Proceedings of AWIC04*, 135-144.

[20] Eichmann, D. and Srinivasan, P. (2002). Adaptive filtering of newswire stories using two-level clustering, *Information Retrieval*, 5 (2/3), 209-237.

[21] Manning, C.D., Raghavan, P., and Schütze, H. (2008), *Introduction to Information Retrieval*, Cambridge University Press.

[22] Tamine-Lechani, L., Boughanem, M. and Zemirli, N. (2007), Exploiting Multi-Evidence from Multiple User's Interests to Personalizing Information Retrieval, *IEEE International Conference on Digital Information Management (ICDIM)*, p. 7-12.

[23] Chirita, P.A., Firan, C.S. and Nejdl, W. (2007), Personalized Query Expansion for the Web, *Proceedings of SIGIR'07*, pp. 7-14.

[24] Degemmis, M., Lops, P., Ferilli, S., Mauro, N.D., Basile, T.M.A. and Semeraro, G. (2005), Learning User Profiles from Text in e-Commerce, *proceedings of the First International Conference on Advanced Data Mining and Applications (ADMA)*, pp. 370-381.

[25] Economides A.A. (2005), Personalized Feedback in CAT, *WSEAS Transactions on Advances in Engineering Education*, Issue 3, Volume 2, pp. 174-181.

[26] Tseng, J.C.R, Chu, H.C., Hwang, G.J. and Tsai, C.C. (2007), Development of an Adaptive Learning System with Two Sources of Personalization Information, *Computers & Education*.