Relevance Feedback with Active Learning for Document Retrieval

Takashi Onoda† Hiroshi Murata† and Seiji Yamada‡
†Central Research Institute of Electric Power Industry, 2-11-1 Iwadokita, Komae-shi, Tokyo, 201-8511 Japan
‡National Institute of Informatics,2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 Japan
{onoda,murata}@criepi.denken.or.jp, seiji@nii.ac.jp

Abstract— **We investigate the following data mining problems from the document retrieval: From a large data set of documents, we need to find documents that relate to human interesting in as few iterations of human testing or checking as possible. In each iteration a comparatively small batch of documents is evaluated for relating to the human interesting. We apply active learning techniques based on Support Vector Machine for evaluating successive batches, which is called *relevance feedback*. Finally, our proposed approach is very useful for document retrieval with relevance feedback experimentally.**

I. Introduction

As progression of the internet technology, accessible information by end users is explosively increasing. In this situation, we can now easily access a huge document database through the WWW. However it is hard for a user to retrieve relevant documents from which he/she can obtain useful information, and a lot of studies have been done in information retrieval), especially document retrieval [20]. Active works for such document retrieval have been reported in TREC(Text Retrieval Conference) [17] for English documents, IREX(Information Retrieval and Extraction Exercise) [4] and NTCIR(NII-NACSIS Test Collection for Information Retrieval System) [8] for Japanese documents.

In most frameworks for information retrieval, a Vector Space Model(which is called VSM) in which a document is described with a high-dimensional vector is used [13]. An information retrieval system using a vector space model computes the similarity between a query vector and document vectors by cosine of the two vectors and indicates a user a list of retrieved documents.

In general, since a user hardly describes a precise query in the first trial, interactive approach to modify the query vector by evaluation of the user on documents in a list of retrieved documents. This method is called *relevance feedback* [12] and used widely in information retrieval systems. In this method, a user directly evaluates whether a document is relevant or irrelevant in a list of retrieved documents, and a system modifies the query vector using the user evaluation. A traditional way to modify a query vector is a simple learning rule to reduce the difference between the query vector and documents evaluated as relevant by a user.

In another approach, relevant and irrelevant document vectors are considered as positive and negative examples, and relevance feedback is transposed to a binary classification problem [9]. For the binary classification problem, SVM shows the excellent ability. And some studies applied SVM to the text classification problems [16] and the information retrieval problems [3].

We propose a relevance feedback framework with SVM as *active learning*. In contrast that a conventional SVM based relevance feedback system indicates a user a list of the most relevant documents, our system provides a user a list of documents which are hard for SVM to classify them and may be relevant for the user. This is a kind of active learning approach and we consider it promising for relevance feedback.

Okabe and Yamada [9] proposed a frame work in which relational learning to classification rules was applied to interactive document retrieval. Since the learned classification rules is described with symbolic representation, they are readable to our human and we can easily modify the rules directly using a sort of editor. However we consider SVM dealing with continuous values can do more precise classification than symbolic classification rules.

The relevance feedback is similar to what is termed active learning in that we try to maximize test performance using the smallest number of documents in the training set [16]. From an active learning point of view, we are interested in maximizing learning performance. Tong et al. proposed SVM based text classification method from an active learning point of view. The method tries to maximize learning performance. Drucker et al. applied SVM to the information retrieval [3]. At each retrieval, their method tries to maximize the number of useful documents, which are displayed to users. But it does not consider the learning performance. Documents are generally represented by the vector space model for the information retrieval. In this method, term frequency(TF) and binary representation are used as the vector space model. However, the conventional relevance feedback information retrieval method is useful in term frequency inverse document frequency(TFIDF) representation [12]. We are interested in comparing the performance between SVM based relevance feedback method and the conventional method in TFIDF representation. The detail of this difference will be described in the third section. And we propose the SVM based relevance feedback method, which can give many relevant documents for users at each retrievaland keep the learning performance.

In the remaining parts of this paper, we explain a SVM

0-7803-7898-9/03/$17.00 ©2003 IEEE

algorithm in the second section briefly, and an active learning with SVM for the relevance feedback in the third section. In the fourth section, in order to evaluate the effectiveness of our approach, we made experiments using a TREC data set of Los Angeles Times and discuss the experimental results. Eventually we conclude our work and discuss open problems in the fifth section.

II. Support Vector Machines

Formally, the Support Vector Machine (SVM) [18] like any other classification method aims to estimate a classification function $f : \mathcal{X} \to \{\pm 1\}$ using labeled training data from $\mathcal{X} \times \{\pm 1\}$. Moreover this function f should even classify unseen examples correctly.

In order to construct good classifiers by learning, two conditions have to be respected. First, the training data must be an unbiased sample from the same source (pdf) as the unseen test data. This concerns the experimental setup. Second, the size of the class of functions from which we choose our estimate f, the so-called capacity of the learning machine, has to be properly restricted according to statistical learning theory [18]. If the capacity is too small, complex discriminant functions cannot be approximated sufficiently well by any selectable function f in the chosen class of functions – the learning machine is too simple to learn well. On the other hand, if the capacity is too large, the learning machine bears the risk of overfitting.

In neural network training, overfitting is avoided by early stopping, regularization or asymptotic model selection [1], [7], [10], [11].

For SV learning machines that implement linear discriminant functions in feature spaces, the capacity limitation corresponds to finding a large margin separation between the two classes. The margin ϱ is the minimal distance of training points $(\mathbf{x}_1, y_1), \ldots, (\mathbf{x}_i, y_i), \mathbf{x}_i \in \mathbf{R}, y_i \in \{\pm 1\}$ to the separation surface, i.e.

$$\varrho = \min_{i=1,\ldots,\ell} \rho(\mathbf{z}_i, f) \tag{1}$$

where $\mathbf{z}_i = (\mathbf{x}_i, y_i)$ and

$$\rho(\mathbf{z}_i, f) = y_i f(\mathbf{x}_i), \tag{2}$$

and f is the linear discriminant function in some feature space

$$f(\mathbf{x}) = (\mathbf{w} \cdot \Phi(\mathbf{x})) + b = \sum_{i=1}^{\ell} \alpha_i y_i (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x})) + b, \tag{3}$$

with $\mathbf{w}$ expressed as $\mathbf{w} = \sum_{i=1}^{\ell} \alpha_i y_i \Phi(\mathbf{x}_i)$. The quantity Φ denotes the mapping from input space $\mathcal{X}$ by explicitly transforming the data into a feature space $\mathcal{F}$ using $\Phi : \mathcal{X} \to \mathcal{F}$. (see Figure 1). SVM can do so implicitly. In order to train and classify, all that SVMs use are dot products of pairs of data points $\Phi(\mathbf{x}), \Phi(\mathbf{x}_i) \in \mathcal{F}$ in feature space (cf. Eq. (3)). Thus, we need only to supply a so-called kernel function that can compute these dot products. A kernel function k allows to implicitly define the feature space (Mercer's Theorem, e.g. [2]) via

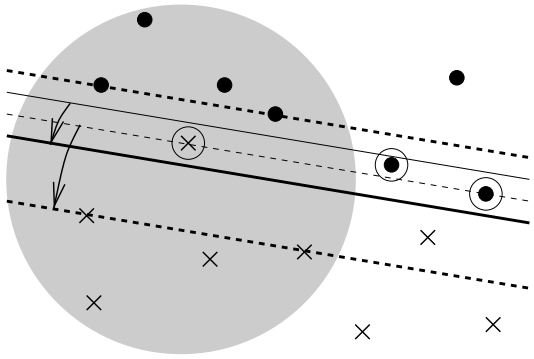

Fig. 1. A binary classification toy problem: This problem is to separate black circles from crosses. The shaded region consists of training examples, the other regions of test data. The training data can be separated with a margin indicated by the slim dashed line and the upper fat dashed line, implicating the slim solid line as discriminate function. Misclassifying one training example(a circled white circle) leads to a considerable extension(arrows) of the margin(fat dashed and solid lines) and this fat solid line can classify two test examples(circled black circles) correctly.

$$k(\mathbf{x}, \mathbf{x}_i) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{x}_i)). \tag{4}$$

By using different kernel functions, the SVM algorithm can construct a variety of learning machines, some of which coincide with classical architectures:

Polynomial classifiers of degree d:

$$k(\mathbf{x}, \mathbf{x}_i) = (\kappa \cdot (\mathbf{x} \cdot \mathbf{x}_i) + \Theta)^d \tag{5}$$

Neural networks(sigmoidal):

$$k(\mathbf{x}, \mathbf{x}_i) = \tanh(\kappa \cdot (\mathbf{x} \cdot \mathbf{x}_i) + \Theta) \tag{6}$$

Radial basis function classifiers:

$$k(\mathbf{x}, \mathbf{x}_i) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}_i\|^2}{\sigma}\right) \tag{7}$$

Note that there is no need to use or know the form of Φ, because the mapping is never performed explicitly The introduction of Φ in the explanation above was for purely didactical and not algorithmical purposes. Therefore, we can computationally afford to work in implicitly very large (e.g. 10^{10}- dimensional) feature spaces. SVM can avoid overfitting by controlling the capacity and maximizing the margin. Simultaneously, SVMs learn which of the features implied by the kernel k are distinctive for the two classes, i.e. instead of finding well-suited features by ourselves (which can often be difficult), we can use the SVM to select them from an extremely rich feature space.

With respect to good generalization, it is often profitable to misclassify some outlying training data points in order to achieve a larger margin between the other training points (see Figure 1 for an example).

This soft-margin strategy can also learn non-separable data. The trade-off between margin size and number of

misclassified training points is then controlled by the regularization parameter C (softness of the margin). The following quadratic program (QP) (see e.g. [18], [15]):

$$\begin{array}{lll} \min & \|\mathbf{w}\|^2 + C\sum_{i=1}^{\ell}\xi_i & \\ \text{s.t.} & \rho(\mathbf{z}_i, f) \geq 1 - \xi_i & \text{for all } 1 \leq i \leq \ell \\ & \xi_i \geq 0 & \text{for all } 1 \leq i \leq \ell \end{array} \quad (8)$$

leads to the SV soft-margin solution allowing for some errors.

III. Active learning with SVM in Information Retrieval

In this section, we describe the information retrieval system using relevance feedback with SVM from an active learning point of view. Fig. 2 shows the concept of the relevance feedback document retrieval. In Fig. 2, the user makes the initial retrieval by inputing the query. The result of the initial retrieval consists of too many documents, which are ranked by similarity between the query and the documents. But the rank of the documents is not usually useful for the user. In the relevance feedback information retrieval, the user can see the top N ranked documents and evaluate whether the documents are relevant or not. Then the evaluated documents with the initial query are given to a supervised learning algorithm to produce a new classifier. The classifier is used to generate the new rank of the initial retrieved documents. The new ranking ranks the actual relevant documents at higher levels than the previous ranking does. This re-rank makes until finding useful documents iteratively. In Fig. 2, the iterative procedure is the gray arrows parts. In the relevance feedback method, the user have to judge the re-ranked documents. Hence, it is difficult to use a large number of user judged documents for supervised learning algorithms because the user can not overcome much effort to judge the many documents. The SVMs have a great ability to discriminate even if the training data is small. Consequently, we propose to apply SVMs as the classifier in relevance feedback method. The retrieval steps of proposed method perform as follows:

Step 1: **Preparation of documents for the first feedback**
The conventional information retrieval system based on vector space model displays the top N ranked documents along with a request query to the user. In our method, the top N ranked documents are selected by using cosine distance between the request query vector and each document vector for the first feedback iteration.

Step 2: **Judgement of documents**
The user then classifiers these N documents into relevant or irrelevant. The relevant documents and the irrelevant documents are labeled. For instance, the relevant documents have "+1" label and the irrelevant documents have "-1" label after the user's classification.

Step 3: **Determination of the optimal hyperplane**
The optimal hyperplane for classifying relevant and irrelevant documents is determined by using a SVM which is learned by labeled documents(see Figure 3).

Step 4: **Discrimination documents and information retrieval**
The documents, which are retrieved in the Step1, are mapped into the feature space. The SVM learned by the previous step classifies the documents as relevant or irrelevant. The documents, which are discriminated relevant and in the margin area of SVM are selected. From the selected documents, the top N ranked documents, which are ranked using the distance from the relevant documents area, are shown to user as the information retrieval results of the system(see Figure 4). If the number of feedback iterations is more than m, then go to next step. Otherwise, return to Step 2. The m is a maximal number of feedback iterations and is determined by the user.

Step 5: **Display of the final retrieved documents**
The retrieved documents are ranked by the distance between the documents and the hyper-plane which is the discriminant function determined by SVM. The retrieved documents are displayed based on this ranking(see Figure 5).

In the reference [3], Drucker et al. selects the higher ranked documents, which are relevant and far from the

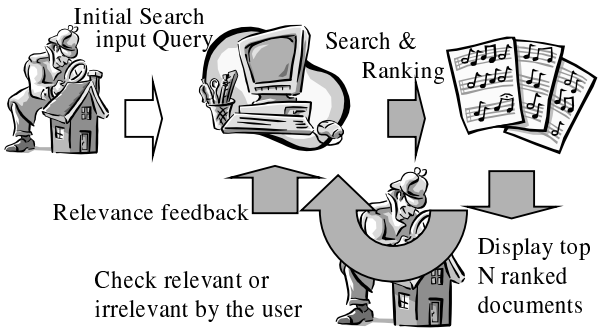

Fig. 2. Image of the relevance feedback documents retrieval: The gray arrow parts are made iteratively to retrieve useful documents for the user. This iteration is called feedback iteration in the information retrieval research area.

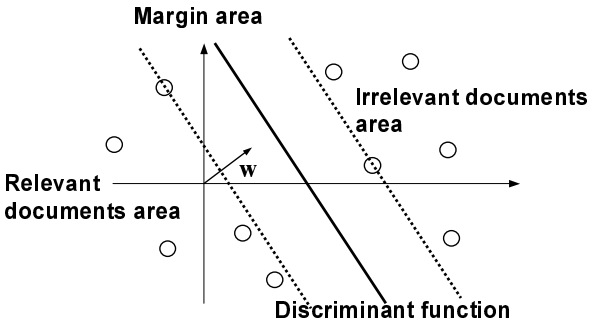

Fig. 3. Discriminant function for classifying relevant or irrelevant documents: Circles denote documents which are checked relevant or irrelevant by a user. The solid line denotes a discriminant function. The margin area is between dotted lines.

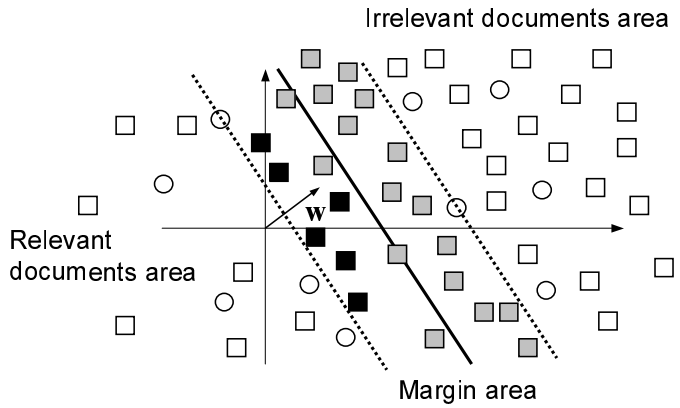

Fig. 4. Mapped non-checked documents into the feature space: Boxes denote non-checked documents which are mapped into the feature space. Circles denotes checked documents which are mapped into the feature space. Black and gray boxes are documents in the margin area. We show the documents which are represented by black boxes to a user for next iteration. These documents are in the margin area and near the relevant documents area.

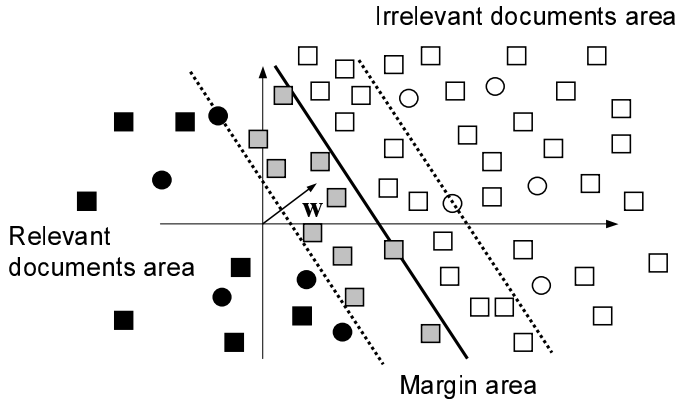

Fig. 5. Displayed documents as the result of document retrieval: Boxes denote non-checked documents which are mapped into the feature space. Circles denotes checked documents which are mapped into the feature space. The system displays the documents which are represented by black circles and boxes as the result of document retrieval to a user.

discriminant function. The selected documents do not neet to be in the margin area. The strategy may be able to show many relevant documents to the user. But it can not keep the efficient learning performance from an active learning point of view. Because the documents, which are on or near the discriminant function, should be selected to get the efficient learning performance [16]. In the reference [16], Tong et al. select the documents, which are on or near the discriminant function. This selection can make the efficient learning. However, users feel stress of the selection because it is difficult for the user to evaluate which the documents are relevant or irrelevant. The feature of our SVM based feedback is the selection of displayed documents to users in Step 4. Our proposed method selects the documents, which are discriminated in relevant and in the margin area, and near the relevant documents area. The documents may the relevant documents for the user, because the documents are near the relevant area. And the documents may be able to keep the learning performance, because the documents are in the margin area and have the useful information to make good learning performance. The margin area means the obscurity area of classification. Therefore, our selection can be expected that the efficient learning can be kept and users do not need to feel stress.

IV. Experiments

A. Experimental setting

We made experiments for evaluating the utility of our interactive document retrieval with active learning of SVM in section III. The document data set we used is a set of articles in the Los Angeles Times which is widely used in the document retrieval conference TREC [17]. The data set has about 130 thousands articles. The average number of words in a article is 526. This data set includes not only queries but also the relevant documents to each query. Thus we used the queries for experiments.

We used TFIDF [20], which is one of the most popular methods in information retrieval to generate document feature vectors, and the concrete equation [14] of a weight of a term t in a document d w_t^d are in the following.

$$
\begin{aligned}
w_t^d &= L \times t \times u \qquad (9)\\
L &= \frac{1+\log(tf(t,d))}{1+log(\text{average of } tf(t,d)ind)} \quad (tf)\\
t &= \log(\frac{n+1}{df(t)}) \quad (idf)\\
u &= \frac{1}{0.8+0.2\frac{uniq(d)}{\text{average of } uniq(d)}} \quad (normalization)
\end{aligned}
$$

The notations in these equation denote as follows:

- w_t^d is a weight of a term t in a document d,
- $tf(t,d)$ is a frequency of a term t in a document d,
- n is the total number of documents in a data set,
- $df(t)$ is the number of documents including a term t,
- $uniq(d)$ is the number of different terms in a document d.

The size N of retrieved and displayed results developed in *Step 1* in section III was set as twenty. The feedback iterations m were 1, 2, 3 and 4. In order to investigate the influence of feedback iterations on accuracy of retrieval, we used plural feedback iterations.

In our experiments, we used the linear kernel for SVM learning, and found a discriminant function for the SVM classifier in this feature space. The VSM of documents is high dimensional space. Therefore, in order to classify the labeled documents into relevant or irrelevant, we do not need to use the kernel trick and the regularization parameter C(see section II). The VSM consists of TFIDF representation. Drucker et al. did not use TFIDF representation for SVM learning [3]. And we used LibSVM [6] as SVM software in our experiment.

For comparison with our approach, two information retrieval methods were used. The first is an information

retrieval method that does not use a feedback. The second is an information retrieval method using conventional Rocchio-based relevance feedback [12] which is widely used in information retrieval research.

The Rocchio-based relevance feedback modifies a query vector Q_i by evaluation of a user using the following equation.

$$Q_{i+1} = Q_i + \alpha \sum_{x \in R_r} x - \beta \sum_{x \in R_n} x, \qquad (10)$$

where R_r is a set of documents which were evaluated as relevant documents by a user at the $i$the feedback, and R_n is a set of documents which were evaluated as irrelevant documents at the i feedback. α and β are weights for relevant and irrelevant documents respectively. In this experiment, we set $\alpha = 1.0$, $\beta = 0.5$ which are known adequate experimentally.

In general, retrieval accuracy significantly depends on the number of the feedback iterations. Thus we changed feedback iterations for 1, 2, 3, 4 and investigated the accuracy for each iteration.

We utilized *precision* and *recall* for evaluating the two information retrieval methods [5][19] and our approach. The following equations are used to compute *precision* and *recall*. Since a recall-precision curve is investigated to each query, we used the average recall-precision curve over all the queries as evaluation.

$$precision = \frac{\text{The No. of retrieved relevant doc.}}{\text{The No. of retrieved doc,}}$$

$$recall = \frac{\text{The No. of retrieved relevant doc.}}{\text{The total No. of relevant doc.}}$$

B. Experimental results

B.1 Comparing of recall-precision performance curves

In this section, we investigated the effectiveness of proposed method, when the user judged the twenty higher ranked documents at each feedback iteration. In the first iteration, twenty higher ranked documents were retrieved using cosine distance between document vectors and a query vector in VSM, which is represented by TFIDF. The query vector was generated by a user's input of keywords. In the other iterations, the user does not need to input keywords for the information retrieval, and the user labels "+1" and "-1" as relevant and irrelevant documents respectively.

Figure 6 show a recall-precision performance curve of our SVM based method, after four feedback iterations. For comparison, this figure also show the recall-precision curves of the conventional feedback method (i.e., Rocchio-based method) and VSM (i.e., without feedback). The thick solid line is the proposed method, the broken line is the conventional feedback method, and the thin solid line was the VSM without feedback.

This figure shows that the retrieval effectiveness of both feedback methods, i.e., proposed and conventional feedback methods, is improved compared with that of the VSM without feedback. In this result, we could confirm that the relevance feedback was useful technique for improving the performance of information retrieval in VSM.

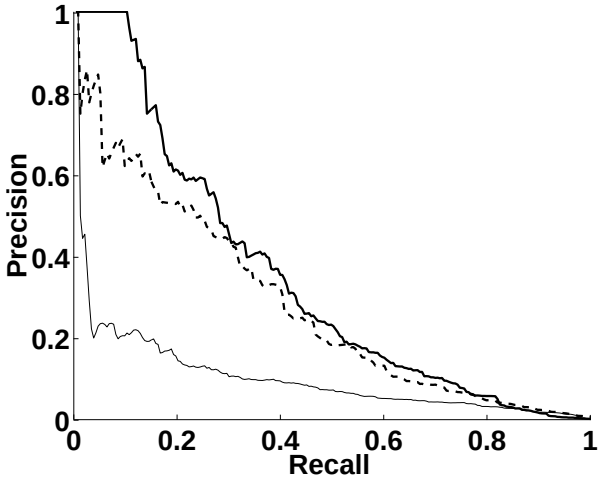

Fig. 6. The effectiveness of SVM based feedback: The lines show recall-precision performance curve by using twenty feedback documents on the set of articles in the Los Angeles Times after 4 feedback iterations. The wide solid line is proposed method, the broken line is conventional feedback method (i.e. Rocchio-based method), and the solid line is the VSM without feedback.

TABLE I
AVERAGE PRECISION USING SVM BASED FEEDBACK METHOD AND ROCCHIO-BASED FEEDBACK METHOD

No. of feedback iterations	Average precision	
	SVM	Rocchio
1	0.2625	0.2250
2	0.3500	0.2500
3	0.6125	0.2350
4	0.6375	0.2250

Furthermore, this figure also shows that the proposed feedback method improves the performance compared with conventional feedback method at all recall points. Consequently, in this experiment, we conclude that the SVM is a useful relevant feedback technique improving performance of information retrieval in VSM, which is represented by TFIDF.

B.2 Relationships between the performance and the number of feedback iterations

Here, we describe the relationships between the performances of proposed method and the number of feedback iterations. Table I gave the average precision result as a function of the number of feedback iterations. We carried out twenty times document retrieval. At each feedback iteration, the system displays twenty higher ranked relevant documents in the margin area's documents for our proposed method. We also show the average precision of Rocchio-based method for comparing to proposed method in table I.

We can see from this table that the SVM based relevance feedback approach gives the higher performance in propor-

TABLE II

IN A SPECIAL CASE, THE RELATIONSHIP BETWEEN THE NUMBER OF FEEDBACK ITERATIONS AND THE NUMBER OF ACTUAL RELEVANT DOCUMENTS IN TWENTY HIGHER RANKED RELEVANT DOCUMENTS IN WHOLE DOCUMENTS.

No. of feedback iterations	No. of relevant documents	
	SVM	Rocchio
1	9	11
2	18	13
3	20	12
4	20	11

tion to increase the number of feedback iterations. On the other hand, the Rocchio-based relevance feedback method degrades the retrieval performance nevertheless the number of feedback iterations increased from three to four. In a case of VSM without feedback, the average precision is 0.15. Hence, we can consider that the more feedback iterations, the better relevance documents can be obtained by using SVM based feedback method. Especially, the proposed method can improve the performance of conventional feedback method at each feedback iteration. We believe that the reason of these results is that the SVM can find a more suitable hyperplane for discriminating between relevant and irrelevant documents as increasing the number of the feedback iterations. After all, we can believe that the proposed method can keep effective learning from active learning point of view.

Furthermore, we compare the performance of proposed method to that of Rocchio-based feedback method from an information retrieval point of view. Table II shows the relationship between the number of feedback iterations and the number of actual relevant documents in twenty higher ranked relevant documents in a special case. In this case, five documents were labeled as relevant documents in twenty documents at the first iteration. In almost case, one or two documents were labeled as relevant documents in twenty documents at the first iteration. We can see from this table that our SVM based feedback can increase the number of actual relevant documents, which are useful for the user in proportion to increase the number of feedback iterations. On the other hand, the Rocchio-based feedback method degrades the number of actual relevant documents nevertheless the number of feedback iterations increased from three to four. Hence, we can consider that the proposed method can give the suitable number of actual relevant documents to the user.

V. CONCLUSION

In this paper, we proposed the relevance feedback method with the support vector machine (SVM) for the information retrieval. Because the SVM has an excellent ability to discriminate even if the training data is small, we applied the SVM to relevance feedback method. Experimental results on a set of articles in the Los Angeles Times showed the proposed method gave a consistently better performance than the conventional feedback method. Therefore our proposed SVM based approach is very useful for the information retrieval with relevance feedback.

In our experiments, we used TFIDF documents representation as VSM. Drucker et al. used binary documents representation and TF representation for estimating the performance of their proposed method. We plan to apply our proposed method to the binary representation and TF representation and compare our method with other SVM based methods(Drucker's method and Tong's method) experimentally. And this paper proposed that the system should display the documents which are discriminated relevant and in the margin area of SVM at each feedback iteration. However, we do not discuss how the selection of documents influence both the effective learning and the performance of information retrieval theoretically. This point is also our future work.

REFERENCES

[1] C. Bishop, *Neural Networks for Pattern Recognition*. Oxford: Clarendon Press, 1995.

[2] B. Boser, I. Guyon, and V. Vapnik, "A training algorithm for optimal margin classifiers," in *5th Annual ACM Workshop on COLT*, (D. Haussler, ed.), (Pittsburgh, PA), pp. 144–152, ACM Press, 1992.

[3] H. Drucker, B. Shahrary, and D. C. Gibbon, "Relevance feedback using support vector machines," in *Proceedings of the Eighteenth International Conference on Machine Learning*, pp. 122–129, 2001.

[4] IREX http://cs.nyu.edu/cs/projects/ proteus/irex/.

[5] D. Lewis, "Evaluating text categorization," in *Proceedings of Speech and Natural Language Workshop*, pp. 312–318, 1991.

[6] K. Machines http://www.kernel-machines.org/.

[7] N. Murata, S. Yoshizawa, and S. Amari, "Network information criterion - determining the number of hidden units for an artificial neural network model," *IEEE Transactions on Neural Networks*, vol. 5, no. 6, pp. 865–872, 1994.

[8] NTCIR http://www.rd.nacsis.ac.jp/ ntcadm/.

[9] M. Okabe and S. Yamada, "Interactive document retrieval with relational learning," in *Proceedings of the 16th ACM Symposium on Applied Computing*, pp. 27–31, 2001.

[10] T. Onoda, "Neural network information criterion for the optimal number of hidden units," in *Proc. ICNN'95*, pp. 275–280, 1995.

[11] J. Orr and K.-R. Müller, eds., *Neural Networks: Tricks of the Trade*. LNCS 1524, Springer Verlag, 1998.

[12] G. Salton, ed., *Relevance feedback in information retrieval*, pp. 313–323. Englewood Cliffs, N.J.: Prentice Hall, 1971.

[13] G. Salton and J. McGill, *Introduction to modern information retrieval*. McGraw-Hill, 1983.

[14] R. Schapire, Y. Singer, and A. Singhal, "Boosting and rocchio applied to text filtering," in *Proceedings of the Twenty-First Annual International ACM SIGIR*, pp. 215–223, 1998.

[15] B. Schölkopf, A. Smola, R. Williamson, and P. Bartlett, "New support vector algorithms," Technical Report NC-TR-1998-031, Department of Computer Science, Royal Holloway, University of London, Egham, UK, 1998. *Neural Computation 2000*.

[16] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," in *Journal of Machine Learning Research*, pp. 45–66, 2001.

[17] TREC Web page http://trec.nist.gov/.

[18] V. Vapnik, *The Nature of Statistical Learning Theory*. Springer, 1995.

[19] I. Witten, A. Moffat, and T. Bell, *Managing Gigabytes: Compressing and Indexing Documents and Images*. New York: Van Nostrand Reinhold, 1994.

[20] R. B. Yates and B. R. Neto, *Modern Information Retrieval*. Addison Wesley, 1999.