2006 International Joint Conference on Neural Networks
Sheraton Vancouver Wall Centre Hotel, Vancouver, BC, Canada
July 16-21, 2006

Non-Relevance Feedback Document Retrieval based on One Class SVM and SVDD

Takashi Onoda, and Hiroshi Murata, and Seiji Yamada

Abstract— **This paper reports a new document retrieval method using non-relevant documents. Especially, this paper reports a comparison of retrieval efficiency between One Class Support vector Machine(SVM) based and Support Vector Data Description(SVDD) based interactive document retrieval method using non-relevant documents only. 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 applied active learning techniques based on Support Vector Machine for evaluating successive batches, which is called *relevance feedback*. Our proposed approach has been very useful for document retrieval with relevance feedback experimentally. The traditional relevance feedback needs a set of relevant and non-relevant documents to work usefully. However, the initial retrieved documents, which are displayed to a user, sometimes don't include relevant documents. In order to solve this problem, we propose a new feedback method using information of non-relevant documents only. We named this method *non-relevance feedback document retrieval*. The non-relevance feedback document retrievals are based on One Class Support Vector Machine and Support Vector Data Description. Our experimental results show that One Class Support Vector Machine based method can retrieve relevant documents efficiently using information of non-relevant documents only.**

I. Introduction

As Internet technology progresses, accessible information by end users is explosively increasing. In this situation, we can now easily access a huge document database through the web. 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 [1]. Active works for such document retrieval have been reported in TREC (Text Retrieval Conference) [2] for English documents, IREX (Information Retrieval and Extraction Exercise) [3] and NTCIR (NII-NACSIS Test Collection for Information Retrieval System) [4] for Japanese documents.

In most frameworks for information retrieval, a vector space model in which a document is described with a high-dimensional vector is used [5]. An information retrieval system using a vector space model computes the similarity between a query vector and document vectors by the cosine of the two vectors and indicates a user a list of retrieved documents.

Takashi Onoda and Hiroshi Murata are with the Department of System Engineering Lab., CRIEPI, 2-11-1, Iwado Kita, Komae-shi, Tokyo 201-8511 JAPAN (phone: +81 3-3480-2111; fax: +81 3-5497-0318; email: {onoda, murata}@criepi.denken.or.jp).

Seiji Yamada is with Natinal Institute of Informatics, 2-1-1, Hitotsubashi, Chiyoda-ku, Tokyo 101-8430 JAPAN (phone: +81 3-4212-2562; fax: +81 3-4212-2562; email: seiji@nii.ac.jp).

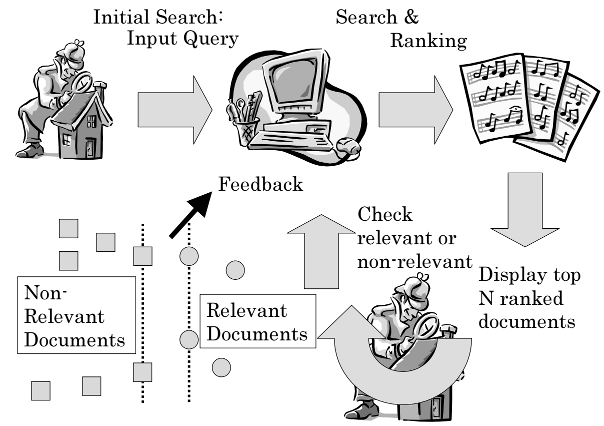

Fig. 1. Image of the relevance feedback documents retrieval: The gray arrow parts are made iteratively to retrieve useful documents for the user. This iteration is called feedback iteration in the information retrieval research area.

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* [6] and used widely in information retrieval systems. In this method, a user directly evaluates whether a document is relevant or non-relevant 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(see Figure 1).

In another approach, relevant and irrelevant document vectors are considered as positive and negative examples, and relevance feedback is transposed to a binary class classification problem [7]. For the binary class classification problem, Support Vector Machines (which are called SVMs) have shown the excellent ability. And some studies applied SVM to the text classification problems [8] and the information retrieval problems [9]. Recently, we have proposed a relevance feedback framework with SVM as *active learning* and shown the usefulness of our proposed method experimentally [10].

The initial retrieved documents, which are displayed to a user, sometimes don't include relevant documents. In this case, almost all relevance feedback document retrieval systems work hardly, because the systems need relevant and non-relevant documents to construct a binary class classifi-

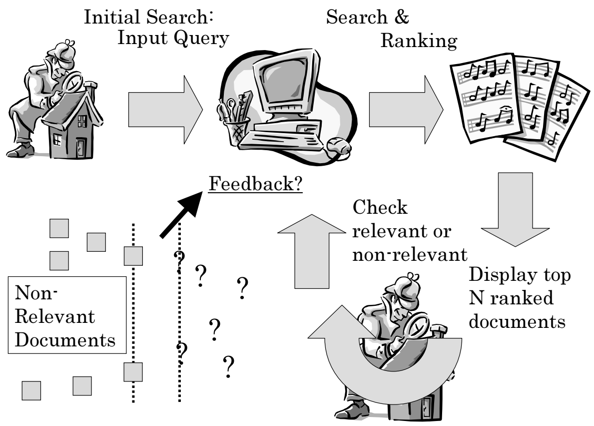

Fig. 2. Image of a problem in the relevance feedback documents retrieval: The gray arrow parts are made iteratively to retrieve useful documents for the user. This iteration is called feedback iteration in the information retrieval research area. But if the evaluation of the user has only non-relevant documents, ordinary relevance feedback methods can not feed back the information of useful retrieval.

cation problem(see Figure 2).

While a machine learning research field has some methods which can deal with one class classification problem. In the above document retrieval case, we can use only non-relevant documents information. Therefore, we consider this retrieval situation is as same as one class classification problems.

In this paper, we propose a framework of an interactive document retrieval using only non-relevant documents information. We call the interactive document retrieval *non-relevance feedback document retrieval*, because we can use only non-relevant documents information. Our proposed non-relevance document retrieval methods are based on the One Class SVM[11] and the SVDD[12]. One Class SVM can generate a discriminant hyper-plane that can separate the non-relevant documents which are evaluated by a user. Our proposed method can display documents, which may be relevant documents for the user, using the discriminant hyper-plane. SVDD can generate a discriminant hyper-sphere that can separate the non-relevant documents which are evaluated by a user. This SVDD based method can display documents, which may be relevant documents for the user, using the discriminant hyper-sphere. Finally, this paper shows the comparison of retrieval efficiency between One Class SVM based and SVDD based interactive document retrieval methods using non-relevant documents only.

In the remaining parts of this paper, we explain the One Class SVM and SVDD algorithms in the next section briefly. In the third section, we propose the framework of non-relevance feedback document retrieval. In order to compare the retrieval efficiency between One Class SVM based and SVDD based non-relevance documents retrievals, we show experimental results using a TREC data set of Los Angels Times in the fourth section. Finally we conclude our work and discuss our future work in the fifth section.

II. One Class SVM and SVDD

In this section, we introduce the One Class SVM and SVDD briefly.

A. One Class SVM

Schölkopf et al. suggested a method of adapting the SVM methodology to one class classification problem. Essentially, after transforming the feature via a kernel, they treat the origin as the only member of the second class. The using *relaxation parameters* they separate the image of the one class from the origin. Then the standard two-class SVM techniques are employed.

One Class SVM [11] returns a function f that takes the value +1 in a *small* region capturing most of the training data points, and -1 elsewhere.

The algorithm can be summarized as mapping the data into a feature space H using an appropriate kernel function, and then trying to separate the mapped vectors from the origin with maximum margin (see Figure 3).

Let the training data be

$$\mathbf{x}_1, \ldots, \mathbf{x}_\ell \tag{1}$$

belonging to one class X, where X is a compact subset of R^N and ℓ is the number of observations. Let $\Phi : X \to H$ be a kernel map which transforms the training examples to feature space. The dot product in the image of Φ can be computed by evaluating some simple kernel

$$k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) \tag{2}$$

such as the linear kernel, which is used in our experiment,

$$k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{y}. \tag{3}$$

The strategy is to map the data into the feature space corresponding to the kernel, and to separate them from the origin with maximum margin. Then, to separate the data set from the origin, one needs to solve the following quadratic program:

$$\begin{aligned} \min_{\mathbf{w}\in H, \xi\in R^\ell \rho\in R^N} \quad & \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho \\ \text{subject to} \quad & (\mathbf{w}\cdot\Phi(\mathbf{x}_i)) \geq \rho - \xi_i, \\ & \xi_i \geq 0. \end{aligned} \tag{4}$$

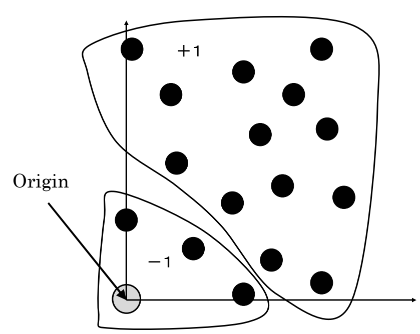

Fig. 3. One Class SVM Classifier: the origin is the only original member of the second class.

Here, $\nu \in (0,1)$ is an upper bound on the fraction of outliers, and a lower bound on the fraction of Support Vectors (SVs).

Since nonzero slack variables ξ_i are penalized in the objective function, we can expect that if $\mathbf{w}$ and ρ solve this problem, then the decision function

$$f(\mathbf{x}) = \operatorname{sgn}\left((\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho\right) \tag{5}$$

will be positive for most examples $\mathbf{x}_i$ contained in the training set, while the SV type regularization term $\|w\|$ will still be small. The actual trade-off between these two is controlled by ν. For a new point $\mathbf{x}$, the value $f(\mathbf{x})$ is determined by evaluating which side of the hyperplane it falls on, in feature space.

Using multipliers $\alpha_i, \beta_i \geq 0$, we introduce a Lagrangian

$$\begin{aligned} &L(\mathbf{w}, \boldsymbol{\xi}, \rho, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \\ &\qquad \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu\ell}\sum_i \xi_i - \rho \\ &\qquad - \sum_i \alpha_i((\mathbf{w} \cdot \mathbf{x}_i) - \rho + \xi_i) - \sum_i \beta_i \xi_i \end{aligned} \tag{6}$$

and set the derivatives with respect to the primal variables $\mathbf{w}, \xi_i, \rho$ equal to zero, yielding

$$\mathbf{w} = \sum_i \alpha_i \mathbf{x}_i, \tag{7}$$

$$\alpha_i = \frac{1}{\nu\ell} - \beta_i \leq \frac{1}{\nu\ell}, \quad \sum_i \alpha_i = 1. \tag{8}$$

In Eqn. (7), all patterns $\{\mathbf{x}_i : i \in [\ell], \alpha_i > 0\}$ are called Support Vectors. Using Eqn. (2), the SV expansion transforms the decision function Eqn. (5)

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_i \alpha_i k(\mathbf{x}_i, \mathbf{x}) - \rho\right). \tag{9}$$

Substituting Eqn. (7) and Eqn. (8) into Eqn. (6), we obtain the dual problem:

$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j k(\mathbf{x}_i, \mathbf{x}_j) \tag{10}$$

$$\text{subject to} \quad 0 \leq \alpha_i \leq \frac{1}{\nu\ell}, \quad \sum_i \alpha_i = 1. \tag{11}$$

One can show that at the optimum, the two inequality constraints Eqn. (4) become equalities if α_i and β_i are nonzero, i.e. if $0 < \alpha \leq 1/(\nu\ell)$. Therefore, we can recover ρ by exploiting that for any such α_i, the corresponding pattern $\mathbf{x}_i$ satisfies

$$\rho = (\mathbf{w} \cdot \mathbf{x}_i) = \sum_j \alpha_j \mathbf{x}_j \cdot \mathbf{x}_i. \tag{12}$$

Note that if ν approaches to 0, the upper boundaries on the Lagrange multipliers tend to infinity, i.e. the second inequality constraint in Eqn. (11) becomes void. The problem then resembles the corresponding *hard margin* algorithm, since the penalization of errors becomes infinite, as can be seen from the primal objective function Eqn. (4). It is still a feasible problem, since we have placed no restriction on ρ, so ρ can become a large negative number in order to satisfy Eqn. (4). If we had required $\rho \geq 0$ from the start, we would have ended up with the constraint $\sum_i \alpha_i \geq 1$ instead of the corresponding equality constraint in Eqn. (11), and the multipliers α_i could have diverged.

In our research we used the LIBSVM. This is an integrated tool for support vector classification and regression which can handle one-class SVM using the Schölkopf etc algorithms. The LIBSVM is available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

B. SVDD

It is required to find a description of a data set containing N data objects, $\{\mathbf{x}_i, i = 1, \ldots, N\}$. We try to find a sphere with minimum volume, containing all (or most of) the data objects. We define the following error function, which is described by a center $\mathbf{a}$ and radius R.

$$F(R, \mathbf{a}) = R^2, \tag{13}$$

with the constraints

$$\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2, \quad i = 1, \ldots N. \tag{14}$$

This is very sensitive to the most outlying object in the target data set. When one or few very remote objects are in the training set, a very large sphere is obtained which will not represent the data very well. Therefore, we allow for some data points outside the sphere and introduce slack variables ξ_i.

We minimize the following function of the sphere, which is described by a center $\mathbf{a}$ and radius R.

$$F(R, \mathbf{a}, \xi) = R^2 + C\sum_{i=1}^{N} \xi_i, \tag{15}$$

where the variable C gives the trade-off between simplicity (or volume of the sphere) and the number of errors (number of target objects rejected)(see Figure 4).

This has to be minimized under the constraints

$$(\mathbf{x}_i - \mathbf{a})^\top (\mathbf{x}_i - \mathbf{a}) \leq R^2 + \xi_i \quad i = 1, \ldots N, \xi_i \geq 0. \tag{16}$$

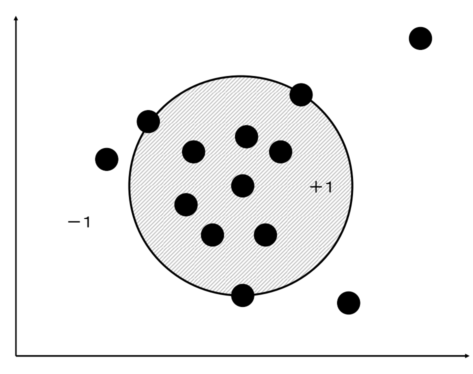

Fig. 4. The concept of SVDD Classifier.

Incorporating these constraints in Eq.(15), we construct the following Lagrangian,

$$L(R, \mathbf{a}, \boldsymbol{\alpha}, \boldsymbol{\xi}) = R^2 + C\sum_{i=1}^{N}\xi_i$$
$$-\sum_{i=1}^{N}\alpha_i\{R^2+\xi_i-(\mathbf{x}_i^2-2\mathbf{a}\cdot\mathbf{x}_i+\mathbf{a}^2)\}-\sum_{i=1}^{N}\gamma_i\xi_i \quad (17)$$

with Lagrange multipliers $\alpha_i \geq 0$ and $\gamma_i \geq 0$. The Lagrangian L should be minimized with respect to $R, \mathbf{a}, \xi_i$ and maximized with respect to α_i and γ_i.

Setting the partial derivatives to 0 gives new constraints.

$$\frac{\partial L}{\partial R} = 0 \rightarrow \sum_{i=1}^{N}\alpha_i = 1 \quad (18)$$

$$\frac{\partial L}{\partial \mathbf{a}} = 0 \rightarrow \mathbf{a} = \frac{\sum_{i=1}^{N}\alpha_i}{\sum_{i=1}^{N}\alpha_i\mathbf{x}_i} = \sum_{i=1}^{N}\alpha_i\mathbf{x}_i, \quad (19)$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow C-\alpha_i-\gamma_i = 0, i = 1,\ldots,N. \quad (20)$$

Since $\alpha_i \geq 0$ and $\gamma_i \geq 0$, we can remove the variables γ_i from the equation (20) and use the constraint

$$0 \leq \alpha_i \leq C, i = 1,\ldots,N. \quad (21)$$

Rewriting Eq.(17) and resubstituting Eqs.(18), (19), (20) give to maximize with respect to α_j

$$L = \sum_{i=1}^{N}\alpha_i(\mathbf{x}_i\cdot\mathbf{x}_j) - \sum_{i,j=1}^{N}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j), \quad (22)$$

with constraints $0 \leq \alpha_i \leq C$, $\sum_{i=1}^{N}\alpha_i = 1$.

When an objects $\mathbf{x}_i$ satisfies the unequality $\|\mathbf{x}_i-\mathbf{a}\|^2 < R^2+\xi_i$, the constraint is satisfied and the corresponding Lagrange multiplier will be zero ($\alpha_i = 0$). For objects, which satisfy the equality $\|\mathbf{x}_i-\mathbf{a}\|^2 = R^2+\xi_i$, the constraint has to be enforced and the Lagrange multiplier will not become zero ($\alpha_i > 0$). Therefore, we can get the following conditions.

$$\|x_i-\mathbf{a}\|^2 < R^2 \rightarrow \alpha_i = 0, \gamma_i = 0, \quad (23)$$
$$\|x_i-\mathbf{a}\|^2 = R^2 \rightarrow 0 < \alpha_i < C, \gamma_i = 0, \quad (24)$$
$$\|x_i-\mathbf{a}\|^2 > R^2 \rightarrow \alpha_i = C, \gamma_i > 0. \quad (25)$$

The objects with $\alpha > 0$ are needed to generate the sphere. Therefore, these objects are called Support Vectors of the description for generating the sphere.

When the following inequality is satisfied, an object $\mathbf{z}$ is accepted.

$$\begin{aligned}
& (\mathbf{z}-\mathbf{a})^\top(\mathbf{z}-\mathbf{a}) \\
= \ & (\mathbf{z}-\sum_i\alpha_i\mathbf{x}_i)^\top(\mathbf{z}-\sum_i\alpha_i\mathbf{x}_i) \\
= \ & (\mathbf{z}\cdot\mathbf{z}) - 2\sum_i(\mathbf{z}\cdot\mathbf{x}_i) + \sum_{i,j}\alpha_i\alpha_j(\mathbf{x}_i\cdot\mathbf{x}_j) \\
\geq \ & R^2 \quad (26)
\end{aligned}$$

Also we can replace all inner products $(\mathbf{x}_i\cdot\mathbf{x}_j)$ by a proper $k(\mathbf{x}_i,\mathbf{x}_j)$ which satisfies Eq.(2) and the problem of finding a data description is now given by

$$L = \sum_{i=1}^{N}\alpha_i k(\mathbf{x}_i,\mathbf{x}_j) - \sum_{i,j=1}^{N}\alpha_i\alpha_j k(\mathbf{x}_i,\mathbf{x}_j), \quad (27)$$

The more detail of SVDD can be seen in the reference [12].

III. Non-Relevance Feedback Document Retrieval

In this section, we describe our proposed method of document retrieval based on Non-relevant documents only.

In relevance feedback document retrieval, the user has the option of labeling some of the top N ranked documents according to whether they are relevant or non-relevant. The labeled documents along with the original request are then given to a supervised learning procedure to produce a new classifier. The new classifier is used to produce a new ranking, which retrieves more relevant documents at higher ranks than the original did(see Figure 1)[10].

The initial retrieved documents, which are displayed to a user, sometimes don't include relevant documents. In this case, almost all relevance feedback document retrieval systems does not contribute to make an efficient document retrieval, because the systems need relevant and non-relevant documents to construct a binary class classification problem Figure 2).

The One Class SVM and SVDD can generate discriminant hyper-plane or hyper-sphere for the one class classification using one class training data. Consequently, we propose to apply One Class SVM and SVDD in a *non-relevance feedback document retrieval methods*. The retrieval steps of proposed method perform as follows (see Figure 5):

Step 1: **Preparation of documents for the first feedback**
The conventional information retrieval system

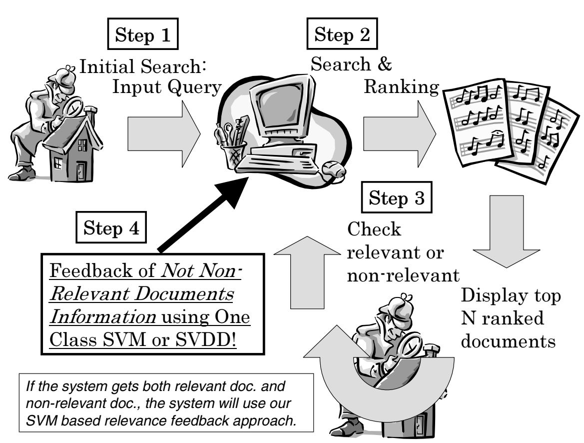

Fig. 5. Outline of the procedure in the non relevance feedback documents retrieval: The gray and black arrow parts from Step 2 to Step 4 are made iteratively to retrieve relevant documents for the user. If the system gets both relevant and non-relevant documents, the system will use our SVM based relevance feedback approach, which is proposed by us previously.

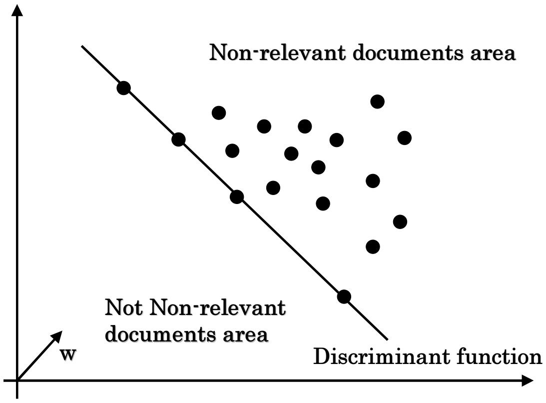

Fig. 6. Generation of a hyper-plane to discriminate *non-relevant documents area* by One Class SVM: Circles denote non-relevant documents which are evaluated by a user. The solid line denotes the discriminant hyper-plane to distinguish between *non-relevant documents area and not non-relevant documents area*.

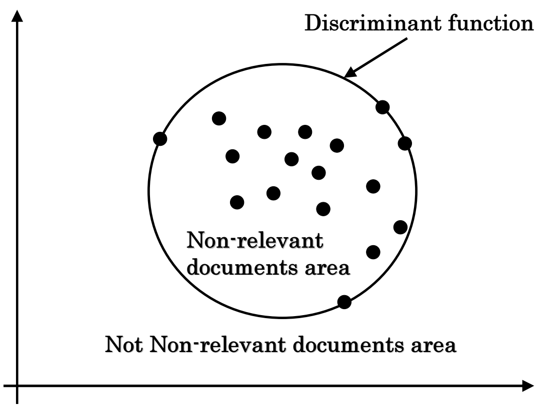

Fig. 7. Generation of a hyper-sphere to discriminate non-relevant documents area by SVDD: Circles denote non-relevant documents which are evaluated by a user. The solid circle denotes the discriminant hyper-sphere to distinguish between *non-relevant documents area and not non-relevant documents area*.

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 vectors for the first feedback iteration (see Figure 5).

Step 2: **Judgment of displayed documents**
The user then classifiers these N documents into relevant or non-relevant. If the user labels all N documents non-relevant, the documents are labeled "+1" and go to the next step. Otherwise, our previous proposed relevant feedback method is adopted[10].

Step 3: **Determination of non-relevant documents area**
The discriminant hyper-plane or hyper-sphere for classifying non-relevant documents area is generated by using One Class SVM or SVDD. In order to generate the hyper-plane, the One Class SVM learns labeled non-relevant documents which are evaluated in the previous step (see Figure 6). Also in order to generate the hyper-sphere, the SVDD learns labeled non-relevant documents (see Figure 7).

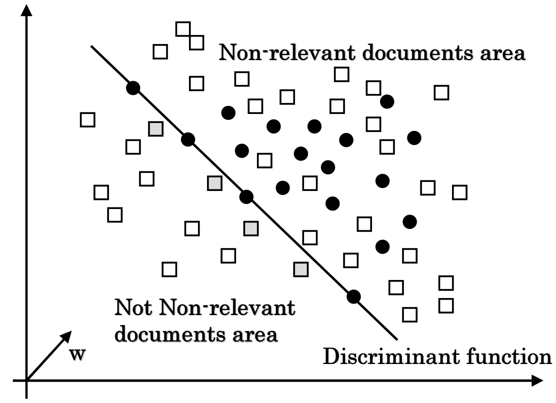

Fig. 8. Mapped non-checked documents into the feature space: Boxes denote non-checked documents which are mapped into the feature space. Circles denotes checked documents which are mapped into the feature space. Gray boxes denotes the displayed documents to a user in the next iteration. These documents are in the *not non-relevant document area* and near the discriminant hyper-plane, which is generated by One Class SVM.

Step 4: **Classification of all documents and Selection of displayed documents**
The One-class SVM or SVDD learned by previous step can classifies the whole documents as non-relevant or not non-relevant. The documents which are discriminated in *not non-relevant area* are newly selected. From the selected documents, the top N ranked documents, which are ranked in the increasing order of the distance from the non-relevant documents area, are shown to user as the document retrieval results of the system (see Figure 8 and Figure 9). These N documents have high existence probability of initial keywords. Then return to Step 2.

The feature of our One Class SVM based or SVDD based non-relevant feedback document retrieval is the selection of displayed documents to a user in Step 4. Our proposed method selects the documents which are discriminated *not non-relevant* and near the discriminant hyperplane between *non-relevant documents* and *not non-relevant documents*. Generally if the system got the non-relevant information from a user, the system should select the information, which is far from the non-relevant information area, for displaying to the user. However, in our case, the classified non-relevant documents by the user includes a request query vector of the user. Therefore, if we select the documents, which are far from the non-relevant documents area, the documents may not include the request query of the user. Our selected documents(see Figure 8 and 9) must have the high probability of the relevant documents for the user is high, because the documents are not non-relevant and may include the query

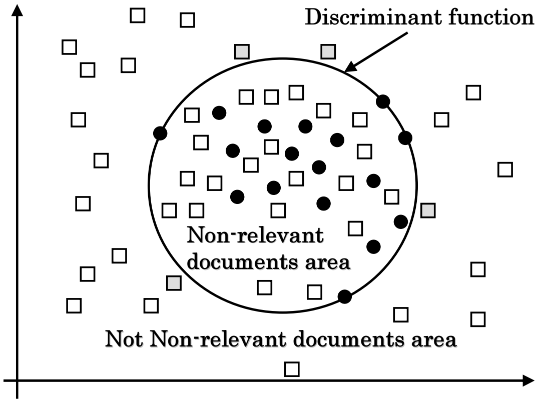

Fig. 9. Mapped non-checked documents into the feature space: Boxes denote non-checked documents which are mapped into the feature space. Circles denotes checked documents which are mapped into the feature space. Gray boxes denotes the displayed documents to a user in the next iteration. These documents are in the *not non-relevant document area* and near the discriminant hyper-sphere, which is generated by SVDD.

vector of the user.

IV. Experiments

In this section, we explain our experimental setting and report our experimental results.

A. Experimental setting

We made experiments for evaluating the utility of our interactive document retrieval based on non-relevant documents using One Class SVM or SVDD described in section II. The document data set we used is a set of articles in the Los Angels Times which is widely used in the document retrieval conference TREC [2]. 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. Our experiment used three topics, which are in the Los Angels Times and table I shows these topics. These topics do not have relevant documents in top 20 ranked documents in the order of cosine distance between the query vector and document vectors. Our experiments set the size of N of displayed documents 10 and 20.

We used TFIDF [1], which is one of the most popular methods in information retrieval to generate document feature vectors, and the concrete equation [13] of a weight of a term t in a document d w_t^d are in the following.

$$w_t^d = L \times t \times u \tag{28}$$

$$L = \frac{1+\log(tf(t,d))}{1+\log(\text{average of } tf(t,d) \text{ in } d)} \quad \text{(TF)}$$

$$t = \log(\frac{n+1}{df(t)}) \quad \text{(IDF)}$$

$$u = \frac{1}{0.8+0.2\frac{uniq(d)}{\text{average of } uniq(d)}} \quad \text{(normalization)}$$

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.

In our experiments, we used the linear kernel for One Class SVM and SVDD learning, and found a discriminant function for the One Class SVM classifier and the SVDD classifier in the feature space. The vector space model of documents is high dimensional space. Moreover, the number of the documents which are evaluated by a user is small. Therefore, we do not need to use the kernel trick, and the parameter ν (see section II) is set adequately small value ($\nu = 0.01$). The small ν means hard margin in the One Class SVM and the SVDD, and it is important to make hard margin in our problem.

For comparison with our approaches, two information retrieval methods were used. The first is an information retrieval method that does not use a feedback, namely documents are retrieved using the rank based on the cosine distance between a query vector and document vectors in vector space model(VSM). The other is an information retrieval method using the conventional Rocchio-based relevance feedback approach [6] 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.

$$Q_{i+1} = Q_i + \alpha \sum_{x \in R_r} x - \beta \sum_{x \in R_n} x, \tag{29}$$

where R_r is a set of documents which were evaluated as relevant documents by a user at the ith feedback, and R_n is a set of documents which were evaluated as non-relevant documents at the i feedback. And x denotes terms included in a document. α and β are weights for relevant and non-relevant documents respectively. In this experiment, we set $\alpha = 1.0$, $\beta = 0.5$ which are known adequate experimentally.

TABLE I
Topics, query words and the number of relevant documents in the Los Angels Times used for experiments

topic	query words	# of relevant doc.
306	Africa, civilian, death	34
343	police, death	88
383	mental, ill, drug	55

B. Experimental results

Here, we describe the relationships between the performances of proposed methods and the number of feedback iterations. Table II gave the number of retrieved relevant documents at each feedback iteration. At each feedback iteration, the system displays ten higher ranked *not non-relevant documents*, which are near the discriminant hyper-plane of One Class SVM or the discriminant hyper-sphere of SVDD, for our proposed methods. We also show the retrieved documents of Rocchio-based method at each feedback iteration for comparing to proposed methods in table II.

Table III gave the number of retrieved relevant documents at each feedback iteration. At each feedback iteration, the system displays twenty higher ranked *not non-relevant documents*, which are near the discriminant hyper-plane of One Class SVM or the discriminant hyper-sphere of SVDD, for our proposed methods. We also show the retrieved documents of Rocchio-based method at each feedback iteration for comparing to proposed methods in table III.

We can see from the table II that our non-relevance feedback approach based on One Class SVM gives the higher performance in terms of the number of iteration for retrieving relevant documents. On the other hand, our non-relevance feedback approach based on SVDD and the Rocchio-based feedback method and SVDD based method can not search a relevant document in all cases. The vector space model without feedback is better than SVDD based and Rocchio-based feedback methods. After all, we can believe that the proposed method based on One Class SVM can make an effective document retrieval using only non-relevant documents, and Rocchio-based feedback method can not work well when the system can receive the only non-relevant documents information. And the proposed method based on SVDD can not work well, too.

TABLE II

THE NUMBER OF RETRIEVED RELEVANT DOCUMENTS AT EACH ITERATION: THE NUMBER OF DISPLAYED DOCUMENTS IS 10 AT EACH ITERATION

topic 308	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	1	0	0	0
2	–	0	0	0
3	–	0	1	0
4	–	0	–	0
5	–	0	–	0
topic 343	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	0	0	0	0
2	1	0	0	0
3	–	0	0	0
4	–	0	0	0
5	–	0	0	0
topic 383	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	0	0	0	0
2	1	0	0	0
3	–	0	0	0
4	–	0	1	0
5	–	0	–	0

TABLE III

THE NUMBER OF RETRIEVED RELEVANT DOCUMENTS AT EACH ITERATION: THE NUMBER OF DISPLAYED DOCUMENTS IS 20 AT EACH ITERATION

topic 308	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	1	0	1	0
2	–	0	–	0
3	–	0	–	0
4	–	0	–	0
5	–	0	–	0
topic 343	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	1	0	0	0
2	–	0	0	0
3	–	0	0	0
4	–	0	1	0
5	–	0	–	0
topic 383	# of retrieved relevant doc.			
# of itrs.	SVM	SVDD	VSM	Rocchio
1	1	0	0	0
2	–	0	1	0
3	–	0	–	0
4	–	0	–	0
5	–	0	–	0

We can observe from the table III that our non-relevance feedback approach based on One Class SVM gives the higher performance than the Rocchio-based method, SVDD based method and VSM without feedback in terms of the number of iteration for retrieving relevant documents. This experimental results are as same as the results observed from the table II.

In table II, a user already have seen twenty documents at the first iteration. Before the fist iteration, the user have to see ten documents, which are retrieved results using the cosine distance between a query vector and document vectors in VSM. In table III, the user also have seen forty documents at the first iteration. Before the fist iteration, the user also have to see ten documents to evaluate the documents, which are retrieved results using the cosine distance between a query vector and document vectors in VSM. When we compare the experimental results of table II with the results of table III, we can observe that the small number of displayed documents makes more effective document retrieval performance than the large number of displayed documents. In table II, the user had to see thirty documents to find the first relevant document about topic 343 and 383 using One Class SVM based method. In table III, the user had to see forty documents to find the first relevant document about topic 343 and 383 using One Class SVM method. Therefore, we believe that the early non-relevance feedback is useful to find the first relevant document for an interactive document retrieval.

We also show a precision and recall curve in figure 10. This figure is the precision and recall curve of topic no. 306 at the second iteration. From this figure, we can understand that all precision-recall curves are not good. However, our proposed approach based on One Class SVM is more efficient than the other three approaches for the interactive documents retrieval. In the figure 10, the shape at the left side

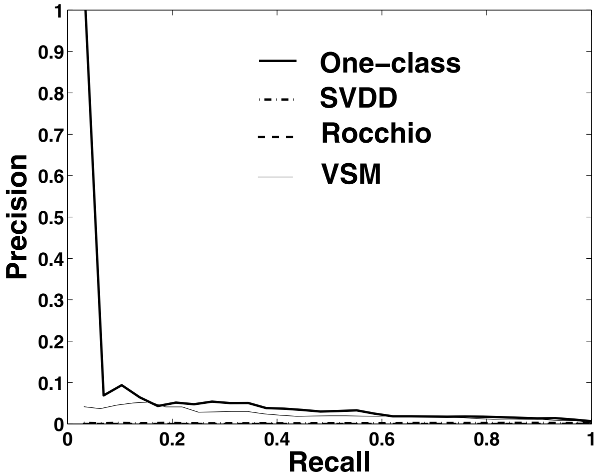

Fig. 10. The precision and recall curve of topic no. 304 at the second iteration.

of the figure is very important. Because if the system can at least display a relevant document to the user, the system will be able to switch from a non-relevance feedback approach to our previous proposed relevant feedback approach based on SVM[10]. Therefore, when the system takes non-relevant documents only, it is important for the user to find relevant documents as soon as possible by the document retrieval system.

V. Conclusion

In this paper, we proposed the non-relevance feedback document retrieval based on One Class SVM or SVDD using only non-relevant documents for a user. In our non-relevance feedback document retrieval, the system use only non-relevant documents information. One-Class SVM can generate a discriminant hyperplane of observed one class information and SVDD can find a discriminant hyper-sphere of observed one class information, so our proposed method adopted One Class SVM and SVDD for non-relevance feedback document retrieval.

The main point of this paper is the comparison of retrieval efficiency between One-Class SVM based and SVDD based non-relevant feedback methods. In our experiments, One Class SVM based non-relevance feedback approach makes better performance than SVDD based approach for an interactive document retrieval. The main reason behind this result is that SVDD can not obtain a very tight sphere description of non-relevant documents. In our experiments, SVDD based approach computes a sphere around a few nonrelevant documents in the input space, which is very high dimension. Generally, these documents are not spherically distributed, even when the most outlying objects are ignored. Therefore, we can not expect to obtain a very tight sphere description of non-relevant documents in the input space. If we use kernel method to SVDD, SVDD may be able to generates a very tight sphere description of non-relevant documents (see figure 11). However, it is difficult for SVDD

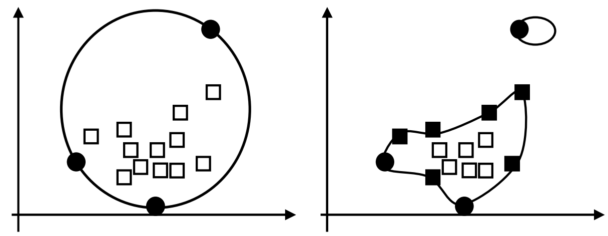

Fig. 11. An image of data description using SVDD: the left side figure shows an image of normal spherical description. the right side figure shows an image data description using a kernel function.

to select appropriate parameters of a kernel function. So, we did not use kernel method for SVDD in our experiments.

After all, our experimental results on a set of articles in the Los Angels Times showed that the proposed method based on One Class SVM gave a consistently better performance than the compared three methods. Therefore, we believe that our proposed One Class SVM based approach is very useful for the document retrieval with only non-relevant documents information.

In our future work, we plan to apply the proposed method to other many topics and make a test including normal human users. In this paper, we do not use kernel method. It is also our future work how to use kernel method effectively for *non-relevance feedback document retrieval*. And we will research more effective non-relevance feedback document retrieval methods, which can use a few non-relevance documents only.

References

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

[2] TREC Web page, "http://trec.nist.gov/."

[3] IREX, "http://cs.nyu.edu/cs/projects/proteus/irex/."

[4] NTCIR, "http://www.rd.nacsis.ac.jp/˜ntcadm/."

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

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

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

[8] S. Tong and D. Koller, "Support vector machine active learning with applications to text classification," in *Journal of Machine Learning Research*, vol. 2, 2001, pp. 45–66.

[9] H. Drucker, B. Shahrary, and D. C. Gibbon, "Relevance feedback using support vector machines," in *Proceedings of the Eighteenth International Conference on Machine Learning*, 2001, pp. 122–129. [Online]. Available: citeseer.nj.nec.com/443262.html

[10] T. Onoda, H. Murata, and S. Yamada, "Relevance feedback with active learning for document retrieval," in *Proc. of IJCNN2003*, 2003, pp. 1757–1762.

[11] B. Schölkopf, J. Platt, J. Shawe-Taylor, A. Smola, and R. Williamson, "Estimating the support for a high-dimensional distribution," Microsoft Research, One Microsoft Way Redmon WA 98052, Tech. Rep. MSR-TR-99-87, 1999.

[12] D. Tax and R. Duin, "Support vector data description," *Machine Learning*, vol. 54, pp. 45–66, 2004.

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