One Class Classification Methods based Non-Relevance Feedback Document Retrieval

Takashi Onoda, Hiroshi Murata
Central Research Institute of
Electric Power Industry
2-11-1, Iwado Kita, Komae-shi,
Tokyo 201-8511 JAPAN
{onoda, murata}@criepi.denken.or.jp

Seiji Yamada
National Institute of Informatics
2-1-1, Hitotsubashi, Chiyoda-ku,
Tokyo 101-8430 JAPAN
seiji@nii.ac.jp

Abstract

We applied active learning techniques based on Support Vector Machine for evaluating documents each iteration, which is called relevance feedback. *Our proposed approach has been very useful for document retrieval with relevance feedback experimentally. 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.*

1. 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 [10]. Active works for such document retrieval have been reported in TREC (Text Retrieval Conference) [9] for English documents, IREX (Information Retrieval and Extraction Exercise) [1] and NTCIR (NII-NACSIS Test Collection for Information Retrieval System) [2] 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.

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* [4] 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. Recently, we have proposed a relevance feedback framework with SVM as *active learning* and shown the usefulness of our proposed method experimentally [3].

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 classification problem(see Figure 1).

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[7] and the SVDD[8].

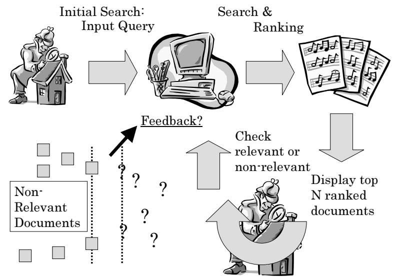

Figure 1. Image of a problem in the relevance feedback documents retrieval.

2. One Class SVM and SVDD

2.1. 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 [7] 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.

Let the training data be $\mathbf{x}_1, \ldots, \mathbf{x}_\ell$ belonging to one class X, where X is a compact subset of R^N and ℓ is the number of observations. Let $\Phi : X \rightarrow H$ 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 $k(\mathbf{x}, \mathbf{y}) = (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y}))$ such as the linear kernel $k(\mathbf{x}, \mathbf{y}) = \mathbf{x}^\top \mathbf{y}$, which is used in our experiment.

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:

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

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

$$f(\mathbf{x}) = \operatorname{sgn}\left((\mathbf{w} \cdot \Phi(\mathbf{x})) - \rho\right) \tag{2}$$

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.

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`.

2.2. 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 error function $F(R, \mathbf{a}) = R^2$, which is described by a center **a** and radius R, with the constraints $\|\mathbf{x}_i - \mathbf{a}\|^2 \leq R^2, \quad i = 1, \ldots N$. 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 **a** and radius R.

$$F(R, \mathbf{a}, \xi) = R^2 + C\sum_{i=1}^{N} \xi_i, \tag{3}$$

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). This has to be minimized under the constraints

$$(\mathbf{x}_i - \mathbf{a})^\top(\mathbf{x}_i - \mathbf{a}) \leq R^2 + \xi_i \quad i = 1, \ldots N, \xi_i \geq 0. \tag{4}$$

When an objects $\mathbf{x}_i$ satisfies the inequality $\|\mathbf{x}_i - \mathbf{a}\|^2 < R^2 + \xi_i$, 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 ($\alpha_i > 0$).

The more detail of SVDD can be seen in the reference [8].

3. Non-Relevance Feedback Document Retrieval

In this section, we describe our proposed method of document retrieval based on Non-relevant documents only. 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(see Figure 1).

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:

Step 1: Preparation of documents for the first feedback
In the initial retrieval, the top N ranked documents are selected by using cosine distance between the request query vector and each document vectors.

Step 2: Judgment of displayed documents
The user classifies 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[3].

Step 3: Determination of non-relevant documents area
The discriminant function for classifying non-relevant documents area is generated by using One Class SVM or SVDD.

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 2). These N documents have high existence probability of initial keywords. Then return to Step 2.

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 2) must have the high probability of the relevant documents for the user, because the documents are not non-relevant and may include the query vector of the user.

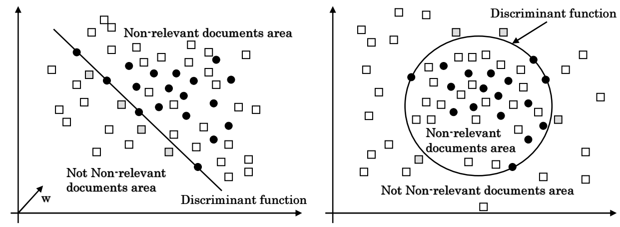

Figure 2. The selected documents by One Class SVM(left side) or SVDD(right side).

4. Experiments

4.1. 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 2. 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 [9]. 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 1 shows these topics. These topics do not have relevant documents in top 10 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.

We used TFIDF [10], which is one of the most popular methods in information retrieval to generate document feature vectors, and the concrete equation of a weight of a term t in a document d w_t^d in the reference [6].

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.

Table 1. Topics used for experiments

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

Table 2. The number of retrieved relevant documents 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

Therefore, we do not need to use the kernel trick, and the parameter ν (see section 2) 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 [4] which is widely used in information retrieval research.

4.2. Experimental results

Here, we describe the relationships between the performances of proposed methods and the number of feedback iterations. Table 2 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 2.

We can see from the table 2 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.

5 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 nonrelevance 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 hypersphere of observed one class information, so our proposed method adopted One Class SVM and SVDD for non-relevance feedback document retrieval.

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.

References

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

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

[3] T. Onoda, H. Murata, and S. Yamada. Relevance feedback with active learning for document retrieval. In *Proc. of IJCNN2003*, pages 1757–1762, 2003.

[4] G. Salton, editor. *Relevance feedback in information retrieval*, pages 313–323. Englewood Cliffs, N.J.: Prentice Hall, 1971.

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

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

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

[8] D. Tax and R. Duin. Support vector data description. *Machine Learning*, 54:45–66, 2004.

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

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