J Syst Sci Syst Eng (Sep 2007) 16(3): 356-369 ISSN: 1004-3756 (Paper) 1861-9576 (Online)
DOI: 10.1007/s11518-007-5050-x CN11-2983/N

TEXT CLASSIFICATION TOWARD A SCIENTIFIC FORUM*

Wen ZHANG[1] **Xijin TANG**[2] **Taketoshi YOSHIDA**[1]

[1] *School of Knowledge Science, Japan Advanced Institute of Science and Technology,*
1-1 Ashahidai, Tatsunokuchi, Ishikawa 923-1292, Japan
{*zhangwen, yoshida*}*@jaist.ac.jp* (✉)

[2]*Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences,*
Beijing 100080, P.R.China
xjtang@amss.ac.cn

Abstract

Text mining, also known as discovering knowledge from the text, which has emerged as a possible solution for the current information explosion, refers to the process of extracting non-trivial and useful patterns from unstructured text. Among the general tasks of text mining such as text clustering, summarization, etc, text classification is a subtask of intelligent information processing, which employs unsupervised learning to construct a classifier from training text by which to predict the class of unlabeled text. Because of its simplicity and objectivity in performance evaluation, text classification was usually used as a standard tool to determine the advantage or weakness of a text processing method, such as text representation, text feature selection, etc. In this paper, text classification is carried out to classify the Web documents collected from XSSC Website (http://www.xssc.ac.cn). The performance of support vector machine (SVM) and back propagation neural network (BPNN) is compared on this task. Specifically, binary text classification and multi-class text classification were conducted on the XSSC documents. Moreover, the classification results of both methods are combined to improve the accuracy of classification. An experiment is conducted to show that BPNN can compete with SVM in binary text classification; but for multi-class text classification, SVM performs much better. Furthermore, the classification is improved in both binary and multi-class with the combined method.

Keywords: Text classification, SVM, BPNN, Xiangshan Science Conference

1. Introduction

As a famous academic activity to promote national basic research in China, XSSC (Xiangshan Science Conference) is made up of a small-scale academic workshop series. During the workshop, experts and scholars in different

* This work is supported by Ministry of Education, Culture, Sports, Science and Technology of Japan under the "Kanazawa Region, Ishikawa High-Tech Sensing Cluster of Knowledge-Based Cluster Creation Project" and the National Natural Science Foundation of China under Grant No.70571078 and 70221001.

disciplines were invited to present their opinions concerning the current situation and future direction for Chinese science and technology in order to foster interdisciplinary cooperation and integrated studies in various areas of excellence and explore new frontiers. After the conference, all the documents which recorded the conference contents were posted on the XSSC Website. Since its foundation in 1992, more than 200 academic symposia with a total participation of about 8,000 scholars were held during the decade from 1993 to 2003. This makes the XSSC Website a gigantic textual data warehouse with respect to Chinese science that can be reused by the organizers of XSSC to manage the conference and for the subsequent participants to prepare their presentations for the conference. In order to make use of the rich knowledge from scientists contributed under a free discussion and debate atmosphere, in-depth studies have been undertaken and a series of intelligent information processing tools are developed to help users to discover, make sense of and share the knowledge embodied in this website and to learn the current trends of basic research in China, aiming at facilitating knowledge creation in XSSC ((Liu, Tang and Li, 2005; Tang, Liu and Zhang, 2005; Liu and Tang, 2006; Zhang and Tang, 2006).

The rest of this paper is organized as follows. Section 2 briefly addresses related work on text classification, SVM and BPNN. Section 3 describes the XSSC Web text representation and textual data normalization for the performance examination. The cosine formula for the measure of similarity between two documents was employed to conduct the text representation. A standard XSSC data set is produced after hierarchical clustering analysis (HCA) and heuristic selection were performed on the raw XSSC data set. Then, in Section 4, the design of the experiment is specified. The training and test data for binary classification and multi-class classification are selected from the standard XSSC data set. SVM kernel function and BPNN net are designed to conduct machine learning as well. Also, the combined method is introduced in this section. Based on the experiments, in Section 5, a comparison between SVM and BPNN was made on both binary classification and multi-class classification. Furthermore, the combined method was utilized to integrate the results of SVM and BPNN, and its performance is presented. Concluding remarks and further research are indicated in Section 6.

2. Related Work and Brief Introduction to SVM and BPNN

This section firstly describes the mechanism of text classification and its applications in state-of-the-art briefly. Then two machines learning methods, SVM and BPNN are introduced for better understanding of the designed experiments.

2.1 Related Work on Text Classification

Text classification, namely text categorization, is defined as assigning predefined categories to text documents, where documents can be news stories, technical reports, web pages, etc., and categories are most often subjects or topics, but may also be based on style (genres), pertinence, etc. Whatever the specific method employed, a text classification task starts with a training set $D = (d_1,\ldots,d_n)$ of documents that are already labeled with a category $L \in C$ (e.g. sports, politics). The task is then to determine a classification model as equation (1) which is able to assign the correct class to a new document d of the domain.

$$f : D \rightarrow C \quad f(d) = L \tag{1}$$

To measure the performance of a classification model, a random fraction of the labeled documents is set aside and not used for training. We may classify the documents of this test set with the classification model and compare the estimated labels with true labels. The fraction of correctly classified documents in relation to the total number of documents is called accuracy, and is a basic performance measure.

Recently, various kinds of research on text classification have been conducted regarding its applications. For instance, Adeva and Atxa (2007) applied Naive bayes (NB), k-nearest neighbour (KNN) and Racchio classifiers to learn the characteristics of both normal and malicious user behaviors from the log entries generated by the web application server and the performance of each classifier was compared. They reported that NB outperformed the other two by more than one percent on both the macro- and micro-average F-measure. Zhang and Jiao (2007) developed an associative classification-based recommendation system for customer profile personalization in B2C e-commerce to predict customer requirements according to the sales records stored in database by evolving the traditional association rule. The regular linear least-squares fit (LSSF) algorithm was used in Hissa et al. (2007) for the automatic classification of texts whose contents concern the nursing care narratives of some diseases. Their results indicated that the free text in nursing documentation can be automatically classified, and this can offer a way to develop electronic patient records. Yang and Liu (1999) used many kinds of statistical learning methods such as SVM, neural network (NNet), etc., on the Reuters-21578 text classification, and reported that SVM, KNN and LLSF outperform NNet and NB when the number of positive training instances per category are small (less than 10), and that all the methods perform comparably when the categories are sufficiently common (over 300 instances per category).

Our motivation to carry out this work is both to provide an applicable text classifier to automatically predict the categories of XSSC documents, and to evaluate the performance of different machine learning methods on text classification. Much manual work will be saved if the performance of this classifier can favorably satisfy the requirements of XSSC. Generally, there is no superior algorithm in the statistical learning area. Even with the same classifier, different performance may be revealed with different types of data sets because until now no statistical analysis was applied to verify the impact of difference in the data on the performance variation of these classifiers.

2.2 Introduction to SVM

SVM is a classifier derived from statistical learning theory by Vapnik and Chervonenkis and it was firstly introduced in 1995 (Mulier, 1999). Based on VC theory and also the kernel theory (Cristianini and Taylor, 2000), SVM was proposed that is equivalent to solve a linearly constrained quadratic programming problem so that the solution of SVM is always globally optimal. An SVM for non-separable case is trained via the following optimization problem.

$$\min_{\omega} \frac{1}{2}\|\omega\|^2 + C\sum_i \xi_i \quad (2)$$

with constraints

$$y_i(x_i w + b) \geq 1 - \xi_i \quad \xi_i \geq 0, \forall i \quad (3)$$

Considering the multi-class classification in this paper, the One-Against-the-Rest approach was adopted (Weston and Watkins, 1999). With this method, k-class pattern recognition was regarded as a collection of $\frac{k(k-1)}{2}$ binary classification problems. The *kth* classifier constructs a hyperplane between class n and the other *k*-1 classes. A majority vote across the classifier or some other measures can be applied to classify a new point. In addition, other methods for *k*-classes (*k*>2) classification are also discussed in (Weston and Watkins, 1999) such as error-correcting output codes, SVM decision tree, etc.

2.3 Introduction to BPNN

Rumelhart, Hilton and Williams (1986) presented a method known as Back Propagation for updating the weights of a multilayered network undergoing supervised training. Back Propagation learns by iteratively processing a data set of training tuples, comparing the network's prediction for each tuple with the known target value. For each training tuple, the weights are modified so as to minimize the mean squared error between the network's prediction and the actual target value. The back propagation algorithm defines two sweeps of the network: a forward sweep from the input layer to the output layer, and then a backward sweep from the output layer to the input layer. The backward sweep is similar to the forward sweep, except that error values are propagated back through the network to control how the weights are changed during training. During training, each input sample will have an associated target vector. The objective of training is to find a set of network weights that provide a solution to the particular problem at hand. For the details of the back propagation algorithm, readers can refer to Han and Kamberl (2006).

3. XSSC Web Document Representation and Data Normalization

The purpose of this section is to describe the documents collected from XSSC website and the preprocessing used to produce the standard data set available for performance examination of SVM and BPNN.

3.1 Dictionary Construction for XSSC Web Documents

Based on our prior work addressed at Zhang and Tang (2006), 192 Web documents were collected from XSSC Website using a web crawler. To mine text, we first need to process it into a form that data mining procedures can use. This typically involves generating features in a spreadsheet format, i.e., constructing a text dictionary (word set) for the text representation using bag of words method. In this paper, we did not conduct any deep analysis of the linguistic content of the documents. ICTCLAS[*] is employed to conduct morphological analysis on

[*] Chinese Lexical Analysis System: ICTCLAS. Online: http://nlp.org.cn/~zhp/ ICTCLAS/codes.html

the Chinese text, to segment it into a collection of individual words. Nouns and substantive expressions were retained as dictionary keyword candidates for Web texts. Fig.1 is our heuristic method to construct the keyword dictionary for text representation. In detail, the 15% words with the highest frequency in the text were selected as initial keywords for each text. Next, we combined all the initial keywords into an overall words collection, and selected only the 5% of highest frequency words of the overall collection to construct domain word collection for all texts. Also, the domain words should be examined by experts of XSSC. Then each initial keyword collection was used to obtain the final keywords for text after being subtracted the domain words collection from it. Finally, all the retained keywords (8392 keywords) of each text were combined to construct the dictionary for text representation. The reason for this method in constructing the feature dictionary is that usually there are some words which have very high frequency in each text, but actually, these words are not so powerful in identifying the text they belong to, such as "science", "system", etc. The motivation for this method of texts analysis is based on Zipf's law (see Zipf and George (1949)). Another reason is that our Chinese word-frequency distribution is not as the same as Luhn's description of that in English in Luhn (1958). This topic was discussed in detail in Zhang (2006).

3.2 XSSC Web Document Representation

After the keyword dictionary of XSSC text collection was established, text representation was conducted using Boolean model. That is expressed in equation (4).

$$Doc(i) = (k_{i,1}, \ldots, k_{i,j}, \ldots, k_{i,m}), \text{ let } k_{i,j} = \begin{cases} 1, \text{ if keyword } j \text{ occurrs in the } i\text{th document} \\ 0, \text{if keyword } j \text{ is absent from the } i\text{th document} \end{cases} \tag{4}$$

$m = 8392$ is the total number of keywords contained in the dictionary for text representation

Thus, 192 Boolean vectors were obtained to represent the 192 Web documents mentioned above initially. Then, a cosine transformation was conducted with these Boolean vectors to represent the documents in another way. That is, let $\overline{k}_{ij} = \dfrac{Doc(i)Doc(j)}{|Doc(i)||Doc(j)|}$ and the 192 text representation vectors were replaced with the newly calculated cosine vectors $\overline{\mathrm{Doc}}(i) = (\overline{k}_{i,1}, \overline{k}_{i,2}, \ldots, \overline{k}_{i,192})$ instead of the original 192 Boolean vectors. Our motivation for using cosine transformation for text representation is specified in Zhang (2006). It should be pointed out here that the following data preprocessing and the latter performance examination were all carried out on these transformed representation vectors.

3.3 Data Preprocessing

One of the significant characteristics of the XSSC documents is its documents length. Although, only 192 documents involved, it has totally 7628 sentences with average 41.5 sentences each text. The widely adopted bench data for text classification Reuters has 19403 valid texts but with only total 103011 sentences at average 4.5 sentences for each text. Then it is more difficult to capture the important features from XSSC documents. Other characteristics of

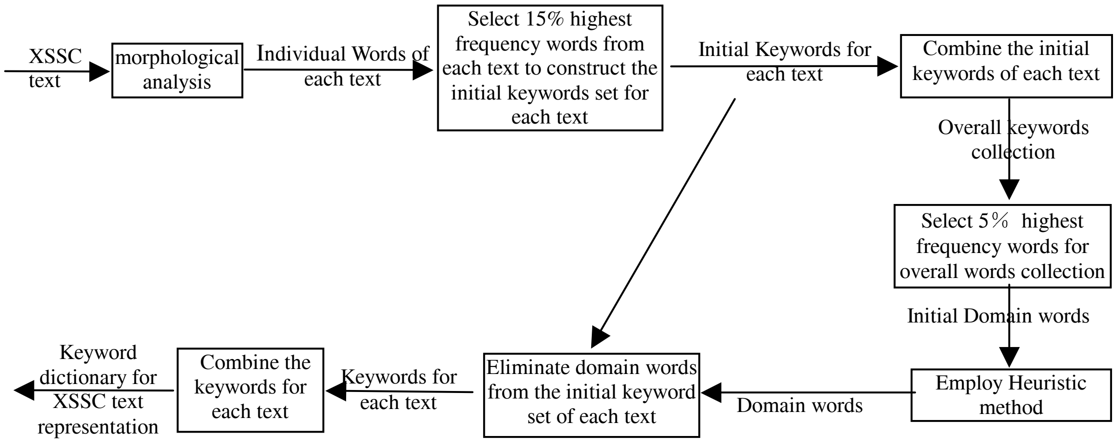

Figure 1 Construction of keyword dictionary for text representation

XSSC texts such the wide rang of its content, the unique terms from academic research which can not identified by the morphological analysis tool, etc., increases the difficulty in the feature selection for text classification on XSSC documents.

In order to obtain the standard data set for the performance examination of SVM and BPNN, the 192 documents were classified into standard categories by both HCA and heuristic tuning method. Clustering techniques apply when there is no class to be predicted and the instances are required to be divided into natural groups. Usually, the clustering techniques can only give us a rough description of the groups hiding in our data set. The heuristic method provided by human experts is necessary if we want to normalize the data set into the standard categories which were provided by the XSSC committee. The standard textual data set was obtained through the following 2 steps.

Step 1: The similarity vectors which represented the XSSC documents were processed by HCA in SPSS, and a dendrogram was generated to describe the overall distribution of the documents.

Step 2: Heuristic method was employed by conducting manual adjustment on the document clusters obtained in Step 1 to normalize them into standard XSSC categories provided by XSSC committee.

Table 1 is the standard document categorization generated by the above processing, and this data set was used as bench mark data for performance examination on SVM and BPNN.

From Table 1, a skewed category distribution and a general research trend currently in XSSC can be seen. For instance, life science has the highest percentage, 31.25%, among all the scientific disciplines; almost two times second one, which is resource and environment science at 16.15%. Further, the top three categories, which account for less than 1/4 of the total 14

XSSC categories, occupied nearly 60% of the total number of XSSC documents. Nevertheless, 5 outliers are detected and clustered as one category with no specified label. Further exploration indicates that their representation vectors are almost zeros when represented by Boolean model. After normalization of the data set, 187 documents were categorized into 13 standard categories.

Table 1 Standard documents classification on XSSC data set

Category ID	Subject of Disciplines	Total	Percentage
1	Life Science	60	31.25
2	Resource and Environment Science	31	16.15
3	Basic Science	21	10.94
4	Scientific Policy	16	8.33
5	Material Science	15	7.81
6	Transportation and Energy Science	11	5.48
7	Information Science	8	4.17
8	Space Science	6	3.13
9	Complexity Science	6	3.12
10	Outliers	5	2.60
11	Aeronautics & Astronautics	4	2.08
12	Micro-electronic Science	3	1.56
13	Safety Science	3	1.56
14	Other	3	1.56

4. Experiments Design

In this section, binary and multi-class text classifications were designed specifically to compare the performance of SVM and BPNN. Here, the problem of unbalanced data was addressed by assigning different amounts of training data and test data.

4.1 Binary Text Classification Experiment Design

For the binary classification, as the documents which belong to the category "life science" is isolated from other documents, two classes (positive class and negative class) of documents are acquired, to test the binary classification of SVM and BPNN. In order to gauge performance for different amounts of training data and attack the skewed data problem, we create nested training sets of 1/4, 1/3, 1/2 and 3/4 of total documents, and left other documents as test data. In detail, training samples of 15, 20, 30, 45 documents from among the "life science" documents, were randomly selected and accordingly 32, 42, 64, and 95 documents are available for other class. Other unselected documents in both classes are used for the test set. Table 2 shows our design for binary classification examination.

4.2 Multi-class Text Classification Experiment Design

As for multi-class text classification, four experiments were designed to examine the performance of SVM and BPNN. For simplification, only 3-class examination is conducted since the classification of more than three classes is similar with that for three classes.

And, the number of document samples was not sufficient to carry out classification for more than three classes. However, the test strategy here is different from the binary classification test mentioned previously. In 3-class examination, number of training data is fixed as two times the number of test data, but the categories of the data sets are selected to be different from each other. Table 3 shows our experiment design for the 3-class examination on SVM and BPNN. It can be seen that the number of training data and test data follows a decreasing trend because we also want to study on the performance of SVM and BPNN for varying amounts of training and test data.

Table 2 Experimental design for binary classification

Test No.	Test 1	Test 2	Test 3	Test 4
Number of Training data	15/32	20/42	30/64	45/95
Number of Test data	45/95	40/85	30/63	15/32

4.3 SVM Kernel Selection and BPNN Specification

Any function which satisfies Mercer's condition can be used as kernel function. In this paper, the polynomial kernel $K(s,t) = ((s \bullet t)+c)d$ (c=1,d=2) is used as the kernel function of SVM classifier. In the 3-class text classification, the One-Against-the-Rest method was adopted, as it has the same computational complexity as the One-Against-One (OAO) in the SVM classifier, and usually performs well (Rennie and Rifkin, 2001).

One of the intriguing aspects of neural networks is that, although they have nodes with very limited computing capability, when many of these nodes are connected together the complete network is capable of performing complicated tasks, and even a single neural network for multi-class pattern classification provides a neural learning process with all class information, which can result in, in theory, an optimal classification (Ou and Murphey, 2007). The BPNN network in this paper is a three-layer fully connected feed-forward network which consists of an input layer, a hidden layer and an output layer. The "tansigmod" function was used in the hidden layer with 5 nodes, and "purelinear" function for the output layer with 1 node*. In our experiment, BPNN network was designed as shown in Figure 2.

4.4 Combined Method

Based on individual classifiers' results, a combined method is in trial for category prediction improvement. The reason for the combined method is that we want to investigate whether an improvement in prediction accuracy can be achieved by combing the prediction results of SVM and BPNN. If the unlabeled sample was predicted with the same label by both SVM and BPNN, it was then labeled with this "agreed" label. Otherwise, it was given no label, and was not assigned to any class. In binary classification, we have two classes: positive and negative. In multi-class classificat-

* Neural Network Toolbox for MATLAB. Online: http://www.mathworks.com/products/ neural-net/

Table 3 Experiment design for multi-classification

Test No.	Test 1	Test 2	Test 3	Test 4
Selected Categories	Life Science/ Environment Science/ Other classes	Environment Science/Basic Science/Scientific Policy	Basic Science/Scientific Policy/Material Science	Scientific policy/ Material Science/Energy Science
Numbers of Training Data	30/20/20	20/14/11	14/11/10	11/10/8
Numbers of Testing Data	15/10/10	10/7/5	7/5/5	5/5/3

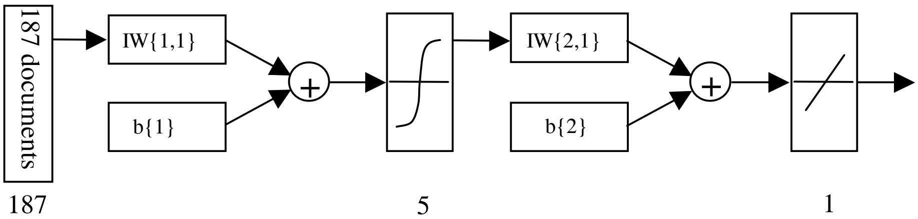

Figure 2 BP network with 5 nodes in hidden layer and 1 node in output layer

ion, we have three classes, Class 1, Class 2 and Class 3. The accuracy of the combined method is calculated by the following formula (see equation 5).

$$Accuracy(Combined\ Method) = \frac{\left|S_{L(\mathrm{SVM})=L(\mathrm{BPNN})=L(\mathrm{Standard})}\right|}{\left|S_{L(\mathrm{SVM})=L(\mathrm{BPNN})}\right|} \qquad (5)$$

where, $S_{L(\mathrm{SVM})=L(\mathrm{BPNN})}$ denotes the set of those test tuples for which SVM and BPNN gave the same label. $S_{L(\mathrm{SVM},\,i)=L(\mathrm{BPNN},\,i)=L(\mathrm{Standard},\,i)}$ denotes the set of those tuples whose labels are given same by both methods and also in accord with the standard labels..

5. Experiment Results

According to the design in Section 4, relevant experiments are then conducted where classifiers of both methods are implemented by mySVM (Stefan, 2000) and Matlab Neural ToolBox. Each designed test is repeated 10 times, and the average values of indicators are given to observe the performance of SVM and BPNN performance.

5.1 The Results of SVM and BPNN on Binary Text Classification

The results of SVM and BPNN on binary text classification are shown in Table 4. The general indicators on information retrieval, accuracy and recall, are adopted here to measure the classification performance of SVM and BPNN. Take Test 1 for BPNN as an example, we obtained the accuracy of 0.8929, which comes from 34 of 45 positive samples and 91 of 95 negative samples, being classified correctly into their corresponding classes by BPNN. And the recall number 38/102 means that 38 of 45 positive samples and 102 negative samples were recalled by BPNN in this test.

From Table 4, it can be seen that BPNN performed favorably, comparable to SVM on the measure of accuracy and recall. Nevertheless, it seems on the whole that BPNN has slightly better performance than SVM in binary text classification, at least as far as the recall is concerned.

5.2 The Results of SVM and BPNN on Multi-class Text Classification

The results of SVM and BPNN on multi-class text classification are shown in Table 5. In the Test 1 for BPNN, the accuracy of 0.7134 is obtained; specifically 10 of 15 from Class 1, 9 of 10 from Class 2 and 6 of 10 from Class 3 are assigned with right labels. And the recall number 10/16/9 means that 10 of 15 class No.1 samples, 16 class No.2 samples and 9 class No.3 samples were recalled by BPNN in this test.

From Table 5, it can be seen that SVM outperforms BPNN on the task of XSSC document 3-class classification, and the result from SVM classifier is convincingly better than that from BPNN on measures of accuracy and recall.

Table 4 The results of SVM and BPNN on binary text classification

Classifier \ Test No.		Test 1	Test 2	Test 3	Test 4
BPNN	Accuracy	0.8929 (34/91)	0.920 (37/78)	0.8710 22/59)	0.9362 (14/30)
	Recall	38/102	44/81	26/67	16/31
SVM	Accuracy	0.8714 (33/91)	0.8640 (27/81)	0.9032 (24/60)	0.9362 (13/31)
	Recall	37/103	31/94	27/66	14/33

Table 5 Results of SVM and BPNN on multi-class text classification

Classifier \ Test No.		Test 1	Test 2	Test 3	Test 4
BPNN	Accuracy	0.7143 (10/9/6)	0.5909 (8/2/3)	0.4706 (2/3/3)	0.6923 (3/3/3)
	Recall	10/16/9	10/5/7	4/8/5	5/4/4
SVM	Accuracy	0.7714 (11/8/8)	0.6364 (9/3/2)	0.4706 (5/1/2)	0.8462 (4/4/3)
	Recall	14/11/10	11/7/4	9/3/5	5/4/4

Table 6 Results from combined method on binary and multi-class text classification

Classification \ Test No.	Test 1	Test 2	Test 3	Test 4
Binary classification	0.9431(116/123)	0.9804(100/102)	0.9186(79/86)	0.9767(42/43)
Multi-class classification	0.9200(23/25)	0.6875(11/16)	0.5714(4/7)	0.8889(8/9)

5.3 The Result with the Combined Method

The combined method introduced in Section 4 was conducted on binary text classification and 3-class text classification. Table 6 is the experiment result from the combined method. Taking the binary text classification with Test.1 as an example, we obtained accuracy of 0.9431, which resulted from 123 texts being given the same label by both BPNN and SVM in Test 1 of binary text classification, and 116 of these 123 texts were categorized with the same label as the standard documents classification on XSSC data set. It can be deduced from Table 6 that the combined accuracy was significantly better than either SVM or BPNN. A particular comparison between combined method and SVM and BPNN is given in Figure 2.

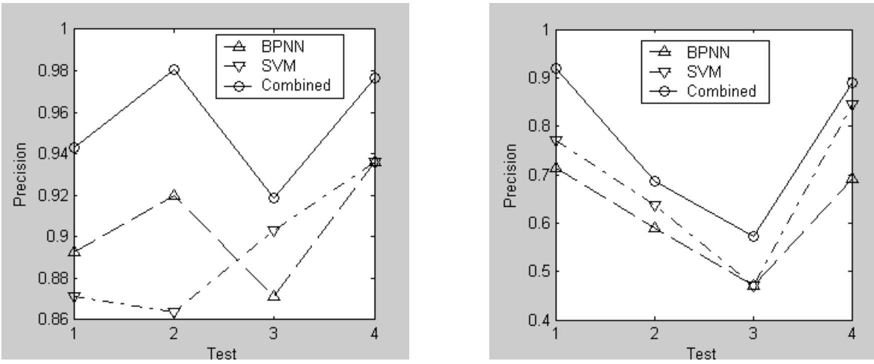

Figure 3 BP network with 5 nodes in hidden layer and 1 node in output layer. Accuracy of combined method, SVM method and BPNN on binary classification and multi-class text classification

6. Concluding Remarks and Further Research

In this paper, we have carried out an experiment on the tasks of binary and multi-class classification with SVM and BPNN on XSSC documents. In order to improve the classification performance of SVM and BPNN, we also developed a combined method to combine the results from SVM and BPNN. The experiment results demonstrated that BPNN and SVM are comparable on the task of binary text classification. However, for multi-class text classification, SVM has obtained better performance than BPNN on the measures of accuracy and recall. This point is very controversial; with the statement in Rennie and Rifkin (2001) that SVM has better performance in multi-class classification because it is more powerful in binary classification than other learning methods. With the adaptation of our combined method, an improvement was achieved in accuracy in both binary and multi-class classification. The motivation to propose the combined method is that we want to verify whether an improvement in accuracy will occur if the prediction results from different classifiers are combined. And the experiment result has validated our hypothesis.

Although the initial results have shown some clues on constructing the XSSC text classifier, we cannot boldly generalize our conclusion from this study, for the reason that our work is on a particular data set, and more investigation should be conducted to make our conclusion more convincing and widely accepted.

One of the promising directions in the text mining field concerns predictive pattern discovery from large quantities of documents. To achieve this goal, not only the required learning algorithms but also semantics, especially ontology techniques such as ontology mapping and ontological indexing, should be introduced into the text mining field (Weiss, Indurkhya and Zhang, 2005; Jacob, Stephen Michael and Alexander, 2006). Since we have conducted an initial survey on the performance of statistical learning methods, in the future, more attention will be concentrated on the area of semantic Web and ontology based knowledge management, especially on work that employs ontologies to describe the existing concepts in text collection, in order to represent document more precisely, and to explore the relationships among concepts from textual resources automatically (John, Dieter and Frank, 2003).

References

[1] Adeva, J.J.G. & Atxa, J.M.P. (2007). Intrusion detection in Web applications using text mining. Engineering Applications of Artificial Intelligence, 20(1): 555-566

[2] Cristianini, N. & Taylor, J. S. (2000). An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press. New York

[3] Hiissa, M. et al. (2007). Towards automated classification of intensive care nursing narratives. International Journal of Medical Informatics. In Press

[4] Han, J.W. & Kamberl, M. (2006). Data Mining Concepts and Techniques (Second Edition). Morgan Kaufmann Publishers. San Francisco

[5] Jacob, K., Stephen, P., Michael, S. & Alexander, R. (2006). Ontology based text indexing and querying for the semantic web. Knowledge-Based Systems, 19: 744-754

[6] John, D., Dieter, F. & Frank, V. H. (2003). Towards the Semantic Web: Ontology-Driven Knowledge Management. John Wiley & Sons, Ltd., New York

[7] Liu, Y.J. & Tang, X.J. (2006). Developed computerized tools based on mental models for creativity support. International Journal of Knowledge and System Sciences, 3(3): 34-40

[8] Liu, Y.J., Tang, X.J. & Li, Z. H. (2005). A preliminary analysis of XSSC as transdisciplinary argumentation. In: Liu, S.F., et al, (eds.), New Development of Management Science and Systematic Science (Proceeding of The 8th Youth Conference on Management Science and System Science). 35-40, Nanjing, May 7-10, Press of HeHai University. (in Chinese)

[9] Luhn, H.P. (1958). The automatic creation of literature abstracts. IBM Journal of Research and Development, 2(2): 159-165

[10]Mulier, F. (1999). Vapnik- Chervonenkis (VC) learning theory and its application. IEEE Transaction on Neural Networks, 10(5): 5-7

[11]Ou, G. & Murphey, Y. (2007). Multi-class classification using neural networks. Pattern Recognition, 40: 4-18

[12]Rennie, J. D. & Rifkin, R. (2001). Improving Multi-class Text Classification with the Support Vector Machine. Master's

thesis. MIT

[13]Rumelhart, D. E., Hinton, G. E. & Williams, R. J. (1986). Learning internal representations by error propagation. In: Parallel Distributed Processing, Exploitations in the Microstructure of Cognition, Vol. 1. Cambridge, MA: MIT Press. 318-362

[14]Stefan, R. (2000). mySVM-Manual. Available via DIALOG. http://www-ai.cs. unidortmund.de /software/ mysvm

[15]Tang, X.J., Liu, Y.J. & Zhang, W. (2005). Computerized support for idea generation during knowledge creating process. In: Khosla, R. J. Howlett, and L. C. Jain (eds.), Knowledge-Based Intelligent Information & Engineering Systems (Proceedings of KES'2005, Part IV), Lecture Notes on Artificial Intelligence, Vol.3684: 437-443, Springer-Verlag, Berlin Heidelberg

[16]Wahba, G. (1999). Support vector machines, reproducing kernel Hilbert spaces and their randomized GACV. In: Advances in Kernel Methods-Support Vector Learning, 69-88. MIT Press

[17]Weiss, S.M., Indurkhya, N., Zhang, T. & Damerau, F. (2005). Text mining - Predictive Methods for Analyzing Unstructured Information. Springer Science and Business Media, Inc. New York

[18]Weston, J. & Watkins, C. (1999). Multi-class support vector machines. In Proceedings ESANN. Brussels

[19]Yang, Y.M. & Lin, X. (1999). A re-examination of text categorization methods. In: Proceedings on the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. 42-49. Berkeley, California, USA

[20] Zhang, Y.Y. & Jiao, J.X. (2007). An associative classification based recommendation system for personalization in B2C e-commerce application. Expert Systems with Applications, 33(1): 357-367

[21] Zhang, W. (2006). Information support tool based on web text mining and its application. Master thesis. Academy of Mathematics and Systems Science, Chinese Academy of Sciences.(in Chinese)

[22]Zhang, W. & Tang, X.J. (2006). Web text mining on a scientific forum. International Journal of Knowledge and System Sciences, 3(4): 51-59, December 2006

[23]Zhang, W. & Tang, X.J. (2006). Information Support tool based on web content mining. Journal of Management Review, 18(9): 21-26 (in Chinese)

[24] Zipf, G. K. (1949). Human Behaviour and the Principle of Least Effort, Addison-Wesely, Cambridge, Massachusetts

Wen Zhang is a PhD student in School of Knowledge Science, Japan Advanced Institute of Science and Technology. His current research interest is in knowledge discovery from text that includes computational linguistics and statistical machine learning. He has published 10 papers until now.

Xijin Tang is an Associate Professor at the Department of Management, Decision-Making and Information System, Institute of Systems Science, Chinese Academy of Sciences. Her current research interests are creativity support systems, expert mining, knowledge synthesis,

modeling and model integration and social network analysis. Her publications until now include 34 Journal papers and 63 conference papers.

Taketoshi Yoshida is Professor of School of Knowledge Science, Japan Advanced Institute of Science and Technology. His current research interests are management information system and evidence based medical information system.