BULGARIAN ACADEMY OF SCIENCES

CYBERNETICS AND INFORMATION TECHNOLOGIES • Volume **14**, Special Issue

Sofia • 2014 Print ISSN: 1311-9702; Online ISSN: 1314-4081
DOI: 10.2478/cait-2014-0041

Research of Two Class Confidence Classification Based on One Class Classifier

Jiang Fangchun[1,2], *Tian Shengfeng*[1]

[1] *School of Computer and Information Technology, Beijing Jiao Tong University, Beijing 100044, China*
[2] *Shen Zhen Institute of Information Technology, Shen Zhen, Guang Dong 518172, China*
Email: 06112061@bjtu.edu.cn

Abstract: *To have simple and efficient confidence machine learning is an important focus in confidence machine researches. Using one class classifier as a tool, the paper applies it twice for two-class classification problems. Setting reject options and a multi-layer ensemble learning method are used in this study. In this method there is no necessity to set up a specific threshold and the confidence computation is omitted. Realizing five experiments, the study proves it as efficient.*

Keywords: *Confidence machine, credibility, one class classifier, ensemble learning, reject option.*

1. Introduction

The confidence machine refers to providing a convincible judgment of the learning result or pre-classification of the learning result in the process of machine learning. The confidence machine is significant in many fields, such as medical diagnoses. It is a branch of machine learning. Since it is new, there are not many relevant theories or methods.

Vovk et al. [1] proposed the transductive learning, a direct way to measure the confidence. This method not only provides a classification result, but also confidence classification. Liu and Nakagawa [2] proposed an indirect way to measure the confidence. The classification distance of the machine learning result was transformed to posterior probability for confidence evaluation. Richard and Lippmann [3] used multiple neural network classifiers to output the expectation that was treated as posterior probability. By setting reject options, we can have pre-classification, like Grandvalet described in [4]. He set reject options in the output to eliminate the results with low confidence to achieve confidence classification.

Usually the machine learning method is used once and the result is obtained. As [5] has suggested, it is better to combine several methods or use one machine learning method several times for confidence evaluation to enhance the confidence level.

Machine learning has some advantages under special conditions. For example, it can make up for the imbalanced data and is able to reach a more balanced classification. As the abnormal conditions are always insufficient, one class classifier works well to distinguish abnormal conditions.

With one class classifier as a tool, this paper uses one class classifier twice to address two-class classifier problems and by setting a reject option and using an ensemble learning method, it designs a confidence machine algorithm, namely, TCCC-OCC algorithm (Two Class Confidence Classification Based on One Class Classifier). Five experimental databases, including the one for ionosphere and sonar are used to test the algorithm.

2. Theory for the algorithm design

The algorithm is based on one class classifier theory, a reject option and ensemble learning methods.

2.1. One class classifier

One class classifier [6] only classifies the first class of objects in the sample. The categorized first class objects are called a positive class. The other objects are called a negative class. In essence, it establishes a super sphere that concludes as many positive classes as possible while making the sphere as small as possible.

Suppose that the sample set for a positive class is $\{x_1, x_2, \cdots, x_n\}$, R refers to the diameter of the super sphere, a refers to the centre of the super sphere. To optimize it, we introduce the slack variable ξ; C refers to the factor of a penalizing parameter adjustment. The problem is transformed to:

(1) $$\min\ L(R) = R^2 + C\sum_{i=1}^{n} \xi_i\ ,$$

(2) $$\text{s. t. } \|x_i - a\|^2 \le R^2 + \xi_i,\ \ \xi_i \ge 0.$$

According to (1) and (2), supposing that Lagrange multiplier is $\alpha \ge 0$, $\gamma \ge 0$, the Lagrange function is

(3) $$L(R, a, \alpha, \gamma, \xi) = R^2 + C\sum_{i=1}^{n}\xi_i - \sum_{i=1}^{n}\alpha_i\left[R^2 + \xi_i - \left(x_i^2 - 2ax_i + a^2\right)\right] - \sum_{i}^{n}\gamma_i\xi_i.$$

Using (3) to get the partial differential by R and a and making the partial differential 0

(4) $$\max\ L = \sum_{i=1}^{n}\alpha_i\left(x_i \cdot x_i\right) - \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\left(x_i \cdot x_j\right),$$

(5) $$\text{s.t.} \sum_{i=1}^{n} \alpha_i = 1 \qquad 0 \le \alpha_i \le C.$$

Expression (4) is a convex quadratic programming. If for x_i there is $\alpha_i > 0$, then x_i is a support vector. All support vectors compose the super sphere. If $\|x_i - a\| \le R^2$ and make $\alpha_i = 0$, then x_i is within the super sphere.

For a new unknown sample if there exists

(6) $$f(z) = \|z - a\|^2 = (z \cdot z) - 2\sum_{i=1}^{n} \alpha_i (z \cdot x_i) + \sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j (x_i \cdot x_j) \le R^2,$$

then such a sample is called a positive class.

As one class classifier [7] only defines first class objects in the sample, it is relevant only to those samples with positive samples.

2.2. A reject option

When taking the machine learning, we usually divide the space into two complementary areas: the reject region R and the acceptance region (classification region) A. The definition is $R = \{x \mid 1 - \max_i p(\omega_i \mid x) > t\}$, $A = \{x \mid 1 - \max_i p(\omega_i \mid x) \le t\}$, in which t refers to the threshold. The smaller t is, the bigger the reject region R is. If sample x is in the acceptance region A, then classify x according to some learning methods. If it is in the reject region R, then we do not have to classify x.

Chow [8] studied classifiers with a reject option and proposed an optimized classifier and the reject rule. According to Bayes learning rule, if $p(\omega_1 \mid x) \ge 1 - t$, then the classification is ω_1. If $p(\omega_2 \mid x) \ge (1 - t)$, the classification is ω_2. If it does not fall into any classifications, reject it. t is the threshold of rejection and used as the threshold for posterior probabilities. It is a constant number and $0 \le t \le 0.5$. If t is 0 and the Posterior Probability is 1, then all samples can be accepted. If t is 0.5, then the Bayes learning rule for a reject option is not efficient. The posterior probability of rejecting any sample is smaller than $1 - t$, which guarantees high reliability for the confidence evaluation.

2.3. Ensemble learning

The ensemble learning method [9] is to learn through learning algorithms. The production and the application of the fundamental learning machines are very necessary. The combination of several fundamental learning machines also matters. To put it simple, many single learning machines consist of a larger integrated one through a certain way, so as to enhance the learning effect.

There are two ways to produce the fundamental learning machine, namely, the heterogeneous method and the homogeneous method. The heterogeneous method refers to applying different learning algorithms to the same data set. The homogeneous method refers to applying the same algorithm to different data sets. This paper adopts the latter one.

The selection and the production of fundamental learning machines are followed by their integration. In [10] three standards of integration are proposed, the confidence level, the rank level and the abstraction level. The confidence level standard can be applied in the situation where the output result of the fundamental factor is subject to probability distribution. The rank level is applied to the situation where the output is according to the classification result so that it sets a ranking order for the output. The abstraction level standard is applied to the situation where the output results are the classifier tags. This paper uses the confidence level standard and the abstraction level standard to design the algorithm. It refers to [5] for the calculation of the parameters, such as the recognition rate, the accuracy rate and the error rate.

The ensemble learning method is widely applied to medical diagnoses, image processing and biological engineering [11].

3. The algorithm design

The core of the algorithm design is to use the one class classifier to achieve two-class confidence classification and to use multilevel ensemble learning to enhance learning result.

3.1. Use of one class classifier to achieve two-class confidence classification

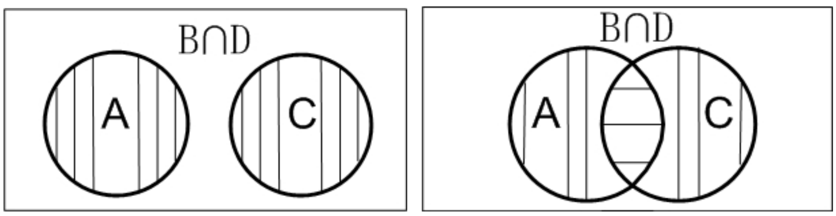
Two-Class Confidence Classification Based on One Class Classifier (TCCC-OCC) refers to the use of one class classifier for positive classification. The result of a positive class is expressed as A and the rest is expressed as B. One class classifier is used for negative classification. The result of a negative class is expressed as C and the rest is expressed as D. There are two situations, the one – A is non-intersect with C, as shown in Fig. 1a; the other is A, intersect with C, as shown in Fig. 1b. After such classification, there are four possible situations for a sample:

① The sample belongs to A only

② The sample belongs to C only

③ The sample belongs to A and C, that is $A \cap C$

④ The sample neither belongs to A, nor C, so it belongs to $B \cap D$



(a) (b)

Fig. 1. A is non-intersect with C after classification (a); comparison of A and C after classification (b)

A sample that belongs to A or C only is called a convincible sample, which is represented by the vertical line in the Fig. 1. A sample that belongs to A and C is called an inconvincible sample, which is represented by the horizontal line in the Fig. 1. A sample that neither belongs to A, nor to C, but belongs to B and D is also called an inconvincible sample, which is represented by the blank in the Fig. 1.

According to the above mentioned definition, we can put the convincible samples into the acceptance region, expressed by the vertical line. And put the inconvincible samples in the reject region, expressed by the horizontal line and the blank.

As a result, the convincible and inconvincible samples are distinguished and put into the reject region and the acceptance region, respectively. Thus we are able to achieve the confidence classification.

3.2. Use multiple confidence sets to enhance the effect of classification

There are several choices to deal with the samples in the reject region after two-class confidence classification. For example, leave it to a human, have later treatment, or have instant response. This algorithm chooses the Ensemble Classifier System to deal with samples in the reject region.

The key lies in the fact of subjecting the samples in the reject region to further confidence classification by a multiple ensemble classification method. In other words, use the one class classifier for a second time. If necessary, use one class classifier for a third time to enhance the effect of classification.

The three-layer confidence ensemble classification is shown in Fig. 2.

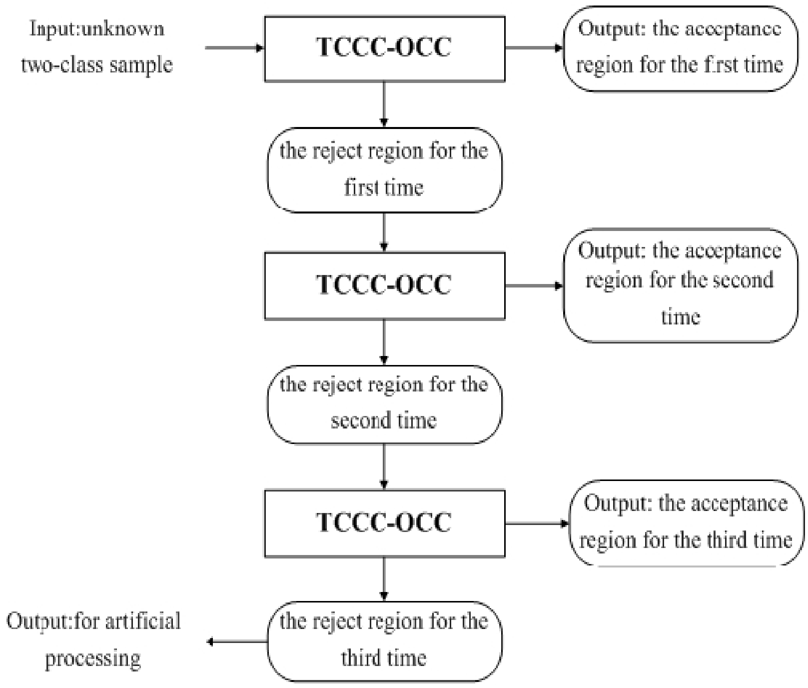


Fig. 2. Three-tier confidence ensemble classification

3.3. Explanation of algorithm computing

To better analyze the result, this paper defines the recognition rate (Recognition rate), the Rejection rate (Rejection rate), Reliability (Reliability), Error rate (Error rate), as follows referring to [5]:

Recognition Rate (RR) = the number of samples that have been recognized correctly/test set samples

Rejection Rate (ReR) = the number of samples that have been rejected/test set samples

REliability (RE) = the number of samples/test set samples that have been recognized correctly

The number of samples that have been rejected)/test set samples

Error Rate (ER) = 100% − RE

Correct Rate (CR) = the number of samples that have been recognized correctly/total number of samples that have been recognized

The definition of the cumulative recognition rate (Recognition rate), cumulative Correct rate (Correct rate), cumulative Rejection rate (Rejection rate), cumulative Reliability (Reliability), cumulative Error rate (Error rate) is described as follows:

Cumulative Recognition Rate (RR) = the cumulative number of samples that have been recognized correctly/test set samples

Cumulative Correct Rate (CR) = the cumulative number of samples that have been recognized correctly/total number of samples that have been recognized

Cumulative Rejection Rate (ReR) = the cumulative number of samples that have been rejected/test set samples

Cumulative REliability (RE) = the cumulative number of samples/test set samples that have been recognized correctly + the cumulative number of samples that have been rejected)/test set samples

Cumulative Error Rate (ER) = 100% − the cumulative RE

According to these definitions, there is: Reliability = recognition rate+ rejection rate. In other words, when the reject option has low confidence, it does not mean that there is no way to recognize the error. It is supposed to take the machine learning again or leave it to an artificial solution.

3.4. The feature of algorithm design

The TCCC-OCC algorithm is able to achieve the confidence classification. It has the following features:

① **Reduces computing.** This algorithm does the analysis by one time of learning to divide the acceptance region and the reject region. So there is no necessity to compute the confidence for every unknown sample. And we do not have to set up a threshold for the reject region.

② **Simplifies the algorithm.** This algorithm adopts one machine learning method, namely, one class classifier. The algorithm is largely simplified.

③ **Flexible control.** With the ensemble and homogenous method, this algorithm is granted with flexibility to set up the ensemble tier and to control the study effect, so that the learning is satisfying.

4. The algorithm realization

The tools and the environment for realization of the algorithm, experiments and analysis are described below.

4.1. Tools and data

LIBSVM [12] is used for one class classifier. MATLAB7.0 is the platform for realizing the algorithm. LIBSVM is based on the hyperplane method that equals to a super sphere method, when the radial basis function kernel is chosen for parameter selection [13]. Four class, sonar, ionosphere, german-numer and australian data set in UCI [14] and others are selected. The parameters are shown in Table.1.

Table 1. Parameters of the data set for the experiment

No	Data set	Type of data set	Eigen-value	Sample number	Number of positive example	Number of negative example	Number of training set	Number of test set
1	Four class	Two-class classification	2	862	307	555	800	62
2	Sonar	Two-class classification	60	208	97	111	150	58
3	Ionosphere	Two-class classification	34	351	225	126	300	51
4	German-numer	Two-class classification	24	1000	300	700	800	200
5	Australian	Two-class classification	14	690	307	383	600	90

First use Python to manage the data set. Then subject the data set to the algorithm. Repeat 10 times in each data set to get an average level. Compare and analyze the results.

4.2. Algorithm

Input:
X : Two-class data sample
Y : Two-class sample tag
Train Set Number: The number of the training set
Test Set Number: The number of the test set
Output:
Test Set A: Acceptance region of the test set
Test Set R: Reject region of the test set
Recognition Rate (RR): Recognition rate

Correct Rate (CR): Correct rate
Rejection Rate (ReR): Rejection rate
REliability (RE): Reliability
Error Rate(ER): Error rate

Process

Step 1. Use the sample set (X, Y). Set up the random function. Draw samples out randomly according to the Train Set Number and Test Set Number to produce a Train Set and a Test Set.

Step 2. Train one class classifier LIBSVM and get parameters relevant to c and g of LIBSVM.

Step 3. Use one class classifier LIBSVM to train the positive samples on the Train Set and get the positive sample recognition model of one class classifier.

Step 4. Use the positive sample recognition model of one class classifier to recognize the positive samples on the Test Set.

Step 5. Use one class classifier LIBSVM to train the negative samples on the Train Set and get the negative sample recognition model of one class classifier.

Step 6. Use the positive sample recognition model of one class classifier to recognize the negative samples on the Test Set.

Step 7. Count the number of the samples that have been recognized correctly and the number of samples that have been recognized wrongly in the Test Set A.

Step 8. Compute the number of samples in the Test Set R.

Step 9. Compute RR, CR, ReR, RE, ER.

Step 10. Judge whether the ensemble learning has been done for three times.

Step 11. If not, continue. If it does, go to Step 14.

Step 12. Train Set=the reject region of Train Set R, Test Set=the reject region of Test Set R.

Step 13. Repeat Steps 2-11.

Step 14. Recycle Steps from 1 up to 13 for 10 times and get the experimental value.

Step 15. End.

4.3. Experiment

The four class data set contains 862 samples, among which 307 are positive samples and 555 are negative samples. The number of the train set is 800 and the test set, 62. Table 2 shows the comparison between LIBSVM and TCCC-OCC for the first classification.

Table 2. Comparison of the classification results of LIBSVM and TCCC-OCC (%)

Algorithm	RR	CR	RE	ER
LIBSVM recognized positive samples	87.10	87.50	70.00	30.00
LIBSVM recognized negative samples	59.68	86.96	47.62	52.38
TCCC-OCC first learning	48.39	96.77	98.39	1.61
TCCC-OCC second learning cumulative	77.42	97.96	98.39	1.61
TCCC-OCC third learning cumulative	87.10	98.18	98.39	1.61

From Table 2 we can see that the correct rate CR and Reliability RE of LIBSVM are lower than those of TCCC-OCC in terms of the recognized positive samples and negative samples. But the Error rate ER of LIBSVM is higher than that of TCCC-OCC. The Recognition rate RR of the first learning of TCCC-OCC is low, because the reject region is deleted. After the second learning or the third learning, RR has increased substantially. At the same time, RE does not change and CR goes up somewhat. Table 3 shows the average value of ten times of the experiment for a fourclass data set.

Table 3. Average value of ten times of the experiment for a four class data set (%)

Item	RR	CR	ReR	RE	ER
First learning	38.71	81.46	52.42	91.13	8.87
Second learning	32.31	79.15	59.43	91.74	8.26
Cumulative value after the second learning	55.65	80.75	31.13	86.77	13.23
Third learning	38.77	97.80	60.15	98.92	1.08

From Table 3 it is seen that the first learning of RR is 50% higher than the second learning and the third learning. The second learning of RR and the third learning are almost the same. The cumulative is on the rise, which means that as the times of learning increases, the number of correct recognition also increases. The first learning of CR is a little higher than the second learning and lower than the third learning, which means that as the times of learning increases, the number of wrong recognition also increases. The first learning, the second learning and third learning of ReR have some variations. But the cumulative decreases, which means that as the times of learning increases, the number of reject decreases. RE is reducing slowly and ER is increasing slowly. So it must be tailored to the real situation to decide how many levels of ensemble learning are to be chosen. A balance needs to be reached among the rejection rate, the reliability and the error rate.

The algorithm is applied to other five data sets. The results are shown in Table 4.

Table 4. Data value of five data sets (%)

<table>
<tr><th>No</th><th>Data set</th><th>Item</th><th>Data</th><th>Max</th><th>Min</th><th>Average</th></tr>
<tr><td rowspan="15">1</td><td rowspan="15">Fourclass</td><td rowspan="5">First learning</td><td>RR</td><td>48.39</td><td>25.81</td><td>38.71</td></tr>
<tr><td>CR</td><td>96.77</td><td>66.67</td><td>81.46</td></tr>
<tr><td>ReR</td><td>64.52</td><td>45.16</td><td>52.42</td></tr>
<tr><td>RE</td><td>98.39</td><td>85.48</td><td>91.13</td></tr>
<tr><td>ER</td><td>14.52</td><td>1.61</td><td>8.87</td></tr>
<tr><td rowspan="5">Cumulative value after the second learning</td><td>RR</td><td>77.42</td><td>46.77</td><td>55.65</td></tr>
<tr><td>CR</td><td>97.96</td><td>72.50</td><td>80.75</td></tr>
<tr><td>ReR</td><td>43.55</td><td>20.97</td><td>31.13</td></tr>
<tr><td>RE</td><td>98.39</td><td>80.65</td><td>86.77</td></tr>
<tr><td>ER</td><td>19.35</td><td>1.61</td><td>13.23</td></tr>
<tr><td rowspan="5">Cumulative value after the third learning</td><td>RR</td><td>87.10</td><td>51.61</td><td>67.58</td></tr>
<tr><td>CR</td><td>98.18</td><td>74.42</td><td>83.15</td></tr>
<tr><td>ReR</td><td>30.65</td><td>11.29</td><td>18.87</td></tr>
<tr><td>RE</td><td>98.39</td><td>80.65</td><td>86.54</td></tr>
<tr><td>ER</td><td>19.35</td><td>1.61</td><td>13.55</td></tr>
</table>

Table 4 (c o n t i n u e d)

No	Data set	Item	Data	Max	Min	Average
2	Sonar	First learning	RR	27.59	13.79	21.38
			CR	84.62	61.54	76.85
			ReR	79.31	65.52	72.24
			RE	96.55	89.66	93.62
			ER	10.34	3.45	6.38
		Cumulative value after the second learning	RR	34.48	18.97	28.10
			CR	86.96	61.11	77.62
			ReR	72.41	56.90	63.97
			RE	94.83	87.93	92.07
			ER	12.07	5.17	7.93
		Cumulative value after the third learning	RR	37.93	18.97	30.69
			CR	84.62	57.89	76.16
			ReR	68.97	50.00	60.00
			RE	93.10	86.21	90.69
			ER	13.79	6.90	9.31
3	Ionosphere	First learning	RR	31.37	17.65	23.73
			CR	100.00	75.00	91.32
			ReR	78.43	66.67	74.12
			RE	100.00	94.12	97.84
			ER	5.88	0.00	2.16
		Cumulative value after the second learning	RR	45.10	33.33	39.02
			CR	100.00	85.00	93.37
			ReR	64.71	50.98	58.24
			RE	100.00	94.12	97.25
			ER	5.88	0.00	2.75
		Cumulative value after the third learning	RR	50.98	39.22	44.31
			CR	100.00	86.96	94.15
			ReR	56.86	45.10	52.94
			RE	100.00	94.12	97.25
			ER	5.88	0.00	2.75
4	German-numer	First learning	RR	18.50	11.50	14.85
			CR	85.29	63.41	78.08
			ReR	83.50	77.50	81.00
			RE	97.50	92.50	95.85
			ER	7.50	2.50	4.15
		Cumulative value after the second learning	RR	23.50	15.00	20.40
			CR	84.31	66.67	76.68
			ReR	77.50	69.50	73.40
			RE	96.00	91.50	93.80
			ER	8.50	4.00	6.20
		Cumulative value after the third learning	RR	26.50	19.50	22.50
			CR	84.13	67.69	76.10
			ReR	75.00	66.00	70.35
			RE	96.00	89.50	92.85
			ER	10.50	4.00	7.15
5	Australian	First learning	RR	58.89	36.67	48.67
			CR	97.62	86.67	91.91
			ReR	60.00	34.44	47.00
			RE	98.89	93.33	95.67
			ER	6.67	1.11	4.33
		Cumulative value after the second learning	RR	74.44	58.89	65.56
			CR	95.45	84.06	90.83
			ReR	35.56	18.89	27.78
			RE	96.67	87.78	93.33
			ER	12.22	3.33	6.67
		Cumulative value after the third learning	RR	85.56	66.67	73.44
			CR	96.00	81.58	89.62
			ReR	26.67	7.78	18.00
			RE	96.67	84.44	91.44
			ER	15.56	3.33	8.56

4.4. Experimental analysis

The algorithm has better performance on the fourclass and Australian data sets. The average CR is between 80% and 90%. ReR can reach from 15% up to 20%, RE, 90% and ER, 10%. However, the result is less satisfying in german-number data set. CR is around 75%, RE, around 90% and ER, 10%. ReR is as high as 70%. The performance of the algorithm on sonar and ionosphere data sets is between the previous two situations.

From the above mentioned discussion, is obvious that the experiment has reached the design requirements and realizes the confidence classification.

5. Conclusion

For TCCC-OCC algorithm, taking one class classifier as a tool, the reject option method is introduced to solve two-class classification problems. The homogeneous multilevel ensemble learning method is used to enhance the learning result. Experiments are taken on five data sets with useful results. This algorithm provides a way of confidence evaluation. It is simple and flexible with little computing.

Further study is needed to figure out the ways to control accurately the reject region under specific requirements on the confidence algorithm.

Acknowledgement: Thanks for the support of Guangdong Natural Science Fund (No S2011010000824) and Shenzhen Science and Technology Collaboration and Innovation Project (GJHS20120627094240003)

References

1. Vovk, V., A. Gammerman, G. Shafer. Algorithmic Learning in A Random World. USA, Springer Science+Business Media, Inc., 2005.
2. Liu, C. L., M. Nakagawa. Precise Candidate Selection for Large Character Set Recognition by Confidence Evaluation. – IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. **22**, 2000, No 6, 636-642.
3. Richard, M. D., R. P. Lippmann. Neural Network Classifiers Estimate Bayesian a Posterior Probabilities. – Neural Computation, Vol. **3**, l991, No 4, 461-483.
4. Grandvalet, Y. Support Vector Machines with a Reject Option. Vancouver B. C. Canada: Advances in Neural Information Processing Systems 21. – In: Proceedings of the 2008 Conference, 2009, 537-544.
5. Ping, Zhang, Tien D. Bui, Ching Y. Suen. Anovel Cascade Ensemble Classifier System with a High Recognition Performance on Handwritten Digits. – Pattern Recognition, Vol. **40**, 2007, 3415-3429.
6. Kyrki, V., J. K. Kamarainen, H. Kalviainen. Simple Gabor Feature Space for Invariant Object Recognition. – Pattern Recognition Letters, Vol. **25**, 2004, No 3, 311-318.
7. Tax, D. M. J. One-Class Classification. Delft, Delft University of Technology, 2001.
8. Chow, C. K. On Optimum Recognition Error and Reject Tradeoff. – IEEE Trans. on Inf. Theory, Vol. **16**, 1970, 41-46.
9. Zhang, Chunxia, Zhang Jiangshe. Review of Selective Ensemble Learning Algorithm. – Chinese Journal of Computers, Vol. **34**, August 2011, No 8.
10. Xu, L., A. Krzyzak, C. Y. Suen. Methods of Combining Multiple Classifiers and Their Application to Handwritting Recognition. – IEEE Transaction on Systems, Man, and Cybernetics, Vol. **22**, 1992, No 3, 418-435.

11. G h o d s e l a h i, A. A Hybrid Support Vector Machine Ensemble Model for Credit Scoring. – International Journal of Computer Applications, Vol. **17**, 2011, No 5, p. 1.
12. **http://www.csie.ntu.edu.tw/~cjlin/**
13. F e n g, A i m i n. Structure-Driven Single Classifier Design and Extension Research. – Nanjing, Nanjing University of Aeronautics and Astronautics, 2011, 3, 14-17.
14. **http://archive.ics.uci.edu/ml/**