A Novel Pattern Rejection Criterion Based on Multiple Classifiers

Wei-Na Wang[1], Xu-Yao Zhang[2], and Ching Y. Suen[1]

[1] CENPARMI, Concordia University, Montreal, Canada
{wein_wa,suen}@encs.concordia.ca
[2] Institute of Automation, Chinese Academy of Sciences, Beijing, China
xyz@nlpr.ia.ac.cn

Abstract. Aiming at improving the reliability of a recognition system, this paper presents a novel SVM-based rejection measurement (SVMM) and voting based combination methods of multiple classifier system (MCS) for pattern rejection. Compared with the previous heuristic designed criteria, SVMM is more straight-forward and can make use of much more information from the training data. The voting based combination methods for rejection is a preliminary attempt to adopt MCS for rejection. Comparison of SVMM with other well-known rejection criteria proves that it achieves the highest performance. Two different methods (structural modification and dataset re-sampling) are used to build MCSs. The basic classifier is the convolution neural network (CNN) which has achieved promising performances in numerous applications. Rejection based on MCS is then evaluated on MNIST and CENPARMI digit databases. Specifically, different rejection criteria (FRM, FTRM and SVMM) are individually combined with MCS for pattern rejection. Experimental results indicate that these combinations improve the rejection performance consistently and MCS built by dataset re-sampling works better than that with structural modification in rejection.

Keywords: Rejection criterion, SVMM, MCS, CNN, soft voting, handwritten digit recognition.

1 Introduction

In pattern recognition, the recognition rate is always an important factor in evaluating the performance of a classifier and plenty of classifiers or multiple classifier systems have achieved high recognition rates based on different datasets like MNIST, CENPARMI and so forth in the past decades. However, although the recognition accuracy of some models has reached error rates of less than 1% on the benchmark MNIST dataset [1, 2, 3, 4] and CENPARMI numeral dataset [5], it is still impossible to reach a 100% recognition accuracy. And a low percentage of errors in recognition could still cause a huge loss in real-life systems, like check-reading in the banks; hence the reliability of a classifier is as important as recognition accuracy, as defined below:

Z.-H. Zhou, F. Roli, and J. Kittler (Eds.): MCS 2013, LNCS 7872, pp. 331–342, 2013.

$$Recognition\ rate = \frac{Number\ of\ correct\ samples}{Total\ number\ of\ testing\ samples}$$

$$Rejection\ rate = \frac{Number\ of\ rejected\ samples}{Total\ number\ of\ testing\ samples}$$

$$Reliability = \frac{Number\ of\ correct\ samples\ among\ non\text{rejected}\ ones}{Total\ number\ of\ testing\ samples - \text{number of rejected samples}}$$

In order to improve the reliability of a classifier, some confusing patterns must be rejected before entering the testing loop in order to prevent errors. That is why some useful rejection criteria are produced to determine and filter out the confusing samples. To evaluate the effectiveness of rejection, we can draw a curve in the coordinate system whose x-axis is the number of rejected samples and the y-axis is reliability. A good rejection criterion can achieve a higher reliability with fewer samples rejected. So in this case, we expect the curve to be as close to the top left corner as possible.

In this paper, our main goal is to improve the reliability of recognition systems by detecting the confusing samples that may easily cause error. To accomplish this goal, we have designed a novel rejection criterion, called SVM-based Measurement (SVMM), which learns the optimal rejection boundary from the training data. Brief descriptions of this criterion as well as several other well-known rejection criteria are presented in Sections 2. After that, we first attempt to use Multiple Classifier System (MCS) for the purpose of pattern rejection. It is implemented by using voting methods to combine decisions from different classifiers. Both hard voting and soft voting are considered and details are followed in Section 3. Section 4 reports all the experimental results and analyses. Specifically, the newly proposed rejection criterion verified and compared with other rejection criteria on MNIST numeral dataset. MCS based rejections with both hard voting and soft voting are evaluated on the same dataset and also CENPARMI numeral dataset with MCSs differing in structural modification and dataset re-sampling. At last, we provide our concluding remarks in Section 5.

2 Rejection Criteria

Pattern rejection can be viewed as a two-class recognition problem, which takes the output values of a classifier as features to recognize a pattern as a confusing one to reject or a clear one to accept. Generally, for a regular classifier, the output is always a vector consisting of confidence values or probabilities of possible classes. Given a pattern x, suppose the output vector of the classification is (c is the number of possible classes)

$$\{f_1, f_2, \dots, f_c\},\ f_i \geq 0,\ i = 1,2,\dots,c \tag{1}$$

After that, this pattern is classified according to $x \in class\ arg\ \max_{1\leq i\leq c} f_i$. In case that the outputs are negative, normalization can be used to guarantee that all the values are positive (e.g. $f_i = f_i - f_{min},\ f_{min} = \min_{1\leq i\leq c} f_i$).

2.1 Existing Rejection Criteria

In the research field of rejection, some traditional rejection criteria have been studied before and have reached high recognition rates as well as high reliability. In this section, some useful criteria are presented.

The first rank confidence value (FR) and the second rank confidence value (SR) can be described as

$$FR = \max_{1 \le i \le c} f_i\,,\quad SR = \max_{1 \le i \le c, f_i \ne FR} f_i \tag{2}$$

FR is expected to be much larger than all the other output values for a clear sample. Besides, the gap between FR and SR is also viewed as a useful index, to reflect the quality of a sample. That is why First Rank Measurement (FRM) and First Two Rank Measurement (FTRM) have been proposed for rejection [6].

FRM is one of the most useful criteria, which takes into account only FR of the output vector. It rejects samples by setting a threshold T_1 to FR and accepts those satisfying $FR \ge T_1$.

FTRM is another important index for rejection. Unlike FRM, it emphasizes the gap between FR and SR. It sets a threshold T_2 to the gap and accepts only the samples satisfying $FR - SR \ge T_2$.

Besides these two well-known rejection criteria, He et al propose a novel LDA measurement (LDAM) in [6, 7], which relies on the principle of Fisher Linear Discriminant Analysis. They apply the principle of LDA on outputs for the rejection option as a one dimensional application which shifts the Fisher criterion to

$$J(w) = \frac{S_B}{S_W} = \frac{(\mu_1 - \mu_2)}{\Sigma_{12}} \tag{3}$$

where μ_1 and μ_2 are the centers of two classes and Σ_{12} is within-class scatter.

Then they define two classes for rejecting and accepting samples: $G^{(1)} = \{\hat{f}_1\}$ and $G^{(2)} = \{\hat{f}_2, \dots, \hat{f}_c\}$, in order to maximize the separation between FR and all the other confidence values. (Here $\hat{f}_{\mathrm{i}}$ are confidence values in a descending order). Thus, in LDA, $J(w)$ can be defined by:

$$J(w) = \frac{\left\{\sum_{i=2}^{c}(\hat{f}_1 - \hat{f}_{\mathrm{i}})\right\}^2}{(c-1)^2\Sigma_{12}} \tag{4}$$

where $\mu_1 = \hat{f}_1, \mu_2 = \frac{1}{c-1}\sum_{i=2}^{c}\hat{f}_{\mathrm{i}},\ \Sigma_1 = 0,\ \Sigma_2 = \frac{1}{c-1}\sum_{i=2}^{c}(\hat{f}_{\mathrm{i}} - \mu_2)^2$ and $\Sigma_{12} = \frac{1}{2}\Sigma_2$.

Then a threshold T_3 is set and samples are accepted if they satisfy $J(w) \ge T_3$. The criterion has been proved to produce a better performance than FRM and FTRM based on eight-direction gradient feature with SVM classifier for handwritten character recognition [6, 7].

2.2 SVM-Based Rejection Measurement (SVMM)

The previous rejection criteria have been designed based on some heuristic ideas. In this section, we propose a new SVM-based rejection measurement (SVMM) to extend the rejection process into a learning based method. Specifically, rejection can be viewed as a two-class recognition problem, one stands for rejected samples and the other for accepted ones. For a classifier, the output of a sample is a vector of confidence values $\{f_1, f_2, \ldots, f_c\}$, $f_i \geq 0$, $i = 1,2,\ldots,c$, as mentioned before. Then these values are extracted as features and sorted into a descending order:

$$\{\hat{f}_1, \hat{f}_2, \ldots, \hat{f}_c\},\ \hat{f}_1 \geq \hat{f}_2 \geq \cdots \geq \hat{f}_c \tag{5}$$

The correctly and wrongly classified samples are labeled differently (correctly classified samples with label "1" while incorrectly classified ones with label "-1") and used to train an SVM classifier. Linear SVM is selected for training to locate the rejection boundary. So the decision boundary is a linear function combining all the components of the output vector, represented in Eq. (6). ($\{w_i\}_{i=0}^{c}$ are the coefficients of SVM)

$$T = \sum_{i=1}^{c} w_i \hat{f}_i + w_0 \tag{6}$$

The reason for choosing a linear kernel for SVM rather than a nonlinear one, like RBF kernel, is based on the following points:

1. A linear kernel works very fast in training and testing and an optimal linear separating boundary is a good way to avoid over-fitting.
2. A linear boundary is more meaningful physically and function (6) includes some special cases in it. For instance, FRM can be viewed as a linear boundary with $w_1 = 1$ and $w_2 = w_3 = \cdots = w_c = w_0 = 0$; while FTRM can be viewed as: $w_1 = 1$, $w_2 = -1$ and $w_3 = w_4 = \cdots = w_c = w_0 = 0$.

Note that in the training process of SVMM, the number of samples in class "1" is always much larger than that of class "-1", because the baseline accuracy of the classifier is high. In this case, the problem is an unbalanced classification problem. To solve this problem, we use different weighting functions for different classes in the "libsvm" software [8]. In the testing process, the same features are extracted and sorted into descending order, and a sample is rejected if T in Eq. (6) is smaller than a pre-defined threshold.

With this new criterion, the linear rejection boundary is located by training an SVM with training data. The main difference between SVMM and other criteria, like FRM, FTRM and LDAM, is that SVMM extends the rejection process from heuristic design to learning based procedure. Using learning based method on the training set to predict the rejection on testing samples is more straight-forward and can make use of much more information from the data.

3 Rejection with Multiple Classifier System

3.1 Construction of Multiple Classifier System (MCS)

Since convolution neural network (CNN), especially MCS based on CNN, works effectively in handwritten character recognition as shown in [4, 9, 10], it is selected as the core classifier and MCS is built on it in our strategy. The CNN classifier is based on the principle of deep learning. It processes the raw images of samples and extracts useful trainable features to classify samples into different categories [1].

Re-sampling the dataset (with Bagging [11], Boosting [12] and so forth) and changing the classifier (in structure or type [13]) are two main ways to produce committees. Many researchers have used these methods to produce a group of classifiers and applied certain combination methods for recognition. Some of them have achieved extremely high recognition rate in handwritten numeral recognition with CNN model on MNIST dataset [9, 10].

For the construction of the MCS, we select the CNN model in [4] as the basis model "M0". It has three convolution layers with 25, 50 and 100 feature maps sequentially, and one output layer which is fully connected to the last convolution layer. Two modifications have been explored: one is changing the number of feature maps in each of three convolution layers in both increasing and decreasing ways to build new models. The other is using "Bagging" method (i.e. dataset re-sampling) to randomly select samples for the training sets to train the same CNN model numerous times. The structures of the modified classifiers are listed in Tables 1 and 2, while the information of re-sampling datasets is listed in Table 3.

3.2 Rejection Based on MCS

MCS for Recognition VS Rejection. MCS with different combination methods are often used in pattern recognition to enhance the recognition rate. In handwritten numeral recognition, some researchers have yielded state-of-the-art performance in recognition based on differently designed MCSs. On the MNIST numeral dataset, a recognition rate of 99.73% is achieved with an MCS consisting of 35 classifiers [9]; Wu et al obtained an even better recognition rate of 99.77% based on a MCS with 5 CNNs based on different training sets as well as different operations of spatial pooling [10].

Although MCS has contributed a lot to recognition, it is seldom used for pattern rejection. As it is so effective in recognition, it is assumed to be useful in rejection as well. Therefore, we attempt to adopt MCS to the rejection problem. In [14, 15], the authors apply MCS for rejection based on the cascading methods and achieve high performances. In this paper, a committee approach for MCS rejection is used.

Voting Based Combination Method for MCS Rejection. For the purpose of combining multiple classifiers, voting is always a good choice for the reason that it is simple and effective. Hard voting is the simplest voting method which assigns equal weight to all votes. Another frequently used method is soft voting, which assigns a weight to each classifier according to its performance [16, 17]. For the weights part, all the rejection criteria mentioned in Section 2 can be selected for the reason that they reflect the rejection performance of a single classifier. A certain type of rejection criterion is assigned to each model in the voting procedure, and the class label with the highest voting value provides the final decision for each sample.

Suppose there are N different classifiers in the MCS, denoted as $g_1, g_2, \dots, g_N$, for a random pattern, each classifier $g_i (i = 1,2, \dots, N)$ would provide a prediction of the label y_i as well as an output vector $\{f_1^i, f_2^i, \dots, f_c^i\}$. Then for each classifier, the selected rejection criterion (FRM, FTRM or SVMM) can be calculated based on the output vector $\{f_1^i, f_2^i, \dots, f_c^i\}$, denoted as $t_i (i = 1,2, \dots, N)$. (For the reason that LDAM does not work as effectively as the other criteria, it is not considered for combination.) The above-mentioned method is the *soft voting*. We also consider the *hard voting* method by simply setting $t_i = 1$. After that, a voting value $V_j (j = 1,2, \dots, c)$ is calculated for each class denoted as:

$$V_j = \sum_{i=1}^{N} t_i I(y_i, j), \quad I(y_i, j) = \begin{cases} 1 & if\ y_i = j \\ 0 & otherwise \end{cases} \tag{7}$$

Among V_j, a maximum voting value $V_{max} = \max_{1 \le j \le c} V_j$ can be found and a threshold T_{com} is searched and determined. A pattern is rejected if V_{max} is smaller than a threshold. As the voting values are sums of all models, the thresholds T_{com} can be any real numbers between 0 and N. But for the hard voting method, the threshold can only be an integer which cannot yield a reliability-rejection curve. The whole procedure of MCS based pattern rejection is shown in Fig. 1.

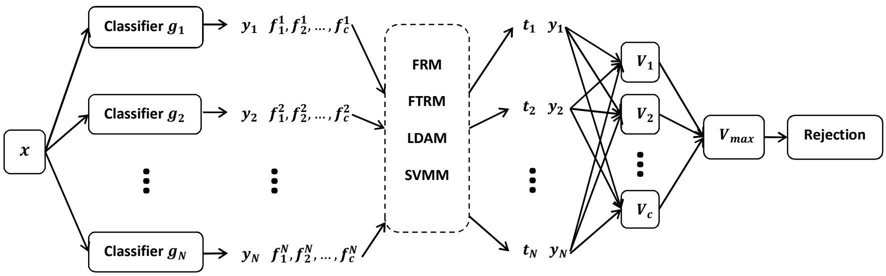

Fig. 1. Flow chart of voting based combination of MCS for pattern rejection

4 Experiments

4.1 Multiple Classifier System

Two well-known datasets are selected for these experiments including CENPARMI [18] and MNIST [19] handwritten numeral datasets. The former contains 4000 training samples and 2000 testing samples with no-fixed size while the latter contains 60000 training samples and 10000 testing samples with identical size of 28 by 28 pixels.

Firstly, structural modification (SM) method [20] is conducted to build committees. For the CENPARMI data, we increase the numbers of feature maps in each convolution layer (C1, C3 and C5) of the basic model and train all the models to 150th epoch as shown in Table 1. For the MNIST data, these numbers are slightly changed in both increasing and decreasing directions as listed in Table 2 below. Secondly, dataset re-sampling (DR) method is used on CENPARMI data. In this phase, model structure is fixed as the basic one. Different training sets are formed by randomly selecting 2000 training samples and distorting them with elastic algorithm [3]. The process is repeated four times to obtain 4 different training sets (G1-G4) with 4000 samples each, as listed in Table 3. The numbers in the first 10 columns represent the numbers of samples selected in different categories for different training sets.

Table 1. Information about modified structures in MCS with CENPARMI dataset

	M0 (basis)	M1	M2	M3
C1	25	50	50	70
C3	50	75	90	75
C5	100	120	100	100
Training Error Rate (%)	0.5	0.38	0.38	0.43
Testing Error Rate (%)	2.45	2.45	2.25	2.45

Table 2. Information about modified structures in MCS with MNIST dataset

	M0	M1	M2	M3	M4	M5	M6
C1	25	25	25	25	25	10	40
C3	50	50	50	30	80	50	50
C5	100	80	120	100	100	100	100
Training Error Rate (%)	0.36	0.34	0.31	0.34	0.26	0.34	0.29
Testing Error Rate (%)	0.62	0.63	0.61	0.6	0.58	0.63	0.61

Table 3. Information about re-sampling training sets with CENPARMI data

	0	1	2	3	4	5	6	7	8	9	Training Error Rate (%)	Testing Error Rate (%)
G1	474	462	416	350	332	394	380	370	400	422	1.65	2.80
G2	450	408	358	404	394	382	424	424	396	360	1.52	3.65
G3	458	482	408	340	372	410	392	426	386	326	1.27	3.50
G4	402	440	380	390	430	426	370	412	350	400	1.77	3.45

4.2 Comparison of Different Rejection Criteria

In the selected CNN model, the output of each sample is a 10-dimention vector consisting of confidence values for possible classes. FRM, FTRM and LDAM are used respectively as rejection criteria with the basic model. Thresholds are searched incrementally. As in CNN model, the outputs are confidence values instead of probabilities, the most appropriate starting point, step and ending point for thresholds searching vary according to different rejection criteria. For the newly proposed SVMM, "libsvm" tools are applied and the same CNN model is used as a feature extractor. Totally, there are 216 out of 60000 samples labeled "-1" while the rest are labeled "1" for the training process. Since the training set is so unbalanced with the number of samples in class "1" almost 300 times that of class "-1", the weight parameter is set to "400" for class "-1". A linear kernel is selected in order to find a linear decision boundary in the feature space. Normalization is conducted on the decision value with SVM of each sample on purpose of making the threshold-setting procedure more convenient. Then different thresholds are set for rejection. All the results are shown by the curves presenting the relationship between the number of rejected samples and reliability in Fig. 2.

Results show that, although LDAM is proved to have a better performance than FRM and FTRM in [7] based on eight-direction gradient feature with an SVM classifier, it is the least useful one in our experiment with the CNN model. The performances of FRM and FTRM which are far different in [7] are insignificantly different in CNN model "M0". So it can be concluded that these pre-defined criteria vary in performance with different classifier models or types of features.

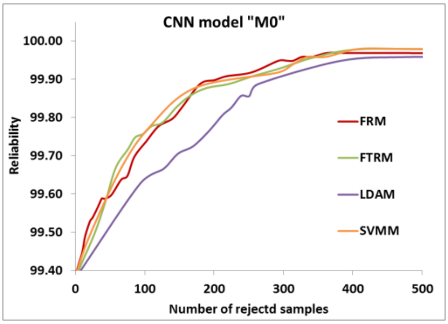

Fig. 2. Relationship between number of rejected sample and reliability in "M0"

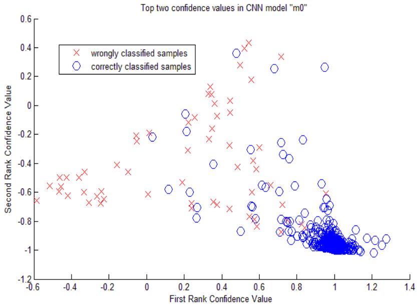

Fig. 3. Samples in FR-SR feature space

From Fig. 3, FR and SR of correctly classified samples are extremely close to 1 and -1 respectively. As a result, a line with slope "1" standing for FTRM is an optimal boundary to separate wrongly and correctly classified samples. That is why FTRM is an effective criterion for rejection. Another effective criterion FRM can also be viewed as a problem of finding a boundary parallel to the y-axis in Fig. 3, which, by observation, is less effective than FTRM. However, it is noticed that although these two criteria can be useful, many correctly classified samples will also be rejected by them no matter where the boundary is.

It is also shown in Fig. 2 that SVMM works as effective as FTRM in rejection and the two are always the relatively best ones among all of the criteria. Similar results appear when we applied these criteria to all the modified CNN models, as displayed in Fig. 4. Besides, it is noticed that the performances of FTRM and SVMM are too close to determine which one is better. The reason for this can be traced back to the training process of CNN model when the expected values in the decision layer are set to be "1" for the true class and "-1" for the other classes. Hence, FTRM is already a distinctively effective criterion to determine the quality of a sample as analyzed with Fig. 3. When we use the SVMM, which uses all the values of the output vector, FR and SR contribute much more than the others since the others are slightly different from SR. Therefore, the rejection boundary of SVMM is very close to that of FTRM. This explains the similar performances of these two criteria.

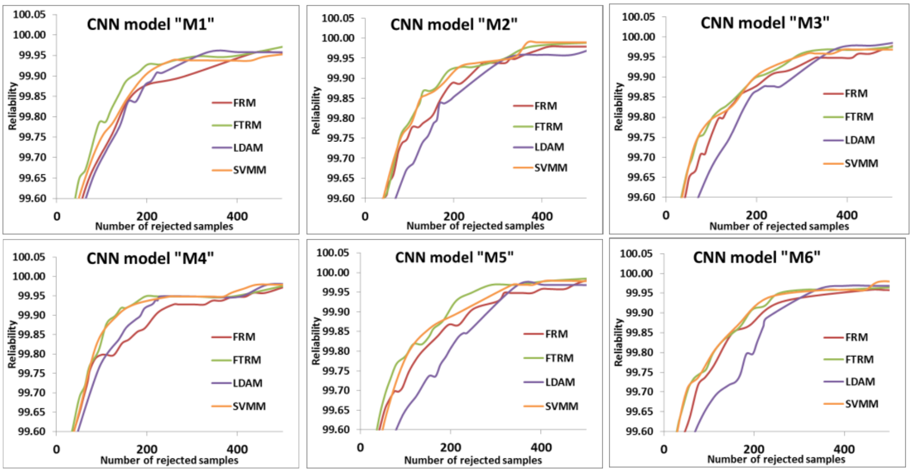

Fig. 4. Relationship between number of rejected sample and reliability in modified models

4.3 Pattern Rejection with MCS

Voting Based Combination Method. In this experiment, hard voting and soft voting rejection methods are both conducted based on the MNIST dataset with MCS built by SM.

In hard voting, a range limitation problem makes the rejection process inflexible for the reason that the thresholds can only be set to several integers. Once the maximum value (number of classifiers in the MCS) is reached, the reliability cannot be improved anymore. The highest reliability is 99.86% with 118 samples rejected when the threshold is set to "7".

In soft voting, the proposed combination method has been applied with FRM, FTRM and SVMM respectively. Since these criteria have different value ranges, different starting points, search steps and ending points are chosen. For FRM and SVMM, the starting and ending points are 0 and 1 respectively; while for FTRM, the starting and ending points are 0 and 2. The search steps for all of them are 0.1 at regular places and 0.01 at the sections where the number of rejected samples changes sharply based on different criteria. The results are shown in Fig. 5. We can find that with the combination of seven CNN models, the rejection performances are consistently improved for all rejection criteria (FTM, FTRM and SVMM).

Structural Modification (SM) and Data Re-sampling (DR). In this section, we adopt the soft voting combination rejection method with MCS on the CENPARMI handwritten numeral dataset. The MCS is constructed in two different ways including SM and DR, as presented in Section 4.1. FTRM is chosen as weight for soft voting combination and thresholds are searched from 0 with an incremental step of 0.05 until suitable reliability values are reached. The results are shown as curves displaying the relationship between number of rejected samples and reliability, presented in Fig. 6.

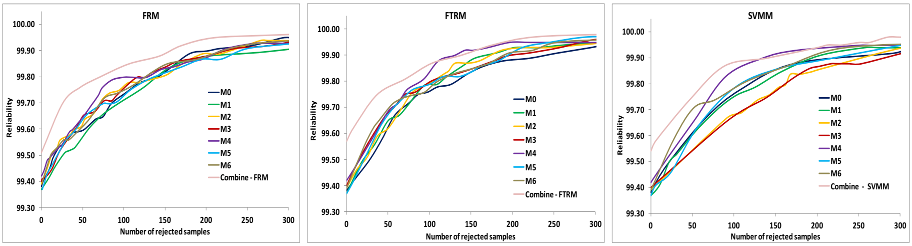

Fig. 5. Relationship between number of rejected sample and reliability with MCS and single models based on different rejection criteria

From these two figures, it is proved again that soft-voting combination method with MCS could improve the rejection performance of the system no matter which method is adopted to construct the MCS. Furthermore, it is shown in Fig. 6 that with our combination method, although MCS does not necessarily improve the recognition rate (without rejection), it can still improve the rejection performance of the whole system.

Table 4 below lists some important information about the performance of different rejection methods based on the CENPARMI dataset. In [7], it is claimed that using LDAM, a reliability of 99.67% is achieved with 175 samples rejected. With our combination methods, the MCS with SM (Com-SM) obtains a reliability of 99.78% with only 164 samples rejected and 99.89% with 180 rejected. The other MCS with DR (Com-DR) achieves the same reliability as LDAM with 6 less samples rejected and 99.73% with 179 samples rejected. Both of these two construction methods with MCS obtain better rejection results than state-of-the-art rejection method based on the same dataset.

Comparing two different construction methods of MCS (SM and DR), it is clear that the system with DR performs better than that with SM. As shown in Table 4, to reach a reliability of 99.94%, DR should reject 257 samples while SM should reject 393 samples, even if the original recognition rate (without rejection) of DR is smaller than that of SM (see Table 1 and 3). This indicates that building MCS with DR makes errors between different classifiers in the system much more diverse.

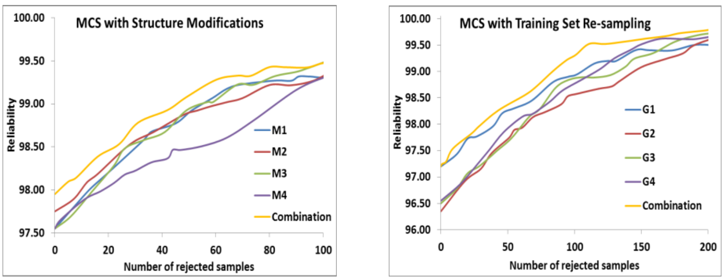

Fig. 6. Relationship between number of rejected sample and reliability with MCS built by different methods

Table 4. Rejection performances of different rejection methods based on CENPARMI dataset

Number of rejected samples	Reliability	Method
175	99.67%	[7]
164	99.78%	Com-SM
180	99.89%	Com-SM
169	99.67%	Com-DR
179	99.73%	Com-DR
393	99.94%	Com-SM
257	99.94%	Com-DR

5 Conclusion

In this paper, a novel SVM-based rejection measurement and voting based combination methods with multiple classifier system (MCS) for rejection are proposed. The main difference between SVMM and other criteria (FRM, FTRM, LDAM and so forth) is that SVMM finds the rejection boundary based on the training data rather than experiences as in those pre-defined criteria. The voting based combination method of MCS is a new attempt to adopt MCS for the purpose of rejection. In the soft voting method, different rejection criteria (FRM, FTRM and SVMM) are used as weights for different models since they reflect their rejection effectiveness. Experiments are conducted on well-known MNIST and CENPARMI digit datasets. Different MCSs are constructed with two different building methods, structural modification and dataset re-sampling. The results show that no matter what building method is chosen or what criterion is selected as weight in soft voting, rejection based on MCS can improve the rejection performance of the system consistently. It is also indicated that MCS built by dataset re-sampling works better than that by structural modification in rejection.

References

1. Lauer, F., Suen, C.Y., Bloch, G.: A trainable feature extractor for handwritten digit recognition. Pattern Recognition 40(6), 1816–1824 (2007)
2. Zhang, P., Bui, T.D., Suen, C.Y.: A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition 40(12), 3415–3429 (2007)
3. Simard, P.Y., Steinkraus, D., Platt, J.C.: Best practice for convolutional neural networks applied to visual document analysis. In: Int'l Conf. on Document Analysis and Recognition, pp. 958–962 (2003)
4. Niu, X.X., Suen, C.Y.: A novel hybrid CNN-SVM classifier for recognition handwritten digits. Pattern Recognition 45(4), 1318–1325 (2012)
5. Liu, C.L., Nakashima, K., Sako, H., Fujisawa, H.: Handwritten digit recognition: investigation of normalization and feature extractor techniques. Pattern Recognition 37(2), 265–279 (2004)

6. He, C.L., Lam, L., Suen, C.Y.: A novel rejection measurement in handwritten numeral recognition on linear discriminant analysis. In: Int'l Conf. on Document Analysis and Recognition, pp. 451–455 (2009)
7. He, C.L., Lam, L., Suen, C.Y.: Rejection measurement based on linear discriminant analysis for document recognition. Int'l J. of Document Analysis and Recognition 14(3), 263–272 (2011)
8. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machine. Journal ACM Transactions on Intelligent Systems and Technology 2(3), 27:1–27:27 (2011)
9. Ciresan, D.C., Meier, U., Gambardella, L.M., Schmidhuber, J.: Convolutional neural network committees for handwritten character classification. In: Int'l Conf. on Document Analysis and Recognition, pp. 1135–1139 (2011)
10. Wu, C., Fan, W., He, Y., Sun, J., Naoi, S.: Cascaded heterogeneous convolution neural networks for handwritten digit recognition. In: Int'l Conf. on Pattern Recognition, pp. 657–660 (2012)
11. Breiman, L.: Bagging predictors. Machine Learning 24(2), 123–140 (1996)
12. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Int'l Conf. Machine Learning, pp. 148–156 (1996)
13. Dietterich, T.G.: Ensemble methods in machine learning. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 1–15. Springer, Heidelberg (2000)
14. Zhang, P., Bui, T.D., Suen, C.Y.: A novel cascade ensemble classifier system with a high recognition performance on handwritten digits. Pattern Recognition 40(12), 3415–3429 (2007)
15. He, C.L., Suen, C.Y.: A hybrid multiple classifier system of unconstrained handwritten numeral recognition. Pattern Recognition and Image Analysis 17(4), 608–611 (2007)
16. Suen, C.Y., Lam, L.: Multiple classifier combination methodologies for different output levels. In: Kittler, J., Roli, F. (eds.) MCS 2000. LNCS, vol. 1857, pp. 52–66. Springer, Heidelberg (2000)
17. Lam, L., Suen, C.Y.: Optimal combinations of pattern classifiers. Pattern Recognition Letters 16(9), 945–954 (1995)
18. Suen, C.Y., et al.: Computer recognition of unconstrained handwritten numerals. Proc. IEEE 80(7), 1162–1180 (1992)
19. LeCun, Y.: The MNIST database of handwritten digits, `http://yann.lecun.com/exdb/mnist`
20. Yuan, A., Bai, G., Jiao, L., Liu, Y.: Offline handwritten English character recognition based on convolutional neural network. In: Int'l Workshop Document Analysis Systems, pp. 125–129 (2012)