Combined mRMR-MLPSVM Scheme for High Accuracy and Low Cost Handwritten Digits Recognition

Mohammad Hassan Shammakhi
Electrical Engineering Department
Amirkabir University of Technology
Email: mh.shammakhi@aut.ac.ir

Ali Mirzaei
Electrical Engineering Department
Amirkabir University of Technology
Email: ali_mirzaei@aut.ac.ir

Parviz Khavari
Electrical Engineering Department
Amirkabir University of Technology
Email: parvizkhavari@aut.ac.ir

Vahid Pourahmadi
Electrical Engineering Department
Amirkabir University of Technology
Email: v.pourahmadi@aut.ac.ir

Abstract—**this paper presents a novel algorithm for handwritten digit recognition which in addition to its high accuracy, enjoys low implementation complexity. The proposed algorithm sorts all features using mRMR (Minimal-Redundancy and Maximal-Relevance) method and selects the best top features by evaluating the training data. The selected features are then used as input of our classifier. The classifier that we used is an MLPSVM classifier which combines the good properties of MLP (Multi-Layer Perceptron) and SVM (Support Vector Machines). The performance of the proposed scheme is then evaluated against ORHD and MNIST datasets, which shows that despite lower complexity compared to existing methods, it can get to high accuracy of 96.1 and 98.14 on the datasets respectively.**

Keywords— ***handwritten digit recognition; mRMR; SVM; MLP; ORHD; MNIST;***

I. Introduction

Handwritten digit recognition is an important problem in computer vision, and many studies have already been conducted in this field. Their main purpose is to reduce classification error, and in some cases they try to decrease computational complexity.

One of the most accurate approaches for recognition are convolutional network based methods [1] [2]. Despite the high accuracy of these methods, they are not practical in most situations, because their complexity are high, and they are time consuming in both train and test stages. In several existing digit recognition schemes, all features of raw images are used for classification. The most prominent disadvantage of these methods is complexity of the learned classifier. Because in high resolution images, using all features results a high dimensional data, and this issue leads to a complex model. Furthermore, when complexity increases, the generalization ability of trained model will decrease.

This paper presents a novel algorithm with high accuracy which also takes advantage of using limited number of features for classification.

The structure of the paper is as follow. First in section II, we propose mRMR as our feature selection method that will be applied on raw images. After exploring how to select the finite features to reduce computational cost, in third section, MPLSVM which is a joint MLP and SVM technique, will be reviewed. In the fourth section, we will present our scheme which combines the benefits of mRMR and MLPSVM. In the fifth section, we will evaluate the performance of the proposed approach, and finally section VI concludes the paper.

II. mRMR Feature Selection

In classification problems, we aim to find the best subset of features to choose appropriate model and minimize classification error. In this paper, to reduce complexity, we propose a technique called mRMR to select a number of top features, and then to perform classification [3]. The main purpose of mRMR as a feature selection method is to acquire features which are the most relevant subset of features between the features and the output class. Clearly, if m equals 1, the selected subset is the best feature, and for $m > 1$ it indicates the best subset of features.

In this paper with respect to our input data (which are raw images of handwritten digits), we have considered the pixels as features, and we will try to select best subset of these features and use them to recognize the handwritten digits.

In fact, mRMR is using mutual information between output class and our selected subset of features as a criterion to obtain maximum dependency. It means we are searching for subset of pixels with maximum dependency with output class:

$$\max D(S,c)\ ,\ D = I(\{x_i, 1, \dots, m\}; c), \qquad (1)$$

Where for any random variable x and y mutual information is defined as:

$$I(x;y) = p(x,y)\log \iint \left(\frac{p(x,y)}{p(x)p(y)}\right) dx\ dy. \qquad (2)$$

Using (2), the mutual information term in (1) can be written as:

$$I(S_m;c) = p(S_m,c) \iint log\left(\frac{p(S_m,c)}{p(S_m)p(c)}\right) dS_m dc$$

$$= \iint p(S_{m-1}, x_m, c) \log\left(\frac{p(S_{m-1}, x_m, c)}{p(S_{m-1}, x_m) p(c)}\right) dS_{m-1} dx_m dc$$

$$= \int \dots \int p(x_1, x_2, \dots, x_m, c) \log\left(\frac{p(x_1, x_2, \dots, x_m, c)}{p(x_1, x_2, \dots, x_m) p(c)}\right) dx_1 dx_2 \dots dx_m dc \quad (3)$$

where S_m means the subset that mRMR has selected with max-dependency between features and output class and min-redundant among S_m selected subset.

It is often difficult to achieve joint probability function for two main reasons: firstly, there is inadequate number of samples (N) to compute joint probability function, and secondly, it is a complicated task to compute the joint probability function which needs inverse of high-dimensional covariance matrix. Furthermore, assume k features are selected, and for each feature there are a states, k features could have a maximum $\min(a^k, N)$ joint states.

Due to these complexities, [3] redefines the max-dependency as:

$$\max D(S, c) \ , \qquad D = \frac{1}{|S|} \sum_{i=1}^{m} I(x_i, c) \qquad (4)$$

While we want to maximize $D(S, c)$, we have to consider searching for a subset of pixels with low mutual information along each two features in the subset. Therefore, we define redundancy:

$$\min R(S) \quad , \quad R = \frac{1}{|S|^2} \sum_{x_i, x_j \in S} I(x_i, x_j) \qquad (5)$$

mRMR selection criterion is $\Phi = D - R$, and we concentrated on maximizing it as:

$$\max \Phi(D, R), \Phi = D - R \qquad (6)$$

As shown in [3], with (6) we will get the best subset which provides max-relevance and minimum-redundant.

To achieve this goal, suppose that we have selected the best m-1 subset of pixels with respect to Φ as a criterion. It has been proved to create best subset of m *principal pixels*, we can find m^{th} pixel by scanning all features except those m-1 features we have selected in our subset and select the one which maximizes the following expression:

$$\max_{x_j \in X - S_{m-1}} \left[I\left(x_j; c\right) - \frac{1}{m-1} \sum_{x_i \in S_{m-1}} I\left(x_j, x_i\right) \right] \qquad (7)$$

It is possible to apply this method with different wrappers. In this paper, forward-selection is used. This leads to series of subset of features with following property according to different m:

$$S_1 \subset S_2 \subset S_3 \subset \cdots \subset S_n \qquad (8)$$

III. MLPSVM CLASSIFIER AND IMPLEMENTATION

One of the most important drawback of handwritten digit recognition methods is incorrect recognition when they are facing poor quality images, and most of their common mistakes are detecting valid label along similar pair such as (1,7) , (6,8).

In this paper, we try to resolve this problem by application of MLPSVM [6]. More accurately, we first train MLP(Multi-Layer Perceptron [4]) with input nodes corresponds to input pixels, a hidden layer with 50 nodes, and 10 output nodes which represent output labels from (0-9). After training MLP network, to resolve misclassification we choose two highest-score, and if their difference was more than the determined threshold (in our experiment we have used 0.4), we select the highest score's label as our test label confidently, but if the difference is less than a threshold, then we apply SVM classification [5] for two highest-score classes. This combined approach as discussed in [6], will select appropriate label with higher accuracy, and it will reduce the classification error significantly.

IV. COMBINED MLPSVM-MRMR METHOD

After reviewing the efficiency of MLPSVM on handwritten digits, we explore a combinational method to select features (pixels) with mRMR and incorporate it with MLP.

Therefore, the algorithm is as the following. Firstly, by using mRMR criterion we select a subset of features, and then with respect to selected subset of features, we train neural network based on MLPSVM method and evaluate error rate on our test set. If the error rate is less than defined threshold, then we will stop feed-forward and choose those pixels subset as our subset of features. Nonetheless, if the error rate is more than the threshold, we use mRMR criterion and increase number of selected features (pixels) in our subset to reduce the error rate. It means we have stopped algorithm on training set when we reached defined error rate, and we will not use all features (all pixels) which may cause over-fitting. With this scheme, we have prevented computational complexity while achieving appropriate error rate in relatively short time.

V. PERFORMANCE EVALUATION

We will apply our proposed method on two datasets with completely different features.

A. ORHD data set

The first dataset, ORHD, contains 5620 handwritten digits image, of size 8 * 8 pixels. We have used 3976 of samples for training and 1797 samples of handwritten digits as test-set in our model. Some of these samples from (0-9) are illustrated below [7]:

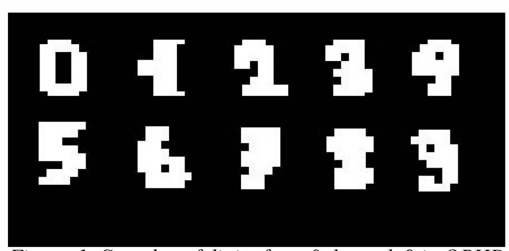

Figure 1: Samples of digits from 0 through 9 in ORHD

To implement our approach on first data set, ORHD, we turn grayscale value of each pixel which is between 0-15 into logical value - 1 or 0. If the value of gray scale of pixel is more than *8*, we map pixels to 1, and if it is less than 8, we map pixel to 0.

Afterward, by using mRMR criterion, we have selected principal pixel with defined threshold error rate equals to 2%, mRMR only has selected 38 pixels out of 64 pixels. To clarify we illustrate error rate and size of selected subset of feature according to mRMR, form 1 through 64 in figure 2. We have used this 38 pixels as input in MLPSVM network. We achieved 96.1% accuracy on our test set which in comparison with other methods in [4] such as EWLDR (93.88%) WLDR (93.93%) ، ePAC (93.88%) and LDA (93.88%) shows a considerable improvement.

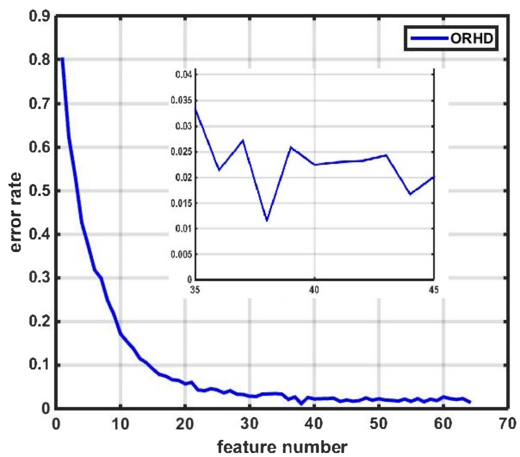

Figure 2: error rate on training data

Note that this error rate is achieved when we only used 38 pixels of 64 pixels, meanwhile previous approaches had used all of pixels and feature extractions such as LDA. In addition, to handle images with poor quality and high error rate, we have used SVM classification after MLP network on each similar pair of digits. This approach has reduced the error rate to 4% which is about half of error rate in other methods.

B. MNIST DATASET

Second dataset, MNIST, is one of the most useful datasets in handwritten digit recognition which includes more than 7000 samples size of 28*28 pixels for handwritten digits. We have used 60000 of them to train and 10000 samples of handwritten digits to test our model [9]. In this dataset, therefore, the proposed feature selection step can potentially decrease the complexity significantly. Some of these samples from (0-9) are illustrated below:

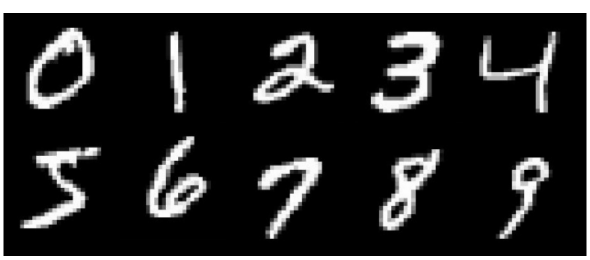

Figure 3: random digits from 0 through 9 in MNIST dataset

If we wanted to use all 60000 images of size 784 pixels for training our model, it will need to compute 5 million data which means heavy computational cost. It is also very time consuming, and therefore it is not perfect for real time applications. Training on MNIST dataset lasts 16 times more than ORHD dataset. Therefore, it is necessary not to use all the pixels to reduce computational cost, and with our proposed approach we have used only half of the pixels although we have achieved to very low error rates.

Table 1 illustrates the performance of proposed method with different selections of error rate threshold (error rate threshold has been defined in Section IV where we have described our approach):

Table 1: test error rate of combined mRMR-MLPSVM approach on MNIST dataset

Error rate threshold (%)	Selected pixels	Test error
1	472	2.83
0.8	487	2.56
0.5	524	**1.86**

To see the performance, we can compare the results with [9]. For example, one of the methods that is used in [10] is a neural network which all 784 pixels were used to recognize the number. The best error rate that was reported there, was 1.4 percent although their model has higher complexity due to higher number of features. There are also methods like [11] which achieve 0.8% error rate by using 12 hidden layer and 100 units in each one. However, in this paper we have used only 524 pixels out of 784 pixels to recognize the number, thus it has much less complexity than [11], and in the meantime we have achieved acceptable error rates.

VI. CONCLUSION

In this paper a novel low complexity method with high accuracy of handwritten digits recognition is presented. The paper has proposed firstly, a feature selection step for dimensional reduction, and secondly, used a classification method by combination of MLP and SVM. The algorithm is designed to use SVM only when the quality of image is poor and this improves performance and pace of algorithm. The proposed scheme is also evaluated against MNIST and OHDR datasets which leads to 96.1 and 98.14 accuracy respectively.

VII. REFRENCES

[1] Ciresan, Dan, Ueli Meier, and Jürgen Schmidhuber. "Multi-column deep neural networks for image classification." Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012

[2] Poultney, Christopher, Sumit Chopra, and Yann L. Cun. "Efficient learning of sparse representations with an energy-

based model." Advances in neural information processing systems. 2006.
[3] Hanchuan Peng et al.," Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-RelevanCe, and Min-Redundancy" IEEE Trans. Pattern Anal. Mach. Intell, vol. 27, NO.8, pp. 1126-1238, 2005
[4] Bishop CM, Pattern Recognition and Machine Learning, Springer, USA, 2006
[5] CORINNA CORTES, VLADIMIR VAPNIK," Support-Vector Networks" Machine Learning, pp. 273-297, 1995
[6] A. Bellili et al.,"An MLP-SVM combination architecture for offline handwritten digit recognition "international journal on document and recognition(IJDAR),pp. 244-252, 2003
[7] "Optical Recognition of Handwritten Digits Data Set". Internet: https://archive.ics.uci.edu/ml/datasets/Optical+Recognition+of+Handwritten+Digits , [Aug.7, 2015]
[8] E.K. Tang et al., "Linear dimensionality reduction using relevance weighted LDA" Pattern Recognition, pp. 485-493, 2005
[9] Yann LeCun et al., "THE MNIST DATABASE of handwritten digits". Internet: http://yann.lecun.com/exdb/mnist , [Aug.7, 2015]
[10] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner: Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE, 86(11):2278-2324, November 1998
[11] Ernst Kussul, Tatiana Baidyk. "Improved method of handwritten digit recognition tested on MNIST database", Image and Vision Computing , pp. 971–981, 2004