2016 International Conference on Information System and Artificial Intelligence

# A New Sparse Representation Classifier (SRC) Based on Probability Judgement Rule

Jiang-shu WEI
Machine Intelligence Laboratory, College of Computer Science
Sichuan University
Chengdu, P. R. China
weijiangshu66@163.com

Jian-cheng LV*(correspongding author)
Machine Intelligence Laboratory, College of Computer Science
Sichuan University
Chengdu, P. R. China
lvjiancheng@scu.edu.cn

Chun-zhi XIE
Machine Intelligence Laboratory, College of Computer Science
Sichuan University
Chengdu, P. R. China
xcz_xihua@sina.com

***Abstract*—Sparse representation classifier (SRC) is a classical method for classification, which was proposed in 2009. SRC method boosted the research of sparse representation, based on SRC, many other new methods were also proposed, such as structured-sparse representation, collaborative representation optimized classifier (CROC), deformable sparse recovery and classification (DSRC) method, misalignment robust representation (MRR) method. In recent years, many classification application problems were solved by using above methods. Although SRC is effective for handling with the classification applications, it also has some disadvantages. The classification criterion of SRC is from the residuals using truncation way. If the *i*th residual is the smallest, then the SRC judges that the test sample belongs to the *i*th class. However, from the truncation way, some test samples are also misclassified by SRC. This paper presents a new probability judgement rule to sparse representation classifier (SRC), which using a probability judgement model instead of using the truncation residual directly. The extensive experimental results show that the proposed new method can achieve higher classification accuracy rates than SRC using truncation residual directly.**

***Keywords-SRC; Probability judgement; Truncation residual; Classification***

## I. INTRODUCTION

In recent years, there has been an increasing interest in sparse representation. Sparse representation classifier (SRC) is a widespread concern method [1].

Based on SRC, many other new methods were also proposed, Elhamifar et al. proposed a structured-sparse representation method. They casted the face recognition problem as a structured sparse recovery problem [2], [3]. Chi et al. proposed a collaborative representation optimized classifier (CROC) [4], Zhang et al. proposed a collaborative representation classification method by using $l_2$ norm minimization [5]. Wagner et al. proposed a deformable sparse recovery and classification (DSRC) method [6]. Yang et al. proposed a misalignment robust representation (MRR) method [7]. A large number of application problems were solved by using these methods [8], [9], [10], [11], [12], [13]. Thus, sparse representation boosted the research of representation learning in recent years [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28].

Although SRC is effective for solving the classification problems, there are also some disadvantages about SRC, a very important thing is that the classification criterion of SRC is that using the truncation residuals. If the *i*th residual is the smallest, then SRC judges that the test sample belongs to the *i*th class. From the truncation residuals, some test samples are misclassified by SRC. However, the classification accuracy rates of SRC can be further improved using other methods instead of using truncation residuals directly.

In this paper, based on probability theory, a new probability judgement rule is proposed. Depend on the probability judgement rule, the classification accuracy rates of SRC can be further improved.

This paper is organized as follows. In Section 2, we review the sparse representation classification (SRC) method. In Section 3, a new SRC based on probability judgement rule is proposed. Some experimental results are presented in Section 4. Section 5concludes this paper.

## II. MOTIVATION

Assume there are $k$ classes, for each class, there are $n_i$ training samples $\{\boldsymbol{a}_{ij} \in R^m\}_{j=1}^{n_i}$ . $\boldsymbol{A_i} \in R^{m\times n_i}$ is denoted by the collection of training sample in the *i*th class:

978-1-5090-1585-6/16 $31.00 © 2016 IEEE
DOI 10.1109/ISAI.2016.55

$$A_i = [a_{i1}, a_{i2}, \dots, a_{in_i}] \in R^{m \times n_i}.$$

$\boldsymbol{A}$ is denoted by concatenating all training samples:

$$\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k] \in R^{m \times n},$$

where $\mathrm{n} = \sum_{i=1}^{k} n_i$ is the total number of all training samples. Given a test sample $\boldsymbol{b} \in R^m$, the aim of the classification problem is to judge the test sample $\boldsymbol{b}$ belongs to which class.

### *A. SRC Method*

With the above classification problem, SRC is effective for solving it. The steps of SRC as shown in algorithm 1 [1]:

---

Algorithm1. sparse representation classifier (SRC)

---

*1)* Input a matrix of training samples $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k] \in R^{m \times n}$ for $k$ classes, a test sample $\boldsymbol{b} \in R^m$.

2) Normalize the columns of $\boldsymbol{A}$ to use unit $l_2$ norm.

3) Solvethe $l_1$ norm minimization problem:

$\boldsymbol{x}_1 = arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 \text{s.t.} \boldsymbol{b} = \boldsymbol{Ax}$

*4)* Compute the residuals $R_i(\boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{A}\delta_i(\boldsymbol{x}_1)\|_2$ for *i=1,…,k.*

*5)* Output: identity$(\boldsymbol{b}) = \mathrm{arg}min_i R_i(\boldsymbol{b})$.

### *B. Disadvantages of SRC*

Although SRC has been proved to be useful for many classification applications, it also has some disadvantages.

*1)* Using $l_1$ norm minimization, it cannot obtain the closed form solution, the computationalcomplexity of SRC is a little high. With the problem of high computationalcomplexity of SRC, some papers have proposed some new methods. Zhang et al. proposeda collaborative representation classification methodby using $l_2$ norm minimization [5], which can obtain closed form solution. Wei et al. proposed a method using distance-based representation with square weights [22].

*2)* The SRC method only looks for the sparsest representation of a test sampleby using $l_1$norm minimization.But the sparsest representation does not meanachievingthe highest recognition rate for classification problems.

## III. The New SRC based on probability judgement rule

In this section, the new probability judgement rule model is proposed. The new probability judgement rule can be combined with SRC for classification. In addition, its performance is analyzed in detail.

### *A. Model of the Probability Judgement Rule*

The probability judgement rule is as follows:

$$p(i) = \frac{e^{-r_i(\boldsymbol{b})}}{\sum_{i=1}^{k} e^{-r_i(\boldsymbol{b})}}, \quad \textit{for i=1,…, k},$$

where $p(\cdot)$denotes the probability that the test sample $\boldsymbol{b}$ belongs to class *i*. *k* denotes the number of the classes. With the test sample $\boldsymbol{b}$, $r_i(\boldsymbol{b})$ denotes the truncation residuals for the *i*th class. From (1), it is very clear that $0 \le \mathrm{p}(i) \le 1, \sum_{i=1}^{k} p(i) = 1$, furthermore, if $r_i(\boldsymbol{b})$is smaller, the probability $p(i)$ is bigger.

From the probability judgement rule, it can substitute the probability value for the truncation residuals. The probability value is a better metric than truncation residuals, which can indicate the probability of one test sample belongs to one class.

### *B. The Probability Judgement Rule Combined with SRC*

The probability judgement rule can be combined with SRC, and then the new SRC method based on probability judgement rule is proposed. The steps of the new method as shown in algorithm 2:

---

Algorithm2. new SRC method based on probability judgement rule

---

*1)* Input a matrix of training samples $\boldsymbol{A} = [\boldsymbol{A}_1, \boldsymbol{A}_2, \dots, \boldsymbol{A}_k] \in R^{m \times n}$for $k$ classes, a test sample$\boldsymbol{b} \in R^m$.

*2)* Normalize the columns of $\boldsymbol{A}$ to use unit $l_2$ norm.

*3)* Solvethe $l_1$ norm minimization problem:

*4)* $x_1 = arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_1 s.t.\boldsymbol{b} = \boldsymbol{Ax}$

*5) C*ompute the residuals$R_i(\boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{A}\delta_i(\boldsymbol{x_1})\|_2$,*for i=1,…,k.*

*6) C*ompute the residuals $r_i(\boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{A}\delta_i(\boldsymbol{x_1})\|_2^2 / \|\delta_i(\boldsymbol{x_1})\|_2^2$ *, for i=1,…,k.*

*7)* Compute the probability $p(i) = \frac{e^{-r_i(\boldsymbol{b})}}{\sum_{i=1}^{k} e^{-r_i(\boldsymbol{b})}}$ ,for *i=1,…, k.*

*8)* If the max$p(i)$ is greater than or equal to 0.99, then output: identity$(\boldsymbol{b}) = argmin_i R_i(\boldsymbol{b})$; if the max$p(i)$ is smaller than 0.99, then the identity$(\boldsymbol{b})$isneeded to be reidentified.

*9)* If the max$p(i)$ is smaller than 0.99, then solve the $l_2$ norm minimization problem:

*10)* $\boldsymbol{x_2} = arg\min_{\boldsymbol{x}} \|\boldsymbol{x}\|_2 \text{s.t.} \boldsymbol{b} = \boldsymbol{Ax}$

*11)* Compute the residuals $r_i(\boldsymbol{b}) = \|\boldsymbol{b} - \boldsymbol{A}\delta_i(\boldsymbol{x_2})\|_2^2 / \|\delta_i(\boldsymbol{x_2})\|_2^2$ , for *i=1,…,k.*

*12)* Output:identity$(\boldsymbol{b}) = argmin_i r_i(\boldsymbol{b})$.

The core idea of the new SRC is the step 5) and step 6), from the step 5), it can transform the truncation residuals to probability $p(i)$, most of the *maxp*($i$) is greater than 0.99. If the *maxp*($i$) is smaller than 0.99, then the classification result may be wrong, thus the classification result is needed to be reidentified.

## IV. Experiments

In this section, some experiments on various databases are presented to show the accuracy of classification. We focus on the performance evaluation of the proposed new SRC based on probability judgement rule and the original SRC. Four databases, including AR, MNIST, USPS and Binary Alphadigits, are used to test the performance of the new SRC and the original SRC.

### *A. Classification on the AR Database*

The AR database contains about 4,000 images corresponding to 126 people's faces (70 men and 56 women) [1], [4], [5], [22]. For each person, 26 images were taken in two separate sessions. These images are captured under different facial expressions, illumination conditions, and occlusions. In our classification experiments: a subset of 50 men and 50 women with only changes of illumination and expressions were used. 700 samples were chosen as training samples (i.e., about 7 images per subject), and 699 samples were chosen as test samples. The dimension of each image is 60×43. Table I and Figure 1 show the recognition rates versus feature dimension by the new SRC and original SRC.

TABLE I. THE RECOGNITION RESULTS OF THE NEW SRC AND ORIGINAL SRC ON THE AR DATABASE

| **Dimension** | **100 (%)** | **150 (%)** | **200(%)** | **250(%)** | **300(%)** | **350(%)** |
|---|---|---|---|---|---|---|
| New SRC | 89.13 | 90.99 | 92.56 | 93.71 | 93.13 | 92.70 |
| Original SRC | 88.70 | 90.70 | 91.56 | 92.13 | 90.99 | 92.56 |

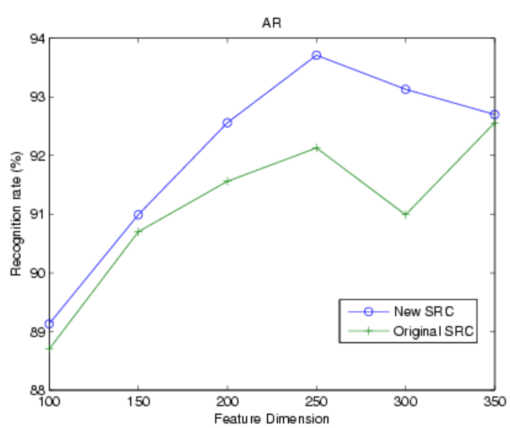


Figure 1. Recognition results on the ar database for different methods.

### *B. Classification on the MNIST Database*

The MNIST database has a training set of 60,000 samples and a test set of 10,000 samples. The dimension of each image is 28×28. Every image is an 8-bit gray scale image of digits "0" through "9" [4], [22]. In our classification experiments: 1,000 samples were randomly chosen as training samples, and 1,000 samples were also randomly chosen as test samples. Table II and Figure 2 show the recognition rates versus feature dimension by the new SRC and original SRC.

TABLE II. THE RECOGNITION RESULTS OF THE NEW SRC AND ORIGINAL SRC ON THE MNIST DATABASE

| **Dimension** | **80 (%)** | **150 (%)** | **200(%)** | **250(%)** | **300(%)** | **350(%)** |
|---|---|---|---|---|---|---|
| New SRC | 87.00 | 86.60 | 84.80 | 84.00 | 82.40 | 81.70 |
| Original SRC | 86.60 | 84.90 | 83.70 | 82.40 | 81.10 | 79.80 |

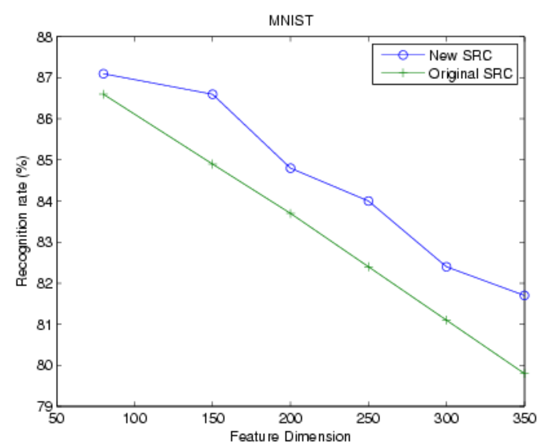


Figure 2. Recognition results on the mnist database for different methods.

### *C. Classification on the USPS Database*

The USPS handwritten digit database has 11,000 samples. The dimension of each sample is 16×16. Every sample, is an 8-bit gray scale image of digits "0" through "9". There are 1,100 samples of each class. In our classification experiments: 1,000 samples were chosen as training samples, there are 100 samples for every class "0"through "9"; and a test set of 1,000 samples was used, there are also 100 samples for every class "0" through "9". Table III and Figure 3 show the recognition rates versus feature dimension by the new SRC and original SRC.

TABLE III. THE RECOGNITION RESULTS OF THE NEW SRC AND ORIGINAL SRC ON THE USPS DATABASE

| **Dimension** | **200 (%)** | **210 (%)** | **220(%)** | **230(%)** | **240 (%)** | **250(%)** |
|---|---|---|---|---|---|---|
| New SRC | 87.00 | 87.50 | 87.00 | 87.30 | 87.20 | 86.50 |
| Original SRC | 86.60 | 86.70 | 85.20 | 85.40 | 85.10 | 84.50 |

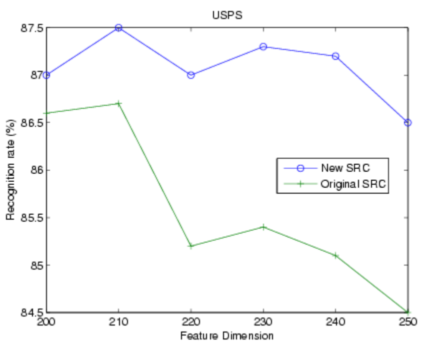


Figure 3. Recognition results on the usps database for different methods.

## D. Classification on the Binary Alphadigits Database

The Binary Alphadigits database contains images of digits of ”0” through “9” and capital letters “A” through “Z”, forming 36 classes. There are 39 samples of each class, and the total number of all the samples is 1,404. The dimension of each sample is $20\times16$. Only 1,014 capital letters were used in our classification experiments: 910capital letters were chosen as the training samples, the other 104 capital letters were chosen as the test samples. Table Ⅳ and Figure 4 show the recognition rates versus feature dimension by the new SRC and original SRC.

TABLE IV. THE RECOGNITION RESULTS OF THE NEW SRC AND ORIGINAL SRC ON THE BINARY DATABASE

| **Dimension** | **30 (%)** | **50 (%)** | **80(%)** | **100(%)** | **200(%)** | **250(%)** |
|---|---|---|---|---|---|---|
| New SRC | 76.92 | 77.88 | 74.04 | 70.19 | 71.15 | 64.42 |
| Original SRC | 74.04 | 75.96 | 72.12 | 68.27 | 67.31 | 59.62 |

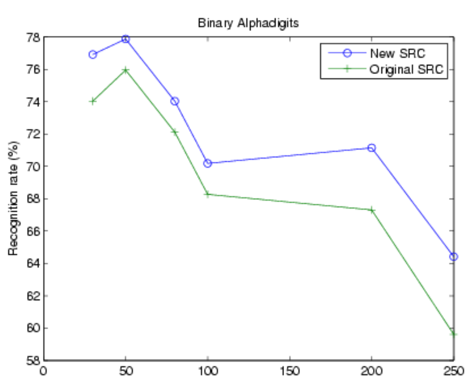


Figure 4. Recognition results on the binary alphadigits database for different methods.

## V. COCLUSIONS

In this paper, a new SRC was proposed, namely the SRC based on probability judgement rule. The original SRC uses the truncation residuals directly for classification. However, some test samples are misclassified by using the truncation residuals from the $l_1$ norm minimization. Unlike the original SRC, the new SRC depends on the probability judgement rule for judging the test sample belongs to which class.

From the classification results obtained by the new SRC clearly show that our new method can obtain very competitive classification results.

## ACKNOWLEDGMENTS

This work was supported by the National Science Foundation of China (Grant No.61375065), partially supported by the State Key Program of National Science Foundation of China (Grant No. 61432012 and 61432014).

## REFERENCES

[1] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry and Y. Ma, “Robust face recognition viasparse representation,”Pattern Analysis and Machine Intelligence, IEEE Transactionson 31 (2), 2009, pp. 210–227.

[2] E. Elhamifar, R. Vidal, “Robust classification using structured sparse representation,”Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conferenceon, IEEE, 2011, pp. 1873–1879.

[3] E. Elhamifar, R. Vidal, “Block-sparse recovery via convex optimization,” SignalProcessing, IEEE Transactions on 60 (8), 2012, pp. 4094–4107.

[4] Y. Chi, F. Porikli, “Connecting the dots in multi-class classification: From nearestsubspace to collaborative representation,” Computer Vision and Pattern Recognition(CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 3602–3609.

[5] L. Zhang, M. Yang and X. Feng, “Sparse representation or collaborative representation:Which helps facerecognition?,”Computer Vision (ICCV), 2011 IEEEInternational Conference on, IEEE, 2011, pp. 471–478.

[6] A.Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi and Y. Ma, “Toward a practicalface recognition system: Robust alignment and illumination by sparse representation,”Pattern Analysis and Machine Intelligence, IEEE Transactions on 34 (2), 2012, pp. 372–386.

[7] M. Yang, L. Zhang and D. Zhang, “Efficient misalignment-robust representation forreal-time face recognition,”Computer Vision–ECCV 2012, Springer, 2012, pp.850–863.

[8] M. Aharon, M. Elad and A. Bruckstein, “K-svd: An algorithm for designing overcompletedictionaries for sparse representation,” Signal Processing, IEEE Transactionson 54 (11) , 2006, pp. 4311–4322.

[9] S. Yan, H. Wang, “Semi-supervised learning by sparse representation,” SDM,SIAM, 2009, pp. 792–801.

[10] P. Sprechmann, G. Sapiro, “Dictionary learning and sparse coding for unsupervisedclustering,”Acoustics Speech and Signal Processing (ICASSP), 2010IEEE International Conference on, IEEE, 2010, pp. 2042–2045.

[11] L. Zhuang, H. Gao, Z. Lin, Y. Ma, X. Zhang and N. Yu, “Non-negative low rankand sparse graph for semi-supervised learning,” Computer Vision and PatternRecognition (CVPR), 2012 IEEE Conference on, IEEE, 2012, pp. 2328–2335.

[12] I. Ramirez, P. Sprechmann and G. Sapiro, “Classification and clustering via dictionarylearning with structured incoherence and shared features,”Computer Visionand Pattern Recognition (CVPR), 2010 IEEE Conference on, IEEE, 2010, pp.3501–3508.

[13] S. F. Cotter, “Recognition of occluded facial expressions using a fusion of localizedsparse representation classifiers,”Digital Signal

Processing Workshopand IEEE Signal Processing EducationWorkshop (DSP/SPE), 2011, pp. 437–442.

[14] S. F. Cotter, “Sparse representation for accurate classification of corrupted and occludedfacial expressions,”Acoustics Speech and Signal Processing (ICASSP),2010 IEEE International Conference on, IEEE, 2010, pp. 838–841.

[15] M. Elad, M. Aharon, Image denoising via sparse and redundant representationsover learned dictionaries, Image Processing, IEEE Transactions on 15 (12) , 2006, pp. 3736–3745.

[16] E. Elhamifar, R. Vidal, “Sparse subspace clustering: Algorithm, theory, and applications,”Pattern Analysis and Machine Intelligence, IEEE Transactions on 35 (11), 2013, pp. 2765–2781.

[17] J. J. Fuchs, “Recovery of exact sparse representations in the presence of boundednoise,” Information Theory, IEEE Transactions on 51 (10), 2005, pp. 3601–3608.

[18] T. Guha, R. K. Ward, “Learning sparserepresentations for human action recognition,”Pattern Analysis and Machine Intelligence, IEEE Transactions on 34 (8), 2012, pp. 1576–1588.

[19] J. Mairal, F. Bach, J. Ponce and G. Sapiro, “Online dictionary learning for sparsecoding,”Proceedings of the 26th annual international conference on machinelearning, ACM, 2009, pp. 689–696.

[20] M. Stojnic, F. Parvaresh and B. Hassibi, “On the reconstruction of block-sparse signalswith an optimal number of measurements,” Signal Processing, IEEE Transactionson 57 (8).

[21] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang and S. Yan, “Sparse representationfor computer vision and pattern recognition,” Proceedings of the IEEE 98 (6), 2010, pp. 1031–1044.

[22] J. Wei, J. C. Lv and Z. Yi, “Robust classifier using distance-based representation withsquare weights,” Soft Computing 19 (2), 2015, pp. 507–515.

[23] S. Zafeiriou, M. Petrou, “Sparse representations for facial expressions recognitionvia $l_1$ optimization,” Computer Vision and Pattern Recognition Workshops(CVPRW), 2010, pp. 32-39.

[24] E. Mohamed, H. A. Khalil, O. F. Hassan, “Image denoising based on sparse representationin a probabilistic framework,”Signal Process. Int. J 8 (3), 2014, pp. 20–42.

[25] L. Xu, J. Li, “Bayesian classification of hyperspectral imagery based on probabilisticsparse representation and markov random field,”Geoscience and RemoteSensing Letters, IEEE 11 (4), 2014, pp. 823–827.

[26] J. C. Lv, K. K. Tan, Z. Yi and S. Huang, “Stability and chaos of a class of learningalgorithms for ica neural networks,” Neural Processing Letters 28 (1), 2008, pp. 35–47.

[27] J. C. Lv, Z. Yi and K. K. Tan, “Global convergence of gha learning algorithm withnonzero-approaching adaptive learning rates,”Neural Networks, IEEE Transactionson 18 (6), 2007, pp. 1557–1571.

[28] J. C. Lv, Z. Yi and Y. Li, “Non-divergence of stochastic discrete time algorithms for PCA neural networks,” Neural Networks and Learning Systems, IEEE Transactionson 26 (2), 2015, pp. 394–399.