



Journal of
**Software
Engineering**

ISSN 1819-4311



Academic
Journals Inc.

www.academicjournals.com

Collaborative Representation Classifier Based on K Nearest Neighbors for Classification

Jiangshu Wei, Xiangjun Qi and Mantao Wang

College of Information and Engineering Technology, Sichuan Agricultural University, Ya'an, 625014, China

Corresponding Author: Jiangshu Wei, College of Information and Engineering Technology, Sichuan Agricultural University, Ya'an, 625014, China

ABSTRACT

The Sparse Representation based Classifier (SRC) is a classical representation method for classification. The solution of SRC is obtained by l_1 norm minimization which cannot obtain the closed form solution. Thus, the computational complexity of SRC is a little high. The Collaborative Representation Classifier (CRC) is another classical method for classification. The solution of CRC is obtained by l_2 norm minimization, from the l_2 norm minimization, it can obtain the closed form solution which makes the computational complexity of CRC is much lower than SRC. Although, CRC is effective for classification, there are also some problems about CRC. Under some conditions, some test samples may be misclassified by CRC. This study proposes a local CRC method which is called KNN-CRC. This method firstly chooses K nearest neighbors of a test sample from all the training samples, then given a test sample, the test sample is represented by these K training samples. The solution of KNN-CRC is obtained by l_2 norm minimization and the size of K is much smaller than the total number of all training samples. Thus, the computational complexity of KNN-CRC is much lower than SRC and CRC. Furthermore, the extensive experiments show that the proposed KNN-CRC can obtain very competitive classification results compared with other methods.

Key words: SRC, KNN-CRC, recognition rate, computational complexity, classification

INTRODUCTION

With the rapid development of computer network technology and information technology, the research of big data is becoming a hot research field recently. Data mining is very important for handling big data application problems. However, the research of data mining includes many aspects. Classification is one of the most important aspects in the field.

Given some training samples from multiple classes, the aim of classification is to assign class labels to a test sample. Classification has been widely used in many fields such as data mining, computer vision, machine learning, pattern recognition, etc (Chi and Porikli, 2012; Roweis and Saul, 2000; Wright *et al.*, 2009; Zhang *et al.*, 2011). There are many conventional methods for handling classification problems such as Linear SVM, Nearest Subspace Classifier, K Nearest Neighbor, etc. Recently, there has been an increasing interest in the research of representation theory. With the rapid development of l_0 and l_1 norm minimization algorithms (Chen and Donoho, 1994; Pati *et al.*, 1993), sparse representation has been applied for solving many data mining problems. These sparse representation methods have been studied in many literatures. Sparse

Representation based Classifier (SRC) is a classical sparse representation among all the representation methods. The SRC was proposed by Wright *et al.* (2009) which is an interesting and effective method for handling the pattern classification problems. A test sample is first sparsely represented by all the training samples and then the classifier computes the residual for every class, if the i th residual is the smallest, the SRC method will judge that the test sample belongs to the i th class. The SRC is a classical representation method which boosts the research of representation theory and applications. Based on SRC, many other representation methods are proposed, a lot of application problems such as face recognition, digit recognition, signal processing etc., are also solved by representation methods.

Although, SRC is effective for handling pattern classification problems (Elhamifar and Vidal, 2011, 2012; Fuchs, 2005; Guha and Ward, 2012; Mairal *et al.*, 2009; Saul *et al.*, 2003; Soltanolkotabi *et al.*, 2014; Stojnic *et al.*, 2009; Tropp, 2004; Wright *et al.*, 2010), there are also some problems about SRC (Chi and Porikli, 2012; Zhang *et al.*, 2011). First, the SRC method only looks for the sparsest representation of a test sample, however, the sparsest representation does not mean obtaining the highest right classification rate. Under some conditions, some test samples are misclassified by SRC. Second, the solution of SRC is obtained by l_1 norm minimization which cannot obtain the closed form solution. Thus, the computational complexity of SRC is a little high.

In view of the advantages and disadvantages of SRC, some study proposed many other representation methods, Chi and Porikli (2012) proposed a Collaborative Representation Optimized Classifier (CROC). The CROC method combines the advantages of Nearest Subspace Classifier (NSC) and Collaborative Representation Classifier (CRC). Elhamifar and Vidal (2011, 2012) proposed a Block-Sparse representation for face recognition. They casted the classification problem as a structured sparse representation problem, they much emphasized the structured property of training samples. Zhang *et al.* (2011) argued that not the sparse representation but the usage of collaborative representation is more important for the success of the SRC. They proposed a kind of Collaborative Representation Classifier (CRC) method, by using l_2 norm minimization, CRC method can obtain closed form solution. Combined the advantages of KNN and SRC, Zhang and Yang (2010) proposed the KNN-SRC method. First, for every test sample, KNN-SRC chooses the K nearest neighbors from all the training samples then KNN-CRC makes these K nearest neighbors as the training samples. Finally, the classifier computes the residual for every class, if the i th residual is the smallest, the classifier will judge that the test sample belongs to the i th class.

These representation methods are all effective for handling classification problems, however, they all have advantages and disadvantages. For the Block-Sparse representation method, the solution of it is also obtained by l_1 norm minimization, the computational complexity of Block-Sparse representation is still a little high. For the CRC method, under some conditions, some test samples are also misclassified by CRC method. For the KNN-SRC method, the solution of KNN-SRC is also obtained by l_1 norm minimization, its computational complexity is also a little high. Furthermore, under some conditions, some test samples are also misclassified by KNN-SRC.

The KNN is a conventional method for classification which is familiar to us. And CRC is another representation method for classification, the solution of CRC is obtained by l_2 norm minimization, thus, the computational complexity of CRC is low. In this study, combined the KNN and CRC, a new classification method is proposed, namely KNN-CRC method. This method combines the advantages of KNN and CRC, KNN-CRC firstly chooses the K nearest neighbors for every test sample as the training samples, then the solution of KNN-CRC is

obtained by l_2 norm minimization, it is a closed form solution and the size of K is much smaller than the total number of all training samples.

Thus, the computational complexity of KNN-CRC is much lower than SRC, KNN-SRC and CRC. Furthermore, the extensive face and digit classification experiments clearly show that the proposed method can obtain very competitive results compared with other methods.

METHODOLGY

Multi-class classification problem: Assume that there are L know classes, for every class, there are n_i training samples $\{b_{ij} \in \mathbb{R}^m\}_{j=1}^{n_i}$ in the i th class which formed a matrix as:

$$B_i \in \mathbb{R}^{m \times n_i} [b_{i1}, b_{i2}, \dots, b_{in_i}]$$

B is denoted by the collection of all training samples:

$$B = [B_1, B_2, \dots, B_L]$$

If given a test sample $y \in \mathbb{R}^m$, the aim of multi-class classification task is to judge the test sample y belongs to which class (Adler *et al.*, 2011; Aharon *et al.*, 2006; Bruckstein *et al.*, 2009; Cheng *et al.*, 2010; Chi and Porikli, 2012; Cotter, 2010; Elad and Aharon, 2006; Wright *et al.*, 2009; Zhang *et al.*, 2011).

K Nearest Neighbors (KNN) classification method: The nearest neighbor classifier was proposed by Hart and Cover for solving the classification problems. It was improved to the K Nearest Neighbors classifier immediately (Zhang and Yang, 2010). Assume that there are K classes, for every class there are n_i training samples $\{b_{ij} \in \mathbb{R}^m\}_{j=1}^{n_i}$ in the i th class:

$$B_i = [b_{i1}, b_{i2}, \dots, b_{in_i}] \in \mathbb{R}^{m \times n_i}$$

constitute the training samples of the i th class. Given a test sample y , it is easy to find its nearest neighbors b_{ir} in every class. The square of Euclidean distance between y and the i th class is obtained as:

$$d_i(y) = \|y - b_{ir}\|_2^2 \quad (1 \leq i \leq L)$$

Suppose the distance between y and the i th class is the minimal distance, then the 1-NN classifier will identify that the test sample y belongs to the i th class. The 1-NN is a conventional and easy method for classification. The K Nearest Neighbor classification (KNN) method is the improvement of the 1-NN algorithm. First, given a test sample y , the KNN classifier chooses the K nearest neighbors between the test sample and all the training samples. Second, assume that there are k_i samples from the i th class, if $k_j = \max_i k_i (1 \leq i \leq L)$, the KNN classifier will identify that the test sample y belongs to the j th class. The KNN is also a conventional method for classification. However, the extensive experiments show that under many conditions, the accurate classification rates of KNN are not high.

Collaborative Representation Classifier (CRC): The Sparsest solution to $y = Bx$ can be obtained by solving the following problem:

$$P_{10}: \min \|x\|_0 \text{ s.t. } y = Bx$$

However, the P_{10} problem is a NP hard problem. Fortunately, Bruckstein *et al.* (2009) proved that P_{10} problem can be substituted by P_{11} problem. From l_1 norm minimization, the sparsest solution of $y = Bx$ can be also obtained. The P_{11} program is as follows:

$$P_{11}: \min \|x\|_1 \text{ s.t. } y = Bx$$

However, from l_1 norm minimization, it cannot obtain the closed form solution. The computational complexity of l_1 norm minimization is a little high. Thus, some study proposed other method by using l_2 norm minimization. Collaborative Representation Classifier (CRC) is proposed by Zhang and Yang (2010) which is a typical method by l_2 norm minimization. The steps of CRC are given in Algorithm 1 (Zhang and Yang, 2010).

Algorithm 1: Step of CRC method for classification

Input: A matrix concatenated by training samples $B = [B_1, B_2, \dots, B_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $y \in \mathbb{R}^m$

Output: Identity $(y) = \arg \min_i r_i(y)$

Input a matrix concatenated by training samples $B = [B_1, B_2, \dots, B_k] \in \mathbb{R}^{m \times n}$ for k classes, a test sample $y \in \mathbb{R}^m$

Solve the l_2 norm minimization problem: $x_2 = \arg \min_x \|x\|_2 \text{ s.t. } y = Bx$

Compute the residuals: $r_i = (y) = \|y - B\delta_i(x_2)\|_2^2$, for $i = 1, 2, \dots, k$

Output: Identity $(y) = \arg \min_i r_i(y)$

From l_2 norm minimization, a closed form solution can be obtained which gives $x = (B^T B + \lambda I)^{-1} B^T y$. Its computational complexity is low. Although, CRC is useful for classification, it also has some problems. Extensive experiments show that, under some conditions, some test samples are misclassified by CRC. Thus, combined the advantages of KNN and CRC, this study proposes a local method, namely Collaborative Representation Classifier based on K Nearest Neighbors (KNN-CRC) method. The computational complexity of KNN-CRC is lower than SRC and CRC. Furthermore, under many conditions, the classification results of KNN-CRC are better than KNN and CRC.

Collaborative representation classifier based on k nearest neighbors:

- **KNN-CRC method:** Combined the advantages of KNN and CRC, this study proposes a local CRC method, namely KNN-CRC method. The basic steps of KNN-CRC are as follows. First, given a test sample, find its K nearest neighbors from all the training samples. Second, make the K nearest neighbors as the dictionary, represent the test sample with the dictionary. It can also obtain the closed solution from l_2 norm minimization. Third, compute the residual between the test sample and the each class in the dictionary. Finally, the KNN-CRC classifier can identify the test sample belongs to which class

Specifically, the representation using KNN-CRC method is defined by Eq. 1:

$$x = \arg \min_x \|x\|_2 \text{ s.t. } y = \tilde{B}x \quad (1)$$

where, y denotes a test sample. \tilde{B} denotes the dictionary which is composed of the K nearest neighbors chosen from all the training samples. Unlike CRC, the KNN-CRC method uses the K nearest neighbors as the dictionary. However, CRC method uses all the training samples as the dictionary. The size of K is much smaller than the total number of all training samples, thus, the computational complexity of KNN-CRC is much lower than SRC and CRC. The extensive experiments also clearly show that the proposed method can obtain very competitive classification results compared with other methods.

From KNN-CRC method, the solution to Eq. 1 is:

$$x = (\tilde{B}^T \tilde{B} + \lambda I)^{-1} \tilde{B}^T y$$

Using Lagrange multiplier, a relaxed form of Eq. 1, we obtained Eq. 2:

$$x = \arg \min_x \|y - \tilde{B}x\|_2^2 + \lambda \|x\|_2^2 \quad (2)$$

Taking derivative of Eq. 2 with respect to x , we obtained Eq. 3:

$$-2\tilde{B}^T y + \tilde{B}^T \tilde{B}x + \tilde{B}^T \tilde{B}x + 2\lambda x = 0 \quad (3)$$

Solution of Eq. 3 is obtained as:

$$x = (\tilde{B}^T + \lambda I)^{-1} \tilde{B}^T y$$

- **KNN-CRC classifier:** From the KNN-CRC method, the solution can be obtained. The solution is represented as:

$$x = [x_{\lambda_1}, \dots, x_{\lambda_2}, \dots, x_{\lambda_k}]$$

where, x_{λ_j} is the part of coefficients corresponding to the j th class in x . The j th block of y is defined as:

$$y_j - \tilde{B}_{\lambda_j} x_{\lambda_j}, 1 \leq j \leq L$$

For $j = 1, \dots, k$, this classifier will compute the residual:

$$r_j = \|y - y_j\|_2^2 = \|y - \tilde{B}_{\lambda_j} x_{\lambda_j}\|_2^2$$

It can find the smallest residual easily. If the j th residual is the smallest, it will identify that the test sample y belongs to the j th class.

EXPERIMENTS

In this study, some experiments on digit recognition and face recognition are presented to show the results of classification. These experiments focus on the property evaluation of KNN-CRC and other methods on the digit recognition and face recognition. Three databases, including AR

(Chi and Porikli, 2012; Wright *et al.*, 2009; Zhang *et al.*, 2011), Extended Yale-B (Chi and Porikli, 2012; Wright *et al.*, 2009; Zhang *et al.*, 2011) and MNIST Handwritten Digits database (Chi and Porikli, 2012) are used to test the performance of KNN-CRC and other methods including NN, SRC and CRC. The KNN-CRC is proposed in this study. The NN is the nearest neighbors algorithm which is a conventional method for classification. The SRC is the classical sparse representation method (Wright *et al.*, 2009). The CRC is the collaborative representation based classification with regularized least square (Zhang *et al.*, 2011). This experiments focus on the property comparison of KNN-CRC and other methods.

Face recognition: The KNN-CRC and other methods are tested for comparing the recognition rate. Recognition rate denotes how many test samples can be classified correctly for all the test samples. Higher recognition rate means the property of this method is better. In this experiments, the Eigenface is used as preprocessing in feature extraction.

- **AR database:** The AR database contains about 4000 frontal images for 126 individuals (Chi and Porikli, 2012; Wright *et al.*, 2009; Zhang *et al.*, 2011). These images are captured under different facial disguises, illuminations and expressions. The images are cropped to size 60×43. A few samples of AR database are shown in Fig. 1. Table 1 show the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC
- **Extended Yale-B database:** The Extended Yale-B database contains 2414 frontal face images of 38 individuals (Chi and Porikli, 2012; Wright *et al.*, 2009; Zhang *et al.*, 2011). These samples were cropped and normalized to 54×48. A few samples of Extended Yale-B database are shown in Fig. 2. Table 2 shows the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC

Table 1: Face recognition rates of different methods on the AR database

Dimension	Percentage							
	45	50	55	60	65	70	75	80
NN	65.52	66.38	66.81	67.24	67.38	67.38	67.38	68.10
SRC	78.97	79.26	81.55	84.12	84.26	84.12	85.69	86.70
CRC	76.40	79.26	80.40	83.26	83.41	84.26	86.55	86.84
KNN-CRC (K = 150)	80.83	82.26	83.12	84.84	85.27	85.69	85.84	86.41
KNN-CRC (K = 200)	80.69	81.69	82.69	84.55	85.55	86.41	86.84	87.55
KNN-CRC (K = 300)	81.16	81.83	83.98	85.12	86.12	87.84	87.84	88.70



Fig. 1: Some training samples from the AR database



Fig. 2: Some training samples from the extended Yale-B database



Fig. 3: Some training samples from the MNIST database

Table 2: Face recognition rates of different methods on the extended Yale-B database

Dimension	Percentage						
	85	90	95	105	110	130	180
NN	70.33	70.78	71.32	72.14	72.60	73.50	75.05
SRC	95.01	95.46	95.64	95.46	95.64	95.82	96.55
CRC	95.01	95.46	95.46	95.64	95.83	96.01	96.82
KNN-CRC (K = 150)	95.19	95.64	95.74	96.01	96.01	96.64	96.73
KNN-CRC (K = 200)	95.55	95.46	95.83	95.28	96.28	96.55	97.19
KNN-CRC (K = 300)	95.10	95.46	95.46	96.10	96.37	96.19	96.92

Table 3: Digit recognition results of different methods on the MNIST database

Dimension	Percentage								
	15	25	35	45	55	65	75	85	95
SRC	59	60	61	55	59	58	60	58	55
CRC	63	60	61	59	59	57	59	57	56
KNN-CRC (K = 95)	61	58	58	57	57	56	56	55	54

- Digit recognition:** The MNIST database is used to test the property of these methods. The dimension of each image is 28×28 . Every image which is an 8 bit gray scale image from 0-9 (Chi and Porikli, 2012). The MNIST handwritten digits database which has a training set of 60,000 samples and a test set of 10,000 samples. For our experiment, 10 training samples are randomly selected from each class, 10 test samples are also randomly selected from each class. A few images of MNIST database are shown in Fig. 3. Table 3 show the recognition rates versus feature dimension by NN, SRC, CRC and KNN-CRC

RESULTES AND DISCUSSION

Experiment results and analysis: From all the above face recognition and digit recognition experiments, the differences of SRC, CRC, NN and KNN-CRC can be seen clearly. The SRC is the classical sparse representation method. However, it cannot obtain the closed form solution, its computational complexity is high. In many cases, the recognition rate of KNN-CRC is higher than

Table 4: Comparison of KNN-CRC and other methods

Algorithm	Advantages	Disadvantages
SRC	Classification results are good	Cannot obtain closed form solution, computational complexity is high
CRC	Computational complexity is low	Classification results are not very good
NN	Computational complexity is low	Classification results are bad
KNN-CRC (this study)	Classification results are good, computational complexity is low	

SRC. The CRC is the l_2 norm constraint method, it can obtain the closed form solution, its computational complexity is lower than SRC. However, the recognition rate of KNN-CRC is also higher than CRC. The NN is a simple method for classification. However, in many cases, the recognition rate of NN is the lowest. Compared to other methods, KNN-CRC has many advantages. The recognition rate of KNN-CRC is higher than other methods, the computational complexity of KNN-CRC is also low. The comparison of KNN-CRC and other methods is listed in Table 4.

CONCLUSION

In this study, a new classification method is proposed which is different from SRC and CRC method, namely collaborative representation classifier based on K nearest neighbors (KNN-CRC) method. The KNN-CRC combines the advantages of KNN and CRC methods, the solution of KNN-CRC is obtained from the l_2 norm minimization, it is a closed form solution. Furthermore with the KNN-CRC method, the training samples of a test sample are from the K nearest neighbors of the test sample. The number of the K nearest neighbors is much fewer than all the training samples. A closed form solution is also obtained from KNN-CRC. Thus, the computational complexity of KNN-CRC is much lower than SRC and CRC.

From the digit recognition and face recognition experiments, it clearly showed that the proposed KNN-CRC method can obtain very competitive results compared with SRC, CRC and other representation methods.

ACKNOWLEDGMENTS

This study was supported by the key fund project of Sichuan Educational department (NO.14ZA0005).

REFERENCES

- Adler, A., M. Elad and Y. Hel-Or, 2011. Fast subspace clustering via sparse representations. Technical Report, Department of Computer Science, Technion-Israel Institute of Technology, Israel.
- Aharon, M., M. Elad and A. Bruckstein, 2006. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process*, 54: 4311-4322.
- Bruckstein, A.M., D.L. Donoho and M. Elad, 2009. From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev.*, 51: 34-81.
- Chen, S. and D. Donoho, 1994. Basis pursuit. *Proceedings of the 28th Asilomar Conference on Signals, Systems and Computers*, Volume 1, October 31-November 2, 1994, Pacific Grove, CA., USA., pp: 41-44.
- Cheng, B., J. Yang, S. Yan, Y. Fu and T.S. Huang, 2010. Learning With L^1 -graph for image analysis. *IEEE Trans. Image Process.*, 19: 858-866.

- Chi, Y. and F. Porikli, 2012. Connecting the dots in multi-class classification: From nearest subspace to collaborative representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 16-21, 2012, Providence, RI., USA., pp: 3602-3609.
- Cotter, S.F., 2010. Sparse representation for accurate classification of corrupted and occluded facial expressions. Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing, March 14-19, 2010, Dallas, TX., USA., pp: 838-841.
- Elad, M. and M. Aharon, 2006. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Trans. Image Process.*, 15: 3736-3745.
- Elhamifar, E. and R. Vidal, 2011. Robust classification using structured sparse representation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2011, Providence, RI., USA., pp: 1873-1879.
- Elhamifar, E. and R. Vidal, 2012. Block-sparse recovery via convex optimization. *IEEE Trans. Signal Process.*, 60: 4094-4107.
- Fuchs, J.J., 2005. Recovery of exact sparse representations in the presence of bounded noise. *IEEE Trans. Inform. Theory*, 51: 3601-3608.
- Guha, T. and R.K. Ward, 2012. Learning sparse representations for human action recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34: 1576-1588.
- Mairal, J., F. Bach, J. Ponce and G. Sapiro, 2009. Online dictionary learning for sparse coding. Proceedings of the 26th Annual International Conference on Machine Learning, June 14-18, 2009, Montreal, QC., Canada, pp: 689-696.
- Pati, Y.C., R. Rezaifar and P.S. Krishnaprasad, 1993. Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition. Proceedings of the 27th Asilomar Conference on Signals, Systems and Computers, November 1-3, 1993, Pacific Grove, CA., USA., pp: 40-44.
- Roweis, S.T. and L.K. Saul, 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290: 2323-2326.
- Saul, L.K., S.T. Roweis and Y. Singer, 2003. Think globally, fit locally: Unsupervised learning of low dimensional manifolds. *J. Machine Learn. Res.*, 4: 119-155.
- Soltanolkotabi, M., E. Elhamifar and E.J. Candes, 2014. Robust subspace clustering. *Ann. Stat.*, 42: 669-699.
- Stojnic, M., F. Parvaresh and B. Hassibi, 2009. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans. Signal Process.*, 57: 3075-3085.
- Tropp, J.A., 2004. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50: 2231-2242.
- Wright, J., A.Y. Yang, A. Ganesh, S.S. Sastry and Y. Ma, 2009. Robust face recognition via sparse representation. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31: 210-227.
- Wright, J., Y. Ma, J. Mairal, G. Sapiro, T.S. Huang and S. Yan, 2010. Sparse representation for computer vision and pattern recognition *Proc. IEEE*, 98: 1031-1044.
- Zhang, N. and J. Yang, 2010. K nearest neighbor based local sparse representation classifier. Proceedings of the Chinese Conference on Pattern Recognition, October 21-23, 2010, Chongqing, China, pp: 1-5.
- Zhang, L., M. Yang and X. Feng, 2011. Sparse representation or collaborative representation: Which helps face recognition? Proceedings of the IEEE International Conference on Computer Vision, November 6-13, 2011, Barcelona, Spain, pp: 471-478.