

Modified Constraint Scores for Semi-Supervised Feature Selection

Jianqiao Wang, Yuehua Li* and Kun Chen

*School of Electronic and Optical Engineering, Nanjing University of Science and
Technology, Nanjing, China*
*cambwang@126.com *nlglyh2013@sina.cn*

Abstract

Semi-supervised constraint scores, which utilize both pairwise constraints and the local property of the unlabeled data to select features, achieve comparable performance to the supervised feature selection methods. The local property is characterized without considering the pairwise constraints and these two conditions are introduced independently. However, the pairwise constraints and the local property may contain conflicting information. In this paper, we utilize the conflicting information to improve the local property. Instead of characterizing the local property by all neighbors, samples which do not appear in the cannot-link constraints can be used. A performance indicator, called neighborhood-cannot-link (NC) coefficient, is proposed to measure the improvement of the local property. We use the improved local property and the pairwise constraints to perform semi-supervised constraint scores algorithm. Experiments on several real world data sets demonstrate the effectiveness of the methods.

Keywords: *cannot-link constraints, local property, semi-supervised, feature selection*

1. Introduction

The class label and the pairwise constraints [1-3] are two kinds of user supervision information which are utilized in supervised learning and semi-supervised learning. The pairwise constraints contain the must-link constraints and the cannot-link constraints. Compared with the class label, considering the pairwise constraints is more practical, because the true labels may be unknown a priori, while it is easier for a user to specify whether some pairs of examples belong to the same class or not [4]. Constraint scores [4, 5], which utilizes the pairwise constraints to select features, have been widely studied in recent years. In order to make use of the substantial unlabeled samples, the semi-supervised versions [6, 7] of constraint scores have also been proposed. These semi-supervised methods utilize both pairwise constraints and the local property of the unlabeled data to select features, and achieve comparable performance to the supervised methods.

In the semi-supervised methods, the pairwise constraints are utilized as an independent component, as well as the local property. However, the pairwise constraints and the local property may contain conflicting information. In other words, the relationship between two samples contained in the pairwise constraints and the local property may be different. For example, if two samples are neighbors, they can be utilized to characterize the local property. If these two samples appear in the cannot-link constraints at the same time, they cannot be utilized to describe the local property, because they belong to the different classes. Thus, one of the known conditions should be modified according to another known condition. In practice, the local property is characterized by the neighbors in the sense of Euclidean distance [8-10]. The ideal neighbors of a sample should have the same label. Unfortunately,

due to the similarity among the samples which are going to be classified, the phenomenon that the neighbors have different labels is quite common. Take Yale face database [11] for example, we characterize the local property by calculating the k -nearest neighbor on the first three individuals. The parameter k is set to 3. The result shows that the first, the fourth and the seventh sample have one, two and three neighbors with different labels, respectively. When the number of nearest neighbors increases to 7, every sample in the first individual has at least one neighbor with different labels. For two samples x_i and x_j , if the pairwise constraints indicate that x_i and x_j belong to the different classes, x_i should not appear in the neighborhood of x_j , even though x_i is one of the k -nearest neighbors of x_j . If we do not consider the pairwise constraints when we calculate the local property, the local property may contain incorrect information.

In this paper, we consider how to improve the known conditions before we perform a learning task. Since the pairwise constraints are given as supervision information, while the local property is calculated from all of the samples, we modified the local property according to the cannot-link constraints. Instead of characterizing the local property by all neighbors, samples which do not appear in the cannot-link constraints can be used in our method. To measure the improvement of the local property, we propose a performance indicator, called neighborhood-cannot-link (NC) coefficient. Then, we perform the constraint scores by using the improved local property and the pairwise constraints for semi-supervised feature selection. We name our method as Modified Constraint Scores (MCS). Several experiments were performed to demonstrate the efficacy of our method.

The remainder of the paper is organized as follows. In Section 2, we review the related work. Our proposed method is presented in Section 3. In Section 4, we perform several experiments on UCI database, Yale face database and millimeter wave radar high resolution range profile (HRRP) to demonstrate the effectiveness of the methods. Finally, we provide some concluding remarks in Section 5.

2. Related Work

2.1. Laplacian Score

Laplacian Score (LS) [12-14] is a “filter” method for feature selection. The basic assumption of LS is that, in many real world classification problems, data from the same class are often close to each other. The importance of a feature is evaluated by its power of locality preserving. For each feature, the locality preserving power can be reflected by its Laplacian score.

Given a set of samples $X = [x_1, x_2, \dots, x_n]$. Let x_{ir} denote the r th feature in the i th sample. LS first constructs a nearest neighbor graph G with n nodes. If nodes i and j are connected, put

$$S_{ij} = \exp\left(\frac{-\|x_i - x_j\|^2}{t}\right) \quad (1)$$

where t is a suitable constant. Otherwise, put $S_{ij} = 0$. The weight matrix S of the graph models the local structure of the data space. Then, for the r th feature, define the graph Laplacian matrix:

$$L = D - S \quad (2)$$

where

$$D_{ii} = \sum_{j=1}^n S_{ij} \quad (3)$$

The Laplacian score of the r th feature can be expressed as

$$L_r = \frac{\tilde{f}_r^T L \tilde{f}_r}{\tilde{f}_r^T D \tilde{f}_r} \quad (4)$$

where

$$\tilde{f}_r = f_r - \bar{f}_r \quad (5)$$

$$\bar{f}_r = \frac{\sum_{i=1}^n x_{ir} D_{ii}}{\sum_{i=1}^n D_{ii}} \quad (6)$$

2.2. Constraint Scores

Constraint scores [4, 5] are also filter methods for feature selection. They utilize pairwise constraints as the supervision information. Pairwise constraints simply mention for some pairs of samples that they are similar (must-link constraints), or that they are dissimilar (cannot-link constraints). The must-link constraints M and the cannot-link constraints C can be defined as

$$M = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ belong to the same class}\} \quad (7)$$

$$C = \{(x_i, x_j) \mid x_i \text{ and } x_j \text{ belong to the different classes}\} \quad (8)$$

We construct two graphs G^M and G^C by using the pairwise constraints in M and C . For each graph, we put an edge between node i and j if there is a constraint between samples x_i and x_j . Once the graphs are constructed, their weight matrices can be defined as

$$S_{ij}^M = \begin{cases} 1 & \text{if } (x_i, x_j) \in M \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

$$S_{ij}^C = \begin{cases} 1 & \text{if } (x_i, x_j) \in C \\ 0 & \text{otherwise} \end{cases} \quad (10)$$

Then, the two kinds of constraint scores can be expressed as

$$C_r^1 = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^M}{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^C} \quad (11)$$

$$C_r^2 = \sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^M - \lambda \sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^C \quad (12)$$

where λ is a regularization parameter whose function is to balance the contributions of the two terms in Eq. (12).

3. NC Coefficient and MCS

3.1. NC Coefficient

Semi-supervised constraint scores use both pairwise constraints and the local property to select features. The local property can be obtained without considering the pairwise constraints. Thus, some given class information may not be utilized when we characterized the local property. To improve the local property by using the class information contained in the cannot-link constraints, we define the neighborhood constraints N as

$$N = \{(x_i, x_j) | x_i \text{ belong to the } k\text{-nearest neighbors of } x_j\} \quad (13)$$

Then we construct the k -nearest neighbor graph G^N , where a connection is established between two nodes x_i and x_j if x_i belongs to the k -nearest neighbors of x_j . The weight matrix of graph G^N can be defined as

$$S_{ij}^N = \begin{cases} 1 & \text{if } (x_i, x_j) \in N \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

For two samples x_i and x_j , if the corresponding S_{ij}^N and S_{ij}^C satisfy

$$S_{ij}^N \times S_{ij}^C = 1 \quad (15)$$

x_i should be discarded from the neighbors of x_j . In other words, s_{ij}^N should be set to 0. Then, we recalculate the k -nearest neighbors of x_j until there is no sample satisfies Eq. (15). The final neighborhood weights are stored in matrix S^{NN} . Obviously, samples which have different labels in the neighborhood can be found through this method.

We define the NC coefficient as the following:

$$p = \frac{\sum_{i=1}^n \sum_{j=1}^n S_{ij}^N \times S_{ij}^C}{\sum_{i=1}^n \sum_{j=1}^n S_{ij}^C} \quad (16)$$

The meaning of the NC coefficient is obvious: the larger the NC coefficient, the greater the local property will be changed. It is easy to demonstrate that $0 \leq p \leq 1$. In the case of $p = 0$, the original neighbors of a sample contains no samples which have different labels and the traditional semi-supervised constraint scores do not need to modify. It is noteworthy that all known labels mentioned here are given by the cannot-link constraints. Since the rest labels are not known, there may still be samples with different labels in the neighborhood. In the case of $p > 0$, we have to discard these nearest neighbor samples and recalculate the neighborhood matrix to make sure $p = 0$. If $p = 1$, all the samples in the cannot-link constraints appear in the neighborhood. In this case, the cannot-link constraints contain the most class information, and the performance of algorithms can be best improved.

3.2. Modified Constraint Scores

Zhao, *et al.*, [6] proposed semi-supervised constraint score which uses both unlabeled data and pairwise constraints. Kalakech, *et al.*, [7] provided another semi-supervised version, which is simple product between the laplacian score processed with the unlabeled samples and the constraint score C_r^l defined in Eq. (11). Now, let us consider these two semi-supervised constraint scores. They are described as the following:

$$C_r^3 = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^W}{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^C} \quad (17)$$

$$C_r^4 = L_r \cdot \frac{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^M}{\sum_{i=1}^n \sum_{j=1}^n (x_{ir} - x_{jr})^2 S_{ij}^C} \quad (18)$$

where

$$S_{ij}^W = \begin{cases} \gamma & \text{if } (x_i, x_j) \in M \\ 1 & \text{if } x_i \text{ or } x_j \text{ is unlabeled but } (x_i, x_j) \in N \\ 0 & \text{otherwise} \end{cases} \quad (19)$$

For the given pairwise constraints, we first calculate the neighborhood weight matrix S^N and NC coefficient p . If $p=0$, we execute the algorithms (11) and (12) directly. If $p>0$, we discard the samples which satisfy Eq. (15) from the neighborhood matrix and recalculate the neighborhood matrix until $p=0$. Then, we execute the algorithms with the new neighborhood weight matrix S^{NN} . We summarize the MCS algorithm in Table 1.

Table 1. MCS Algorithm

Input: the sample set X , the pairwise constraints M and C , the number of nearest neighbors k , the number of selected features nf .
Output: the selected features sf .
Algorithm: 1: calculate the neighborhood matrix S^N ; 2: calculate the NC coefficient p ; 3: if $p=0$, execute step 4; if $p>0$, discard the samples which satisfy Eq. (15) from the neighborhood matrix and recalculate the neighborhood matrix until $p=0$, store the new neighborhood weight in matrix S^{NN} ; 4: calculate Eq. (17) or Eq. (18) with S^{NN} . The features corresponding to the nf smallest scores are the selected features.

4. Experimental Results

4.1. Experimental Setting

For convenience, we denote the algorithm using (17) as SSCS-1, denote the algorithm using (18) as SSCS. The corresponding modified algorithms are denoted as MCS-1 and MCS-2, respectively. In the following experiments, we use Yale face database [15], ‘Wine’ and ‘Vehicle’ data set from the UCI repository [16] and millimeter wave radar high resolution range profile (HRRP) to test the performance of these methods. We randomly select pairs of samples from the data set and create must-link and cannot-link constraints. In order to demonstrate the improvement of the modified methods, we perform experiments with different NC coefficient (0.25, 0.50, 0.75, 1.00). We first randomly select pairs with different labels from neighborhood weight matrix as a part of cannot-link constraints, the rest part of cannot-link constraints are obtained by random selection from the whole data except pairs

which are contained in neighborhood weight matrix. The ratio of these two parts of cannot-link constraints is determined by NC coefficient. The number of must-link constraints is as same as the cannot-link constraints for each data set, respectively. For each NC coefficient, we repeat 50 runs to obtain the average classification accuracy. The classifier used here is nearest neighbor classifier.

4.2. Yale Face Database

The Yale face database is constructed at the Yale center for computational vision and control. It contains 165 grayscale images of 15 individuals. Each individual has 11 images. The images demonstrate variations in lighting condition (left-light, center-light, right-light), facial expression (normal, happy, sad, sleepy, surprised, and wink) and with/without glasses. All the face images are manually aligned and cropped. The size of each cropped image is 32×32 pixels, with 256 gray levels. Figure 1 shows some sample images from the Yale face database. Before we perform the experiment, each image is represented as a 1024-dimensional vector.



Figure 1. Sample Images from the Yale Face Database

We create 20 must-link and cannot-link constraints. The recognition accuracy versus different number of selected features on Yale face database is shown in Figure 2. The average recognition accuracy and stand deviations are shown in Table 2.

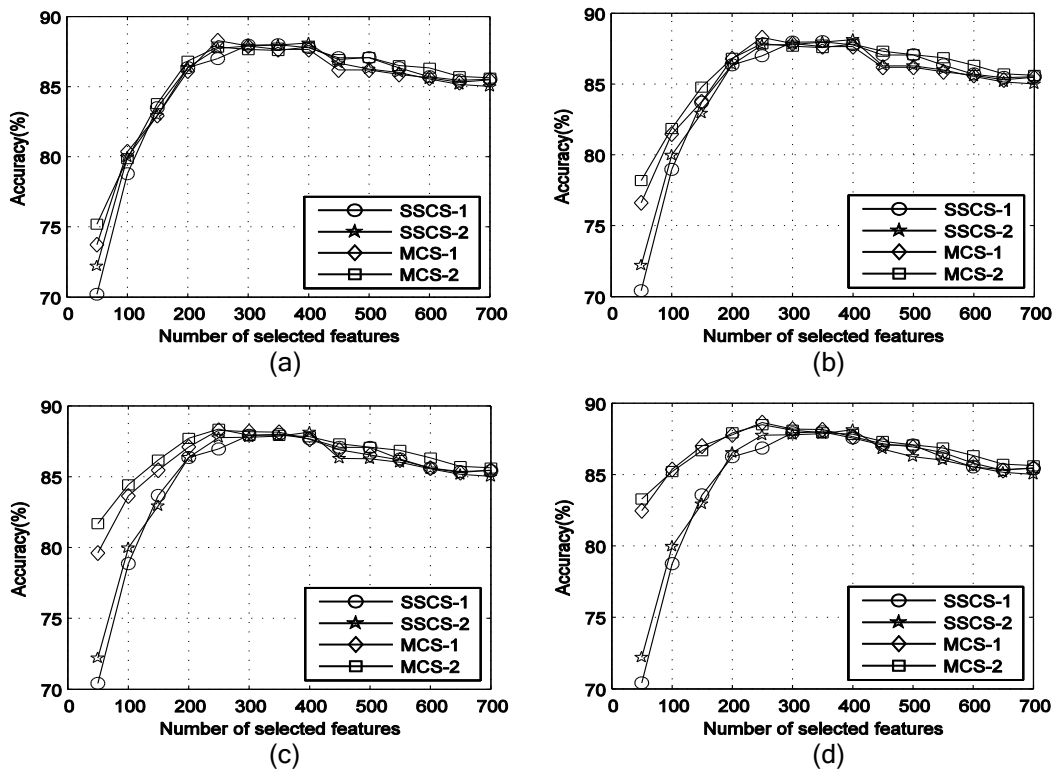


Figure 2. Accuracy vs. Different Number of Selected Features on Yale Face Database with NC Coefficient (a) $p = 0.25$ (b) $p = 0.50$ (c) $p = 0.75$ (d) $p = 1.00$

Table 2. Classification Accuracy and Stand Deviations (Percent) on Yale Face Database with Different NC Coefficient

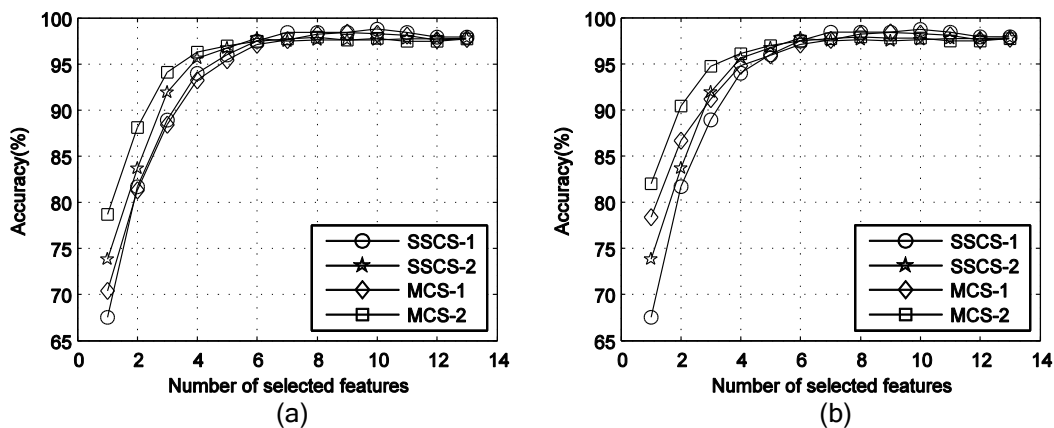
NC coefficient	Accuracy(%)			
	SSCS-1	SSCS-2	MCS-1	MCS-2
0.25	84.76 ± 4.81	84.83 ± 4.25	84.93 ± 3.83	84.71 ± 3.59
0.50	84.80 ± 4.75	84.80 ± 4.23	85.32 ± 3.07	84.84 ± 2.71
0.75	84.75 ± 4.74	84.79 ± 4.23	85.99 ± 2.27	86.58 ± 1.77
1.00	84.71 ± 4.74	84.84 ± 4.25	86.48 ± 1.63	86.73 ± 1.42

From Figure 2 and Table 2, we see that the performance of SSCS-1 and SSCS-2 is similar in the case of four different NC coefficients. While that of MCS-1 and MCS-2 improves when the NC coefficient increases. This increasement is remarkable when the number of selected features is small. Moreover, in the task of race recognition, the recognition accuracy does not always increase with the increasement of the number of selected features. For Yale face database, the most suitable number of selected features is between 250 and 400.

4.3. UCI Database

We utilize ‘Wine’ and ‘Vehicle’ database from the UCI repository to test the performance of our methods. The wine data are the results of a chemical analysis of wines grown in the same region in Italy but derived from three different cultivars. It contains 178 instances and 13 attributes for each instance. These instances are from three classes. Each class contains 59, 71 and 48 instances, respectively. The vehicle data 846 instances from four classes and 18 attributes for each instance. The four classes of vehicles are Opel, Saab, bus and van, respectively. Each class has 212, 217, 218 and 199 instances, respectively.

We create 50 must-link and cannot-link constraints for both wine data set and vehicle data set. The recognition accuracy versus different number of selected features is shown in Figure 3 and Figure 4. The average recognition accuracy and stand deviations are shown in Table 3 and Table 4.



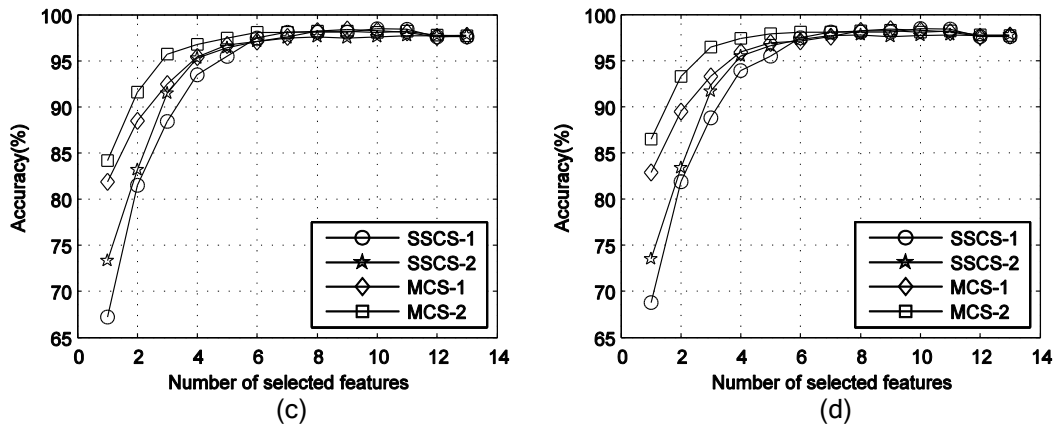


Figure 3. Accuracy vs. Different Number of Selected Features on Wine Data Set with NC Coefficient (a) $p = 0.25$ (b) $p = 0.50$ (c) $p = 0.75$ (d) $p = 1.00$

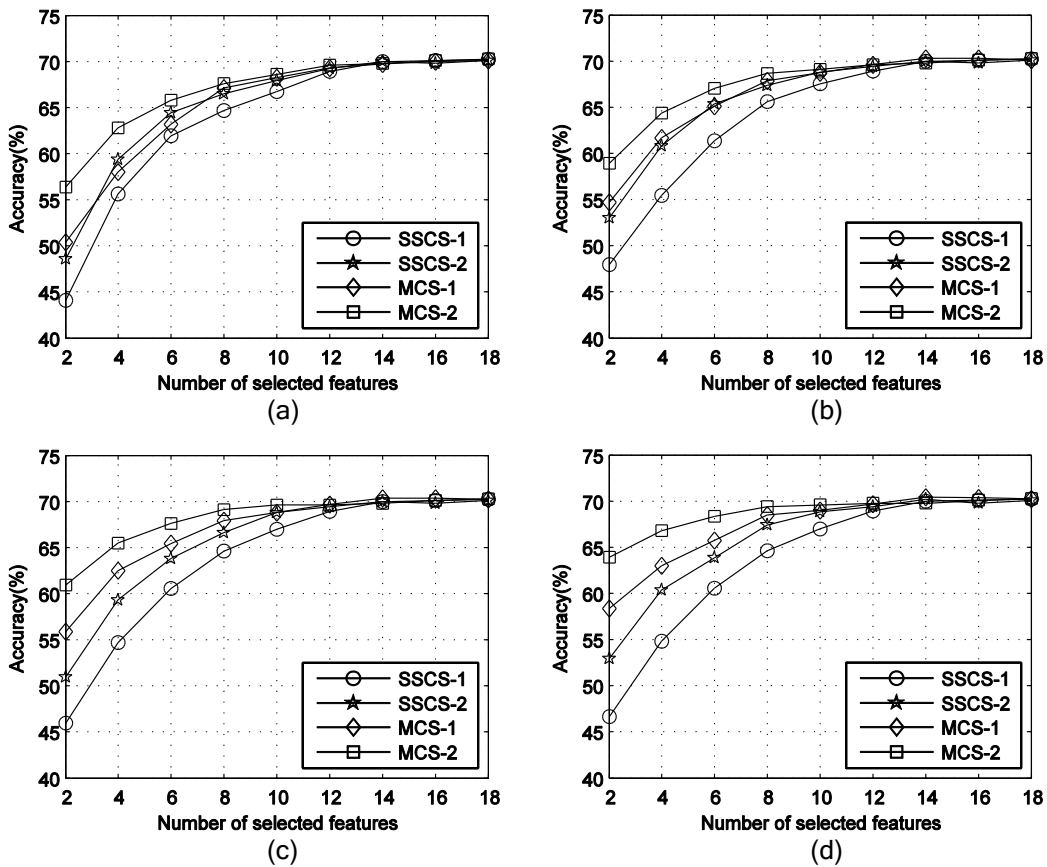


Figure 4. Accuracy vs. Different Number of Selected Features on Vehicle Data Set with NC Coefficient (a) $p = 0.25$ (b) $p = 0.50$ (c) $p = 0.75$ (d) $p = 1.00$

Table 3. Classification Accuracy and Stand Deviations (Percent) on Wine Data Set with Different NC Coefficient

NC coefficient	Accuracy(%)			
	SSCS-1	SSCS-2	MCS-1	MCS-2
0.25	93.37 ± 9.23	94.08 ± 7.28	93.21 ± 8.49	95.03 ± 5.60
0.50	93.36 ± 9.22	94.07 ± 7.27	94.62 ± 5.95	95.51 ± 4.56
0.75	93.08 ± 9.24	93.86 ± 7.42	95.23 ± 4.92	96.17 ± 4.03
1.00	93.29 ± 8.81	94.01 ± 7.39	95.50 ± 4.55	96.62 ± 3.32

Table 4. Classification Accuracy and Stand Deviations (Percent) on Vehicle Data Set with Different NC Coefficient

NC coefficient	Accuracy(%)			
	SSCS-1	SSCS-2	MCS-1	MCS-2
0.25	63.58 ± 8.75	65.10 ± 7.10	65.11 ± 6.81	66.78 ± 4.60
0.50	64.13 ± 7.81	66.07 ± 5.76	66.48 ± 5.26	67.55 ± 3.73
0.75	63.56 ± 8.42	65.42 ± 6.53	66.80 ± 4.87	68.06 ± 3.08
1.00	63.65 ± 8.23	65.88 ± 5.89	67.24 ± 4.16	68.67 ± 2.09

From Figure 3, Figure 4, Table 3 and Table 4, we see that the performance of the four methods is similar with that on Yale face database. The performance of SSCS-1 and SSCS-2 on wine data set and vehicle data set is similar in the case of four different NC coefficients. While that of MCS-1 and MCS-2 improves when the NC coefficient increases. Different from the result of the face recognition, the recognition accuracy increases with the increasement of the number of selected features. If more features are selected, the performance becomes better.

4.4. HRRP Data Set

For high range resolution radar, in a resolution cell, the target can be seen as an assembly of many scatterers, which usually locate on the position where the curvature is large, such as corner and vertex. In this experiment, we use corner reflectors with different position to simulate three different targets. The experimental environment is as follows. The background is cement floor. The bandwidth of transmit signal is 400MHz, thus the range resolution is 0.375m. For each target, we obtain a 256-point HRRP every 1° from azimuth 0° to 179°. Thus each target has 180 HRRPs, and each HRRP is a 256-dimensional vector. Figure 5 shows 8 HRRPs (azimuth 0°, 25°, 50°, 75°, 100°, 125°, 150° and 175°) for each target.

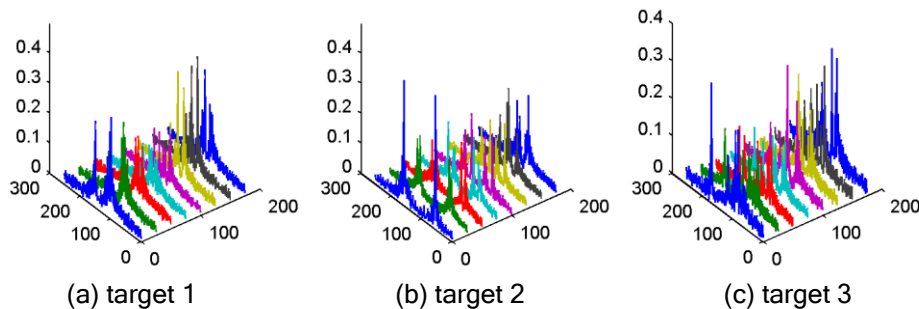


Figure 5. Sample HRRP for Each Target with Azimuth 0°, 25°, 50°, 75°, 100°, 125°, 150° and 175°

We create 50 must-link and cannot-link constraints. The recognition accuracy versus different number of selected features is shown in Figure 6. The average recognition accuracy and stand deviations are shown in Table 5.

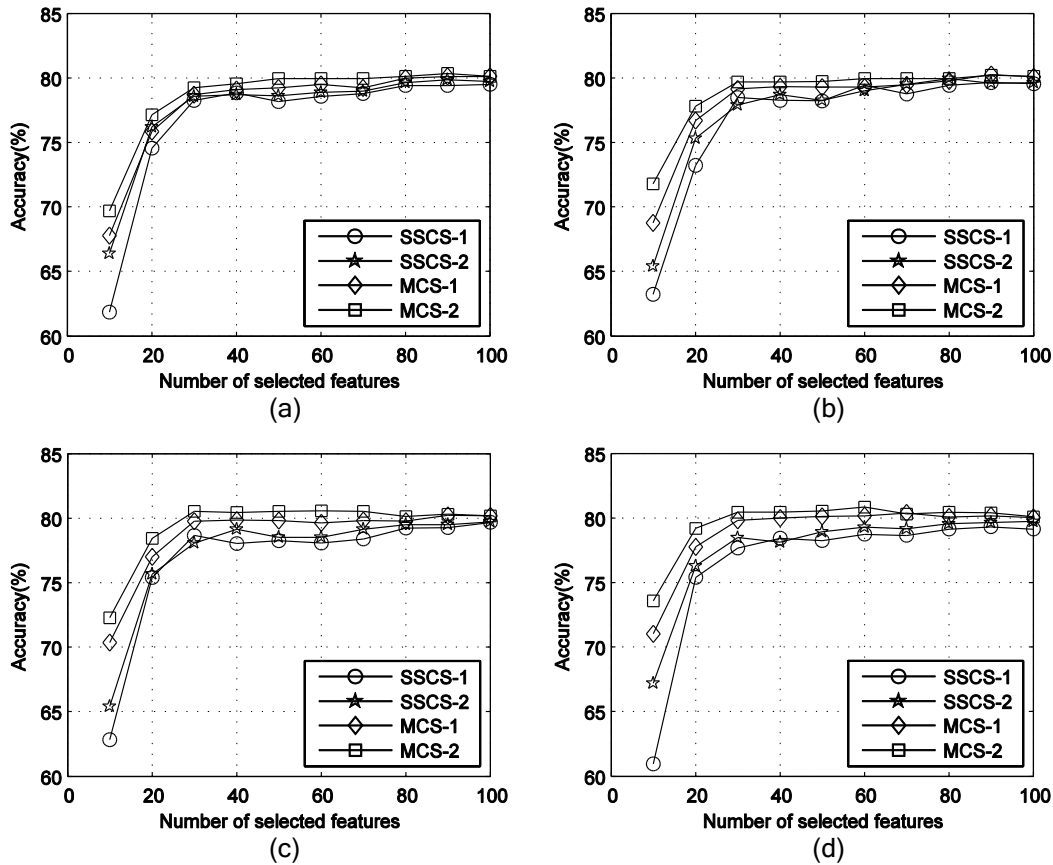


Figure 6. Accuracy vs. Different Number of Selected Features on HRRP Data Set with NC Coefficient (a) $p = 0.25$ (b) $p = 0.50$ (c) $p = 0.75$ (d) $p = 1.00$

Table 5. Classification Accuracy and Stand Deviations (Percent) on HRRP Data Set with Different NC Coefficient

NC coefficient	Accuracy(%)			
	SSCS-1	SSCS-2	MCS-1	MCS-2
0.25	76.73 ± 5.41	77.55 ± 4.05	77.96 ± 3.78	78.60 ± 3.27
0.50	76.81 ± 5.12	77.32 ± 4.39	78.20 ± 3.46	78.88 ± 2.59
0.75	76.78 ± 5.04	77.31 ± 4.34	78.64 ± 3.05	79.38 ± 2.58
1.00	76.57 ± 5.60	77.63 ± 3.90	78.96 ± 2.88	79.61 ± 2.17

From Figure 6 and Table 4, we see that the performance of SSCS-1 and SSCS-2 is similar in the case of four different NC coefficients. While that of MCS-1 and MCS-2 improves when the NC coefficient increases. The recognition accuracy increases with the increasement of the number of selected features. When the number of selected features is larger than 30, the performance tends basically to steady.

5. Conclusion

In this paper, we propose a performance indicator, called NC coefficient, to indicate whether the cannot-link constraints are good or not, which are widely used in pairwise constraints based algorithms. On this basis, we make use of the class information introduced by cannot-link constraints to improve the local property and proposed two modified constraint scores for semi-supervised feature selection. Experimental results on several data sets demonstrate the effectiveness of our method.

Acknowledgements

The authors would like to thank the anonymous reviewers and editors for their helpful suggestions. This work is supported by National (Jiangsu) Natural Science Foundation of China under Grants 60901008, 61001010 and 61371038 (BK2010490), National Ministry Foundation of China under Grants 51305050102.

References

- [1] H. Zeng and Y. Cheung, "Semi-Supervised Maximum Margin Clustering with Pairwise Constraints", *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 5, (2012), pp. 926-939.
- [2] S. Chen and D. Zhang, "Semi supervised Dimensionality Reduction with Pairwise Constraints for Hyperspectral Image Classification", *IEEE Geosci. Remote Sens. Lett.*, vol. 2, no. 8, (2011), pp. 369-373.
- [3] R. Yan, J. Zhang, L. Yang and A. G. Hauptmann, "A Discriminative Learning Framework with Pairwise Constraints for Video Object Classification", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 4, no. 28, (2006), pp. 578-593.
- [4] D. Q. Zhang, S. C. Chen and Z. H. Zhou, "Constraint Score: A New Filter Method for Feature Selection with Pairwise Constraints", *Pattern Recognit.*, vol. 5, no. 41, (2008), pp. 1440-1451.
- [5] D. Sun and D. Zhang, "Bagging Constraint Score for Feature Selection with Pairwise Constraints", *Pattern Recognit.*, vol. 6, no. 43, (2010), pp. 2106-2118.
- [6] J. D. Zhao, K. Lu and X. F. He, "Locality Sensitive Semi-Supervised Feature Selection", *Neurocomputing*, vol. 10-12, no. 71, (2008), pp. 1842-1849.
- [7] M. Kalakech, P. Biela, L. Macaire and D. Hamad, "Constrain Scores for Semi-Supervised Feature Selection: A Comparative Study", *Pattern Recognit. Lett.*, vol. 5, no. 32, (2011), pp. 656-665.
- [8] J. B. Tenenbaum, V. de Silva and J. C. Langford, "A Global Geometric Framework for Nonlinear Dimensionality Reduction", *Science*, vol. 5500, no. 290, (2000), pp. 2319-2323.
- [9] S. T. Roweis and L. K. Saul, "Nonlinear Dimensionality Reduction by Locally Linear Embedding", *Science*, vol. 5500, no. 290, (2000), pp. 2323-2326.
- [10] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for Dimensionality Reduction and Data Representation", *Neural Computation*, vol. 6, no. 15, (2003), pp. 1373-1396.
- [11] A. S. Georghiades, P. N. Belhumeur and D. J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 6, no. 23, (2001), pp. 643-660.
- [12] X. F. He, D. Cai and P. Niyogi, "Laplacian Score for Feature Selection", *Proc. Conf. Advances in Neural Information Processing System*, vol. 18, (2005), December 5-8, pp. 507-514, Vancouver, British Columbia, Canada.
- [13] K. Benabdeslem and M. Hindawi, "Constrained Laplacian Score for Semi-Supervised Feature Selection", *Machine Learning and Knowledge Discovery in Databases, Lecture Notes in Computer Science*, vol. 6911, (2011), pp. 204-218.
- [14] L. Zhu, L. Miao and D. Zhang, "Iterative Laplacian Score for Feature Selection. *Pattern Recognition*", *Communications in Computer and Information Science*, vol. 321, (2012), pp. 80-87.
- [15] P. N. Belhumeur, J. Hespanha and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 7, no. 19, (1997), pp. 711-720.
- [16] C. Blade, E. Keogh and C. J. Merz, "UCI Repository of Machine Learning Databases", (1998), Available: <http://archive.ics.uci.edu/ml/datasets.html>.

Author



Jianqiao Wang, he received the Bachelor Degree in information countermeasure technology from Nanjing University of Science and Technology in 2009. His current research interests on radar target recognition, feature selection and manifold learning.