LETTER

Local Reconstruction Error Alignment: A Fast Unsupervised Feature Selection Algorithm for Radar Target Clustering

Jianqiao WANG[†a)], *Nonmember*, **Yuehua LI**[†b)], *Member*, *and* **Jianfei CHEN**[†], *Nonmember*

SUMMARY Observed samples in wideband radar are always represented as nonlinear points in high dimensional space. In this paper, we consider the feature selection problem in the scenario of wideband radar target clustering. Inspired by manifold learning, we propose a novel feature selection algorithm, called Local Reconstruction Error Alignment (LREA), to select the features that can best preserve the underlying manifold structure. We first select the features that minimize the reconstruction error in every neighborhood. Then, we apply the alignment technique to extend the local optimal feature sequence to a global unique feature sequence. Experiments demonstrate the effectiveness of our proposed method.

key words: *feature selection, local reconstruction, manifold, alignment*

1. Introduction

In wideband radar applications, one is often confronted with very high-dimensional data. The redundancy present in the object's data only slightly improves classifier performance [1]. Besides, time and space requirements will increase significantly when the data contains a large number of redundancy features. Feature selection techniques endeavor to select a meaningful feature subset that contains fewer relevant features, while as much class information is retained as possible [2].

Traditional feature selection algorithms treat the data as globally linear. However, in reality, most of the observed data are represented as nonlinear points in high dimensional Euclidean space. Recently, several manifold based [3] feature selection algorithms have been presented, such as Laplacian Score [4], Laplacian regularized A-optimal feature selection (LapAOFS) and Laplacian regularized D-optimal feature selection (LapDOFS) [5]. These algorithms utilize the graph Laplacian matrix as a component of the criterion to preserve the local structure of the high dimensional data. Thus, the features in the low dimensional data can reflect the similar underlying manifold structure of the observed data.

In this paper, we present a novel manifold based feature selection algorithm, called Local Reconstruction Error Alignment (LREA). Inspired by manifold learning, we select the features that can best preserve the underlying manifold structure. We first calculate the weight matrix that characterizes the manifold structure in the observed data space. Then, for each neighborhood, we select these features which minimize the reconstruction error. As the selected feature sequence may not be the same for every data, we apply the alignment technique to extend the local optimal feature sequence to a globally unique feature sequence.

Manuscript received September 11, 2013.
Manuscript revised October 21, 2013.

[†]The authors are with School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing, China.
a) E-mail: cambwang@126.com
b) E-mail: hmb4507@mail.njust.edu.cn (Corresponding author)

DOI: 10.1587/transinf.E97.D.357

2. Local Reconstruction Error Alignment

2.1 The Object Function

Given the original sample matrix X, each column indicates an observed sample and each row denotes a feature. Suppose that there are n samples in X and each sample has m features. Thus, we can define the original feature set as $F = \{f_1, f_2, \cdots, f_m\}$. Then, we select a feature subset $F^S = \{f_1^S, f_2^S, \cdots, f_q^S\} \subset F$ which contains q features. The sample matrix after the feature has been selected can be presented as X^S.

LLE [6] assumes that each sample can be reconstructed by the linear combination of its neighbors. The combination coefficient, in other words, the weight, contains the local structure information. The basic idea of our algorithm is that the samples in X^S share the same local geometric structure of the samples in X. For each x_i, let $X_i^{nb} = [x_{i1}, x_{i2}, \cdots, x_{ik}]$ be a matrix consisting of its k-nearest neighbors, say in terms of the Euclidean distance. We define the reconstruction weight as w_{ij}, which indicates the contribution of x_j to x_i. The optimal weight can be obtained by the following objective function minimizing the reconstruction error

$$\arg\min_{w_{ij}} \sum_{i=1}^{n} \|\varepsilon_i\|^2 \tag{1}$$

where ε_i is the reconstruction error of sample x_i and

$$\varepsilon_i = x_i - \sum_{j=1}^{k} w_{ij} x_{ij} \tag{2}$$

Let $X_i^{Snb} = [x_{i1}^S, x_{i2}^S, \cdots, x_{ik}^S]$ be the corresponding neighborhood matrix in X^S. In order to keep the same local geometric structure, the local reconstruction of x_i^S can be written as

$$x_i^S = w_{i1} x_{i1}^S + w_{i2} x_{i2}^S + \cdots + w_{ik} x_{ik}^S + \varepsilon_i^S \tag{3}$$

where ε_i^S indicates the reconstruction error of sample x_i^S and

$$\varepsilon_i^S = x_i^S - \sum_{j=1}^{k} w_{ij} x_{ij}^S \tag{4}$$

In order to get the optimal X^S, we should minimize the reconstruction error in Eq. (4). Then, the objective function of LREA can be presented as the following

$$\begin{aligned} &\arg\min_{X^S} \sum_{i=1}^{n} \left\| \varepsilon_i^S \right\|^2 \\ &s.t.\ \ F^S \subset F \end{aligned} \tag{5}$$

2.2 Optimization Scheme

Obviously, it is a NP-hard problem if we solve the objective function directly. At the same time, it is noteworthy that the locally optimal solution may not be optimal in global. In order to overcome these shortcomings, we apply the alignment technique [7] to optimize the objective function.

Let Λ_i be an $m \times m$ diagonal matrix whose diagonal entry Λ_{ii} is 1 if the ith feature is selected and 0 otherwise. It is easy to demonstrate that Λ_i has the following property.

$$\Lambda_i^T \Lambda_i = \Lambda_i \tag{6}$$

The sample in X and the corresponding sample in X^S have the following relationship.

$$x_i^S = \Lambda_i x_i \tag{7}$$

Then, the objective function (5) can be expressed as a function of Λ_i. In general, each locally optimal solution, in other words, Λ_i, may not be the same. However, the goal of our algorithm is to select features from the whole samples, which means the sequence of the selected features in each sample must be the same. Thus, we have

$$\Lambda_1 = \Lambda_2 = \cdots = \Lambda_n = \Lambda \tag{8}$$

The reconstruction error of x_i^S can be rewritten as

$$\begin{aligned} \varepsilon_i^S &= x_i^S - \sum_{j=1}^{k} w_{ij} x_{ij}^S \\ &= \Lambda x_i - \sum_{j=1}^{k} w_{ij} \Lambda x_{ij} \\ &= \Lambda (x_i - \sum_{j=1}^{k} w_{ij} x_{ij}) \\ &= \Lambda \varepsilon_i \end{aligned} \tag{9}$$

Then, the optimization problem Eq. (5) can be expressed as

$$\begin{aligned} \sum_{i=1}^{n} \left\| \varepsilon_i^S \right\|^2 &= \sum_{i=1}^{n} \|\Lambda \varepsilon_i\|^2 \\ &= \sum_{i=1}^{n} \varepsilon_i^T \Lambda^T \Lambda \varepsilon_i \\ &= \sum_{i=1}^{n} \varepsilon_i^T \Lambda \varepsilon_i \end{aligned} \tag{10}$$

Then, the objective function can be represented as

$$\begin{aligned} &\arg\min_{\Lambda} \sum_{i=1}^{n} \varepsilon_i^T \Lambda \varepsilon_i \\ &s.t.\ \ \Lambda \text{ is diagonal} \\ &\Lambda_{ii} \in \{0, 1\},\ i = 1, 2, \cdots, m \\ &\sum_{i=1}^{m} \Lambda_{ii} = q \end{aligned} \tag{11}$$

According to the definition of Λ, we have

$$\varepsilon_i^T \Lambda \varepsilon_i = \sum_{j=1}^{m} \Lambda_{jj} \varepsilon_{ij}^2 \tag{12}$$

where ε_{ij}^2 indicates the jth component in ε_i. The optimization problem Eq. (11) can be expressed as

$$\begin{aligned} \sum_{i=1}^{n} \varepsilon_i^T \Lambda \varepsilon_i &= \sum_{i=1}^{n} \sum_{j=1}^{m} \Lambda_{jj} \varepsilon_{ij}^2 \\ &= \sum_{j=1}^{m} (\Lambda_{jj} \sum_{i=1}^{n} \varepsilon_{ij}^2) \end{aligned} \tag{13}$$

The optimization problem can be solved by using Eq. (13) directly. It is obvious that the selected feature sequence is the q smallest value in $\sum_{i=1}^{n} \varepsilon_{ij}^2$. LREA algorithm is summarized in Table 1.

2.3 Computational Complexity Analysis

Suppose we have n samples and each sample has m features. We need to select q features. The number of nearest neighbors is k. From the derivation of LREA, we can see that the most time-consuming parts are step 1 and step 2. Computing nearest neighbors scales in the worst case as $O(mn^2)$, or linearly in the input dimensionality, n, and quadratically in the number of data points, n. However, for many distributions of data, constructions such as K-D trees or ball trees can be used to compute the neighbors in $O(n \log n)$ [6]. The cost of computing the reconstruction weights is $O(mnk^3)$. Compared with the first two steps, the cost of the last two steps is much smaller. Both step 3 and step 4 have a complexity of $O(mn)$. Thus, the total cost of LEAR is $O(mnk^3)$.

Table 1 LREA algorithm.

Input: the original sample matrix X, the number of nearest neighbor k, the number of selected features q.
Output: the indexes of selected features SF.
Algorithm:
1: For each sample x_i, calculate the k-nearest neighborhood matrix X_i^{nb};
2: Calculate the reconstruction weight w_{ij} by using Eq.(1);
3: For each sample x_i, calculate the reconstruction error ε_i by using Eq. (2);
4: Calculate the quadratic sum of reconstruction error $\sum_{i=1}^{n} \varepsilon_{ij}^2$, the index corresponding to the q smallest value in the quadratic sum is selected as sequence of the selected feature.

By contrast, The total cost of the sequential scheme of LapAOFS is $O(qmn^2)$ [5]. It is noteworthy that, in practice, in order to preserve the local structure, the number of nearest neighbors k is usually set much smaller than the number of samples n. In order to compare the time complexities between LREA and LapAOFS, simulation experiments with MATALB were performed in the next section.

3. Experimental Results

High range resolution profile (HRRP) is a kind of general signal in wideband radar system. It contains the information which indicates the radial location of the scatterers in the observation target. 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 small, such as corner and vertex. In this section, we carry out clustering experiments on millimeter wave radar HRRP. In order to demonstrate the effectiveness of our proposed algorithm, we compare our LREA with LapAOFS, Laplacian Score and Data Variance. We select a feature subset from the original data with these four feature selection algorithms. Then, we carry out clustering algorithm on the selected features. The clustering algorithm we used here is k-means.

The experimental environment is as follows. We use corner reflectors with different position to simulate three different targets. The distance from radar to target is less than 10 meters. The background is cement floor. The bandwidth of transmit signal is 400MHz, thus the range resolution is 0.375m. For each simulate target, we obtain a 256 point HRRP every 1° form azimuth 0° to 89°. Thus each target has 90 HRRPs, and each HRRP has 256 features. For each HRRP, we select $10, 20, \ldots, 250$, features, respectively. Then, we apply k-means algorithm to cluster these selected features. The average recognition rate of 20 tests is utilized to measure the performance of each algorithm. The recognition rate curves are shown in Fig. 1. Table 2 shows the average recognition rate, as well as the stand deviations. Due to the achievement of LREA is relatively simplicity, the mean time consumption of each algorithm is presented in Table 3 as reference.

As can be observed from Fig. 1, the proposed LREA achieve the highest recognition rate when the number of selected features is 70. The performances of LREA and LapAOFS are similar. From Table 2, we can get that the average recognition rate of LREA is a litter better than that of LapAOFS. However, the time consumption shown in Table 3 indicates that LapAOFS needs more time to select features than the three other algorithms. As LREA, Laplacian Socre and Data Variance sort the features according to importance, the feature sequence can be obtained one time. Thus, the time consumption is similar when various features are selected. The features selected by LapAOFS are obtained through iterative approach. The number of selected features indicates the number of iterations. So, the time consumption rise significantly with the number of selected features increases.

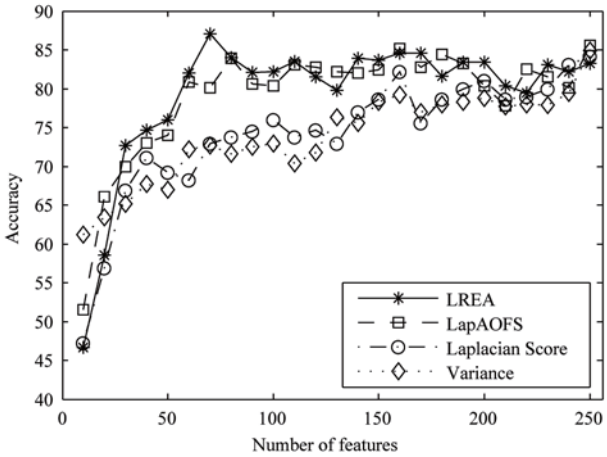

Fig. 1 Performance comparison of recognition rate.

Table 2 Average recognition rate and stand deviations.

	LREA	LapAOFS	Laplacian Score	Variance
mean	79.38	79.06	74.24	73.92
std	8.05	7.44	8.18	5.73

Table 3 Time consumption (s).

number of features	LREA	LapAOFS	Laplacian Score	Variance
10	0.0423	0.4127	0.0366	0.0012
20	0.0421	0.7551	0.0357	0.0013
30	0.0457	1.1002	0.0360	0.0013
40	0.0419	1.4366	0.0353	0.0013
50	0.0420	1.7853	0.0365	0.0013
60	0.0424	2.0809	0.0352	0.0014
70	0.0421	2.3973	0.0346	0.0013
80	00465	2.6710	0.0346	0.0014
90	0.0430	2.9496	0.0358	0.0014
100	0.0429	3.1961	0.0353	0.0013

4. Conclusion

In this paper, we proposed a fast manifold based feature selection algorithm, called LREA, to select the features that can best preserve the underlying manifold structure. The manifold structure is characterized by the reconstruction weight matrix. For each neighborhood, we minimize the reconstruction error to obtain the optimal features. Finally, we apply the alignment technique to extend the local optimal feature sequence to a global unique feature sequence. Our algorithm yields better performance more rapidly than existing manifold-based feature selection algorithms.

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 and 61001010 (BK2010490), National Ministry Foundation of China under Grants 51305050102.

References

[1] S. Boutemedjet, N. Bouguila, and D. Ziou, "A hybrid feature extraction selection approach for high-dimensional non-gaussian data clustering," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.8, pp.1429–1443, 2009.

[2] D. Tao, X. Li, X. Wu, and S.J. Maybank, "Geometric mean for subspace selection," IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.2, pp.260–274, 2009.

[3] M. Belkin, P. Niyoni, and V. Sindhwani, "Manifold regularization: A geometric framework for learning from examples," J. Machine Learning Research, vol.7, pp.2399–2434, 2006.

[4] X. He, D. Cai, and P. Niyogi, "Laplacian score for feature selection," in Advances in Neural Information Processing Systems 18, MIT Press, 2005.

[5] X. He, M. Ji, C. Zhang, and H. Bao, "A variance minimization criterion to feature selection using Laplacian regularization," IEEE Trans. Pattern Anal. Mach. Intell., vol.33, no.10, pp.2013–2025, 2011.

[6] S. Roweis and L. Saul, "Nonlinear dimensionality reduction by locally linear embedding," Science, vol.290, no.5500, pp.2323–2326, 2000.

[7] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimension reduction via local tangent space alignment," SIAM J. Sci. Comput., vol.26, no.1, pp.313–338, 2004.