Shrinkage-divergence-proximity locally linear embedding algorithm for dimensionality reduction of hyperspectral image

Qin Luo (罗 琴)[1], Zheng Tian (田 铮)[1,2], and Zhixiang Zhao (赵志祥)[1]

[1]*Department of Applied Mathematics, Northwestern Polytechnical University, Xi'an 710072*

[2]*State Key Laboratory of Remote Sensing Science, Institute of Remote Sensing Applications, Chinese Academy of Sciences, Beijing 100101*

Received September 29, 2007

Existing manifold learning algorithms use Euclidean distance to measure the proximity of data points. However, in high-dimensional space, Minkowski metrics are no longer stable because the ratio of distance of nearest and farthest neighbors to a given query is almost unit. It will degrade the performance of manifold learning algorithms when applied to dimensionality reduction of high-dimensional data. We introduce a new distance function named shrinkage-divergence-proximity (SDP) to manifold learning, which is meaningful in any high-dimensional space. An improved locally linear embedding (LLE) algorithm named SDP-LLE is proposed in light of the theoretical result. Experiments are conducted on a hyperspectral data set and an image segmentation data set. Experimental results show that the proposed method can efficiently reduce the dimensionality while getting higher classification accuracy.

OCIS codes: 100.0100, 070.4340, 100.5010.
doi: 10.3788/COL20080608.0558.

Hyperspectral data has high spectral dimension and altitudinal interband redundancy which greatly abate the performance of common analysis techniques. To solve this problem, researchers have proposed many methods to reduce data volume, such as image compression[1], endmember extraction[2], and dimensionality reduction[3]. Manifold learning is one of the novel and efficient nonlinear dimensionality reduction techniques. Researches have shown the great potential of manifold learning for dimensionality reduction of hyperspectral data[4]. However, because of the "curse of dimensionality" phenomenon[5], the distance function will become unstable in high-dimensional space even with the commonly used Minkowski metrics on the Euclidean space[6].

Hsu *et al.* provided the sufficient and necessary condition of unstability of distance function and designed a novel distance function called shrinkage-divergence-proximity (SDP)[7]. In this paper, we firstly prove the stability of SDP. Then in light of the theoretical result, we propose an improved locally linear embedding (LLE) algorithm, named SDP-LLE. Experiments are conducted on two data sets, one is hyperspectral data from the Airborne Visual/Infrared Imaging Spectrometer (AVIRIS), and the other is image segmentation data from the University of California, Irvine (UCI) machine learning repository. Experimental results show that the proposed method can efficiently reduce the dimension while getting higher classification accuracy.

The measurement of similarity or distance is fundamental in data mining field. Especially during the manifold learning process, most algorithms begin with the calculation of a matrix of distance. It has been proved that the ratio of distances of nearest and farthest neighbors to a given query in high-dimensional space is almost unit for Minkowski metrics L_p $(p \geq 1)$[6]. Hinneburg *et al.* examined the unstable behavior of L_p norm[8]. SDP is a meaningful distance function defined as follows.

Let f be a non-negative real function defined on set of non-negative real numbers such that

$$f_{a,b}(x) = \begin{cases} 0, & 0 \leq x < a \\ x, & a \leq x < b \\ e^x, & \text{otherwise} \end{cases}. \tag{1}$$

For any m-dimensional data point $\vec{x} = (x_1, \cdots, x_m)$ and $\vec{y} = (y_1, \cdots, y_m)$, we define the SDP function as

$$\text{SDP}^G(\vec{x}, \vec{y}) = \sum_{i=1}^{m} \omega_i f_{s_{i1}, s_{i2}}\left(d_1\left(x_i, y_i\right)\right), \tag{2}$$

where the distance function $f_{s_{i1}, s_{i2}}$ is defined between $\vec{x}$ and $\vec{y}$ on each individual attribute, ω_i $(i = 1, \cdots, m)$ implies the importance of attribute i, s_{i1} and s_{i2} are dependent on the distribution of data points projected on ith dimension, and d_1 is the distance function of a one-dimensional (1D) data space. Equations (2) is a general form of SDP.

We provide a proof in the following theory where the stability of SDP distance function is derived, showing that SDP distance function is stable in any high-dimensional space.

Let $X = \{\vec{x}_1, \cdots, \vec{x}_N\}$ be a data set in high-dimensional space (N is the number of data points), and take on certain clustering structure, with the class number and the attribution of each data point unknown. The distance function SDP is stable if all attributes of the data set is independent of each other. This theory is proved as follows.

According to Ref. [7], the result will be proved if we show that the extremal ratio $D_{\max m}/D_{\min m}$ will not converge to 1 with increasing m ($D_{\max m}$, $D_{\min m}$ are the farthest and nearest distances between data points in m-dimensional space). Assume that the distance between $\vec{x}_{i0}$ and $\vec{x}_{j0}$ is maximal, and the distance of $\vec{x}_{i1}$ to $\vec{x}_{j1}$ is minimal. It is obvious that $\vec{x}_{i0}$ and $\vec{x}_{j0}$ are in different classes, $\vec{x}_{i1}$ and $\vec{x}_{j1}$ are in the same class.

Set

$$\begin{aligned}|\vec{x}_{i0}-\vec{x}_{j0}| &= \{|x_{i0,1}-x_{j0,1}|,\cdots,|x_{i0,m}-x_{j0,m}|\}\\ &= (r_1,\cdots,r_m),\\ |\vec{x}_{i1}-\vec{x}_{j1}| &= \{|x_{i1,1}-x_{j1,1}|,\cdots,|x_{i1,m}-x_{j1,m}|\}\\ &= (r'_1,\cdots,r'_m),\end{aligned} \quad (3)$$

$$D_{\max m}=D(\vec{x}_{i0},\vec{x}_{j0}),\quad D_{\min m}=D(\vec{x}_{i1},\vec{x}_{j1}). \quad (4)$$

For any $\varepsilon>0$, let $s^*_{k_1}$ be the maximum of s_{k_1} which satisfies $P\{d_1(x_{ik},x_{jk})\le s_{k_1}\}\le\varepsilon$ in the condition that $\vec{x}_i$ and $\vec{x}_j$ fall into different classes, that is,

$$s^*_{k_1}=\max\{s_{k_1}|P\{d_1(x_{ik},x_{jk})\le s_{k_1}\}\le\varepsilon\}. \quad (5)$$

Using the same argument as shown before, let $s^*_{k_2}$ be the minimum of s_{k_2} which satisfies $P\{d_1(x_{ik},x_{jk})\ge s_{k_2}\}\le\varepsilon$ in the condition that $\vec{x}_i$ and $\vec{x}_j$ are in the same class:

$$s^*_{k_2}=\min\{s_{k_2}|P\{d_1(x_{ik},x_{jk})\ge s_{k_2}\}\le\varepsilon\}. \quad (6)$$

Defining: $s_1=\min\{s^*_{k_1}|k=1,2,\cdots,m\}$ and $s_2=\max\{s^*_{k_2}|k=1,2,\cdots,m\}$, it is easy to check that s_1 decreases and s_2 increases respectively with the increase of dimensionality m.

Considering that all attributes of the data are independent of each other, it is natural based on the definition of SDP that: when $\vec{x}_i$ and $\vec{x}_j$ are in different classes,

$$P\{\mathrm{SDP}_{s_1,s_2}(x_i,x_j)=0\}\le(\varepsilon)^m \xrightarrow[m\to\infty]{P} 0, \quad (7)$$

when $\vec{x}_i$ and $\vec{x}_j$ are in the same classes,

$$P\{\mathrm{SDP}_{s_1,s_2}(x_i,x_j)\ge m\mathrm{e}^{s_2}\}\le(\varepsilon)^m \xrightarrow[m\to\infty]{P} 0, \quad (8)$$

where $\vec{x}_n \xrightarrow{P} \vec{c}$ represents that a vector sequence $\vec{x}_1,\vec{x}_2,\cdots$ converges in probability to a constant vector $\vec{c}$ if $\forall\varepsilon>0$, $\lim_{n\to\infty}P\{|\vec{x}_n-\vec{c}|\le\varepsilon\}=1$.

According to the analysis above, we have

$$P(f_{s_1,s_2}(r_i)=0)\xrightarrow[m\to\infty]{P} 0, \quad (9)$$

$$P\left(f_{s_1,s_2}(r'_i)=\mathrm{e}^{r'_i}\right)\xrightarrow[m\to\infty]{P} 0. \quad (10)$$

Without loss of generality, let

$$P(f_{s_1,s_2}(r_i)=r_i, s_1\le r_i<s_2)\xrightarrow[m\to\infty]{P} t_1,$$

$$P(f_{s_1,s_2}(r_i)=\mathrm{e}^{r_i}, r_i\ge s_2)\xrightarrow[m\to\infty]{P} 1-t_1, \quad (11)$$

$$P(f_{s_1,s_2}(r'_i)=0, 0\le r'_i<s_1)\xrightarrow[m\to\infty]{P} 1-t_2,$$

$$P(f_{s_1,s_2}(r'_i)=r'_i, s_1\le r'_i<s_2)\xrightarrow[m\to\infty]{P} t_2, \quad (12)$$

where $0<t_1,\ t_2<1$.

It is clear that

$$D_{\max m}\ge mt_1\cdot s_1+m(1-t_1)\mathrm{e}^{s_2},\quad m\to\infty, \quad (13)$$

$$D_{\min m}<mt_2\cdot s_2,\quad m\to\infty. \quad (14)$$

Based on expressions (13) and (14), we get

$$\frac{D_{\max m}}{D_{\min m}}>\frac{mt_1\cdot s_1+m(1-t_1)\cdot\mathrm{e}^{s_2}}{mt_2\cdot s_2} =\frac{t_1\cdot s_1+(1-t_1)\cdot\mathrm{e}^{s_2}}{t_2\cdot s_2}\gg 1,\quad m\to\infty.$$

Thus we arrive to the conclusion that SDP distance function is stable.

Among the existing manifold learning algorithms such as isometric feature mapping (Isomap)[9], LLE[10], Laplacian eigenmap[11], and stochastic neighbor embedding (SNE)[12], LLE has some advantages[13]: optimizations not involving local minima, recovering global nonlinear structure from locally linear fits, lower computation load because it only considers local relation of data points. This algorithm consists of three steps:

1) Find k nearest neighbors for each point $\vec{x}_i$ in space R^D by using Euclidean distance to measure similarity.

2) Calculate the reconstructing linear coefficients by minimizing the reconstruction error[10] $\varepsilon(w)=\sum\limits_{i=1}^{N}\left\|\vec{x}_i-\sum\limits_{j=1}^{N}w_{ij}\vec{x}_j\right\|^2$ (N is the number of data points). Subject to two constrains, $\sum\limits_{j=1}^{N}w_{ij}=1$ and $w_{ij}=0$, if $\vec{x}_j$ is not the neighbor of $\vec{x}_i$.

3) Compute low-dimensional embeddings which best preserve the local geometry by minimizing the embedding cost function for the fixed weights[9]: $\delta(\vec{y})=\sum\limits_{i=1}^{N}\left\|\vec{y}_i-\sum\limits_{j=1}^{N}w_{ij}\vec{y}_j\right\|$. Under two constraints, $\frac{1}{N}\sum\limits_{i=1}^{N}\vec{y}_i\vec{y}_j'=1$ (normalized unit covariance) and $\sum\limits_{i=1}^{N}\vec{y}_i=0$ (translation-invariant embedding), a unique solution is guaranteed.

We use SDP to measure the similarities in the first step. Furthermore, the hierarchical method[14] is used to automatically chose appropriate neighbor parameter k.

We performed experiments on hyperspectral image data generated by AVIRIS. Figure 1(a) is a sub image chosen from the source reference map with rows from 345 to 468 and columns from 267 to 311 by TNTmips software. Using LLE and SDP-LLE for feature extractors respectively and c-means for classifier, we display the classification results with different gray levels for different classes. In comparison, the results of SDP-LLE (Fig. 1(b)) can describe more details than the results of LLE (Fig. 1(c)). It proves that SDP-LLE can preserve more information than LLE during the dimensionality reduction process.

In order to use classification accuracy to quantificationally compare SDP-LLE and LLE, we got an image

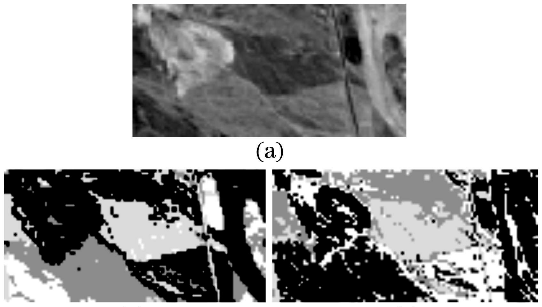

Fig. 1. Comparison of classification by LLE and SDP-LLE. (a) Sub image from source reference map; (b) result of SDP-LLE; (c) result of LLE.

Table 1. Confusion Matrix of Test Data Reduced by LLE

		Referenced Data						
		B	S	F	C	W	P	G
	B	212	0	182	40	67	37	2
	S	0	277	0	23	0	0	0
Classified	F	0	0	35	9	4	15	21
Data	C	19	19	9	202	21	0	0
	W	69	0	74	11	208	9	28
	P	0	4	0	9	0	239	6
	G	0	0	0	6	0	0	243

For simplicity, clusters of Brickface, Sky, Foliage, Cement, Window, Path, and Grass are represented by B, S, F, C, W, P, G, respectively.

Table 2. Confusion Matrix of Test Data Reduced by SDP-LLE

		Referenced Data						
		B	S	F	C	W	P	G
	B	241	0	146	48	62	20	22
	S	0	299	0	9	6	0	0
Classified	F	0	0	89	2	5	9	1
Data	C	19	1	4	213	16	0	0
	W	40	0	61	12	208	0	24
	P	0	0	0	12	0	269	0
	G	0	0	0	4	3	2	253

image segmentation data set from UCI machine learning repository and use confusion matrices[15] to visualize the predicted classes of referenced data. The dimensionality of this data set is 19. It consists of 2100 data points and 7 clusters with equal data points each. The names of the clusters are Brickface, Sky, Foliage, Cement, Window, Path, and Grass. The confusion matrices for LLE and SDP-LLE are illustrated in Tables 1 and 2.

The overall accuracy is computed as the percentage of correctly classified pixels among all considered pixels. In confusion matrix, it is the sum of values in the main diagonal divided by the total number of pixels. As illustrated, the overall accuracy of Table 1 is 0.74857, higher than that of 0.67429 in Table 2.

In this paper, we concentrated on meaningful distance function in high-dimensional space and its application to dimensionality reduction by manifold learning algorithms. We proved the stability of SDP distance function and provided SDP-LLE algorithm by using SDP to measure similarities instead of Euclidean norm of original LLE. Experiments demonstrated that the improved LLE algorithm is more capable to discover the true structure of high-dimensional data sets.

This work was supported by the Graduate Starting Seed Fund of Northwestern Polytechnical University (No. Z200760) and the Innovation Fund of Northwestern Polytechnical University. Q. Luo's e-mail address is lqkybs@sina.com. Z. Tian's e-mail address is ztian@nwpu.edu.cn.

References

1. J. Huang, R. Zhu, J. Li, and Y. He, Chin. Opt. Lett. **5,** 393 (2007).
2. L. Xun, Y. Fang, and X. Li, Acta Opt. Sin. **27,** 1178 (2007).
3. Y. Zhu, P. K. Varshney, and H. Chen, in *Proceedings of IEEE International Conference on Image Processing 2007* **4,** 97 (2007).
4. Y. Chen, M. M. Crawford, and J. Ghosh, in *Proceedings of 2005 IEEE International Geoscience and Remote Sensing Symposium* **6,** 4311 (2005).
5. D. L. Donoho, "High dimensional data analysis: the curses and blessings of dimensionality" Lecture at the "Mathematical Challenges of the 21st Century" Conference of the American Math. Society (2000). Currently available on the web at http://www.stat.stanford.edu/∼donoho/Lectures/AMS2000/AMS2000.html (April 25, 2008).
6. K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, Lecture Notes in Computer Science **1540,** 217 (1999).
7. C.-M. Hsu and M.-S. Chen, in *Proceedings of the 6th SIAM International Conference on Data Mining* 12 (2006).
8. A. Hinneburg, C. Aggarwal, and D. A. Keim, in *Proceedings of the 26th International Conference on Very Large Data Bases* 506 (2000).
9. J. B. Tenenbaum, V. de Silva, and J. C. Langford, Science **290,** 2319 (2000).
10. S. T. Roweis and L. K. Saul, Science **290,** 2323 (2000).
11. M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering" in *Advances in Neural Information Processing Systems vol.14* (MIT Press, Cambridge, 2002) pp.585 – 591.
12. G. Hinton and S. Roweis, "Stochastic neighbor embedding" in *Advances in Neural Information Processing Systems vol.15* (MIT Press, Cambridge, 2003) pp.833 – 840.
13. L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding" http://www.cs.toronto.edu/∼roweis/lle/papers/lleintro.pdf (July 18, 2007).
14. O. Kouropteva, O. Okun, and M. Pietikainen, in *Proceedings of the 1st International Conference on Fuzzy Systems and Knowledge Discovery* 359 (2002).
15. F. Provost, T. Fawcett, and R. Kohavi, in *Proceedings of the 15th International Conference on Machine Learning* 445 (1998).