

文章编号:1001-9081(2010)11-2917-04

基于小世界模型的流形学习算法

石陆魁¹, 杨庆新²

(1. 河北工业大学 计算机科学与软件学院, 天津 300401; 2. 天津工业大学 电气工程与自动化学院, 天津 300160)

(lkshi@eyou.com)

摘要:等距特征映射 (ISOMAP) 不仅计算复杂度很高, 而且缺乏对新样本的学习能力。基于标志点的 ISOMAP (L-ISOMAP) 通过只保持一些标志点之间的测地线距离有效地降低了复杂度, 然而标志点集的随机选择常常会导致较差的嵌入结果。为此, 提出了一种基于小世界模型的流形学习算法。根据小世界模型的原理, 该算法仅仅保持每个样本点与其 k 个最近邻和一些随机选择的远点之间的测地线距离, 采用最速梯度下降法优化来得到数据的低维表示。理论分析表明, 该算法的计算复杂度远远低于 ISOMAP 的复杂度。利用应力函数和剩余方差对 3 个算法进行了比较。实验结果表明, 从该算法得到的结果与从 ISOMAP 得到的结果相近, 且优于从 L-ISOMAP 得到的结果。同时, 该算法可以实现对新样本的学习, 对噪声也不太敏感。

关键词:流形学习; 等距特征映射; 最速梯度下降; 小世界模型; 标志点

中图分类号: TP181; TP301.6 **文献标志码:** A

Manifold learning algorithm based on the small world model

SHI Lu-kui¹, YANG Qing-xin²

(1. School of Computer Science and Engineering, Hebei University of Technology, Tianjin 300401, China)

(2. School of Electrical Engineering and Automation, Tianjin Polytechnic University, Tianjin 300160, China)

Abstract: Isometric Feature Mapping (ISOMAP) not only has high complexity but also can not learn new samples. L-ISOMAP has lower complexity by only preserving the geodesic distances between some landmark points. However, landmark point set randomly selected often leads to worse embedding results. A manifold learning algorithm based on the small world model was proposed, which only preserve the geodesic distances between each point and its k nearest neighbors as well as some distant points randomly chosen according to the small world model. The deepest gradient descent method was used to optimize the iterative process to obtain the low dimensional representation of data. The theoretic analysis demonstrates that the complexity of the proposed algorithm is far below one of ISOMAP. The stress function and the residual variance were used to compare the three methods. The experiments show that the results from the new method are close to those from ISOMAP and are superior to those from L-ISOMAP. Moreover, the algorithm can treat new data and is also not sensitive to noise.

Key words: manifold learning; Isometric Feature Mapping (ISOMAP); deepest gradient descent; small world model; landmark point

0 引言

在实际应用和研究中, 人们会不可避免地遇到大量的高维数据, 如全球气候模型、人类基因分布、文档词频数据、用户评分数据、Web 使用数据及多媒体数据等。研究发现高维数据的特性完全不同于低维数据, 数据维数的大幅度提高给随后的数据处理工作带来了前所未有的困难, 产生了所谓的“维数灾难”问题。因此, 对高维数据进行降维处理是非常必要的。流形学习作为一种数据降维技术, 能够有效地发现隐藏在低维数据中的非线性流形, 引起了广大研究人员的极大兴趣, 并得到了广泛应用^[1-9]。比较具有代表性的流形学习算法有等距特征映射 (Isometric Feature Mapping, ISOMAP)^[10]、局部线性嵌入 (Locally Linear Embedding, LLE)^[11]、拉普拉斯特征映射 (Laplacian Eigenmaps, LE)^[12]、局部切空间校正 (Local Tangent Space Alignment, LTSA)^[13]、多层自编码网络 (Multilayer Autoencoder)^[14-15] 等。

ISOMAP 保持了所有点之间的测地线距离, 通过求解一个完全 $N \times N$ 阶的特征值问题来达到数据的低维表示, 具有极高的复杂度, 不适合求解大规模高维数据。同时, ISOMAP

没有在原空间和嵌入空间之间建立一个映射或模型, 缺乏对新样本的学习能力, 不适合动态数据环境和在线学习。为了降低算法的复杂度, 基于标志点的 ISOMAP (ISOAMP based on Landmark points, L-ISOMAP)^[16] 仅仅保持了一些标志点之间的测地线距离。但是, 这些标志点通常随机地选择, 常常不能反映流形的真实的拓扑结构, 从而导致了较差的嵌入结果。

为了在降低计算复杂度的同时保证嵌入的质量, 提出了基于小世界模型的流形学习算法 (Manifold Learning Algorithm based on the Small World Model, SWM-MLA)。按照小世界模型的行为特征, 仅仅保持每个点与它的 k 个最近邻和一些远点之间的测地线距离, 通过迭代求解来得到数据的低维表示。

1 相关工作

1.1 ISOMAP 算法

ISOMAP 是一种将主成分分析法 (Principal Component Analysis, PCA) 和多维标度变换法 (MultiDimensional Scaling, MDS) 结合起来的全球优化算法, 其基本思想是用测地线距离代替欧氏距离来克服经典 MDS 的限制。ISOMAP 的关键是估计数据点之间的测地线距离。算法如下^{[10]2322}:

收稿日期: 2010-05-05; 修回日期: 2010-06-22。 **基金项目:** 天津市应用基础及前沿技术研究计划项目 (10JCZDJC16000)。

作者简介: 石陆魁 (1974-), 男, 河北邯郸人, 副教授, 博士, 主要研究方向: 机器学习、数据挖掘; 杨庆新 (1961-), 男, 河北唐山人, 教授, 博士生导师, 博士, 主要研究方向: 工程电磁场数值计算、电磁装置、电磁传感与检测。

1) 确定邻域,构造邻域图。对于输入空间 X 中的点 X_i 和 X_j , 设其欧氏距离为 $d_E(X_i, X_j)$, 或者是将距离每个点最近的 k 个点作为邻域, 或者是将每个点的一个指定半径 ε 内的所有点作为邻域。将邻域关系表示成一个以数据点为节点的赋权图, 即如果点 X_j 是点 X_i 的邻域点, 则将点 X_i 与点 X_j 连接起来, 边长赋值为 $d_E(X_i, X_j)$, 这样得到有权图 G 。

2) 估计测地线距离 $d_M(X_i, X_j)$ 。用点 X_i 和 X_j 之间的最短路径 $d_G(X_i, X_j)$ 近似流形 M 上的测地线距离 $d_M(X_i, X_j)$, 得到距离矩阵 $D_G = \{d_G(X_i, X_j)\}$ 。最短路径可用 Dijkstra 算法计算。

3) 应用经典的 MDS 算法构造能够最好地保持流形真实结构的低维嵌入。

1.2 L-ISOMAP 算法

ISOMAP 通过求解一个完全 $N \times N$ 阶特征值问题来得到高维数据的低维嵌入, 其计算复杂度为 $O(N^3)$, 同时也有较大的存储需求。为了降低算法的复杂度, L-ISOMAP 选择一些标志点并保持这些点之间的测地线距离。在 L-ISOMAP 中, 首先随机地选择一些标志点, 并计算这些标志点之间的测地线距离, 然后利用 MDS 得到标志点集的低维表示, 最后根据标志点的低维坐标和标志点与其他点之间的测地线距离来估计其他点的低维坐标。假设高维数据集有 N 个元素, 选择 n 个标志点 (n 远远小于 N), L-ISOMAP 求解特征值问题的复杂度为 $O(n^3)$ 。显然, 算法的复杂度大大降低。

1.3 小世界模型

小世界现象又称为六自由度分离, 首先由 Stanley Milgram 于 20 世纪 60 年代通过对社会现象的研究提出。到目前为止, 包括社会和自然科学在内的许多领域中, 小世界现象仍然是一个非常活跃的开放问题。小世界网络是一种以大量的短程连接为主、少量的长程连接为辅的稀疏网络结构。根据小世界原理, 一个网络的结构可以这样来设计, 网络中的每个节点仅仅与它的若干个最近邻点和少数几个远程节点相连接, 相连的远程节点用来产生短路径。研究表明, 这种网络可以产生具有较小半径的图, 而且任意两个节点之间都可以有一个较短的路径^[17]。

2 基于小世界模型的流形学习算法

2.1 基本原理

在 ISOMAP 中, 求解完全 $N \times N$ 阶的特征值问题导致了极高的计算复杂度。与 ISOMAP 不同, L-ISOMAP 仅仅保持一些标志点之间的测地线距离, 通过求解一个阶数小得多的 $n \times n$ 阶的特征值问题来达到数据的低维表示, 因此获得了较高的计算效率。然而, 由于标志点集的随机性, 经常导致较差的嵌入结果。那么, 如何在保证嵌入质量的前提下降低算法的计算复杂度呢?

这个问题可以通过结合流形的局部结构和全局结构来解决。在小世界模型中, 网络中的每个节点仅仅连接到它的若干个最近邻和少数几个远程节点, 从而结合了全连接网络和随机网络的优点。根据小世界模型的原理, 将流形的局部特征和全局特征结合起来, 只保持每个点与它的 k 个最近邻和一些随机选择的远点之间的测地线距离, 从而达到在保证嵌入质量的情况下提高算法计算效率的目的。

考虑一个包含 N 个样本的高维数据集 X , 设 d_{ij} 表示输入空间中的测地线距离, d_{ij}^* 表示嵌入空间中相应的欧氏距离, 降维的目的就是对所有数据点对尽可能地使 d_{ij}^* 与 d_{ij} 相匹配, 这可以通过最小化应力函数来实现。应力函数用来评判原

空间和嵌入空间中所有点对之间相似性或非相似性的误差, 克鲁斯克尔 (Kruskal) 给出了应力函数的一种表示形式:

$$ks = \sqrt{\frac{\sum_{ij} (d_{ij} - d_{ij}^*)^2}{\sum_{ij} d_{ij}^2}} \quad (1)$$

根据应力函数, 引入如下的能量函数:

$$E = \frac{1}{2} \lambda \sum_{ij} (d_{ij} - d_{ij}^*)^2 \quad (2)$$

其中: λ 是学习速率参数, 其作用是避免在迭代的过程中产生振荡现象。由于现在仅仅保持一些点之间的测地线距离, 因此, 能量函数可以修改为:

$$E' = \frac{1}{2} \lambda \sum_i \sum_{j \in \Omega} (d_{ij} - d_{ij}^*)^2 \quad (3)$$

其中 Ω 是包含 k 个最近邻和随机选择的一些远点的集合。式(3)可以重新表示为如下形式:

$$E' = \sum_i E'_i \quad (4)$$

其中:

$$E'_i = \frac{1}{2} \lambda \sum_{j \in \Omega} (d_{ij} - d_{ij}^*)^2 \quad (5)$$

为了得到最优解, 用最速梯度下降法来最小化能量函数。设 $\nabla E'_i$ 表示 y_i 的梯度, 则:

$$\Delta y_i = -\nabla E'_i = -\lambda \sum_{j \in \Omega} \frac{d_{ij} - d_{ij}^*}{d_{ij}^*} (y_j - y_i) \quad (6)$$

由此可知, 在算法迭代过程中, 第 $l+1$ 步的更新方法可以表示为:

$$y_i^{l+1} = y_i^l + \Delta y_i \quad (7)$$

式(7)给出了每一迭代步中更新低维坐标的方法。在该方法中, 仅仅考虑了若干最近邻和一些远点对嵌入结果的影响, 具体算法描述如下:

输入 包含 N 个数据点的高维数据集 X , 邻域大小 k_1 , 随机选择的远点数目 k_2 , 嵌入维数 d , 最大学习步数 L , 应力函数最小值 s 和学习速率 λ 。

输出 X 的低维表示 Y 。

1) 构造邻域图并计算测地线距离。

a) 计算所有点对之间的欧氏距离。

b) 根据欧氏距离为每个样本点选择 k_1 个最近邻和 k_2 个随机远点。

c) 用邻域图上的最短路径近似每个点与上一步骤中选择的近邻点和远点之间的测地线距离。

2) 随机初始化 Y 。

3) 设置 $l = 1$ 。

4) 计算应力函数值。若应力小于最小值 s , 则转步骤 7)。

5) 更新 Y , 对每个点执行下列步骤:

a) 对每个点利用式(6) 计算 Δy_i 。

b) 用式(7) 更新 y_i 。

6) 设置 $l = l + 1$ 。如果 l 小于最大迭代步数 L , 那么转 4)。

7) 返回低维坐标 Y 。

上述算法在迭代过程中有可能陷入局部极小。不过, 第 3 章的实验表明该算法通常会收敛到一个较好的结果。算法中有两个重要参数需要用户选择, 一个是邻域大小 k_1 , 另一个是远点数目 k_2 。邻域大小的选择既与数据的自身结构有关, 也与样本的稠密程度有关。而远点数目的多少将会影响算法的执行效率和低维嵌入质量。太少的远点会导致较差的嵌入质量, 而太多的远点又会降低算法的执行效率。根据小世界模型原理, 远点的数目应该远远小于样本数。

2.2 算法的复杂性分析

在上述算法中,计算花费主要包括两方面:估计测地线距离和迭代过程。测地线距离用邻域图上的最短路径近似,最短路径用 Dijkstra 算法计算,其复杂度为 $O(k_1(k_1 + k_2)N \log N)$, 通常 $k_1 + k_2 \ll N$ 。而 ISOMAP 中估计测地线距离的复杂度为 $O(k_1 N^2 \log N)$ 。迭代过程的复杂度为 $O(c(k_1 + k_2)N)$, c 是一个常数。此外,在 SWM-MLA 中,只需要存储一个 $(k_1 + k_2) \times N$ 阶的测地线距离矩阵,其空间复杂度为 $O((k_1 + k_2)N)$, 而 ISOMAP 的空间复杂度为 $O(N^2)$ 。显然,SWM-MLA 具有较低的时间和空间复杂度。

2.3 新样本的学习

上述算法通过保持每个点与它的 k 个最近邻和一些随机选择的远点之间的测地线距离来发现其低维坐标,各个点的迭代过程相对独立。对于新样本点,其低维坐标也可用同样的方法得到,即在训练集中选择它的 k 个最近邻和一些远点,然后保持它和这些点之间的测地线距离。设 x_{new} 表示新样本, y_{new} 表示新样本的低维坐标,则式(7)可以改写为:

$$y_{\text{new}}^{i+1} = y_{\text{new}}^i + \Delta y_{\text{new}} \quad (8)$$

算法的具体步骤与映射训练集的算法相似。

3 实验

为了测试 SWM-MLA 算法的性能,在 4 个数据集上执行该算法。实验包括 3 个部分:不含噪声的数据集上的性能测试、含噪声的数据集上的性能测试和算法学习新样本的能力测试。在实验中,输入空间使用测地线距离,嵌入空间使用欧氏距离。所有程序均用 Matlab 实现。

为了将 SWM-MLA 与 ISOMAP 和 L-ISOMAP 进行比较,利用应力函数和剩余方差作为性能指标进行评价。应力函数如式(1)所示。剩余方差在 ISOMAP 中用来确定低维维数^{[10]2321},其定义如下:

$$r = 1 - \rho_{D_X D_Y}^2 \quad (9)$$

其中: ρ 表示相关系数, D_X 和 D_Y 为行向量,它们的元素分别表示两个空间中的测地线距离和欧氏距离,在下面的实验中用 r 表示。应力和剩余方差值越小,说明嵌入质量越好。

3.1 不含噪声的数据集上的性能测试

在 3 个不含噪声的数据集上进行性能测试,分别是包含 1000 个样本的 Swiss roll 数据集^{[10]2321} (图 1(a)), 包含 1000 个样本的 S-curve 数据集^{[11]2323} (图 1(b)) 和合成 face images 数据集^{[10]2320}。合成 face images 数据集包含 698 个不同姿势和光照条件的图像样本,每个样本由 4096 个像素组成。

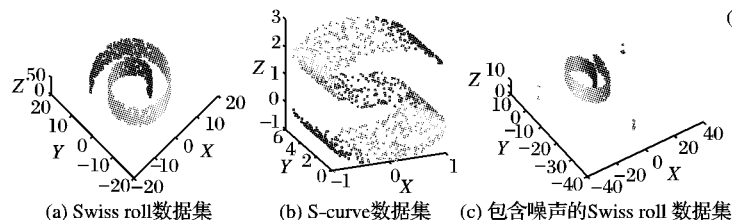


图1 3个包含1000个样本的数据集

首先,在 3 个数据集上分别执行 ISOMAP、L-ISOMAP 和 SWM-MLA 3 个算法,然后根据执行结果计算相应的应力和剩余方差,结果如表 1~3 所示。其中,ISOMAP 在每个数据集上执行 1 次。L-ISOMAP 在每个数据集上分 4 组执行,标志点数分别为 5, 10, 15 和 20,每组都执行 100 次,每次执行都随机地选择标志点集,计算 100 次执行的最小值、最大值、均值和方差。SWM-MLA 在每个数据集上执行 4 次,远点数 k_2 分别为 5, 10, 15 和 20。对于同一数据集 3 个算法中邻域大小 k_1 一

样,其中对于 Swiss roll 数据集 $k_1 = 7$, S-curve 数据集 $k_1 = 20$, 合成 image faces 数据集 $k_1 = 7$ 。

如表 1~3 所示,在 3 个数据集上,从 SWM-MLA 得到的应力值总是远小于从 L-ISOMAP 得到的结果,而与从 ISOMAP 得到的结果相近甚至更小。从 SWM-MLA 得到的剩余方差与从 ISOMAP 得到的结果相近,小于或接近从 L-ISOMAP 得到的最小值,而远小于其均值。因此,SWM-MLA 可以获得比 L-ISOMAP 更好的嵌入结果,而具有比 ISOMAP 低得多的计算复杂度。对于 L-ISOMAP 算法,随机选择标志点集是造成嵌入结果较差的主要原因,如果选择了一个好的标志点集,就可以得到一个较好的结果。同时,实验还表明,对于 SWM-MLA 算法,增加远点对嵌入结果的改善并不明显,因此,在实际中并不需要选择太多的远点。

3.2 含噪声的数据集上的性能测试

为了检验 SWM-MLA 算法的鲁棒性,构造了一个包含 1000 个样本点的含噪声的 Swiss roll 数据集,如图 1(c) 所示,在该数据集上分别执行上述 3 个算法,根据执行结果计算相应的应力和剩余方差,结果如表 4 所示。其中,ISOMAP 执行 1 次。L-ISOMAP 分 4 组执行,标志点数分别为 5, 10, 15 和 20,每组执行 100 次,计算 100 次执行的最小值、最大值、均值和方差。SWM-MLA 执行 4 次,远点数 k_2 分别是 5, 10, 15 和 20。对于 3 种方法,邻域大小 k_1 都是 8。实验结果表明,ISOMAP 比 L-ISOMAP 和 SWM-MLA 对噪声更敏感,而 L-ISOMAP 比 SWM-MLA 对噪声更敏感。这主要是由于 SWM-MLA 在选择远点和 L-ISOMAP 在选择标志点时会过滤掉一部分噪声。

3.3 新样本的学习能力测试

为了测试 SWM-MLA 对新样本的学习能力,从 Swiss roll 数据集中选择 2000 个点作为样本集。将该数据集分为两个包含 1000 个样本的子集,一个作为训练集,如图 1(a) 所示,另一个作为测试集,如图 2(a) 所示。首先,在训练集上执行 SWM-MLA 得到训练集的低维表示,接着根据训练集的低维坐标和式(8)学习测试集中每个样本的低维坐标,其中邻域大小 $k_1 = 7$,远点数 $k_2 = 10$,结果如图 2(b) 所示。实验表明,SWM-MLA 能够较好地实现对新样本的学习。

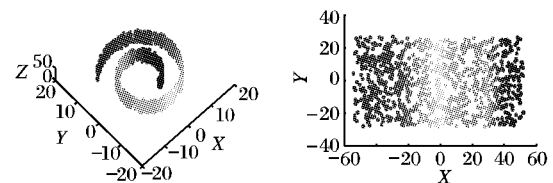


图2 Swiss roll 数据集上的执行结果

4 结语

针对 ISOMAP 和 L-ISOMAP 算法的缺点,提出了一种基于小世界模型的流形学习算法。根据小世界行为模式,只保持了每个点与它的若干个近邻点和一些远点之间的测地线距离,克服了 ISOMAP 计算复杂度高的缺点。通过近邻点和远点将流形的局部特征和全局特征结合起来,克服了 L-ISOMAP 随机选择标志点集造成的嵌入结果较差的问题。实验结果表明,SWM-MLA 不但得到了与 ISOMAP 相近的嵌入结果,而且具有较高的执行效率。同时,SWM-MLA 可以有效地将新样本映射到低维空间中,可以适用于动态数据环境中的降维处理,并且对噪声也不太敏感。下一步工作是探讨远点数对嵌入结果的影响,并在实际应用中检验 SWM-MLA 算法的性能。

表1 3个算法在 Swiss roll 数据集上的执行结果

ISOMAP		SWM-MLA				L-ISOMAP							
k_s	r	远点数 k_2	k_s	r	标志 点数	k_s				r			
						最小值	最大值	均值	方差	最小值	最大值	均值	方差
0.0256	0.0027	5	0.0321	0.0043	5	0.6918	0.8425	0.7568	0.00120	0.0046	0.2797	0.0450	0.00310
		10	0.0267	0.0030	10	0.6607	0.7514	0.7001	0.00038	0.0031	0.1120	0.0192	0.00025
		15	0.0780	0.0272	15	0.6163	0.7212	0.6634	0.00042	0.0035	0.0448	0.0126	0.00010
		20	0.0232	0.0023	20	0.6014	0.7027	0.6376	0.00026	0.0030	0.0392	0.0097	0.00003

表2 3个算法在 S-curve 数据集上的执行结果

ISOMAP		SWM-MLA				L-ISOMAP							
k_s	r	远点数 k_2	k_s	r	标志 点数	k_s				r			
						最小值	最大值	均值	方差	最小值	最大值	均值	方差
0.0066	0.0002	5	0.0220	0.0021	5	0.7028	0.8281	0.7589	0.00088	0.0008	0.1584	0.0342	0.00350
		10	0.0133	0.0008	10	0.6638	0.7519	0.6990	0.00046	0.0005	0.0675	0.0130	0.00023
		15	0.0124	0.0007	15	0.6269	0.7037	0.6594	0.00027	0.0003	0.0943	0.0106	0.00062
		20	0.0119	0.0007	20	0.5986	0.6647	0.6309	0.00031	0.0003	0.0309	0.0053	0.00068

表3 3个算法在合成 face images 数据集上的执行结果

ISOMAP		SWM-MLA				L-ISOMAP							
k_s	r	远点数 k_2	k_s	r	标志 点数	k_s				r			
						最小值	最大值	均值	方差	最小值	最大值	均值	方差
0.0668	0.0235	5	0.0676	0.0280	5	0.6869	0.8598	0.7431	0.00056	0.0648	0.6564	0.1521	0.00390
		10	0.0657	0.0266	10	0.6427	0.7515	0.6753	0.00031	0.0366	0.1803	0.0670	0.00065
		15	0.0633	0.0247	15	0.5940	0.6989	0.6372	0.00016	0.0332	0.1192	0.0518	0.00032
		20	0.0592	0.0214	20	0.5654	0.6477	0.6049	0.00015	0.0293	0.0873	0.0426	0.00079

表4 3个算法在含噪声的 Swiss roll 数据集上的执行结果

ISOMAP		SWM-MLA				L-ISOMAP							
k_s	r	远点数 k_2	k_s	r	标志 点数	k_s				r			
						最小值	最大值	均值	方差	最小值	最大值	均值	方差
0.0471	0.0053	5	0.0296	0.0028	5	0.6791	0.8661	0.7560	0.00120	0.0042	0.0756	0.0127	0.00047
		10	0.0295	0.0028	10	0.6275	0.7842	0.6941	0.00069	0.0032	0.0276	0.0090	0.00034
		15	0.0280	0.0025	15	0.6091	0.7191	0.6614	0.00061	0.0033	0.0275	0.0095	0.00021
		20	0.0421	0.0051	20	0.5817	0.6822	0.6309	0.00055	0.0033	0.0209	0.0098	0.00025

参考文献:

- [1] VLACHOS M, DOMENICONI C, GUNOPULOS D, *et al.* Non-linear dimensionality reduction techniques for classification and visualization [C]// Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM Press, 2002: 645-651.
- [2] JENKINS O C, MATARIC M J. A spatio-temporal extension to ISOMAP nonlinear dimension reduction [C]// Proceedings of the International Conference on Machine Learning. New York: ACM Press, 2004: 441-448.
- [3] 翁时锋, 张长水, 张学工. 非线性降维在高维医学数据处理中的应用[J]. 清华大学学报, 2004, 44(4): 485-488.
- [4] ZHANG J, LI S, WANG J. Manifold learning and applications in recognition [C]// Intelligent Multimedia Processing with Soft Computing. Berlin: Springer-Verlag, 2004: 281-300.
- [5] 谭璐, 易东云, 吴翊, 等. 基于非线性降维的图像识别[J]. 计算机工程, 2005, 31(13): 54-55.
- [6] ZENG X, LUO S, WANG J. Auto-associative neural network system for recognition [C]// Proceedings of the 6th International Conference on Machine Learning and Cybernetics. Washington, DC: IEEE Press, 2007: 2885-2890.
- [7] 尹峻松, 肖健, 周宗潭, 等. 非线性流形学习方法的分析与应用[J]. 自然科学进展, 2007, 17(8): 1015-1025.
- [8] PLESS R, SOUVENIR R. A survey of manifold learning for images [J]. IPSJ Transactions on Computer Vision and Applications, 2009, 1: 83-94.
- [9] 王自强, 钱旭. 基于流形学习和 SVM 的 Web 文档分类算法[J]. 计算机工程, 2009, 35(15): 38-40.
- [10] TENENBAUM J B, SILVA V D, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(5500): 2319-2323.
- [11] ROWEIS S T, SAUL L K. Nonlinear dimensionality reduction by locally linear embedding [J]. Science, 2000, 290(5500): 2323-2326.
- [12] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation [J]. Neural Computation, 2003, 15(6): 1373-1396.
- [13] ZHANG Z, ZHA H. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment [J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313-338.
- [14] HINTON G E, SALAKHUTDINOV R R. Reducing the dimensionality of data with neural networks [J]. Science, 2006, 313(5786): 504-507.
- [15] ERHAN D, MANZAGOL P A, BENGIO Y, *et al.* The difficulty of training deep architectures and the effect of unsupervised pre-training [C]// Proceedings of 20th International Conference on Artificial Intelligence and Statistics. Clearwater: JMLR, 2009: 153-160.
- [16] SILVA V D, TENENBAUM J B. Global versus local methods in nonlinear dimensionality reduction [C]// Proceedings of the 6th Annual Conference on Neural Information Processing Systems. Cambridge: MIT Press, 2003: 705-712.
- [17] NEWMAN M E J. Models of the small world [J]. Journal of Statistical Physics, 2000, 101(3/4): 819-841.