

局部测地距离估计的 Hessian 局部线性嵌入

文贵华¹, 江丽君², 文 军³

(1. 华南理工大学 计算机科学与工程学院, 广东 广州 510641; 2 华南理工大学 电子材料科学与工程系, 广东 广州 510641; 3 湖北民族学院 理学院, 湖北 恩施 445000)

摘要:为处理极度弯曲的数据流形,提出了基于局部测地距离估计的 Hessian 局部线性嵌入算法. 算法采用 Hessian 局部线性嵌入 (HLL E) 的概念框架, 采用局部估计的测地距离而不是欧氏距离来确定每个点的邻域, 从而减少数据流形弯曲对邻域选择的影响. 算法可认为是全局和局部方法的综合, 在性能上不仅比 HLL E 显著提高, 有更强的鲁棒性, 而且时间增加不明显. 标准数据集上的实验结果验证了所提方法的有效性.

关键词:流形学习; Hessian 变换; 局部线性嵌入; 测地距离

中图分类号: TP391 **文献标识码:** A **文章编号:** 1673-4785 (2008) 05-0429-07

Using locally estimated geodesic distances to improve Hessian local linear embedding

WEN Gui-hua¹, JIANG Li-jun², WEN Jun³

(1. School of Computer Science and Engineering, South China University of Technology, Guangzhou 510641, China; 2 Department of Electronic Material Science and Engineering, South China University of Technology, Guangzhou 510641, China; 3. School of Mathematical Science, Hubei Institute for Nationalities, Enshi 445000, China)

Abstract: To deal with highly curved data manifolds, the Hessian locally linear embedding (HLL E) algorithm was modified based on a locally estimated geodesic distance. It used the general conceptual framework of HLL E to guarantee correct setting of local isometry to an open connected subset. It employs the locally estimated geodesic distance instead of the Euclidean distance to determine the neighborhood of any point, so that it reduces the distorting influence of curvature of the data manifold on determining the neighborhood. This approach can be regarded as the integration of a local method and a global method, so that it has better performance and stability than HLL E, with only a slight increase in computational time. Experiments conducted on benchmark data sets verified efficiency of the proposed approach.

Keywords: manifold learning; Hessian transformation; locally linear embedding; geodesic distance

很多高维数据如遥感、气候等常常分布于较低维的流形上, 自从国际著名期刊《Science》在 2000 年发表 2 种最有代表性的方法 ISOMAP^[1]和 LLE^[2]以来, 寻找描述这样低维流形的参数空间就成为最近的研究热点. ISOMAP 在降维过程中通过计算点对之间的测地距离, 并采用 MDS 方法来获取全局最优的几何结构, 获得了较好的效果, 目前已发展了很

多改进算法. LLE 在降维嵌入过程中保持局部的几何结构不变, 并能避免局部极小, 最终获得一个全局的低维嵌入系统, 效果也很好. 目前发展的改进算法包括利用拉普拉斯 (Laplacian)^[3]、赫森 (Hessian) 变换改进的算法 HLL E^[4]、利用数据分类信息改进的监督 LLE^[5]、增量式 LLE^[6]、利用 Fisher 改进的 LLE^[7]、以及利用 PCA 改进的鲁棒性 LLE^[8]等. 性能上, HLL E 是对 LLE 的较大改进, 它在某些情况下超越了 ISOMAP 的能力. ISOMAP 的基本假设是全局等距映射和凸的参数空间, 这在很多情况下难以满

收稿日期: 2007-11-25.

基金项目: 广东省科技攻关资助项目 (2007B030803006); 湖北省科技攻关资助项目 (2005AA101C17).

通信作者: 文贵华. E-mail: crghwen@scut.edu.cn

足. 而 HLLC 只要求局部等距映射和开的连通参数空间, 因而应用范围更宽. 但是 HLLC 需要保持邻域的线性化, 当数据流形比较弯曲时, 难以满足, 不可避免地面临短路问题, 因此需要邻域优化. 目前有 3 类与邻域相关的工作, 第 1 类是构造连通的邻域图^[9-10]; 第 2 类采用新的测度来选择邻域点^[11-15], 例如监督 LLE 利用分类信息构造新的测度^[15], 但需要分类信息, 替代策略是采用自动聚类确定分类信息^[13]. 同时图代数还可以用来定义新的测度并优化邻域^[14-15], 以及利用测地距离确定邻域^[11, 18]等; 第 3 类是研究邻域的大小问题, 包括如何选择合适的邻域大小^[16], 以及适用于非均匀分布流形的邻域大小自适应确定算法^[17]. 本文的工作属于第 2 类, 采用局部估计的测地距离来确定邻域, 以减少流形弯曲对邻域选择的影响. 此前华南理工大学曾采用测地距离来确定 HLLC 的邻域^[18], Claudio Varini 采用测地距离来确定 LLE 的邻域^[11], 但它们都需要计算所有点对之间的测地距离, 然后用于确定邻域, 增加的时间复杂度太高, 接近 $O(N^3)$. 为此本文改变策略, 只估计每个点与其局部区域内的点之间的测地距离, 并用之确定 HLLC 的邻域, 不仅增加的时间不明显, 而且效果仍然很好. 最后的实验表明此改进算法在降维性能上比 ISOMAP、HLLC 和 LLE 都有效, 更适合于处理弯曲的数据流形.

1 Hessian 局部线性嵌入

假定有一个参数空间 $C \subset R^d$ 和一个光滑映射 $f: C \rightarrow R^n$, 其中嵌入空间 R^n 满足 $n > d$, 则称 $M = f(C)$ 为流形. 流形学习的目的是根据观察数据确定参数空间 $C \subset R^d$. ISOMAP 采用等距嵌入 (isometric embedding) 和等角嵌入 (conformal embedding) 来实现流形学习. 其基本假设是: 1) 等距, 即测地距离在等距嵌入映射下是不变的, 流形上任意点之间的测地距离在等距嵌入变换下获得的欧氏空间中仍然保持 $G(x, y) = \|x - y\|$, $x \in M, y \in M$; 2) 凸性, 即参数空间 $C \subset R^d$ 是凸的. 对任意 $x, y \in C$, 线段 $\{(1-t)x + ty \mid t \in (0, 1)\}$ 仍然属于 C .

ISOMAP 算法利用等距嵌入的这种不变性, 在没有任何关于观察数据测度知识的基础上构造测地

距离, 方法是假定当两点非常近时, 测地距离等于欧氏距离, 而对较远的点对之间的测地距离则根据近邻点之间测地距离的累加实现. 当观察数据集足够密集且内在的参数空间是凸的, 则 ISOMAP 能够成功地获得参数空间. ISOMAP 面临的问题是, 在很多情况下参数空间并不是凸的, 此时它得不到正确的结果.

HLLC 采用局部线性方法实现流形学习. 它只要求局部等距映射的子集是开的且连通, 而不必是凸的. 其理论依据来源于流形切空间上的 Hessian 变换. 假定流形 $M \subset R^n$ 是光滑的, 其任意点 $m \in M$ 都有切空间 $T_m(M)$, 在此空间引入欧氏空间的内积就可以建立局部坐标系, 且有原点 $O \in T_m(M)$. 设 N_m 是 $m \in M$ 的邻域, 对任意 $m \in N_m$ 都有惟一的最近点 $v \in T_m(M)$ 使得映射 $m \rightarrow v$ 是光滑的, 因此 N_m 具有局部坐标系, 记为 $x_1^{(\tan, m)}, \dots, x_d^{(\tan, m)}$. 设 $F: M \rightarrow R$ 在 m 附近是 C^2 光滑的, $G: U \rightarrow R, U \subset R^d$ 是零点 O 的一个邻域. 令 $g(x) = f(m + x)$, 因 $m \rightarrow x$ 是光滑的, g 是 C^2 光滑的, 则 f 的 Hessian 变换为

$$(H_f^{\tan}(m))_{ij} = \frac{\partial}{\partial x_i} \frac{\partial}{\partial x_j} g(x) \Big|_{x=0}$$

$$\text{二次型 } H(f) = \sum_M H_f^{\tan}(m) \sum_F d_m^2 \text{ 定义了 } f: M$$

R 在 M 上的平均弯曲率, 其中 $\sum_F \cdot^2$ 是 Frobenius 范数. 可以证明若参数空间 $C \subset R^d$ 是开的连通子集, 则 $H(f)$ 有一个 $(d+1)$ 维的零空间 (null space), 参数空间能够通过计算此零空间的一个合适的基坐标来发现, 这就是 HLLC 的核心. 由此可以发现: 1) 框架上 HLLC 与 LE 和 LLE 一致, 不同的是用 Hessian 变换取代了 LE 的 Laplacian 算子, 而 LLE 是 LE 理论框架下的一种经验实现; 2) HLLC 需要对每个数据点计算 $n \times n$ 次偏导数, 当观察数据的维非常高时, 计算量不小; 3) 每个点的切空间要求其邻域是线性的, 当邻域高度弯曲时, 极易面临短路威胁, 这是 HLLC 需要克服的另一个主要问题.

2 改进算法

HLLC 对极度弯曲的数据流形难以得到正确的结果, 原因是被当作线性化的局部区域实际上包含

了很多弯曲的曲面,从而容易导致短路问题.如图 1 所示,HLLE采用欧氏距离确定点 x 的 5 邻域就导致短路问题,其邻域为虚线园内的点,点的类别分别为 {square, square, square, circle, circle},这显然发生了短路现象,因为类别为 square 的点并不是点 x 在流形上的最近邻,由此确定的邻域扭曲了数据真实的邻域结构,进而产生不正确的低维嵌入,尤其对稀疏或含噪音的数据集更是如此.相反,用测地距离确定的邻域则为虚曲线内的点,点的类别为 {circle, circle, circle, circle, circle},这是正确的结果.因此只要测地距离的估计是正确的,那么确定的邻域就能导致正确的低维嵌入.

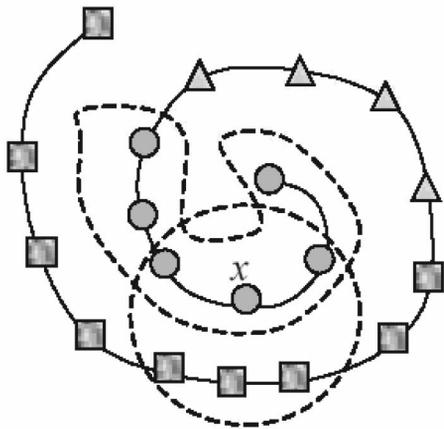


图 1 欧氏距离和测地距离确定的不同邻域

Fig 1 Different neighborhoods using Euclidean distance and geodesic distance

以前采用 ISMAP 计算测地距离的方法计算输入数据之间的测地距离,并用之确定每个数据点的邻域,然后用 HLLE 方法计算每个点的低维嵌入^[15], ISMAP 计算测地距离主要包括 2 步:

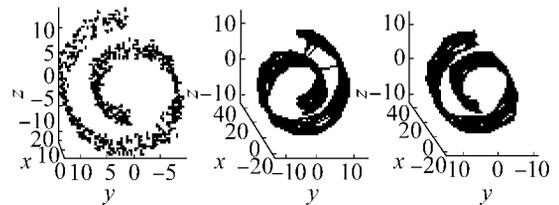
- 1)根据观察数据 x 和邻域大小 k 构造邻域权重图 $G = (V, E)$. V 对应于 X 中的观察数据, E 为连接 2 点的边的集合, $(x_i, x_j) \in E$ 若 x_i 是 x_j 的 k 最近邻, x_i 与 x_j 之间的距离为欧氏距离 $d_G(x_i, x_j)$.
- 2)通过求 G 上任意两点之间的最短距离来估计流形上所有点对之间的测地距离 $d_G(x_i, x_j)$.

初始化

$$d_G(x_i, x_j) = \begin{cases} d_E(x_i, x_j), & \text{if } (x_i, x_j) \in E \\ \infty, & \text{else} \end{cases}$$

利用所有 t 迭代计算所有的 $d_G(x_i, x_j) = \min\{d_G(x_i, x_j), d_G(x_i, x_t) + d_G(x_t, x_j)\}$.

这样计算测地距离的时间复杂度近似为 $O(|X|^3)$,复杂度太高.从图 1 可以看出,用测地距离确定 x 点的邻域,不必要计算 x 到所有点的测地距离,只需要计算 x 附近的点就可以了.具体方法是首先确定局部区域大小 L , 然后以 x 为中心,选择其 L 个最近邻点构成该点的局部区域,进而在此局部区域内估计 x 到其他 L 个点之间的测地距离.最后根据估计的测地距离来选择最小的 k 个最近邻作为点 x 的邻域,由此形成的邻域可显著减少短路边,如图 2 所示,其中 $k = 12, L = m \times k = 5 \times 12$,而用于计算测地距离的邻域大小 $k_g = 6$ 图 2(a)的样本是从 Swiss roll surface 随机采样 600 个点,但叠加了均值为 0 和方差为 0.3 的高斯噪音后的含噪音数据集,图 2(b)是利用 Euclidean 距离确定的邻域图,其中明显出现了短路边,它们将靠得很近的 2 个不同曲面连接了起来,而图 2(c)是用局部估计的测地距离确定的邻域,其中没有任何短路边出现,这说明了利用局部估计的测地距离改进 HLLE 将取得更好的嵌入结果.将据此改进的 HLLE 记为 GL-HLLE,以区别于利用全局估计的测地距离改进的 HLLE,记为 GA-HLLE 为保持算法的清晰性,给出完整的 GL-HLLE 算法如下:



(a) 噪音数据集 (b) Euclidean (c) 局部估计

图 2 利用 Euclidean 距离和局部估计的测地距离确定的不同邻域图

Fig 2 Different neighborhood graph using Euclidean distance and locally estimated geodesic distance

算法 GL-HLEE 算法 (X, k, k_g, d, m) :

/* 输入中 X 是高维观察数据, k 是用于确定切空间的局部邻域大小, d 是低维参数空间的维数,而 k_g 是用于估计测地距离的邻域大小, m 确定局部区域的参数. 输出是低维参数空间 W */

1) 首先计算每个点的 $m \times k$ 个最近邻构成该点的局部区域, 然后以 k_g 为邻域大小估计 x 到局部区域内其他任意点之间的测地距离, 然后根据估计的测地距离来选择最小的 k 个最近邻作为点 x 的邻域, 并将每个点的邻域表达为中心化的行向量 $k \times n$ 矩阵 M^i ;

2) 采用奇异值分解每个邻域矩阵 M^i , 将其正交向量 V 的前 d 个分量作为其切空间;

3) 求切空间的 Hessian 矩阵. 当 $d=2$ 时, 根据切空间中的点形成如下矩阵 $X^i = [1 \quad V_{\cdot,1} \quad V_{\cdot,2} \quad (V_{\cdot,1})^2 \quad (V_{\cdot,2})^2 \quad (V_{\cdot,1} \times V_{\cdot,2})]$, 其中 $V_{\cdot,1}$ 表示切空间中所有点的第一维的值, 对 $d>2$ 采用相同的方法创建 $1+d+d(d+1)/2$ 列的矩阵, 然后用 Gram-Schmidt 正交化 X^i 产生新的正交矩阵 \bar{X}^i , 将其转置后取最后的 $d(d+1)/2$ 列构成 Hessian 矩阵 H^i ;

4) 首先构造二次型 $H_{ij} = \sum_l \sum_r ((H^l)_{xi} (H^l)_{xj})$, 对 H 进行特征分析, 获取其 $d+1$ 个最小特征值对应的 $d+1$ 维子空间, 第 1 个特征值 0 对应于常函数, 接下来的 d 个特征向量就构成 d 维空间, 对其选择一个正交基, 变换就可获得要恢复的参数空间 w .

算法 GL-HLLE 中的 1) 利用局部估计的测地距离确定每个点的邻域, 余下的步骤与 HLLE 相同, 因此复杂的数学推导和更详细的算法描述见 HLLE 原文^[4]. GL-HLLE 增加了估计局部测地距离的时间 $O(N)$, $N = |X|$, 但与 HLLE 一样都只需要 N 个 $k \times k$ 稀疏的特征问题计算, 而 ISOMAP 需要一个 $N \times N$ 密集的特征问题求解. 如果 N 很大时, 显然 GL-HLLE 比 ISOMAP 有优越性, 而从后面的实验结果看, GL-HLLE 的降维性能比 HLLE、ISOMAP 和 LLE 都要好.

3 实验分析

实验比较 GL-HLLE 与 GA-HLLE、HLLE、LLE 和 ISOMA 方法的性能和时间复杂度. 几种方法均采用 Matlab 编程实现, 其中 HLLE、LLE 和 ISOMAP 采用原作者提供的 Matlab 代码, GA-HLLE 和 GL-HLLE 由作者编程实现. 实验中的参数设置为线性重构的邻域 $k=12$, 局部区域大小为 $5 \times k$, 而估计测地距离的邻域参数取 6. 实验数据采用 HLLE、

LLE、ISOMAP 及其许多改进算法都采用的标准测试数据, 包括 Swiss roll surface、S-curve、3D clusters、Torioidal Helix 和 Twin Peaks, 在每个数据集上均采样 600 个点构造实验数据.

3.1 性能分析

GL-HLLE 等 5 种方法在几种标准数据集上的运行结果如图 3~7 所示. 图 3 中的实验数据是采用 HLLE 的方法从 Swiss roll surface 上采样一个数据规模为 600 个点的长方形但从其中心移去一个小的长方形而形成的非凸数据集, 这也是 HLLE 实验中采用的实验数据^[4]. 不难发现在采样的这个数据集上 LLE 和 HLLE 不能够完美地将数据嵌入在二维空间. ISOMAP 将去除的区域强烈膨胀, 并扭曲其余的数据点. 而 GA-HLLE 和 GL-HLLE 能够较完美地将数据嵌入在二维空间, 其中心移去的一个小长方形也能在嵌入的二维空间中正确体现, 因此在这个数据集上 GL-HLLE 能够解决 ISOMAP 等其他方法不能解决的问题. 同时在剩余的 4 个数据集上, GL-HLLE 和 GA-HLLE 都胜过 HLLE 和 LLE, 而 GL-HLLE 和 GA-HLLE 的效果无明显区别. 在有的数据集如图 5 上它们的性能略差于 ISOMAP, 而在另外一些数据集如图 3 和图 7 上它们的性能则优于 ISOMAP. 因此从整体性能上 GL-HLLE 和 GA-HLLE 是相对较好的方法, 特别是它们都比 HLLE 好.

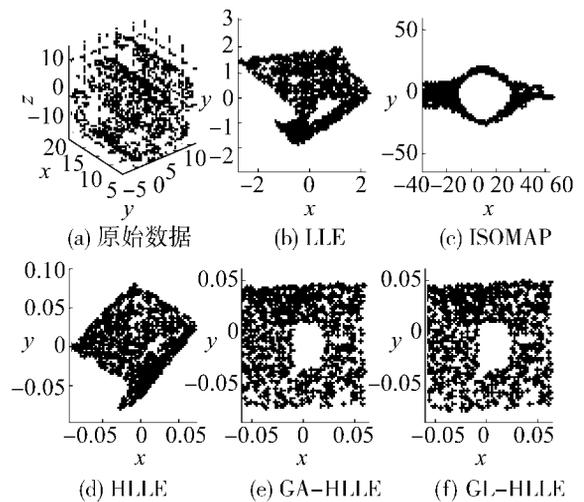


图 3 几种算法在数据集 Swiss roll surface 上的结果
Fig 3 Embedding results of the compared approaches on Swiss roll surface data

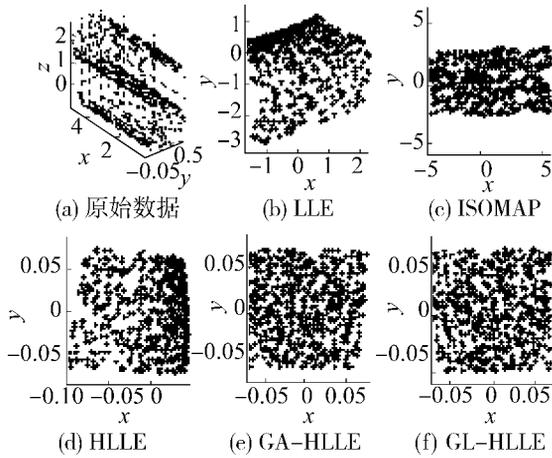


图 4 几种算法在数据集 S-curve上的结果

Fig 4 Embedding results of the compared approaches on S-curve data

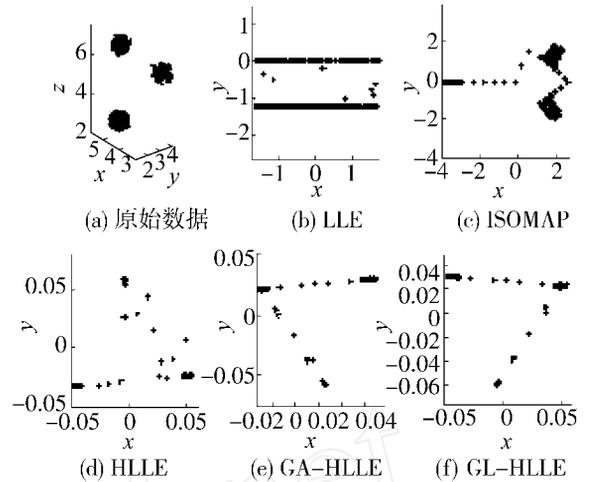


图 6 几种算法在数据集 3D clusters上的结果

Fig 6 Embedding results of the compared approaches on 3D clusters data

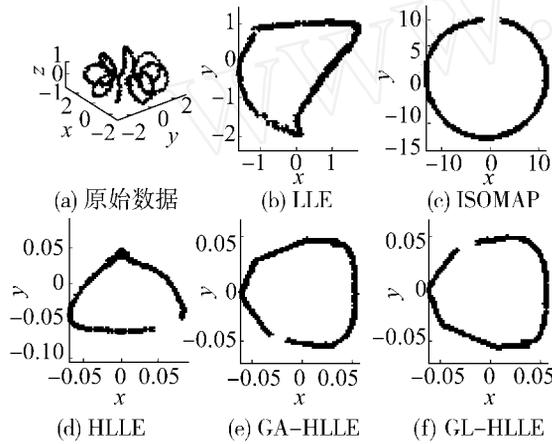


图 5 几种算法在数据集 Toroidal Helix上的结果

Fig 3 Embedding results of the compared approaches on Toroidal Helix data

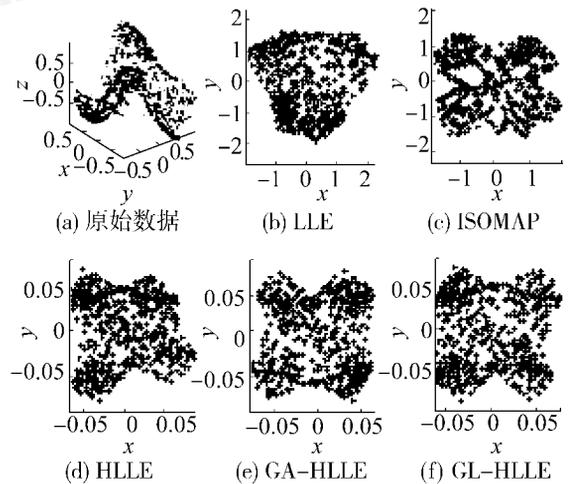


图 7 几种算法在数据集 Twin Peaks上的结果

Fig 7 Embedding results of the compared approaches on Twin Peaks data

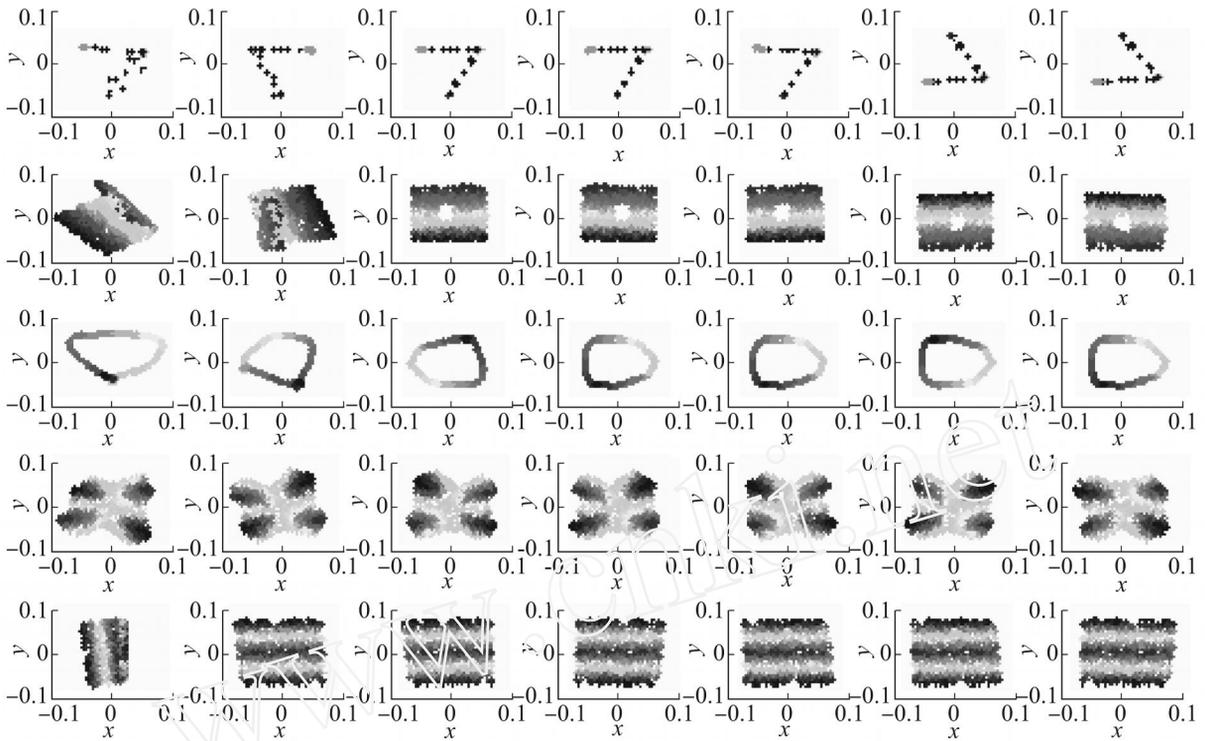
3.2 局部区域参数的选择

相对于 GA-HLLC, GL-HLLC引入了确定局部区域的参数 m , 与估计测地距离的局部区域大小 L 的关系是 $L = m \times k$, 其中 k 为用于线性重构造的邻域. 幸运的是 m 的取值一般小于 10 图 8 的结果表明 $m = 3$ 时 GL-HLLC在几个数据集上均已经取得较好的效果, $m > 3$ 时增加的效果不明显, 这说明 m 取很小的值就可获得满意的效果, 这样 m 参数的选取就非常容易, 同时在局部区域 L 内估计测地距离也自然只需要少量的时间.

3.3 运行时间比较

从 Swiss Roll Surface 上依次采样 500、1 000、

1 500、2 000、2 500个点的 5类规模的数据样本, 每类规模的样本随机采样 5次, 记录 LLE、HLLC、ISOMAP、GA-HLLC和 GL-HLLC5种方法分别运行这 5个样本的平均时间作为该规模的时间, 则 5种方法在 5类规模数据上的平均时间如表 1所示. 不难发现, GL-HLLC和 HLLC的运行时间很接近, 但是 GA-HLLC的运行时间显著高于 GL-HLLC, 是几种方法中运行最慢的, 这说明从嵌入性能上 GL-HLLC和 GA-HLLC相当, 但在时间上 GL-HLLC却有显著优势, 特别是它在嵌入性能上比 HLLC有显著优越, 但时间只有很细微的增加.



(a) localsize = 1 \times 12 (b) localsize = 2 \times 12 (c) localsize = 3 \times 12 (d) localsize = 4 \times 12 (e) localsize = 5 \times 12 (f) localsize = 7 \times 12 (g) localsize = 10 \times 12

图 8 嵌入效果与局部区域参数的关系

Fig 8 The relationship of Embedding results and local region parameters

表 1 5种方法在 5种样本规模上的平均运行时间比较

Table1 Average running time of the compared approaches on five kinds of data sets /s

算法	样本规模				
	500	1000	1500	2000	2500
LLE	0.587 6	2.540 6	2.562 8	9.368 6	7.174 8
HLLE	1.653 0	18.237 6	18.234 2	72.715 6	55.537 2
GL - HLLE	3.540 18	59.681 0	60.081 6	245.822 0	182.172 0
GA - HLLE	6.3121 6	139.674 6	141.837 4	584.624 8	442.328 0
ISOMAP	10.250 1	279.850 0	283.041 7	1 151.033 3	862.050 0

4 结束语

提出了基于局部测地距离估计的 Hessian局部线性嵌入算法 GL-HLLE 算法采用 Hessian局部线性嵌入的概念框架,但采用局部估计的测地距离而不是欧氏距离来确定局部邻域,从而减少数据流形弯曲对邻域选择的影响.它与基于全局测地距离估计的算法相比,性能上没有明显差别,但时间复杂度上却显著减少. GL-HLLE在性能上比 HLLE明显增强,但增加的运行时间不明显.算法可认为是全局和局部方法的综合,比现有方法如 HLLE、LLE和 ISO-

MAP等有更强大的鲁棒性.标准数据集上的实验结果验证了所提方法的有效性. GL-HLLE的思想还可用于改进.其他基于邻域的流形学习算法包括 LLE和 ISOMAP. GL-HLLE的缺陷是与 HLLE一样,对观察数据的维数敏感,对文本之类高达上万维的观察数据不合适,需要采取进一步的措施.

参考文献:

[1] TENENBAUM J B, De SILVA V, LANGFORD J C. A global geometric framework for nonlinear dimensionality reduction [J]. Science, 2000, 290(12): 2319-2323.

- [2] ROWE S T, SAUL L K Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(12): 2319-2323.
- [3] BELKN M, N IYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computing, 2003, 15(6): 1373-1396.
- [4] DONOHO D L, GR MES C. Hessian eigenmaps: locally linear embedding, techniques for high-dimensional data[J]. PNAS, 2003, 100(10): 591-596.
- [5] De R DDER D, KOUROPTOVA O, OKUN O, et al Supervised locally linear embedding[C]//Artificial Neural Networks and Neural Information Processing Istanbul, Turkey: Springer, 2003: 333-341.
- [6] KOUROPTOVA O, OKUN O, PIETIKA NEN M. Incremental locally linear embedding[J]. Pattern Recognition, 2005, 38(10): 1764-1767.
- [7] De R DDER D, LOOG M, RENDERS M J T. Local fisher embedding[C]//Proceedings of the 17th International Conference on Pattern Recognition [S 1], 2004: 295-298.
- [8] CHANG Hong, YENG Dityan Robust locally linear embedding[J]. Pattern Recognition, 2000, 39(6): 1053-1065.
- [9] YANG L. Building k-connected neighborhood graphs for isometric data embedding[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(5): 827-831.
- [10] YANG L. Building connected neighborhood graphs for locally linear embedding[C]//18th International Conference on Pattern Recognition Hong Kong, China, 2006: 1680-1683.
- [11] VAR NI C, DEGENHARD A, NATTKEMPER T W. ISOLLE: LLE with geodesic distance[J]. Neurocomputing, 2006, 69(15): 1768-1771.
- [12] 王和勇,郑杰,姚正安,李磊. 基于聚类和改进距离的 LLE方法在数据降维中的应用[J]. 计算机研究与发展, 2006, 43(8): 1485-1490.
WANG Heyong, ZHENG Jie, YAO Zhengan, LI Lei Application of dimension reduction on using improved LLE based on clustering[J]. Journal of Computer Research and Development, 2006, 43(8): 1485-1490.
- [13] WEN Guihua, JANG Lijun, WEN Jun, NIGEL R S Clustering-based nonlinear dimensionality reduction on manifold[C]//Lecture Notes in Artificial Intelligence (LNAI). Springer, 2006: 444-453.
- [14] 文贵华,江丽君,文军. 基于邻域优化的局部线性嵌入[J]. 系统仿真学报, 2007, 19(13): 3119-3122.
WEN Guihua, JANG Lijun, WEN Jun Locally linear embedding based on optimization of neighborhood[J]. Journal of System Simulation, 2007, 19(13): 3119-3122.
- [15] WEN Guihua, JANG Lijun, NIGEL R S Using graph algebra to optimize neighborhood for Isometric mapping[C]//20th International Joint Conference on Artificial Intelligence (IJCAI07). India, 2007: 2398-2403.
- [16] SAMKO O, MARSHALL A D, ROSN P L. Selection of the optimal parameter value for the Isomap algorithm[J]. Pattern Recognition Letters, 2006, 27(9): 968-979.
- [17] 文贵华,江丽君,文军. 邻域参数动态变化的局部线性嵌入[J]. 软件学报, 2008, 19(7): 1666-1673.
WEN Guihua, JANG Lijun, WEN Jun Dynamically determining neighborhood parameter for locally linear embedding[J]. Journal of Software, 2008, 19(7): 1666-1673.
- [18] WEN Guihua, JANG Lijun Globalizing local neighborhood for locally linear embedding[C]//Proceedings of 2006 IEEE International Conference on System, Man and Cybernetics Taipei, China, 2006: 3491-3496.
- [19] De R DDER D, LOOG M, RENDERS M J T. Local fisher embedding[C]//Proceedings of the 17th International Conference on Pattern Recognition Cambridge, UK, 2004: 295-298.
- [20] HONG Chang, YEUND Yang Robust locally linear embedding[J]. Pattern Recognition, 2003, 39(6): 1053-1065.

作者简介:



文贵华,男,1968年生,主要研究方向为创新计算、认知几何、机器学习与数据挖掘。



江丽君,女,1971年生,主要研究方向为机器学习与图形 CAD。



文军,男,1964年生,主要研究方向为创新计算与应用。