保持局部邻域关系的增量 Hessian LLE 算法

高翠珍 胡建龙 李德玉

(山西大学计算机与信息技术学院 太原 030006)

摘 要 Hessian LLE 算法是一种经典的流形学习算法,但该方法是以批处理的方式进行的,当新的数据点加入时,必须重新运行整个算法,计算所有数据点低维嵌入,原来的运算结果被全部丢弃。鉴于此,提出了一种保持局部邻域关系的增量 Hessian LLE(LIHLLE)算法,该方法通过保证流形新增样本点在原空间和嵌入空间局部邻域的线性关系不变,用其已有邻域点的低维坐标线性表示新增样本点,来得到新增点的低维嵌入,实现增量学习。在 Swiss roll with hole 和 frey_rawface 数据集上的实验表明,该方法简便、有效可行。

关键词 流形学习,Hessian LLE,增量学习

中图法分类号 TP391.4 **文献标识码** A

Incremental Hessian LLE by Preserving Local Adjacent Information between Data Points

GAO Cui-zhen HU Jian-long LI De-yu

(Department of Computer and Information Technology, Shanxi University, Taiyuan 030006, China)

Abstract Hessian LLE algorithm is a classical manifold learning algorithm. However, Hessian LLE is a batch mode. If only new samples are observed, the whole algorithm must run repeatedly and all the former computational results are discarded. So, incremental Hessian LLE(LIHLLE) algorithm was proposed, which preserves local neighborhood relationship between the original space and the embedding space. New sample points were linearly reconstructed with existing embedding results of local neighborhood samples. The proposed method can learn manifold in an incremental way. Simulation results in Swiss roll with hole and frey_rawface database testify the efficiency and accuracy of the proposed algorithms.

Keywords Manifold learning, Hessian LLE, Incremental learning

1 引言

流形学习是一种非线性降维技术,它通过分析数据集的外在结构来认识其本质,已经成为机器学习、模式识别、数据挖掘等领域的研究热点之一。近年来,流形学习已经得到了快速的发展,产生了大量的研究成果。Tenenbaum 等人[1]提出的 ISOMAP 算法首先使用最近邻图中的最短路径得到近似的测地线距离,用其代替不能表示内在流形结构的欧式距离,然后输入到多维尺度分析(MDS)中处理,进而发现嵌入在高维空间的低维坐标。Rowels 和 Saul[2]提出的 LLE 算法能够将高维输入数据点映射到一个全局低维坐标系,同时保持了邻接点之间的关系,这样原有的几何结构就能够得到保留。基于 LLE 的发展,人们提出了一些改进的算法,包括利用拉普拉斯(Laplacian)算子变化改进的算法 LE[3]、利用赫森(Hessian)变换改进的算法 HLLE[4]、基于切空间的改进算法 LTSA[5]、利用数据分类信息改进的监督 LLE[6,7]、增量式 LLE 等。然而,在性能上,HLLE 对 LLE 有较大的改进,甚至在某些情况下已超越了 ISOMAP 的能力。ISOMAP 是在假设全局等距和凸参数空间下进行的,这在多数情况下难以满足,而 HLLE 只要求局部等距映射和开的连通参数空间,因而适用范围更广。

以上算法已经被广泛应用,然而这些方法都是批处理的模式。当新样本不断地加入时,批处理方法必须重新计算所有的样本,计算是复杂的。为了克服此问题,一些科研工作者已经致力于研究增量学习算法。2005 年,Martin and Anil 提出了 ISOMAP 的增量算法[8],该方法首先更新测地距离,然后将问题转化为子空间的特征分解来实现全局的增量学习;Kouropteva 等人在假定原有特征值不变的基础上提出了 LLE 的增量算法[9],即在低维空间实现最优化;2006 年,Liu 等人提出了 LTSA 的增量算法[10];2007 年,曾宪华等人提出了一种动态增殖流形学习算法[11];2009 年,Peng Jia 等人用邻域点的低维坐标表示新增点的坐标,提出了 LE 的增量算法[12];2010 年,李厚森和成礼智受文献[13]启发,提出了增量的 HLLE 算法[14]等等。

到稿日期:2011-07-09 返修日期:2011-09-30 本文受国家自然科学基金资助项目(60875040,60970014,61175067),教育部高等学校博士点基金(200801080006),山西省自然科学基金资助项目(2010011021-1),山西省科技攻关项目(20110321027-02),太原市科技局明星专项(09121001)资助。

高翠珍(1986—),女,硕士,主要研究方向为数据挖掘、流形学习,E-mail:gcz2005241018@126.com;胡建龙(1981—),男,博士生,主要研究方向为流形学习、数据挖掘等;李德玉(1965—),教授,博士生导师,主要研究方向为粗糙集理论、模式识别、人工智能等。

通过以上分析，不难发现大多数已有的增量学习方法都是将原空间的特征分解问题转化为低维空间的特征分解来实现的，这些方法虽然降低了算法的复杂性，但每次增加一个样本点均需重新计算所有样本点的低维嵌入。事实上，由LLE、HLLE等局部流形学习算法的特性可知，当有单个新样本点加入时，大多数样本点的结果几乎不发生改变，因此这样的计算也造成了浪费。鉴于此，本文基于人的认知，利用流形结构的局部线性特性，提出了保持局部邻域关系的增量HLLE算法(LIHLLE)。当有新的样本加入时，保证流形新增样本点在原空间和嵌入空间局部邻域的权值不变，用其邻域点的低维坐标线性表示新增样本点，得到新增点的嵌入结果。

2 Hessian局部线性嵌入(HLLE)

HLLE算法是LLE算法的一种改进，将LLE的局部带权线性表示方法用局部等距代替，实现了数据降维。具体的算法步骤如下：

(1)邻域选取。获取每个样本点 x_i 的邻域点。记 $X_i=[x_{i1},\cdots x_{ik}]$ 为样本点 x_i 的 k 个最近的邻域点。

(2)计算切空间坐标。对每个样本点的邻域，计算中心化矩阵 $X_i-\overline{x_i}1_k^{\mathrm{T}}$ 的前 d 个最大的特征值对应的特征向量，并将这 d 个特征向量组成矩阵 V_i。

(3)估计Hessian矩阵。有

$$M_i=[1,V_i,(V_i(:,s)*V_i(:,l)_{1\leqslant s\leqslant l\leqslant d})]$$

式中，矩阵共有 $1+d+d(d+1)/2$ 列：前 $d+1$ 列由分量为1的列向量和 V_i 组成，$V_i(:,s)*V_i(:,l)$ 表示矩阵 V_i 的第 s 列和第 l 列的点积。对矩阵 M_i 进行Gram-Schmidt正交化，得到列正交阵 $\overline{M_i}$，则Hessian矩阵

$$H^i=\overline{M_i}(:,d+1:1+d+d(d+1)/2)^{\mathrm{T}}$$

(4)构造二次项。利用每个邻域的Hessian矩阵 H^i，$i=1,\cdots,N$ 来构造对称矩阵 H，其元素为

$$H_{ij}=\sum_{s=1}^{N}\sum_{l=1}^{d(d+1)/2}(H^s)_{l,i}(H^s)_{l,j}$$

(5)计算 H 的零空间。计算 H 的 $d+1$ 个最小特征值对应的特征向量 $u_1,\cdots,u_d$，则 $U=[u_2,\cdots,u_{d+1}]$ 就是所求的零空间。

(6)计算低维嵌入。记矩阵

$$R_{ij}=\sum_{l\in J_1}U_{l,i}U_{l,j}(i,j=1,\cdots,d)$$

式中，J_1 表示某个样本点的邻域，则 $T=R^{-1/2}U^{\mathrm{T}}$ 为低维嵌入。

由以上算法可以发现：1)HLLE框架上与LLE一致，不同的是，HLLE用Hessian变换取代了LLE的局部带权线性表示；2)每个点的切空间要求其邻域是线性的，即选择合适的邻域，保证邻域的局部线性特性，这是HLLE需要克服的一个问题；3)HLLE需要对每个数据点计算 $d\times d$ 次偏导数，当观察数据的维数非常高时，计算量较大，这是HLLE需要克服的另一个问题。因此，针对增量的HLLE算法，本文着重考虑了这两个问题，保证随着数据点的增多，邻域的大小也自适应地发生变化，同时尽可能降低算法的复杂度。

3 增量Hessian LLE算法(LIHLLE)

在假设原样本点计算结果准确的基础上，本文提出的保持局部邻域关系的增量HLLE算法分3步来计算新样本的低维嵌入：①根据流形结构的局部线性特性，自适应地确定新增点的邻域，邻域的线性程度用PCA来度量；②通过使得新增点在原空间中的邻域线性表示和其本身的误差最小，计算新增点在原空间中的邻域权值；③利用原样本的嵌入结果，保持原空间和嵌入空间邻域权值不变，计算新增点的投影值。

3.1 自适应的邻域选择

设原样本集为 $X=\{x_1,\cdots,x_n\}$，用经典的HLLE算法得到的 X 的嵌入结果为 $Y=\{y_1,\cdots,y_n\}$，新增样本点记为 x_{n+1}。用 $KNN(x_{n+1})$ 表示 x_{n+1} 的 K 最近邻，用 $RN(x_{n+1})$ 表示 x_{n+1} 的合理邻域集，初始值为空。新增样本 x_{n+1} 加入后，其合理邻域的计算过程如下：

(1)计算距离其最近的 $K(K>d)$ 个点，将其按边长度递增的顺序排列，记为 $KNN(x_{n+1})=\{x_{n+1,1},\cdots,x_{n+1,K}\}$，同时令 $RN(x_{n+1})=\{x_{n+1,1},\cdots,x_{n+1,d}\}$；

(2)用PCA方法度量其局部线性特性，满足邻域线性条件的点为合理邻域点，即依次计算 $x_{n+1,i}\cup RN(x_{n+1})(i=d+1:K)$ 样本集的协方差矩阵的特征值 $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_D$，如果特征值满足 $\sum_{i=1}^{i=d}\lambda_i/\sum_{i=1}^{i=D}\lambda_i>\eta$($\eta$ 为经验值)，$x_{n+1,i}$ 就记为 x_{n+1} 的合理邻域，$RN(x_{n+1})=x_{n+1,i}\cup RN(x_{n+1})$

3.2 计算邻域权值

通过 $RN(x_{n+1})$ 来计算新增样本点 x_{n+1} 的重构权值 $w_{n+1,i}$，代价函数为

$$\varepsilon(W)=|x_{n+1}-\sum_{i=1}^{size(RN(x_{n+1}))}w_{n+1,i}x_{n+1,i}|^2 \tag{1}$$

式中，$x_{n+1,i}\in RN(x_{n+1})$ 权值 $w_{n+1,i}$ 满足条件

$$\sum_i w_{n+1,i}=1 \tag{2}$$

这样，求最优权值就是对于式(1)在约束条件(2)下求解最小二乘问题。

3.3 计算低维嵌入

邻域权值 $w_{n+1,j}$ 确定后，新样本点 x_{n+1} 的低维表示可以通过以下公式得到，即

$$x_{n+1}\rightarrow y_{n+1}=\sum_{i=1}^{size(RN(x_{n+1}))}w_{n+1,i}y_{n+1,i} \tag{3}$$

式中，$y_{n+1,i}\in Y$。

以上算法可以计算出单个新增点的低维嵌入，当有多个样本同时加入时，循环计算每个数据点，将每次计算出的结果加入已有的嵌入结果中作为原始点，依次计算出所有新增点的低维嵌入(本实验中 K 取30，η 取0.93)。

4 实验结果及分析

4.1 实验

实验选取Swiss roll with hole和frey_rawface两个数据集，分别用批处理HLLE和增量HLLE两种方法计算样本点的投影值，通过比较两种方法的实验结果验证本算法的有效性。Swiss roll with hole数据集为3维空间的面包圈，将其嵌入到2维空间；frey_rawface数据集由1965幅像素为 $20*28$ 的不同姿态和表情的人脸图像组成，即原始空间为 $20*28=560$ 维，嵌入到2维空间分别表示人脸姿态和表情。图1是frey_rawface数据集的一些样本。为了比较批处理HLLE算法和增量HLLE两种算法嵌入值的差别，定义了误差来进行度量。误差的表示形式为

$$error=\sqrt{\frac{1}{N}\sum_{i=1}^{i=N}\frac{\| y_i-\hat{y}_i \|^2}{\| \hat{y}_i \|^2}} \tag{4}$$

式中，y_i 和 $\hat{y}_i$ 分别表示第 i 个点用 HLLE 和 LIHLLE 计算的低维嵌入值，误差越小，LIHLLE 的嵌入结果越接近于 HLLE 的嵌入结果。同时通过比较两种方法所耗的时间来度量算法的效率。

图 1　frey_rawface 数据库中的 10 个样本

先在每个数据集上各随机抽取 500 个点作为原始样本点，用 HLLE 计算它们的低维嵌入，然后再将新样本点一个接一个地加入，用 LIHLLE 算法计算它们的投影值，直到 2000 个数据点。

图 2 显示了在 Swiss roll with hole 数据集上的效果图，(a. b. c)，(d. e. f)，(g. h. i) 3 组图分别表示样本点数 N 从 500 增加到 1000、1500、2000 时对应的原始数据图、HLLE 算法的低维嵌入图及 LIHLLE 算法的低维嵌入图。图 3 用误差曲线图定量地表示了两种方法的低维嵌入结果，图 3(a) 为 Swiss roll with hole 数据集上，LIHLLE 算法对应的误差曲线图，图 3(b) 为 frey_rawfac 数据集上，LIHLLE 算法对应的误差曲线图，其中横轴均表示样本数，纵轴均表示用两种方法映射产生的误差。同时表 1 记录了原始样本点数 N 从 500 依次增加为 1000、1500、2000 个数据点时，用两种方法映射到低维空间时所消耗的时间。

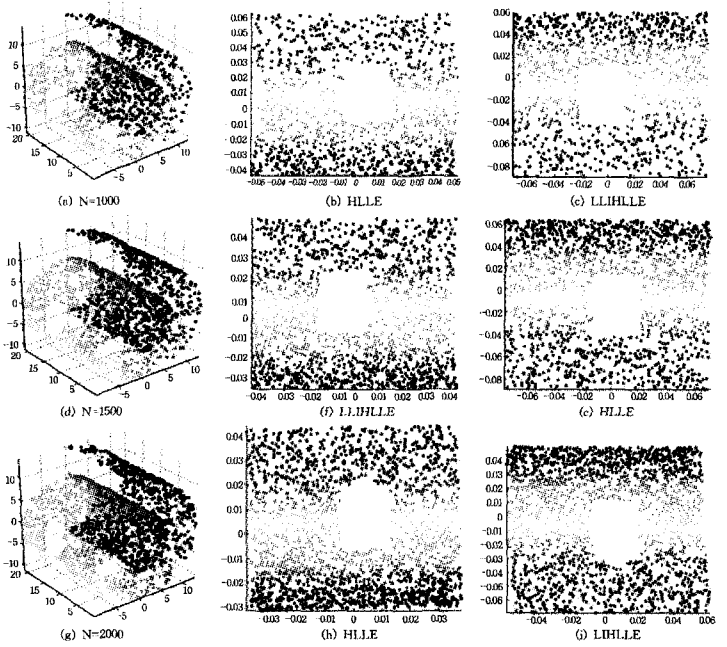
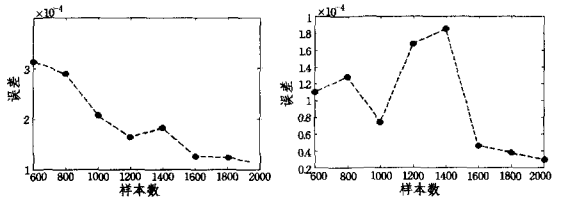


图 2　Swiss roll with hole 数据集上，样本点分别为 1000、1500、2000 时对应的原始图、HLLE 嵌入图、LIHLLE 嵌入图



(a) Swiss roll with hole　　(b) frey_rawface

图 3　LIHLLE 算法在两个数据集上的误差曲线图

表 1　HLLE 和 LIHLLE 两种方法的运行时间(单位：s)

	Swiss roll with hole(3 to 2)		frey_rawface(560 to 2)	
	HLLE	*LIHLLE*	*HLLE*	*LIHLLE*
500	1.51		5.96	
1000	837.3	0.14	4013.9	0.46
1500	4951.1	0.37	12573.7	1.32
2000	13530.2	0.65	N/A	N/A

4.2　结果分析

通过图 2 可以直观地发现所提增量 HLLE 算法可以很好地将 Swiss roll with hole 数据集展开，和原始 HLLE 算法的展开结果几乎一样。图 3(a) 和 3(b) 进一步用误差定量地证明了本文中算法的有效性，该方法在 Swiss roll with hole 数据集上的平均误差为 9.7153e-005，在 frey_rawface 数据集上的平均误差为 1.8955e-004，这两个误差均很小，可见该方法的处理精度完全满足要求。

表 1 的数据表示本文的增量 HLLE 算法和原始的 HLLE 算法在两种数据集上的运行时间，由此可知，本文中算法的速度明显快于批处理的 HLLE 算法。此外，文献[14]中提出的增量 HLLE 学习方法虽然比原始的 HLLE 算法效率高，但仍需要对每个点计算 Hession 矩阵，计算过程仍然是复杂的，而本文提出的算法巧妙地避免了这些复杂的计算，大大减少了算法运行的时间。因此，本文在保证嵌入结果准确的前提下，大大提高了算法的效率。

结束语　本文基于流形的局部线性特性，在假设原来的映射结果准确的基础上提出了增量的 HLLE 算法。该算法避免了重复计算所有样本点的嵌入，利用原有的嵌入结果简单高效地计算出新增样本点的嵌入结果，实现了 HLLE 的增量学习。在 Swiss roll with hole 和 frey_rawface 数据集上的实验证明，该算法得到的结果与批处理方法得到的结果相近，然而算法的效率得到了很大的提高。

(下转第 226 页)

表 3 EF 所离散化的数据集上的分类结果

决策树算法	建模所使用的属性个数	检测率(%)	建模时间(s)
DTRDE	15	99.304	0.792
ID3	41	99.023	0.83
C4.5	41	98.924	0.74

从表 3 可以看出，在由 EF 所离散化的数据集上，DTRDE 的性能也要好于 ID3 和 C4.5。这也证明了所提算法的有效性。

结束语 本文将粗糙集与决策树有机地结合在一起，基于相对决策熵提出一种新的分离属性选择标准，并给出相应的决策树算法。在建树之前，该算法利用数据约简技术进行预剪枝，以有效降低决策树的规模。在从 KDD Cup99 中所随机抽取的数据子集上的实验表明，该算法的入侵检测效果要好于其他算法。但由于实验数据集的规模有限，该算法的真实检测性能还有待进一步的考证。

在后续工作中，计划将该算法扩展到邻域粗糙集模型中[26]，设计一种不需要离散化就能够直接处理连续型和离散型属性的算法。

参 考 文 献

[1] Quinlan R. Induction of decision trees[J]. Machine Learning, 1986,1(1):81-106

[2] Quinlan R. C4.5:Programs for Machine Learning[M]. Morgan Kaufmann,1993

[3] Huang L J,Huang M H,Guo B. A new method for constructing decision tree based on rough set theory[C]// 2007 IEEE Int. Conf. on Granular Computing. 2007:241-244

[4] Li X P,Dong M. An algorithm for constructing decision tree based on variable precision rough set model[C]// The 4th Int. Conf. on Natural Computation. 2008:280-283

[5] Wei J M,Huang D,Wang S Q. Rough set based decision tree[C]// Proc. of the 4th World Congress on Intelligent Control and Automation. 2002:426-431

[6] Bai J S,Fan B,Xue J Y. Knowledge representation and acquisition approach based on decision tree[C]// Int. Conf. on Natural Language Processing and Knowledge Engineering. 2003:533-538

[7] Pawlak Z. Rough Sets[J]. Int. J. Comput. Informat. Sci. ,1982, 11(5):341-356

[8] Pawlak Z. Rough Sets:Theoretical Aspects of Reasoning about Data[M]. Kluwer Academic Publishing,1991

[9] 苗夺谦，胡桂荣. 知识约简的一种启发式算法[J]. 计算机研究与发展，1999,36(6):681-684

[10] 王国胤. Rough 集理论与知识获取[M]. 西安：西安交通大学出版社，2001

[11] 王国胤，于洪，等. 基于条件信息熵的决策表约简[J]. 计算机学报，2002,25(7):759-766

[12] 刘少辉，盛秋戬，等. Rough 集高效算法的研究[J]. 计算机学报，2003,26(5):525-529

[13] 徐章艳，刘作鹏，等. 一个复杂度为 max(O(|C||U|),O(|C|^2|U/C|))的快速属性约简算法[J]. 计算机学报，2006,29(3):391-399

[14] Anderson J P. Computer Security Threat Monitoring and Surveillance[M]. James P. Anderson Co. ,USA,1980

[15] Li X Y,Ye N. Decision tree classifiers for computer intrusion detection[J]. Journal of Parallel and Distributed Computing Practices,2001,4(2):179-190

[16] Kruegel C,Toth T. Using decision trees to improve signature-based intrusion detection[C]// Symp. on Recent Advances in Intrusion Detection. 2003:173-191

[17] Amor N B,Benferhat S,Elouedi Z. Naive Bayes vs decision trees in intrusion detection systems [C]// ACM Symp. on Applied Computing. 2004:420-424

[18] Shannon C E. The mathematical theory of communication[J]. Bell System Technical Journal,1948,27(3/4):373-423

[19] Liang J Y,Shi Z Z. The information entropy,rough entropy and knowledge granulation in rough set theory[J]. Int. Journal of Uncertainty,Fuzziness and Knowledge Based Systems,2004,12(1):37-46

[20] Düntsch I,Gediga G. Uncertainty measures of rough set prediction[J]. Artificial Intelligence,1998,106:109-137

[21] Witten I H,Frank E. Data Mining:Practical Machine Learning Tools and Techniques with Java Implementations[M]. Morgan Kaufmann,2000

[22] KDD Cup 99 Dataset[OL]. http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html

[23] Bay S D. The UCI KDD repository[OL]. http://kdd.ics.uci.edu,1999

[24] 陈仕涛，陈国龙，等. 基于粒子群优化和邻域约简的入侵检测日志数据特征选择[J]. 计算机研究与发展，2010,47(7):1261-1267

[25] Øhrn A. Rosetta Technical Reference Manual. 2000

[26] Hu Q H,Yu D R,Liu J F,et al. Neighborhood rough set based heterogeneous feature subset selection [J]. Information Scien-ces,2008,178(18):3577-3594

(上接第 219 页)

参 考 文 献

[1] Tenenbaum J B,De Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000,290:2319-2323

[2] Roweis S T,Saul L K. Nonlinear Dimensionality Reduction by Locally Linear Embedding[J]. Science,2000,290:2323-2326

[3] Belkin M,Niyogi P. Laplacian eigenmaps for dimensionality reduction and data rep resentation[J]. Neural Computing,2003,15(6):1373-1396

[4] Donoho D L,Grimes C. Hessian Eigenmaps:Locally Linear Embedding Techniques for High Dimensional Data[J]. Proceedings of the National Academy of Sciences of the United States of America,2003,100(10):5591-5596

[5] Zhang Z,Zha H. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment[J]. Journal of Shanghai University(English Edition),2004,8(4):406-424

[6] de Ridder D, Kouropteva O, Okun O, et al. Supervised locally linear embedding[J]. Artificial Neural Networks and Neural Information Processing,2003,2714:333-341

[7] Kouropteva O, Okun O, Pietikainen M. Supervised Locally Linear Embedding Algorithm for Pattern Recognition[J]. Pattern Recognition and Image Analysis,2003,2652:386-394

[8] Martin H C L,Anil K J. Incremental Nonlinear Dimensionality Reduction by Manifold Learning[J]. IEEE Trans. Pattern Analysis and Machine Intelligence,2006,28(3):377-391

[9] Kouropteva O,Okun O. Incremental locally linear embedding algorithm[J]. Pattern Recognition,2005,38(10):1764-1767

[10] Liu X, Yin J, Feng Z, et al. Incremental manifold learning via tangent space alignment[J]. ANNPR,2006,4087:107-121

[11] 曾宪华，罗四维. 动态增量流形学习算法[J]. 计算机研究与发展，2007,44(9):1462-1468

[12] Jia P,et al. Incremental Laplacian eigenmaps by preserving adjacent information between data points[J]. Pattern Recognition, 2009,30:1457-1463

[13] Abdel-Mannan O,Hamza A B,Youssef A. Incremental Hessian Locally Linear Embedding Algorithm[C]// Proc. of the 9th International Symposium on Signal Processing and Its Applications. 2007

[14] 李厚森，成礼智. 增量 Hessian LLE 算法研究[J]. 计算机工程，2010,37(6):159-161