

一种改进的有间隔流形 ISOMAP 算法

王洪元 刘爱萍 程起才 刘锁兰

(常州大学信息科学与工程学院常州市过程感知与互联技术重点实验室 常州 213164)

摘要 ISOMAP 算法成功应用的潜在条件是要求数据集均匀抽样于单个的内在流形。如果数据集均匀抽样于某个内在流形,但内部出现了一个间隔,ISOMAP 算法可能失效。提出了 G-ISOMAP (ISOMAP with a Gap) 算法,该算法充分利用了数据集中的间隔特性。首先检测被间隔的子流形间最短欧氏距离对应的数据点,然后将这些数据点互相设置为邻域点,最后用 ISOMAP 算法找到低维嵌入结果。对 G-ISOMAP 与 ISOMAP 算法的区别与联系进行了详细的理论说明,得出 ISOMAP 算法是 G-ISOMAP 算法的一个特例,G-ISOMAP 算法是 ISOMAP 算法扩充的结论。实验结果验证了该算法比其他常用的流形学习算法在有间隔的数据集上更有效。

关键词 流形学习,等测距映射,有间隔流形

Improved ISOMAP for a Single Manifold with a Gap

WANG Hong-yuan LIU Ai-ping CHENG Qi-cai LIU Suo-lan

(Changzhou Key Laboratory for Process Perception and Interconnected Technology, School of Information Science and Engineering, Changzhou University, Changzhou 213164, China)

Abstract ISOMAP algorithm could have been applied successfully on uniform-density dataset drawn from a single manifold. However, given a uniform-density dataset with a gap, ISOMAP fails possibly. In this paper, G-ISOMAP (ISOMAP with a Gap) algorithm was presented, which exploits characteristic of the gap in the dataset. The algorithm first finds pairs of data points, whose Euclidean distances are shortest between the separated sub-manifolds, and then makes them neighbors each other. At last, ISOMAP algorithm is applied to find low dimensional embedding structure. The theoretical discussion on difference and relationship between G-ISOMAP and ISOMAP can be given and it is concluded that ISOMAP is a special case of G-ISOMAP algorithm and G-ISOMAP is an extension of ISOMAP algorithm. The experimental results show that the proposed algorithm is best among the frequently used manifold learning algorithms on several datasets with a gap.

Keywords Manifold learning, ISOMAP, Manifold with a gap

1 引言

在高维空间中进行各种处理需要的样本数量会呈指数增加,样本间距离的价值也越来越小,即面临维数灾难问题。所幸的是,在很多实际问题中,大部分高维观测数据变量可以用少量几个影响因素来表示,这说明其中包含着大量冗余信息,各成分之间通常有着较强的相关性,这种现象在几何学上表现为数据分布在低维流形上,或者是在低维流形附近。2000 年 Seung 等人从认知的角度在 Science 上发表文章,提出感知可能以流形方式存在,视觉记忆也可能是以稳态的流形存储^[1]。从这以后,研究和模拟人的这种感知能力,即从有限的样本数据集中学习潜在流形的问题,成为国内外许多机器学习方面专家的研究热点,出现了许多经典的科学文献。例如基于古典的 MDS 算法的等距映射 (ISOMAP) 算法^[2]以流形上近似

的测地线距离作为 MDS 的输入,使得在嵌入空间中原始样本间的近似测地线距离得以保持;基于一个流形在很小的局部邻域上可以近似看成是欧氏的局部线性嵌入 LLE (Locally Linear Embedding) 算法^[3]在最小二乘意义上将每个样本表示为相邻样本的线性组合,然后要求相邻样本在嵌入空间中仍尽量保持同样的线性组合,从而求出最终的映射;通过考察局部 Hessian 的二次型来得出结论的 Hessian LLE 算法^[4]、基于图的嵌入 (graph embedding) 来找到数据的低维表示的 LE (Laplacian Eigenmaps) 算法^[5],都是保持样本间的邻近关系,即相距较近的样本在映射后应离得较近,而相距较远的样本在映射后仍离得较远;将 LE 算法线性化的保局投影算法 LPP (Locality Preserving Projections)^[6]、局部切空间排列算法 LTSA (Local Tangent Space Alignment)^[7],首先求出各点的切空间,然后对准相邻切空间来决定全局嵌入坐标;最大方

到稿日期:2011-01-18 返修日期:2011-04-29 本文受国家自然科学基金项目(61070121, 60973094),江苏省自然科学基金项目(BK2009538),江苏省产学研前瞻性联合研究项目(BY2009117)资助。

王洪元(1960—),男,博士,教授,硕士生导师,主要研究领域为模式识别与智能系统, E-mail: hywang@cczu.edu.cn;刘爱萍(1986—),女,硕士生,主要研究领域为机器学习;程起才(1981—),男,硕士,讲师,主要研究领域为机器学习、数字图像处理;刘锁兰(1980—),女,博士,讲师,主要研究领域为机器学习与模式识别。

差展开 MVU(Maximum Variance Unfolding)算法^[8],在保持局部近邻点的欧氏距离不变的约束下,尽量使映射后数据分布的方差最大。其中在 ISOMAP 算法基础上,国内外学者提出了许多改进的算法,如基于保形映射的 C-ISOMAP 和基于少量标志点的 L-ISOMAP^[9]、基于核方法的鲁棒 ISOMAP 算法^[10]、有监督框架下的 S-ISOMAP 算法^[11]和基于核密度估计的短路边检测的 P-ISOMAP 算法^[12]。

然而,这些 ISOMAP 算法及其变种总体上是基于通过精心选择邻域参数并一次性构造好邻域图的思路,很明显这是很耗时的。如果事先知道抽样的数据集中间可能存在一个较大的间隔(gap),那么利用数据集这个间隔特性来发现数据集内的低维流形结构将变得简单。基于这种想法,本文提出了一种称之为带有间隔的流形 ISOMAP 算法 G-ISOMAP(ISO-MAP with a Gap),该算法能够成功发现带有间隔数据集内在的低维流形结构特征。

本文第 1 节简要概述 ISOMAP 算法;第 2 节提出 G-ISOMAP 算法,讨论 G-ISOMAP 算法与 ISOMAP 算法的区别与联系;第 3 节给出实验结果,最后得出结论并提出未来值得研究的几个问题。

2 ISOMAP 概述

人们通常用欧氏距离来表达数据间的相异度,但这是建立在全局线性结构假设基础之上的^[13]。如果数据不是来源于一个线性结构,而是来源于一个高度扭曲的非线性结构,此时用欧氏距离来度量数据间的相异度就不合适了。为了刻画流形上的点之间的相异度,ISOMAP 采用由数据点形成的无向加权邻接图中点与点之间的最短路径长度来刻画^[14]。

但是 ISOMAP 算法是否得以成功运用,关键在于两点:一是邻域图是否构造正确,例如短路边的产生将使邻域图完全扭曲^[15];二是构造邻域图时,是否产生了多个子邻域图,因为 ISOMAP 算法对多个子邻域图是无法同时产生低维嵌入的。如果被采样的数据集中间出现一个较大的间隔,此时出现短路边的概率非常大,甚至是不可避免的。以类似于多峰高斯分布形式的二维数据为例,图 1(a)给出了嵌入在二维欧氏空间的一维流形,图 1(b)抽样于图 1(a),很明显,除中间有间隔(虚线表示间隔,该间隔处没有数据产生)外,其它位置的数据分布较均匀。对于该数据集,若采用 ISOMAP 算法,那么当邻域参数值取得较小时,会形成两个子邻域图;为了让它形成一个邻域图,必须使得邻域参数的值一直变大,直到间隔处一端的数据自动成为间隔另一端数据的邻域点为止,但是此时在该流形的两个峰附近必然会产生大量的短路边。例如图 1(b)中的 B 点和 F 点就会产生短路边,所以对于图 1(b)的数据集,无论怎么精心选择参数 K(或 ϵ)的值,ISOMAP 算法均得不到正确的嵌入结果。

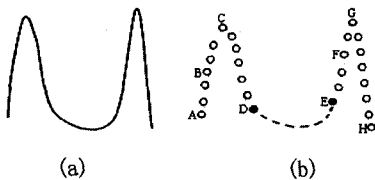


图 1 原流形与有间隔的数据集

3 G-ISOMAP

3.1 G-ISOMAP 理论分析

本文针对 ISOMAP 算法对类似于图 1(b)的数据集不能产生正确低维嵌入的缺点,提出了 G-ISOMAP 算法。为了更好地描述 G-ISOMAP 算法的适用场合和算法步骤,称 \widehat{ABCD} 与 \widehat{EFGH} 为被间隔后的子流形,并引入如下符号。

1)子流形内:子流形内部第 i 个数据点 X_i 到该子流形上其他数据点的欧氏距离记为 $W(X_i)$ 。不失一般性,假设 $W(X_i)$ 里每个元素是按照升序排列的,那么 $[W(X_i)]_1, [W(X_i)]_2, \dots, [W(X_i)]_K$ 表示离 X_i 最近的 K 个欧氏距离, X_i^K 表示 $[W(X_i)]_K$ 对应的另一端数据点,即 $\|X_i - X_i^K\| = [W(X_i)]_K$ 。

很明显,如果 X_i 与 X_i^K 的弧长远远大于其对应的欧氏距离,即满足式(1),则认为 X_i^K 与 X_i 产生了短路边。

$$|\widehat{X_i X_i^K}| \gg \|X_i - X_i^K\| = [W(X_i)]_K \quad (1)$$

2)子流形间:一个子流形上第 i 数据点 X_i 到另一个子流形上所有数据点之间的欧氏距离记为 $B(X_i)$ 。同样,也假设 $B(X_i)$ 里每个元素是按照升序进行排列的。很明显, $[B(X_i)]_1$ 表示数据点 X_i 到位于另外一个子流形上所有数据点的最短欧氏距离。

应用上面的符号定义,以图 1(b)所示的数据集为例,分别讨论 ISOMAP 算法与 G-ISOMAP 算法的适用场合以及它们之间的关系。因为在 ISOMAP 算法中选择某个点的邻域,通常有两类方法,所以考虑两种情形。

①邻域参数采用半径为 ϵ 的超球:对于图 1(b)所示流形的数据集,如果应用 ISOMAP 算法, ϵ 取值必须满足 $\epsilon > |DE|$ 才能够保证在整个数据集上产生一个邻域图。但不幸的是,此时在每个子流形内部产生了大量短路边。这是因为 $|DE|$ 比 $W(B)$ 绝大多数元素要大,而且这些元素均满足条件(1),所以 ISOMAP 算法在邻域参数取 ϵ 的超球的情况下,无法得到正确嵌入结果。

②邻域参数采用 K 近邻:对于图 1(b)所示流形的数据集,很明显,两个子流形间最短的欧氏距离为 $|DE|$,即满足 $|DE| = [B(D)]_1 = [B(E)]_1$ 。那么,如果应用经典的 ISOMAP 算法,必须满足如下条件才能够使得在整个数据集上产生唯一的邻域图。

$$[B(D)]_1 < [W(D)]_K \text{ 或者 } [B(E)]_1 < [W(E)]_K \quad (2)$$

满足条件(2)即意味着在邻域参数 K 下, E 点被设置成 D 点的邻域点,或者 D 点被设置成 E 点的邻域点,故可以构造出唯一的邻域图。但是在满足条件(2)的情况下,若子流形内部的数据点又满足了条件(1),即在流形内部产生了短路边,此时 ISOMAP 算法也无法得到正确的嵌入。

而本文提出的 G-ISOMAP 算法能够有效解决这种情形的问题。本文算法的主要思想是:在基于数据集中可能包含一个间隔、位于其他位置的数据是均匀分布的假设的前提下,不是企图增大邻域参数的值,以使得数据集内的所有样本点产生一个邻域图,而是首先在被隔开的子数据集上选择一个合适的、不产生短路边的邻域参数 ϵ 或 K ,这样就产生了两个子邻域图。然后求出两个子邻域图之间最短的 γ 个欧氏距离来代替间隔区域边缘数据点之间的测地线距离,这样就使得

两个子邻域图重新变成一个邻域图,同时在子流形内部不会出现短路边现象。很明显,对于嵌入在高维空间的一维流形只需要 $\gamma=1$,对于二维及以上流形的数据集, $\gamma>1$,因为这样可以避免间隔区域边缘数据点间的最短路径必须经过这条边而影响测地线距离估计的问题。例如在图 1(b)数据集中,在邻域参数取合适值时,会产生两个分散的流形,然后求出该分散流形间的最短($\gamma=1$)欧氏距离所对应的数据点,即图中的 D 点和 E 点。此时将 D 点和 E 点互相设置为邻域点,这样被间隔的两个子邻域图重新被拼接成一个邻域图,最后应用 ISOMAP 算法来发现数据集的内部结构。

所以 ISOMAP 与 G-ISOMAP 的适用场合均在不满足式(1)的条件下的所有邻域参数取值中,

①如果条件(2)得到满足,此时 ISOMAP 算法与 G-ISOMAP 算法均能得到正确的低维嵌入;

②如果条件(2)不满足,此时 ISOMAP 算法失效,但 G-ISOMAP 算法能得到正确的低维嵌入。

不过,在寻找带有间隔数据集的低维流形结构时,G-ISOMAP 的效果与间隔区域边缘数据集之间的测地线距离估计有关。如果所选取的 γ 条边两端的数据点恰好位于被间隔区域的边缘,则此时算法能够成功发现数据集的内在低维流形结构, γ 条边两端的数据点离被间隔区域的边缘越远,效果将会越差。

3.2 G-ISOMAP 学习算法

G-ISOMAP 算法具体描述为:假设输入数据集为 $\{X_1, X_2, \dots, X_N\}$,其中 $X_i \in R^n$,相应的低维嵌入输出数据集为 $\{Y_1, Y_2, \dots, Y_N\}$, $Y_i \in R^m$ (一般 $m \ll n$),那么算法步骤如下:

Step1 构造邻域图。选取合适的邻域参数 ϵ 或 K ,求出 X_i 对应的邻域点集合 $N(X_i)$,并计算出邻域图个数 p 。如果 $p=1$,则进行 Step5-Step6,否则从 Step2 开始执行。

Step2 在 Step1 所选择的邻域参数值下,求出产生的两个子邻域图以及每个子邻域图由哪些数据点组成:以 S_1, S_2 分别表示构成这两个子邻域图的数据。

Step3 求出子邻域图数据集间最短的 γ 条欧氏距离分别对应的数据点,并将该数据点按照其所属的子邻域图进行分类,分别记为 E_1, E_2 。显然, $E_1 \subset S_1, E_2 \subset S_2$ 。

Step4 将 E_1, E_2 中的数据点互相设置为邻域点,即对于 $\forall X_i \in E_1, X_j \in E_2$,邻域点集合 $N(X_i)$ 被更新为 $N(X_i) \cup E_2$ 。

Step5 求出数据点间的最短路径。Step4 结束,整个数据集上形成了单个邻域图,求一个以欧氏距离加权的无向邻域图中数据点间的最短路径,可以使用经典的 Floyd 算法或 Dijkstra 算法。

Step6 构建 m 维的低维嵌入。以 Step5 所得到的点对点之间的最短路径作为古典的 MDS 算法的输入,得到 m 维的低维嵌入 $\{Y_1, Y_2, \dots, Y_N\}$,其中 $Y_i \in R^m$ 。

从上述算法步骤可以看出 G-ISOMAP 算法与 ISOMAP 算法的区别与联系:

① G-ISOMAP 算法可调参数是邻域参数和子邻域图间最短的欧氏距离条数 γ , ISOMAP 算法只有一个邻域参数可调。

②如果选择的邻域参数值能够在整个数据集上形成一个

邻域图且又不出现短路边,那么上述 Step2 到 Step4 就可以省略,此时 G-ISOMAP 算法的步骤完全与 ISOMAP 算法等价,ISOMAP 算法是 G-ISOMAP 算法的一个特例,G-ISOMAP 算法是 ISOMAP 算法的扩充。也就是说,G-ISOMAP 算法不仅适用于均匀抽样于某个潜在流形的数据集,而且适用于抽样于单个流形但中间带有一个间隔的数据集。

4 实验结果与分析

为了验证本文提出的算法的性能,分别在人造数据集和人脸数据集上进行了实验,编程环境为 MatlabR2008a。

4.1 人造数据集

给出了两种类型的人造数据集,如图 2 所示,图 3 是其对应的中间含有间隔情况的数据集。图 4 分别给出了其理想的低维嵌入结果¹⁾,图 5 给出了每个数据集上形成的两个子邻域图之间 γ 条最短欧氏距离边。将各种算法在这两个数据集上产生的实验结果与该理想结果进行比较,作为评价算法性能好坏的标准。

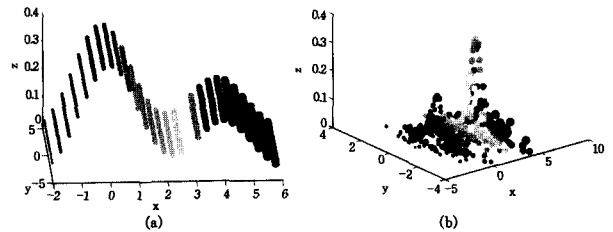


图 2 均匀抽样于单个流形且中间没有间隔的数据集

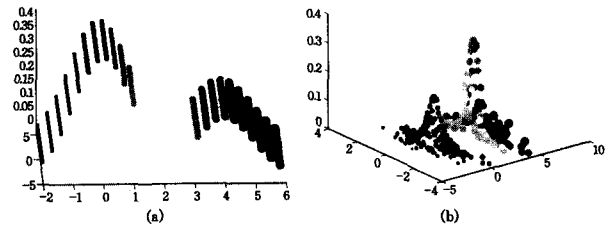


图 3 均匀抽样于单个流形且中间有间隔的数据集

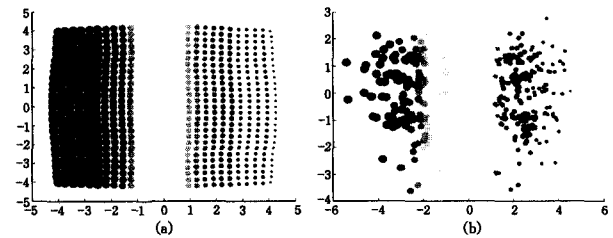


图 4 理想结果

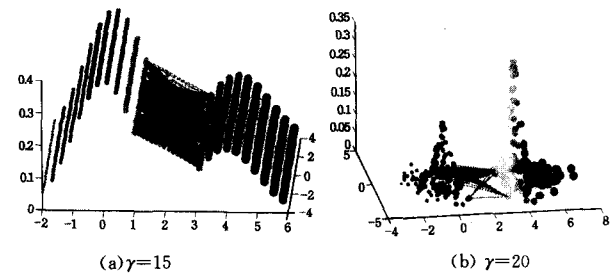


图 5 γ 条最短欧氏距离对应的边

¹⁾ 这个理想的结果是首先把 ISOMAP 算法分别应用在图 2 数据集上,得到整个数据集上的低维嵌入,然后在这个低维嵌入结果的基础上去掉间隔部分,所得对应的低维嵌入就是理想结果。

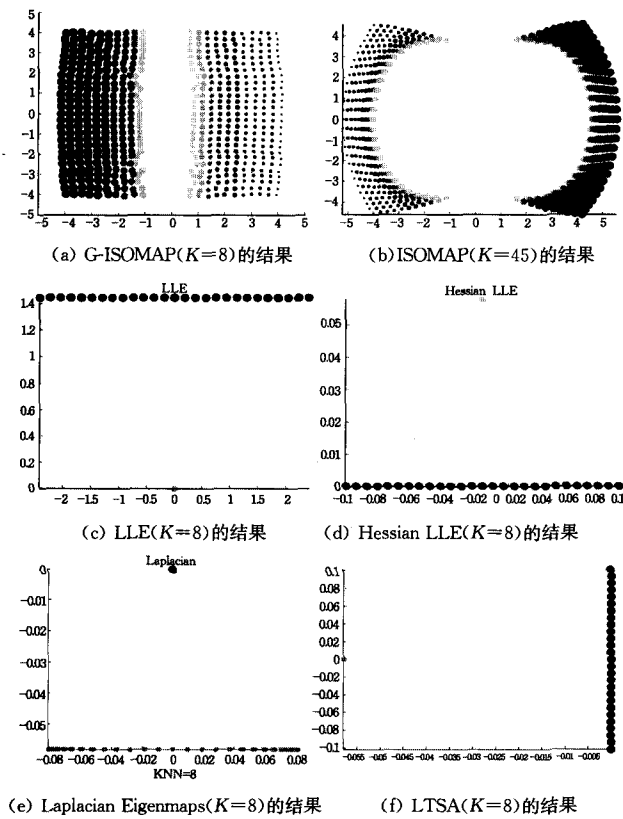


图6 常用流形学习算法在图3(a)数据集上的结果

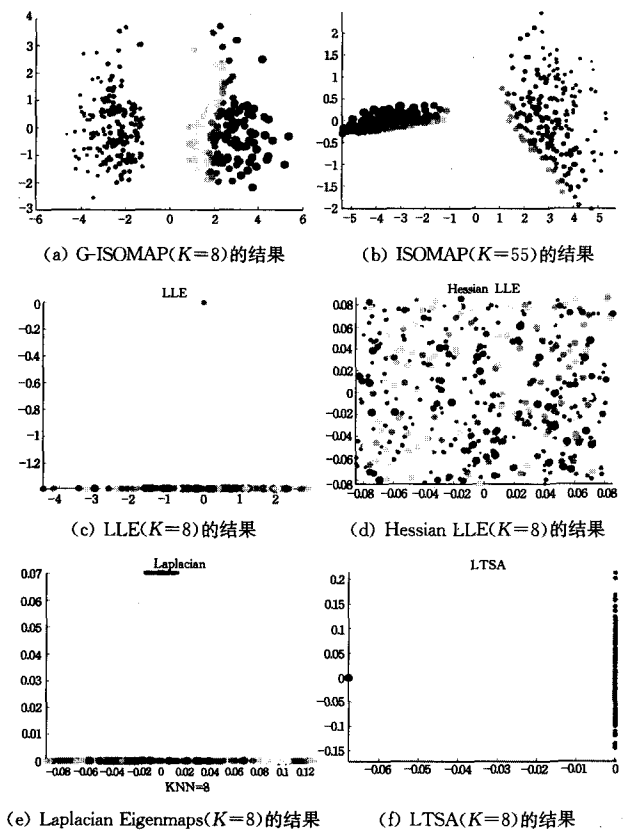


图7 常用流形学习算法在图3(b)数据集上的结果

本小节的实验以图3为数据集,将本文所提出的 G-ISOMAP 算法与 ISOMAP, LLE, Hessian LLE, Laplacian Eigenmaps 和 LTSA 进行嵌入结果比较,邻域参数采用 K 近邻方法,实验结果如图6、图7所示。其中对于图3(a)数据集,G-ISOMAP 算法的参数值为 $K=8, \gamma=15$ 。除 ISOMAP 算法之

外,邻域参数 K 的值与 G-ISOMAP 一样。ISOMAP 算法的邻域参数 $K=45$,原因是对于图2(b)的数据集,从实验中发现 K 的取值至少取45才能保证在整个数据集上产生一个邻域图。很显然,当 $K>45$ 时,将有更多的短路边出现,实验结果将会更差。同样,对于图3(b)数据集,G-ISOMAP 算法的参数值为 $K=8, \gamma=20$,ISOMAP 中的 $K=55$,其他算法的邻域参数 $K=8$ 。从实验结果可以看出,只有 G-ISOMAP 算法的结果与理想结果非常相似。

4.2 人脸数据集

采用 ISOMAP 算法中所使用的人脸姿势数据集^[16],该数据集共698幅,人脸姿势从面部朝左到朝右,从仰视、平视到俯视,部分典型人脸图像如图8所示。在本小节实验中,人工去掉在正面附近(包括正面)的图像序列,即去掉图8中红色虚线内的图像序列,共去掉131幅图像,将剩下的567幅图像作为 G-ISOMAP 和 ISOMAP 算法的输入,并进行效果比较。

通过实验发现,G-ISOMAP 算法在该数据集上取得最佳效果的参数值是 $K=4, \gamma=50$ 。为使 ISOMAP 算法在该567幅人脸图像上产生唯一邻域图,邻域参数 $K \geq 10$ 。实验结果分别如图9、图10所示。

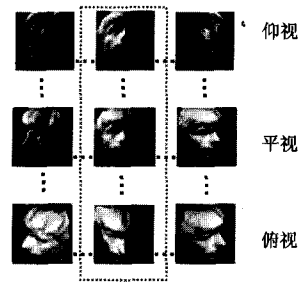


图8 典型人脸姿势图像

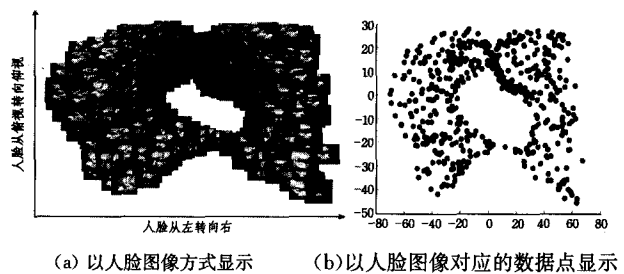


图9 G-ISOMAP 算法在人脸数据集的效果($K=4, \gamma=50$)

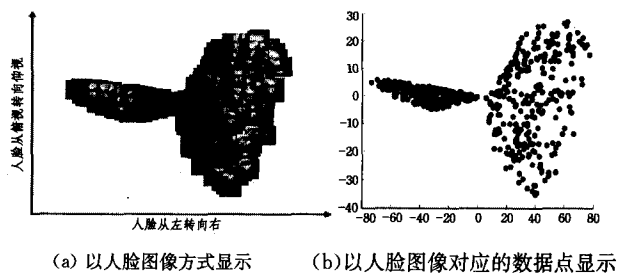


图10 ISOMAP 算法在人脸数据集效果($K=10$)

图9(a)说明了 G-ISOMAP 算法能够有效地发现该数据集的内在特征,横轴揭示人脸面朝左边到右边的变化趋势,纵轴揭示人脸仰视到俯视的变化趋势。图9(b)将图9(a)中的

每幅图像以空间中点的形式输出;图 9 分别以图像和空间中点的形式显示了 ISOMAP 算法在该数据集上的实验效果,可明显发现 ISOMAP 算法只能大致发现人脸面朝左边到右边的变化趋势,即沿图 10(a)的横轴方向。但图 10(a)的左半部分在纵轴方向上出现了过分收缩,即没有发现人脸从仰视到俯视的姿势变化趋势。

结束语 本文针对 ISOMAP 算法的不足,提出了一种适合于被间隔的流形学习算法 G-ISOMAP。该算法不仅对现有 ISOMAP 算法所适合的数据集有效,更重要的是它对于抽样于同一流形但中间有间隔的数据集也能得到比其他算法更满意的结果,所以本文的 G-ISOMAP 算法比 ISOMAP 算法更灵活,应用面更广。

不过需要提及的是:①如果被间隔的子流形间最短的欧氏距离对应的数据点恰好位于间隔区域的边缘,那么本文算法能够成功发现内在低维流形的结构,然而,所发现的数据点离间隔区域越远,效果就越不佳。②本文的假设是被间隔的数据是采样于一个流形的。如果数据集来源于多个类,每个类的数据集抽样于一个单独的流形,那么如何将本文的方法扩展到该情形。③目前主流的 ISOMAP 算法及其变种大多属于无监督范畴,如果完成了第②个工作,那么在此基础上,应用神经网络找到原始的高维数据到低维嵌入之间的映射函数,最后应用现有的分类器(如 K 近邻分类器等)来完成分类任务,从而可以使 ISOMAP 算法适合有监督框架。这 3 点也许是未来值得研究的问题。

参考文献

- [1] Seung H S, Lee D D. Manifold ways of perception[J]. Science, 2000, 290(5500): 2268-2269
- [2] Tenenbaum J B, de Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323
- [3] Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326
- [4] Donoho D, Grimes C. Hessian Eigenmaps: new locally linear embedding techniques for high-dimensional data[J]. PNAS, 2003, 100(10): 5591-5596
- [5] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396
- [6] He X, Niyogi P. Locality preserving projections[C]// Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2004: 327-334
- [7] Zhang Z, Zha H. Principal manifolds and nonlinear dimension reduction via local tangent space alignment[J]. Journal of Scientific Computing, 2005, 26(1): 313-338
- [8] Weinberger K Q, Saul L K. Unsupervised learning of image manifolds by semidefinite programming[J]. International Journal of Computer Vision, 2006, 70(1): 77-90
- [9] de Silva V, Tenenbaum J. Global versus local methods in nonlinear dimensionality reduction[C]// Becker S, Thrun S, Obermayer K, eds. Proc. NIPS. 2003, 15: 721-728
- [10] Choi H Y, Choi S J. Robust kernel isomap[J]. Pattern Recognition, 2007, 40(3): 853-862
- [11] Geng X, Zhan D C, Zhou Z H. Supervised nonlinear dimensionality reduction for visualization and classification[J]. IEEE Transactions on Systems, Man and Cybernetics, 2005, 35(6): 1098-1107
- [12] 邵超, 黄厚宽, 赵连伟. 一种更具拓扑稳定性的 ISOMAP 算法[J]. 软件学报, 2007, 18(4): 869-877
- [13] 周志华, 曹存根. 神经网络及其应用[M]. 北京: 清华大学出版社, 2004
- [14] Bernstein M, de Silva V, Langford J C, et al. Graph approximations to geodesics on embedded manifolds[R]. Stanford University, 2000
- [15] Balasubramanian M, Schwartz E L. The ISOMAP algorithm and topological stability[J]. Science, 2002, 295(5552): 7
- [16] <http://waldron.stanford.edu/~isomap/datasets.html>
- [5] Lee Y J, Hsieh W F, Huang C M. e-SSVR: a smooth support vector machine for e-insensitive regression[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(5): 678-685
- [6] Johan A K S, Tony V G, Jos D B, et al. Least Squares Support Vector Machine [M]. World Scientific, 2002: 182-183
- [7] 杜喆, 胡廷锋, 刘三阳. 用于回归的临近支持向量机[J]. 计算机科学, 2009, 36(1): 126-128
- [8] Peng X J. TSVR: An efficient Twin Support Vector Machine for Regression[J]. Neural Networks, 2010, 23: 365-372
- [9] Mangasarian O L, Musicant D R. Lagrangian support vector machine[J]. Journal of Machine Learning Research, 2001, 1: 161-177
- [10] Mangasarian O L, Musicant D R. Finite Newton method for Lagrangian support vector machine[J]. Neurocomputing, 2003, 55: 39-55
- [11] Balasundaram S K. On Lagrangian support vector regression[J]. Expert Systems with Applications, 2010, 37: 8784-8793
- [12] Bodyfat DataSets [EB/OL]. <http://lib.stat.cmu.edu/datasets/>
- [13] Mackey-Glass DataSets [EB/OL]. <http://www.cse.ogi.edu/~ericwan>
- [14] Trizaines DataSets [EB/OL]. <http://www.liaad.up.pt/~ltor-go/Regression/DataSets.html>
- [15] UCIdatabases [EB/OL]. <http://www.ics.uci.edu/~mllearn/ML-Repository.html>

(上接第 249 页)