第7卷第5期　　智能系统学报　　Vol.7 №.5
2012年10月　　CAAI Transactions on Intelligent Systems　　Oct. 2012

DOI:10.3969/j.issn.1673-4785.201204028
网络出版地址:http://www.cnki.net/kcms/detail/23.1538.TP.20120917.1702.002.html

增量与演化流形学习综述

谈超[1],关信红[1],周水庚[2,3]
(1.同济大学 计算机科学与技术系,上海 201804;2.复旦大学 计算机学院,上海 200433;3.复旦大学 上海市智能信息处理重点实验室,上海 200433)

摘　要:流形学习的目标是发现观测数据嵌入在高维数据空间中的低维光滑流形.近年来,在线或增量地发现内在低维流形结构成为流形学习的研究热点.从增量学习和演化学习2个方面入手,对该领域已有研究进展进行综述.增量流形学习较之传统的批量流形学习方法具有动态增量的能力,而演化流形学习能够在线地发现海量动态数据的内在规律,有利于进行维数约简和数据分析.文中对主要的增量与演化流形学习算法的基本原理、特点进行了阐述,分析了各自的优点与不足,指出了该领域的开放问题,并对进一步的研究方向进行了展望.

关键词:流形学习;增量流形学习;演化流形学习
中图分类号: TP181　**文献标志码**:A　**文章编号**:1673-4785(2012)05-0377-12

Incremental and evolutionary manifold learning: a survey

TAN Chao[1], GUAN Jihong[1], ZHOU Shuigeng[2,3]
(1. Department of Computer Science and Technology, Tongji University, Shanghai 201804, China; 2. School of Computer Science, Fudan University, Shanghai 200433, China; 3. Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai 200433, China)

Abstract: Manifold learning is to find the low-dimensional smooth manifold of observation data embedded in high-dimensional data space. In recent years, exploring the intrinsic low-dimensional manifold structure online or incrementally becomes a hot research topic in manifold learning area. This paper surveys the state of the art of incremental and evolutionary manifold learning, including the mechanisms and features of major existing incremental and evolutionary manifold learning methods, their advantages and disadvantages, and highlights the open research issues and future research directions.

Keywords: manifold learning; incremental manifold learning; evolutionary manifold learning.

流形学习是近10余年来发展起来的一种非线性维数约简方法,旨在发现嵌入在高维非线性数据空间的低维光滑流形的内在几何结构或内在维数,便于对数据的深层理解和进一步处理.近年来,流形学习在许多研究领域,包括数据挖掘、机器学习、图像处理和计算机视觉等都引起了广泛的关注.基于流形学习的非线性维数约减方法成为了机器学习中的研究热点[1].流形学习的基本前提是:观测空间中的点在高维观测空间张成一个流形,通过有效地展开观测空间中卷曲的流形或发现内在的主要变量,可以对数据集进行降维[2].流形学习的意义在于把人的认知流形规律引入到机器学习研究之中[3].当采样数据处于高维流形空间时,要从采样数据中学习并发现低维流形的内在几何结构,这意味着流形学习比传统的维数约减方法更能体现出事物的本质规律.

近10余年来出现了一些知名的线性与非线性的数据降维算法,如主成分分析(principle component analysis, PCA)[4]、等距映射(isometric mapping, Isomap)[5]、拉普拉斯特征映射(Laplacian eigenmap, LE)[6]、局部线性嵌入(local linear em-

收稿日期:2012-05-02.　**网络出版日期**:2012-09-17.
基金项目:国家自然科学基金资助项目(61173118).
通信作者:关信红. E-mail: jhguan@tongji.edu.cn.

bedding, LLE)[7]等. 尽管这些算法的提出大大推动了流形学习的发展,但它们都不具有动态增量处理数据的能力,难以满足计算效率的实际需求.

为了从高维数据流和大规模海量数据集中探索有价值的信息,人们迫切需要增量地发现内在低维流形结构. 许多实际应用如数据流挖掘、视频监控和语音识别都要求高维数据的实时嵌入. 一个简单的做法是:每当一个新数据点进来时,在所有已存在数据点上运行数据嵌入算法. 考虑到大多数数据嵌入算法具有至少需要 $O(n^2)$ 的时间复杂度(n 是数据点的数量),因此其时间复杂度过高,运算量过大. 增量学习方法就是为了解决上述问题而提出的. 实施增量流形学习主要考虑以下 2 个方面的问题:1) 如何保证在线状态下实时动态地处理增量数据;2) 如何有效地处理大规模海量数据的嵌入[8]. 当前,增量地处理新加入样本是流形学习和动态数据流分析的研究热点.

流形学习的主要目的是对高维数据集的观察或测量值建立预测模型,当有新的输入值时,这些预测模型就可以预测出相对的输出值. 多数流形学习算法都是通过优化某个代价函数来求得输出坐标,而演化学习是基于模仿生物演化行为而发展的最优化方法,可以适用于许多最优化问题;因此可以用于解决许多困难的机器学习问题,比如用在流形学习中的演化流形学习,为发现高维观察值的低维流形提供了新的途径.

本文立足于流形学习,对增量流形学习和演化流形学习的最新进展进行综述. 首先,对流形学习的研究背景和现状进行简要介绍,在此基础上,对增量流形学习算法进行分类,并对增量流形学习的主要算法进行综合对比与分析. 然后介绍了流形学习的另外一个方向:演化流形学习,概括了该领域的主要算法,包括非监督演化学习及监督演化学习. 最后讨论了增量流形学习及演化流形学习可扩展和待解决的问题,以及进一步的研究方向.

1 流形学习:概念与方法

1.1 流形学习的基本概念

流形学习的主要目标是发现嵌入在高维数据空间中观测数据的低维光滑流形. 流形学习对维数约减的过程可概括为:设数据是均匀采样于一个高维欧氏空间中,流形学习就是找到高维空间中的低维流形,并求出相应的嵌入映射,以实现维数约减[9].

流形学习的数学定义:设 $Y \subset \mathbf{R}^d$, $f: Y \to \mathbf{R}^M$ 是一个光滑嵌套, $M > d$, 流形学习的目标是基于 $\mathbf{R}^M$ 上的一个给定被观测数据集合 $\{\boldsymbol{x}_i\}$ 恢复 Y 与 f, 在 Y 中隐藏的数据 $\{\boldsymbol{y}_i\}$ 被映射到观测空间 $\mathbf{R}^M$, 使得 $\boldsymbol{x}_i = f(\boldsymbol{y}_i)$[10].

1.2 传统流形学习算法

近年来,流形学习领域出现了很多研究成果,在包括数据挖掘、机器学习、图像处理[11]和计算机视觉等领域得到了广泛应用[12-15]. 流形学习中的很多典型算法都是针对线性和非线性降维来进行的. 主成分分析算法(PCA)是线性降维方法中的代表;非线性方法中,发表在 2000 年《Science》上的等距映射算法(Isomap)[5]和局部线性嵌入算法(LLE)[7]是 2 个著名的流形学习降维算法. Isomap 算法使用最近邻图中的最短路径得到近似的测地线距离,代替不能表示内在流形结构的欧氏距离,然后用多尺度分析[16](multidimensional scaling analysis, MDS)进行处理,得到嵌入在高维空间中的低维坐标. LLE 能实现高维输入数据映射到一个全局低维坐标系中,同时保留了邻接点之间的关系和固有的非线性结构. 另外还有拉普拉斯特征映射算法(LE)、局部切空间排列算法(local tangent space alignment, LTSA)[17]等一系列著名的流形学习算法,在非线性降维方面均取得了显著的效果.

目前,流形学习中的非线性维数约减算法大部分都是应用于数据可视化,并已在人脸图像处理、手写数字图像及语言处理等方面取得了良好的效果.

1.3 传统流形学习算法分类

按原始观察空间与经过仿射变换后的嵌入空间保持邻域结构的不同方式,传统流形学习算法可划分为全局嵌入法和局部嵌入法.

1)全局嵌入法. 如等距离映射算法(Isomap),将流形上邻近的点映射到低维空间中的邻近点,同时将流形上距离远的点映射到低维空间中距离远的点,从而保持低维空间中点之间的距离关系及流形的结构.

2)局部嵌入法. 如局部线性嵌入算法(LLE)、拉普拉斯特征映射算法(LE)、局部切空间排列算法(LTSA)等,这些方法将流形上距离近的点映射到低维空间中的邻近点,得到局部空间的低维坐标,再通过线性嵌入、拉普拉斯映射及切空间排列调准等方法得到全局的低维坐标,从而实现流形学习降维.

1.4 传统流形学习算法的不足

传统流形学习方法在寻找规模不断增加的数据

集的内在规律时，新数据集到来后不是利用已经获得的低维流形结构，而是把新数据集和已有数据集合并成更大规模的数据集，通过重新学习来发现整个数据集的低维流形[8]. 这些方法的共同特点是以批量或者离线的方式处理数据，不具有增量学习的能力. 因此，传统流形学习算法不适用于增量学习.

2　增量流形学习

增量流形学习是针对传统批量流形学习算法的不足而发展起来的一种新兴流形学习算法，它在动态数据处理过程中新数据加入后，构建与原来数据集之间的邻域关系，重新表达加入新数据点后的高维数据集的嵌入空间，从而揭示高维空间中数据点之间的本质关系.

增量流形算法的一个优点是可以将数据流形的演化进行可视化. 当获得越来越多数据点时，流形变化的可视化能显示出数据流的一些性质. 适应性也是增量流形学习的一个优点——算法可以在数据逐渐变化中调整流形. 例如，假设学习 N 个个体的面部图像的流形. 经过一段时间后，不同人的脸部逐渐改变，这称为时间效应，是人脸识别中一个最具挑战性的研究工作. 如果面部图像的流形可以根据这些面部变化调整，系统的性能就能提高[17]. 实际应用中大量流数据的产生为增量流形学习提供了广阔的发展前景. 在数据挖掘中，数据通常是从一个数据流中有序地收集. 在这种情况下，如果能用新增数据点对已有学习结果进行有效的更新，那将会非常有用. 例如在图像检索[18]、人脸识别、数据可视化等应用领域，该技术能更好地描述图像的内蕴结构和语义关系.

2.1　增量流形学习的分类

目前增量流形学习方法主要可以分为 2 种：样本独立训练和样本依赖训练. 前者将新样本嵌入到新构建的子空间中，是全局嵌入法的增量形式；而后者则侧重保持局部的邻域结构，求解在局部邻接信息约束下的优化问题. 前者的优势是易于从理论角度进行理解，在表达数据全局结构的基础上进行新样本点的增量嵌入；后者一般只需要进行增量谱方法的计算，计算量上具有一定的优势. 表 1 所列为现有主要的增量流形学习方法，如 IDR(incremental dimension reduction algorithm)[19]、增量 IAM(incremental alignment method)[20]、谱嵌入增量流形学习算法[21]及增量 LLE 算法[22]等. 本文分别对其中主要的方法进行详细介绍.

表 1　主要的增量流形学习算法

Table 1　Major incremental manifold learning algorithms

分类	核心思想	主要算法
样本独立训练	从已经存在或是一个新的种类中计算新样本的低维嵌入，是子空间方法的增量版本.	IDR、谱嵌入增量流形学习算法、增量 PCA 算法、增量 Isomap 算法
样本依赖训练	保持数据集内部的局部邻接信息，通过已存在样本的邻接信息取得低维的嵌入.	IAM、增量 LLE 算法、基于正交迭代的增量 LLE 算法、增量 Laplacian 映射算法、增量 LTSA 算法

2.2　针对数据流的增量 Isomap 算法

考虑数据流的特点，Law 等提出了增量式的 Isomap 算法[1,23]，增量式学习不仅能够更有效地计算，同时能够发现流形结构演化的过程.

2.2.1 增量 Isomap 算法

增量 Isomap 的思想是通过更新坐标来保持最佳的测地距，其算法主要过程分为以下几步[1].

1)更新测地距. 就 Isomap 算法而言，对每个新增的数据点 y_{n+1} 都将引入一个新的顶点 v_{n+1} 到图 G 中. 然而，新增的顶点不仅会改变原有的邻域结构和一些已知的最短路径，也增加了新的路径. 在算法中，首先增加或移去某条与 v_{n+1} 相关的边. 点对之间的测地距离需要重新计算，这里使用一种改进的 Dijkstra 算法. 对于增加的边，需要检查是否存在新的最短路径；对于移去的边，需要重新计算所有曾经基于该边来计算的点对.

2)寻找新样本 x_{n+1} 的坐标. 匹配其与最接近目标值的样本 x_i 的内积形式，尽可能与目标值接近来确定其相应的位置.

3)全局坐标校准. 根据调整后的测地距离矩阵 $\boldsymbol{G}_{\text{new}}$，更新内在低维空间的数据点坐标. 这里存在 2 种更新方法，一种是对损失函数构造梯度下降算法来更新；另一种是子空间迭代，直接对调整后的矩阵作相应的特征分解，可视为求解增量的特征值问题，通过特征分解得到坐标.

2.2.2 增量 Isomap 算法的不足、扩展与改进

当加入一个样本点，可能会引起"短路边"出现，样本点对之间的测地距发生很大的变化，导致点的坐标产生很大的偏差. 一些扩展和改进增量 Isomap 算法的研究包括如下[23].

1)一个增量的测地距更新规则. 该测地距被用在增量 Isomap 中，通过改变测地距的稀疏性来提高坐标更新的效率.

2)增量地更新拓扑空间坐标的方法. 使用子空

间迭代方法来增量地更新插入新点以后的全局坐标,并使用 Rayleigh-Ritz 加速[24]. 该方法独立于测地结构的定义,故也可以被用在其他增量非线性维数约减方法中.

3)对已有的增量式方法的改进[23]. 由于 Isomap 的测地线计算是全局算法,因此,改进的算法并非是完全在线的. Isomap 是一个全局的算法,对任何新样本,需要考虑它是如何与其他样本相互影响的,为了找到其坐标,需要将所有数据点的几何信息都保存起来,不适宜在大数据集上使用. 解决该问题有 2 种方法:一种办法是当累积了足够数量的样本时忽略最旧的样本点,这给算法带来了适应性的优势;另一种是维持一组固定大小的“标记点”(landmark points),并只考虑新样本点与标记点之间的关系,最后,可以通过沿着流形的高斯分布来压缩数据点集,无需存储额外的信息.

在文献[25]中,作者提出了一种基于小世界模型的增量流形学习算法,将 Isomap 算法应用于增量处理新样本中. 首先,对于新样本点,在训练集中找出它的 k 近邻点及一些距离较远的点,接下来通过保持新样本和周围这些点的测地距离来获得新样本点的低维嵌入. 从而新样本可以有效地映射到低维空间中,该算法具有较低的复杂度.

2.3 针对 LLE 的 2 种增量算法

L. K. Saul 等提出的线性化的 LLE 算法假设流形是局部线性的,训练样本的投影值不会因新样本的加入而发生改变[26]. 这是一种线性近似的方法,实际上,当新样本点加入时,由于原始数据集中一些样本的 k 近邻点会发生改变,它们投影以后的值也会随之改变. 故 LLE 不适用于顺序到来的实时数据. 另外 LLE 进行降维的原始数据集必须是数量固定的,当新点加入时,LLE 必须在扩展后的数据集上重新运行,对新点不具有一般性,这使得 LLE 不能用于动态系统的高维数据集,时间复杂度过高. 针对 LLE 的这些不足,近年来出现了 2 种方法来实现增量计算.

1) O. Kouropteva 等提出的一种增量 LLE 算法[27].

当加入新样本点后,假设新代价矩阵 $\boldsymbol{M}_{\text{new}}$ 与原代价矩阵 $\boldsymbol{M}$ 的前 d 个最小特征值近似相等,通过最小化 $(\boldsymbol{Y}_{\text{new}}\boldsymbol{M}_{\text{new}}\boldsymbol{T}^{\text{T}}_{\text{new}}-\text{diag}(\boldsymbol{\lambda}_1,\boldsymbol{\lambda}_2,\cdots,\boldsymbol{\lambda}_d))$ 求出所有样本的低维映射. 该算法成功地求解了增量特征问题,但是存在 2 个缺点:①最优化问题. LLE 已将最优化问题转化为求解矩阵特征向量的问题,而这种增量 LLE 方法又重新回到了求解最优化问题,处理不便;②病态条件下的特征问题. 由于假设 $\boldsymbol{M}_{\text{new}}$ 的特征值与原代价矩阵 $\boldsymbol{M}$ 的前 d 个最小的特征值近似相等,随着新增样本数目的增加,它们之间的差值将越来越大,从而导致特征值和特征向量解的误差对微小的扰动非常敏感.

2)朱明旱等提出的一种基于正交迭代的增量 LLE 算法[28].

该算法分为 2 步:首先更新代价矩阵,在这个步骤中避免了一些重复的权值计算;然后不断地利用前一次的处理结果来计算各样本的投影值,避免了新样本加入时的重复计算,并将求投影坐标时所需的对高阶矩阵的分解转化为对低阶矩阵的分解. 此方法降低了分解矩阵的阶数和数据的运算量,从而提高了计算效率,较好地解决了在新样本不断加入的情况下总体流形不断更新的问题.

2.4 动态增殖流形学习算法

对于不断增加的海量观测数据集序列,不可能一次获得嵌入到高维数据空间上的所有数据点集,对于新来的观测数据子集,如何把其包含的几何信息融合到以前所获信息中去是一个要解决的问题. 对于已获得的观测数据集,要对整个数据集进行流形学习以发现其内在规律. 目前的流形学习算法对于海量观测数据往往计算复杂度过高. 流形学习的增量非线性维数约减和增量 LLE 是对新增单个数据点进行逐个更新,但逐点更新计算代价较高,并且新的观测区域数据的出现会破坏原有的几何结构.

为了解决以上问题,曾宪华等提出了一种动态增殖流形学习算法[8]. 这是通过整合重叠邻域中的信息来发现全局几何结构的一种方法,保证嵌入空间和内在低维空间对应数据点与局部邻域内的点保持相同的序关系,任何数据点和它的近邻点具有旋转、平移与伸缩不变性. 先用 LLE 计算各子集的流形,再将各子集的流形整合,得到整体流形结构,这是一种分批处理的思想. 利用这一思想,增殖流形学习算法处理不断增加的数据集时,对稠密的近邻或重叠数据子集(稠密是指该子集中的点能够反映嵌入流形上某一部分的内在几何结构)分块,发现其低维流形结构. 然后固定一块,对另一块施加平移、旋转和伸缩变换,使得两者的低维流形具有观察数据子集间相同的近邻关系,这样随着观测数据集的增多,通过平移、旋转及伸缩变换整合得到的流形更加逼近高维数据空间的内在低维全局结构. 这是一个增殖并保持原有信息的过程,因此称之为动态增殖流形学习.

增殖流形学习利用分治的方法和嵌入机理,把大任务分成多个不同的小任务,利用一种或多种流形学习算法实现,用于实际应用中大数据集或超大

数据集的内在几何结构及其内在规律的动态发现和整合,通过不断动态发现局部数据集的内在几何结构,拓扑成几何结构更加完善的低维流形.这种动态拓扑低维流形的方法是在线或者增量的方法,可以克服现有流形学习算法存在的缺陷[8].

该算法学习所需的数据子集必须是相邻或重叠的,否则发现的低维流形将有较大的扭曲,即整个数据集是非稠密的.非稠密的数据集上的增殖流形学习是该算法进一步的研究方向.

2.5 增量 LTSA 算法

如前所述,等距映射(Isomap)[5]、局部线性嵌入(LLE)[7]和局部切空间调准(LTSA)[17]等算法在一些人工合成的数据集上取得了满意的可视化效果,并在一些分类问题中得到了应用.但上述方法都是批处理模式,即在降维时所有数据都要准备好.而在监控等应用中,图像数据是逐步得到的,批处理模式需要大量的计算量,重复地进行批处理降维极其耗时[29].

针对批处理模式算法的不足,国内外学者开始考虑增量式的流形学习算法,这里讨论的是非监督式的非线性流形降维问题,对于属于非参数化的增量流形学习算法来说,很适合数据的逐步获取.增量式算法的另一个好处是能检测数据流中的渐进变化,可以很容易地修改以达到“遗忘”的效果[29].增量 Isomap 及标志点的增量式 Isomap 方法[23]可以降低在处理新数据点时的时间复杂度.通过使用标志点,可以降低 Isomap 算法的空间要求,能较好地支持大数据量的应用.但由于 Isomap 算法基于最短测地距,在处理新数据时,需要进行耗时的图重构,时间复杂度依然比较高[29].Liu 等提出的增量 LTSA 算法能快速处理新数据点,通过最小化新数据点与已有数据点的低维坐标的最小重构误差,得到新数据点的低维嵌入坐标[30].受标志点 Isomap 算法的启发,Yin 等的标志点 LTSA 算法能降低对内存的要求[31].下面介绍几种对应的增量式算法.

1)增量 LTSA 算法[30].

假设给定 n 个数据点 x_i 的全局低维坐标 t_i,对于新观测到的样本点 x_{n+1},更新现有的坐标 t_i 并给出 x_{n+1}的映射 t_{i+1}.算法由 3 步组成:首先,对于新数据点 x_{n+1},更新已有数据点的局部几何信息;接着使用 x_{n+1} 相对于已有数据点的局部几何信息估计 t_{n+1};最后更新所有数据点的全局低维坐标 t_i.算法具体步骤见文献[30].根据文献[30],增量 LTSA 算法最大的运算量是计算实对称半正定的排列矩阵 $\boldsymbol{B}$ 的最小特征向量集.当新数据点到达时,只改变部分数据点的近邻结构,而受轻度扰动的实对称矩阵的特征值与特征向量不会产生大的变化.这使得重用当前的变换矩阵和坐标来更新更加有效.更确切地说,因为 LTSA 算法中的排列矩阵 $\boldsymbol{B}$ 更具有局部性,新数据点的影响也更局部化,使得矩阵的更新更简单[29].相对于文献[23]中增量式的 Isomap 算法,LTSA 的增量式算法不需要费时的图重构过程而显得更有效.不过,相对于原始 LTSA 算法时间复杂度较高,增量 LTSA 算法使用了 Rayleigh-Ritz 加速方法来提高算法效率.

2)带标记点的增量 LTSA 算法.

在标记点 Isomap 算法[23]中,寻找保持一组标记点的测地距的映射,代替所有成对点的测地距,减少计算复杂度.标记点增量 LTSA 算法与之类似,令最先的 u 个点为标志点,在其中找 k 个最小距离的近邻点,构建局部切空间.新数据点的局部切空间使用标记点的局部几何信息构建,新点的低维全局坐标通过最小二乘问题求解.此方法解决了测地距离矩阵存储空间大及计算复杂度高的问题,为在大数据集上的使用提供了便利.

2.6 增量拉普拉斯特征映射算法

Jia 等提出的增量拉普拉斯特征映射算法[32],通过一定意义上局部邻域信息的理想化保持计算数据集的低维表示,提出子流形分析算法及线性增量的模式来增量地学习新样本点并将其映射到低维空间.该算法将局部线性重构机制用于更新已存在样本的低维嵌入结果,并加入新的邻接信息.更新机制类似于迭代算法,每当观测到一个新样本,只更新邻域发生变化的样本点,局部地改进现有样本的嵌入结果.算法的主要步骤如下.

1)更新邻接矩阵 $\boldsymbol{w}$.构造新样本点 x_{n+1}与已知样本点之间的权重,重构样本点之间因新点插入而改变的权重.

2)在低维空间映射新点.这里作者给出了 2 种方法,第 1 种是线性增量法,最小化加权距离目标函数从而得到新点的低维映射 y_{n+1}.第 2 种是在新点的 k 个最近邻域即子流形上应用拉普拉斯映射,通过构建子邻接矩阵并计算特征值和特征向量来得到新点在子流形中的低维坐标.接着计算新点的全局坐标 y_{n+1},计算过程可看作是坐标变换问题,通过变换保持新点与其 k 个最近邻点之间的关系.通过在子流形上使用拉普拉斯映射,算法检测到新点 x_{n+1}与已知点之间的内在结构信息,再计算它们之间的约束权重矩阵来最小化重构误差.服从该约束的最佳权重可通过解最小二乘问题得到,再由权向量得出 x_{n+1}的全局坐标 y_{n+1}.

3)更新已存在样本点的低维嵌入坐标. 若已存在样本点的邻域由于 x_{n+1} 的插入而改变,除了计算 y_{n+1},它们的嵌入坐标也需要被更新. 这里使用局部线性重构机制来加入新的邻接信息,并修改已存在样本的低维嵌入结果. 更新分为 2 步:首先找到每个已有点 x_i 周围的 k 个最近邻点并计算重构权重来最小化耗费函数 ε;然后基于这些权重来找每个点的低维坐标 y_i. 这里,某个点有可能不能立即得到精确的低维嵌入坐标,但随着增量学习的更新步骤,低维坐标可不断被调整,直到近似最优化.

该算法的增量模式可以被一般化到其他非线性流形学习算法中去,例如 LLE 和 Isomap. 与一般迭代算法不同的是,增量拉普拉斯特征映射算法很简单,当依次观测样本时可以实现在线学习.

2.7 增量等距嵌入算法

Zhao 等提出了增量等距嵌入的方法[33-34],通过只映射新的数据点,并调整存在的嵌入结果,用增量的方法来产生新的嵌入结果. 此外,这些方法也可以删除已存在的数据点并相应地调整嵌入的结果,通过在数据流上建立"滑动窗口"并嵌入增量数据,来约减一个无界数据流的维数.

2.8 增量校准算法

与已有流形学习算法相比,为保持度量,避免陷入局部最小的问题,Han 等提出了增量校准算法[20]. 该算法属于局部保持映射,其主要思想是增量地校准输入数据的局部坐标,通过成块地对齐邻域信息,迭代产生整个高维数据集的低维映射. 该方法包括 2 个步骤:增量和校准. 第 1 个步骤增量地寻找邻域块用于接下来的校准;第 2 个步骤迭代地对齐邻域块的低维坐标,产生整个数据集的低维嵌入. 该算法与增量 LTSA 算法和增量等距嵌入算法都具有一定的相似性,都是通过局部校准邻域信息来增量学习高维数据,从而得到其低维的嵌入坐标. 这种增量校准算法能适应不断增加的观测数据集的处理需求,对在线数据流处理、图像检索等方面具有良好的应用价值.

2.9 高维增量在线学习

在最近的统计学习中,高维输入数据的非线性近似函数是一个重要的问题,特别是在增量和实时的情况中. 越来越多的问题领域中,增量和实时这 2 个特性都很重要. 例如动态进程的在线模型,由可视化监督所观测,高级计算机接口的用户模型和数值函数学习,控制模型特别是在高维移动系统如人类或仿真机器人的背景中. 针对这些任务的理想算法需要避免潜在数量上的问题,如输入数据的冗余,消除不恰当的输入维数,在数据保持有效的情况下学习更新的计算复杂度较低,允许在线增量学习,并获得准确的函数近似及足够的一般性.

局部加权回归映射(locally weighted projection regression,LWPR)[35]是高维空间中具有冗余输入维度的增量非线性近似函数的新算法,其核心是应用局部线性模型的非参数化回归. 为了保持计算效率和数量的鲁棒性,每个局部模型以部分最小二乘回归意义,在输入空间里用所选方向的一个单边变化回归的较小数,预先形成了回归分析. Sethu 等讨论了局部学习技术是如何成功地用在高维空间中,并回顾了各种局部维数约减技术,提出了 LWPR 算法. LWPR 的特点是:1)应用基于增量训练的第二顺序学习法快速学习;2)使用统计上有效的随机 leave-one-out 交叉确认来学习,无需存储训练数据;3)仅依赖于局部信息来调整其加权核,最小化增量学习负面干扰的危险;4)具有与输入数量呈线性的计算复杂度;5)可以处理大量的(可能是冗余的)输入,例如上升到 90 维的数据集中的各种经验估计,一般解释为产生了预变异和信任距离.

文献[35]提出,LWPR 是第一个真正的增量空间局部学习方法,可以成功并有效地用于非常高维的空间中.

2.10 增量流形学习算法小结

1)技术分类.

①邻接矩阵的更新. 更新邻接信息矩阵对于增量学习来说是必不可少的步骤. 一些增量算法如增量 LLE、增量拉普拉斯特征映射和增量 LTSA 等,保持数据集内部的局部邻接信息并通过已存在样本的邻接信息取得低维的嵌入.

②迭代的使用. 增量校准算法是一种典型的代表,它在找到邻域块之后,对齐它们的低维坐标,多次使用迭代的方法来达到增量学习的目的.

2)算法的区别和相似性.

①全局和局部的区别. 增量 Isomap 保持了原 Isomap 算法的全局性质,而增量 LLE 仍然是研究新样本点与邻域之间的局部关系,这一点在增量拉普拉斯特征映射和增量 LTSA 中的第 2、第 3 步也有体现,基于局部线性映射的增量学习算法[36]也属于这种类型.

②步骤的相似性. 通过上面几种主要增量流形学习降维算法的阐述,可以看出,增量 Isomap 算法、增量 LTSA 和增量拉普拉斯特征映射都包括 3 个步骤. 其中增量 Isomap 包括:首先更新测地线距离;接着更新坐标;最后寻找新样本 x_{n+1} 的坐标. 增量拉普拉斯特征映射为:第 1 步,更新邻接矩阵 $\boldsymbol{w}$;第 2 步,

在低维空间映射新点；第 3 步，更新已存在样本点的嵌入结果. 增量 LTSA 首先根据新数据点 x_{n+1} 更新已有数据点的局部几何信息；接着使用 x_{n+1} 相对于已有数据点的局部几何信息估计全局低维坐标；最后更新所有数据点的全局低维坐标. 可以看出，增量拉普拉斯特征映射和增量 LTSA 的步骤非常类似.

③处理数据的类型以及数学表达的方式. 增量 Isomap 是等距嵌入的算法，它主要是从保持测地线的角度来处理新加入的样本. 增量 LLE 是通过解协方差矩阵的特征值和特征向量来求新点在低维坐标下的映射. 增量拉普拉斯特征映射和增量 LTSA 都是从数据点的局部几何信息出发，根据新点与其相邻的已有点之间的邻接关系来估计其全局坐标，是通过研究邻域关系进行的嵌入算法. 表 2 对上文提到的增量流形学习算法进行了对比分析.

3）主要存在的问题.

Isomap 的测地线计算是一个全局的算法，因此，增量 Isomap 算法并非是完全在线的，对任何样本，在找到其坐标之前需要考虑它是如何与其他样本相互影响的；增量 LLE 牵涉到最优化问题，LLE 本身是将最优化问题转化为求解矩阵特征向量的问题，而增量 LLE 实际上又重新回到了最优化问题，并且由于新点的特征值没有更新，随着新增样本数目的增加，新点与已知点的前 d 个最小的特征值间的差值会越来越大，从而导致误差会越来越大. 表 3 对部分增量流形学习算法的优缺点及适用范围进行了整体的概括和评价.

表 2　增量流形学习算法的对比分析

Table 2　Comparative analysis of incremental manifold learning algorithms

算　法	全局/局部	处理数据类型	数学表达方式
增量 Isomap	全局性	保持测地线	等距嵌入
增量 LLE	局部性	协方差矩阵	最优化问题
动态增殖流形学习算法	局部性	拓扑几何结构	分治和嵌入
增量 LTSA	局部性	数据点局部几何信息	嵌入算法
增量 LE	局部性	数据点局部几何信息	嵌入算法
Incremental K-MST	局部性	局部邻域信息	等距嵌入
LWPR	局部性	输入空间所选方向的较小数	部分最小二乘回归

表 3　部分增量流形学习算法的评价及适用范围

Table 3　Evaluation and applicable field of incremental learning methods

算　法	优　势	缺　陷	适用范围
增量 Isomap	能够发现流形结构演化的过程	属于全局算法，并非完全在线	基于可视化数据流进行增量的流形学习
增量 LLE（分为 ILLE 和基于正交迭代的增量 LLE 算法）	基于正交迭代的增量 LLE 算法有效利用原处理结果，避免重复计算，降低了数据的运算量，提高了处理速度	对于最优化问题处理不便，误差会随着新增样本数目的增加而增大	针对现实世界的非线性流形学习，处理非均匀分布的数据集
增量拉普拉斯特征映射	实现容易，当样本依次被观测时可以实现在线学习	时间复杂度高	适用于在线学习
增量 LTSA	局部化算法，使矩阵的更新更简单，新数据点的影响更局部化	时间复杂度较高，为提高算法效率需使用 Rayleigh-Ritz acceleration 方法加速	能检测数据流中的渐进变化，适合数据的逐步获取

3　演化流形学习

3.1　演化学习的概念

实际系统往往是动态的，数据沿着时间不断变化，静态数据上的常规学习并不适合. 针对动态问题的解法包括：1）在数据集上应用静态算法；2）在每个时间点上对数据应用静态算法，包括时间进化机制和动态预测.

演化学习是在时间进化数据上的学习问题，学习任务包括静态状态和时间进化机制的学习，可分为半监督学习（分类/回归）和非监督学习（聚类/密度估计）等.

1）增量学习：根据有限内存或在线设备，数据组织成批量或流状模型. 与演化学习的不同在于：增量学习假设所有数据来自同一分布，对整个数据集的输出单一，如分类、回归、数据分类或密度模型. 增

量算法的期望对数据顺序敏感,因此不对临时数据进行推进.

2)在线学习:实际学习环境中的一种特殊类型,即连续循环的一组序列. 当在时间 t 预测时,学习者只能观察到 t 为止的样本,故动态地更新以寻求较低的遗憾值(regret). 与演化学习不同,在每个时间点 t,在线学习观察单个样本,而演化学习观察一组样本,故演化学习可以在线或离线地处理数据[37-38]. 在线学习要求每个样本点的回馈,而演化学习中,一组样本可能是标记的、未标记的、或部分标记的. 最小化在线学习与离线学习相应的遗憾值在每个时间点不存在一般化问题. 而演化学习旨在达到 2 个目标:在每个时间点的损失一般化及时间进化机制的学习.

演化学习与增量学习、在线学习[35]的比较如表 4 所示.

表 4 演化学习算法与其他学习算法的比较

Table 4 Comparison of evolution learning algorithms with other learning methods

算法	算法思想	性能特点
增量学习	基于流形学习的增量方法,对动态增量的数据流模型进行学习	假设所有数据来自同一分布,对整个数据集的输出单一,如分类、回归、数据分类或密度模型. 不对临时数据进行推进,增量算法期望对数据顺序敏感.
在线学习	当在时间 t 预测时,学习者只观察到 t 为止的样本,动态地更新来寻求较低的遗憾值(regret)	要求每个样本点的回馈,观察单个样本.
演化学习	在时间进化数据上的学习	在时间 t 对样本 x_t 的分布 P_t 进行学习. 可以在线或离线处理数据,在每个时间点上对数据应用静态算法,包括时间进化机制和动态预测.

前面提到,演化学习是一种基于演化发展的最优化方法,可以用于解决机器学习中的最优化问题,用在流形学习中,为发现高维数据集的低维流形提供了新的途径. 首先,当数据分布移动时,可能不适合采用传统的聚类算法来处理整体数据;其次,如果对每个新样本点独立使用传统的聚类算法,不能保留聚类结果沿着时间的连贯性. 这里就体现出演化流形学习的必要性,下面将从非监督和半监督 2 个方面来对演化流形学习算法进行讨论.

3.2 非监督演化学习

近年来演化聚类成为数据挖掘中提取信息的一项新的研究热点. 在流形学习领域,聚类是一个基本的问题[39-41]. 传统的聚类算法,如 K-means[42] 和谱聚类[43]都是基于静态数据,并假设所有数据是独立同分布(independent and identically-distributed, I. I. D)的,即样本来自同一个潜在的分布. 在这种数据上的聚类任务产生出了演化聚类问题[44]. 此外,与增量演化聚类[45]不同的是,由于数据的分布时刻彼此相邻,沿着时间顺序的聚类结果应该是连续的. 演化聚类的目标包括:对每个时间点 t,输出 x_t 上的一个分布 π_t;保持分类的质量和沿 t 分布 $\{\pi_t\}_{t=1}^{T}$ 的光滑性;聚类追踪等. 演化聚类的意义在于:可解释机器学习及数据挖掘算法;保持沿着时间的聚类结果的连贯性,特别是在可视化方面[44];聚类追踪为动态网络行为的分析提供有力对策. 由于种种原因,例如数据大小的变化、聚类数目的变化、聚类沿着时间的动态行为(出生、交叉和死亡等),演化聚类具有很大的挑战性. 前面一节提到,当数据分布移动时,采用传统的聚类算法处理整体数据时,不能保留聚类结果沿着时间的连贯性. 故将演化聚类用在流形学习中,对于处理动态高维数据集的低维流形并得出其聚类结果具有积极意义.

演化谱聚类和演化聚类是 2 种基本的数据组成方式[46-47]. 目前研究的主要问题包括:1)什么是演化;2)光滑代表什么;3)演化 K-均值、演化 GMM 等方法是否有一般化的方法;4)聚类行为;5)聚类数目和数据大小的变化.

Zhang 等提出了在线演化幂簇混合算法(online evolutionary exponential family mixture)[48],该方法提出了对聚类问题的密度估计的观点,针对幂簇混合模型(K-均值模型和多项式混合模型)进行密度估计,属于非监督的算法. 它采用了期望-最大化(expectation-maximization, EM)方法进行计算:E 步属于幂簇混合(exponential family mixture, EFM)的聚类问题;M 步使用 EFM 的闭合形式解法. Zhang 等还提出了 2 个框架:第 1 个框架依赖历史数据,即当前的模式分布近似于历史数据的分布;第 2 个框架依赖于历史模型,即当前的模型分布近似于历史模型分布. 这 2 种框架都基于 EFM 模型[44],该方法用于流形学习中,对动态高维数据集进行数据降维以后的聚类,属于演化流形学习的非监督方法,有待解决的是聚类追踪的问题.

3.3 半监督演化学习

演化分类是指学习一条针对不同时期的演化分类链. 它的意义在于当使用历史分类器或所有历史数据都无用的情况下，历史分类信息可能有助于分布缓慢进化，可以使用而不是抛弃历史标记. Jia 等提出了半监督演化分类算法（semi-supervised evolutionary classification）[49]，将一般学习问题中的光滑假设扩展到演化数据中. 光滑假设是指：若 2 个数据点靠得很近，容易有近似的标记. 而演化光滑假设是指：若时间点 t_1 和 t_2 靠得很近，两分类函数 f_{t_1} 和 f_{t_2} 容易近似. 实现以上假设的直接方法是对时间 t 积分，当 t 不连续时，用反向差异来近似以上积分，这是学习算法的一种时间调整.

学习数据的自然演化信息在机器学习研究中是一种新的挑战. 上述算法属于演化流形学习的半监督方法，应用在真实世界的流形学习中，处理经过降维以后的实时数据，并对数据集学习了一系列演化的分类函数，无论是稳定性还是精确性方面都产生了更好的性能.

3.4 演化流形学习小结

一种超越了独立同分布假设（I. I. D）[42-43]的学习问题的新类型称为演化学习. 最近又提出了一些非监督算法[50-53]、半监督算法[54-55]和有监督的流形学习算法[56-57]，在线增量学习的向量量化策略[58]，以及增量谱聚类[59]等演化流形学习算法，取得了令人满意的结果. 但这些仅仅是探索工作，未提供理论分析. 演化学习主要是基于模仿生物演化行为而发展的最优化算法，用于学习高维流形时，可以帮助人们解决许多困难的流形学习问题. 无论是非监督演化聚类还是半监督演化分类，都是建立在已有监督或非监督学习算法之上，引进了时间进化的条件. 相信通过进一步的理论探索，无论是增量学习、在线学习还是演化学习都将在流形学习领域产生新的研究点和研究价值.

4 总结与展望

目前，流形学习在许多领域都有广泛的应用，例如模式识别、统计回归、智能控制、生物信息、数据挖掘、数据压缩、时间序列分析等，是近年来机器学习领域的研究热点. 而增量流形学习能发现大规模海量数据集的内在低维信息，更好地解决在线和增量的问题；演化学习在流形学习中的应用，解决了传统流形学习存在的不足，比如观察数据集的移动分布，及时间连贯性等问题，学习数据的自然演化信息，建立更接近真实世界的优化模型.

现有流形学习算法主要面临的问题是在线环境下的误差、优化及收敛性等问题. 这与邻域图上各点邻域点的选取、坐标映射过程中方程的求解以及算法收敛性等方面都有一定关系. 未来工作可以从这几个方面展开，以解决存在的这些问题. 另外还有以下几个方面值得进一步地研究.

1）减少算法中参数的依赖性. 目前的增量流形学习算法中近邻数是变量，如何依据要处理的问题和数据的分布自适应地选择适当的局部近邻数值得进一步研究.

2）目前的增量流形学习算法大多属于非监督学习，如何提高新增样本点的学习能力是一个重要的问题. 虽然标记点技术已用于增量 Isomap 和增量 LTSA 等算法中，但它们还不具有监督意义. 可以考虑将监督算法与增量流形学习领域的算法相结合，提高在线学习的能力.

3）提高算法的效率. 增量流形学习一般比批量算法的时间复杂度更高，如何对已有算法进行加速，使之应用到对计算时间要求较高的新领域，也是值得深入研究的问题.

4）当样本点包含噪声点时，原始的流形学习算法（如 ISOMAP 和 LLE）受到了很大的影响，增量 Isomap 算法也存在类似的问题. 如何减少噪声干扰在增量流形学习中也是一个重要的问题.

5）就目前来看，许多实际应用中的数据集都存在规律性，因此研究流形与数据集的关系成为了一个重要领域. 在将来的工作中，必须找到一个行之有效的方法来分析高度相关、复杂分布的数据集的内在结构. 最近 Zhang 等提出的自适应流形学习算法通过自适应地选择局部邻域结构，将局部结构拼接在一起产生全局参数信息，解决了局部依赖问题[60]. 怎样使用增量流形的性质来模拟未知的概念和设计新的算法是今后研究的另一个方向.

6）演化学习的产生是为了解决现有学习算法与实际系统中的困境，目前出现的成果还仅是一些探索工作，包括非监督演化学习及半监督演化学习. 对这些算法进行更深入的理论分析，并在更多的数据集上进行实验来评价算法的性能和价值，是值得进一步探讨的问题.

参考文献：

[1] LAW M, ZHANG Nan, JAIN A K. Nonlinear manifold learning for data stream[C]//Proceedings of the Fourth SIAM International Conference on Data Mining. Lake Buena Vista, USA, 2004: 33-44.

[2]徐蓉,姜峰,姚鸿勋,等.流形学习概述[J].智能系统学报, 2006, 1(1): 44-51.
XU Rong, JIANG Feng, YAO Hongxun, et al. Overview of manifold learning[J]. CAAI Transactions on Intelligent Systems, 2006, 1(1): 44-51.

[3]SEUNG H, LEE D. The manifold ways of perception[J]. Science, 2000, 290(5500): 2268-2269.

[4]PEARSON K. On lines and planes of closest fit to systems of points in space[J]. Philosophical Magazine, 1901, 2(6): 559-572.

[5]TENENBAUM J, DE SILVA V, LANGFORD J. A global geometric framework for nonlinear dimensionality reduction[J]. Science, 2000, 290(5500): 2319-2323.

[6]BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[J]. Neural Computation, 2003, 15(6): 1373-1396.

[7]ROWEI S, SAUL L. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323-2326.

[8]曾宪华,罗四维.动态增殖流形学习算法[J].计算机研究与发展, 2007, 44(9): 1462-1468.
ZENG Xianhua, LUO Siwei. A dynamically incremental manifold learning algorithm[J]. Journal of Computer Research and Development, 2007, 44(9): 1462-1468.

[9]DE SILVA V, TENENBAUM J B. Global versus local methods in nonlinear dimensionality reduction[M]//BECKER S, THRUN S, OBERMAYER K. Advances in Neural Information Processing Systems. Cambridge, USA: The MIT Press, 2003: 721-728.

[10]BERGER M, GOSTIAUX B. Differential geometry: manifolds, curves and surfaces[M]. [S.l.]: Springer-Verlag, 1988: 474.

[11]PLESS R, SOUVENIR R. A survey of manifold learning for images[J]. IPSJ Transactions on Computer Vision and Applications, 2009, 1: 83-94.

[12]BREGLER C, OMPHUNDRO S M. Nonlinear manifold learning for visual speech recognition[C]//Proceedings of the 5th International Conference on Computer Vision. Washington, DC, USA: IEEE Computer Society, 1995: 494-499.

[13]HADID A, KOUROPTEVA O, PIETIKANINEN M. Unsupervised learning using locally linear embedding: experiments in face pose analysis[C]//Proceedings of the 16th International Conference on Pattern Recognition. Quebec City, Canada, 2002: 111-114.

[14]JENKINS O C, MATARIC M J. A spatiotemporal extension to Isomap nonlinear dimension reduction[C]//Proceedings of the 21th International Conference on Machine Learning. New York, USA, 2002: 2551-2556.

[15]NISKANEN M, SILVEN O. Comparison of dimensionality reduction methods for wood surface inspection[C]//Proceedings of the 6th International Conference on Quality Control by Artificial Vision. Gatlinburg, USA, 2003: 178-188.

[16]KRUSKAL J B, WISH M. Multidimensional scaling[M]. Beverly Hills, USA: Sage Publications, 1977.

[17]ZHANG Zhenyue, ZHA Hongyuan. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment[J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313-338.

[18]LU Ke, HE Xiaofei. Image retrieval based on incremental subspace learning[J]. Pattern Recognition, 2005, 38(11): 2047-2054.

[19]YE Jieping, LI Qi, XIONG Hui, et al. IDR/QR: an incremental dimension reduction algorithm via QR decomposition[J]. IEEE Transactions on Knowledge and Data Engineering, 2005, 17(9): 1208-1222.

[20]HAN Zhi, MENG Deyu, XU Zongben, et al. Incremental alignment manifold learning[J]. Journal of Computer Science and Technology, 2011, 26(1): 153-165.

[21]LI Housen, JIANG Hao, BARRIO R, et al. Incremental manifold learning by spectral embedding methods[J]. Pattern Recognition Letters, 2011, 32(10): 1447-1455.

[22]KOUROPTEVA O, OKUN O, PIETIKAINEN M. Incremental locally linear embedding algorithm[C]//Proceedings of the 14th Scandinavian Conference Image Analysis. Joensuu, Finland, 2005: 521-530.

[23]LAW M H C, JAIN A K. Incremental nonlinear dimensionality reduction by manifold learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(3): 337-391.

[24]GOLUB G H, VAN LOAN C F. Matrix computations[M]. Baltimore, USA: Johns Hopkins University Press, 1996: 1-694.

[25]SHI Lukui, YANG Qingxin, LIU Enhai, et al. An incremental manifold learning algorithm based on the small world model[C]//Proceedings of the 2010 International Conference on Life System Modeling and Intelligent Computing, and 2010 International Conference on Intelligent Computing for Sustainable Energy and Environment. Wuxi, China, 2010: 324-332.

[26]SAUL L K, ROWEIS S T. Think globally, fit locally: unsupervised learning of low dimensional manifolds[J]. Journal of Machine Learning Research, 2003, 4: 119-155.

[27]KOUROPTEVA O, OKUN O, PIETIKANEN M. Incremental locally linear embedding[J]. Pattern Recognition, 2005, 38(10): 1764-1767.

[28]朱明旱,罗大庸,易励群,等.基于正交迭代的增量LLE算法[J].电子学报, 2009, 37(1): 132-136.
ZHU Minghan, LUO Dayong, YI Liqun, et al. Incremental locally linear embedding algorithm based on orthogonal

iteration method[J]. Acta Electronica Sinica, 2009, 37(1): 132-136.

[29]刘小明.数据降维及分类中的流形学习研究[D].杭州:浙江大学, 2007: 1-108.
LIU Xiaoming. Research on data dimension reduction and manifold learning in classification[D]. Hangzhou: Zhejiang University, 2007: 1-108.

[30]LIU Xiaoming, YIN Jianwei, FENG Zhilin, et al. Incremental manifold learning via tangent space alignment[C]//Proceedings of the Second International Conference on Artificial Neural Networks in Pattern Recognition. Ulm, Germany, 2006: 107-121.

[31]YIN Jianwei, LIU Xiaoming, FENG Zhilin, et al. A local tangent space alignment based transductive classification algorithm[C]//Proceedings of the Second International Conference on Artificial Neural Networks in Pattern Recognition. Ulm, Germany, 2006: 93-106.

[32]JIA Peng, YIN Junsong, HUANG Xinsheng, et al. Incremental Laplacian eigenmaps by preserving adjacent information between data points[J]. Pattern Recognition Letters, 2009, 30(16): 1457-1463.

[33]ZHAO Dongfang, YANG Li. Incremental isometric embedding of high-dimensional data using connected neighborhood graphs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(1): 86-98.

[34]ZHAO Dongfang, YANG Li. Incremental construction of neighborhood graphs for nonlinear dimensionality reduction[C]//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China, 2006: 177-180.

[35]VIJAYAKUMAR S, DSOUZA A, SCHAAL S. Incremental online learning in high dimensions[J]. Neural Computation, 2005, 17(12): 2602-2634.

[36]FRITZKE B. Incremental learning of local linear mappings[C]//Proceedings of the International Conference on Artificial Neural Networks. Paris, France, 1995: 217-222.

[37]WANG Yi, LIU Shixia, FENG Jianhua, et al. Mining naturally smooth evolution of clusters from dynamic data[C]//Proceedings of the SIAM International Conference on Data Mining. Minneapolis, USA, 2007: 125-134.

[38]AHMED A, XING E. Dynamic non-parametric mixture models and the recurrent Chinese restaurant process: with applications to evolutionary clustering[C]//Proceedings of the SIAM International Conference on Data Mining. Atlanta, USA, 2008: 219-230.

[39]SOUVENIR R, PLESS R. Manifold clustering[C]//Proceedings of the 10th IEEE International Conference on Computer Vision. Beijing, China, 2005: 648-653.

[40]CAO Wenbo, HARALICK R. Nonlinear manifold clustering by dimensionality[C]//Proceedings of the 18th International Conference on Pattern Recognition. Hong Kong, China, 2006: 920-924.

[41]XU Rui, WUNSCHLL D. Survey on clustering algorithms[J]. IEEE Transactions on Neural Networks, 2003, 16(3): 645-678.

[42]HARTIGAN J A, WONG M A. A k-means clustering algorithm[J]. Applied Statistics, 1979, 28(1): 100-108.

[43]NG A Y, JORDAN M I, WEISS Y. On spectral clustering: analysis and an algorithm[C]//Neural Information Processing Systems: Natural and Synthetic. Vancouver, Canada, 2001: 849-856.

[44]CHAKRABARTI D, KUMAR R, TOMKINS A. Evolutionary clustering[C]//Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Philadelphia, USA, 2006: 554-560.

[45]CHARIKAR M, CHEKURI C, FEDER T, et al. Incremental clustering and dynamic information retrieval[C]//Proceedings of the Twenty-Ninth Annual ACM Symposium on the Theory of Computing. El Paso, USA, 1997: 626-635.

[46]CHI Yun, SONG Xiaodan, ZHOU Dengyong, et al. Evolutionary spectral clustering by incorporating temporal smoothness[C]//Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Jose, USA, 2007: 153-162.

[47]TANG Lei, LIU Huan, ZHANG Jianping, et al. Community evolution in dynamic multi-mode networks[C]//Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA, 2008: 677-685.

[48]ZHANG Jianwen, SONG Yangqiu, CHEN Gang, et al. Online evolutionary exponential family mixture[C]//Proceedings of International Joint Conference on Artificial Intelligence. Pasadena, USA, 2009: 1610-1615.

[49]JIA Yangqing, YAN Shuicheng, ZHANG Changshui, et al. Semi-supervised classification on evolutionary data[C]//Proceedings of International Joint Conference on Artificial Intelligence. Pasadena, USA, 2009: 1083-1088.

[50]SANGER T D. Optimal unsupervised learning in a single-layer linear feed-forward neural network[J]. IEEE Transactions on Neural Networks, 1989, 1(2): 459-473.

[51]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.

[52]GOH A, VIDAL R. Segmenting motions of different types by unsupervised manifold clustering[C]//IEEE Conference on Computer Vision and Pattern Recognition. Minneapolis, USA, 2007: 1-6.

[53]GOH A, VIDAL R. Unsupervised riemannian clustering of probability density functions[C]//Proceedings of the 2008 European Conference on Machine Learning and Knowledge Discovery in Databases. Antwerp, Belgium, 2008: 377-392.

[54] HE Xiaofei. Incremental semi-supervised subspace learning for image retrieval[C]//Proceedings of the 12th Annual ACM International Conference on Multimedia. New York,USA,2004: 2-8.

[55] BELKIN M,NIYOGI P. Semi-supervised learning on riemannian manifolds[J]. Machine Learning,2004,56(1/2/3): 209-239.

[56] 孟德宇,徐宗本,戴明伟.一种新的有监督流形学习方法[J].计算机研究与发展,2007,44(12): 2072-2077.
MENG Deyu,XU Zongben,DAI Mingwei. A new supervised manifold learning method [J]. Journal of Computer Research and Development,2007,44(12): 2072-2077.

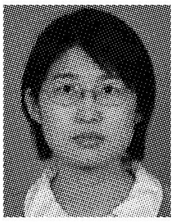
[57] CHENG Miao,FANG Bin,TANG Yuanyan,et al. Incremental embedding and learning in the loal discriminant subspace with application to face recognition[J]. IEEE Transactions on Systems,Man,and Cybernetics,Part C: Applications and Reviews,2010,40(5): 580-591.

[58] XU Ye,SHEN Furao,HASEGAWA Q,et al. An online incremental learning vector quantization[C]//Proceedings of the 13th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. Bangkok,Thailand,2009: 1046-1053.

[59] NING Huazhong,XU Wei,CHI Yun,et al. Incremental spectral clustering by efficiently updating the eigen-system [J].Patterm Recognition,2010,43(1): 113-127.

[60] ZHANG Zhenyue,WANG Jing,ZHA Hongyuan. Adaptive manifold learning[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2012,34(2): 253-265.

作者简介:



谈超,女,1983年生,博士研究生,主要研究方向为机器学习与数据挖掘.

关佶红,女,1969年生,教授,博士生导师,中国计算机学会数据库专委会委员、开发系统专委会委员.主要研究方向为空间数据库、数据挖掘、生物信息学等.主持和参与国家自然科学基金、国家"863"计划项目、省部级以及其他科研项目30余项.2011年获教育部科技进步二等奖,发表学术论文200余篇.

周水庚,男,1966年生,教授,博士生导师,中国计算机学会数据库专委会和人工智能与模式识别专委会委员,中国人工智能学会机器学习专委会常委.主要研究方向为数据库、数据挖掘、生物信息学等.主持或参与国家"973"计划子项目、国家"863"计划项目、国家自然科学基金重大项目与面上项目及其他省部级科研项目20余项.获部级自然科学奖/科技进步奖二等奖6项、三等奖1项,发表学术论文150余篇.