数据降维方法分析与研究*

吴晓婷，闫德勤
（辽宁师范大学 计算机与信息技术学院，辽宁 大连 116081）

摘 要：全面总结现有的数据降维方法，对具有代表性的降维方法进行了系统分类，详细地阐述了典型的降维方法，并从算法的时间复杂度和优缺点两方面对这些算法进行了深入的分析和比较。最后提出了数据降维中仍待解决的问题。

关键词：数据降维；主成分分析；局部线性嵌入；等度规映射；计算复杂度
中图分类号：TP301 文献标志码：A 文章编号：1001-3695(2009)08-2832-04
doi:10.3969/j.issn.1001-3695.2009.08.008

Analysis and research on method of data dimensionality reduction

WU Xiao-ting, YAN De-qin
(School of Computer & Information Technology, Liaoning Normal University, Dalian Liaoning 116081, China)

Abstract: This paper gave a comprehensive summarization of existing dimensionality reduction methods, as well as made a classification to the representative methods systematically and described some typical methods in detail. Furthermore, it deeply analyzed and compared these methods by their computational complexity and their advantages and disadvantages. Finally, it proposed the crucial problems which needed to be resolved in future work in data dimensionality reduction.

Key words: data dimensionality reduction; principal component analysis(PCA); locally linear embedding(LLE); isometric mapping; computational complexity

近年来，数据降维在许多领域起着越来越重要的作用。通过数据降维可以减轻维数灾难和高维空间中其他不相关属性，从而促进高维数据的分类、可视化及压缩。所谓数据降维是指通过线性或非线性映射将样本从高维空间映射到低维空间，从而获得高维数据的一个有意义的低维表示的过程。数据降维的数学描述如下：a) $X=\{x_i\}_{i=1}^N$是 D维空间中的一个样本集，$Y=\{y_i\}_{i=1}^N$是 d($d<<D$)维空间中的一个数据集；b)降维映射，$M: X\to Y$，$x\to y=M(x)$，称 y为 x的低维表示。

目前已经提出了许多降维方法[1~6]，主要包括主成分分析（PCA）、多维尺度分析（multidimensional scaling, MDS）以及近年来提出的基于流形学习的算法，如 Isomap、局部线性嵌入（LLE）、拉普拉斯特征映射（Laplacian Eigenmaps）等。对现有的降维方法，可以从不同角度进行分类。从待处理的数据的性质角度考虑可分为线性和非线性的；从算法执行的过程可分为基于特征值求解的方法和迭代方法；从几何结构的保留角度考虑可分为全局方法和局部方法。本文依据降维方法间的主要区别，将现有的降维方法进行了系统的分类，如图 1所示，并对几种典型的线性和非线性降维方法进行了详细的阐述，最后对这些降维方法进行了系统的分析比较。

典型的降维方法

. 线性降维方法

1) PCA

PCA[1]是通过对原始变量的相关矩阵或协方差矩阵内部结构的研究，将多个变量转换为少数几个综合变量即主成分，从而达到降维目的的一种线性降维方法。这些主成分能够反映原始变量的绝大部分信息，它们通常表示为原始变量的线性组合。

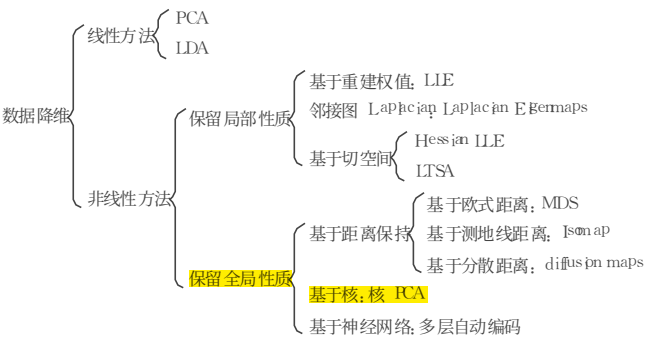

图 1 现有降维方法分类

设 $X=(X_1, X_2, \cdots, X_n)^T$是一个 n维随机变量，$C=1/(n-1)\sum_{i=1}^{n}(X_i-\overline{X})(X_i-\overline{X})^T$为样本协方差矩阵。假设存在如下线性变换：

$$\begin{cases} Y_1=a_{11}X_1+a_{21}X_2+\cdots+a_{N1}X_N=a_1^TX \\ Y_2=a_{12}X_1+a_{22}X_2+\cdots+a_{N2}X_N=a_2^TX \\ \vdots \\ Y_N=a_{1N}X_1+a_{2N}X_2+\cdots+a_{NN}X_N=a_N^TX \end{cases} \quad (1)$$

若用 Y_1代替原来的 n个变量，则要求 Y_1尽可能多地反映原来 n个变量的信息。而方差 $var(Y_1)$越大则表示 Y_1包含的信息越多，因此要求最大化 $var(Y_1)$，同时限定 $a_1^Ta_1=1$以消

收稿日期：2008-11-26；修回日期：2009-01-24 基金项目：国家自然科学基金资助项目（60372071）；中国科学院自动化研究所复杂系统与智能科学重点实验室开放课题基金资助项目（20070101）；辽宁省教育厅高等学校科学研究基金资助项目（2004C031）

作者简介：吴晓婷（1985-），女（蒙古族），内蒙古呼伦贝尔人，硕士研究生，主要研究方向为数据降维、模式识别等（xiaotingwu85@hotmail.com）；闫德勤（1962-），男，博士，主要研究方向为模式识别、数字水印和数据挖掘等。

除方差最大值的不确定性。根据上述条件易求得 $\mathrm{var}(Y_1)=a_1^{\mathrm{T}}Ca_1$，因此，求解方差 $\mathrm{var}(Y_1)$最大问题可转换为在约束 $a_1^{\mathrm{T}}a_1=1$下求以下最优问题：

$$\begin{cases}\max\ a_1^{\mathrm{T}}Ca_1\\ \mathrm{s.t}\ \ a_1^{\mathrm{T}}a_1=1\end{cases}\tag{2}$$

通过拉格朗日乘子法求解，有 $Ca_1=\lambda a_1$。设 $\lambda=\lambda_1$为 C的最大特征值，则相应的特征向量 a_1即为所求。如果 Y_1不能代表 n个变量的绝大部分信息，则可以用同样的方法求得 Y_2甚至 Y_3，Y_4等。一般地，求 X的第 i个主成分可通过求 C的第 i大特征值对应的特征向量得到。为了使它们所含信息互不重叠，通常要求它们相互独立，即 $\mathrm{cov}(Y_i,Y_j)=a_i^{\mathrm{T}}Ca_j=0(i\neq j)$。

通过上述方法就可以找到线性变换（式（1））的一组线性基，从而找到原始变量的一组综合变量（主成分）来代替原始变量。在实际应用中通常不会使用所有 n个主成分，而选取 m（$m<<n$）个主成分。m的选取根据前 m个主成分的累计贡献率 $\sum_{i=1}^{m}\lambda_i/\sum_{i=1}^{n}\lambda_j$来选取。

2）LDA

Fisher在 1936年提出著名的 Fisher准则，对于二类（分别称为正类和负类）问题，希望投影后得到的 $y=w^{\mathrm{T}}x$能够使得 $J(w)$最大：

$$J(w)=\|m_1-m_2\|^2/(\sigma_1^2-\sigma_2^2)\tag{3}$$

其中：m_1、m_2分别是正、负样本在投影方向上的均值；σ_1、σ_2是正、负样本在投影方向上方差。可将其推广到多类问题，此时希望找到的优化方向是使得在低维空间中同类数据尽量靠近，而非同类数据尽量分离，从而保留丰富的辨别信息，使投影后的数据具有最大的可分性。此时，Fisher准则可修正为

$$W_{\mathrm{opt}}=\arg\max_{w}|w^{\mathrm{T}}S_Bw|/|w^{\mathrm{T}}S_ww|\tag{4}$$

其中：S_B、S_w分别是类间分散和类内分散，定义为

$$S_w=\sum_c p_c\mathrm{cov}_{X^c-\bar{X}^c},\quad S_B=\mathrm{cov}_{X-\bar{X}}-S_w\tag{5}$$

其中：p_c是类标 c的预先类；$\mathrm{cov}_{X^c-\bar{X}^c}$表示分配给类 $c\in C$（C为可能的类的集合）的零均值数据点 x_i的协方差矩阵，且 $\mathrm{cov}_{X-\bar{X}}$是零均值数据 X的协方差矩阵。

最大化过程可以通过计算 $S_w^{-1}S_B$（在必要条件 $d<|C|$下）的 d个主特征向量完成。求出特征向量后，原始数据 X在这些特征向量上的投影系数就是其低维嵌入坐标。

. 非线性降维方法

1）核主成分分析（KPCA）

核方法是一系列非线性数据处理技术的总称，它们的共同特征是这些数据处理方法均用到了核映射。近几年，使用核函数[6]对线性方法的重建提出一些成功方法，如支持向量机回归、核 PCA、核 Fisher分析等。

核 PCA是线性 PCA的推广，主要思想是把输入数据 x经由一个非线性映射 $\Phi(x)$映射到特征空间 F，然后在特征空间 F上执行线性 PCA。基本原理如下：

设给定高维数据观测集 $X=\{x_1,x_2,\cdots,x_N\}$，$x_i\in R^D$。通过非线性映射函数 $x\to\Phi(x)\in F$（F称为特征空间），将每个数据点 x映射到一个高维的特征空间。对原始空间中任意两个数据点 x_i、x_j在 F空间中的距离用它们的内积 $\Phi(x_i)\Phi(x_j)$表示，定义核函数 $k(x_i,x_j)=\Phi(x_i)\Phi(x_j)$。假设 $\sum_{i=1}^{N}\Phi(x_i)=0$，则在特征空间 F上映射数据的协方差矩阵为 $C=(1/N)\sum_{i=1}^{N}\Phi_i\Phi_i^{\mathrm{T}}$，$\Phi_i=\Phi(x_i)$。求 C的特征值 $\lambda(\lambda\geq0)$和特征向量 v：

$$Cv=\lambda v\tag{6}$$

即有 $\Phi_kCv=\lambda\Phi_kv(k=1,2,\cdots,N)$。因为 v是在 $\{\Phi_i\}$生成的空间中，所以 v可以表示为

$$v=\sum_i\alpha_i\Phi_i\tag{7}$$

将式（7）带入式（6），有

$$\lambda\sum_{i=1}^{N}\alpha_i(\Phi_k\Phi_i)=(1/N)\sum_{i=1}^{N}\alpha_i(\Phi_k\sum_{k=1}^{N}\Phi_j)(\Phi_j\Phi_i)$$

即

$$K\alpha=\bar{\lambda}\alpha\tag{8}$$

其中：$K_{ij}=\Phi_i\Phi_j$为核矩阵，$\bar{\lambda}=N\lambda$。对式（8）求解可获得要求的特征值和特征向量。但假设 $\sum_{i=1}^{N}\Phi(x_i)=0$一般情况下不成立，因此可用

$$\tilde{K}_{ij}=K_{ij}-(1/N)\sum_{i=1}^{N}K_{il}-(1/N)\sum_{j=1}^{N}K_{lj}-(1/N^2)\sum_{l,m=1}^{N}K_{lm}\tag{9}$$

代替式（8）中的 K。为了获取低维表示，数据被投影到协方差矩阵的特征向量 v_i上，投影结果（即低维数据表示 Y）由

$$Y=\{\sum_j\alpha_1\Phi(x_j)\Phi(x),\sum_j\alpha_2\Phi(x_j)\Phi(x),\cdots,\sum_j\alpha_d\Phi(x_j)\Phi(x)\}\tag{10}$$

给出。

2）MDS

MDS[8]是保留数据点间相似性或距离的一种非线性降维方法。MDS可分为度量性 MDS和非度量性 MDS。度量 MDS利用数据点间的距离或相似性获得数据的低维几何表示，而非度量 MDS仅利用原始数据点间的顺序信息来获得其低维表示。前者将距离平方阵转换为内积阵，通过求内积阵的特征值和特征向量获取低维表示；后者采用迭代方法。下面主要介绍度量性 MDS。

设 $X=(X_1,X_2,\cdots,X_N)$是 D维空间中的一个包含 N个样本点的数据集，$d(X_i,X_j)$表示数据点 X_i与 X_j之间的欧式距离，即

$$d(X_i,X_j)=\|X_i-X_j\|_2=(\sum_{k=1}^{D}(x_{ki}-x_{kj})^2)^{1/2}\tag{11}$$

MDS的目的就是从距离集合中重构出 X的坐标表示，为了使低维嵌入坐标的中心在原点并与坐标轴对齐，可以假设原始数据已经被中心化，即 $\sum_{i=1}^{N}X_i=0$。令距离平方阵为

$$D=(d^2(X_i,X_j))_{N\times N}=[(X_i-X_j)^2]=$$
$$[\|X_i\|^2-2X_i^{\mathrm{T}}X_j+\|X_j\|^2]=Be^{\mathrm{T}}-2X^{\mathrm{T}}X+eB^{\mathrm{T}}\tag{12}$$

其中：$B=(\|X_1\|^2,\cdots,\|X_N\|^2)^{\mathrm{T}}$。

令 $J=I-ee^{\mathrm{T}}/N$，其中 e为 N维全 1向量，I为 N维单位矩阵，则有 $Je=0$，$J^{\mathrm{T}}=J$，且对于任意的 N维向量 $X=(X_1,X_2,\cdots,X_N)^{\mathrm{T}}$，有 $JX=[X_i-\mu]$，$\mu=1/N\sum_{i=1}^{N}X_i$。利用 J对距离平方阵双中心化，有

$$JDJ=JBe^{\mathrm{T}}J-2JX^{\mathrm{T}}XJ+JeB^{\mathrm{T}}J=-2X^{\mathrm{T}}X$$

设 $H=X^{\mathrm{T}}X=-JDJ/2$对 H作特征分解：$H=U\Lambda U^{\mathrm{T}}$。设 $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_N\geq0$为 H的特征值，$U_1,U_2,\cdots,U_N\in R^N$为对应的特征向量，取前 d个特征值和对应的特征向量，得到 X的低维表示：

$$Y=\mathrm{diag}(\sqrt{\lambda_1},\sqrt{\lambda_2},\cdots,\sqrt{\lambda_d})[U_1,U_2,\cdots,U_d]^{\mathrm{T}}\tag{13}$$

3）Isomap

Tenenbaum等人[3,9]提出的 Isomap算法是对经典 MDS的一种推广。但 MDS是基于欧式距离的且没有考虑邻近数据点的分布。假如高维数据点分布或近似分布于一个弯曲的流形上，如 Swiss roll数据集[9]，MDS可能将两个数据点看做是近邻点，然而它们沿着流形的距离要远远大于它们的输入距离。Isomap的基本思想是首先使用最近邻图中的最短路径得到近

似的测地线距离(图 2),代替不能表示内在流形结构的 Euclidean距离,然后应用 MDS算法,进而发现嵌入在高维空间的低维坐标。测地线距离是两点之间沿着流形的距离。具体算法如下:

a)构建输入空间 X中流形 M上所有数据点 x_i($i=1,2,\cdots,N$), $x_i\in R^D$的邻接图 G,距离定义为 Euclidean距离 $d_x(i,j)$,邻接关系定义为 ε近邻或 k近邻。

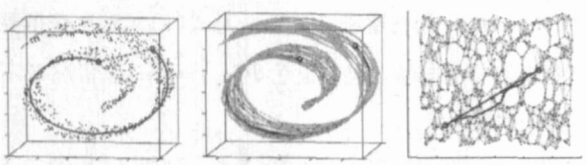

图2 Isomap基本思想

b)通过计算图 G上两点间的最短路径 $d_G(i,j)$估计流形 M上测地线距离 $d_M(i,j)$,得到的矩阵 $D_G=\{d_G(i,j)\}$为图 G上任意两点间的最短路径距离(其中最短路径可由 Dijkstra算法求得)。

c)应用 MDS算法,构建 d维 Euclidean空间 Y上的嵌入。

4) LLE

局部线性嵌入[2](LLE)是与 Isomap相似的一种局部降维方法。但与 Isomap不同的是, Isomap中建立了数据点的邻接图表示,而 LLE只试图保留数据点的局部性质,这使它对短环路问题没有 Isomap敏感。此外,局部性质的保留允许非凸流形的成功嵌入。其基本思想是假设每个数据点与它的邻近点位于流形的一个线性或近似线性区域中,将全局非线性转换为局部线性,而相互重叠的局部邻域能够提供全局结构的信息。具体步骤如下:

a)局部近邻选取。对于给定的数据集 $X=\{x_1,x_2,\cdots,x_N\}$, $x_i\in R^D$在高维空间中寻找每个样本点 x_i的 k($k<N$)个近邻点,距离公式为

$$d_{ij}=\left[\sum_{k=1}^{D}|x_k-x_{jk}|^2\right]^{1/2} \tag{14}$$

b)计算样本点的局部重建权值矩阵。定义误差函数

$$\min\varepsilon(W)=\sum_{i=1}^{N}\left\|X_i-\sum_{j=1}^{N}w_{ij}x_j\right\|_2^2 \tag{15}$$

其中: x_{ij}($j=1,2,\cdots,k$)为 x_i的 k个近邻点; w_{ij}是 x_i与 x_{ij}之间的权值。为了保证重构与坐标的平移、旋转和缩放等无关,限制 $\sum_{j=1}^{k}w_{ij}=1$。结合限制条件,式(15)可改写为

$$\min\varepsilon(W)=\sum_{i=1}^{N}\left\|\sum_{j=1}^{k}w_{ij}(x_i-x_{ij})\right\|^2=\sum_{i=1}^{N}\left\|(x_i-x_{ij})w_i\right\|^2=\sum_{i=1}^{N}(w_i)^TZ_iw_i \tag{16}$$

其中: $Z_i=(x_i-x_{ij})^T(x_i-x_{ij})$为第 i个样本点的局部协方差矩阵; $w_i=[w_{i1},w_{i2},\cdots,w_{ik}]^T$为第 i个样本点的局部重建权值。引入 Lagrange乘子求解此约束问题,则有

$$L(W)=\sum_{i=1}^{N}(w_i)^TZ_iw_i+\lambda\left(\sum_{j=1}^{k}w_{ij}-1\right)\Rightarrow$$
$$(\partial L)/\partial w_{ij}=2Z_iw_i+\lambda\times1\Rightarrow Z_iw_i$$

通常采取简单的求解方法,令 $Z_iw_i=1$,然后调整权值使其和为 1来求得 w_i。

c)利用权值矩阵 W寻找样本集的低维嵌入 Y。通过最小化重构误差和函数 $\min\Phi(Y)=\sum_{i=1}^{N}\left\|y_i-\sum_{j=1}^{N}w_{ij}y_j\right\|_2^2$来实现。为了固定 Y和避免数据集在低维使坍塌到坐标原点,可简单地对 Y加以限制: $\sum_{i=1}^{N}y_i=0$, $1/N\sum_{i=1}^{N}y_iy_i^T=I$。其中: I表示 N维单位矩阵。相应地,优化问题转换为下列约束优化问题:

$$\begin{cases}\min\Phi(Y)=\sum_{i=1}^{N}\|YI_i-YW_i\|^2=\\ \sum_{i=1}^{N}\|Y(I_i-W_i)\|^2=\min\ \mathrm{tr}\,YMY^T\\ s.t.\quad YY^T=I\end{cases} \tag{17}$$

其中: $M=(I-W)^T(I-W)$。使用 Lagrange乘子法,解得 $MY^T=\lambda Y^T$,则 M的特征向量即所求的嵌入坐标。取 M的值最小的 d个非零特征值所对应的特征向量作为低维坐标 Y。通常最小特征值几乎为零,因此取 2~(d+1)间的特征值所对应的特征向量作为输出结果。

5) Laplacian Eigenmaps

Laplacian Eigenmaps[4,10]是由 Belkin等人于 2001年提出的。类似于 LLE, Laplacian Eigenmaps也是通过保留流形的局部特性发现数据低维表示的一种数据降维方法。 Laplacian Eigenmaps寻求一个能在平均意义下保持流形局部特性的映射,而其局部特性基于每对邻近点间的距离。具体算法如下:

a)对于给定的高维观测数据集 $X=\{x_1,x_2,\cdots,x_N\}$, $x_i\in R^D$,首先构建数据点间的邻接图 G(使用 k近邻或 ε近邻),若两点 x_i x_j近邻,则 $G_{ij}=1$,否则为 0。

b)定义邻接权值矩阵 W。使用热核(heat kernel)方法,即如果 $G_{ij}=1$,则 $w_{ij}=e^{-\|x_i-x_j\|^2/2\sigma^2}$,否则为 0。或者使用更简单的方法,若 $G_{ij}=1$,则 $w_{ij}=1$,否则为 0。

c)构建低维表示。设图 G为简单连接图(否则对每一个连接部分), M和 W分别为 G的顶点度对角矩阵和邻接矩阵($M_{ii}=\sum_j w_{ji}$),则 $L=M-W$为 G的 Laplacian矩阵。构造目标函数,则有

$$\begin{cases}\min\phi(Y)=\sum_{ij}(y_i-y_j)^2w_{ij}=\mathrm{tr}(YLY^T)\\ s.t.\ YDY^T=I\end{cases} \tag{18}$$

其中: $Y=\{y_1,y_2,\cdots,y_N\}$。因此最小化 $\phi(Y)$等价于最小化 YLY^T。利用 Lagrange乘子法求解上述问题,等价于计算 Laplacian矩阵 L的特征值及对应的特征向量:

$$LY=\lambda MY \tag{19}$$

取与最小的 d个特征值对应的特征向量(最小特征值 0对应的特征向量除外),即得到了全局低维坐标

$$Y=\{y_1,y_2,\cdots,y_N\},\ y_i\in R^d$$

6) Local tangent space alignment(LTSA)

2004年,浙江大学的张振跃等人根据非线性流形的全局非线性结构来自于局部线性分析和局部线性信息的全局整合这一思想提出了局部切空间排列[5,11](LTSA)算法。 LTSA具体可概括为两点,即投影和整合。算法通过逼近每个样本点的切空间来构建低维流形的局部几何,观测数据点在局部切空间的投影获得局部低维坐标,交叠的局部低维坐标被局部仿射变换后获得全局低维嵌入坐标。算法描述如下:

a)设给定高维观测数据集 $X=\{x_1,x_2,\cdots,x_N\}$, $x_i\in R^D$。对每个样本点 x_i选取包含 x_i在内的 k($k<N$)个近邻 $X_i=\{x_1,x_2,\cdots,x_k\}$。

b)计算每个近邻数据集 X_i的 d维局部切空间的基向量和观测数据的局部切坐标 Θ_i。

为了使近邻数据点在切空间上的投影方差最小,每个数据点的 d维切空间的标准正交基 Q_i可由观测数据矩阵 $X_i(I-k^{-1}ee^T)$进行 SVD分解的 d个最大奇异值对应的 d个左奇异特征向量给出[11];而局部坐标由观测数据在基 Q_i上的投影坐标 $\Theta_i=Q_i^TX_i(I-k^{-1}ee^T)$给出。

c)根据 N个局部投影坐标 $\Theta_i=[\theta_1,\cdots,\theta_k]$($i=1,\cdots,$

N)，通过局部仿射变换 $L_i \in R^{k\times d}$整合得到全局坐标 $\{y_i\}_{i=1}^N$。

设 $Y=[y_1, \cdots, y_N]$，$Y_i=[y_{i_1}, \cdots, y_{i_k}]$，$\bar{y}_i$为 $y_{i_1}, \cdots, y_{i_k}$的均值，通过最小化全局排列的重构误差得到全局坐标：

$$\min_{L_i, Y_i} \sum_i \| \varepsilon_i \|_2^2 = \min_{L_i, Y_i} \sum_i \| Y_i - y_i e^T - L_i \Theta_i \|_2^2 \qquad (20)$$

上述最小化问题能够等价地转换为一个求特征值的简单问题。

算法分析与比较

降维算法的复杂度大小严重地影响着算法的实施。算法的复杂度或所需存储空间过大会导致算法不可行。下面简单分析一下上述各类降维算法的时间复杂度。

降维算法的复杂度主要由数据点的个数 n，原始维数 D，目标维数 d，以及算法涉及的参数如近邻点个数 k(对基于谱图理论的方法)和迭代次数 i(对迭代方法)来决定。PCA和 LDA中计算协方差矩阵需要 $O(nD)$，而对 $D\times D$协方差矩阵进行特征分析需要 $O(D^3)$；Isomap和 KPCA对 $n\times n$矩阵进行特征分析需要 $O(n^3)$，KPCA计算核需要 $O(Dn^2)$，而 Isomap中 n次寻找近邻点需要 $O(Dn \log n)$，在邻近图上应用 Dijkstra算法需要 $O(nk+n \log n)$。类似于 KPCA，局部降维算法也需要对 $n\times n$矩阵进行特征分析，但局部方法中的 $n\times n$矩阵通常是稀疏的，这就大大降低了特征分析的时间复杂度。对稀疏矩阵进行特征分析需要的时间复杂度为 $O(pn^2)$(这里 p是稀疏矩阵中非零元和零元的比率)。对 LLE，Laplacian Eigenmaps和 LTSA建立稀疏的邻近图需要 $O(Dn \log n)$；另外，LLE解 n个大小为 $k\times k$的方程的复杂度为 $O(nk^3)$，Laplacian Eigenmaps计算高斯核需要 $O(pnD)$；而 LTSA对 $k\times k$正交矩阵的特征分解需要 $O(nk^3)$。综合上述分析，表 1列出了各类算法的时间复杂度。

表 1　算法的时间复杂度

算法	时间复杂度
PCA/LDA	$O(nD)+O(D^3)$
KPCA	$O(Dn^2)+O(n^3)$
MDS	$O(n^3)$
Isomap	$O(Dn \log n)+O(nk+n \log n)+O(n^3)$
LLE	$O(Dn \log n)+O(nk^3)+O(pn^2)$
Laplacian Eigenmaps	$O(Dn \log n)+O(pnD)+O(pn^2)$
LTSA	$O(Dn \log n)+O(nk^3)+O(pn^2)$

另外，虽然上述算法在不同领域都得到了一定的应用，但目前还没有一种算法能够适用于所有的领域，它们都存在着一定的不足。本文对这些算法的优缺点作了一个比较详细的总结，如表 2所示。

结束语

本文以降维数据集间的主要区别(线性/非线性)为主线，对已有典型的降维方法进行了详细的阐述和分析比较。比较中发现现有降维方法存在以下有待解决的问题：a)现有的非线性降维方法对于个别的人造数据效果很好，但对于现实数据往往并不优于传统的线性方法，因而要进一步研究这些非线性降维方法使其得到最大程度的改进；b)流形学习的提出为数据降维提供了非常有利的框架，如 LLE，Laplacian Eigenmaps都是具有代表性的流形学习方法，但它们大多为局部方法，局部方法的一个很大的缺陷就是受噪声影响大，如何减少噪声的干扰、提高算法的鲁棒性一直以来都是研究的方向；c)现有降维方法不具有增值能力，对动态增加的观测数据点不能快速明确地映射到低维空间，学习改进增量算法具有一定的研究价值。

表 2　降维算法比较

算　法	优　点	缺　点
PCA	理论完善、概念简单、计算方便，具有最优线性重构误差	对于非常高维的数据特征向量的计算可能不可行，主成分个数的确定没有明确的准则，不能用于处理非线性数据
LDA	是监督方法，已成功应用于大量的分类工作	当各类的类别中心重叠时，不可使用；当各类的协方差不同时，不能找到最好的投影变换，其找到的变换是次优的；最大维数限制为 $k-1$(这里 k为类别数)，不足以处理复杂问题
KPCA	核空间内 PCA的重建相对简单，可处理非线性数据	算法性能依赖于核的选择，核矩阵大小是数据中样本数的平方
MDS	能够较好地保持数据间的差异性	对高维非线性数据无能为力，没有统一的标准评价所得到的嵌入维数的质量
Isomap	以流形上测地线距离代替欧式距离，更好保留数据的几何结构	具有拓扑不稳定性[12]；短环路[13]会严重影响其执行；要求流形是凸的，否则会发生变形；不能解决流形上有洞的问题
LLE	具有平移、旋转等不变性；可变参数少，不包含局部极小，对短环路没有 Isomap敏感	要求所学习的流形只能是不闭合的且在局部是线性的，且要求样本采样稠密；参数 k和 d有过多的选择，对噪声敏感
LTSA	利用局部切空间可以很好地反映流形的局部几何特征	不能处理样本数较大的样本集及不能增量学习，对于高曲率的流形学习效果不好

参考文献：

[1] HOTELLING H. Analysis of a complex of statistical variables into principal components[J]. Journal of Educational Psychology, 1933, 24: 417-441.

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

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

[4] BELKIN M, NIYOGI P. Laplacian Eigenmaps and spectral techniques for embedding and clustering[C]//Proc of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2001: 585-591.

[5] ZHANG Zhen-yue, ZHA Hong-yuan. Principal manifolds and nonlinear dimensionality reduction via local tangent space alignment[J]. SIAM Journal of Scientific Computing, 2004, 26(1): 313-338.

[6] SCHOLKOPF B, SMOLA A J, MULLER K R. Nonlinear component analysis as a kernel eigenvalue problem[J]. Neural Computation, 1998, 10(5): 1299-1319.

[7] FISHER R A. The use of multiple measurements in taxonomic problems[J]. Annals of Eugenics, 1936, 7(2): 179-188.

[8] COX T, COX M. Multidimensional scaling[M]. London: Chapman & Hall, 1994.

[9] TENENBAUM J B. Mapping a manifold of perceptual observations[C]//Proc of Advances in Neural Information Processing Systems. Cambridge: MIT Press, 1998: 682-688.

[10] BELKIN M, NIYOGI P. Laplacian eigenmaps for dimensionality reduction and data representation[R]. Chicago: University of Chicago, 2001.

[11] ZHANG Zhen-yue, ZHA Hong-yuan. Linear low-rank approximations and nonlinear dimensionality reduction[J]. Science in China Series A-Mathematics, 2005, 35(3): 273-285.

[12] BALASUBRAMANIAN M, SCHWARTZ E L. The Isomap algorithm and topological stability[J]. Science, 2002, 295(5552): 7.

[13] LEE J A, VERLEYSEN M. Nonlinear dimensionality reduction of data manifolds with essential loops[J]. Neurocomputing, 2005, 67(1): 29-53.