

谱聚类的算子理论研究进展

管涛 王杰

(郑州航空工业管理学院计算机科学与应用系 郑州 450015)

摘要 谱聚类来源于算子理论研究成果,在大数据降维和分类中发挥着重要的作用,但是目前国内的研究多注重应用算法设计,很少见到谱聚类理论方面的研究。为弥补这方面的一些不足,较为系统地总结了这些理论,侧重于阐述与谱聚类的算子理论紧密相关的最新理论研究成果,并简要介绍了一些具体的谱聚类算法、原理及其性能。从积分算子、图谱理论、流形学习出发,评述和分析了谱聚类的最新理论原理、收敛性结论、发展现状以及与流形学习的内在联系,最后指出了理论研究的一些方向。

关键词 谱聚类,积分算子,流形学习,Laplacian 特征映射(LEM),算子收敛性

中图分类号 TP311 文献标识码 A

Operator Theory Frontier of Spectral Clustering

GUAN Tao WANG Jie

(Department of Computer Science and Application, Zhengzhou Institute of Aeronautical Industry Management, Zhengzhou 450015, China)

Abstract Spectral clustering comes from operator theory and is able to efficiently decrease data dimension and classify data. However, current domestic researches pay more attention to applied algorithm design, and there is lack of theoretical outcome. In order to make up the shortcomings in theory, this paper systematically summaries the theory of spectral clustering, pays more attention to study the newest foreign operator theory research achievements. Besides, we briefly discuss and analyze some specific spectral clustering algorithms. We introduce and analyze the principle, convergence, current status of spectral clustering and its inherent association with manifold learning in terms of integral operator, spectral graph theory and manifold learning. At last, we suggest some possible directions.

Keywords Spectral clustering, Integral operator, Manifold learning, Laplacian eigenmap(LEM), Convergence of operator

1 引言

谱聚类是大数据分析、图像和视频分割中的重要非线性降维方法之一,是一种基于图谱理论的无监督学习方法^[1-3]。当前的理论研究围绕着图相似矩阵、积分算子、核函数、数据降维等领域开展,如特征函数学习问题、矩阵的低秩分解与扰动分析、抽样分析。在国内,除了个别数学基础研究成果之外,谱聚类的物理原理、收敛性、渐进理论等与机器学习理论相关的要点很少被涉及。另一方面,谱聚类与流形学习中 Laplacian 特征映射(LEM)关系紧密,国外学者在这方面的工作深入而透彻,相比而言国内研究多侧重于应用算法的设计。为加深学习理论中这两个方面的认识,分析各种相关概念之间的内在联系,下面结合算子理论和再生核 Hilbert 空间理论,从热力学扩散方程、正定积分算子、流形学习和 Laplacian 特征映射的角度,介绍当前谱聚类的特征函数学习问题、收敛性分析工作,比较分析谱聚类与 Laplacian 特征映射的数学物理本质及内在的联系,为将来深入研究和设计应用型算法奠定理论基础。

2 热核与热算子

在物理学中, $f=f(x, y, z, t)$ 表示关于空间和时间的温度函数,空间的热传导方程表示为:

$$\frac{\partial f}{\partial t} = k(\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2}) = k \nabla^2 f \quad (1)$$

或者

$$\frac{\partial f}{\partial t} = k \Delta f$$

式中, ∇^2 或 Δ 表示 Laplacian 算子, k 为常数。该方程也称为扩散方程,是 Fokker-Plank 方程的原型,描述了粒子或者热量的空间分布随时间的变化。该方程的解析解定义了热算子形式,表达为热核函数 $G(t, x)$ (即高斯分布)与 f 的卷积的形式,即:对于 $x \in R^d$,

$$L(f)(x) = G(t, x) * f = \int G(t, x, \tau) f(\tau) d\tau \quad (2)$$

由式(2)得到的高斯微分算子是式(1)的最优解,刻画了热量等流体在空间随时间的传播过程。式(2)在去噪、滤波等领域具有广泛的应用。由式(2)可知, L 关于 $x \in R^d$ 的导数可

本文受国家自然科学基金项目(41171341),教育部新世纪优秀人才支持计划(NCET-09-0126),河南省科技厅项目(122102210227,092300410140,102102210447),河南省教育厅项目(2011B520038,2010B520032),郑州市科技局项目(112PPTGY248-6)资助。

管涛(1974-),博士,CCF会员,主要研究领域为统计机器学习、数据挖掘、图像分析与处理,E-mail:timmm.guan@gmail.com.

以表达为:

$$L_{x^m}(f)(x) = \frac{\partial^m G(t, x)}{\partial x^m} * f \quad (3)$$

式中, $\frac{\partial^m G(t, x)}{\partial x^m} = \frac{\partial^m G(t, x)}{\partial x_1^{m_1} \cdots \partial x_d^{m_d}}, \sum_{j=1}^d m_j = m$, 则 L_{x^m} 构成了高斯微分算子。它定义了尺度空间理论中的基本变换, t 为尺度参数。典型的微分算子即为高斯 Laplace 算子 $\Delta G(t, x) = \frac{\partial^2 G(t, x)}{\partial x_1^2} + \cdots + \frac{\partial^2 G(t, x)}{\partial x_d^2}$, 常用于图像多尺度不变特征提取, 如图像的 SIFT 特征。

3 积分算子特征函数学习问题

谱聚类的数学本质问题是正定积分算子的特征函数学习问题。给定正定核函数 K , 定义 Hilbert 空间中的连续积分算子 $T: L^2(X, \rho) \rightarrow L^2(X, \rho)$,

$$Tf(x) = \int_X K(x, y) f(y) d\rho_y \quad (4)$$

式中, ρ_y 为关于 $y \in X$ 的概率测度。式(4)亦称为卷积算子。当 K 取高斯核函数时, 除了时间因素之外, 式(4)与式(2)一致。此时, 线性算子 T 的特征函数刻画了热力学方程中热量流动的方向, 而特征值刻画了流动的速度。

算子 T 的特征值分布与核函数的性质有关。数学学者 H. Weyl, J. B. Reade, Chung-Wei Ha 等人研究了不同空间正定核函数所对应特征值的极限分布, 一般情况下: 当 $K(x, y) \in C^p$ 且正定, $p \geq 1$, 有 $\lambda_n = o(1/n^{p+1})$ 。在正定核的条件下, 由 T 的特征函数和特征值构造的集合 $\{\sqrt{\lambda_k} f_k, \lambda_k > 0\}$ 构成了 $K(x, y)$ 生成的再生核 Hilbert 空间的标准正交集^[4]。进而, 假设 $\text{supp} \rho_y = X$, 根据 Mercer 定理, $K(x, y)$ 可以表达为 T 的特征函数的傅里叶展开, 即:

$$K(x, y) = \sum_{k \geq 1} \lambda_k f_k(x) f_k(y) \quad (5)$$

一般核函数下积分算子的特征值和特征函数很难求解。在正定核的情况下, 方程(4)的离散特征值通过数据集 X 近似计算, 由独立同分布的样本 $A = \{x_1, x_2, \dots, x_q\} \subset X, |A| \ll |X|$, 得到的特征系统表示为^[5]:

$$T_q f(x) = \frac{1}{q} \sum_{j=1}^q K(x, x_j) f(x_j) = \lambda_k f_k(x) \quad (6)$$

此时, 核函数 K 构成了相似性 Gram 矩阵 $M, M_{ij} = K(x_i, x_j)$ 。若 $x \in A$, 则有 $MU = U\Lambda, \Lambda = \text{diag}\{\lambda_1^{(q)}, \dots, \lambda_q^{(q)}\}, U = (U_{ij}^{(q)})$ 。由 Williams 和 Seeger 的结果知^[6]:

$$f_k(x_j) \simeq \sqrt{q} U_{jk}^{(q)}, \lambda_k \simeq \lambda_k^{(q)} / q$$

代入式(6)即得任意数据的 Nyström 近似:

$$f_k(y) \simeq \frac{\sqrt{q}}{\lambda_k^{(q)}} \sum_{j=1}^q K(y, x_j) U_{jk}^{(q)} \quad (7)$$

4 谱聚类

4.1 原理与典型算法

从黎曼几何的角度看, 无序离散点的近邻图表示了低维流形, 为了寻找其到低维空间的同胚, 定义图上的 Laplacian 算子 $L = D - W$, 它是连续流形 Laplace-Beltrami 算子(将在后面介绍)的离散形式, 其中 W 为图的相似度矩阵, 由核函数定义, $D = \text{diag}(D_{11}, D_{22}, \dots, D_{mm}), D_{ii} = \sum_{j=1}^n W_{ij}$ 。算子 L 的特

征函数 f 最小化了能量 $\frac{1}{2} \sum_{i,j} (f_i - f_j)^2 W_{ij} = f^T L f$ 。在约束 $\|f\| = 1$ 的条件下, 由拉格朗日最优性条件得到特征值问题: $Lf = \lambda Df$ 或者 $Wf = (1 - \lambda) Df$ 。可见 L 和 W 对应的特征向量一致, 而特征值满足和为 1。因此, 可以针对标准化的 W 研究广义特征函数学习问题, 即在 Hilbert 空间由对称核函数定义的线性算子及其特征函数问题: $Tf = \lambda f$ 。

典型的谱聚类算法有 NJW、NCut 和基于 Nyström 方法的谱聚类^[1]、Ratio Cut。NJW 计算标准图 Laplacian 矩阵的非零特征值对应的特征向量, 实际上得到了全部数据在若干特征向量构成的空间上的投影, 是一种非线性降维方法。从矩阵原理上看 NJW 构造了 Gram 矩阵的低秩逼近, 体现了当今的大数据降维思想。NCut 算法将标准化图分割转化为广义特征值问题: $(D - W)y = \lambda D y$, 该方程的解与图相似矩阵的特征值和特征向量具有一一对应关系, 马普研究所的 von Luxburg 等人清晰地指出了它们的关系。Nyström 方法是一种基于采样的求解积分方程的近似方法, 来源于 Baker 对积分方程的数值求解研究。文献[5, 7]的谱聚类算法是通过随机采样构造样本的 Gram 矩阵, 然后利用 Nyström 方法计算数据的近似低维特征向量。而非样本数据(out-of-sample)在低维空间的投影通过特征函数计算。表 1 给出几种典型谱聚类算法计算性能的比较。

表 1 几种典型谱聚类算法复杂度比较

算法	复杂度	典型应用
NJW	$O(n^3)$	小样本聚类分析
Nyström-based	$O(n^2)$	图像与视频分割、聚类
Ncut	$O(n^3)$	图像与视频分割、聚类
Ratio cut	多项式	图像分割、边界提取

注: n 为样本数。

4.2 采样和输入数据分布

采样降低了数据量, 降低了谱聚类的计算复杂度, 当前有一些典型的理论研究。Nyström 方法通过随机采样点计算系统的特征函数, 在保证聚类质量的条件下有效地降低了算法复杂度, 在大数据应用中具有显著的优势。但是采样本身对于 Nyström 方法的影响至今少见报道。利用谱聚类对图像多尺度进行分割, 存在着计算复杂度高和类不平衡问题。因此, 李小斌和田铮研究了基于采样技术的多尺度谱聚类算法, 提出了保证每类均有样本的采样定理, 但是文中没有探讨采样对于近似质量的影响^[8]。

采样量、采样方式和分布函数 μ 影响近似结果, 那么影响程度、近似误差等问题值得研究。文献[8]是根据类别研究采样量的较早文献, 指出每类具有样本的条件下的最低采样量。经典 Nyström 方法进行随机采样可得到独立同分布的样本, 但可能导致样本矩阵奇异和结果的随机性。最近两年, 围绕 Nyström 方法的采样有不少深入的研究。Drineas 等人系统地分析了 Gram 矩阵的低秩近似方法, 提出了基于非均匀采样的 Nyström 方法的低秩矩阵近似算法^[7]。与通常的直接分析特征值和特征函数方式不同, Kai Zhang 等人分析了 Nyström 方法中采样方案的近似矩阵的误差, 指出基于 k -means 算法的采样能达到最小的矩阵近似误差^[9]。2012 年, Kumar 等人研究了各种固定的和适应性采样方法的效率, 提

出了一族集成采样算法。在一系列对比实验分析的同时,也证明了集成 Nyström 算法的收敛性^[11]。在输入数据分布方面,文献[6]描述积分算子中输入数据分布的影响,说明了高斯分布的数据输入可用于类别任务。文中以高斯分布为例进行了分析,解释了非均匀分布情况下产生的数据分类效果。Tao Shi 等人研究了分布 μ 和积分算子特征值和特征函数的关系,指出当 μ 为高斯(混合)分布时,其参数可以通过积分算子的特征空间计算^[10]。

4.3 算子的收敛性

积分算子的收敛性是验证谱聚类算法合理性的重要依据,当前多在算子理论框架下研究谱聚类中标准化经验算子的极限性质、经验算子及对应的连续形式的特征值差异和特征函数关系、经验算子的收敛率等问题。国内数学学者韩彦彬就曾研究了在不同空间中积分算子的特征值、高维正定核的特征值分布^[12]。在谱聚类中,通常假定算子在 L^2 、 R^n 或再生核 Hilbert 空间(RKHS)运算。经验算子 T_n 和 T 的近似程度与收敛关系、算子特征值和特征函数的收敛性等问题是近几年国外学者研究的重点之一^[13-15]。最近,在谱聚类方面有几个重要的工作。von Luxburg 等人从有界线性算子的扰动理论和谱理论出发,描述了算子相关的若干收敛性,研究了标准化谱聚类中连续算子及其离散形式之间的收敛关系、收敛性^[13]。对于标准 Laplacian 算子的收敛率,文献[13]给出了经验算子的特征向量收敛的上界。在高斯核的情况下,其收敛率为 $O(1/\sqrt{n})$, $n \rightarrow \infty$ 。与此不同, Rosasco 等人利用了 Hilbert 空间的中心不等式,在文献[13]的基础上简化了积分算子收敛性的证明过程^[12]。文中采用了对称正定的再生核函数,由样本定义了再生核 Hilbert 空间 \mathcal{H} 中的自伴算子:

$$T_{\mathcal{K}}: \mathcal{H} \rightarrow \mathcal{H}, T_{\mathcal{K}} = \int \langle \cdot, K_y \rangle K_y d\rho_y, K_y = K(\cdot, y)$$

$$T_n: \mathcal{H} \rightarrow \mathcal{H}, T_n = \frac{1}{n} \sum_{i=1}^n \langle \cdot, K_{x_i} \rangle K_{x_i}$$

它们均为 Hilbert-Schmidt 算子,并且依 $1 - 2e^{-\tau}$ 的置信度满足 $\|T_{\mathcal{K}} - T_n\|_{HS} \leq 2\kappa \sqrt{2\tau}/\sqrt{n}$ 。进而, Rosasco 等人利用奇异值分解方法说明了 $T_{\mathcal{K}}$ 和 T 具有相同的非零特征谱,且特征函数满足一定的等价关系,详见文献[14]命题 8。 T 和核矩阵 K 的特征值之差的上线通过这种关系和中心不等式定义。2011 年, Pelletier 和 Pudlo 研究了谱聚类在水平集上算子范数的收敛性,提出了基于密度估计方法的谱聚类算法,并证明了标准化经验算子的必然收敛性^[15]。

5 Laplacian 特征映射

谱聚类与 LEM 紧密相关,而 LEM 来源于相关的物理背景,即 Maupertuis 所谓的自然界的“最小作用原理”。Lagrange 将这类问题的最优目标定义为物体动能 $\int mv^2 dt$, 其中 m, v, t 分别表示质量、速度和时间。在黎曼几何中该目标的最优解可以通过流形学习方法解决。考虑 R^n 中流形 \mathcal{M} 上的光滑函数 $f: [0, T] \rightarrow R^n$, $\|f\|_{L^2(\mathcal{M})} = 1$, 它将时间映射到空间, $\int_{\mathcal{M}} \|\nabla f(x)\|^2$ 即表示了单位质量物体的动能,其中 ∇ 表示梯度^[15]。Laplace-Beltrami 算子 Δ 的特征函数即为该目标的最优解。该算子定义为函数 f 的梯度的散度,即: $\Delta f = \nabla$

$(\nabla f) = \nabla^2 f$, 由于 ∇ 与 $-\text{div}$ 为伴随算子(Stokes 定理), 从而优化目标转化为: $\int_{\mathcal{M}} \|\nabla f(x)\|^2 = \int_{\mathcal{M}} \langle \nabla f(x), \nabla f(x) \rangle = - \int_{\mathcal{M}} \text{div}(\nabla f(x)) f = - \int_{\mathcal{M}} \Delta f(x) f$ 。

根据 Rosenberg 的结论, Δ 是正半定子伴随算子,其特征函数是无限可微函数。Belkin 和 Niyogi 构造了流形上 Laplace-Beltrami 算子的泛函近似 L_t ^[17]:

$$L_t(f)(x) = \frac{1}{t(4\pi t)^{\frac{k}{2}}} \left(\int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(x) d\mu(y) - \int_{\mathcal{M}} e^{-\frac{\|x-y\|^2}{4t}} f(y) d\mu(y) \right) \quad (8)$$

并证明了 L_t 及其离散形式的特征值和特征函数将随着 $t \rightarrow 0$ 收敛于 $\frac{1}{\text{vol}(\mathcal{M})} \Delta$ 的对应值,通常 $\text{vol}(\mathcal{M}) = 1$ 。式(8)采用了方差为 $\sqrt{2t}$ 的高斯核函数, $tL_t(f)$ 的前半部分与式(4)一致,后半部分为常数。他们证明了算子的扰动定理(见文献[15]定理 4.3),指出 $(1 - H_t)/t$ 的特征值和特征向量收敛于连续积分算子 L_t 的对应值^[15,16],其中: H_t 为热算子,定义如式(2)。进而,证明了存在收敛的时间序列,使得经验算子 $\hat{L}_{t,n}$ 的特征值和特征向量收敛于流形上 Laplace-Beltrami 算子的对应值。在 Laplacian 点态收敛(pointwise convergence)方面, Lafon 研究了流形上任意分布下算子的广义点态收敛性,推广了 Belkin 的结论^[18]。Hein 等人在点态收敛性方面开展了系列研究,给出了图 Laplacian 算子与流形的强(弱)点态协调性定理^[19],而 Singer 对该结果给出了更快的收敛率^[20]。

6 数据降维及应用

降维是大数据分析面临的主要工作,目的是在保证结果质量的前提下提高运算效率。谱聚类均能有效地实现数据的非线性降维,除此之外,还包括(核)PCA、奇异值分解(SVD)等。(核)PCA 实现了(非)线性降维,其特征值和特征向量与谱聚类的对应量存在等式关系, Bengio 等人在技术报告中详细地阐明了这一点^[21]。SVD 在谱聚类分析中起到了重要的作用,用来建立经验算子和积分算子的特征值和特征向量之间的联系^[14]。

谱聚类计算了 Gram 矩阵的低秩逼近,与许多当前流行的矩阵降维目标一致。也有一些相关的研究值得参考,张振跃和查宏远结合国外最新的研究趋势,利用 SVD 和流形学习研究了 $m \times n$ 矩阵的低秩逼近方法。算法中矩阵被按列分块,在已知每个子块的最佳低秩逼近的条件下,近似得到原始矩阵的奇异值分解^[22]。Frieze 等人采用 Monte Carlo 算法寻找 $m \times n$ 矩阵 A 的最优低秩逼近。该算法按照假设的概率选择长方阵的行和行内单元,找到最多秩 k 的矩阵 D^* ,使得 $\|A - D^*\|_F^2 \leq \min_{D, \text{rank}(D) \leq k} \|A - D\|_F^2 + \epsilon \|A\|_F^2$ 至少以 $1 - \delta$ 的概率成立。算法是参数 $k, 1/\epsilon$ 和 $\log(1/\delta)$ 的多项式时间复杂度,与矩阵大小无关^[23,24]。

在应用方面,除了适用于任意分布的数据集聚类之外,谱聚类算法主要用于图像与视频分割。分割的基本思想是通过 Laplace 变换将像素的特征点投影到低维空间中(为多对一的映射),形成低维空间聚集。像素点根据这些聚集进行标识分类。

结束语 谱聚类是当前流形的数据嵌入方法之一,可以有效实现数据的降维,当前有一些关于算法及应用方面的综述性文献^[25-27]。本文不注重讨论、分析和介绍现有算法及其应用,更注重围绕谱聚类相关的数学基础开展研究和阐述。

当前的谱聚类算子理论的研究将当前的热点领域,即再生核 Hilbert 空间、核函数、正定积分算子、算子奇异值分解、算子的收敛性理论等内容有机地联系在一起,有效地解释了谱聚类的降维本质。通过一定的变换方法可知,谱聚类与主成分分析^[21]、判别分析^[28]、核 k-means 等算法也具有一定的内在联系。谱聚类算子理论的研究仍有一些工作,如不同论域算子特征值的变化规律、特征函数的分类能力、算子收敛性、矩阵的低秩逼近等。

在应用基础方面,目前仍然存在值得研究的问题:(1)对于一般形式的数据分布 ρ_y 或 μ ,积分算子的性质和特征函数的形式需要研究。(2)随机图上 Laplacian 算子的特征函数学习问题。随机图在社会网络、动态系统中广泛存在,其特征值问题亦与积分算子特征函数学习有关,那么现有的一些结论能不能简单推广,值得深入研究。(3)Nyström 方法是一种基于采样的低秩逼近算法,不仅在谱聚类中效果良好,在文本挖掘、语音处理中也多有应用,值得在应用基础领域扩展研究。(4)利用谱聚类方法估计(高斯)混合模型的参数和研究高斯分支之间的可分性^[29]。但是,这种计算的复杂性和准确性需要研究。(5)在大数据环境下,相似度矩阵在变量的相关性上可能是稀疏的,如何寻找有效的正交基对矩阵进行变换,发现最稀疏矩阵或者近似最优矩阵是一个问题,其中难点在于如何找到合适的正交基。

参 考 文 献

- [1] von Luxburg U. A tutorial on spectral clustering [J]. *Statistics and Computing*, 2007, 17(4): 1-32
- [2] Bath F R, Jordan M I. Learning spectral clustering[R]. Technical report, UC Berkeley, 2003
- [3] Ng A, Jordan M, Weiss Y. On spectral clustering: analysis and an algorithm [C] // *Neural Information Processing Systems*, 2001, 14: 849-856
- [4] 盛宝怀,等. 核学习中的非光滑分析法[M]. 北京: 科学出版社, 2012
- [5] Fowlkes C, Belongie S, Chung F, et al. Spectral grouping using the Nyström method[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, 26(2): 214-225
- [6] Williams C K I, Seeger M. The effect of the input density distribution on kernel-based classifiers[C] // *Proceedings of the 17th International Conference on Machine Learning*. 2000: 1159-1166
- [7] Drineas P, Mahoney M W. On the Nyström method for approximating a Gram matrix for improved kernel-based learning[J]. *Journal of Machine Learning Research*, 2005, 6: 2153-2175
- [8] 李小斌, 田铮. 基于谱聚类的图像多尺度随机树分割[J]. *中国科学 E 辑: 信息科学*, 2007, 37(8): 1073-1085
- [9] Zhang K, Tsang I W, Kwok J T. Improved Nyström low-rank approximation and error analysis[C] // *Proceedings of the 25th International Conference on Machine Learning*. 2008: 1232-1239
- [10] Shi Tao, Belkin M, Yu Bin. Data spectroscopy: eigenspaces of convolution operators and clustering[J]. *The Annals of Statistics*, 2009, 37(6B): 3960-3984
- [11] Kumar S, Mohri M, Talwalkar A. Sampling methods for the Nyström method[J]. *Journal of Machine Learning Research*, 2012(13): 981-1006
- [12] 韩彦彬. 高维正定核的本征值[J]. *数学学报*, 1993, 32(2): 188-194
- [13] von Luxburg U, Belkin M, Bousquet O. Consistency of spectral clustering[J]. *The Annals of Statistics*, 2008, 36: 555-586
- [14] Rosasco L, Belkin M, Vito E D. On learning with integral operators[J]. *Journal of Machine Learning Research*, 2010, 11: 905-934
- [15] Pelletier B, Pudlo P. Operator norm convergence of spectral clustering on level sets[J]. *Journal of Machine Learning Research*, 2011, 12: 385-416
- [16] Belkin M, Niyogi P. Laplacian eigenmaps and spectral techniques for embedding and clustering[C] // *Neural Information Processing Systems*. 2001, 14: 585-591
- [17] Belkin M, Niyogi P. Convergence of Laplacian eigenmaps[C] // *Neural Information Processing Systems*. 2006: 129-136
- [18] Lafon S S. Diffusion Maps and Geometric Harmonics[D]. PhD thesis, Yale University, 2004
- [19] Hein M, Audibert J-Y, von Luxburg U. From graphs to manifolds-weak and strong pointwise consistency of graph Laplacians [C] // *18th Annual Conference on Learning Theory*. 2005: 470-485
- [20] Singer A. From graph to manifold Laplacian; the convergence rate[J]. *Applied and Computational Harmonic Analysis*, 2006, 21(1): 135-144
- [21] Bengio Y, et al. Spectral clustering and kernel PCA are learning eigenfunctions[R]. Technical Report 1239, Département d'Informatique et Recherche Opérationnelle, Université de Montréal, Montréal, 2004
- [22] 张振跃, 查宏远. 线性低秩逼近与非线性降维[J]. *中国科学 A 辑, 数学*, 2005, 35(3): 273-285
- [23] Frieze A, Kannan R, Vempala S. Fast Monte-Carlo algorithm for finding low-rank approximations [J]. *Journal of the ACM*, 2004, 51(6): 1025-1041
- [24] Frieze A, Kannan R. Quick approximation to matrices and applications[J]. *Combinatorica*, 1999, 19(2): 175-220
- [25] Filippone M, Camastra F, Masulli F, et al. A survey of kernel and spectral methods for clustering[J]. *Pattern Recognition*, 2008, 41: 176-190
- [26] De la Torre F. A least-squares unified view of PCA, LDA, CCA, and spectral graph methods[R]. CMU-RI-TR-08-29, Robotics Institute, Carnegie Mellon University, 2008
- [27] Yan S, Xu D, Zhang B, Zhang H. Graph embedding and extensions: A general framework for dimensionality reduction [J]. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 2007, 29(1): 40-51
- [28] Cai D, He X, Han J. SRDA: An efficient algorithm for large-scale discriminant analysis[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2008, 20(1): 1-12
- [29] Belkin M, Sinha K. Toward learning Gaussian mixtures with arbitrary separation[C] // *COLT*. 2010: 407-419