

改进的自适应谱聚类 NJW 算法

大连理工大学 李金泽 徐喜荣 潘子琦 李晓杰
(大连理工大学计算机科学与技术学院 辽宁 大连 116024)

徐喜荣 副教授

《计算机科学》 2017 年 6 月

摘要: 聚类算法是近年来国际上机器学习领域的一个新的研究热点,为了能在任意形状的样本空间上聚类,学者们提出了谱聚类和图论聚类等优秀的算法。本文首先介绍了图论聚类算法中的谱聚类经典NJW算法和NeiMu图论聚类算法的基本思路,提出了改进的自适应谱聚类NJW算法。本文提出的自适应NJW算法的优点在于无需调试参数,即可自动求出聚类个数,克服了经典NJW算法需要事先设置聚类个数且需反复调试参数 δ 才能得出数据分类结果的缺点。在UCI标准数据集及实测数据集上对自适应NJW算法与经典NJW算法、自适应NJW算法与NeiMu图论聚类算法进行了比较。实验结果表明,自适应NJW算法具有方便快捷,较好的实用性。

关键词: 谱聚类;图论聚类;本征间隙;

Abstract: Clustering algorithm is a new research hotspot in the field of machine learning in recent years. In order to be able to cluster in the sample space of any shape, the scholars have put forward the excellent algorithms such as spectral clustering and graph theory clustering. In this paper, we first introduce the classical NJW algorithm of graph theory clustering algorithm and the basic idea of NeiMu algorithm, and then we put forward with a new algorithm named Improved adaptive spectral clustering NJW algorithm. This method overcomes the shortcomings of the classical NJW algorithm, which need to set the number of clusters in advance and need to be repeated to get the data classification results. We compare the adaptive NJW algorithm with the classical NJW algorithm, the adaptive NJW algorithm and the NeiMu graph theory clustering algorithm on the UCI standard data set and the measured data set. The experimental results show that the adaptive NJW algorithm is convenient and well practicable.

一 引言

聚类分析是机器学习领域中的一个重要分支^[1],是人们认识和探索事物之间内在联系的有效手段。所谓聚类(clustering)就是将数据对象分组成为多个类或簇(cluster),使得在同一簇中的对象之间具有较高的相似度,而不同簇中的对象差别较大。传统的聚类算法,如k-means 算法、EM 算法等都是建立在凸球形的样本空间上,但当样本空间不为凸时,算法会陷入局部最优。

为了能在任意形状的样本空间上聚类,学者们提出了谱聚类和图论聚类等优秀的算法。国内的张亚平等人也发表了一篇基于密度敏感的自适应谱聚类算法^[2],提出了一种基于密度敏感的相似性度量,有效地描述了数据的实际聚类分布。王娜等人提出了一种主动式半监督的谱聚类算法^[3],使得类内各点紧凑,类间散布。谱聚类经典算法NJW算法可以有效的对数据进行聚类分类,但其相似性矩阵的参数需大量人工调试方可确定,并且传统的NJW算法的聚类个数需人为规定,不能自动计算分类个数。为此,Zelnik-Manor和Perona提出了Self-Tuning谱聚类算法^[2],他们利用一种称为“Local Scaling”的思想很好的解决了相似性矩阵的参数问题。国内的孔万增等人提出了基于本征间隙与正交特征向量的自动谱聚类^[4],解决了经典谱聚类NJW算法需要人为输入聚类个数的问题。基于图论的聚类算法 NeiMu (Neighboring Mutually)首先分析数据中的对象,寻找每个对象的k近邻,根据k近邻关系构造k近邻有向图,然后通过k近邻有向图中的k-互邻居关系构造k-聚类图,发现数据中的自然聚类^[5]。

二 图论聚类的基本理论

作者简介: 李金泽(1995-),男,主要研究方向为深度学习、神经网络。

图论聚类法最早是由Zahn提出来的，又称作最大(小)支撑树聚类算法^[6]。图论聚类法的前提是要把聚类的数据表示成一个带权的无向图。图是由一些节点以及连接节点之间的边或线所构成的几何图形。在图论中，它可以反映一些对象之间的关系。两点之间不带箭头的连线称为无向边 e_i 。由点和无向边构成的图叫无向图，记作 $G=(P, E)$ ，其中 P 是图 G 的点集合， E 是图 G 的边集合。对无向图 G 的每一条边 (p_i, p_j) ，相应地有一个数 w 表示边 (p_i, p_j) 上的权。权函数一般定义为两节点之间的相似度^[7]。

谱聚类算法最早是从谱图划分理论演化而来。设每一个样本数据看成是图中的顶点 V ，根据样本两两之间的相似度将相应顶点之间的连接边 E 赋权重 W ，于是就得到基于样本的相似度的无向加权图 $G=(V, E, W)$ 。从图论的最优划分理论来看，类似于线性判别算法，就是使划分成的若干子图之间相似度最小，子图内部相似度最大^[8]。常用的划分准则有最小割集、规范割集、平均割集、比例割集以及最小最大割集等准则^[9~11]。图划分问题的最优解是一个NP难题，一个较好的求解方法是考虑问题的连续松弛形式，于是便可将原问题转换成相似度矩阵或Laplacian矩阵的谱分解。

(一) 典谱聚类NJW算法

根据不同的准则函数及谱映射方法，谱聚类算法发展了很多不同的具体实现方法，但是都可以归纳为下面三个主要步骤^[12]：第一步是构建表示样本集的矩阵 Z ；第二步通过计算矩阵 Z 的前 k 个特征值与特征向量，构建特征向量空间；第三步利用 k -means或其它经典聚类算法对特征向量空间中的特征向量进行聚类。

经典谱聚类NJW算法的主要步骤如下：

Step 1. 构建相似矩阵 A 。 A 中的元素为 $A_{ij} = \exp(-\frac{d^2(v_i, v_j)}{2\sigma^2})$ ，其中， v_i 是数据样本点， $d(v_i, v_j)$ 代表两样本点之间的欧式距离 $\|v_i - v_j\|$ ， σ 是尺度参数，决定样本点之间衰减速度。

Step 2. 构造规范性Laplacian矩阵 L_{sym} 。首先构造度矩阵 D ，度矩阵 D 的主对角线上的元素 $D(i, j)$ 为相似性矩阵 A 的第 i 行元素之和，其它元素均为0；然后构造拉普拉斯矩阵 $L = D - A$ 为 $L = D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ ；进一步再构造规范性Laplacian矩阵 $L_{sym} = D^{-\frac{1}{2}}LD^{-\frac{1}{2}} = E - D^{-\frac{1}{2}}AD^{-\frac{1}{2}}$ 。

Step 3. 计算矩阵 L_{sym} 的前 k 个最大特征值对应的特征向量 x_1, \dots, x_k （必要时需作正交化处理），此 k 值需人为输入，并据此构造矩阵 $X=[x_1, \dots, x_k] \in R_n \times k$ ，将矩阵 X 的行向量转变为单位向量，得到矩阵 Y ： $Y_{ij} = \frac{x_{ij}}{\sqrt{\sum_j x_{ij}^2}}$ 。将矩阵 Y 的每一行看作是 R_k 空间中的一个点，对其使用 k 均值算法或任意其它经典聚类算法，得到 k 个聚类；将数据点 y_i 划分到聚类 j 中，当且仅当 Y 的第 i 行被划分到聚类 j 中。

(二) NeiMu图论聚类算法

基于NeiMu (Neighboring Mutually) 图论的聚类算法首先分析数据中的对象，寻找每个对象的 k -近邻，根据 k -近邻关系构造 k -近邻有向图，然后通过 k -近邻有向图中的 k -互邻居关系构造 k -聚类图，发现数据中的自然聚类。该算法的特点是根据数据之间的互为 k -近邻关系确定数据中的自然簇，而不必引入其他方法来划分小簇，从而能够保证对象不会被错误聚类，仅会与其他小簇一起融合到一个大簇中。这一优点可以有效保证NeiMu 算法的聚类质量。NeiMu算法的步骤如下所示^[4]：

Step 1. 构造距离矩阵 D

$$D = \begin{bmatrix} 0 & d(1,2) & \cdots & d(1,n) \\ d(2,1) & 0 & \cdots & d(2,n) \\ \vdots & \vdots & \ddots & \vdots \\ d(n,1) & d(n,2) & \cdots & 0 \end{bmatrix} \quad (1)$$

其中， $d(i, j)$ 为对象 i 和对象 j 之间的距离，满足 $d(i, j) \geq 0$ ， $d(i, j) = d(j, i)$ 且 $d(i, i) = 0$ 。在计算和存储中， k 近邻有向图 $kNN_G = (W, E)$ 可以表示为邻接矩阵的形式，矩阵的元素 $A_{kNN_G}(i, j)$ 为：

$$A_{KNN_G}(i, j) = \begin{cases} 1, & \text{如果边} W_j \text{是} W_i \text{的} k \text{近邻} \\ 0, & \text{否则} \end{cases}$$

Step 2. 构造k-聚类图

$$A_{KNN_G}(i, j) = \begin{cases} 1, & \text{如果} A_{KNN_G}(i, j) * A_{KNN_G}(j, i) = 1 \\ 0, & \text{如果} A_{KNN_G}(i, j) * A_{KNN_G}(j, i) = 0 \end{cases} \quad (3)$$

Step 3. 遍历k-聚类图

k-聚类图包含若干个由k-互邻居组成的连通子图（当子图仅包含一个结点时子图退化成孤立点）。因此只需对k-聚类图进行深度优先搜索遍历，即可得到最终聚类结果，使得k-聚类图中每个连通子图的遍历结果构成一个聚类。

三 改进的自适应谱聚类NJW算法

传统谱聚类NJW算法能比较准确地对数据进行聚类，但仍有很多不足之处。在Step1中，高斯核函数的尺度参数 σ 需要人为规定，需反复调试方可确定其最优值。在Step 3中需要人为规定聚类个数k，无法自动获得聚类的个数。为此，Zelnik-Manor和Perona提出了Self-Tuning谱聚类算法，他们利用一种称为“Local Scaling”的思想很好的解决了Step1中函数的参数 σ 无法准确确定的问题。国内的卜德云在Self-Tuning的基础上做了关于自适应谱聚类算法的相关研究[13]。

本文采取Self-Tuning的有关算法，对经典谱聚类NJW算法中Step1改进，将Step1中的 $A_{ij} = \exp(-\frac{d^2(v_i, v_j)}{2\sigma^2})$ 改进为 $A_{ij} = \exp(\frac{-d^2(s_i, s_j)}{\sigma_i \sigma_j})$ ，其中 $\sigma_i = d(s_i, s_T)$ 是样本点i 到其第T个近邻的点的欧式距离，文献[14]、[15]、[16] 测试多个人工数据集和实际数据集认为T取7时算法的效果最佳，这里不做多余讨论。

对经典谱聚类NJW算法中Step3采取了基于本征间隙的自动谱聚类算法，解决了传统NJW算法无法自动计算聚类个数的问题。计算规范Laplacian矩阵的特征值并按顺序从大到小排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ ，计算本征间隙序列 $\{g_1, g_2, \dots, g_{n-1} \mid g_i = \lambda_i - \lambda_{i+1}\}$ ，在本征间隙序列中依次寻找第一个极大值，则该值对应的下标即为类个数。即类别数 $k = \text{argmin}\{g_i - g_{j \mid j < i} > 0 \& \& g_i - g_{i+1} > 0\}$ [17]。

改进的自适应谱聚类NJW算法如下：

Step 1. 构建相似矩阵A。 A中的元素为 $A_{ij} = \exp(\frac{-d^2(s_i, s_j)}{\sigma_i \sigma_j})$ ，其中， $\sigma_i = d(s_i, s_T)$ 是样本点到其第T个近邻的点的欧式距离， v_i 是数据样本点， $d(v_i, v_j)$ 代表两样本点之间的欧式距离 $\|v_i - v_j\|$ ；

Step 2. 首先构造度矩阵D。度矩阵D的主对角线上的元素D(i, j)为相似性矩阵A的第i行元素之和，其它元素均为0。然后再构造拉普拉斯矩阵 $L = D - A$ 为 $L = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$ 。进一步再计算规范性Laplacian矩阵

$$L_{\text{sym}} = D^{-\frac{1}{2}} L D^{-\frac{1}{2}} = E - D^{-\frac{1}{2}} A D^{-\frac{1}{2}};$$

Step 3. 计算规范Laplacian矩阵的特征值并按顺序从大到小排列 $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ ，计算本征间隙序列 $\{g_1, g_2, \dots, g_{n-1} \mid g_i = \lambda_i - \lambda_{i+1}\}$ ，在本征间隙序列中依次寻找第一个极大值，则该值对应的下标即为类个数k；计算矩阵 L_{sym} 的前k个最大特征值对应的特征向量 x_1, \dots, x_k （必要时需作正交化处理）。据此构造矩阵 $X = [x_1, \dots, x_k] \in R^n \times k$ ，将矩阵X的行向量转变为单位向量，得到矩阵Y： $Y_{ij} = \frac{x_{ij}}{\sqrt{\sum_j x_{ij}^2}}$ ，将矩阵Y的每一行看作

是 R_k 空间中的一个点，对其使用k均值算法或任意其它经典聚类算法，得到k个聚类；将数据点 y_i 划分到聚类j中，当且仅当Y的第i行被划分到聚类j中。

四 实验分析与比较

谱聚类是基于谱图划分理论的一种聚类算法，是图论聚类的一个分支，自适应 NJW 算法是对经典谱聚类 NJW 算法的一种改进，为了得出改进的谱聚类自适应 NJW 算法与经典的谱聚类 NJW 算法以及图论聚类 NeiMu 算法之间的区别，本文分别在 UCI 标准数据中以及实测数据进行了相关的比较实验。

(一) 基于UCI标准数据的自适应NJW与经典NJW聚类的比较

本实验在国际通用数据库 UCI 数据库(专门用于测试分类、聚类算法)中的 Tea、Iris 以及 Ionesphere 三个数据集进行了改进的自适应 NJW 算法与经典 NJW 算法的比较实验。本次实验中经典 NJW 算法参数 δ 的调试步长为 0.01, 参与比较的自适应 NJW 算法的参数 T 取经验值 7. 本次实验在计算机内存 4.0G、CPU 3.2GHz 的硬件条件以及 Matlab 2016b 环境下运行。

本次实验中的经典 NJW 算法参数的调整步长为 0.01, k 值取数据集的固有聚类个数, 准确率平均值是指参数 δ 在较好聚类效果区间内以 0.01 为步长所得到的所有准确率的平均值, 因此本次比较的实验结果因为参数的取值范围而具有局限性。从表 1 可以看出自适应 NJW 算法在数据分类准确率上略微提高。自适应 NJW 算法最大的优点在于无需调试参数, 只需将参数 T 设置为经验值 7 即可自动的、较为准确地求出聚类个数, 而经典 NJW 算法不仅需要事先设置聚类个数, 且参数 δ 也很大程度上影响数据分类准确率, 需反复调试 δ 才能得出最好的数据分类结果。

(二) 基于UCI标准数据的自适应NJW与图论聚类NeiMu算法的比较

本实验在国际通用数据库 UCI 数据库(专门用于测试分类、聚类算法)中的 Satimage、Iris、Ionesphere 及 Segmentation 四个数据集进行了谱聚类算法与图论聚类算法的比较实验, 本次实验在计算机内存 4.0G、CPU 3.2GHz 的硬件条件以及 Matlab 2016b 环境下运行。

从表 2 可以看出, 多维数据用谱聚类的自适应 NJW 算法比图论聚类 NeiMu 算法效果要好。在用 Satimage 数据集和 Segmentation 数据集做算法比较实验时, NeiMu 算法出现严重的分类失误。具体体现在个别簇包含过多的点而某些簇包含极少数的点, 这就体现出图论聚类的不足。基于图论聚类的 NeiMu 算法对点间距离较大或者密度比较稀疏的数据集能起到很好的聚类效果, 对于点间距离较小或者数据分布与点间距离没有关系的数据集分类效果较差。

(三) 基于UCI标准数据的经典NJW算法与图论聚类NeiMu算法的比较

本实验在国际通用数据库 UCI 数据库(专门用于测试分类、聚类算法)中的 Tea、Ionesphere 以及 Satimage 三个数据集进行了改进的图论聚类 NeiMu 算法与经典 NJW 算法的比较实验, 实验参数 δ 是聚类结果最好时的取值. 本次实验在计算机内存 4.0G、CPU 3.2GHz 的硬件条件以及 Matlab 2016b 环境下运行。

从表 3 可以看出, 多维数据用谱聚类的经典 NJW 算法比图论聚类 NeiMu 算法效果要好。在对 Satimage 数据集进行图论聚类时, 出现较为严重的失误。此失误表现在 Sattimage 数据集的固有类数较多, 个别簇包含过多被错误聚类的点而某些簇包含极少被正确聚类的点。基于图论聚类的 NeiMu 算法对点间距离较大或者密度比较稀疏的数据集能起到很好的聚类效果, 对于点间距离较小或者数据分布与点间距离没有关系的数据集分类效果较差。

表1 自适应NJW算法与经典NJW算法在UCI数据集上的比较实验

	δ 的取值范围	聚类效果	经典NJW算法	自适应NJW算法
Iris	$0.16 < \delta < 0.25$	较好	K=3时, δ 在 $0.16 \sim 0.25$ 得到的准确率平均值为 0.903	无需调试参数 聚类准确率为 0.920
	$\delta < 0.16$ 或 $\delta > 0.25$	较差		
Tea	$0.08 < \delta < 0.11$	较好	K=3 时, δ 在 $0.08 \sim 0.11$ 得到的准确率平均值为 0.384	无需调试参数 聚类准确率为 0.411
	$\delta < 0.08$ 或 $\delta > 0.11$	较差		
Ione	$0.15 < \delta < 0.25$	较好	K=2时, δ 在 $0.15 \sim 0.25$ 得到的准确率平均值为 0.721	无需调节参数 聚类准确率为 0.746
	$\delta < 0.15$ 或 $\delta > 0.3$	较差		

表2 图论聚类算法与谱聚类算法在UCI数据集上的比较实验结果

数据属性	算法性能 (分类准确率)
------	--------------

	样本数	维数	固有类数	NeiMu算法(k值)	自适应NJW算法
Iris	150	4	3	0.673 (k=16)	0.913
Ionesphere	351	34	2	0.644 (k=24)	0.741
Satimage	444	36	6	0.340 (k=30)	0.858
Segmentation	210	19	7	0.343 (k=9)	0.779

表3 图论聚类算法NeiMu与经典NJW算法的比较实验

	数据属性			算法效率(分类准确率)	
	样本数	维数	固有类	NeiMu算法(k值)	经典NJW算法(δ 值)
Iris	150	4	3	0.673 (k=16)	0.920 ($\delta=0.17$)
Ionesphere	351	34	2	0.644 (k=24)	0.721 ($\delta=0.20$)
Satimage	210	19	6	0.340 (k=30)	0.865 ($\delta=0.13$)

(四) 基于实测数据的自适应NJW、经典NJW算法与NeiMu图论聚类之间的比较

高场不对称波形离子迁移谱(High-field Asymmetric Waveform Ion Mobility Spectrometry, FAIMS)是一种大气环境下利用不同离子在强电场中离子迁移率非线性变化特点来实现离子空间分离和识别的检测技术。FAIMS 利用不同离子在强电场中离子迁移率非线性变化特性来实现离子的分离,从而实现很多领域的数据采集工作。本文数据是由FAIMS 仪器采集的,获得三维空间坐标下的谱图数据

(x, y, z), 其中 x 轴: 补偿电压 CV, 共有 512 个采样点, 从 $-6V \sim +6V$; y 轴: 分离电压 DV, 共有 26 个采样点, 范围从 $0\% \sim 100\%$, 步长为 4%; z 轴: 离子流强度 Ion Current, 即每一幅谱图共有 $512 \times 26 = 13312$ 个点(x, y, z)。

本文待处理的数据是针对空间中的 26 条曲线, 采用谱图峰识别算法识别出每条曲线上的若干个谱峰点所得到的谱峰点集合, 一共 66 个三维数据点。

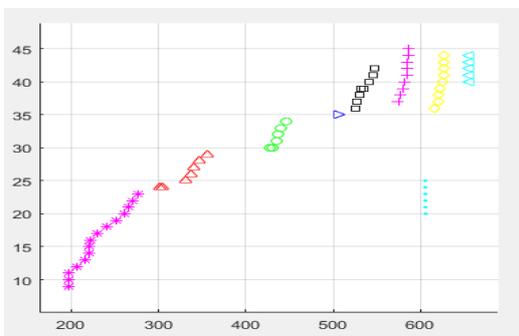


图2 NeiMu 算法的实验结果(K=4)

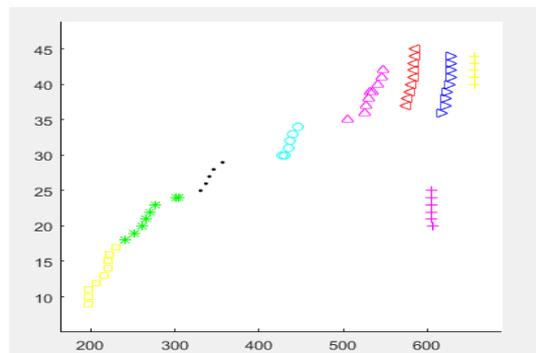


图3 经典 NJW 算法的聚类结果($\sigma=0.16$)

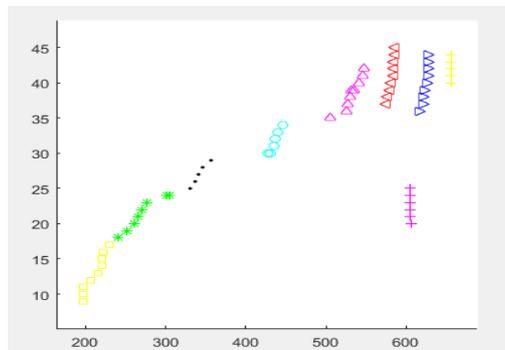


图4 自适应 NJW 算法的聚类结果

图 2 为 NeiMu 算法参数 $K=4$ 时的实验结果，由图 2 可以看出数据被分为 9 类，并且第四个聚类（从左向右）只含有一个孤点，可见图论聚类 NeiMu 算法对孤立点比较敏感。图 3 为经典 NJW 算法参数 $\sigma = 0.16$ ， $k=9$ 时的实验结果，由图 3 可以看出经典 NJW 算法与自适应 NJW 算法的聚类效果无异，但经典 NJW 算法需要反复调节参数 σ 的值，且需手动输入参数 k 的值，直到聚类效果较好为止。图 4 为自适应 NJW 算法的实验结果，由图 4 可知数据点被分成 9 类，与图 2 不同之处在于自适应 NJW 算法将孤立点划分到其他聚类之中，并且不需

表 4 图论聚类与谱集上的对比

	孤立点敏感	运算时间	参数
NeiMu	敏感	0.203s	$K=4$ 最优
经典 NJW	不敏感	1.601s	$\sigma = 0.16$ 最优
自适应 NJW	不敏感	1.430s	无需参数

调试参数，就可以准确得出聚类个数。
聚类在 FAMIS 数据

由表 4 易知自适应 NJW 算法和经典耗时比较长，且对孤立点不敏感。NJW 算法的突出优势在于能自动确定聚类个数，不用人工调试参数，这是图论聚类 NeiMu 算法和经典 NJW 算法所不能达到的。相比图论聚类 NeiMu 算法，基于谱聚类的自适应 NJW 算法则能很好地对高维数据进行聚类，无需人工调试输入参数，即可较为准确的确定聚类个数，这些是图论聚类 NeiMu 算法所不能及的。

五 结论

本文将经典 NJW 算法做了两项改进，一方面在数据相似性矩阵引入了 Self-Turning 算法思想，解决了高斯核函数参数 σ 难以确定的问题。另一方面采用孔万增等人提出的基于本征间隙的自动谱聚类算法，使得 NJW 算法可以较为准确的自动计算聚类个数。通过若干次改进后的自适应 NJW 算法与经典 NJW 算法和图论聚类 NeiMu 算法三者的对比试验，我们可以得出以下结论：

(1) 谱聚类基于谱图的划分理论，是图论聚类的一个分支。自适应 NJW 算法是基于谱聚类的、对经典 NJW 算法的改进。自适应 NJW 算法无需调试参数的优点，决定了它方便快捷、适用于现场检测的实用特点。

(2) 图论聚类适用于密度变化大的数据、大小相差大的数据、维数较低的数据集，尤其在二维数据处理上，能快速准确的对数据进行聚类。自适应谱聚类 NJW 算法适用于存在维数较多，数据量较大的数据集。

(3) 从实测数据上可以看出，图论聚类对孤立点敏感，谱聚类对孤立点不敏感。在对数据进行定性分析时，谱聚类优于图论聚类。在对特殊点处理时，图论聚类优于谱聚类。

参考文献：

[1] Jain A, Murty M, Flynn P. Data clustering: A Review [J]. ACM Computing Surveys, 1999, 31(3):264-323
 [2] 张亚平, 杨明. 一种基于密度敏感的自适应谱聚类算法 [J]. 数学的实践与认识, 2013, 43(20):150-156
 [3] 王娜, 李霞. 基于监督信息特性的主动半监督谱聚类算法 [J]. 电子学报, 2010, 38(1):172-176. Wang Na, Li Xia. Active semi-supervised spectral clustering based on pairwise constraint [J]. Acta Electronica

Sinica, 2010, 38(1):172-176. (in Chinese)

- [4] 孔万增, 孙志海, 杨 灿, 戴国骏, 孙昌思核. 基于本征间隙与正交特征向量的自动谱聚类[J]. 电子学报, 2010, 38(8): 1880- 1891
- [5] 应德全, 应晓敏, 叶继华. 一种基于图论的聚类算法NeiMu[J]. 计算机工程与应用, 2009, 45(3): 47 -50
- [6] 钱云涛, 赵荣椿, 谢维信. 鲁棒聚类—基于图论和目标函数的方法[J] . 电子学报, 1998 , 26 (2):91 -94
- [7] 刘锁兰, 王江涛, 王建国, 杨靖宇. 一种新的基于图论聚类的分割算法[J]. 计算机科学, 2008, 35 (9) : 245 -247
- [8] J Shi, J Malik .Normalized cuts and image segmentation[J] .IEEE Transactions on Pattern Analysis and Machine Intelligence, 2000 , 22 (8):888-905.
- [9] 蔡晓妍, 戴冠中, 杨黎斌. 谱聚类算法综述[J] . 计算机科学, 2008 , 35(7):14-18. CAI Xiao-yan, DAI Guan-zhong, YANG Li-bin. Survey on Spectral Clustering Algorithms[J] .Computer Science , 2008 ,35(7):14 -18 .
- [10] M Filipponea , F Camastrab, F Masullia , S Rovetta .A survey of kernel and spectral methods for clustering[J] .Pattern Recognition, 2008, 41(1):176 -190 .
- [11] U Luxburg. A tutorial on spectral clustering[J] .Statistics and Computing, 2007, 17(4):395-416 .
- [12] Verma D, Meila M .A comparison of spectral clustering algorithms .Technical report , 2003 .UW CSE Technical report 03-05-01
- [13] 卜德云, 张道强. 自适应谱聚类算法研究[J]. 山东大学学报, 2009, 39 (5) : 22 -39
- [14] Zhao F, Jiao L C, Liu H Q, et al, Spectral Clustering with Eigenvector Selection Based on Entropy ranking. Neurocomputing, 2010, 73(10 /11 /12): 1704 - 1717
- [15] Zhao F, Jiao L C, Liu H Q, et al. Semi-Supervised Eigenvector Selection for Spectral Clustering. Pattern Recognition and Artificial Intelligence, 2011, 24(1): 48 - 56 (in Chinese)
(赵凤, 焦李成, 刘汉强, 等. 半监督谱聚类特征向量选择算法. 模式识别与人工智能, 2011, 24(1): 48 - 56)
- [16] Zelnik-Manor L, Perona P. Self-Tuning Spectral Clustering // Proc of Advances in Neural Information Processing Systems 17. Vancouver, Canada, 2004: 1601 - 1608
- [17] Y Ng , M I Jordan, Y Weiss .On spectral clustering:Analysis and an algorithm[A] .Proceedings of the 14th Advances in Neural Information Processing Systems (NIPS 2002) [C] .Cambridge, MIT Press , 2002:849-856.