中国经济文库 · 应用经济学精品系列（二）

基于谱聚类的金融时间序列数据挖掘方法研究

苏木亚◎著

Research on Financial Time Series Data Mining Method Based on Spectral Clustering

中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE

中国经济文库·应用经济学精品系列（二）

苏木亚◎著

基于谱聚类的金融时间序列数据挖掘方法研究

Research on Financial Time Series
Data Mining Method
Based on Spectral Clustering

中国经济出版社
CHINA ECONOMIC PUBLISHING HOUSE
北京

图书在版编目（CIP）数据

基于谱聚类的金融时间序列数据挖掘方法研究/苏木亚著.
北京：中国经济出版社，2013.8
ISBN 978-7-5136-2582-1
Ⅰ.①基… Ⅱ.①苏… Ⅲ.①金融—数据收集—方法研究 Ⅳ.①F830.41
中国版本图书馆 CIP 数据核字（2013）第 118947 号

责任编辑　崔清北　陈　洁　杨元丽
责任审读　霍宏涛
责任印制　张江虹
封面设计　华子图文设计公司

出版发行　中国经济出版社
印 刷 者　北京市媛明印刷厂
经 销 者　各地新华书店
开　　本　710mm×1000mm　1/16
印　　张　12.75
字　　数　200 千字
版　　次　2013 年 8 月第 1 版
印　　次　2013 年 8 月第 1 次
书　　号　ISBN 978-7-5136-2582-1/F·9786
定　　价　45.00 元

中国经济出版社 **网址** www.economyph.com **社址** 北京市西城区百万庄北街 3 号 **邮编** 100037
本版图书如存在印装质量问题，请与本社发行中心联系调换（联系电话：010-68319116）

序

近年来，大数据已成为互联网信息技术行业的流行词汇，并引起社会的高度关注。2011 年麦肯锡全球研究院（MGI）发布了研究报告《大数据：下一轮创新、竞争和生产力的前沿》，率先从经济和商业维度诠释了大数据发展的巨大潜力。麦肯锡在研究报告中指出："数据已经渗透到当今每一个行业和业务职能领域，成为重要的生产因素。人们对于海量数据的挖掘和运用，预示着新一波生产率增长和消费者盈余浪潮的到来。"目前，大数据时代已经来临，在商业、经济及其他领域中，许多决策日益基于数据和分析，而非基于经验和直觉。

虽然不同行业数据的存储量、类型和内容会有所不同，但越来越多的行业都呈现出大数据现象，特别是金融领域、信息服务领域、公共事业服务等数据强度高的行业更加具有通过大数据来创造价值的潜力。大数据的价值密度相对较低，如何通过强大的数据挖掘算法迅速地完成数据的价值"提纯"，是大数据时代亟待解决的难题。

在大数据分析、挖掘以及应用中，存在大量的聚类问题。聚类是将对象划分到不同簇的一个过程，以使同一个簇中的对象有很大的相似性，而不同簇间的对象有很大的相异性。聚类分析技术经过多年的发展，产生了大量解决聚类问题的相关算法。谱聚类算法是建立在谱图理论基础上的一类新的聚类算法，其本质是将聚类问题转化为图的最优划分问题，具有能在任意形状的样本空间上聚类、思想简单、易于执行等优点，应用前景广阔。

苏木亚所著《基于谱聚类的金融时间序列数据挖掘方法研究》一书，以博士论文的研究工作为基础，围绕谱聚类方法的理论、算法及其在金融时间序列数据挖掘中的应用，对自己近几年的科研成果进行了系统的总结。我认为本书的价值在于：一是对谱聚类方法的理论基础和算

法设计进行了一些有益的探讨，得到了一些有意义的结论；二是基于谱聚类方法，对股票收益率数据、股指数据、开放式基金数据进行了探索式挖掘，挖掘结果表明，谱聚类方法应用于金融数据挖掘领域是可行的。

金融领域具有大量的大数据挖掘问题，本书中谱聚类方法所处理的金融数据从数据量上可能还不够“大”，欲将谱聚类方法真正应用于海量的金融数据挖掘，还有很长的路要走。被誉为“大数据时代的预言者”维克托·麦而－舍恩伯格认为：大数据中的“大”并非绝对意义上的大，而是说要尽可能多地掌握数据，利用所掌握的数据进行分析、论证假设，从而获得有价值的见解。从这点看来，本书在金融“大”数据挖掘方面也算是一次有益的探索和尝试。

作为苏木亚攻读博士学位期间的博士生指导教师，我愿意为这本专著作序。希望本书的出版，能为谱聚类方法的理论研究及金融数据挖掘的应用研究起到抛砖引玉的作用。

郭崇慧

大连理工大学

2013 年 7 月 15 日

摘　要

数据挖掘是商务智能的核心技术之一。近年来，数据挖掘已经被广泛应用于金融管理、客户关系管理、工作流管理、风险管理等管理领域，为企业的决策支持、成本控制、组织协同等提供了极大帮助。

聚类分析是数据挖掘研究的一个重要组成部分。聚类是把对象的集合分组成为多个簇的过程，使同一个簇中的对象具有较高的相似度，而不同簇的对象差别较大。聚类分析已在股票数据分析、市场细分、生产监管、异常检测等领域发挥重要作用。在聚类分析的众多算法中，谱聚类是基于谱图理论的一类新的聚类方法，具有能够对任意形状的数据进行划分、易于执行等优点。许多文献已经对谱聚类算法的特点进行了深入研究，并提出了一些改进方法。然而，无论从理论、算法还是实践层面，仍有很多问题有待解决，例如：谱聚类方法中如何确定数据集的既合理又稳定的聚类数目？如何选取包含聚类信息的特征向量组？从矩阵扰动理论角度看多路归一化割谱聚类方法是否合理？利用成分分析法对单变量时间序列降维的原理是什么？如何利用谱聚类方法对实际的金融时间序列数据进行分析？

有鉴于此，本书围绕谱聚类方法及其在金融时间序列数据挖掘中的应用做了以下工作：

（1）针对经典谱聚类的聚类数目估计问题，提出了基于稳定性的非唯一聚类数目确定方法。对候选聚类数目 k，该方法利用本书提出的 $Ratio(k)$ 指标评价与其对应的划分结果的合理性。进一步，通过改变高斯核参数的大小来确定划分结果的稳定性。所提方法能够找出一组既合理又稳定的聚类数目。

（2）针对谱聚类方法中选择包含聚类信息的特征向量组问题，提出了谱聚类中自动选择包含聚类信息的特征向量组方法。通过该方法找

出的包含聚类信息的特征向量组关于高斯核参数的稳定性较好、其聚类特征比较明显，而且该方法易于执行。

（3）以矩阵扰动理论为工具，对多路归一化割谱聚类方法的合理性进行了分析。分析结果表明，从矩阵扰动理论角度看，在理想情形下设计谱聚类方法并将其推广到一般情形的做法是合理可行的。

（4）针对主成分分析法对单变量时间序列降维原理问题，从线性空间中向量、基向量和系数矩阵间关系角度对其进行解释。在此基础上，提出了一种基于主成分分析的单变量时间序列谱聚类方法。该方法体现了在线性空间中同一组基下，用系数之间的相似性来反应对应向量之间相似性的思想。

（5）针对独立成分分析法对单变量时间序列降维原理问展开讨论，考虑了独立成分分析法的含混性对聚类结果的影响。在理论分析的基础上提出了一种基于独立成分分析的时间序列多路归一化割谱聚类方法。该方法首先选用独立成分分析法对时间序列数据进行特征提取，然后利用本文提出的广义特征值法估计聚类数目，最后利用多路归一化割谱聚类方法对提取出的特征数据进行聚类，从而完成对原单变量时间序列的聚类任务。

（6）采用多路归一化割谱聚类方法，对欧洲主权债务危机背景下的全球主要股指进行了联动性与稳定性分析。首先分别实证考察了全球主要股指在欧洲主权债务危机开始前、开端、发展、蔓延、升级、调整、再升级以及复苏八个不同阶段内的联动性及各相邻阶段之间的变化，即稳定性特征。其次考虑了全球主要股指在欧洲主权债务危机不同阶段的聚集情况。

（7）采用多路归一化割谱聚类方法和独立成分分析法对国内开放式基金进行了投资风格识别研究。为此，首先，利用独立成分分析法对所选出的开放式基金进行特征提取。其次，采用本书提出的广义特征值法估计聚类数目并运用多路归一化割谱聚类方法对提取出的特征进行划分，从而完成对原开放式基金的投资风格分类。最后，选用本书提出的基于 Sharpe 系数间隙判断投资风格归属的方法判断各类代表元基金投资风格的具体类型。

Abstract

Data mining is one of the core techniques for business intelligence. In real – world applications, data mining technique has been widely used in financial management, customer relationship management, workflow management, risk management, and so on. It is of great use for the success of enterprises in strategic decision making, cost control, and business collaboration.

Cluster analysis is one of the key components of data mining research. The process of grouping a set of objects into classes of similar objects is called clustering. A cluster is a collection of data objects that are similar to one another within the same cluster and are dissimilar to the objects in other clusters. Cluster analysis has long played an important role in a wide variety of fields such as stock data analysis, market segmentation, production supervision, anomaly detection. Spectral clustering is a novel clustering method which based on the spectral graph theory. Spectral clustering has main advantages of easy implementation and can be used to cluster data with arbitrary shape. Many studies have been devoted to the research on spectral clustering. However, further study still need to be addressed for some important questions in the theory, algorithm and real application of spectral clustering. The questions including how to determine reasonable and stable cluster numbers in spectral clustering? How can we select the informative eigenvectors in spectral clustering? Do we actually compute a reasonable clustering from matrix perturbation theory point of view? What is the principle of using component analysis in dimension reduction of univariate time series? How can we make use of spectral clustering to analyze real financial time series data?

This book thus focuses on spectral clustering methods and their applica-

tion in financial time series data mining as follows:

(1) We propose non - unique cluster numbers determination methods based on stability in spectral clustering. For a candidate cluster number$k$, first we used index $Ratio(k)$ to judge its rationality. Then, by varying the scaling parameter in the Gaussian function to judge whether the reasonable cluster number k is also a stability one. The algorithm mentioned above can determine not only reasonable but also stable cluster numbers of the given data set.

(2) For choosing informative eigenvectors in spectral clustering, we propose an algorithm called automatic selection of informative eigenvectors in spectral clustering (ASIESC). ASIESC differs from previous approaches in that it can distinguish informative eigenvectors remarkably from uninformative ones, easy to be implement and more stable than existing algorithms.

(3) Using matrix perturbation theory to analyze the spectral clustering matrix used in the multiway normalized cut spectral clustering method. The results show that, multiway normalized cut spectral clustering method is reasonable from matrix perturbation theory point of view.

(4) We analyze the principle of dimension reduction for univariate time series via principal component analysis from the linear algebraic point of view. Based on theoretical analysis, we propose univariate time series spectral clustering method based on principal component analysis. The main idea is that, similarities among the univariate time series can be reflected by similarities among the corresponding coefficients under the same basic vectors of linear space.

(5) We discuss the principle of making use of independent component analysis (ICA) to reduce the dimension for univariate time series. Especially, we analyze the impact of ambiguity of the independent components to the clustering results. We propose a spectral clustering method based on independent component analysis for time series according to our theoretical analysis. In the algorithm mentioned above, first, we use ICA to re-

duce the dimension. Then estimate the cluster number via generalized eigenvalues method. At last cluster the feature data by multiway normalized cut spectral clustering method.

(6) We make use of spectral clustering method to analyze comovement and stability of the global stock indices during the European sovereign debt crisis. First, we propose a cluster number determination method based on stability. Then analyze comovement and differences of adjacent time stages of the global stock indices during eight stages, including before crisis, beginning, developing, spreading, uploading, adjusting, re - uploading and recovering. At last, we analyze the distribution of the global stock indices in different time stages.

(7) Study the investment styles recognition of the Chinese open - end funds by the multiway normalized cut spectral clustering method and the ICA. First, we make use of the ICA to extract features, then estimate the cluster number via the generalized eigenvalues method, cluster the feature data by the multiway normalized cut spectral clustering method. At last, judge investment styles according to our investment styles recognition method based on Sharpe's coefficients.

目录 Contents

序 …… 1
摘 要 …… 1
Abstract …… 1

第1章 绪 论

1.1 研究背景及研究意义 …… 1
1.2 国内外相关研究进展 …… 3
1.3 主要研究内容与结构 …… 20

第2章 谱聚类中基于稳定性的非唯一聚类数目确定方法

2.1 引 言 …… 25
2.2 预备知识 …… 25
2.3 聚类结果的合理性与稳定性度量 …… 30
2.4 算法提出 …… 31
2.5 数值实验 …… 33
本章小结 …… 43

第3章 谱聚类中包含聚类信息的特征向量组自动选取方法

3.1 引 言 …… 45
3.2 预备知识 …… 46
3.3 算法原理 …… 47
3.4 算法提出 …… 49
3.5 数值实验 …… 50
本章小结 …… 57

第4章 谱聚类矩阵的扰动分析

4.1 引 言 …… 59

4.2 矩阵扰动理论 …… 59
4.3 规范 Laplace 矩阵的扰动分析 …… 63
4.4 数值实验 …… 76
本章小结 …… 97

第 5 章 基于成分分析的单变量时间序列谱聚类方法

5.1 引 言 …… 99
5.2 基于主成分分析的单变量时间序列谱聚类方法 …… 99
5.3 基于独立成分分析的单变量时间序列多路归一化割谱聚类方法 …… 110
5.4 基于成分分析的单变量时间序列聚类方法在股票数据集上的实验 …… 120
本章小结 …… 131

第 6 章 谱聚类方法在金融时间序列数据挖掘中的应用

6.1 引 言 …… 133
6.2 基于谱聚类的欧洲主权债务危机下全球主要股指联动性分析 …… 133
6.3 基于独立成分分析 - 谱聚类 - Sharpe 模型的开放式基金投资风格识别方法 …… 147
本章小结 …… 163

第 7 章 结论与展望

7.1 主要结论 …… 165
7.2 主要创新点 …… 166
7.3 研究展望 …… 167

参考文献 …… 169
索 引 …… 187

第1章 Chapter 1

绪 论

1.1 研究背景及研究意义

1.1.1 研究背景

商务智能（Business Intelligence，BI）从20世纪90年代开始，已经在众多企业中引起关注，成为业界关注的热点。商务智能自产生以来发展较快，但目前还不成熟，关于商务智能还没有统一的定义，人们只是从不同的角度表达了对商务智能的理解。作为企业界的典型代表，IBM公司认为商务智能是一种能力，通过使用企业的数据资源来制定更好的商务决策。企业的决策人员以数据仓库为基础，通过使用各种查询分析工具、进行联机分析处理或者数据挖掘，再利用决策人员的行业知识，从数据仓库中获得有用的信息，进而帮助企业提高利润，增强竞争力。学术界的观点是：商务智能实际上是帮助企业提高决策水平和运营能力的概念、方法、过程以及软件的集合，其主要目标是将企业所掌握的信息转换成竞争优势，提高企业决策能力、决策效率、决策准确性[1]。目前国内外商务智能的一些应用领域包括生产制造、金融、电信、市场营销和Web应用等[2]。

数据挖掘（Data Mining，DM）是商务智能的核心技术之一[2]。它主要基于人工智能、数据库、统计学等技术，高度自动化地分析企业原有的数据，做出归纳性的推理，从中挖掘出潜在的模式，预测客户的行为，帮助企业决策者调整营销策略，减小风险，做出正确的决策[2-4]。实践表明，数据挖掘已经被广泛应用于金融管理（如贷款偿还预测、

顾客信用分析、客户分类与聚类以及金融犯罪侦破等）、零售业（如促销活动的有效性分析、顾客保持力—顾客忠诚度分析和产品推荐）、电信业（如电信数据多维分析、盗用模式分析和多维关联模式分析等）[5-8]以及一些其他领域[9,10]。随着信息技术的不断发展，数据挖掘将为商务智能的进一步发展发挥更大的作用[1,2]。

聚类分析（Clustering Analysis）是数据挖掘的一个重要分支。聚类分析为探索未知的数据结构提供帮助，并能成为一系列数据分析的起点[11]。聚类分析已在金融管理、市场营销、生产监管、信息检索与分类等商业领域发挥重要作用[12,13]。

谱聚类（Spectral Clustering）是一类新型聚类分析方法，具有能够处理任意形状的数据集、易于执行等优点，从而克服了 k- 均值等传统聚类方法的缺点[14]。然而，无论从理论、算法还是实际应用角度看，考虑到商务实践中大量数据的复杂分布特征，谱聚类方法仍有很多问题有待解决。因此，探讨谱聚类的理论、方法和应用，对商务智能的进一步发展具有较高的理论研究价值和实用价值。

1.1.2 研究意义

本书以金融时间序列数据挖掘为应用背景，针对谱聚类方法的理论、模型及算法进行研究，并探讨如何从实际的金融时间序列中有效地挖掘和发现新的知识和隐含的规律，从而拓展金融时间序列分析的理论与方法，进而拓展商务智能的理论、方法以及实际应用。

本书的选题具有以下几层意义：

（1）谱聚类方法是综合性能较好的一种聚类方法。谱聚类具有易于执行、不对数据分布进行假设等优点。但是不论从理论分析、算法设计以及实际应用等角度看，谱聚类方法都有诸多问题尚未解决。研究成果有助于进一步丰富商务智能的理论和方法。

（2）金融时间序列数据广泛存在于实际的商业活动和经济生活中。金融时间序列数据分析的研究一直是学术界的热点之一。证券市场是最为活跃的金融市场之一，也是一个国家的“经济晴雨表”。股票和基金是两种重要的证券交易工具。与此同时，极其丰富并且公开的股票和基金历史数据的获取比较方便，因而本书的应用领域选择股票数据分析和

基金数据分析。

(3) 数据挖掘是一门极具发展前景的新兴交叉学科。金融时间序列数据挖掘方法能够有效地发现丰富的金融时间序列数据当中蕴含的知识和规律，为金融时间序列的深入分析提供了新的思路和视野。

综上所述，从研究角度看，本书的成果对于丰富商务智能的理论和方法具有一定价值；从实用角度看，本书的研究成果对金融监管部门和金融机构以及投资者进一步了解金融市场的变化规律、进行有效的金融监管、提高投资效率等提供理论支持和技术支撑，并对促进金融市场技术分析理论与方法的不断创新与深入发展，丰富金融管理与投资分析的新方法和新思路等具有一定理论与现实意义。

1.2 国内外相关研究进展

本节由五部分组成。在第一部分介绍谱聚类方法研究进展；由于利用谱聚类方法对时间序列数据聚类时可能需要降维，因此，在第二部分介绍主成分分析法和独立成分分析法应用于时间序列降维的研究进展；本书的应用部分利用谱聚类方法和独立成分分析法对金融时间序列数据进行分析。首先，对欧洲主权债务危机下国际股指进行联动性分析。其次，对开放式基金投资风格识别进行研究。因此，本节第三部分和第四部分分别对股市联动性定量分析和基金投资风格识别方面的相关研究进行综述。在第五部分总结已有相关研究存在的主要问题。

1.2.1 谱聚类的研究进展

聚类分析是数据挖掘研究的一个重要领域。聚类是把对象的集合分组成为多个簇的过程，使同一个簇中的对象具有较高的相似度，而不同簇的对象差别较大。同一簇内相似度越大，不同簇间差别越大，表明所得到的聚类结果越好。聚类分析在社会科学、生物学、统计学、模式识别、信息检索、机器学习和数据挖掘等广泛的领域扮演着重要角色[12,13]。

根据数据集中每个元素在所属类中的隶属程度将聚类分为软聚类和硬聚类两大类。本书只考虑硬聚类的情形，即数据集中每个元素属于且

只属于一个类。根据聚类方法的算法模式和聚类结果的表示方式可将聚类方法粗略地分成层次聚类和划分聚类两大类。经典的层次聚类方法通过凝聚或分裂得到簇（Clusters）。层次聚类通常使用树状图（Dendrogram）展示聚类结果。划分聚类将数据集划分为 $k(k \leqslant n)$ 个类，其中 n 是数据集所包含数据点的个数。k- 均值是经典的划分聚类方法之一[11]。

k- 均值具有简单并且可以用来对多种类型的数据进行聚类等优点。但是 k- 均值不能处理密度不均匀或大小不均匀的簇，而且对非球形簇也失效[11]。谱聚类方法能够克服 k- 均值等经典方法的一些缺点[11,14]。

谱聚类方法建立在图论中经典的谱图理论基础上[15,16]，本质是将聚类问题转化为图的最优划分问题。基于图论的最优划分准则使得划分出的子图之间的连边的权重之和较小，而子图内连边的权重之和较大。图划分准则的质量对划分结果有直接影响，图划分准则越合理所得到的划分结果越好[17-19]。代表性的图划分准则有最小割集准则[20]、比率割集准则[21,22]、平均割集准则[23]、最小最大割集准则[24,25]和归一化割集准则[26-28]等。在诸多图划分准则中，从定义的合理性以及划分效果等角度来看归一化割集准则是一种相对较好的准则[29]。对于给定的划分准则和聚类数目 k，可以通过多次运行递归法将数据集划分成 k 个簇或者利用多路谱聚类法同时得到 k 个簇。而从算法执行难易角度看，多路谱聚类方法比递归谱聚类方法更容易实现[30]。因此，本书主要围绕多路归一化割谱聚类方法的一些理论、方法及应用展开讨论。

求解图划分问题的最优解是一个 NP 难问题。一种较好的解决方法是将原来图的最优划分问题转化成连续松弛形式，从而可将原问题转化为求解某些特殊矩阵的谱分解问题[14,26,31]。本书将这些用来对数据集进行划分的矩阵简称为谱聚类矩阵，将这类聚类方法统称为谱聚类方法。

常用的谱聚类矩阵有相似矩阵、Laplace 矩阵、规范相似矩阵、对称规范 Laplace 矩阵、转移概率矩阵和非对称规范 Laplace 矩阵[14]。Luxburg（2007）[14]从 Laplace 矩阵的特征值和特征向量的性质、连通子图的数目与 Laplace 矩阵的谱之间的关系、非对称规范 Laplace 矩阵的特征值和特征向量的性质以及连通子图的数目与非对称规范 Laplace 矩阵的谱之间的关系角度对几种谱聚类矩阵的性质进行了总结。Luxburg 还

对转移概率矩阵与非对称规范 Laplace 矩阵和规范相似矩阵与对称规范 Laplace 矩阵的特征值和特征向量之间的关系进行了总结。由 Luxburg 的总结可知，从谱聚类角度看，转移概率矩阵和规范相似矩阵分别与非对称规范 Laplace 矩阵和对称规范 Laplace 矩阵等价，区别在于所用到的特征值的大小不同。

Meilă 等（2001，2003）[27,28] 给出了 Lumpability 定理。Lumpability 定理在理论上揭示了利用转移概率矩阵或非对称规范 Laplace 矩阵对数据集进行划分的原理。

代表性的谱聚类方法有 Ng 等（2001）[30] 以及 Bach 和 Jordan（2004）[32] 提出的多路谱聚类方法、Meilă 等（2001，2003）[27,28] 提出的多路归一化割谱聚类方法和田铮等（2007）[32] 提出的基于权矩阵的无监督谱聚类算法。Ng 和 Bach 等提出的谱聚类方法选择规范相似矩阵为谱聚类矩阵[33]、Meilă 等选用非对称规范 Laplace 矩阵为谱聚类矩阵[27,28]、田铮等对相似矩阵进行理论分析的基础上提出了谱聚类方法[33]。

除以上典型谱聚类方法之外还有其他一些谱聚类方法。Zhang 和 Jordan（2008）[34] 从另外一个角度分析谱聚类方法并提出了一种新算法。贾建华（2011）[35] 研究了谱聚类及其集成理论和应用，主要讨论了空间约束谱聚类算法、基于成分数据的谱聚类集成、选择性谱聚类集成算法的理论及其在数据聚类和图像分割中的应用。Chi 等（2009）[36] 在 Chakrabarti 等（2009）[37] 的工作基础上提出了演化谱聚类（Evolutionary Spectral Clustering）方法。Jin 等（2006）[38] 提出“软分割”（“Soft Cut”）谱聚类方法。该方法能够实现软划分，是对归一化割谱聚类方法的一种推广。Meilă 和 Pentney（2007）[39] 以及 Zhou 等（2005）[40] 研究了对加权有向图（Weighted Cuts in Directed Graph）的谱聚类问题。Chen 等（2006）[41] 提出了一种 SRcut 谱聚类方法，SRcut 谱聚类方法可结合关于簇大小的先验知识对数据集进行划分。

由于其较好的划分效果和易于执行等优点，Ng 等提出的多路谱聚类方法和 Meilă 等提出的多路归一化割谱聚类方法被诸多学者采纳。然而以上两种谱聚类方法也存在许多问题有待进一步完善，例如：在谱聚类方法中如何选择和确定较合理的聚类数目？如何选择包含聚类信息的