

## 分形技术在案例库维护中的应用

倪志伟<sup>1,2</sup>,倪丽萍<sup>1,2</sup>,杨葛钟啸<sup>1,2</sup>

(1. 合肥工业大学 管理学院,合肥 230009; 2. 合肥工业大学 过程优化与智能决策教育部重点实验室,合肥 230009)  
(lilacnlp@126.com)

**摘要:**阐述了分形技术运用于案例库维护的可行性,并提出了一种基于分形技术的案例库维护模型,该模型利用分形技术中的盒维算法对案例库进行维护。实验表明该模型能够较为明显地降低案例库的规模,增强案例库数据全局分布的均匀性,进而提高案例库的检索效率。

**关键词:**分形;分形维数;案例库;案例库维护

**中图分类号:** TP18 **文献标志码:** A

## Application of fractal in case base maintenance

NI Zhi-wei<sup>1,2</sup>, NI Li-ping<sup>1,2</sup>, YANGGE Zhong-xiao<sup>1,2</sup>

(1. School of Management, Hefei University of Technology, Hefei Anhui 230009, China;  
2. Key Laboratory of Process Optimization and Intelligent Decision-making, Ministry of Education,  
Hefei University of Technology, Hefei Anhui 230009, China)

**Abstract:** The feasibility of fractal approach used in Case Base Maintenance (CBM) was elaborated, and then a model of case base maintenance based on fractal was proposed. The maintenance model utilized box dimension algorithm to maintain the case base. Experimental results show that this model can obviously reduce the size of case base, strengthen the uniformity of case base, and thus improve the speed of the case base retrieval process.

**Key words:** fractal; fractal dimension; case base; Case Base Maintenance (CBM)

### 0 引言

案例库维护(Case Base Maintenance, CBM)中最重要的问题是如何防止案例库在学习过程中的无限增大<sup>[1]</sup>,进而引起沼泽问题(Swamping Problem)<sup>[2]</sup>。近年来,对案例库维护的研究集中表现为以下两种思想<sup>[3]</sup>:第一是提高案例的检索效率,对案例库结构和索引机制的调整是目前实现这种思想的主要途径,但这种方法目的是降低每一次案例检索过程中所涉及的案例数,因而并不删除案例,从而无助于防止案例库无限增大;第二是约简案例库,包括案例的删除、案例属性约简等,采用的方法主要有分类、聚类、粗糙集和神经网络等。但是运用这些方法往往需要人为制定某些标准(如相似度阈值、删除策略等),且这些方法大都是基于欧氏距离度量机制的,这虽然在低维数据分析中获得了优良的性能和令人满意的结果,但当案例维数较高时,无法很好地发现案例之间的关系,因而可能降低学习的精度和效率,甚至会给进一步的数据分析工作带来误导,从而降低了案例推理(Case-based Reasoning, CBR)的效果。分形是自然界中普遍存在的现象,分形理论为解决现实生活中存在的非线性问题和看似混沌、复杂、无规律的问题提供了一种全新的方法。近年来,分形理论被逐渐应用于数据挖掘领域,如属性选择<sup>[4-5]</sup>、聚类<sup>[6]</sup>、分类<sup>[7]</sup>和关联规则挖掘等<sup>[8]</sup>。同时相关的研究表明<sup>[4,9-10]</sup>:现实中的大部分数据集都是具有分形特征的,运用分形方法进行数据挖掘,克服了传统的数据挖掘方法不能很好地处理高维数据以及分布孤立的数据点缺陷,在增量、任意形状聚类及聚类内部密度不均匀和聚类非邻接的情况中展现出了较大的

优越性,并成功地应用于数据集的建模及压缩中<sup>[11-12]</sup>。但是分形方法必须在数据集的无标度空间内才有效,而目前尚无算法对数据集的无标度空间进行界定。案例库可以看成是特殊意义的数据集,其中的数据由案例组成,这些案例是过去的一些经历,往往具有相似性,因而可以将分形技术运用于案例库中。

本文的主要贡献可以归纳为:根据案例库在统计意义上的标度空间的下界,提出了一种基于分形技术的案例库维护模型,该模型结合了分形技术中的盒维算法对案例库进行维护,并通过实验进行了验证。

### 1 问题定义

基于对约简案例库方法的分析,可以对该方法进行如下的描述:

设:案例库为 $S = \{P_1, P_2, \dots, P_i, \dots, P_N\}$ ,其中 $P_i$ 为第 $i$ 行案例 $i \in [1, N]$ (每行案例有 $M$ 个属性)。对 $S$ 的约简就是尽可能地减少 $S$ 中 $P_i$ 的数量,或者在保持 $S$ 原有特性的基础上约简 $M$ 的个数。

现有的有关约简案例库的研究工作大都是基于欧氏距离度量机制,这会导致那些极有价值但在空间分布孤立的案例被误删。

分形方法应用于案例库维护主要存在以下几个难点:

- 1) 分形方法如何与案例库维护技术相结合;
- 2) 如何寻找能够展现案例库具有分形特征的度量尺度;
- 3) 如何在找到的度量尺度基础上最大限度地案例库

所要表达知识的整体结构完整保留下来。

收稿日期:2008-12-12;修回日期:2009-02-20。

基金项目:国家863计划项目(2007AA04Z116);国家自然科学基金资助项目(70871033)。

作者简介:倪志伟(1963-),男,安徽桐城人,教授,博士生导师,主要研究方向:人工智能、机器学习;倪丽萍(1981-),女,安徽合肥人,博士研究生,主要研究方向:数据挖掘、机器学习、分形数据挖掘;杨葛钟啸(1984-),男,江苏淮安人,硕士研究生,主要研究方向:分形技术。

## 2 分形及其相关理论

分形理论的提出是为了研究自然界和非线性系统中的一些非规则的几何体,但随着它的发展,其中的一些基本思想和观点已经形成了一种方法论,用于指导人们解决实际问题中一些看似混沌、无规律和复杂的问题,以下是分形理论中的一些相关的定义。

**定义1** 理想的分形体  $F$  具有精细结构,即在任意小的比例尺度内包含整体; $F$  是不规则的,以至于不能用传统的几何语言来描述; $F$  通常具有某种自相似性,或许是近似的,或许是统计意义下的。

分形理论的创始人 Mandelbrot 认为分形集这类奇异集合的性质不能用欧氏测度来衡量,但分形维数恰是此类集合尺度变化下的不变量,因此主张用分形维数来刻画这类集合。

**定义2** 分形维数是对数据集分布自由程度的估计,它反映了数据在多维空间中的分布特性和对空间的填充能力<sup>[13]</sup>。

度量分形维数的方法有很多种,例如: Hausdorff 维数、盒维数、信息维数、关联维数和模糊维数等。盒维数由于计算简便因而应用范围较广,其基本思想是<sup>[14]</sup>:  $S$  是  $R^n$  上任意非空的有界子集,  $D$  为分形维数,  $N(r)$  是直径最大为  $r$ , 并且可以覆盖  $S$  集的最少个数,则  $S$  的下、上盒维数分别定义为:

$$\underline{D} = - \lim_{r \rightarrow 0} \frac{\log N(r)}{\log r} \quad (1)$$

$$\overline{D} = - \lim_{r \rightarrow 0} \frac{\log N(r)}{\log r} \quad (2)$$

如果这两个值相等,则称这共同的值为  $S$  的盒维数,记为

$$D = - \lim_{r \rightarrow 0} \frac{\log N(r)}{\log r} \quad (3)$$

显然为计算  $S$  的盒维数  $D$ , 可以构造一些边长为  $r$  的  $n$  维“盒子”, 然后计算不同  $r$  值的“盒子”和  $S$  相交的个数  $N(r)$ , 这个维数是当  $r \rightarrow 0$  时,  $N(r)$  增加的对数斜率。

对于分形事物而言, 它一定具有标度不变性。

**定义3** 所谓标度不变性, 是指在分形上任选一局部区域, 对它进行放大或缩小, 这时得到的放大或缩小图又会显示出原图的形态特性<sup>[13]</sup>。

**定义4** 通常把标度不变性适用的空间称为该分形体的无标度空间。

分形事物只有在其无标度空间内才具有分形特征, 超过这个空间之外就不为分形了<sup>[15]</sup>。

## 3 分形技术与案例库维护方法的内在联系

定义2 说明作为区别于欧氏测度的一种方法, 分形维数  $D$  能够通过其所反映的数据空间的填充能力, 很好地展现数据集中某些隐含的规律, 并将其作为整体呈现出来, 而不是将其硬性地割裂为几块, 这是用欧氏空间无法体现的。因此  $D$  描述了数据集中数据点的分布情况, 表现出整个数据集的结构特征。案例库约简的目的在于约简后的案例库能够保持原有案例库的特征, 因而保证维护后的案例库与原始案例库具有相同的分形维数, 即以  $D$  作为衡量案例库分形特征的尺度, 删除  $S$  中的冗余  $P_i$ , 将会保持案例库中数据的分布特征, 大大减少重要但孤立的  $P_i$  被误删的情况, 从而使整个  $S$  中的  $P_i$  具有分布的均匀性。

一个案例库的分形维数  $D$  可以采用盒维数进行度量。式

(3) 为该维数的计算公式。由定义1 可知在分形定义中,  $F$  是具有任意小比例尺度的细节即精细结构, 也就是说从理论上讲真正的自相似分形是无限的,  $F$  的标度空间下界趋近于零, 因而式(3) 中的  $r$  趋近于零。但现实中  $S$  中案例的个数是有限的, 不具有精细结构, 其相似性体现在统计意义上, 即只有在标度空间内案例库  $S$  才具有分形特征, 在运用式(3) 时  $r \in [r_{\min}, r_{\max}]$ <sup>[16]</sup>, 其中  $r_{\min}$  为案例库无标度空间的下界。

从式(3) 中还可以看出, 盒维数  $D$  仅与  $N(r)$  有关, 而与  $N(r)$  中所包含的数据点个数无关。因而为了保证维护后的案例库在标度空间中仍然保持原有的分形维数  $D$ , 则可以在  $r = r_{\min}$  时, 删除此半径下  $N(r)$  个盒子内的多余点 (即每个非空盒子内只保留一个点), 那么当  $r_{\min} < r \leq r_{\max}$  时不会影响  $N(r)$  的数目 (小半径下在一个盒子中的点, 在大半径下仍会在一个盒子中)。这样删除这些多余的点不仅可以减少案例库中的案例条数, 而且还不会影响整个案例库  $S$  的分形维数。由于在删除了  $S$  在  $r_{\min}$  下对应的  $N(r_{\min})$  个盒子中的多余数据点后每个盒子内都具有相同的点数, 因而这样的维护方式保证了案例库中数据全局分布的均匀性。图1、2 表明了这一思想: 图1 的科赫曲线用连续的实线展现了该曲线的迭代规律, 假设用式(3) 计算图1 的  $D$ , 那么可以想象图1 中科赫曲线的实线部分将布满正方形的小盒子, 显然每个小盒子里将是一段连续的实线, 而连续的实线又可以看作是无穷多点的集合, 如果用一个点来取代每个小盒子内的线段, 其效果便如图2 所示, 虽然图1 的连续实线变成了点集, 但这并没有影响到图2 对科赫曲线迭代规律的展现。因此可以看出去除  $N(r)$  个盒子内多余的点数不仅去除了  $S$  中的冗余数据, 而且保持了  $S$  原有的结构特征, 同时数据的分布在结构上也非常均匀。



图1 科赫曲线(连续的实线)



图2 科赫曲线(点集)

通过上述分析可知, 利用分形维数进行案例库维护关键是找到案例库的无标度空间下界  $r_{\min}$ 。式(3) 说明具有分形特征的事物满足幂律分布即  $N(r) = r^{-D}$ , 因而在实际计算盒维时, 可以通过设置多个  $r$ , 得到每一个  $r$  对应的  $N(r)$ , 绘制  $\log r - \log N(r)$  曲线图, 再用最小二乘法拟合直线  $\langle \log r, \log N(r) \rangle$  求出斜率的相反数作为分形维数  $D$ 。由于案例库仅在一定标度空间内体现出分形特征因而只有在  $r \in [r_{\min}, r_{\max}]$  内, 案例库才满足分形特征, 在该区间范围内的点连接起来应形成一条直线, 直线斜率的相反数为分形维数, 而在该范围之外的点则应该排除。根据该规则可以确定案例库的无标度空间下界  $r_{\min}$ 。具体方法为设定盒子的半径为  $r (r = 2^i)$ , 计算出不同  $r_i$  下的  $N(r_i)$ , 从而形成一系列点  $\langle \log r_i, \log N(r_i) \rangle$ , 将这些点进行分组拟合, 选取拟合误差最小 (拟合的直线基本上通过所选取的点) 的直线, 将这段直线斜率的相反数作为  $D$ <sup>[18]</sup>。然后将这段直线所经过点中最小的半径  $r$  作为  $r_{\min}$ 。

## 4 基于分形技术的案例库维护模型的建模流程

设  $P_i$  的盒子编号为  $Num(i) = \{num(i, 1), num(i, 2)\}$ ,

$\dots, num(i, j), \dots, num(i, M)\}$  其中  $num(i, j)$  为每个属性的盒子编号。

第1步 设盒子半径为  $r_i = 2^i$ , 其中  $i = -20, -19, \dots, 19, 20$ ; 对于每一个半径  $r_i$  执行第2和第3步。

第2步 计算  $P_{ij}/r = num(i, j)$  其中  $i \in [1, N], j \in [1, M]$ , 得到每个  $P_i$  的  $Num(i)$ 。

第3步 遍历  $Num(i)$  得到  $N(r_i)$ 。

第4步 将所得的  $\langle \log r_i, \log N(r_i) \rangle$  点进行分组, 并分别拟合成直线, 直线斜率的相反数作为分形维数  $D$  的备选值, 其中分组方法可以描述如下:

首先对 41 个点进行有效性处理, 处理后设得到的点系列为  $\langle \log r_m, \log N(r_m) \rangle, \dots, \langle \log r_n, \log N(r_n) \rangle$ , 共有  $n - m + 1$  个点。令  $deltastep = \lfloor \frac{n - m + 1}{2} \rfloor$ , 将  $n - m + 1$  个点进行分组, 分为  $K$  组, 分别为:

$$\begin{aligned} l_1 &= (\langle \log r_m, \log N(r_m) \rangle, \dots, \\ &\quad \langle \log r_{m+deltastep}, \log N(r_{m+deltastep}) \rangle) \\ l_2 &= (\langle \log r_{m+1}, \log N(r_{m+1}) \rangle, \dots, \\ &\quad \langle \log r_{m+1+deltastep}, \log N(r_{m+1+deltastep}) \rangle) \\ &\dots \\ l_k &= (\langle \log r_{m+k-1}, \log N(r_{m+k-1}) \rangle, \dots, \\ &\quad \langle \log r_{m+k-1+deltastep}, \log N(r_{m+k-1+deltastep}) \rangle) \end{aligned}$$

其中  $m + k - 1 + deltastep = n$ 。对这  $K$  组点分别进行拟合得到  $D$  的备选值  $D_x = \{D_1, \dots, D_k\}$ ;

第5步 从  $K$  组  $D$  的备选值  $D_x = \{D_1, \dots, D_k\}$  中选出一个拟合误差最小的  $D_x$  作为  $S$  的  $D$ ;

第6步 将得到  $D_x$  的直线经过的点设为:

$$(\langle \log r_{m+x-1}, \log N(r_{m+x-1}) \rangle, \langle \log r_{m+x}, \log N(r_{m+x}) \rangle, \dots, \langle \log r_{m+x-1+deltastep}, \log N(r_{m+x-1+deltastep}) \rangle)$$

其中  $r_{m+x-1}$  为最小半径。

第7步 删除  $N(r_{m+x-1})$  个盒子内多余的点, 并得到新的案例库  $S_0$ 。

上述算法中第2步用于计算每条案例每一维的位置, 从而得到每条案例落入的盒子号, 进而统计出非空盒子数。第4步和第5步是为了获得一个在统计意义上最优的分形维数  $D$  作为该案例库的分形维数, 并通过分形维数的计算得到案例库无标度空间的下界。

## 5 实验分析

为对基于分形技术的案例库维护模型进行验证和评价, 本文用 UCI 数据集(表1)来模拟案例库, 并用提出的基于分形技术的案例库维护模型对表1中案例库进行维护。

根据本文所提出的案例库维护方法对表1中的4个案例库进行冗余数据的排除。由于这4个实验案例库都含有类别属性列, 这些类别属性列是案例库的决策属性, 因而在对这4个案例库进行案例库约简时, 忽略样本的类别属性。其中案例库1中有16条案例缺失属性, 将它们忽略, 故案例库1共有683条案例。

表2和图3说明了 bupa 案例库在求解分形维数时所得到的点对  $\langle \log r_i, \log N(r_i) \rangle$  (其中当  $r \leq 2^1$  时,  $\log N(r)$  为 8.413628,

当  $r \geq 2^9$  时,  $\log N(r)$  为 0, 因而在图3和表2中忽略这两段无效部分)。从图3中可以看出, 在不同盒子半径下所得的点并不都在一条直线上。将点对进行分组并分别拟合最后得出当  $r_{\min} = 2^4$  时所拟合出的直线误差最小, 此时的分形维数为 1.68396。在此半径下删除每个非空盒子中多余的点, 最终得到的案例库规模为 159 条。

表1 测试案例库

数据集	样本数	类别数	属性数
breast-cancer	699	2	9
bupa	345	2	6
segment	2310	7	19
sat	4435	7	36

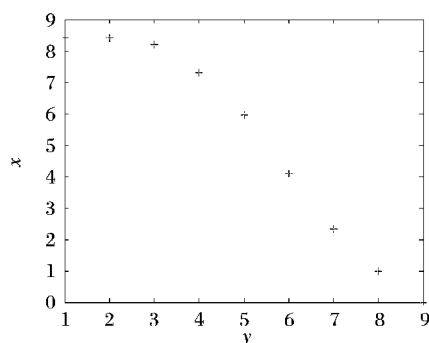


图3 bupa 在不同盒子半径下得到的非空盒子数

表2 bupa 案例库为例的维护结果

$r$	$\log N(r)$	$r$	$\log N(r)$	$r$	$\log N(r)$
$2^1$	8.413628	$2^4$	7.312883	$2^7$	2.321900
$2^2$	8.405141	$2^5$	5.954196	$2^8$	1.000000
$2^3$	8.204571	$2^6$	4.087463	$2^9$	0.000000

为了验证案例库的效果, 本文采用 SVM<sup>[17]</sup> 和 K-fold 交叉验证相结合的方法对 4 个案例库维护前后的准确率进行测试, SVM 中核函数选择的是 RBF。测试结果如表3所示。其中案例库约简率是案例库约简的样本数与维护前的样本数的比值。从表3中可以看出, 使用分形方法对案例库进行维护后案例库的规模显著减少, 并且维护后的分类准确率与维护前的分类准确率相当。

## 6 结语

本文论述了分形技术运用于案例库维护的可行性, 提出了一种基于分形技术的案例库维护模型, 该模型的建模思想如下: 首先通过拟合的方式找到案例库的分形维数在统计意义上的近似值  $D$ , 再由  $D$  确定案例库标度空间的下界  $r_{\min}$  及其所对应的  $N(r_{\min})$ , 最后去除  $N(r_{\min})$  个盒子中的冗余数据。实验结果表明该方法能大大降低案例库的存储空间, 使整个案例库中的案例分布具有分形特征下的均匀性, 同时维护后的

表3 案例库维护效果

数据集	$r_{\min}$	维护后的 样本数	案例库 约简率/%	维护前最优 分类准确率/%	维护后 分类最优准确率/%
breast-cancer	$2^1$	326	52.27	97.0674	94.4615
bupa	$2^4$	159	53.91	73.8372	74.2138
segment	$2^4$	1321	42.81	97.4026	96.2150
sat	$2^5$	3044	31.36	92.2418	92.1788

出最大可能出现的下一个搜索语句集,用户只需要在这些语句后键入空格就可以继续显示下一个可能的搜索语句集,以帮助用户快速匹配自己的搜索目标并减少用户输入工作量。这种模式对供应链成员统计服装流行元素的构成非常快捷。图 6 为搜索结果的界面图,图的上部分为基于 Ajax 模式的用户输入框,图的下部分为选择“夏奈尔 连衣裙 粉色”后所精确搜索到的商品。

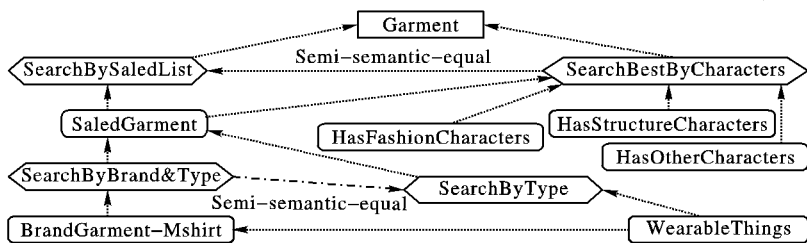


图 5 服装本体的 SWS 模型

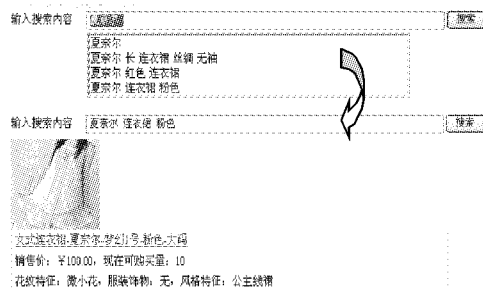


图 6 服装搜索引擎搜索结果的界面

与目前实际使用的搜索引擎相比,本文所介绍的引擎的搜索结果更加精确和高效,呈现出良好的性能特征:1) 基于语义树的搜索模式通过逐步缩小搜索域,从逻辑层次上实施多关键词搜索,其算法效率要高于全文搜索模式;2) 支持合作伙伴之间因业务需要而产生的搜索,可以灵活地根据协作

策略规定搜索结果的排列顺序;3) 原则上,本服装搜索引擎支持语义明确的任意长语句输入搜索形式。

## 5 结语

服装搜索引擎与目前通用的搜索引擎相比,由于使用了明确的语义定义 XML 标签,能让搜索程序充分理解概念的意义以及概念之间的关系,在分布式环境下,表现出良好的数据理解适应性。对于消费者而言,搜索引擎能根据消费者的个性化穿着需求,在整个行业内进行搜索,以最大限度地满足需求;对于供应链成员,需要在语义上理解消费者的需求特点和变化,协同操作,快速响应市场需求。这种分布式搜索引擎系统为电子商务背景下的供应链协同运作提供新的工具。由于目前在不同平台的数据表达统一预处理方面采用的仍然是穷举推理方法,灵活性受到限制,进一步的研究方向是智能处理方法在服装搜索引擎中的应用。

## 参考文献:

- [1] SHU B, KAK S. A neural network-based intelligent metasearch engine [J]. Information Sciences, 1999, 120(1/4): 1-11.
- [2] CNET 科技资讯网. Google 面临新技术挑战, 语义搜索潜力大 [EB/OL]. [2008-06-25]. <http://www.cnetnews.com.cn/2008/0618/933251.shtml>.
- [3] LEE T B, HENDLER J, LASSILA O. The semantic Web [J]. Scientific American, 2001(5): 34-43.
- [4] STUDER R, BENJAMINS V R, FENSEL D. Knowledge engineering, principles and methods [J]. Data and Knowledge Engineering, 1998, 25(1/2): 161-1971.
- [5] 顾宁, 刘家茂, 柴晓路. Web Services 原理与研发实践[M]. 北京: 机械工业出版社, 2007: 1-11.
- [6] 宋庭新, 黄必清, 熊健民, 等. 语义 Web 服务在业务协同与供应链集成中的应用[J]. 中国机械工程, 2008, 19(4): 410-413.

(上接第 1600 页)

案例库与维护前的案例库在分类准确率上基本相当。

## 参考文献:

- [1] SMYTH B, KEANE M T. Remembering to forget: A competence-preserving case deletion policy for case-based reasoning systems [EB/OL]. [2008-10-10]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.17.2767>.
- [2] FRANCIS A G, RAM A. The utility problem in case-based reasoning [EB/OL]. [2008-10-10]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.50.639>.
- [3] 耿焕同, 肖明军, 邹翔, 等. 聚类算法在范例库维护中的应用研究 [J]. 计算机工程, 2005, 31(12): 166-168.
- [4] 鲍玉斌, 王琢, 孙焕良, 等. 一种基于分形维的快速属性选择算法 [J]. 东北大学学报: 自然科学版, 2003, 24(6): 527-530.
- [5] LEE H D, MONARD M C, WU FENG-CHUNG. A fractal dimension based filter algorithm to select features for supervised learning [EB/OL]. [2008-10-10]. [http://www.icmc.usp.br/~iam2006/sbia/apresentacoes/IBERAMIASBIA\\_TS1/IBERAMIASBIA\\_TS1\\_A2.pdf](http://www.icmc.usp.br/~iam2006/sbia/apresentacoes/IBERAMIASBIA_TS1/IBERAMIASBIA_TS1_A2.pdf).
- [6] BARBARA D, CHEN P. Using the fractal dimension to cluster datasets [C]// ACM-SIGKDD: Proceedings of the 2000 International Conference in Knowledge Discovery and Data Mining. Menlo Park, California: AAAI Press, 2000: 260-264.
- [7] HIPPENSTIEL R, EL-KISHKY H, RADEV P. On time-series analysis and signal classification - part I: Fractal dimensions [C]// Conference Record of the 38th Asilomar Conference on Signals, Systems and Computers. Washington, DC: IEEE Press, 2004, 2:

2121-2125.

- [8] BARBARA D, NAZERI Z. Fractal mining of association rules over interval data [R]. Fairfax: George Mason University, 2000.
- [9] 闫光辉, 李战怀, 党建武. 基于多重分形的聚类层次优化算法 [J]. 软件学报, 2008, 19(6): 1283-1300.
- [10] BARBARA D. Chaotic mining: Knowledge discovery using the fractal dimension [EB/OL]. [2008-10-10]. [http://www.isse.gmu.edu/techrep/1999/99\\_03\\_barbara.ps](http://www.isse.gmu.edu/techrep/1999/99_03_barbara.ps).
- [11] BORGAT P, FLANDRIN P, AMBLARD P O. Stochastic discrete scale invariance [J]. IEEE Signal Processing Letters, 2002, 9(6): 181-184.
- [12] MAZEL D S, HAYES M H. Using iterated function systems to model discrete sequences [J]. IEEE Transactions on Signal Processing, 1992, 40(7): 1724-1734.
- [13] 张济忠. 分形[M]. 北京: 清华大学出版社, 2001.
- [14] 姜灵敏, 周峰. 上证指数盒维数的计量与特性研究[J]. 系统工程学报, 2006, 21(4): 434-437.
- [15] 倪志伟, 李锋刚, 毛雪岷. 智能管理技术与方法[M]. 北京: 科学出版社, 2007.
- [16] 倪丽萍, 倪志伟, 吴昊, 等. 基于分形维数的数据挖掘技术研究综述[J]. 计算机科学, 2008, 35(1): 187-189.
- [17] CHANG C C, LIN C J. LIBSVM - A library for support vector machines [EB/OL]. [2008-10-10]. <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [18] Christos Faloutsos - Released software [EB/OL]. [2008-10-10]. <http://www.cs.cmu.edu/~christos/software.html>.