

# 基于粗糙度的一种分形维数计算方法

张学友<sup>1</sup>, 苗强<sup>2</sup>, 毛军军<sup>1,2</sup>

(1. 安徽大学 数学科学学院, 安徽 合肥 230039;

2. 安徽大学 智能计算与信号处理教育部重点实验室, 安徽 合肥 230039)

**摘要:**在股票市场中, 计算分形维数有助于投资者把握股市的相对复杂性, 分析股市的动向或趋势, 利用一种新的属性选择分类标准——集合分类粗糙度, 并结合粗糙集中决策表的数据离散化方法, 针对计算股票市场中的股票价格的分形维数问题, 选用香港恒生指数, 把开盘价、最高价、最低价、收盘价视为条件属性, 交易量视为决策属性, 在不影响分类结果的前提下, 得出: 不同的证券市场数据, 应选择不同的属性指标计算分形维数。该方法是可行的、有效的, 从而为分形维数的计算, 选择合适的属性提供了一条新的途径。

**关键词:**证券市场; 分形维数; 粗糙集; 集合分类粗糙度

**中图分类号:** TP18

**文献标识码:** A

**文章编号:** 1673-629X(2010)05-0136-03

## A Calculation Method of Fractal Dimension Based on Roughness

ZHANG Xue-you<sup>1</sup>, MIAO Qiang<sup>2</sup>, MAO Jun-jun<sup>1,2</sup>

(1. School of Mathematics and Computation Science, Anhui University, Hefei 230039, China;

2. Ministry of Education Key Laboratory of Intelligent Computing & Signal Processing,  
Anhui University, Hefei 230039, China)

**Abstract:** The fractal dimension can help investors know the relative complexity of the stock market and analyse its movements or trends in the stock market. In this paper, a new attribute selection classification criteria——sets classify roughness was used, which combined with data discretization of decision table to compute the fractal dimension of the stock price. It applies to the raw data of the stock price of Heng Sheng Index. The opening price, highest price, lowest price and the closing price are regarded as condition attribute and the volume is regarded as decision attribute. The result is that using the opening price to compute the fractal dimension is reasonable. The method is feasible and effective so that pave a new way for computing fractal dimension and choosing appropriate attribute.

**Key words:** stock market; fractal dimension; rough set; sets classify roughness

## 0 引言

证券市场中股票波动是一个有多种因素组成的复杂系统, 传统的投资金融学认为, 股票的价格遵循“有效市场假说”, 这意味着现在的价格与过去和将来的价格没有关系<sup>[1]</sup>, 这明显与事实不符, 股票价格有很强的相关性, 且观测到分形的特征<sup>[2]</sup>, 因此计算股票价格的分形维数, 来对资本市场特性进行研究, 维数越大, 代表其股票波动的复杂程度大, 所以风险大, 反之, 风险小<sup>[3]</sup>。

股票价格数据来源中, 给出了不同时间标度下的

开盘价、最高价、最低价、收盘价和交易量, 近年来, 大多学者在计算股票市场的分形维数时, 一般都对收盘价进行分形维数的各种测量<sup>[4]</sup>, 以此来表示股票市场的复杂程度。

但事实上, 开盘价、最高价、最低价, 尤其是交易量都是股票市场的重要属性, 不同的证券市场数据, 应选择不同的属性指标计算分形维数, 才能更合理地反映市场的复杂度以及不规则程度。

文中利用一种新的属性选择分类标准——集合分类粗糙度, 并结合粗糙集中决策表的数据离散化方法, 针对计算股票市场中的股票价格的分形维数问题, 选用香港恒生指数为实证数据, 把开盘价、最高价、最低价、收盘价视为条件属性, 交易量视为决策属性, 分别计算它们的集合分类粗糙度, 值小的分类能力越强, 其对应的属性, 计算它的分形维数来表示股票市场的复杂程度将更合理。

收稿日期: 2009-09-11; 修回日期: 2009-12-25

基金项目: 国家自然科学基金(60675031); 安徽省高校省级自然科学基金(2008B093); 安徽大学人才队伍(2006)建设经费资助项目

作者简介: 张学友(1982-), 男, 安徽舒城人, 硕士研究生, 研究方向为运筹学; 毛军军, 副教授, 博士, 研究方向为人工智能、信息优化。

## 1 分形维数的定义

设  $F$  是  $R^n$  上任意非空的有界子集,  $N_\delta(F)$  是直径最大为  $\delta$ , 可以覆盖  $F$  集的最少个数, 则  $F$  的下、上分形维数分别定义为

$$\underline{\text{Dim}}_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \quad (1)$$

$$\overline{\text{Dim}}_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta} \quad (2)$$

如果这两个值相等, 则称这共同的值为  $F$  的分形

维数, 记为  $\text{Dim}_B F = \lim_{\delta \rightarrow 0} \frac{\log N_\delta(F)}{-\log \delta}$

## 2 粗糙集基本理论与方法

### 2.1 基本概念

定义 1<sup>[5]</sup> 一个信息系统  $S$  可表示为  $S = (U, A, V, f)$ , 其中:  $U$  为对象的集合, 即论域;  $A$  为属性的集合;  $V = \bigcup V_a, a \in A, V_a$  为属性  $a$  的值域, 为信息函数;  $f = U \times A \rightarrow V$ 。

如果属性  $A$  可分为条件属性集  $C$  和决策属性集  $D$ , 即  $C \cup D = A, C \cap D = \emptyset$ , 则该信息系统被称为决策系统或决策表, 表示为  $S(U, C \cup D)$ 。

定义 2<sup>[5]</sup> 在信息系统  $S$  中, 对任何一个属性子集  $B \subseteq A$ , 可以定义一个不可分辨关系  $\text{IND}(B) = \{X \mid X \subseteq U \wedge f(x, b) = f(y, b), x \in X, y \in X, b \in B\}$ 。用不可分辨关系  $\text{IND}(B)$  可将论域  $U$  划分为若干等价类, 记为  $U/\text{IND}(B)$ , 简记为  $U/B$ 。

定义 3<sup>[5]</sup> 在信息系统  $S$  中, 对任一对象子集  $X \subseteq U$  和属性子集  $B \subseteq A, B$  的下近似集定义为  $B_-(X) = \{Y_i \mid Y_i \in U/B \wedge Y_i \subseteq X\}$ 。

### 2.2 集合分类粗糙度 (Sets Classify Roughness)

假设  $a$  中有  $n$  个等价类,  $D$  中有  $m$  个等价类, 那么属性  $a$  对决策属性  $D$  的分类能力则为  $a$  中  $n$  个等价类分别对  $D$  中  $m$  个等价类的知识表示能力之和。

定义 4 在 1 个集合  $A$  中, 存在一个划分  $A_1, A_2, \dots, A_i$ , 则称  $i$  为集合  $A$  的等价类个数, 记为  $\text{ECN}(A)$ 。

定义 5 对 1 集合  $A$ , 其元素个数记为  $\text{card}(A)$ 。

定义 6 在信息系统  $S$  中, 对任一对象子集  $X \subseteq U$  和属性子集  $B \subseteq A$ , 属性子集  $B$  对子集  $X$  的集合分类粗糙度定义为:

$$\text{SCR}(B, X) = 1 -$$

$$\left( \sum_{j=1}^m \frac{\text{ECN}(B_-(X_j))}{\text{ECN}(U/B)} + \gamma \sum_{j=1}^m \frac{\text{card}(B_-(X_j))}{\text{card}(X_j)} \right)$$

其中:  $\frac{\text{ECN}(B_-(X_j))}{\text{ECN}(U/B)}$  表示  $B$  属性中等价类将  $X$  集合进行分类的能力, 并且这个因素占主导地位;

$\frac{\text{card}(B_-(X_j))}{\text{card}(X_j)}$  表示  $B$  的上近似集在  $X_j$  中的比重, 占

次要地位, 它是对  $\frac{\text{ECN}(B_-(X_j))}{\text{ECN}(U/B)}$  的补充, 主要用于当

2 个属性的  $\frac{\text{ECN}(B_-(X_j))}{\text{ECN}(U/B)}$  一样时, 微调集合分类精度作为微调数只能占微小比重, 所以用  $\gamma$  表示其所占  $\text{SCR}(B, X)$  比例, 取  $\gamma = 0.1$ ;  $j$  表示决策属性的第  $j$  个等价类,  $j = 1, \dots, m$ ; 表示决策属性第  $j$  个等价类集合<sup>[6,7]</sup>。

当某个属性的  $\text{SCR}(B, X)$  越小时, 说明它对决策属性的分类更加确定<sup>[8]</sup>。

### 2.3 粗糙集属性离散化方法

属性  $a$  的值域  $V_a$  上的一个断点可以记为  $(a, c)$ , 其中  $a \in R, c$  为实数集, 在值域  $V_a = [l_a, r_a]$  上的任意一个断点集合  $\{(a, c_1^a), (a, c_2^a) \dots (a, c_{k_a}^a)\}$  定义了  $V_a$  上的一个分类  $P_a$ 。

$$P_a = \{[c_0^a, c_1^a], [c_1^a, c_2^a], \dots, [c_{k_a}^a, c_{k_a+1}^a]\}$$

$$l_a = c_0^a < c_1^a < c_2^a < \dots < c_{k_a}^a < c_{k_a+1}^a = r_a$$

$$v_a = [(c_0^a, c_1^a) \cup (c_1^a, c_2^a) \cup \dots \cup (c_{k_a}^a, c_{k_a+1}^a)]$$

分别取  $(c_0^a, c_1^a) = 1, (c_2^a, c_3^a) = 2, \dots, (c_{k_a}^a, c_{k_a+1}^a) = k_{a+1}^a$ 。

## 3 算法实例

选用恒生指数中 2008 年第三季度的 (时间标度为日)。数据来源于 <http://finance.cn.yahoo.com>, 如表 1 所示。

表 1 恒生指数 2008 年度第三季度日价格

日期	开盘价	最高价	最低价	收盘价	交易量
2008-7-2	21785.39	21938.2	21555.53	21704.45	2548582600
2008-7-3	21389.49	21742.07	21163.57	21242.78	2725284800
2008-7-4	21402.17	21534.05	21344.85	21423.82	1796344200
2008-7-7	21402.7	21916.21	21402.7	21913.06	1964579200
2008-7-8	21632.7	21684.21	21098.84	21220.81	2015196800
2008-7-9	21740.86	21954.17	21531.97	21805.81	2691241400
2008-7-10	21562.07	22020.66	21498.87	21821.78	2512108400
2008-7-11	21834.64	22225.38	21761.04	22184.55	2163732400
2008-7-14	22205	22360.29	21871.6	22014.46	1494919500
2008-7-15	21644.04	21644.04	21077.24	21174.77	2140148000
2008-7-16	20988.74	21334.38	20988.74	21223.5	1784065200
2008-7-17	21825.24	21892.53	21672.27	21734.72	2289947600
2008-7-18	22010.94	22010.94	21677.15	21874.19	1842076400
2008-7-21	22523.28	22645.52	22455.01	22532.9	2571576800
2008-7-22	22430.59	22690.74	22393.14	22527.48	1505886500
2008-7-23	22900.77	23134.55	22871.04	23134.55	2586745600
2008-7-24	23330.89	23369.05	23062.62	23087.72	2589708600
2008-7-25	22751.51	22843.2	22542.08	22740.71	1917652600
2008-7-28	22801.85	22862.03	22619.23	22687.21	1081764400
2008-7-29	22265.93	22265.93	22089.09	22258	1403232700
2008-7-30	22637.33	22751.04	22573.18	22690.6	1968748000
2008-7-31	22878.76	22878.76	22695.73	22731.1	1572662100
2008-8-1	22497.9	22881.27	22207.31	22862.6	1800206000
2008-8-4	22630.59	22713.58	22425.13	22514.92	1293079100

(续表 1)

日期	开盘价	最高价	最低价	收盘价	交易量
2008-8-5	22225.06	22225.06	21739.22	21949.75	1946666000
2008-8-7	22403.26	22424.54	21915.28	22104.2	2172200800
2008-8-8	21997.64	22230.55	21690.6	21885.21	1737934600
2008-8-11	22020.54	22235.51	21859.34	21859.34	1549134100
2008-8-12	21992.18	22309.33	21640.89	21640.89	1868403600
2008-8-13	21270.86	21665.75	21223.38	21293.32	3083083600
2008-8-14	21302.7	21453.48	21109.01	21392.71	2250392800
2008-8-15	21383.71	21383.71	20994.54	21160.58	1450461200
2008-8-18	21163.01	21206.6	20751.14	20930.67	1318434000
2008-8-19	20675.75	20902.51	20484.37	20484.37	1387873500
2008-8-20	20388.79	20971.19	20388.79	20931.26	1838994400
2008-8-21	20762.65	20762.65	20350.48	20392.06	1711519200
2008-8-22	20392.06	20392.06	20392.06	20392.06	0
2008-8-25	20739.48	21108.25	20739.48	21104.79	1915426400
2008-8-26	20849.08	21173.56	20785.8	21056.66	1382900700
2008-8-27	21104.56	21464.72	21104.56	21464.72	1945726000
2008-8-28	21546.94	21546.94	20857.03	20972.29	2187599200
2008-8-29	21289.7	21474.31	21223.99	21261.89	2047897200
2008-9-1	20999.32	21031.08	20844.15	20906.31	1213874700
2008-9-2	20956.94	21066.58	20595.59	21042.46	1618410500
2008-9-3	20964.75	20964.75	20526.73	20585.06	1666102200
2008-9-4	20545.38	20621.48	20356.51	20389.48	1561120600
2008-9-5	19833.87	19987.15	19708.39	19933.28	2582614600
2008-9-8	20840.69	20840.69	20637.99	20794.27	3035311000
2008-9-9	20439.47	20543.15	20299.97	20491.11	1942634000
2008-9-10	20114.86	20283.99	19951.36	19999.78	2697584000
2008-9-11	19854.82	19854.82	19220.28	19388.72	2913054800
2008-9-12	19432.72	19525.55	19157.73	19352.9	2445765000
2008-9-16	18325.65	18538.51	18019.2	18300.61	5213601200
2008-9-17	18691.3	18699.18	17637.19	17637.19	4390107200
2008-9-18	17120.23	17849.97	16283.72	17632.46	6237659200
2008-9-19	18878.26	19327.73	18588.11	19327.73	6690248800
2008-9-22	19869.02	19869.02	19137.67	19632.2	3731644400
2008-9-23	19178.49	19302.76	18872.85	18872.85	2546635600
2008-9-24	18954.32	19291.02	18862.9	18961.99	2084468000
2008-9-25	19003.22	19248.72	18870.11	18934.43	1682466400
2008-9-26	18909.59	18936.94	18500.11	18682.09	1895439600
2008-9-29	18742.25	18742.25	17796.34	17880.68	2578483600
2008-9-30	16898.33	18029.77	16799.29	18016.21	3226857200

对于开盘价至收盘价,视为条件属性,它们的最低值为 16283.72,最高值为 23369.05,取  $(16000.00, 17000.00) = 1, (17001.00, 18000.00) = 2, \dots, (23001.00, 24000.00) = 8$ 。对于交易量,视为决策属性,它的最低值为 1081764400,最高值为 6690248800,取  $(1000000000, 1700000000) = 1, (1700000000, 2400000000) = 2, \dots, (5900000000, 6700000000) = 8$ , 条件属性分别用  $a_1, a_2, a_3, a_4$  来表示,决策变量用  $D$  来表示,日期采用样本编号。

对于原始数据表中空值的,采用平均值方法进行决策表补齐,样本容量  $n = 63$ ,数据处理后的决策表如表 2 所示。

决策属性  $D$  分为 8 个等价类:  $\{9, 15, 19, 20, 22, 24, 28, 32, 33, 34, 39, 43, 44, 45, 46, 60\} \dots \{55, 56\}$ , 条件属性  $a_1$  的 8 个等价类  $\{63\}, \{55\}, \{53, 54, 56, 59, 61,$

$62\} \dots \{9, 13, 14, 15, 16, 18, 19, 20, 21, 22, 23, 24, 25, 26, 28\}, \{17\}, a_2, a_3, a_4$  的等价类略。计算可得  $SCR(a_1, D) = 0.5183, SCR(a_2, D) = 0.6867, SCR(a_3, D) = 0.8683, SCR(a_4, D) = 0.8617$ 。明显看到条件属性  $a_1$  的集合分类粗糙度最小,所以它的分类能力更确定,所以在计算本股票市场时,选择开盘价计算该市场的分形维数,值为 1.3183,这说明市场的不规则度和复杂度较高。

表 2 实例信息表

$n$	$a_1$	$a_2$	$a_3$	$a_4$	$D$	$n$	$a_1$	$a_2$	$a_3$	$a_4$	$D$	$n$	$a_1$	$a_2$	$a_3$	$a_4$	$D$
1	6	6	6	6	3	22	7	7	7	7	1	43	5	6	5	5	1
2	6	6	6	6	3	23	7	7	7	7	2	44	5	6	5	6	1
3	6	6	6	6	2	24	7	7	7	7	1	45	5	5	5	5	1
4	6	6	6	6	2	25	7	7	6	6	2	46	5	5	5	5	1
5	6	6	6	6	2	26	7	7	6	7	2	47	4	4	4	4	3
6	6	6	6	6	3	27	6	7	6	6	2	48	5	5	5	5	3
7	6	7	6	6	3	28	7	7	6	6	1	49	5	5	5	5	2
8	6	7	6	6	2	29	6	7	6	6	2	50	5	5	4	4	3
9	7	7	6	7	1	30	6	6	6	6	3	51	4	4	4	4	3
10	6	6	6	6	2	31	6	6	6	6	2	52	4	4	4	4	3
11	5	6	5	6	2	32	6	6	5	6	1	53	3	3	3	3	7
12	6	6	6	6	2	33	6	6	5	5	1	54	3	3	2	2	5
13	7	7	6	6	2	34	5	5	5	5	1	55	2	2	1	2	8
14	7	7	7	7	3	35	5	5	5	5	2	56	3	4	3	3	8
15	7	7	7	7	1	36	5	5	5	5	2	57	4	4	4	4	4
16	7	8	7	8	3	37	5	5	5	5	5	58	4	4	3	3	3
17	8	8	8	8	3	38	5	6	5	6	2	59	3	4	3	3	2
18	7	7	7	7	2	39	5	6	5	6	1	60	4	4	3	3	1
19	7	7	7	7	1	40	6	6	6	6	2	61	3	3	3	3	2
20	7	7	7	7	1	41	6	6	5	5	2	62	3	3	2	2	3
21	7	7	7	7	2	42	6	6	6	6	2	63	1	3	1	3	4

4 结束语

开盘价、收盘价、最高价、最低价、交易量都是股票市场的重要属性,不同的证券市场数据,应选择不同的属性指标计算分形维数,才能更合理地反映市场的复杂度以及不规则程度。文中利用一种新的属性选择分类标准——集合分类粗糙度,并结合粗糙集中决策表的数据离散化方法,讨论计算股票市场中的股票价格的分形维数问题,为分形维数的计算,选择合适的属性提供了一条新的途径。

参考文献:

[1] Osborne M F M. The Random Character of Stock Market Prices[M]//Cootner P. Cambridge, MA: MIT Press, 1964.  
[2] Peters Edgar E. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics[M]. New York: John Wiley & Sons, Inc, 1994.

当地规划办提供,是一种预测值;人口数是从该地区的城市规划资料中获得。最后采用改进型的 RBFANN 对该地区的电力负荷进行预测。三种方法的预测结果如表 1 所示。

表 1 三种预测方法的结果

年度	灰色理论	RBFANN 法	改进型 RBFANN	实际负荷
2004 年	13.83	12.98	12.68	12.51
2005 年	15.18	14.29	14.58	14.01
2006 年	17.31	16.47	15.94	15.41
2007 年	20.55	17.65	17.55	16.94
2008 年	23.92	18.35	19.87	19.48

注:表中各产值的单位为亿元,人口的单位为万人,负荷的单位为万千瓦。

由表 1 可以看出:三种方法的预测值有一定的差

距。可见较之灰色理论和 ANN 法,该改进型 RBFANN 在实际应用中是较精准的。

5 结束语

文中提出一种混合式最优化储层参数识别预测方法,该方法引入自适应机制的浮点数编码的遗传算法,并将其与梯度下降法混合交互运算,作为径向基函数网络的学习算法,形成了基于改进遗传算法的径向基函数网络,它克服了径向基函数网络的学习算法上的缺陷,在实际工作中能够良好地进行应用。文中最后给出了该算法在某地区电力负荷的预测,实例证明了该算法的优越性,值得进一步探讨和研究。

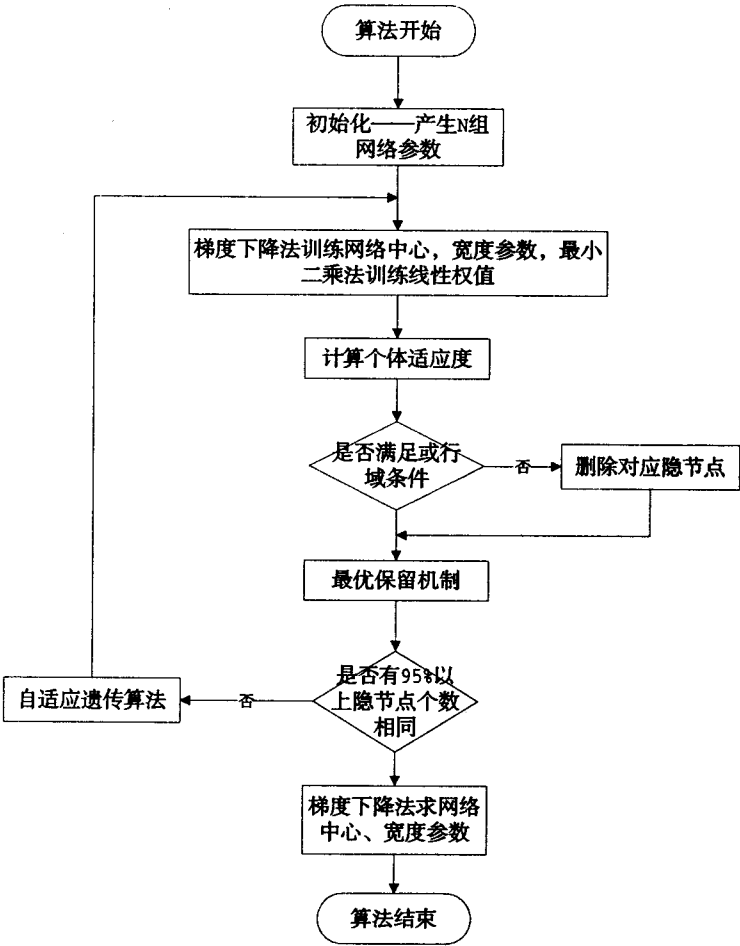


图 1 算法流程图

参考文献:

[1] 顾庆雯,陈 刚.基于遗传算法和径向基函数神经网络的短期边际电价预测[J].电网技术,2007,30(7):153-155.

[2] 李敏强,林 丹.遗传算法的基本理论与应用[M].北京:科学出版社,2003.

[3] 赵志刚,单晓虹.一种基于遗传算法的 RBF 神经网络优化方法[J].计算机工程,2007,33(6):211-212.

[4] Roy A, Govil S, Miranda R. A Neural Network Learning Theory and a Polynomial Time RBF Algorithm[J]. IEEE Trans on Neural Network, 1997,8(6):1301-1313.

[5] 张明辉,王尚锦.具有自适应交叉算子的遗传算法及其应用[J].机械工程学报,2002,38(1):74-78.

[6] 董艳昌,王志远.基于浮点制编码自适应遗传算法在模型识别中的应用[J].Application and Practice,2006,40(1):20-23.

[7] 谢培元,游大海,曾次玲.基于遗传算法优化 BP 网络的提前一天市场清算电价预测[J].电力自动化设备,2004,24(3):53-56.

[8] 陈泽淮,张 尧,武志刚. RBF 神经网络在中长期负荷预测中的应用[J].电力系统及其自动化学报,2006,18(1):15-19.

(上接第 138 页)

[3] Taqqu M S, Willinger W, Sherman R. Proof of a fundamental result in self-similar trac modeling[J]. Computer Communication Review,1997,27:5-23.

[4] 姜灵敏,周 锋.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):434-436.

[5] 张文修,梁 怡,吴伟志.信息系统与知识发现[M].北京:

科学出版社,2003:22-33.

[6] 韦萍萍.结合 ROUGH 集的决策树构建方法[J].重庆工学院学报,2007,21(9):101-103.

[7] 蒋 芸.一种基于粗糙集构造决策树的新方法[J].计算机应用,2004,24(8):21-23.

[8] 邵峰晶,于忠清.数据挖掘原理与算法[M].北京:中国水利水电出版社,2003:101-127.