灰色-周期外延模型的动态关联规则元规则挖掘

张忠林 石皓尹 宋 航
(兰州交通大学电子与信息工程学院 兰州 730070)

摘 要 提出了一种把灰色-周期外延模型应用到动态关联规则元规则挖掘中并提高元规则预测精度的方法。该方法首先对动态关联规则元规则支持度计数建立GM(1,1)模型;然后对建模时产生的残差序列建立周期外延模型并且提取最优周期;其次将周期外延模型作为GM(1,1)模型的残差补偿,把GM(1,1)模型以及周期外延模型得到的结果进行叠加得到最终的预测模型。最后通过实验证明了该方法有较高的预测精度。

关键词 GM(1,1)模型,周期外延模型,动态关联规则,元规则
中图法分类号 TP273 **文献标识码** A

Research of Mining Meta-association Rules for Dynamic Association Rule Based on Gray-periodic Extensional Model

ZHANG Zhong-lin SHI Hao-yin SONG Hang
(School of Electronic and Information Engineering, Lanzhou Jiaotong University, Lanzhou 730070, China)

Abstract A method applying Gray-Periodic extensional combinatorial model to Meta-rule mining for dynamic association rules to improve the prediction accuracy was proposed. In this method, firstly a GM(1,1) model is established by counting the support of the Meta-rule for dynamic association rules and then the residuals sequence generated during modeling is integrated to build the periodic extensional model and extract the optimal cycle. Secondly, the periodic extensional model is used as residual compensation for the GM(1,1) model. So the final prediction model is presented when superimposing the GM(1,1) model and results of the periodic extensional model together. At last, the results show that the method has higher prediction accuracy.

Keywords GM(1,1) model, Periodic extensional model, Dynamic association rules, Meta-rule

1 引言

关联规则挖掘是数据挖掘中主要的研究方法[1],主要用于发现事物数据集中项与项之间的关系[2]。由于事务数据通常具有时间特性,因此提出了考虑时间因素的序列模式挖掘;为了描述关联规则随时间变化的特点,提出了动态关联规则的定义[3];在以上基础上进一步改进了动态关联规则的定义及挖掘算法[4,5];文献[6]首先提出了动态关联规则元规则的定义,且文献[6-8]中已经提出了几种不同的动态关联规则元规则的挖掘方法。鉴于灰色-周期外延模型不仅能够反映出系统的总体变化趋势,而且能够体现出变化的周期性和动态性,因而成为了一种重要的分析工具和方法[9,10]。灰色-周期外延模型已被广泛应用于电力负荷预测、货运量的预测、地下水的动态预测等各个领域中,并且取得了很好的效果。本文提出了一种把灰色-周期外延模型应用到动态关联规则元规则挖掘的方法,该方法在一定程度上可提高预测的精准度。

2 动态关联规则

动态关联规则的描述如下:设 $I=\{i_1,i_2,\cdots,i_n\}$ 是项集合,数据集 D 是在时间段 t 内收集到的,t 为相等不相交的长度为 n 的时间序列,即有 $t=\{t_1,t_2,\cdots,t_n\}$。根据时间段的划分将数据集 D 分为 n 个子数据集,其中数据子集 $D_i(i\in\{1,2,\cdots,n\})$ 是在 $t_i(i\in\{1,2,\cdots,n\})$ 时间段内收集的项集 T,其满足 $T\subseteq I$。若 A 和 B 为项集,$A\subset I$,$B\subset I$,并且 $A\cap B=\varnothing$,则有如下相关定义:

定义1 支持度向量(SV)是动态关联规则 $A\Rightarrow B$ 的向量,有如下表示形式:

$$SV=[s_{(A\cup B)_1},s_{(A\cup B)_2},\cdots,s_{(A\cup B)_n}]$$

其中,$s_{(A\cup B)_i}=f_{(A\cup B)_i}/|D_i|(i\in\{1,2,\cdots,n\})$,$f_{(A\cup B)_i}$ 为 $A\cup B$ 在数据子集 $D_i(i\in\{1,2,\cdots,n\})$ 中出现的频数,$|D_i|$ 为 D_i 中的事务数。

定义2 设项集 $A\cup B$ 的支持度为 s,则:

$$s=s_{(A\cup B)}=f_{(A\cup B)}/M=\sum_{i=1}^{n}f_{(A\cup B)_i}/M$$

其中,M 是 D 中的事务数。项集的支持度向量为 $SV=[f_1,f_2,\cdots,f_n]$。

定义3 动态关联规则 $A\Rightarrow B$ 的置信度向量具有以下表示形式:

到稿日期:2013-06-09 返修日期:2013-08-06 本文受国家自然科学基金资助项目(61163010),甘肃省科技支撑计划项目(1104GKCA016)资助。

张忠林(1965—),男,博士,教授,CCF会员,主要研究方向为智能信息处理、软件工程,E-mail:zhangzl@mail.lzjtu.cn;石皓尹(1989—),男,硕士生,主要研究方向为数据挖掘;宋 航(1989—),男,硕士生,主要研究方向为数据挖掘。

$CV=[c_{(A\cup B)_1},c_{(A\cup B)_2},\cdots,c_{(A\cup B)_n}]$

其中，$c_{(A\cup B)_i}=s_{(A\cup B)_i}/s_{A_i}$，$s_{(A\cup B)_i}$为$A\cup B$的$SV$中的第$i$个元素，$s_{A_i}$为$A$的$SV$中的第$i$个元素。

定义4 设项集$A\cup B$的置信度为c，则：

$$c=\sum_{i=1}^{n}s_{(A\cup B)_i}/\sum_{i=1}^{n}s_{A_i}=\sum_{i=1}^{n}c_{(A\cup B)_i}$$

动态关联规则表述为具有支持度向量SV、置信度向量CV、支持度s、置信度c 4个参数的关联规则。表示形式为：$A\Rightarrow B(SV,CV,s,c)$。元规则形如$A_1\wedge A_2\wedge\cdots\wedge A_l\Rightarrow B_1\wedge B_2\wedge\cdots\wedge B_r$，其中$A_i(i=1,2,\cdots,l)$和$B_j(j=1,2,\cdots,r)$是示例谓词或是谓词变量。

定义5 $D=\{D_1,D_2,\cdots,D_n\}(i=1,2,\cdots,n)$上，规则$A\Rightarrow B$的支持度向量定义为：

$$SV=[s_{(A\cup B)_1},s_{(A\cup B)_2},\cdots,s_{(A\cup B)_n}]$$

只有当$S(A\cup B)_i\geqslant$min_sup，其中min_sup为最小支持度。则动态关联规则在数据集D上的支持度元规则为$A\Rightarrow B$：SV。

3 灰色-周期外延模型

3.1 GM(1,1)模型

我国学者邓聚龙教授首次提出了灰色系统理论。灰色系统理论是建立系统运行趋势模型的有效理论方法[11,12]，适用于动态预测，且只需少量已知信息就可建立预测模型[13]。灰色GM(1,1)预测模型的具体步骤如下：

(1)设序列$\{x^{(0)}=(x^{(0)}(1),x^{(0)}(2),\cdots,x^{(0)}(n)),n=1,2,\cdots,n\}$为原始序列，对序列$x^{(0)}$进行一次累加生成处理得到序列$x^{(1)}$：

$$x^{(1)}(i)=\sum_{t=1}^{i}x^{(0)}(t),i=(1,2,\cdots,n)\tag{1}$$

(2)$x^{(1)}=(x^{(1)}(1),x^{(1)}(2),\cdots,x^{(1)}(n))$满足单变量常微分方：$\frac{dx^{(1)}}{dt}+ax^{(1)}=b$。$-a$为发展系数，$b$为灰色作用变量。此方程的解为：

$$\hat{x}^{(1)}(k+1)=[x^{(0)}(1)-\frac{b}{a}]e^{-ak}+\frac{b}{a}\tag{2}$$

系数a和b运用灰色系统生成理论按最小二乘法可求解：

$$\begin{bmatrix}\hat{a}\\ \hat{b}\end{bmatrix}=(B^{T}B)^{-1}B^{T}\bar{Y}\tag{3}$$

其中：

$$\bar{Y}=\begin{bmatrix}x^{(0)}(2)\\ x^{(0)}(3)\\ \vdots\\ x^{(0)}(n)\end{bmatrix}\tag{4}$$

$$B=\begin{bmatrix}-\frac{1}{2}(x^{(1)}(2)+x^{(1)}(1)) & 1\\ -\frac{1}{2}(x^{(1)}(3)+x^{(1)}(2)) & 1\\ \vdots & \vdots\\ -\frac{1}{2}(x^{(1)}(n)+x^{(1)}(n-1)) & 1\end{bmatrix}\tag{5}$$

(3)把估计值$\hat{a},\hat{b}$代入式(2)中得到时间影响方程：

$$\hat{x}^{(1)}(k+1)=[x^{(0)}(1)-\frac{\hat{b}}{\hat{a}}]e^{-\hat{a}k}+\frac{\hat{b}}{\hat{a}}\tag{6}$$

(4)对序列$\hat{x}^{(1)}(k+1)$做还原处理可得原始列$x^{(0)}$的灰色预测模型：

$$\hat{x}^{(0)}(k)=(1-e^{\hat{a}})(x^{(0)}(1)-\frac{\hat{b}}{\hat{a}})e^{-\hat{a}(k-1)}\tag{7}$$

当$k=1,2,\cdots,n$时，由式(7)得到的是拟合值，而当$k\geqslant n+1$时，得到的是预测值。

在运用灰色系统理论建立GM(1,1)模型时需要注意两点：

①序列$x^{(0)}$的级比$\sigma(k)=\frac{x(k-1)}{x(k)}$必须落在可行域：$\sigma(k)\in(e^{\frac{-2}{n+1}},e^{\frac{2}{n+1}})$，其中$n$为序列$x^{(0)}$的长度。

②GM(1,1)模型的适用范围与发展系数$-a$相关[14]，具体如表1所列。

表1 GM(1,1)模型适用范围表

发展系数$-a$	模型适用范围
$-a<0.3$	可用于中短期预测
$0.3<-a\leqslant 0.5$	可用于短期预测
$0.5<-a\leqslant 0.8$	可用于短期预测，应十分谨慎
$-a>1$	不宜采用

3.2 周期外延模型

周期外延模型的优势在于它能很好地反映出系统的周期波动特点[9,10]。建立数列周期外延模型的具体步骤如下：

1)计算序列$x(k)$的均值生成函数：

$$\bar{x}_m(i)=(\sum_{j=0}^{n_m-1}x(i+jm))/n_m(i=1,2,\cdots,m(1\leqslant m\leqslant M))$$

其中，n为样本序列长度；n_m为小于n/m的最大整数；M为小于$n/2$的最大整数。可得均值生成函数矩阵：

$$\begin{bmatrix}\bar{x}_1(1) & \bar{x}_2(1) & \bar{x}_3(1) & \cdots & \bar{x}_M(1)\\ & \bar{x}_2(2) & \bar{x}_3(2) & \cdots & \bar{x}_M(2)\\ & & \bar{x}_3(3) & \cdots & \bar{x}_M(3)\\ & & & \ddots & \cdots\\ & & & & \bar{x}_M(M)\end{bmatrix}$$

对均值生成函数$\bar{x}_m(i)$作周期性延拓，即令$f_m(k)=\bar{x}_m(k)$，$k=i[\mathrm{mod}(m)]$，$k=1,2,\cdots,n$，式中mod表示同余；$f_m(k)$称作均值生成函数的延拓函数。

2)提取优势周期。目前有两种方法：

①可用下式来检验序列$x(k)$是否隐含长度为m的周期。

$$F^{(m)}=(n-m)S^{(m)}/((m-1)S)$$

其为服从自由度$(m-1,n-m)$的F分布。其中

$$S^{(m)}=\sum_{i=1}^{m}n_i(\bar{x}_m(i)-\bar{x})^2$$

$$n_i=n/i,\bar{x}=(\sum_{i=1}^{n}x(i))/n$$

$$S=\sum_{i=1}^{m}\sum_{j=n}^{n}(x(i+(j-1)m)-\bar{x}_m(i))^2$$

对于事先给定的置信水平α，若$F^{(m)}>F_\alpha(m-1,n-m)$，则认为$x(k)$隐含长度为$m$的优势周期。

②确定长度为m的优势周期，只需取：

$$S(m)/m=\max S(m)/m,2\leqslant m\leqslant M$$

3)序列$x(k)$减去周期m所对应的延拓函数构成一新序列，即$x'(k)=x(k)-f_m(k)$，再对新序列$x'(k)$重复2)、3)，进

一步提取优势周期。

4)将不同周期同一时刻的值叠加记为 $f(k)$，$f(k)=\sum_{i=1}^{m} f_i(k)$，称此模型为周期外延模型，$x(k)$可近似地取为 $f(k)$。

4 方法描述

现有的关联规则挖掘方法大多是基于静态数据库的，是将现在之前的所有数据作为一个整体，通过给定最小支持度和最小置信度来计算的静态挖掘算法。且在挖掘中认为所有的项目和事物都是同等重要的，其次对挖掘所得到的规则进行分析发现大多规则在时间上的分布不均匀。综上所述，静态关联规则已经不适应当代各方面发展的需求。在对关联规则挖掘深入的研究之后出现了动态关联规则。动态关联规则挖掘与之前的静态关联规则挖掘的本质区别在于挖掘过程中充分考虑到时间因素的重要性。通过对时间段进行划分，对每个子时间段内的数据通过设置最小支持度和最小置信度的方法进行挖掘，可以得到连续且分布均匀的规则。因此动态关联规则挖掘在一定程度上克服了静态关联规则挖掘的不足，其参考价值和指导意义相对静态关联规则挖掘有了显著的提高。

但是动态关联规则依然有不足之处：不同的时间段划分方法可能会得到不同的结果；在挖掘过程中所有项目和事物不能具有相同的重要性，因此结果的参考价值也会不同；在大多数挖掘方法中我们都需提前设定最小支持度阈值和最小置信度阈值，因此如果给定的阈值不同，那么所得到的挖掘结果也必将不同。

元规则简单地说就是“规则”的规则，元规则的挖掘就是要发现规则本身随时间变化的趋势，进一步预测下一个或者多个时间段的支持度值。因此为动态关联规则建立元规则，可以对关联模式进行更加准确有效的挖掘。针对动态关联规则元规则的挖掘目前提出的方法主要有：基于灰色-Markov模型的元规则挖掘方法、基于 AR-Markov 模型的元规则挖掘方法、基于时间序列的元规则挖掘方法以及基于小波变换理论的元规则挖掘方法等。以上几种挖掘方法可以针对具有不同特点的动态关联规则元规则进行挖掘，但是没有充分考虑关联规则在实际挖掘中的周期性。例如针对某商场或者超市的销售数据进行数据挖掘之后，我们可能会发现部分关联规则具有很强的周期性或者季节性。这时如果我们还用以上提出的几种方法来进行元规则挖掘，那么预测的精度会大大地下降，从而得到可靠性不高的结果。

本文所采用的灰色-周期外延模型的挖掘方法既能很好地反映序列的总体趋势，也能很大程度地反映出周期变化这一特点，具体过程如下：

①$SV=[f_1, f_2, \cdots, f_n]$为经过挖掘所得到的支持度向量，将 SV 中各数据组成的支持度计数序列$\{f_1, f_2, \cdots, f_n\}$作为原始序列 $x(k)$进行级比 $\sigma(k)$验证，并且对其建立 GM(1,1)模型。

②运用 GM(1,1)模型建模方法得到序列$\{f_1, f_2, \cdots, f_n\}$的预测公式。

③根据公式得到序列的拟合值$\hat{x}(k)$，由拟合值和真实值进一步得到残差序列 $x'(k)$。

④运用周期外延模型的相关知识，将 $x'(k)$进行周期外延建模。

⑤确定 $x'(k)$的优势周期。

⑥将序列 $x'(k)$中不同周期中同一时刻的值相加：$f(k)=\sum_{i=1}^{m} f_i(k)$。

⑦将$\hat{x}(k)$与 $f(k)$之和作为原始序列 $x(k)$的最终拟合值：$\bar{x}(k)=\hat{x}(k)+f(k)$，进一步得到最终的预测值。

5 实验分析过程

为了更加直观地说明上述挖掘方法在动态关联规则的元规则挖掘的具体过程，下面将针对实例进行说明。本文实例分析部分所采用的数据来源于文献[11]，将其处理后作为某超市 2010 年 1 月至 2011 年 3 月的销售数据来挖掘出动态关联规则的支持度序列，$SV=\{278, 256, 273, 273, 255, 269, 268, 254, 264, 261, 248, 258, 257, 243, 255\}$。

为了验证本文所用方法的预测准确性，用 2010 年全年 12 个月的数据(SV中前 12 个数据)建立模型，并用此模型预测 2011 年第一季度支持度计数的预测值。在此基础上，将预测结果和 2011 年第一季度的实际结果(SV中后 3 个数据)进行对比。

具体步骤如下：

1)将 SV 中前 12 个数据作为原始序列$\{x^{(0)}=(x^{(0)}(1), x^{(0)}(2), \cdots, x^{(0)}(n)), n=1,2,\cdots 12\}$，建立 GM(1,1)模型。

2)得到原始序列 GM(1,1)模型的发展系数 $-a$ 和灰色作用变量 b 的值。其中 $-a=-0.004242$，$b=269.061410$(注：为简化期间，只取小数点后 6 位，而在具体建模时以计算出的实际值进行运算)，得到时间影响公式：

$$\hat{x}(k)=268.451027e^{-0.004242(k-1)}$$

3)根据时间影响公式得出原始序列的拟合值、残差以及残差百分比，如表 2 所列。

表 2 预测模型拟合值分析表

月份	真实值	拟合值	残差	百分比
1	278	268.451027	9.548972	3.434882%
2	256	267.314504	−11.314504	−4.419728%
3	273	266.182793	6.817206	2.497145%
4	273	265.055873	7.944126	2.909936%
5	255	263.933724	−8.933724	−3.503421%
6	269	262.816326	6.183673	2.298763%
7	268	261.703659	6.296340	2.349380%
8	254	260.595702	−6.595702	−2.596733%
9	264	259.492436	4.507563	1.707410%
10	261	258.393841	2.606158	0.998528%
11	248	257.299896	−9.299896	−3.749958%
12	258	256.210584	1.789415	0.693572%

由上表可以看出：GM(1,1)能反映出序列总体变化趋势，即有逐步降低之势；负误差每隔两个数据出现一次，即周期出现，说明 GM(1,1)模型没有反映出序列周期波动的现象。

4)计算残差序列 $x'(k)$的均值生成函数，可得均值生成函数矩阵：

$$\begin{bmatrix} \bar{x}_1(1) & \bar{x}_2(1) & \bar{x}_3(1) & \cdots & \bar{x}_6(1) \\ & \bar{x}_2(2) & \bar{x}_3(2) & \cdots & \bar{x}_6(2) \\ & & \bar{x}_3(3) & \cdots & \bar{x}_6(3) \\ & & & \ddots & \cdots \\ & & & & \bar{x}_6(6) \end{bmatrix} =$$

$$\begin{bmatrix} 0.7958 & 1.4894 & 6.5988 & 1.7076 & 3.2163 & 7.9226 \\ & 0.1021 & -9.0359 & -0.8415 & -1.6143 & -8.9551 \\ & & 4.8244 & 1.2712 & 0.1107 & 5.6623 \\ & & & 1.0459 & 6.2258 & 5.2751 \\ & & & & -3.1637 & -9.1168 \\ & & & & & 3.9865 \end{bmatrix}$$

5)提取残差序列 $x'(k)$ 的周期。本文采用 $S(m)/m=\max S(m)/m$ 的方法确定周期，可以得到 $S(3)/3=\max S(m)/m$。据此，残差序列 $x'(k)$ 的周期为 3。

6)周期叠加建立的周期外延预测模型为 $f(k)=f_3(k)$。经计算，均值生成函数 $\bar{x}_3(1)=6.5988$，$\bar{x}_3(2)=-9.0359$，$\bar{x}_3(3)=4.8244$。

7)叠加生成灰色-周期外延组合模型：

$$\overline{\hat{x}}(k)=\hat{x}(k)+f(k)=268.451027e^{-0.004242(k-1)}+f_3(k)$$

该模型的拟合效果见表 3。

表 3　组合模型拟合值分析表

月份	真实值	拟合值	残差	百分比
1	278	275.049926	2.950073	1.061177%
2	256	258.278547	−2.278547	−0.890057%
3	273	271.007258	1.992741	0.729942%
4	273	271.654773	1.345226	0.492757%
5	255	254.897767	0.102232	0.040091%
6	269	267.640791	1.359208	0.505281%
7	268	268.302559	−0.302559	−0.112895%
8	254	251.559745	2.440254	0.960730%
9	264	264.316901	−0.316901	−0.120038%
10	261	264.992740	−3.992740	−1.529785%
11	248	248.263939	−0.263939	−0.106427%
12	258	261.035048	−3.035048	−1.176375%

两种模型的拟合精度分析如图 1 所示。

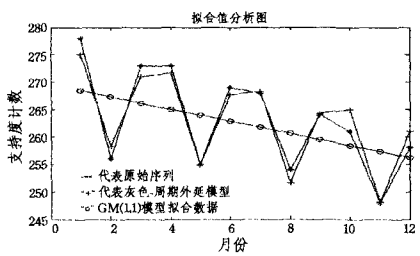

图 1　拟合值分析图

由表 3 和图 1 可知，GM(1,1)模型预测的误差百分比绝对值之和是灰色-周期外延模型预测的误差百分比绝对值之和的近两倍，这主要是因为单纯地使用 GM(1,1)模型时结果只具有下降的趋势，而灰色-周期外延组合模型考虑了数据周期波动性，而且 GM(1,1)模型只能表现出变化的总体趋势。因此，对原始数据的周期性变化进行分析可以在一定程度上提高拟合和预测的精度和准确性，可以准确地把握动态规则支持度的未来趋势。

根据表 2 和表 3 可得：灰色-周期外延模型的误差和 GM(1,1)模型的误差都是在可接受范围内，我们分别用两种模型来对 2011 年第一季度的数据做预测，之后用实际数据验证预测精度。具体结果如表 4 所列。

由表 4 可知：灰色-周期外延模型的预测精度要大大高于单纯使用 GM(1,1)模型的预测精度，因此当原始数据具有明显的周期变化特点时，使用灰色-周期外延模型可以准确地把握动态规则支持度的未来趋势。

表 4　两种模型预测分析表

月份		1	2	2
真实值		257	243	255
GM(1,1)模型	预测值	255.125882	254.045774	252.970237
	误差	1.874117	−11.045774	2.029762
	百分比	0.07292%	−4.5455%	0.07959%
灰色—周期外延	预测值	261.724782	245.009816	257.794702
	误差	−4.724782	−2.009816	−2.794702
	百分比	−1.83843%	−0.0827%	−1.0959%

6　方法对比

我们已经通过实例验证了本文所提出的基于灰色-周期外延模型的动态关联规则元规则挖掘方法的有效性，并且本文第 5 节将本方法和 GM(1,1)模型进行了纵向对比。文献[6-8]中已经提出了几种动态关联规则元规则挖掘的方法，且已有的方法都取得了良好的效果。下面把本文所提出的方法和文献[8]中提出的基于 AR-Markov 模型的挖掘方法进行横向对比，进一步研究本方法适用的范围和建模特点。

下面将使用本文的方法对文献[8]中的实例数据进行建模分析，从而与基于 AR-Markov 模型的动态关联规则元规则挖掘方法做详细对比。

对序列[162,175,162,156,174,157,154,177,159,171,166,162,171,183,163,152,165,175,167,170,170,167,171,164]建立灰色-周期外延模型。

首先我们得到 GM(1,1)模型的预测公式：$\hat{x}(k)=164.474853e^{0.001050(k-1)}$。得到拟合值之后通过对残差序列建立周期外延模型，可得残差序列的最优周期为 12。进一步可以得到最终的预测公式：

$$\hat{x}(k)=164.474853e^{0.001050(k-1)}+f_{12}(k)$$

拟合值、残差以及残差百分比的具体结果如表 5 所列(与文献[8]一样，我们只对 2007 年的数据进行拟合分析)。

表 5　拟合值分析表

月份	真实值	拟合值	残差	百分比
1	171	167.5431	3.4568	2.0215%
2	183	180.0442	2.9557	1.6151%
3	163	163.5453	−0.5453	−0.3345%
4	152	155.0464	−3.0464	−2.0042%
5	165	170.5475	−5.5475	−3.3621%
6	175	167.0486	7.9513	4.5436%
7	167	161.5497	5.4502	3.2636%
8	170	174.5508	−4.5508	−2.6769%
9	170	165.5519	4.4480	2.6164%
10	167	170.0530	−3.0530	−1.8281%
11	171	169.5541	1.4458	0.8455%
12	164	164.0552	−0.0552	−0.0337%

两种方法的具体对比结果如图 2 所示。

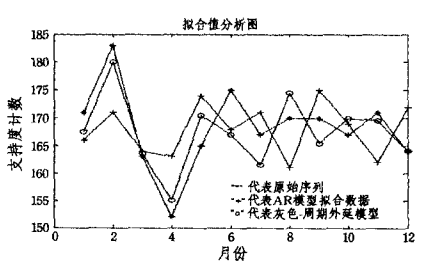

图 2　拟合值分析图

通过表 5 和图 2 我们发现在拟合过程中，本文所提方法

的拟合精度要高于文献[8]中所用挖掘方法的精度，且使用AR模型时会出现个别误差较大的情况。而使用灰色-周期外延模型时没有出现误差过大的情况，并且误差分布得比较平稳。再次验证了本方法具有较好的有效性和较高的预测精度。

结束语 本文提出了一种基于灰色-周期外延模型的动态关联规则元规则挖掘方法。通过一个实例说明本方法的一般过程：首先建立序列GM(1,1)模型，然后对残差序列建立周期外延模型，作为灰色GM(1,1)模型的残差补偿。通过比较得出此模型比直接用灰色模型预测要可靠、合理，总体效果要优于灰色GM(1,1)模型，能更准确地把握规则的变化趋势，从而使动态关联规则挖掘在合理的元规则指导下得到更精确的结果。今后的主要研究方向是将灰色-周期外延模型应用到动态关联规则趋势度的挖掘中[15]；当原始数列波动性过大且无明显规律时如何保证预测的精度；对象为大数据库或者海量数据时，如何提高挖掘算法的效率和准确性[16]。

参 考 文 献

[1] Agrawal R, Imeielinksi T, Swami A. Mining Assocaiation Rules Between Sets of Items in Large Databases [C]// Proceedings of the 1993 ACM SIGMOD Conference Washington DC. Washington, 1993: 207-216

[2] Au W-H, Chan K C C. Mining changes in association rules: a fuzzy approach[J]. Fuzzy Sets Syst, 2005, 149(1): 87-104

[3] 荣冈，刘进锋，顾海杰. 数据库中动态关联规则的挖掘[J]. 控制理论与应用，2007，24(1)：129-133

[4] 沈斌，姚敏. 一种新的动态关联规则及其挖掘算法[J]. 控制与决策，2009，24(9)：1310-1315

[5] Shen Bin, Yao Min, Wu Zhao-hui, et al. Mining dynamic association rules with comments [J]. Knowledge and Information Systems, 2010, 23(1): 73-98

[6] 刘俊，谢彦峰，张忠林. 基于灰色Markov模型动态关联规则元规则挖掘[J]. 计算机应用，2008，28(9)：2353-2356

[7] 张忠林，许凡. 基于小波变换的动态关联规则元规则GM(1,1)挖掘[J]. 计算机科学，2013，40(5)：209-212，246

[8] 张忠林，刘俊，谢彦峰. AR-Markov模型在动态关联规则挖掘中的应用[J]. 计算机工程与应用，2010(14)：135-137，147

[9] 任峰，李伟，丁超. 基于灰色-周期外延组合模型的电力负荷预测[J]. 电网技术，2007，24：52-54

[10] 杨俊祥，程盛芳. 灰色-周期外延组合模型在煤炭需求预测中的应用[J]. 统计与决策，2010，13：162-163

[11] 刘思峰，党耀国，方志耕，等. 灰色系统理论及其应用[M]. 北京：科学出版社，2004，142-146

[12] Zhang Yi, Wei Yong, Zhou Ping. Improved Approach of Gray Derivative in GM (1,1) Model [J]. The Journal of Grey System, 2006, 116(10): 160-162

[13] Sun Yan-na. Optimization of Grey Derivative in GM(1,1) Based on the Discrete Exponential Sequence, 2009 [C]// Proceeding of the 2nd International Symposium on Information Processing(ISTP2009). Huangshan, P. R. China, 2009: 313-315

[14] 刘思峰，邓聚龙. GM(1,1)模型的适用范围[J]. 系统工程理论及实践，2000，20(5)：121-124

[15] 张忠林，曾庆飞，许凡. 动态关联规则的趋势度挖掘方法[J]. 计算机应用，2012，32 (1)：196-198

[16] Angel K-M, Fatima R-E. A search space reduction methodology for data mining in large databases[J]. Engineering Applications of Artificial Intelligence, 2009, 22(1): 57-65

(上接第 232 页)

从上面的几组实验中可以看到，所给算法对非线性回归问题具有较好的拟合和预测能力。

结束语 分类和回归技术在实际中具有广泛的应用领域，比如医学中的疾病诊断就是一种模式分类问题，而根据历史的经济数据建立回归模型来预测未来的经济趋势是统计工作者经常应用的方法。本文根据一类分类思想和核方法提出一种新的非线性支持向量回归算法，该算法归结为求解一个线性规划，因此其运算速度要明显优于基于二次规划的标准支持向量回归算法。另外，该算法还展示了一类分类和回归之间的关系，从而拓宽了建立支持向量回归的途径，算法有助于直观地理解支持向量回归的几何意义。在两个人工数据和一个实际数据上的实验结果显示了所给算法具有良好的泛化能力。

参 考 文 献

[1] Vapnik V N. The Nature of Statistical Learning Theory [M]. Berlin: Springer-Verlag, 1995

[2] Burges C J C. A tutorial on support vector machines for pattern recognition[J]. Knowledge Discovery and Data Mining, 1998, 2 (2): 34-38

[3] Smola A J, Scholkopf B. A tutorial on support vector regression [R]. NeuroCOLT TR NC-TR 98-030. Royal Holloway College University of London, UK, 1998

[4] Suykens J A K, Vandewalle J. Least squares support vector machines classifiers[J]. Neural Processing Letters, 1999, 9(3): 293-300

[5] 陶卿，孙德敏，范劲松，等. 基于闭凸包收缩的最大边缘线性分类器[J]. 软件学报，2002，13(3)：404-409

[6] 陶卿，曹进得，孙德敏. 基于支持向量机分类的回归算法[J]. 软件学报，2002，13(5)：1024-1028

[7] Scholkopf B, Williamson R C, Smola A J, et al. Support vector method for novelty detection[J]. Advances in Neural Information Processing Systems, 2000, 12(3): 582-588

[8] Scholkopf B, Platt J C, Shawe-Taylor J, et al. Estimating the support of a high-dimensional distribution[J]. Neural Computation, 2001, 13(7): 1443-1471

[9] 徐磊，赵光宙. 基于模糊一类支持向量机的核聚类算法[J]. 控制与决策，2008，23(9)：1030-1034

[10] 田江，顾宏. 孤立点一类支持向量机算法研究[J]. 电子与信息学报，2010，32(6)：1284-1288

[11] 张庆，徐光华，华成，等. 一类支持向量机的设备状态自适应报警方法[J]. 西安交通大学学报：自然科学版，2009，11：61-65

[12] 陈伟，余旭初，张鹏强，等. 基于一类支持向量机的高光谱影像地物识别[J]. 计算机应用，2011，31(8)：2092-2096

[13] Mangasarian O L. Arbitrary-norm separating plane[J]. Operations Research Letters, 1999, 1(24): 15-23