

时间序列短期趋势信号模型研究

宫 正, 刘晓燕

(昆明理工大学 信息工程与自动化学院, 云南 昆明 650051)

摘 要:时间序列是按固定的时间间隔对数据采样,按照时间顺序依次排列的一组被观测数据或信息。随着我国金融证券市场的不断发展和日渐完善,金融时间序列的研究对金融投资者具有越来越重要的意义,受到越来越多的研究者和投资者的关注。文中的研究目的是通过金融时间序列的短期的趋势信号估计金融时间序列的短期趋势,提出短期趋势为信号的模型,用最小二乘法对所估计出的短期趋势建立趋势模型,并做了短期趋势信号的模型的实证研究。通过对上证 A 股的时间序列进行实证分析,实验表明所建立的模型是有效的,能够为投资者提供参考。

关键词:时间序列;短期趋势信号;最小二乘法;趋势模型

中图分类号:TP31

文献标识码:A

文章编号:1673-629X(2011)12-0105-04

Study on Time Series of Short-Term Trend Signal Model

GONG Zheng, LIU Xiao-yan

(Faculty of Information Engineering and Automation, Kunming University of Science and Technology,
Kunming 650051, China)

Abstract: Time series is based on a fixed time interval of data sampling. Time series is arranged in chronological order a set of data or information which has been observed. With the development of financial market in our country, financial time series are becoming more important for financial investors. More and more scholars and investors pay attention to the study of time series. It is on the trend of financial time series. Detect the short term signals and predict the short term trend. Detect the rising trend and falling trend in financial time series. Classify the trend in time series as strict trend, flexible trend and interval trend. Also define the strength of trend. Use actual value of time series to construct a model to modify the forecast value of the financial time series. Predict the trend rightly. So it is useful to investors.

Key words: time series; short term signals; least square method; model of trend

0 引 言

我国的证券交易市场已成为社会参与程度最高的金融市场,要获得较高的收益必须对股票时间序列的趋势有比较准确的预测,因而对股票时间序列的研究成为投资者和金融领域的研究者尤其重视的课题。

在时间序列分析中主要有两类方法^[1]:

其一是频谱分析,这种方法最主要的假设就是任何没有趋势的时间序列都可以被分解成为一定数量的具有周期性的谱图。

第二类方法是时域分析法,可以追溯到英国统计学家 G. U. Yule 于 1927 年提出的自回归模型。

此外,George E. P. Box 和 Gwilym M. Jenkins 提出的时序分析模型被称作博克斯-詹金斯(Box-Jenkins)

法,其基本模型有三种,自回归模型(AR)、移动平均模型(MA)和自回归移动平均模型(ARMA)^[2,3],这些方法对稳定的时间序列预测有着较好的结果。但是这些模型在理论上和应用上还存在着许多局限性,所以统计学家开始转向多变量条件、异方差条件以及非线性条件的时间序列分析方法的研究,在异方差条件下,美国的计量经济学家 Robert F. Engle 提出的自回归条件异方差模型(ARCH)是使用最为广泛的一类模型^[4]。

条件异方差模型是把时间序列的动态模型加以推广,用自回归过程来刻画扰动项的条件方差动态变化的特征^[5,6]。经济学家迈克尔·斯宾塞提出了信号传递模型,其本质就是一个动态不完全信息对策,即在信息不完全为决策制定者所掌握的情况下需要做出决策^[7]。文中的研究正是基于信号传递理论,在信息不全的情况下根据趋势信号对金融时间序列的趋势做出估计^[8],检测金融时间序列中的短期趋势信号并据此估计时间序列未来的趋势,再通过建立金融时间序列模型对所估计的时间序列值进行实验,证明其有效性。

收稿日期:2011-05-22;修回日期:2011-08-27

基金项目:云南省教育科学研究基金项目(07C10799)

作者简介:宫 正(1984-),男,山东烟台人,硕士研究生,研究方向为数据挖掘、信号处理;刘晓燕,副教授,硕士生导师,研究方向为模式识别、实时系统。

1 时间序列

1.1 时间序列定义

时间序列是按固定的时间间隔对数据采样,按照时间顺序依次排列的一组被观测数据或信息。人们常常以过去的历史数据为依据,预测未来的变化。

对某一个或一组变量 $x(t)$ 进行记录,将时刻 t_1, t_2, \dots, t_n (t 为自变量且 $t_1 < t_2 < \dots < t_n$) 所得到的数据组成的序列集合 $\{x(t_1), x(t_2), \dots, x(t_n)\}$, 称之为时间序列, 记为 $X = \{x(t_1), x(t_2), \dots, x(t_n)\}$ 。这种有时间意义的序列也被称为动态数据。

时间序列的定义并没有统一的表达方式, 尽管这些定义在表述上存在差异, 但基本含义是相同的, 时间序列就是一系列各种社会自然现象数量指标的统计数据, 按照均匀的实际间隔排列起来的序列。例如, 某地区每月的降雨量、股票市场中某只股票每天的收盘价, 以及河水流量等, 这些都是时间序列。

1.2 时间序列组成成分

时间序列由四部分构成, 即趋势变化、季节变化、循环变动、不规则变动。下面以金融时间序列为例讨论时间序列的组成成分。

(1) 趋势变化。趋势变化是时间序列随着时间的变化而逐渐增加或减少的变化过程。反映了时间序列一般的变化方向, 在时间序列图形上的显示是在较长的时间间隔上数据的上升或下降的变化, 其一般表现为一种趋势曲线。

(2) 季节变化。时间序列数据往往呈现出一种周期性, 把这种特性成为季节性。时间序列的季节变化是指在固定的一段时间内, 表现出的规律性变动。它反映了每隔一段固定的时间序列数据都会重复出现的变化规律。

(3) 循环变动。循环变动指的是趋势曲线呈现摆动的现象。循环变动既可以是周期性变动, 也可以不是周期性变动。在相同的间隔之内, 循环并不一定按照相同的模型演进。循环变动的周期一般较长, 变动的原因较多, 周期与幅度也都不一致。时间序列的循环变动通常是由多个小规模的时间序列循环组合而成。

(4) 不规则变动。不规则变动是指变量波动的不可预测性, 反映的不确定因素引起的变化。不规则变动是把趋势量、季节量和循环量等成分从时间序列中分离后, 剩下的随机波动的部分。在数据预测时, 应先除去不规则变动, 然后再进行预测。

2 短期趋势作为信号的模型

2.1 建模思想

信号传递模型在本质上是一个动态不完全信息对

策。这个对策包括两个参与人, 一个是发送者, 一个是接收者, 发送者一般来说具有一些接收者所不了解的, 与参与人的效用或者支付相关的信息。对策分为两个阶段: 第一个阶段, 发送者向接收者发出一个信息 (message), 或者叫一个信号 (signal); 第二个阶段, 接收者接收到信号后做出一个行动, 对策结束。这时, 两个参与人的效用就得到决定。他们的效用既是私人信息, 又是 message, 同时也是接收者所选择的行动的函数。注意, 第一阶段接收者只能看到发送者发出的信号, 而看不到发送者所拥有的私人信息。

在对时间序列的研究中, 时间序列同样会作为信号的发送者发出信号, 短期上涨趋势和短期下降趋势都是一种信号。我们便利用短期信号建立模型^[9]。对确定证券时间序列的分析的特点是认为数据去掉随机扰动外, 剩下的部分可以用确定时间函数表示。用 Y 表示时间序列, 则该时间序列可以分解成下面几个部分:

$$Y = f(T, C, S, e) \quad (1)$$

其中:

T : 趋势项, 时间序列逐渐增长或逐渐减少的变化。

C : 循环项, 时间超过一年的周期性波动。

S : 季节项, 一年内的周期性变化。

e : 随机项, 不可预测的偶然因素对时间序列的影响。

证券时间序列可能包括以上四个因素中的全部或者其中的几个。因为循环和季节因素都是在大于一年的周期中出现的, 所以文中对这两个因素不作考虑, 主要是针对趋势项和随机项进行建模。用公式表示为:

$$Y = f(T, e) \quad (2)$$

在实际的应用过程中, 公式(2)是不能使用的, 因为函数 f 没有具体的形式, 通常使用以下两类模型:

$$\text{加法模型: } Y = T + e \quad (3)$$

$$\text{乘法模型: } Y = T \times e \quad (4)$$

如果时间序列中的数据偏离趋势部分的大小随时间的改变而改变, 则用加法模型; 如果偏离趋势部分的大小随时间的改变而增加, 则使用乘法模型。增长趋势和下降趋势都是一个短期的趋势信号, 可以根据短期趋势信号对证券时间序列的影响建立模型^[10]。即根据识别出的增长趋势或下降趋势对证券时间序列的趋势进行拟合, 拟合的结果就是时间序列的线性趋势模型。

2.2 趋势模型

对证券时间序列的增长趋势和下降趋势模型的建立是使用时间 t 作为解释变量的回归模型, 线性趋势模型可以表示为:

$$y_i = c_0 + c_1 t \quad (5)$$

增长的数量是常数 c_0 , $t+1$ 比 t 时刻增加 c_1 。对未知参数 c_0, c_1 的估计可以使用最小二乘法来估计^[11,12]。

2.3 参数估计

随机选择 600630(龙头股份)在 2011-03-08 出现的上涨趋势的收盘数据如表 1 所示。

表 1 600630 上涨趋势收盘数据

| 时间 | 收盘价 | 时间 | 收盘价 | 时间 | 收盘价 |
|------------|------|------------|------|------------|------|
| 2011-03-04 | 8.44 | 2011-03-07 | 8.49 | 2011-03-08 | 8.61 |

分别设第一天、第二天、第三天的时间为 $t_1 = 1, t_2 = 2, t_3 = 3$ 。

使用最小二乘法对式(5)中的参数进行估计。最小二乘法的思路是使得因变量的估计值与实际的观察值之间的差异最小。用数学公式表示这一思想更明了。对 N 个观察值来讲,希望 N 个残差的和越小越好。因为残差可能为正也可能为负,直接把这 N 个残差的值相加,可能它们的和很小,但是残差的绝对值都很大,所以为了避免这种情况,把残差的平方求和:

$$\sum_{i=1}^N \hat{\mu}_i^2 = \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (6)$$

上式计算了残差的平方和,使得最小的参数估计是普通最小二乘估计量。求解下面的最优化问题:

$$\min_{\hat{\beta}} \sum_{i=1}^N \hat{\mu}_i^2 = \min_{\hat{\beta}} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (7)$$

可以解得最优化问题的解为:

$$\hat{\beta} = (T^T T)^{-1} T^T Y$$

把趋势点代入线性趋势模型可得:

$$\begin{cases} y_1 = c_0 + c_1 t_1 \\ y_2 = c_0 + c_1 t_2 \\ y_3 = c_0 + c_1 t_3 \end{cases} \quad (8)$$

用矩阵表示上面的方程组:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} \quad (9)$$

简化为

$$Y = T\beta$$

其中

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}, T = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ 1 & t_3 \end{bmatrix}, \beta = \begin{bmatrix} c_0 \\ c_1 \end{bmatrix}$$

用最小二乘估计量计算公式如下:

$$\hat{\beta} = (T^T T)^{-1} T^T Y \quad (10)$$

通过计算得到 $c_0 = 8.3433, c_1 = 0.0850$ 。

即线性趋势模型为:

$$y_i = 8.3433 + 0.085t \quad (11)$$

3 实验分析

根据参数估计中所介绍的方法对检测出的短期趋势建立模型,并预测 10 天的数据,对上证 A 股进行预测的部分结果如图 1 所示。

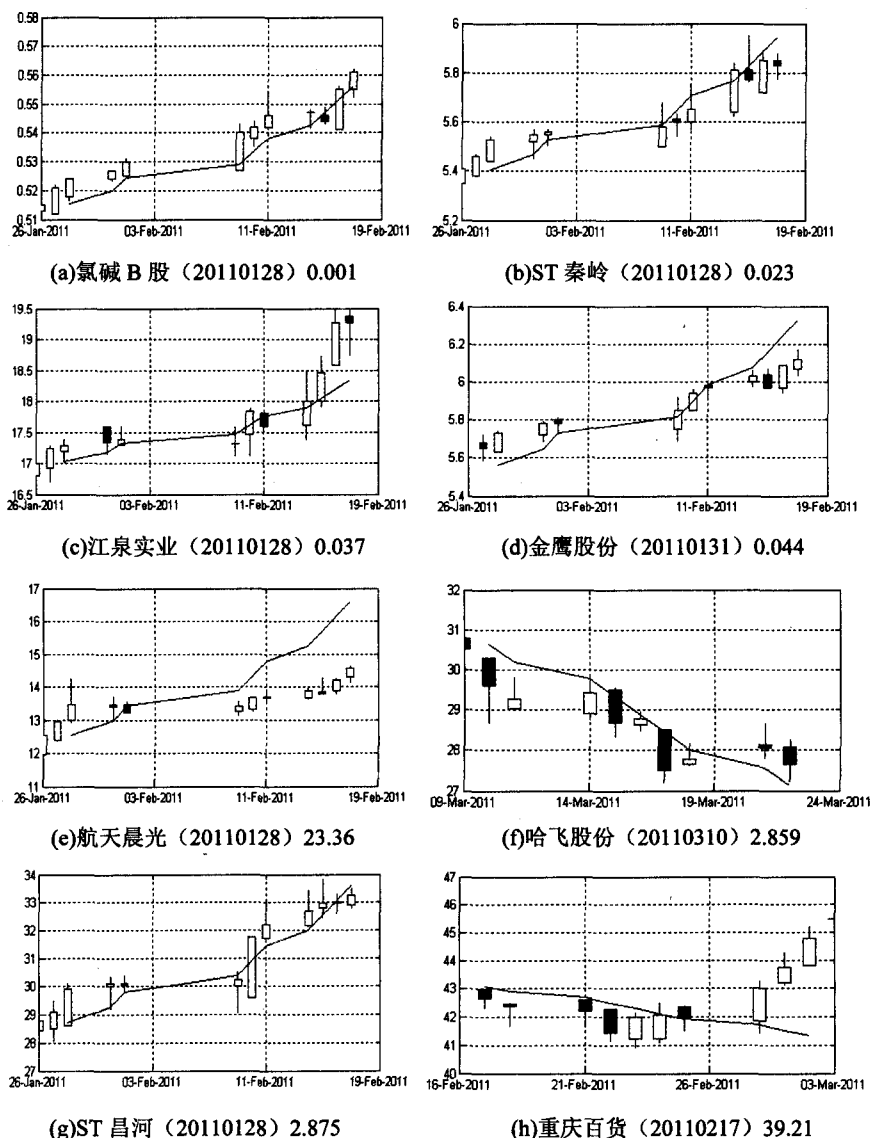


图 1 上证 A 股部分股票预测图

由实验结果显示方差小于 40 时,对趋势的估计基

本正确,具有较好的预测效果,统计检测出的 2203 个短期趋势点,并对其数据进行预测,计算预测值与真实值的差的平方统计结果如表 2 所示。

表 2 计算预测值与真实值的差的平方

| 方差范围 | 趋势数量 | 比例 |
|-----------|------|--------|
| [0,1] | 678 | 30.78% |
| [1,10] | 799 | 36.27% |
| [10,40] | 412 | 18.70% |
| [40,1228] | 314 | 14.25% |

从上表的统计结果可以看出,方差小于 40 的比例达到 85.25%,说明根据短期趋势对时间序列进行预测是一种可行的预测方法,对于误差较大的信息点,可以进一步地研究其趋势差异较大的原因。由于对时间序列的影响是各类信息的累积,因此根据趋势信号预测长期趋势的误差存在差异是合理的。对定义的不同类型的趋势进行检测结果如图 2~图 4 所示。

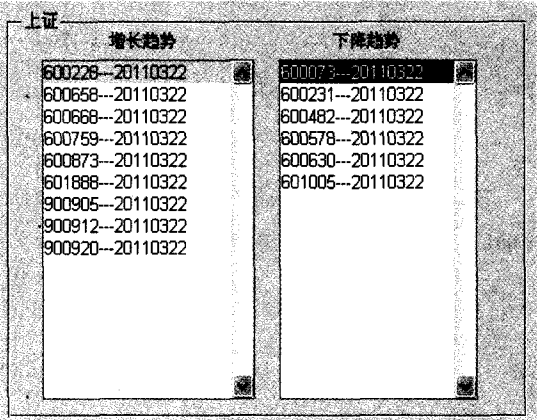


图 2 严格三级

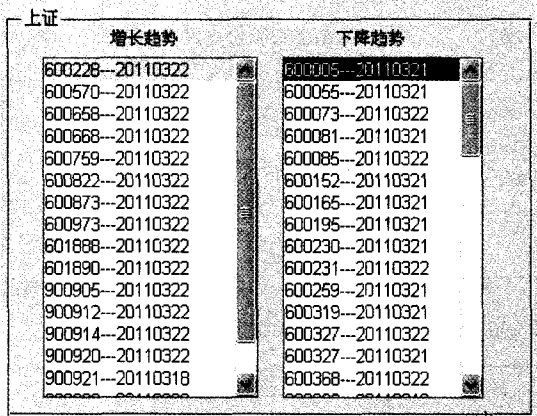


图 3 宽松三级

增长趋势是指证券价格止跌后逐渐递增的趋势。然而在实际的证券时间序列中很少存在严格的增长趋势,并且由于增长趋势之前的下跌的时间有长有短,因此对增长趋势进行分类,分别是严格增长趋势、宽松的增长趋势和区间增长趋势。下降趋势与增长趋势相反,是指证券价格停止上涨后逐渐下降的趋势。下降

趋势也分为三类,严格下降趋势、宽松下降趋势和区间下降趋势。从实验结果可以看出,严格、宽松和区间三种定义对增长和下降趋势的条件不同,因此检测出的结果不同。其中,严格趋势要求最为苛刻,必须满足收盘价上涨或下降的条件,宽松趋势要求次之,区间又宽松一些。实验表明这样分类定义是非常有必要的,如图 5 所示区间定义的趋势,虽然没有满足严格的上涨下降趋势,但检测出的信号有明显的趋势。如果没有区间趋势这种定义,将会忽略掉很多的有明显趋势的信号点。

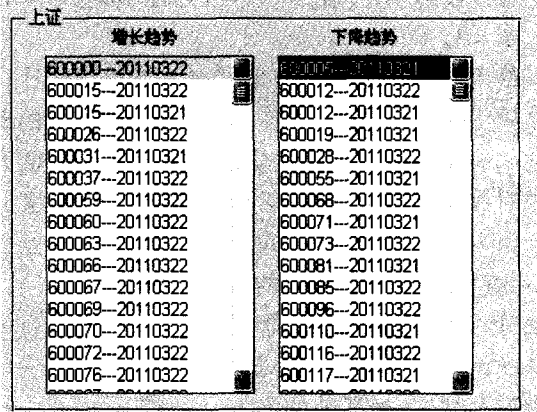


图 4 区间三级

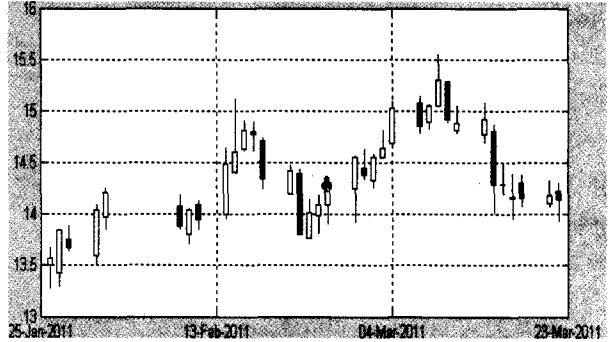


图 5 区间趋势显示图

4 结束语

文中主要是基于信息传递理论在时间序列中检测短期趋势来展开讨论的。把金融时间序列中出现的短期趋势分为增长趋势和下降趋势,又分别对增长趋势和下降趋势定义了严格、宽松和区间等不同的类型,并且根据历史数据定义了不同的强度,然后对趋势进行预测。并通过时间序列短时间内的增长或下降趋势提出了短期趋势估计方法,由此设计了短期趋势为信号的模型,用最小二乘法对所估计出的短期趋势建立趋势模型。最后做了短期趋势信号的模型的验证工作,实验表明所建立的模型是有效的。在未来的工作中,将会对短期趋势预测长期趋势的时间段作进一步的研究,进而构建长期趋势模型以及对长期趋势模型实证分析。

13-12-11-16-21-22-17-18-23-24-25-20-19-14-15-10-5-4-9-8-3-2-1-6-7。

三个算例的数值模拟实验结果如表 1 所示。

表 1 算例结果

| 算例序号 | 序列长度 | 最大迭代次数 | 最优适应度值 | 平均用时(s) |
|------|------|--------|--------|---------|
| 1 | 20 | 1000 | 12 | 18 |
| 2 | 36 | 2000 | 21 | 37 |
| 3 | 25 | 1000 | 8 | 13 |

结果表明,改进的算法具有较强的可行性和有效性,并且在时间上有了较大的优势。

4 结束语

文中提出了一种改进的 PSO 算法,并借此来求解基于二维 HP 格点模型的蛋白质结构预测问题,数值模拟实验表明,改进的算法具有较强的可行性和有效性,且在时间上有较大的优势。改进的 PSO 算法在解蛋白质结构预测问题时,初始解以及初始调整序的选择决定了下一代解的好坏以及能否很快收敛到最优解。因此,如何恰当地解决个体间的协作和个体经历最优及群体经历最优对个体的影响是解决这一问题的关键,也是进一步研究的问题。

参考文献:

- [1] 王菲露,宋杰,宋杨. BP 神经网络在蛋白质二级结构预测中的应用[J]. 计算机技术与发展,2009,19(5):217-218.
- [2] Luo L F. The time scale of protein folding simple model of chaperous[J]. Acta SNON,1994(25):52-56.
- [3] Dill K A, Bronnberg S, Yue K, et al. Principles of protein folding—a perspective from simple exact models[J]. Protein science,1995(4):561-602.

(上接第 108 页)

参考文献:

- [1] Franses P H. 商业和经济预测中的时间序列模型[M]. 北京:中国人民大学出版社,2002.
- [2] 张善文. Matlab 在时间序列分析中的应用[M]. 西安:西安电子科技大学出版社,2007.
- [3] de Gooijer J G, Hyndman R J. 25 years of time series forecasting[J]. International Journal of Forecasting,2006,22:443-473.
- [4] 于俊年. 计量经济学[M]. 第 2 版. 北京:对外经贸大学出版社,2007.
- [5] Kumar K, Jain V K. Autoregressive integrated moving averages (ARIMA) modeling of a traffic noise time series[J]. Applied Acoustics,1999,58:283-294.
- [6] Tseng Fang-Mei, Yu Hsiao-Cheng, Tzeng Gwo-Hsiung. Combining neural network model with seasonal time series ARIMA

- [4] Kennedy J, Eberhart R C. Particle Swarm Optimization[C]//Proceedings of IEEE International Conference on Neural Networks. Piscataway:IEEE,1995:1942-1948.
- [5] Shi Y, Eberhart R C. Empirical study of particle swarm optimization[C]//Proceedings of the IEEE Congress on Evolutionary Computation. Piscataway:IEEE,1999:1945-1990.
- [6] 唐俊. PSO 算法原理及应用[J]. 计算机技术与发展,2010,20(2):215-216.
- [7] 李绍新,张延娇. 改进的遗传算法在蛋白质结构预测中的应用[J]. 华南师范大学学报(自然科学版),2009(1):56-60.
- [8] Guo Yuzhen, Feng Enmin, Wang Yong. Exploration of two-dimensional HP lattice model by combining local search with elastic net algorithm[J]. The Journal of Chemical Physics,2006,125(15):154102.
- [9] Guo Yuzhen, Feng Enmin, Zhao Jingcheng, et al. Optimal HP Configurations of Protein by Combining Local Search with Elastic Net Algorithm[J]. Journal of Biochemical and Biophysical Method,2007,70(3):335-340.
- [10] 王翠茹,张江维. 改进粒子群优化算法求解旅行商问题[J]. 华北电力大学学报,2005,32(6):47-51.
- [11] Hart W E, Istrail S. Robust proofs of NP-hardness for protein folding general lattices and energy potentials[J]. Journal of Computational Biology,1997,4(1):1-22.
- [12] 刘任任. 算法设计与分析[M]. 武汉:武汉理工大学出版社,2003:121-124.
- [13] Wang K P, Huang L, Zhou C G, et al. Particle swarm optimization for traveling salesman problem[C]//The 2nd International Conference on Machine Learning and Cybernetics. Xi'an: [s. n.],2003:1583-1586.
- [14] Yanikoglu B, Erman B. Minimum energy configurations of the 2-dimensional HP-model of proteins by self-organizing networks[J]. Journal of Computation Biology,2002,9(4):613-620.

model[J]. Technological Forecasting & Social Change,2002,69:71-87.

- [7] Spence M. Job Market Signaling[J]. Quarterly Journal of Economics,1973,87(3):355-374.
- [8] 王琼,刘国祥. 金融时间序列的趋势路径的提取[J]. 南京师大学报(自然科学版),2002,25(2):105-109.
- [9] 刘小君,张立臣. 基于 UML 的实时系统开发[J]. 微机发展(现更名:计算机技术与发展),2003,13(5):81-83.
- [10] 赵纪涛,马莉,王现君. 一种自适应的模糊关联规则挖掘算法[J]. 计算机技术与发展,2008,18(5):64-66.
- [11] 辛治运,顾明. 基于最小二乘支持向量机的复杂金融时间序列预测[J]. 清华大学学报,2008,48(7):1147-1149.
- [12] 但志平,郑胜. 最小二乘向量机在说话人识别中的应用[J]. 计算机技术与发展,2007,17(5):30-32.