

基于 ARIMA - SVM 组合模型的股票价格预测

程昌品, 陈 强, 姜永生

(广东第二师范学院计算机科学系, 广东 广州 510303)

摘要: 现有的股票价格准确预测方法各有优缺点, 为了发挥各种预测方法的优点, 提出二进正交小波变换和 ARIMA - SVM 方法的非平稳时间序列预测方案。使用小波分解算法对数据进行分解, 分离出非平稳时间序列中的低频信息和高频信息; 然后对高频信息构建自回归模型 ARIMA 预测, 对低频信息则用 SVM 模型进行拟合; 最后将各模型的预测结果进行叠加, 从而得到原始时间序列的预测值。将预测结果与实际值比较, 组合模型具有较好的预测效果。经实验证明, 小波分解的 ARIMA - SVM 组合模型较单一的预测模型效果更为理想。

关键词: 小波变换; 非平稳时间序列; 支持向量机组合模型; 预测

中图分类号: TP391 **文献标识码:** B

Research on Stock Price Prediction Based on Wavelet Decomposition and ARIMA - SVM Combined Model

CHENG Chang - pin, CHEN Qiang, JIANG Yong - sheng

(Guangdong Institute of Education, Guangzhou Guangdong 510303, China)

ABSTRACT: Existing stock forecasting methods have advantages and disadvantages. In order to make full uses of the advantages of stock forecasting methods, a non-stationary time series prediction method based on wavelet transform and ARIMA - SVM Combined Model was proposed. By wavelet decomposition and reconstruction, the non-stationary time series were decomposed into a low frequency signal and several high frequency signals. The high frequency signals were predicted with auto-regression models ARIMA, and the low frequency was predicted with SVM. The prediction result of the original time series was the superimposition of the respective prediction. The method in this paper is better than the traditional methods, and the obtained result in this paper is close to the actual value.

KEYWORDS: Wavelet decomposition; Non-stationary time series; ARIMA - SVM combined model; Prediction

1 引言

股票价格变化是一种多自变量的时间序列数据, 由于受公司经营状况、财务状况、宏观经济状况和制度政策等多种因素影响, 具有高度的非线性、时变性, 是一种高度不稳定、复杂且难以预测的时间序列数据。目前股票预测的方法很多, 比较常用的方法有回归分析^[1]、时间序列分析^[2]等, 如自回归移动平均 ARIMA、神经网络模型(ANN)及支持向量机(SVM)等。这些预测方法在一定程度上揭示了股市的运行规律, 但是股价系统内部结构的复杂性、外部因素的多变性限制了股市预测的准确性, 现有的分析预测方法应用效果并不理想。如 ARIMA 模型, 贝叶斯网络预测精度高, 技术成

熟, 但是建模方法繁琐; 人工智能和神经网络适合非线性关系的拟合, 但容易陷入局部极小点, 收敛速度慢; 支持向量机适合小样本的快速预测, 能得到全局最优, 但预测效果对核函数非常敏感。

为了发挥各种预测方法的优点, 人们提出了组合预测模型, 就是将参与组合的各种预测方法通过适当的组合, 在组合预测模型中发挥参与组合的各种预测方法的优势。目前通过组合模型进行股市预测主要有: ARIMA 模型与神经网络组合预测^[3], 小波包与神经网络组合预测^[4], 神经网络与支持向量机的组合预测等^[5]。

在参考已有文献的基础上, 本文提出了基于小波分析的组合模型, 用 ARIMA 模型捕捉股票价格的线性规律, 用 SVM 模型捕捉股票价格的非线性规律, 从而创新性地用基于小波分解的 ARIMA - SVM 组合模型^[7]对股票收盘价格这一典型的非平稳时间序列进行预测。根据小波分析理论, 可以将信

基金项目: 广东省科技计划资助项目(2010B010600018); 2011 年广东省现代信息服务业发展专项资金项目(13090)

收稿日期: 2011-06-30 修回日期: 2011-08-08

号分解到不同的频率通道上,分解后的信号在频率成分上比原始信号单一,小波分解后对信号作平滑处理,然后重构分解信号,这样非平稳时间序列经过处理后,成为近似意义上的平稳时间序列,这样就得到了原始数据的近似信号,然后用自回归移动平均 (ARIMA) 和支持向量机 (SVM) 的组合预测模型对这些信号预测,将预测结果与实际值进行比较,用实例验证组合模型的预测有效性。

2 预测原理

2.1 二进正交小波变换的 Mallat 算法^[8]

2.1.1 分解算法

若 c_0 为待分解的离散信号,则根据分解算法有:

$$c_{j+1} = Hc_j, d_{j+1} = Gd_j, j = 0, 1, \dots, J \quad (1)$$

其中 H 和 G 分别为低通滤波器和高通滤波器; c_{j+1} 和 d_{j+1} 分别为原始信号在分辨率 $2^{-(j+1)}$ 下的低频信号和高频信号,是原始信号在相邻不同频率段上的成分; J 表示最大的分解层数。最终将待分解信号 c_0 分解为 d_1, d_2, \dots, d_J 和 c_J 。

该分解算法利用二抽取,使每层分解比分解前的信号数据长度减半,而总输出数据长度与输入数据 c_0 长度保持一致。经 Mallat 算法分解后的信号可进行二插值重构。

2.1.2 重构算法

$$C_j = H^* C_{j+1} + G^* D_{j+1}, j = J-1, J-2, \dots, 0 \quad (2)$$

其中 H^* 和 G^* 分别是 H 和 G 的对偶算子,采用式 (2) 对小波分解后的信号进行重构可以增加信号个数。

对 d_1, d_2, \dots, d_J 和 c_J 分别进行重构,得到 D_1, D_2, \dots, D_J 和 C_J ,有

$$X = D_1 + D_2 + \dots + D_J + C_J \quad (3)$$

其中 D_1, D_2, \dots, D_J 为第 1 层至第 J 层的高频信号重构, C_J 为第 J 层低频信号重构。 $D_i = \{d_{i1}, d_{i2}, \dots, d_{iN}\}$, $C_j = \{c_{j1}, c_{j2}, \dots, c_{jN}\}$, $X_j = \{x_{j1}, x_{j2}, \dots, x_{jN}\}$ 为第 J 层逼近信号的重构结果,因此:

$$X_{ji} = d_{1i} + d_{2i} + \dots + d_{ji} + c_{ji} \quad (4)$$

2.2 ARIMA - SVM 预测模型

2.2.1 ARIMA - SVM 原理

ARIMA 被认为是一种精确度较高的线性时间序列预测模型,而支持向量机 SVM 则是一种性能优异的非线性分析模型,被视为替代人工神经网络的较好算法。

由于股票价格具有复杂的非线性动力系统特征,单纯使用 SVM 或 ARMA 模型进行预测都有可能误差过大。因此,首先利用 Mallat 算法,对非平稳时间序列进行分解和重构;然后对低频和高频信息分别建模,由 (3) 知 D_1, D_2, \dots, D_J 为高频信号重构,可近似看作平稳时间序列,可建立 ARIMA 模型预测股票价格历史数据;由于低频信号表示长期趋势,且具有非线性,可采用 SVM 模型对其进行建模和预测,使非线性规律包含在 SVM 的预测结果中。原始序列的预测结果为低频和高频预测结果的叠加预测框架如图 1 所示。

2.2.2 基于小波分解与重构的 ARIMA - SVM 预测模型

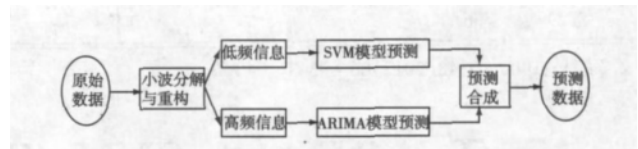


图 1 基于 Mallat 算法的 ARIMA - SVM 组合模型预测框架

算法

Step1: 把一组时间序列的数据 γ_i 通过小波变换的 Mallat 算法分解和重构后得到数据 $\{X_{ji}\}$, 分为线性分解部分 D_{ji} 和非线性分解部分 c_{ji} , 如式 (4) 所示, 最后得到

$$X_{ji} = D_{ji} + c_{ji}$$

Step2: 用 ARIMA 模型对 D_{ji} 进行预测。设预测结果为 D_{ji}^*

Step3: 利用 SVM 模型对 c_{ji} 进行预测。设预测结果为 c_{ji}^*

Step4: 利用两种模型的预测结果组合成为最终的预测结果 $\hat{\gamma}_i$, 结果为 $\hat{\gamma}_i = D_{ji}^* + C_i^*$

2.3 参比模型及评价指标

为了比较 ARIMA - SVM 模型的有效性,选择 SVM, AR-MA 为参比模型,所有模型均采用一步预测法。ARIMA 由 DPS6.55 得到, SVM 由 MATLAB 7.0 通过调用 SVM 工具箱实现。为了评价模型预测性能的优劣,使用均方根误差 (Root Mean Square Error, RMSE) 作为模型的评价指标。RMSE 定义如下:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\gamma_i - \hat{\gamma}_i)^2} \quad (5)$$

其中 γ_i 为实际值, $\hat{\gamma}_i$ 为预测值, n 为预测样本数。RMSE 适用于同一数据集不同模型间的比较。

3 股票价格预测仿真

3.1 数据来源

本文选用四川圣达股票 2005 年 11 月 7 号到 2006 年 12 月 12 号的每交易日的收盘价格作为研究对象,共 250 个样本进行建模分析,前 230 个样本作为训练样本进行算法模型的构建,后 20 个样本作为测试样本,对未来 20 天的股票收盘价格进行预测并与实际数据进行对比,图 2 为四川圣达股票前 230 个样本收盘价格变化图,是非常典型的非平稳时间序列。

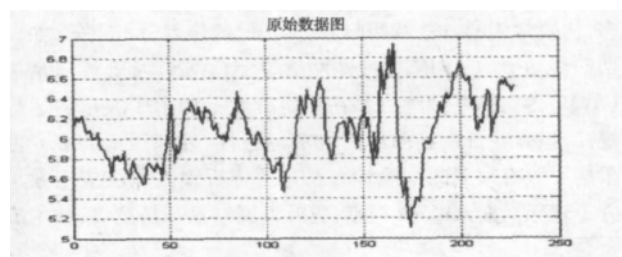


图 2 四川圣达股票 2005 ~ 2006 年 230 个交易日股价走势图

3.2 组合预测方法实现

3.2.1 数据的小波分析

由图 2 可知,该曲线是一个具有趋势性的非平稳时间序列过程,必须进行平稳化处理,对上述数据用 db4 小波进行分解,分解水平为 3,并对分解后的数据作平滑处理,然后重构,得到原数据的近似信号,这时数据走势如图 3 所示。

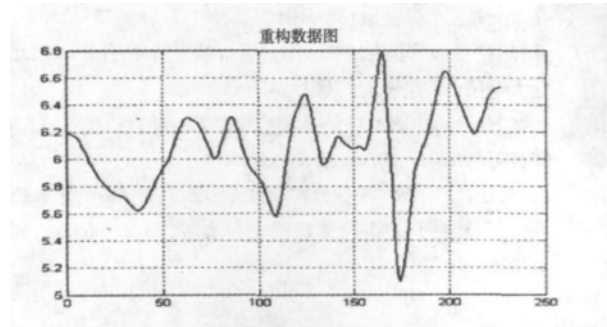


图 3 小波重构后股价走势图

通过比较可以看出进行小波变换后,时间序列趋于平稳。重构后得到线性部分 d 序列和非线性部分 c 序列。

3.2.2 线性部分的 ARIMA 预测

对线性部分 d 采用 ARIMA 模型进行预测,ARIMA 的预测步骤如下:

Step1: 线性部分 d 需要对其进行平稳化处理。对数据序列进行 1、2、3 阶差分,发现 3 阶差分后数据已经基本平稳化,所以设定 ARIMA 模型参数 $d=3$ 。

Step2: 借助 DP56.55 构建 ARIMA 模型,采用从低阶到高阶逐步试探法来识别模型的类型和阶数,经过比较分析,发现选择 $P=2, q=1$ 时,ARIMA 模型的预测效果较好,预测样本的 RMSE 为 51.3。

Step3: 利用已经确定的最优参数得到最优模型 ARIMA (3 2 1),利用最优模型对最后 20 个样本进行预测,预测结果待用。

3.2.3 非线性部分的 SVM 预测

1) 模型的定阶

股票价格变化是一种多维时间序列,股票价格不仅与当各种影响因子有关,还与上一日甚至前几日股票价格有关,有必要把前几天的股票价格作为当天股票价格变化的影响因子来考虑,到底前多少日对当前股票价格变化有影响,可通过模型的阶数来确定。

设一个股票价格数据序列有 N 个样本、一个因变量、 $m-1$ 个自变量,用 SVM 由低阶到高阶逐步进行非线性拓阶,模型每拓一阶,自变量相应增加 m 个,每次拓阶之后进行交叉验证测试,并以测试结果均方误差是否最小为标准决定拓展与否:假设 $SVM(n)$ 为拓阶 n 次后的模型, $SVM(n+1)$ 为拓阶 $n+1$ 次后的模型,比较两模型的 RMSE 大小,若 $RMSE_{svm}(n) > RMSE_{svm}(n+1)$, x 继续拓阶,反之, $RMSE_{svm}(n) \leq RMSE_{svm}(n+1)$,拓阶终止,取 $SVM(n)$ 为定阶后模

型。最后取阶数为 4。

2) 序列的重构及预测

通过模型阶数的确定,得到模型的阶数为 $n=4$,说明当天股票的价格与前 4 天的股票价格有着紧密联系,这就意味着将前 4 天股票价格的残差作为 SVM 的输入来预测当天股票价格的残差。在 MATLAB7.0 平台下自编调用 LIBSVM 工具箱来实现 SVM 建模,核函数为高斯核函数,采用 5 折交叉验证,经 gridre-gression.py 自动搜索确定模型最优参数,使用最优参数下训练得到的模型对图 3 非线性部分 c 序列进行预测。

3.3 仿真分析

根据 ARIMA 模型得到的线性预测结果和 SVM 模型的非线性预测结果,进行简单的相加得到 ARIMA-SVM 模型的预测结果,如图 4 所示,图中用“*”画出的曲线代表真实数据,用“o”画出的曲线代表预测值,图中包含 230 个训练样本,20 个测试样本。

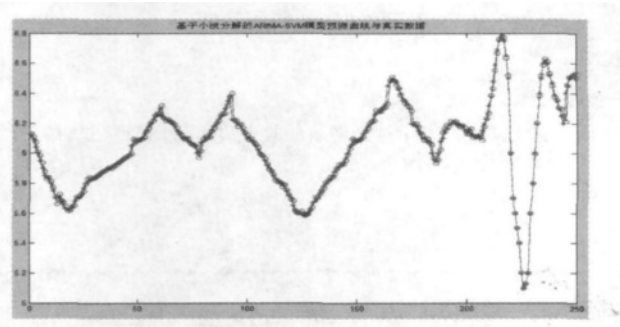


图 4 四川圣达股票价格预测图

从图 4 可以看到,预测的总体趋势和实际股票价格的走势一致,预测效果较好。表 1 列举了 9 个测试样本的预测数据,表 2 则给出了三种模型的 RMSE 比较。

表 1 三种模型预测结果对比表

日期	实际值	ARIMA	SVM	组合模型
2006-11-30	6.44	6.54	6.51	6.48
2006-12-1	6.47	6.58	6.54	6.52
2006-12-4	6.38	6.45	6.42	6.41
2006-12-5	6.67	6.83	6.74	6.71
2006-12-6	6.39	6.53	6.46	6.43
2006-12-7	6.27	6.39	6.33	6.32
2006-12-8	6.08	6.21	6.15	6.13
2006-12-11	6.14	6.27	6.21	6.18
2006-12-12	6.11	6.25	6.18	6.15

表 2 三种模型的 RMSE 对比表

模型	RMSE
ARIMA	51.3
SVM	21.43
组合模型	7.58

观察图 4 及表 1、表 2 发现:

1) 小波分解和重构对股票价格进行平稳化处理,使股票价格的价格本质及变化规律能够很好反映出来,说明在进行股票价格预测前,对其进行小波分析是必要的。

2) 组合模型集成 SVM 和 ARIMA 的优点,达到了互补的效果,其非线性逼近能力明显优于单一模型,得出的预测结果更为精确。

4 结论

由于股市的不稳定性及金融市场的动荡,目前传统算法的预测准确率相当低。本文在现有预测模型的基础上提出基于小波分析的组合模型,用 ARIMA 模型捕捉股票价格的线性规律,用 SVM 预测股票价格的非线性规律,并对四川圣达股票进行仿真预测,从预测结果可知:该组合模型能刻画各种随机因素的影响,对股票价格的非线性部分和线性部分采用不同的预测模型,比单一模型的预测更精确、更可靠,是一种有效的股票价格时间序列预测模型。

参考文献:

- [1] 赵建,黄炯. 基于神经网络的股市预测[J]. 计算机研究与发展,1996,33(9):692-69.

- [2] 邓凯,赵振勇. 基于遗传 BP 网络的股市预测模型研究与仿真[J]. 计算机仿真,2009,26(5):316-319.
- [3] 甘昕艳,张钰玲. 基于股价指数预测的仿真研究[J]. 计算机仿真,2010,27(10):297-300.
- [4] 陈振伟,郭拯危. 小波神经网络预测模型的仿真实现[J]. 计算机仿真,2008,25(6):147-150.
- [5] 张玉川,张作泉. 支持向量机在股票价格预测中的应用[J]. 北京交通大学学报(自然科学版),2007,31(6):73-76.
- [6] 孔祥凤. 小波分析在股市数据分析中的应用研究[D]. 西北工业大学硕士学位论文,2002-6.
- [7] 吴虹,尹华. ARIMA 与 SVM 组合模型的石油价格预测[J]. 计算机仿真,2010,27(5):264-266.
- [8] 张华,任若恩. 基于小波分解和残差 GM(1,1)AR 的非平稳时间序列预测[J]. 系统工程理论与实践,2010,30(6):1017-1018.

[作者简介]



程昌品(1963-),男(汉族),湖北监利人,副教授,主要从事软件工程及数据库系统方向应用研究。

陈强(1963-),男(汉族),广东化州人,研究员,主要从事知识管理与知识发现应用研究。

姜永生(1966-),男(汉族),新疆吐鲁番人,讲师,主要从事计算机应用研究。

(上接第 338 页)

其中,横坐标是汽车刹车时的初始速度,纵坐标是汽车在刹车过程中的抱死系数,用来衡量抱死的可能性大小。根据图 5 能够得知,利用本文算法得到的刹车抱死系数要小于传统的方法,也就说明本文方法的刹车抱死概率要低于传统方法。其相关数据用表 1 描述。

表 1 实验数据对比

	传统算法	本文算法
刹车初始车速	100 公里/小时	100 公里/小时
抱死系数	9.7	7.4

根据表 1 能够得知,利用本文算法的刹车抱死系数远低于传统方法,大幅降低刹车抱死的可能,提高了汽车的安全性能。

5 结束语

本文提出了一种基于模糊系数动态调整技术的防刹车抱死系统,利用模糊控制规则,获取汽车刹车设备之间的关系,获取刹车相关参数之间的关系,然后利用控制系数隶属度函数,获取路况情况,从而动态调整刹车系统相关参数设置,大幅降低刹车抱死可能。实验证明,利用本文算法能够降低抱死可能,有很大的应用空间和商业价值。

参考文献:

- [1] 申荣卫. 汽车电子技术[M]. 北京:机械工业出版社,2002-11.
- [2] 齐晓杰. 防抱死制动和牵引力控制系统[M]. 北京:化学工业出版社,2005.
- [3] 张新. 汽车液压防抱制动系统(ABS)的理论与实践[M]. 长沙:中南大学出版社,2005.
- [4] 宋明,刘昭度,梁鹏霄. 汽车制动防抱死系统的 ADAMS/SIMULINK 联合仿真[J]. 计算机仿真,2004-10:163-166.
- [5] 吕立亚,朱伟兴. 模糊智能控制在 ABS 中的应用[J]. 计算机测量与控制,2003-8.
- [6] S Jung, S Kim. Improvement of scanning accuracy of PZT piezoelectric actuators by feed-forward model-reference control[J]. Precision Engineering, 1994, 16(1):49-55.



[作者简介]

陈靖宇(1973-),男(汉族),广东兴宁人,实验师,主要研究方向:移动计算与物联网技术、计算机网络与网络安全。