

基于小波变换域的 SVM 股市 时间序列预测算法

杨 稣^{1,2} 史耀媛¹ 宋 恒³

(西安电子科技大学¹, 西安 710075; 西北工业大学², 西安 710072; 海军航空工程学院³, 烟台 264001)

摘 要 研究利用小波变换技术提高基于 SVM 的股市时间序列预测算法的效果。采用结构相同的若干 SVM 同步预测股市时间序列数据在不同尺度下的小波变换系数, 通过对各预测值进行加权组合预测股市变化趋势。其中所有 SVM 的核参数采用遗传算法同时自动进化调整。通过实证分析, 以及同基于 SVM 的股市时间序列预测算法进行对比, 结果证明结合小波变换技术能够更深入揭示股市规律。

关键词 股市 时间序列预测 支持向量机 小波变换
中图法分类号 TP181 文献标志码 A

股市是一个自适应复杂系统, 股市中的各种因素交织在一起, 使得股价时序变得复杂, 因此难以预测。因此, 预测这样复杂的股票市场要有特殊的方法^[1-3]。Peter^[4]等指出金融市场包括股票市场是由不同投资时间水平的交易者组成, 如短期、中期和长期交易者等等。不同类型的交易者以不同的时间尺度来看待和影响市场, 如短期交易者只关注市场短期的价格变化, 其行为所引起的价格波动只具有短期的记忆; 而长期交易者只关注市场长期范围内的价格变化, 其行为所引起的价格波动具有长期的记忆。由于每个个体都具有趋利避害的趋势, 各有自己的投资理念、受影响因素及投资策略也各不相同, 结果引起的股票价格波动特征也完全不同, 分散反映在相应的不同时间尺度上。这也是造成股票市场是混沌的原因。

小波变换是 20 世纪 80 年代后期发展起来的应用数学分支, 是泛函分析、Fourier 分析、样条分析、调和分析、数值分析的结晶。小波变换既是一种信号的时间尺度分析, 同时又是一种信号的时间频

率分析, 具有优良的多分辨率分析特性^[5,6]。它的最大的优点是能将时间序列按不同尺度分解成不同的层次, 展现不同层次的信息特征, 这一方面使问题变得简单, 便于分析和预测; 另一方面使正常信号中夹带的瞬态反常现象得到探测, 并使其成份得到展现。小波变换的这些特点正是提取股市时间序列变化特征所需要的。

利用小波变换能够检测信号的奇异性和自相似的特性, 在文献[7]研究的基础上, 提出基于小波变换域的 SVM 股市时间序列预测算法。通过实证分析, 并与文献[7]的方法作对比, 证明基于小波变换域的 SVM 股市分析算法的确能够更深入揭示股市规律。

1 基于 SVM 的股市时间序列预测算法设计

1.1 算法结构

利用小波变换具有多分辨率的特点, 首先将股市时间序列数据进行小波变换, 求出不同尺度下的小波变换系数; 然后分别采用几个结构相同的 SVM 同时训练并预测各尺度下的小波系数; 最后对各个 SVM 的预测值进行加权组合, 得到股市时间序列数据的预测值。其中所有 SVM 的核参数采用遗传算法同时自动进化调整。基于小波变换域的 SVM 股

2008 年 3 月 11 日收到 国家自然科学基金 (70602089) 资助

第一作者简介: 杨 稣, (1971—) 男, 四川合川人, 博士, 西安电子科技大学副教授, 西北工业大学博士后研究人员, 研究方向: 系统工程、区域经济学。Email: yangsu2008@msn.com

(C)1994-2020 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

市时间序列预测算法的结构如下图 1 所示。

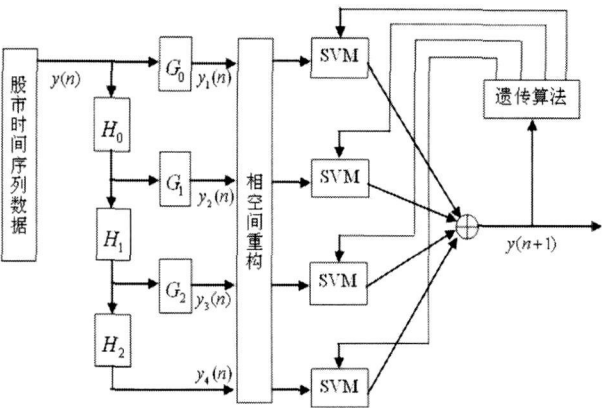


图 1 基于小波变换域的 SVM 股市时间序列预测算法结构图

设获取的股市时间序列数据为 $\{y(n)\}$ ，对其进行三级小波变换，将所得各级小波变换系数分别嵌入重构相空间 R^{D_E} 中（嵌入时间延迟 τ ，嵌入维数 D_E ）得：

$$\begin{aligned} \varphi_R^m(n) &= [y_m(n), y_m(n-\tau), \dots, \\ &y_m(n-(D_E-1)\tau)]; m=1, 2, 3, 4 \end{aligned} \quad (1)$$

利用重构的状态矢量对各尺度小波系数进行预测，即构造映射（回归估计函数） $f_m: R^{D_E} \rightarrow R$ ，使得

$$y_m(n+1) = f_m(\varphi_R^m(n)); m=1, 2, 3, 4 \quad (2)$$

设当前数据时刻为 n ，训练数据容量为 N （训练数据容量是可增加的），则训练数据可表示为

$$\begin{aligned} &(\varphi_R^m(i), y_m(i+1))_{i=n-N+1}^{i=n} \\ &m=1, 2, 3, 4; i=D_E, D_E+1, \dots \end{aligned} \quad (3)$$

采用 SVM 方法构造回归估计函数 f_m ，将 D_E 维的输入空间变换到另一个高维特征空间，并在该特征空间中根据结构风险最小化原理，综合考虑函数复杂度和回归误差，求取最优回归曲线。所有的 SVM 结构及算法参阅文献 [7]。

所有 SVM 并行计算，得到股市数据的各尺度小波变换系数的回归估计为

$$\begin{aligned} y_m(n+1) &= f_m(\varphi_R^m(n)) = \\ &\sum_{i=n-N+1}^n (\alpha_i^* - \alpha_i) k(\varphi_R^m(i), \varphi_R^m(n)) + b(n) \end{aligned} \quad (4)$$

(4) 式中 $\alpha_i^{(*)} - \alpha_i$ 为最优的 Lagrange 乘子， $b(n)$ 为阈值。

将所有 SVM 的输出 $y_m(n+1)$ 进行加权组合，得到对股市时间序列的预测值

$$y(n+1) = \sum_{m=1}^4 \mu_m y_m(n+1) \quad (5)$$

(5) 式中， μ_m 为加权系数。

1.2 SVM 核参数的自适应调节算法

采用高斯核函数

$$k(\varphi_R^m(i), \varphi_R^m(n)) = \exp\left\{-\frac{\|\varphi_R^m(i) - \varphi_R^m(n)\|^2}{2\sigma_m^2}\right\} \quad (6)$$

(6) 式中， $\|\cdot\|$ 表示欧氏距离， σ_m 表示核宽度。采用遗传算法进化同时调整所有 SVM 的核参数 σ_m 。

关于遗传算法的基本原理及调节 SVM 的核参数的方法文献 [7] 已经详细阐述过。因为基于小波变换域的 SVM 股市时间序列预测算法中有多个 SVM 并行工作，现在设计用遗传算法同时调整所有 SVM 的核参数 σ_m ，具体过程如下。

将 SVM 的核参数 σ_m 进行编码作为个体，具体编码方式：

- (1) 将核参数 σ_1 值的范围选定为 0.00 ~ 10.23 共 1024 个个体，采用 10 位二进制编码，见表 1；
- (2) 将 $\sigma_m - \sigma_1$ ($m=2, 3, 4$) 值的范围均选定为 0.00 ~ 1.27 共 128 个个体，采用 7 位二进制编码，见表 2；
- (3) 将 σ_1 和 $\sigma_m - \sigma_1$ ($m=2, 3, 4$) 的编码串连起来，得到所有 SVM 的核参数的 31 位二进制编码。

表 1 核参数编码 (1)

核参数 σ_1	二进制编码
0.01	0000000000
0.02	0000000001
\vdots	\vdots
10.22	1111111110
10.23	1111111111

表 2 核参数编码 (2)

核参数 $\sigma_m - \sigma_1$	二进制编码
0.01	0000000
0.02	0000001
\vdots	\vdots
1.26	1111110
1.27	1111111

在核参数 σ_m 的初始值附近, 随机生成 25 个样本 (样本为 31 位二进制码), 作为初始群体。

综合考虑支持向量数量和回归误差, 设计的适应度函数定义如下。

$$f_d = \eta_1 \frac{N-1}{N} + \eta_2 (1 - |y(n+1) - \hat{y}(n+1)|)$$

(7)

(7) 式中, $y(n)$ 表示真实值, N 表示训练样本总数, N 表示支持向量数量, η_1 和 η_2 表示权系数。

采用轮盘赌选择算子、“与/或”交叉算子、基本位变异算子^[8,9]进行遗传操作。设定适应度大于预定值 θ_{ad} 或进化代数大于 100 时, 结束进化。

2 实证分析

为了便于与文献[7]相比较, 仍然选取上证指数和深圳成指两种大盘指数, 选取 G 五粮液等 5 种个股指数进行实证分析 (从 <http://www.stockstar.com> 下载)。

将上证指数 1997 年 1 月 2 日 ~ 2001 年 4 月 16 日共 1030 个交易日的收盘价, 深圳成指 1997 年 1 月 2 日 ~ 2001 年 4 月 12 日共 1030 个交易日的收盘价, 五粮液等 5 种个股 1998 年 4 月 27 日 ~ 2002 年 8 月 20 日共 1030 个交易日的收盘价, 进行指数平滑 (取指数平滑因子 $\eta=5$) 后, 作为股指数据。采用具有紧支撑和对称性的 Coiflet 小波系中的 Coif5^[10] 对信号进行离散小波变换。根据 Takens 定理^[11], 对各级小波变换系数进行相空间重构 (对上证指数和深圳成指, 嵌入时间延迟 $\tau=3$ 和估计嵌入维 $D_E=4$; 对 G 五粮液等个股, 嵌入时间延迟 $\tau=2$ 和估计嵌入维 $D_E=4$)。在重构相空间中, 取前 1 000 个交易日数据作为训练数据, 同时训练所有 SVM (训练过程中遗传算法自动同时调整所有核参数 σ_m , 加权系数 $\psi_m=0.25$, $m=1, 2, 3, 4$)。将后 30 个交易日作为预测区, 进行单步预测实验, 将结果同文献[7]的结果相比较, 单步预测的预测误差比较见表 3。

表 3 股市时间序列预测算法比较

股指种类	股市时间序列预测算法	
	SVM	结合小波变换域的 SVM
上证指数 000001	0.198 9	0.092 0
深圳成指 399001	0.201 2	0.123 2
G 五粮液 000858	0.299 1	0.152 1
G 长虹 600839	0.339 2	0.197 8
上海石化 600688	0.397 1	0.242 1
乐山电力 600644	0.345 8	0.193 8
茂化实华 000637	0.367 8	0.210 1

从表 3 中可以明显看出, 提出的基于小波变换域的 SVM 的股市时间序列预测算法提高了预测精度。这正是因为股市比纯粹的混沌系统更复杂, 小波变换将股市信号在不同的分辨率下展示, 使得 SVM 能够更准确的辨识股市系统特征的结果。

3 结论

利用小波变换多分辨率分析的优点, 本文提出了基于小波变换域的 SVM 股市时间序列预测算法。通过实证分析, 与文献[7]进行对比, 证明结合小波变换技术的确有助于深入揭示股市规律。混沌理论是股市预测的理论基础, 更进一步, 在利用小波变换等技术研究哪些因素对股市混沌行为起着关键性的控制作用, 最终建立一套股市混沌行为模拟和控制系统。

参 考 文 献

1 杨 健. 股票市场技术分析手册. 北京: 中国宇航出版社, 2002

2 Yang Y J, Liu G Z, Zhang Z P. Stock market trend prediction based on neural networks, multiresolution analysis and dynamical reconstruction. IEEE/ IAFI Conference on Computational Intelligence for Financial Engineering. Proceedings (CIFER), 2000, Mar 26- 28. 155- 157

3 姚洪兴, 盛昭瀚. 股市预测中的小波神经网络方法的研究. 管理工程学报, 2002, 16(2): 32- 37

4 Peters E. Fractal market analysis: applying chaos theory to investment and economics. New York: John Wiley and Sons, 1994

5 仲崇霞, 杨廷高. 小波域中的维纳滤波在综合脉冲星时算法中的应用. 物理学报, 2007, 103(10): 598- 604

6 林 海, 刘连寿. 非二进制离散小波的研究. 高能物理与核物

- 理, 2006 13(7): 60—65
- 7 杨 稣, 史耀媛, 宋 恒. 基于支持向量机的股市时间序列预测算法. 科学技术与工程, 2008 8(2): 381—386
- 8 云庆夏. 进化算法. 北京: 冶金工业出版社, 2000
- 9 王小平, 曹立明. 遗传算法. 西安: 西安交通大学出版社, 2000
- 10 胡昌华, 张军波, 夏 军, 等. 基于 MATLAB的系统分析与设计——小波分析. 西安: 西安电子科技大学出版社, 2001
- 11 Takens F. On the numerical determination of the dimension of an attractor. Eds Rand D, Young L S. Dynamical Systems and Turbulence. Springer-Verlag, 1981, 898, 316—381

Stock Market Time Series Prediction Method Based on SVM and Wavelet

YANG Su², SHI Yao Yuan¹, SONG Heng

(Xidian University, Xi'an 710075, P. R. China; Northwestern Polytechnical University, Xi'an 710072, P. R. China;

Navy Aviation Engineering College, Yantai 264001, P. R. China)

[Abstract] How to improve the effect of stock market time series prediction algorithm based on SVM by making use of wavelet transform technology is studied. The model predicts stock market trend by adding wavelet coefficients of stock market time series data in some different scale which are predicted by some same SVM. The method can synchronously adjust parameters of all SVM by means of genetic algorithm. After simulated and compared with the prediction method based on SVM, the method shows that the SVM stock market time series prediction algorithm combining wavelet transform further opens out the stock market rule.

[Key words] stock market; time series prediction; SVM; wavelet transform

(上接第 3161 页)

- 2 Han Shichen, Wang Hsiaoquan. Improvement of speaker recognition by combining residual and prosodic features with acoustic features. Acoustics, Speech, and Signal Processing, 2004 (ICASSP'04), IEEE International Conference on, Volume 1, 2004 1: 1—93—6
- 3 Yegnanarayana B, Reddy S K, Kishore S P. Source and system features for speaker recognition using AANN models. ICASSP Salt Lake City 2001, 409—413
- 4 Plumpe M D, Quatieri T F, Reynolds D A. Modeling of the glottal flow derivative waveform with application to speaker identification. IEEE Trans Speech Audio Processing, 1999, 7(5): 569—585
- 5 赵 力, 邹采荣, 吴镇扬. HMM在说话人识别中的应用. 电路与系统学报, 2001, 9: 51—57
- 6 李晓宇, 李虎生, 刘 加, 等. 利用 MCE算法提高说话人识别性能. 电路与系统学报, 2000, 3: 46—49

Speaker Recognition Based on HOCOR and MMCE

FAN Xiaochun, QIU Zhengquan

(School of Information and Electrical Engineering, Hunan University Science and Technology, Xiangtan 411201, P. R. China)

[Abstract] A new method to extract discriminative features from the linear prediction (LP) residual signal is proposed, which are closely related to the glottal excitation of individual speaker. Rather than taking Fourier transform, HAAR transform is applied to the residual signal, which is computationally simpler. A novel feature HOCOR is acquired by applying LP residue with HAAR transform. In order to improve the performance of recognition and the training velocity, modified maximum likelihood Error (MMCE) in the recognizing stage is applied. Experiment is showed that a better recognition result is acquired with the proposed novel feature and MMCE.

[Key words] LP residue; HAAR; HOCOR; MMCE