Computer Science and Application 计算机科学与应用, 2018, 8(6), 944-948
Published Online June 2018 in Hans. http://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2018.86105

Hans汉斯

Prediction of Finance Data Based Intelligent Support Vector Regression

Tian Luo

The School of Finance, Renmin University of China, Beijing
Email: MarcoLouis@163.com

Received: Jun. 6th, 2018; accepted: Jun. 21st, 2018; published: Jun. 28th, 2018

Abstract

Aiming at nonlinear, time variant, random, fuzziness and uncertainty of finance data, we propose a new intelligent support vector regression model and use new genetic algorithm to optimize the model's parameters. Experiment results show that intelligent support vector regression has higher accuracy and runs faster than BP Neural Networks.

Keywords

Support Vector Machine, Intelligent Genetic Algorithm, Finance Data, Prediction

基于智能支持向量机回归模型的金融数据预测

罗　添

中国人民大学，财政金融学院，北京
Email: MarcoLouis@163.com

收稿日期：2018年6月6日；录用日期：2018年6月21日；发布日期：2018年6月28日

摘　要

针对金融数据的非线性、时变性、随机性、模糊性、不确定性等特点，提出一种崭新的智能支持向量回归模型，并且运用一种新型的遗传算法优选模型参数。实验结果表明，所提出的智能支持向量回归模型预测金融数据比BP神经网络模型预测精度高、速度快。

关键词

支持向量回归，智能遗传算法，金融数据，预测

文章引用: 罗添. 基于智能支持向量机回归模型的金融数据预测[J]. 计算机科学与应用, 2018, 8(6): 944-948.
DOI: 10.12677/csa.2018.86105

1. 引言

金融数据由于受政策、技术等多种因素的影响，普遍具有非线性、时变性、随机性、模糊性、不确定性等特性。在计算经济学基础上建立的大部分金融线性模型，虽然能够比较直观地解决问题，但是在解决实际问题时也会有很大的不足。其根本原因在于金融数据具有非线性特征，运用线性模型来预测肯定会有较大的误差。所以，人们正在努力寻求高精度的预测工具进行金融数据建模[1] [2] [3]。

为了更加精确地进行金融数据预测，本文提出一种崭新的智能支持向量回归模型进行金融数据预测。支持向量机于 1995 年由 Vapnik 等人正式提出，已经成功地应用到回归等问题。然而，到目前为止，支持向量回归模型仍然没有好的参数优选方法[4]。本文运用一种新型遗传算法[5]来进行支持向量回归模型的参数优选。该新型遗传算法可以解决一般的遗传算法所带来的早熟问题和进化缓慢问题，具有较强的搜索能力，能够寻找全局最优解。因此，本文运用新型遗传算法对支持向量回归模型进行最优参数设置，从而得到一种崭新的智能支持向量回归模型。最后，将所建立的智能支持向量回归模型应用于金融数据预测，通过与 BP 神经网络模型比较，得到本文所提出的模型预测精度比较高，是进行金融数据预测的一种有效方法。

2. 支持向量回归模型

假设数据集 $S=\{(x_i,y_i),i=1,2,\cdots,n\}$，$x_i$ 是输入向量，y_i 是一个实数观测值，n 是数据的总数。支持向量回归函数的一般形式为

$$g(x)=w\phi(x)+b \tag{1}$$

式中 $\phi(x)$ 是一个高维的特征函数，b 是常数。

为求解 w 和 b，引入正松弛变量 ξ_i,ξ_i^*，可以得到如下规划问题[4]:

$$\begin{aligned}\min\quad & \frac{1}{2}\|w\|_1^2+C\sum_{i=1}^{n}(\xi_i+\xi_i^*)\\ \text{s.t.}\quad & y_i-w\phi(x_i)-b\le\varepsilon+\xi_i\\ & w\phi(x_i)+b-y_i\le\varepsilon+\xi_i^*,\xi^{(*)}\ge 0\end{aligned} \tag{2}$$

运用拉格郎日乘数法可以得到二次规划(2)的对偶规划为：

$$W(\alpha_i,\alpha_i^*)=\sum_{i=1}^{n}y_i(\alpha_i-\alpha_i^*)-\varepsilon\sum_{i=1}^{n}(\alpha_i+\alpha_i^*)-\frac{1}{2}\sum_{i=1}^{n}\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)(\phi(x_i)\cdot\phi(x_j)) \tag{3}$$

使得 $\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)=0$ 和 $\alpha_i,\alpha_i^*\in[0,C]$；

进而可以得到决策函数：

$$g(x)=\sum_{i=1}^{n}(\alpha_i-\alpha_i^*)\phi(x_i)\cdot\phi(x)+b \tag{4}$$

通过引进核函数，方程(4)可以写成如下形式

$$g(x)=\sum_{i=1}^{n}\left(\alpha_i-\alpha_i^*\right)K\left(x_i,x\right)+b \tag{5}$$

在知识发现理论中，高斯核函数已经被证明能够提供好的泛化能力。因此，本文采用高斯核函数 $K\left(x_i,x_j\right)=\exp\left(-\frac{\left\|x_i-x_j\right\|_1^2}{2\sigma^2}\right)$ 作为支持向量回归模型的核函数。另外，本文采用的损失函数是 ε-不敏感损失函数，即

$$L\left(g\left(x_i\right),y_i\right)=\begin{cases}\left|g\left(x_i\right)-y_i\right|-\varepsilon, & \left|g\left(x_i\right)-y_i\right|\geq\varepsilon\\ 0, & \text{others}\end{cases} \tag{6}$$

式中 ε 称为管道大小，它反映函数逼近的精确程度。参数 ε 属于自定义的参数。

3. 新型遗传算法

3.1. 适应度函数

训练数据上的 10-fold 交叉验证(CV)后的平均值的负值定义为适应度函数，即

$$g=-\overline{MAPE_{CV}} \tag{7}$$

$$MAPE_{CV}=\frac{\sum_{i=1}^{n}\frac{\left|y_i-g_i\right|}{2y_i}}{n}\times100\% \tag{8}$$

在这里 $MAPE$ 是平均绝对百分误差，n 是训练数据样本的数目，y_i 是确切值的数目，g_i 是预测值的数目。

3.2. 编码方式

因为浮点数编码不受维数限制，不需编码、解码操作，可以有效提高计算速度和求解精度，同时，浮点编码比二进制编码在变异操作上能更好地保持种群多样性，所以，采用浮点数编码方式[6]。

3.3. 选择操作

本文采用基于排序的适应度分派准则。按照适应度值对种群内的个体排序，然后按下式确定选择第 i 个个体的概率：

$$P_i=c\left(1-3c\right)^{i-1} \tag{9}$$

在这里 i 为个体排序序号；c 为排序第一的个体的选择概率。

3.4. 交叉和变异操作

交叉概率 P_c 和变异概率 P_m 根据文献[7]所提出的自适应度遗传算法来进行选择：

$$P_c=\begin{cases}P_{c_1}-\left(P_{c_1}-P_{c_2}\right)\left(g'-g_{avg}\right)\Big/\left(g_{\max}-g_{avg}\right), & g'\geq g_{avg}\\ P_{c_1}, & g'<g_{avg}\end{cases} \tag{10}$$

$$P_m=\begin{cases}P_{m_1}-\left(P_{m_1}-P_{m_2}\right)\left(f_{\max}-g\right)\Big/\left(g_{\max}-g_{avg}\right), & g'\geq g_{avg}\\ P_{m_1}, & g'<g_{avg}\end{cases} \tag{11}$$

在这里 $P_{c_1}=0.92, P_{c_2}=0.65, P_{m_1}=0.08, P_{m_2}=0.001$，$g_{\max}$ 为群体中最大的适应度值；g_{avg} 为每代群体

的平均适应度值； g' 为交叉的两个个体中较大的适应度值；g 为变异个体的适应度值。

进化代数按下式自适应变化：

$$P_c^t = \begin{cases} P_{c_1} \cdot \sqrt{1-\left(t/t_{\max}\right)^2}, & P_c^t < P_{c_2}, \\ P_{c_2}, & P_c^t \geq P_{c_2} \end{cases} \tag{12}$$

$$P_m^t = \begin{cases} P_{m_1} \cdot \exp\left(-\lambda \cdot t/t_{\max}\right), & P_m^t > P_{m_2}, \\ P_{m_2}, & P_m^t \leq P_{m_2} \end{cases} \tag{13}$$

在这里 t 为遗传代数， $t_{\max}$ 为最大遗传代数， λ 为常数，这里取 10。

4. 智能支持向量回归模型模型及应用

4.1. 智能支持向量回归模型流程图

智能支持向量回归模型流程图见图 1。

4.2. 数据收集

本例选取的是中小企业的华邦制药从 2010 年 1 月 4 日至 2012 年 3 月 8 日共 515 个交易日的收盘价作为研究数据，全部数据分为两部分，其中 2010 年 1 月 4 日至 2012 年 3 月 1 日的数据用于建立模型，剩下的 5 个数据用于预测模型的检验。

4.3. 数据标准化处理

数据标准化处理中采用的常用办法是转换数据的尺度，将全部数据线性映射到区间[0, 1]。

4.4. 实验结果

实验结果表 1 所示：

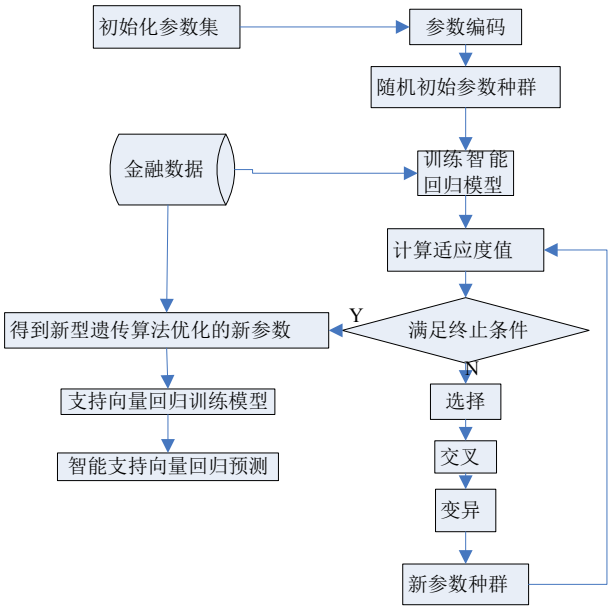

Figure 1. Flow chart of intelligent support vector regression model
图 1. 智能支持向量回归模型流程图

Table 1. The comparison of intelligent support vector regression model
表 1. 模型结果比较

日期	实际值	智能支持向量回归模型		BP 神经网络	
		预测值	相对误差(%)	预测值	相对误差(%)
2012-3-2	36.57	36.5682	−0.00492	36.4975	−0.19821
2012-3-5	36.73	36.7148	−0.00414	36.8759	0.39724
2012-3-6	36.25	36.2457	−0.01186	36.3871	0.37820
2012-3-7	35.56	35.5513	−0.02446	35.5841	0.06777
2012-3-8	35.99	35.9827	−0.02028	34.2518	−4.8296

表 1 是智能支持向量回归模型和 BP 神经网络模型预测结果的比较，可以看出，我们提出的智能支持向量回归模型明显优于 BP 神经网络模型。

5. 结论

本文提出的智能支持向量回归模型具有四大优点：1) 该模型精度高，是进行金融数据预测的一种很好方法；2) 智能支持向量回归具备较强的非线性映射能力，可以针对少量有限样本情况；3) 智能支持向量回归模型最终转化成为一个二次型寻优问题，得到的是全局最优解，解决了在神经网络方法中无法避免的局部极值问题；4) 智能支持向量回归模型将实际问题通过非线性变换转换到高维特征空间，在高维空间中构造线性判别函数来实现原空间中的非线性判别函数，使得模型具备较强的泛化能力。

参考文献

[1] 盛丹姝, 王德辉, 刘书丽. 基于融合估计的金融数据预测[J]. 吉林师范大学学报(自然科学版), 2013(1): 38-41.

[2] 廖丽芳, 蔡如华. 基于 MODWT 在金融数据预测的应用[J].计算机工程与设计, 2013, 34(4): 1346-1350.

[3] 徐喆. 逻辑回归模型在互联网金融 P2P 业务信用风险的应用[J]. 统计科学与实践，2015(11): 26-29.

[4] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning with Applications in R. Springer New York Heidlberg Dordrecht London. https://doi.org/10.1007/978-1-4614-7138-7

[5] 杨从锐, 钱谦, 王锋, 孙铭会. 改进的自适应遗传算法在函数优化中的应用[J]. 计算机应用研究, 2017, 35(4): 1-5.

[6] Lin, C.D., Anderson-Cook, C.M., Hamada, M.S., *et al.* (2015) Using Genetic Algorithms to Design Experiments: A Review. *Quality & Reliability Engineering International*, **31**, 155-167. https://doi.org/10.1002/qre.1591

[7] Goren, H.G., Tunali, S. and Jans, R. (2010) A Review of Applications of Genetic Algorithms in Lot Sizing. *Journal of Intelligent Manufacturing*, **21**, 575-590. https://doi.org/10.1007/s10845-008-0205-2

知网检索的两种方式：

1. 打开知网页面 http://kns.cnki.net/kns/brief/result.aspx?dbPrefix=WWJD
   下拉列表框选择：[ISSN]，输入期刊 ISSN：2161-8801，即可查询
2. 打开知网首页 http://cnki.net/
   左侧“国际文献总库”进入，输入文章标题，即可查询

投稿请点击：http://www.hanspub.org/Submission.aspx
期刊邮箱：csa@hanspub.org