

文章编号: 1001-0920(2013)12-1915-06

区间时间序列的混合预测模型

岳继光¹, 杨臻明¹, 孙强¹, 王晓保²

(1. 同济大学 电子与信息工程学院, 上海 201804; 2. 上海申通轨道交通研究咨询有限公司, 上海 201103)

摘要: 提出一种基于自回归求和移动平均 (ARIMA) 与人工神经网络 (ANN) 的区间时间序列混合模型, 并用混合模型分别对区间中值序列和区间半径序列建模. 采用 Monte Carlo 方法生成模拟区间序列, 分别用 ARIMA、ANN 和混合模型 3 种方法进行建模和预测实验, 并用统计学方法检验模型误差. 最后分别采用 3 种方法对 H 市轨道交通某号线牵引能耗区间序列进行了建模和预测, 实验结果表明混合模型的建模精度和预测性能均优于单一模型.

关键词: 区间分析; 时间序列; 混合模型; ARIMA 模型; 人工神经网络; 城市轨道交通; Monte Carlo 方法
中图分类号: TP273 **文献标志码:** A

Hybrid model for interval-valued time series

YUE Ji-guang¹, YANG Zhen-ming¹, SUN Qiang¹, WANG Xiao-bao²

(1. College of Electronic and Information Engineering, Tongji University, Shanghai 201804, China; 2. Shanghai Shentong Rail Transit Research and Consultancy Co Ltd, Shanghai 201103, China. Correspondent: YANG Zhen-ming, E-mail: yangzhenmingtk@163.com)

Abstract: A hybrid model based on the autoregressive integrated moving average (ARIMA) model and the artificial neural network (ANN) model is proposed to model and predict interval-valued time series. The interval-valued time series are converted to the mid-point and the half-range series, the forecasting of which is accomplished through a hybrid model, respectively. The evaluation of the ARIMA, ANN and hybrid models is based on the estimation of the average behavior of the mean squared error with synthetic and real interval-valued series in the framework of a Monte Carlo experiment. The experimental results show that the hybrid model is an effective way to improve the forecasting accuracy achieved by any one of the models separately.

Key words: interval analysis; time series; hybrid model; ARIMA model; artificial neural networks; urban rail transit; Monte Carlo method

0 引言

在时间序列研究领域, 人们对区间数据的分析和预测越来越重视. 一组按顺序排列的区间数便构成了区间时间序列^[1]. 传统的时间序列分析和预测方法大多基于统计学理论, 如自回归求和移动平均 (ARIMA) 类模型^[2] 和自回归条件异方差 (ARCH) 类模型等^[3]. 目前, 人工神经网络 (ANN) 已经成为通用的非线性时间序列建模预测方法^[4-5], 但尚未见到直接的区间时间序列建模和预测方法. 一般大都先将区间序列转换为普通序列, 然后再用传统时间序列方法对其进行分析和建模^[6], 例如 Roque 等^[4] 将区间序列转换为区间上下限序列后用神经网络建模.

本文提出一种基于混合模型的区间时间序列建模和预测方法. 首先, 简要回顾了 ARIMA、ANN 等时

间序列模型, 介绍了普通时间序列的混合模型; 然后, 将区间序列表示为区间中值序列和区间半径序列, 将普通时间序列方法推广至区间序列的建模和预测, 并在普通时间序列混合模型的基础上提出了区间序列的混合模型; 最后, 用 Monte Carlo 模拟方法生成合成区间序列, 分别用 ARIMA、ANN 和混合模型 3 种方法进行建模和预测实验, 用平均区间误差平方和来评价建模和预测误差, 同时对 H 市轨道交通某号线牵引能耗区间序列进行了建模和预测, 并比较了 3 种方法之间的差异.

1 时间序列混合预测模型

ARIMA 模型是常用的时间序列模型. 如果时间序列 y_t 平稳, 则可直接建立自回归移动平均 (ARMA) 模型, 其表达式为^[2]

收稿日期: 2013-01-05; 修回日期: 2013-10-06.

作者简介: 岳继光(1961—), 男, 教授, 博士生导师, 从事过程控制、计算机控制等研究; 杨臻明(1982—), 男, 博士生, 从事非线性系统、神经网络的研究.

$$y_t = \theta_0 + \sum_{i=1}^p \phi_i y_{t-i} + \varepsilon_t + \sum_{j=1}^q \theta_j \varepsilon_{t-j}. \quad (1)$$

其中: ε_t 是 t 时刻的随机误差项, ϕ_i 和 θ_j 是模型待估参数 ($i=1, 2, \dots, p, j=1, 2, \dots, q$), θ_0 是常数项, p 和 q 是 ARMA 模型的阶数. 随机误差项 ε_t 相互独立, 且服从均值为 0、方差为 σ^2 的正态分布. ARIMA (p, d, q) 模型可以对经过 d 阶逐期差分后平稳的时间序列 y_t 建模.

ANN 可以对非线性时间序列进行建模和预测. 在常用的多层感知器 (MLP) 网络基础上, 将网络输出经延迟后反馈到输入端, 得到非线性自回归 (NARX) 网络^[7-9], 其结构如图 1 所示.

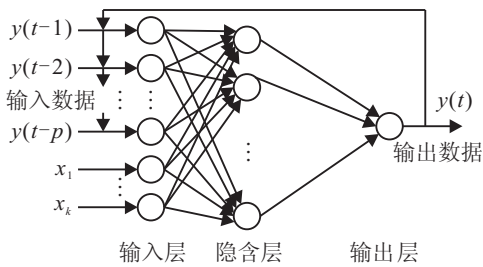


图 1 非线性自回归 (NARX) 网络结构

NARX 网络的表达式可以写为

$$y_t = \alpha_0 + \sum_{j=1}^q \alpha_j \cdot g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} + \sum_{i=p+1}^{p+m} \beta_{ij} x_m \right) + \varepsilon_t. \quad (2)$$

其中: α 和 β 是模型参数 (网络权值), p 是输出反馈延时步数, m 是外部输入个数, q 是隐含层节点个数, g 是隐含层传递函数, NARX 网络总的输入节点个数为 $p+m$. 当确定某些外部变量对时间序列有显著影响时, 可以将其作为 NARX 网络的输入变量.

ARIMA 模型和 ANN 模型分别在线性和非线性领域取得了很大成功. 然而, 将 ARIMA 模型用于解决复杂非线性系统建模问题并不合适; 同样, 使用 ANN 对线性系统建模会得到相互矛盾的结果^[10]. 若能将两者结合起来, 则能更好地描述时间序列的复杂结构. 在实际应用中, 可以使用混合模型同时对时间序列的线性和非线性过程建模^[10-12].

可以认为时间序列由线性自相关结构与非线性结构两部分组成^[10], 即

$$y_t = L_t + N_t, \quad (3)$$

其中 L_t 和 N_t 分别为时间序列的线性和非线性部分. ARIMA 与 ANN 的混合模型建模过程分为两步. 首先, 利用 ARIMA 模型对线性部分 L_t 建模, 其剩余残差部分

$$e_t = y_t - \hat{L}_t, \quad (4)$$

包含了时间序列的非线性特征; 然后, 利用 ANN 模型描述残差

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-p}) + \varepsilon_t. \quad (5)$$

在上述表达式中, p 是 ANN 输入层的节点个数, f 是 ANN 的非线性传递函数, ε_t 是随机误差项. 通过 ANN 模型得到的残差序列 \hat{e}_t 代表了时间序列非线性部分的估计值 \hat{N}_t , 据此, 混合模型对时间序列的预测值可表示为

$$\hat{y}_t = \hat{L}_t + \hat{N}_t. \quad (6)$$

混合模型利用 ARIMA 和 ANN 分别对时间序列的线性和非线性部分建模, 充分发挥了两者的优势, 其建模精度和预测性能优于单一模型.

2 区间时间序列预测模型

区间时间序列是普通时间序列的扩展, 其每个时刻的观测值是二元区间数. 因此, 普通时间序列分析方法不能直接应用于区间序列, 需要进行相应的改进. 一种思路是将区间序列转化为普通时间序列, 然后再用普通时序分析方法对其建模与预测.

2.1 区间中值与区间半径

表示区间数有多种方法, 除了区间上限和区间下限外, 还可用区间中值和区间半径表示区间数^[13]. 将区间数根据时间顺序排列起来, 得到区间时间序列. 设 n 是区间序列的长度, 在各个时间点 $t=1, 2, \dots, n$ 上, 令 X_{U_t} 为区间上限, X_{L_t} 为区间下限, 并且有 $X_{L_t} \leq X_{U_t}$, 整个区间时间序列为 $[X_{L_1}, X_{U_1}], [X_{L_2}, X_{U_2}], \dots, [X_{L_n}, X_{U_n}]$. 该区间序列可分解为两个时间序列: 区间中值序列 X^c 和区间半径序列 X^r . 区间中值序列与区间半径序列可以分别用区间上下限表示为

$$X_t^c = \frac{X_{U_t} + X_{L_t}}{2}, \quad X_t^r = \frac{X_{U_t} - X_{L_t}}{2}, \quad t=1, 2, \dots, n; \quad (7)$$

反之, 也可将区间序列表示成区间中值与区间半径的形式, 即

$$[X_{L_t}, X_{U_t}] = [X_t^c - X_t^r, X_t^c + X_t^r]. \quad (8)$$

只要用 ARIMA、ANN 或其他方法对区间中值序列和区间半径序列建立合适的模型, 即可对区间序列进行建模和预测.

用区间中值与区间半径表示二元区间有利于提高区间序列的建模精度. 设某种时间序列预测方法对于区间上下限序列、区间中值和半径序列的相对预测误差为

$$e_c = \frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{X}_{c_t} - X_{c_t}}{X_{c_t}} \right)^2, \\ e_r = \frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{X}_{r_t} - X_{r_t}}{X_{c_t}} \right)^2,$$

$$e_L = \frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{X}_{L_t} - X_{L_t}}{X_{L_t}} \right)^2, \\ e_U = \frac{1}{n} \sum_{t=1}^n \left(\frac{\hat{X}_{U_t} - X_{U_t}}{X_{U_t}} \right)^2. \quad (9)$$

假定该方法对各序列有相同的预测精度, 则有

$$e_c = e_r = e_L = e_U = e. \quad (10)$$

用上下限表示的平均区间误差平方和 (MSE) 为

$$\text{MSE}_1 = \frac{1}{n} \sum_{i=1}^n (\hat{X}_{L_t} - X_{L_t})^2 + \frac{1}{n} \sum_{i=1}^n (\hat{X}_{U_t} - X_{U_t})^2 = \frac{X_{L_t}^2}{n} \sum_{i=1}^n \frac{(\hat{X}_{L_t} - X_{L_t})^2}{X_{L_t}^2} + \frac{X_{U_t}^2}{n} \sum_{i=1}^n \frac{(\hat{X}_{U_t} - X_{U_t})^2}{X_{U_t}^2} = X_{L_t}^2 e_L + X_{U_t}^2 e_U = (X_{L_t}^2 + X_{U_t}^2) e; \quad (11)$$

用区间中值与区间半径表示的 MSE 为

$$\text{MSE}_2 = \frac{1}{n} \sum_{i=1}^n (\hat{X}_{c_t} - X_{c_t})^2 + \frac{1}{n} \sum_{i=1}^n (\hat{X}_{r_t} - X_{r_t})^2 = (X_{c_t}^2 + X_{r_t}^2) e. \quad (12)$$

将式 (7) 代入 (12), 得到

$$\text{MSE}_2 = \frac{1}{2} (X_{L_t}^2 + X_{U_t}^2) e. \quad (13)$$

这说明在预测方法相同的情况下, 用区间中值与区间半径表示的区间序列预测误差仅为用区间上下限表示的一半.

2.2 区间序列的混合模型

区间序列混合模型的基本思想是, 用普通混合模型分别对区间中值序列 X^c 和区间半径序列 X^r 建模, 得到 X^c 和 X^r 的预测值 \hat{X}^c 和 \hat{X}^r , 然后根据式 (8) 计算出区间序列的预测值 $[\hat{X}_{L_t}, \hat{X}_{U_t}]$.

设混合模型中的线性部分, 即 ARIMA 模型对区间中值序列 X^c 的估计值为 \hat{L}^c , 则根据线性模型的残差公式 (4), 可得到区间中值的残差序列

$$e_{X_t^c} = X_t^c - \hat{L}_t^c, \quad (14)$$

其中 X_t^c 是区间中值序列的观测值. 同理, 设 ARIMA 模型对区间半径序列 X^r 的估计值为 \hat{L}_t^r , 则区间半径的残差序列为

$$e_{X_t^r} = X_t^r - \hat{L}_t^r. \quad (15)$$

由此, 在建立线性模型后, 得到 2 个含有原区间序列非线性信息的残差序列 $e_{X_t^c}$ 和 $e_{X_t^r}$.

下面用混合模型的非线性部分, 即 ANN 模型分别对残差序列 $e_{X_t^c}$ 和 $e_{X_t^r}$ 建模, 得到残差序列的估计值 $\hat{e}_{X_t^c}$ 和 $\hat{e}_{X_t^r}$, 该估计值分别代表了混合模型中的非线性部分 \hat{N}_t^c 和 \hat{N}_t^r . 根据混合模型 (6), 时间序列的最

终预测值可以表示为线性与非线性两部分预测结果的和, 则区间中值序列与区间半径序列的预测值分别为

$$\hat{X}_t^c = \hat{L}_t^c + \hat{N}_t^c, \\ \hat{X}_t^r = \hat{L}_t^r + \hat{N}_t^r. \quad (16)$$

根据式 (8), 得到原区间序列的区间上下限分别为

$$\hat{X}_{L_t} = \hat{X}_t^c - \hat{X}_t^r = (\hat{L}_t^c - \hat{L}_t^r) + (\hat{N}_t^c - \hat{N}_t^r), \\ \hat{X}_{U_t} = \hat{X}_t^c + \hat{X}_t^r = (\hat{L}_t^c + \hat{L}_t^r) + (\hat{N}_t^c + \hat{N}_t^r). \quad (17)$$

由此即建立了区间序列的混合模型. 建模过程中没有对原区间序列作出任何假设, 该混合模型可以对任何区间序列建模.

3 区间序列预测实验

3.1 合成区间序列

为了评价 ARIMA、ANN 和混合模型的建模精度和预测性能, 模拟了 2 个长度为 200 的区间序列, 序列的生成步骤如下:

- 1) 假设区间中值序列 X^c 由已知随机过程产生, 如 AR(1) 过程;
- 2) 假设区间半径序列 X^r 服从已知均匀分布, 如 $X^r \sim U[10, 20]$;
- 3) 通过式 (8) 构建对应的区间序列 $X_t = [X_{L_t}, X_{U_t}]$.

每个合成区间序列 200 个观测值中的前 160 个为训练集, 后 40 个为测试集. 生成 2 个区间序列的随机过程见表 1. S_1 的区间中值 X^c 是由随机游走过程生成的, 这是典型的 AR(1) 过程, 序列当前值与前期值之间是线性关系, 随机误差项 ε_t 服从标准正态分布, 即 $\varepsilon_t \sim N(0, 1)$. S_2 的区间中值 X^c 是由非线性过程生成的, 序列表现出复杂的混沌特征. 序列 S_1 和 S_2 分别用来考察 3 种方法对线性和非线性区间序列的建模能力.

表 1 模拟合成的区间序列

| 区间序列 | X^c 过程 | X^r 过程 |
|-------|--|------------|
| S_1 | $X_t^c = X_{t-1}^c + \varepsilon_t$ | $U[5, 10]$ |
| S_2 | $X_t^c = 4X_{t-1}^c (1 - X_{t-1}^c) + \varepsilon_t$ | $U[2, 5]$ |

3.2 实验结果评估

设区间序列的观测值 $X_t = [X_{L_t}, X_{U_t}]$ 对应的预测值为 $\hat{X}_t = [\hat{X}_{L_t}, \hat{X}_{U_t}] (t = 1, 2, \dots, n)$, 区间序列模型的性能用平均区间上限误差平方和 (MSE_U) 和平均区间下限误差平方和 (MSE_L) 评价, 即

$$\text{MSE}_U = \frac{1}{n} \sum_{i=1}^n (\hat{X}_{U_t} - X_{U_t})^2, \quad (18)$$

$$\text{MSE}_L = \frac{1}{n} \sum_{i=1}^n (\hat{X}_{L_t} - X_{L_t})^2. \quad (19)$$

用 Monte Carlo 方法分别生成 1 000 个序列 S_1 和 S_2 的样本, 分别用 ARIMA、ANN 和混合模型对每个样本的训练集建模, 计算模型误差; 然后用该模型进行 12 步超前预测, 并用测试集计算相应的预测误差. Monte Carlo 方法的收敛性和误差是 2 个重要的指标^[14], 因此在模拟 1 000 个序列样本完成后将计算 MSE_U 和 MSE_L 的最终均值和标准差. 由序列的自相关和偏自相关特征, 依据 AIC 准则确定序列 S_1 和 S_2 的模型, 分别如下: 序列 S_1 的模型 ARIMA (1, 1, 0), ANN 输入层节点数 = 2, 隐含层节点数 = 3; 序列 S_2 的模型 ARIMA(2, 1, 1), ANN 输入层节点数 = 3, 隐含层节点数 = 5. 所用混合模型的结构如图 2 所示.

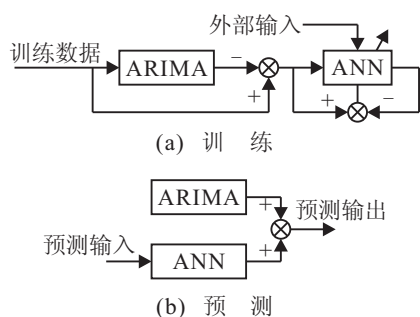


图 2 混合模型结构示意图

表 2 和表 3 显示了 Monte Carlo 方法得到的序列 S_1 和 S_2 的建模和预测结果, 括号内的数值是误差的标准差. 用 t 检验法检验 3 种方法建模的 MSE 的差异是否显著, 取显著性水平为 0.05, 结果见表 4 和表 5.

表 2 ARIMA、ANN 和混合模型对合成区间序列建模的 MSE 均值与标准差

| 区间序列 | 模型 | MSE_U | MSE_L |
|-------|-------|------------------|------------------|
| S_1 | ARIMA | 1.850 0(1.123 2) | 1.889 3(1.125 2) |
| | ANN | 1.500 6(1.156 5) | 1.499 8(1.155 5) |
| | 混合模型 | 1.110 4(1.155 9) | 1.133 0(1.155 4) |
| S_2 | ARIMA | 0.846 8(0.011 3) | 0.916 6(0.012 0) |
| | ANN | 0.348 9(0.012 2) | 0.376 2(0.015 4) |
| | 混合模型 | 0.146 7(0.012 9) | 0.148 6(0.012 7) |

表 3 ARIMA、ANN 和混合模型对合成区间序列 12 步超前预测的 MSE 均值与标准差

| 区间序列 | 模型 | MSE_U | MSE_L |
|-------|-------|------------------|------------------|
| S_1 | ARIMA | 3.242 6(0.656 4) | 3.044 7(0.558 7) |
| | ANN | 2.154 3(0.405 6) | 2.240 9(0.497 6) |
| | 混合模型 | 1.433 7(0.450 8) | 1.470 0(0.455 4) |
| S_2 | ARIMA | 3.616 4(0.309 7) | 3.667 2(0.298 3) |
| | ANN | 1.885 5(0.134 4) | 1.750 4(0.147 9) |
| | 混合模型 | 1.257 8(0.137 3) | 1.274 8(0.124 9) |

表 4 ARIMA、ANN 和混合模型对合成区间序列建模的 MSE 比较

| 区间序列 | MSE_U | MSE_L |
|-------|----------------|----------------|
| S_1 | ARIMA>ANN>混合模型 | ARIMA>ANN>混合模型 |
| S_2 | ARIMA>ANN>混合模型 | ARIMA>ANN>混合模型 |

表 5 ARIMA、ANN 和混合模型对合成区间序列 12 步超前预测的 MSE 比较

| 区间序列 | MSE_U | MSE_L |
|-------|----------------|----------------|
| S_1 | ARIMA>ANN>混合模型 | ARIMA>ANN>混合模型 |
| S_2 | ARIMA>ANN>混合模型 | ARIMA>ANN>混合模型 |

从表 2 和表 4 可看出, 无论对于线性区间序列 S_1 还是非线性区间序列 S_2 , 混合模型对训练集的建模精度都全面超过 ARIMA 模型和 ANN 模型, 而 ANN 模型又优于 ARIMA 模型. 值得注意的是, 即便对线性序列建模, 混合模型的表现也要优于线性 ARIMA 模型.

从表 3 和表 5 可看出, 在 12 步超前预测的情况下, 混合模型无论是对线性区间序列 S_1 还是非线性区间序列 S_2 , 其预测结果都是 3 种方法中最好的. 对于非线性区间序列的建模和预测, ANN 和混合模型都明显优于 ARIMA 模型, 这样的结果与预期相符. 综合 ANN 与混合模型在建模和预测两方面的表现, 混合模型优于 ANN 模型, 更优于 ARIMA 模型.

4 城市轨道交通能耗区间序列建模与预测

城市轨道交通系统的牵引能耗数据来源于运营企业的统计报表. 统计牵引能耗除了实际运送乘客所消耗的电能外, 还包括了车辆空驶、检测维修等非实际营运因素所消耗的电能. 因为种种原因, 这些额外能耗无法精确统计, 所以统计牵引能耗数据存在着不确定性^[15].

为了描述牵引能耗的不确定性, 引入牵引能耗区间数, 并进一步扩展为牵引能耗区间序列. 根据 H 市轨道交通某号线 2007 年~2010 年的牵引能耗数据, 用 ARIMA、ANN 和混合模型 3 种方法分别建立牵引能耗区间模型, 并预测 2011 年的牵引能耗区间.

4.1 牵引能耗数据与模型参数

表 6 是 H 市轨道交通某号线 2007 年 1 月~2010 年 12 月的牵引能耗区间上下限和月平均气温数据.

根据式 (7) 将牵引能耗区间序列转化为区间中值序列和区间半径序列. 由于序列非平稳且具有季节性, 做 1 阶逐期差分与季节长度 $S = 12$ 的 1 阶季节差分. 由序列的自相关和偏自相关特征, 依据 AIC 准则确定模型 ARIMA (3, 1, 1)(1, 1, 1)¹². 用 Eviews 实现上述的 ARIMA 模型.

根据城市轨道交通系统牵引能耗的特性, 选取表 7 中的变量作为 ANN 的输入输出变量^[15].

采用 NARX 网络对轨道交通能耗建模^[7]. 网络输入变量包括牵引能耗 E_q (输出反馈信号) 和月平均气温 T (外部输入信号); 输出变量为牵引能耗 E_q , 其中 E_q 的滞后步长为 12; 网络输入层的节点个数 = 输入变量的个数 = 13, 输出层的节点数 = 1. 用 Matlab 实

现上述 NARX 网络, 整个网络结构参见图 1.

表 6 H 市轨道交通某号线牵引能耗上下限和月平均气温

| 年/月 | 牵引能耗下限 10 ⁴ kW·h | 牵引能耗上限 10 ⁴ kW·h | 月平均气温 °C |
|---------|--------------------------------|--------------------------------|-------------|
| 2007/1 | 592.90 | 624.10 | 5.9 |
| 2007/2 | 555.09 | 584.30 | 9.8 |
| 2007/3 | 572.47 | 602.60 | 12.1 |
| 2007/4 | 663.96 | 698.90 | 15.9 |
| 2007/5 | 754.30 | 794.00 | 22.9 |
| 2007/6 | 771.02 | 811.60 | 25 |
| 2007/7 | 946.11 | 995.90 | 30.4 |
| 2007/8 | 964.35 | 1 015.10 | 29.7 |
| 2007/9 | 857.00 | 902.10 | 25.4 |
| 2007/10 | 835.91 | 879.90 | 20.6 |
| 2007/11 | 771.31 | 811.90 | 14.2 |
| 2007/12 | 838.00 | 882.10 | 9.8 |
| 2008/1 | 866.78 | 912.40 | 4.5 |
| 2008/2 | 841.32 | 885.60 | 4.2 |
| 2008/3 | 895.76 | 942.90 | 11.6 |
| 2008/4 | 902.12 | 949.60 | 16.1 |
| 2008/5 | 985.15 | 1 037.00 | 21.8 |
| 2008/6 | 995.79 | 1 048.20 | 24.2 |
| 2008/7 | 1 210.87 | 1 274.60 | 30.4 |
| 2008/8 | 1 195.48 | 1 258.40 | 28.6 |
| 2008/9 | 1 099.91 | 1 157.80 | 26 |
| 2008/10 | 1 036.45 | 1 091.00 | 21 |
| 2008/11 | 945.16 | 994.90 | 13.3 |
| 2008/12 | 976.32 | 1 027.70 | 7.9 |
| 2009/1 | 941.20 | 990.74 | 4.3 |
| 2009/2 | 855.37 | 900.39 | 9.3 |
| 2009/3 | 934.67 | 983.86 | 10.8 |
| 2009/4 | 939.05 | 988.47 | 16.7 |
| 2009/5 | 1 018.39 | 1 071.99 | 22.5 |
| 2009/6 | 1 083.63 | 1 140.66 | 26.3 |
| 2009/7 | 1 232.19 | 1 297.04 | 29 |
| 2009/8 | 1 212.31 | 1 276.12 | 28.1 |
| 2009/9 | 1 105.73 | 1 163.93 | 25.4 |
| 2009/10 | 1 057.05 | 1 112.68 | 21.4 |
| 2009/11 | 976.80 | 1 028.21 | 12.4 |
| 2009/12 | 971.00 | 1 022.10 | 6.9 |
| 2010/1 | 1 020.99 | 1 074.73 | 5.7 |
| 2010/2 | 915.35 | 963.53 | 7.7 |
| 2010/3 | 1 058.95 | 1 114.68 | 9.6 |
| 2010/4 | 1 020.37 | 1 074.07 | 13.3 |
| 2010/5 | 1 108.96 | 1 167.33 | 20.9 |
| 2010/6 | 1 219.26 | 1 283.44 | 24.1 |
| 2010/7 | 1 407.89 | 1 481.99 | 28.8 |
| 2010/8 | 1 556.00 | 1 637.89 | 30.9 |
| 2010/9 | 1 327.36 | 1 397.22 | 26.2 |
| 2010/10 | 1 141.43 | 1 201.50 | 19.3 |
| 2010/11 | 1 005.84 | 1 058.78 | 14.2 |
| 2010/12 | 994.50 | 1 046.85 | 5.7 |

表 7 牵引能耗神经网络输入输出变量表

| 符号 | 牵引能耗 E_q | 月平均气温 T | 牵引能耗 E_q |
|----|----------------------|--------------|----------------------|
| 单位 | 10 ⁴ kW·h | °C | 10 ⁴ kW·h |

4.2 预测结果与分析

用 2007 年 1 月~2010 年 12 月的牵引能耗区间

数据建立模型, 对 2011 年各月的牵引能耗区间进行预测. ARIMA、ANN 和混合模型的模型误差和预测误差如表 8 所示.

表 8 H 市轨道交通某号线牵引能耗区间序列的模型误差和预测误差

| | 模型 | MSE _U | MSE _L |
|------|-------|------------------|------------------|
| 模型误差 | ARIMA | 4.658 | 4.873 |
| | ANN | 2.895 | 2.812 |
| | 混合模型 | 2.187 | 2.156 |
| 预测误差 | ARIMA | 5.283 | 5.344 |
| | ANN | 4.198 | 4.265 |
| | 混合模型 | 3.231 | 3.328 |

从表 8 可以看出, 无论是建模还是预测, 混合模型的误差都明显小于 ARIMA 模型和 ANN 模型, 这与合成区间序列的实验结果吻合, 展示了混合模型对于实际区间时间序列建模与预测的性能.

5 结 论

本文提出了一种区间时间序列建模和预测的新方法. 将区间序列表示为区间中值序列和区间半径序列, 用普通时间序列方法分别对这两个序列进行建模和预测. 在普通时间序列混合模型的基础上, 提出了区间时间序列的混合模型, 分别用 ARIMA 和 ANN 模型对区间序列的线性和非线性部分建模. 区间序列的预测值由区间中值与区间半径预测值的线性组合表示.

本文分别用 ARIMA、ANN 和混合模型 3 种方法对 2 个合成区间序列进行了建模和预测实验, 用平均区间误差平方和评价模型误差; 采用 Monte Carlo 方法生成合成区间序列, 并用统计学方法对实验结果进行了检验. Monte Carlo 模拟结果表明, 3 种方法都具有较好的预测性能, 其中混合模型的综合性能是 3 种模型中最好的, 即便对于线性区间序列, 混合模型的表现也优于 ARIMA 模型. 对 H 市轨道交通某号线牵引能耗区间序列进行了建模和预测, 结果表明, 当区间序列具有复杂特性时, 混合模型的建模精度和预测性能均显著好于 ARIMA 模型和 ANN 模型, 混合模型展现出了良好的适应能力. 混合模型为区间时间序列的建模提供了一种新的方法, 在复杂区间序列预测方面具有良好的应用前景.

参考文献(References)

[1] Moore R E. Interval arithmetic and automatic error analysis in digital computing[M]. Starford: Stanford University Press, 1962: 1-35.
 [2] Box G E P, Jenkins G M, Reinsel G C. Time series analysis: Forecasting and control[M]. 4th ed. New York: John Wiley & Sons, 2008: 21-224.

- [3] Engle R F. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom Inflation[J]. *Econometrica*, 1982, 50(4): 987-1007.
- [4] Roque A M, Mate C, Arroyo J, et al. iMLP: Applying multi-layer perceptrons to interval-valued data[J]. *Neural Processing Letters*, 2007, 25(2): 157-169.
- [5] 曹玉苹, 田学民. 基于 SVM 和 Kalman 预测的非线性系统故障预报[J]. *控制与决策*, 2009, 24(3): 477-480.
(Cao Y P, Tian X M. Nonlinear system fault prognosis based on SVM and Kalman predictor[J]. *Control and Decision*, 2009, 24(3): 477-480.)
- [6] Arroyo J, Maté C. Introducing interval time series: Accuracy measures[C]. *Proc in Computational Statistics*. Heidelberg: Springer, 2006: 1139-1146.
- [7] Zemouri R, Gouriveau R, Zerhouni N. Defining and applying prediction performance metrics on a recurrent NARX time series model[J]. *Neurocomputing*, 2010, 73(13/14/15): 2506-2521.
- [8] Ma Q L, Zheng Q L, Peng H, et al. Multi-step-prediction of chaotic time series based on co-evolutionary recurrent neural network[J]. *Chinese Physics B*, 2008, 17(2): 536-542.
- [9] 李松, 刘力军, 解永乐. 遗传算法优化 BP 神经网络的短时交通流混沌预测[J]. *控制与决策*, 2011, 26(10): 1581-1585.
(Li S, Liu L J, Xie Y L. Chaotic prediction for short-term traffic flow of optimized BP neural network based on genetic algorithm[J]. *Control and Decision*, 2011, 26(10): 1581-1585.)
- [10] Zhang G. Time series forecasting using a hybrid ARIMA and neural network model[J]. *Neurocomputing*, 2003, 50(1/2/3): 159-175.
- [11] Araujo R, Ferreira T A E. An intelligent hybrid morphological-rank-linear method for financial time series prediction[J]. *Neurocomputing*, 2009, 72(10/11/12): 2507-2524.
- [12] Omer F D. A hybrid neural network and ARIMA model for water quality time series prediction[J]. *Engineering Applications of Artificial Intelligence*, 2010, 23(4): 586-594.
- [13] 徐惠莉, 吴柏林, 江韶珊. 区间时间序列预测准确度探讨[J]. *数量经济技术经济研究*, 2008, 12(1): 133-140.
(Xu H L, Wu B L, Jiang S S. On forecasting efficiency evaluation for interval time series[J]. *The J of Quantitative & Technical Economics*, 2008, 12(1): 133-140.)
- [14] Kaastra I, Boyd M. Designing a neural network for forecasting financial and economic time-series[J]. *Neurocomputing*, 1996, 10(3): 215-236.
- [15] 杨臻明, 岳继光, 王晓保, 等. 基于回归模型的城市轨道交通能耗预测[J]. *城市轨道交通研究*, 2010, 13(12): 22-25.
(Yang Z M, Yue J G, Wang X B, et al. Prediction of urban rail transit power consumption based on regression model[J]. *Urban Mass Transit*, 2010, 13(12): 22-25.)