

基于先验知识的移动通信话务量预测

彭 宇¹, 雷 苗¹, 郭 嘉¹, 彭喜元¹, 于 江², 陈 强²

(1. 哈尔滨工业大学自动化测试与控制研究所, 黑龙江哈尔滨 150080; 2. 中国移动黑龙江有限公司, 黑龙江哈尔滨 150028)

摘 要: 本文提出了一种基于先验知识引导的极大重叠离散小波变换的移动通信话务量预测方法. 采用傅里叶谱分析作为小波分解子成分先验知识降低小波分解的盲目性. 利用具有明确物理意义且更易提取子层的极大重叠离散小波变换对话务量序列进行分解. 分解后仍以傅里叶谱先验知识为参考, 合并相关子层形成趋势项和周期项两部分, 并采用季节性求和自回归滑动平均 (ARIMA) 模型对二者分别建模和预测. 采用真实数据测试的结果表明: 本文方法可实现多步预测, 且预测精度优于单纯的季节性 ARIMA 模型.

关键词: 移动通信; 话务量预测; 极大重叠离散小波变换; 先验知识

中图分类号: TP391 **文献标识码:** A **文章编号:** 0372-2112 (2011) 01-0190-05

Mobile Communication Traffic Forecasting with Prior Knowledge

PENG Yu¹, LEI Miao¹, GUO Jia¹, PENG Xi-yuan¹, YU Jiang², CHEN Qiang²

(1. Automatic Test and Control Institute, Harbin Institute of Technology, Harbin, Heilongjiang 150080, China;

2. China Mobile Heilongjiang Co. Ltd, Harbin, Heilongjiang 150028, China)

Abstract: This paper proposed a methodology of forecasting for mobile communication traffic with maximal overlap discrete wavelet transform (MODWT) according to priori knowledge. Fourier spectrum was chosen as the priori knowledge to avoid the blindness of wavelet decomposition. Then, MODWT which is easy to extract components with obvious physical meaning was employed to decompose the communication traffic data. Moreover, prior knowledge of fourier spectrum was taken as reference to synthesize relevant sublayers, leading to the trend and seasonal components. Further, seasonal autoregressive integrated moving average (ARIMA) model was applied to model and predict the previous trend and seasonal components, respectively. The results tested with real communication traffic data indicate: the methodology proposed in this paper can realize multistep prediction and the forecasting accuracy is superior to that of seasonal ARIMA models.

Key words: mobile communication; traffic forecasting; maximal overlap discrete wavelet transform; prior knowledge

1 引言

随着通信业务的快速发展, 对话务量的预测需求日益增加. 准确的话务量预测对网络管理, 规划和设计具有重要意义, 可为网络拥塞、覆盖和干扰等提供决策支持^[1]. 目前, 实际应用的方法主要是移动平均、指数平滑、线性回归等. 这些方法对话务量的趋势预测比较准确, 但对剧烈波动成分的预测则比较粗糙. 也有研究者采用传统的时间序列预测方法 ARMA、ARIMA 模型^[2]实现话务量预测, 收到了较好的效果. 但这些研究工作往往是针对平稳或差分后平稳的数据进行的, 针对非平稳、非线性数据, 则预测效果较差. 目前, 神经网络^[3,4]、支持向量机^[5]等智能算法也广泛应用于话务量等时间序列的预测中, 并取得了一定的效果. 然而, 在实际应用中, 数据的采集和汇总往往存在较大的时间延迟问题,

如: 移动通信公司的网络话务量数据虽然每小时自动采集一次, 但汇总延迟最大可达 24h. 而对移动业务峰值预警等实际网络管理需求的分析表明: 在日常的移动通信网络维护中, 对中长期预测的需求最为迫切, 如: 一周左右 (168h) 的预测. 由此可见, 若采用适合单步或少量多步预测的智能算法解决实际问题还存在一定的困难.

针对 IP 骨干网网络流量中长期预测需求, 文献[6]以 1.5h 为采样间隔, 利用离散小波变换 (DWT) 算法的二进伸缩能力, 准确提取出了傅里叶谱里显示的周期为 12h、24h 的细节项, 并以尺度为 96h 的信号近似傅里叶谱周期为 168h 的趋势项信号, 在半年至一年的长期预测中取得了良好的效果. 借鉴文献[6]的研究思路, 在此基础上, 本文提出一种基于先验知识引导的极大重叠离散小波变换 (MODWT) 的移动通信话务量中期预测方法, 简记为 MMODWT (Modified MODWT) 方法. 其中, 极大

重叠离散小波变换^[7]是对原始序列及其所有可能的循环平移序列经离散小波变换加权而得,因此可能更容易地提取出话务量中物理意义明晰的特征,而不会像文献[6]所采用的 DWT 方法那样,局限于采样尺度上时间尺度的二进伸缩,对趋势项的提取存在原理误差.本文利用傅里叶谱分析的结果为小波分析提供先验知识,用以确定 MODWT 中小波基和分解层数的选取,实现话务量趋势和细节的分离.

本文其它部分结构安排如下.第 2 部分介绍话务量序列傅里叶分析,极大重叠离散小波分解算法,基于季节性 ARIMA 模型的话务量预测方法和 MMODWT 方法流程.采用实际话务量数据进行的测试和结果分析在第 3 部分给出.最后,第 4 部分总结全文.

2 采用 MMODWT 的话务量中期预测

近年来,基于小波分解等局部模型预测方法,将复杂时间序列分解为一定频段的子序列,分层预测,降低了每层的预测难度,取得了良好的效果.但是,由于实际数据的多样性和复杂性,以及先验知识的缺乏常常导致分解的盲目性.本文将引入傅里叶谱分析的结果作为先验知识,引导 MODWT 的分解过程.

2.1 移动通信话务量数据的傅里叶分析

本文采用 FFT 算法对话务量数据频谱成分进行分析,并将“频率—幅度”特性转化为“周期—幅度”特性,作为后续小波分解的先验知识和分解后模型精简整合的依据.

在黑龙江省移动通信话务量数据集随机抽取的 16 个小区的实验中,傅里叶谱显示话务量数据在 6、8、12、24、84、168h 等成分上具有明显周期.选取其中具有典型代表意义的哈尔滨市某小区,其“周期—幅度”特性如图 1.这个结果也符合人们的作息是以日、周等为周期的生活规律.

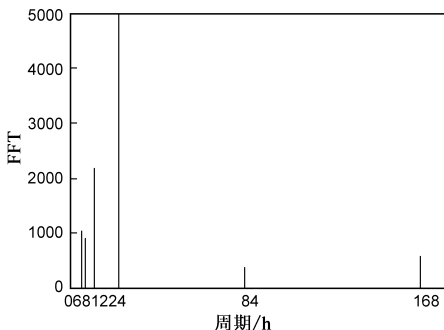


图1 哈市某小区话务量频谱

2.2 话务量序列的极大重叠离散小波分解

设 X 为 N 点实序列: $X = \{X_t, t = 0, 1, \dots, N-1\}$, J_0 为小波分解的层数,通过利用第 j 层小波滤波器 $\{\tilde{h}_{j,t}\}$ 和第 j 层尺度滤波器 $\{\tilde{g}_{j,t}\}$ 对 X 进行循环滤波,也即

$$\tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,t} X_{t-l \bmod N} \quad (1)$$

$$\tilde{V}_{j_0,t} = \sum_{l=0}^{L_{j_0}-1} \tilde{g}_{j_0,t} X_{t-l \bmod N} \quad (2)$$

可得 J_0 个 MODWT 的小波系数和 1 个尺度系数,记为

$$\tilde{W}_j = \{\tilde{W}_{j,0}, \tilde{W}_{j,1}, \dots, \tilde{W}_{j,N-1}\}^T, j = 1, 2, \dots, J_0 \quad (3)$$

$$\tilde{V}_{j_0} = \{\tilde{V}_{j_0,0}, \tilde{V}_{j_0,1}, \dots, \tilde{V}_{j_0,N-1}\}^T \quad (4)$$

其中, $j = 1, 2, \dots, J_0, t = 0, 1, 2, \dots, N-1, L_j = (2^j + 1)(L - 1) + 1, L_{j_0} = (2^{J_0} + 1)(L - 1) + 1$.

Percival 等(1997)又给出了极大重叠离散小波逆变换公式以实现序列 X 的多分辨分析和精确重构,利用式(3)和(4)得到的小波系数向量和尺度系数向量,合成 MODWT 总的变换系数向量:

$$\tilde{W} = (\tilde{W}_1^T, \dots, \tilde{W}_{J_0}^T, \tilde{V}_{j_0}^T)^T \quad (5)$$

同时,结合满足 $\tilde{W} = \tilde{w}X$ 的 MODWT 矩阵 $\tilde{w} = (\mathbf{w}_1^T, \mathbf{w}_2^T, \dots, \mathbf{w}_{j_0}^T, \mathbf{v}_{j_0}^T)^T$, 得到 X 的重构公式为:

$$X = \tilde{w}^T \tilde{W} = \sum_{j=1}^{J_0} \tilde{w}_j^T \tilde{W}_j + \mathbf{v}_{j_0}^T \tilde{V}_{j_0} = \sum_{j=1}^{J_0} \tilde{D}_j + \tilde{S}_{j_0} \quad (6)$$

其中 MODWT 矩阵 \tilde{w} 可由第 j 层小波滤波器 $\{\tilde{h}_{j,t}\}$ 和第 j 层尺度滤波器 $\{\tilde{g}_{j,t}\}$ 产生; $\tilde{S}_{j_0} = \mathbf{v}_{j_0}^T \tilde{V}_{j_0}$ 为趋势项, $\tilde{D}_j = \mathbf{w}_j^T \tilde{W}_j$ 为第 j 层小波细节.这样就得到了序列 X 的一个基于 MODWT 的多分辨分析,实现了序列的极大重叠离散小波分解.

实际应用中,取 X 为移动通信话务量序列,黑龙江省移动话务量采样尺度为 1h.具体分解过程中,选择合适的小波基,增大分解层数 J_0 ,直到恰好分解出傅里叶谱里显示的周期最长(本文取 $T = 168h$)的成分,而比其周期更长,更平稳的成分则可视趋势项.

2.3 话务量季节性 ARIMA 模型预测方法

本文采用 Box 和 Jenkins 提出的季节性 ARIMA 模型对子层分别建模和预测.该模型结合了季节差分运算处理和 ARMA 模型,并在实际应用中取得了良好的效果^[8].关于模型的理论和应用分析可参考文献[8]和[9].其中,在后文对子层进行建模的过程中,设定差分阶数为 1,AR 和 MA 阶数的上限均设为 2,针对不同的数据遍历上述两个参数获取相应最佳的模型阶数,并进一步确定相应系数.

2.4 MMODWT 方法流程

MMODWT 方法总的预测流程图如图 2 所示.首先,傅里叶分析发现其在 6、8、12、24、84、168h 等时间尺度上具有显著的周期特性.其次,采用 MODWT,提取出 $T = 168h$ 的成分,并利用 FFT 算法进行验证.对分解后的子序列,整合为趋势项和周期项两部分^[10].最后,两者分别采用季节性 ARIMA 模型进行建模和预测,等权相

加即得总的预测结果。

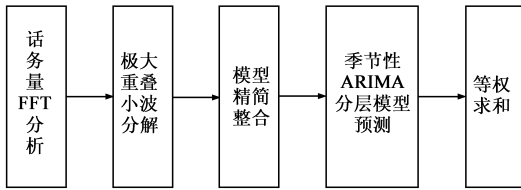


图2 话务量中期预测流程

3 测试及结果分析

实验用计算机 CPU 为 Intel Pentium D 2.8GHz, 内存 1GB. 其中季节性 ARIMA 模型建模和预测过程在统计分析软件 SAS 环境下运行; 其它过程在 MATLAB 运行。

3.1 极大重叠离散小波分解

选取 Haar 小波基(后文将在算法相关问题部分给出原因), 当分解层数增加为 7 层时, 话务量序列的多分辨分析如图 3. 这里选取哈尔滨市某小区 2007 年 11 月和 12 月份前两个星期共 1008 点话务量数据作为建模数据, 进行极大重叠离散小波分解, 为了便于观察, 图 3 仅给出最后 672 点数据的多分辨分析。

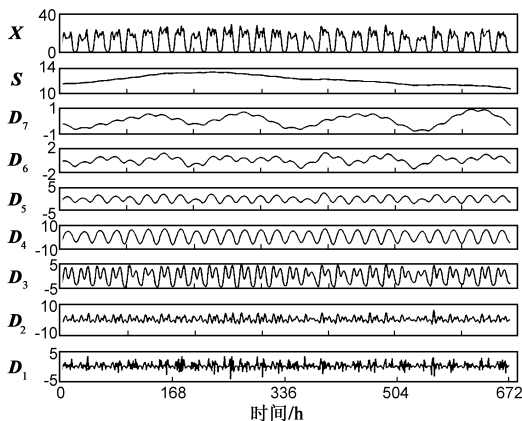


图3 哈尔滨市某小区话务量多分辨分析

如图 3 所示, 原始序列 X 每 24 个采样点完整起伏一次, 与之对照, 可发现在分解得到的细节项 D_7 的一个周期里, 包含原始序列 X 的七个完整起伏, 可能是傅里叶谱里显示的 $T = 168h$ 的成分; 而 S 和原始序列 X 的轮廓相吻合, 可能是趋势项成分. 为了进一步论证, 我们对 D_7 和 S 分别进行傅里叶分析, 结果显示 D_7 的最主要周期成分为 $T = 168h$, S 无显著周期, 证明了我们的猜测。

3.2 四种预测方法

MMODWT 方法以傅里叶谱分析的结果作为先验知识, 采用 MODWT 算法对原始话务量进行分解. 而后, 为了减少预测的子层数, 将所有细节项整合为一个周期项, 也即 $D = D_1 + D_2 + D_3 + D_4 + D_5 + D_6 + D_7$, 于是总的分解效果相当于 $X = D + S$; 对趋势项采取 $T = 24$ 作

为季节差分的周期, 对周期项采用公共周期 ($T = 168h$) 作为季节差分的周期, 分别利用季节性 ARIMA 模型建模, 并一次性预测 168 步. 作为对比, 利用季节性 ARIMA 模型, 选取 $T = 168$ 作为季节差分的周期, 直接对原始话务量进行预测, 后文简记为季节性 ARIMA 模型方法。

为了验证 MMODWT 的有效性, 同时给出 DWT 算法分解后逐层预测的预测精度参数作为对比, (分解过程中也选取 Haar 小波基、分解层数为 7 层), 后文简记为 DWT 方法; 为了验证精简整合过程的效果, 也给出先验知识引导下 MODWT 算法分解后逐层预测的预测精度参数, 后文简记为无整合 MMODWT。

3.3 评价参数

为了更加客观地评价预测性能, 本文采用平均绝对误差 MAE (Mean Absolute Error)、规范化均方根误差 NRMSE (Normalized Root Mean Square Error) 以及标准化误差 NE (Normalized Error) 三种评价标准。

设 \hat{X}_k 为预测值, X_k 为其对应真值, 其中 $k = 1, 2, 3, \dots, M$, 本文取 $M = 168$. 定义:

$$MAE = \frac{1}{M} \left(\sum_{k=1}^M |X_k - \hat{X}_k| \right) \quad (7)$$

$$NRMSE = \sqrt{\frac{1}{M} \sum_{k=1}^M \left(\frac{X_k - \hat{X}_k}{X_k} \right)^2} \quad (8)$$

$$NE = \frac{\sum_{k=1}^M (X_k - \hat{X}_k)^2}{\sum_{k=1}^M (X_k - \bar{X})^2} \quad (9)$$

$$\bar{X} = \frac{1}{M} \sum_{k=1}^M X_k \quad (10)$$

3.4 测试结果及分析

下面给出 MMODWT、季节性 ARIMA、DWT、无整合 MMODWT 等四种方法的预测性能评价, 见表 1。

表 1 哈尔滨市某小区预测 168h

预测方法	MAE	NRMSE	NE
MMODWT	1.66	0.34	0.12
无整合 MMODWT	1.91	0.37	0.14
季节性 ARIMA	2.35	0.46	0.23
DWT	4.48	0.65	0.58

由表 1 可见:

(1) MMODWT 方法对话务量总体趋势预测相对更为准确, 与仅采用季节性 ARIMA 模型的方法相比, 3 个误差指标均降低 25% 以上。

(2) DWT 方法预测效果最差, 比不分解直接预测的精度还低. 这说明盲目的分解并不能提高预测精度。

(3) 无整合 MMODWT 方法预测精度与 MMODWT 方法相当, 但需要建模的子层数较多。

直接采用季节性 ARIMA 模型的预测效果如图 4,

MMODWT 方法的预测效果如图 5. 对比图 4 和图 5, 可发现 MMODWT 在星期三、星期四、星期五上预测误差明显降低.

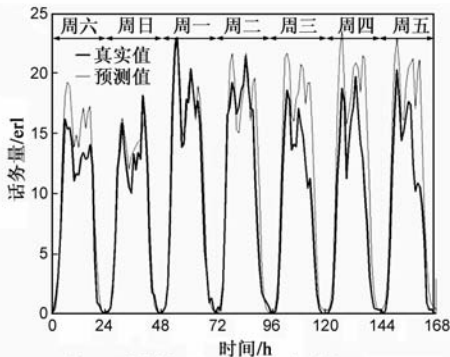


图4 季节性ARIMA模型预测168h

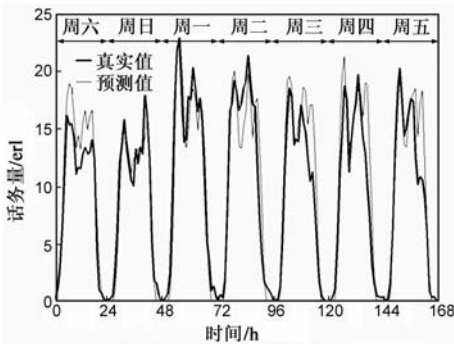


图5 MMODWT方法预测168h

为了进一步验证 MMODWT 方法的有效性,下文再抽取大庆市某小区 2008 年 4 月和 5 月前两周共 1008 点作为建模数据.

傅里叶分析显示主要的频率成分,除少一个周期为 84h 成分外,与哈市某小区一致.预测精度评价表格见表 2.由表 2 可知,MMODWT 方法比仅采用季节性 ARIMA 模型,3 个误差指标也降低 20%.

表 2 大庆市某小区预测 168h

预测方法	MAE	NRMSE	NE
MMODWT	1.71	0.40	0.16
无整合 MMODWT	1.79	0.43	0.18
季节性 ARIMA	2.38	0.49	0.26
DWT	2.91	0.57	0.38

3.5 算法相关问题

本节主要阐述先验知识引导下 MODWT 算法分解过程中选取 Haar 小波基的原因,并对预测算法的复杂度进行分析.

首先给出选取 Haar 小波基的原因:对图 3 中细节项 D_3 进行傅里叶分析,发现其最主要周期成分为 $T = 12h$,而 Haar 小波基提取出的周期为 12h 的成分和原始序列吻合良好,如图 6 所示.

算法复杂度方面,DWT 为 $O(N)$,而 MODWT 为

$O(N \log_2 N)$.虽然与 DWT 相比,MODWT 的计算复杂性有所增加,但仍然和 FFT 相同,因此是可以接受的.

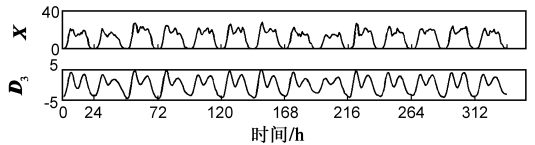


图6 Haar小波基对细节的提取

4 结论

针对移动通信话务量预测需求,本文提出了一种采用先验知识的移动通信话务量预测方法.实验结果表明,本文方法在一周左右的中期预测上,较好的预测出了话务量整体趋势,可以作为峰值预警等的依据,与单纯采用季节性 ARIMA 模型方法相比,算法复杂度增加不大,而预测精度提高明显.此外,本文方法对于如网络流量、居民用水量、城市公交车流量、电梯人流量等其他实际多周期时间序列的预测也具有一定的参考价值.

需要指出的是,本文方法对话务量及其它具有多周期特性的实际时间序列具有一般意义,但对于周期特性并不明显的时间序列并不适用.如何获取非周期序列的先验知识,以分析出原始序列中隐藏的不同尺度成分有待研究.同时,由于本文采取的是线性方法进行预测,在局部的预测效果较差,如何既能保证较长的预测步数,又能兼顾趋势和细节的预测效果有待进一步研究.

参考文献:

- [1] 张伟,张新波.移动 GSM 网话务量的 ARIMA 模型的建立及其预测模型[J].数学理论与应用,2008,28(2):70-74.
Zhang Wei, Zhang Xin-bo. Modeling and predicting wireless GSM traffic load on the ARIMA model[J]. Mathematical Theory and Applications, 2008, 28(2): 70-74. (in Chinese)
- [2] 薄今纲,于敏芳,刘嘉,等.基于季节 ARIMA 模型的 GSM 话务量建模和预报[J].数理统计与管理,2004,24(06):19-24.
Bo Jin-gang, Yu Min-fang, Liu Jia-kun, et al. Modeling and predicting wireless GSM traffic on the seasonal ARIMA model [J]. Application of Statistics and Management, 2004, 24(06): 19-24. (in Chinese)
- [3] 邓波,李健,孙涛,等.基于神经网络的话务量预测[J].成都信息工程学院学报,2008,23(05):518-521.
Deng Bo, Li Jian, Sun Tao, et al. Mobile telephone traffic prediction model based on neural network[J]. Journal of Chengdu University of Information Technology, 2008, 23(05): 518-521. (in Chinese)
- [4] 马君,刘小东,孟颖.基于神经网络的城市交通流量预测研究[J].电子学报,2009,37(5):1092-1094.

- Ma Jun, Liu Xiao-dong, Meng Ying. Research of urban traffic flow forecasting based on neural network[J]. Acta Electronica Sinica, 2009, 37(5): 1092 – 1094. (in Chinese)
- [5] 刘大同, 彭宇, 彭喜元. 基于残差预测修正的局部在线时间序列预测方法[J]. 电子学报, 2008, 36(12A): 81 – 85.
- Liu Da-tong, Peng Yu, Peng Xi-yuan. Local online time series prediction based on the residual compensation with online SVR [J]. Acta Electronica Sinica, 2008, 36(12A): 81 – 85. (in Chinese)
- [6] Konstantina Papagiannaki, Nina Taft, et al. Long-term forecasting of internet backbone traffic[J]. IEEE Transactions on Neural Networks, 2005, 16(5): 1110 – 1124.
- [7] Xu Qi-fa, Zhang Shi-ying. Research on multiresolution persistence and common persistence[J]. Systems Engineering Theory Practice, 2007, 27(2): 36 – 45.
- [8] George E P Box, Gwilym M Jenkins, Gregory C Reinsel. Time Series Analysis: Forecasting and Control[M]. 3rd ed. Englewood Cliffs, NJ; Prentice-Hall, 1994. 377 – 391.
- [9] Peter J Brockwell, Richard A Davis. Time Series: Theory and Methods[M]. 2nd ed. New York; Springer-Verlag, 1991. 250 – 253.

- [10] Viviana Fernandez. Wavelet and SVM-based forecasts: An analysis of the U.S metal and materials manufacturing industry[J]. Resources Policy, 2007, 32: 82 – 89.

作者简介:



彭宇 男, 教授、博士生导师, 1973年6月生于陕西西安, 哈尔滨工业大学自动化测试与控制研究所副所长, 主要研究方向为测试诊断技术、无线传感器网络技术和数据挖掘技术等。

E-mail: pengyu@hit.edu.cn



雷苗 男, 1985年8月生于重庆, 现为哈尔滨工业大学自动化测试与控制系博士研究生, 主要研究方向为时间序列小波分析、复杂网络等。

E-mail: leimiao_hit@126.com