

文章编号：1005-8451 (2016) 03-0031-04

基于EEMD-GA-BP的组合客流预测算法研究

翁涅元¹，单杏花²

(1.中国铁道科学研究院 铁道技术研修学院，北京 100081；

2.中国铁道科学研究院 电子计算技术研究所学院，北京 100081)

摘要：以高速铁路泰安站到达客流为研究对象，从客流数据的时频特性角度分析客流的特征，并结合经验模态分解法的时频分析优势以及遗传算法优化的神经网络的拟合能力，探索可行组合预测算法，以泰安站到达客流数据为例进行了实例分析，比较不同的IMF分量重构方法并确定了较优方案。

关键词：经验模态分解；遗传算法；BP神经网络；统计

中图分类号：U293.13 TP39 文献标识码：A

Combined passenger flow prediction algorithm based on EEMD-GA-BP

WENG Shengyuan¹, SHAN Xinghua²

(1. Railway Technology Research College, China Academy of Railway Sciences, Beijing 100081, China

2. Institute of Computing Technologies, China Academy of Railway Sciences, Beijing 100081, China)

Abstract: This article analyzed the passenger flow time and frequency characteristic of Tai'an Station, explored a feasible combination forecasting algorithm combining with EEMD and GA-BP Algorithms, taken the travelers of Tai'an Station as example to analyze and compare different reconstruction methods of IMFs, determine the optimal one.

Key words: Empirical Mode Decomposition(EMD); Genetic Algorithm; Back Propagation(BP) neural network; statistics

客流数据具有复杂的变化规律，从客流的时频特性上进行分析有利于分解这些规律，并为预测工作提供更有效的信息^[1]。噪声辅助经验模态分解法(EEMD)在分析非平稳及非线性数据上效果出色，结合遗传算法优化的神经网络(GA-BP)^[2]建模工具形成的组合预测方法(以下简称EEMD-GA-BP算法)较传统的单一预测方法有明显的优势^[3]。本文以泰安站到达客流数据为例，对基于EEMD的组合预测方法以及传统单一预测方法的效果进行对比，并针对EEMD的本征模函数预测值，选择出较优的重构方式，以保证最终预测精度。

1 算法描述

1.1 EEMD-GA-BP组合预测算法综述

基于EEMD的组合预测算法的一般步骤如下：

将原始数据 $x(t)$ ($t=1, 2, \dots, T$) 通过EEMD分解为若干本征模函数 $C_j(t)$ ($j=1, 2, \dots, N$) 以及趋势项 $r(t)$ 。

收稿日期：2015-08-31

基金项目：国家自然科学基金(U1334201)。

作者简介：翁涅元，在读硕士研究生；单杏花，研究员。

(1) 使用遗传算法优化的神经网络(GA-BP)对本征模函数进行建模与预测。

(2) 将分量的预测结果相加使其还原为原序列的预测结果。

综上所述，传统基于EEMD的预测算法步骤如图1所示。

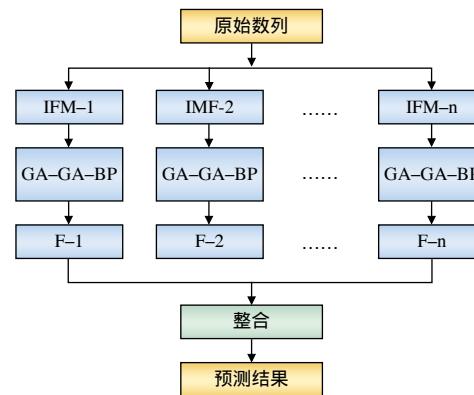


图1 EEMD-GA-BP组合预测算法流程图

1.2 噪声辅助经验模态分解算法

经验模态分解法(EMD)将非平稳序列分解成有限本征模函数(IMF)分量和一个趋势项，这些分量包含的频率成分随序列的变化而变化，通过对

分量的频率和幅值进行分析，可以准确地反映出原有序列的时频特性。经过 EMD 处理后的数据表示如式 (1)：

$$x(t) = \sum_{j=1}^N c_j(t) + r(t) \quad (1)$$

其中， N 表示本征模函数的个数， $c_j(t)$ ($j=1, 2, \dots, N$) 表示本征模函数，频率范围由高到低， $r(t)$ 是趋势项，代表原序列的主要趋势。

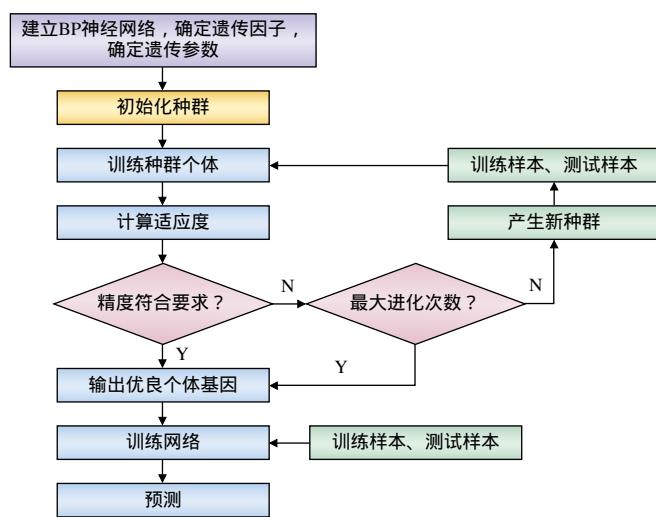
在处理真实数据时存在的外部干扰因素，EMD 分解容易出现模式混叠的现象而影响本征模函数的分析。因此引入噪声辅助经验模态分解法 (EEMD)^[4]，即在分析时对原序列加入随机白噪声信号后再进行分解，如此重复 M 次，取分解结果的平均值。IMF 的分量表达式由式 (2) 表示：

$$c_j(t)' = \frac{\sum_{i=1}^M c_{ji}(t)}{M} \quad (i=1, 2, \dots, M) \quad (2)$$

其中， $c_{ji}(t)$ 代表第 i 次 EMD 分解加入噪声后的数据的 IMF 分量。

1.3 GA-BP算法

遗传算法优化的 BP 神经网络 (GA-BP) 算法，即在传统的 BP 神经网络基础上根据神经网络个体的拟合精度作为适应度指标，以 BP 神经网络隐藏层节点数、节点间的权重系数作为遗传因子利用遗传算法进行参数优化^[5]，最终生成拟合精度较高的神经网络的算法。其流程如图 2 所示。



1.4 IMF分量重构

对于 IMF 分量的重构^[6] 有多重方法，若简单的

相加会影响预测的精度。GA-BP 神经网络本身具有非线性映射的功能，因此可以使用 GA-BP 神经网络进行训练来搜索 IMF 的最优权重组。在重构时是否需要选择所有的分量，以及往年同期数据的加入是否对预测有所帮助将在章节 3 中做出讨论。

2 客流特性分析

2.1 高速铁路泰安站客流特性

高速铁路泰安站是京沪高速铁路 24 个站点之一，毗邻著名旅游风景区泰山，在一年中不同时段有着明显不同的特点^[7]。泰安站 2013 年全年的到达客流序列如图 3 所示。

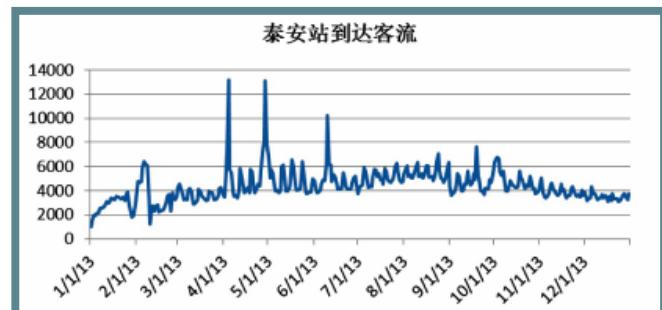


图3 高速铁路泰安站2013年日到达客流示意图

可以看到 4 月 4 日 (清明)、4 月 29 日 (五一)、6 月 10 日 (端午)、9 月 19 日 (中秋)、10 月 1 日 (国庆) 均表现为客流迅速攀升达到尖峰，可以推测为旅游流的集中爆发。2 月 9 日 (除夕) 以前的客流表现为逐渐攀升至高点而在 2 月 10 日 (初一) 突降至最低点，可以推测增加的客流为返乡流。由图 3 可以看出高铁泰安站表现出明显的非平稳特征，部分时间点的客流表现出相较于平时明显的差异性。

2.2 客流序列的聚类分析

通过对客流序列进行聚类分析为客流分类提供依据，将客流分为若干段长度为 7 天的片段，并使用 K 均值聚类方法分为具有明显区别的 4 类，各类片段在全年的分布如图 4 所示。

由图 4 可以看出除节假日外绝大部分日期均被归类为类别 1，说明泰安站的平日客流具有相似的特征，因此对于平日客流的预测可以使用较为统一的方法。

2.3 客流序列数据EEMD分析

将客流序列进行 EEMD 分解后的 IMF 以及趋势

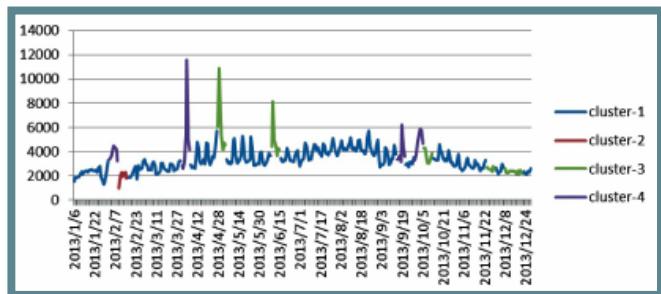


图4 客流聚类分布图

项曲线如图5所示。

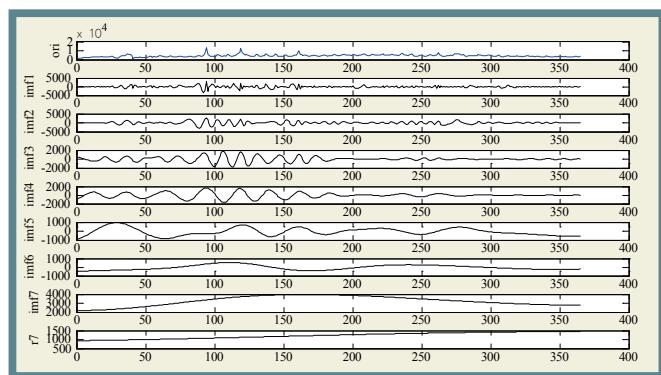


图5 客流数据EEMD分解结果

将 IMF 分量以及趋势项与原序列进行 Pearson 相关性分析, 结果如下:

$$P(i) = [-0.0075 \ 0.0505 \ 0.0555 \ -0.0557 \ 0.0519 \\ 0.3146 \ 0.4794 \ -0.0915]$$

可以看出 IMF-1 与原序列的相关性极小, 可以推测序列间的相关性不大。

其经过希尔伯特变换 (HHT)^[8] 后的边际图谱反映了原数据的频率与幅值的对应关系, 为了更好的反应数据中的主要频率分布, 我们从原序列中去除不相关的 IMF 分量以及能量较对大趋势项后绘制其边际图谱如图 6 所示。

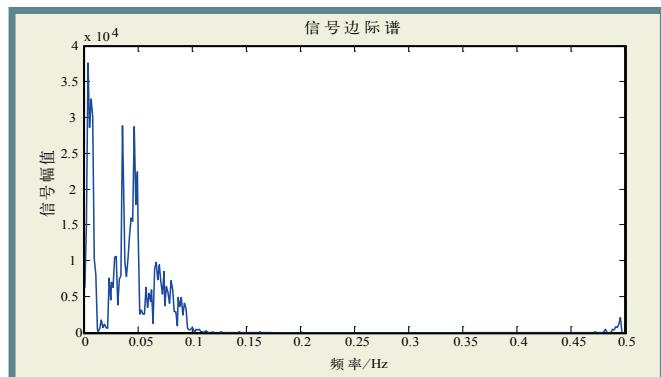


图6 客流数据边际图谱示意图

可以看出其幅值的局部极值点分布频率为: 0.002 76 Hz、0.034 53 Hz、0.066 29 Hz、0.498 62 Hz 处, 对应的周期分别为 362 天、30 天、15 天、2 天, 即全年客流序列大致呈现出以 1 年、1 月、15 日以及 2 日的周期分布。其中, 以 2 日为周期的能量幅值虽不高, 但可以明显与周围幅值分布区分出来, 因此推测为节假日的短时旅游客流大幅增加所致。

3 数据实测与分析

本文以 2013 年 1 月 1 日 ~12 月 31 日数据作为历史数据, 预测的目标时间范围为 2014 年 1 月 1 日 ~3 月 30 日。将 GA-BP 神经网络预测与 EEMD-GA-BP 组合预测方法的预测效果进行对比。选用绝对误差率 (MAPE) 以及皮尔逊相关系数 (PEARSON) 作为比较依据, IMF 的 GA-BP 预测结果如表 1 所示。

表1 IMF分量预测结果

	MAPE	PERSON
IMF1 (GA-BP)	31.26%	0.48
IMF2 (GA-BP)	8.03%	0.94
IMF3 (GA-BP)	0.58%	0.99
IMF4 (GA-BP)	0.17%	1.00
IMF5 (GA-BP)	0.01%	1.00
IMF6 (GA-BP)	0.02%	1.00
IMF7 (GA-BP)	0.01%	1.00
R7 (GA-BP)	0%	1.00

由表 1 可以看出高频成分 IMF1 的预测结果不理想, 但考虑到 IMF1 分量本身与原数据的相关性不高, 因此可以考虑将其剔除。其余分量的预测误差在 10% 以内, 同时 Pearson 相关系数呈现强正相关。其中 IMF3 ~ 7, R7 的分量预测结果已经十分精确, 由此可以看出对于分量的预测精度随分量的频率下降而提高。

为了探究分量重构的在该情景下的最佳方法, 我们做如下尝试:

- (1) 将所有 8 个 IMF 分量求和作为预测结果;
- (2) 使用所有 8 个 IMF 分量, 将其作为 GA-BP 神经网络的输入, 对当日客流作为输出进行训练并测试;
- (3) 使用所有 8 个 IMF 分量, 以及 GA-BP 预测

(下转 P41)

问题，采用科学的方法，善于进行总结和分析，这样才能把定测工作做好，保证安装数据的准确性。

参考文献：

- [1] 中华人民共和国建设部，中华人民共和国质量监督检验检疫局. GB 50308—2008 城市轨道交通工程测量规范 [S]. 北京：中国建筑工业出版社，2008.

- [2] 郑国华. 地铁车站设备安装调试技术 [M]. 北京：中国建筑工业出版社，2010，5.

- [3] 李应洪. 高速铁路定测线位控制探讨 [J]. 铁道勘测与设计，1998（4）.

- [4] 冉龙华. 线路定测曲线调整的方法及其应用 [J]. 铁道标准设计，2005（2）.

责任编辑 徐侃春

（上接 P33）

的客流数据作为输入，对当日客流作为输出进行训练并测试。

（4）使用相关性较强的 IMF2 ~ 7, R7, 以及 GA-BP 预测的客流数据作为输入，对当日客流作为输出进行训练并测试。

最终的预测结果如表 2 所示。

表2 各重构方案预测结果

方案	MAPE	PEARSON
GA-BP直接预测	11.99%	0.772 4
方案（1）	6.02%	0.890 0
方案（2）	5.51%	0.960 1
方案（3）	8.36%	0.892 2
方案（4）	5.80%	0.772 7

观察表 2，可以看出使用 GA-BP 算法对客流数据直接预测的精度是最差的，对 IMF 简单求和的方案效果亦不理想。对比方案（3）与方案（4），可以看出剔除 IMF1 分量可以显著减少分量预测误差对最终精度的影响。对比方案（2）与方案（3），可以看出在输入集中加入 GA-BP 直接预测的结果并无必要，反而会因为预测的中间误差使最终精度下降。方案（2）与方案（4）的预测精度相差不多，但是相关性相差较大。可以推测 IMF1 作为高频分量，预测结果虽不理想，但是在相关性方面的贡献较大，其中包含了较多原数据的高频波动信息。

综上，方案（2）是较优的方案。方案（2）的预测结果与实际值的对比如图 7 所示。

4 结束语

本文描述了 EEMD-GA-BP 的组合预测算法，并以泰安站到达客流数据为例验证了其相较于传统预测算法（如 GA-BP 算法）具有更高的预测精度，在

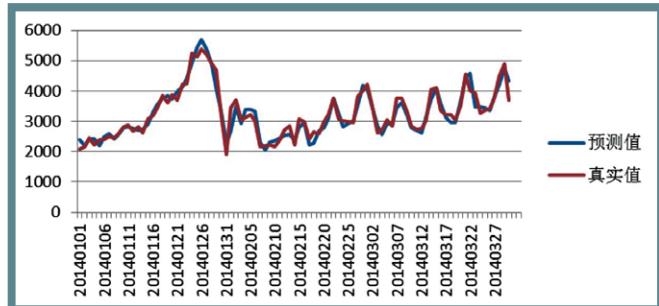


图7 方案（2）预测结果示意图

其基础上对比了多个 IMF 的重构方案，并确定了较优方案，使 IMF 的中间预测误差对最终结果影响较小，保证了最终预测结果的准确。

参考文献：

- [1] 许廷霆，李克平. 应用 EMD 方法分析交通流量的动态特性 [J]. 科学技术与工程，2009（11）：3003-3008.
- [2] 黄建国，罗航，王厚军，等. 运用 GA-BP 神经网络研究时间序列的预测 [J]. 电子科技大学学报，2009, 38（5）：687-692.
- [3] 白玮莉，刘志刚，彭权威，等. 基于 HHT 和神经网络组合的负荷预测模型研究 [J]. 电力系统保护与控制，2009（19）：31-35.
- [4] 潘亮. 基于 EEMD、GSVM 的高速铁路短期客流预测 [D]. 北京：北京交通大学，2012.
- [5] 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.
- [6] 杨柳青. 基于 EEMD 的航空客流需求预测技术及其应用研究 [D]. 武汉：华中科技大学，2009.
- [7] 甄静. 京沪线铁路客流规律分析 [J]. 中国铁道科学，2002, 23（2）：122-126.
- [8] 钟佑明，秦树人. 希尔伯特-黄变换的统一理论依据研究 [J]. 振动与冲击，2006, 25（3）：40-43.

责任编辑 徐侃春