基于时间序列模型下组合预测方法的研究及应用

李晨　邵明月

（商丘工学院　河南　商丘　476000）

[摘　要]本文从时间序列的研究背景及意义出发，主要介绍了时间序列模型的发展状况，并列举了 $ARIMA(p,d,q)$ 模型的具体建模步骤以及灰色预测模型 $GM(1,1)$ 及方法体系，最后给出了在时间序列模型下如何构造组合预测模型。从而说明了此研究的可行性。

[关键词]时间序列；　模型；　模型；组合预测

【DOI】10.12252/j.issn.2096-6288.2021.09.1843

一、研究的背景与意义

自从人类诞生以来，不管人们是否已经认识到它，预测活动就已经存在了，但是预测科学在二十世纪四十年代才真正进入萌芽时期，当时的预测研究还只是停留在哲学概念的纯理论与学术的研究上，谈不上什么具体应用；到了二十世纪后期，对于预测方面的研究工作开始从之前的纯理论逐步上升到应用上。科学技术的重要作用，尤其是它在给社会带来巨大利益的同时，又会给社会带来一些令人担心的不良后果，于是，预测研究就更引起了人们的关注，预测研究的领域也在不断扩大，研究的方法也在逐渐完善。直至今日，预测科学已经成为一门发展迅速、应用十分广泛的新学科[1]。

然而在实际预测过程中，由于不同的建模结构和思考问题的方式，一般情况下，不同种类的预测方法可能适用于相同的预测问题，在这种情况下很难选出最好的预测方法，但将不同种类的预测方法进行适当的组合形成新的预测方法，新的预测方法含有单个方法所包含的信息，在一定程度上提高了模型上午准确度。

在最近这几十年期间，组合预测模型作为一种相对前卫的预测方法，伴随着不同单一模型的优化，组合预测模型在理论研究和实际应用中在不断更新，如今基本成为各种预测方法中的一个比较常用的方法。如今在不同领域，存在着很多比较成熟组合预测方法，比如经济领域[2]、计算机领域[3]、社会领域[4]、工业领域[5]等等。由于组合预测模型在不同领域之间的广泛使用，使得它逐渐成为一个比较热点的问题在现有的预测方法中。不仅仅在国内，在国外也是如此。

二、主要的预备知识

（一）时间序列

时间序列就是数据按照某种给的指标进行排列而成的序列，时间序列分析则是将之前所形成的序列进行进一步的研究及分析，从而找出来一个模型，而这个模型恰到好处的呈现出给定现象的发展规律，然后应用这个模型来预测未来事物发展的规律。

在数理统计的研究中，人们常把具有时间顺序的一个随机事件排成一组随机变量的形式例如：

$$X_1,X_2,X_3,\cdots,X_t,\cdots \qquad (2\text{-}1)$$

来表示，简记为 $\{X_t,t\in T\}$ 或 $\{X_t\}$，T 为时间的集合。

该随机序列的 n 个观察值用

$$X_1,X_2,\cdots,X_n \text{或} \{X_t,t=1,2,\cdots,n\} \qquad (2\text{-}2)$$

表示，而这个观察值的序列长度表示为 n，其中（2-1）式叫做（2-2）式的一个实现。

（二） $ARIMA(p,d,q)$ 模型的具体建模步骤如下：

1. 数据白噪声检验

假如一开始的序列具有白噪声的特点，那么将不作其他处理，直接通过序列图和有关统计量作进一步的研究。

2. 数据异方差检验

假如一开始的序列不具有白噪声的特点，那么将对其进行异方差检验，假如该序列有异方差，那么对其进行Box-Cox转换。

3. 去除资料趋势和季节趋势以及模型识别

假如一开始的序列不具有平稳性，可通过Box-Cox转换使其具有平稳性，这个时候这组数据就不具有平稳性这一假设，便可以进行差分法让其符合假设，从而可应用 $ARIMA(p,d,q)$ 模型。

4. 模型的检验与修正

假如 $\{y_n\}$ 被估计为 $ARIMA(p,d,q)$ 序列，即 $\Phi_p(B)\nabla^d y_n=\Theta_q(B)\varepsilon_n$，且模型是平稳的和可逆的，那么 $\varepsilon_n=\Theta_q^{-1}(B)\Phi_p(B)\nabla^d y_n$ 就应当为白噪声序列。因此如果能从样本序列 $y_1,y_2,\cdots,y_n$ 中求得 ε_n 的样本值 $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_n$，那么就可以对“$\varepsilon_n$是白噪声序列”这一概念进行数理统计中的假设检验。若命题为肯定的，则估计模型拟合成功；否则模型拟合失败。

5. 模型的参数估计

如果改进的模型可以通过验证的 $ARIMA(p,d,q)$ 序列，那么将可利用最大概拟法和非线性最小平方法进行参数校对。也就是依据现有的n个数据对 $\{y_n\}$ 的模型 $\varphi_1,\cdots,\varphi_p,\theta_1,\cdots,\theta_q,\sigma_\varepsilon^2$ 等参数做出估计。进行参数估计的模型中所需的每一个变量的系数都应该是经过t检验而显著的。

（三）灰色预测模型 $GM(1,1)$ 及方法体系[29]

1. $GM(1,1)$ 模型的基本形式

在灰色系统中GM（1，1）模型是其基础，如果序列 $y^{(0)}$ 为非负准光滑序列，则 $y^{(0)}$ 1-AGO序列 $y^{(0)}$ 具有指数规律。

定义2.3.1 称 $y^{(0)}(k)+az^{(1)}(k)=b$ 为GM（1，1）模型的灰色微分方程[30]，其中 $az^{(1)}(k)=0.5y^{(1)}(k)+0.5y^{(1)}(k-1)$。

定理2.3.1 设 $y^{(0)}$ 非负序列：

$$y^{(0)}=\left(y^{(0)}(1),y^{(0)}(2),\cdots,y^{(0)}(n)\right),$$

其中 $y^{(0)}(k)\geq 0,k=1,2,3,\cdots,n$；

$y^{(1)}$ 为 $y^{(0)}$ 的1-AGO序列：

$$y^{(1)}=\left(y^{(1)}(1),y^{(1)}(2),\cdots,y^{(1)}(n)\right)$$

其中 $y^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(k),k=1,2,\cdots,n$；

$z^{(1)}$ 为 $y^{(1)}$；的紧邻均值生成序列：

$$z^{(1)}=\left(z^{(1)}(2),z^{(1)}(3),\cdots,z^{(1)}(n)\right)$$

其中 $z^{(1)}(k)=0.5y^{(1)}(k)+0.5y^{(1)}(k-1),k=2,3,\cdots,n$

若$\hat{a}=[a,b]^T$为参数列，且

$$Y=\begin{bmatrix} y^{(0)}(2) \\ y^{(0)}(3) \\ \cdot \\ \cdot \\ \cdot \\ y^{(0)}(n) \end{bmatrix} \qquad B=\begin{bmatrix} -w^{(1)}(2) & 1 \\ -w^{(1)}(3) & 1 \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ -w^{(1)}(n) & 1 \end{bmatrix}$$

则灰色微分方程$y^{(0)}(k)+az^{(1)}(k)=b$的最小二乘估计参数列满足$\hat{a}=(B^TB)^{-1}B^TY$，则称

$$\frac{dy^{(1)}}{dt}+ay^{(1)}=b$$

为灰色微分方程$y^{(0)}(k)+az^{(1)}(k)=b$的白化方程，也叫影子方程。

定理2.3.2 设$B,Y,\hat{a}$如定理2.3.1所述，$\hat{a}=[a,b]^T=(B^TB)^{-1}B^TY$，则

（1）白化方程$\frac{dy^{(1)}}{dt}+ay^{(1)}=b$的解也称时间响应函数为

$$y^{(1)}(t)=\left(y^{(1)}(0)-\frac{b}{a}\right)e^{-at}+\frac{b}{a}$$

（2）GM（1，1）灰色微分方程[32]$y^{(0)}(k)+az^{(1)}(k)=b$的时间响应序列为

$$\hat{y}^{(1)}(t)=\left(y^{(1)}(0)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a};k=1,2,\cdots,n$$

（3）取$y^{(1)}(0)=y^{(0)}(1)$则

$$\hat{y}^{(1)}(k+1)=\left(y^{(1)}(1)-\frac{b}{a}\right)e^{-ak}+\frac{b}{a};k=1,2,\cdots,n$$

（4）还原值$\hat{y}^{(0)}(k+1)=y^{(1)}(k+1)-y^{(1)}(k);k=1,2,\cdots,n$

（四）模糊时间序列

定义2.4.1　定义U为论域，将论域U划分为有限个有序子集，即为模糊区间，可表为：$U=\{u_1,u_2,\cdots,u_n\}$。假定A为论域U上的语义变量集，即模糊集，定义为：

$$A=\frac{f_A(u_1)}{u_1}+\frac{f_A(u_2)}{u_2}+\cdots+\frac{f_A(u_n)}{u_n}$$

其中f_A为模糊集A上的模糊隶属度函数f_A：$U\to[0,1]$，$f_A(u_i)$表示u_i在模糊集A上的模糊隶属度的值，并且$f_A(u_i)\in[0,1],1\le i\le n$。

在此基础上进行扩展，给出模糊时间序列的概念及相关定义。

定义2.4.2假设$Y(t)(t=0,1,2,\cdots)$为一个实数子集，定文$f_i(t)(t=1,2,\cdots)$为$Y(t)$论域上的一组模糊集，如果$F(t)$包含其中所有的模糊集$f_i(t)(t=1,2,\cdots)$，即$F(t)=\{f_1(t),f_2(t),\cdots\}$，则把$F(t)$定义为$Y(t)$上的一个模糊时间序列。

定义2.4.3设$R(t,t-1)$为定义在$F(t-1)$到$F(t)$的模糊关系，且满足$F(t)=F(t-1)\circ R(t,t-1)$，则称$F(t)$是由$F(t-1)$通过模糊关系$R(t,t-1)$推导得到的，且可以用模糊逻辑关系$F(t-1)\to F(t)$来表示。其中，“o”代表合成运算，$F(t)$和$F(t-1)$都是模糊集，关系$R$称为定义$F(t)$上的一阶模糊关系。

定义2.4.4令$F(t-1)=A_i,F(t)=A_j$，一个模糊逻辑关系可以表示在两个连续的观测值$F(t-1)$和$F(t)$之间，记为$A_i\to A_j$　这里A_i就是人们常说的该关系的前件（或者左件），而A_j则叫做该关系的后件（或者右件）。

定义2.4.5假设$F(t)$为一个模糊时间序列，且$F(t)$是由$F(t-1),F(t-2),\cdots,F(t-n)$通过模糊逻辑关系推导得到的，且$F(t)=A_j,F(t-1)=A_{i_1},F(t-2)=A_{i_2},\cdots,F(t-n)=A_{i_n}$那么当模糊时间序列模型为$n$阶时模糊逻辑关系表示为：$A_{i_n},\cdots,A_{i_2},A_{i_1}\to A_j$。

三、主要方法

首先，论文把*ARIMA*模型和残差自回归模型采用适当的方法构成一种新的组合预测模型，并利用新生儿人口的数据进行了预测。

其次，论文利用*ARIMA*模型对时间序列数据进行模型识别和拟合，之后由灰色预测模型与优化后的灰色$GM(1,1)$模型对时间序列数据进行拟合与预测，最后通过赋予合理的权重再结合这些方法提出组合预测模型，并进行了实例验证。

再次，将模糊的时间序列和经典的时间序列进行有效结合，提出一个结合趋势预测和自回归模型的模糊时间序列预测模型。新的模型能够呈现出时间序列中所包含的明显趋势，接着依据*ARIMA*模型来求出预测数据的波动量，进而得出最后的预测值。将前文给出的两个预测模型分别与到现实中的时间序列相结合，然后将实验得出的数据与别的同类型的预测模型进行对比。

最后大胆尝试将*ARIMA*模型、灰色预测模型、模糊时间序列通过赋予合理的权重再次组合，并进行实例验证，比较各种发法的优缺点，以寻找到最合理的预测模型。

四、总结

本文通过大量阅读相关领域的文献，认真研究关于时间序列、灰色预测模型和模糊序列的相关文献，综述本课题研究背景及意义，给出了时间序列、灰色模型以及模糊序列的定义及相关的基本性质，依据建立新的组合模型的方法去预测模型，为今后论文的编写工作打下了牢固的基础。

参考文献

[1]赵仁义，朱玉辉.关于时间序列预测法的探讨[J].科技信息，2011(15)：192-194.

[2]温青，张筱慧.杨旭基于负荷误差和经济发展趋势的组合预测模型在中长期负荷预测中的应用[J].电力系统保护与控制，2011(03)：57-61.

[3]程跃华，崔艳.组合模型在软件可靠性预测中的建模与仿真[J].计机仿真，2011(06)：371-374.

[4]Zhou Q，Ren H J，Li J，Zhang Y，et al.Variable Weight Combination Method for Mid-long Term Power Load Forecasting Based on Hierarchical Structure[J]. Zhongguo Dianji Gongcheng Xuebao/Proceedings of the Chinese Society of Electrical Engineering，2010(16)：47-52.

[5]Thordarson F O，Madsen H，Nielsen H A，et al.Conditional Weighted Combination of Wind Power Forecasts[J].Wind Energy，2010(08)：751-76\3.

作者简介:

李晨(1987-)，女，河南省商丘市人，所在院校：商丘工学院，讲师，硕士，研究方向：数理统计；

邵明月(1993—)，女，河南省商丘市人，所在院校：商丘工学院，助教，硕士，研究方向：传染病稳定性方面。