

# 时间序列分析的理论与应用综述

罗芳琼, 吴春梅

(柳州师范高等专科学校 数学与计算机科学系, 广西 柳州 545004)

**摘 要:** 时间序列分析提供的理论和方法是进行大型高难度综合课题研究的工具之一。其预测和评估技术相对比较完善, 其预测情景也比较明确。近年来已有很多学者对于时间序列的研究取得了极其丰硕的成果, 有的甚至在时间序列分析方法的基础上, 研究出新的预测方法, 在应用中求创新求发展。笔者从基本理论与应用等方面对时间序列分析进行了综述, 同时阐述了它未来的发展趋势。

**关键词:** 时间序列分析; 非线性; 数据挖掘

**中图分类号:** O236

**文献标识码:** A

**文章编号:** 1003 - 7020(2009)03 - 0113 - 05

时间序列分析不仅可以从数量上揭示某一现象的发展变化规律或从动态的角度刻划某一现象与其他现象之间的内在数量关系及其变化规律性, 达到认识客观世界之目的, 而且运用时间序列模型还可以预测和控制现象的未来行为。许多经济、金融、商业等方面的数据都是时间序列数据, 对这些数据进行分析、处理和研究, 从中挖掘有用信息是广大工作者当前研究的焦点之一。目前时间序列的预测和评估技术相对比较完善, 其预测情景也比较明确, 综合他人的智慧、借助各种资料, 本文介绍了时间序列分析的基本理论及其进展, 阐述了它目前的应用领域及未来的发展趋势。

## 1 时间序列分析产生的背景

7000 年前的古埃及人把尼罗河涨落的情况逐天记录下来, 就构成所谓的时间序列。对这个时间序列长期的观察使他们发现尼罗河的涨落非常有规律。象古埃及人一样按照时间的顺序把随机事件变化发展的过程记录下来就构成了一个时间序列, 对时间序列进行观察、研究, 找寻它变化发展的规律, 预测它将来的走势就是时间序列分析。早期的时间序列分析通常都是通过直观的数据比较或绘图观测, 寻找序列中蕴含的发展规律, 这种分析方法就称为描述性时序分析。古埃及人发现尼罗河泛滥的规律就是依靠这种分析方法。但随着研究领域的不断拓广, 在很多研究领域随机变量的发展通常会呈现出非常强的随

机性, 人们发现依靠单纯的描述性时序分析已不能准确地找出随机变量发展变化的规律, 为了更准确地估计随机序列发展变化的规律, 从 20 世纪 20 年代开始, 学术界利用数理统计学原理分析时间序列, 研究的重心从表面现象的总结转移到分析序列值内在的相关关系上, 由此开辟了一门应用统计学科——时间序列分析<sup>[1]</sup>。

时间序列分析方法最早起源于 1927 年数学家 Yule 提出建立自回归模型 (AR 模型) 来预测市场变化的规律。1931 年, 另一位数学家在 AR 模型的启发下, 建立了移动平均模型 (MA 模型), 初步奠定了时间序列分析方法的基础。20 世纪 60 年代后, 时间序列分析方法迈上了一个新的台阶, 在工程领域方面的应用非常广泛。近几年, 随着计算机技术和信号处理技术的迅速发展, 时间序列分析理论和方法更趋完善。

## 2 时间序列分析的基本思想与理论进展

不论是经济领域中每年的产值、国民收入、某一商品在某一市场上的销量、价格变动等, 或是社会领域中某一地区的人口数、医院患者人数、铁路客流量等, 还是自然领域的太阳黑子数、月降水量、河流流量等等, 都形成了一个时间序列。根据这些时间序列, 较精确地找出相应系统的内在统计特性和发展规律

[收稿日期] 2008 - 11 - 25

[基金项目] 广西自然科学基金 (0832092); 广西教育厅科研项目 (200707MS061); 柳州师专基金项目 (LSZ2008A002)

[作者简介] 罗芳琼 (1971 —), 女 (壮族), 广西忻城人, 讲师, 研究方向: 计算机网络及神经网络应用; 吴春梅 (1970 —), 女, 讲师, 研究方向: 计算机应用及神经网络应用。

性,从中提取人类所需要的准确信息的方法就是时间序列分析。它是一种根据动态数据揭示系统动态结构和规律的统计方法。其基本思想是根据系统的有限长度的运行记录(观察数据),建立能够比较精确地反映时间序列中所包含的动态依存关系的数学模型,并借以对系统的未来行为进行预报。

## 2.1 时间序列分析基本特征

2.1.1 时间序列分析法是根据过去的变化趋势预测未来的发展,它的前提是假定事物的过去延续到未来

时间序列分析,正是根据客观事物发展的连续规律性,运用过去的历史数据,通过统计分析,进一步推测未来的发展趋势。事物的过去会延续到未来这个假设前提包含两层含义:一是不会发生突然的跳跃变化,是以相对小的步伐前进;二是过去和当前的现象可能表明现在和将来活动的发展变化趋向。这就决定了在一般情况下,时间序列分析法对于短、近期预测比较显著,但如延伸到更远的将来,就会出现很大的局限性,导致预测值偏离实际较大而使决策失误。

2.1.2 时间序列数据变动存在着规律性与不规律性

时间序列中的每个观察值大小,是影响变化的各种不同因素在同一时刻发生作用的综合结果。从这些影响因素发生作用的大小和方向变化的时间特性来看,这些因素造成的时间序列数据的变动分为四种类型。

(1)趋势性:某个变量随着时间进展或自变量变化,呈现一种比较缓慢而长期的持续上升、下降、停留的同性质变动趋向,但变动幅度可能不相等。

(2)周期性:某因素由于外部影响随着自然季节的交替出现高峰与低谷的规律。

(3)随机性:个别为随机变动,整体呈统计规律。

(4)综合性:实际变化情况是几种变动的叠加或组合。预测时设法过滤除去不规则变动,突出反映趋势性和周期性变动。

## 2.2 时间序列模型

时间序列中的模型,常见的有:

### 1. 自回归 AR(p) 模型

仅通过时间序列变量的自身历史观测值来反映有关因素对预测目标的影响和作用,不受模型变量相互独立的假设条件约束,所构成的模型可以消除普通回归预测方法中由于自变量选择、多重共线性等造成的困难。

### 2. 移动平均 MA(q) 模型

用过去各个时期的随机干扰或预测误差的线性组合来表达当前预测值。AR(p)的假设条件不满足时可以考虑用此形式。

### 3. 自回归移动平均 ARMA(p, q) 模型

使用两个多项式的比率近似一个较长的 AR 多项式,即其中  $p+q$  个数比 AR(p) 模型中阶数  $p$  小。

前二种模型分别是该种模型的特例。一个 ARMA 过程可能是 AR 与 MA 过程、几个 AR 过程、AR 与 AR-MA 过程的迭加,也可能是测度误差较大的 AR 过程。

## 4. 自回归综合移动平均 ARIMA(p, d, q) 模型

模型形式类似 ARMA(p, q) 模型,但数据必须经过特殊处理。特别当线性时间序列非平稳时,不能直接利用 ARMA(p, q) 模型,但可以利用有限阶差分使非平稳时间序列平稳化,实际应用中  $d$  一般不超过 2。

若时间序列存在周期性波动,则可按时间周期进行差分,目的是将随机误差有长久影响的时间序列变成仅有暂时影响的时间序列。即差分处理后新序列符合 ARMA(p, q) 模型,原序列符合 ARIMA(p, d, q) 模型。

时间序列建模基本步骤是:用观测、调查、统计、抽样等方法取得被观测系统时间序列动态数据。

根据动态数据作相关图,进行相关分析,求自相关函数。辨识合适的随机模型,进行曲线拟合,即用通用随机模型去拟合时间序列的观测数据。对于短的或简单的时间序列,可用趋势模型和季节模型加上误差来进行拟合。对于平稳时间序列,可用通用 ARMA 模型(自回归滑动平均模型)及其特殊情况的自回归模型、滑动平均模型或组合 ARMA 模型等进行拟合。当观测值多于 50 个时一般都采用 ARMA 模型。

## 2.3 时间序列分析的理论进展

非线性模型理论和单位根理论是时间序列分析在理论上的进展的主要表现。非线性模型理论的进展集中在几何遍历性问题和非线性过程的平稳性这两方面。对于简单模型 TAR(1),Chen 和 Tsay (1991),Petrucelli 和 Woolford 得出了一些有趣的结论。在非线性过程的稳定性研究中作出杰出贡献的有 Meyn 和 Tweedie 以及 Tong(1990)等人<sup>[2]</sup>。

近年来,在时间序列分析理论中发展比较快的是单位根理论。这一理论主要研究随机漫步过程统计量的非对称性质。单位根问题已经引起了越来越多计量经济学家和统计学家的关注。它不但为决定 ARIMA 模型差分的阶提供了正式的检验方法,也为一些统计量的检验开辟了新的领域。Tsay 和 Tiao (1990)将单位根检验扩展到多元情形,这就是所谓的协整检验。

在非线性时间序列分析方面我国学者也取得了一定的成果。姚琦伟教授基于信息量,首次提出了描述一般随机系统对初始条件敏感性的度量及估计方法。在高维模型领域,姚琦伟教授提出用复系数线性模型近似高维非线性回归函数的新方法,以此克服高维非参数回归中样本量短缺的困难问题。此方法在生物、经济、金融等应用中获得了成功。在时间序列

模型的最大似然估计方法的研究中,他完整地建立了在金融风险管理中有直接应用的 ARCH 和 GARCH 模型为最大似然估计的极限理论。对于重尾部 (heavy - tailed) 分布模型,提出了基于 bootstrap 的新的估计方法以及稳健统计方法。他还首次建立了在空间域上空间 ARMA 过程的最大似然估计理论。这一工作同时也对 Hannah1973 年给出的关于时间序列的最大似然估计理论首次给出了一个完整的时域上的证明。汤家豪教授将有关非线性时间序列分析的研究与动力系统科学的模型连接而深受人们的赞赏。安鸿志、朱力行、陈敏关于非线性自回归模型的平稳性、遍历性和高阶矩的成果,获得了有这些性质的最弱条件。他们研究了条件方差为非常数的回归和自回归模型的平稳性、遍历性和检验方法<sup>[3]</sup>。

### 3 时间序列分析的应用进展

统计学的一个重要任务就是从变动的数据中找出事物的内在本质和运动规律,并最终进行预测和控制。由于现实中各种因素错综复杂,运用多元回归等静态因果结构模型进行分析预测,往往比较困难,而根据事物自身变动情况建立动态模型——时间序列分析,则是一种行之有效的方法。近些年来,时间序列分析已广泛应用于心理学、地球科学、数据挖掘、数字化误差等各个方面。

#### 3.1 时间序列分析方法在心理学中的应用

20 世纪末时间序列分析的方法开始引入到心理学领域,目前应用最多的主要集中在心理治疗领域、心理动力学的研究、管理心理学的研究、社会心理和家庭心理的研究等研究领域。

时间序列分析方法应用在心理学中的重要价值在于它不仅能比较个体间心理量发生发展的趋势,确定动态变化变量间的因果关系,而且可以实现对人类心理和行为的预测和控制。它既是一种研究思想也是一种新的数据分析方法,比较贴近人类心理的真实情境,有利于心理学工作者对心理学问题开展更加深入的研究。在国外,时间序列分析方法作为研究手段在心理学研究中不管是理论还是应用已经取得了一系列的进展。姚懿、李小平的时间序列分析方法在心理学中的应用,大致代表了近几年时间序列分析在国外心理学界的发展趋势。常用的模式从单变量的时间序列分析,到两变量的时间序列分析,再到更复杂的多变量的时间序列分析。对时间序列分析的应用也从定性的现象描述到定量的建立模型,从分析简单的问题到复杂的问题<sup>[4]</sup>。可见,时间序列分析的方法在心理学领域里正在全面发展,这也应该引起国内心理学工作者的重视、学习和借鉴新的研究范式,对推动我国心理学的发展有着不可估量的作用。因此,我们可以预见时间序列分析方法在心理学中的研究领

域将越来越广,影响将越来越大,心理学也将逐渐与科学发展的总体趋势保持一致,尤其是在借鉴了新的研究方法后,将使我们的研究越来越接近真实的人的心理。

#### 3.2 时间序列分析方法在数据挖掘中的应用

各种类型的数据都可以作为数据挖掘的对象,时间序列在数据集中十分普遍,对时间序列进行数据挖掘已成为当前研究的焦点之一。

近年来,时间序列分析方法在数据挖掘中的应用也取得了一定的进展。学者们利用数据挖掘对象,根据时间序列分析方法,提出了基于模糊集合的数据挖掘时间序列模式算法;根据某些时间序列所具有的分形特征,分析了利用分形理论中的 R/S 分析,发现具有分形特征的时间序列模式的方法;对大型数据库的海量数据分析提出了进行时间序列模式挖掘的算法,为用户的决策支持和趋势预测提供了依据。尤其是刘劲松的数据挖掘中的现代时间序列分析方法,是目前处理海量时间序列数据挖掘的一种新的非常适用的方法,其文中提出要解决时变参数系统的自适应预测问题,目前最理想、最有效、最适合的工具就是现代时间序列分析方法<sup>[5]</sup>。同时文献[6]和文献[7]中给出大量的应用实例,也充分说明该方法的有效性。

#### 3.3 时间序列分析在数字化误差处理中的应用

通常数字化数据的获取是按某种时间顺序或空间顺序来实现的,而这些数据的采集不是完全随机的,而是始终围绕着某目标线段进行的,因此得到的数据误差序列具有序列相关的性质,可以采用时间序列分析方法来处理。朱光、张保钢应用时间序列分析理论中的 AR(p) 自回归模型研究了数字化误差的数学模型,并用实例说明该方法是有效的<sup>[8]</sup>。由于 AR(p) 模型的参数可以调整,具有简单的解析表达形式,应用灵活,因此也适用于对各种图形的误差模拟。除此之外,使用 AP(p) 模型来描述数字化误差,不仅可以对各种地图精度标准进行比较,而且还可以对多边形面积或弧段长度的影响进行评价以及对图形复杂性的影响进行评估。

#### 3.4 非线性时间序列分析在地球科学中的应用

地球科学中有大量由观测得到的时间序列,人们需要通过这些序列认识各种地学现象的内在规律,对各种序列进行预报,而这些地学现象中,大部分是非线性的,因此,非线性时间序列分析方法在地球科学的应用已经成为研究者进行研究的一个新领域。

非线性时间序列分析的理论和方法的研究与应用紧密地联系在一起。1991 年,著名的圣塔菲研究所专门组织了一次“圣塔菲时间序列预报与分析竞赛”(Santa Fe Time Series Prediction and Analysis Competition),事后,又召开了专门的学术讨论会,出版了文集,推动了这一领域的发展。非线性时间序列

分析在地球科学中早期应用的一个著名例子是气候吸引子的研究<sup>[9]</sup>。

### 3.5 时间序列分析在股市中的应用

在证券领域中得到的观测数据列一般都具有较强的时间变化趋势,股票价格的数据都是以时间序列的形式出现的。因此,采用时间序列分析法对股市数据进行分析预测是可行的,很多文献都应用了具体的事例来说明它的有效性。

目前,采用非线性时间序列分析技术对股市指数序列和指数收益率序列的研究已取得了可喜的成绩。学者们用上证指数高频数据分析了价格波动的非线性特征,通过重构相空间方法重构了上证指数时间序列的奇怪吸引子,确认了上证指数时间序列的混沌行为;部分学者在对股票价格时间序列进行混沌性分析中,还提出了一种新的基于转折指标量的股票价格时间序列拐点预测模型,该模型在股票时间序列预测中有一定的实用价值。

### 3.6 时间序列分析在地下水位预报中的应用

在实际问题中,地下水水位动态常常与大气降水量和开采量等因素密切相关,而这些因素表现了一定的趋势性、周期性和随机性,因此,采用时间序列分析法来预报地下水水位动态是一种可靠有效的方法。近20年来,时序分析理论已应用于地下水资源评价、预报和管理之中。常用于地下水位预报的时序模型有自回归模型(AR)、滑动平均模型(MA)、自回归滑动平均模型(AR-MA)。近年来随着计算机技术的发展和广泛应用又发展了多种模型,如求和自回归滑动平均模型(ARIMA)、乘积型季节性模型、组合模型、混合模型以及门限模型、混沌时间序列模型等。这些模型应用灵活方便且具有较高的模拟精度,给大区域地下水动态预报分析带来了极大的便利<sup>[10]</sup>。赵杰、卞玉梅应用时间序列分析法,对近年来沈阳市地下水位动态变化进行分析和预测得出了近年来地下水位经历了由逐渐下降到逐渐回升的过程<sup>[11]</sup>。周晓君、兰双双、王滨运用时间序列分析方法建立了地下水位预报模型,该预报模型为制定该地区水资源可持续利用政策提供了重要的参考依据<sup>[12]</sup>。

### 3.7 时间序列分析在空气污染分析中的应用

众所周知,许多数学方法和计算机技术已应用于大气环境质量评价中,如模糊数学、GIS技术、灰色系统等。这些方法大多是针对某一时刻或某一时间段内的大气污染物进行评价,而大气污染物的浓度和分布具有时间变化特性。因此,数学方法中的时间序列能够反映大气污染物随时间的推移而呈现的变动,为大气环境质量评价提供了有力的支持,并且得到了广泛的应用<sup>[13]~[15]</sup>。柴微涛、宋述军采用时间序列分析方法对成都市2001~2005年的空气污染指数进行分析,探讨成都市空气污染指数的变化规律,通过对

ARMA模型的具体分析,表明应用时间序列分析大气污物状况是比较有效的,也是可行的<sup>[16]</sup>。

## 4 未来展望

虽然目前时间序列分析理论的研究已有了一些进展,在分析预测领域也取得了可喜的成绩,但分析预测是一项艰辛和复杂的工作,同时由于目前已有模型的固有缺陷,这些都极大地影响分析预测的结果。因此,在这方面的研究还需继续深入和广阔拓展。笔者认为进一步的研究和创新工作可在以下几个方面进行:

(1) 现在时间序列数据挖掘研究中提出很多的挖掘算法都是具有普适性的算法,算法的适用面广,但算法的预测效果相对较弱,比如张保稳的时间序列数据挖掘研究中提出的挖掘算法,若能将时序进行分类,然后从每类时序中抽取该类时序的特征并将其引入到算法中,形成各类时序的专用挖掘算法可能会提高算法的性能和效果。随着人工智能、机器学习等学科的发展,普通的基于回归模型、聚类模型和统计模型的挖掘技术需要结合人工智能技术的发展而给出更多更有效的数据挖掘技术,从而更好的支持大规模非线性系统的决策。当前对于数据挖掘中不确定性问题的研究正在成为数据挖掘的一个新的方向,比如模糊数据挖掘研究,现在这方面大部分的研究都集中在关联规则挖掘上,有关时间序列模糊挖掘方面的研究尚亟待开展。

(2) 近几年来,尽管产生了大量的预测方法,但是却没有哪一种方法在任何情况下都能有最好的表现。复杂的方法虽然有“完善的”理论支持,在实践中却不一定有令人满意的结果。朴素的方法以其意外的成功证明着:朴素并非意味着简陋。理论和实践的结果,都要求我们正确考虑预测的策略。目前,实际应用的预测方法有几百种,但如何正确评价预测方法与预测模型的功效并依据具体实际选用合适的预测模型与方法也是广大预测者面临的实际问题。

(3) 对于混沌时间序列的预测问题,用基于混沌吸引子的时间序列预测方法进行短期预测能够获得比较满意的结果。但该方法的应用也是有一定缺陷的,如要求数据样本量要大,必须借助计算机建模并花费很多的上机时间。在实际应用中,预测点邻域半径的选取以及数据中噪声的处理都是值得认真研究的。应用混沌理论研究预测,国内外的相关研究大多局限于应用混沌理论的某些概念作定性的分析与具体方法的开发及应用,从系统演化发展角度,对预测的基本假定、思维定势及框架等的范式转换和建模原理进行系统化研究的成果并不多,理论上的深度、体系上的广度均有待拓展。

(4) 部分研究者用不同的数据将神经网络模型与

Box-Jenkins 模型进行了比较,结果发现对于短期预测,Box-Jenkins 模型要优于神经网络模型,而对于长期预测,则是神经网络模型要优于 Box-Jenkins 模型。还有许多学者也试图利用神经网络模型来进行预测问题的研究,并提出了不同的方法,大多数结果都表明神经网络模型至少与传统的预测方法具有相同的精度,有些超过传统的预测方法。从以上看出,不同的研究结果之间似乎存在着矛盾。实际上,造成这些矛盾的原因与不同的研究采用不同的网络结构有关,此外,数据序列的类型(静态或动态)以及时间序列数据间的关系也会对最后的结果产生影响。因此,应针对某一具体问题来评价神经网络模型的好坏,而不能泛泛地将其与传统的预测方法相比较。

总之,在笔者看来未来时间序列的研究将会朝几个方向发展:海量时间序列数据大多是非平稳的,其特征参数和数学分布是随着时间的推移而发生变化,模型必须跟踪这种变化才能适应当前的数据,准确预测未来;由于人们要求对实际问题讨论越来越精确,越来越深入,使得在实际预测工作中,采用时变参数模型和自适应预测技术的现代时间序列方法已经成为必然;出于研究不同变量间动态关系的需要以及计算机硬件的发展,多元模型在向量 ARMA 模型中或在状态空间模型中的应用会日益增加;BP 神经网络模型引入到混沌相空间,更能反映时间序列的变化趋势。

### [参 考 文 献]

[1] 王燕. 应用时间序列分析[M]. 北京:中国人民大学出版社,2005:3.  
 [2] 程振源. 时间序列分析:历史回顾与未来展望[J]. 统计与决策,2002(9):45-46.  
 [3] 李锐,向书坚. 我国时间序列分析研究工作综述[J]. 统计教育,2006(7):6-8.

[4] 姚懿,李小平. 时间序列分析方法在心理学中的应用[J]. 徐州师范大学学报,2007(5):111-114.  
 [5] 刘劲松. 数据挖掘中的现代时间序列分析方法[J]. 信息技术,2007(7):100-101.  
 [6] 邓自立,郭一新. 现代时间序列分析及其应用——建模、滤波、去卷、预报和控制[M]. 北京:知识出版社,1989:415-418.  
 [7] 刘劲松. 实用经济预测方法[M]. 哈尔滨:哈尔滨出版社,1996:214-217.  
 [8] 朱光,张保钢. 时间序列分析在数字化误差处理中的应用[J]. 北京建筑工程学院学报,1996,12(4):35-45.  
 [9] 洪时中. 非线性时间序列分析的最新进展及其在地球科学中的应用前景[J]. 地球科学进展,1999(14):562.  
 [10] 张小娟,蒋云钟,秦长海,沈媛媛. 时间序列分析在地下水位预报中的应用[J]. 南水北调与水利科技,2007(8):40-42.  
 [11] 赵杰,卞玉梅,周晓君. 时间序列分析法在沈阳市地下水位动态预报中的应用[J]. 东北水利水电,2007(8):31-34.  
 [12] 周晓君,兰双双,王滨. 时间序列分析在四平地区地下水位预报中的应用[J]. 东北水利水电,2007(8):35-37.  
 [13] LEE Chung-kung. Multifractal characteristics in airpollutant concentration time series[J]. Water, Air, and Soil Pollution, 2002(13):389-409.  
 [14] 李天宇,毕经宝. 时间序列在秦皇岛港煤尘污染分析中的应用[J]. 交通环保,2003,24(4):22-23.  
 [15] 黄磊,王踢震. 青岛市大气污染时间序列分析预报方法研究[J]. 青岛海洋大学学报,2001,31(1):14-20.  
 [16] 柴微涛,宋述军,宋学鸿. 成都市城区空气污染指数的时间序列分析[J]. 成都理工大学学报,2007,34(4):485-488.

(责任编辑:谭少班)

## Study of Theory and Application of Time Series Analysis

LUO Fang-qiong, WU Chun-mei

(Department of Mathematics and Computer Science, Liuzhou Teachers College, Liuzhou, Guangxi 545004, China)

**Abstract:** The theory and methods provided by time series analysis is one of the tools to carry out large-scale sophisticated research projects. Its predicting and evaluating technology is relatively perfect, predicting scene clear. In recent years, many scholars have achieved a lot in the study of time series, some of whom even have explored new predicting methods and put them into application. This paper is to carry out a summary study on time series from perspectives of its theory and application, and illuminate its development tendency.

**Key words:** time series analysis; non-linear; data mining