

# 时间序列理论在电信行业预测决策系统中的应用

杨晓波<sup>1</sup> 胡黎伟<sup>2</sup>

<sup>1</sup>(浙江财经学院信息学院,杭州 310012)

<sup>2</sup>(浙江鸿程计算机系统公司,杭州 310005)

E-mail: yxb@zjhcssoft.com

**摘要** 文章介绍了时间序列理论的各个模块算法和流程概况,以及该理论在电信行业预测决策系统中的应用情况,实践证明,时间序列理论作为一种数据挖掘工具,可以较为准确地预测数据未来发展趋势,为企业的经营决策提供帮助。

**关键词** 时间序列 测决策 数据挖掘

文章编号 1002-8331-(2004)28-0070-03 文献标识码 A 中图分类号 TP311.131

## Application of Time Series Theory in the Telecom Forecast and Decision-making System

Yang Xiaobo<sup>1</sup> Hu Liwei<sup>2</sup>

<sup>1</sup>(Dept. of Information, Zhejiang University of Finance & Economics, Hangzhou 310012)

<sup>2</sup>(Zhejiang Hongcheng Computer System Co. LTD., Hangzhou 310005)

**Abstract:** Time series theory is proposed in this paper, which includes several module algorithms and flows survey. The application in the telecom forecast and decision-making system are based on this theory. It proves that time series analysis algorithm can predict the data tendency correctly, which can supply help for the enterprise management and decision-making.

**Keywords:** time series, forecast and decision-making, data mining

### 1 引言

时间序列理论是从具有先后顺序的数据中提取有用信息的一门科学,它是数理统计的一个重要分支<sup>[1-3]</sup>。应用范围非常广泛,遍及自然界、工程界、社会界以及生物学等众多领域。

时间序列通过分析对象(包括总类和细类)的历史数据,预测该对象在将来一段时间内的可能数据,并分析数据的未来发展趋势等等。该文将时间序列理论应用于电信行业的预测决策系统之中,以期从大量的业务数据中,分析并挖掘数据间潜在的有用信息,为企业的经营决策提供依据,该项技术是数据挖掘<sup>[4-6]</sup>技术的进一步延伸。

### 2 时间序列分析算法及流程概述

#### 2.1 输入数据定义

时间序列分析算法是依次分析每个用户对象的。输入数据是从数据仓库中按照要处理的对象取出该对象的数据,它是一组按时间顺序的有序数据。其涉及的数据在数据仓库中包括以下一些表:商业大客户信息表,计费事实表,时间维信息表和计费费用种类表。具体说明如下:

(1)商业大客户信息表。用来确定所要分析的客户对象,搜索相应的客户数据。其过程是在使用者输入一个用户的姓名和客户标识号 ID 后,寻找客户的计费事实表。

(2)计费事实表。计费事实表中保存了客户的所有历史数据,通过一定的限制输出需要的历史数据,限制的形式有:时间的幅度,即输入数据的时间跨度。理论上来说,时间越长效果越

好,但考虑到市场的变化,特别是剧变发生后,该时间以前的历史数据会变得没有价值,甚至有干扰的可能。因此,一般取一到两年的历史数据为宜。还有一个限制是数据的粒度,数据粒度越大,越笼统,但基本上还是表现出重要的数据特征价值;数据粒度越小,效果会越好。因此,根据能提供的数据粒度程度,尽量做到粒度的最小化,但也要在运行时间许可的范围内做到最精。

(3)时间维信息表。用来控制数据的粒度大小。一般来说,数据是以月为单位的。

(4)计费费用种类表。提供按费用分类进行预测的参考。

输入数据的处理过程如图 1。

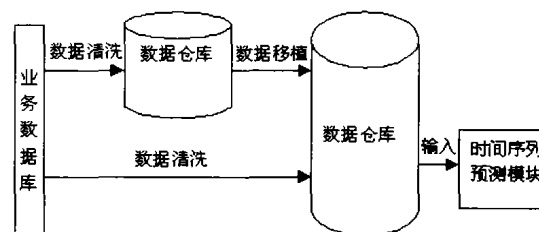


图 1 输入数据的处理流程

从图中可知,业务数据经清洗或移植后,送入数据仓库之中,数据仓库可以进行数据清理和数据集成,为统一的历史数据分析提供坚实的平台;处理后的数据被输入到时间序列预测模块中进行预测分析。

**作者简介:**杨晓波(1971-),男,博士研究生,浙江财经学院副教授,主要研究领域:数据仓库和数据挖掘应用研究。

## 2.2 算法及流程介绍

采用时间序列模块算法的总体结构如下:

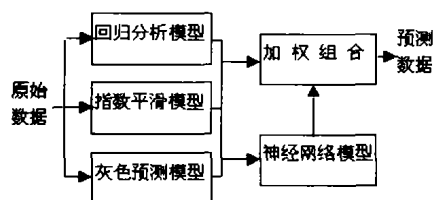


图2 组合预测核心算法模型

预测模型中,回归分析模型、指数平滑模型和灰色预测模型同时接受原始数据(即待预测量的历史数据序列),产生各自的预测值,如 $f(t), g(t), h(t), t=1, 2, \dots$ 。接下来,这些预测值经过加权组合,就得到最后的预测数据,即:

$$C(t)=w_f f(t)+w_g g(t)+w_h h(t), t=1, 2, \dots \quad (1)$$

其中 $w_f, w_g, w_h$ 为加权值, $C(t)$ 是最终的预测值。加权组合运算中的加权值是利用人工神经网络的非线性变换能力并通过学习来确定的。

将组合预测核心算法嵌入到应用软件之中,应用软件的体系结构图如下:

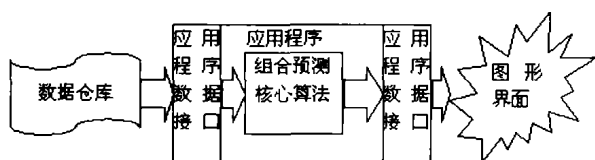


图3 应用软件体系结构图

该软件结构的流程为:首先从数据仓库中提取数据,利用组合预测核心算法计算预测值,然后将算法的输出数据以图形的形式展现给用户,一般以曲线的形式表现,附加一定的解释。接口设计包括与外系统的接口和内部接口,外部接口将应用程序做成程序包的形式,直接运行,输入数据为数据文件,文件中保存单个或多个用户的历史数据。内部接口以DLL动态连接库形式提供,如果脱机运行,则DLL的输入接口为文件路径名,输出接口也是文件路径名。如果是在线运行,输入接口则为数组。

应用程序中数据结构采用数组,数据文件中的数据结构也是以数组的形式提供。如果文件中包含多个用户,则以记录中包含用户数据数组的形式保存数据。

算法设计主要包括:线性回归预测、指数平滑预测、灰色预测和神经网络,现将各部分的算法流程逐一介绍:

### 2.2.1 线性回归预测流程

一元线性回归预测的基本表达式为:

$$\hat{y}=a+bx \quad (2)$$

主要的任务就是要确定 $a, b$ 的值是误差最小,利用误差的平方和最小来求得 $a, b$ 的值。误差平方和为:

$$\sum_{i=1}^n (y_i - \hat{y})^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \quad (3)$$

分别对 $a, b$ 求微商,得:

$$b = \frac{\sum x_i y_i - \bar{x} \sum y_i}{\sum x_i^2 - \bar{x} \sum x_i} \quad a = \bar{y} - b\bar{x} \quad (4)$$

这里分别是客户历史数据的时间和费用。

线性回归的流程如下:

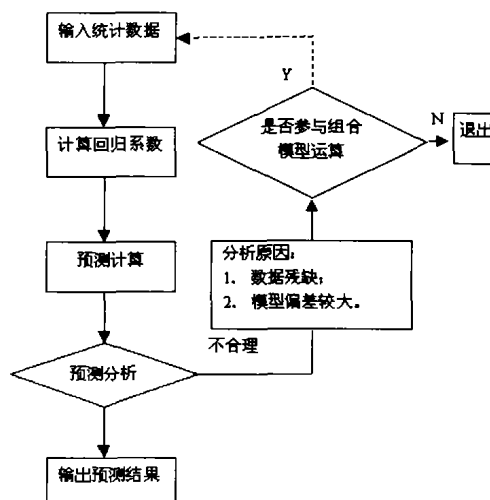


图4 线性回归流程简图

### 2.2.2 指数平滑预测流程

指数平滑预测方法的基本思想是:在预测下一周期的指标时,既考虑这个周期的指标,又不忘记前面的指标,最重要的一点是:离现在的时间越长的历史指标,其赋予的权值越小。从下面的推倒中可以看出。

对时间序列 $y_1, y_2, \dots, y_t$ ,一次指数平滑的公式为:

$$s_t^{(1)} = ay_t + (1-a)s_{t-1}^{(1)} \quad (5)$$

该公式求得的结果就是下一周期的预测值, $a$ 为加权系数, $0 < a < 1$ 。

假定 $y_t$ 的序列长度无限,将上式中的 $t$ 分别以 $t-1, t-2, t-3, \dots$ 一次代入可得:

$$\begin{aligned} s_t^{(1)} &= ay_t + (1-a)[ay_{t-1} + (1-a)s_{t-2}^{(1)}] \\ &= ay_t + a(1-a)y_{t-1} + (1-a)^2[ay_{t-2} + (1-a)s_{t-3}^{(1)}] \\ &\dots \\ &= ay_t + a(1-a)y_{t-1} + a(1-a)^2y_{t-2} + \dots + a(1-a)^j y_{t-j} + \dots \\ &= a \sum_{j=0}^{\infty} (1-a)^j y_{t-j} \end{aligned} \quad (6)$$

可见预测值是历史数据的加权平均,加权系数分别为 $a, a(1-a), a(1-a)^2, \dots$ ,即首项为 $a$ ,公比为 $(1-a)$ 的等比数列。加权系数的和为:

$$a = \sum_{j=0}^{\infty} (1-a)^j = a \frac{1}{1-(1-a)} = 1 \quad (7)$$

所以,指数平滑实际上是一种以时间定权的加权平均。越近的数据,加权系数也越大;数据越远,加权的系数越小。

指数平滑的预测流程如图5。

### 2.2.3 灰色预测流程

灰色数列利用GM(1,1)模型来对未来的数据进行预测。其步骤如下:

(1)确定原始数列。

(2)以某一点为参考点,取领域系。

所谓领域是指包含参考点的等时空距、相邻、非跳跃去定的数列。所有的领域构成领域系。

(3)取领域系中的部分领域建立GM(1,1)模型。GM(1,1)

中  $a$  为模型发展参数,  $b$  为模型的内生控制参数。

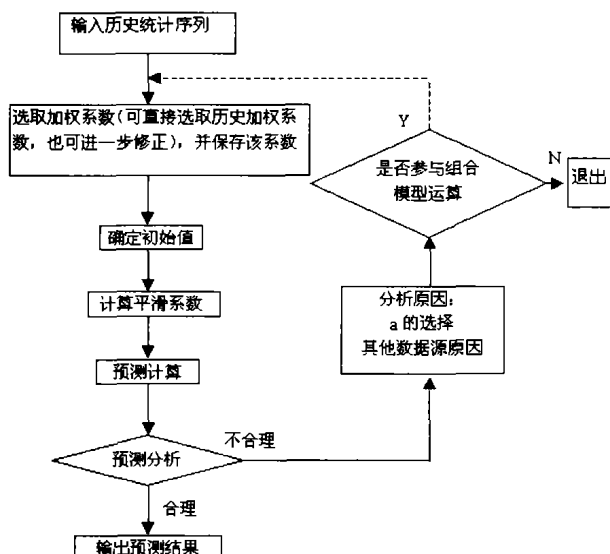


图5 指数平滑预测流程

记  $A$  为 GM 中  $ai$  的全体:

$A=\{ai|i \in I\}$ ,  $I$  为指标集。

记  $U$  为 GM 中  $ui$  的全体:

$U=\{ui|i \in I\}$

(4)确定时间响应函数:

$$\hat{X}^{(1)}(k+1)=(X_i^{(0)}(1)-\frac{u_i}{a_i})e^{-ak}+\frac{u_i}{a_i} \quad (8)$$

(5)确定未来某一时刻的值。

## 2.2.4 神经网络组合流程

神经网络用来控制上述各种算法的权值分配的调整。在初始分配权值时,可以依靠专家来确定,也可均分。神经网络的调整正是通过新数据的出现,以此来和预测数据进行比较,预测越准确的算法,分配越高的权值,反之,则权值较低。运用神经网络,需要保存一些历史的权值,通过一些预定的规则对历史权值进行修正,这些历史权值及其效果保存在数据库里,相当于知识。规则体现在算法之中,而不存在数据库,因为神经网络的是一个不断的训练过程,在修正和训练之后,如果准确率不能通过验证,需要进一步的训练;其示意图如图6。

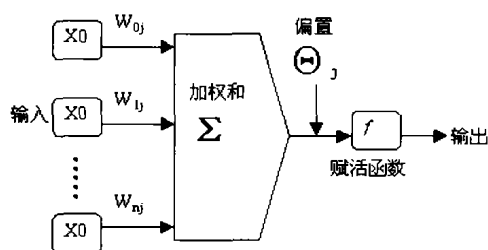


图6 神经网络权值训练

后向传播算法的神经网络学习流程如图7。

## 3 时间序列算法的实际应用

以电信行业 1999 年 1 月到 2001 年 11 月的业务数据为例,将该业务数据作为历史数据,数据量达到 10000 条,应用前述的时间序列算法,预测未来 12 个月的数据变化情况,精度要求控制在 10% 误差以内。预测结果见图 8。

72 2004.28 计算机工程与应用

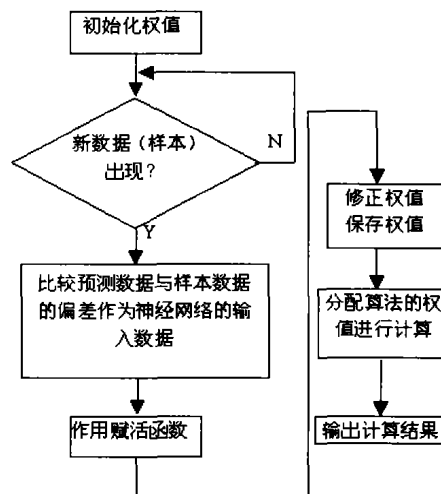


图7 神经网络学习流程

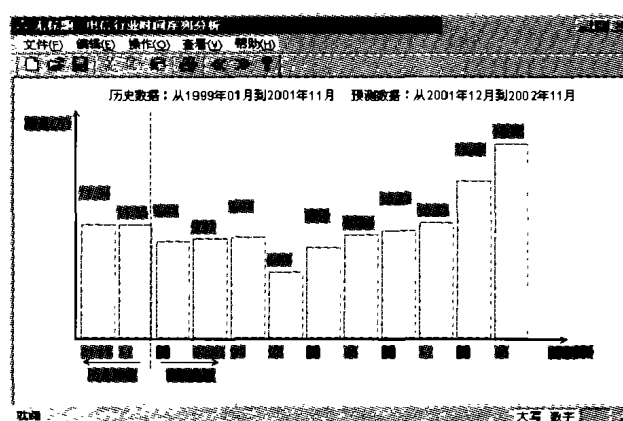


图8 时间序列预测结果

为了了解预测的准确性,首先利用时间序列分析算法对历史数据进行预测,然后预测未来数据,预测结果如图9。

从图中可以看出,时间序列分析算法对历史数据预测的平均误差为 5.3%,预测值与真实值非常接近,这说明时间序列分析算法完全可以用于对未来数据的预测,且结果较为准确、可靠,为企业的经营决策提供了事实依据。

## 4 结束语

时间序列分析算法融合了线性回归模型、指数平滑模型、灰色预测模型和神经网络模型等多个模型结构,模型之间相互关联,且互为一体。将该算法应用于电信行业的业务数据分析,并预测未来数据的发展趋势;实践表明,预测结果较为准确合理,从而为企业的经营决策提供了参考和依据。如果进一步优化和调整权值,可望得到更好的预测结果。

(收稿日期:2003 年 4 月)

## 参考文献

- 1.P J Brockwell,R A Davis.Time Series:Theory and Methods[M].2 edition,Berlin :Springer-Verlag,1991
- 2.Mohsen Pourahmadi.Foundation of Time Series Analysis and Prediction Theory[M].2001
- 3.Katsuto Tanaka.Time Series Analysis;Nonstationary and Noninvertible

(下转 143 页)

SLEE Management Interfaces、Timer、Alarm、Log 六部份构成:

Service Container 业务容器,管理业务实例,进行业务逻辑属性交互、事件通知;

Service Management Interface 业务管理接口,业务逻辑与业务数据的接口;

SLEE Management Interface SLEE 管理接口,与 Dispatcher 交互,管理 SLEE 负载均衡;

Timer 定时器,与业务生命周期管理交互;

Alarm 报警,业务逻辑启动、运行异常告警;

Log 系统日志,记录系统维护信息,运行异常信息。

#### (4)Adapter

协议适配,物理上位于一台高性能服务器上;

目前开发 SIP、INAP 两个适配器。

#### (5)OAM

系统操作维护接口;

逻辑上位于一台 PC 机上,实现对应用服务器的总体监控。

## 4 结论

软交换的概念最早由 Lucent 公司贝尔实验室于 90 年代中期提出。其基本指导思想在于软件系统不但独立于所控制的硬件,而且与自身所运行的平台无关。在下一代网络中,由于采用前面所讲的软交换分层结构,使得呼叫控制与业务提供相互独立,网络运营商与业务提供商相互独立,因而网络中的业务控制环境可以由多个属于不同运营商的域组成,业务可能是跨平台由多个运营商共同提供,也可能由独立的第三方应用/业

务提供商提供。

软交换系统是一个技术热点,现在有众多的研究者正在研究,人们的想法基本是一致的。也就是说,采用软交换的模式,把呼叫控制与呼叫传输分离。软交换所面临的首要的也是最主要的挑战就是多协议,即软交换所支持的协议很多,而且这些协议分别来自不同的标准化组织,这为技术研究、设备生产通信运营等各方面都带来了协调方面的困难。软交换系统中的应用服务器是专为增值业务而引入的。应用服务器负责各种增值业务的逻辑产生和管理,并且还提供各种开放的 API,为第三方业务的开发提供开发平台。应用服务器是一个独立的组件,与控制层的软交换无关,从而实现了业务与呼叫控制的分离,有利于新业务的引入。笔者设计的软交换系统是基于 SIP 协议的,SIP 协议具有简单灵活、分布式控制等优点,所以在 SIP 基础上来实现软交换系统中的应用增值业务具有实现意义。(收稿日期:2003 年 5 月)

## 参考文献

- 1.Charlotte Wolter.Softswitch Evolution[M].Virgo Publishing,Inc,2001
- 2.Bill Douskalis.Softswitch Network Design and Testing[M].Prentice Hall PTR,2002
- 3.John Jones.The Softswitch in the Call Center[J].connections magazine,2001
- 4.TrueTel Communications.TrueConverge(tm)Solution.2002
- 5.杨放春,孙其博.智能网技术及其发展[M].北京邮电大学出版社,2002
- 6.赵慧玲,叶华.以软交换为核心的下一代网络技术[M].人民邮电出版社,2002

(上接 72 页)

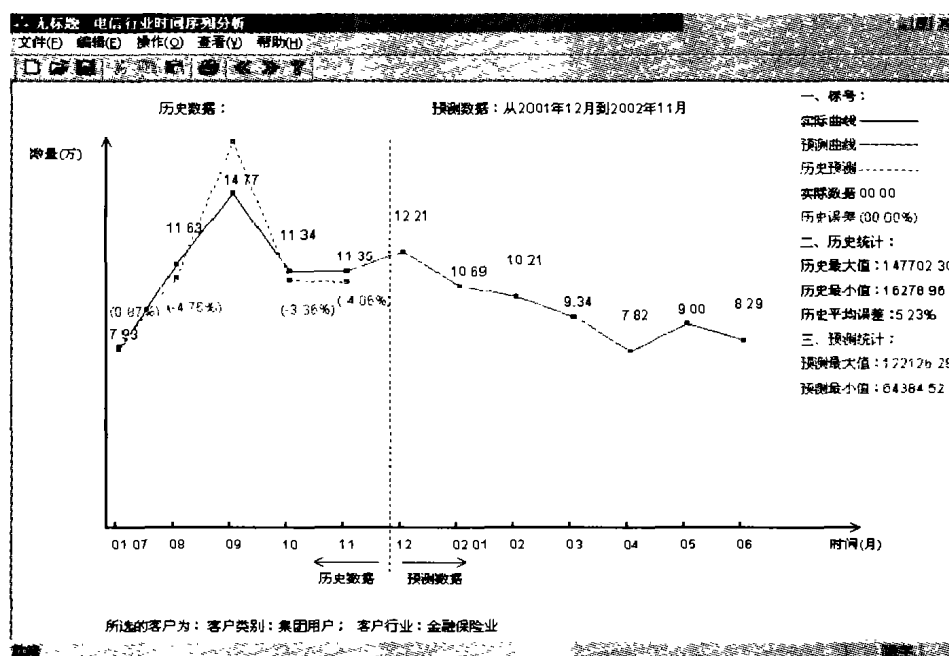


图 9 历史数据和未来数据的预测分析

Distribution Theory[M].1996

4.M Ankerst,M Breunig.OPTICS:Ordering points to identify the clustering structure[C].In:Proc 1999 ACM-SIGMOD Int Conf Management of Data(SIGMOD'99),Philadelphia,PA,1999:49~60

5.T Anand,G Kahn.Opportunity explorer:Navigating large databases

using knowledge discovery templates[C].In:Proc AAAI-93 Workshop Knowledge Discovery in Databases,Washington,DC,1993:45~51

6.A Berson,S J Smith.Data Warehousing,Data Mining,and OLAP[M].New York:McGraw-Hill,1997