

基于指数平滑法的动态预测机制

沈海迪, 万振凯

(天津工业大学 计算机科学与软件学院, 天津 300387)

摘要:对数据进行实时预测,从而制定下阶段的工作计划,对于生活中的工作生产具有很大意义。现阶段存在很多预测模型,其中,指数平滑法在短期预测中应用较为广泛。针对传统预测模型中存在的方法单一,预测模式固定,不能更好地追随数据的变化趋势,具有局限性等问题,提出了一种基于传统模型的动态预测机制。该预测机制在预测过程中选取一定步长的数据,根据数据的实际变化趋势得到可变的平滑系数,此平滑系数能够根据实际数据的变化趋势进行自动调整,使预测模型在预测过程中能追踪到数据的实际走势,具有更高的自适应性。在仿真实验中,以某高校多媒体教室的实际使用频率作为实验数据,对修正后的模型和传统模型进行对比,结果显示,修正后的动态预测机制具有更高的预测精度,且模型简单易用,能够满足实际情况的需要。

关键词:预测模型;指数平滑法;传统模型;动态平滑系数;多媒体教室

中图分类号:TP393

文献标识码:A

文章编号:1673-629X(2017)07-0006-04

doi:10.3969/j.issn.1673-629X.2017.07.002

A Dynamic Prediction Mechanism Based on Exponential Smoothing Method

SHEN Hai-di, WAN Zhen-kai

(College of Computer Science and Software, Tianjin Polytechnic University, Tianjin 300387, China)

Abstract:Real time prediction for data and developing the work plan in next phase is significant for the production of life. There are many prediction models in the present stage in which the exponential smoothing method is widely used in short-term forecasting. In view of problems of the traditional forecasting model like single way, fixed forecasting mode, not better following the trend of the change of the data and the limitation and so on, a dynamic prediction mechanism based on the traditional model is proposed. It selects a certain step of data in the prediction process and acquires variable smoothing coefficient according to the actual changes of data. The smoothing coefficient can be adjusted automatically according to the change trend of the actual data, making the forecast model can be traced to the actual trend of the data in the prediction process, which has a higher adaptability. In the simulation experiment, the actual usage frequency of the multimedia classroom in a college is used as the experimental data to compare the modified model and the traditional model. The result shows that the modified dynamic prediction mechanism has higher prediction accuracy, and the model is simple and easy to use, and can meet the needs of the actual situation.

Key words:prediction model;exponential smoothing method;traditional model;dynamic smoothing coefficient;multimedia classroom

0 引言

在现阶段的工作生产中,数据的实时预测极为重要,对于提高生产效率具有重大意义。

目前,针对数据预测,国内外学者提出了一些成型的理论模型,这些理论模型涉及到生活中的各个方面,如交通流预测、网络流量预测、无线传感器、网络管理以及操作模式的生产率等^[1-6]。在实际生活中,大多数情况下,数据是随着时间的改变而改变的,这使得数

据在一定程度上符合时间序列的一般规律。在时间序列方面,目前已存在一些可用的模型。例如,李程等结合 ARIMA 与 GM 模型对民航货邮周转量进行研究^[7];C. R. Rivero 等采用平滑技术对数据进行完善,通过不完整的数据对短期时间序列进行处理,实现了降雨量的预测^[8];另外,可以通过机器学习法,根据之前的历史数据得到 SVM 预测模型,或者根据序列间的耦合度,找到相似的子序列,通过相似子序列的趋势

收稿日期:2016-08-22

修回日期:2016-11-29

网络出版时间:2017-04-28

基金项目:天津市科技计划项目(15JCTPJCS100)

作者简介:沈海迪(1991-),男,硕士研究生,研究方向为计算机应用;万振凯,教授,研究方向为计算机应用。

网络出版地址: <http://www.cnki.net/kcms/detail/61.1450.TP.20170428.1704.096.html>

点建立预测模型^[9-10];文献[11]采用多输出支持向量回归和多输入多输出预测策略,实现了时间序列的多步预测。除此之外,关于时间序列方面的预测模型还有很多,国内外学者不断进行完善和更新,在这些理论模型中,涉及到数据的短期、中期以及长期预测,应用时可以根据实际情况选取。在实际工作生产中,有时无需对数据进行长期预测,短期及时的数据更具有参考价值,管理者通常会根据短期的预测数据制定最近的工作计划。在短期的数据预测中,有很多成熟的理论模型,其中 Anton Bezuglov 等应用灰色理论模型进行短期高速公路参数的预测^[12];另外,指数平滑法在短期数据预测中具有简单易用的特点,所以在实际生活中,此模型得到了广泛应用,如芮海田等采用指数平滑法和马尔可夫模型相结合的形式对公路客运量进行预测^[13]。除此之外,指数平滑法在其他方面也得到了广泛应用,但是实际情况中,传统指数平滑法往往不能追踪到数据的实际变化趋势,从而产生预测精度不高等问题。

针对传统指数平滑预测模型的缺陷,采用动态平滑系数修正传统模型,得到一种动态预测机制,使预测模型能够及时准确地追踪到实际数据的变化趋势,从而提高预测精度。

1 传统模型预测

1.1 理论基础

指数平滑法是一种对数据进行预处理的方法^[14],主要应用于中短期预测,现已成为一种很常用的预测方法。其可以分为一、二、三次指数平滑法,分别适合时间序列无明显变化、线性变化以及非线性变化三种情况的预测,应用时应根据数据的实际变化情况。下面以三次平滑法为例进行说明。

设某个时间序列的各期值为 $x_i (i = 1, 2, \dots)$, 第 t 期的三次指数平滑值为: s_t^1, s_t^2, s_t^3 , 对应的计算公式为:

$$\begin{cases} s_t^1 = \sigma x_t + (1 - \sigma) s_{t-1}^1 \\ s_t^2 = \sigma s_t^1 + (1 - \sigma) s_{t-1}^2 \\ s_t^3 = \sigma s_t^2 + (1 - \sigma) s_{t-1}^3 \end{cases} \quad (1)$$

其中, x_t 为第 t 期的数据实际值; σ 为平滑系数,取值范围为 $(0, 1)$ 。

采用三次指数平滑法对 t 期之后的 q 期数据进行预测,公式为:

$$x_{t+q} = a_t + b_t q + c_t q^2 \quad (2)$$

其中, a_t, b_t, c_t 为第 t 期的预测参数,计算公式为: 万方数据

$$\begin{cases} a_t = 3s_t^1 - 3s_t^2 + s_t^3 \\ b_t = \frac{\sigma}{2(1 - \sigma)^2} [(6 - 5\sigma)s_t^1 - 2(5 - 4\sigma)s_t^2 + (4 - 3\sigma)s_t^3] \\ c_t = \frac{\sigma^2(s_t^1 - 2s_t^2 + s_t^3)}{2(1 - \sigma)^2} \end{cases} \quad (3)$$

1.2 确定数据

若想对数据进行预测,在理论公式的基础上,还需要确定出两个数据,即初始值及平滑系数。

对于整体的预测过程来说,需要知道预测的初始值,即 s_0^1, s_0^2, s_0^3 , 在传统指数平滑法的模式中,按照初始值选取的原则,当数据较多时,可以用第一期的实际值代替,即:

$$s_0^1 = s_0^2 = s_0^3 = x_0 \quad (4)$$

当数据较少时,选取前三期数据的平均数,即:

$$s_0^1 = s_0^2 = s_0^3 = \frac{x_1 + x_2 + x_3}{3} \quad (5)$$

平滑系数的选取对于整体的预测过程十分重要, σ 的取值范围是 $0 \sim 1$, 一般情况下,根据经验,其取值范围是 $0.3 \sim 0.7$ 。在传统模型中,平滑系数的选取大都是根据经验或者是大量实验,最终确定出一个相对较好的值。

1.3 模型缺陷

传统的指数平滑法虽然能够简单易用地对数据进行预测,但是缺陷明显。通过之前所介绍的理论基础可以看出,在预测过程中,平滑系数在整个预测过程中是一个常数,不能随数据改变的趋势进行调整,整体预测过程显得僵硬,无法进行动态评估,难以适应数据出现的特殊跳跃等情况。

2 动态预测机制

在预测过程中,平滑系数反映的是数据的变化趋势^[15]。传统模型中,平滑系数一旦确定即为一个常数,使其难以追踪整个过程中数据的变化,使得整个预测过程的适应性降低^[16]。对于大多数应用场景,随着技术的发展,数据的实际变化情况会随着环境的影响发生变化,这也就说明,仅采用一个平滑系数,难以适应整体的变化过程,所以采用动态平滑系数对传统模型进行修正。

2.1 动态平滑指数

采用动态平滑系数就是让系数随着实际数据动态化,在预测过程中自动适应序列的变化^[17]。实验表明,当数据呈现水平趋势时,平滑系数应选取 $0.1 \sim 0.3$, 当数据持续上升或下降时,平滑系数则应选取 $0.3 \sim 0.5$, 当波动较大时,应选取 $0.6 \sim 0.8$ 。所以,当

数据前后的变化趋势不同时,一个平滑系数难以满足整体的预测需求。

为了避免传统模式中的弊端,在动态预测机制中采用动态平滑系数,即让整体预测过程中的平滑系数随着数据变化趋势的改变而自动调整,这样预测机制能够更好地根据当前的数据变化趋势进行动态调整,平滑系数也能够更加准确地追踪到数据的实际变化情况。

2.2 动态预测过程

在对整个时间序列采用动态平滑系数进行预测时,首先确定出数据的期数,即假设已知 i 期实际数据,选取其中 T 期,根据这 T 期确定出数据基本走势;再采用迭代法搜寻最优平滑系数,从而进行 $T+1$ 期数据的预测。采用动态预测机制的整体预测过程为:

(1)对平滑指数进行步长为 0.000 1 的迭代,求出这 T 期数据采用三次指数平滑法预测得到的误差平方和,根据误差平方和最小原则选取最优的平滑指数,即:

$$e = \min \sum_k^N (x'_k - x_k)^2, k = 1, 2, 3 \cdots, N = k + T - 1$$

(6)

由此得到最优的 σ 。

(2)根据步骤(1)得到的最优系数进行 $N+1$ 期预测,得到预测结果后,去掉之前的第一个值,采用 $k+1$ 到 $k+T-1$ 期的数据,重复步骤(1),进行 $N+2$ 期数据的预测。

图 1 为迭代时的程序流程。

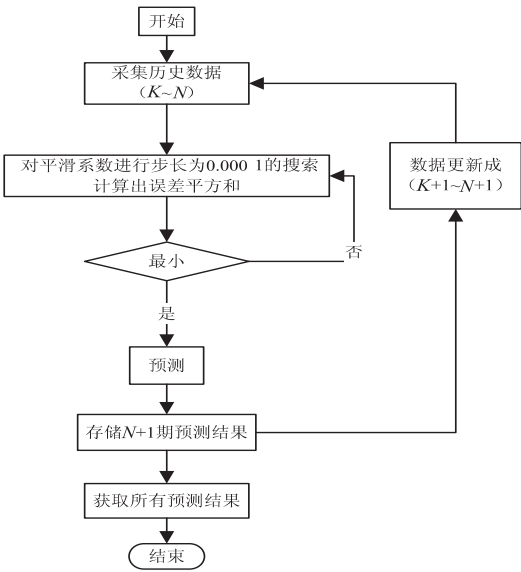


图 1 动态平滑系数寻优流程图

对于初始值的选取,采用指数平滑法初始值的选取原则,当数据较多时,即第一期的实际值为初始值,当数据较少时,采用前三期数据的平均值作为初始值,选取时应根据实际情况进行判定。

3 模型验证及分析

3.1 验证过程

采用某高校多媒体教室的实际使用频率(见表 1)作为实验数据对动态预测机制进行验证。

表 1 某校多媒体实际使用频率

期数	实际值
1	3 325
2	3 391
3	3 406
4	3 575
5	3 762
6	4 091
7	4 607
8	5 096
9	5 530
10	6 136
11	6 875
12	7 790

对数据进行分析可以发现,数据呈现非线性变化趋势,可以采用三次指数平滑法,分别采用传统模型和动态预测机制对数据进行预测。在传统模型中,分别选取不同的平滑系数值进行预测,得到的相对误差分布如图 2 所示。

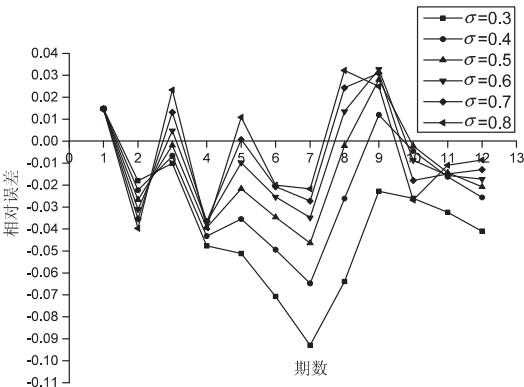


图 2 各系数相对误差分布

根据相对误差分布以及考虑到整体的误差平方和,最终确定平滑系数为 0.7,采用此平滑系数确定平滑值,对第 13、14 期数据进行预测。

接下来采用动态预测机制进行预测,根据数据的实际情况,选取 5 个数据为一阶段,得到的 6 期及以后的预测结果和传统模型相比如图 3 所示。

其中采用两种不同形式对第 13、14 期的预测结果见表 2。

3.2 结果分析

根据图 2 可以看出,从第 6 期开始,动态平滑系数的优势渐渐显示出来,而且采用动态系数预测出的后五期的误差平方和要小于普通方法的误差平方和。综上,采用动态平滑指数之后,效果得到了一定改善。

表 2 第 13 期和第 14 期的预测结果

预测形式	期数	实际值	预测值	相对误差/%
平滑系数 为 0.7	11	9 095	8 815	-3.08
	12	10 660	9 975	-6.43
动态系数	11	9 095	8 905	-2.09
	12	10 660	10 212	-4.20

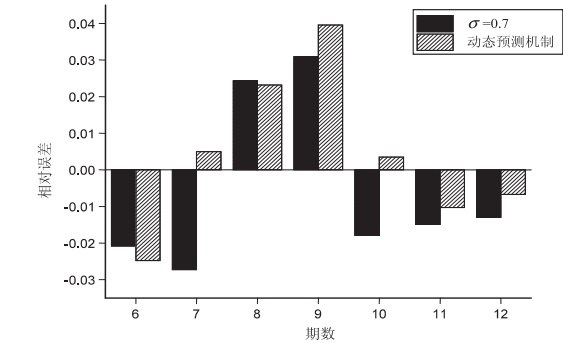


图 3 动态系数和传统模型对比

传统方法预测出后两期的相对误差为-3.08%和-6.43%,采用动态系数对传统模型进行修正之后,得到的相对误差为-2.09%和-4.20%,预测精度得到提升。随着多媒体教室的使用频率逐个学期增加,减少误差很有必要,所以采用动态系数对多媒体教室的使用频率进行预测具有一定的可行性。

采用实际数据对动态预测机制进行验证之后可知,改进后的动态预测机制具有一定的实用性。

4 结束语

针对现阶段生产工作中对于数据预测的必要性,经过对各种预测理论模型的分析,基于传统的指数平滑法进行修正,通过阶段性数据判断出数据的近期走势,根据迭代的方式,以误差平方和最小为依据搜寻出最佳平滑系数,从而得到能够随着数据改变自动调整的动态预测机制,改善了传统模型中系数单一、预测过程僵硬等缺点。采用某高校多媒体教室的实际使用频率为实验数据对模型进行验证,验证过程中,分别采用传统模型和动态机制对实际数据进行预测,并且将动态预测机制得到的预测结果和传统模型中的最优结果进行比较。结果表明,动态预测机制得到的预测结果在预测精度上得到了一定提升,预测结果能够在误差允许的范围之内,证明了修正后的动态预测机制具有更好的实用性。

参考文献:

[1] 钟慧玲,邝朝剑,黄晓宇,等. 基于历史频繁模式的交通流

预测算法[J]. 计算机工程与设计,2012,33(4):1547-1552.

[2] 邱 婧,夏靖波,吴吉祥,等. 网络流量预测模型研究进展[J]. 计算机工程与设计,2012,33(3):865-869.

[3] 冯华丽,刘 渊,陈 冬,等. QPSO 算法优化 BP 网络的网络流量预测[J]. 计算机工程与应用,2012,48(3):102-104.

[4] 邹长忠. 无线传感器网络中基于 SVR 的节点数据预测算法[J]. 计算机应用,2010,30(1):127-129.

[5] 夏 怒,李 伟,罗军舟,等. 一种基于波动类型识别的路由节点行为预测算法[J]. 计算机学报,2014,37(2):326-334.

[6] Ding Jinliang, Chai Tianyou, Cheng Weijian, et al. Data-based multiple-model prediction of the production rate for hematite ore beneficiation process[J]. Control Engineering Practice, 2015, 45:219-229.

[7] 李 程,徐 琪. 基于 ARIMA-GM 的短期民航货邮周转量研究[J]. 计算机工程与应用,2015,51(2):259-264.

[8] Rivero C R, Patiño H D, Pucheta J A. Short-term rainfall time series prediction with incomplete data[C]//International joint conference on neural networks. Killarney: [s. n.], 2015:1-6.

[9] 吴湘宁,胡 炫,胡光道,等. Oracle 中使用支持向量机的时间序列预测方法[J]. 计算机工程与应用,2013,49(14):121-125.

[10] 冯凯文,孟凡荣,牛 强,等. 基于趋势点状态模型的时间序列预测算法[J]. 计算机应用研究,2011,28(12):4510-4512.

[11] Bao Yukun, Xiong Tao, Hu Zhongyi. Multi-step-ahead time series prediction using multiple-output support vector regression[J]. Neurocomputing, 2014, 129:482-493.

[12] Bezuglov A, Comert G. Short-term freeway traffic parameter prediction: application of grey system theory models[J]. Expert Systems with Applications, 2016, 62:4545-4551.

[13] 芮海田,吴群琪,袁华智,等. 基于指数平滑法和马尔科夫模型的公路客运量预测方法[J]. 交通运输工程学报, 2013, 13(4):87-93.

[14] Chan K Y, Dillon T S, Singh J, et al. Traffic flow forecasting neural networks based on exponential smoothing method [C]//IEEE conference on industrial electronics and applications. Beijing: IEEE, 2011:376-381.

[15] 高春雷,安 泰,向 兵,等. 基于粒子群算法的指数平滑系数优化方法[J]. 西北大学学报:自然科学版,2014,44(3):383-386.

[16] 王国权,王 森,刘华勇,等. 基于自适应的动态三次指数平滑法的风电场风速预测[J]. 电力系统保护与控制, 2014, 42(15):117-122.

[17] 齐 驰,侯忠生. 自适应单指数平滑法在短期交通流预测中的应用[J]. 控制理论与应用,2012,29(4):465-469.