

# 基于核函数技术的时尚服装需求预测方法

孟志青<sup>1</sup> 马珂<sup>1</sup> 郑英<sup>2</sup>

(浙江工业大学经贸管理学院 杭州 310023)<sup>1</sup> (大红鹰学院基础学院 宁波 315327)<sup>2</sup>

**摘要** 短生命周期的服装需求预测问题一直是服装品牌公司无法解决的问题,运用了非线性机器学习的核函数技术,提出了一个适合短生命周期时尚类服装的预测方法。结合服装公司的产品特征和服装数据仓库应用研究,建立了一种基于核函数技术的服装需求预测模型,提出了一个计算算法,通过实际数据进行了分析验证,结果表明所提出的方法对于时尚服装需求预测具有较高的动态预测精度,适合服装公司进行动态补货,对于品牌公司控制库存具有重要的实际意义。

**关键词** 时尚服装,需求,核函数,预测模型

中图分类号 TP181 文献标识码 A

## Forecasting Method for Fashion Clothing Demand Based on Kernel Functions Technology

MENG Zhi-qing<sup>1</sup> MA Ke<sup>1</sup> ZHENG Ying<sup>2</sup>

(College of Economics and Management, Zhejiang University of Technology, Hangzhou 310023, China)<sup>1</sup>

(Basic College, Dahongying University, Ningbo 315327, China)<sup>2</sup>

**Abstract** Fashion clothing demand forecasting of short life cycle has plagued the production and inventory of garment brand company, and it can not be solved. In this paper, we used the kernel function technique of nonlinear machine learning, and put forward a prediction method for short life cycle fashion clothing. In combination with the characteristics of garment companies and the application of data warehouse, a model of garment demand forecasting based on kernel function was established. We gave a calculation algorithm and carried out the analysis and verification through the actual data. Data experiments show that the proposed method for fashion clothing demand forecasting has higher prediction accuracy, and it is suitable for clothing company dynamic replenishment. It also has important practical significance for the company inventory control.

**Keywords** Fashion clothing, Demand, Kernel function, Prediction model

## 1 引言

服装销售预测一直是困扰许多服装品牌公司的难题,尽管品牌公司都有销售历史数据,但是运用这些销售数据进行销售预测是一项非常复杂和困难的工作,导致当前许多服装品牌公司销售预测停留在拍脑袋过程(商品计划部或供应链部);而且多数品牌公司一直沿用零售商、代理商和加盟商的订货会销售模式,一次订货将一个季节的服装订完,后续补货的情况非常少,结果造成年年库存增加,平均产销率大约为 50%~65%,经常导致有的品种大量缺货,有的品种大量库存,大大地降低了服装企业的竞争力。事实上,服装公司非常渴望有准确的销售预测方法,帮助他们制定生产计划和控制库存,为此,提出和研究实用的服装销售预测方法对于提高品牌公司的运营水平具有重要的实用价值。

时尚服装具有明显的季节性,季节性对应于服装产品的生命周期,产品生命周期(Product Life Cycle, PCL)最早出现于经济领域,20 世纪 50 年代,Dean 和 Levirt 先后提出了这一

理论。服装的生命周期即其投入市场到退出市场的全过程,根据生命周期理论,可以把服装的销售过程大致划分为 4 个阶段:导入期、成长期、成熟期、衰退期。针对我国服装产业的特点,胡敏敏<sup>[19]</sup>提出服装供应链呈现出结构短、订货提前期长、销售周期短、在市场上易被竞争者模仿、竞争优势降低导致边际利润下降、产品销售价格随时间下降快等特点。随着人们生活质量的改善、流行趋势变化的加速,服装产品的生命周期也在越来越短,一件新品从上市到退出可能只有一年甚至几个月时间。因此服装产品销售预测属于短生命周期的产品预测。

国外针对销售预测的研究比较丰富,创建和研究了包括投入产出法、BP 神经网络法、Bass 扩散模型、多元回归分析、核函数技术等常见的预测模型和方法。投入产出方法创立于 20 世纪 30 年代,由诺贝尔经济学获奖者 W. Leontief 首次提出,经过六十多年的研究和发展,其理论和方法日渐成熟,并已得到广泛应用。在最初的应用中,它是简单易用的静态投入产出模型,能够反映一段时间的经济发展和结构条件。

本文受浙江省自然科学基金资助项目(LY15G010007)资助。

孟志青(1962—),男,博士,教授,主要研究方向为数据挖掘;马珂(1990—),女,硕士生,主要研究方向为数据挖掘;郑英(1983—),女,讲师,主要研究方向为优化与供应链。

1987年神经网络方法由Lapedes和Farber在经济预测领域中推出,该方法可以预测短期经济前景,也可用于中长期经济预测。Frank Bass于1969年针对耐用品创建了Bass扩散模型,其可应用于某一类产品或者某一行业的预测,欧美许多企业也成功利用Bass模型进行新产品的扩散分析和销售预测。Gardner<sup>[11]</sup>通过研究指数平滑预测法的现状,重点突出其经济预测特点以及鲁棒性(robust)特征,但该方法对变量间的因果关系未进行考虑。Chandra和Grabis<sup>[5]</sup>将自回归模型应用于多级供应链的需求预测,基于库存管理提出物料需求计划预测方法,以减小订货量的方差,即降低牛鞭效应。Edward和Thomas在2004年根据美国威斯康辛州彩票销售数据,利用线性混合的纵向数据模型预计销量。核函数作为非线性机器学习技术,1985年Adamowskik引入了核函数技术进行预测,核函数预测对于波动性巨大的非线性样本具有较好的预测稳定性,所以近年来被广泛应用于非线性支持向量机器学习模型中进行非线性预测研究。事实证明,核函数技术对于非线性预测具有较高的精度,因此核函数适合于波动较大的短周期服装销售需求预测。

在服装预测方面,国内外学者进行了大量研究,例如刘咏梅<sup>[20]</sup>认为服装销售情况可以反应出每款产品的生命周期变化曲线,可以利用数学公式和销售历史记录对产品进行销售预测,以此确定产品处于生命周期哪一阶段并做出相应的库存管理反应;她利用移动平均法对一部分产品的销售信息进行分析,对该产品在生命周期各阶段的企业策略进行评价。胡觉亮等<sup>[17]</sup>以产品生命周期理论为基础,将服装销售数据结合到生命周期变化模型中,用男装西服销售的实例分析和肯定了改进的BASS模型用于描述服装生命周期曲线的合理性。乔亮亮<sup>[22]</sup>由已知的青年女装某时期内的季度销售额及其相关影响因素的数据,利用多元线性回归法进行预测,求出其数学模型,并进行误差和精度分析。徐贤浩等<sup>[24]</sup>采用正交矩阵主轴旋转的预测建模方法,实现对叉积阵的预测建模。运用多元线性回归的预测模型,可以在无须对未来系统采样的情况下,推测未来的回归系数及模型精度。刘铮等<sup>[21]</sup>通过对Bass模型的研究,基于平均服务的影响因子以及商品的平均重复购买率来修正原始的预测模型,以便更好地实现服务-购买意愿双渠道短生命周期产品的需求预测,以反映实际需求。Donohue<sup>[6]</sup>采用预测更新研究时尚品的补货,采用了二次订货策略以更好地适应市场需求的波动,提高供应链收益,使得分销商可以根据更新的销售信息来调整其订货水平。Hsu<sup>[12]</sup>和Zhou<sup>[14]</sup>等人通过讲述美国服装行业的发展,研究短生命周期产品在不同阶段的订购策略和定价策略,实证研究了需求预测模型和市场实际需求的偏差。胡觉亮等<sup>[18]</sup>基于需求信息更新进行服装供应链最优订货决策的探索,认为服装产品销售期具有明显的三阶段特点,得出零售商进行三次订货获益更多,并且销售期开始前零售商对市场需求的预测准确性对供应链效率有很大的影响。盛方正等<sup>[23]</sup>主要研究了顾客的需求预测信息更新的情况下,利用三阶段订货模型制定零售商的最优订货策略,在第一阶段订货数量与信息无关,在第二阶段根据需求信息调整第一阶段的订货水平,第三阶段实现顾客需求。

上述文献大多考虑了具有普适性的各预测模型的整合,针对一类服装产品的特点,分析并量化其需求影响因素,构建

出有效和针对性强的需求预测综合模型,其具有一定的适用性。由于大多数的预测方法需要辅以大量的数据才能进行相对准确的预测,并且如果只考虑到时间推移或有限的几项影响因素,很难控制其预测误差。然而服装短生命周期产品的生命周期较短,往往缺乏甚至没有历史数据,而且在出现大量需求数据时一般已经进入了成熟期甚至衰退期,所以以上方法在短生命周期产品预测的实践中很难行之有效。所以需要针对服装短生命周期产品的需求预测实用方法进一步研究。由于服装短生命周期产品时效性强、需求波动大、采购提前期长及与外界环境存在着复杂的非线性关系,在生命周期结束后,其剩余价值将会丧失或变得很低,因此服装的销售预测的准确性很难提高。由于品牌公司每年的品类变化基本不大,品类的销售历史数据可以应用在下一年度的预测,另一方面品牌公司实际要求在一定的销售周期内进行动态预测,并且根据上季度和本季度的销售数据进行预测,提高预测的准确性。通过对实际数据采用的预测方法的比较研究,发现核函数估计预测对于影响因素较多的短生命周期服装销售预测具有较好的预测精度和预测稳健性,核函数技术运用在预测中,大多数采用支持向量机模型,这时需要求解一个非常困难的非线性优化问题<sup>[27,28]</sup>,另一方面,核函数也可以直接进行模拟预测,具有较好的预测精度,并且比其它许多预测技术计算要简单很多,避免求解优化问题。因此,本文提出了一个基于核函数技术的动态销量预测模型,通过核函数估计学习,直接获得预测曲线,通过实际数据实验分析,能够优化企业的采购和库存控制能力,改善预测不准确、供需不平衡的现状,及时有效地响应顾客需求。本文的研究成果可以为品牌服装企业的经营发展提供具有实践意义的指导方案,同时对短生命周期的季节性、流行性产品等的生产订货计划具有较好的借鉴意义,本方法思想简单,早期预测较其他方法更为准确。此外,本文对服装短生命周期的研究和销量预测模型的探索,为今后该方向的研究提供了参考,具有一定的理论价值。

## 2 非参数密度估计动态预测模型

服装销售预测一般是通过已经发生的销售数据建立一个回归模型,然后进行未来的预测,假设观测到一组销售数据 $x'_i (i=1,2,3,\dots,Q')$ ,作为自变量,设 $Y'$ 为因变量,可以得到如下函数关系:

$$Y' = f_i(x', \theta') + \epsilon_i \quad (1)$$

其中,称 $f_i(x', \theta')$ 为回归函数, $\epsilon$ 为误差,我们的目的是估计 $f_i(x', \theta')$ 。当假定 $f_i(x', \theta')$ 形式已知,参数估计方法可以根据先验条件估算出未知的 $\theta'$ ,即可以得到 $f_i(x', \theta')$ ,包括线性和非线性回归模型,其中 $t$ 是时间(天),随着销售日的变化,预测函数也是变动的,这是一个动态回归问题。

非参数核估计的原理简单,就是在已知某一事物的概率分布的情况下,若某一观测值在观察区域中出现,则认为该数值的概率密度很大,距离观测值较近的数的概率密度也会较大,而距离观测值较远的数则概率密度较小。具体来讲,针对观测数据将数据的值域分为若干个相等的区间,数据就按落入的区间分为若干组,每组数据和总观测值个数的比率就是每个区间的概率值,所以对于任意点 $x$ 的核密度估计为:

$$f_i(x) = \frac{1}{Q'h} \sum_{i=1}^{Q'} K\left(\frac{x-x'_i}{h}\right) \quad (2)$$

可以看出式(2)是一个加权平均,其中, $K(x)$ 称为窗或核函数,可以作为权重函数影响图形的形状; $h$ 称为窗宽或带宽,作为平滑系数,影响曲线的光滑程度,对于有 $Q'$ 个实例的样本集 $x_i'$ ,所有实例都对销量估计有影响,并且影响随距离增大而减小。 $K(x)$ 和 $h$ 所组成的核密度函数 $f_i(x)$ 用于估计观测中的点 $x$ 所用数据点的个数和利用程度,通过 $f_i(x-x_i)$ 拟合出符合远小近大的概率密度。

核密度估计是概率论中用来估计未知的密度函数,属于非参数检验方法之一,由 Rosenblatt 和 Emanuel Parzen 提出,又名 Parzen 窗(Parzen Window)。核方法是解决非线性模式分析问题的一种有效途径,其核心思想是:首先,通过某种非线性映射将原始数据嵌入到合适的高维特征空间;然后,利用通用的线性学习器在这个新的空间中分析和处理模式。在多种多样的非参数估计方法中,核密度估计是理论发展最完善、使用最广泛的已知方法。

在函数模型(2)中, $K(\cdot)$ 称为核函数(Kernal Function),通常满足对称性,可取:

$$K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right) \quad (3)$$

对于式(1),利用 Nadaraya-Waston 核估计得到:

$$f_h(x) = \frac{\sum_{i=1}^Q K_h\left(\frac{x-x_i}{h}\right) Y_i}{\sum_{i=1}^Q K_h\left(\frac{x-x_i}{h}\right)} \quad (4)$$

核函数 $K(\cdot)$ 可以利用观测点 $x_i$ 到 $x$ 的距离决定 $x_i$ 在估计点 $x$ 的密度时所起的作用,距离越近作用越大, $h$ 体现划分的值域的宽度,其大小影响区间个数, $K$ 和 $h$ 的选择共同对预测的准确性起重要影响。

#### (1)核函数选择

在核估计方法中,常用的核函数包括:均匀核、三角核、四方核、高斯核等<sup>[27]</sup>。核函数的选择取决于根据距离分配的各个样本点对密度的不同贡献,Hardle 于 2005 年指出核函数的选取通常不是核估计中最关键的因素,因为选择任何核函数对于核估计的拟合效果和稳定性都有一定的保证,对估计的渐进性质影响很小,通过对各种核函数进行了拟合实验,发现正态核函数对于预测拟合具有较好的稳定性。因此,选用正态窗函数:

$$K_h(t) = \frac{1}{h} K\left(\frac{t}{h}\right) = \frac{1}{h} \times \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{t}{h}\right)^2} \quad (5)$$

#### (2)窗宽 $h$ 的选择

窗宽 $h$ 也称为平滑参数,顾名思义窗宽对拟合曲线的光滑程度影响较大,在对许多品类数据进行拟合时, $h$ 对密度估计的相合性和估计值的准确性具有较大影响。

根据式(5)可以看出,当 $h$ 趋于0时,除了 $x=x_i, \frac{x-x_i}{h} \rightarrow \infty$ ,则使得 $K_h\left(\frac{x-x_i}{h}\right)$ 趋于0,即 $f_i(x) \rightarrow Y_i$ ,因此所得预测结果 $f_i(x)$ 只是源数据的重复表示;而当 $h \rightarrow +\infty$ 时, $\frac{x-x_i}{h} \rightarrow 0, f_i(t) \rightarrow \frac{1}{Q} \sum Y_i = \bar{Y}$ ,故为一条水平直线。

选择4种窗宽 $h=0.01, h=0.1, h=0.03, h=0.3$ 代入式(4),通过 Matlab 进行算法编程,对样本数据中的夏季小品净色男 POLO 生命期内的销售数据进行模拟试算。分别画出

图1—图4等4个的核估计拟合图,图中虚线即为核密度估计拟合曲线。观察发现当 $h=0.01$ 时拟合程度很好,但曲线的光滑性较差,峰谷值突出;而随着带宽 $h$ 的增加,核估计拟合曲线也变得越来越大光滑平顺,当 $h=0.3$ 时甚至近似于一条直线,光滑度很高但是拟合效果最差,与销量实际曲线(实线)的走势形状相差太大,几乎无法描述数据特征。

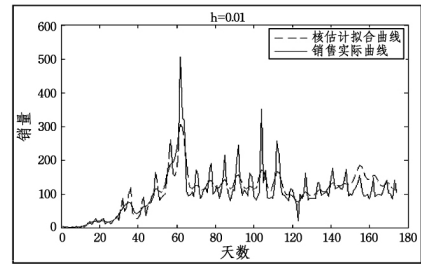


图1  $h=0.01$  的核估计拟合曲线图

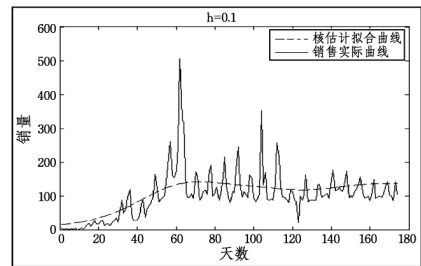


图2  $h=0.1$  的核估计拟合曲线图

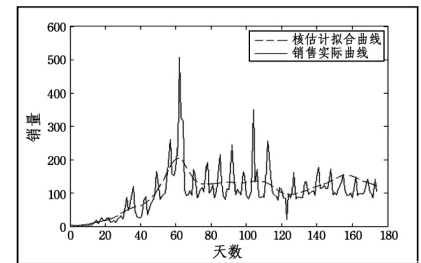


图3  $h=0.03$  的核估计拟合曲线图

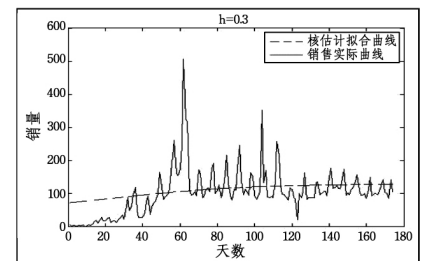


图4  $h=0.3$  的核估计拟合曲线图

由此可见当 $h$ 越大,销量预测曲线越光滑,平均化作用越明显,可能导致拟合不足,掩盖了概率密度的细节部分,无法达到预测效果;当 $h$ 越小,销售预测曲线波动越大,随机性影响太大,光滑度较差,可能造成过度拟合。所以窗宽 $h$ 的选择既不是越大越好,也不是越小越好,需要适中选择。

可以通过对估计偏差和方差作检验来判断带宽的好坏,大多数文献对带宽选择的研究一般遵循的最优原则包括极小化均分误差法(Mean Square Error, MSE)、极小化积分均方误差法(Mean Integrated Square Error, MISE)、极小化渐进积分均方误差法(Asymptotic Mean Integrated Square Error,

AMISE), 其中

$$AMISE(h) = \frac{\int K^2(x) d(x)}{Qh} + \frac{h^4 \sigma^4 \int [f''(x)]^2 dx}{4} \quad (6)$$

其中,  $\sigma$  为样本方差, 要最小化渐进均方积分误差, 必须选择恰当的  $h$  来平衡  $AMISE(h)$  中的偏差值和方差值, 因为一般  $h$  偏大时, 过平滑致使估计值的方差较小但偏差会很大, 反之则会欠平滑使偏差较小而方差太大。基于最小化  $AMISE(h)$  原则, 可以得到最优窗宽:

$$h = \left( \frac{\int K^2(x) dx}{Q\sigma^4 \int [f''(x)]^2 dx} \right)^{\frac{1}{5}} \quad (7)$$

在取  $K$  为正态核函数时, 利用便于实际操作的窗宽估计方法, 求出式(7)为:

$$h = \frac{1}{\sqrt{2\pi}} \times \frac{3}{(\sqrt[5]{35Q})\sigma} \quad (8)$$

选用的是极大光滑法对净色男 POLO 的夏季生命期销售数据进行核估计试算, 利用式(8)得到  $h = 0.03$ , 绘制拟合曲线(见图 3), 此时的密度估计曲线拟合效果较好, 在实际预测时, 选择  $h$  在  $[0.01, 0.03]$  之间。

### (3) 基于数据更新的预测模型

基于数据更新的预测主要表现在每天都要进行新的预测, 已经发生的数据将对预测准确性起到重要作用, 假设取数据库中某一季度某一品类中小品品类所在生命期每天的销售数据为样本数据  $x_i, i = 1, 2, 3, \dots, Q$ , 给定时间序列  $t_1, t_2, \dots, t_Q$ , 其中  $Q$  为生命期长度,  $x_1, x_2, \dots, x_Q$  即对应的销售值, 初始预测方程为:

$$f_0(t) = \frac{1}{Q} \sum_{i=1}^Q W_i(t) x_i \quad (9)$$

对于任意观测值  $t \in t_i, W(t)$  的核估计公式为:

$$W_i(t) = \frac{K_h\left(\frac{t-t_i}{h}\right) \times Q}{\sum_{j=1}^Q K_h\left(\frac{t-t_j}{h}\right)} \quad (10)$$

其中, 选择  $K$  为正态核函数,  $h$  使用极大光滑法则, 由样本数量和方差确定。

动态预测的总体思路: 利用非参数核估计模型预测当季销量, 假定同类产品所在销售季度的需求情况相似, 可以利用某品类服装产品上一季(2014)相似品类的销售数据作为参考, 即获取当前(2015)对应季度某个品类中小品品类的已知日销量  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k$ , 去更新过去(2014年)生命期里的每日销售数据  $x_i, i = 1, 2, \dots, Q$ , 替换前  $k$  个数据, 其中  $k < Q$ , 并以  $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k, x_{k+1}, x_{k+2}, \dots, x_Q)$  组成新的样本数据, 进行新的曲线拟合。

可以得到第  $k$  天的动态更新预测模型:

$$f_k(t) = \frac{1}{Q} \sum_{i=1}^k W_i(t) \hat{x}_i + \frac{\alpha_k}{Q} \sum_{i=k+1}^Q W_i(t) x_i \quad (11)$$

其中,  $\alpha_k (\alpha_0 = 1)$  是调整系数, 用来保证预测增长趋势与实际一致, 表明上季的销售与当季的销售变化趋势, 可以取  $\alpha_k = \frac{\sum_{i=1}^k \hat{x}_i}{\sum_{i=1}^k x_i}$ , 这个调整参数非常重要, 刻画了当季发生的数据对未来趋势的影响程度, 在实际的数据预测测试中, 随着时间

的推移, 预测精度越来越准。当前根据非参数核密度估计函数得出的销售预测模型的具体预测算法思想如下:

第 1 步 从当前汇总销售中提取当年某一季度某品类当前已知日销量数据  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k$ , 写入预测增量销售表。

第 2 步 将  $(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k, x_{k+1}, x_{k+2}, \dots, x_Q)$  所有日销量数字用最小-最大标准化法进行线性变换, 将原始值归一化映射成区间  $[0, 1]$  中的值:

$$\hat{x}_i' = (\hat{x}_i - \hat{x}_{\min}) / (\hat{x}_{\max} - \hat{x}_{\min}), i = 1, 2, 3, \dots, k$$

$$x_i' = (x_i - x_{\min}) / (x_{\max} - x_{\min}), i = k+1, k+2, \dots, Q$$

并写入当季销量标准化表, 计算预测调整系数  $\alpha_k$  和销量标准值。

第 3 步 从上年度对应季度销售表中提取该品类在生命期内的日销量估计数据, 该数据为标准化后的结果, 计算公式见式(2), 选择不同核函数和窗口  $h$ , 选择相对误差最小的核函数, 保证上季趋势的稳定性。

第 4 步 对提取的该小品品类已知数据  $\hat{x}_1, \hat{x}_2, \dots, \hat{x}_k$  的标准化结果, 替换上季度生命期内第  $k$  日前的估计数据, 保留  $x_{k+1}, x_{k+2}, \dots, x_Q$  的数据, 计算公式见式(11), 获得第  $k$  日的预测公式。

第 5 步 令  $t \rightarrow t+1, k \rightarrow k+1$ , 按组成新的数组利用非参数核估计模型预测当季第  $k$  天后的日销售量:  $f_k(t_{k+1}), f_k(t_{k+2}), \dots, f_k(t_Q)$ , 输入预测结果表, 输出结果时需要将标准化数据还原。

$$\bar{x}_i = f_k(t_i)(x_{\max} - x_{\min}) + x_{\min}, i = k+1, k+2, \dots, Q$$

### (4) 预测误差估计

预测精度是衡量预测模型拟合程度的一项指标, 即由预测模型试算产生的模拟值与现实发生的实际值之间拟合的优劣程度。简单来说, 拟合程度越低, 误差越大, 预测精度就越低; 反之拟合程度越高, 误差越小, 预测就越精准。预测精度对于检验预测模型的有效性和适用性具有重要价值, 下面是预测总量相对误差公式:

$$\varepsilon = \frac{\sum_{i=1}^Q (f_i - \hat{x}_i)}{\sum_{i=1}^Q \hat{x}_i} \times 100\%$$

其中,  $f_i$  是预测值,  $\hat{x}_i$  是实际值, 利用总量相对误差指标的衡量将误差百分比化, 表 1 列出了预测评价的一般性要求作为判断参考标准。

表 1 预测精度标准

| 相对误差             | 预测精确度 |
|------------------|-------|
| $\leq 10\%$      | 高精度   |
| $10\% \sim 20\%$ | 良好    |
| $20\% \sim 50\%$ | 一般    |
| $> 50\%$         | 不准确   |

最后, 根据非参数核估计预测模型, 可以利用预测结果对可以做出的补货行为, 判断当前库存是否满足销售预测总量, 可能产生缺货量或库存剩余量, 针对不同的库存状况提出相应的建议, 即补货或促销, 对于滚动补货方式非常有用。这些建议有助于设计灵活、机动的补货计划。

### 3 数值实验结果分析

按上面的动态预测模型对某品牌公司的夏季 41 个品类服装进行了实际预测分析,以小品净色(男 POLO)为例,假设当前日期是生命期第 150 天(当日日期为 2015 年 6 月 27 日),即  $k=150$ ,则需要计算从第 151 天开始至生命期结束(2015 年 6 月 28 日至 2015 年 8 月 21 日)这段时间内的预测销量,输出结果如图 5—图 8 所示,生命期内每天都可得出预测参考方案。

通过 Matlab 对 2015 年夏季 41 个品类进行了核密度估计模型的预测实验,并对各个小品品类的预测结果进行统计分析和偏差计算,通过分析其每日预测的变化趋势和误差比例,对比检验利用非参数核估计方法建立的预测模型的预测精度。

抽选 41 个夏季小品品类来检验预测精度,发现销量在一万以下的 20 个品类的预测准确性偏低,不过一万以下销量的品类由于销售量级总体较低,一般通过订货会一次订货完成,不再进行补货动作,以及数量较小导致预测的不稳定性较大,所以对于一万以下的品类销量预测的结果不建议作为参考。本次重点分析对比 21 款销量在一万以上的夏季小品品类的预测结果。

夏季小品品类的在架销售时间较长,从 3 月初到 9 月初,生命周期跨度约有 170 日左右,其中试销期约 30 天,旺销点大多出现在 5~6 月,这 60 天左右是服装进入二次补货动作的重要时期,预测结果对于补货生产计划具有参考作用,而在 6 月底到 7 月初,即生命期的 120 天左右,是店铺开始进入打折促销活动的阶段,预测结果可以为促销期间的折扣力度和定价提供一定的建议。根据试销期和已发生的销售情况,在出现热销前把握好补货机会合理安排补货数量,对于防止缺货损失和库存过量具有重要意义,合理的定价促销安排对于提高销量、降低库存、提高利润具有积极价值。本文在检验时决定选取第 30、60、90、120 天这 4 个较为关键的时间点,来观察预测销量偏差和变化。

下面 4 幅图分别为 4 个品类在第 30、60、90、120 天的非参数核估计拟合曲线与实际销量曲线的对比图,可以较为直观地看出核密度估计拟合的曲线在一定程度上可以描述实际销量的走势,其中净色男 POLO 和条纹男 POLO 两个品类的曲线从第 30 天的预测到 120 天的预测有较为明显的愈来愈准确趋势,净色女休闲长裤和牛仔女衬衫在各个测试日都拥有较好的拟合效果,能较早地获得更为准确的预测结果。这 4 个品类利用核估计的预测方法在拟合程度上有一定区别,但总体上都能够判断产品在未来的销售趋势,且判断效果随着获得实际数据的增加越来越准确。

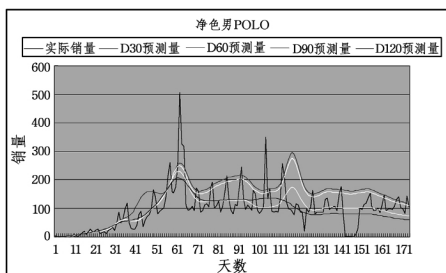


图 5 净色男 POLO 在不同日期的预测图

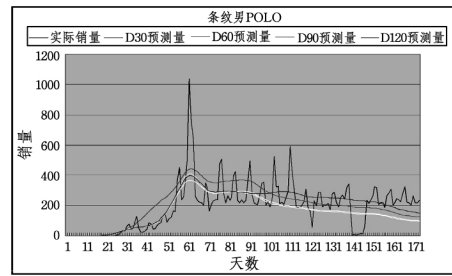


图 6 条纹男 POLO 在不同日期的预测

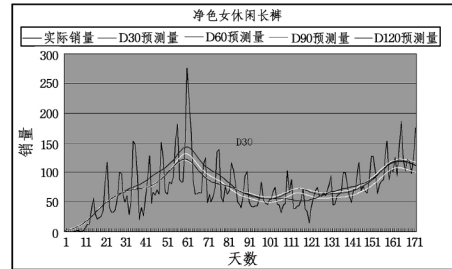


图 7 净色女休闲长裤在不同日期的预测

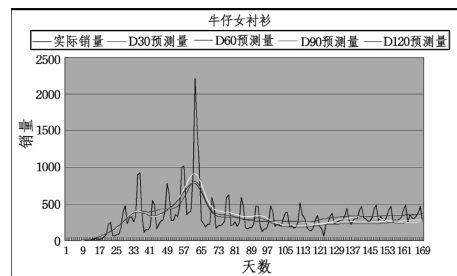


图 8 牛仔女衬衫在不同日期的预测

从上述 4 个测试日的预测数据误差中可以更为直观地看出,预测精度在不断提升,误差量呈现出了不断缩小的趋势。不论是品类销量的平均绝对误差还是均方误差,都在不断变小,尤其是均方误差的降低趋势更为明显,第 60 天的预测精度提升十分显著。由于品类间总销量量级的不同以及个体间预测效果差距影响,品类间也存在明显的误差大小之别。不过,两张误差表中同一品类基本都表现出了误差趋小的态势,且在 D90、D120 天基本处于较小的误差范围内,在 D60 日的预测结果相对于销售量级也还都比较理想。结合图片和表格,基本可以肯定本文设计的非参数核估计预测模型能够利用更新的销售数据对未来销量的预测提升预测精度,在关键的时期得到可靠的预测结果作为参考。

在通过非参数核估计方法进行预测实验的同时,本文使用提供样本数据的服装公司原先的销量预测方法给出的结果以及通过时间序列法中的简单移动平均法对夏季品类历史数据进行销量预测,计算了这两种销量预测方法得出的预测总量的相对误差,将其与使用非参数核估计方法在 D60 日时的总量相对误差  $\epsilon$  进行比较。

表 2 中正数比例表示预测值高于实际值,负数比例表示预测值低于实际值。从表中横向比较 3 种方法计算得到的各个小品品类的误差比例可以看出,该公司原来使用的预测方法所产生的预测误差最大,预测精度最低,简单移动平均法的预测效果相对于原来的方法有所改善,但预测效果最好的是非参数核估计方法,本次试算的 21 个品类中有 15 个品类预测的总量相对误差小于其他两类预测方法。

表 2 不同销量预测法的品种预测总量相对误差表

| 品类名称        | 传统<br>预测方法 | 简单<br>移动平均法 | 非参数<br>核估计法 |
|-------------|------------|-------------|-------------|
| 净色(男 POLO)  | -37.43%    | 43.32%      | 35.18%      |
| 条纹(男 POLO)  | -31.90%    | -13.92%     | -17.15%     |
| 条纹(女翻领 T 恤) | -29.95%    | 18.43%      | -13.56%     |
| 印花(女翻领 T 恤) | -63.07%    | -47.42%     | -26.14%     |
| 拼接(男 T 恤)   | 78.11%     | -26.11%     | 22.28%      |
| 条纹(男 T 恤)   | -30.94%    | -24.42%     | -17.65%     |
| 两件套(女 T 恤)  | 61.70%     | 21.32%      | 1.59%       |
| 条纹(女 T 恤)   | 19.88%     | -13.72%     | -18.46%     |
| 牛仔(男衬衫)     | 36.07%     | 25.03%      | 14.66%      |
| 净色(男衬衫)     | -24.77%    | -18.44%     | 15.08%      |
| 牛仔(女衬衫)     | 29.46%     | 23.00%      | -8.07%      |
| 格子(女衬衫)     | 46.08%     | -11.21%     | -8.56%      |
| 基本款(男休闲长裤)  | 28.13%     | 11.34%      | 2.80%       |
| 印花(男休闲长裤)   | -28.62%    | -22.05%     | 27.06%      |
| 西裤(男休闲长裤)   | 51.49%     | -12.40%     | -14.96%     |
| 基本款(女休闲长裤)  | 39.12%     | 14.26%      | 6.43%       |
| 净色(女休闲长裤)   | -10.83%    | -13.68%     | 6.19%       |
| 梭织(连衣裙)     | -33.17%    | -33.03%     | -33.58%     |
| 针织(连衣裙)     | -17.19%    | -11.43%     | -12.34%     |
| 修身小脚(男牛仔长裤) | -23.16%    | -16.90%     | -15.76%     |
| 修身小脚(女牛仔长裤) | -33.22%    | 20.19%      | -8.10%      |

从总量相对误差的数字本身来看,对照相对误差的标准表格,过去的预测方法和简单移动平均法预测的小品销量都无法达到高精度的预测结果,其中过去的预测方法在 21 个样本中只有两个品类的误差在 20% 以内,还有 4 个小品预测不准确。而简单移动平均法的预测结果虽然好于原来的预测方法,但预测精度不高,有近  $\frac{1}{2}$  的小品相对误差只能达到合理范围,结果不够理想。非参数核估计方法预测的小品中有 7 个相对误差小于 10%,达到高精度的预测效果,占  $\frac{1}{3}$ ;有 9 个小品相对误差在 10% 到 20% 之间,预测精度良好;还有 5 个品类相对误差属于合理范围,所有小品误差都不大于 50% 的不合理状态,大多数预测精度能达到良好以上。

为进一步确定相对误差所体现的预测模型的拟合精度,还可以从误差的均值和稳定性两方面进行评估。均值即是将所有样本小品的误差比例的绝对值加上总计算平均相对误差绝对值,如表 3 可以分别得到对应的 35.92%、21.03%、15.50% 的结果。过去的预测方法的误差比例接近于该公司目前每年的过剩库存比例,而非参数估计法的预测精度良好,且大幅缩小了原来的预测误差,明显优于其他两种预测方法,对于提高时尚休闲服饰的销量预测精度具有实际价值。稳定性指标可以通过计算所有样本小品的误差比例的绝对值的标准差和方差来观察,当结果越接近于零,表明预测模型预测出的销量结果拟合精度的稳定性越好,将估计值用于指导公司制定生产和订货计划的可靠性越高;如果标准差越大,则预测结果准确程度的稳定性越低,作为参考的可靠性就越差。表中非参数核估计的标准差和方差都是最小的,预测精度的稳定性明显优于其他两种方法。

表 3 品类总量相对误差的均值与稳定性指标

|     | 传统预测方法  | 简单移动平均法 | 非参数核估计法 |
|-----|---------|---------|---------|
| 均值  | 0.3592  | 0.2103  | 0.1550  |
| N   | 21      | 21      | 21      |
| 标准差 | 0.16297 | 0.09969 | 0.09273 |
| 方差  | 0.027   | 0.010   | 0.009   |

结束语 本文给出了一种非参数核估计动态预测模型,其可以应用于品牌品类服装实际的销量预测中的预测模型,并解释了预测过程中所需要的数据以及对数据的处理方式,对整个实例应用的预测过程进行了分析。最后为了验证预测结果的可靠性,通过计算预测精度的相对误差指标进行检验。通过将理论模型与实际数据实验相结合,该方法实现容易,预测稳健性好,利用非参数核密度估计函数建立的预测模型在实际的销量预测中可以发挥重要的指导作用。

利用 Matlab 进行编程试算非参数核估计函数的预测模型,基于 2014 年的夏季销售数据和 2015 年不断更新得到的已发生的夏季实际销量,用较少的数据就能预测出夏季服装的日销量和总体销量,为公司二次订货提供参考意见。对预测精度的各项指标进行检验以及不同预测方法的对比分析可以发现,非参数核估计法计算出的预测结果精确度良好且明显优于其他两种方法,对于提高服装销量预测的精确度和降低公司库存具有积极意义,为公司订货提供的参考和指导也具有一定的可靠性。

### 参考文献

- [1] Adamowskik. Nonparametric kernel estimation of frequencies [J]. WaterResources, Research, 1985, 21(11): 1585-1590
- [2] Efron B. Bootstrap Methods; Another look at the jackknife [J]. The Annals of Statistics, 1979, 7(1): 1-26
- [3] Silverman B W, Young G A. The Bootstrap: To smooth or not to smooth [J]. Biometrika, 1987, 7(4): 469-479
- [4] Bayus, Barry L. High Definition Television; Assessing Demand Forecasts for Next Generation Consumer Durable [J]. Management Science, 1993, 39: 1319-1333
- [5] Chandra C, Grabis J. Application of multi-steps forecasting for restraining the bullwhip effect and improving inventory performance under autoregressive demand [J]. European Journal of Operational Research, 2005, 166(2): 337-350
- [6] Donohue K L. Efficient supply contracts for fashion goods with forecast updating and two production modes [J]. Management Science, 2000, 46: 1397-1411
- [7] Frees E W, Miller T W. Sales forecasting using longitudinal data models [J]. International Journal of Forecasting, 2004, 20: 99-114
- [8] Parzen E. On estimation of a probability density function and mode [J]. The Annals of Mathematical Statistics, 1962, 33(3): 1065-1076
- [9] Bass F M. A new product growth model for consumer durables [J]. Management Science, 1969, 15: 215-227
- [10] Gallego G Moon I. The distribution free newsboy problem; review and extensions [J]. Journal of Operational Research Society, 1993, 44(1): 825-834
- [11] Gardner J E S. Exponential smoothing; The state of the art-Part II [J]. International Journal of Forecasting, 2006, 22(4): 637-666
- [12] Hsu P, Teng Hui-ming. Coordinated ordering decisions for products with short lifecycle and variable selling price [J]. Computer & Industrial Engineering, 2008, 54: 602-612
- [13] Vemon R. International Investment and International Trade in the Product Cycle [J]. Quarterly Journal of Economics, 1966, 80: 190-207

(下转第 465 页)

由于本文的研究样本没有按照以往研究的配对抽样法进行抽样,因此在验证预测准确率方面,不适合建立以传统 ST 方法为基础的预测模型进行对比验证。但是本文的预测结果包含 4 种不同的财务危机程度,与以往只能预测是否具有财务危机的模型相比,具有更好的应用价值。而且根据目前的研究文献来看,以往大部分的财务危机预测模型的准确率在 70%到 85%之间。根据表 6 可以看出,本文方法在预测“危机”与“安全”时,具有较高的准确率,尽管在预测“较安全”与“一般”时准确率不够,但这对该方法应用价值的影响相对较小。因此本文的研究在保证较高的准确率的基础上,不但可以预测企业是否具有财务危机,还能预测财务危机的程度。

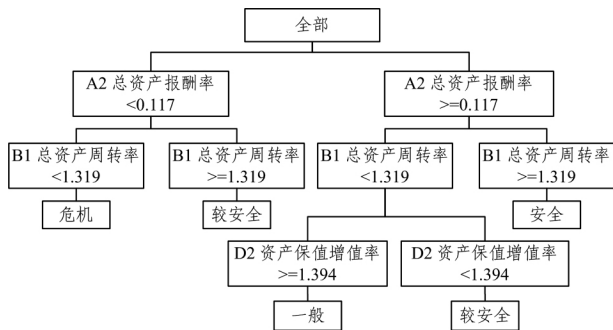


图 1 决策树结果

在企业财务危机预测中,错误主要有两类:第一类是预测的结果比实际结果要好,第二类是预测的结果比实际结果要差。Sinkey 证实了,在预测模型中,总体预测正确率高但是其中第一类错误率高的错误成本,要高于总体预测正确率低而其中第一类错误率低的错误成本<sup>[13]</sup>。本文的预测模型的错误主要来源于第二类错误,此类错误要优于上面提到的第一类错误。

结束语 本文将数据包络引入财务危机预测模型,对财务危机程度进行了细分,取得了良好的效果。就目前研究的结果来看,总资产报酬率是一个相当稳健的指标,该指标不但是反映企业总体盈利能力的重要指标,同时对企业财务危机的预测也有着重要的参考意义。无论是管理者,还是外部投资者,都应着重关注企业总资产报酬率的情况。而相关的企

业财务综合评价模型,也可适当地调高该指标的权重,以获取更为精准的评价结果。

## 参考文献

- [1] 边海容,万常选,刘德喜,等.考虑 Web 金融信息的上市企业财务危机预测模型研究[J]. 计算机科学,2013,40(11): 295-298
- [2] Altman E I. Discriminate Analysis and Prediction of Corporate Bankruptcy[J]. Financial Ratios,1968(9):589-609
- [3] Lau A H. A five state financial distress prediction model[J]. Journal of Accounting Research,1987,25(1):127-138
- [4] Ohlson J S. Financial Ratios and the Probabilistic Prediction of Bankruptcy[J]. Journal of Accountant Research,1980,18(19): 103-131
- [5] Aziz A D. Emanuel and G Lawsom. Bank Prediction: An Investigation of Cash Flow Based Models[J]. Journal of Management Studies,1988(25):419-437
- [6] 周敏,王新宇.基于模糊优选和神经网络的企业财务危机预警[J]. 管理科学学报,2002(3):86-90
- [7] 吕长江,韩慧博.上市公司财务困境预测方法的比较研究[J]. 吉林大学社会科学学报,2005(6):99-109
- [8] 文拥军.基于多类 Logistic 的财务困境预警模型实证研究[J]. 中国海洋大学学报,2009(2):46-50
- [9] 陈磊,任若恩,曹汉平.公司多阶段财务危机动态预警研究[J]. 系统工程理论与实践,2008,28(11):29-35
- [10] 邱玉莲,朱琴.基于支持向量机的财务预警方法[J]. 统计与决策,2006(16):153-155
- [11] 柏云燕.数据包络分析法在高校财务绩效评价中的应用[J]. 电子商务,2012(9):58-62
- [12] 李小程.基于数据包络分析的房地产投资项目财务效率评价研究[J]. 湖南财政经济学院学报,2011(8):94-98
- [13] 张坤.基于数据包络分析:决策树的企业财务评价与预警——以西部四省上市企业为例[J]. 财务与金融,2014(3):36-40
- [14] 徐凤江.数据包络分析法分析企业财务[J]. 现代商贸工业,2009,21(4):276-278
- [15] Sinkey. A Multivariate Statistical Analysis of the Characteristics of Problem Bank[J]. Journal of Finance,1975,3(1):21-38

(上接第 460 页)

- [14] Zhou Jian-heng, Shu Ling. Strategy for supply chain coordination based on fashion life cycle[C]//2010 International Conference on Management and Service Science, Wuhan, 2010:1-5
- [15] 陈希孺,柴根象.非参数统计教程[M].上海:华东师范大学出版社,1993
- [16] 程秩.常用预测方法及评价综述[J]. 四川师范大学学报,2002,22:5-6
- [17] 胡觉亮,何秋霞,韩曙光,等.基于改进的 BASS 模型的服装产品生命周期研究[J]. 浙江理工大学学报,2010,1(27):69-73
- [18] 胡觉亮,孔云鹏,韩曙光.需求信息更新下服装供应链的最优订货决策研究[J]. 纺织学报,2012,33(11):136-140
- [19] 胡敏敏.析 SPA 型服装品牌零售模式——以 ZARA 为例[J]. 浙江理工大学学报,2007,7:404-406
- [20] 刘咏梅,王朝晖,王咏.基于产品生命周期的服装库存管理智能决策[J]. 东华大学学报,2004,30(2):34-38
- [21] 刘铮,徐琪.基于服务一购买意愿的双渠道短生命周期产品需求

预测模型——以快时尚服装行业为例[J]. 东华大学学报,2013,5(39):668-681

- [22] 乔亮亮.多元线性回归预测法在服装制造中的应用[J]. 长江大学学报,2010,1(7):145-160
- [23] 盛方正,季建华.需求预测信息更新的零售商三阶段最优采购策略[J]. 管理工程学报,2009,2:156-159
- [24] 徐贤浩,陈雯,廖丽平,等.基于需求预测的短生命周期产品订货策略研究[J]. 管理科学学报,2013,4:22-23
- [25] 楼俊钢,蒋云良,申情,等.软件可靠性预测中不同核函数的预测能力评估[J]. 计算机学报,2013,36(6):1303-1311
- [26] 王静,蔡勇,蒋刚.基于多尺度核函数的铆接件腐蚀疲劳预测[J]. 计算机应用研究,2015,32(4):1074-1077
- [27] 汪廷华,陈峻婷.核函数的选择研究综述[J]. 计算机工程与设计,2012,33(3):1181-1186
- [28] 梅玲,张亚丽.结构化数据核函数综述[J]. 计算机工程,2010,36(23):199-201