

基于支持向量机的建筑企业利润总额预测 ——以辽宁省建筑企业为例

王秋菲,唐冰洁

(沈阳建筑大学商学院,辽宁 沈阳 110168)

摘要:对比分析了指数平滑法、BP神经网络分析法及支持向量机3种预测方法,并以辽宁省建筑企业利润总额的预测为例,比较分析了这3种方法的预测值及实际值。通过分析可知,支持向量机具有更好的短期预测效果,能较好地模拟并预测建筑企业利润总额变化,为准确预测企业利润总额提供了重要依据。

关键词:利润总额;预测方法;支持向量机;建筑企业

中图分类号:F424.7

文献标志码:A

随着产业结构的优化升级以及国家宏观经济政策和区域产业政策的深入推进,中国经济进入“新常态”发展阶段,企业也同样进入新的发展阶段。利润是业主的投资回报增加和工人报酬提高的来源,也是企业资本积累和业务规模扩张的源泉。利润不但可以使企业的管理者保持自己的地位,而且也是企业蓬勃发展所必需的。追求利润是每个企业生存和持续经营的动力,利润最大化是每个企业所追求的目标。在社会主义市场经济条件下,现代企业要建立自己的管理制度,而优胜劣汰的竞争机制是最常见的,企业要想在竞争中保持优势,就要保证获得较高的利润。

因此,对企业未来有关数据进行预测是企业应该充分考虑的问题。项目或企业可以根据预测出的相关数据有针对性地调整后续发展策略,规避风险,赢得利润。而科学的预测方法是数据准确性的必要保障,现有的预测方法包括指数平滑法、时间序列预测法、BP神经网络法以及支持向量机等,这些方法

在不同领域得到了广泛的应用。笔者将对指数平滑法、BP神经网络算法及支持向量机在建筑企业利润预测方面的优劣进行分析,并以辽宁省建筑企业利润预测为例,提出合理的意见和建议。

一、预测方法对比分析

随着科技的发展,预测方法也在不断更新调整,不同领域会应用不同的预测方法来进行科学合理的分析,以便得到更为准确的数据。

1. 指数平滑法

指数平滑法是基于移动平均法开发的时间序列分析和预测方法,它通过计算指数平滑值和构建一定的时间序列预测模型来预测现象的未来。指数平滑法是加拿大人霍尔特于1957年首先提出的,他将所提出的布朗指数平滑模型用于非线性模型的预测。1965年,美国人首先提出时间序列模型的最佳预测结果可以由指数平滑法产生。20世纪70

年代后期,在考虑数据季节性和趋势变化等问题后,出现了经过优化的温特线性 and 季节性指数平滑模型,至此,指数平滑法的应用范围得到了扩展。随后加德纳于1985年发表的专门论述指数平滑法的文章影响至今,为后来指数平滑模型的应用指引了方向,将比较健全的指数平滑模型推广到需要预测的各个学科领域。

2010年,张根保^[1]等运用动态二次指数平滑法观察和预测各类用户需求随时间变化趋势;2013年,谢晓燕^[2]等根据呼、包、鄂三角地区物流的实际需求,结合指数平滑预测模型对物流需求量进行了短期预测。指数平滑方法按时间先后顺序对之前的观察值进行加权,充分合理利用之前的数据,获得合理、准确的预测结果。根据实际情况分析可知,指数平滑方法易于掌握,操作简单,更适合进行计算量较少的预测。

由于指数平滑方法可以根据较少的资料预测所需的结果,而且包含移动平均和全期平均的优点,被广泛接受和使用。但该方法不会丢弃过去的的数据,只给出一个渐进的影响程度,即给出逐渐趋于零的平滑指数,因而这种方法给出的长期比例较小,近期比例较大,所以更适合短期的预测。

2. BP神经网络算法

BP神经网络预测方法,又称为反向传播神经网络(Backpropagation Neural Network),简称BP神经网络,是一种用来训练多层感知机的监督学习算法。BP神经网络在1986年由Rumelhart和McClelland等科学家提出,是一种由误差反向传播算法训练的多层前馈神经网络。由于BP神经网络成功地解决了非线性连续函数求解时多层次前馈神经网络调整权重的问题,成为目前应用最为广泛的神经网络,近90%的神经网络模型是BP神经网络及其变形。其在矿产开采、农业害虫预测和农产品质量检测等领域得到了广泛的应用,取得了较好的预测效果。

宋宇辰^[3]等以包头市二氧化硫、二氧化氮和可吸入颗粒物为例,利用BP神经网络

预测了这些颗粒物2010年的年均浓度以及2011年3月—2012年12月的月平均浓度。在预测空气污染物的年平均浓度方面,BP神经网络简单有效,可以很好地找到经济因素和大气污染物的规律性。因此,BP神经网络在空气预报领域势必会得到更广泛的应用。然而,在BP神经网络训练过程中,各层神经元节点数、层间传递函数与训练方法只能通过大量的试验才能得到,因此,需要改进模型以适应综合应用。在对价格预测方法的研究中,贺清哲^[4]构建了BP神经网络时间序列预测模型,以钢材价格为实例进行研究,发现该方法预测钢材价格变化方面效果良好。虽然BP神经网络在应用上存在难以解释、缺乏稳定性等缺陷,但在一定程度上突破了传统预测方法的许多局限性,在实际应用中有较广泛的应用,是一种非常具有前景的预测方法。

BP神经网络是一个复杂的非线性系统,可以用来处理非常复杂的经济问题。然而,其恒定的学习速度导致网络收敛速度较慢,在对企业利润总额进行预测时,运用该方法需要很长的训练时间才能收敛到最优状态^[5]。由于BP神经网络没有明确的理论依据对隐含层网络节点和网络结构进行选择,所以在大多数情况下,都是根据经验或模仿其他模型来确认,增加了实验人员的主观因素,也增加了网络学习的难度,而且大部分BP神经网络的连接权值都需要人工设定,所有这些原因都会导致最终的仿真结果误差过大,这种人为误差也会对算法的泛化能力、鲁棒性和容错性产生影响^[6]。总体而言,BP神经网络在网络理论和性能上都已经比较成熟,其突出的优点是具有强大的非线性映射能力和灵活的网络结构。但是,BP神经网络有学习速度慢、容易陷入局部极小值的缺点,在网络层数和神经元选择等方面也存在着缺乏相应的理论指导和网络推广能力有限的缺陷^[7]。

3. 支持向量机

支持向量机(Support Vector Machine,

SVM)最早由 Corinna Cortes 和 Vapnik 等人于 1995 年提出。它是一种在数据挖掘方面通过优化方法来解决机器学习问题的工具,是一种基于统计学习理论的模式识别方法。它克服了“维数灾难”,解决了过度学习和欠学习等问题,在求解非线性和高维模式识别问题上显示出很多独特的优势,可以广泛应用于其他机器学习,如函数拟合问题。

王平等^[8]通过支持向量机,利用武汉市的 SVM 预测模型和实际数据对房地产进行了单指标预测,并将结果与 BP 神经网络的预测结果进行比较,结果表明在房地产单指标预测方面支持向量机有更高的精度,说明基于支持向量机回归的房地产单指标预测模型具有很强的泛化能力。贺本岚^[9]使用支持向量机模型,结合商业银行客户损失样本数据对客户损失进行预测,采用随机抽样的方法对支持向量机模型进行优化,在与 Logistic 回归模型预测结果进行对比中发现支持向量机的预测效果更优。

支持向量机是一种由扎实的理论作支撑的新型小样本的学习方法,它不涉及大数定律以及大数概率测度等,避免了传统过程中从归纳到演绎这一程序,实现了训练样本“转导推理”至预测样本,极大地简化了常见的分类和回归问题。支持向量机的最终决策仅由少数支持向量决定,因此,计算复杂度取决于支持向量的数量,而不是样本空间的维数。从某种意义上说,这样可以避免“维度灾难”。

通过对 3 种预测方法的分析比较可知,在进行小样本预测时,支持向量机模型通过权衡复杂性与学习能力,使该模型具有较强的准确性和泛化能力。该模型在非线性和高维模式识别方面同样具有很大的优势,克服了神经网络在过度学习、学习不足、局部最小等问题上的局限性,而且克服了指平滑法仅适用于较少计算量的情况,已成为继神经网络后研究机器学习理论领域的新热点。因此,笔者将以建筑企业利润总额为例,利用支持向量机进行预测分析。

二、支持向量机的相关理论及预测过程

1. 支持向量机的理论基础

支持向量机回归模型(SVM)可以较好地处理非线性回归问题^[10]。该方法的优点包括学习速度较快、全局最优以及泛化能力强等,并且该方法的学习结果较其他传统的回归预测方法也更加精确。

利用 SVM 进行预测分类的基本思想是通过非线性映射 φ 将数据 x_i 映射到高纬度特征空间 F 中去,在此基础上进行线性回归。

对于样本数据 $\{x_i, y_i\}, i = 1, 2, \cdots, m;$
 $x_i \in R^m, y_i \in R^m$, 回归函数为

$$f(x) = (\omega \cdot \varphi(x)) + b$$

R^m 空间到 F 空间的非线性映射用 $\varphi(x)$ 表示,权向量用 ω 表示,偏差水平用 b 表示。传统的预测方法是在集合中查找 $f \in F$,使结构风险最小化。结构风险公式为

$$R_{reg} = \lambda \|\omega\|^2 + R_{emp}[f] = \sum_{i=1}^s C(e_i) + \lambda \|\omega\|^2$$

置信风险用 $\|\omega\|^2$ 来表示,经验风险用 $R_{emp}[f]$ 来表示, λ 是用来平衡功能复杂性和损失误差的常数,模型经验损失用 $C(e_i)$ 来表示,样本误差值用 e_i 来表示,样本量用 S 来表示。对于给定的损失函数,该式可以转化为最小化的线性风险泛函问题

$$\min \eta = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^S (\zeta_i + \zeta_i^*)$$

$$s. t. \begin{cases} y_i - (\omega, \varphi(x_i)) - b \leq \varepsilon + \zeta_i \\ (\omega, \varphi(x_i)) + b - y_i \leq \varepsilon + \zeta_i \\ \zeta_i, \zeta_i^* \geq 0 \end{cases}$$

$C = 1/\lambda$,估计精度用 ε 来表示,松弛变量用 ζ_i, ζ_i^* 来表示,具体取值为

$$\xi_i^* = \begin{cases} 0, & f(x_i) - y_i < \varepsilon \\ |f(x_i) - y_i| - \varepsilon, & f(x_i) - y_i \geq \varepsilon \end{cases}$$

为了在求解时更加方便快捷,通过引入拉格朗日乘子 α_i 和 α_i^* ,可以将这类问题转化为对偶问题

$$\max \mu = -\frac{1}{2} \sum_{i,j=1}^S (\alpha_i - \alpha_i^*)(\alpha_j^* - \alpha_j)(\varphi(x_i) \cdot \varphi(x_j)) + \sum_{i=1}^S \alpha_i^*(y_i + \varepsilon) - \sum_{i=1}^S \alpha_i(y_i + \varepsilon)$$
$$s. t. \begin{cases} \sum_{i=1}^S \alpha_i^* = \sum_{i=1}^S \alpha_i \\ 0 \leq \alpha_i \leq C \\ 0 \leq \alpha_i^* \leq C \end{cases}$$

通过求解,将得到 ω 权向量,将求解结果带入 $f(x) = (\omega \cdot \varphi(x)) + b$,得到非线性函数^[11]

$$f(x) = \sum_{i=1}^S (\alpha_i - \alpha_i^*)(\varphi(x_i) \cdot \varphi(x)) + b$$

2. 支持向量机模型的预测过程

支持向量机模型预测的过程共分 6 步(见图 1):第一步,选取用于预测的指标和数据;第二步,数据预处理;第三步,给出指定参数;第四步,对具有不同核函数的 SVM 进行训练,核函数包括线性函数、多项式核函数、sigmoid 核函数和径向机核函数等,针对不同的核函数会有不同的训练方式及过程;第五步,使用前一步训练的支持向量机模型进行回归预测;第六步,根据回归预测结果,选择出最佳的方案,并以此建立相关回归预测模型。

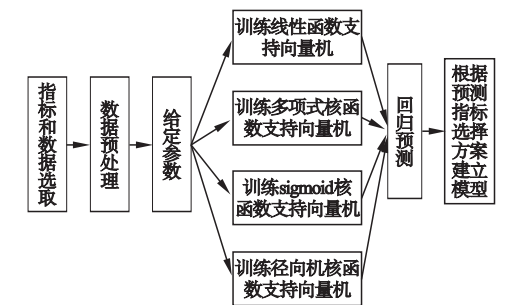


图1 支持向量机模型的预测过程

三、辽宁省建筑企业利润预测分析

1. 辽宁省建筑企业利润总额基本情况

进入 21 世纪之后,随着国民经济的发展,辽宁省建筑企业也日益发展,成为经济发展至关重要的一部分。国家统计局数据显示,辽宁省建筑企业利润总额由 1997 年的 -0.87 亿元增长到 2016 年的 121.14 亿元,20

年间增长迅速,具体情况如表 1 所示。

表 1 1997—2016 年辽宁省建筑企业利润总额
亿元

年份	利润总额	年份	利润总额
1997	-0.87	2007	83.68
1998	-4.25	2008	88.35
1999	3.10	2009	118.89
2000	6.67	2010	173.96
2001	14.82	2011	224.31
2002	12.70	2012	256.08
2003	16.51	2013	315.70
2004	29.35	2014	254.68
2005	54.39	2015	168.79
2006	58.17	2016	121.14

2. 对建筑企业利润总额的预测过程

- (1)选取年份作为输入参数,各年建筑企业利润总额作为输出参数。
- (2)采用 mapminmax 函数进行试验数据的归一化处理。为使数据更容易进行学习和训练,在使用支持向量机回归模型进行预测时,要事先归一化处理相应指标,将指标值控制在 $[-1,1]$,使得处理后的数据更容易训练和学习。
- (3)采用 K - CV 法优化惩罚参数 c 和径向基核参数 g ,从而搜寻到网络性能最佳时的 c, g 数值。
- (4)借助最佳参数 c, g 训练支持向量机预测模型,实现对数据的预测。预测流程如图 2 所示。

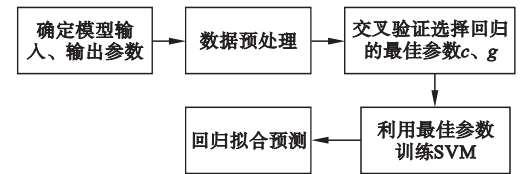


图2 支持向量机算法流程图

3. 对建筑企业利润总额的预测结果

针对建筑企业利润,笔者选用指数平滑法、BP 神经网络分析法以及支持向量机 3 种方法进行了预测分析,相关预测结果如表 2 所示。

笔者利用指数平滑法、BP 神经网络分析法及支持向量机 3 种方法,对 1997—2016 年辽宁省建筑企业利润总额进行了预测,并将

表 2 辽宁省建筑企业利润总额 3 种预测方法比较分析

亿元

年份	实际值	指数平滑法		BP 神经网络		支持向量机	
		预测值	相对误差	预测值	相对误差	预测值	相对误差
1997	-0.87	-1.59	0.72	-8.26	7.39	-2.60	1.73
1998	-4.25	-3.72	-0.53	-3.55	-0.70	-4.21	-0.03
1999	3.10	-7.45	10.55	33.87	-30.77	2.85	0.25
2000	6.67	8.29	-1.62	34.24	-27.57	7.05	-0.38
2001	14.82	11.62	3.20	34.38	-19.56	15.15	-0.33
2002	12.70	22.17	-9.47	35.45	-22.75	14.92	-2.22
2003	16.51	12.79	3.72	35.48	-18.96	19.71	-3.20
2004	29.35	18.63	10.72	36.43	-7.08	32.23	-2.88
2005	54.39	40.42	13.97	38.63	15.76	55.57	-1.18
2006	58.17	77.71	-19.54	48.54	9.63	60.81	-2.65
2007	83.68	67.24	16.44	57.22	26.46	85.32	-1.65
2008	88.35	103.96	-15.61	88.71	-0.36	91.59	-3.24
2009	118.89	97.77	21.12	135.10	-16.21	121.21	-2.32
2010	173.96	143.66	30.30	165.69	8.29	173.61	0.35
2011	224.31	225.09	-0.78	205.02	19.28	221.90	2.41
2012	256.08	277.84	-21.76	230.59	25.49	252.92	3.16
2013	315.70	292.11	23.59	299.16	16.54	308.06	7.64
2014	254.68	368.44	-133.76	253.30	1.38	255.83	-1.15
2015	168.79	218.71	-49.92	163.69	5.10	179.92	-11.13
2016	121.14	140.77	-19.63	134.57	-13.43	173.31	-52.17

3 种方法的预测结果与实际值进行了比较。可以清楚地看出,3 种方法对辽宁省建筑企业利润总额的预测都有一定的准确性。计算均方误差可以得到,指数平滑法的均方误差为 982.38,BP 神经网络预测法的均方误差为 297.39,支持向量机的均方误差为 148.75。通过比较也可看出,支持向量机的均方误差最小,准确性更强。因此,运用支持向量机对建筑企业利润总额进行进一步的预测,具体情况如表 3 所示。

表 3 利用支持向量机对辽宁省建筑企业利润总额的预测

亿元

年份	SVM 预测值	年份	SVM 预测值
2017	165.83	2019	214.46
2018	235.76	2020	398.24

由表 3 可知,基于支持向量机的预测结果呈现波动趋势,这恰好符合经济波动状况,由此可以说明该方法的预测结果有一定的可信度。通过对辽宁省建筑企业利润总额的高精度预测,可以针对不同阶段制定准确高效的经营战略,对各个影响因素进行有针对性的调整,降低经营成本,提高经营效率,从而提高企业利润。

四、结 语

笔者采用指数平滑法、BP 神经网络分析法及支持向量机方法的回归预测模型对辽宁省建筑企业利润总额进行了预测,对这 3 种方法的比较分析结果证明支持向量机模型用于建筑企业利润回归预测能够取得很好的效果。建筑企业利润的影响因素较多,容易受到各种因素的干扰,所以预测难度相对于其他领域更大,但仍取得了较好的预测效果,因此,该模型也适用于其他企业利润方面的预测。但是,目前该方法只是从技术的角度对建筑企业利润总额进行了短期回归预测,还有一定的局限性。下一步,从影响建筑企业利润总额的各种指标入手,对建筑企业利润总额进行长期预测将是笔者的研究重点。

综上所述,针对不同领域的预测分析提出如下建议:

(1) 支持向量机的建立对样本的数量和结构有一定的要求,而实际预测时,往往又会出现历史样本收集不够准确、不充足的情况,这就会影响到预测结果的稳定性和准确性。因此,一些领域在进行数据分析时应注意所

选样本的数量,尽可能做到足够充足和准确。

(2)由于偶然因素的影响,历史数据样本的某些区域可能会扭曲某些预测模型的预测,因此,在针对这些数据容易受到影响的领域进行数据分析时要将该因素充分考虑进去。

(3)在针对与政府宏观调控有关的数据进行分析预测时,同样要考虑政府的宏观调控等因素,以便确定准确有效的预测模型。

参考文献:

[1] 张根保,李玲,纪富义. 基于成分数据动态指数平滑的用户需求变化趋势预测模型[J]. 统计与决策,2010(14):32-35.

[2] 谢晓燕,韦学婷,王霖. 基于指数平滑法的呼、包、鄂三角区物流需求量预测[J]. 干旱区资源与环境,2013(1):58-62.

[3] 宋宇辰,甄莎. BP神经网络和时间序列模型在包头市空气质量预测中的应用[J]. 干旱区资源与环境,2013,27(7):65-70.

[4] 贺清哲. 基于人工神经网络的钢材价格预测[J]. 现代商贸工业,2015(26):65-66.

[5] 负永峰,范永慧,孙扬. 基于BP神经网络的隧道围岩力学参数反分析方法[J]. 沈阳建筑大学学报(自然科学版),2011,27(2):292-296.

[6] 刘莉,叶文. 基于BP神经网络时间序列模型的降水量预测[J]. 水资源与水工程学报,2010(5):156-159.

[7] 赵永未,杨力纲,袁兴明. 基于BP神经网络的矿山GPS数据时间序列预测分析[J]. 科技创新导报,2011(34):123-124.

[8] 王平,师青. 基于支持向量机回归的房地产单指标预测[J]. 统计与决策,2011(21):165-167.

[9] 贺本岚. 支持向量机模型在银行客户流失预测中的应用研究[J]. 金融论坛,2014,19(9):70-74.

[10] 马北玲,游达明,胡小清. 基于支持向量机的企业突破性创新识别模型研究[J]. 软科学,2013(1):109-111.

[11] 张水波,康飞,李祥飞. 基于支持向量机的建设工程项目经理胜任力评价[J]. 中国软科学,2013(11):83-90.

Total Profits Prediction of Construction Enterprises Based on Support Vector Machine: Taking Construction Enterprises in Liaoning Province as an Example

WANG Qiufei, TANG Bingjie
(Business School, Shenyang Jianzhu University, Shenyang 110168, China)

Abstract: This paper compares and analyzes the three forecasting methods: exponential smoothing, BP neural network and support vector machines. Taken the total profit forecast of construction enterprises in Liaoning Province as an example, this paper compares and analyzes the actual values and the predicted values by these three methods. The results show that SVM has better short-term forecasting effect, and can better simulate and predict the total profit change of construction enterprises in Liaoning Province, which provides an important method for the accurate prediction of total profit in enterprises.

Key words: total profit; forecast method; support vector machine; construction enterprise