

区域物流需求预测的应用研究

万 励, 李余琪, 吴洁明

(梧州学院 计算机科学系, 广西 梧州 543002)

摘 要: 为了提高区域物流需求精度, 综合考虑区域物流需求线性、周期性和非线性信息, 提出了采用组合模型的预测方法. 通过将各种单一预测模型看作代表不同信息的片段, 通过 BP 神经网络对不同信息进行集成, 充分利用各单项预测方法的有用信息, 从而提高区域物流需求预测精度. 通过上海市的物流需求数据对组合模型进行测试, 实验结果表明, 组合模型很好揭示了上海物流需求的变化规律, 提高了物流需求的预测精度, 为区域物流需求预测提供了一种新的思路和方法.

关键词: 区域物流需求; 线性回归模型; 支持向量机; 神经网络

中图分类号: F252

文献标识码: A

文章编号: 1000-7180(2011)09-0160-05

Application Research on Prediction of Regional Logistics Demand

WAN Li, LI Yu-qi, WU Jie-ming

(Department of Computer Science, Wuzhou University, Wuzhou 543002, China)

Abstract: In order to improve the precision of regional logistics demand, comprehensive consideration of the regional logistics demand linear, periodic and nonlinear information, the paper proposed a composite model based on all sorts of single forecast model, the integration of different information, make full use of every single forecasting method, so as to improve the useful information regional logistics demand forecasting precision. The logistics demand data from Shanghai to combination model test, the experimental results indicate that the combined model is very good reveals the change rule of Shanghai logistics demand, improved logistics demand forecasting precision for the regional logistics demand forecasting, provides a new thinking and method.

Key words: regional logistics demand; MLR; SVM; ANN

1 引言

随着区域经济快速发展, 单个企业物流向区域物流方向发展是一种必然趋势, 区域物流成为区域经济的重要构成要素, 对规划和建立完善的区域物流网络体系十分重要, 然而物流需求预测是最为关键的环节. 准确的物流需求预测能为物流管理部门制定规划和开拓市场提供参考依据^[1].

国内外学者对物流需求预测进行大量的研究, 建立了许多预测模型. 传统预测模型有: 回归分析、时间序列法(指数平滑、移动平滑)、货运强度法、灰色预测方法和马尔可夫链等, 这些方法都是仅仅采

用采用物流需求历史数据进行线性建模, 而现代物流需求与地区经济、物流服务水平、人口数量密切相关, 具有很大的随机性和周期性, 因此传统预测模型不能全面、科学反映区域物流需求的变化规律, 预测精度不理想^[2-4]. 近年来, 神经网络等人工智能预测方法得到了快速发展, 尤其是神经网络具有传统预测模型不具备的高度非线性函数映射能力, 很好的提高了物流需求预测精度, 但是神经网络是基于大样本预测理论的方法, 对于小样本存在收敛速度慢和局部极小等不足. 然而区域物流需求研究是近些年才开始的, 在实际预测过程中, 区域物流需求数据样本非常少, 因此神经网络预测精度与实际要求有

一定的差距^[5]. 支持向量机是一种基于小样本的预测方法,可以很好对区域物流需求进行预测^[6]. 对于同一个区域物流需求问题,不同的预测模型包含各自独立的有用信息,如果将单个预测模型合理组合起来,最大化地利用有用信息,可以提高预测精度. 在进行模型组合时,有两个关键的问题要解决:第一,如何选择合理的单一模型,盲目组合会导致预测精度反而降低. 第二,如何合理确定模型之间的关系,由于单一模型间不是简单的线性关系,因此应该采用非线性模型来确定各模型间的权重^[7].

针对当前单一预测模型无法全面反映区域物流需求的变化规律,文中提出一种组合的区域物流需求模型. 首先采用线性回归模型对区域物流需求的线性变化规律进行预测,采用移动自回归模型对区域物流需求的周期性变化规律进行预测,采用支持向量机对区域物流需求的非线性变化规律进行预测,然后采用 BP 神经网络确定三种预测模型之间关系,最后得到区域物流需求预测结果.

2 区域物流需求组合预测的数学模型

区域物流需求预测是利用区域物流需求历史资料和市场信息,对未来区域物流需求状况进行科学分析、估算和推断,指导调节人们的物流管理活动^[8]. 物流需求与影响物流需求的各种因素密切相关,因素与物流需求之间的关系不是简单的线性关系,而是一种高度非线性关系,使物流需求波动相当大,难用简单明确的模型来描述,但是可以采用抽象的数学表达式来描述:

$$y = f(x_1, x_2, \cdots, x_n) \tag{1}$$

式中, y 表示物流需求量; $x_1, x_2, \cdots, x_n$ 表示影响因素.

从式(1)可知,多个影响因素对物流需求影响程度不同且错综复杂,这样物流需求与影响因素之间是一种多元非线性映射关系,导致区域物流需求具有线性、周期性和非线性,因此表示为:

$$y = f_1(x) + f_2(x) + f_3(x) \tag{2}$$

其中, $f_1(x)$ 表示线性部分, $f_2(x)$ 表示周期部分, $f_3(x)$ 表示非线性部分.

从式(2)可知,由于传统单一预测模型只能反映区域物流需求部分信息,因此预测精度低. 要获得高精度的区域物流需求预测结果,必须对其线性、周期性和非线性变化规律同时预测,因此本文提出采用组合模型对其进行预测,组合模型包括:线性回归模型(MLR)、移动自回归模型(ARMA)、支持向

量机(SVM),即可以表示为:

$$\begin{cases} y_1 = f(x_n) \\ y_2 = f(x_n) \\ y_3 = f(x_n) \end{cases} \tag{3}$$

其中, y_1, y_2, y_3 分别表示三种模型的预测值. 组合模型关键是如何确定单一模型预测结果之间的关系,由于 MLR、ARMA 和 SVM 不是简单的非线性关系,因此本文采用非线性函数映射能力的 BP 神经网络来确定 3 个模型的关系^[9],即:

$$(r_1, r_2, r_3) = f(\text{MLR}, \text{ARMA}, \text{SVM}) \tag{4}$$

这样可以得到区域物流需求的组合模型预测结果为:

$$\hat{y} = f(y_1, y_2, y_3) \tag{5}$$

其中, $\hat{y}$ 为区域物流需求的最终预测结果, $f()$ 表示 BP 神经网络模型.

3 区域物流需求组合预测模型

3.1 组合预测模型的结构图

区域物流需求预测是一个复杂系统,文中采用最优线性预测模型、周期预测模型和非线性预测模型组成的组合模型,其模型结构如图 1 所示.

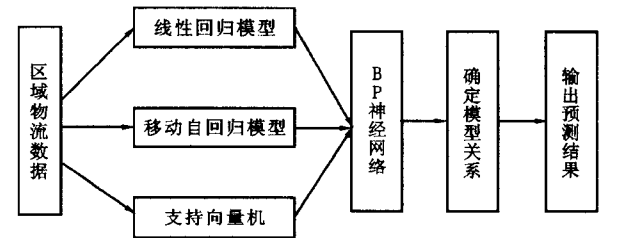


图 1 区域物流需求的组合预测模型

3.2 物流需求的影响因素选择

文中选择一下地区的“货物运输量”作为物流需求量,由于统计数据的限制,本着可操作性原则,选择影响货物运输量影响如表 1 所示.

表 1 物流需求的影响因素

代号	影响因素
x_1	第一产业产值
x_2	第二产业产值
x_3	第三产业产值
x_4	人均消费水平
x_5	区域零售总额
x_6	区域外贸总额

3.3 物流数据归一化处理

物流需求影响因素多,其取值范围各不相同,有

时相差很大,而 MLR、ARMA、SVM 对数据相当的敏感,为了防止由于数据对物流需求预测结果有不利影响,在进行预测之前,对数据进行归一化处理,加快预测模型的学习速度,提高物流需求预测精度,归一化公式为:

$$x'_i = \frac{x_i - \max(x_i)}{\max(x_i) - \min(x_i)} \quad (6)$$

其中, x_i 表示物流需求的影响因素, x'_i 表示归一化后的物流需求量, $\max(x_i)$ 、 $\min(x_i)$ 表示 x_i 的最大值和最小值。

3.4 物流需求的线性部分预测

多元回归分析方法是一种处理多个变量之间相互关系线性数学模型。设获得 n 组物流需求数据,即: $\{x_{i1}, x_{i2}, \dots, x_{ip}, y_i\}$, $i = 1, 2, \dots, n$, 且满足如下关系:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon_i \quad (7)$$

其中, $\beta_0, \beta_1, \dots, \beta_p$ 为模型待估计参数, ϵ_i 是均值为 0 的随机变量, 式(7) 就表示一个多元线性回归模型。多元线性回归模型可以写成矩阵形式:

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \dots & x_{np} \end{bmatrix} \quad (8)$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \quad (9)$$

通过最小二乘法对多元线性回归模型的系数进行估计,并采用 t 检测对其性能进行检验,如果参数合理,就采用该模型对将来的区域物流需求量进行预测,得到物流需求的线性部分预测结果。

3.5 物流需求的周期性部分预测

ARMA 模型是一种描述平稳随机过程的自回归滑动平均模型,是回归模型与滑动平均模型的组合,是一种经典的时间序列预测模型。ARMA 模型的数学表达式为:

$$x(k) = \alpha_1 x(k-1) + \alpha_2 x(k-1) + \dots + \alpha_n x(k-1) + c_0 e(k) - c_1 e(k-1) - c_m e(k-m) \quad (10)$$

其中, α_i 表示待定参数, m, n 表示模型的阶数, $e(t)$ 表示模型残差。

将模型的参数进行简化,ARMA 模型就可以表示为:

$$Y = AX + BX' + CE \quad (11)$$

区域物流需求具有周期性变化规律,因此需要

对 ARMA 模型进行改进,在原模型中加入周期因子项,使 ARMA 模型具有周期性预测功能,因此加入周期因子后 ARMA 模型可以表示为:

$$Y = AX + BX' + CE \quad (12)$$

其中, B 表示周期系数矩阵。

同样,采用最小二乘法对 ARMA 模型的参数进行估计,建立周期性区域物流需求预测模型,并对其周期性变化规律进行预测。

3.6 物流需求的非线性部分预测

支持向量机是一种结构风险最小化原则的人工智能学习方法,具有很强非线性预测能力,克服了神经网络中存在的过拟合、易陷入局部极值等缺点。支持向量机预测是通过一个非线性映射函数 φ , 数据 x_i 映射到高维特征空间 F , 并在这个空间进行线性回归,即

$$f(x) = \omega^T \varphi(x) + b \quad (13)$$

其中, ω 为超平面的权值向量, b 为偏置量。

支持向量机预测为了使训练集上获得的回归模型具有更好的推广能力,不但要考虑经验风险的最小化,同时还要设法降低模型的复杂度,即:

$$\min_{\omega, b, \xi_i, \xi_i^*} = \frac{1}{2} \omega^T \omega + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (14)$$

式中, ξ_i, ξ_i^* 为松弛变量, C 惩罚参数,其值越大,拟合程度越高; ϵ 为回归允许的最大误差,控制支持向量的个数和泛化能力,其值越大,支持向量越少,相应支持向量回归估计函数为:

$$f(x) = \sum_{i=1}^l (a_i - a_i^*) k(x - x_i) + b \quad (15)$$

将区域物流需求数据输入到支持向量机中,支持向量机通过一定的参数优化方法,找到最优参数,利用最优参数对区域物流需求数据进行训练,建立相应的预测模型,然后对将来的区域物流需求进行预测,从而得到区域物流需求的非线性变化规律。

3.7 组合模型的最终预测结果

当得到区域物流需求的线性、周期性和非线性预测结果后,如果将其结果好好的组合在一起,是获得高精度预测结果的关键,由于线性、周期性和非线性预测结果不是一种简单的线性关系,因此文中采用非线性关系映射能力强的 BP 神经网络确定 3 个单一模型间的关系,即将 MLR、ARMA 和 SVM 模型预测结果作为 BP 神经网络的输入,真实物流需求结果作为 BP 神经网络的输出,通过 BP 神经网络的自组织、自学习功能建立一种模型,然后得到组合模型物流需求预测结果。综合上述可知,区域物流需

求组合模型的预测流程如图 2 所示。

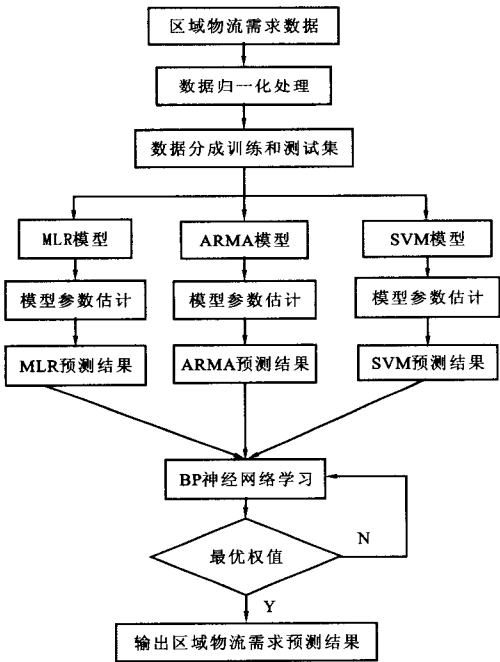


图 2 区域物流需求组合模型的预测流程

4 组合模型在区域物流需求中的应用

为了检验组合模型在区域物流需求预测中的效果,采用上海市物流需求数据作为研究对象,进行验证性测试,物流需求的影响因素和物流需求数据具体如表 2 所示。

表 2 上海物流需求数据及影响因素

年份	x_1	x_2	x_3	x_4	x_5	x_6	y
1978	11	211.05	50.76	54.1	30.26	442	19645
1979	11.39	221.21	53.83	68.28	38.78	512	19613
1980	10.1	236.1	65.69	80.43	45.06	553	20037
1981	10.58	244.34	69.84	88.73	41.5	585	20150
1982	13.31	249.32	74.44	89.8	38.93	576	21153
1983	13.52	255.32	82.97	100.68	41.4	615	21594
1984	17.26	275.37	98.22	123.72	44	726	23121
...	...	...	...	...	...	...	...
2000	83.2	2163.68	2304.27	1722.27	547.1	11546	52206
2001	85.5	2355.53	2509.81	1861.3	608.98	12562	54049
2002	88.24	2564.69	2755.83	2035.21	726.64	14295	56652
2003	90.27	2977.61	2976.3	2220.41	1098.68	14867	61073
2004	96.7	3788.2	3565.3	2454.6	1568	16683	62986

4.1 模型的实现

将表 2 中的数据分成两部分,训练集和测试集,训练集用于对区域物流需求数据建模,测试集用于

对建立的模型性能进行检验.训练集为:1978~1999 年的物流数据,测试集 2000~2004 年物流数据,采用一步预测方法。

首先对表 2 数据进行归一化处理,然后分别采用 MLR、ARMA、SVM 模型对其进行学习建模,建立 3 线性、周期性和非线性预测模型,并采用 3 种模型分别对上海物流需求量进行预测,得到区域物流需求的线性、周期性和非线性预测结果,最后将 3 个模型的结果输入到 BP 神经网络学习建模,并对 2000~2004 年的物流需求数据进行预测,得到组合模型的物流需求预测结果如表 3 所示。

表 3 物流需求预测结果

年份	物流需求实际值	组合模型的预测值
2000	52206	52105
2001	54049	54110
2002	56652	57000
2003	61073	60988
2004	62986	62795

4.2 与单一模型预测结果比较

为了检验组合预测模型与单一预测模型的性能的优劣,分别采用 MLR、ARMA 和 SVM 对表 2 的数据进行预测,数据处理方式与组合模型一样,其预测结果的对比如图 3 所示。

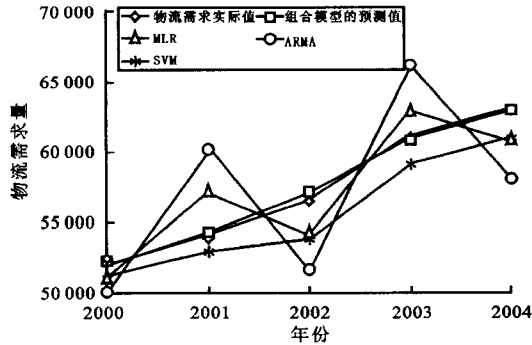


图 3 各种模型的预测结果

从图 3 可知,组合模型的预测值与区域物流需求的实际值相当的接近,而单一预测模型的预测结果与实际相差比较大,说明组合模型提高了预测区域物流需求的预测精度,减少了预测误差,主要是因为组合模型综合了 MLR、ARMA、SVM 的线性、周期性和非线性预测优点,充分挖掘了区域物流需求的信息,保证了组合模型能获得比单一预测模型更好的预测精度,因此组合模型是一种泛化能力,预测误差低的区域物流需求预测模型。

5 结束语

为了提高区域物流需求预测的精度,文中结合线性预测模型、周期性预测模型和非线性预测模型优点,提出一种基于 BP 神经网络的区域物流需求组合预测型.通过对上海区域物流需求数据进行验证性实验,实验结果表明,组合模型很好揭示了区域经物流需求与影响因素间的非线性映射关系,提高了区域物流需求的预测精度.组合预测模型比单一预测模型更适合于非线性特征的复杂系统,可以准确描述区域物流未来发展趋势,为区域物流需求预测提供了一种新的思路和方法.

参考文献:

- [1] Rodrigo A, Hani S. Forecasting freight transportation demand with the space-time multinomial probity model [J]. Transportation Research Part B, 2000;403-418.
- [2] 李莉,张建华,周海燕.物流产业发展与国民经济整体水平提升的相关性分析[J].中国机械工程,2003,14(10):884-887.
- [3] 过秀成,谢实海,胡斌.区域物流需求分析模型及其算法

[J]. 东南大学学报:自然科学版,2001,31(3):1-5.

- [4] 潘英英,宋国喜.基于灰色系统理论的广西物流需求预测[J].广西财经学院学报,2008,21(1):26-30.
- [5] 黄毅,夏国恩.基于支持向量回归机的广西物流需求预测[J].科技管理研究,2011,12(2):142-146.
- [6] 杨海光,夏国恩.基于灰色预测模型的广西物流需求预测[J].中国管理信息化,2011,14(5):39-41.
- [7] 陈以,万梅芳.RBF神经网络在物流系统中的应用[J].计算机仿真,2010,21(4):41-44.
- [8] 后锐,张毕西.基于 MLP 神经网络的区域物流需求预测方法及其应用[J].系统工程理论与实践,2005,12(1):43-47.
- [9] 孙晓玲,王宁,梁艳.应用 BP 神经网络的教学评价模型及仿真[J].计算机仿真,2010,21(11):84-86.

作者简介:

万 励 女,(1972-),硕士,工程师,讲师.研究方向为数据库和计算机应用.

李余琪 男,(1980-),硕士研究生,讲师.研究方向为数据分析预测.

吴洁明 女,(1959-),研究生,副教授.研究方向为计算机软件与理论.

(上接第 159 页)

由表格可以看出,用户一的主机安全性权值 W 为 0.46,用户二的主机安全性权值 W 为 0.38.经比较,用户一主机安全性比用户二的主机安全性低,这验证了用户行为习惯对主机安全性是有影响这一结论.实验证明:该检测方法可行.

4 结束语

文中通过建立拓展攻击树模型,利用贝叶斯算法判断主机安全性的方法是有效可行的,并且该方法可以作为判断未知文件恶意性.当来源地集合越完善时,该方法对主机安全性的检测越精确.

随着来源地集合的不断完善,对拓展攻击树的进一步改造以及对贝叶斯算法的优化将是接下来的

工作的重点.

参考文献:

- [1] 张国力,马军.主机安全技术研究[J].信息安全与技术,2010(8):56-58.
- [2] 夏冰,徐飞,郑秋生.一种主机安全评估系统的研究和设计[J].中原工学院学报,2009,20(5):24-29.
- [3] 刘巍伟,石勇,郭煜,等.一种基于综合行为特征的恶意代码识别方法[J].电子学报,2009,37(4):696-700.
- [4] 张波云,殷建平,蒿敬波,等.基于多重朴素贝叶斯算法的未知病毒检测[J].计算机工程,2006,32(10):18-21.

作者简介:

余林琛 女,(1977-),讲师.研究方向为网络安全、对等计算.