

基于主导因子法的装备维修保障人员调度值预测

单黎黎*, 张宏军, 张睿, 程恺, 王之腾

(解放军理工大学 工程兵工程学院, 南京 210007)

(*通信作者电子邮箱 zaishuiyiren1981@sina.com)

摘要:为实现装备维修保障人员调度数量的准确简单预测,提出一种通用的支持向量机(SVM)输入变量特征的确定方法——主导因子法。该算法在定义了“主导因子”、“驱动因子”、“主动性行为”和“行为载体”等相关术语的基础上,通过“极大关联性”准则和“行为目的”法设置主动性行为预测变量的主导因子,然后根据该主导因子和“驱动因子设置法则”提炼出各驱动因子作为 SVM 输入变量的特征。实际应用及与其他方法比较后表明:将主导因子法确定的各装备维修保障人员调度值驱动因子作为 SVM 输入变量特征对相关值进行预测的平均相对误差低至 0.0109,相对于其他特征确定方法具有更高的预测准确率。

关键词:支持向量机;回归预测;输入向量特征;装备维修保障;人员调度

中图分类号: TP181 **文献标志码:** A

Prediction on dispatching number of equipment maintenance people based on main factor method

SHAN Li-li*, ZHANG Hong-jun, ZHANG Rui, CHENG Kai, WANG Zhi-teng

(Engineering Institute of Engineer Corp, PLA University of Science and Technology, Nanjing Jiangsu 210007, China)

Abstract: In order to forecast the number of equipment maintenance people more easily and validly, a common approach of selecting the features of input vector in Support Vector Machine (SVM) named Main Factor Method (MFM) was proposed. The relevant terms of "main factor", "driving factor", "voluntary action" and "actions' carrier" were defined, based on which the theoretical MFM was constructed. Firstly, the predicting vector's main factor of voluntary actions was setup by "infinitely related principle" and "action purpose" method. Then the driving factors which can be looked as the characteristics of SVM input vector were refined through the selected main factor and "selecting principles of driving factors". The experimental results and comparison with other congeneric methods show that the proposed method can select the more accurate prediction with the value of relative average error 0.0109.

Key words: Support Vector Machine (SVM); regression prediction; input-vector features; equipment maintenance; people dispatching

0 引言

在各中继级维修单位实施巡修前,如能准确地预测出参与保障的人员数量,就能有效地避免多配、少配等现象,从而优化了与人员相关的各种保障资源的配置性能,提高了巡修的整体效能。目前,确定维修人员数量 N_a 的算法主要有直接计算法和分析计算法^[1]。直接计算法是通过统计各维修工作所需的工时数来推算 N_a 的值,用此方法实现对 N_a 的有效预测,需事先准确统计多种参量值,大多数参量值的计量在实际操作中往往存在可行性差、误差大等不足;分析计算法是先通过分析确定各维修任务类别,然后预测出各类任务对应的工时数,最后根据公式求得 N_a ,计算中要用到多种预测值,从而降低了预测的准确度,该方法适合对 N_a 的粗略估算。

支持向量机(Support Vector Machine, SVM)是一种基于统计学习理论的机器学习算法,追求在有限信息的条件下得到最优结果。以往困扰机器学习方法的很多问题,如模型选择

和学习问题、非线性和维数灾难问题、局部极小问题等,均可通过支持向量机在一定程度上得到解决^[2]。目前, SVM 已成功应用于分类和回归领域,通过查阅大量文献资料发现:将 SVM 应用于装备维修保障人员调度值(Dispatching Number of Equipment Maintenance People, DNEMP)预测是一种有益的尝试。

是否合理选取 SVM 输入变量特征 x_u 是决定输出准确与否的前提,一直以来,对于 x_u 的选取多采用遍历或随机尝试法,存在低效、费时、可靠性差和主观性强等不足^[3]。本文提出了一种确定 x_u 的方法——主导因子法,该方法的核心思想是寻找与输入变量的各特征 x_u 间最大限度符合某些规则的因子——主导因子 f_m ,然后根据该主导因子确定 x_u 。本文据此方法去提取 DNEMP 的主导因子和它的各个特征,并通过实例验证所选变量特征的准确性以证明所提方法的合理性。

1 SVM 回归预测思想及 ε -SVR

回归预测是指根据某种影响因素的变动来推测所研究对

收稿日期:2012-02-22;修回日期:2012-04-11。 基金项目:国家自然科学基金资助项目(70971137)。

作者简介:单黎黎(1981-),女,江苏宿迁人,工程师,博士研究生,主要研究方向:装备维修保障信息化;张宏军(1963-),男,江苏姜堰人,教授,博士,主要研究方向:军事运筹;张睿(1977-),男,山东威海人,副教授,博士,主要研究方向:军事仿真、作战模拟;程恺(1983-),男,河南郑州人,博士研究生,主要研究方向:作战模拟、效能评估;王之腾(1982-),男,黑龙江依兰人,博士研究生,主要研究方向:面向对象服务技术。

象的变动方向和程度^[4]。

1.1 SVM 回归预测思想

支持向量机回归思想描述如下: 给定的数据样本集合为 $\{(x_i, y_i) \cdots (x_l, y_l)\}$, 其中 $x_i \in \mathbf{R}^n, y_i \in \mathbf{R}, i = 1, 2, 3, \dots, l$ 。寻找 \mathbf{R}^n 上的函数 $f(x)$, 以使用 $f(x)$ 来推断任一输入 x 所对应的 y 值。为实现对非线性空间函数的预测, SVM 通过核函数将非线性输入空间变换到高维空间来寻找输入变量和输出变量的线性关系。

1.2 ε -SVR

本文选择 ε -SVR 来实现对 DNEMP 的预测。在 SVM 分类算法中引入如式(1)所示的 ε 不敏感损失函数可以构成 ε 型支持向量回归算法 (ε -SVR) 来实现 SVM 的回归预测功能, 如此, 便可将大量估计指示函数(模式识别问题)中得到的结果推广到实函数的估计(回归问题)。 ε -SVR 已经成功应用到大量非线性函数的回归问题^[5-6]。

$$L(y, f(x, \alpha)) = L(|y - f(x, \alpha)|_\varepsilon) \quad (1)$$

其中:

$$|y - f(x, \alpha)|_\varepsilon = \begin{cases} 0, & |y - f(x, \alpha)| \leq \varepsilon \\ |y - f(x, \alpha)| - \varepsilon, & \text{其他} \end{cases} \quad (2)$$

ε -SVR 通过核函数 $K(x)$ 将非线性空间中的输入向量 x 映射到高维特征空间 Z 中, 用 Z 中的线性函数集合 $f(x, \alpha) = (\omega K(x)) + b$ 来预测回归函数 $f(x)$, 这是一个凸二次优化问题, 引入拉格朗日函数后, 该优化问题转化为在式(4)中

(6) 约束下最大化函数:

$$W(\alpha, \alpha^*) = -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon \quad (3)$$

$$\omega = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (4)$$

$$\alpha_i \geq 0, 0 \leq \alpha_i^* \leq C, \alpha_i, \alpha_i^* \in \mathbf{R}; i = 1, 2, \dots, l \quad (5)$$

$$\frac{\partial L}{\partial \omega} = 0, \frac{\partial L}{\partial b} = 0, \frac{\partial L}{\partial \xi_i} = 0, \frac{\partial L}{\partial \xi_i^*} = 0 \quad (6)$$

要求的回归函数 $f(x)$ 可表示为:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (7)$$

可求得:

$$\begin{cases} \varepsilon - y_i + f(x_i) = 0, & \alpha_i \in (0, C) \\ \varepsilon + y_i - f(x_i) = 0, & \alpha_i^* \in (0, C) \end{cases} \quad (8)$$

根据式(8)可求出 b , 选择适当的核函数从而最终可以求出 $f(x)$ 。式(6)中 ξ_i, ξ_i^* 为松弛因子, 式(8)中 ε 是误差边界, C 为正则化参数, $C, \varepsilon \in \mathbf{R}$ 。

2 主导因子法

SVM 输入变量的各分量称为特征, 记为 $x_u (i = 1, 2, \dots, n)$ 。一个输入变量一般有多个特征, 特征是描述变量特性或影响变量值的最关键因素。在利用 SVM 进行预测之前, 首先要确定待测变量的各个特征以据此去采集样本数据, 选择合理的变量特征是实现准确预测的关键。本文提出了一种 x_u 的确定方法——主导因子法(Main Factor Method, MFM)。

2.1 相关定义及主体思想

定义1 主导因子。在用支持向量机实现预测的过程中, 与待预测量数值变化关联最紧密的因素称为主导因子, 记为 f_m 。

f_m 可以是时间、距离、气温等不同量纲的量, 对于一个变量可以有一个或多个主导因子。因主导因子与驱动因子存在极大关联, 其值是否便于采集直接关系到最终确定的驱动因子能否成功应用于实际。因此, 选择主导因子时, 其值的采集便捷性和可行性也是重要的考虑因素^[7]。

定义2 驱动因子。为描述方便, 将 SVM 输入变量的特征称为驱动因子, 记为 x_u 。一个 n 维输入变量的 n 个驱动因子可以具有不同量纲。

设定驱动因子确定准则: 1) 变量的所有驱动因子均应与该变量的各个主导因子极大程度上关联。理想状态下, 满足 $x_{ui} = F_i(f_{m1}, f_{m2}, f_{m3} \cdots f_{mn}, f_{oi})$, F_i 为关联函数, f_{oi} 为影响 x_{ui} 的其他因素。2) 驱动因子值的变化能够直接迅速地引起变量值的变化, 即二者间的变化存在“一致性关联”。3) 驱动因子的模型应尽量简单。4) 驱动因子模型中各中间量数值应便于采集。

主导因子法的主体思想: 在对待预测量所处环境(行为过程、变量关系等)进行详细分析的基础上尽量提取出与该变量关联性最强的因素, 将这些因素设立为主导因子。然后根据驱动因子确定准则逐一确定变量的驱动因子, 各主导因子和所有驱动因子间存在极大关联性是主导因子法的核心准则。

由图1可以看出, 每个驱动因子 x_{ui} 都通过各自对应的关联函数 F_i 与所有主导因子 f_{mi} 相关联, 即 x_{ui} 与 f_{mi} 间存在最大关联性, 该状态是理想条件下的情形, 在实际应用中, 一般遵循“极大关联性”原则, 即尽量找到 x_{ui} 与大部分 f_{mi} 的关联函数。

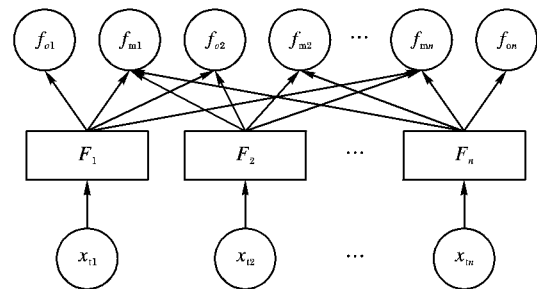


图1 主导因子法示意图

2.2 主动性行为的主导因子确定方法

现实生活中, 有很多行为是由行为主体为实现一定的目的而去执行的, 比如本文所讨论的装备维修保障行动, 显然, 在用 SVM 对这些行为的某些特征值进行预测时, 也需要提取与这些特征值相关的指标。

定义3 主动性行为。由行为执行者为实现某种目的主动去操作的行为称为主动性行为。

经过分析, 我们认为主动性行为是由行为目的 G 、行为载体 C 、行为过程 P 和行为结果 R 组成的关系体, 即 $A = \langle G, C, P, R \rangle$ 。

定义4 行为载体。行为过程中所涉及的实体(如人、物)和非实体(如数量、时间、气温、速度)的资源总和称为行为载体, 记为 C 。

用SVM对主动性行为进行预测一般是对行为载体的相关量的预测。

为实现一个行为目的,一般将行为划分为若干个重要程度不同的子行为 A_i ,各个子行为对应其各自的行为载体集合,记为 C_i ,行为过程 P 可看成是通过某些关系将所有子行为联系起来的机制,子行为之间的函数关系可视为一个子行为过程,记为 P_i 。

由定义1知,主导因子是与待预测量数值变化关联最紧密的因素,其在整个体系中起“主导”作用。毫无疑问,行为目标的实现程度是评判行为效能的关键指标,因此,可从分析行为目的入手来提炼行为载体的主导因子。

设待预测行为载体为 C_x ,主动性行为的主导因子确定步骤如下:1)确定主动性行为的目的 G ;2)根据 G 确定主体子行为 A_m 和次级子行为 A_{si} ;3)由 A_m 确定主体行为载体 C_m ,由 A_{si} 确定次级行为载体 C_{si} ;4)分析影响 G 的主要因素及与 A_m 、 C_m 、 A_{si} 、 C_{si} 间的各种过程关系 P_i 找到与 C_m 和 C_{si} 均关联的影响因素 $f_{mi}(i \in N)$,将其设为 C_x 的主导因子。

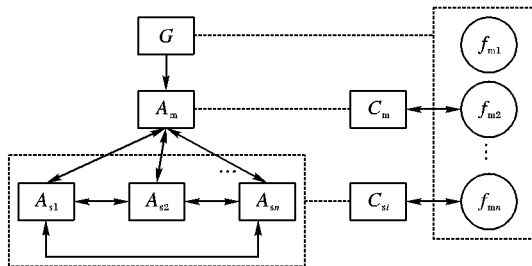


图2 基于主次思想的主动性行为主导因子确定法示意图

图2中:实线箭头代表“关联关系”,设定存在“关联关系”的两个主体之间的关系可用某种函数表示;虚线代表“对应关系”,设定存在“对应关系”的两个主体之间可通过分析、归纳等方式从一个主体得到另外一个主体;虚线框内的每个元素均与框外与其相连的元素存在线段类型所示的关系。由图2可见, C_m 和 C_{si} 与驱动因子 $f_{m1}, f_{m2}, \dots, f_{mn}$ 中的每个元素都存在“关联关系”,即 $C_m = \langle f_{m1}, f_{m2}, \dots, f_{mn} \rangle, C_{si} = \langle f_{m1}, f_{m2}, \dots, f_{mn} \rangle$ 。

3 DNEMP的主导因子和驱动因子的确定

装备维修保障是为了实现对部队装备的维修、保养而实施的一种主动性行为,因而可用第2章所述方法确定其行为载体的主导因子和驱动因子。

3.1 DNEMP主导因子的确定

根据2.2节可得DNEMP主导因子确定步骤如下:

步骤1 确定行为目的 G 。装备维修保障的根本目的是为了在尽可能合理配置保障资源的前提下实现对部队装备的高效率维修和保养,装备保养可视为一种特殊的维修方式,为描述方便我们将“维修”和“保养”统称为“维修”。

步骤2 确定 A_m 、 A_{si} 。由行动目的不难确定装备维修保障行动的主要任务为“维修”,记为 R_w ，“维修人员”是该任务的实施者。为了实现任务的顺利、高效完成，“行政管理人员”需对参与保障的维修人员及其他人员进行合理地编排与管理,其任务 $R_g = \{R_{gi} | i \in N\}$ 。此外,还需由“后勤保障人员”负责装备托运、检修后送、人员护理工作,其任务 $R_h = \{R_{hi} | i \in N\}$ 。由此可得出: A_m 和 A_{si} 分别为: $A_m = \{R_w\}, A_{si} =$

$\{R_g, R_h\}$ 。

步骤3 确定 C_m 和 C_{si} 。因最终是要实现对参与装备维修保障的所有人员数目 N_a 的预测,因此行为载体可设定为“人数”。由步骤2确定的主体子行为和次级子行为可得: $C_m = N_w, C_{si} = \{N_g, N_h\}$, N_w 表示维修人员数, N_g 表示行政管理人员数, N_h 表示后勤人员数。

步骤4 确定主导因子 f_m 。维修保障行动是由各类人员协同完成的,是各子行为间通过各种交互关系实现的。装备维修保障各子行为间的任务关系如图3所示。

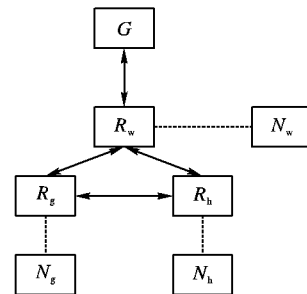


图3 维修保障任务关联图

对图3进行分析可以看出:维修作为整个保障行动的主体任务,维修任务量 N_w 的多少直接影响到主要任务 R_w 的繁重程度,相同工作效率下, N_w 越大, R_w 越繁重,所需调配的维修人员数 N_w 越大, N_w 和 N_g 之间呈正相关关系,即 $N_w = F_{hw}(N_g)$, $F_{hw}(x)$ 为单调递增函数。

N_w 的大小又制约着后勤保障人员数 N_h 和管理人员数 N_g 。虽然 N_h 与 N_g 之间存在相互制约关系,但它们二者均随 N_w 的变化而变化。 N_w 、 N_h 和 N_g 之间的关系用如下公式表示: $N_h = F_{hw}(N_w, N_g), N_g = F_{gh}(N_w, N_h)$ 。

参与保障的总人数 $N_a = N_w + N_h + N_g = F_{nw}(N_w) + F_{hw}(N_w, N_g) + F_{gh}(N_w, N_h)$,可见: N_w 与 N_h 和 N_g 均存在关联,由2.2节可知 N_w 是影响 N_a 的关键因素,从而提炼出“维修任务量 N_w ”作为DNEMP的主导因子。

3.2 DNEMP驱动因子的确定

选出主导因子后,如何确定驱动因子是建立高效SVM预测模型的另外一个关键环节。由2.1节所介绍的驱动因子确定准则可知:DNEMP驱动因子应与维修任务量 N_w 极大程度上关联;DNEMP驱动因子的数值变化直接影响到人员调度值 N_a 的变化,模型简单且数值方便采集。

目前国际上普遍采用的三级维修保障方式中,中继级维修保障部门需定期(一般以季度为一个巡修周期)对其所属的保障点实施巡修以保证各保障点的装备运行稳定,各保障点的装备出现故障,送至中继级实施修理,中继级对季度送修任务进行定性定量统计^[8-10]。理想情况下,本文所建立的各种驱动因子模型中的中间参数值均可从中继级季度统计值中获取。

季度装备送修量 N_s :周期时间内,装备的送修量 N_s 是战地装备损坏情况的直观反映,是巡修中的维修任务量 N_w 的直接度量值,二者间具有极强的关联性, N_s 越大, N_w 越大。它们之间的关系如下:

$$N_s = \sum_{i=1}^n N_{si} = F_{sw}(N_w); i \in N \quad (9)$$

其中: N_{si} 为不同型号装备的送修量,函数 $F_{sw}(x)$ 为单调递增函数。

季度装备修复率 r_a : 一般而言, 投入执行任务的装备越多, 装备可能出现故障的概率越大, 对于送修的装备, 其修复率越大, 返回回战地的数量越大, 巡修任务量越大。因此, 季度装备修复率 r_a 与巡修任务量 W_n 之间也存在正相关性。即 $r_a = F_{an}(W_n)$, $F_{an}(x)$ 为单调递增函数。

$$r_a = n_{rd}/N_s \times 100\% \quad (10)$$

其中: n_{rd} 表示季度内修理好的装备数量, N_s 表示季度装备送修量。

季度装备修复时间加权平均值: 假设待维修装备种类数为 i ($i \in N$), 抽样采集每种待维修装备的数量为 N_i , 第 i 种抽样装备每台的维修时间记为 t_{ij} , 因为不同种类装备具有不同的配备数量和维修复杂度, 因此单台不同装备的平均维修时间 T_{ij} 也不同, 可得 $T_{ij} = \sum_{j=1}^{N_i} t_{ij}/N_i$, 设 M_i 为第 i 种装备的实际维修数量, 则季度装备修复时间的加权平均值为:

$$T_a = \sum_{i=1}^n (T_{ij} \times M_i) / \sum_{i=1}^n M_i \quad (11)$$

可见, T_a 可直观衡量维修技术人员的修理水平, 对于同等维修任务量而言, T_a 越大, 所需维修人数越少, 即 T_a 与主体任务人群数 N_w 呈负相关性, 而 N_w 与 W_n 呈正相关性, 则有 $T_a = F_{tw}(W_n)$, $F_{tw}(x)$ 为单调递减函数。

季度装备报废量 N_d : 对于难以修复或已到退役年限的装备应做报废处理, 并做及时补充^[6]。新装备需要维修的概率相对于老装备要小很多, 因此, 报废量 N_d 越大, 需要补充的新装备越多, 巡修任务量 W_n 也就越小, 二者呈负相关性, 即 $W_n = F_{nd}(N_d)$, $F_{nd}(x)$ 为单调递减函数, 季度装备报废量可根据统计直接获得。

可见, N_s 、 r_a 、 N_d 和 T_a 均与主导因子 W_n 有直接关联, 且模型简单, 所有中间变量值获取便捷, 符合驱动因子设置原则, 其准确性可通过实例得到检验。

4 实例验证

SVM 的输入端为变量 $x = \{N_s, r_a, N_d, T_a\}$, 输出端为即将参与巡修的人员数 N_a 。

根据美军某装备维修保障机构 1996—2003 年巡修资源调度统计数据^[1]并结合 3.2 节所建立的各类驱动因子模型, 经过计算后得如表 1 所示的训练样本和表 2 所示的测试样本。因为 SVM 可实现小样本条件下的预测, 因此只需采集有限的样本数据, 本文采集 16 次巡修资源统计值, 按时间先后, 设定前面的 10 组数据样本用来训练支持向量机, 后面 6 组作为测试样本来检测训练完成的 SVM 学习机, 用平均相对误差 E_{mape} 作为预测精度的参考值。

$$E_{\text{mape}} = \frac{1}{n} \sum_{i=1}^n |(y_i - f(x_i)) / y_i| \quad (12)$$

本测试程序是通过 LibSVM 实现的, 设置 LibSVM 中的“svmtrain”命令如下: svmtrain -s 3 -p 0.21 -c 10 -t 2 -d 1 data_file, 其含义为以径向基核函数, $C = 10$, 损失函数 $\varepsilon = 0.24$, 敏感度 $\sigma = 1$, 构造 ε -SVR 型支持向量机来实现回归预测^[11-12]。

实验过程如下: 首先, 根据表 1 的数据, 构造训练样本点 $(x_1, y_1), (x_2, y_2), \dots, (x_{10}, y_{10})$, $x_i = \{N_s, r_a, N_d, T_a\}$, $y_i = N_a$, x_i 的各特征值为从准备第 i 次巡修时刻起之前一个季度

(3 个月) 的统计数据, 将 x_i 作为输入, y_i 作为输出, 对构造好的 SVM 进行训练, 每一个样本实现对 SVM 的一次训练, 根据训练样本数量, 共进行 10 次训练, 随着训练次数的增多, SVM 的预测性能越趋良好。将经过 10 次训练后的 SVM 称为学习机, 记为 $f(x)$ 。然后, 根据表 2 构造测试样本点 $(x_1, y_1), (x_2, y_2), \dots, (x_6, y_6)$, $x_i = \{N_s, r_a, N_d, T_a\}$, $y_i = N_a$, 将 x_i 作为 $f(x)$ 的输入, 得到的输出为预测值 $f(x_i)$ (如表 3 所示), 将 y_i 、 $f(x_i)$ 代入式 (4) 得平均相对误差值为 0.0109。

为了说明主导因子法的实用性与先进性, 本文将其与常用的特征提取算法——GMDH (Group Method of Data Handling) 进行对比。GMDH 是一种基于自组织和外推原理的数据挖掘算法, 将遗传、变异、选择和进化四个过程进行反复迭代以最终得到最优复杂度模型为止^[13-15]。由 GMDH 所得的输入变量特征为: 季度装备损坏率、季度装备平均修复率、季度装备损坏种类^[16-18]。表 3 给出了通过 GMDH 得到的人员调度预测值和平均相对误差值为 0.0353。

表 1~3 中, n 代表巡修的次序, 例如: 1 到 10 代表按照时间先后进行的 10 次巡修行动。通过比较不难看出, 主导因子法较 GMDH 具有更高的预测准确性。

表 1 训练样本

n	N_s	$r_a/\%$	T_a/h	N_d	N_a
1	3011	91.2109	1.3103	58	207
2	2708	98.7045	1.2607	42	187
3	3426	95.6370	1.3300	48	243
4	3200	94.3623	1.0701	39	220
5	3403	89.2700	0.9803	60	246
6	2978	99.6389	1.1900	79	193
7	3644	97.3620	1.1913	53	256
8	3030	97.5707	0.9921	67	220
9	2973	93.6312	1.5050	63	209
10	2890	88.6320	1.4309	90	199

表 2 测试样本

n	N_s	$r_a/\%$	T_a/h	N_d	N_a
1	3504	87.2479	1.2911	46	233
2	4001	96.4932	1.2302	58	289
3	3783	97.1012	1.1700	81	277
4	3651	95.4343	1.6778	87	283
5	4230	83.6356	1.9833	65	267
6	3987	89.7020	1.3667	70	275

表 3 MFM 和 GMDH 的真实值与预测值对比

n	真实值	MFM		GMDH	
		预测值	相对误差	预测值	相对误差
1	227	233	-0.0264	239	-0.0528
2	291	289	0.0069	280	0.0378
3	270	277	-0.0259	277	-0.0259
4	287	283	0.0139	280	0.0243
5	259	267	-0.0309	269	-0.0386
6	279	275	0.0143	270	0.0322
E_{mape}		0.0109		0.0353	

5 结语

为了对装备维修保障人员调度值 (DNEMP) 进行预测, 本文提出了一种新的确定支持向量机变量特征的方法——主导因子法, 据此方法确定了季度装备送修量 N_s 、季度装备修复

率 r_a 、季度装备修复时间加权平均值和季度装备报废量 N_d 作为 DNEMP 的变量特征并建立了它们的数学模型,通过实例和比较验证了这些特征选择的合理性、模型构建的准确性以及所提方法的先进性,该方法可视为是对目前已有 SVM 算法研究成果在一定程度上的补充,也说明了支持向量机是一种实现 DNEMP 预测的有效工具。本文所提出的理论方法只适用于单目标主动性行为的行为载体特征确定,针对多目标的研究还有待进一步深入。

参考文献:

- [1] 郝杰忠,杨建军,杨若鹏. 装备技术保障运筹分析[M]. 北京: 国防工业出版社, 2006: 23-24.
- [2] VAPNIK V N. The nature of statistical learning theory [M]. 2nd ed. New York: Springer-Verlag, 1999.
- [3] CHALIMOURDA A, SCHÖLKOPF B, SMOLA A J. Experimentally optimal V in support vector regression for different noise and parameter settings [J]. Neural Networks, 2005, 18(2): 205-205.
- [4] SUN JUN, FENG BIN, XU WENBO. Particle swarm optimization with particles having quantum behavior [C]// CEC 2004: Proceedings of the Congress on Evolutionary Computation. Piscataway: IEEE, 2004, 1: 325-331.
- [5] SUYKENS J A K, van GESTEL T, de BRABANTER J, et al. Least squares support vector machines [M]. 3rd ed. Singapore: World Scientific Publishers, 2003.
- [6] 李仁兵, 李艾华, 李亮, 等. 支持向量机在导弹动力系统推力预测中的应用[J]. 系统仿真学报, 2010, 22(4): 934-937.
- [7] CRISTIANINI N, SHAW E. An introduction to support vector machines and other kernel-based learning methods [M]. 2nd ed. Cambridge, UK: Cambridge University Press, 2000.
- [8] 白鹏, 张喜斌, 张斌. 支持向量机理论及工程应用实例 [M]. 4 版. 西安: 西安电子科技大学出版社, 2008: 79-81.
- [9] 杨俊燕, 张优云, 朱永生. ϵ -不敏感损失函数支持向量机分类性能研究 [J]. 西安交通大学学报, 2007, 41(11): 1315-1320.
- [10] 于青, 赵辉. 基于 GA 的 ϵ -支持向量机参数优化研究 [J]. 计算机工程与应用, 2008, 44(15): 139-141.
- [11] 孙祥, 徐流美, 吴清. Matlab 7.0 基础教程 [M]. 3 版. 北京: 清华大学出版社, 2005: 37-39.
- [12] LIBSVM—A library for support vector machines [EB/OL]. [2010-08-20]. <http://www.csie.ntu.edu.tw/~cjlin/>
- [13] IVAKHNENKO A G. Sorting methods for modeling and clustering (survey of the GMDH papers for the years 1983-1988) the present stage of GMDH development [J]. Soviet Journal of Automation and Information Sciences, 1988, 21(4): 1-13.
- [14] KIM D, SEO S-J, PARK G-T. Hybrid GMDH-type modeling for nonlinear systems: Synergism to intelligent identification [J]. Advances in Engineering Software, 2009, 40: 1087-1094.
- [15] 朱帮助, 魏一鸣. 基于 GMDH-PSO-LSSVM 的国际碳市场价格预测 [J]. 系统工程理论与实践, 2011, 31(12): 2264-2271.
- [16] 邹昊飞, 夏国平, 杨方廷. 基于两阶段优化算法的神经网络预测模型 [J]. 管理科学学报, 2006, 9(5): 28-35.
- [17] DEJONG K F, MÜLLER J-A. Self-organizing data mining [J]. Systems Analysis Modeling Simulation, 2003, 43(2): 231-240.
- [18] HUANG C-L, WANG C-J. A GA-based feature selection and parameters optimization for support vector machines [J]. Expert Systems with Applications, 2006, 31(2): 231-240.
- [19] 25(9): 1965-1969.
- [4] WU XIAOYUAN, BOLIVAR A. Keyword extraction for contextual advertisement [C]// WWW'08: Proceedings of the 17th International Conference on World Wide Web. New York: ACM, 2008: 1195-1196.
- [5] 刘远超, 王晓龙, 刘秉权, 等. 信息检索中的聚类分析技术 [J]. 电子与信息学报, 2006, 28(4): 606-609.
- [6] 陈竹敏. 面向垂直搜索引擎的主题爬行技术研究 [D]. 济南, 山东大学, 2008.
- [7] 李晓明, 闫宏飞, 王继民. 搜索引擎——原理、技术与系统 [M]. 北京: 科学出版社, 2006: 98-103.
- [8] 韩客松, 王永成, 滕伟. Web 页面中文文本主题的自动提取研究 [J]. 情报学报, 2001, 20(2): 217-222.
- [9] 任克强, 赵光甫, 张国萍. 基于带权语言网络的网页关键词抽取 [J]. 计算机工程与应用, 2008, 44(8): 155-157.
- [10] BUYUKKOKTEN O, GARCIA-MOLINA H, PAEPCKE A. Seeing the whole in parts: text summarization for Web browsing on handheld devices [C]// WWW'01: Proceedings of the 10th International Conference on World Wide Web. New York: ACM, 2001: 652-662.
- [11] 王琦, 唐世渭, 杨冬青, 等. 基于 DOM 的网页主题信息自动提取 [J]. 计算机研究与发展, 2004, 41(10): 1786-1791.
- [12] CAI DENG, YU SHIPENG, WEN JI-RONG, et al. VIPS: a vision-based page segmentation algorithm, MSR-TR-2003-79 [R]. Redmond: Microsoft Research Corporation, 2003.

(上接第 2363 页)

4 结语

网页的关键词自动抽取是聚焦爬虫研究中非常重要的问题。主题型网页的关键词自动抽取已经有了较多的研究,但是,导航型网页的关键词自动抽取问题研究还不多,仍是一个技术难点。我们注意到,网页作者往往会利用各种视觉强化手段,把一些与主题相关的词汇突出地显示出来。借鉴人类善于发现图像中“奇异点”的原理,将 Web 网页表示为由普通区域和关键区域构成的平面图,提出了一个新的视觉表示模型 PIX-PAGE。通过有效的视觉量化方法,准确地表达并强化了与网页主题关键词相关的特殊区域的视觉“奇异性”。PIX-PAGE 的视觉量化方法和遵循“强者愈强、弱者愈弱”的重要性传递规则,有效地提升了模型的抗主题噪声的能力。因此,在此基础上提出的关键词自动抽取算法 P-KEA 的性能得到了较大的提高。

参考文献:

- [1] CHAKRABARTI S, van den BERG M, DOM B. Focused crawling: a new approach to topic-specific Web resource discovery [J]. Computer Networks, 1999, 31(11-16): 1623-1640.
- [2] CHAU M, CHEN H. Incorporating Web analysis into neural networks: an example in Hopfield net searching [J]. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 2007, 37(3): 352-358.
- [3] 周立柱, 林玲. 聚焦爬虫技术研究综述 [J]. 计算机应用, 2005,