

文章编号:1005-0523(2018)03-0001-07

轨道不平顺预测随机模型的 SVM-MC 求解方法

许玉德¹,刘一鸣¹,沈坚锋²

(1.同济大学道路与交通工程教育部重点实验室,上海 201804; 2.中交投资有限公司,北京 100029)

摘要:实现铁路轨道科学管理的前提是对轨道几何不平顺的发展趋势进行有效预测,预测模型从确定性向随机性模型转变,其重点是如何进行模型的求解。论文对轨道高低不平顺的预测随机模型建立了一种支持向量机—蒙特卡洛(SVM-MC)两阶段求解方法,第一阶段利用 ϵ -SVM 算法确定属于小样本集的模式参数,第二阶段运用蒙特卡洛模拟对随机过程进行仿真,得到高低不平顺标准差的预测值。与以往的轨道不平顺预测方法相比,所建立的两阶段求解方法解决了预测中小样本、非线性的问题,且预测精度在计算机容量和速度足够时可以得到保证。在沪昆有砟线路的应用表明,所提出的随机预测方法及求解算法,预测效果良好,平均相对误差为 4.63%,可满足现场的工程应用,为养护维修计划决策提供技术支持。

关键词:轨道几何不平顺;恶化预测随机模型;支持向量机;蒙特卡洛模拟

中图分类号:U216

文献标志码:A

铁路轨道是行车的基础,支承并引导列车按照允许的速度安全有序运行,在列车荷载、外部环境等众多因素的作用下,轨道几何状态发生变化,形成轨道几何不平顺^[1],影响行车安全。为实现列车的高效运行、减少轮轨的相互作用,需要对轨道几何形位进行控制,控制方法有“事后修”和“预防修”,实现“预防修”的前提是可以准确预测轨道几何不平顺的发展状态。

从 20 世纪 80 年代起,国内外学者对轨道高低不平顺恶化预测模型展开大量的研究^[2-3],其特点是一个固定的表达式来表示几何状态随时间的变化。由于轨道系统的复杂性和特殊性,即便是在相同的轨道结构和运输条件下,运用相同的养护作业手段,轨道几何状态的变化也未必相同,最终导致了现有确定性模型在进行预测时精度不足。2010 年前后起,国外研究重点开始转向以随机性为主的恶化模型^[3-5],减少误差、提高精度。

据现场养护工作的调查统计^[2],大部分线路的高低不平顺和捣固作业直接相关,而捣固作业周期一般在 1~1.5 年之间,因此不平顺恶化随机预测模型中的待定参数属于小样本集,需要采用适用于小样本的预测方法。现有研究中对于小样本事件的预测方法,主要包括:

1) 多元线性回归、偏最小二乘回归:这类方法会由于各变量间存在多重相关性,或主成分分析时损失有效变量信息而影响到模型的鲁棒性。

2) BP 神经网络方法、灰色理论:前者需要较多的历史数据,且模型的收敛速度比较慢,受样本容量影响较大,后者可以预测较少的样本数据且预测准确度较高,但对波动性较大的时间序列样本的预测结果往往较差。

收稿日期:2017-12-21

基金项目:国家自然科学基金项目(51678445)

作者简介:许玉德(1965—),男,教授,博士,博士生导师,研究方向为轨道管理,轨道养修技术。

3) Bayes 方法、统计学习理论:前者对先验分布的准确度要求颇高,这关系到预测的可靠性,后者在被提出之初就是为了解决非线性、小样本高维数据的识别问题,具有较强的泛化能力,而其所引出的支持向量机(Support Vector Machine, SVM) [6-7]已被广泛应用于各个高科技领域。

本文选取支持向量机 SVM 作为小样本参数的预测算法,并运用蒙特卡洛模拟[8-9]对恶化阶段进行随机过程仿真,得到轨道几何不平顺标准差的预测值。

1 不平顺恶化随机模型

现有国内外研究[10-11]发现,在相邻的两次养护维修工作之间,高低不平顺恶化随机模型采用非线性指数函数表述是合适的,通过对沪昆线 2010 年至 2015 年轨道检测车实测数据分析,轨道不平顺随时间呈非线性恶化的规律符合如下的指数预测模型[4-5]。

计划区段 i 在某开始时刻 t_0 的轨道不平顺标准差为 σ_0 ,在不进行养护维修作业的情况下,经过 t 时间后在 T 时刻的轨道不平顺标准差 $\sigma_i(T)$

$$\sigma_i(T) = \sigma_0 \cdot e^{k_i t} + \varepsilon(t) \quad (1)$$

式中: σ_0/mm 为计划区段 i 在时间 $T=0$ 的轨道不平顺标准差; $\sigma_i(T)/\text{mm}$ 为计划区段 i 在时间 T 的轨道不平顺标准差; k_i 为服从对数正态分布的待定随机变量参数[4-5]; $\varepsilon(t)/\text{mm}$ 为服从正态分布的随机噪声[4-5]。

式(1)中, k_i 和 $\varepsilon(t)$ 为模型的随机变量,其中 k_i 服从对数正态分布,如式(2)所示; $\varepsilon(t)$ 服从正态分布,其随大机作业次数 n 的变化不明显,可按照区段长期检测数据取值

$$k_i \sim LN(\mu_{k_i}(n), \sigma_{k_i}^2(n)) \quad (2)$$

式中: n 为大型养路机械养护维修作业的次数(以大机捣固作业为主)。

可见,随机恶化模型中,首先要确定随机变量 k_i 的概率分布参数 μ_{k_i} 和 $\sigma_{k_i}^2$ 与大机养护维修作业次数 n 之间的相互关系,线路捣固作业周期较长,养护维修次数 n 属于典型的小样本数据集,因此,对应的统计参数 μ_{k_i} 和 $\sigma_{k_i}^2$ 也属于小样本数据集。

2 轨道不平顺随机预测求解方法

对轨道不平顺标准差的随机预测模型求解分为两个阶段进行,第一阶段是基于支持向量机理论预测下一恶化阶段时模型中随机变量 k_i 所服从概率分布的参数,第二阶段运用蒙特卡洛模拟对下一恶化阶段进行随机过程仿真,得到几何不平顺标准差的预测值。

2.1 基于支持向量机的随机变量参数预测

支持向量机的拓展算法 ε -SVR[12]在解决小样本、非线性及高维模式识别中具有优势,本文选其作为随机变量参数的预测算法。以某个长 1 km 的区段 i 为例,介绍应用支持向量机进行参数预测的 6 个主要步骤。

1) 实测数据处理。国内一般用 200 m 单元区段内的几何不平顺标准差来评价轨道几何状态 [13],故按 200 m 长度将其划分为 5 个单元区段。第 1 次恶化阶段中,通过检测车实测,获得区段 i 内 5 个单元区段的轨道不平顺标准差数据,按照上述不平顺恶化随机模型,用最小二乘法进行数据拟合,得到 5 个单元区段各自的模型参数 k , 5 个 k 的均值和方差就是 $\mu_{k_i}(1), \sigma_{k_i}^2(1)$;同理,对于后续历次恶化阶段进行分析,可得到参数 μ_{k_i} 和 $\sigma_{k_i}^2$ 的样本集,如下式(3)、式(4)所示

$$\mu_{k_i} = (\mu_{k_i}(1), \mu_{k_i}(2), \dots, \mu_{k_i}(n)) \quad (3)$$

$$\sigma_{k_i}^2 = (\sigma_{k_i}^2(1), \sigma_{k_i}^2(2), \dots, \sigma_{k_i}^2(n)) \quad (4)$$

2) 数据规范化。数据的规范化是利用函数变换,将数据按比例缩放,把原始数据映射到某个特定区间内。合理的数据规范化可提高回归效果且大幅加快仿真速度。

3) 核函数选择。常见的核函数有线性核函数、多项式核函数、高斯径向基核函数、Sigmoid 核函数。通过计算各核函数下所得回归数据和实测数据间的误差,选择误差最小的高斯径向基核函数作为模型使用的核

函数。

4) 参数寻优。采用遗传算法对模型参数 c, p 和高斯径向基核函数参数 g 进行确定。 c 是 ε -SVM 函数的惩罚系数; p 表示函数 ε -SVM 函数的 ε 值; g 决定了高斯径向基核函数的幅度, 反映局部小领域内样本的差异。

5) 样本集训练。将前 n 次恶化阶段中参数 μ_{ki} 和 σ_{ki}^2 的样本集作为训练集样本, 并用参数寻优得到的最优 c, p 和 g 参数来训练模型。

6) 测试集预测。用训练好的模型预测测试集数据, 从而得到 $n+1$ 次恶化阶段时随机变量的概率分布参数 $\mu_{ki}(n+1)$ 和 $\sigma_{ki}^2(n+1)$ 。

支持向量机进行参数预测的流程示意图如图 1 所示。

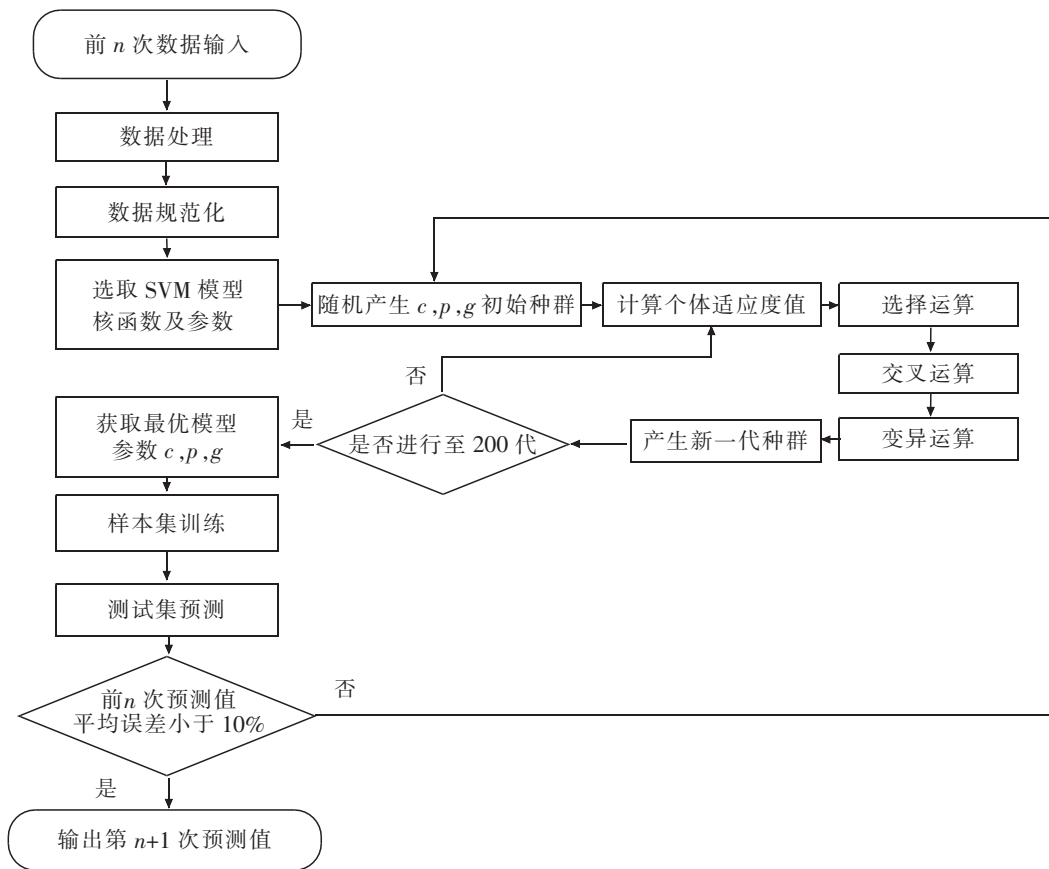


图 1 参数预测流程示意图

Fig.1 Parameter prediction flow chart

2.2 轨道不平顺恶化阶段的蒙特卡洛模拟

恶化随机模型中的参数实际上是服从概率分布的随机变量, 运用蒙特卡洛模拟(monte carlo simulation)对下一恶化阶段进行随机过程仿真, 得到轨道不平顺标准差的预测值。蒙特卡洛模拟主要步骤包括: 构造一个随机模型, 随机产生样本, 对每个随机变量进行抽样, 求出问题的随机近似解。

在进行模拟时, 设置模拟结束时间 T_{end} , 仿真步长 T_{sim} , 模拟次数 M , 不平顺恶化发展的阈值 σ_{th} , 首先输入前一阶段基于支持向量机预测得到的随机变量 k_i 和 $\varepsilon(t)$ 所服从的概率分布参数 $\mu_{ki}(n+1), \sigma_{ki}^2(n+1), \sigma_{\varepsilon}^2(n+1)$, 输入不平顺初值 σ_0 , 然后每个随机变量都随机产生所需的样本集, 根据所建立的随机模型进行仿真计算, 同时对模拟时间是否达到结束时间, 或仿真计算得到的不平顺值是否超过阈值进行判断, 直至仿真终止, 并输出模拟计算得到的随机近似解。在计算机上进行蒙特卡洛模拟的流程示意图如图 2 所示。

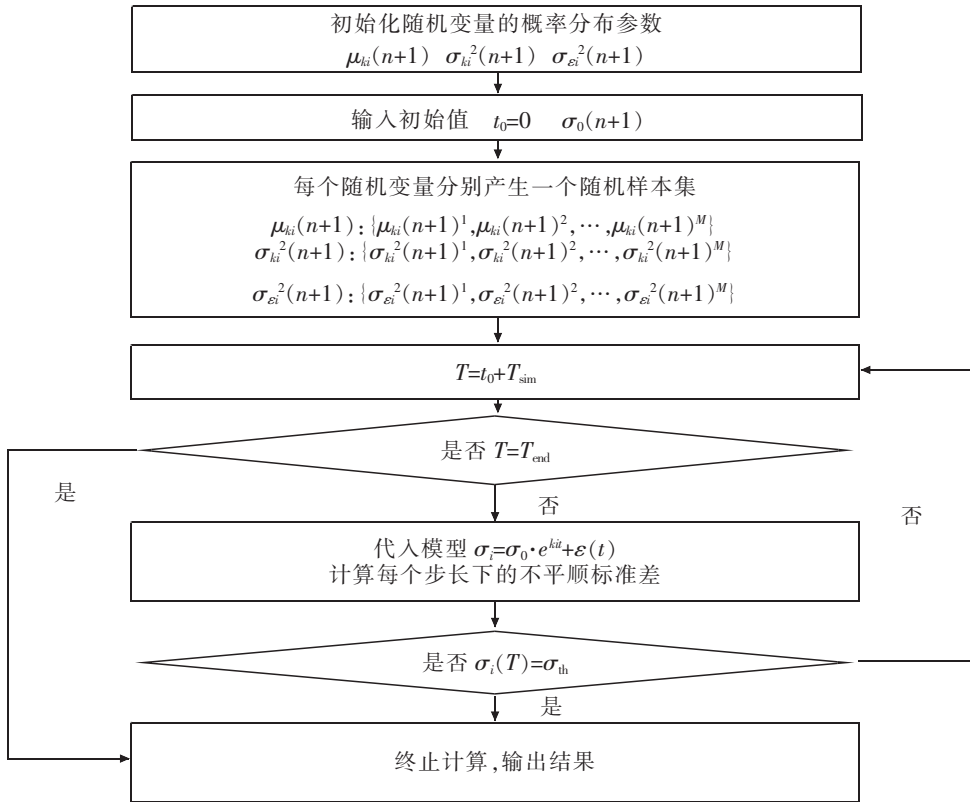


图2 蒙特卡洛模拟流程图

Fig.2 Monte Carlo simulation flow chart

3 案例分析

沪昆线东起上海,西至昆明,全长 2 690 km,是一条双线电气化铁路。以沪昆线 K487+200~K488+200 圆曲线区段为例,对其轨道高低不平顺进行预测。该区段于 2009 年 3 月 6 日至 2015 年 5 月 17 日共采集了 82 次轨检车数据,在所选时间范围内进行了 6 次的大型养路机械捣固作业,对该区段的第 7 次恶化阶段进行轨道高低不平顺标准差预测。

区段按 200 m 划分为若干单元区段,计算各单元区段的标准差,图 3 是前 6 次恶化阶段中各单元区段高低不平顺标准差随时间恶化的序列图。将区段内各单元区段高低不平顺标准差的平均值作为该区段的高低不平顺标准差数据。

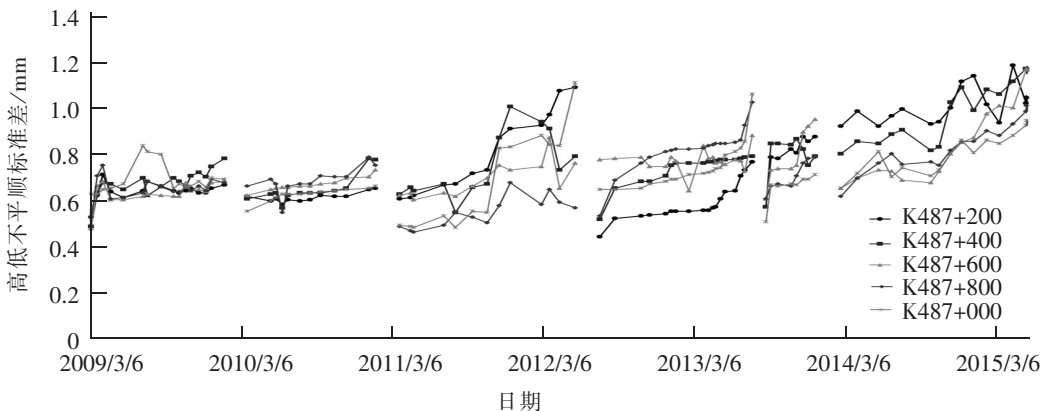


图3 区段高低不平顺标准差恶化趋势

Fig.3 Deterioration of standard deviation of longitudinal irregularity in the section

3.1 随机变量参数确定

利用公式(1)所示指数函数对图3的实测高低不平顺标准差数据进行拟合,获得前6次恶化阶段各自模型随机变量 k 的平均值 $\mu_{ki}=(\mu_{ki}(1),\mu_{ki}(2),\dots,\mu_{ki}(6))$ 和标准差 $\sigma_{ki}^2=(\sigma_{ki}^2(1),\sigma_{ki}^2(2),\dots,\sigma_{ki}^2(6))$,利用 ε -SVR算法预测得到第7次恶化阶段模型随机变量 k 的平均值 $\mu_{ki}(7)$ 和标准差 $\sigma_{ki}^2(7)$ 。 μ_{ki} 与 σ_{ki}^2 的实测值与预测值如图4,图5。 μ_{ki} 的预测值与实测值相比,误差最大的是第3阶段标准差预测的11.50%,平均误差为5.57%; σ_{ki}^2 的预测值与实测值相比,误差最大的是第2阶段标准差预测的17.73%,平均误差为9.73%。

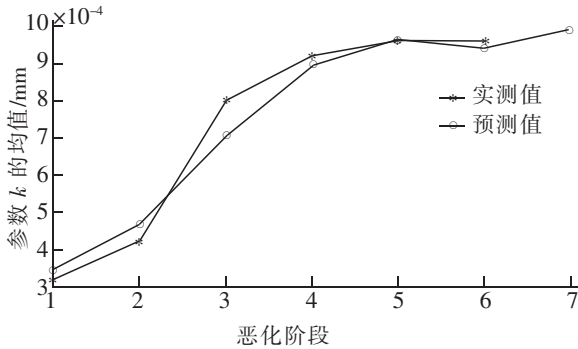


图4 μ_{ki} 实测值和预测值

Fig.4 Measured value and predicted value of μ_{ki}

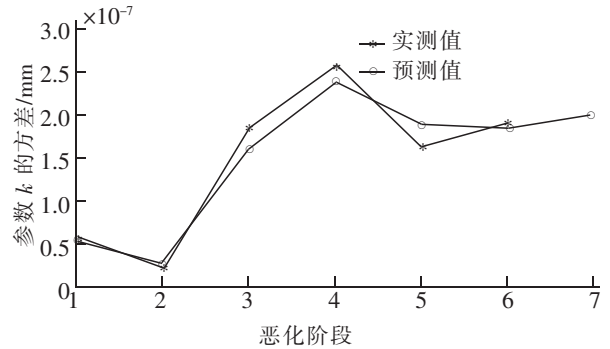


图5 σ_{ki}^2 实测值和预测值

Fig.5 Measured value and predicted value of σ_{ki}^2

3.2 轨道不平顺恶化阶段模拟

确定了恶化模型中随机变量的参数之后,对轨道不平顺恶化阶段进行蒙特卡洛模拟:

模拟结束时间 $T_{end}=360$ 天,即从下一次恶化阶段开始,仿真模拟一年;仿真步长 $T_{sim}=15$ 天,与现场轨检车的检测周期相一致;模拟次数 $M=100000$;不平顺初值 $\sigma_0(7)$ 取各单元区段的初始高低不平顺标准差的均值;随机噪声 $\varepsilon(t)$ 所服从正态分布的参数 $\sigma_\varepsilon^2(7)$,取前6次恶化周期内 σ_ε^2 的均值;对于阈值 σ_{th} ,因为恶化随机模型不涉及何时进行养修的问题,故阈值 σ_{th} 不设定上限。

比较2015年6月29日至2016年1月16日期间高低不平顺标准差的预测值与实测值如图6所示。

由图6可以看到,区段上高低不平顺标准差的恶化发展总体发展较为平稳,在年计划制定的时间跨度内只增加了0.676 mm,仿真模拟结束时的高低不平顺标准差恶化量并未超过1.6 mm的限值,11次实测高低不平顺标准差与预测值的差距均不大,相对误差最大是2015年10月29日为10.6%,相对误差最小是2015年11月16日为0.19%,平均相对误差为4.63%。

在恶化随机模型中,除了参数 k 外,初值 σ_0 的选择也会对最后的预测精度产生影响。以下表1所列的就是区段在训练模型其余参数都一致的情况下,所选初值不同对预测误差的影响。表中前5个初值分别是1 km计划区段内5个单元区段上的高低不平顺标准差,第6个初值是这5个单元区段不平顺标准差的平均值。

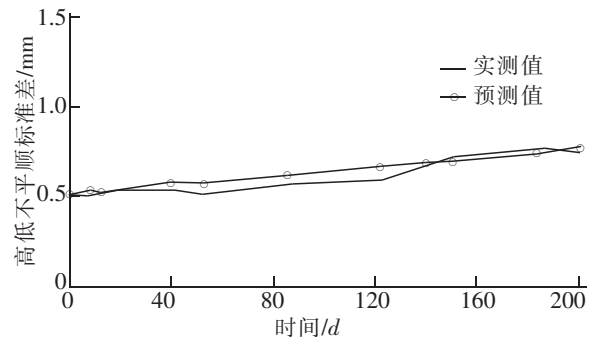


图6 区段第7次恶化阶段高低不平顺标准差预测曲线与部分实测值

Fig.6 Prediction curve and measured value of standard deviation of longitudinal irregularity in the seventh deterioration stage in the section

表 1 模型初值选择对预测相对误差的影响
Tab.1 Influence of initial value on prediction relative error

初值/mm	0.445	0.46	0.54	0.555	0.575	0.515
平均误差/%	12.47	10.36	10.03	12.28	15.33	4.63

由表 1 可以看到,当模型的初值选择 1 km 计划区段内的平均值时,其恶化模型的预测相对误差都是优于其余的初值选择方式的。因此,选择将 5 个单元区段上高低不平顺标准差的平均值作为该计划区段的代表值是合适的,该方法不会对整个恶化随机预测模型造成很大的预测误差。

4 结论

轨道不平顺恶化模型的建立是制定铁路轨道养护计划的基础,对于相邻两次养修之间单元区段上轨道不平顺的恶化发展规律,宜选用非线性函数进行描述,建立支持向量机-蒙特卡洛模拟(SVM-MC)两阶段方法实现了恶化预测随机模型:

1) 模型中随机变量 k_i 的概率分布参数和大机养修作业次数 n 密切相关,因大机养修作业周期长,故随机变量 k_i 的概率分布参数属于小样本数据集,首先运用支持向量机理论进行参数 k_i 的预测问题;在对参数 k_i 的预测过程中,参数寻优过程是训练预测模型最重要的部分,利用遗传算法对模型及高斯径向基核函数作中参数 c, p 和 g 进行寻优操作。在完成了参数 k_i 的预测后,又利用蒙特卡洛模拟方法,通过构造随机样本对轨道不平顺恶化发展的趋势进行仿真并模拟计算得到不平顺恶化量的随机近似解。

2) 目前国内外对轨道不平顺的预测多采用确定性模型,忽略了轨道结构自身特殊性带来的预测差异,且当预测样本数量很小时无法保证预测精度。论文建立的支持向量机-蒙特卡洛模拟(SVM-MC)两阶段方法解决了这些问题。支持向量机方法的使用解决了轨道不平顺预测中小样本、非线性的问题,而蒙特卡洛模拟利用随机数进行统计试验,可以较好的模拟不平顺的随机恶化过程,蒙特卡洛模拟的误差属于概率误差,与当计算机的计算速度和容量够高时,蒙特卡洛模拟的精度可以得到保证。

3) 沪昆线案例表明:所建的轨道不平顺恶化随在恶化随机模型中,除了参数 k 外,初值 σ_0 的选择也会对最后的预测精度产生影响。以下表 1 所列的就是区段在训练模型其余参数都一致的条件下,所选初值不同对预测误差的影响。表中前五个初值分别是 1 km 计划区段内 5 个单元区段上的高低不平顺标准差,第 6 个初值是这 5 个单元区段不平顺标准差的平均值。预测模型预测效果良好,平均相对误差为 4.63%,且将计划区段中各单元区段不平顺的平均值作为该计划区段不平顺的代表值是合适的。

参考文献:

- [1] 杨震,王效堂,练松良,等. 合武客专无砟轨道曲线段不平顺谱分析[J]. 华东交通大学学报,2010,27(5):11-16.
- [2] 佐藤吉彦,徐涌,李德浚. 新轨道力学[M]. 北京:中国铁道出版社,2001:98-194.
- [3] 三和,雅史,内田,雅夫. 轨道状态推移モデルの設定と轨道保守施策決定法(特集 轨道技術)[J]. Rtri Report,1997,11:7-12.
- [4] QUIROGA L,SCHNIEDER E. Modelling of high speed railroad geometry ageing as a discrete-continuous process[C]. Proceedings of the stochastic motle Ling techniques cmd daita analysis iuternatiand Conference, Crete, Chani 2010:655-666.
- [5] QUIROGA L,SCHNIEDER E. Heuristic forecasting of geometry deterioration of high speed railway tracks[J]. 2011, 6928:609-616.

- [6] CHOLETTE M E, BORGESANI P, GIALLEONARDO E D, et al. Using support vector machines for the computationally efficient identification of acceptable design parameters in computer-aided engineering applications[J]. *Expert Systems with Applications*, 2017, 81: 39–52.
- [7] 袁义生, 钟青峰. 基于支持向量机的电力电子电路损耗预测[J]. *华东交通大学学报*, 2017, 34(4): 122–128.
- [8] ZHANG D, HU H, ROBERTS C. Rail maintenance analysis using Petri nets[J]. *Structure & Infrastructure Engineering*, 2017: 1–11.
- [9] 赵振宇, 游维扬, 吕乾雷. 基于遗传算法和蒙特卡洛模拟的并行工程设计工序优化[J]. *土木工程学报*, 2009(2): 139–144.
- [10] 韩晋, 杨岳, 陈峰, 等. 基于非等时距加权灰色模型与神经网络的轨道不平顺预测[J]. *铁道学报*, 2014(1): 81–87.
- [11] CHANG H, LIU R, LI Q. A multi-stage linear prediction model for the irregularity of the longitudinal level over unit railway sections[J]. *Computers in Railways XII*, 2010, 114: 641–650.
- [12] ANDREW A M. An introduction to support vector machines and other kernel based learning methods[J]. *Kybernetes*, 2001, 32(1): 1–28.
- [13] 徐伟昌, 仲春艳, 许玉德, 等. 高速铁路无砟轨道几何不平顺区段管理长度研究[J]. *华东交通大学学报*, 2016, 33(2): 9–14.

SVM-MC Method for Solving Stochastic Model of Track Irregularity Prediction

Xu Yude¹, Liu Yiming¹, Shen Jianfeng²

(1. Key Laboratory of Road and Traffic Engineering of the Ministry of Education, Tongji University, Shanghai 201804, China;

2. CCCC Investment Company Ltd., Beijing 100029, China)

Abstract: The precondition of realizing the scientific management of railway track is to predict the development trend of track geometric irregularity. The prediction model changes from certainty model to randomness model, which puts emphasis on the solution method for this model. In this paper, a two-stage method of support vector machine-Monte Carlo simulation (SVM-MC) for solving the prediction stochastic model of track irregularity was established. In the first stage, the ε -SVM algorithm was used to determine the model parameters which belong to small sample set, and in the second stage, the Monte Carlo simulation was used to simulate the stochastic process, and the standard deviation of the height irregularity was obtained. Compared with the previous method of track irregularity prediction, the established two-stage solution method solves the problems of small samples and non-linear prediction, and the prediction accuracy can be guaranteed when the computer capacity and speed are sufficient. A case in Shanghai and Kunming ballasted line shows that the proposed random prediction method and algorithm have good prediction results, and the average relative error is 4.63%. This method can meet the engineering application in the field and provide technical support for the maintenance plan decision.

Key words: track geometric irregularity; prediction stochastic model of deterioration; support vector machine; Monte Carlo simulation