

基于随机森林的虚拟路谱准确度 智能主观评价法

邱 宇¹, 李文魁¹, 李 冕^{2,3}

(1. 上海汽车集团股份有限公司技术中心试验认证部, 上海 201804)

(2. 上海交通大学密西根学院, 上海 200240)

(3. 上海交通大学电子信息与电气工程学院, 上海 200240)

[摘要] 针对当前虚拟路谱准确度评价方法中的局限性, 利用随机森林建立了一个虚拟路谱的智能主观评分模型, 以路谱的客观统计值作为模型输入, 以主观评价的分值作为输出。在建模过程中全面考虑信号的多方面特征, 利用特征重要度筛选出最重要的 6 个客观统计值作为模型输入; 通过模型训练固化多位专家的专业评分经验, 避免每次进行主观评分时由人为因素导致的结果波动。工程应用显示, 该模型的精度高、泛化能力强, 是对虚拟路谱评价的一个全面、高效、准确的智能工具。

[关键词] 路谱采集, 机器学习, 随机森林, 时域信号, 主观评价

[中图分类号] U467.4+97 **[文献标志码]** A **[文章编号]** 1672-1292(2022)04-0045-10

Intelligent Subjective Evaluation of Virtual Road Load Data Accuracy Based on Random Forest

Qiu Yu¹, Li Wenkui¹, Li Mian^{2,3}

(1. Test & Validation Department, SAIC Motor Technical Center, Shanghai 201804, China)

(2. UM-SJTU Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China)

(3. School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China)

Abstract: Considering the limitations of current methods for virtual road load data evaluation, an intelligent subjective evaluation model of virtual road load data signals based on random forest is proposed. The objective statistical parameters of time history signals are taken as the inputs and the scores of subjective evaluations from experts as the output. The most important six objective statistical parameters are selected through feature importance ranking, which takes various features of the time history signals into a comprehensive consideration. The professional evaluation experiences from many experts are solidified into the model by training, and the fluctuations caused by human factors is avoided. The intelligent subjective evaluation model has high accuracy and good generalization, and it is proven a comprehensive, efficient and accurate intelligent tool for virtual time history signal evaluation for real-world applications.

Key words: road load data, machine learning, random forest, time series signals, subjective evaluation

时域信号描述的是数据随时间的变化关系, 在很多领域都有广泛应用。在汽车的开发过程中, 最常用的时域信号就是道路载荷谱以及由道路载荷谱衍生的台架载荷谱、仿真载荷谱等, 所涉及的数据是诸如力、应变、加速度等物理量^[1]。道路载荷谱简称路谱, 是由试验车辆进行实际测量所获得的真实的时域信号。在实际的工程开发中, 通常还会产生一些虚拟的时域信号: 利用路谱作为目标信号进行道路模拟多通道试验^[2]产生的迭代载荷谱; 利用虚拟试验场结合多体动力学模型产生的虚拟路谱^[3]; 利用数据模型和机器学习算法所预测出的虚拟时域信号等。这些虚拟时域信号在应用之前, 均需进行一系列的评价, 满足相关要求后才能使用。在达到迭代精度之后, 王灵龙等^[2]从最大值、最小值、均方根值 3 个方面对台架的

响应信号和实测路谱进行了对比. 赵强等^[4]则从穿级计数结果、频谱、损伤等方面对台架的驱动信号和实测路谱进行了对比. 孙成智等^[3]未系统评价虚拟路谱和实际路谱之间的差别,而是从数据的主要峰值、RMS 比值、伪损伤比值等方面进行了对比,但未给出明确的标准. 王长明等^[5]在此基础上增加了数据的标准差、穿级计数、功率谱密度等参数的对比,完善了这一工作,但仍未给出明确的评价标准,仅给出模糊评价“一致性较好”. 综上所述,当前对虚拟时域信号的评估方法不够全面,且无明确的数值评价标准.

主观评价是在工程开发中常用的一种方法. 在一些以人的感觉为主要评价依据的领域,例如对噪声水平、声音品质的评价^[6]等,以及一些难以完全用客观数据进行评估的场景,例如对汽车整车腐蚀状态的评估^[7]、对车内挥发性气味的评估^[8]等,都可以用主观评估的方法完成对试验结果或产品性能的评估. 但主观评价存在的最大问题是评价结果受评价者本身的影响较大,评价者的专业能力、评分时的状态波动等都会对评价结果产生不确定的影响.

随机森林(random forest)是一种被广泛应用的非线性机器学习算法,是一种通过集成学习算法(ensemble learning algorithm)所组成的集成评估器(ensemble estimator),其本身并非一个单独的算法,而是先利用特定的算法构建多个模型,这些模型被称为基评估器(base estimator),再用某种方法将这些模型的结果集成在一起,以获取比单一基评估器更好的表现. 随机森林的决策过程可通过树状图清晰地展示出来,有助于问题的理解及模型调参. 集成算法可避免决策树作为单一基评估器的不稳定性,利用“有放回抽样(bootstrap)”的方法,形成不同的训练数据,建立多个不同的相互独立的决策树模型作为基评估器,对所有独立树模型的结果进行求平均作为集成评估器的结果,从而提高模型的鲁棒性.

客观特征和主观评分作为评价虚拟时域信号的两种手段,通过机器学习,固化专家的经验,避免人为噪声的持续输入,从而形成一种可靠有效的虚拟时域信号评分方法.

1 时域信号的客观特征及其局限性

实际路谱是通过路谱采集试验所获取的真实道路载荷谱,可作为汽车开发过程中很多关键工程开发环节的输入,如多体动力学分析、有限元仿真分析、多通道台架试验等. 实际采集的路谱是真实的时域信号,可作为对各类虚拟时域信号进行评判的参考依据,例如评价虚拟试车场所产生的虚拟路谱即可采用同样工况下的实际路谱进行对比.

本文所进行的路谱采集试验,采集了 70 种工况数据,每个工况均有 74 个数据通道. 记某个工况下的某个通道的实际路谱的时域信号为 $S_{i,j}$,所形成的实际时域信号矩阵为 \mathbf{S} ;同样工况下的虚拟时域信号为 $S'_{i,j}$,形成的实际时域信号矩阵为 \mathbf{S}' :

$$\mathbf{S} = \begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,74} \\ S_{2,1} & S_{2,2} & \cdots & S_{2,74} \\ \vdots & \vdots & \ddots & \vdots \\ S_{70,1} & S_{70,2} & \cdots & S_{70,74} \end{bmatrix}_{70 \times 74}, \quad (1)$$

$$\mathbf{S}' = \begin{bmatrix} S'_{1,1} & S'_{1,2} & \cdots & S'_{1,74} \\ S'_{2,1} & S'_{2,2} & \cdots & S'_{2,74} \\ \vdots & \vdots & \ddots & \vdots \\ S'_{70,1} & S'_{70,2} & \cdots & S'_{70,74} \end{bmatrix}_{70 \times 74}. \quad (2)$$

每一条时域信号均可产生诸多统计值,如最大值、最小值、均方根值等. 将虚拟时域信号的统计值 $f'_{S_{i,j},n}$ 和实际时域信号的统计值 $f_{S_{i,j},n}$ 的相对差作为评价虚拟时域信号的客观特征,记为 F_n :

$$F_n = \frac{f'_{S_{i,j},n} - f_{S_{i,j},n}}{f_{S_{i,j},n}} \times 100\%. \quad (3)$$

由特征值组成特征矩阵 \mathbf{F} . 根据工程经验,本文共选择 14 个统计值,故每个虚拟时域信号拥有 14 个特征值,如表 1 所示. 合并试验工况和数据通道这两个维度,使 $\mathbf{F}_{5180 \times 14}$ 形成一个二维矩阵,方便运算和描述. 直接用 \mathbf{F} 中的某一个或几个特征对虚拟时域信号进行评价存在两个问题:一是不够全面,二是没有合理标准. 此外,有些特征并不客观.

表 1 特征值列表
Table 1 List of all features

序号	特征缩写	特征含义	序号	特征缩写	特征含义
1	Max	最大值	8	F_Mse	频域数据均方根误差
2	Min	最小值	9	50%_Max_MSE	时域数据大于 50%最大值的均方根误差
3	Range	变程值	10	50%_Min_MSE	时域数据小于 50%最小值的均方根误差
4	RMS	均方根值	11	75%_Max_MSE	时域数据大于 75%最大值的均方根误差
5	Crest Factor	峰谷系数	12	75%_Min_MSE	时域数据小于 75%最小值的均方根误差
6	Kurtosis	峰态系数	13	Damage($K=-3$)	损伤(SN 斜率取 $K=-3$)
7	Variance	方差	14	Damage($K=-5$)	损伤(SN 斜率取 $K=-5$)

2 时域信号的主观评分

主观评价是一种人的心理判断活动. 专业研究^[9]认为,在处理这种心理判断问题时,判断等级划分为 7 ± 2 个等级是最佳范围. 本文对时域信号的评分标签划为 5 个等级,如表 2 所示. 选择 M 名有丰富经验的专家以 S 内的信号作为参考,对 S' 内的所有信号进行打分. 第 m 位专家对每条时域信号的评分记为 $p_{i,j}^m$,所有信号的评分保存在评分矩阵 \mathbf{P}_m 中. 对所有专家的评分进行一致性检验,即衡量 M 个变量的相关度,可用 Kendall 和谐系数 W 进行评判^[10]:

$$W = \frac{12 \cdot S}{M^2(N^3 - N) - M \sum_{i=1}^M T_m}, \quad (4)$$

其中,

$$S = \sum_{i=1}^N (R_i - \bar{R})^2,$$

$$T_m = \sum_{j=1}^J (n_{mj}^3 - n_{mj}).$$

式中, M 为专家数; N 为待评价信号数; R_i 为每条虚拟时域信号的评分之和; \bar{R} 为同一位专家对所有虚拟时域信号的评分平均值; J 为第 m 位专家出现重复评分的级数; n_{mj} 为第 m 位专家第 j 级重复评分的个数.

表 2 虚拟时域信号主观评分依据

Table 2 Principals of subjective evaluation on virtual time history signals

评分	等级定义	等级描述
0	很差	被评价信号与目标信号几乎没有任何相似之处
1	差	被评价信号与目标信号在某些特征上有一定的相关性,但被评价信号不能用来代替目标信号
1.5	合格-	介于 1 和 2 之间
2	合格	被评价信号与目标信号整体看起来还不错,用被评价信号去代替目标信号是可以接受的
2.5	合格+	介于 2 和 3 之间
3	好	被评价信号和目标信号的主要特征都吻合得非常好
4	非常好	被评价信号与目标信号几乎完全一致

利用 SPSS 计算 Kendall 系数,得到 $W=0.674$. 本文采用大样本近似法计算卡方值 χ^2 对该结果进行显著性检验^[11-12]:

$$\chi^2 = M(N-1)W. \quad (5)$$

代入 $M=12$, $N=5180$, $W=0.674$, 求得 χ^2 为 41 887.75. 对于卡方分布,当 n 充分大时近似于正态分布,其界值可按下式估算:

$$\chi_{\alpha,N}^2 = (Z_{\alpha} + \sqrt{(2N-1)})^2/2. \quad (6)$$

当 $\alpha=0.01$ 时,在该显著性水平上的卡方分布临界值 $Z_{\alpha}=2.326$, $\chi_{\alpha,N}^2$ 为 5 409.58,显然 $\chi^2 > \chi_{0.01,5180}^2$,故评分结果的一致性显著的. 对所有专家的主观评分求平均值,同样对试验工况和数据通道两个维度进行合并,所得的评分平均值作为向量 $\bar{\mathbf{P}}_{5180 \times 1}$ 存在,该向量可作为机器学习模型的标签向量. 主观评分的流程如图 1 所示.

结合 $\bar{\mathbf{P}}$ 和 \mathbf{F} 可以看出,客观特征与主观评分之间呈现出很明显的非线性关系,如图 2 所示. 且客观特征的分布量级跨度很大,所呈现出的分布形态多变,不易进行参数化描述,如图 3 所示. 这就要求接下来所选用的机器学习算法是非线性算法,且对特征数据的量级和分布形态无严格要求.

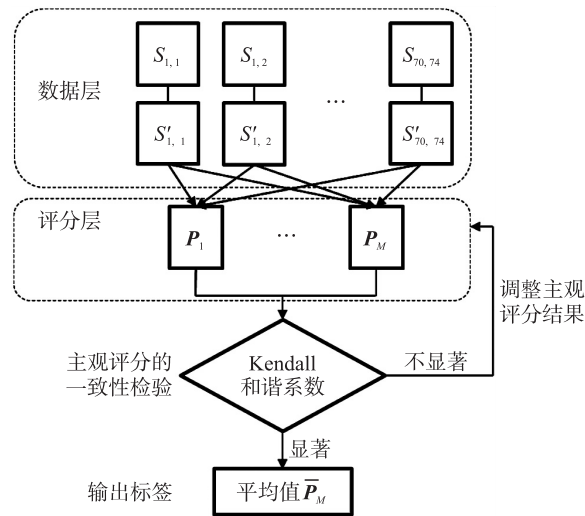


图 1 虚拟时域信号主观评分流程图

Fig. 1 Flow chart of subjective evaluation on virtual time history signals

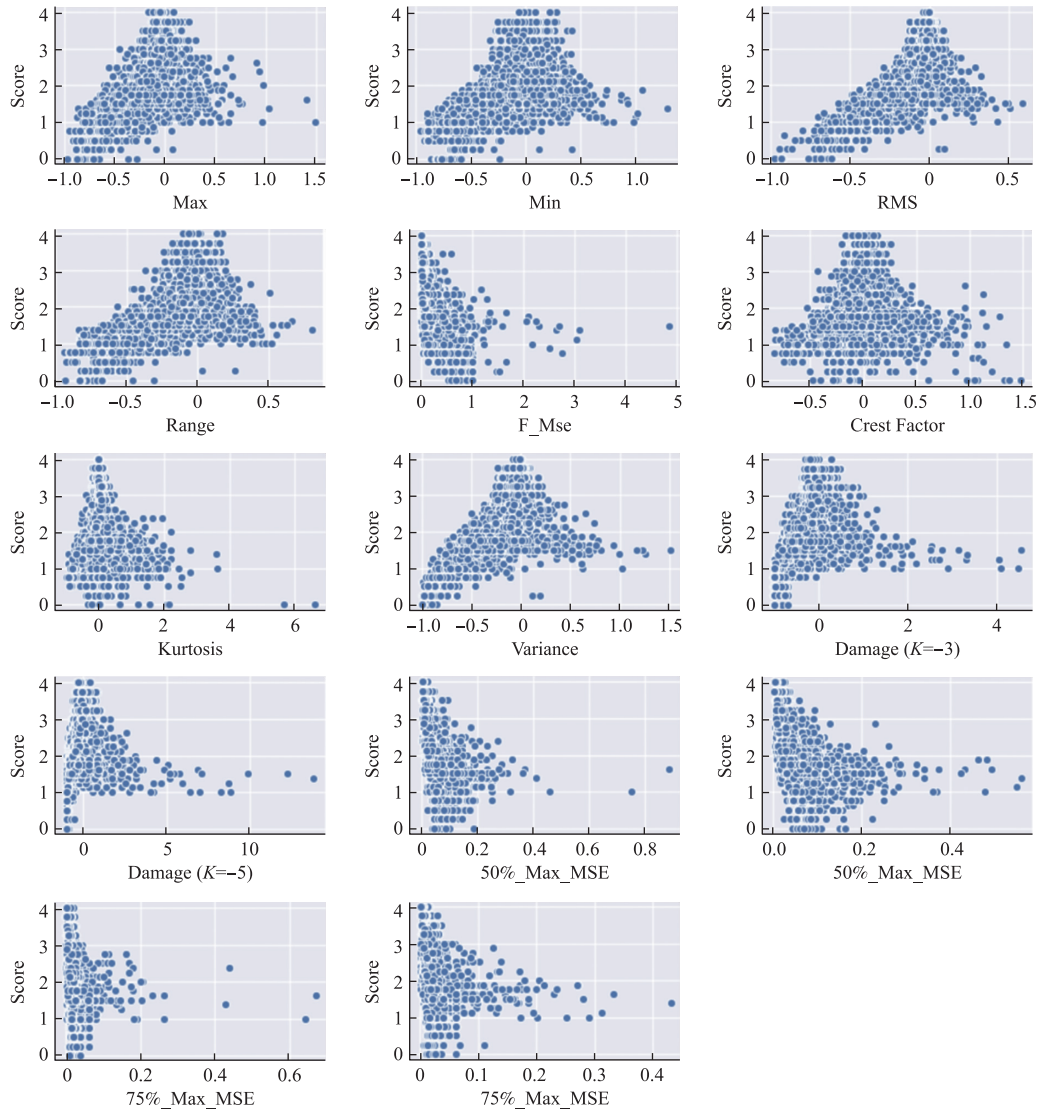


图 2 客观特征与主观评分的关系散点图

Fig. 2 Relationship between subjective score and objective features

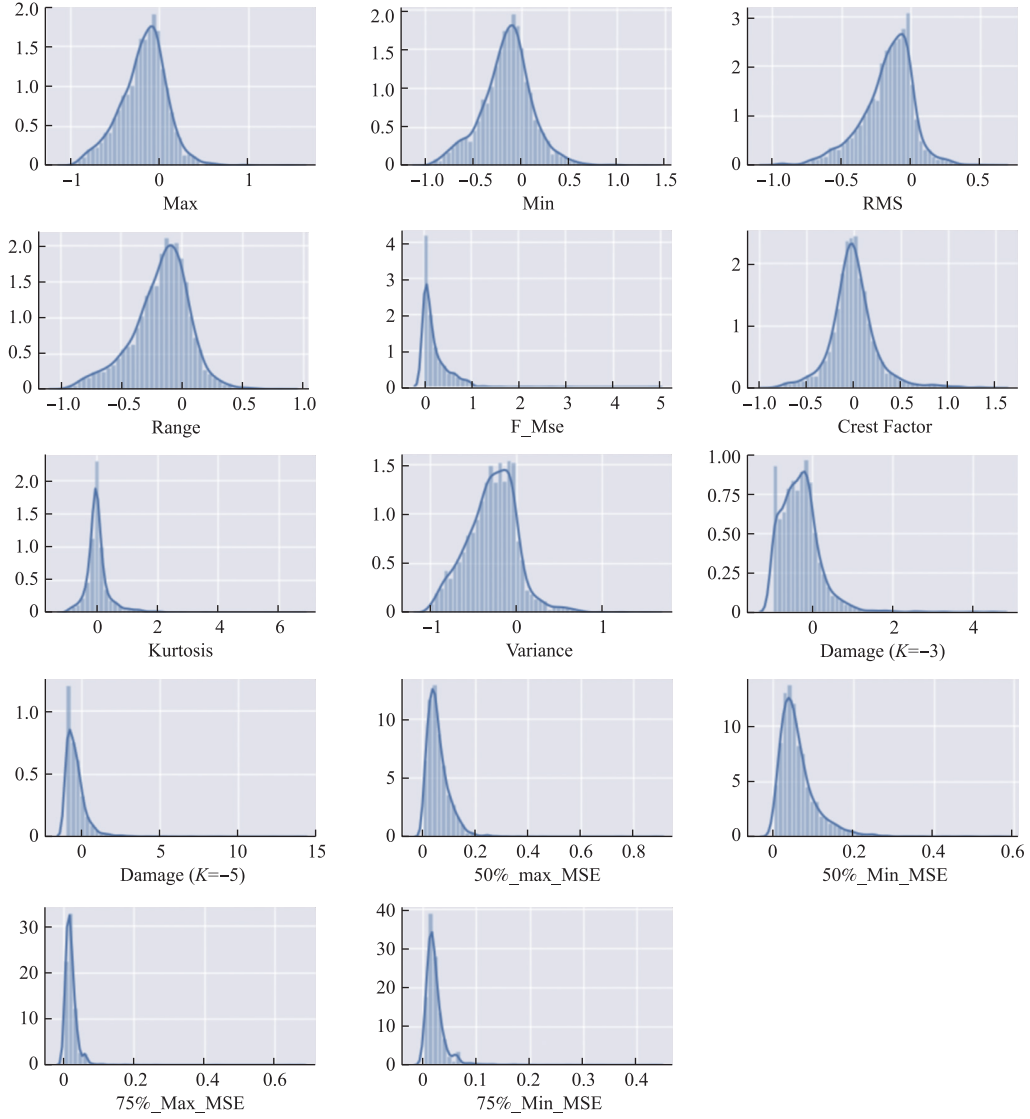


图3 客观特征的概率分布图

Fig. 3 Distributions of objective features

3 虚拟时域信号智能评价模型

3.1 随机森林方法

决策树(decision tree)是随机森林的基评估器. 从某个特征 F_n 出发, 以该特征的某个值 s 作为分类的阈值, 整个数据被分成两个区域:

$$R_{\text{left}}(n, s) = \{F_n \mid F_n \leq s\}, \quad (7)$$

$$R_{\text{right}}(n, s) = \{F_n \mid F_n > s\}. \quad (8)$$

落在 $R_{\text{left}}(n, s)$ 区域的数据有 K_{left} 个, 落在另外一个区域的数据有 K_{right} 个, 且有:

$$K_{\text{left}} + K_{\text{right}} = K. \quad (9)$$

本文中 $K=5180$. 遍历所有的 N 个特征, 若所有数据的标签值的均方误差 MSE 之和最小, 则认为该特征 F_n 可作为根节点, 对应的数值 s 可作为阈值:

$$(F_n, s) = \arg \min_{n, s} \left[\frac{1}{K_{\text{left}}} \sum_{F_n \in R_{\text{left}}} (p_k - \bar{p}_1)^2 + \frac{1}{K_{\text{right}}} \sum_{F_n \in R_{\text{right}}} (p_k - \bar{p}_2)^2 \right]. \quad (10)$$

继续在两个子空间 R_{left} 和 R_{right} 上进行上述划分操作, 以获得更多的叶子节点, 直至各个叶子节点上的 MSE 都达到较小的水平无法继续进行空间划分, 或触及了事先设定的树的停止生长条件, 则该过程停止,

该决策树完成生长.

此过程即为最常用的分类与回归树(CART)方法^[13]. 本文选择回归树(decision tree regressor)作为基评估器,对所有独立树模型的结果进行求平均作为集成评估器的结果,也即袋装集成算法(Bagging)的思路. 由袋装集成算法所形成的模型即随机森林模型. 采用有放回抽样方法保证了各个基评估器的多样性,从而显著降低随机森林模型对训练样本的依赖度,提高了模型的稳定性.

随机森林模型的预测效果可用决定系数 R^2 和均方误差 MSE 来评估. 其中, R^2 按下式计算:

$$R^2 = 1 - \frac{\text{MSE}}{\text{Var}},$$

(11)

式中,Var 为方差;MSE 代表的是残差平方和,Var 代表的是总平方和. R^2 越接近 1,则模型的预测效果越好. 在 Var 一定的前提下,MSE 越小, R^2 越大. 以上两个指标可基于一组全新的验证数据进行计算.

此外,bootstrap 有放回抽样的特点,还提供了另外的验证数据来源. 由于是有放回抽样,必然有一些样本数据从未参与过任何基评估器的训练过程,这些数据可认为是在袋装集成算法的考虑范围内,属袋外数据(out of bag),可利用这些数据对模型的效果进行评估,评估结果称为袋外数据得分.

3.2 模型训练

随机森林模型的参数众多,其中基评估器数量(即组成随机森林的单棵决策树的数量)越大,模型效果通常就越好,但代价是模型计算量也会越大. 所以要先确定基评估器的数量,用绘制学习曲线的方法,先进行大范围的粗略学习,如图 4 所示,初步确定基评估器数量在 110 个附近时模型效果最好. 然后在 110±10 这个区域内再进行精细学习,结果如图 5 所示,故本文选择的基评估器数量为 115 个.

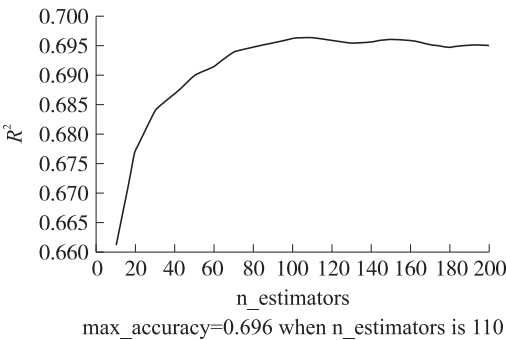


图 4 基评估器数量学习曲线(大范围粗略学习)
Fig. 4 Learning curve of estimators(rough learning over large scope)

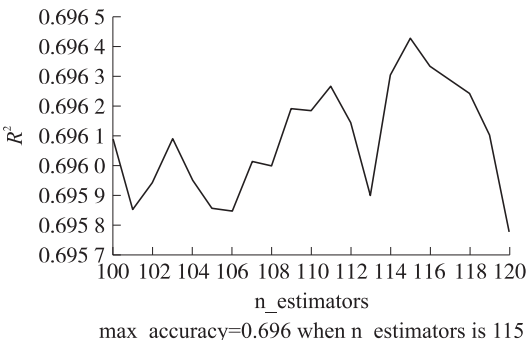


图 5 基评估器数量学习曲线(小范围精细学习)
Fig. 5 Learning curve of estimators(fine learning over small scope)

最大深度、最小分枝叶片数和最小分枝样本数等 3 个参数是决定随机森林中各决策树生长的参数,即决策树的生长停止条件. 这 3 个参数相互影响,应采用网格搜索法^[14-15]进行同步调节,但若搜索范围太大会导致计算量太大,严重影响搜索效率. 故先采用单参数学习曲线的方法,找到各参数的最优区间,然后再进行局部区域的网格搜索. 调参结果如表 3 所示.

利用实验获取的 5 180 组数据进行模型训练,并采用不同方法评价训练效果. 本模型的袋外数据 R^2 得分为 0.769. 另一种常用的评价方法是交叉验证,本文选用随机排列交叉验证方法(shuffle split cross validation)进行交叉验证,所得结果 R^2 平均分为 0.765,MSE 为 0.122.

3.3 模型结构优化

在训练过程中,可随机在森林中的每棵树上分析每个特征形成的分枝节点上 MSE 的下降程度,然后对所有树的结果进行平均,其结果可以评价该特征在模型中的重要程度,即特征重要度(feature importance, FI)^[16-17]. 对各个特征按重要度进行排序,并对模型进行降维优化. 在建模时,增加了一个随机特征,命名为 Random,保留大于 10 倍于 Random 重要度的特征,即保留特征的依据为:

表 3 随机森林模型参数

Table 3 Parameters of random forest model

模型参数	单参数最优值	网格搜索区域	网格搜索最优值
最大深度	7	[5,15]	13
最小分枝叶片数	8	[5,15]	9
最小分枝样本数	113	[60,160]	69

$$FI_i > 10 \times FI_{\text{Random}} \quad (12)$$

本文所保留的前6个参数如表4所示。这6个参数的特征重要度之和达到了97.55%。

表4 降维优化保留的特征及重要度

Table 4 Retained features after dimension reduction and the importance

序号	特征名称	FI _i /%	序号	特征名称	FI _i /%	序号	特征名称	FI _i /%
1	Variance	26.00	3	50%_Max_MSE	17.15	5	50%_Min_MSE	14.07
2	RMS	21.04	4	Damage(K=-3)	16.98	6	F_Mse	2.32

3.4 建模结果

将完成降维优化的数据继续进行模型训练,按3.2节方法调节模型参数,得到最终的模型结果,如表5所示。

表5 随机森林模型参数优化

Table 5 Parameters of random forest model after dimension reduction

模型参数	原始最优值	降维优化后的最优值
基评估器数量	115	160
最大深度	13	7
最小分枝叶片数	9	21
最小分枝样本数	69	84

优化后的模型袋外得分为0.766,交叉验证平均分为0.756,MSE为0.125。和降维优化之前相比,模型的精度几乎未受影响,但模型的复杂度显著降低。

本模型进行训练时,对训练数据进行了划分,其中10%的数据未参与训练,因此可用这部分数据对模型的泛化能力进行验证。如图6所示,模型的输出评分和实际评分的分布很接近。进一步分析模型评分和实际评分的差值,发现85.3%以上的数据评分差值在0.5分以内,说明在大部分数据上模型的输出评分和实际评分的差别很小。

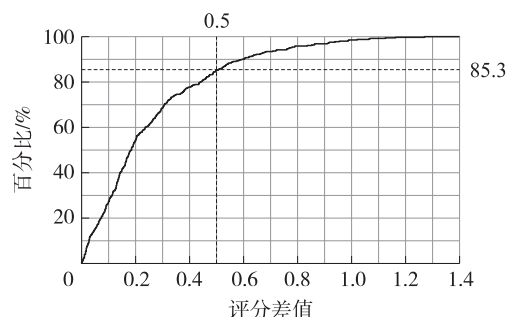


图6 验证数据的模型评分和实际评分的差值

Fig. 6 Difference between model output scores and experts subjective scores on validation data

4 模型效果验证

4.1 时域信号智能评分模型封装

保存训练好的随机森林模型,封装在nCodeGlyphworks平台上,形成一个自动化的智能评分流程,实际时域信号和待评分的虚拟时域信号同时输入该流程,完成客观特征统计,然后利用随机森林模型进行运算,最后输出评分结果,如图7所示。

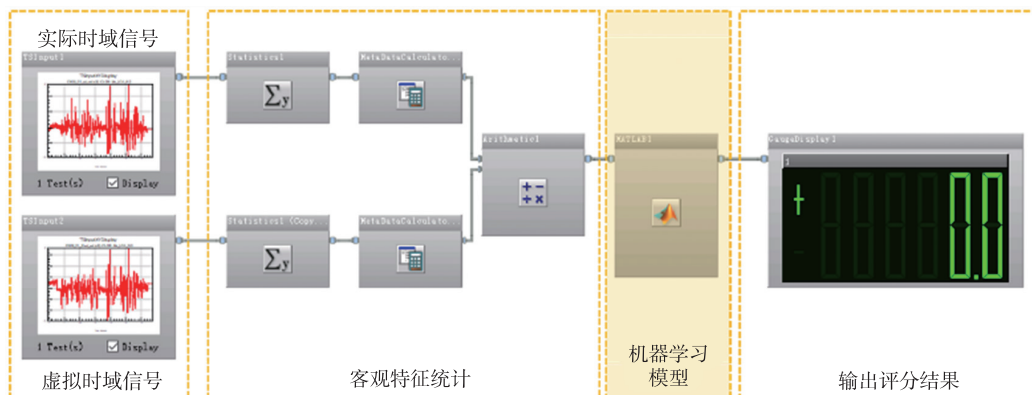


图7 基于nCodeGlyphworks平台的虚拟时域信号评分流程

Fig. 7 Work flow chart based on nCodeGlyphworks software platform

4.2 模型运行效果

选取5组不通工况的路谱及对应的用两种不同方法获得的虚拟路谱,利用智能评分模型对这些虚拟时域信号进行评分,结果如图8-12所示。从智能评分模型的输出结果来看,基本能反映出专家的主观判断,且能对不同的虚拟信号进行区分,说明利用随机森林开发的虚拟时域信号智能评分模型可以代表专家

的主观评价开展评分工作.

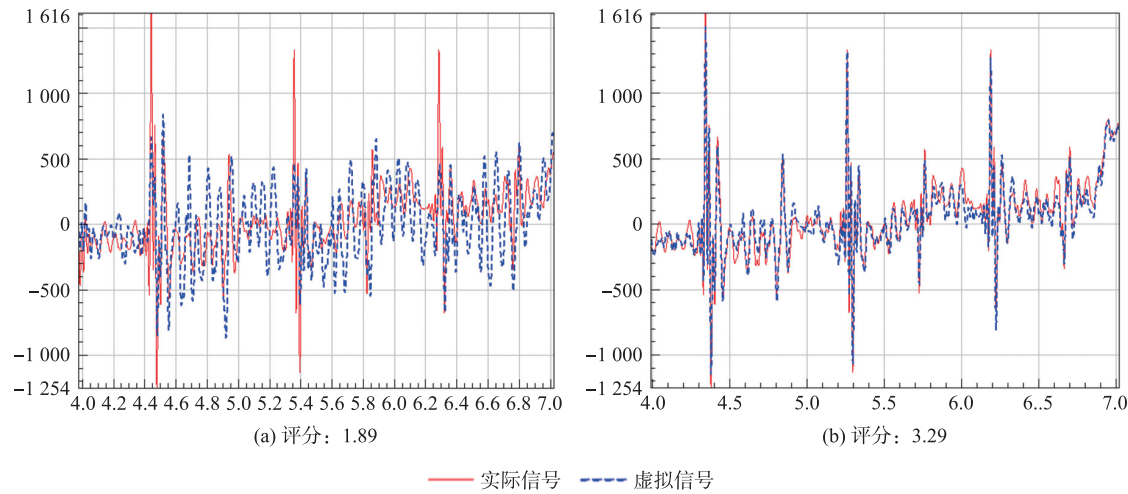


图 8 水泥接缝路信号及评分

Fig. 8 Signals and scores on cement joints road

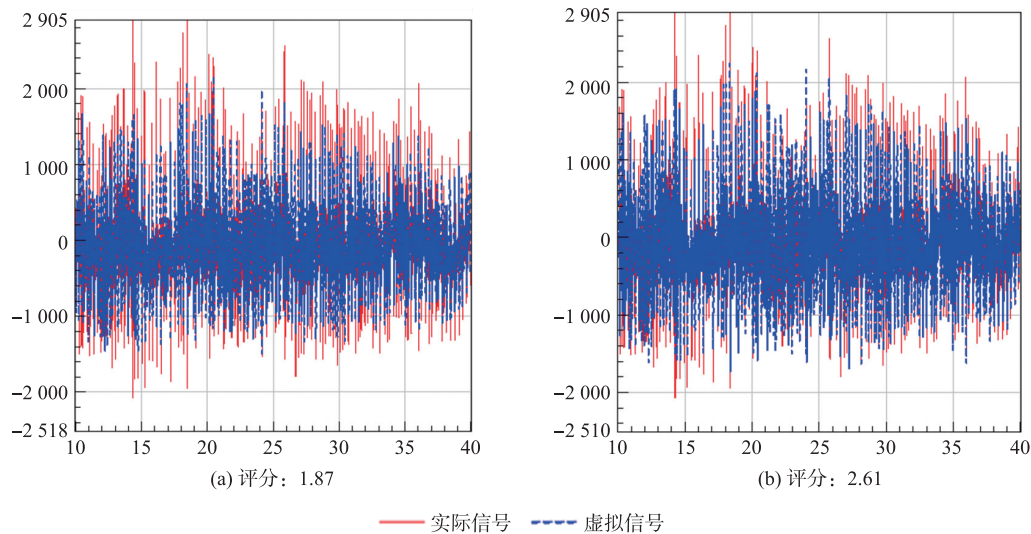


图 9 共振路信号

Fig. 9 Signals and scores on resonance road

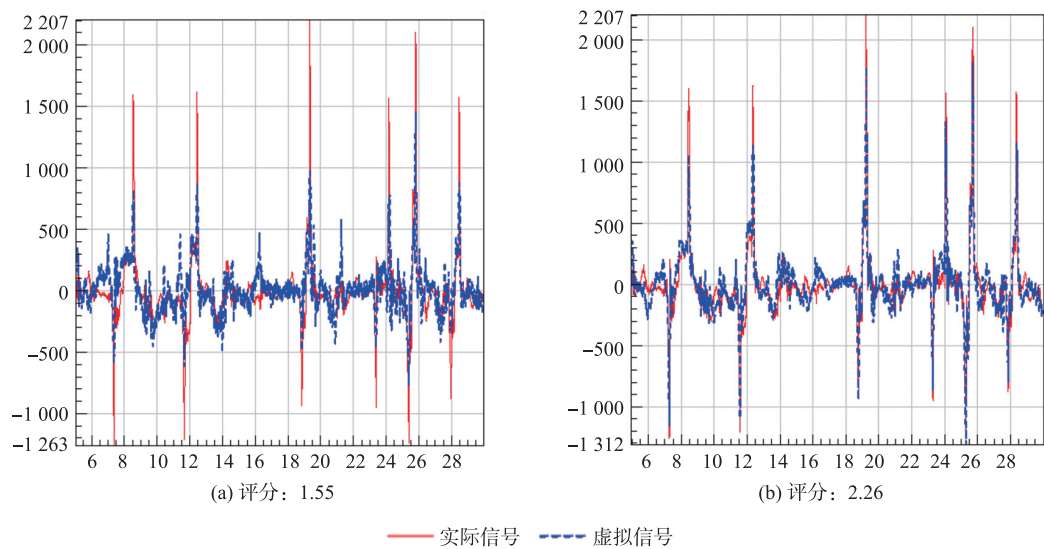


图 10 坑洼路信号

Fig. 10 Signals and scores on corrugation road

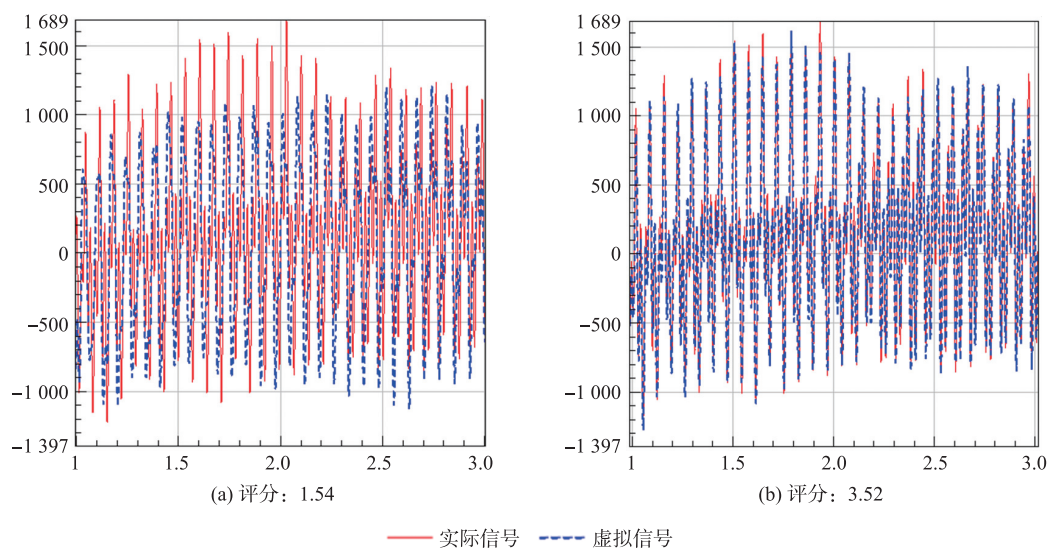


图 11 绳索路信号

Fig. 11 Signals and scores on ropes road

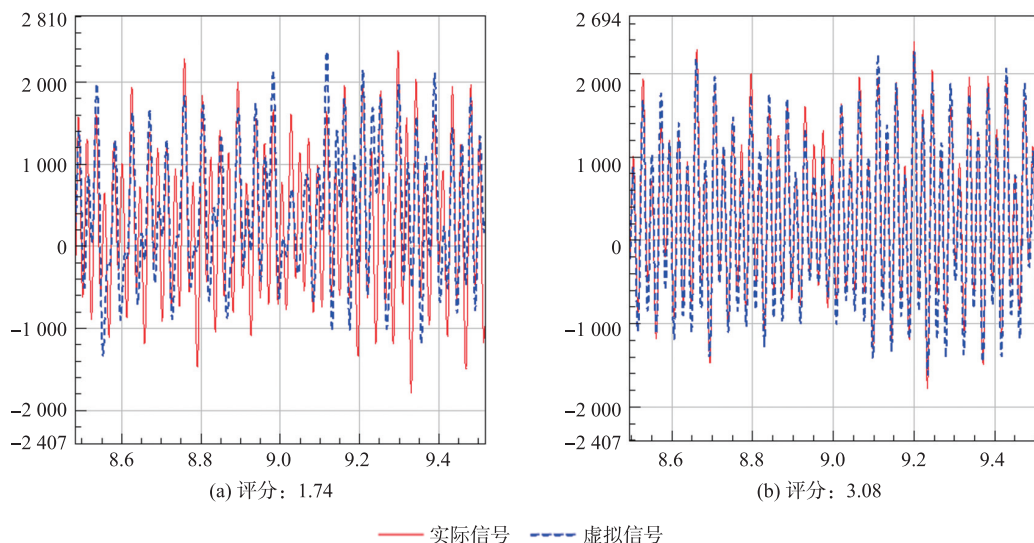


图 12 搓板路信号

Fig. 12 Signals and scores on washboard road

5 结论

本文通过分析当前虚拟时域信号评价方法中的局限性,确立了利用机器学习方法开发虚拟时域信号智能评分系统的目标. 首先提取出评价虚拟时域信号的多个客观统计值,形成客观特征 F ,然后组织多位专家进行主观评分,利用 Kendall 和谐系数确保多位专家意见的一致性,进而形成标签向量 P . 而后利用随机森林方法,以 F 和 P 为训练数据进行模型训练,通过对模型参数的优化,最终形成一个智能评价模型. 该模型以时域信号的客观统计值为输入,以主观评分为输出,在进行虚拟时域信号评价时具有以下优点:

- (1) 兼顾多个客观统计值,能全面考虑时域信号各方面的指标;
- (2) 输出结果体现了专家经验,又可避免人的情绪、状态等不确定因素所导致的评分结果波动;
- (3) 模型的泛化能力强,对新的虚拟时域信号的评分准确可靠;
- (4) 可快速执行大量时域信号的评分工作,效率高.

[参考文献] (References)

- [1] 邱宇,王磊,张皓清. 基于客户数据的混动车散热器热疲劳性能优化与认证[J]. 汽车零部件,2018(12):8-14.

- [2] 王灵龙,刘伟,李伟. 基于整车室内道路模拟的汽车零部件疲劳试验研究[J]. 汽车实用技术,2019(24):97-100.
- [3] 孙成智,段向雷,翁洋,等. 基于 3D 数字路面的整车耐久性能评价方法研究[J]. 汽车工程,2017,39(10):1211-1216.
- [4] 赵强,霍福祥,杨立峰,等. 发动机悬置支架多轴加载道路模拟试验方法研究[J]. 汽车文摘,2020(1):36-40.
- [5] 王长明,曹正林,彭亮,等. 虚拟试验载荷分析与试验验证研究[J]. 汽车技术,2016(8):43-47.
- [6] 左言言,张海峰,庄婷. 车内声品质主客观评价模型对比分析[J]. 江苏大学学报(自然科学版),2017,38(4):403-409.
- [7] 王添琪,马琼,李易. 整车防腐性能量化评价方法研究[J]. 汽车实用技术,2018(8):148-150.
- [8] 黄文杰,吕孟强,高鹏,等. 基于用户主观评价的车内气味调查研究[J]. 暖通空调,2020,50(4):14-20.
- [9] 毛东兴. 声品质研究与应用进展[J]. 声学技术,2007,26(1):159-164.
- [10] 柯蕙新,黄京华,沈浩. 调查研究中的统计分析法[M]. 北京:北京广播学院出版社,1992.
- [11] DALIA V, ROSA E L, JUAN R. A Kendall correlation coefficient between functional data[J]. Advances in Data Analysis and Classification, 2019, 13(4):1083-1103.
- [12] 程琮,刘一志,王如德. Kendall 协调系数 W 检验及其 SPSS 实现[J]. 泰山医学院学报,2010,31(7):487-490.
- [13] AURÉLIEN G. Hands-On Machine Learning with Scikit-Learn and TensorFlow Concepts, Tools, and Techniques to Build Intelligent Systems[M]. Cambridge, USA: O'Reilly Media Inc, 2017.
- [14] 奚杏杏,刘宇红,张荣芬. 基于网格搜索与交叉验证的 SVR 血压预测[J]. 计算机与现代化,2020(3):44-48.
- [15] 王兴玲,李占斌. 基于网格搜索的支持向量机核函数参数的确定[J]. 中国海洋大学学报(自然科学版),2005,35(5):859-862.
- [16] 彭漂. 基于随机森林的变量重要性度量和核密度估计算法研究[D]. 厦门:厦门大学,2017.
- [17] BOULESTEIX A L, BENDER A, BERMEJO J L, et al. Random forest Gini importance favours SNPs with large minor allele frequency: impact, sources and recommendations[J]. Briefings in Bioinformatics, 2012, 13(3):292-304.

[责任编辑:严海琳]