Proceeding of the 18th INFATS & 27th SAE-C/STC, 2024, pp1-6
2024.606

New Virtual Evaluation Technology for Seatbelt Chest Displacement Based on Machine Learning

Lihua ZHU[1], Ze CHENG[2],WANG[3], , Hui YANG[4]
[1]*AOS, ZF LIFETEC, Shanghai, China*
[2] *AOS, ZF LIFETEC, Shanghai, China*
[3] *AOS, ZF LIFETEC, Shanghai, China*
[4] *AOS, ZF LIFETEC, Shanghai, China*
Email: *lihua.zhu@zf-lifetec.com, ze.cheng@zf-lifetec.com, zhenfei.wang@zf-lifetec.com, hui.yang@zf-lifetec.com*

Abstract: Explore the possibility of dummy chest displacement in seatbelt sled tests by establishing four machine learning algorithm models in the field of automotive virtual evaluation. Optimized by PCA, the result indicated that PCA could enhance the prediction accuracy of the four models. Among the four algorithm models, XGBoost and LightGBM showed similar performance, achieving prediction accuracy of over 89%, which was relatively high. Additionally, the importance of data governance in improving data quality and the predictive performance of algorithm models was emphasized.

Keywords: Virtual evaluation; Machine learning; Seatbelt sled; Chest displacement prediction; Data governance

基于机器学习的安全带胸部位移虚拟测评技术研究

朱利华 [1]，成泽 [2]，王振飞 [3]，杨慧 [4]
[1] 采埃孚亚太汽车安全系统（上海）有限公司，上海，中国，200000
[2] 采埃孚亚太汽车安全系统（上海）有限公司，上海，中国，200000
[3] 采埃孚亚太汽车安全系统（上海）有限公司，上海，中国，200000
[4] 采埃孚亚太汽车安全系统（上海）有限公司，上海，中国，200000
Email: lihua.zhu@zf-lifetec.com, ze.cheng@zf-lifetec.com, zhenfei.wang@zf-lifetec.com, hui.yang@zf-lifetec.com

摘　要:本文建立了四种机器学习算法模型，对安全带滑台实验中假人胸部位移虚拟测评进行研究，经 PCA 优化后的结果表明：PCA 能够提高四种模型的预测精度，且在四种算法模型中 XGBoost 和 LightGBM 的预测结果性能类似，达到 89%以上，预测精度较高。本研究采用的数据经过数据治理，提升了数据质量，对提高算法模型预测性能起到关键作用

关键词: 虚拟测评；机器学习；安全带滑台；胸部位移预测；数据治理

1 引言

汽车虚拟测评是一种利用计算机模拟技术来评估汽车性能和安全性的方法，该方法可以在没有实际制造和测试汽车的情况下，对汽车的设计、性能和安全性进行全面的分析和评估[1]。而虚拟测评的应用可以大幅降低研发成本和时间，同时提高汽车设计的安全性和可靠性[2]。

中国汽研汽车安全技术中心在 2022 年 11 月 17 日召开的会议上讨论了安全虚拟测评的研究进展，强调了虚拟测评在汽车安全与测评技术发展中的重要性。此外，C-NCAP 2024 版在全球首次引入了离位乘员保护虚拟测评，这表明虚拟测评技术已经开始应用于实际的车辆安全评估标准中[3]。

虚拟测评的技术基础包括计算机数值仿真、车辆动力学模型、传感器模型和决策规划算法等，这些技术可以模拟车辆在各种工况下的性能，包括在危险场景中智能安全系统的表现，以及在碰撞工况中假人或

人体模型的运动和损伤响应[4]。

随着技术的发展，虚拟测评有望成为汽车安全评估的重要工具，为实现道路交通“零伤亡愿景”提供支持。

GB14166 规定汽车被动安全系统-安全带性能滑台测试中，安全带性能需满足假人胸部位移（带预紧）50mm-300mm, 100mm-300mm（非预紧）的要求[5]。

影响假人胸部位移的因素与假人胸部位移结果之间具有高度且复杂的非线性关系以及较强的噪音， 且涉及因素较多， 用传统的多项式拟合方法难以取得较好的预测结果。本文以采埃孚亚太汽车安全系统（上海）有限公司的 2018-2024 年的有效安全带滑台实验数据 583 组（训练集-518 组，测试集-65 组），分别尝试使用 BP 神经网络+LSTM、SVR、XGBoost、LightGBM 来对多款车型下不同座位的 GB14166 安全带滑台实验假人胸部位移的预测，以此用于虚拟测评快速评价安全带性能。

2 设计变量选取及相关数据治理

2.1 设计变量的选取

基于对历史数据的基本分析和安全带系统的专业知识， 选定影响安全带滑台假人胸部位移高度相关的数据特征共包括 23 个，主要来自四类：碰撞强度包括碰撞速度、碰撞加速度； 安全带产品固定点位置，包括卷收器位置、锁扣位置等；安全带相关产品，包括卷收器类型、限力杆直径、动态锁舌使用情况等；实验参数，包括点火时间、安全带剩余量等。

2.2 数据治理

数据资产是现代化企业非常重要的组成部分，它们是由企业拥有或控制的，能够为企业带来未来经济利益的数据资源[6]。如何将数据资产化，即将数据转化为经济价值的资产的过程，设计到数据资源化、产品化和可视化三个阶段[7]。而数据治理可以确保数据的质量和安全性，包括数据质量、数据安全、数据合规性、数据的存储和访问等方面[8]，通过数据治理提升数据质量的提升，可以有效地提升数据的资产化。在假人胸部位移预测应用中，数据的质量是影响算法模型预测性能的关键因素。

2.2.1 安全带滑台数据收集及数据清洗

安全带滑台数据保存在企业公共实验盘上，数据现状为数字化不完全，数据保存格式不统一。产品信息以照片形式、固定点信息为手工记录并以照片形式进行保存，整体上看数据无法使用数字化技术进行自动提取筛选， 很难直接用于算法学习。目前的算法开发是人工将上述选取的 23 个设计变量及相关胸部位移量结果数据进行整合，并对失效无效数据进行清洗，对冗余数据进行剔除，这部分工作需要有经验的工程师完成，才能做到既不丢失数据，又不造成数据污染，使算法精度大幅提升。比如历史实验数据有些未标注实验失效，可以通过传感器检测到的假人胸部位移曲线分析进行筛选，如图 1， 此项目重复实验共 15 组，其中曲线 12 数据明显变异，经分析由于实验过程中假人滑脱导致。通过以上过程共筛选出 2018-2024 年的有效安全带滑台实验数据 583 组（训练集-518 组，测试集-65 组）。

2.2.2 未来安全带滑台数据治理

未来的实验数据应该是用数字化手段直接自动进行提取及筛选，直接用于算法训练和预测。传统的安全带滑台数据的收集过程耗费大量的人力， 造成企业资源的浪费，亟需进行改革通过数据治理提高数据的质量、安全性和合规性。

以下为安全带滑台数据治理工作的主要内容：

- 定义全面化、标准化的数据输入输出模板及数据保存框架。
- 实施无纸化电子化办公。
- 落实相关数据责任人及提升数据监管力度。

- 实施考核制度。

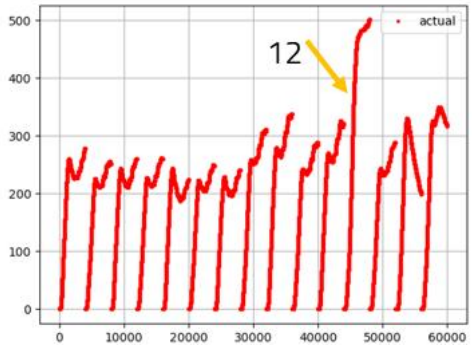

Figure1. Dummy chest displacement curve

图 1. 假人胸部位移曲线分析图

2.2.3 集成安全数据平台搭建

搭建集成安全数据平台的目的：

- 将标准化记录的实验相关数据进行归集，包括数据表、时序数据、视频图片等等，该系统对实验数据进行分类管理，提高实验数据的保存质量及安全性。
- 可以进行自动扫描提取、格式转换、在线查询浏览等功能，方便用户进行查阅、筛选、分析实验。
- 自动提取、整合、筛选算法模型需要的数据，用于算法模型的更新迭代。
- 嵌入算法模型， 用户可进行新项目方案设计及结果可视化展示。

3 机器学习及算法模型优化

机器学习（Machine Learning, ML）是人工智能（Artificial Intelligence, AI）的一个核心分支，它涉及构建和使用算法，使计算机系统能够从数据中学习并提升性能，而无需进行明确的编程[9]。随着技术的不断进步， 机器学习在汽车领域内解决复杂问题和提高决策质量方面发挥着越来越重要的作用。本文分别尝试用 BP 神经网络+LSTM、SVR、XGBoost、LightGBM 四种模型进行来对多款车型下不同座位的 GB14166 安全带滑台实验假人胸部位移的预测，以此用于虚拟测评快速评价安全带性能或者企业的新方案设计。

3.1 BP 神经网络 + LSTM

反向传播(Back Propagation, BP)网络模型采用误差反向传播网络[10]，长短期记忆网络（LSTM，Long Short-Term Memory）是一种时间循环神经网络。将两种网络进行结合，进行回归处理时，可以得到比较好的效果。假设（X_p， Y_p）为第p个样本，x_k为第p个样本中的第k个参数。W^l为第l层连接权重矩阵，ω_{ki}为第k个神经元对第i个神经元的连接权重，则第i个神经元的网络输入为：

$$neti = x_1\,\omega_{1i} + x_2\,\omega_{2i} + \ldots + x_n\,\omega_{ni}$$

神经元的输出为：

$$Output = f(net) = \frac{1}{1 + e^{-net}}$$

对于 LSTM，这是一种时间循环神经网络，可以对数据有效的进行拟合。

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$
$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$
$$\hat{c_t} = \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$
$$c_t = f_t \odot c + i_t \odot \hat{c_t}$$
$$h_t = o_t \odot \sigma_h(c_t)$$

其中初始值$c_0 = 0$和$h_0 = 0$，运算符表示 Hadamard 积(元素积)。下标 t 表示时间步长。其中x_t为输入向量，f_t为遗忘门的激活向量，i_t为更新门的激活向量，o_t为输出门的激活向量，h_t为输出向量。$\hat{c}_t$为单元输入激活向量，c_t为单元状态向量。W，U 和 b 为需要学习的权重[11]。

将整个网络的结果输入到 *LSTM* 中，就可以得到最终的结果：

$$Predicted\_value = LSTM(Output)$$

3.2 SVR

基于 SVR(support vector regression)支持向量回归，是一种适用于离散数据的回归模型。在高维度下，参数可能会进行分离。这样，在超平面下，就可以进行数据分类，从而进行分类以及回归[12]。图 2 为 SVR 的示意图。

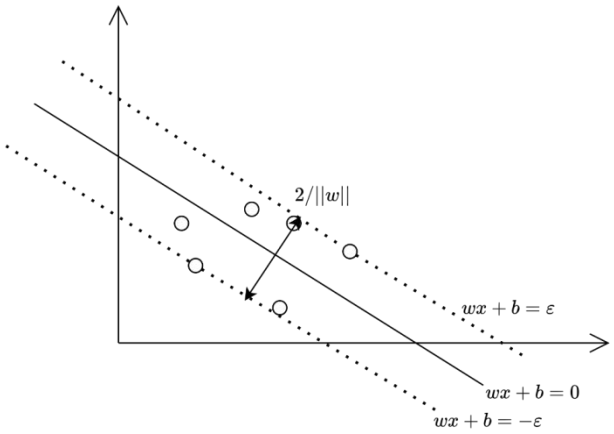

Figure2. SVR schematic diagram

图 2. SVR 示意图

对于任意的 SVR，只需要满足$min\frac{1}{2}||\omega||^2$，在$|y_i - (\omega x + b)| \leq \varepsilon$ 条件下。对于ω参数，一般选取核函数（RBF， Linear 等）进行高维映射。

3.3 XGBoost

XGBoost(Extreme Gradient Boosting)极端梯度提升, XGBoost 采用了一部分是损失函数，一部分是正则化（用于控制模型的复杂度），来进行分类以及回归，由于考虑了正则化项，大大提高了模型的鲁棒性，减少了过拟合[13]。

XGBoost 会训练多颗树，对于第 t 颗树，第 i 个样本，模型的预测值：

$$\hat{y}_i(t) = \sum_{k=1}^{t} f_k(x_i) = \hat{y}_i^{(t-1)} + f_t(x_i)$$

其中$\hat{y}_i(t)$ 是第 t 次迭代之后样本 i 的预测结果，$f_t(x_i)$是第 t 颗树的模型预测结果，$\hat{y}_i^{(t-1)}$是第 t – 1 颗树的预测结果。

这样，整体的目标函数可以定义如下：

$$Object = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{i=1}^{t} \Omega(f_j)$$

其中$l(y_i, \hat{y}_i)$是损失函数，$\hat{y}_i$是整个模型对第 i 个样本的预测值，y_i是第 i 个样本的真实值。$\Omega(f_j)$是全部第 t 颗树的复杂度求和，可以当作正则项。

3.4 LightGBM

LightGBM（Light Gradient Boosting Machine）是一种基于 Histogram 的决策树算法。将浮点特征离散化成 k 个整数，同时构造一个宽度为 k 的直方图。然后根据直方图的离散值，遍历寻找最优的分割点。同时，LightGBM 采用带有深度限制的按叶子生长(leaf-wise)算法，可以降低更多的误差，得到更好的精度[14]。图 3 为 LightGBM 计算过程示意图。

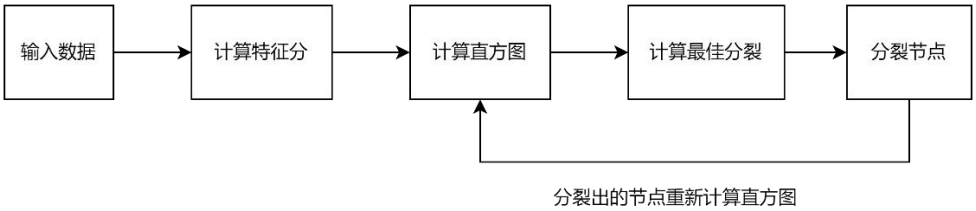

Figure3．LightGBM processing schematic diagram

图 3. LightGBM 计算过程示意图

3.5 算法模型优化

本文算法模型数据来自于历史物理实验，由于实验数据的限制，会出现数据量不足，同时，由于实验存在着噪音，需要对数据进行有效的清洗，且物理实验会有数据不均衡现象即有些特征会出现不均衡的情况，所以，需要对上诉 23 个数据特征进行优化处理。

目前采用 PCA 降维，PCA（principal Component Analysis），即主成分分析方法，是一种使用广泛的数据降维算法[15]。由于多变量之间可能存在相关性，从而增加了问题分析的复杂度。所以可以将输入参数进行降低维度，大大提高运算速度，减少运算时间。

$$Z = X * U$$

其中 X 是 m*n 的矩阵，U 是 n*k 的矩阵，Z 是 m*k 的矩阵，就可以得到 X 降维之后的降维矩阵 Z。这样就从 n 个参数，变成了 k 个参数。

同时，可以采用一些先验知识，对相关关键特征，加大其权重因子，可以有效地增加相关关键特征对应的数据量，增强模型的鲁棒性。

本文采用先验知识，对相关关键影响因子进行数据增强，然后采用 PCA +（BP+LSTM, SVR, XGBoost, LightGBM）结合，优化整体算法模型。表 1 为四类算法模型采用 PCA 优化前后的预测结果正确性对比。从结果上可以看出，采用 PCA 方法降维后，四种模型的预测精度都有提高 3%左右。在四种算法模型 XGBoost 和 LightGBM 的预测结果性能类似，达到 89%以上， 表明该两种算法模型对假人胸部预测的效果良好。

Table1.Algorithm performance comparison without/with optimization

表 1. 算法模型优化前后预测正确性对比

算法模型	基础模型预测精度	采用 PCA 后模型预测精度
BP+LSTM	76.1%	78.2%
SVR	83.1%	86.3%
XGBoost	86.0%	89.5%
LightGBM	86.5%	89.9%

4 结论

1) 通过数据治理，可以提高数据质量，这对于算法模型的预测性能至关重要。

2) 尝试使用 BP 神经网络+LSTM、SVR、XGBoost、LightGBM 等机器学习算法对安全带滑台实验中假人胸部位移进行预测，在数据量少、不均衡且存在噪音时，XGBoost 和 LightGBM 预测水平类似且预测效果良好。

3) PCA 降维和先验知识的应用可使算法模型优化，进一步提高算法预测精度。

4) 本文提出未来的研究方向，包括进一步采集数据优化算法模型，探索新的数据治理策略，以及将虚拟测评技术应用于更广泛的汽车安全评估场景。

参考文献 (References)

[1] 虚拟测评在车辆安全性能评价中的应用.汽车测试网, https://www.auto-testing.net/news/show-118365.html,2023-04-25.

[2] ZHU Bing, ZHANG Pei-xing, ZHAO Jian, CHEN Hong, XU Zhi-gang, ZHAO Xiang-mo, DENG Wei-wen,Review of Scenario-based Virtual Validation Methods for Automated Vehicles[J]. China Journal of Highway and Transport, 2019, 32(6): 1-2.

[3] 中国新车评价规程（C-NCAP）2024 版.

[4] HAN Fei-fei,虚拟测评被动安全未来发展路线新技术应用.汽车测试网, https://www.auto-testing.net/news/show-115959.html.

[5] GB14166-2013 《机动车乘员用安全带、约束系统、儿童约束系统和 ISOFIX 儿童约束系统》.

[6] 梅宏等.数据治理之论.中国人民大学出版社,2020.

[7] 翁翕.数据资产化的内涵、国际经验及政策建议.研究简报第 289 期.

[8] 数据治理. https://cloud.tencent.com/developer/techpedia/1545,2023-7-24.

[9] 什么是机器学习. https://blog.csdn.net/qq_36459893/article/details,2018-08-25.

[10] 周政,BP 神经网络的发展现状综述. CNKI:SUN:SXDS.0.2008-02-037.

[11] 长短时记忆网络（LSTM）完整实战：从理论到 PyTorch 实战演示.腾讯云, https://cloud.tencent.com/developer/article/2348486.

[12] ZHANG Fan, Lauren J. O'Donnell, Support vector regression. ScienceDiect, 2020.p.123-140.

[13] KANG Xiao-fei,ZENG Xuan, QIAO Wei, Indoor Positioning Algorithm Based on XGBoost Prediction and Elastic Net Error Compensation. System Simulation Journal,2022.34(4):p.719-726.

[14] 肖迁,穆云飞, 焦志鹏, 基于改进 LightGBM 的电动汽车电池剩余使用寿命在线预测. 电工技术学报, 2022.37(17):p.5-6.

[15] SUN Ping-an, WANG Bei-zhan, A Research on PCA Dimension Reduction with its Application in Machine Learning.Journal of Human University of Technology,2019.33(1):p.2-3.