

Crash Pulse Prediction based on Gaussian Process Regression and Principle Component Analysis

Jiawen LIANG, Ping JIANG, Jing LU

GAC Automotive research & development center, Guang zhou, 511434

Email: liangjiawen@gacrnd.com

Abstract: In order to provide the comprehensive and adaptive protection for the drivers and passengers under the development of intelligent and autonomous vehicle, it is essential to build up an accurate real-time occupant injury prediction system and form a collision mitigation strategy before the crash. Crash pulse prediction technique is capable to produce the crash pulse based on the near crash conditions, which can be used for occupant injury prediction or we straightforward construct a collision mitigation strategy based on. Unfortunately, because the acquisition of crash pulse with high quality is extremely expensive, we are available to obtain small amount of data for training. In this paper, we propose a novel crash pulse prediction method based on Gaussian Process Regression and Principle Component Analysis, which is suitable for the tasks with a small number of samples and takes the relatively low computation and storage cost. The statistics on the finite element simulation dataset show that this method can obtain satisfactory prediction.

Keywords: Crash pulse prediction; Gaussian process regression; Principle component analysis

基于高斯过程回归与主成分分析的碰撞波形预测技术

梁嘉文, 蒋 平, 卢 静

广州汽车集团股份有限公司汽车工程研究院, 广州, 511434

Email: liangjiawen@gacrnd.com

摘 要: 为了在汽车智能化发展的趋势下实现对驾驶员与乘客提供全面、自适应的保护, 需要建立准确实时乘员损伤预测系统, 并在碰撞前形成碰撞减轻策略。碰撞波形预测技术可以作为乘员损伤预测的前置过程, 或基于预测波形直接构建碰撞减轻策略。但由于高质量碰撞波形数据的获取成本较高, 因此可以用于模型训练的数据量较少。针对这个问题, 本文提出了一种基于高斯过程回归与主成分分析的碰撞波形预测技术, 它适用于样本较少的回归任务, 并使用较低的计算与存储资源。在有限元仿真数据集的测试结果显示, 该方法可以得到令人满意的预测结果。

关键词: 碰撞波形预测; 高斯过程回归; 主成分分析

1 引言

在汽车发展的电动化、智能化、网联化的趋势驱动下, 传统被动安全技术也在向智能化转型, 为自动驾驶汽车的驾驶员和乘客提供更全面、多场景、自适应的保护。为了实现该目标, 需要建立准确实时的临碰撞感知模型和乘员损伤预测技术, 并在碰撞前形成碰撞减轻策略。乘员损伤预测技术指基于智能车辆的车载感知系统, 提取与碰撞事故严重性相关的信息, 实现对于事故中舱内乘员损伤严重性的准确评估。其中对乘员损伤情况具有关键影响的变量可分为舱内和舱外两方面, 具体包括: 舱内信息——乘员的生理学特征(如年龄、性别、身高、体重等)、乘坐姿态(座椅朝向、椅背角度、身体姿态等)与约束系统信息(安全带、气囊等); 舱外信息——车辆物理参数(质量、尺寸、刚度等)、障碍物种类(汽车、行人、建筑物等)与碰撞速度与碰撞位置等。碰撞波形预测技术可以作为乘员损伤预测的前置过程, 利用舱外信息预测碰撞后舱内又惯性力产生的碰撞信号/波形, 结合碰撞信号与舱内信息可以对乘员不同部位产生的孙损伤进行预测。另外, 基于该碰撞波形也能直接构建碰撞减轻策略。

机器学习方法已经广泛应用在碰撞波形的预测任务上，为了建立一个可靠的波形预测模型，应尽量使用高质量的数据进行训练。最高质量的数据是实车碰撞数据，但获得实车碰撞数据的成本极高。质量稍低的是使用有限元仿真得到的碰撞数据，可以通过增加有限元网格数量以及缩短计算步长提升仿真的精度，但代价是消耗更多的计算资源。为了得到一例高质量的仿真数据，即使在高性能计算集群上进行仿真也需数小时之久。由于高质量碰撞波形数据的获取成本较高，可以用于模型训练的数据较少。针对上述问题，本专利使用碰撞角度、碰撞速度、重叠率作为输入，基于高斯过程回归（Gaussian Process Regression, GPR）以及主成分分析（Principle Component Analysis, PCA）建立波形预测模型。该模型可以用于多个方面：（1）利用预测的波形以及舱内信息，得到乘员伤害的预测；（2）利用预测的波形计算与伤害相关的指标，作为车辆路径规划问题中目标函数的一部分；（3）利用波形预测模型提升日常仿真分析的效率；（4）预测的波形可用于安全气囊点爆风险预估。

2 技术现状分析

机器学习方法已经广泛应用在碰撞波形的预测任务上。Huang^[1]提出了响应逆滤波（Response Inverse Filtering）的方法进行碰撞波形预测。Julian^[2]将高斯过程回归模型应用于碰撞波形预测任务上，在 100 例 LSDyna 生成的仿真数据上进行训练和测试，发现预测的波形与真实的波形的 OLC 差距比较小。Chen^[3]提出了时域增强网络（Temporal Augmentative Network, TAN）进行碰撞波形预测，该网络包含了嵌入层（Embedding layer）和残差结构，并使用 VCS 工具箱（Visual Crash Studio）中的 Camry 模型以不同的初始状态与刚性壁障碰撞进行仿真实验，采集 B 柱上 X 向的加速度生成了 5000 例数据。TAN 模型在仿真数据上的 CORA 值达到 0.8。Mira Arana^[4]对多种不同的机器学习模型进行了比较，包括神经网络、支持向量回归、梯度提升树、随机森林。数据集来自沃尔沃汽车安全中心的 LS_Dyna 仿真数据集，包含车辆碰撞后重心沿不同方向的加速度与角速率。Sequeira^[5]建立了一个由输入到输出的线性模型，并在包含有限元、弹簧质量模型、动量碰撞模型生成的共 917 例数据上进行仿真与测试。Höschele^[6]使用半正弦函数、小波作为固定的基函数，并认为碰撞加速度波形都能表达为基函数的线性组合。建立线性回归模型可以得到碰撞初始状态到基函数线性系数的映射，以此完成碰撞加速度的任务。

3 基于高斯过程回归与主成分分析的碰撞波形预测模型研究

建立碰撞波形预测模型最大的困难是缺少充足的高质量数据。最高质量的数据是实车碰撞数据，但获得实车碰撞数据的成本极高，一次高速的碰撞实验就会损耗掉一台样车。质量稍低的碰撞数据是使用有限元仿真得到的数据，通过增加有限元网格以及缩短计算步长，可以提升仿真的精度使得仿真数据更接近真实碰撞数据。但代价是消耗更多的计算资源，为了得到一例高质量的仿真数据，即使在高性能计算集群上进行仿真也需数小时之久。因此，[2]中的仿真也只使用了 100 例 LSDyna 的数据。质量更低的数据则由多刚体、宏单元以及弹簧质量等模型生成，通常一例数据生成的时间在数分钟到数秒钟不等，但仿真数据与实车数据的差异较大。由于仿真时长较短，[3]中使用了 5000 例由 VCS 生成的数据训练模型。

为了建立一个可靠的波形预测模型，应使用质量较高的数据进行训练，并避免使用精度较低的仿真数据。针对训练数据较少这个特点，本文提出了基于高斯过程回归与主成分分析的碰撞波形预测的模型。

3.1 高斯过程回归

高斯过程回归是一种典型的非参数模型，它与神经网络这种参数模型不同，它并不是在构建输入输出变量间的映射关系，而是在构建样本间的关系。如图 1 所示，高斯过程回归可以理解作为一种广义插值方法，但不同于使用固定参数的 K 近邻或径向基函数插值等算法，高斯过程回归中的模型参数可以从训练样本中学习。

高斯过程回归模型的灵活性较高，可以通过使用不同的核函数构建样本间的关系，本文使用带有自动相关性确定（Automatic relevance determination, ARD）机制的高斯核函数：

$$k(\mathbf{x}, \mathbf{x}') = \theta_0 \exp\left(-\frac{1}{2} \sum_{i=1}^M \eta_i (x_i - x'_i)^2\right) \quad (1)$$

其中 $k(\cdot, \cdot)$ 为核函数, $\mathbf{x}, \mathbf{x}' \in \mathbb{R}^D$ 为表征碰撞状态的向量, 包括碰撞速度、角度、重叠率等, D 表示状态的数量, 也即状态向量的维度。 $k(\mathbf{x}, \mathbf{x}')$ 可以理解为样本向量 \mathbf{x} 与 \mathbf{x}' 的协方差, θ_0 与 $\{\eta_i\}_{i=1}^M$ 为待定参数。若有训练样本 $(\mathbf{x}_1, t_1), \dots, (\mathbf{x}_N, t_N)$, 则待优化的目标函数为对数边缘似然函数:

$$\log p(\mathbf{t}|\boldsymbol{\theta}) = -\frac{1}{2} \log |\mathbf{C}_N| - \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \mathbf{t} - \frac{N}{2} \log 2\pi \quad (2)$$

其中 \mathbf{C}_N 表示样本间的协方差矩阵, 协方差矩阵的元素为 $C(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \beta^{-1} \delta_{nm}$, δ_{nm} 为示性函数, 只有当 $m = n$ 时取值为1, 否则为0。该目标函数可以通过基于梯度的优化方式进行优化, 常见的优化方式为拟牛顿法, 如BFGS方法等。因此需要计算目标函数对优化参数的偏导数:

$$\frac{\partial}{\partial \theta_i} \log p(\mathbf{t}|\boldsymbol{\theta}) = -\frac{1}{2} \text{Tr} \left(\mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \right) + \frac{1}{2} \mathbf{t}^T \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \mathbf{C}_N^{-1} \mathbf{t} \quad (3)$$

其中 θ_i 表示核函数中的待定参数, 通过优化待定参数来提升边缘似然函数, 直到目标函数收敛。高斯过程回归适用于训练样本较少的回归任务, 训练过程稳定性高, 并且可以通过构造特征来融入工程师的先验知识。

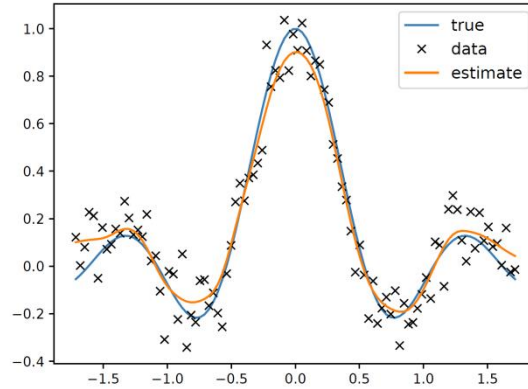


图1 高斯过程回归

3.2 主成分分析

虽然高斯过程针对样本较少的任务具有稳定性高、过拟合风险小等优点, 但由于它是一种非参数模型, 在进行预测时需要调用所有的训练样本。由于碰撞加速度数据的时间维度都较高, 这导致每次预测时都会耗费较多的内存资源, 同时也较难应用于实时预测。针对时效性较高的任务, 需要考虑使用计算资源与存储资源消耗较低的方案。主成分分析 (Principle Component Analysis, PCA) 是一种广泛使用的用于数据压缩的算法, 将数据分解为主成分的线性组合, 则在预测时只需要将固定数量的主成分放到内存中, 而不需要将所有训练数据都放入内存, 这样可以同时节省计算资源和存储资源。如图2所示, 可以将数据投影到主成分的向量上并保留原数据尽量多的特征。

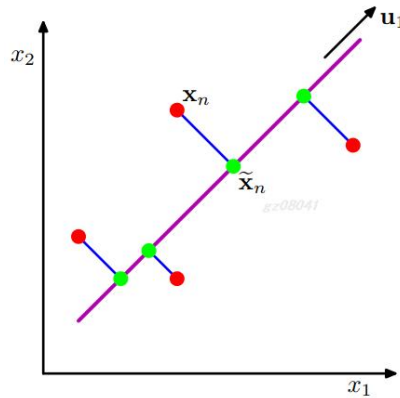


图2 主成分分析示意图

主成分分析可以通过奇异值分解实现，现有波形数据矩阵 $\mathbf{S} \in \mathbb{R}^{N \times T}$ ，其中 N 表示样本的数量， T 表示时间维度。对数据中心化后再进行奇异值分解得：

$$\tilde{\mathbf{S}} = \mathbf{U}\mathbf{D}\mathbf{V}^T \quad (4)$$

其中 $\tilde{\mathbf{S}}$ 表示中心化后的数据矩阵， $\mathbf{V} \in \mathbb{R}^{T \times T}$ 为主成分向量矩阵，每一列表示一个主成分， \mathbf{D} 为对角阵，对角元素表示主成分方差（Variance），方差越大表示对应的主成分越能表征数据的特征。从 \mathbf{V} 中筛选出固定数量的主成分向量 $\tilde{\mathbf{V}} \in \mathbb{R}^{T \times K}$ ， K 表示筛选后的主成分数量。可以由工程师指定或者基于一定的准则进行筛选。基于 $\tilde{\mathbf{V}}$ ，数据矩阵 $\tilde{\mathbf{S}}$ 可以分解为：

$$\tilde{\mathbf{S}} \approx \mathbf{W}\tilde{\mathbf{V}}^T \quad (5)$$

其中 $\mathbf{W} \in \mathbb{R}^{N \times K}$ 表示不同波形样本在主成分上的分量，上式表示利用主成分 $\tilde{\mathbf{V}}$ 以及对应的分量矩阵 \mathbf{W} 可以以一定的精度重构数据矩阵 $\tilde{\mathbf{S}}$ 。如图 3 所示，我们可以将预测的目标转化为主成分上的分量 \mathbf{W} 而非波形本身 $\tilde{\mathbf{S}}$ ，并且在预测时并不需要调用所有的训练数据，而是使用固定数量的主成分。当 $K \ll D$ 时，可以大大节省计算与存储资源。

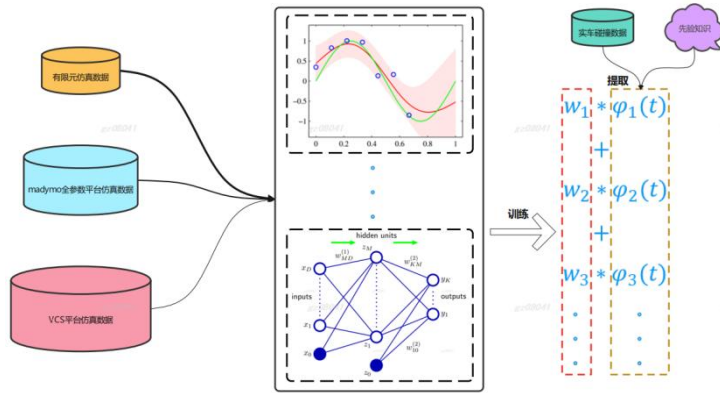


图 3 高斯过程回归+主成分分析示意图

3.3 算法流程

假设有训练数据 $\{\mathbf{x}_n, \mathbf{S}_n\}_{n=1}^N$ ，其中 \mathbf{x}_n 表示样本 n 的特征，包括速度、重叠率等， \mathbf{S}_n 表示对应的波形。以下为训练流程：

- (1) 计算所有碰撞波形的均值 $\bar{\mathbf{S}} = \frac{1}{N} \sum_{n=1}^N \mathbf{S}_n \in \mathbb{R}^T$ ；
- (2) 使用式(4)所示的主成分分析对波形 $\{\mathbf{S}_n\}_{n=1}^N$ 提取主成分 $\{\mathbf{v}_i\}_{i=1}^K$ ，其中 $\tilde{\mathbf{S}} = [\mathbf{S}_1 - \bar{\mathbf{S}}, \dots, \mathbf{S}_N - \bar{\mathbf{S}}]^T \in \mathbb{R}^{N \times T}$ ，表示已经中心化的碰撞波形矩阵； $\tilde{\mathbf{V}} = [\mathbf{v}_1, \dots, \mathbf{v}_K] \in \mathbb{R}^{T \times K}$ ，表示筛选后的主成分向量。
- (3) 使用 $\mathbf{W} = \tilde{\mathbf{S}}\tilde{\mathbf{V}}(\tilde{\mathbf{V}}^T\tilde{\mathbf{V}})^{-1}$ 计算所有训练样本在不同主成分上的分量系数，其中分量矩阵按行列可以分解如下：

$$\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_N]^T = [\boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_K] \in \mathbb{R}^{N \times K}$$

$\mathbf{w}_n \in \mathbb{R}^K$ 表示第 n 个样本在所有主成分上的分量， $\boldsymbol{\omega}_k \in \mathbb{R}^N$ 表示所有训练样本在第 k 个主成分上的分量；

- (4) 基于式(1)计算不同训练样本间的协方差：

$$\mathbf{C}(\mathbf{x}_n, \mathbf{x}_m) = k(\mathbf{x}_n, \mathbf{x}_m) + \beta^{-1} \delta_{nm} \quad (5)$$

其中 δ_{nm} 为示性函数，当且仅当 $m = n$ 时取值为1，否则为0。记协方差矩阵 \mathbf{C}_N 为 N 行 N 列对称矩阵，第 m 行 n 列的元素为 $\mathbf{C}(\mathbf{x}_n, \mathbf{x}_m)$ 。

- (5) 根据协方差矩阵 \mathbf{C}_N 与 $\{\mathbf{w}_n\}_{n=1}^N$ 计算目标函数 \mathcal{L} ：

$$\mathcal{L} = \log p(\mathbf{W}|\boldsymbol{\theta}) = \sum_{k=1}^K \log p(\mathbf{W}|\boldsymbol{\theta}) = -\frac{1}{2} (K \log |\mathbf{C}_N| + \sum_{k=1}^K \boldsymbol{\omega}_k^T \mathbf{C}_N^{-1} \boldsymbol{\omega}_k) \quad (6)$$

其中 $\boldsymbol{\omega}_k$ 为 \mathbf{W} 的第 k 个列向量， $\boldsymbol{\theta}$ 泛指所有待定参数；

- (6) 计算目标函数 \mathcal{L} 对待定参数 θ_i 的偏导数：

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = \frac{\partial}{\partial \theta_i} \log p(\mathbf{W}|\boldsymbol{\theta}) = -\frac{K}{2} \text{Tr} \left(\mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \right) + \frac{1}{2} \sum_{k=1}^K \boldsymbol{\omega}_k^T \mathbf{C}_N^{-1} \frac{\partial \mathbf{C}_N}{\partial \theta_i} \mathbf{C}_N^{-1} \boldsymbol{\omega}_k \quad (7)$$

其中 θ_i 泛指所有待定模型参数，梯度 $\nabla \mathcal{L} = [\frac{\partial \mathcal{L}}{\partial \theta_1}, \dots, \frac{\partial \mathcal{L}}{\partial \theta_M}]^T$ 由对所有的待定参数的偏导数构成；

(7) 根据梯度 $\nabla \mathcal{L}$ ，使用BFGS方法更新待定参数 θ_i ，其中BFGS是一种基于梯度的牛顿法，广泛应用于求解优化问题。；

(8) 重复(4)，(5)，(6)和(7)，直到目标函数 \mathcal{L} 的相对变化小于给定阈值(如 $1e-6$)；

当模型已经训练完成，现有预测数据 $\tilde{\mathbf{x}}$ ，则使用如下方法进行预测：

(1) 使用完成训练的模型参数 θ_0 ， $\{\eta_i\}_{i=1}^D$ ， β ，使用式(5)计算训练数据与 $\{\mathbf{x}_i\}_{i=1}^N$ 与 $\tilde{\mathbf{x}}$ 的协方差矩阵，并记列向量 $\mathbf{k} = [k(\tilde{\mathbf{x}}, \mathbf{x}_1), \dots, k(\tilde{\mathbf{x}}, \mathbf{x}_N)]^T$ ；

(2) 预测的波形 $\hat{\mathbf{S}}$ 可以表示为：

$$\hat{\mathbf{S}} = \mathbf{k}^T \mathbf{C}_N^{-1} \mathbf{W} \tilde{\mathbf{V}} + \bar{\mathbf{S}} \quad (7)$$

完成预测过程。

4 实验分析

4.1 实验流程说明

本文进行两个仿真实验：第一个仿真实验为 4500 例 VCS 仿真数据集，使用模型内置的 Camry 模型与刚性壁障进行碰撞，仿真变量包括碰撞速度、角度、重叠率与摩擦系数，如图 4 所示。碰撞波形为 B 柱 X 向的加速度，从碰撞时刻及之后 150ms 内的加速度，共 150 个数据点；第一个仿真实验使用 100 例有限元仿真数据集进行测试，仿真工况为研发车型与静止的 MPDB 壁障进行碰撞。仿真变量包括碰撞速度、角度和重叠率。碰撞波形使用左侧 B 柱上 X、Y 方向上从碰撞时刻及之后 130ms 内的加速度，共 1300 个数据点。

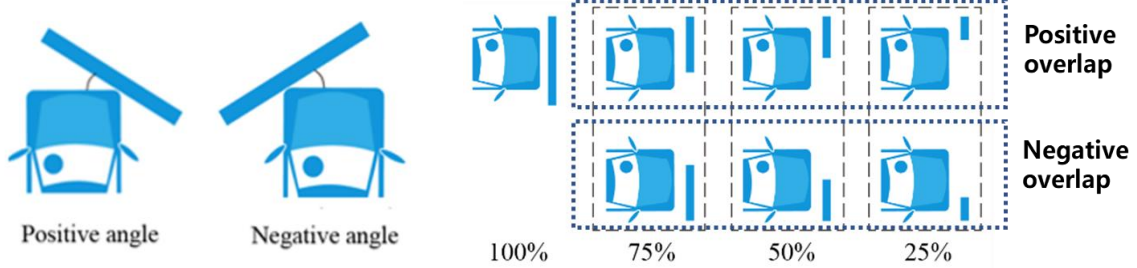


图 4 碰撞角度、重叠率示意图

另外，本文使用 10 折交叉验证来对模型进行测试与验证。具体地，将所有的仿真数据分为 10 组，随机选取其定数量的分组构成训练集，剩余的分组成测试集，对不同的数量重复进行实验可以测试数据量大小对模型性的影响。图 5 表示交叉验证的过程，前者表示使用较多的样本进行训练，后者表示使用较少的样本进行训练。使用乘员载荷准则（Occupant Load Criterion, OLC）评价模型对 X 向加速度波形的预测精度，并使用波形的均方误差（Root Mean Square Error, RMSE）对评价 Y 向的加速度进行评价。

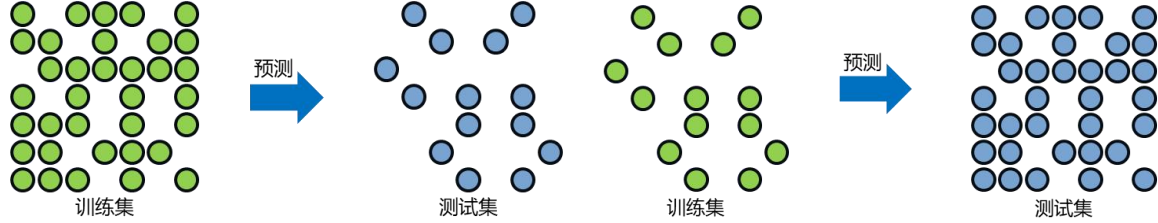


图 5 交叉检验示意图

4.2 实验结果分析

对第一个仿真实验，使用主成分分析从 B 柱 X 方向加速度数据集中提取出来的前三个主成分，结果如图 6 所示。可以发现这些波形能反映加速度变化的不同趋势，波峰的数量、出现的时间各不相同。

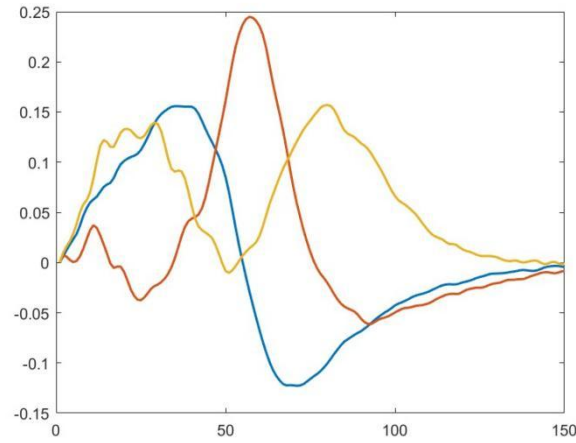


图 6 主成分分析提取的主成分

本文使用[3]中的深度神经网络与本文提出的高斯过程回归模型进行对比，仿真结果如图 7 所示，图中统计了碰撞加速度与模型预测加速度的 OLC 值的平均绝对误差（Mean Absolute Error, MAE），横轴表示使用不同比例的数据作为训练数据，从左到右表示使用越来越多的训练数据。从图中可以看出，随着训练样本的增加，高斯过程回归模型与神经网络预测波形的 OLC 与真实波形的差距都越来越小，显示模型的预测精度越来越高。在每一组的比较中，GP 模型的 OLC-MAE 都比 NN 模型的更小，显示预测的波形与仿真波形的拟合程度更高。另外，当训练样本较少时，只使用 20%的数据进行训练时，GP 能保持较好的预测效果，OLC-MAE 的平均值约为 1，而深度神经网络的 OLC-MAE 的平均值约为 5。图 8 是两例仿真波形与预测波形的对比图，同样可以发现随着训练样本的增加，模型预测的波形与仿真波形的拟合程度越来越高，且具有一定的平滑效果。

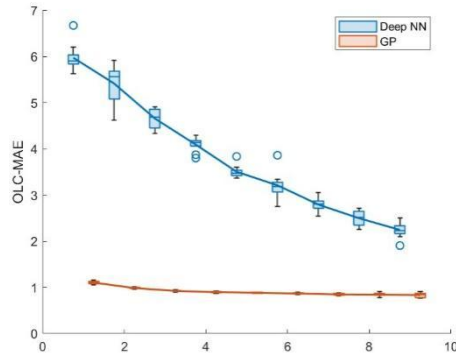


图 7 GP 模型与深度神经网络模型的 OLC-MAE 统计结果对比

图 9 是在第二个仿真实验上，GP 模型对 B 柱 X 方向加速度数据的测试结果，图中统计了碰撞加速度与模型预测加速度的 OLC 值的平均绝对误差。从图 9 中可以看出，随着数据量的增多，MAE 数值越来越低，表示模型的预测效果越来越好。当使用 40 例样本进行训练时，MAE 可以下降到 1.5 左右，达到令人满意的效果。图 10 是仿真波形与预测波形的对比图，同样可以发现随着训练样本的增加，模型预测的碰撞曲线与仿真曲线的贴合程度越来越高，但预测波形会丢失一些波峰。

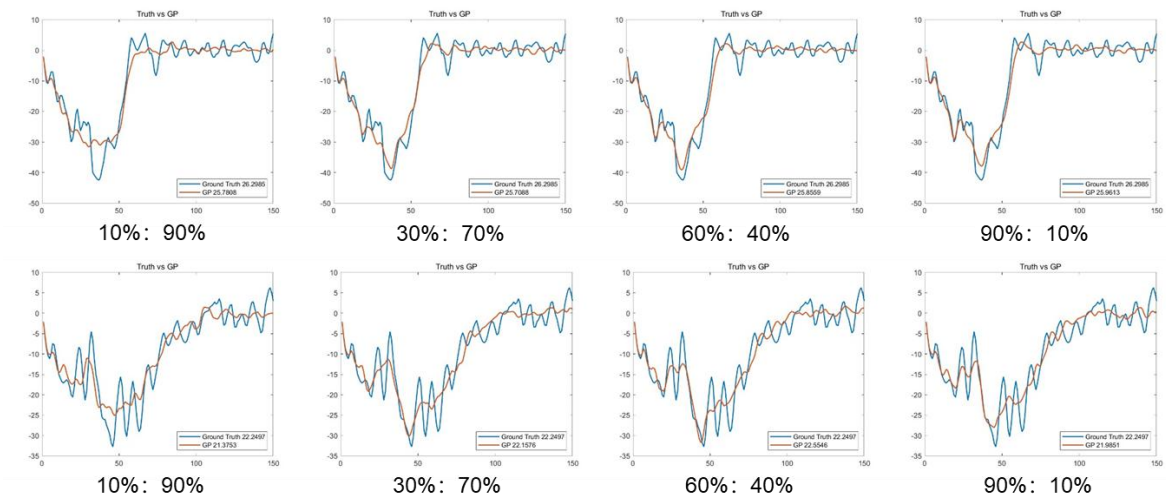


图 8 GP 模型预测波形与仿真波形对比图

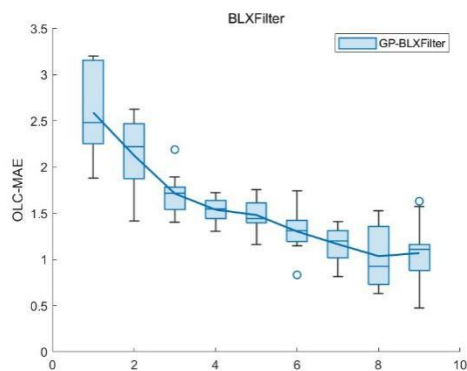


图 9 OLC-MAE 统计结果

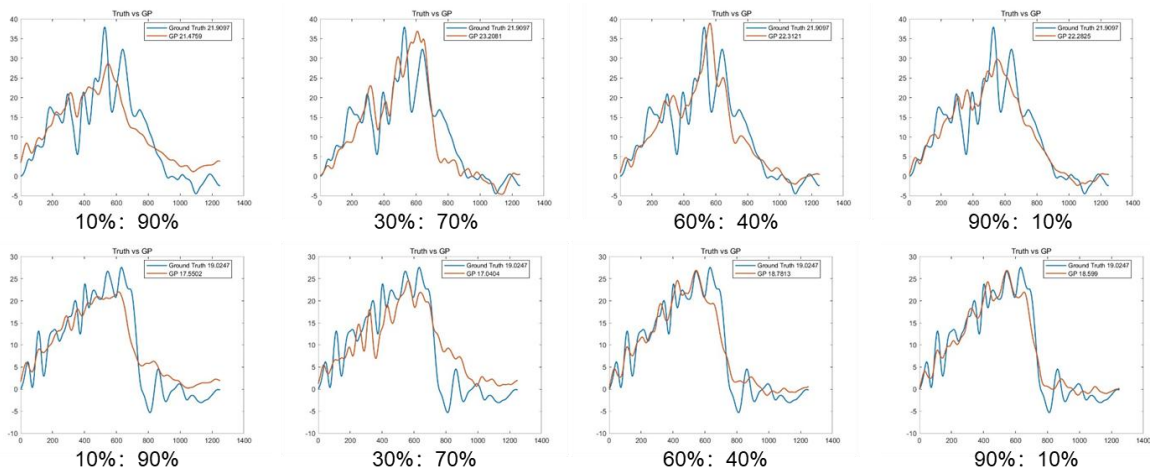


图10 GP模型预测波形与仿真波形对比图

本文还对比了模型对 X、Y 不同方向的预测性能的差异，由于 OLC 不再是评价 Y 方向上人体损伤的主要指标，此处只使用均方误差作为波形评价指标。从图 11 可以看出，由于仿真工况为正碰，Y 方向上的碰撞信号包含更多的噪声与扰动，模型对 Y 方向加速度数据的预测精度不如 X 方向的加速度。

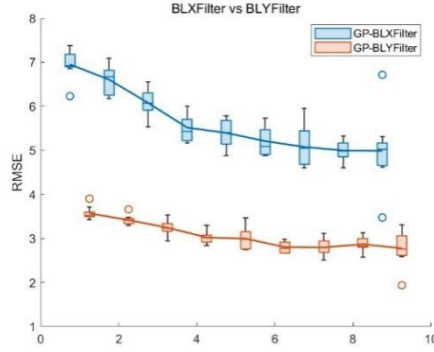


图 11 GP 模型对 X、Y 方向信号预测结果对比

同时，本文还对比了使用 PCA 与不使用 PCA 的性能差异。图 12 为 GP 模型与 GP+PCA 模型在不同训练数据数量下，预测波形与仿真波形的线性相关系数的结果对比。从图中可以看出，当只使用 10 个样本进行训练时，GP+PCA 的预测结果比单独使用 GP 的结果更好。但随着训练样本的增加，单独使用 GP 的性能会提升得更快并超过 PCA，两者的差异较小。另一方面，GP+PCA 的优势在于占用更少的计算资源与存储资源，可以实现更快的响应。如图 13 所示，在 Matlab 上使用 GP+PCA 模型进行预测的平均时间约为 0.4ms，而 GP 模型进行预测时则会消耗更多的时间，且随着训练样本数量的增加而提升。

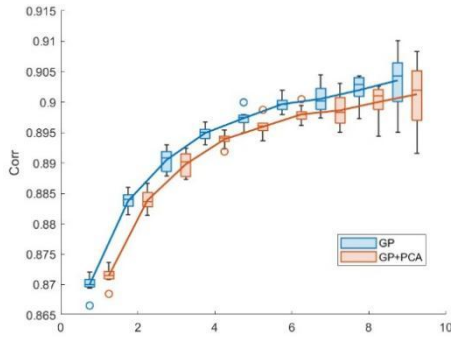


图 12 GP+PAC 与 GP 预测结果对比

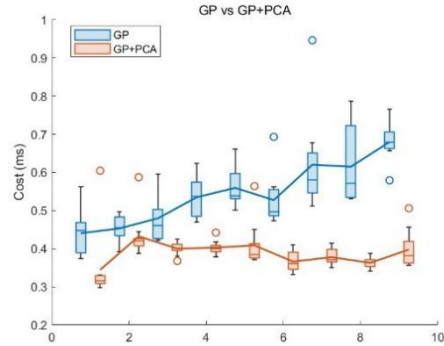


图 13 GP+PAC 与 GP 时间消耗对比

5 结论

本文提出了一种基于高斯过程回归与主成分分析的碰撞波形预测模型，从实验结果中可以发现：

- (1) 高斯过程回归的拟合效果比神经网络更好。
- (2) 训练样本较少时，高斯过程回归能保持较好的预测效果
- (3) 训练样本增多时，神经网络的预测精度会显著提升
- (4) 重构的波形会丢失某些波峰（较难拟合高频信号）
- (5) OLC 拟合程度较高（对低频信号拟合效果较好）

但该模型仍存在如下的不足：（1）当前 FE 仿真数据较少，高斯过程（GP）是样本不充足时较好的解决方案，当有更多数据时可以考虑使用神经网络模型；（2）未使用实车碰撞数据进行训练和测试（可将部分工况的仿真数据替换为实车碰撞数据进行训练）；（3）当前提取特征的方式简单；（4）使用的变量较少，针对单一车型。因此模型可以从如下几个方向进行改进：（1）优化数据预处理、特征提取的方式；（2）优化模型、算法；（3）找到具有代表性的工况，使用这些工况的数据构建训练集，而不是随机生成样本进行训练；（4）研究模型对不同工况的预测效果的差异；（5）研究某个已经训练好的模型能否迁移到另一个车型上。

参考文献

- [1] Huang, Matthew and Raj Jayachandran. Crash pulse prediction via inverse filtering. SAE Technical Paper, 2001, No. 2001-01-3110.
- [2] Hay, Julian, Jörg Fehr and Lars Schories. Crash Pulse Prediction for Scenario-Based Vehicle Crash FE-Simulations. International Research Council on Biomechanics of Injury (IRCOBI) Conference 2020, 2020.
- [3] Chen W, Wang Q, Wang Y, et al. A Generation Model of Vehicle Crash Pulse for Injury Severity Prediction Based on a Typical Sedan. International Research Council on Biomechanics of Injury (IRCOBI) Conference 2021, 2021.
- [4] Mira Arana, P. Crash Pulse Prediction applied to Frontal Crash Configurations. 2021.
- [5] Sequeira, Gerald Joy Alphonso, et al. Crash pulse prediction using regression algorithm with gradient descent optimization method for integrated safety systems. SAE International journal of transportation safety 10.2 (2022): 163-184.
- [6] Höschele, Patrick, et al. Generic crash pulses representing future accident scenarios of highly automated vehicles. SAE International journal of transportation safety 10.2 (2022): 185-210.