

# 面向抗倒塌地震动强度指标选取的特征 选择算法性能评估

胡进军<sup>1,2</sup>, 刘亦恒<sup>1,2</sup>, 刘巴黎<sup>3</sup>

(1. 中国地震局工程力学研究所 地震工程与工程振动重点实验室, 黑龙江 哈尔滨 150080; 2. 地震灾害防治应急管理部重点实验室, 黑龙江 哈尔滨 150080; 3. 湖南科技大学 土木工程学院, 湖南 湘潭 411201)

**摘要:**为了筛选有效预测结构倒塌能力的地震动强度指标,对比分析了 MIC、ReliefF、XGBoost 和 Lasso 这 4 种常见特征选择算法用于地震动强度指标筛选时的性能。基于单自由度结构增量动力分析结果及地震动强度指标建立特征选择回归模型,根据回归模型输出权重及频数得到欧氏距离大小排序并筛选地震动强度指标,利用筛选结果对特征选择算法的性能进行评估。同时基于 2 层、4 层、8 层和 12 层钢筋混凝土框架结构的增量动力分析结果对筛选后强度指标建立最小二乘回归模型,以残差的标准差变化衡量不同特征选择算法筛选出的地震动强度指标对结构倒塌的预测能力。结果表明:基于 Lasso 回归算法筛选的地震动强度指标比其他算法用于结构倒塌预测时准确率提高 31%。结果可为基于性能地震工程(performance-based earthquake engineering, PBEE)框架下结构易损性分析中及地震动不确定性分析中地震动强度指标筛选的特征选择算法提供参考,也可作为结构倒塌预测的地震动强度指标筛选提供有效特征选择算法参考。

**关键词:**地震动强度指标;特征选择算法;结构倒塌分析;增量动力分析;最小二乘回归

中图分类号:P315.9

文献标识码:A

## Performance evaluation of feature selection algorithm for selection of collapse estimated ground motion intensity measures

HU Jinjun<sup>1,2</sup>, LIU Yiheng<sup>1,2</sup>, LIU Bali<sup>3</sup>

(1. Key Laboratory of Earthquake Engineering and Engineering Vibration, Institute of Engineering Mechanics, China Earthquake Administration, Harbin 150080, China; 2. Key Laboratory of Earthquake Disaster Mitigation, Ministry of Emergency Management, Harbin 150080, China; 3. School of Civil Engineering, Hunan University of Science and Technology, Xiangtan 411201, China)

**Abstract:** To identify an efficient and accurate feature selection algorithm for filtering seismic intensity indicators, the performance of four common feature selection algorithms, MIC, ReliefF, XGBoost and Lasso, was compared and analyzed. Based on the incremental dynamic analysis results of single-degree-of-freedom structures and the ground motion features, the feature selection regression model was established, the ground motion features was sorted and screened according to the Euclidean distance, the performance of the feature selection algorithm was evaluated according to the screening results, and the least squares regression model was established based on the incremental dynamic analysis results of the 2-storey, 4-storey, 8-storey and 12-storey reinforced concrete frame structures, and the standard deviation change of residual was used to measure the prediction ability of ground motion intensity measure selected by different feature selection algorithms for structural collapse. The results show that the accuracy of the ground motion features screened by the Lasso regression algorithm is 31% higher than that of other algorithms when used for structural collapse prediction. The results can be used as a feature selection algorithm reference for the selection of ground motion intensity measures in the uncertainty analysis of ground motion

收稿日期:2023-11-07; 修回日期:2023-12-27

基金项目:国家自然科学基金面上项目(52478568); 国家自然科学基金青年项目(52408529)

作者简介:胡进军(1978—),男,研究员,博士,主要从事地震动特征分析和模拟方面的研究。E-mail:hu-jinjun@163.com

in the structural vulnerability analysis under the performance-based earthquake engineering (PBEE) framework, and can also be used as an effective feature selection algorithm reference for the selection of ground motion intensity measure s suitable for structural collapse prediction.

**Key words:** ground motion intensity measure; feature selection algorithm; structure collapse analysis; incremental dynamic analysis; least squares regression

## 0 引言

作为表征地震动破坏势的参数以及联系地震危险性和结构地震响应的桥梁,地震动强度指标(intensity measure, IM)自从以地震动峰值加速度  $A_{pg}$  为代表的第1个参数提出以来受到了研究人员的广泛关注。地震动 IM 的形式主要有3种<sup>[1]</sup>:①以  $A_{pg}$  为代表的峰值参数;②以地震动加速度反应谱  $S_a$  为代表的反应谱参数;③以滤波增量速度  $V_{\dot{u}}$ <sup>[2]</sup> 为代表的地震动时程参数。目前研究中地震动 IM 已经超过60种,从大量的地震动 IM 中筛选出反映地震动破坏强度的地震动 IM,不仅有助于预测地震动破坏强度,更有助于进一步减少结构性态评估中的地震动不确定性。

基于结构响应与地震动 IM 的相关性评价方法是筛选地震动 IM 的主要思路。韩建平<sup>[3]</sup>采用线性拟合的方法,研究不同周期单自由度(single-degree of freedom, SDOF)体系时程分析最大响应与地震动 IM 的相关性,给出不同周期结构应选取的 IM 建议。陈健云<sup>[4]</sup>对3栋不同的周期框架结构进行倒塌分析,并选取典型的整体和局部响应参数与加速度型、速度型和位移型 IM 采用线性拟合进行相关性研究,给出短、中、常3种周期框架结构分别与加速度型、速度型和位移型 IM 相关性好的结论。李雪红<sup>[5]</sup>对 SDOF 结构及减隔震桥梁系统,采用结构最大地震响应与地震动 IM 之间的线性相关性的方法对 IM 的敏感性进行评价,得出对于给定桥梁结构建议采用  $S_a(T_1)$  强度指标作为衡量地震动强度的指标参数的结论。李爽<sup>[6]</sup>采用皮尔森相关系数对 IM 与结构整体破坏指数进行相关性研究,得出近场地震动参数之间的相关性普遍弱于远场的结论,并给出了用来估计结构破坏程度的 IM。胡进军<sup>[7]</sup>采用皮尔森相关系数方法研究了不同场地条件和断层距下结构响应参数与地震动参数的相关性变化,得出研究结构响应与地震动参数相关性时需要考虑场地条件与断层距的影响。采用基于相关性的方法探究了结构损伤指标(damage measure, DM)与 IM 之间的关系,虽然机制简明,可操作性强,但地震动的随机性以及结构系统的复杂特性使得地震作用下结构响应与地震动参数之间呈现复杂的非线性关系<sup>[8-9]</sup>,基于线性相关性评价方法筛选地震动参数存在一定的局限性。

目前,机器学习中的特征选择算法已经在各领域得到了广泛的应用。特征选择是指从输入特征集中选择对机器学习分类或回归算法有益的最优特征子集<sup>[10]</sup>。近年来,基于特征选择算法的地震动 IM 筛选得到了研究人员的关注。刘巴黎<sup>[11]</sup>采用弹性网络回归方法建立大量 SDOF 模型与 IM 之间的模型,给出了地震动 IM 的排序和比选。吴梓楠<sup>[11]</sup>使用极端随机森林算法开展敏感性研究评估了地震动参数对结构损伤的影响程度。XU 等<sup>[12]</sup>基于支持向量机、逻辑回归以及决策树3种机器学习算法对多个地震动参数进行了比选。上述相关研究多基于单一的特征选择算法,基于多种特征选择算法的地震动 IM 筛选对比分析研究较少。此外,相关研究也未见应用于结构倒塌预测地震动 IM 的比选,合理选择地震动 IM 有助于减小结构抗地震倒塌能力分析的不确定性<sup>[13-14]</sup>。

本文利用基于特征选择算法进行 IM 筛选的结构倒塌预测,采用4种特征选择算法建立回归模型,得到 IM 敏感性和频数分析结果并进行排序,利用 SDOF 体系及钢筋混凝土(reinforced concrete, RC)框架结构增量动力分析方法(incremental dynamic analysis, IDA)结果对经过排序的 IM 进行最小二乘回归分析,比较排序及最小二乘回归分析结果,给出基于地震动 IM 预测结构倒塌的最有效特征选择算法,以期能为基于结构倒塌的地震动不确定性分析提供参考。

## 1 常见特征选择算法介绍

特征选择算法主要分为以下几类:过滤法(filter method)、包装法(wrapper method)和嵌入法(embedded method)<sup>[15]</sup>。过滤法是一种独立于分类器的特征选择算法,它通过一定的标准对特征进行排序,根据排序结

果来选择最佳特征子集<sup>[16]</sup>,过滤法计算效率高。常用的过滤法包括 ReliefF 及其变体、基于互信息的方法和 Fisher 评分等。包装法将特征选择看作是一个搜索问题,这种方法通过生成不同的特征子集,使用分类器性能作为特征子集的评估准则,使用搜索策略来搜索可能的特征子集空间,并通过机器学习算法的表现优异程度来评估每个子集,从而选择最佳的特征子集。包装法具有计算成本高、需要自定义停止标准(通常达到以下 3 个训练情况:性能提升、性能降低和达到预定义的特征数量)及训练结果偏向于预定义分类器<sup>[17]</sup>等特性。常用的包装法包括顺序搜索(如正向选择、反向消去和穷举搜索等),随机搜索(如进化计算算法等)。嵌入式算法结合了过滤法和包装法,将特征选择与分类器训练过程融合在一起,通过在分类器训练过程中调整特征的权重或选择特定的特征子集来进行特征选择<sup>[18]</sup>。嵌入法具有直接嵌入模型训练过程,提高算法效率、自动进行选择特征、模型与特征关联性强以及可能无法处理高度相关的特征等特性。常见的嵌入法包括正则化算法(L1 正则化、L2 正则化、弹性网络等)、主成分分析及基于树的方法(决策树、随机森林和梯度提升树等)。

由于包装法自身计算成本高同时普适性差的劣势,使得其在地震动 IM 比选研究中较少被应用,本文主要选取 4 种特征选择算法进行研究:MIC 算法(过滤法)、ReliefF 算法(过滤法)、XGBoost 算法和 Lasso 回归算法。4 种算法的基本原理和特点介绍见表 1。表 2 给出不同特征选择算法关键参数及其设定方法。

表 1 特征算法原理及特点

Table 1 Feature selection algorithm principle and characteristics

特征选择算法	基本原理	特点
MIC <sup>[19]</sup> (Maximal Information Coefficient)	将任意两特征组成的散点图的数据点分割开,通过穷举所有的网格的方法计算 MIC	算法稳定,但非线性分析效果较差
ReliefF <sup>[20]</sup>	计算各个特征与类的关联性,赋予相应的权值,选取权值较大的特征	算法简便效果好,但不能剔除无关特征
XGBoost(eXtreme Gradient Boosting) <sup>[21]</sup>	构建多个决策树并计算特征重要性得分来确定特征的重要程度	准确性高但消耗大量计算资源
Lasso (Least absolute shrinkage and selection operator) <sup>[22]</sup>	回归模型引入 L1 正则化项构造惩罚函数逐步减小无关特征权重至 0,实现对特征的筛选	能剔除无关特征提高筛选准确率,模型复杂度低

表 2 不同特征选择算法关键参数与设定

Table 2 Key parameters and settings of different feature selection algorithm

特征选择算法	关键参数	参数设定方法
MIC <sup>[19]</sup> (Maximal Information Coefficient)	划分网格数的上限 $B(n)$	实证检验 <sup>[19]</sup> $B(n) = n^\alpha$ , 参数 $\alpha$ 由样本量决定
ReliefF <sup>[20]</sup>	迭代次数 $m$ , 近邻数 $k$	实证检验 <sup>[20]</sup> , $m$ 及 $k$ 由样本量确定
XGBoost(eXtreme Gradient Boosting) <sup>[21]</sup>	迭代次数 $n$ , 最大深度 $\max\_depth$ , 学习率 $\eta$ , 子节点二阶导数和 $\min\_child\_weight$	$n$ 通过改进学习曲线方法确定; $\max\_depth$ 通过交叉验证调整默认值确定; $\eta$ 通常取默认值; $\min\_child\_weight$ 通常在默认值附近调整
Lasso (Least absolute shrinkage and selection operator) <sup>[22]</sup>	正则化参数 $\lambda$	交叉验证方法确定 $\lambda$ 值,使 Lasso 回归模型的均方误差最小

## 2 地震动记录及地震动 IM 选取

### 2.1 地震动记录选取

本文选取 BAKER 等<sup>[23]</sup>通过匹配目标谱从 PEER NGA-West2 地震动数据库中挑选出的 Set #1A 地震动集,Set #1A 地震动集由 40 组水平双向地震动分量组成,以设定地震事件矩震级  $M_w$  为 7.0,断层距  $R$  为 10 km,30 m 土层剪切波速  $V_{s30}$  为 250 m/s 与目标加速度反应谱进行匹配。地震动基本信息见表 3。

表 3 地震事件及地震动个数

Table 3 Earthquake events and the number of ground motion records

地震事件编号	NGA 序号	地震事件	日期	震级 $M_w$	地震动条数/条
1	231	Mammoth Lakes-01	1980-05-25	6.1	2
2	1 203	Chi-Chi	1999-09-20	7.6	2
3	829	Cape Mendocino	1992-04-25	7.0	2
4	169	Imperial Valley-06	1979-10-15	6.5	2
5	1 176	Kocaeli	1999-08-17	7.5	2

续表

地震事件编号	NGA 序号	地震事件	日期	震级 $M_w$	地震动条数/条
6	163	Imperial Valley-06	1979-10-15	6.5	2
7	1201	Chi-Chi	1999-09-20	7.6	2
8	1402	Chi-Chi	1999-09-20	7.6	2
9	1158	Kocaeli	1999-08-17	7.5	2
10	281	Trinidad	1980-11-08	7.2	2
11	730	Spitak	1988-12-07	6.8	2
12	768	Loma Prieta	1989-10-18	6.9	2
13	1499	Chi-Chi	1999-09-20	7.6	2
14	266	Victoria	1980-06-09	6.3	2
15	761	Loma Prieta	1989-10-18	6.9	2
16	558	Chalfant Valley-02	1986-07-21	6.2	2
17	1543	Chi-Chi	1999-09-20	7.6	2
18	2114	Denali	2002-11-03	7.9	2
19	179	Imperial Valley-06	1979-10-15	6.5	2
20	931	Big Bear-01	1992-06-28	6.5	2
21	900	Landers	1992-06-28	7.3	2
22	1084	Northridge-01	1994-01-17	6.7	2
23	68	San Fernando	1971-02-09	6.6	2
24	527	N. Palm Springs	1986-07-08	6.1	2
25	776	Loma Prieta	1989-10-18	6.9	2
26	1495	Chi-Chi	1999-09-20	7.6	2
27	1194	Chi-Chi	1999-09-20	7.6	2
28	161	Imperial Valley-06	1979-10-15	6.5	2
29	1236	Chi-Chi	1999-09-20	7.6	2
30	1605	Duzce	1999-11-12	7.1	2
31	1500	Chi-Chi	1999-09-20	7.6	2
32	802	Loma Prieta	1989-10-18	6.9	2
33	6	Imperial Valley-02	1940-05-19	7.0	2
34	2656	Chi-Chi	1999-09-20	6.2	2
35	982	Northridge-01	1994-01-17	6.7	2
36	2509	Chi-Chi	1999-09-20	6.2	2
37	800	Loma Prieta	1989-10-18	6.9	2
38	754	Loma Prieta	1989-10-14	6.9	2
39	1183	Chi-Chi	1999-09-20	7.6	2
40	3512	Chi-Chi	1999-09-22	6.3	2

## 2.2 地震动 IM 选取

本文选取 46 个地震动 IM 作为研究对象,按照其表征意义可以将这 46 个地震动 IM 分为幅值型、频谱型、持时型和混合型四大类,地震动 IM 分类见表 4。其中幅值型包括:峰值加速度( $A_{pg}$ )、峰值速度( $V_{pg}$ )、峰值位移( $D_{pg}$ );持时型包括峰值速度与峰值加速度比( $R_{va}$ )、峰值位移与峰值速度比( $R_{dv}$ )、重要持时( $T_d$ )、频谱型包括谱加速度均值( $S_{a,avg}$ )、谱速度均值( $S_{v,avg}$ )、谱位移均值( $S_{d,avg}$ )、有效峰值加速度( $A_{ep}$ )、有效峰值速度( $V_{ep}$ )、有效峰值位移( $D_{ep}$ )、地震动反应谱烈度( $A_{si}$ 、 $V_{si}$ 和 $D_{si}$ )、Housner 谱强度( $I_s$ )、地震动加速度曲线上单位时间通过零点的次数( $V_o$ )、最大谱加速度( $A_{ps}$ )、最大谱速度( $V_{ps}$ )、最大谱位移( $D_{ps}$ )、滤波增量速度( $V_f$ )、混合型包括 Arias 强度( $I_A$ )、修正的 Arias 强度( $I_{am}$ )、Housner 强度( $P_a$ 、 $a_{rms}$ 、 $P_v$ 、 $v_{rms}$ 、 $P_d$ 、 $d_{rms}$ )、Nau 和 Hall 指标( $E_a$ 、 $E_v$ 、 $E_d$ 、 $A_{rs}$ 、 $V_{rs}$ 和 $D_{rs}$ )、累计绝对速度( $V_{ca}$ )、累计绝对位移( $D_{ca}$ )、累计绝对动量( $I_{ca}$ )、均方根加速度( $A_{rms}$ )、均方根速度( $V_{rms}$ )、均方根位移( $D_{rms}$ )、Fajfar 指标( $I_f$ )、Park-Ang 指标( $I_c$ )、Riddell 指标( $I_a$ 、 $I_v$ 和 $I_d$ )。

表 4 地震动 IM 分类

Table 4 Classification of ground motion intensity measure

类型	强度指标
幅值型	$A_{pg}、V_{pg}、D_{pg}$
持时型	$R_{va}、R_{dv}、T_d$
频谱型	$S_{a,avg}、S_{v,avg}、S_{d,avg}、A_{ep}、V_{ep}、D_{ep}、A_{si}、V_{si}、D_{si}、I_s、V_o、A_{ps}、V_{ps}、D_{ps}、V_{fi}$
混合型	$I_A、J_{am}、P_a、a_{rms}、P_v、v_{rms}、P_d、d_{rms}、E_a、A_{rs}、E_v、V_{rs}、E_d、D_{rs}、V_{ca}、D_{ca}、J_{ca}、A_{rms}、V_{rms}、D_{rms}、J_c、J_f、J_a、J_v、J_d$

### 3 结构建模及抗地震倒塌能力分析

本文 SDOF 体系恢复力模型采用考虑了刚度退化的改进 IMK 模型<sup>[24]</sup>,改进 IMK 模型本构曲线见图 1。图中参数定义如下: $F_c$ 为峰值力; $K_e$ 为弹性刚度; $K_s$ 为强化刚度; $K_c$ 为退化刚度; $\delta_y$ 为屈服变形; $\delta_c$ 为峰值变形。其表达式为

$$\begin{cases} K_s = \alpha_s \cdot K_e \\ K_c = \alpha_c \cdot K_e \end{cases} \quad (1)$$

式中: $\alpha_s$ 为强化刚度系数; $\alpha_c$ 为退化刚度系数。SDOF 体系自振周期  $T$  取值为 0.6、0.8、1.0、1.2、1.4、1.6、1.8、2.0、2.2 s。屈服强度系数分别取 0.1、0.2、0.3、0.4、0.5、0.6、0.7、0.8。第 2 刚度系数  $\alpha_s$  取值为 0、0.05、0.10、0.15、0.2。由上可知共获得 360 个不同的 SDOF 体系。

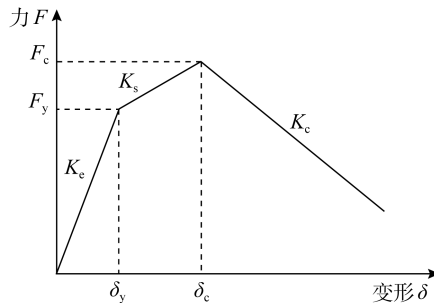


图 1 SDOF 恢复力模型(改进 IMK 本构模型)

Fig. 1 SDOF restoring force model (Modified-IMK constitutive model)

本文 RC 框架结构采用 HASLTON 等<sup>[25]</sup>基于洛杉矶场地危险性设计的 4 座典型建筑,层数分别为 2、4、8、12 层,包含了常见框架结构层数。利用 OpenSees 建立结构有限元模型,结构一阶自振周期  $T_1$  分别为 0.63、0.94、1.80、2.14 s。其中 2、4 层结构平面为 36.60 m×54.98 m,8、12 层结构平面为 36.6 m×36.6 m。首层层高 4.6 m,标准层高 4.0 m,梁柱塑性铰为集中塑性铰,8 层 RC 框架模型平立面图见图 2。

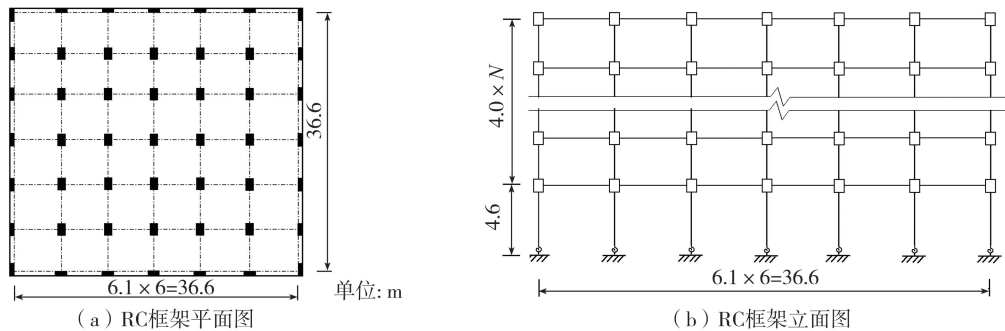


图 2 RC 框架结构的平面和立面图

Fig. 2 Plan and elevation drawings of RC frame structure

采取增量动力分析(incremental dynamic analysis, IDA)<sup>[26]</sup>对 SDOF 体系和 RC 框架结构进行结构地震倒塌分析。采用结构 1 阶自振周期对应的加速度谱  $S_a(T_1, \xi)$  表示地震动强度值 IM,采用结构最大层间位移角  $\theta_{max}$  作为 DM。IDA 分析时结构倒塌标准采用 IM-DM 混合判别方法<sup>[26]</sup>,取屈服后刚度  $K_{ult}$  小于初始刚度  $K_0$  的 20%作为结构倒塌依据,见图 3。

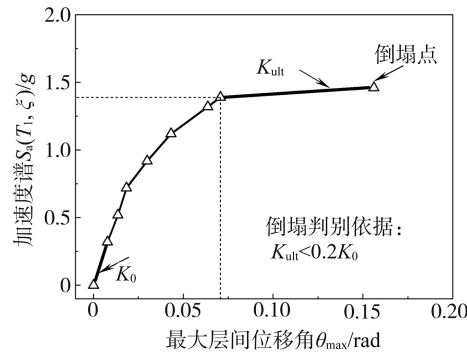


图3 单条地震动记录的 IDA 曲线

Fig. 3 IDA curve of a single ground motion record

## 4 特征选择算法性能对比

### 4.1 地震动 IM 筛选方法及算法性能对比验证

为验证第 1 节中的 4 种特征选择算法对地震动强度指标筛选时的性能,分别利用每种特征选择算法进行强度指标筛选,筛选对象为第 3 节中的 360 个不同的 SDOF 体系。由于特征选择算法属于机器学习中的回归类算法,因此对于每种特征选择算法,可以得到 360 个回归模型。考虑到 360 个 SDOF 体系地震动 IM 权重统计问题,本文采用刘巴黎等<sup>[1]</sup>使用的基于敏感性系数和频数定义欧氏距离方法对 IM 排序。敏感性系数定义为每种特征选择算法 360 个回归模型中地震动 IM 的权重值平方和,频数定义为每种特征选择算法 360 个回归模型中地震动 IM 权重值非零个数。如果将每个 IM 看作二维空间中的点,频数和敏感性系数为该点的横纵坐标,原点距即为该 IM 的欧氏距离,为方便计算,对敏感性系数和频数按最大值归一化,图 4 为欧氏距离示意图。

采用欧氏距离评价特征选择算法性能虽然能展示不同算法对 IM 的筛选,但无法判断算法是否能快速且准确预测结构倒塌点,需要引入其他性能评价方法。IM 与 DM 之间在取对数前提下具有较高的线性相关性<sup>[27-29]</sup>。因此,可以通过建立对数最小二乘回归模型,将基于特征选择算法筛选出的 IM 按欧氏距离排序依次加入最小二乘回归模型,得到最小二乘回归结果与真实结果的残差的标准差。通过残差的标准差随 IM 个数增加的曲线,可以判断该种特征选择算法筛选出的地震动 IM 排序是否可以准确且快速对结构倒塌进行预测。

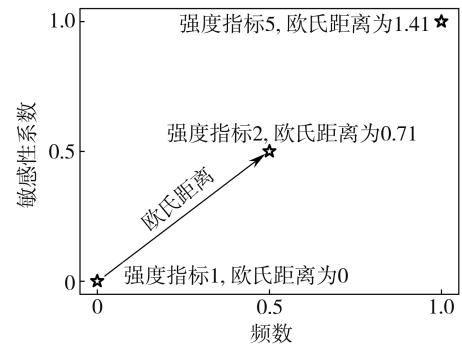


图4 欧氏距离示意图

Fig. 4 Schematic diagram for Euclidean distance

### 4.2 基于单个 SDOF 结构地震动 IM 筛选方法验证

本节利用单个 SDOF 体系,采用不同特征选择算法对 IM 进行欧氏距离排序并利用最小二乘回归对 4.1 节筛选及评价原则进行验证。选取自振周期  $T$  为 1.4 s,屈服强度系数为 0.1,第二刚度系数为 0.2 的 SDOF 体系进行验证。

为更好地表征每个 IM 间欧氏距离差值变化及特征选择算法性能优劣,采用欧氏距离相对误差的均值及标准差衡量。相对误差的计算公式为

$$\delta = \frac{E_n - E_{n+1}}{E_{\max}} \times 100\% \quad (2)$$

式中: $\delta$  为相对误差; $E_n$  为第  $n$  个 IM 的欧氏距离; $E_{\max}$  为最大欧氏距离。需要指出,当欧氏距离降至 0,表示该 IM 在预测结构倒塌上无贡献,因此相对误差计算截止。根据式(2),相对误差的均值越大,各 IM 间的欧氏距离差值越大,相对误差的标准差越小,各 IM 间的欧氏距离差值离散程度越小,对于 IM 筛选有帮助。图 5 给出不同特征选择算法筛选出 IM 的欧氏距离分布径向条形图,图中算法名称周围圆环为坐标轴,46 个强度指标以  $x$  轴正方向开始环绕 360° 分布。每个强度指标对应的条形图长度为欧氏距离,右上角刻度线为欧氏距离刻度。当算法筛选强度指标性能较好,即各强度指标间欧氏距离有明显差距时,径向条形图呈现海螺型;反之则呈现近圆形。计算相对误差, MIC 特征选择算法的欧氏距离相对误差均值为 0.5%,标准差为

0.0060;Relieff 特征选择算法的欧氏距离相对误差均值为 0.6%,标准差为 0.017。XGBoost 特征选择算法的欧氏距离相对误差均值为 0.9%,标准差为 0.046。Lasso 特征选择算法的欧氏距离相对误差均值为 1.9%,标准差为 0.034。

对于单个 SDOF 体系,4 种特征选择算法表现各异。MIC、Relieff 这 2 种特征选择算法得出的 IM 欧氏距离相对误差标准差较小,但相对误差均值亦小,无法对特征进行有效筛选。相较之下,XGBoost、Lasso 这 2 种特征选择算法得出的 IM 欧氏距离相对误差标准差虽然小于过滤法,但其相对误差均值较大,且无关特征欧氏距离为零,能有效对 IM 进行筛选。

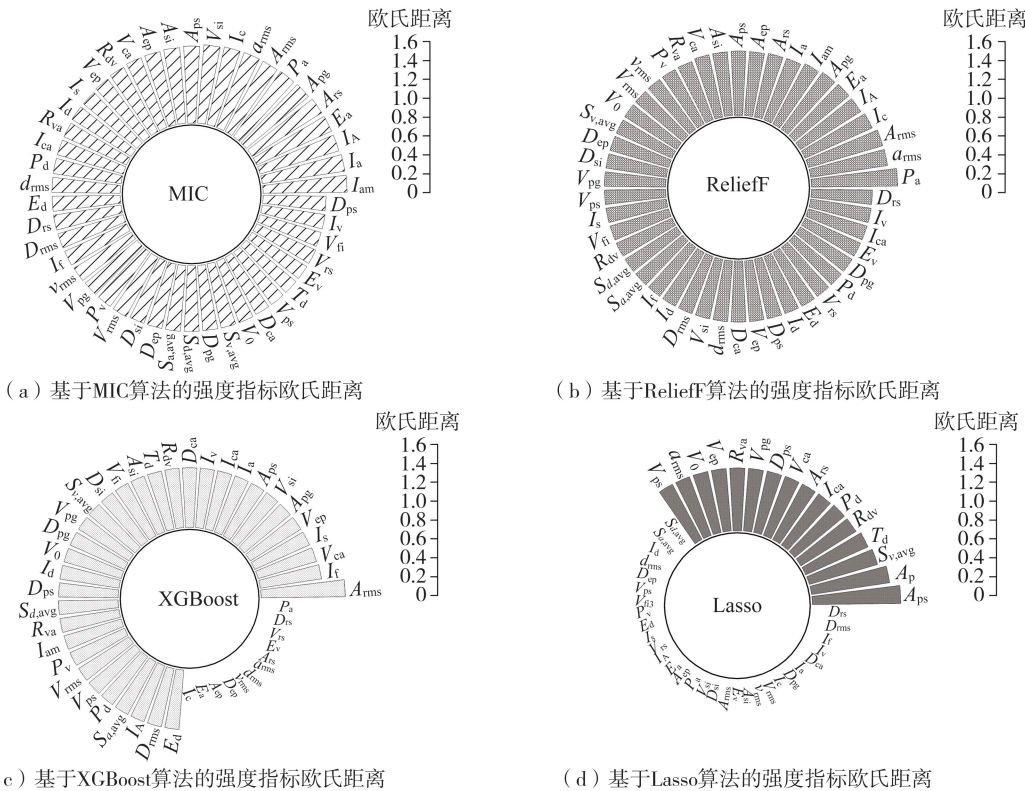


图 5 单个 SDOF 欧氏距离排序径向条形图

Fig. 5 Single SDOF Euclidean distance sorted radial bar chart

图 6 给出单个 SDOF 结构的结构倒塌预测值与真实值的残差标准差随 IM 数量增加变化图,横轴为 IM 累积加入最小二乘回归模型数量,纵轴为残差标准差。当没有 IM 加入最小二乘回归模型时,残差的标准差为 1。随着 IM 按照欧氏距离大小依次加入最小二乘回归模型,残差的标准差降低。下降幅值在早期较大,在后期较小,这是加入的 IM 贡献逐渐变小。取 10 个 IM 作为解释变量加入最小二乘回归时残差的标准差降低值作为比较依据<sup>[1]</sup>,MIC、Relieff、XGBoost 和 Lasso 这 4 种方法残差标准差分别下降 49.5%、49.6%、56.5%、62.9%。随着 IM 数目的增加,Lasso 方法残差标准差下降最快,说明该方法筛选出的 IM 能更快更准确地预测结构倒塌,性能优于其他 3 种方法。根据径向条形图,当 IM 的欧氏距离降为 0 后,2 种嵌入式算法的残差的标准差几乎变化不大,与欧氏距离降为 0 则 IM 对结构倒塌预测几乎无贡献相印证。

根据单个 SDOF 模型对 4.1 节筛选及评价原则的实践,证明 4.1 节的筛选及评价原则能有效用于不同特征选择算法性能对比,可以应用于大量 SDOF 体系及 RC 框架体系。

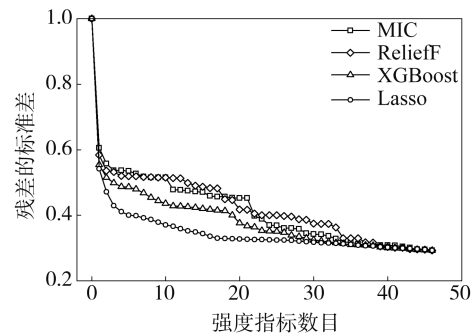


图 6 单个 SDOF 模型 4 种特征选择算法回归分析中残差与地震动 IM 关系

Fig. 6 Relationship between standard deviations of the residuals by the least square regressions and ground motion parameters of a single SDOF the amount of intensity measures

### 4.3 基于 SDOF 体系特征选择算法筛选结果及性能对比分析

本节利用大量 SDOF 体系进行强度指标筛选及特征选择算法性能比较。根据 4.1 节强度指标筛选方法对强度指标进行筛选。图 7 给出基于不同特征选择算法筛选出的 IM 的欧氏距离分布的径向条形图,由图可知,MIC、ReliefF 算法的径向条形图呈现近圆形,XGBoost 算法在排序靠前强度指标变化上呈现近圆形、靠后逐渐呈现海螺型,而 Lasso 算法呈现出明显的海螺型。根据欧氏距离排序结果,MIC 方法筛选出的强度指标为  $A_{ep}$ 、 $A_{si}$ 、 $I_A$ 、 $R_{va}$ 、 $A_{pg}$ 、 $P_a$ 、 $I_c$ 、 $A_{ps}$ 、 $a_{rms}$ 、 $I_a$ ;ReliefF 方法筛选出的强度指标为  $P_a$ 、 $P_d$ 、 $A_{ps}$ 、 $E_a$ 、 $I_A$ 、 $I_c$ 、 $a_{rms}$ 、 $A_{rms}$ 、 $A_{pg}$ 、 $I_a$ ;XGBoost 方法筛选出的强度指标为  $A_{si}$ 、 $V_{ca}$ 、 $A_{rms}$ 、 $D_{ca}$ 、 $I_a$ 、 $I_A$ 、 $V_{si}$ 、 $V_{ep}$ 、 $R_{va}$ 、 $V_{pg}$ ;Lasso 方法筛选出的强度指标为  $A_{ps}$ 、 $S_{a,avg}$ 、 $T_d$ 、 $a_{rms}$ 、 $V_{ep}$ 、 $R_{dv}$ 、 $S_{v,avg}$ 、 $I_d$ 、 $P_a$ 、 $P_d$ 。计算相对误差,MIC 特征选择算法的欧氏距离相对误差均值为 0.5%,标准差为 0.008。ReliefF 特征选择算法的欧氏距离相对误差均值为 0.6%,标准差为 0.016。XGBoost 特征选择算法的欧氏距离相对误差均值为 2.2%,标准差为 0.051。Lasso 特征选择算法的欧氏距离相对误差均值为 2.2%,标准差为 0.034。

对于大量 SDOF 体系,MIC、ReliefF 这 2 种特征选择算法虽然相对误差的标准差较小,但是相对误差的均值也同样较小,不能很好地起到特征选择。XGBoost、Lasso 这 2 种特征选择算法的相对误差均值较大,且标准差也较小,其中 Lasso 特征选择算法不论是在均值还是在标准差均符合能有效进行特征筛选的评判标准。

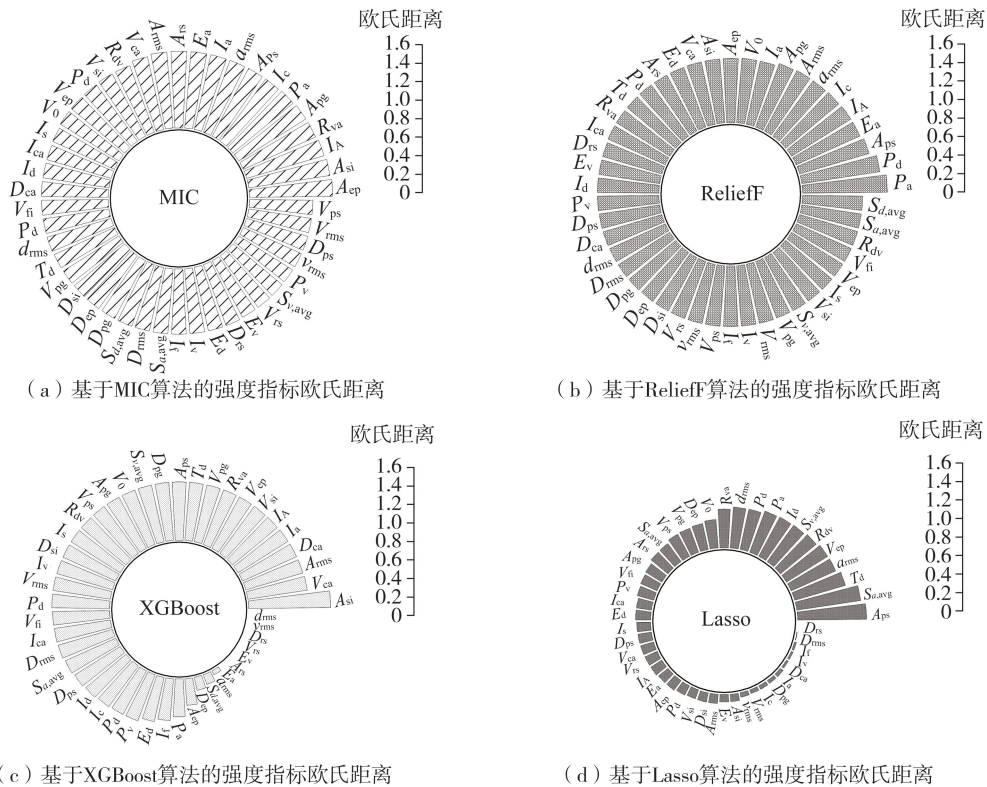


图 7 全部 SDOF 欧氏距离排序径向条形图

Fig. 7 All SDOF Euclidean distance sorted radial bar chart

图 8 给出了所有 SDOF 结构的结构倒塌预测值与真实值的残差标准差均值随 IM 的变化关系图,随着 IM 按照欧氏距离大小依次加入最小二乘回归模型,残差的标准差降低。取 10 个 IM 作为解释变量加入最小二乘回归时残差的标准差均值降低值作为比较依据,MIC、ReliefF、XGBoost 和 Lasso 所给出的地震动 IM 排序建立的最小二乘回归模型残差标准差分别下降 51.1%、48.7%、54.4%、61.1%。随着 IM 数目的增加,Lasso 方法残差标准差下降最快,说明该方法筛选出的 IM 能更快更准确地预测结构倒塌,性能优于其他 3 种方法,与 4.2 节结果一致。

值得注意的是,单个 SDOF 体系的最小二乘回归结果相较于大量 SDOF 体系的最小二乘回归结果更好,这是因为大量 SDOF 体系覆盖了全部的 SDOF 体系,不同 SDOF 体系单体的最小二乘回归效果各异,但是总体符合规律,可以作为后续 RC 框架结构的 IM 排序筛选依据。

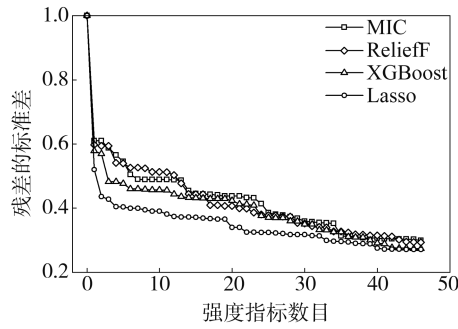


图 8 全部 SDOF 模型 4 种特征选择算法回归分析中残差与地震动 IM 关系

Fig. 8 Relationship between standard deviations of the residuals by the least square regressions and ground motion parameters of all SDOF models

### 4.4 基于 4 个 RC 框架结构特征选择算法性能对比分析

为验证基于大量 SDOF 体系筛选出的 IM 排序是否能够有效预测 RC 框架结构的倒塌,本节利用第 3 节介绍的 4 个 RC 框架模型进行倒塌分析,根据 4.3 节给出的 IM 排序与本节进行的倒塌分析结果建立最小二乘回归模型,以残差的标准差变化作为衡量不同特征选择算法筛选出结果的标准。RC 框架结构倒塌分析,地震动采用 2.1 节给出的 Set #1A 地震动集。以倒塌点作为因变量,将基于 4 种方法得到的地震动 IM 依次放入最小二乘回归模型作为自变量得到 RC 框架结构 4 种特征选择算法回归分析中残差与地震动 IM 关系。

图 9 分别给出不同结构中,4 种特征选择算法回归分析中残差与地震动 IM 关系,其中地震动 IM 个数从 0 增加到 46。由图可知,在 IM 数量为 0 时,各模型的残差标准差均为 1。随着 IM 数目的增加,对于不同结构的不同特征选择方法,残差的标准差均呈现下降趋势。Lasso 方法在前期迅速降低之后缓慢降低,其他 3 种方法均随着地震动 IM 的加入稳步下降。

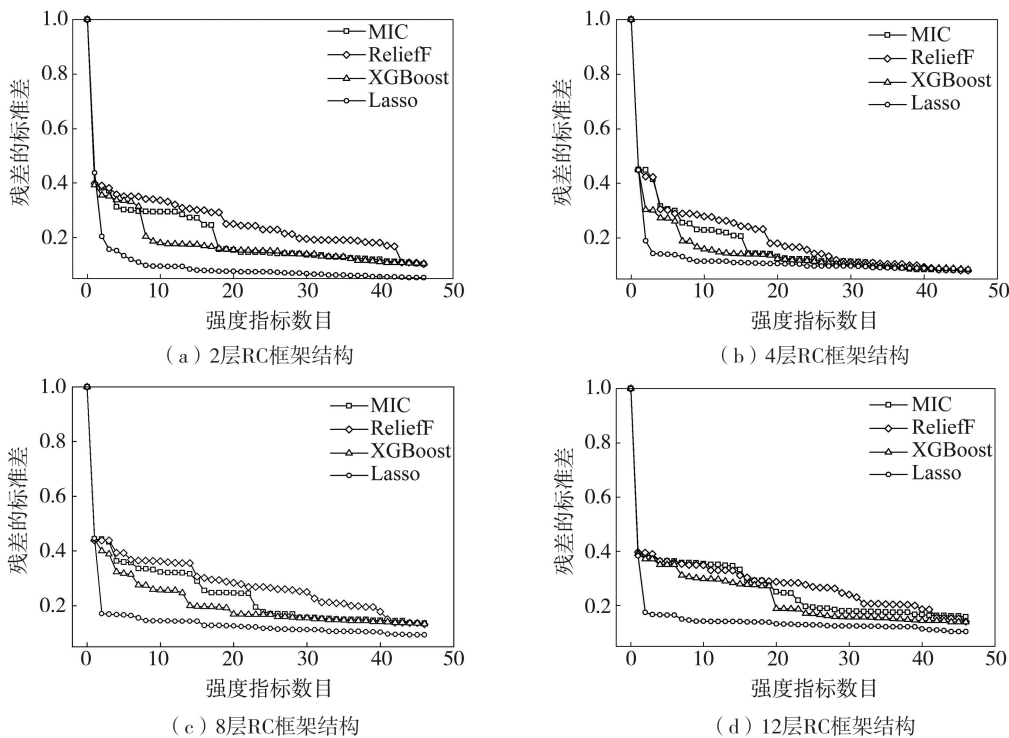


图 9 RC 框架结构 4 种特征选择算法回归分析中残差与地震动 IM 关系

Fig. 9 Relationship between standard deviations of the residuals by the least square regressions and ground motion parameters of RC structures

取 10 个 IM 作为解释变量加入最小二乘回归时残差的标准差降低值作为比较依据,对于 2 层结构, MIC、Relieff、XGBoost 和 Lasso 所给出的地震动 IM 排序建立的最小二乘回归模型残差标准差分别下降 71.6%、66.4%、82.9%、90.6%。对于 4 层结构,4 种方法分别降低 77.2%、72.2%、84.1%、88.5%。对于 8 层结

构,4种方法分别降低67.7%、63.7%、74.3%、85.6%。对于12层结构,4种方法分别降低64.6%、65.1%、70.2%、85.8%。以上数据表明,Lasso特征选择算法不论对于哪种结构,在将10个IM作为解释变量加入最小二乘回归时差的标准差降低值均在85%以上。综上所述,Lasso回归特征选择算法能有效地筛选预测结构倒塌的地震动IM。

## 5 结论

基于SDOF模型的IDA结果,利用不同特征选择算法对地震动IM进行筛选,利用单个SDOF结构分析结果对比不同特征选择算法用于地震动IM筛选时的性能,根据SDOF结构分析结果推荐给出用于地震动IM筛选时性能最好的特征选择算法,并基于RC框架结构倒塌分析结果验证了推荐特征选择算法的有效性。得到主要结论如下:

1) 基于Lasso回归筛选的10个IM作为解释变量加入最小二乘回归时,4个RC框架结构抗地震倒塌能力预测分析时残差标准差降低值分别为90.6%、88.5%、85.6%、85.8%,比ReliefF预测准确率平均高31%,比MIC预测准确率平均高25%,比XGBoost预测准确率平均高13%。这说明基于Lasso特征选择算法筛选出的IM能更有效预测结构抗地震倒塌能力。

2) 采用Lasso特征选择算法进行特征筛选时,单个SDOF结构欧氏距离相对误差均值为1.9%,标准差为0.034。大量SDOF结构欧氏距离相对误差均值为2.2%,标准差为0.034。其结果均优于ReliefF、MIC以及XGBoost特征选择算法,这说明基于Lasso回归的特征选择算法筛选结果稳定,能有效筛选地震动IM。

3) 基于Lasso回归的特征选择算法能有效地筛选用于结构抗地震抗倒塌能力预测的地震动IM,结果可为地震动IM筛选及结构地震倒塌风险分析中地震动不确定性的研究提供参考。

## 参考文献:

- [1] 刘巴黎,胡进军,谢礼立.基于弹性网络回归的地震动参数排序与比选[J].哈尔滨工业大学学报,2024,56(1):54-62.  
LIU Bali, HU Jinjun, XIE Lili, et al. Ranking and comparison of ground motion parameters based on elastic net regression[J]. Journal of Harbin Institute of Technology, 2024, 56(1): 54-62. (in Chinese)
- [2] DÁVALOS H, MIRANDA E. Filtered incremental velocity: A novel approach in intensity measures for seismic collapse estimation[J]. Earthquake Engineering & Structural Dynamics, 2019, 48(12): 1384-1405.
- [3] 韩建平,周伟,李慧.基于汶川地震数据的地震动强度指标与中长周期SDOF体系最大响应相关性[J].工程力学,2011,28(10):185-196.  
HAN Jianping, ZHOU Wei, LI Hui. Correlation between ground motion intensity indices and SDOF system responses with medium-to-long period based on the Wenchuan earthquake data[J]. Engineering Mechanics, 2011, 28(10): 185-196. (in Chinese)
- [4] 陈健云,李静,韩进财,等.地震动强度指标与框架结构响应的相关性研究[J].振动与冲击,2017,36(3):105-112,144.  
CHEN Jianyun, LI Jing, HAN Jincai, et al. Correlation between ground motion intensity indexes and seismic responses of frame structures[J]. Journal of Vibration and Shock, 2017, 36(3): 105-112, 144. (in Chinese)
- [5] 李雪红,李晔暄,吴迪,等.地震动强度指标与结构地震响应的相关性研究[J].振动与冲击,2014,33(23):184-189.  
LI Xuehong, LI Yexuan, WU Di, et al. Correlation between ground motion intensity and structural seismic response[J]. Journal of Vibration and Shock, 2014, 33(23): 184-189. (in Chinese)
- [6] 李爽,谢礼立,郝敏.地震动参数及结构整体破坏相关性研究[J].哈尔滨工业大学学报,2007,39(4):505-509,560.  
LI Shuang, XIE Lili, HAO Min. Correlation between seismic ground motion parameters and their relationship with overall damage to structure[J]. Journal of Harbin Institute of Technology, 2007, 39(4): 505-509, 560. (in Chinese)
- [7] 胡进军,赵泽锋,谢礼立.考虑场地类别和断层距的地震动及结构响应参数相关性分析[J].地震工程与工程振动,2020,40(2):13-22.  
HU Jinjun, ZHAO Zefeng, XIE Lili. Correlation analysis of ground motion parameters and structural response parameters considering the site condition and fault distance[J]. Earthquake Engineering and Engineering Dynamics, 2020, 40(2): 13-22. (in Chinese)
- [8] AFARANI M H C, NICKNAM A. Seismic response of mass irregular steel moment resisting frames (SMRF) according to performance levels from IDA approach[J]. Gazi University Journal of Science, 2012, 25(3): 751-760.
- [9] MAISON B F, KASAI K, DEIERLEIN G. PBEE evaluation of E-defense collapse experiment [C]//The 14th World Conference on Earthquake Engineering. Beijing: China Earthquake Administration, 2008: 15-22.
- [10] 张丽新.高维数据的特征选择及基于特征选择的集成学习研究[D].北京:清华大学,2004.  
ZHANG Lixin. Study on feature selection and ensemble learning based on feature sselection for high-dimensional datasets[D]. Beijing: Tsinghua University, 2004. (in Chinese)
- [11] 吴梓楠,韩小雷,马建峰,等.基于机器学习的地震动强度指标敏感性分析与破坏势评估[J].建筑结构学报,2023,44(11):216-225,235.

- WU Zinan, HAN Xiaolei, MA Jianfeng, et al. Sensitivity analysis of ground motion intensity measures and evaluation of potential damage based on machine learning[J]. *Journal of Building Structures*, 2023, 44(11): 216–225, 235. (in Chinese)
- [12] XU Y J, LU X Z, TIAN Y, et al. Real-time seismic damage prediction and comparison of various ground motion intensity measures based on machine learning[C]//IABSE Congress Reports, IABSE Congress, Christchurch 2021: Resilient Technologies for Sustainable Infrastructure. Christchurch, New Zealand. Zurich, Switzerland; International Association for Bridge and Structural Engineering (IABSE), 2021.
- [13] DENIZ D. Stochastic prediction of collapse of building structures under seismic excitations[D]. Urban: University of Illinois at Urbana Champaign, 2015.
- [14] 查军龙, 刘洋, 戴靠山, 等. 工业厂房结构及非结构构件抗震性能参数研究综述[J]. *地震工程与工程振动*, 2021, 41(4): 196–208. ZHA Junlong, LIU Yang, DAI Kaoshan, et al. A review on seismic performance parameters of industrial plant structures and non-structural components[J]. *Earthquake Engineering and Engineering Dynamics*, 2021, 41(4): 196–208. (in Chinese)
- [15] TANG J, ALELYANI S, LIU H. Feature selection for classification: A review[M]. In *Data Classification: Algorithms and Applications*. Boca Raton: Chapman & Hall/CRC, 2014: 37–64.
- [16] LIU H, MOTODA H. Computational methods of feature selection[M]. Boca Raton: Chapman & Hall/CRC, 2007.
- [17] LIU H, YU L. Toward integrating feature selection algorithms for classification and clustering[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(4): 491–502.
- [18] MA S G, HUANG J. Penalized feature selection and classification in bioinformatics[J]. *Briefings in Bioinformatics*, 2008, 9(5): 392–403.
- [19] RESHEF D N, RESHEF Y A, FINUCANE H K, et al. Detecting novel associations in large data sets[J]. *Science*, 2011, 334(6062): 1518–1524.
- [20] KONONENKO I. Estimating attributes: Analysis and extensions of RELIEF[M]. *Lecture notes in computer science*. Berlin, Heidelberg: Springer Berlin Heidelberg, 1994: 171–182.
- [21] CHEN T Q, GUESTRIN C. XGBoost: A scalable tree boosting system[C]//Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco California USA. New York: ACM, 2016: 785–794.
- [22] TIBSHIRANI R. Regression shrinkage and selection via the lasso: A retrospective[J]. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 2011, 73(3): 273–282.
- [23] BAKER J W, LIN T, SHAHI S K, et al. New ground motion selection procedures and selected motions for the peer transportation research program: Peer Report 2011/3[R]. California: Pacific Earthquake Engineering Research Center, College of Engineering, University of California, 2011.
- [24] IBARRA L F, MEDINA R A, KRAWINKLER H. Hysteretic models that incorporate strength and stiffness deterioration[J]. *Earthquake Engineering & Structural Dynamics*, 2005, 34(12): 1489–1511.
- [25] HASELTON C B, LIEL A B, DEIERLEIN G G, et al. Seismic collapse safety of reinforced concrete buildings. I: Assessment of ductile moment frames[J]. *Journal of Structural Engineering*, 2011, 137(4): 481–491.
- [26] VAMVATSIKOS D, CORNELL C A. Incremental dynamic analysis[J]. *Earthquake Engineering & Structural Dynamics*, 2002, 31(3): 491–514.
- [27] LIU T T, YU X H, LU D G. An approach to develop compound intensity measures for prediction of damage potential of earthquake records using canonical correlation analysis[J]. *Journal of Earthquake Engineering*, 2020, 24(11): 1747–1770.
- [28] OZMEN H B. Developing hybrid parameters for measuring damage potential of earthquake records: Case for RC building stock[J]. *Bulletin of Earthquake Engineering*, 2017, 15(7): 3083–3101.
- [29] 刘亭亭, 于晓辉, 吕大刚. 地震动多元强度参数主成分与结构损伤的相关性分析[J]. *工程力学*, 2018, 35(8): 122–129, 137. LIU Tingting, YU Xiaohui, LÜ Dagang. Analysis of correlation between principal components of multivariate earthquake intensity measures and structural damage[J]. *Engineering Mechanics*, 2018, 35(8): 122–129, 137. (in chinese)