基于可解释机器学习和全局优化算法的电化学阻抗谱解析

金　莹，赵朝阳，赖召贵

b20170427@xs.ustb.edu.cn

北京科技大学 国家材料服役安全科学中心

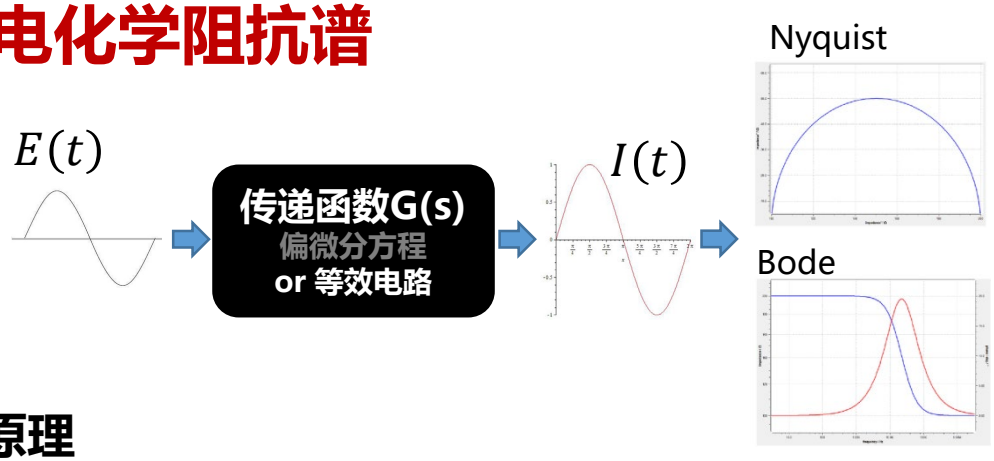
2021年12月16日

主要内容

第1部分	研究背景
第2部分	基于机器学习算法的等效电路预测
第3部分	基于全局优化算法的等效电路参数识别
第4部分	总结与展望

研究背景

➢ 电化学阻抗谱



- **原理**

（Electrochemical Impedance Spectroscopy，EIS）通过向待测**稳定**的电化学系统施加一个小幅的正弦电位激励信号$E(t)$，随后检测响应信号$I(t)$, $Z = \frac{E(t)}{I(t)}$ 即为系统阻抗

- **必要条件**

系统稳定、系统线性、因果性、测试频率无限性（数据验证）

- **优点**

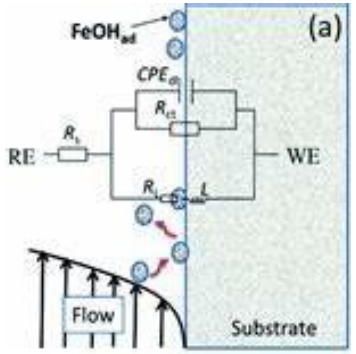
可以获取多个、不同反应速率下物理、化学过程**耦合的**动力学信息

- **常规解析手段**
  - 偏微分方程（过程清晰、难以图形展示、数值解析困难）
  - 等效电路模型（Equivalent Circuit Model，ECM）（图形直观、数值解析简单，依赖经验）

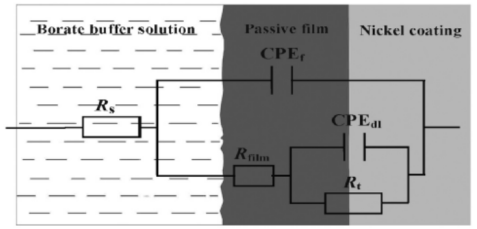
- **广泛的应用场景**

✓ **金属腐蚀相关**

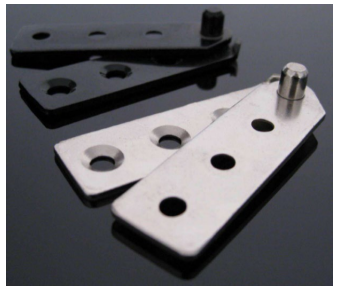
金属腐蚀



涂层保护

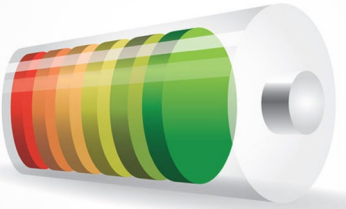


电镀



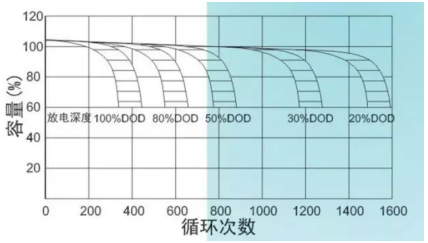
✓ **电池性能健康状态相关**

电池容量

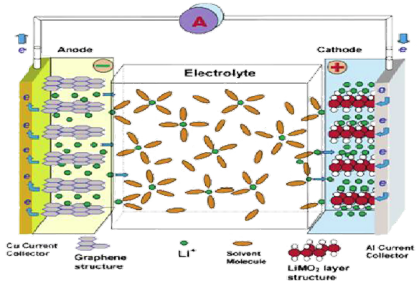


充电状态

剩余使用寿命



导电物质传输



✓ **生物传感器、缓蚀剂等**

研究背景

➢ 基于等效电路模型的电化学阻抗谱解析

● 第一步：等效电路模型选择

第一种：基于经验选择

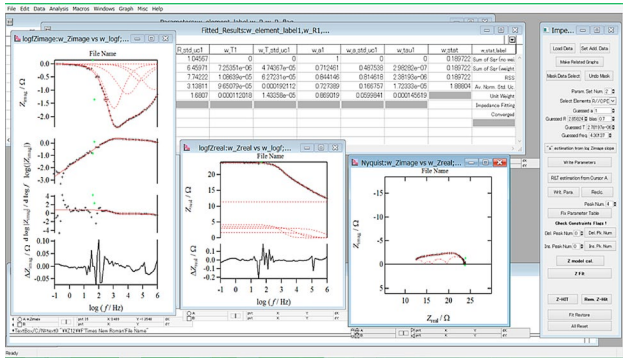
1、根据EIS的Nyquist和Bode中曲线形状和走势来选择电路元件的种类、数量和链接顺序

2、确保每个电路元件代表具体的物理、化学过程含义，链接后的整个电路能近似代表一个完整的电化学过程

3、ECM的参数拟合要满足一定精度

4、多种ECM可选时，选择最简单的一个

第二种：基于软件交互辅助选择



Kiyoshi KOBAYASHI NIMS[1]
基于测量模型Vogit电路

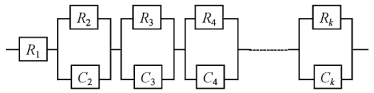
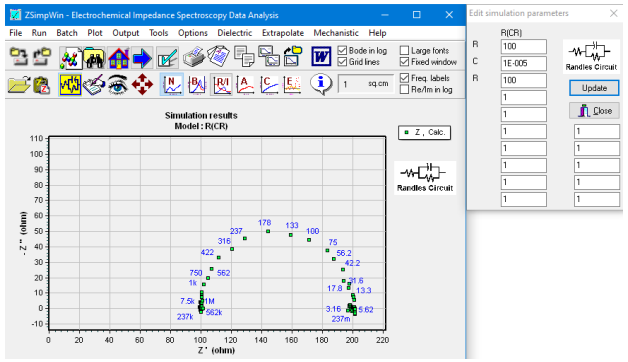
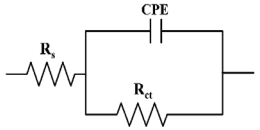


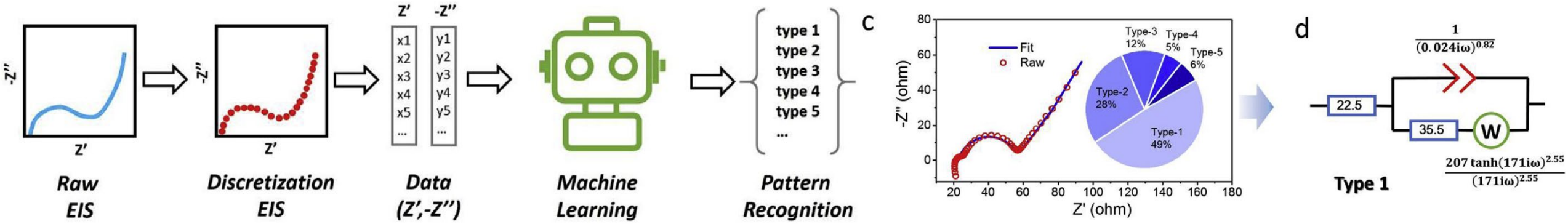
Fig. 1. Equivalent circuit model of the measurement model.



AMETEK, Gamry，…
过程模型



第三种：基于数据驱动的机器学习[2]



最佳机器学习模型为**支持向量机**

在包含**5种**EIS数据的测试集上的**分类正确率为0.48**，**距离1还有很大的提升空间（更多的数据，更优的超参数，更好的算法）**

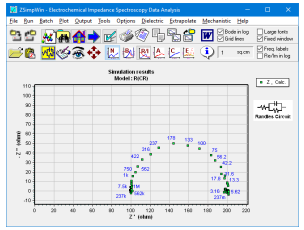
[1] Kobayashi, K., Sakka, Y. & Suzuki, T. S. Development of an electrochemical impedance analysis program based on the expanded measurement model. *J. Ceram. Soc. Jpn*. **124**, 943-949 (2016).
[2] Zhu, S., Sun, X., Wang, Y., Zhao, N. & Sha, J. Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning. *J. Electroanal. Chem*. **855**, 113627 (2019).

研究背景

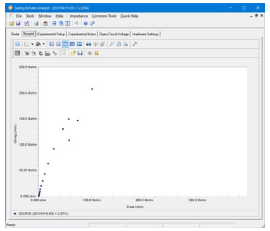
➢ 基于等效电路模型的电化学阻抗谱解析

● 第二步：等效电路模型参数拟合

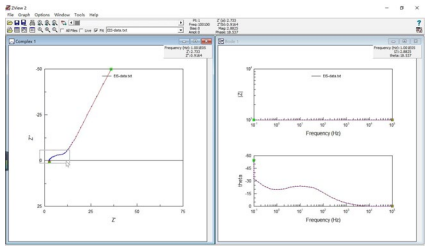
✓ **商用软件**



AMETEK **ZSimpWin**
全局优化



Gamry **EIS300™**
NLLS



Scribner **ZView**
NLLS

✓ **开源软件**

LEVM/LEVMW
J. R. Macdonald[1]
NLLS

PyEIS
Kristian Knudsen[2]
NLLS

Impedance★
ECS Hack Week[3]
NLLS

ECM参数拟合：

- ➢ 拟合初始值的选取、复杂的电路对NLLS拟合结果影响较大
- ➢ 商用软件的数据处理通量一般在有限的1~10条，开源软件尚处于逐条处理

✓ **局部优化算法**

(主要是非线性最小二乘法)

Nonlinear Least Squares Fitting (N)LLS **NLLS**

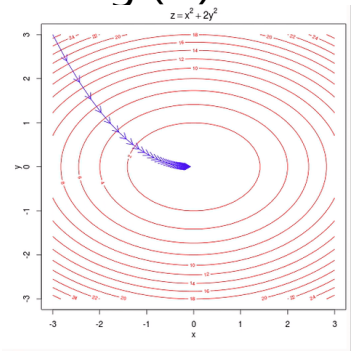
- ➢ 牛顿法
- ➢ 梯度下降（下方动图）
- ➢ **Levenberg-Marquardt**

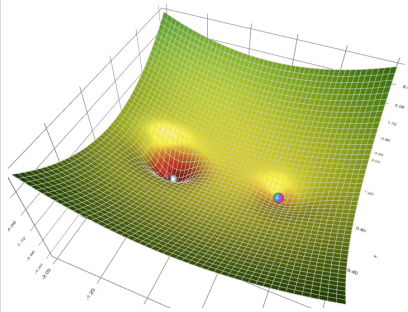
优点

- ➢ 收敛迅速

缺点

- ➢ 需求导
- ➢ 初始值影响大





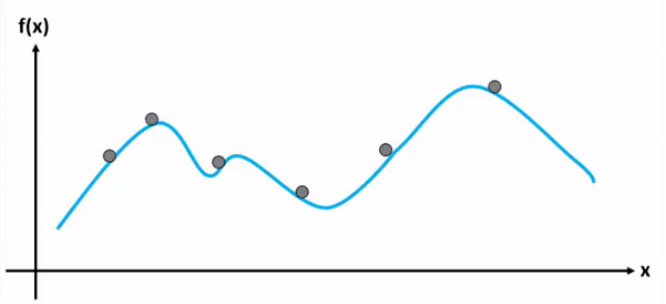
✓ **元启发式全局优化算法**

优点

- ➢ 无需求导
- ➢ 无需初始值

缺点

- ➢ 收敛缓慢



- ➢ **无法理论证明某种算法的优越性，需广泛的对比实验**

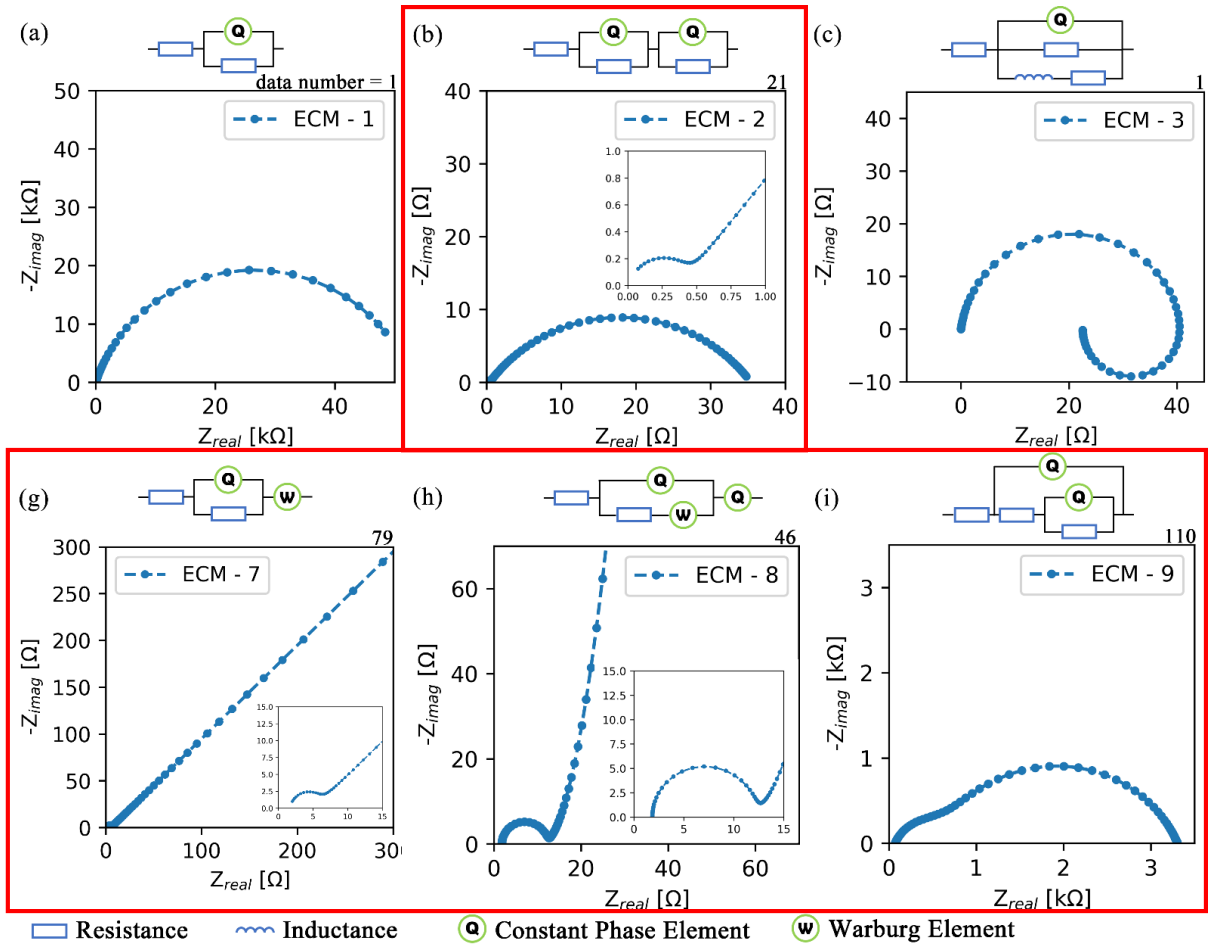
[1] https://jrossmacdonald.com/levmlevmw/; [2] https://github.com/kbknudsen/PyEIS; [3] https://www.electrochem.org/233/hack-week

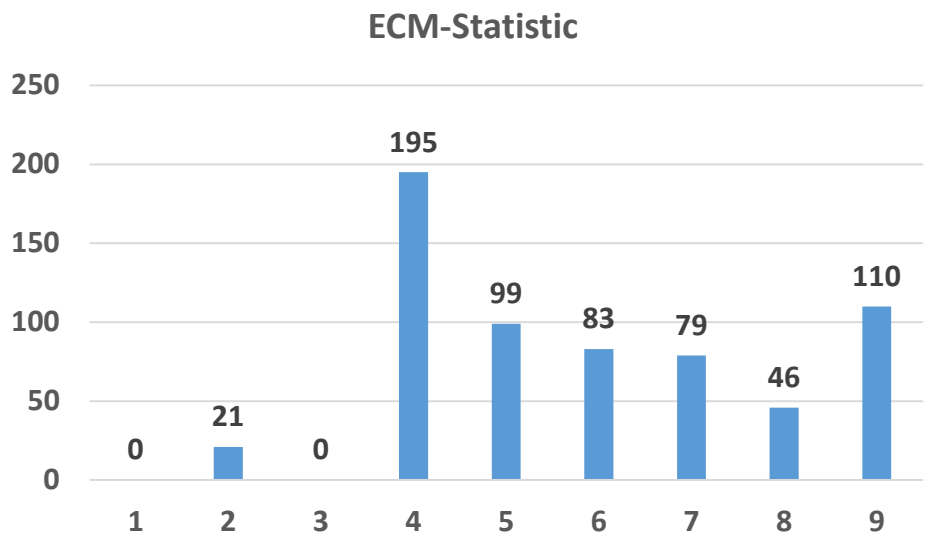
主要内容

第1部分	研究背景
第2部分	基于机器学习算法的等效电路预测
第3部分	基于全局优化算法的等效电路参数识别
第4部分	总结与展望

基于机器学习算法的等效电路预测

➢ 构建机器学习数据集





数据来源：文献收集(锂离子电池测试)[1]+高通量微区EIS实验[2]

数据统计：**7/9个类别，1000+**条EIS（**更多的数据**）

随机划分数据集：训练集(70%)、验证集(15%)、测试集(15%)

[1] Zhu, S., Sun, X., Wang, Y., Zhao, N. & Sha, J. Equivalent circuit model recognition of electrochemical impedance spectroscopy via machine learning. *J. Electroanal. Chem*. 855, 113627 (2019).
[2] Lai, Z., Bi, P., Lei, W., Xue, Y. & Jin, Y. Local electrochemical properties of fusion boundary region in sa508-309l/308l overlay welded joint. *Corros. Sci*. 155, 75-85 (2019).

基于机器学习算法的等效电路预测

➢ 训练多种机器学习模型——通过网格搜索调节超参数

- **机器学习（监督学习）定义**

利用有标签的训练数据不断修正模型来进行(多)分类任务

- **设置网格搜索的参数空间**

机器学习	超参数	重复训练次数	评价指标
K最近邻	最近邻数，距离计算方法	1	Accuracy, Kappa, AK
多分类逻辑斯蒂回归分类器	多分类策略、梯度下降（学习率、迭代次数）	10	$\overline{Accuracy}, \overline{Kappa}, \overline{AK}$
决策树	剪枝策略	1	Accuracy, Kappa, AK
多分类支持向量机	多分类策略、核函数、迭代次数	10	$\overline{Accuracy}, \overline{Kappa}, \overline{AK}$
随机森林	随机树数量	10	$\overline{Accuracy}, \overline{Kappa}, \overline{AK}$
AdaBoost	基学习器数量	10	$\overline{Accuracy}, \overline{Kappa}, \overline{AK}$
朴素贝叶斯	-	1	Accuracy, Kappa, AK

- **网格搜索**

是一项模型**超参数优化**技术，常用于优化三个或者更少数量的超参数，是一种**穷举法**。使用每组超参数训练模型，挑选验证集误差最小的超参数作为最好的超参数

- 网格搜索过程设置
  - ■ 在训练集上训练算法
  - ■ 在验证集上评价算法性能（准确率Accuracy、Kappa、Accuracy+Kappa=AK，三种）

- 算法性能评价指标
  - ✦ 准确率 $Accuracy$ 较片面
  - ✦ 平均准确率 $\overline{Accuracy} = \frac{\text{重复10次的Accuracy}}{10}$
  - ✦ $Kappa$：衡量多分类精度的一种指标 更全面
  - ✦ $\overline{Kappa} = \frac{\text{重复10次的}Kappa}{10}$
  - ✦ $AK = Acuuracy + Kappa$
  - ✦ $\overline{AK} = \overline{Accuracy} + \overline{Kappa}$

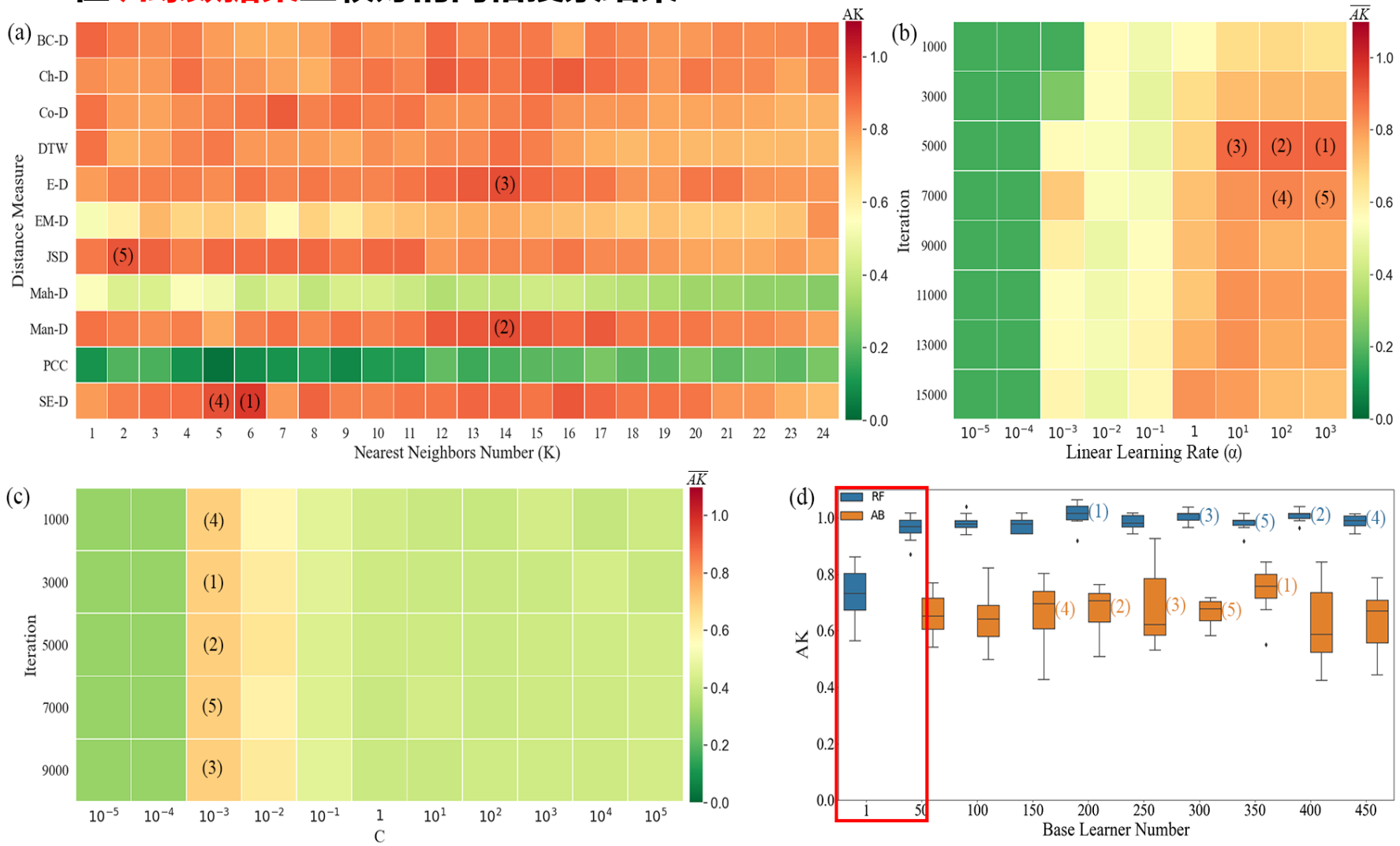
各种标准均是数值越大越好

- 性能计算举例

基于机器学习算法的等效电路预测

➢ 训练多种机器学习模型——通过网格搜索调节超参数

- **在训练数据集上较好的网格搜索结果**



(a) K最近邻

- ➢ 选EM-D或PCC时：AK随K增大而增大
- ➢ 选Mah-D时，AK随K增大而减小
- ➢ 选其他方法时，AK随K呈随机变化

(b) 多分类逻辑斯蒂回归

- ➢ 当固定α，AK随迭代次数的增加先增加再减小
- ➢ 当固定迭代次数，AK随α的增大而持续增大

(c) 多分类支持向量机

- ➢ 迭代次数对AK影响不明显
- ➢ 当迭代次数固定，AK随常数C增大先增大再减小

(d蓝) 随机森林

AK随基学习器的数量增加先增大再趋于平稳，且波动较小

(d黄) AdaBoost

AK与基学习器数量无明显关联趋势，但波动较大

其余的网格搜索结果

基于机器学习算法的等效电路预测

➢ 训练多种机器学习模型——通过网格搜索调节超参数

- **各种机器学习在测试集上的泛化性能**

机器学习	Accuracy	Kappa	AK	按AK排序
AdaBoost	**0.571**	**0.434**	**1.045**	**1**
随机森林	**0.582**	**0.460**	**1.043**	**2**
K最近邻	0.538	0.384	0.923	3
多分类逻辑斯蒂回归分类器	0.495	0.381	0.875	4
朴素贝叶斯	0.495	0.375	0.87	5
决策树	0.495	0.371	0.865	6
多分类支持向量机	0.462 / **0.48**	0.248	0.709	7

最优的机器学习模型

机器学习模型的选择流程

1 性能较优

2 推断依据具备可解释性—
算法做出预测的理由和人类经验、公式、常识等一致

明确为什么好?

机器学习的可解释性

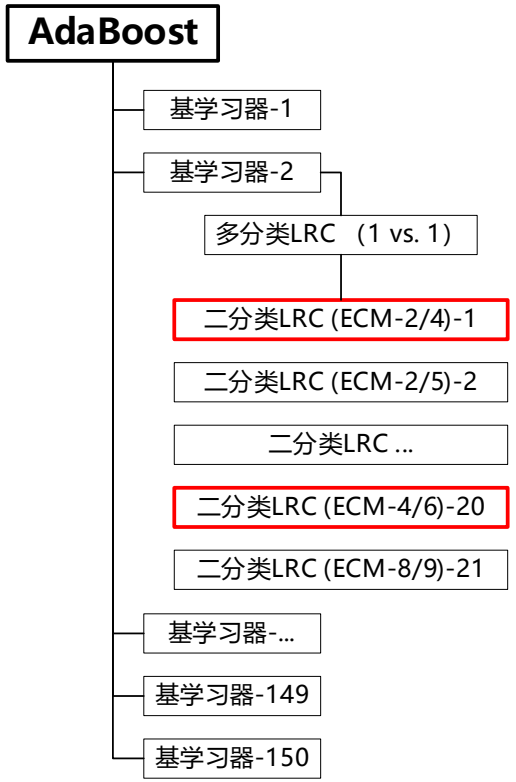
- **人工经验**

ML的数据集中有近70%的数据来自锂离子电池的测试，其EIS的低频、中频上阻抗与锂离子多种特性（离子传输、电极嵌入）相关，是实验人员较为关注的部分。所以针对本工作的数据集，我们倾向于EIS各个频率范围内阻抗数据对ML影响的权重大致为：低频>中频>高频

基于机器学习算法的等效电路预测

➢ 机器学习的可解释性

● Adaboost的可解释性

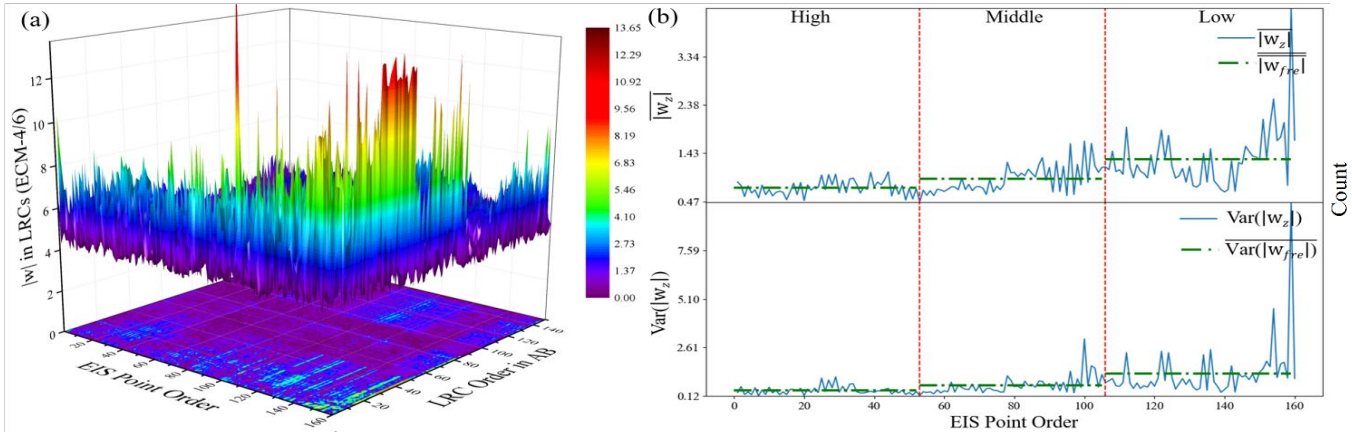


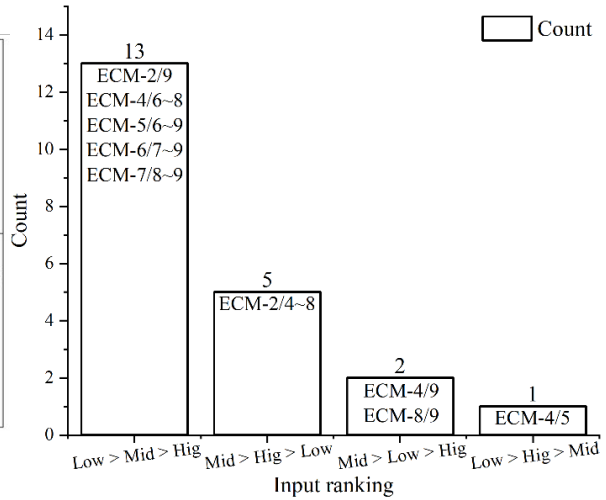
二分类器=逻辑斯蒂回归

$$y = \frac{1}{1 + e^{-(w^T x + b)}}$$

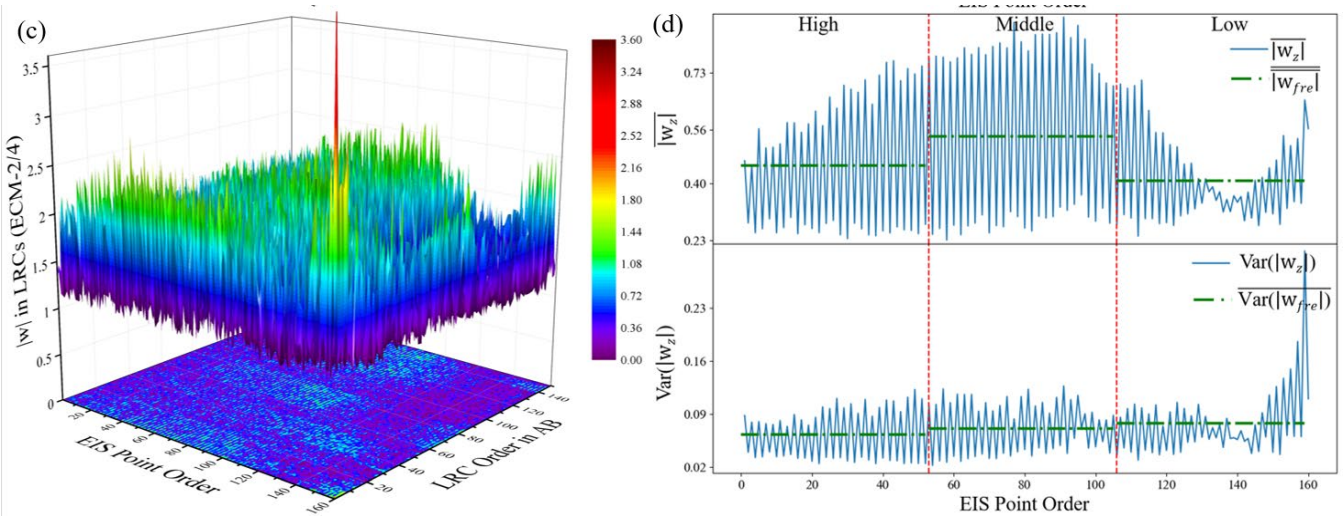
|权重| ≈ 重要性

第一种主要的权重分布（左a）及统计规律（右b）
低频>中频>高频





第二种主要的权重分布（左c）及统计规律（右d）
中频>高频>低频



统计结果显示模型内权重主要分布为：

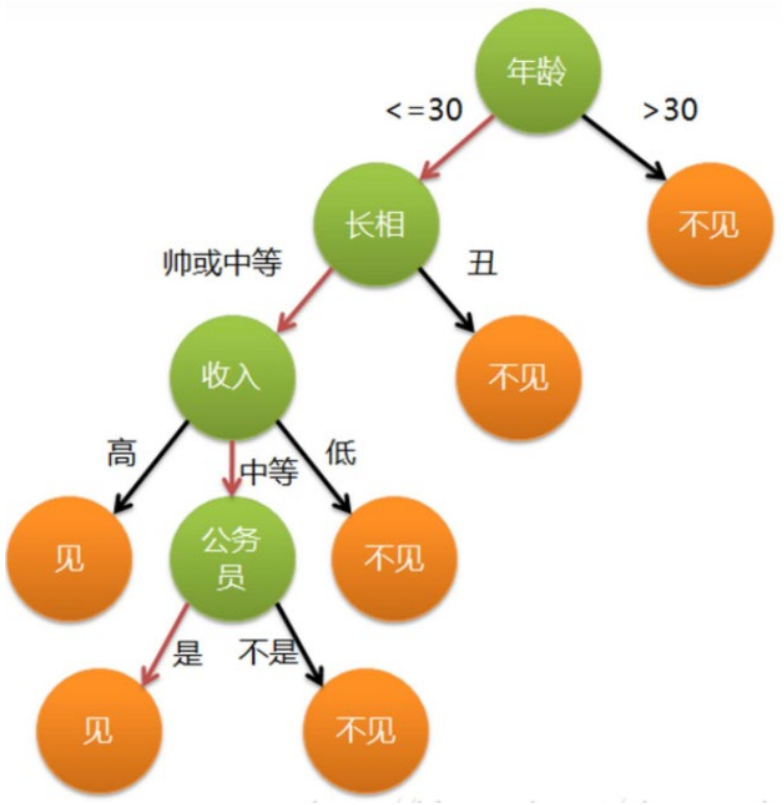
低频>中频>高频

基于机器学习算法的等效电路预测

➢ 机器学习的可解释性

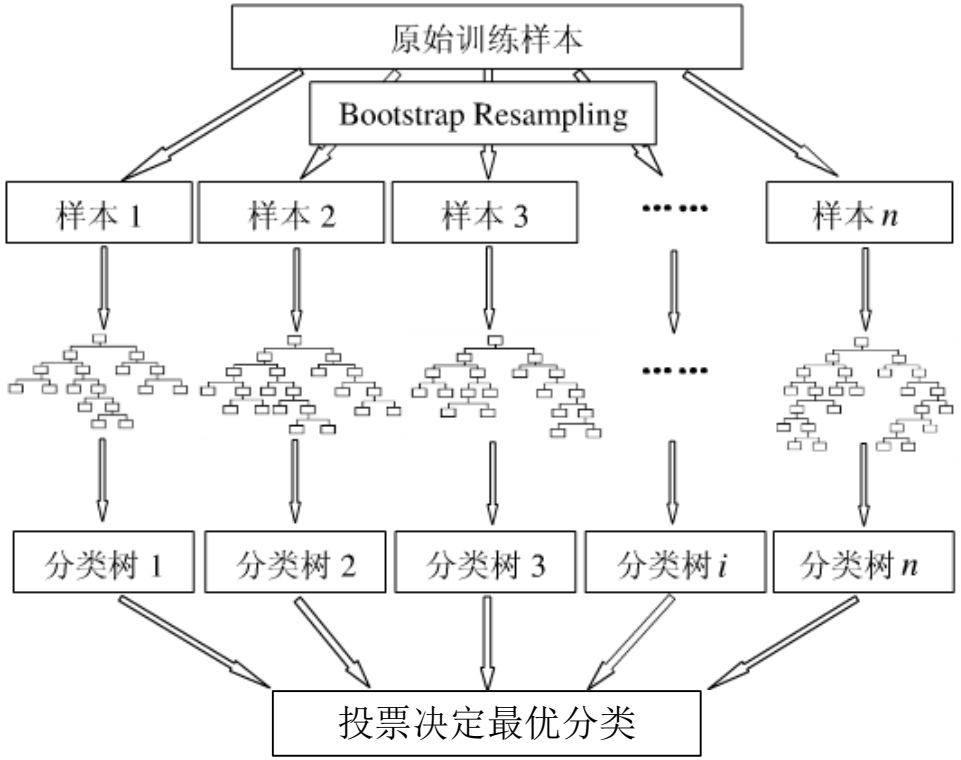
● 随机森林的可解释性

决策树
选择男朋友

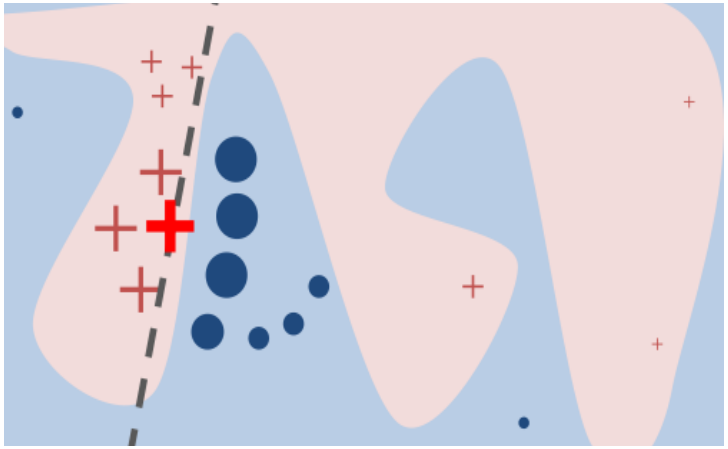


决策过程清晰可见

随机森林



Local Interpretable Model-Agnostic Explanations[1] (LIME)
不可知模型的局部解释



如上图是一个非线性的复杂模型，蓝/粉背景的交界为决策函数，选取关注的样本点，如图粗线的红色十字叉为关注的样本点X，定义一个相似度计算方式，以及要选取的K个特征来解释，在该样本点周围进行扰动采样（细线的红色十字叉），按照它们到X的距离赋予样本权重，用原模型对这些样本进行预测，并训练一个线性模型（虚线）在X的附近对原模型近似

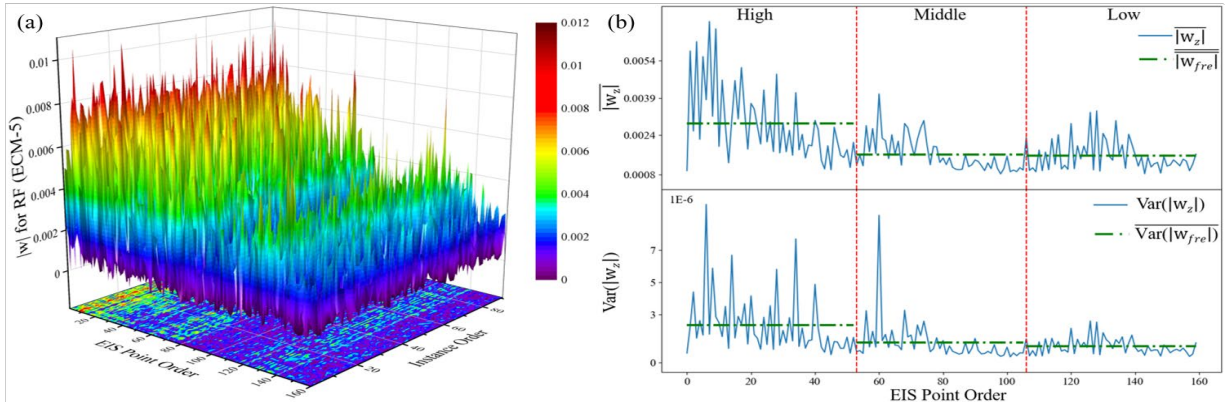
[1] “Why Should I Trust You?” Explaining the Predictions of Any Classifier https://doi.org/10.1145/2939672.2939778

基于机器学习算法的等效电路预测

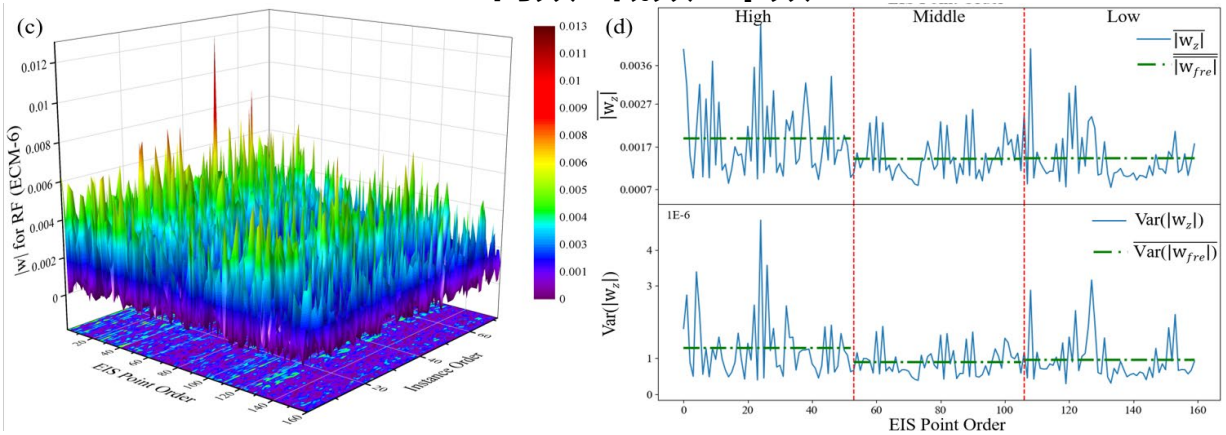
➢ 机器学习的可解释性

● 随机森林的可解释性

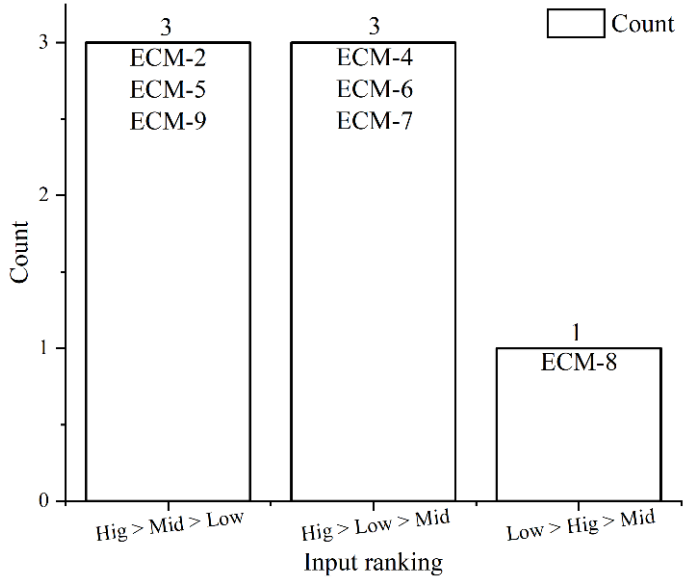
第一种主要的权重分布（左a）及统计规律（右b）
高频>中频>低频



第二种主要的权重分布（左c）及统计规律（右d）
高频>低频>中频



|权重| ≈ 重要性



统计结果显示模型内主要的权重分布为：

低频<中频≈高频

基于机器学习算法的等效电路预测

➢ 机器学习的可解释性

● Fuzzy Curve[1]

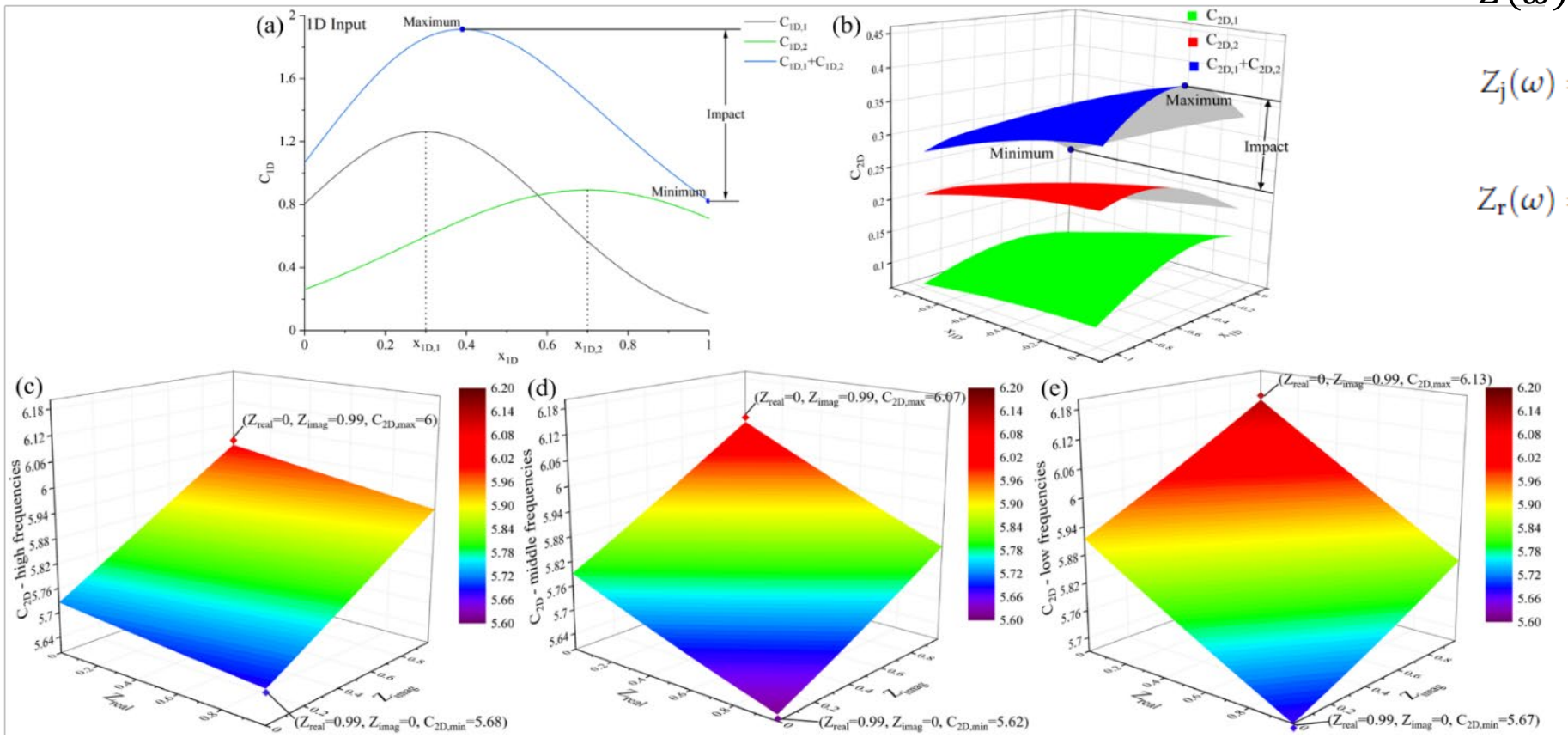
能识别对分类任务重要的输入变量，计算简单，广泛用于定量的评价一维变量的影响(Impact) **Impact ≈ 重要性**

一维变量Fuzzy Curve计算示例 **二维关联变量Fuzzy Curve计算示例**

$$Z(\omega) = Z_r(\omega) + i * Z_j(\omega)$$

$$Z_j(\omega) = \frac{2\omega}{\pi} \int_0^\infty \frac{Z_r(x) - Z_r(\omega)}{x^2 - \omega^2} dx$$

$$Z_r(\omega) = Z_r(\infty) - \frac{2}{\pi} \int_0^\infty \frac{x Z_j(x) - \omega Z_j(\omega)}{x^2 - \omega^2} dx$$



低频>中频>高频

$Impact_{高频阻抗}$ = 0.32 < **$Impact_{中频阻抗}$ = 0.45 <** **$Impact_{低频阻抗}$ = 0.46**

[1] A Fuzzy Approach to Input Variable Identification 10.1109/FUZZY.1994.343528

➢ 小结

✓ 选择最佳的机器学习算法

使用七种机器学习算法为EIS预测ECM，通过网格搜索得到每种算法的最优超参数，最终各种算法在测试集上的性能排名为：AdaBoost>随机森林>K最近邻>多分类逻辑斯蒂回归>朴素贝叶斯>决策树>多分类支持向量机

✓ 探索最佳机器学习算法的可解释性

通过分析AdaBoost和随机森林中权重的分布得到二者的权重分布分别为【低频>中频>高频】和【低频<中频≈高频】，再通过Input Ranking——Fuzzy Curve技术识别模型的重要输入排序【低频>中频>高频】，最后判断AdaBoost的推断依据更符合人工经验

主要内容

第1部分	研究背景
第2部分	基于机器学习算法的等效电路预测
第3部分	基于全局优化算法的等效电路参数识别
第4部分	总结与展望

基于全局优化算法的等效电路参数识别

➢ 评价GOA在理想EIS上参数识别的性能

- **全局优化算法（Global Optimization Algorithm，GOA）**

优点：参数搜索范围广，只需要计算函数目标值，对目标函数不需要进行求导，实现简单

缺点：种类繁多、没有理论证明某种算法在具体问题上的优越性，只有实验证明

基于自然进化

差分进化
Differential Evolution (DE)

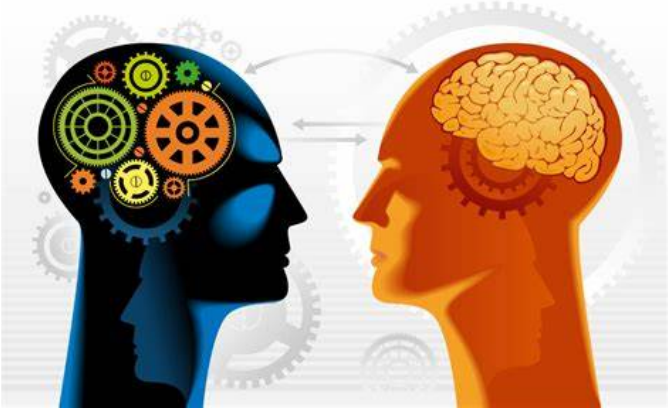
分布估计算法
Estimation of Distribution Algorithm (EDA)

进化规划
Evolution Programming (EP)

进化策略
Evolution Strategy (ES)

遗传算法
Genetic Algorithm (GA)

基于人类活动



群搜索算法
Group Search Optimizer (GSO)

和弦搜索
Harmony Search (HS)

帝国竞争算法
Imperialist Competitive Algorithm (ICA)

室内搜索算法
Interior Search Algorithm (ISA)

教与学优化算法
Teaching Learning Based Optimization (TLBO)

基于动物群体行为

蚁群算法
Ant Colony Optimization (ACO)

人工蜂群算法
Artificial Bee Colony (ABC)

灰狼优化
Grey Wolf Optimizer (GWO)

粒子群算法
Particle Swarm Optimization (PSO)

鲸群优化算法
Whale Optimization Algorithm (WOA)

基于物理定律

大爆炸-大收缩
Big Bang Big Crunch (BBBC)

黑洞
Black Hole (BH)

充电系统搜索
Charged System Search (CSS)

重力搜索算法
Gravitational Search Algorithm (GSA)

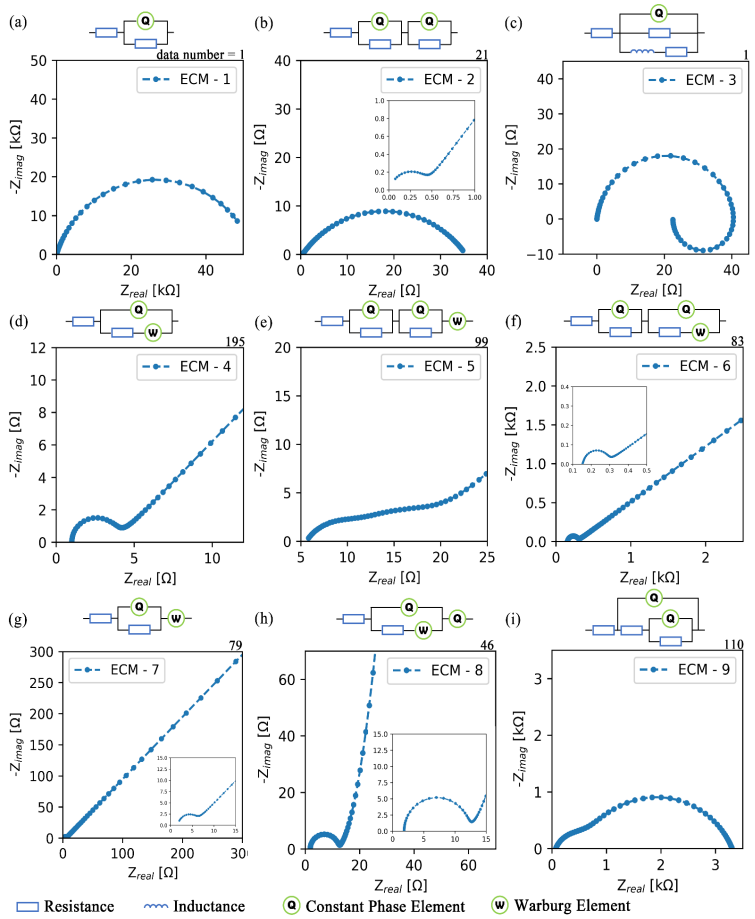
平行宇宙优化器
Multi-verse Optimizer (MVO)

基于全局优化算法的等效电路参数识别

➢ 评价GOA在理想EIS上参数识别的性能

● **GOA评价流程：** 每种GOA拟合一条理想的EIS得到一组结果（卡方误差χ^2、计算时间t），重复100次

● **测试数据：9种理想的EIS**



● **GOA拟合等效电路模型(ECM)的性能评价标准**

✓ **卡方误差**

$$\chi^2 = \sum_{i=1}^{N}\left[\frac{Z_{Re}(\omega_i) - \hat{z}_{Re}(\omega_i)}{|Z(\omega_i)|}\right]^2 + \left[\frac{Z_{Im}(\omega_i) - \hat{Z}_{Im}(\omega_i)}{|Z(\omega_i)|}\right]^2$$

✓ **平均拟合精度**

$$\overline{\chi^2} = \frac{\sum_{i=1}^{100}\chi_i^2}{100}$$

✓ **稳定性**

$$RMSE(\chi^2) = \sqrt{\frac{\sum_{i=1}^{N}\left(\chi_i^2 - \overline{\chi^2}\right)^2}{N}}$$

✓ **平均计算时间**

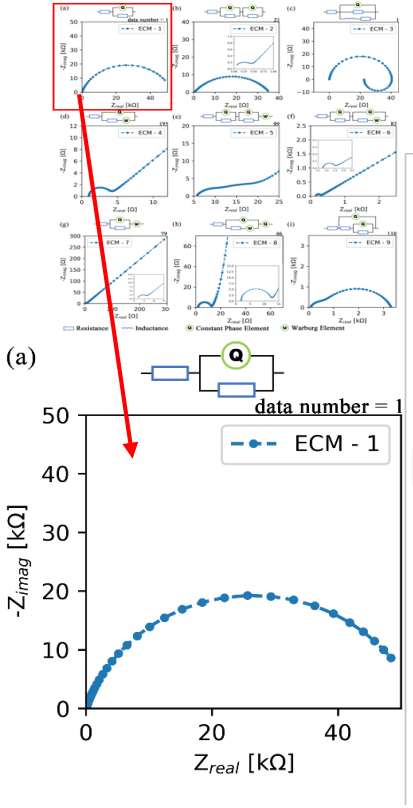
$$\bar{t} = \frac{\sum_{i=1}^{100} t_i}{100}$$

各种指标均是越小越好

基于全局优化算法的等效电路参数识别

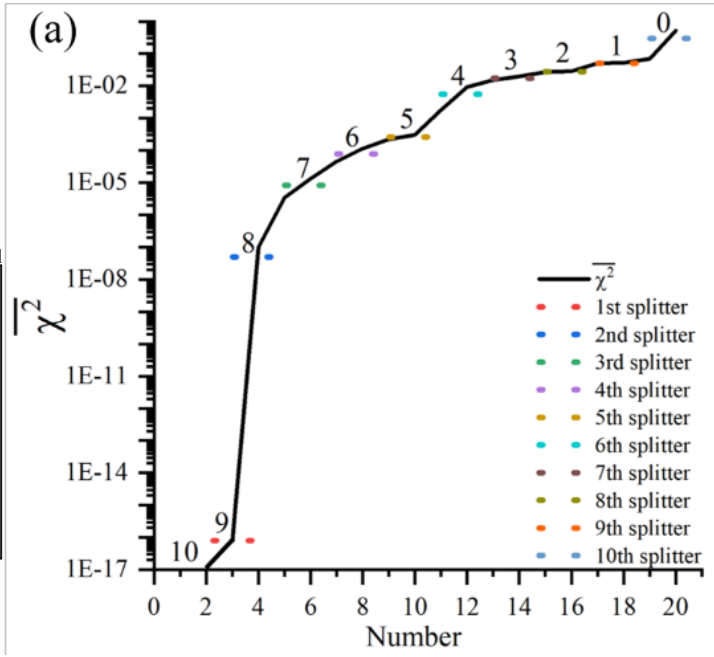
➢ 评价GOA在理想EIS上参数识别的性能

● 20种GOA对于第一条理想的EIS的评价过程

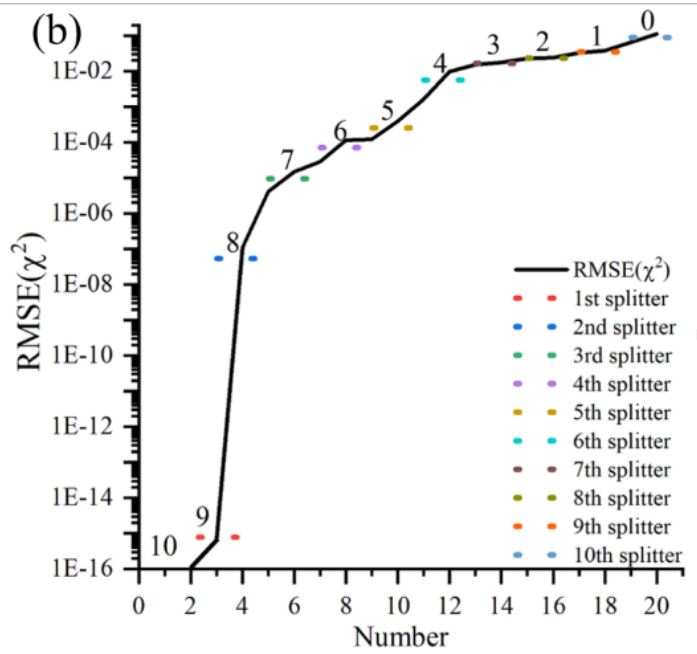


将20种GOA的三种指标从小到大排序，最小的指标得10分，最大的指标得0分

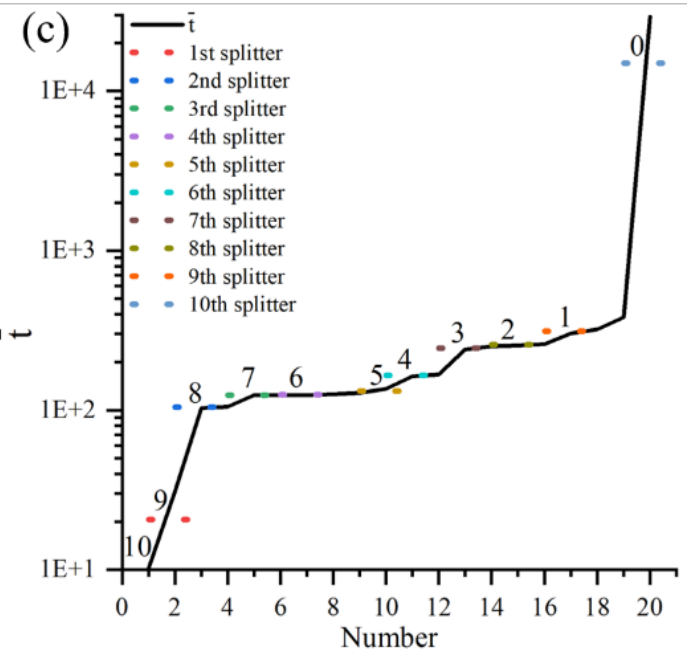
平均拟合精度



稳定性



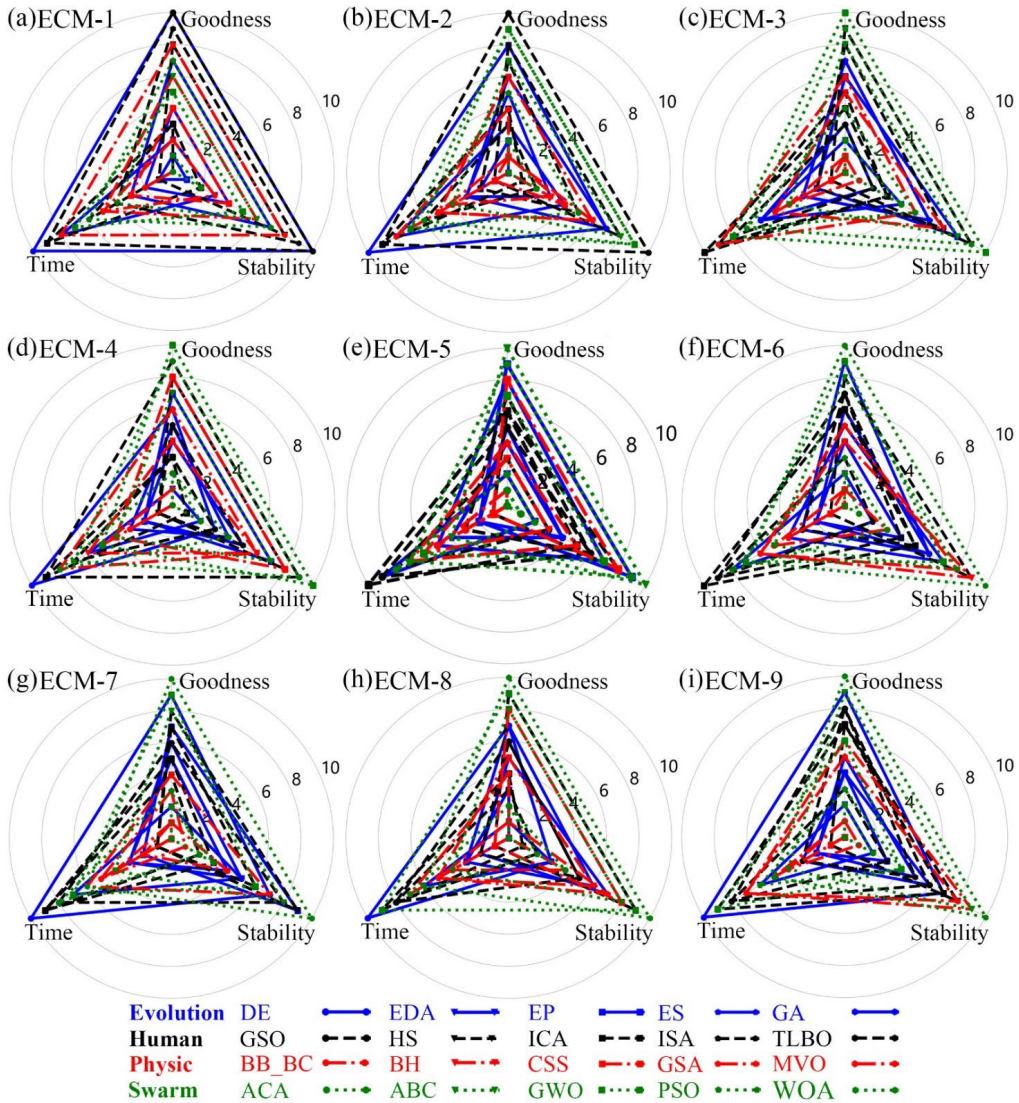
平均计算时间



基于全局优化算法的等效电路参数识别

➢ 评价GOA在理想EIS上参数识别的性能

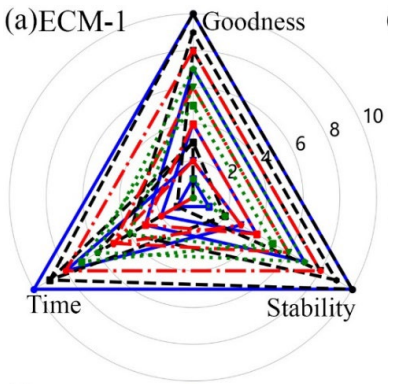
- **20种GOA在9种ECM上的性能打分**



- **为一种ECM挑选合适的GOA**

以ECM-1的拟合结果为例

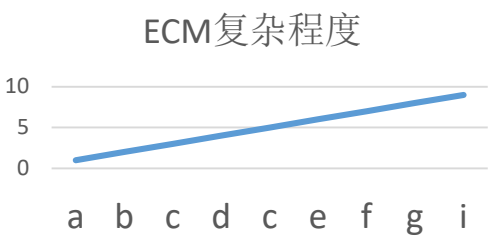
雷达图中最大的三角形为blue 直线，即DE算法，所以DE拟合ECM-1的综合性能最好



- **一类GOA适合拟合哪些ECM**

以基于动物群体活动启发的GOA为例：计算时间较长，但平均拟合精度和稳定性较好，适合处理较为复杂的ECM

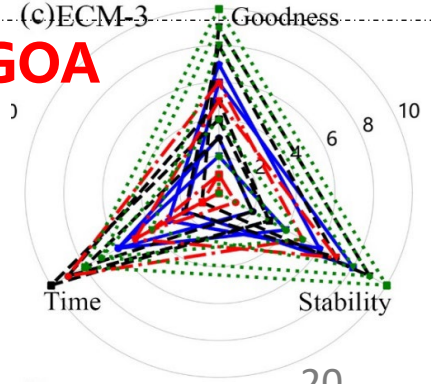
平均拟合精度	稳定性	平均计算时间
(a) (b) (c)	(a) (b) (c)	(a) (b) (c)
(d) (e) (f)	(d) (e) (f)	(d) (e) (f)
(g) (h) (i)	(g) (h) (i)	(g) (h) (i)
好	不好	



- **为特定ECM拟合的特定需求挑选合适的GOA**

以ECM-3的拟合结果为例

以平均拟合精度选	➡	GWO
以平均计算时间	➡	ICA
以稳定性选	➡	GWO
以综合型性能（三角形面积）选	➡	GWO

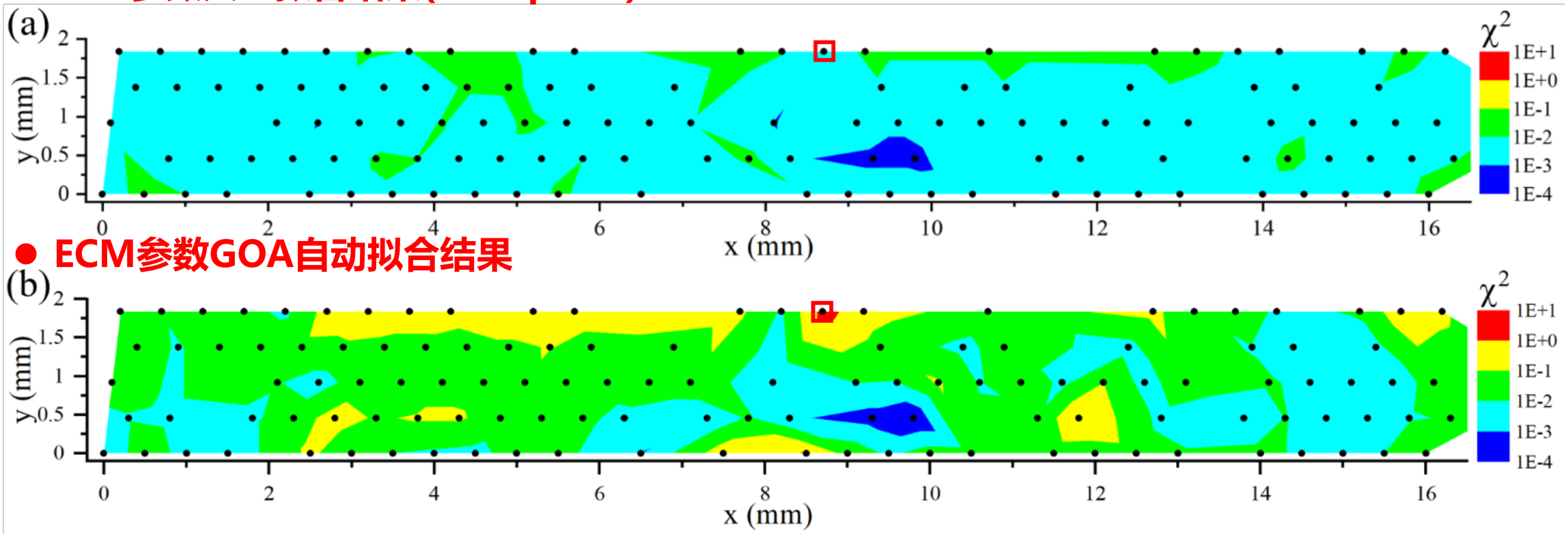


基于全局优化算法的等效电路参数识别

➢ 验证选出的GOA自动拟合参数的能力，并与人工结果对比

EIS实验背景：微区，高通量，SA508-309L/308L焊接接头+盐水溶液，5行*30列约150个测试点

- **ECM参数人工拟合结果(ZSimpWin)**



- **ECM参数GOA自动拟合结果**

✓ EIS数据质量较高的情况下：(a)和(b)深蓝色区域一样，人工和GOA拟合取得精度相近的结果

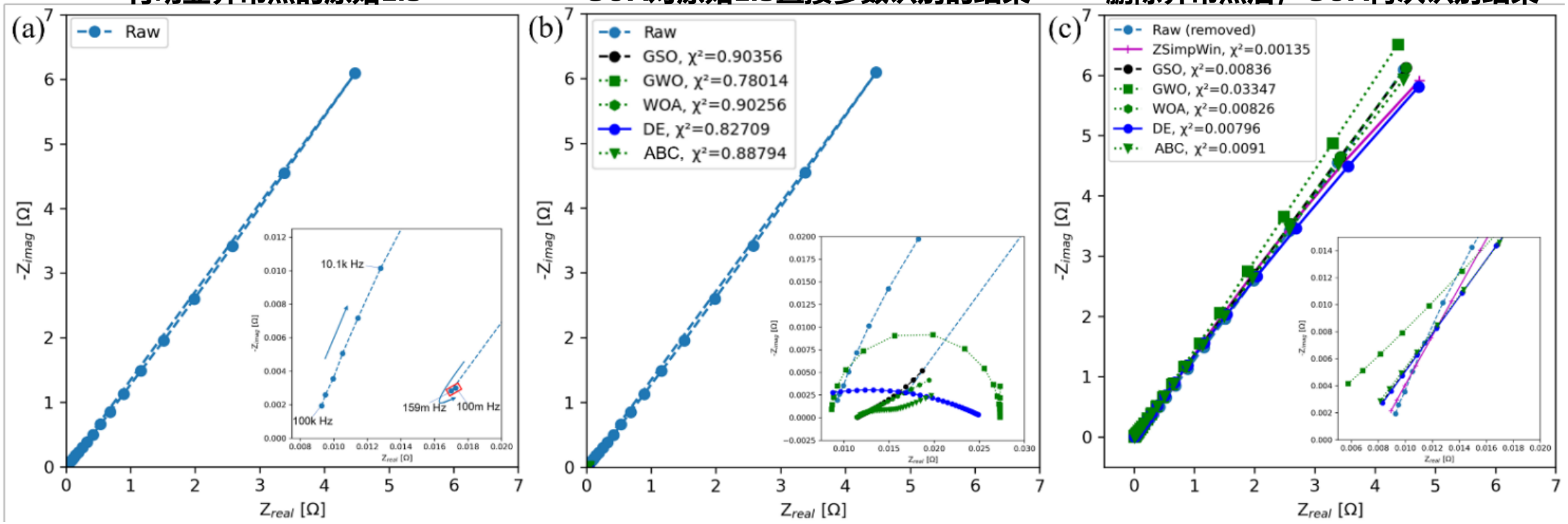
✓ EIS数据质量一般的情况下：(a)大部分的浅蓝色和(b)大部分的绿色分布接近：人工拟合优于GOA

✓ EIS数据质量较差的情况下：(a)和(b)中样品边缘区域：人工拟合显著优于GOA

基于全局优化算法的等效电路参数识别

➢ 异常点对GOA参数拟合结果的影响

有明显异常点的原始EIS　　GOA对原始EIS直接参数识别的结果　　删除异常点后，GOA再次识别结果



微区EIS测试 50 ~ 500 um、较小的导电离子扩散速率，EIS经常出现类似直线的扩散控制

- ✓ 异常点的存在会严重干扰GOA拟合ECM参数结果
- ✓ 在无明显异常点的情况下，GOA能和人工使用ZSimpWin软件取得相近的拟合精度

➢ 小结

✓ 为ECM选择最佳的GOA

- ➢ 通过对4大类20种GOA在9种理想的EIS上进行平均计算时间、稳定性、平均拟合精度三方面测评，选出了每种EIS拟合最适宜的GOA
- ➢ 为GOA拟合ECM的性能评比问题上提供一个比较基准

✓ 在实验数据上验证拟合精度

通过选出的GOA在实验获得的EIS上进行测评，在无明显异常点干扰的条件下，GOA能和人工使用ZSimpWin软件取得相近的拟合精度

主要内容

第1部分	研究背景
第2部分	基于机器学习算法的等效电路预测
第3部分	基于全局优化算法的等效电路参数识别
第4部分	总结与展望

总结

- **基于可解释的机器学习算法预测等效电路**

  ✓ **选择最佳的机器学习算法**

  使用七种机器学习算法为EIS预测ECM，通过网格搜索得到每种算法的最优超参数，最终各种算法在测试集上的性能排名为：**AdaBoost>随机森林**>K最近邻>多分类逻辑斯蒂回归>朴素贝叶斯>决策树>多分类支持向量机

  ✓ **探索最佳机器学习算法的可解释性**

  通过分析AdaBoost和随机森林中权重的分布得到二者的权重分布分别为：【低频>中频>高频】和【低频<中频≈高频】，再通过Input Ranking——Fuzzy Curve技术识别模型的重要输入排序【低频>中频>高频】，最后判断**AdaBoost的推断依据更符合人工经验**

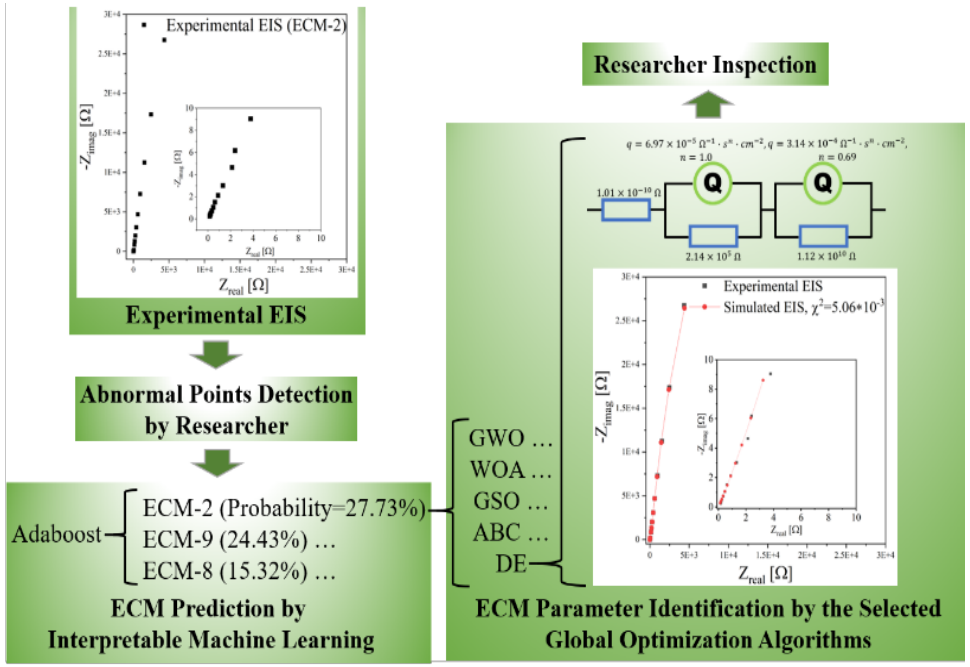
- **基于全局优化算法识别等效电路的参数**

  ✓ **为ECM拟合选择最佳的GOA**

  通过对20种GOA在9种理想的EIS上进行平均计算时间、稳定性、平均拟合精度三方面测评，我们选出每种EIS拟合最适宜的GOA

  ✓ **在实验数据上进一步验证拟合精度**

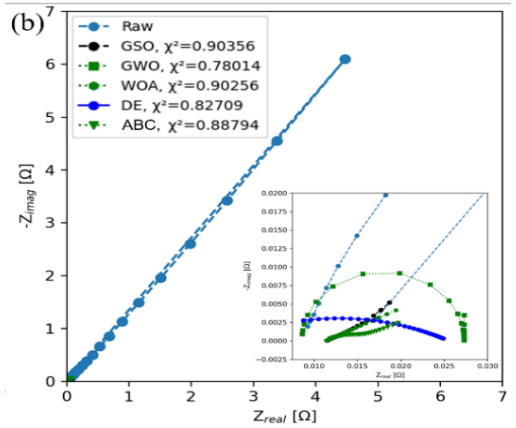
  通过选出的GOA在实验的EIS上进行测评，在无明显异常点干扰的条件下，GOA能和人工使用ZSimpWin软件取得相近的拟合精度



整体工作的流程图

展望

➢ EIS异常点定量化检测与排除



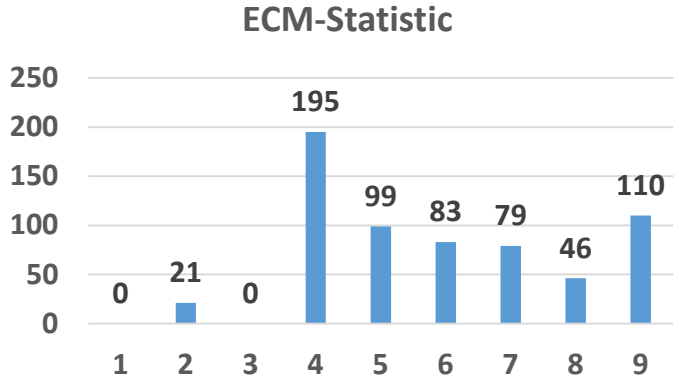
异常点的存在会干扰ECM参数拟合

➢ 提升机器学习的能力

➢ 处理多源数据的能力

EIS测试的频率范围不同、每Decade的测试点数不同造成各种长度不同的EIS数据，目前一个ML模型只能接受固定长度的EIS数据

➢ 处理体量小、分布不平衡数据的能力



➢ 利用ML直接求解被研究系统

偏微分方程：

$$\frac{\partial^2 f}{\partial x^2}+\frac{\partial^2 f}{\partial y^2}=0$$

$$\frac{\partial f}{\partial t}=\alpha\frac{\partial^2 f}{\partial x^2}$$

$$\frac{\partial^2 f}{\partial t^2}=c^2\frac{\partial^2 f}{\partial x^2}$$

➢ 搭建在线EIS智能处理平台

智能

开放共享

不断迭代优化后台代码

欢迎各位同仁莅临国家科学中心交流指导！

敬请指正！

补充材料

➢ 机器学习评价标准计算示例

Actual Class / Predicted Class	A_{act}	B_{act}	C_{act}	Sum of row
A_{pre}	239	21	16	276
B_{pre}	16	73	4	93
C_{pre}	6	9	280	295
Sum of column	261	103	300	

Assess Criterions:

$$\boldsymbol{Accuracy} = \frac{Number\ of\ Correct\ Predictions}{Number\ of\ Samples}$$

$$\boldsymbol{Kappa} = \frac{p_o - p_e}{1 - p_e}$$

$$p_o = \text{Accuracy(Multiple Class)} = \frac{239 + 73 + 280}{664}$$

$$p_e = \frac{261 \times 276 + 103 \times 93 + 300 \times 295}{664 \times 664}$$

$$AK = Accuracy + Kappa$$

AK = 0 (Worst) ~ 2 (Best)

返回

补充材料

➢ 其余的网格搜索结果一

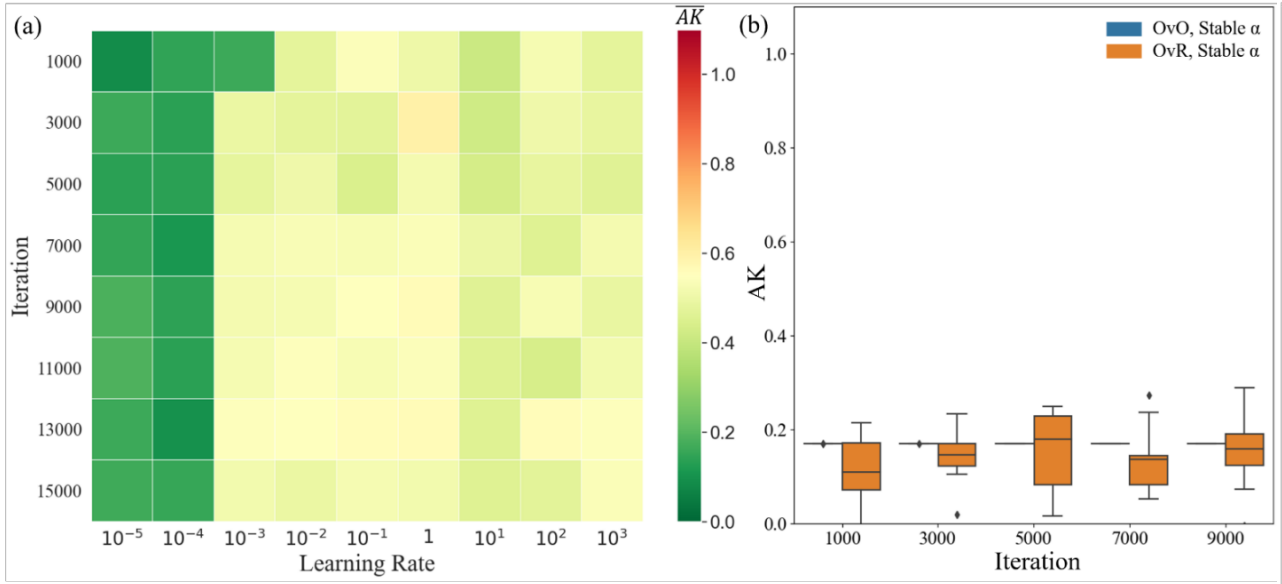


Fig. S1 (a) The grid search results of multiclassification LRC that takes OvR decomposition strategy, different iteration, and different initial values for linear learning rate in gradient descent. (b) The grid search results of multiclassification LRC that, respectively, takes OvO (blue) and OvR (yellow) decomposition strategies and stable learning rates in gradient descent.

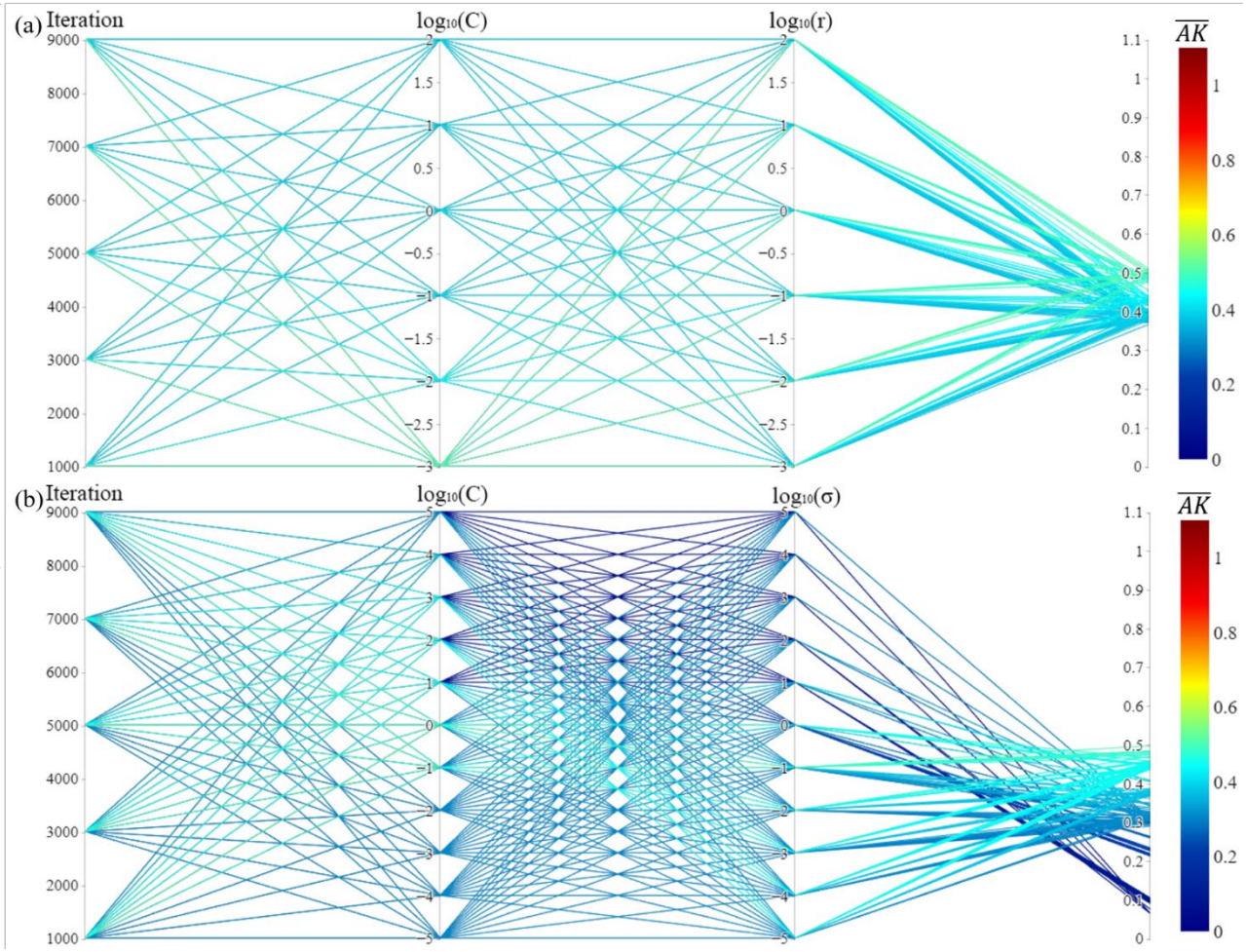
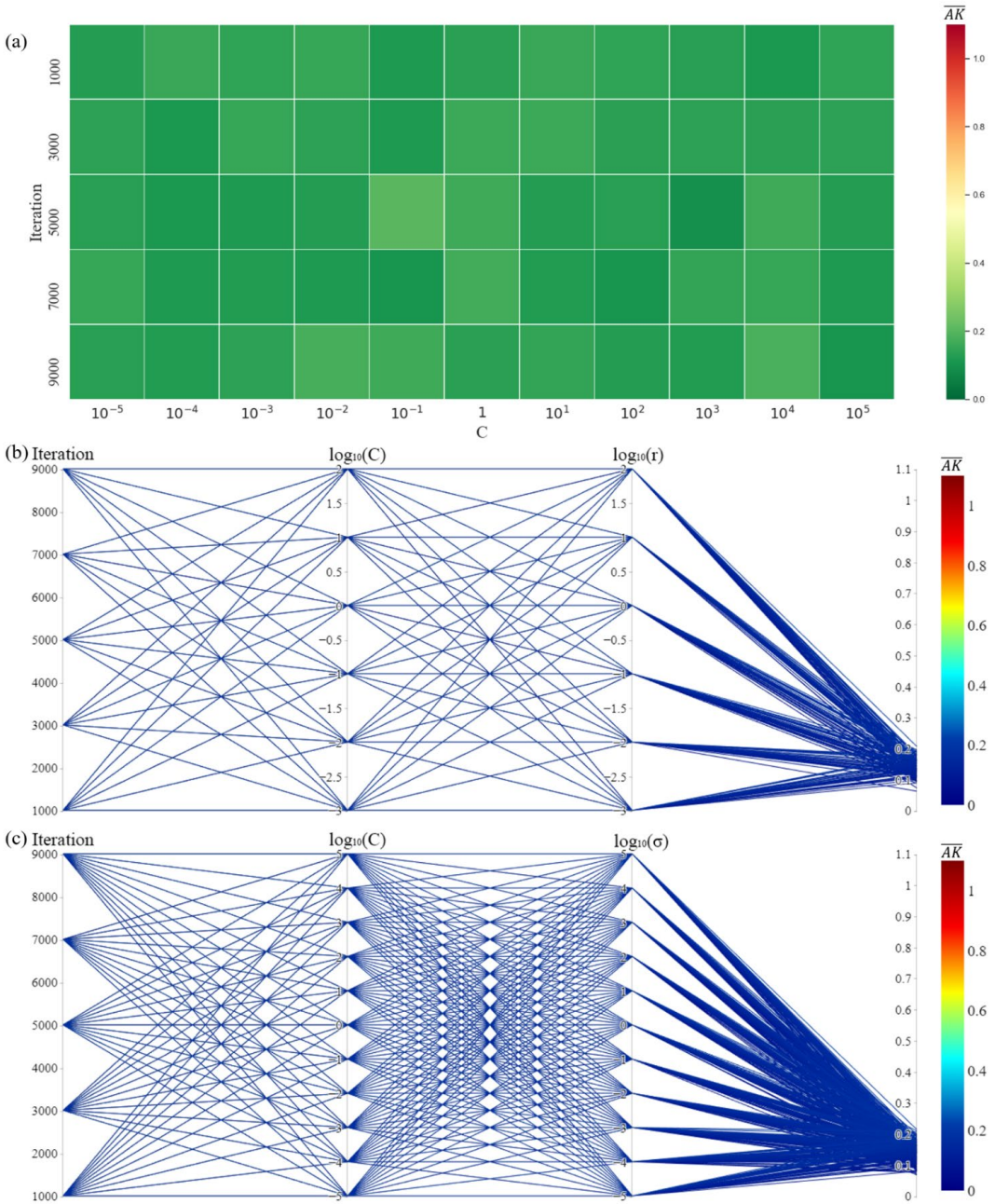


Fig. S2 The grid search results of multiclassification SVM that takes OvO decomposition strategy and, respectively, has (a) polynomial and (b) radial basis function kernels.

补充材料

➢ 其余的网格搜索结果二

Fig. S3 The grid search results of multiclassification SVM that takes OvR decomposition strategy and, respectively, has (a) linear, (b) polynomial, and (c) radial basis function kernels.



返回

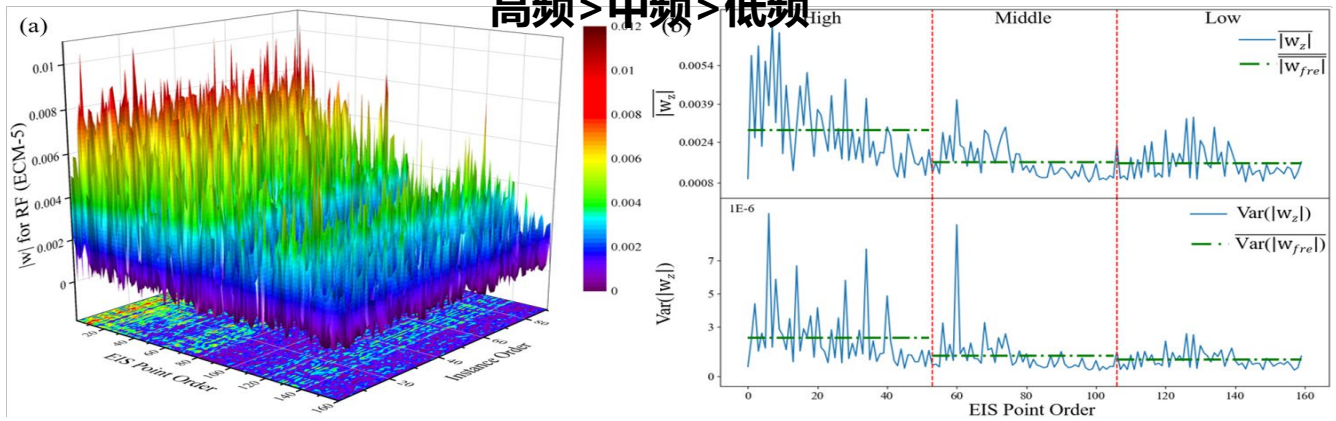
基于机器学习算法的等效电路预测

➢ 机器学习的可解释性

- **随机森林的可解释性**

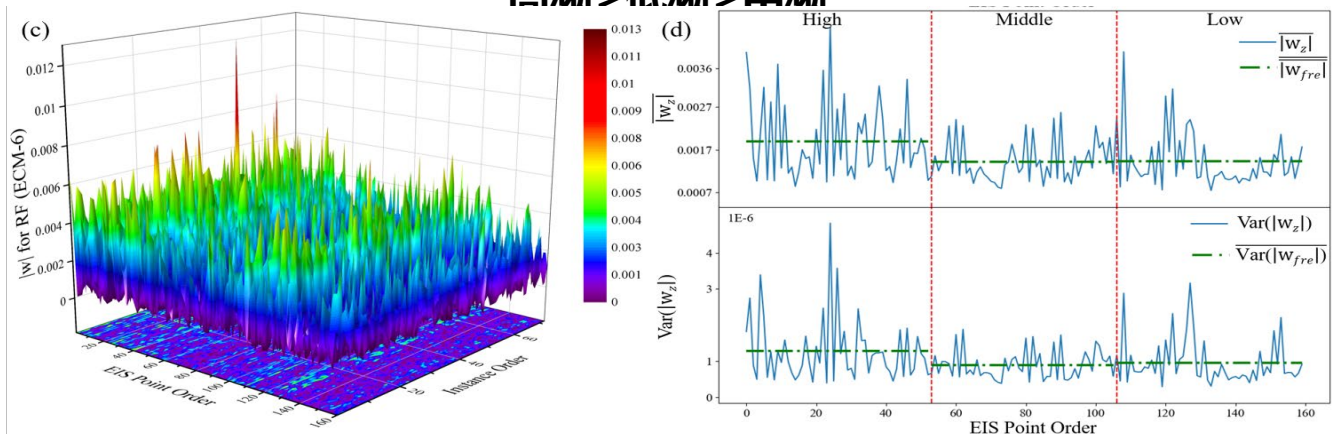
第一种主要的权重分布（左a）及统计规律（右b）

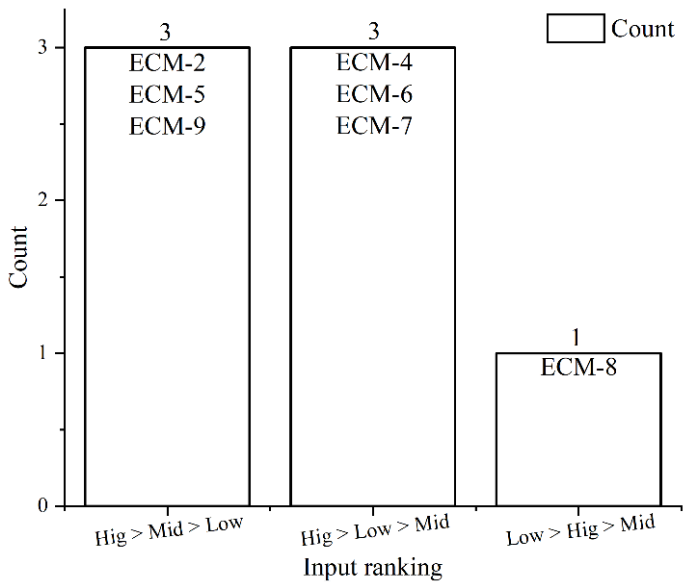
高频>中频>低频



第二种主要的权重分布（左c）及统计规律（右d）

高频>低频>中频





权重 ≈ 重要性

统计模型内主要的权重分布
低频<中频≈高频