



中国研究生创新实践系列大赛
“华为杯”第二十一届中国研究生
数学建模竞赛

学 校

浙江工商大学

参赛队号

24103530064

王东江

队员姓名

武昱泽

徐龙哲

中国研究生创新实践系列大赛

“华为杯”第二十一届中国研究生

数学建模竞赛

题 目： 数据驱动下磁性元件的磁芯损耗建模

摘 要：

随着电能变换技术在诸多新兴领域的广泛应用，作为电力电子系统中的关键组件，降低磁性元件的磁芯损耗问题是解决实际应用需求的关键痛点。因此，从磁芯损耗数据出发，考虑损耗的诸多影响因素，构建具有普适性和稳定性的损耗预测模型，可以为磁性元件设计提供精确的损耗评估，是推动电力电子技术进一步发展的重要手段。

针对问题一，首先绘制依据温度和磁芯材料分组的各采样点励磁波形曲线图，分析波形曲线在不同温度、磁芯材料下呈现出的增减及波动模式；其次，通过磁通密度数据计算励磁波形的 8 个时域特征以及 4 个频域特征，以作为励磁波形分类的特征依据；然后，利用随机森林方法绘制特征重要性图并提取出峭度、重心频率、偏度、均方根频率、最高峰与最低峰的时间差、频率方差以及最高峰的时间点共 7 个特征；最后，通过逻辑回归、支持向量机、随机森林、XGBoost 等机器学习算法分别构建励磁波形分类模型，实验得到四个模型的分类准确率均超过了 0.99，故选取随机森林模型对附件二测试数据进行分类，得到分类结果为三角波 44 个、正弦波 20 个、三角波 16 个。

针对问题二，首先根据附件一材料 1 中 1067 条正弦波数据，采用最小二乘法拟合斯坦麦茨（Steinmetz-equation,SE）方程，以 8:2 的比例划分训练集和测试集，通过分析拟合结果与实际数据，计算出 MAE、MSE、MRE、 R^2 四个模型性能评价指标，其中 R^2 为 0.9448；下一步猜想将温度因素引入 SE 方程的三种形式：幂函数、多项式函数、指数函数，再次利用最小二乘法分别拟合以上三种形式，发现以幂函数形式将温度变量引入 SE 方程的效果最好，此时 R^2 为 0.9953，故将其选作修正后的方程；最后，对比 SE 方程修正前后的各评价指标，并绘制 SE 方程修正前后各样本点下实际损耗与预测损耗的曲线图、误差曲线图以及实际损耗与预测损耗的对比图，可视化方程修正前后的预测效果，修正后的方程从各评价指标来看均优于原方程。

针对问题三，首先根据数据分别绘制不同温度、不同材料、不同励磁波形下磁芯损耗的箱线图以及磁芯损耗均值的直方图，进行探索性分析；其次，根据不同波形和材料将数据划分为 12 组，通过计算各组 Spearman 相关系数，说明数值变量温度和磁芯损耗存在显著的相关性；通过多重 Kruskal-Wallis 检验，说明了类别变量磁芯材料、励磁波形各自与磁芯损耗存在显著相关性；接下来，利用双因素方差分析，说明温度与励磁波形、磁芯材料与励磁波形对磁芯损耗具有协同作用，而温度与磁芯材料对磁芯损耗不具有协同作用。再者由问题二的修正 SE 方程启发，将磁芯材料与励磁波形进行独热编码，以磁芯损耗为

因变量，温度、温度的平方项、磁芯材料、励磁波形、温度与励磁波形的交互项以及磁芯材料与励磁波形的交互项为自变量，建立多元回归模型。最后，将该多元回归模型作为目标函数，以温度、磁芯材料以及励磁波形的独热编码作为决策变量，将温度的取值范围作为约束条件，通过遗传算法求得磁芯损耗的最优值为 23413，取得最优值的条件是温度为 70℃、励磁材料为材料 4、励磁波形为正弦波。

针对问题四，首先考虑问题一中统计的 8 个时域特征、4 个频域特征以及温度、磁芯材料、励磁波形和频率，共得到 16 个特征，作为预测磁芯损耗的初步依据特征；其次，通过随机森林算法绘制特征重要性图，提取出磁通密度峰值、频率、温度、磁芯材料和励磁波形 5 个特征；接着，对磁芯材料与励磁波形两个类别变量进行独热编码，利用随机森林、GBDT、XGBoost 和 LightGBM 四种机器学习模型拟合数据，构建磁芯损耗的预测模型；结果显示四种模型的 R^2 均高于 0.97，其中 XGBoost 模型的拟合优度最高， R^2 为 0.9918，故选择此模型来预测附件三中的数据。其中样本序号 16 的磁芯损耗为 1126.3，样本序号 379 的磁芯损耗为 1297.1。

针对问题五，首先对类别变量磁芯材料以及励磁波形进行独热编码，分别设置 3 个以及 4 个 0-1 变量，并在此基础上添加温度、频率、最大磁通密度，总计 10 个决策变量，接着以最小磁芯损耗、最大传输磁能为目标函数，基于 0-1 规划，以温度、频率、磁通密度峰值的取值范围为约束条件，利用模拟退火算法建立双目标优化模型，最终求得磁性元件实现最优性能的条件为材料一、正弦波、90℃温度、129700HZ 频率、0.0173 磁通密度峰值，此时磁芯损耗为 854.58，传输磁能为 2240.06。

关键词：磁性元件、磁芯损耗、机器学习、独热编码、双目标优化、随机森林、模拟退火算法

目录

1	问题重述	5
1.1	引言	5
1.2	问题的提出	5
2	基本假设与符号说明	6
3	问题一的分析与求解：励磁波形分类	7
3.1	问题分析	7
3.2	数据探索性分析与预处理	7
3.2.1	探索性分析	7
3.2.2	数据预处理	8
3.3	特征提取	8
3.3.1	随机森林	8
3.3.2	提取结果	9
3.4	建立励磁波形分类模型	10
3.4.1	3.2.4 预测结果	13
4	问题二的分析与求解：斯坦麦茨方程（Steinmetz-equation）修正	15
4.1	问题分析	15
4.2	问题求解	15
4.2.1	普通最小二乘法（Ordinary Least Squares, OLS）	15
4.2.2	传统斯坦麦茨模型的拟合	16
4.2.3	修正的斯坦麦茨模型的建立	18
4.2.4	两种斯坦麦茨模型预测磁芯消耗的结果对比	19
5	问题三的分析与求解：磁芯损耗因素分析	20
5.1	问题分析	20
5.2	影响磁芯损耗因素的探索性分析	21
5.2.1	单一影响因素分析	21
5.2.2	协同影响因素分析	22
5.3	影响磁芯损耗因素的相关性检验	25
5.3.1	单因素相关性检验	26
5.3.2	因素间协同作用的分析	27
5.4	多元回归分析	29
5.4.1	数据预处理	29
5.4.2	构建模型	29
5.4.3	优化回归模型	30
6	问题四的分析与求解：基于数据驱动的磁芯损耗预测模型	32
6.1	问题分析	32
6.2	模型建立	33
6.2.1	数据预处理	33
6.2.2	特征提取	33
6.2.3	预测模型的建立	34
6.2.4	L2 正则化	36
6.2.5	预测结果	36
7	问题五的分析与求解：磁性元件的最优化条件	37
7.1	问题分析	37

7.2	双目标优化模型	37
7.3	模拟退火算法	38
7.4	优化模型求解	39
8	模型的评价与改进	39
8.1	模型优点	39
8.2	模型缺点	39
9	参考文献	40

1 问题重述

1.1 引言

随着现代电力电子技术的快速发展，磁性元件被广泛应用于各类电子设备，尤其作为电力电子变换器中的重要组成器件，是现代电能变换技术的基石，而电能变换技术在信息通讯、汽车制造、医疗设备、工业自动化等社会经济的各个领域，均起到了非常重要的作用。磁性元件是电力电子系统中用于存储、传递和调节磁能的关键组件，主要包括变压器和电感。变压器通过磁场传递电能，实现电压和电流的变换；电感则用于储存和释放磁能，滤波和稳定电流。它们在电力变换器中不仅决定了系统的体积、重量和损耗，还影响整体效率。

如今，电能变换技术的高速发展对于功率变换器的性能带来了更大的挑战，既需要满足高效率、高功率密度的要求，还需要使其体积、重量尽可能减小，损耗尽可能减少。而根据数据统计，磁性元件的体积、重量一般占到整个变换装置的二到三成，损耗更是占到了装置总损耗的三成以上，对于整机的效率和功率密度有着决定性的影响。因此，对磁性元件进行研究分析及优化，对提高功率变换器的性能，减少成本有着十分重要的意义。本文的着重点在于磁性元件损耗的建模分析，而磁性元件的损耗又分为绕组损耗和磁芯损耗。绕组损耗已经可以通过电磁场有限元仿真技术准确模拟，但对于磁性材料在高频交变磁通下产生的磁芯损耗（此处仅指功率损耗），其与实际工况（工作频率、磁通密度、励磁波形、工作温度等）以及磁芯材料本身的微观结构等诸多因素有关，成为了亟待解决的行业难题。因此，在不同工况条件下进行损耗实验，记录磁芯材料损耗的数据，建立磁芯材料损耗的数学模型，对于减少磁性元件损耗，提高电力电子设备的功效和性能具有至关重要的指导意义。

1.2 问题的提出

题目共给出了 4 个 excel 表格，其中附件一为训练集，包含 4 个工作表，分别代表 4 种不同的磁芯材料，每个工作表的结构均相同，包括温度、频率、励磁波形类型三种工况变量，磁芯损耗变量以及用于识别励磁波形类型的磁通密度变量（1024 个采样点采集）；附件二与附件三均为测试集，分别用来预测励磁波形类型和磁芯损耗；附件四为答案文件，最终预测的结果将填在其中。要求利用上述数据解决以下问题：

问题一：根据训练集中每一条在 1024 个采样点上采样的磁通密度数据（共 12400 条），分析磁通密度的分布特征及不同励磁波形的形状特征，提取出各种能反映这些特征的特征变量；利用这些特征变量建立励磁波形分类模型，识别出测试集中未标注的 80 条数据各自的波形，并分析所建立分类模型的合理性、有效性。

问题二：将斯坦麦茨方程与实验数据进行拟合，从而推导出斯坦麦茨系数 k_1 、 α_1 和 β_1 ；

由于传统的斯坦麦茨模型受不同温度影响，预测得到的磁芯损耗结果有所偏差，不能适应于实验温度环境的变化，因此将温度变量以不同形式引入到传统的 SE 方程，得到一种在各种温度条件下均适用的修正 SE 方程。利用附件一材料 1 的正弦波形为测试集，对比实际磁芯损耗结果与这两种模型的预测损耗结果，计算各自的误差。

问题三：已知温度、励磁波形以及磁芯材料是常见的对磁芯损耗有较大影响的三个要

素，利用探索性分析以及相关性检验等不同方法，剖析不同要素对于磁芯损耗的独立影响以及两两协同影响，将影响程度量化出来；然后利用各种优化算法，找出使得磁芯损耗最低的最优条件集合。

问题四：利用特征选取方法找出对磁芯损耗具有决定性影响的因素，以这些因素为自变量，利用各种机器学习算法建立数据驱动模型；比较不同模型的预测精度与泛化能力，找出效果最好的模型，对附件三中样本的磁芯损耗进行预测；分析所建立的模型对于业界有何种指导意义。

问题五：在磁性元件设计中，如何找到最佳的设计方案，该方案应在降低磁芯损耗的同时，最大化传输磁能。为此，将使用已有的磁芯损耗预测模型，并结合传输磁能的评估方法（即频率与磁通密度峰值的乘积），对附件一中提供的实验数据进行分析。得到实现最低的磁芯损耗和最高的传输磁能时，温度、频率、磁芯材料、励磁波形以及磁通密度峰值分别为何值。

2 基本假设与符号说明

表 1 符号说明

符号	意义
T	温度
W	励磁波形
M	磁芯材料
f	频率

3 问题一的分析与求解：励磁波形分类

3.1 问题分析

通过问题一的描述与附件一数据集的深入分析，我们从中提取了三种不同励磁波形的磁通密度数据，并绘制了相应的数据曲线。观察发现，正弦波形的分布特征与标准的正弦函数曲线高度相似，而三角波和梯形波的磁通密度曲线则变化陡峭，且存在多个明显的拐点。基于此，我们首先根据各采样点的磁通密度数据绘制分布曲线，细致分析不同波形曲线在增长、衰减及波动模式上的差异；随后，计算一系列能全面反映曲线分布特征的统计量，作为识别励磁波形的重要依据。为进一步精简特征并提高识别精度，我们采用特征提取方法对这些统计量进行重要性排序和提取，力求在保持高精度识别励磁波形类型的同时，尽量减少所需的统计特征数量。最终，我们运用机器学习算法构建励磁波形分类模型，以对测试数据进行有效的波形分类与预测。

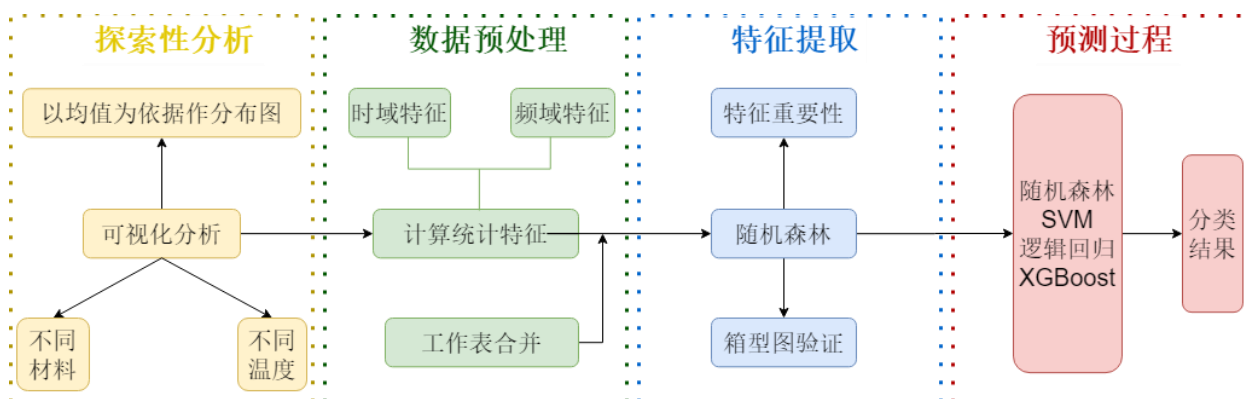


图 3-1 问题一的具体流程

3.2 数据探索性分析与预处理

3.2.1 探索性分析

首先考虑绘制数据表中各个磁通密度的分布曲线进行初步观察，由于数据量较大，不能一一绘制，因此我们依据磁通密度在每个采样点处的均值，分别根据不同温度以及不同材料绘制曲线图，如图所示

由图 3-2 可以看出，材料上的差异对三角波和正弦波而言，磁通密度数据均值并没有较大的差异；对于正弦波而言，材料一的磁通密度数据相比于其他材料存在更大的极差，材料三和材料四存在较高的相似性；

同时，由下图 3-3 我们可以直观地发现，温度上的差别并没有给波形的磁通密度数据带来较大的影响，因此我们在后续的影响特征选择上并没有考虑温度的作用。

不同材料下三种波形的对比（以均值为依据作图）

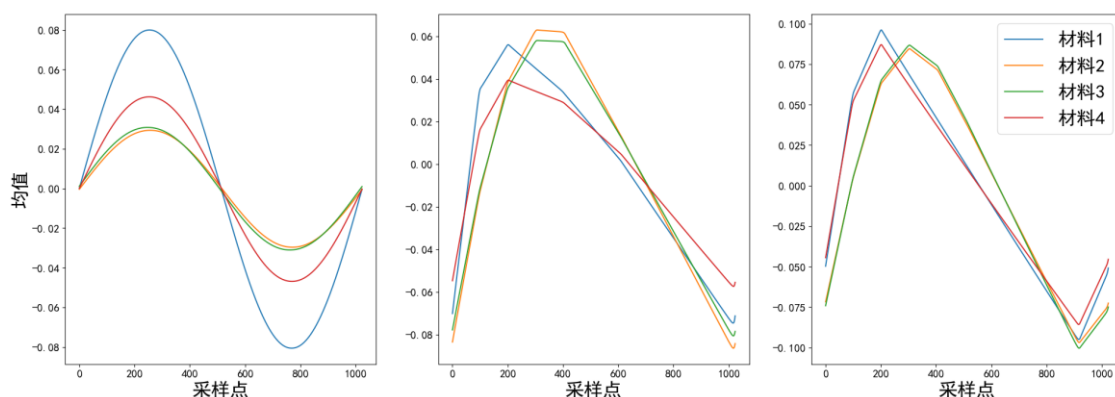


图 3-2 不同材料下三种波形的对比

不同温度下三种波形的对比（以均值为依据作图）

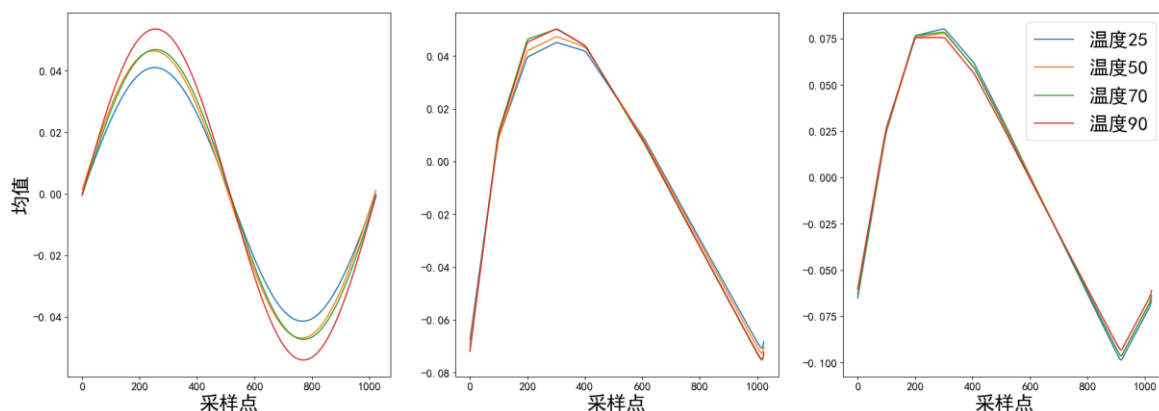


图 3-3 不同温度下三种波形的对比

3.2.2 数据预处理

3.3 特征提取

3.3.1 随机森林

随机森林是一种基于树模型的集成学习方法，它通过在训练过程中引入随机性，使得最终的结果具有较高的准确性和稳定性。具体的实现过程如下：

- **Step1: 数据采样:** 从原始数据集中进行有放回的随机采样，形成一个新的数据集。这个数据集的大小与原始数据集相同。
- **Step2: 决策树构建:** 对于每个采样得到的数据集，训练一个决策树。在训练决策树的时候，每次分裂，随机选择一个特征的子集，根据这部分特征选择最优的分裂点。
- **Step3: 重复步骤 a-b,** 直到达到停止条件，生成了多棵决策树，这些决策树共同构成了一个随机森林。
- **Step4: 随机森林通过投票的方式来决定最终的分类结果，**即每棵树对测试样本进

行分类，最后选择类别最多的作为最终分类。

3.3.2 提取结果

将数据集划分为测试集（80%）和测试集（20%），使用随机森林算法训练分类模型，得到如下特征重要性图：

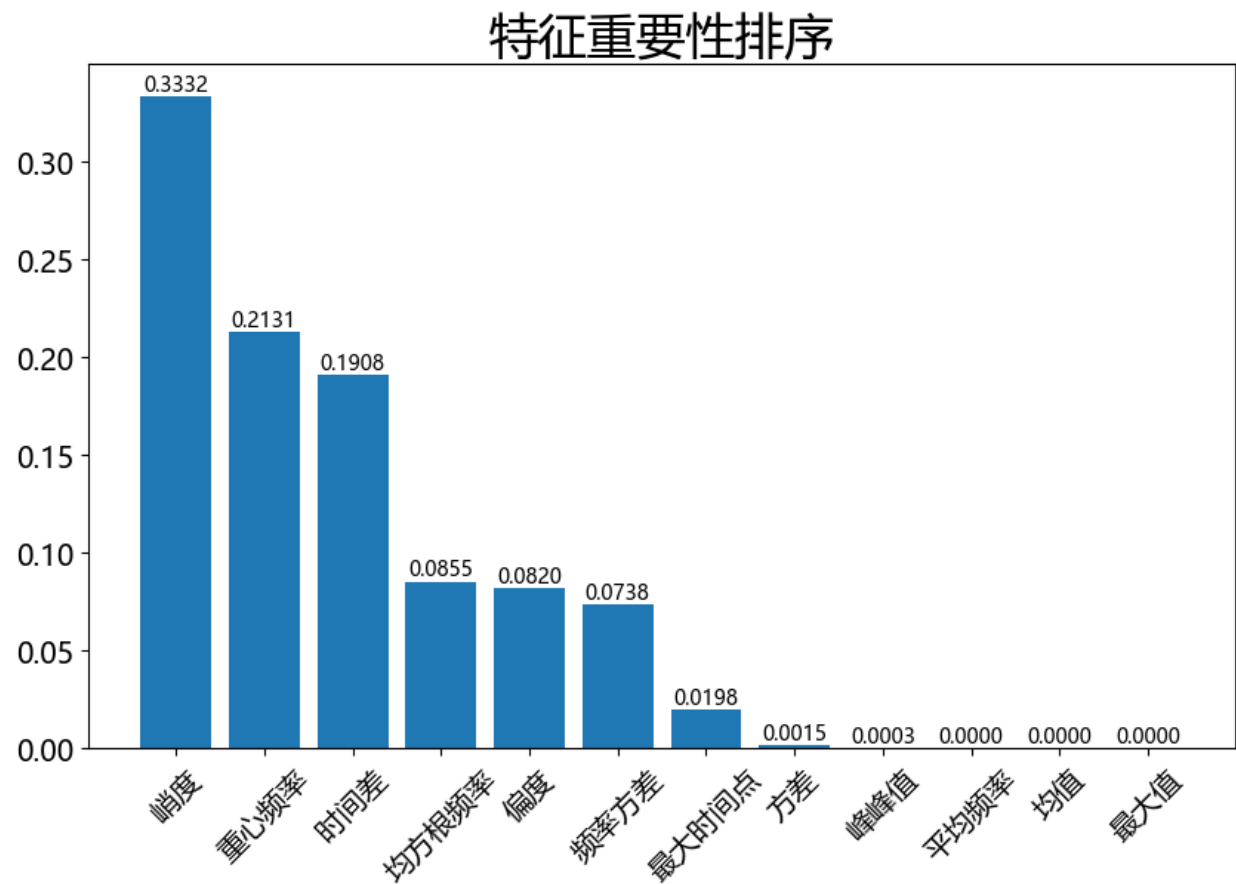
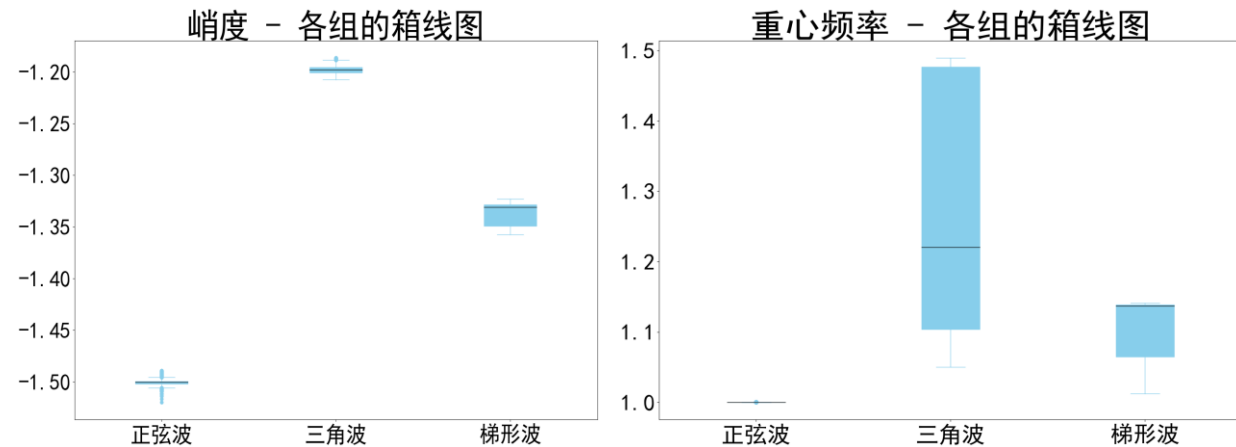


图 3-6 随机森林特征重要性排序

在 12 个特征中，我们选取了特征重要性排名前 7 的统计特征，包含 4 个时域统计特征和 3 个频域统计特征。从图 7 可知，峭度的重要性达到 0.3332，排名前 7 的统计特征总计贡献了 $33.32+21.31+19.08+8.55+8.2+7.38+1.98=99.82\%$ 的信息。另外，对于每个特征，我们绘制了其在 3 种波形上的箱线图分布。如图 3-7 所示。



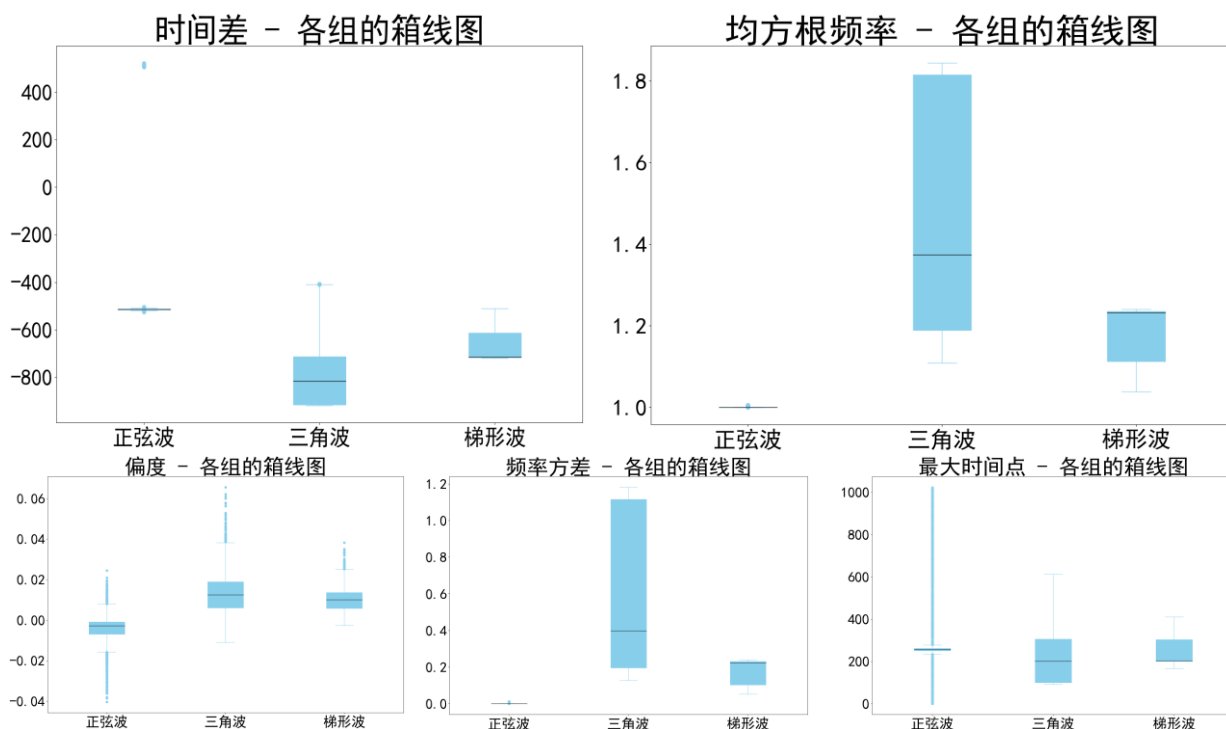


图 3-7 波形重要特征分布箱型图

从各个特征的箱型图可以看出，对于重要性较高的这 7 个特征，不同波形的数据之间重合度较低，可以认为从这些特征出发，能够对波形的分类起到充分的贡献作用。这也和我们之前的猜想相符。

对于重要性最高的峭度特征而言，不同波形的数据分离度非常高，三种波形的数据之间没有产生任何的重复，这说明三种波形数据的峭度值有明显差异，适合作为区分不同波形的依据。接下来我们以选取的 7 个特征构建励磁波形分类模型。

3.4 建立励磁波形分类模型

为了充分利用数据并评估模型的性能，本文考虑使用支持向量机、逻辑回归、随机森林以及 XGBoost 这四种机器学习方法，最后进行预测模型的效果对比。

A、支持向量机

支持向量机旨在通过最优化一个分隔界限，将不同类别的数据在数据集中明确区分开来。该算法通过扩大分类器的边际空间来提升对新数据的适应能力，进而确保分类器的推广性能。具体的实现过程如下：

- 选择核函数。如果数据是线性可分的，可以选择线性核函数；若不是线性可分的，可以选择多项式核函数、径向基函数或其他非线性核函数。
- 确定惩罚参数 C 。 C 值决定了误分类点的重要性。 C 值大，意味着分类器会选择一个较小的间隔，以减少训练样本上的错误分类。
- 构造优化问题并求解。SVM 的目标是找到一个超平面，使得不同类别的数据点之间的间隔最大，同时最小化分类错误。对于线性 SVM，优化问题可以表示为：

$$\begin{aligned} \min_{w,b} \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i (w^T x_i + b) \geq 1 - \xi_i \\ & \xi_i \geq 0 \end{aligned} \quad (3.13)$$

其中， w 是超平面的法向量， b 是截距， ξ_i 是松弛变量， C 是惩罚参数， y_i 是类别标签（取值 ± 1 ）， x_i 是特征向量。使用序列最小优化算法或者二次规划来求解上述优化问题。

d. 找到支持向量（那些位于间隔边界上的点）。

e. 计算偏置，并构建决策函数。一旦找到支持向量，可以通过求解以下方程来计算偏置 b ：

$$b = y_i - wx_i \quad (3.14)$$

其中， y_i 和 x_i 是支持向量的类别标签和特征向量。

决策函数可以表示为：

$$f(x) = \text{sign}(wx + b) \quad (3.15)$$

其中， $\text{sign}()$ 是符号函数，如果 $wx + b > 0$ ，则 $f(x) = 1$ ；否则， $f(x) = -1$ 。

B、逻辑回归

逻辑回归算法是一种广泛应用于分类问题的监督学习算法，它基于线性回归的基础上，通过引入逻辑函数将线性回归的输出映射到 0 和 1 之间，从而实现二分类或多分类。具体来说，逻辑回归首先构建一个线性模型，形式为：

$$z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n \quad (3.16)$$

其中 X_i 是输入特征， β_i 是对应的权重参数， z 是线性模型的输出。接下来，逻辑回归使用

逻辑函数 $S(z) = \frac{1}{(1 + e^{-z})}$ 来将线性模型的输出 z 映射到一个介于 0 和 1 之间的概率值。这

个概率值表示样本属于正类的可能性。对于二分类问题，当 $S(z)$ 大于某个阈值（通常取 0.5）时，模型预测样本为正类；否则，预测为负类。决策边界是由权重参数和特征值决定的，它将特征空间划分为两个区域，一个对应正类，一个对应负类。逻辑回归的输出 $S(z)$ 可以解释为样本属于正类的概率，即 $P(y=1|X)$ 。在训练过程中，逻辑回归通常使用最大似然估计来找到一组权重参数，使得观察到的样本标签的概率最大。这等价于最小化负对数似然损失函数，即

$$L(\beta) = -\sum_{i=1}^n [y_i \log(S(z_i)) + (1 - y_i) \log(1 - S(z_i))] \quad (3.17)$$

其中 y_i 是第 i 个样本的实际标签。逻辑回归模型的训练通常涉及对损失函数进行优化，找到使损失函数最小的权重参数。

C、XGBoost

XGBoost（极端梯度提升）是一种可扩展的分布式梯度提升决策树（GBDT）机器学习库，广泛应用于回归和分类问题。其核心思想是通过组合多个弱分类器（通常是决策树）来构建一个强大的集成模型。XGBoost 的独特之处在于采用了梯度提升算法，并结合了一系列创新技术，以提升模型的准确性和效率。

具体而言，XGBoost 首先用一棵弱分类器（单棵决策树）初始化模型，然后计算每个样本的残差（实际值与预测值之间的差异）。随后，模型会训练一棵新的决策树，以拟合这些残差，从而修正之前模型的错误。这个过程会反复进行，每轮都聚焦于先前模型的不足，使得每个新模型的预测更加精确。对于包含 n 条 m 维的数据集，XGBoost 模型可表示为：

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), \quad f_k \in F(i=1, 2, \dots, n) \quad (3.18)$$

其中， $F = \{f(x) = w_{q(x)}\} (q: R^m \rightarrow \{1, 2, \dots, T\}, w \in R^T)$ 是 CART 决策树的结构集合， q 为样本映射到叶节点的树结构， T 为叶节点个数， w 为叶子节点的实数分数。

一般而言，损失函数描述的是预测值 \hat{y} 与真实值 y 之间的关系。这里使用平方损失函数，对于 n 个样本，可以写成：

$$L = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) \quad (3.19)$$

更进一步，目标函数可以写成：

$$Obj^{(t)} = \sum_{i=1}^n l(y_i, \hat{y}_i^{(t)}) + \Omega(f_t) \quad (3.20)$$

其中 Ω 代表模型的复杂度，若基模型是树模型，则树的深度、叶节点数等指标均可以反映树的复杂度。

D、参数设置

需要注意的是，SVM 与逻辑回归模型一般适用于二分类问题，而励磁波形共有三种类型，这就涉及到将二分类任务扩展至多分类任务的问题。

在使用 SVM 算法对模型进行训练之前，使用的 Sklearn 库中的 SVC() 函数，需要设置参数 `decision_function_shape='ovr'`。这一参数命令 SVM 模型使用“一对多”（One-vs-Rest, OVR）分类策略，即对于每一个类别均创建一个二分类器，将本类别视为正类样本，其他类别均视为负类样本进行训练。在进行预测时，所有的二分类器同时预测，每个分类器会输出一个分数或概率，表示输入样本属于该类别的可能性，最终选择得分最高的分类器对应的类别作为模型的预测结果。

同样地，在使用逻辑回归算法对模型进行训练之前，使用 LogisticRegression() 函数，需要设置参数 `multi_class='multinomial'`。这一参数命令逻辑回归模型使用 Softmax 函数进行多项式回归。Softmax 函数将一个向量中的每个元素转化为一个概率分布，确保所有输出的和为 1。对于一个包含 K 个类别的向量 $\mathbf{z} = [z_1, z_2, \dots, z_k]$ ，Softmax 函数定义为：

$$P(y = k | \mathbf{x}) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}} \quad (3.21)$$

其中， $P(y = k | \mathbf{x})$ 表示输入 \mathbf{x} 属于类别 k 的概率。

3.4.1 3.2.4 预测结果

采用随机森林、支持向量机、逻辑回归和 XGBoost 建立分类模型。在训练过程中将数据集划分成训练集和测试集，训练集占比 80%，测试集占比 20%。预测得到的结果通过混淆矩阵可以直观地反映出预测效果，同时打印出分类报告，其中包括了精确率(precision)、召回率(recall)、准确率(accuracy)、f1 得分(f1-score)等一系列评价分类模型预测效果的指标，取值范围均为[0,1]，且越接近 1 效果越好。

除了上述常见的预测结果评价指标之外，又计算了科恩-卡帕系数(Cohen Kappa Score)，用于衡量分类模型的准确性，同时考虑随机猜测的影响。它的值范围从-1 到 1，其中 1 表示完全一致，0 表示与随机猜测相同，负值表示一致性低于随机猜测。公式为：

$$k = \frac{P_o - P_e}{1 - P_e} \quad (3.22)$$

其中， P_o 是观察到的准确率(accuracy)； P_e 是随机情况下预期的准确率。

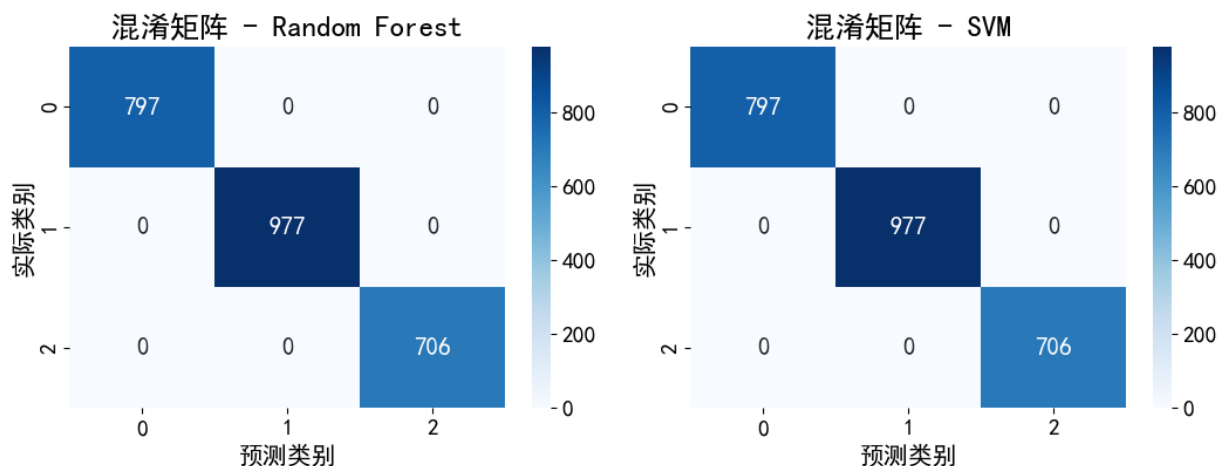
由于本题目给出的数据量大而规范，三种波形的区别也显而易见，仅凭观测就可以准确识别出不同波形。因此，对于问题一的预测结果，除了使用 XGBoost 构造的预测模型将本属于三角波的励磁波形错误地识别成了梯形波，导致各评价指标有些微的降低之外，其余的分类模型进行预测的结果，所有选取的评价指标均为 1。

以随机森林模型为例，计算各个评价指标如下：

表 3-1 随机森林模型的评价指标

特征	precision	recall	accuracy	f1-score	Cohen's Kappa
1.00	1.00	1.00	1.00	1.00	1.00

四个模型预测结果的混淆矩阵如下图所示。



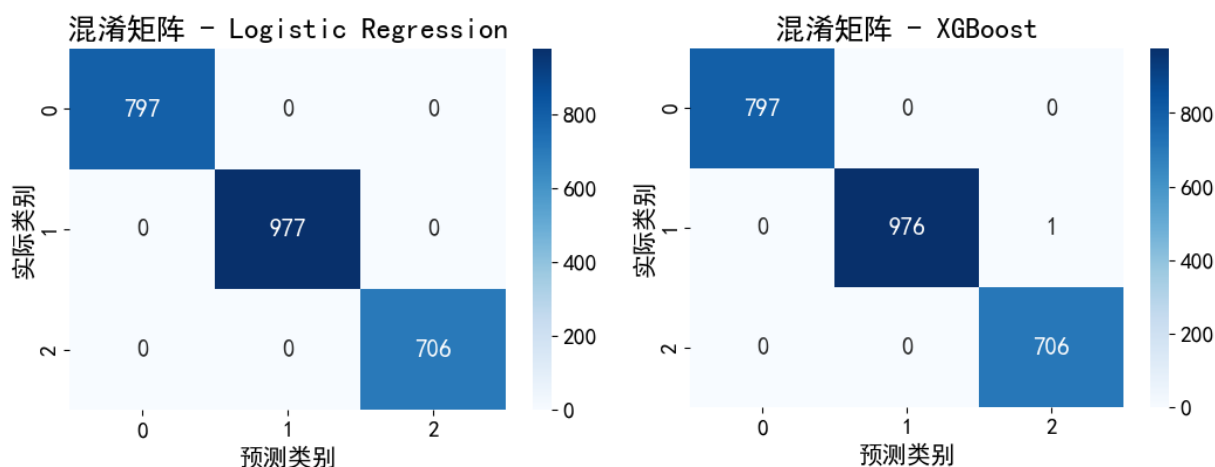


图 3-8 混淆矩阵

以附件二中三种波形各自的数量绘制的玫瑰图如 3-9 所示；题目给定序号样本的分类结果如表 3-2 所示。

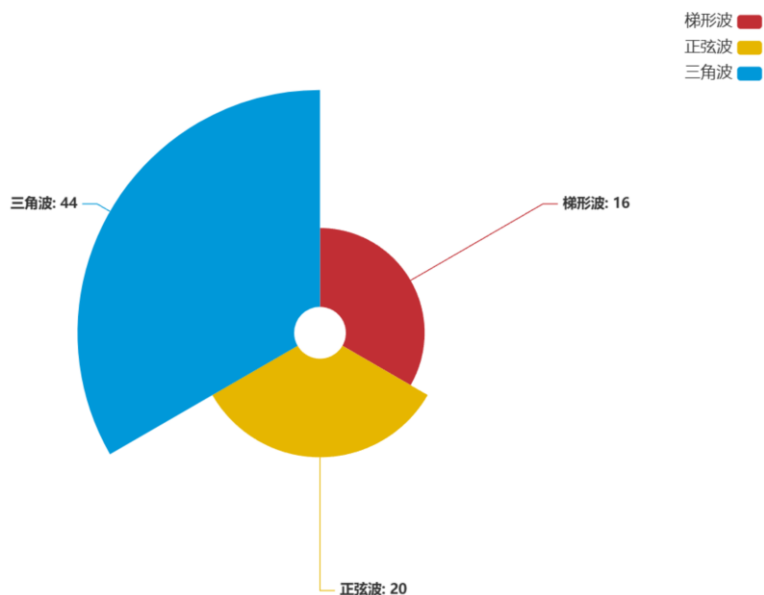


图 3-9 附件二中三种波形的各自数量

表 3-2 指定序号的样本分类结果

样本序号	分类结果	样本序号	分类结果
1	三角波	45	梯形波
5	三角波	55	三角波
15	正弦波	65	三角波
25	三角波	75	三角波
35	梯形波	80	正弦波

4 问题二的分析与求解：斯坦麦茨方程（Steinmetz-equation）修正

4.1 问题分析

斯坦麦茨方程（Steinmetz-equation）是最著名的经验计算模型之一，在正弦波励磁下，磁芯损耗 P 可通过该方程计算得到：

$$P = k_1 \cdot f^{\alpha_1} \cdot B_m^{\beta_1} \quad (4.1)$$

其中，频率 f 已在附件一训练数据中列出，磁通密度峰值 B_m 可以通过简单计算求得，而

剩余的三个方程系数 k_1 、 α_1 和 β_1 （称为斯坦麦茨系数）则需要根据实验数据进行拟合，

因此我们需要根据训练数据中使用交流功率法实际测量出的磁芯损耗值来拟合方程，求出斯坦麦茨系数。

如果在不同温度下使用相同系数的斯坦麦茨方程，会导致预测磁芯损耗的精度较低，故问题二旨在探索一种高精度且适用于不同温度条件下的磁芯损耗修正方程，并将其与传统的斯坦麦茨方程的拟合效果进行对比。本文首先利用附件一材料 1 中正弦波数据（1067 条），采用最小二乘法拟合斯坦麦茨方程，并给出使用方程预测的 1067 条数据的磁芯损耗结果，计算 MAE、MSE、MRE、 R^2 四种预测模型的性能评价指标；接着通过查阅相关论文，发现将温度因素引入磁芯损耗的形式可能为幂函数、多项式函数或者指数函数的形式，因此分别采用以上三种形式将温度变量引入 SE 方程，再次利用数据进行拟合，选择各评价指标最好的方程作为修正后的方程；最后，对比方程修正前后的四个评价指标，通过绘制方程修正前后各样本点下实际损耗与预测损耗的曲线图、误差曲线图以及实际损耗—预测损耗的对比图，可视化方程修正前后的预测效果。

4.2 问题求解

4.2.1 普通最小二乘法（Ordinary Least Squares, OLS）

普通最小二乘法是最常见的拟合方法，简单易用，且适用范围广泛。利用 OLS 拟合多项式的基本步骤为：

Step1: 确定多项式的阶数。多项式的一般形式为：

$$P(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0 \quad (4.2)$$

其中， n 是多项式的阶数， $a_n, a_{n-1}, \dots, a_1, a_0$ 是多项式的系数。

Step2: 收集一系列的 (x, y) 数据点，将这些数据点用于拟合多项式模型。

Step3: 将拟合问题转化为一个最小二乘问题，即找到一组 $a_n, a_{n-1}, \dots, a_1, a_0$ 使得拟

合函数 $P(x)$ 与实际数据 y_i 之间的误差平方和最小：

$$\min_{a_n, a_{n-1}, \dots, a_1, a_0} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.3)$$

通过求解上式这一最小二乘问题，即可得到 $a_n, a_{n-1}, \dots, a_1, a_0$ ，从而拟合函数 $P(x)$ 。

4.2.2 传统斯坦麦茨模型的拟合

利用附件一材料 1 中 1067 条正弦波数据，采用最小二乘法拟合斯坦麦茨方程，得到方程如下：

$$P = 1.50 \times f^{1.43} \times B_m^{2.47} \quad (4.4)$$

即斯坦麦茨系数 $k_1 = 1.50$ ， $\alpha_1 = 1.43$ ， $\beta_1 = 2.47$ 。

采用以下指标评价拟合模型效果：

(1) 均方误差 (MSE)：表示预测值与实际值之间差异平方的平均值，MSE 越小，模型性能越好，计算公式如下：

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.5)$$

(2) 平均绝对误差 (MAE)：表示预测值与实际值之间绝对差异的平均值，MAE 越小，模型性能越好，计算公式如下：

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.6)$$

(3) 平均相对误差 (MRE)：表示预测值与实际值之间相对差异的平均值，MRE 越小，模型性能越好，计算公式如下：

$$\text{MRE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (4.7)$$

(4) 拟合优度 (R^2)：反映因变量的全部变异能通过回归关系被自变量所解释的比例，即模型对数据的解释程度， R^2 越接近 1，模型性能越好，计算公式如下：

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.8)$$

四个公式中， n 代表样本个数， y_i 代表第 i 个观测值， \hat{y}_i 代表第 i 个拟合值， \bar{y} 代表均值。

利用以上四个指标对传统斯坦麦茨拟合模型进行评估，结果如下表所示：

表 4-1 SE 方程拟合模型效果

MSE	MAE	MRE	R^2
1638190472	20616	0.3566	0.9448

以各采样点为横轴，实际损耗为纵轴，绘制在各点处实际的磁芯损耗以及 SE 方程预测的磁芯损耗分布图，如图（）所示，实际磁芯损耗与 SE 方程预测磁芯损耗在个别点处误差较大；以预测损耗为横轴，实际损耗为纵轴，绘制散点图，散点落在 $y = x$ 这条直线附近，说明实际损耗与预测损耗基本一致，则该模型拟合效果较好，如图 4-1 所示，当实际磁芯损耗较小时，SE 方程的预测点基本落在 $y = x$ 这条直线周围，预测结果与实际结果基本一致；随着实际损耗的增大，可发现 SE 方程的预测点逐渐偏离该直线，预测误差不断增大。

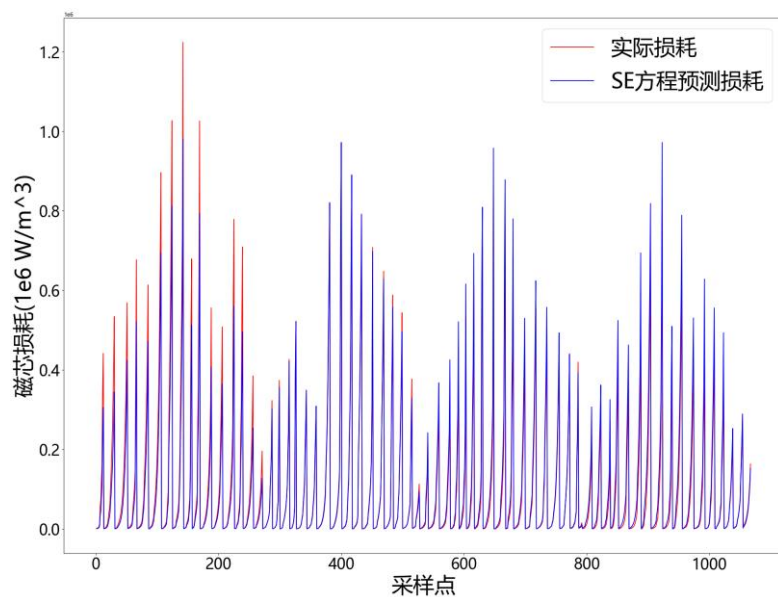


图 4-1 各采样点处的实际损耗与 SE 方程预测损耗分布图

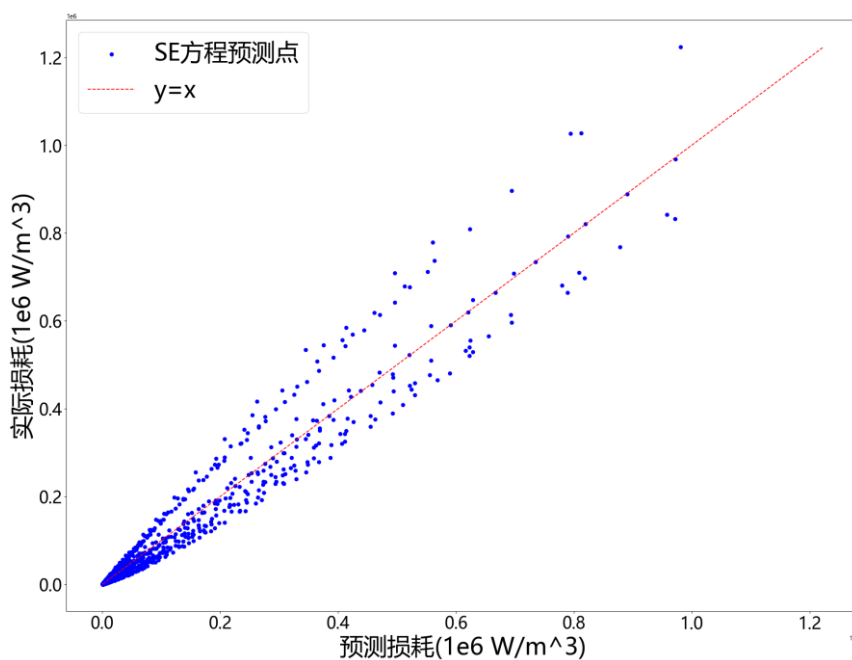


图 4-2 实际损耗—预测损耗散点图

4.2.3 修正的斯坦麦茨模型的建立

假设修正的斯坦麦茨方程的形式为：

$$P = k_2 \cdot f^{\alpha_2} \cdot B_m^{\beta_2} \cdot H(T) \quad (4.9)$$

其中 $H(T)$ 为温度变量。

通过查阅相关文献，得知 $H(T)$ 的引入形式一般有三种：幂函数形式、多项式函数形式以及指数函数形式。因此我们分别根据这三种引入形式建立三种修正的斯坦麦茨模型。

(1) 指数函数形式：

$$P = k_2 \cdot f^{\alpha_2} \cdot B_m^{\beta_2} \cdot e^{\gamma_0 T} \quad (4.10)$$

其中 γ_0 为温度系数。通过普通最小二乘法拟合得到的具体方程为：

$$P = 1.85 \times f^{1.45} \times B_m^{2.44} \times e^{-0.008 \times T} \quad (4.11)$$

(2) 多项式函数形式（以二次多项式为例）：

$$P = k_2 \cdot f^{\alpha_2} \cdot B_m^{\beta_2} \cdot (\gamma_0 + \gamma_1 \cdot T + \gamma_2 \cdot T^2) \quad (4.12)$$

其中 γ_0 、 γ_1 和 γ_2 为温度系数。通过普通最小二乘法拟合得到的具体方程为：

$$P = -0.004 \times f^{1.47} \times B_m^{2.45} \times (-388.68 + 4.82 \times T + (-0.027) \times T^2) \quad (4.13)$$

(3) 幂函数形式：

$$P = k_2 \cdot f^{\alpha_2} \cdot B_m^{\beta_2} \cdot T^{\gamma_0} \quad (4.14)$$

其中 γ_0 为温度系数。通过普通最小二乘法拟合得到的具体方程为：

$$P = 4.74 \times f^{1.46} \times B_m^{2.45} \times T^{-0.40} \quad (4.15)$$

计算以三种不同形式引入温度变量的拟合模型的评价指标，并于传统的 SE 方程拟合模型对比，结果如下表所示：

表格 4-2 以三种不同形式引入温度变量的 SE 方程与传统 SE 方程评价指标对比

	MSE	MAE	MRE	R^2
SE 方程	1638190472	20016	0.3566	0.9448
指数函数修正	211145205	8435	0.2428	0.9929
多项式函数修正	134806625	6775	0.2219	0.9955
幂函数修正	138596508	6850	0.2238	0.9953

通过对比可以发现，无论以哪种形式将温度引入 SE 方程，各评价指标方面都比原来有了较大的提升。其中，以幂函数、多项式函数形式将温度变量引入的 SE 方程指标要明

显优于以指数函数形式将温度变量引入的 SE 方程；另外，以多项式形式将温度变量引入的 SE 方程拟合参数要多于以幂函数形式将温度变量引入的 SE 方程，有一定的过拟合风险，因此本文选择以幂函数形式将温度变量引入的 SE 方程作为修正后的 SE 方程，即：

$$P = 4.74 \times f^{1.46} \times B_m^{2.45} \times T^{-0.40} \quad (4.16)$$

绘制在各采样点下实际磁芯损耗以及修正后的 SE 方程预测磁芯损耗分布图，如下图所示，可发现与图 4-1 相比，实际损耗与预测损耗间的误差大大变小。

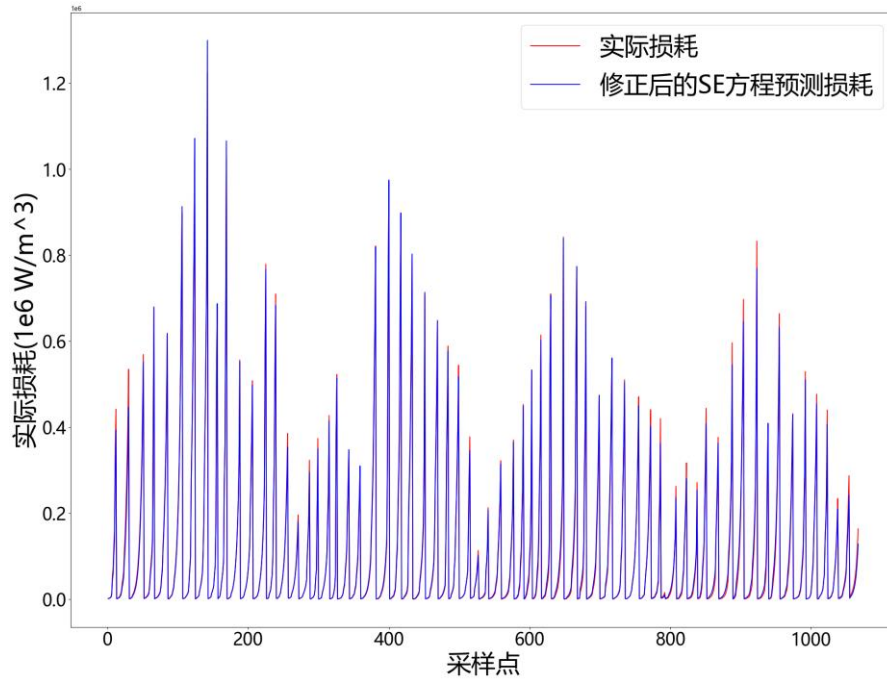


图 4-3 各采样点处的实际损耗与修正的 SE 方程预测损耗分布图

4.2.4 两种斯坦麦茨模型预测磁芯消耗的结果对比

以预测损耗为横轴，实际损耗为纵轴，在一张图上同时两种方程的实际损耗—预测损耗散点图，用颜色进行区分，如图 4-4 所示。从图中可以发现，无论在实际损耗值较小处还是较大处，修正后的 SE 方程散点明显更接近于 $y = x$ 这条直线，没有像传统的 SE 方程那样出现偏离的情况。分别计算修正前后 SE 方程在各采样点处的误差，以采样点为横轴，误差为纵轴，绘制曲线图，如图 4-5 所示，可发现修正后的 SE 方程误差远远小于修正前。

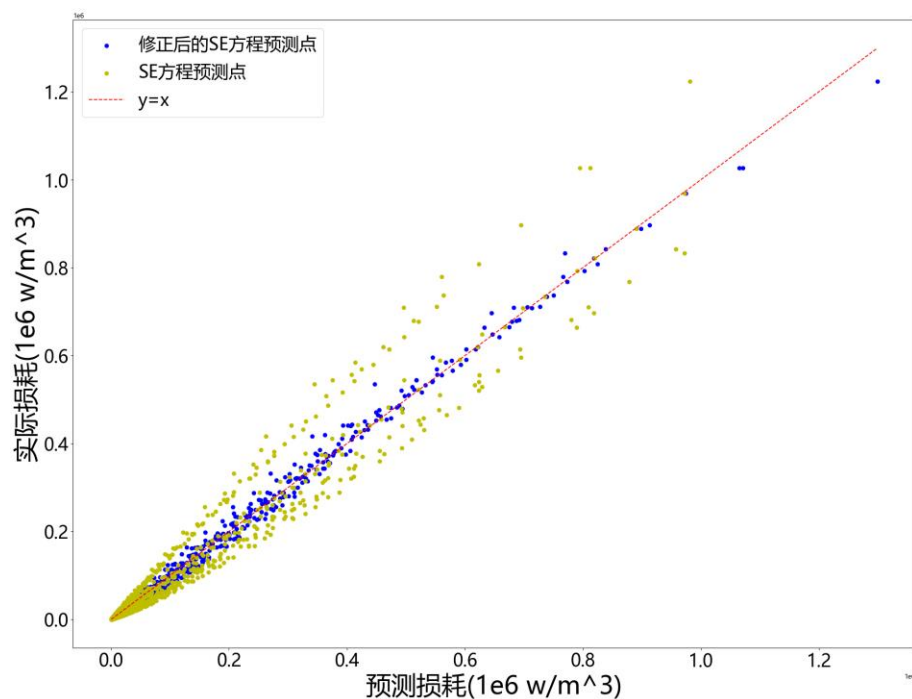


图 4-4 修正前后 SE 方程实际损耗—预测损耗散点图

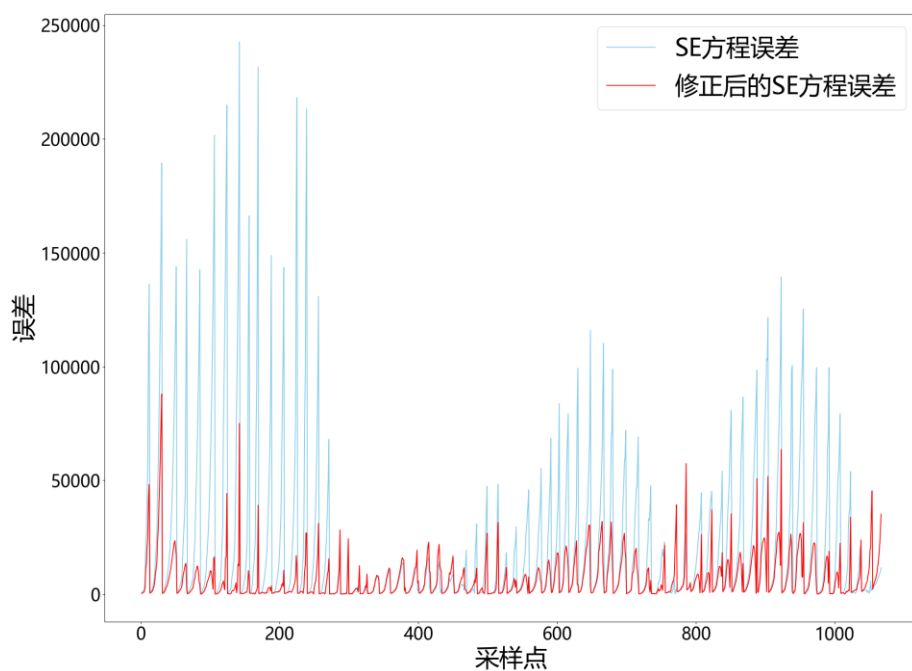


图 4-5 修正前后各采样点处误差对比

5 问题三的分析与求解：磁芯损耗因素分析

5.1 问题分析

问题三旨在分析温度、励磁波形和磁芯材料这三个因素如何独立及协同影响磁芯损耗，

并给出当三个因素为何值时，磁芯损耗达到最小。本题的解题思路为：探索性数据分析、因素间独立影响分析、因素间协同影响分析、建立多元回归模型、通过遗传算法求得最小值。首先根据数据分别绘制不同温度、不同材料、不同励磁波形下磁芯损耗的箱线图以及磁芯损耗均值的直方图，进行初步分析；接着计算 Spearman 相关系数，检验温度和磁芯损耗的相关性，利用 Kruskal-Wallis 检验来探究材料、励磁波形各自与磁芯损耗的相关性；再次，利用双因素方差分析检测温度、材料、励磁波形两两之间对磁芯损耗的协同影响；最后，建立磁芯损耗关于温度、材料、励磁波形以及其两两之间交互项的多元回归模型，通过遗传算法找到磁芯损耗最优值以及此时的温度、材料、励磁波形。

问题三的流程如图如下：

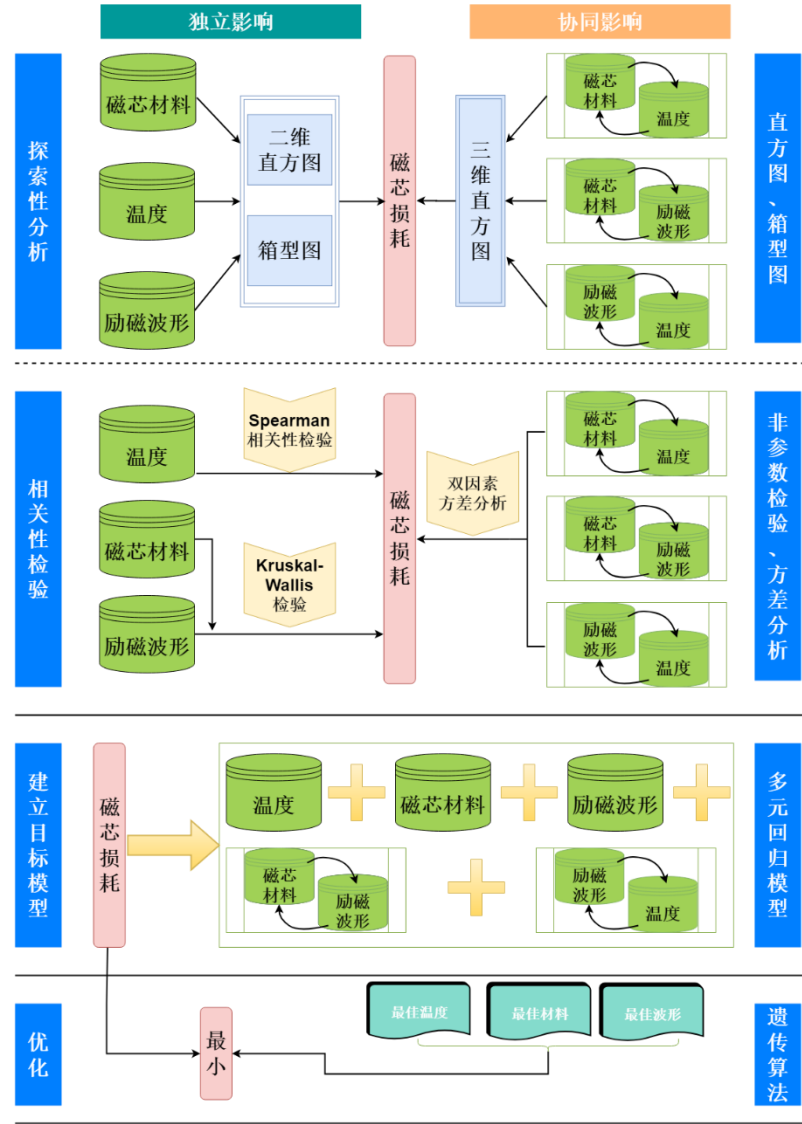


图 5-1 问题三流程图

5.2 影响磁芯损耗因素的探索性分析

5.2.1 单一影响因素分析

磁芯损耗为连续型变量，温度、励磁波形、磁芯材料均为分类型变量。为了初步探究

单一因素对磁芯损耗的影响，首先分别绘制不同温度、不同材料、不同励磁波形下磁芯损耗均值直方图和箱线图，如图 5-2 所示。

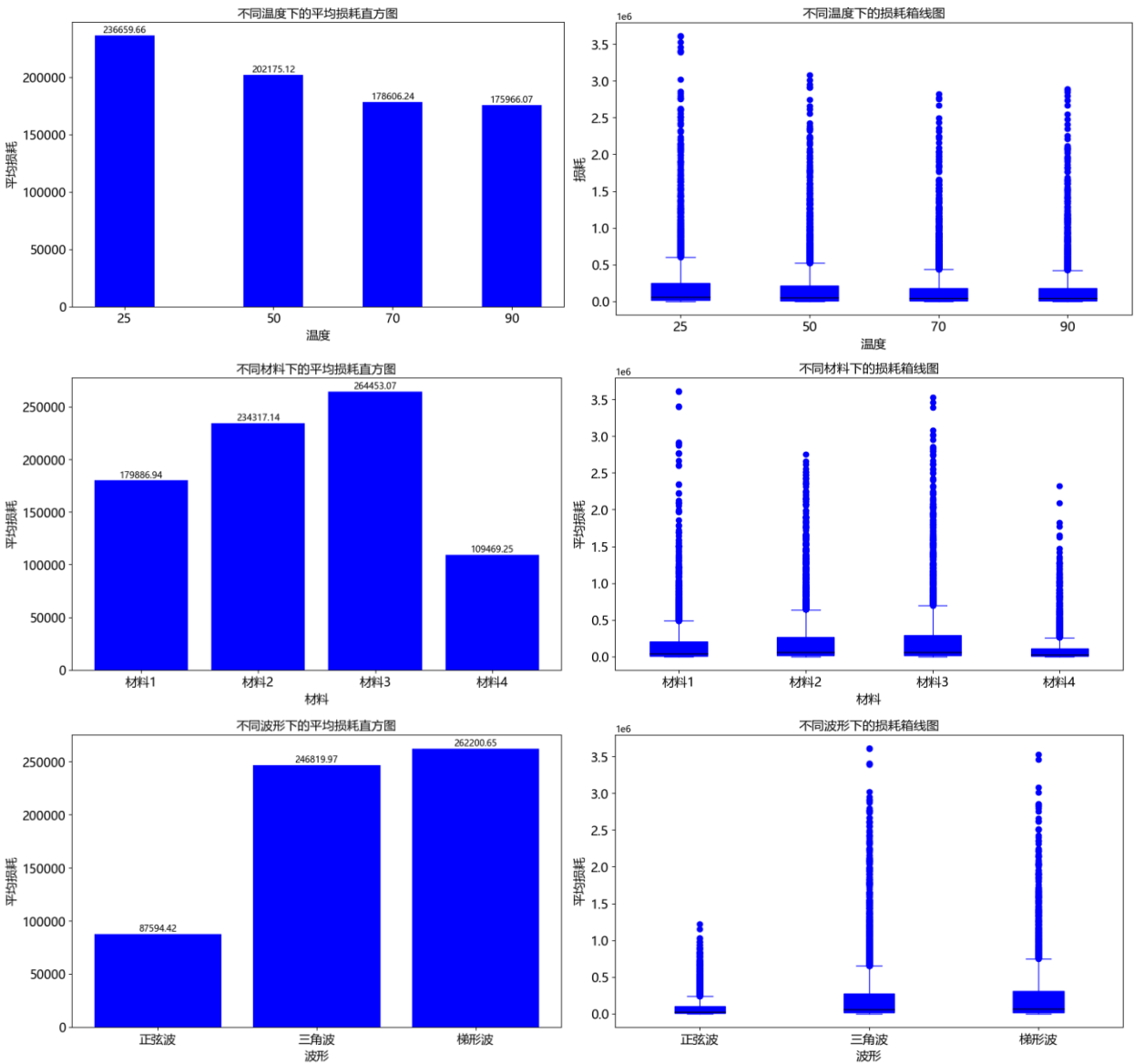


图 5-2 不同条件下的平均损耗直方图和箱线图

从温度单一因素的作用来看，温度越高，磁芯损耗越低，但随着温度的升高，磁芯损耗降低的速度变缓，70 度和 90 度的磁芯损耗非常接近；从材料单一因素的作用来看，不同磁芯材料对于磁芯损耗的影响是显著的。对于本题目所提供数据中的 4 种磁芯材料而言，材料 4 显然是性质最好，稳定性最强的材料，在同样的实验周期中，该材料的损耗要明显低于另外三种，然后是材料 1、2，最不耐用的是材料 3；从波形单一因素来看，对磁性元件施加正弦励磁波进行实验，磁芯损耗要明显低于三角波以及梯形波，而三角波和梯形波的磁芯损耗则比较接近。

5.2.2 协同影响因素分析

为了初步探究分类变量间对磁芯损耗的协同作用，需要一次性使用两个特征与磁芯损耗平均值绘制三维立体直方图。

(1) 温度—励磁波形—磁芯损耗

以 4 种不同温度和 3 种不同励磁波形，共 12 组变量为横纵坐标，以各组的磁芯消耗平均值为竖坐标绘制立体直方图如下：

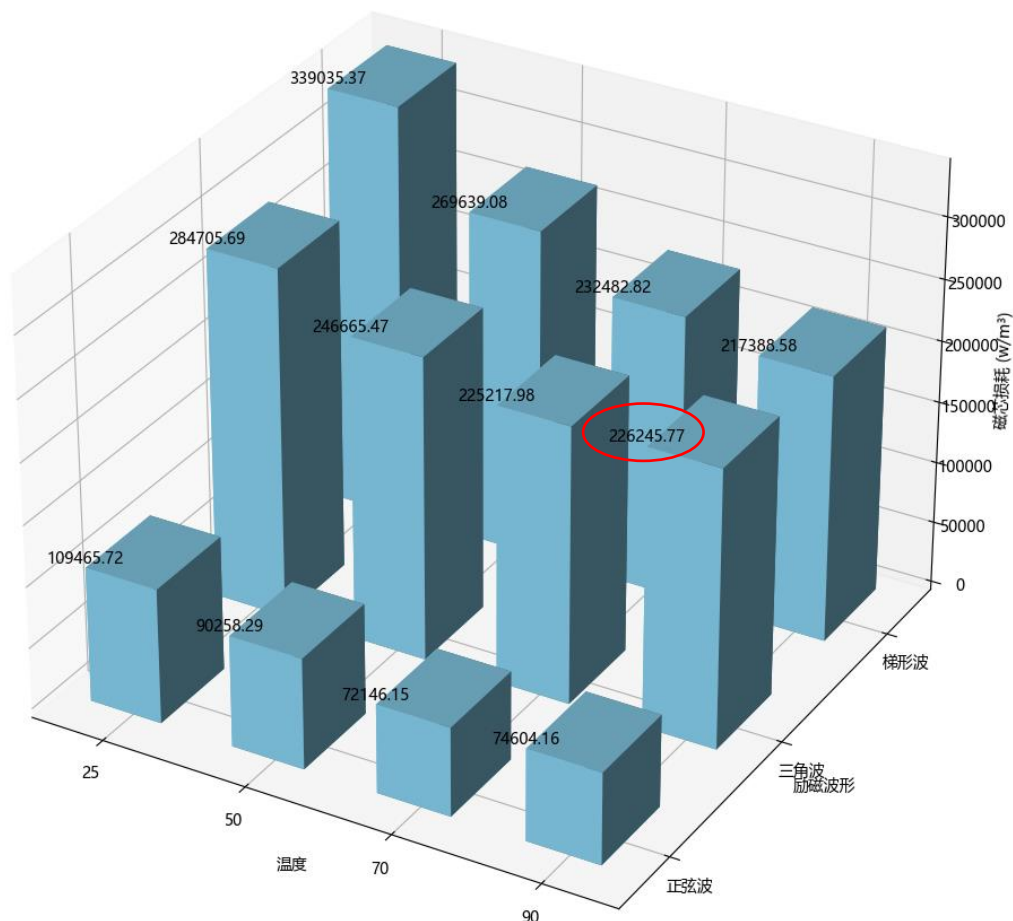


图 5-3 温度—励磁波形—磁芯损耗立体直方图

从上图可以看出，除了三角波形之外，施加其他两种励磁波形，在温度不断增高的情况下磁芯损耗均不断下降，在 25 度时损耗最高，在 90 度时损耗最低，这与我们在单一影响因素分析时得出的结论一致；除了温度为 90 度之外，其他三种温度情况下，均是施加正弦波的磁芯损耗最小且远小于另外两种励磁波形，施加梯形波的磁芯损耗最大且略大于施加三角波的磁芯损耗，这也符合我们在单一因素分析得出的结论。而在温度为 90 度时，施加三角波得到的磁芯损耗超过了梯形波；当励磁波形为三角波时，90 度下的磁芯损耗超过了 70 度，一次性违背了两条我们之前得出的结论。因此可以初步推断，温度和励磁波形对于磁芯损耗在一定条件下是存在显著的协同特征的。

(2) 温度—磁芯材料—磁芯损耗

以 4 种不同温度和 4 种不同磁芯材料，共 16 组变量为横纵坐标，以各组的磁芯消耗平均值为竖坐标绘制立体直方图如下：

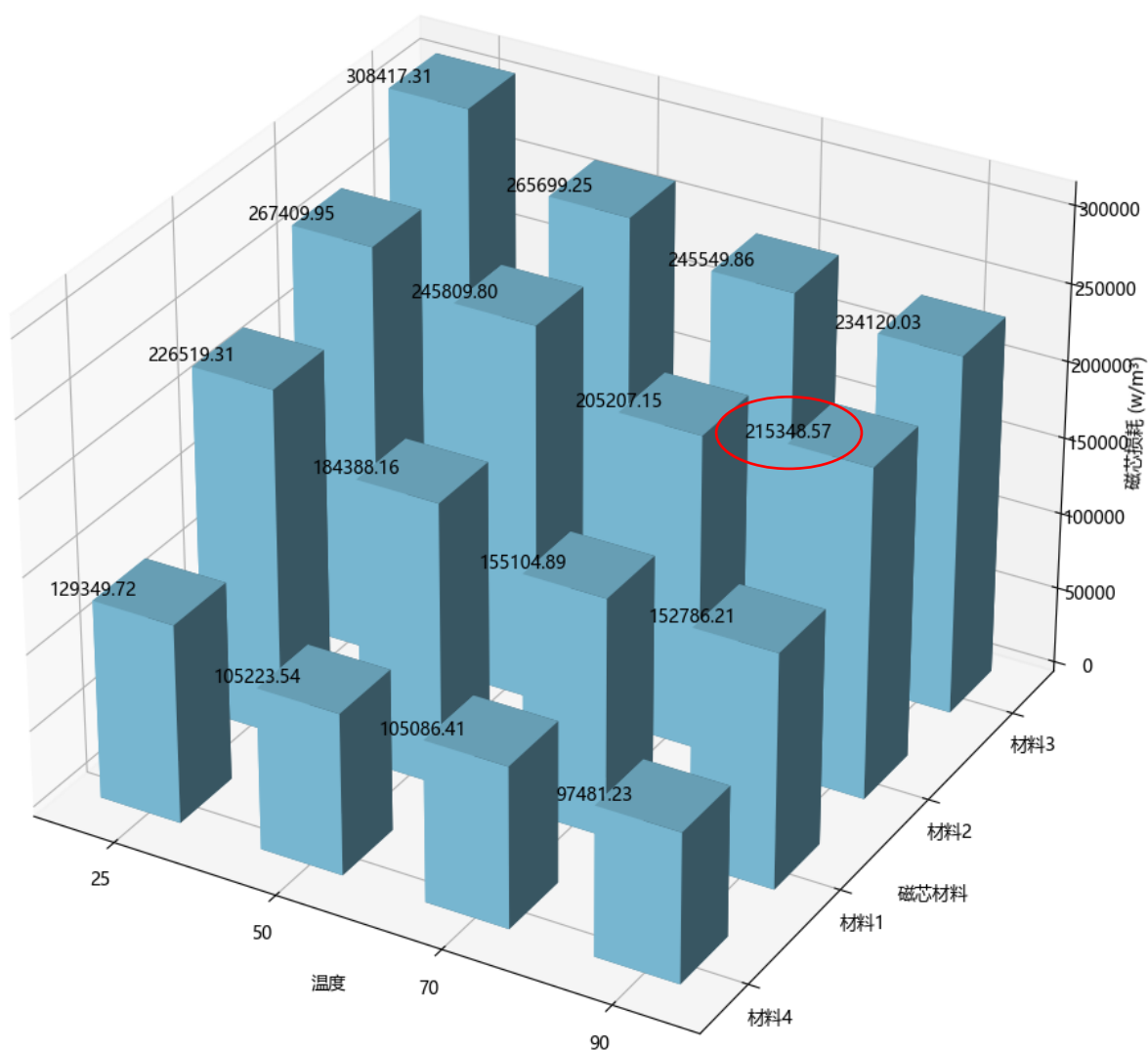


图 5-4 温度—磁芯材料—磁芯损耗立体直方图

从上图可以看出,不管在何种温度下,使用材料 4 的磁芯损耗均是所有磁芯材料中最小的,其次是材料 1、材料 2,损耗最高的是材料 3,这符合单一影响因素分析的结论:除了使用材料 2 之外,磁芯损耗随着温度的升高总是在逐渐下降的,仅在使用材料 2 作为磁芯材料时,90 度下的磁芯损耗略微超过了 70 度。因此可以初步推断,温度和磁芯材料之前可能存在协同作用,但是并不知道这种协同作用是否显著。

(3) 励磁波形—磁芯材料—磁芯损耗

以 3 种不同励磁波形和 4 种不同磁芯材料,共 12 组变量为横纵坐标,以各组的磁芯消耗平均值为竖坐标绘制立体直方图如下:

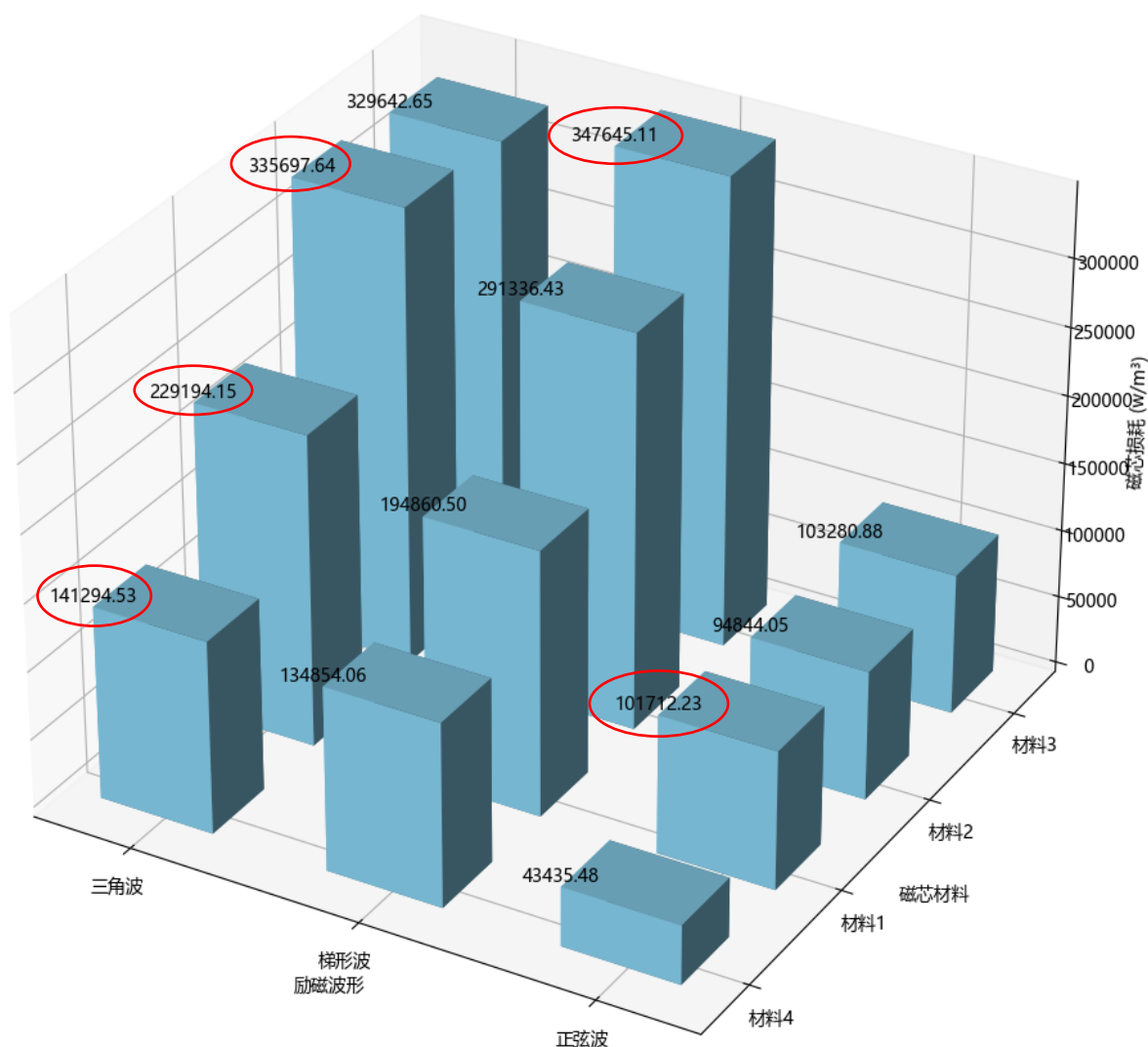


图 5-5 励磁波形—磁芯材料—磁芯损耗立体直方图

在此种特征组合对磁芯损耗的直方图中，出现了明显且众多的不符合单一因素分析结论的情况：除了材料 3 之外，使用其他三种材料的磁芯，在施加三角波时的磁芯损耗全部超过了梯形波；使用材料 1 和材料 2 的磁芯，在不同励磁波形下均出现了超过材料 3 磁芯损耗的情况。可以初步推断，励磁波形和磁芯材料对于磁芯损耗的影响存在明显的协同作用，且这种两两协同的作用在所有的三种组合内是最显著的。

5.3 影响磁芯损耗因素的相关性检验

在上一节对影响磁芯损耗的因素进行初步分析之后，本节将使用各种相关性检验将影响程度进行量化，深入了解不同变量对磁芯损耗的独立及协同影响的作用机理。

5.3.1 单因素相关性检验

具体的检验方法选择与独立变量、因变量的数据类型以及分布有关。探索多分类变量对于连续变量的影响程度时，如果因变量服从正态分布，常见的方法有 Pearson 相关性分析、方差分析、 η^2 效应量指标等；如果因变量不服从正态分布，常见的方法有 Spearman 相关性分析、Kruskal-Wallis 检验、Wilcoxon 符号秩检验等。

A、温度对磁芯损耗的相关性检验

由于磁芯损耗不满足正态分布，且温度为多分类有序变量，所以本文通过计算 Spearman 相关系数分析温度对磁芯损耗的影响。Spearman 相关系数，是一种非参数的秩相关系数，用于衡量两个变量的依赖性。计算步骤如下：

Step1.将两个变量的观测值分别按照大小顺序进行排序，并赋予相应的秩。

Step2.计算每个变量的原始观测值与其秩之间的差值，并求平方。

Step3.分别对两个变量的差值平方求和。

Step4.使用以下公式计算 Spearman 相关系数：

$$\rho = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)} \quad (5.1)$$

其中， d_i 是两个变量秩的差， n 是数据对的数目。当 $\rho=0$ 时，两个变量完全不相关， ρ 越接近于 1，说明两个变量正相关性越强， ρ 越接近于-1，说明两个变量负相关性越强。

首先，本着控制变量的原则，将数据按照不同的励磁波形、不同的材料分成 12 组，使得在探究温度对磁芯损耗独立影响的过程中，其他两个变量保持不变；接着计算各组的 Spearman 系数以及系数所对应的 P 值，若 P 值小于 0.05，则说明不同的温度下磁芯损耗有着显著性差异，结果如表 5-1 所示。

表 5-1 不同材料波形条件下温度和磁芯损耗的 Spearman 相关系数

材料	波形	斯皮尔曼系数	P 值
材料 1	正弦波	-0.087	0.004
材料 1	三角波	-0.075	0.005
材料 1	梯形波	-0.088	0.007
材料 2	正弦波	-0.122	0.000
材料 2	三角波	-0.050	0.113
材料 2	梯形波	-0.089	0.007
材料 3	正弦波	-0.080	0.011
材料 3	三角波	-0.085	0.005
材料 3	梯形波	-0.101	0.001
材料 4	正弦波	-0.080	0.018
材料 4	三角波	-0.009	0.736
材料 4	梯形波	-0.176	0.000

(1) 无论磁芯材料、励磁波形为哪一种，温度与磁芯损耗始终呈负相关

(2) 材料 2 的正弦波、材料 3 的梯形波、材料 4 的梯形波 P 值均接近于 0，说明这三组磁芯损耗与温度具有较强的统计显著性。

(3) 材料 2 中的三角波、材料 4 的三角波 P 值大于 0.05，且斯皮尔曼系数均较小，说明这两组磁芯与温度线性相关性较小，结合 5.2 中温度和波形以及温度和材料的协同作用探索分析来看，材料 2 的磁芯损耗与温度的关系、励磁波形为正弦波的磁芯损耗与温度的关系确实不呈负相关，90℃时的平均磁性损耗已经高于 70℃时的平均磁性损耗。

(4) 总体来看，各组的斯皮尔曼系数均不小于 -0.2，说明温度与磁芯损耗并不呈明显的负相关，出现该情况也合理，从问题二中修正的斯坦麦茨方程来看，温度与磁芯损耗呈幂指数负相关，为 5.4 节中建立磁芯损耗与温度、材料、波形的多元回归模型提供了思路。

B、波形、材料对磁芯损耗的相关性检验

由于磁芯损耗数据不服从正态分布，且波形和材料均为离散无序变量，所以本文选用 Kruskal-Wallis 检验分析不同波形以及材料对磁芯损耗的影响。

Kruskal-Wallis (K-W) 检验是一种非参数检验方法，用于比较三个或更多独立样本的差异。K-W 检验的基本思想是：如果各组样本来自同一总体，那么各组样本的秩和应该相近。具体步骤如下：

Step1. 将所有样本的磁芯损耗数据合并，并按照从小到大的顺序排列，记录每个数据点的秩。

Step2. 对每个组内的秩求和，计算每个组的平均秩。

Step3. 如果各组样本来自同一总体，那么各组的平均秩应该相近。如果某些组的平均秩明显偏高或偏低，则可能表明这些组与其他组存在显著差异。

Step4. 计算 Kruskal-Wallis 检验的统计量 H ，并根据 H 的值和自由度，得到 P 值，若 P 值小于 0.05，则说明不同的材料或波形下磁芯损耗有着显著性差异。

需要注意的是，Kruskal-Wallis 检验只能判断是否存在显著差异，但无法确定具体是哪些组别之间存在差异。所以，对于励磁波形而言，一共有正弦波、三角波、梯形波三组，两两之间做一次 K-W 检验，共需要进行 3 次；同样地，对于材料而言，一共有 4 组，两两之间做一次 K-W 检验，共需要进行 6 次。通过 K-W 检验，励磁波形以及材料对磁芯损耗的影响分别如下表 5-2、表 5-3 所示。

表 5-2 不同励磁波形的 K—W 检验

组别	H 统计量	P 值
三角波 vs 梯形波	0.052	0.820
三角波 vs 正弦波	668.523	0.000
梯形波 vs 正弦波	532.559	0.000

正弦波与梯形波、三角波的 H 检验统计量较大， P 值接近于 0，故正弦波的磁芯损耗与梯形波、三角波存在显著差异，从而说明励磁波形对磁芯损耗存在影响，但梯形波与三角波的 H 检验统计量较小， P 值为 0.820，远远大于 0.05，故梯形波与三角波的磁芯损耗并不存在显著差异。

各材料组别的 K-W 检验 H 统计量均较大， P 值均小于 0.05，故材料 1、材料 2、材料 3、材料 4 中的磁芯损耗均存在显著差异，从而说明材料对磁芯损耗存在影响。

以上结论均与 () 中单一变量对磁芯损耗影响的探索性数据分析结论一致。

5.3.2 因素间协同作用的分析

基于 () 节温度、材料以及励磁波形两两之间的协同因素对磁芯损耗的探索性分析后，

本节将深入探索协同因素对磁芯损耗的影响。当探索多分类变量间协同作用对连续变量的影响时，常用的方法有多因素方差分析、交互作用分析、协方差分析等。由于题目限定仅讨论两两之间的协同影响，故选用双因素方差分析。

A、双因素方差分析

双因素方差分析用于分析两个分类型自变量和一个数值型因变量之间的关系，判断这两个因素是否对因变量有显著影响，以及这两个因素之间是否存在交互作用。数据中观测值与整体平均值之间的差异总量被定义为总“变异”，衡量的是所有观测值与整体平均值之间的总距离的平方和，数学表达式为：

$$SST = \sum_{i=1}^n (Y_i - \bar{Y})^2 \tag{5.2}$$

其中，SST 是总平方和， Y_i 是单个观测值， \bar{Y} 是所有观测值的平均值， n 是观测值的总数。这种变异可以归因于不同的来源，包括由研究中的因素引起的变异、因素之间交互作用引起的变异，以及随机误差引起的变异。在双因素方差分析中，总变异（SST）被进一步分解为：由因素 A 的不同水平引起的变异、由因素 B 的不同水平引起的变异、由因素 A 和因素 B 的交互作用引起的变异。使用 F 检验确定每个因素或交互作用是否对因变量有显著影响，比较因素或交互作用引起的变异与误差变异的比率，数学表达式为：

$$F = \frac{MS_1}{MS_2} \tag{5.3}$$

其中， MS_1 是因素或交互作用的均方，通过该因素或交互作用的平方和除以相应的自由度所得， MS_2 是误差的均方，是误差平方和除以误差的自由度。若 F 值大于 1，且对应的 P 值小于 0.05，则可认为该因素或交互作用对因变量有显著影响。

B、协同作用分析结果

依次进行磁芯材料和温度、磁芯材料和励磁波形、温度和励磁波形的协同作用方差分析，结果如下表（）所示。

温度和磁芯材料的交互作用 F 值为 0.675，小于 1，P 值为 0.732，明显大于 0.05，故认为温度和磁芯材料的交互作用对磁芯损耗没有显著影响，只存在温度以及磁芯材料对磁芯损耗的独立影响，该结论均与（）中温度与材料的交互作用对磁芯损耗影响的探索性分析结论一致；磁芯材料和励磁波形的交互作用 F 值为 15.090，P 值约为 0，说明励磁波形和磁芯材料的交互作用对磁芯损耗有显著性影响；同样，温度和励磁波形的交互作用 F 值为 2.518，P 值约为 0，说明温度和励磁波形的交互作用对磁芯损耗有显著性影响。

表 温度和磁芯材料的交互作用方差分析		
	F 值	P 值
温度	17.233	0.000
磁芯材料	98.549	0.000
温度~磁芯材料	0.675	0.732

表 磁芯材料和励磁波形的交互作用方差分析	
	P 值

磁芯材料	108.971	0.000
励磁波形	293.223	0.000
磁芯材料~励磁波形	15.090	0.000

表 温度和励磁波形的交互作用方差分析

	F 值	P 值
温度	22.467	0.000
励磁波形	284.807	0.000
温度~励磁波形	2.518	0.019

5.4 多元回归分析

根据（）节与（）节的结论，磁芯损耗的大小受到温度、磁芯材料、励磁波形、温度与励磁波形的交互项以及磁芯材料与励磁波形的交互项影响。本文构造以磁芯损耗为因变量，励磁波形、温度、材料为自变量的多元回归模型。

5.4.1 数据预处理

由于励磁波形和磁芯材料均为分类变量，如果要将其引入回归模型，需要将其转换成数值变量，我们在这里使用独热编码（One-hot Encoding）进行处理。独热编码通常用于处理非数值属性，将分类变量转换为 0-1 变量，既不会引入任何类别之间的无序关系，也不会损失信息。对于一个具有 N 个可能类别的分类变量，独热编码会创建 N 个新的二进制列，对于每个类别，只有当前类别列的值是 1 其余列的值都是 0。该多元回归模型的自变量中励磁波形、磁芯材料均为分类变量，在进行模型构建之前，将三种励磁波形转化为 3 个 0-1 变量，四种磁芯材料则转化为 4 个 0-1 变量，下表（）、表（）分别为波形变量、材料变量的独热编码结果。

表 波形变量的独热编码

	W_1	W_2	W_3
正弦波	1	0	0
三角波	0	1	0
梯形波	0	0	1

表 材料变量的独热编码

	M_1	M_2	M_3	M_4
材料 1	1	0	0	0
材料 2	0	1	0	0
材料 3	0	0	1	0
材料 4	0	0	0	1

5.4.2 构建模型

由于波形变量通过独热编码分别转化为 3 个 0-1 变量,材料变量转化为 4 个 0-1 变量,所以温度和波形的交互项转化为 3 个 0-1 变量,材料和波形的交互项转化为 12 个 0-1 变量。加上原有的温度项以及常数项,共有 23 个自变量。本文将构建以磁芯损耗关于这 23 个自变量的回归模型。受到问题二修正斯坦麦茨方程的启发,磁芯损耗与温度具有非线性关系,故再添加一个自变量:温度的二次项,最终的模型共包含 24 个自变量。

多元回归模型如下:

$$P = f(T, T^2, W_i, M_j, T \times W_i, W_i \times M_j) \quad (5.4)$$

假设 f 为简单的线性相关,利用 Lasso 方法拟合该多元回归模型,得到各自变量的系数以及常数项,各自变量的系数可以理解为影响程度,结果如下:

$$\begin{aligned} P = & 18830 - 2242 \times T + 16.42 \times T^2 \\ & + 109900 \times W_1 + 127900 \times W_2 - 54480 \times W_3 \\ & + 30200 \times M_1 + 78950 \times M_2 + 93440 \times M_3 - 19240 \times M_4 \\ & - 734 \times T \times W_1 - 1326 \times T \times W_2 - 181 \times T \times W_3 \\ & + 18790 \times W_1 \times M_1 - 39160 \times W_1 \times M_2 - 44420 \times W_1 \times M_3 + 10310 \times W_1 \times M_4 \\ & + 13550 \times W_2 \times M_1 + 72530 \times W_2 \times M_2 + 51360 \times W_2 \times M_3 - 27530 \times W_2 \times M_4 \\ & - 2143 \times W_3 \times M_1 + 45580 \times W_3 \times M_2 + 86490 \times W_3 \times M_3 - 2015 \times W_3 \times M_4 \end{aligned} \quad (5.5)$$

5.4.3 优化回归模型

(1) 单目标优化模型

基于 5.5.2 节中构建的多元回归模型,以磁芯损耗最小为目标,以 T 、 T^2 、 W_1 、 W_2 、 W_3 、 M_1 、 M_2 、 M_3 、 M_4 为决策变量,以各决策变量的取值范围为约束条件建立单目标优化模型。其中 W_i 为是否对该磁芯施加第 i 种励磁波形, M_j 为该磁芯是否为第 j 种材料,则有

$$W_i = \begin{cases} 0, & \text{不是第 } i \text{ 种波形} \\ 1, & \text{是第 } i \text{ 种波形} \end{cases} \quad i = 1, 2, 3 \quad (5.6)$$

$$M_j = \begin{cases} 0, & \text{不是第 } j \text{ 种材料} \\ 1, & \text{是第 } j \text{ 种材料} \end{cases} \quad j = 1, 2, 3 \quad (5.7)$$

目标函数 1 最小化磁芯损耗,即

$$\min P = G(T, W_i, M_j) \quad (5.8)$$

其中 G_i 为 5.5.2 节中构建的多元回归模型。

约束条件如下:

- (1) 根据附件 1 中的数据,温度只能有四个取值: $T = 25$ 、 $T = 50$ 、 $T = 70$ 和 $T = 90$
- (2) 每一种磁性元件只能属于一种材料,且呈现一种励磁波形,即

$$\sum_{i=1}^3 W_i = 1, \sum_{j=1}^4 M_j = 1$$

综上，实现磁芯损耗最小化的模型为：

$$\begin{aligned} \min P &= G(T, W_i, M_j) \\ \text{s.t.} \quad &\left\{ \begin{array}{l} T = \begin{cases} 25 \\ 50 \\ 70 \\ 90 \end{cases} \\ \sum_{i=1}^3 W_i = 1, \sum_{j=1}^4 M_j = 1 \end{array} \right. \end{aligned} \quad (5.9)$$

(2) 遗传算法

本文利用遗传算法对磁芯损耗最小化的模型进行求解。遗传算法通过模拟自然选择和遗传机制来解决优化问题。具体步骤如下：

Step1.编码。遗传算法中的每个解都被视为一个“个体”，一组解组成“种群”。个体通常用一串数字（二进制）来表示，这串数字编码了问题的潜在解。

Step2.生成初始种群，定义适应度函数。定义一个适应度函数来评估个体的适应度，本题中，适应度函数即为目标函数。适应度越高，个体被选择繁殖的概率越大。

Step3.选择、交叉、变异。首先根据个体的适应度，选择较优的个体进入下一代，本文选用轮盘赌机制，接着随机配对选中的个体，按照交叉概率进行交叉操作，产生新的个体。交叉点可以是单点或多点，最后对新生成的个体以一定的变异概率进行变异操作，以引入新的遗传多样性，本文选用随机位翻转。

Step4.形成新一代种群。根据适应度选择一定数量的个体保留到下一代，形成新的种群。若达到最大迭代次数，将停止优化。最优个体解码则为问题的解。

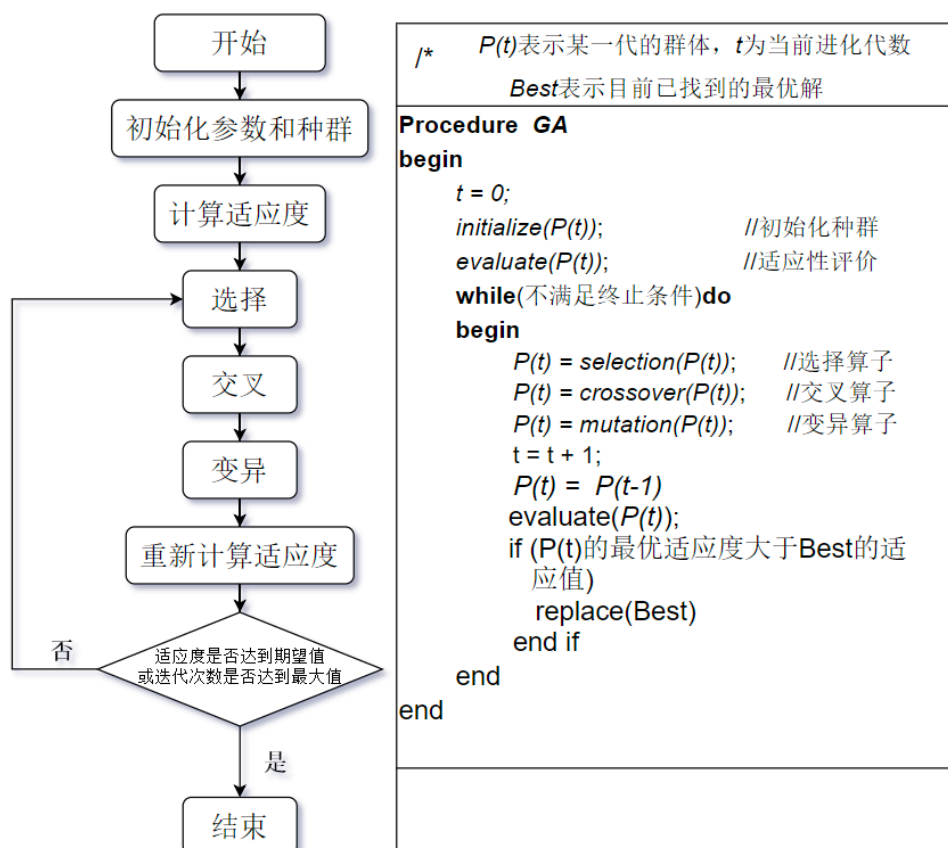


图 遗传算法流程图

6 问题四的分析与求解：基于数据驱动的磁芯损耗预测模型

6.1 问题分析

在磁芯损耗预测领域，传统的模型因精度不足或适用范围有限而难以满足现代电力电子技术对精确性日益增长的需求。故问题四旨在开发一种基于数据驱动的新型模型，该模型应能够跨越不同材料类型与复杂工况条件，提供高精度且广泛适用的磁芯损耗预测。本题的解题思路为：首先将问题一中计算的 12 个时域和频域特征与磁芯材料、温度、频率、励磁波形以及因变量磁芯损耗合并，将磁芯材料与励磁波形两个定量变量进行数据编码，共得到 16 个特征，来作为预测磁芯损耗的依据；然后，使用随机森林算法绘制特征重要性图，提取出最大值（即磁通密度峰值）、频率、温度、磁芯材料和励磁波形 5 个特征；接下来利用随机森林、GBDT、XGBoost 和 LightGBM 四种机器学习模型拟合数据，构建磁芯损耗预测模型；选出效果最好的模型，来完成对附件三数据的预测。

6.2 模型建立

6.2.1 数据预处理

首先对磁芯材料和励磁波形两个分类变量进行数据编码。“材料 1”~“材料 4”分别编码为 1~4；“正弦波”、“三角波”和“梯形波”分别编码为 1、2、3。

由于在问题一中，我们已经得其所提取的 8 种时域特征与 4 种频域特征，可用于识别和预测励磁波形；而在问题三中又得出“励磁波形对磁芯消耗有影响”这一结论，因此我们推断在前文计算的 12 个用于波形识别的特征同样可以用来预测磁芯损耗。再加上已知对磁芯消耗有影响的 4 种特征，则一共计算出 16 种特征初步用于磁芯消耗的预测。

6.2.2 特征提取

与问题一的思路相同，我们仍然使用随机森林算法对磁芯消耗进行回归，绘制特征重要性图进行特征提取。特征重要性图如（）所示。

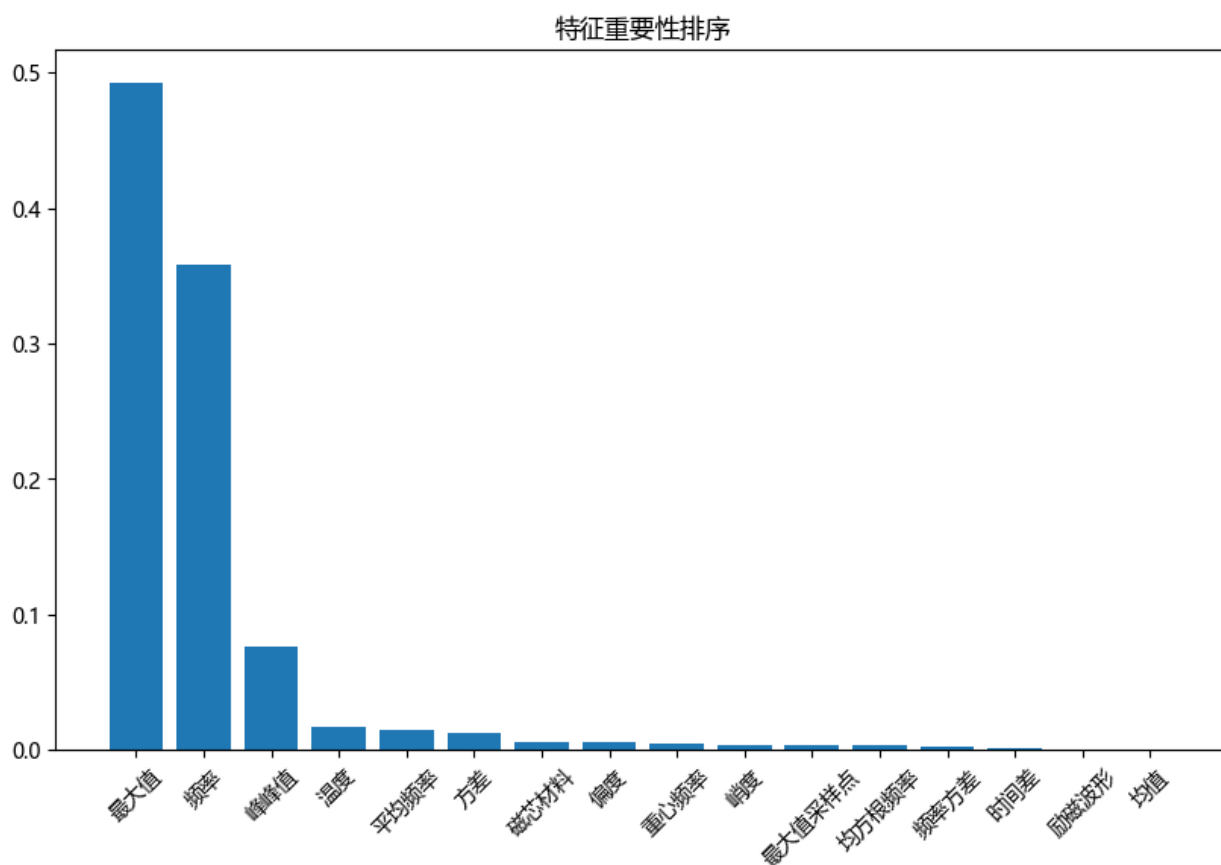


图 随机森林特征重要性排序

从图中可以看出，最大值（即磁通密度峰值 B_m ）对于磁芯消耗有着决定性的作用，占到全部特征重要性的 50%，其次是频率和峰峰值。剩余的特征对于预测磁芯损耗的作用不大。如果选取前三个特征进行建模，并对磁芯损耗进行预测，计算出模型的 $R^2 = 0.98$ 。

由于在第三题中已经探究出了不同磁芯材料与励磁波形对于磁芯损耗有着各自独立且两两协同的影响，因此最终我们选取最大值、频率、峰峰值、磁芯材料与励磁波形共五个特征作为预测磁芯损耗的特征依据。

6.2.3 预测模型的建立

A、梯度提升决策树（Gradient Boosting Decision Tree,GBDT）

GBDT 是一种通过逐步构建决策树来进行预测的算法。它采用梯度提升的思想，即通过不断减少预测残差来改进模型。每一棵新树都是通过拟合前一棵树的残差（即实际值与预测值之间的差异）来训练的。GBDT 算法的步骤为：

- Step1：初始化：通常用一个常数值（如训练集的均值）进行初始化。
- Step2：迭代：在每一轮中，计算当前模型的预测结果与实际值之间的残差，然后训练一棵新的决策树来拟合这个残差。
- Step3：更新：将新树的预测结果乘以学习率（通常小于 1），然后加到当前模型上，从而得到更新后的模型。
- Step4：重复上述过程直至达到预定的树数量或满足其他停止条件。

GBDT 算法能够处理非线性关系，并且可以自动处理特征之间的交互作用；但是算法复杂，尤其是在树的数量很多时，训练时间较长，且对于大型数据集和高维数据，容易过拟合。

B、LightGBM

LightGBM 是对 GBDT 的优化版本，它采用基于直方图的决策树算法，通过将特征值分段来构建直方图，代替原始特征，能够显著提高效率，减少计算量。与传统的 GBDT 按层生长不同，LightGBM 按叶子生长，可以提高模型的准确性。并且该算法采用了高效的并行和分布式训练技术，能够有效应对大规模数据集。LightGBM 算法的步骤为：

- Step1：定义类别特征，并建立特征的直方图。
- Step2：利用直方图进行特征的快速分裂，构建决策树。
- Step3：按照迭代过程逐步优化模型，每一轮都有新的决策树加入，不断调整和减少残差。

除上述两种算法外，我们还使用了随机森林算法和 XGBoost 算法，共建立四种磁芯损耗预测模型，使用合并后的含有 16 个特征的数据表格作为训练集对模型进行训练，最终的效果如表（）所示。

表 4 种预测模型的评价指标

模型	MSE	MAE	MRE	R^2
随机森林	1827277360	17361.4350	9.04%	0.9864
GBDT	3379518101.2775	22532.5767	11.74%	0.9749
XGBoost	1108022219.1002	14554.2045	7.58%	0.9918
LightGBM	1814307175.4280	17535.7711	9.13%	0.9865

可以看出，4 种模型的 R^2 均达到了 97%以上，模型效果非常好。以效果最好的 XGBoost 模型为例，分别以预测值和真实值为横轴和纵轴绘制散点图，以及在测试集上的相对误差图如（）和（）所示。

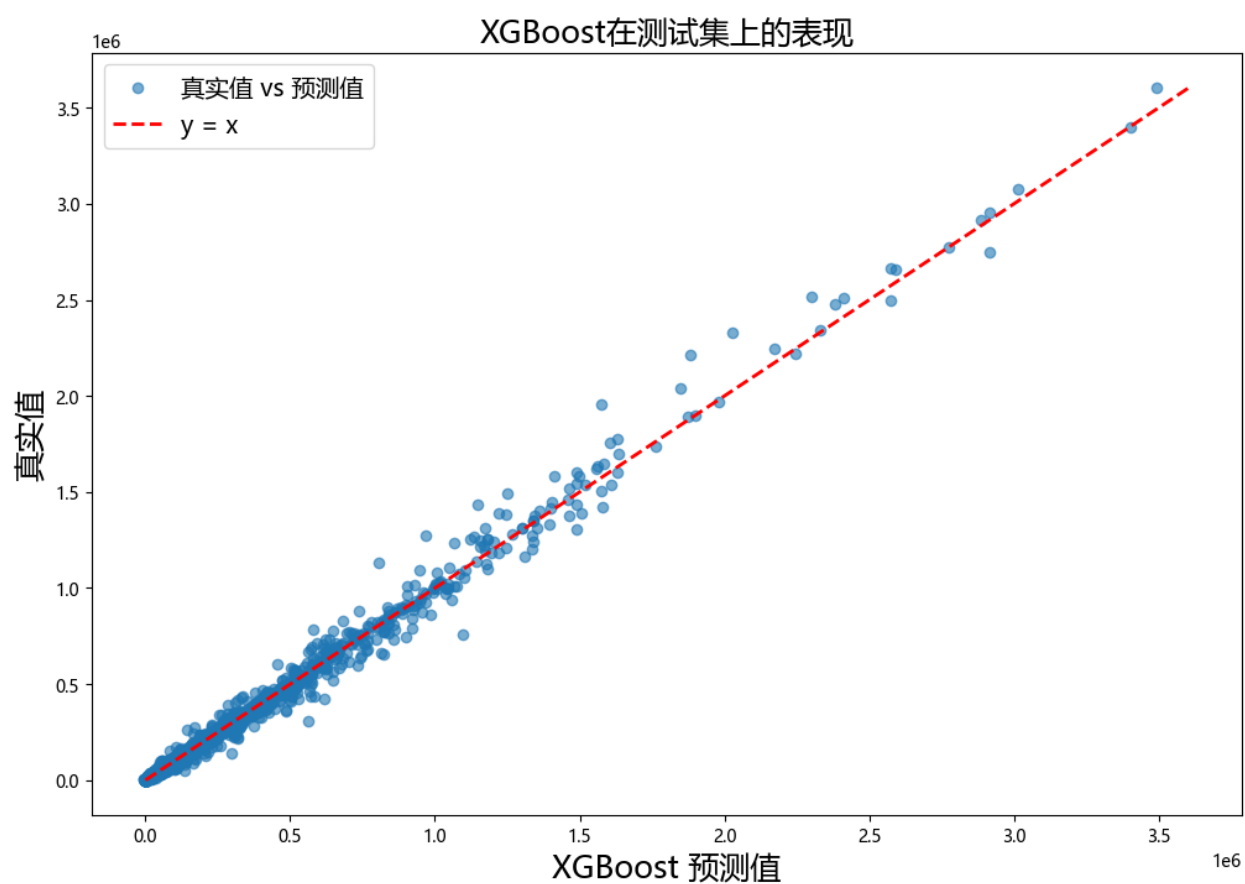


图 XGBoost 预测值—真实值散点图

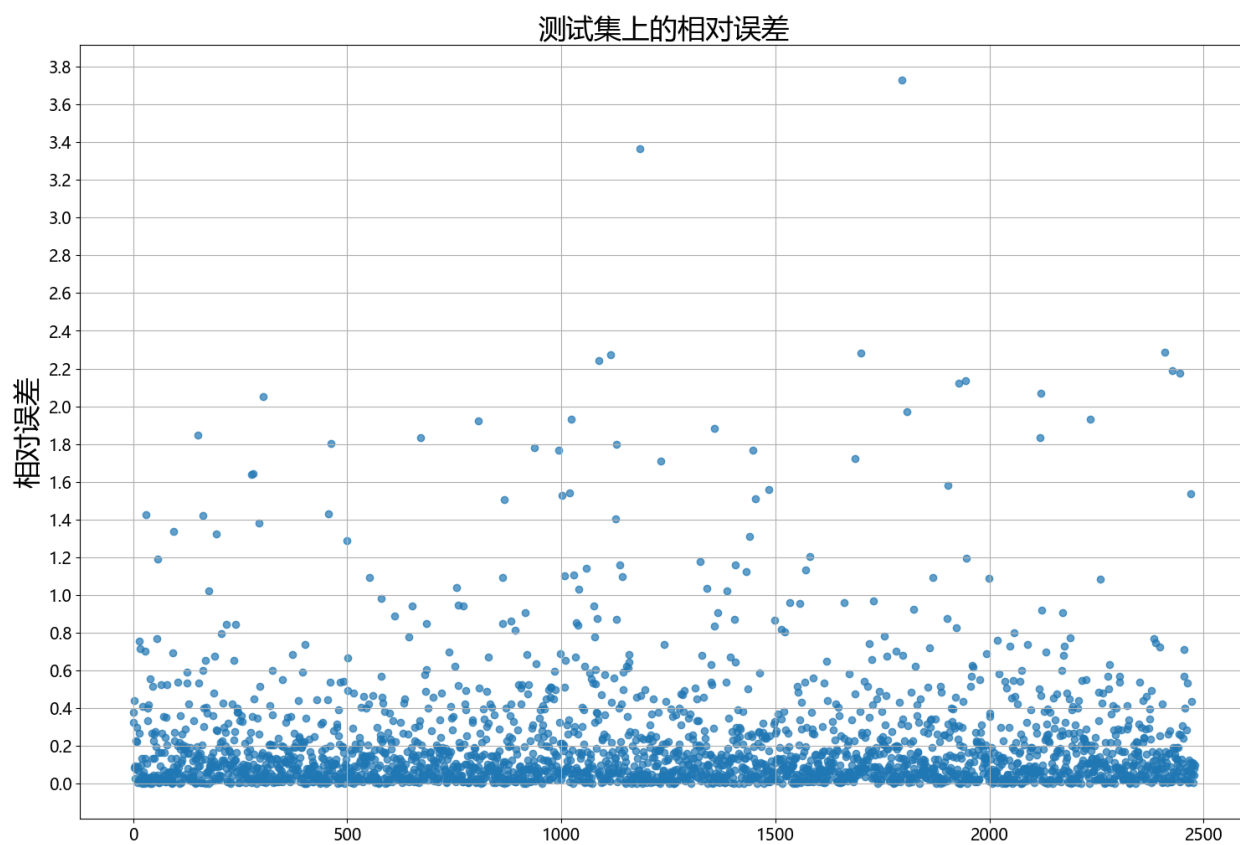


图 XGBoost 模型在测试集上的相对误差

6.2.4 L2 正则化

虽然在上一节中建立的四种预测模型的 R^2 都非常高，但是像 MSE 这种指标也变得非常大，甚至在进行训练的时候出现了预测磁芯损耗结果为负数的情况，初步断定模型中存在一定的过拟合现象。

为了预防这种现象，一般在进行模型的训练之前可以设置模型超参数来降低过拟合的风险。例如对于树模型，可以减少树的最大深度，或者进行剪枝处理等。本文则是使用了 L2 正则化的方法。L2 正则化又称为岭回归(Ridge Regression)或权值衰减(Weight Decay)，是一种在机器学习模型训练过程中常用的正则化技术。其主要目的是通过向损失函数中添加一个与模型参数平方和成正比的惩罚项，来控制模型参数的复杂度，从而防止过拟合现象的发生。

具体来说，在原始的损失函数基础上，L2 正则化引入了一个额外的项，该项等于所有模型参数平方和乘以一个正则化系数 λ （也称为惩罚因子）。这个正则化项的存在，使得模型在训练过程中不仅要最小化原始的预测误差，还要尽量减小模型参数的规模，即倾向于选择那些较小权重的参数组合。通过这种方式，L2 正则化能够有效地降低模型的复杂度，提高模型的泛化能力，使模型在面对新数据时表现更加稳健。

数学上，若以 L 表示原始的损失函数， W 表示模型参数向量，那么加入 L2 正则化后的损失函数可以表示为：

$$L_{reg} = L(W) + \lambda * \|W\|^2 \quad (6.1)$$

其中 $\|\cdot\|^2$ 表示参数向量的 L2 范数（即参数的平方和）。通过调整正则化系数 λ 的大小，可以控制正则化的强度： λ 值越大，正则化效果越强，模型的复杂性越低，但也可能因此牺牲掉一部分对训练数据的拟合度。

6.2.5 预测结果

问题中指定序号的样本预测结果如下表所示

表 附件三指定序号的样本预测结果

序号	预测结果	序号	预测结果
16	1126.3	230	69692.6
76	1924735.4	271	1896996
98	15911.1	338	12297.7
126	1894.3	348	833808.3
168	110094.5	379	1297.1

7 问题五的分析与求解：磁性元件的最优化条件

7.1 问题分析

问题五旨在分析磁性元件的最优化条件，要求磁芯损耗小的同时传输磁能（频率与磁通密度峰值的乘积）尽可能大，故本文建立双目标优化模型。首先由于材料以及波形为多分类变量，故对二者进行独立编码，分别设置 3 个以及 4 个 0-1 变量，在此基础上再加上温度、频率、最大磁通密度，从而共有 10 个决策变量，接着以最小磁芯损耗、最大磁通密度峰值为目标，基于 0-1 规划，以各决策变量的取值范围为约束条件，利用模拟退火算法建立双目标优化模型。

7.2 双目标优化模型

对于磁芯材料而言，只有材料 1，材料 2，材料 3，材料 4 四种情况，对于励磁波形而言，只有正弦波、三角波、梯形波三种情况，故可将这两个变量进行独立编码，基于 0-1 规划的思想建立模型，设 W_i 为是否对该磁芯施加第 i 种励磁波形， M_j 为该磁芯是否为第 j 种材料，则有

$$W_i = \begin{cases} 0, & \text{不是第 } i \text{ 种波形} \\ 1, & \text{是第 } i \text{ 种波形} \end{cases} \quad i = 1, 2, 3 \quad (7.1)$$

$$M_j = \begin{cases} 0, & \text{不是第 } j \text{ 种材料} \\ 1, & \text{是第 } j \text{ 种材料} \end{cases} \quad j = 1, 2, 3, 4 \quad (7.2)$$

磁性元件的最优化原则为磁芯损耗尽量小，传输磁能尽量大。

目标函数 1 最小化磁芯损耗，即

$$\min P = G(T, f, B_m, W_i, M_j) \quad (7.3)$$

其中 P 为磁芯损耗， T 为温度， f 为频率， B_m 为磁通密度的峰值， G 为问题四四种磁芯损耗与该 10 个自变量所构建构建的 XGBoost 模型。

目标函数 2 最大化传输磁能，可转化为最小化传输磁能的相反数，即

$$\min E = -f \times B_m \quad (7.4)$$

约束条件如下：

(1) 根据附件 1 中的数据，温度只能有四个取值： $T = 25$ 、 $T = 50$ 、 $T = 70$ 和 $T = 90$

(2) 根据附件 1 中的数据，频率的取值范围为 50000HZ 至 500000HZ，且为 10 的倍数，即

$$50000 \leq f \leq 500000, \quad f = 10k, k \in \mathbb{Z}^+$$

(3) 根据附件 1 中的数据，磁通密度峰值的取值范围为 0.0096 至 0.3133，即

$$0.0096 \leq B_m \leq 0.3133$$

(4) 每一种磁性元件只能属于一种材料，且呈现一种励磁波形，即

$$\sum_{i=1}^3 W_i = 1, \sum_{j=1}^4 M_j = 1$$

综上，考虑磁芯损耗均衡传输磁能的模型为：

$$\begin{aligned} \min P &= G(T, f, B_m, W_i, M_j) \quad , \quad \min E = -f \times B_m \\ s.t. \quad &\begin{cases} T = \begin{cases} 25 \\ 50 \\ 70 \\ 90 \end{cases} \\ 50000 \leq f \leq 500000 \quad , \quad f = 10k, k \in Z^+ \\ 0.0096 \leq B_m \leq 0.3133 \\ \sum_{i=1}^3 W_i = 1, \sum_{j=1}^4 M_j = 1 \end{cases} \end{aligned} \quad (7.5)$$

7.3 模拟退火算法

模拟退火算法的核心思想是允许算法在搜索过程中以一定的概率接受劣质解，从而跳出局部最优解，最终寻找到全局最优解。算法通过控制一个称为“温度”的参数来实现这一点。在较高温度下，算法更有可能接受劣质解，随着温度的降低，接受劣质解的概率逐渐减小。具体的算法求解过程如下：

Step1.初始化：选择一个初始解作为当前解，在该问题中该解为一个 10 维的向量，每一维度表示一个决策变量。并设置初始温度 T_0 ，以及降温参数 c ，迭代次数 n ，用于控制温度的下降速率。

Step2.迭代过程：在当前温度下，首先通过一定策略产生一个新的候选解，本文通过随机扰动的方式，接着计算当前和候选解之间的能量差，记为 ΔE ，如果 ΔE 为负值，即候选解优于当前解，则接受候选解作为新的当前解，但如果 ΔE 为正值，即候选解劣于当前解，则以一定的概率 P 接受候选解。这个概率通常由下式给出：

$$p = e^{\left(-\frac{T}{\Delta E}\right)} \quad (7.6)$$

其中， T 为当前温度。

Step3.降温。完成一定数量的迭代后，通过下式更新温度：

$$t' = t \times c \quad (7.7)$$

Step4.终止条件。若达到预定的迭代次数，则结束迭代，否则，返回步骤 2 继续迭代。

7.4 优化模型求解

本文运用 7.2.1 中所述的模拟退火算法求解磁芯损耗与传输磁能的双目标优化问题，其中初始解为随机生成的 10 维向量，模拟退火算法的初始温度 T_0 设为 100，迭代次数为 1000000 次，并且用加权求和的方式将多目标优化问题转化为单目标优化问题，即

$$\begin{aligned} \min & k_1 \times P - k_2 \times E = G(T, f, B_m, W_i, M_j) - f \times B_m \\ \text{s.t.} & \begin{cases} T = \begin{cases} 25 \\ 50 \\ 70 \\ 90 \end{cases} \\ 50000 \leq f \leq 500000, \quad f = 10k, k \in \mathbb{Z}^+ \\ 0.0096 \leq B_m \leq 0.3133 \\ \sum_{i=1}^3 W_i = 1, \quad \sum_{j=1}^4 M_j = 1 \end{cases} \end{aligned} \quad (7.8)$$

求得结果为：磁性元件的最优条件为材料一、励磁波形为正弦波、温度为 90、频率为 129700HZ、最大磁通密度为 0.0173，此时磁芯损耗为 854.58，传输磁能为 2240.06。

8 模型的评价与改进

8.1 模型优点

1. 在构建励磁波形分类模型时，不仅考虑了时域因素，还考虑了频域因素，且频域因素对模型的分类效果起到重大作用，使得建立的模型更加精确可信。
2. 将温度变量引入 SE 方程时，考虑了幂函数、多项式、指数函数等多种形式，使得修正的 SE 方程更有说服力。
3. 分析温度、励磁波形和磁芯材料这三个因素如何独立及协同影响着磁芯损耗时，做了探索性分析，使得结果可视化，也为之后检验相关性起到铺垫作用。
4. 构建磁芯损耗预测模型时，考虑随机森林、XGBoost、GBDT、LightGBM 等多种机器学习模型，有利于增强模型拟合效果。

8.2 模型缺点

1. 构建磁芯损耗与温度、材料、波形间的多元回归模型时，拟合优度为 0.08，精度较小，考虑到因为磁芯损耗数据波动大、特征较多等多方面影响造成的。

9 参考文献

- [1] 李莹莹,基于波形形态特征的单频机载激光雷达测深全波形数据分类,红外与激光工程,第 52 卷第 9 期: 1-10, 2023。
- [2] 俞武扬,方俊俊,制造单元布局问题的双目标模型与算法研究,运筹与管理,第 27 卷第 6 期: 1-7, 2018
- [3] 魏泽辉,基于多元非线性回归分析的光伏发电系统经济模型预测研究,电站系统工程,第 40 卷第 3 期: 1-5, 2024
- [4] 戴金辉,袁靖,单因素方差分析与多元线性回归分析检验方法的比较,统计与决策,2016 年第 9 期: 1-4, 2016
- [5] 尚小晶,田彦涛,李阳,王立刚,基于改进概率神经网络的手势动作识别,吉林大学学报(信息科学版),第 28 卷第 5 期: 1-8, 2010