

机器学习大作业——使用切比雪夫多项式优化的人工神经网络

何俊辰 320200945791

实验背景

随着核磁共振成像技术的愈发广泛应用，接受MRI的患者也在逐步增加，与此同时，使用植入式医疗设备的患者数量同样也在快速且持续的增长，这两部分人群的交集存在一个严重的安全问题。无源植入式医疗设备如图1所示：

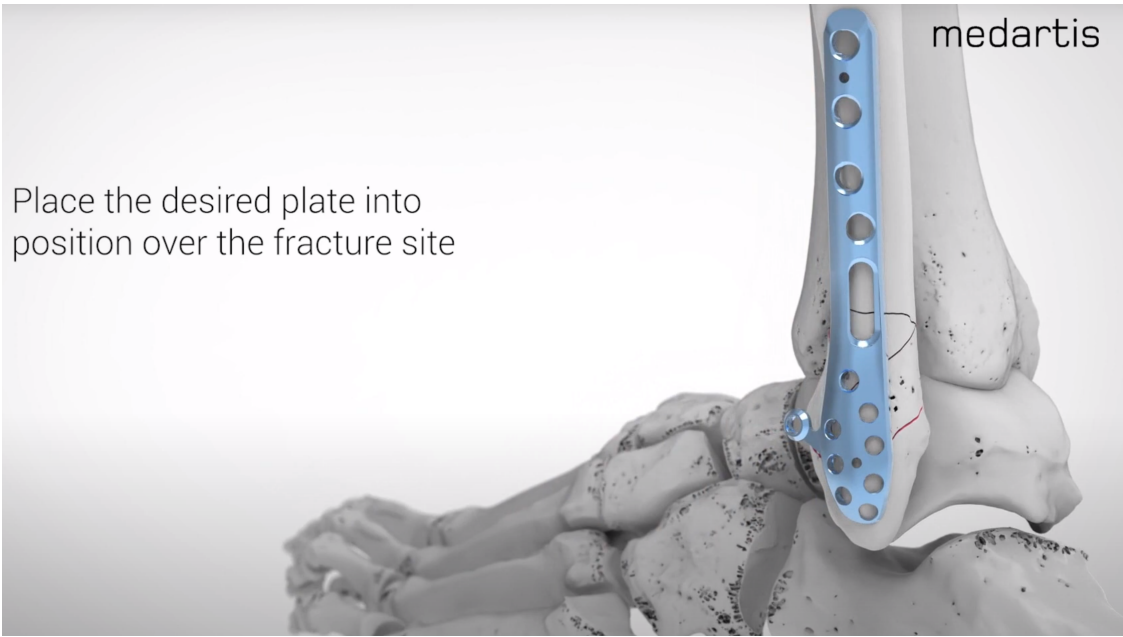


图1 无源植入物示例

在核磁共振的过程中，射频场会在人体内产生强电场并导致人体组织温度升高超过几度，这可能会对患者造成不可逆的健康损伤。本次作业希望使用一种简单但有效的方法：建立一个神经网络模型，通过输入植入物的简单几何特征，预测常见无源植入式医疗器械附近由于MRI过程中由于射频加热而导致的温升。

由于部分无源植入式医疗器械所具有的高维几何特征，本次实验中尝试引入切比雪夫多项式将植入物的特征放大，再将使用切比雪夫多项式处理后的特征作为输入输入神经网络进而获得更加优秀的模型性能以及较小的误差。

数据集介绍

对于无源植入式医疗器械中常见的板状植入物和钉状植入物，本次实验使用的数据集将其特征分别简化为长、宽、高（板状植入物）与半径、长度（钉状植入物），并且记录每个植入物暴露在射频加热中的温度升高情况。

本次实验共使用两个数据集，分别为板状植入物数据集（**screw**）以及钉状植入物数据集（**plate**）。板状数据集与钉状数据集的部分数据如表1、表2所示：

表1 板状植入物部分训练数据

长 (mm)	宽 (mm)	高 (mm)	最大温升 (mK)
225	25	2.0	0.175

长 (mm)	宽 (mm)	高 (mm)	最大温升 (mK)
25	15	3.5	0.294
25	5	2.5	0.0301
150	25	1.5	0.227
150	25	2.0	0.198

其中，板状植入物的长、宽、高三个特征在处理后将作为输入 X 输入神经网络进行训练，板状植入物的最大温升作为 y 训练神经网络的输出结果。该数据集内数据为300组，其中210组数据作为神经网络的训练集，另外90组数据作为神经网络的测试集。

表2 钉状植入物部分训练数据

半径 (mm)	长度 (mm)	最大温升 (mK)
3.5	170	0.285
2.5	200	0.323
0.5	90	0.242
4	10	0.0176
4	240	0.22

其中，钉状植入物的半径、长度两个特征在处理后将作为输入 X 输入神经网络进行训练，钉状植入物的最大温升作为 y 训练神经网络的输出结果。该数据集内数据为300组，其中210组数据作为神经网络的训练集，另外90组数据作为神经网络的测试集。

本次作业分别对板状植入物（**screw**）与钉状植入物（**plate**）数据集分别构建普通神经网络与切比雪夫多项式优化神经网络，并在后续部分展示构建的神经网络的性能并进行比较。

实验方法

数据归一化

对于植入式物的特征，首先使用 *MinMaxScaler* 对数据进行归一化，将数据的特征调整到0到1之间，从而避免特征间的量级差异对模型的表现产生影响，防止过拟合问题。

```
scaler = MinMaxScaler()
data_screw_nor = scaler.fit_transform(data_screw)
data_plate_nor = scaler.fit_transform(data_plate)
```

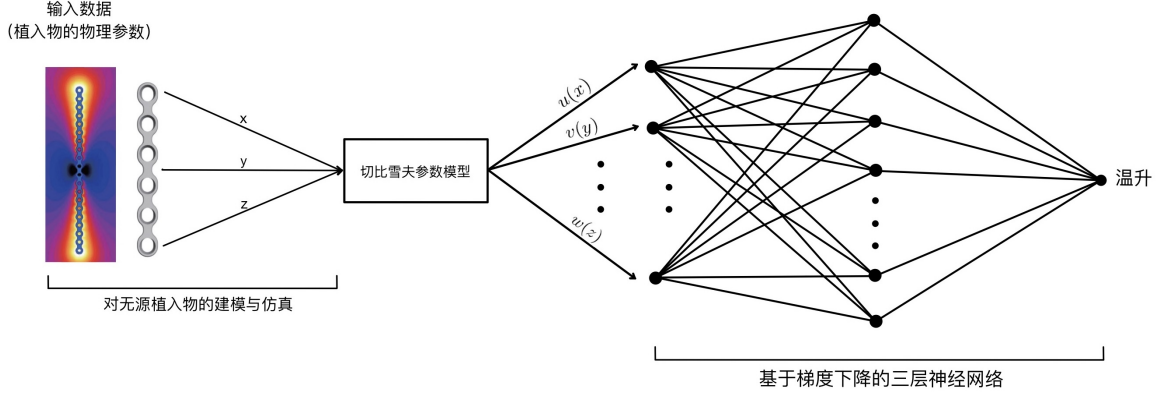
对于每一个特征 X ，做出以下处理：

$$X_{\text{norm}} = \frac{X - X_{\min}}{X_{\max} - X_{\min}}$$

其中， X_{norm} 是缩放后的特征值， X 是原始特征值， X_{\min} X_{\max} 分别是特征值的最小值和最大值。通过 *MinMaxScaler* 不会改变原始数据的分布，并且能够维持数据的稳定性，使得神经网络可以更快的收敛，降低模型学习的复杂性。

构建切比雪夫多项式优化的神经网络

构建切比雪夫多项式网络模型如 图2 所示：



构建切比雪夫多项式

使用切比雪夫多项式对特征进行缩放，创建一个基于特征的高维空间，从而更好地提供符合植入物几何形状的信息，提高模型的收敛速度，增加模型学习效率。以对钉状模型建立切比雪夫多项式为例：

```
def cheby_polynorm_screw(input_data):
    height = input_data.shape[0]
    output = np.zeros((height, 5))
    output[:,0] = 1
    output[:,1] = input_data[:,1]
    output[:,2] = 2*input_data[:,1] - 1
    output[:,3] = 4*input_data[:,1]*input_data[:,1]-3*input_data[:,1]
    output[:,4] = input_data[:,0]
    return output
```

切比雪夫多项式如下式所示：

$$T_0(x) = 1$$

$$T_1(x) = x$$

$$T_r = 2xT_{r-1}(x) - T_{r-2}(x), r = 2, 3, \dots$$

对于无源植入物的不同集合特征 x, y, z （长、宽、高）构建切比雪夫多项式，以提升输入的维度，增强模型的性能：

$$\mu(x) = [\phi_1 T_1(x), \phi_2 T_2(x), \dots, \phi_L T_L(x)]^T = \Phi^T U$$

$$v(y) = [\psi_1 T_1(y), \psi_2 T_2(y), \dots, \psi_M T_M(y)]^T = \Psi^T V$$

$$\varpi(z) = [\omega_1 T_1(z), \omega_2 T_2(z), \dots, \omega_N T_N(x)]^T = \Omega^T W$$

其中， $\mu(x)$ 表示某无源植入式医疗器械的物理特征 x （长度）的切比雪夫多项式模型， $v(y)$ 表示某无源植入式医疗器械的物理特征 y （宽度）的切比雪夫多项式模型， $\mu(z)$ 表示某无源植入式医疗器械的物理特征 z （高度）的切比雪夫多项式模型， $[\bullet]^T$ 表示矩阵的转置。将构建的多个切比雪夫多项式模型在下一步中作为神经网络的输入，从而进行神经网络的训练。

神经网络结构

使用Keras创建一个三层的全连接神经网络：

1. 首先创建一个**Sequential**模型，使用对叠层来完成神经网络的创建。

```
screw_model_chevby = Sequential()
```

2. 创建网络的输入层，在第一层中定义输入的形状，表示输入无源植入物的几何特征的输入维度。

```
screw_model_chevby.add(Dense(units = 5, kernel_initializer='he_normal',  
bias_initializer='zeros', activation='relu', input_shape=(5,)))
```

3. 创建网络的隐藏层，定义十个神经元。

```
screw_model_chevby.add(Dense(units = 10, kernel_initializer='he_normal',  
bias_initializer='zeros', activation='relu'))
```

4. 创建网络的输出层，创建一个神经元，用于预测植入物在MRI检测后的最大温升。

```
screw_model_chevby.add(Dense(units = 1, kernel_initializer='he_normal',  
bias_initializer='zeros', activation='relu'))
```

5. 训练神经网络，将迭代值定义为4000。

```
history_plate_ANN = plate_model_ANN.fit(X_train_plate, y_train_plate,  
epochs=4000)
```

初始化权重

在创建的神经网络中，使用 **he_normal** 方法进行权重初始化，该方法使得当前神经元的权重根据前一层的神经元数量来进行初始化，使得每个神经元的输出的方差应该尽可能保持不变。具体来说，如果一个神经元有n个输入，那么它的权重将被初始化为一个均值为0、标准差为 $\sqrt{2/n}$ 的正态分布中的随机值。

he_normal 初始化适合ReLU作为激活函数的神经网络，该激活函数使得神经元的输出在正数部分是线性的，可以保证输出的方差和输入的方差相等。这种初始化方法可以有效地防止深度网络中的梯度消失或爆炸问题，从而加快网络的收敛速度，提高训练效率。

在定义初始化权重前，训练的神经网络中出现了神经网络在经过迭代后损失值 (*loss*) 无改变的问题，推测是由于神经网络的权重在训练开始之前通常被随机初始化，而不同的初始化可能会导致网络收敛到不同的局部最小值，从而导致神经网络的损失值无法下降问题，迭代过程如 **图3** 所示：

```
Epoch 100/4000  
7/7 [=====] - 0s 3ms/step - loss: 0.0025  
Epoch 101/4000  
7/7 [=====] - 0s 2ms/step - loss: 0.0025  
Epoch 102/4000  
7/7 [=====] - 0s 2ms/step - loss: 0.0025  
Epoch 103/4000  
7/7 [=====] - 0s 2ms/step - loss: 0.0025  
Epoch 104/4000  
7/7 [=====] - 0s 2ms/step - loss: 0.0025
```

图3 迭代过程loss不变问题

在定义 **he_normal** 方法后，该问题得到有效解决。

Adam优化器

使用Adam方法作为优化器，其为每个参数独立地维护一个学习率，这些学习率根据参数的一阶矩（均值）和二阶矩（未标准化的方差）的估计值来调整，从而根据每个参数的梯度动态调整学习率。另外，Adam包含一个偏置修正的步骤，能够纠正偏置估计值在训练初期的低估。

ReLU激活函数

该神经网络的激活函数使用ReLU激活函数，该函数对负的输入值返回0，对于正的输入值返回输入本身：

$$f(x) = \max(0, x)$$

在该表达式中， x 是输入， $f(x)$ 是函数的输出。

使用ReLU激活函数能够有效缓解梯度消失问题，ReLU的梯度只有0与1。这使得在反向传播过程中，正向传播的激活值不会影响梯度的大小，从而缓解了深度神经网络中常见的梯度消失问题。

反向传播

神经网络在每一次迭代过程中均使用反向传播算法将神经网络每个节点的权值进行修正直至误差处于可接受的范围内：

$$\delta_n = P_n - y_n$$

$$e = AM + b$$

$$A'_n = A_n + \eta \delta_n \frac{df(e)}{d(e)} M$$

其中， y_n 表示植入物在MRI中的实际温升， p_n 表示每层神经网络预测得到的值， δ_n 表示单个节点得到的误差， A_n 表示节点的权值， A'_n 表示由反向传播算法得到的该节点的新的权值。

构建普通神经网络

构建普通神经网络模型如 图4 所示：

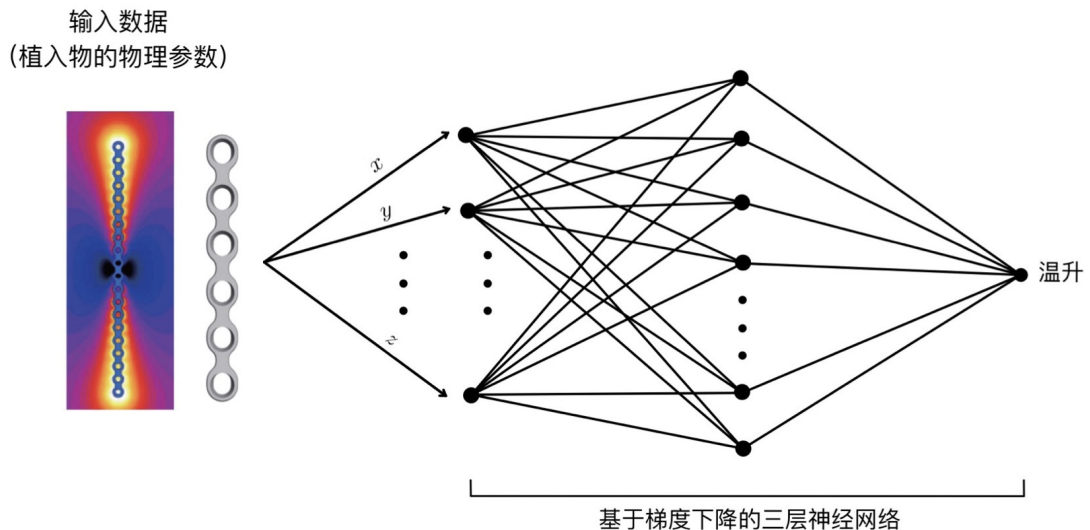


图4 普通神经网络模型

除了基于切比雪夫多项式优化的神经网络外，还构建一个普通神经网络，用于与优化后的神经网络进行对比，该神经网络结构与切比雪夫多项式优化神经网络相同，但将输入值更改为无源植入物的几何特征。神经网络输入层代码如下所示：

```
screw_model_ANN.add(Dense(units = 2, kernel_initializer='he_normal',
bias_initializer='zeros', activation='relu', input_shape=(2,)))
```

在该神经网络中，迭代次数同样被设定为4000次，用来与优化后的神经网络进行比较。

实验结果与分析

钉状植入物（screw）数据集实验结果：

在上述步骤中，分别使用切比雪夫多项式优化神经网络与普通神经网络对钉状植入物进行预测，两个网络的评价指标如 **表3** 所示：

表3 钉状植入物神经网络评价指标

	MSE	MAE
Screw_ANN	0.01217	0.08225
Screw_Chebby	0.00069	0.01964

由 **表3** 中数据可以看出，**Screw_Chebby**（切比雪夫多项式优化模型）表现远好于**Screw_ANN**（普通神经网络模型），该优化模型的MSE与MAE均远小于另一模型，表示该模型能够更加精确的预测钉状植入物的输入特征，提供更加稳定的预测结果。

预测值vs真实值散点图：

在上述过程中对钉状植入物构建的预测值vs真实值散点图如**图5**所示：

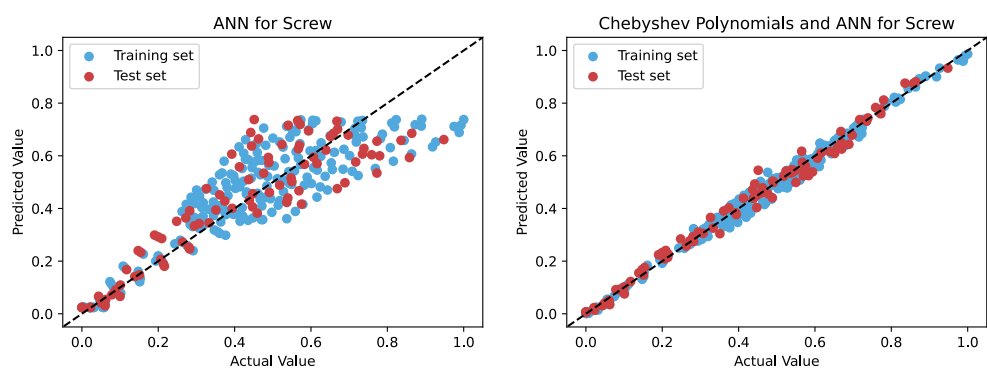


图5 钉状植入物真实值vs预测值

在该图中，对角线是一条 $y = x$ 函数，当图中点越接近对角线，表示模型的预测效果越好。当模型的点分布在对角线附近时，表示模型的预测结果与真实值非常接近。如果散点分布得很开，这可能表示模型的预测能力较差。

在图中，蓝色点表示用于训练的数据的预测效果，红色点表示被用于测试的数据的预测效果。从图中可以看出，两个模型在训练过程中均不存在过拟合问题。

当迭代次数为4000（epochs=4000）时，切比雪夫多项式优化的神经网络模型预测效果远好于普通神经网络。对于普通神经网络，真实值和模型预测值间存在着较大的误差，模型无法较好表征无源植入物几何参数与MRI最大温升的关系。而钉状植入物经由切比雪夫多项式进行高维表征后，预测值与真实值间差距很小，能够很好的表现模型输入值 X 与输出值 y 间关系。

残差图：

在上述过程中对钉状植入物构建的残差图如图6所示：

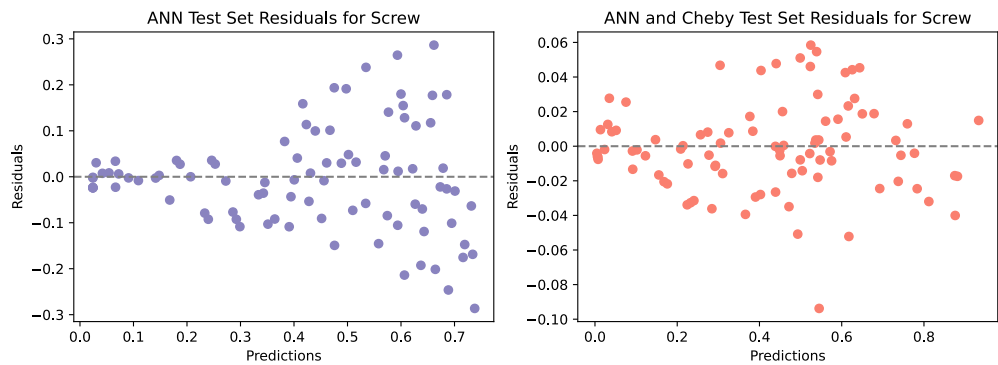


图6 钉状植入物残差图

对于钉状植入物构建的两个模型，为测试集创建了残差图，通过绘制每个观测值的残差（即预测值与真实值之差）与预测值（或观测值的其他度量）的关系，以评价回归模型的性能。

由残差图所示，切比雪夫多项式优化的神经网络模型的残差均匀地分布在零线的两侧，没有可识别的模式，说明模型没有遗漏任何重要的信息或趋势。同时，模型的残差值越小，表示模型的性能越好。而普通神经网络的残差分布不均匀，说明普通神经网络的模型存在较大误差。

从两个模型对比可以看出，切比雪夫多项式优化的神经网络模型较之普通的神经网络模型，残差图的残差点分布更为均匀，且残差均远小于普通神经网络残差值，说明切比雪夫多项式优化的神经网络性能更好。

板状植入物（plate）数据集实验结果：

在上述步骤中，分别使用切比雪夫多项式优化神经网络与普通神经网络对板状植入物进行预测，两个网络的评价指标如 表4 所示：

表4 板状植入物神经网络评价指标

	MSE	MAE
Plate_ANN	0.00204	0.03592
Plate_Chevby	0.00040	0.01608

由 表4 中数据可以看出，Plate_Chevby（切比雪夫多项式优化模型）表现远好于Plate_ANN（普通神经网络模型），该优化模型的MSE与MAE远小于另一模型，表示该模型对于板状植入物输入的几何特征更加敏感，优化后的模型拥有更好的预测精确度。

预测值vs真实值散点图：

在上述过程中对板状植入物构建的预测值vs真实值散点图如图7所示：

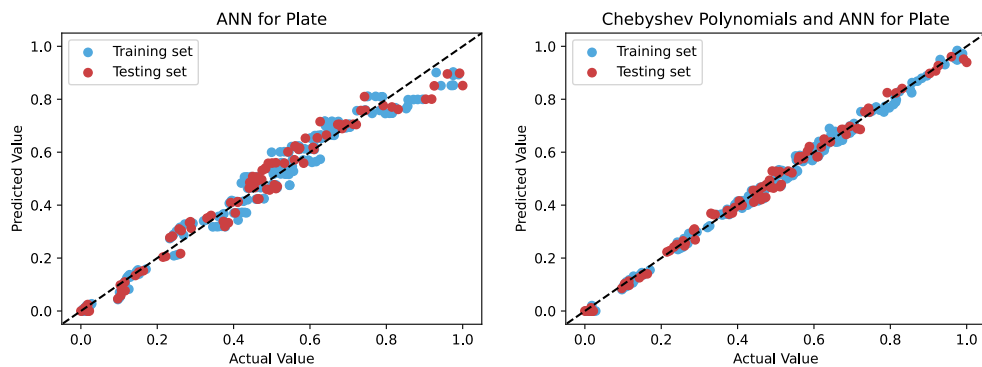


图7 板状植入物真实值vs预测值

在图中，蓝色点表示用于训练的数据的预测效果，红色点表示被用于测试的数据的预测效果。从图中可以看出，两个模型在训练过程中均不存在过拟合问题。

当迭代次数同样为4000 (epochs=4000) 时，从图中可以看出，对于板状植入物数据集，切比雪夫多项式优化后的神经网络比普通神经网络更贴近于图中对角线，表示优化后神经网络能够更好地拟合植入物几何特征与温升间关系。

残差图：

在上述过程中对板状植入物构建的残差图如图8所示：

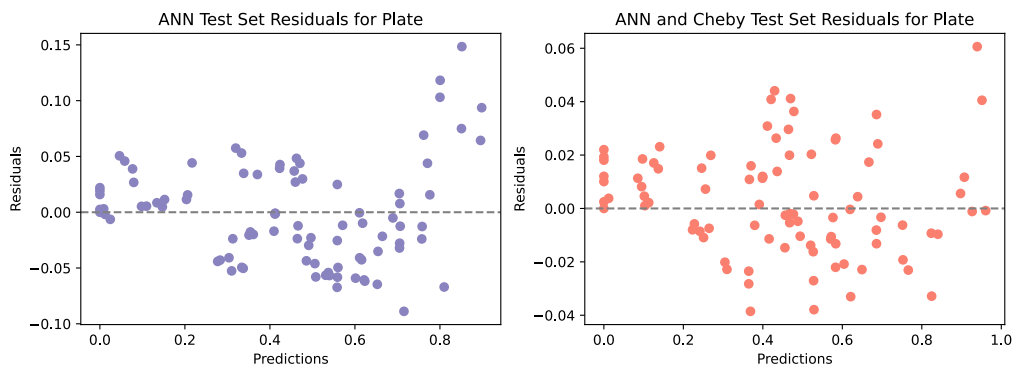


图8 板状植入物残差图

对于板状植入物构建的两个模型，同样为测试集创建了残差图，通过绘制每个观测值的残差（即预测值与真实值之差）与预测值（或观测值的其他度量）的关系，以评价回归模型的性能。

由残差图所示，两个模型的残差点均较均匀的分布在零线两侧，说明两个模型均对植入物几何信息有良好的表征。但由切比雪夫多项式优化后的神经网络模型的分布更加均匀，表示该模型的性能更好。同时，切比雪夫多项式优化的神经网络模型的残差同样远小于普通神经网络残差，表面其具有更好地性能。

总结

在本次作业中，探讨了神经网络模型与切比雪夫多项式结合的应用。着重研究的是无源植入物，这类物体的几何形状通常复杂多变。因此，我尝试采取切比雪夫多项式来对这些无源植入物的几何特征进行有效的表征。这种表征方法将几何特征转化为一系列可以量化的参数，这些参数之后可以被神经网络模型利用，为无源植入物的温升预测提供依据。

为了进一步验证该方法，构建了普通神经网络，比较了传统神经网络模型和切比雪夫多项式优化的神经网络模型在不同数据集上的预测性能。结果显示，切比雪夫多项式优化的神经网络模型在预测性能上明显优于传统的神经网络模型，这强烈地表明了切比雪夫多项式在处理这类复杂几何形状问题上的优势。

总结来说，通过将神经网络模型与切比雪夫多项式结合，能够更准确地预测无源植入物的温升，是一种有效的处理该类问题的神经网络优化方法。