

2024 年第二届大湾区杯科技竞赛

题 号 CG2407

标 题 基于机器学习的超导转变温度预测

| | | | |
|------|-----|----------------|----------------------|
| 成员信息 | 姓名: | 单位: 厦门 | 邮箱: |
| | 帅多 | 大学王亚南 经济研究院 | shuaiduo2022@163.com |
| | 姓名: | 单位: 厦门 | 邮箱: |
| | 寇子孚 | 大学人工智 能研究院 | kouzifu@163.com |
| | 姓名: | 单位: 厦门 | 邮箱: |
| | 黄军杰 | 大学王亚南 经济研究院 | 18064422269@163.com |

签名 (可电子签名): 帅多、寇子孚、黄军杰、

基于图神经网络的超导转变温度预测

帅多¹ 寇子孚² 黄军杰¹

¹王亚南经济研究院，厦门大学，厦门，361005

²人工智能研究院，厦门大学，厦门，361005

摘要： 高温超导材料的研究因其在高效能量传输、先进电磁学和量子计算等领域的应用前景而广受关注。本报告利用图神经网络以实现基于晶体结构信息的超导转变温度的预测，在有序结构数据上提出了 Bi-SPGCN(Bi-level Superconductivity Prediction Graph Convolutional Network)，在无序结构数据上使用改进的 DeeperGATGNN (Deep Global Attention Graph Neural Network)。我们使用考虑晶体周期特殊性的图特征嵌入方法，并针对赛题数据特殊性，引入迁移学习、样本均衡处理等相关技术。最终我们在有序结构数据测试集上的 MAE 和 R2 达到了 4.59/0.95，在全数据测试集上的 MAE 和 R2 达到了 5.93/0.93，远优于基准模型 BSGNN 和其他现有模型。最后，基于共形推断方法，我们给出了模型预测结果的置信度估计。相关代码见 <https://github.com/Polaris0421/AI4SC>。

关键词： 晶体；超导转变温度；图神经网络；迁移学习

1.问题背景

高温超导材料的研究因其在高效能量传输、先进电磁学和量子计算等领域的应用前景而广受关注。近年来，大量不同结构的图神经网络被用于材料性能预测任务，并取得优异效果，其中包括超导材料的超导转变温度预测任务。与传统理论模型相比，深度学习方法拓展了可研究的晶体或化合物的范围，且表现出了更为优良的预测性能。

本赛题聚焦晶体材料，如何基于晶体最小重复单元（晶胞）的原子位置信息构建图，如何选择合适的图神经网络，是本赛题需要着重考虑的问题。同时，与由有限数量原子组成的分子或蛋白质不同，晶体的特点在于它们在无限的三维空间中周期性重复，如何在特征嵌入时考虑晶体的特殊性，也是本赛题的重要挑战。

本报告利用图神经网络以实现基于晶体结构信息的超导转变温度的预测，从有序结构、无序结构两方面分别构建神经网络模型。我们使用考虑晶体周期特殊性的图特征嵌入方法，并针对赛题数据特殊性，引入迁移学习、样本均衡处理等相关技术。我们最终在有序结构数据测试集上的 MAE 和 R2 达到了 4.59/0.95，在全数据测试集上的 MAE 和 R2 达到了 5.93/0.93，远优于基准模型 BSGNN 和其他现有模型。最后，基于共形推断方法，我们给出了模型预测结果的置信度估计。

2.有序结构 Tc 预测

2.1 问题分析

有序结构晶体的性能预测是近几年的热门方向，经典的模型如 CGCNN[1]、MEGNET[2]、GATGNN[3]、ALIGNN[4]等在不同的材料性质预测任务上已经达到了很好的效果。然而在本赛题的超导临界温度预测任务上，存在任务新、数据集小、样本不平衡等问题，直接套用过往的基准模型并不能达到满意的效果。

Liang 等人 [5]提出了基于晶胞结构和原子信息预测超导临界温度的 BSGNN 深度学习模型，BSGNN 模型利用最近邻图表示将晶体中的原子与化学键进行编码，并通过图神经网络信息传递与图注意力机制提升模型能力，并取得了不错的预测效果，最终在其设定的测试集上取得了 $R^2=0.85$ 的结果。

为了验证 BSGNN 模型效果，我们在其提供的数据集上进行了复现，并取得了相同精度的结果，可以认为在后续的实验中我们发挥了 BSGNN 模型的最优效果。然而，BSGNN 模型在赛题数据集上表现不佳，显示出明显的性能局限。

基于现有模型的不足和大量实验，本报告提出了一种用于超导材料临界温度预测的图神经网络结构 Bi-SPGCN(Bi-level Superconductivity Prediction Graph Convolutional Network)，将晶胞图结构信息和晶格大小、晶体种类等晶胞全局信息分别进行学习，再进行聚合，同时利用预训练模型提升模型的理解能力。与 BSGNN 及其他现有模型相比，Bi-SPGCN 在赛题数据集的 R^2 和 MAE 上达到了最优水平，且训练、推理效率上有很大优势。

2.2 实验设计

2.2.1 特征嵌入介绍

与 CrystalNet[6]的特征嵌入方法类似，在节点特征嵌入时我们直接指定能够反应不同节点元素的特征向量表。不同于 BSGNN 中对每个原子使用 15 维的嵌入，我们直接使用 100 维的 One-hot 向量作为每个原子节点的特征。虽然会增加计算复杂度，但是 100 维 One-hot 向量彼此之间的区分度远高于 BSGNN 的 15 维的嵌入，节省了模型学习如何区分原子的过程。对赛题数据集这样的样本量较小的数据集，简化模型的学习内容是必要的。同时，参考 DeeperGATGNN [7],我们将每个节点的入度（每个节点所连的边数目）也加入到节点特征。

与现有几何图神经网络主流做法类似[1-8]，我们使用节点之间的欧氏距离作为边的特征，并用高斯展开成多维向量，以更好的捕捉边特征之间的区别。我们使用 Radius-based 方法[8,11]来确认节点之间的连接关系。

与由有限数量原子组成的分子或蛋白质不同，晶体的特点在于它们在无限的三维空间中周期性重复[12]。为捕捉这一周期性，在建立边关系时，我们不止考虑一个晶胞内的节点，避免因切割方式不同而导致同一晶体被不同表示，如图 2 所示。对同一对原子，我们寻找它们在晶体中最近的距离作为对应边的特征。

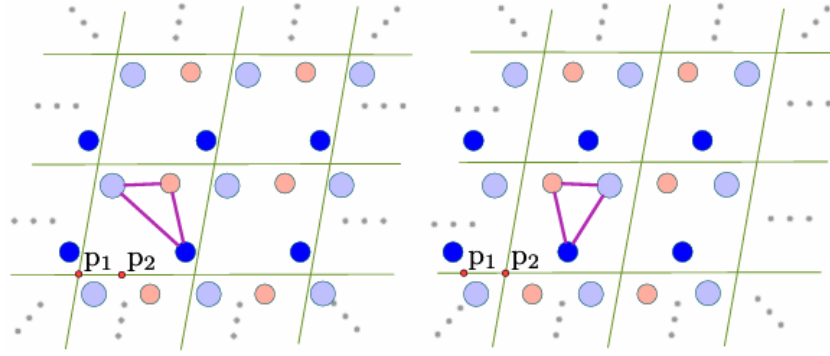


图1 不同最小重复单元切割影响展示

此外，样本中存在最小重复单元仅有一个原子的情况。对这部分样本，我们无法建模边关系以进行信息传递，会影响模型对晶体结构的学习。借鉴 Kaba 等人[14]的做法，我们将这部分样本扩充为 $2 \times 2 \times 2$ 的超级晶胞（Supercell），成为新的含有八个节点的图，作为其模型的输入。

结构信息之外我们使用全局信息，包括晶格边长，晶格夹角，晶格体积，晶格密度以及晶体所在族（H-based、Cuprate、Iron-based、Heavy-Fermion、Others）。给定晶体结构，这些信息都可以直接推断得到。一些现有文章会将全局信息直接加入到节点特征中，如将晶格边长直接作为节点自连接的特征[3]。我们认为全局信息与图结构信息属于不同层次的信息，混合在一起会降低信息传递层对结构信息的学习效果，也会影响模型对全局信息的学习，因此我们选择对两种信息分别处理，最后再拼接的操作。

2.2.2 模型结构介绍

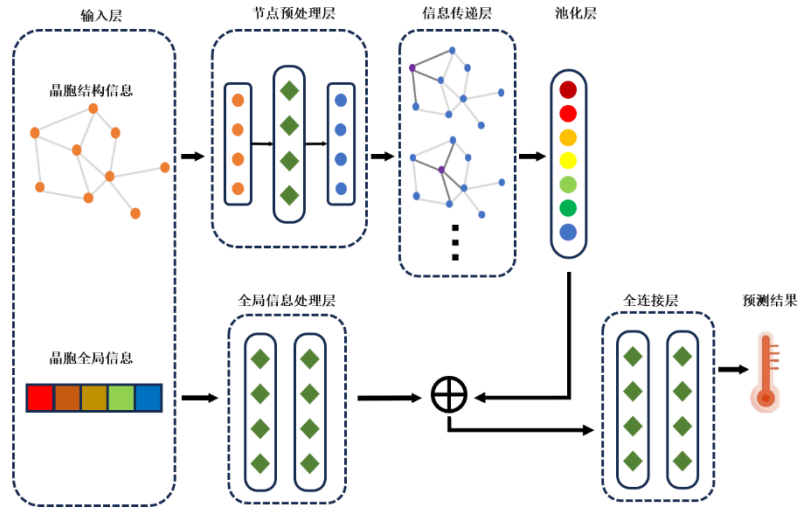


图2 Bi-SPGCN 模型结构图

图2展示了 Bi-SPGCN 模型结构。Bi-SPGCN 主要由晶胞特征提取与全局信息提取两部分组成，晶胞特征提取部分首先使用全连接层对每个节点特征进行预处理，获取节点特征向量后使用图卷积网络进行信息传递，并通过池化层获取图特征向量，与全局信息提取部分得到的全局信息特征进行拼接，最后经过全连接层获得转变温度预测结果。

我们并没有使用近来较为流行的图注意力机制来进行信息传递，因为在实验中，图注意力块并没有带来效果提升，反而会降低训练和推理的效率。

2.2.3 训练方法介绍

有序结构数据的 T_c 转变温度分布展现出严重不均衡，87.20%的数据 T_c 集中在 0-10，而使用不平衡数据集会导致模型学习结果有偏，难以学习到少数样本的特征。为了平衡数据分布，我们将 $T_c > 10$ 的数据重复 k 次添加至初始训练集中进行训练，增强对这部分数据的学习效果。

由于赛题数据集数据量小，我们希望利用迁移学习技术，加强模型对晶体结构的特征提取能力。在[10]中，作者利用结构化晶体数据设计了节点特征还原、节点属性预测与晶胞图数据重建的预训练任务，同时作者设计了基于掩码的对比学习模块，使得模型能够更注重晶胞的结构信息而非特定原子特征，增强了模型的泛化能力。为了提升模型对晶胞结构的理解能力，我们利用[10]中所输出的原子嵌入特征 e_p 设计了预训练特征损失。

最终，我们的损失函数为：

$$L = L_T + \alpha \cdot L_e = MAE(\hat{T}_c, T_c) + \alpha \cdot MSE(e_B, e_p),$$

其中 T_c, \hat{T}_c 分别为晶胞转变温度的真实值与预测值， e_B, e_p 分别代表 Bi-SPGCN 模型与预训练模型 CrysGNN 的原子特征， α 为预训练损失权重。

在训练时，我们使用 AdamW 优化器训练了 600 个 Epoch，Batch Size 设为 63。初始学习率为 0.002，采用 ReduceLROnPlateau 策略，即 10 个 Epoch 内损失若无明显降低，则自动降低学习率，衰减规则为每次乘以 0.9，最低学习率为 $1e-5$ 。我们将训练集中 $T_c > 10$ 的数据重复 5 次。

2.3 结果与分析

我们最终的模型参数量为 175073。我们在一张 RTX3060 6g 上进行训练，训练 100 个 Epoch 耗时约 1 分钟。作为对比，BSGNN 的参数量为 15864833，同一台电脑上每 100 个 Epoch 耗时约 6 分钟。我们的模型展现出足够的训练效率优势。

表 1 展示了迁移学习 (pt)、全局信息(gf)与样本平衡(ag)对模型效果的提升，以重复 5 次结果得到的平均值和标准差的形式展现。

表 1 消融实验结果

| Model | Train MAE | Train R2 | Val MAE | Val R2 | Test MAE | Test R2 |
|-------------|-----------|-----------|-----------|-----------|-----------|-----------|
| no ag pt gf | 1.25±0.15 | 0.99±0.00 | 5.61±0.25 | 0.78±0.04 | 7.12±0.34 | 0.84±0.02 |
| no pt gf | 2.27±0.68 | 0.96±0.01 | 5.88±0.02 | 0.72±0.04 | 6.67±0.25 | 0.82±0.02 |
| no pt | 1.73±0.88 | 0.98±0.02 | 5.76±0.41 | 0.81±0.04 | 5.90±0.70 | 0.90±0.03 |
| with all | 1.55±0.40 | 0.99±0.01 | 5.52±0.29 | 0.83±0.03 | 4.89±0.34 | 0.95±0.02 |

为了验证我们模型的效果，我们复现了 BSGNN 模型，表 2 展示 5 次重复试验的 MAE 与 R2 指标对比，相较于 BSGNN 模型，我们的模型展现出明显优势。

表 2 有序数据 Bi-SPGCN 与 BSGNN 效果对比

| Model | Train MAE | Train R2 | Val MAE | Val R2 | Test MAE | Test R2 |
|----------|-----------|-----------|-----------|-----------|-----------|-----------|
| BSGNN | 3.96±0.45 | 0.90±0.03 | 6.33±0.29 | 0.68±0.02 | 5.91±0.33 | 0.87±0.02 |
| Bi-SPGCN | 1.55±0.40 | 0.99±0.01 | 5.52±0.29 | 0.83±0.03 | 4.89±0.34 | 0.95±0.02 |

我们选用 5 次实验中验证集上效果最好的一次实验结果进行效果展示，该次试验于验证集、测试集上的 MAE 与 R2 指标分别为 5.42/0.84 与 4.59/0.95。我们将测试集上 Bi-SPGCN、BSGNN 的预测残差绘制比较，图 4 中前半轴展示了 Bi-SPGCN 预测效果更好时，BSGNN 预测中最大的十个残差以及对应的 Bi-SPGCN 的残差，后半轴展示了 BSGNN 预测效果更好时，Bi-SPGCN 预测结果中最大的十个残差以及对应的 BSGNN 的残差。可以看到，相较于 BSGNN 模型，Bi-SPGCN 预测结果残差更加稳定，说明 Bi-SPGCN 在预测能力上优于 BSGNN。

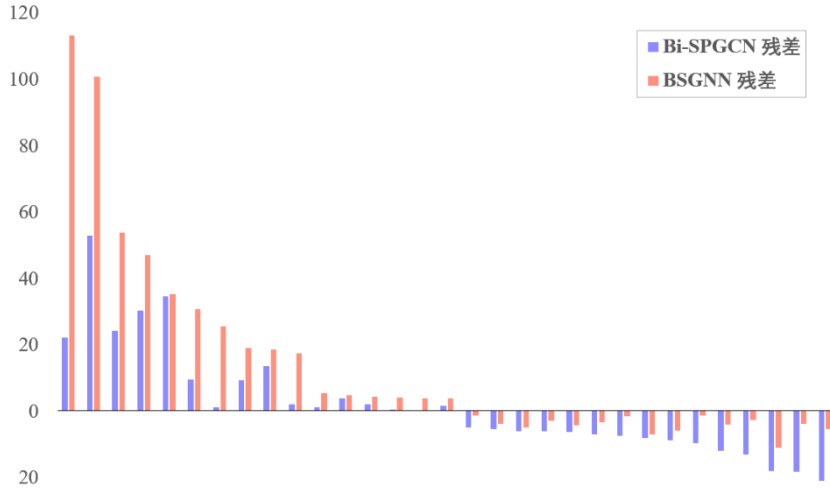


图 3 Bi-SPGCN 与 BSGNN 残差结果对比图

3.无序结构 Tc 预测

3.1 问题分析

无序结构数据分布更复杂、数据量更小，且在现有研究也有限。无序结构主要分为位置无序和替代无序，由于无序结构的存在，晶体不能像有序结构一样方便分割得到最小重复单元。为了得到最小重复单元的简便表示，无序结构数据在每个位置引入通过采样计算得到占有率(Occupancy)。

我们将有序结构和无序结构数据混合，在一些模型上进行了训练尝试。我们发现，基于全量数据集训练的模型并不能在无序结构数据上取得满意的预测效果，且由于有序结构数据更好学习且数据量更大，混合训练会导致在指标平均结果上掩盖模型对无序结构的学习不足。因此针对无序结构数据拟合困难的问题，我们决定将有序结构数据与无序结构数据分别单独进行训练，以提升预测效果。

3.2 实验设计

3.2.1 模型选择

经过实验，我们最终选择使用 DeeperGATGNN [7]模型作为无序数据预测框架，其模型框架如图 5 所示。DeeperGATGNN 使用强化图注意力层（Augmented GAT Layer）来进行信息传递。强化图注意力层是一种改良的图注意力层[3]，改良方式是在节点特征中拼接上该节点所连接的边的特征，以增强图注意力层的学习能力。

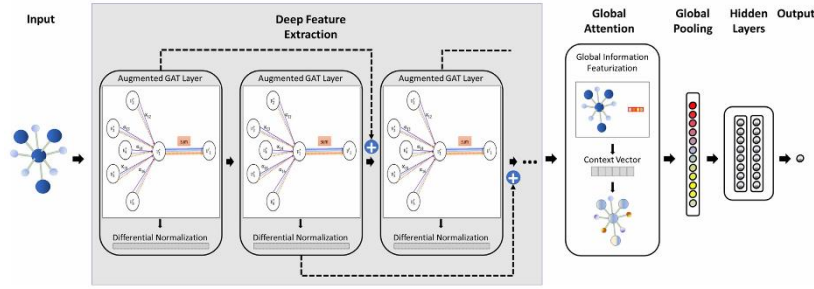


图 4 DeeperGATGNN 模型结构图[7]

DeeperGATGNN 认为更深的网络能带来更好的学习能力。它使用可微分归一化（**Differentiable Group Normalization**）和跳跃链接，以达到增加所使用的强化图注意力层层数的同时，避免梯度消失等问题。

在完成信息传递之后，DeeperGATGNN 将全局特征（各个元素的占比向量）拼接 to 各个节点的特征里，使用一个晶胞所有拼接后的节点信息经过全连接层处理得到包含晶胞结构信息的上下文向量（Context Vector），利用该上下文向量对所有未拼接的节点信息进行调整，然后通过池化和全连接层得到最终的预测结果。

考虑到无序结构和数据集的特殊性，我们的改进方式是将上文提到的晶格边长、晶格夹角、晶格体积、晶格密度、晶格所在族组成的全局信息加入到 DeeperGATGNN 的全局特征里，同时额外添加，有序比例、位置无序比例、替代无序比例作为表示晶体无序程度的三维向量。

对位置无序节点的特征嵌入，我们与[6]的表示方法一致，见下式：

$$W_{disorder} = \sum_{i=1} x_i W_{Z_i},$$

其中 x_i 是第 i 个元素的占有率， W_{Z_i} 是第 i 个原子的特征表示。该表示方法也是现有文献中最主流的表示无序结构原子表示方法。其他的特征嵌入与有序数据上的处理保持一致。

3.2.2 训练细节

在无序结构数据上，我们依然使用调整训练集的方法，将训练集中 $T_c > 17$ 的数据重复 2 次。现有的用于迁移学习的预训练模型并不支持无序结构数据，因此我们这里并不使用迁移学习。

训练时我们使用 AdamW 优化器训练了 500 个 Epoch，Batch Size 设为 31。初始学习率为 0.05，采用 ReduceLR-OnPlateau 策略，10 个 Epoch 内损失若无明显降低，则自动降低学习率，衰减规则为每次乘以 0.8，最低学习率为 $1e-5$ 。

3.3 结果与分析

表 3 展示了无序数据上添加了全局信息（gf）的 DeeperGATGNN 与 BSGNN 效果对比。DeeperGATGNN 在训练集上展现了明显的拟合优势，并且能泛化到验证集上，而 BSGNN 在验证集上效果很差。意外的是，在测试集上 DeeperGATGNN 的效果反而略差于 BSGNN。由于数据量小，模型预测效果的波动和泛化能力的受限不可避免，我们最终还是根据验证集的效果，选择添加了全局信息的 DeeperGATGNN 作为我们这部分的最终模型。

表 3 无序数据改进后的 DeeperGATGNN 与 BSGNN 效果对比

| Model | Train MAE | Train R2 | Val MAE | Val R2 | Test MAE | Test R2 |
|--------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| BSGNN | 6.01 ± 0.21 | 0.87 ± 0.01 | 8.42 ± 0.20 | 0.44 ± 0.03 | 7.77 ± 0.19 | 0.87 ± 0.01 |
| DeeperGAT GNN with gf | 1.96 ± 1.05 | 0.97 ± 0.02 | 5.49 ± 0.29 | 0.70 ± 0.05 | 8.38 ± 0.60 | 0.82 ± 0.03 |

对于 deepGATGNN，我们选用五次实验中验证集上效果最好的一次实验结果作为进行效果展示，该次试验于验证集、测试集上的 MAE 与 R2 指标分别为 5.57/0.72 与 8.54/0.81。

最终，我们将有序结构数据上的最优模型与无序结构数据上的最优模型放在一起使用。我们在全数据测试集上的 MAE 与 R2 指标分别为 5.93 与 0.93。

4. Tc 值置信度计算

4.1 问题分析

在机器学习和深度学习模型中，由于其模型的复杂程度高和灵活性强的特性，模型的输出结果相比参数模型往往受训练数据集的影响更大。因此，如何在有限数据下衡量模型的稳健性和模型输出结果的可信程度就成为了一大难题。

在分类问题中，我们可以较直观地定义分类结果的置信度[13]，但在回归任务上，如何得到回归结果的置信度始终没有一个较为直观的结论。一个较为直接的做法是利用 Bootstrap 等重抽样技术，计算多次拟合下的同一训练样本的拟合值结果，用拟合值结果对真实值的均方误差进行模型置信度的度量。

上述方式能很好地体现模型预测结果的可信程度，但当模型本身较为复杂，训练耗费算力较多时，该方法往往很难有实际应用。

本报告在对模型置信度进行估计时，使用了统计学近年来发展较为迅速的共形推断方法，该方法相比上述重抽样方法在时间和算力上更为高效，且在一些条件下可以达到理论上的置信度保证。

4.2 实验设计

在回归任务的置信度定义上，由于目标值连续分布的特点，我们无法像分类任务一样通过准确率等指标去构建置信度的估计值。因此此处我们考虑利用近邻区域概率来描述模型输出结果的置信度。在给定容忍度水平 δ 下，我们定义回归任务结果的置信度为

$$P(|\hat{y} - y| \leq \delta),$$

即在给定的容忍度下，因变量的预测值落在真实值容忍邻域内的概率，若该概率的取值越高，则说明模型对其预测结果的自信心越强，表现为对同一目标的预测值标准差较小。

在共形推断中，我们需要找到关于某一样本预测值的区间估计满足：

$$P(y \in \hat{C}_n(x)) \geq 1 - \alpha),$$

这一目标往往通过校准集上的残差分位数进行估计实现。不论是在完全有序晶体的预

测任务，还是混合样本的预测任务，由于每个晶体隶属的晶体族不同，或者是晶体存在有序无序的差别，这导致了晶体的生成分布可能存在差异。因此我们考虑利用权重对分位数进行校准得到更为合理的置信度结果。

在得到校准后的经验分布后，我们求得置信度为

$$Conf = \min_p \{p: Q(p, \tilde{F}) \geq \delta\},$$

其中 Q 为分位数函数， \tilde{F} 为经过校准后的经验分布函数。

4.3 结果与分析

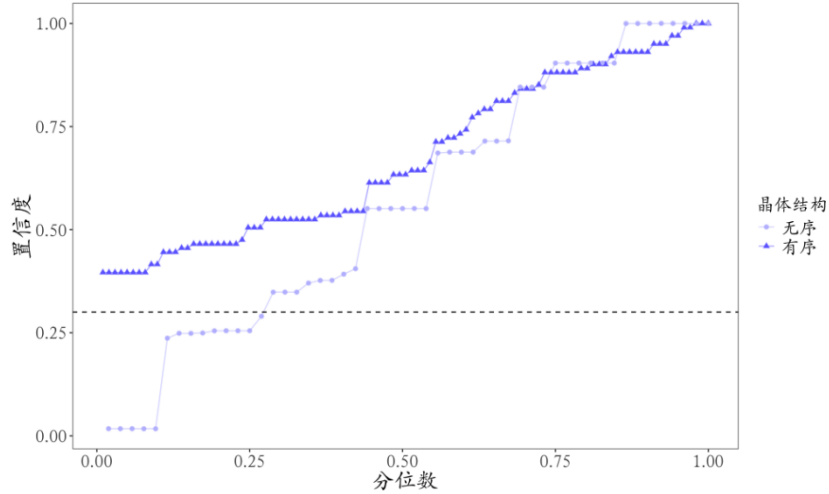


图 5 置信度结果图

上方图像展示了在有序和无序晶体预测任务下模型给出的置信度水平，图中虚线指代的是 0.3 的水平线。在设定容忍度为 1 时，可以看到，在有序情况下，所有测试集预测值的置信度均在 0.3 以上，模型整体平均置信度为 0.67。在无序晶体的预测任务中，整体平均可靠性为 0.55。当容忍度提升为 5 时，有序情况下的平均置信度提升到了 0.85，无序情况下提升到了 0.65。

5. 总结

本报告实现了通过图神经网络对超导材料的超导转变温度的预测。对有序结构数据，我们提出了一种高效的图神经网络结构 Bi-SPGCN；对无序结构数据，我们使用优化后的 DeeperGATGNN。在有序结构数据和无序结构数据上我们均取得了远好于基准模型 BSGNN 的效果。最后，基于共形推断方法，我们给出了模型预测结果的置信度估计。

我们的工作有以下创新点：

一、特征嵌入上，我们查阅大量文献，综合现有方法、晶体特性和赛题数据特殊性，最终形成所使用的特征嵌入方法，得到的边特征不随晶胞周期和最小重复单元切割方法变化。此外，我们考虑到了晶格边长、晶格夹角、晶格体积、晶格密度以及晶体所在族等全局信息。

二、模型上，我们提出了 Bi-SPGCN 来实现有序数据的预测，能够考虑晶胞的全局信息，并能够将不同层次的全局信息与结构信息分别处理，且具有高效的训练与推理过程；在无序结构数据上，我们将全局信息加入到原始的 DeeperGATGNN，达到了更好的效果。

三、训练方法上，针对赛题数据数据量小、样本不平衡等问题，我们引入前沿文献的

迁移学习方法，并采用训练样本调整、超级晶胞扩充、有序无序数据分离等训练方法，有效地增强了训练效果。

四、在置信度上，我们利用了容忍度作为连续值的识别边界限制，定义了类似于分类任务中的置信度指标。我们提出的模型在给定容忍度下的置信度表现较好，说明了我们模型预测结果的可信程度较高。

参考文献

- [1] Xie T, Grossman J C. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties[J]. Physical review letters, 2018, 120(14): 145301.
- [2] Chen C, Ye W, Zuo Y, et al. Graph networks as a universal machine learning framework for molecules and crystals[J]. Chemistry of Materials, 2019, 31(9): 3564-3572.
- [3] Louis S Y, Zhao Y, Nasiri A, et al. Graph convolutional neural networks with global attention for improved materials property prediction[J]. Physical Chemistry Chemical Physics, 2020, 22(32): 18141-18148.
- [4] Choudhary K, DeCost B. Atomistic line graph neural network for improved materials property predictions[J]. npj Computational Materials, 2021, 7(1): 185 .
- [5] Gu L, Liu Y, Chen P, et al. Predicting Transition Temperature of Superconductors with Graph Neural Networks[J]. arXiv preprint arXiv:2308.11160, 2023.
- [6] Chen C, Zuo Y, Ye W, et al. Learning properties of ordered and disordered materials from multi-fidelity data[J]. Nature Computational Science, 2021, 1(1): 46-53.
- [7] Omee S S, Louis S Y, Fu N, et al. Scalable deeper graph neural networks for high-performance materials property prediction[J]. Patterns, 2022, 3(5).
- [8] Yan K, Liu Y, Lin Y, et al. Periodic graph transformers for crystal material property prediction[J]. Advances in Neural Information Processing Systems, 2022, 35: 15066-15080 .
- [9] Sommer T, Willa R, Schmalian J, et al. 3DSC-a dataset of superconductors including crystal structures[J]. Scientific Data, 2023, 10(1): 816.
- [10] Das K, Samanta B, Goyal P, et al. Crysgnn: Distilling pre-trained knowledge to enhance property prediction for crystalline materials[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2023, 37(6): 7323-7331 .
- [11] Ying C, Cai T, Luo S, et al. Do transformers really perform badly for graph representation?[J]. Advances in neural information processing systems, 2021, 34: 28877-28888.
- [12] Han J, Cen J, Wu L, et al. A Survey of Geometric Graph Neural Networks: Data Structures, Models and Applications[J]. arXiv preprint arXiv:2403.00485, 2024.
- [13] Guo C, Pleiss G, Sun Y, et al. On calibration of modern neural networks[C]//International conference on machine learning. PMLR, 2017: 1321-1330.
- [14] Kaba O, Ravanbakhsh S. Equivariant networks for crystal structures[J]. Advances in Neural Information Processing Systems, 2022, 35: 4150-4164.