二、项目研究内容、研究方法及技术路线

（一）项目的主要研究内容

（限1500字以内）

单晶高温合金的发展趋势有两方面：提高合金的使用温度和延长使用寿命，其中改进凝固条件和热处理工艺是两条重要的途径。本团队引入机器学习方法，分析影响单晶高温合金性能的各种因素及其关系，构建性能预测模型，从而提高性能预测的精度和效率，促进单晶高温合金的制备。

1.1 基于特征选择的单晶高温合金性能影响因素分析

由于单晶高温制合金备过程中的复杂性，数据中往往存在大量稀疏、不相关、冗余等特征，特征选择的方法能够确定对单晶高温合金性能影响较大的因素，从而减少制备过程的复杂性，节约人力和财力。

对单晶高温合金数据，本团队拟使用双轨并行的多层级过滤式特征选择方法和强关系关联规则方法选择合适的特征。综合利用单晶高温合金的高通量计算数据和实验数据，分析属性值之间的关联关系、属性与单晶高温合金性能的关系，寻找在单晶高温合金制作过程中影响性能较大的属性，进一步缩短实验的时间和复杂性。

1.2 基于集成学习的单晶高温合金性能预测方法研究

在单晶高温合金的研究中，现有的性能预测方法主要使用单一模型，泛化能力不强。集成学习是集成学习方法是指通过组合多个模型，以获得更好的效果，使集成的模型具有更强的泛化能力。每个算法有不同的适用范围，例如有处理线性可分问题的，有处理线性不可分问题。

本课题拟采用集成学习方法，“取长补短”，发挥各个算法的优势应对不同的数据类型。对未知的非线性关系的逼近，找到合金制备过程中实验控制参数等因素影响下，成分、组织及性能的变化，对其变化进行曲线拟合，建立从成分到性能的预测模型以及高通量计算数据和反应实验数据内在规律的模型，从而对单晶高温合金制备过程的参数控制进行优化，以达到指导试验的目的。

1.3单晶高温合金性能预测模型的可解释性研究

（二）项目拟采取的研究方法

1、项目研究拟解决的问题，及拟采用的方法、原理、机理、算法、模型等

（研究方法、技术路线可用图表描述。限800字以内）

（1）基于主动学习的交互式多层级特征选择方法

多层级过滤式特征选择方法主要是针对材料属性数据原始特征集中存在的稀疏性、不相关性、冗余性问题提出的，其思想是采用分层级方式进行特征过滤，同时专家的特征表校对，确保最终的特征子集中条件属性与决策属性具有较强的相关性，条件属性之间关联较弱。该方法主要分为如图2所示的四个层次。

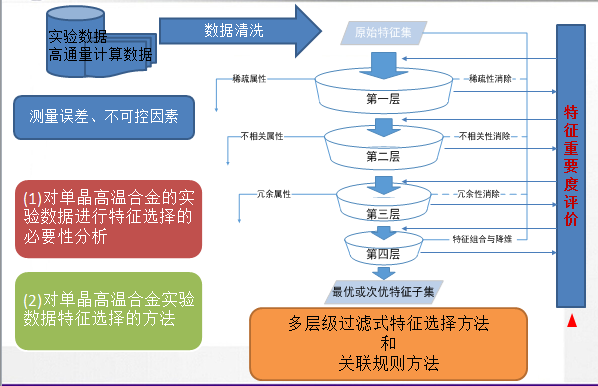


图2 多层级过滤式特征选择方法流程

该方法是按照特征集的流向依次进行特征选择的，综合了封装式特征选择和过滤式特征选择的思想。横向上，对于每一层的筛选结果使用验证模型进行评估，增加了特征选择结果的有效性；纵向上，每一层特征选择的结果作为下一层输入，各层次的执行过程是相互独立的，减小了对验证模型的依赖性。在此过程前，专家将构建的材料属性特征集合，每一层特征选择结合模型验证（）和专家的集合校对进行过滤和阈值调整，获取当前最优特征子集。

（2）基于集成学习的自适应混合式性能预测模型

2.1 单晶高温合金的性能预测模型构建方法

在分析、预测材料性能中，影响单晶高温合金材料的组织和性能的因素多而且复杂，基于集成学习方法建立的模型能从己有的实验数据中自动归纳出规律,并可以利用经训练好模型直接进行推理，用于对单晶高温合金材料组织结构的设计和性能预测。

本团队拟利用集成学习的思想，借助目标最优的选择性集成策略完成对单晶高温合金性能预测方法的建模。单一模型选择支持向量机、神经网络等机器学习算法。通过多个机器学习算法共同对合金性能决策。

2.2 单晶高温合金的性能预测模型选择策略

在机器学习领域已证明基于集成学习的预测优于单模型预测。本团队拟用基于目标最优的选择性集成策略对多个模型的组合优化，其主要思想是通过问题驱动自动选择最适合数据分布的模型，利用集成策略将模型合金数据的性能预测。具体过程如下图，其关键技术是模型的选择策略和模型的集成策略。在模型选择策略方面，本团队拟使用在验证数据集上找到表现最好的模型作为最终的预测模型；在模型集成方面，设置选择函数，判断是否需要对模型集成，当满足集成条件后，通过对择优选择出的模型使用bagging或boosting方式集成，对预测结果进行投票或者取平均值和对多个模型的预测结果做加权平均。

实验数据

预测结果

图2 基于集成学习的性能预测

2、项目研究方法（技术路线）的可行性、先进性分析

（限500字以内）

材料属性数据集中往往存在大量稀疏、不相关、冗余等特征，这些属性会大大降低机器学习模型的预测精度和泛化能力，同时还会增加模型的训练时间。由于单晶高温制合金备过程中的复杂性，通过实验很难获达到目的。本团队通过对基于主动学习的交互式多层级特征选择方法和基于集成学习的自适应混合式性能预测模型的研究，我们的方法能够有效的改善实验的效率与精度。本团队使用双轨并行的多层级过滤式特征选择方法，综合利用单晶高温合金的高通量计算数据和实验数据，分寻找在单晶高温合金制作过程中影响性能较大的属性；采用集成学习方法，“取长补短”，发挥各个算法的优势应对不同的数据类型。本团队的方法有着良好的泛化能力，并且有良好的可解释性。