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。

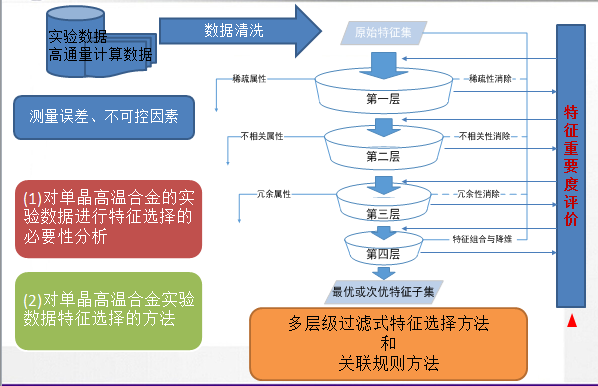


图2 多层级过滤式特征选择方法流程

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

本团队拟采用集成学习的方法，集成的模型拟选择支持向量机、神经网络。贝叶斯等模型，已证明支持向量机一定可以将两类问题区分，而神经网络在这几年的各个应用领域都有着显著的成果和超高的准确率，并且有人已经将神经网络用到了材料性能预测方面。

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。本团队拟用三种集成策略对多个模型的组合优化，分别是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、对多个模型的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。以AdaBoost算法为例，首先需要提取Haar特征；然后将Haar特征转化成对应的弱分类器；最后从大量的弱分类器中迭代选择出最优弱分类器；具体过程如下图1所示：

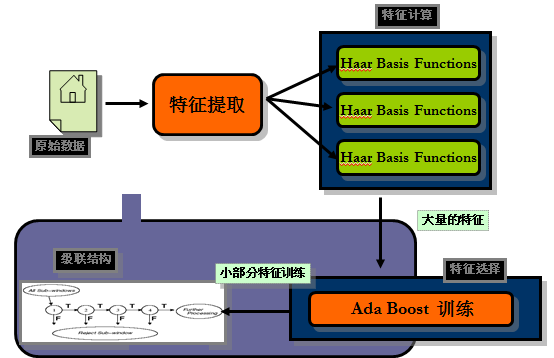


图1：AdaBoost算法的流程

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。