随机模糊粒度决策树

一.成果完成人及所在单位： 马啸宇，计算机与信息工程学院 指导教师：李伟

二.成果简介：

以决策树算法为研究对象，将其与粒计算相结合，通过将分类问题等价地转化为模糊粒度空间来求解，在自适应全局随机聚类（AGRC）算法的基础上，设计了一个并行模型，在模糊粒度空间中，通过设计 RFGDT 来对模糊粒度进行分类，避免了基于修剪算法的过度拟合问题，提高了分类算法的计算效率与准确性。

三. 主要研究内容及创新点：

在本研究中，从粒度计算的角度解决了分类问题。即将分类问题等价地转化为模糊粒度空间来求解。大多数分类算法仅用于处理数字数据；随机模糊粒度决策树（RFGDT）不仅可以处理数值数据，而且可以处理信息粒度等非数值数据。可通过以下四种方式采取措施。首先，提出了一种自适应全局随机聚类（AGRC）算法，该算法能够自适应地找到最优聚类中心，最大化类间标准差与类内标准差之比，避免陷入局部最优解；其次，在 AGRC 的基础上，设计了一个并行模型，用于数据的模糊粒度化，构造粒度空间，与串行粒度化相比，大大提高了效率；第三，在模糊粒度空间中，我们设计了 RFGDT 来对模糊粒度进行分类，它可以根据信息增益比选择重要特征作为树节点，避免了基于所提出的修剪算法的过度拟合问题。最后，我们使用 UC Irvine 机器学习库中的数据集进行验证。理论和实验结果证明，RFGDT 具有较高的效率和准确性，并且在解决分类问题时具有鲁棒性。

创新点：

（1）我们提出了 AGRC，它可以自适应地给出最优的聚类中心，这是一种全局优化方法，可以避免陷入局部优化解。

（2）我们基于上述聚类算法设计了并行粒度方法，解决了传统串行粒度的高复杂性问题，提高了粒度效率。

（3）在粒度空间中，我们定义了模糊粒度和相关算子，并根据信息增益比选择特征，构造了一个用于分类的模糊粒度决策树。为了避免过度学习，我们还设计了相应的修剪算法。该方法可以解决二值分类或多值分类问题，并根据生成树节点的顺序给出特征的重要性。

四. 主要研究成果及应用情况

在 SCI 4 区发表论文一篇。

联系人：马啸宇， 电 话：15603650203 Email:1171205977@qq.com