

题目：基于机器学习的流量数据分析技术研究

一、研究内容

1. 流量数据的收集和预处理。
2. 协议的工作原理及网络数据包数据结构。
3. 各类机器学习算法的工作原理、特性。
4. 特征选择以及机器学习模型的优化。
5. 提出一种改进的分类方法，可以较好地分类网络流量。

二、拟研究解决的关键问题

1. 流量数据的特征选择。
2. 机器学习算法的优化和改进。

三、拟采取的研究方案

1. 首先收集流量数据，从数据中获取全部可用特征，并对数据进行预处理，包括去除无效数据、冗余数据以及数据缺失值处理等步骤。
2. 数据处理过后，首先对数据样本的组成和分布进行初步分析，并确定合理的评价指标，用于评价后续的实验结果，之后对各类机器学习算法的特点进行分析，并进行实验对比，选择合适的算法。
3. 确定算法后，进行数据的特征选择工作，利用python实现不同的特征选择方法，并分别用于特征筛选，对不同方法的筛选结果进行对比，并根据先前确定的指标，选择最优的特征集合。
4. 选定使用的数据特征集合后，进行算法模型的参数调优工作，首先在较大的参数范围内进行粗略搜索，之后再缩小搜索范围，并缩小搜索步长，最后基于交叉验证的方法，得到最优的算法模型。
5. 通过对数据的分析，并结合算法自身特点，提出一种改进的方法，从而达成更良好的数据分类效果。