**doi:**10.1631/FITEE.1601761

**题目：**一种用于文本分类的去冗余特征选择新方法

**概要：**特征选择是文本分类领域一种重要降维方法。针对传统特征选择方法所选特征集常包含冗余信息的问题，提出一种能够有效去除冗余信息的特征选择新方法。首先，为衡量两个词之间的关系，引入基于词频的相关性和相对冗余词集的概念；接着，选择一种最优特征选择方法并用其获得一个临时特征子集；最后，为提高算法执行效率，结合预设阈值去除临时特征子集中的冗余特征，并将结果存储在链表结构中。实验以支持向量机和朴素贝叶斯作为分类器，并以 WebKB、20-Newsgroups 和 Reuters-21578 作为测试数据集。实验结果表明，该方法分类精度高于传统特征选择方法；相对于基于互信息的方法而言，该方法能够在保证分类精度的同时，有效提高运行效率。

**关键词：**特征选择；降维；文本分类；冗余特征；支持向量机；朴素贝叶斯；互信息