doi:10.1631/FITEE.1601322

题目：一种面向软件缺陷预测的相似性度量特征选择方法

概要：软件缺陷预测旨在通过历史数据和能反映软件模块特性的软件特征来发现潜在缺陷。然而，有的特征可能与类别（有缺陷或无缺陷）的相关性较高，有的特征可能是冗余的或无关的。针对软件缺陷预测中不同特征与类别的相关性差异，本文提出一种基于相似性度量（similarity measure, SM）的特征选择方法。首先，根据不同类样本间的相似性来更新特征权重；然后，按照特征权重值降序排列生成特征排序列表，并依次选取特征排序列表中的所有特征子集；最后，在 KNN（*k*-nearest neighbor）模型上验证所有特征子集的分类性能，并采用 AUC （area under curve）指标进行度量。在 11 个美国航空航天局（NASA）数据集上进行实验验证，结果表明，与其它四种特征选择方法相比，本文方法具有与之相当甚至更高的分类性能。

关键词：软件缺陷预测；特征选择；相似性度量；特征权重；特征排序列表