**doi:**10.1631/FITEE.1500283

**题目：**一种观点挖掘新词语权重过程性能分析

**概要：**论坛和博客的普及为大量信息的处理带来了挑战和机遇。基于不同主题的信息通常包含了主观的定性词语，需要经过统计分析转换为可用的定量数据。这些数据如不恰当处理则会影响观点的正确表达。每个观点相关词的主要表义各有不同。为将词的语义转换为数据并加强对观点挖掘的分析，我们提出了一种新颖的加权方案，称为词权重推测法（inferred word weighting, IWW）。IWW 通过对语境下和表义中词语重要性的计算对算法进行增强。相对已有的方法，本文提出的加权方法从分析的视角上为词语提供了合适的权重。此外，通过对包含停用词的文本分类的性能研究，提供了另一种校验方法，作为对所提出的新加权方法的补充。而通常这些停用词都会在文本处理时移除。将包含停用词这一新概念应用于本文提出的加权方法和已有加权方法，可观察到 2 个现象：（1）文本分类性能增强；（2）包含停用词与否，所造成的文本处理结果的差异在所提出的方法中较小，而在已有方法中较大。进而，从这 2 种现象得出推论。基于基准数据集的实验结果表明所提出的方法在分类精度上具有优化潜力。

**关键词：**词权重推测法；观点挖掘；监督分类法；支持向量机；机器学习