基于粗糙集与信息增益的情感特征选择方法

蒲国林

(四川文理学院 计算机学院,四川 达州 635000)

摘 要:为了提高情感特征提取的准确率,为高性能情感分析打下坚实的基础,提出了一种融合粗糙集与信息增益的情感特征选择方法.该方法借助信息增益判据选出高相关性的特征子集,再通过粗糙集剔除高冗余性的特征,从而得到最优的特征子集.在多个数据集上的测试表明,该方法可将若干经典方法的准确率提高 4~9 个百分点,是一种优秀的特征选择方法,对提升情感分析的整体性能有明显意义.

关键词:情感分析;特征选择;粗糙集;信息增益

A Sentiment Feature Selection Method Based on Rough Set and Information Gain

PU Guo-lin

(School of Computer Science, Sichuan University of Arts and Science, Dazhou 635000, China)

Abstract: A Rough Set and Information Gain based on sentiment feature selection method is proposed for building a solid foundation in sentiment analysis. The novel method firstly uses Information Gain to select a feature subset which has high relativity with the class attribute. Secondly, the features which have high redundancy will be eliminated by Rough Set. Experimental results on several datasets reveal the method makes accuracy increase 4-9 percentages than other methods. It is an outstanding feature selection method and has significance in sentiment analysis.

Key words: sentiment analysis; feature selection; Rough Set; Information Gain

作者简介:

蒲国林 男,(1971-),博士,副教授.研究方向为人工智能.
E-mail: puguolsc@126.com