计算机工程与应用 2009 45 (16): 138-140 ISSN: 1002-8331 CN: 11-2127/TP

数据库、信息处理

粗糙集与决策树在电子邮件分类与过滤中的应用

邓春燕[1,3], 陶多秀[2], 吕跃进[3]
1.广西河池学院 计算机与信息科学系，广西 宜州 546300
2.广西大学 电气工程学院，南宁 530004
3.广西大学 数学与信息科学学院，南宁 530004

收稿日期 2008-12-15 修回日期 2009-2-25 网络版发布日期 2009-5-27 接受日期

摘要 垃圾邮件的识别与过滤是目前研究的热点问题之一。而粗糙集是一种新的处理模糊和不确定性知识的数据分析工具，已被成功地应用到许多有关分类的领域。将粗糙集与决策树结合，提出一个基于RS-DT的邮件分类方案与模型，并进行了实验及结果分析。通过与朴素贝叶斯模型及SVM的比较，表明提出的基于RS-DT的模型可以降低把正常邮件错分为垃圾邮件的比率，提高过滤系统的自学习能力。

关键词 垃圾邮件 粗糙集 数据挖掘 决策树
分类号

Application of rough set and decision tree in e-mail classification and filtering

DENG Chun-yan[1,3],TAO Duo-xiu[2],LV Yue-jin[3]

1．Department of Computer and Information Science，Hechi University，Yizhou，Guangxi 546300，China
2.College of Electrical Engineering，Guangxi University，Nanning 530004，China
3.College of Mathematics and Information Science，Guangxi University，Nanning 530004，China

Abstract
Spam identification and filtering is one of the hot issues.And the rough set is a new data analysis tool to deal with ambiguity and uncertainty of knowledge，it has been successfully applied to many areas of classification.Combining rough sets with decision tree，a spam filtering solution based on rough sets and decision tree（RS-DT） was proposed.The feasibility of the solution was indicated by the experiments on the public email corpus.Comparison experiments were also made between SVM classifier，Bayes classifier and RS-DT model.The results show that the RS-DT model can not only reduce the error rate of judging the normal email as spam，but also improve adaptive learning of the filtration system.

Key words spam rough set data mining decision tree

DOI: 10.3778/j.issn.1002-8331.2009.16.040

通讯作者 邓春燕 dchy7072@126.com