

基于中心向量法与万有引力模型相结合的文本分类

王 宇, 刘 斌

(南京工业大学 计算机科学与技术学院, 江苏 南京 210000)

摘 要: 提出一种中心向量模型与万有引力模型相结合的文本分类的方法 (CGMV). 引入了类内信息熵和类间边界置信度, 解决了传统模型下类间分布不均匀与类内个别样本分布稀疏从而导致分类准确度下降的问题. 经过实验证明 CGMV 算法提高了文本分类的准确率.

关键词: CMV; 文本分类; CGMV

Text Classification Based on Central Vector Method and Universal Gravitation

WANG Yu, LIU Bin

(School of Computer Science and Technology, Nanjing Tech University, Nanjing 210000, China)

Abstract: This paper presents a text categorization method (CGMV) that combines the central vector model with the gravitation mode. The intra - class information entropy and inter-class boundary confidence degree are introduced to solve the problem that the classification accuracy is degraded due to the heterogeneity of inter - class distribution and the sparse distribution of individual samples in the traditional model. Experiments show that CGMV algorithm improves the efficiency of text categorization.

Key words: CMV; text classification ; CGMV

作者简介:

王 宇 男, (1992-), 硕士研究生. 研究方向为数据挖掘.

E-mail:15062282096@163.com.

刘 斌 男, (1966-), 博士, 硕士生导师. 研究方向为模糊技术、人工智能、数据挖掘.