一种改进CHAMELEON算法的聚类算法COCK

朱烨行[1]，李艳玲[2]，杨献文[3]
（1 西安邮电大学 经济与管理学院，陕西 西安 710121；
2 第二炮兵工程大学，陕西 西安 710025；
3 西安财经学院 信息与教育技术中心，陕西 西安 710061）

摘 要：通过对现有的CHAMELEON算法进行改进，并借鉴ROCK算法的计算步聚，提出了一个新的层次聚类算法COCK.改进之处在于：合并的簇的内部紧密性、合并的簇的内部互连性、相对紧密性和相对互连性的计算方法进行了改变，并取消了CHAMELEON算法原有两个阶段的第一个阶段.由簇U和簇V合并构成的簇W的内部紧密性由两个因素决定，一个是簇U和簇V本身的内部紧密性的加权和，另一个是簇U和簇V之间的绝对紧密性；簇W的内部互连性由两个因素决定，一个是簇U和簇V本身的内部互连性的加权和，另一个是簇U和簇V之间的绝对互连性；由簇W和簇J合起来的簇的相对紧密性的计算方法是，先求出簇W和簇J内部紧密性的加权和，用这个加权和去除簇W和簇J之间的绝对紧密性；计算两个簇W和J之间的相对互连性的方法是，把簇W和簇J之间的绝对互连性除以簇W和簇J内部互连性的加权和.

关键词：文本聚类；文档聚类；CHAMELEON；ROCK；算法
中图分类号：TP391.4 **文献标识码**：A **文章编号**：1000－7180(2015)12－0173－04

Clustering Algorithm COCK Improved from CHAMELEON Algorithm

ZHU Ye-hang[1], LI Yan-ling[2], YANG Xian-wen[3]
(1 School of Economics and Management, Xi'an University of Posts and Telecommunications, Xi'an 710061, China;
2 Department of Electronic Engineering, The Second Artillery Engineering University, Xi'an 710025, China;
3 Information and Educational Technology Center, Xi'an University of Finance and Economics, Xi'an 710061, China)

Abstract: By improving the existing algorithm CHAMELEON, and using the calculate steps of ROCK algorithm for reference, we propose a new hierarchical clustering algorithm COCK, the improvement comprise that we change the calculation method of the combined cluster internal closeness, the combined cluster internal interconnectivity, relative closeness, relative interconnectivity. And cancel the first phase of CHAMELEON original two-stage algorithm. Two factors determine internal closeness of Cluster W which is constituted of cluster U and V, one is the weighted sum of internal closeness of cluster U and V, another is absolute closeness between cluster U and V; internal interconnectivity of cluster W is determined by two factors, one is weighted sum of internal interconnectivity of cluster U and V, another is the absolute interconnectivity between cluster U and V. Calculation method of relative closeness of the cluster combined by cluster W and J is, at first calculate the weighted sum of internal closeness of cluster W and J, then divide the absolute closeness between cluster W and J by this weighted sum, calculate method of relative interconnectivity between cluster W and J is, divide the absolute interconnectivity between cluster W and J by the weighted sum of internal interconnectivity of cluster W and J.

Key words: text clustering; document clustering; CHAMELEON; ROCK; algorithm

收稿日期：2014－04－09；**修回日期**：2014－05－28
基金项目：国家自然科学基金青年项目(71102149)