基于聚类中心优化的 k-means 最佳聚类数确定方法

贾瑞玉，宋建林

(安徽大学 计算机科学与技术学院，安徽 合肥 230601)

摘　要：　k-means 聚类算法是在确定的聚类数　k　下对数据集进行聚类，通常凭借先验规则假定一个　k　值，取值具有很大主观性；此外，k-means 初始聚类中心的选择一般也是随机的，这使得 k-means 聚类经常出现聚类指标局部最优化，聚类结果不稳定.针对以上两个问题，结合密度法改进了 k-means 初始聚类中心点的选择，并在此基础上提出了一种确定 k-means 最佳聚类数的方法.实验结果证明，该方法可以得到更好的聚类结果，具有更高的准确性、更好的稳定性以及更优的收敛性.

关键词：　k-means 聚类；初始聚类中心；样本密度；聚类数

K-means Optimal Clustering Number Determination Method

Based on Clustering Center Optimization

JIA Rui-yu，SONG Jian-lin

(College of Computer Science and Technology, Anhui University, Hefei 230601, China)

Abstract：　The traditional K-means clustering algorithm requires a certain number of clustering, usually the value of clustering number is assumed by priori rules. Besides, the choice of K-means initial clustering centers is also randomized, which made the clustering result instability. In view of the above factors, improved the selection of initial centers of K-means clusters, and a new method to determine the optimal clustering number is proposed based on that. The experimental results show that the proposed method can get better clustering results and have higher accuracy, better stability and better astringency.

Key words：　words k-means clustering；initial clustering centers；density of samples；clustering number

作者简介：

贾瑞玉　女，（1965-），副教授，硕士生导师.研究方向为智能算法与数据挖掘.

宋建林（通讯作者）　　男，（1990-），硕士研究生.研究方向为数据挖掘.E-mail:435247994@qq.com.