2018届研究生博士学位论文　　学校代码: 10269

学　　号: 52131201004

華東師範大學

演化算法中分类预选择策略的研究

院　　系: 计算机科学与技术系

专　　业: 计算机软件与理论

研 究 方 向: 演化计算

指 导 教 师: 张桂戌教授，周爱民教授

博士研究生: 张晋媛

完成

2018 Doctoral Dissertation

School Code: 10269

Student Number: 52131201004

East China Normal University

THE STUDY OF PRESELECTION VIA CLASSIFICATION ON EVOLUTIONARY ALGORITHMS

Department: Department of Computer Science and Technology

Major: Computer Software and Theory

Subject: Evolutionary Computation

Supervisor: Prof. Guixu Zhang, Prof. Aimin Zhou

Author: Jinyuan Zhang

华东师范大学学位论文原创性声明

郑重声明：本人呈交的学位论文《演化算法中分类预选择策略的研究》，是在华东师范大学攻读硕士/博士（请勾选）学位期间，在导师的指导下进行的研究工作及取得的研究成果。除文中已经注明引用的内容外，本论文不包含其他个人已经发表或撰写过的研究成果。对本文的研究做出重要贡献的个人和集体，均已在文中作了明确说明并表示谢意。

作者签名: _____ **日期:** 年 月 日

华东师范大学学位论文著作权使用声明

《演化算法中分类预选择策略的研究》系本人在华东师范大学攻读学位期间在导师指导下完成的硕士/博士（请勾选）学位论文，本论文的研究成果归华东师范大学所有。本人同意华东师范大学根据相关规定保留和使用此学位论文，并向主管部门和相关机构如国家图书馆、中信所和“知网”送交学位论文的印刷版和电子版；允许学位论文进入华东师范大学图书馆及数据库被查阅、借阅；同意学校将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于（请勾选）

（ ）1.经华东师范大学相关部门审查核定的“内部”或“涉密”学位论文*，于 年 月 日解密，解密后适用上述授权。

（ ）2.不保密，适用上述授权。

导师签名 _____ 本人签名 _____

年 月 日

* “涉密”学位论文应是已经华东师范大学学位评定委员会办公室或保密委员会审定过的学位论文（需附获批的《华东师范大学研究生申请学位论文“涉密”审批表》方为有效），未经上述部门审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权）。

张晋媛 博士学位论文答辩委员会成员名单

姓名	职称	单位	备注
			主席

摘　　要

聚类分析作为一种非监督学习方法，已经广泛的应用到机器学习、数据挖掘、人工智能和图像处理等领域，并成为该领域的研究热点。随着聚类分析的备受关注，存在的一些问题需要研究者去解决，这些问题包括如何自动确定聚类的数目，如何获得全局最优解，如何对任意形状的数据集合进行聚类，如何将多种比较成熟的聚类算法有机的结合在一起，以及如何将聚类算法应用到如数字图像处理等特定领域中。本文针对这几个问题进行探讨，主要包括以下几个方面：

1) 在基于超球形数据的聚类上，本文提出了一种基于进化算法的自适应聚类方法，该方法的基本思想是将聚类问题转化成一个全局优化问题，然后使用进化算法对其优化。其中包括一个可以平衡类内紧凑度和类间分离性的有效性指标函数作为自适应进化聚类的目标函数。在编码方式上，设计了三种可以表示不确定聚类数目的编码方法。结合提出的编码方式和提出的有效性指标自动确定最佳聚类数目。最终经过大量的实验以及与其它算法对比可知，提出的方法不仅可以自动确定最佳聚类数目，并且得到了比较稳定的聚类结果。

2) 在任意形状数据的聚类上，本文提出了四种自动确定任意形状数据最佳聚类数目的方法。一、由于Path距离度量非常适合对任意形状的数据聚类，但较高的时间复杂度制约了其应用范围，设计了两种可以缩短Path距离聚类的计算方法：(1)使用改进的Floyd算法降低Path距离的计算时间复杂度，(2)使用一种预聚类算法降低问题的规模来降低Path距离聚类的时间复杂度，并且还设计了一种Box聚类算法可以自动的确定最佳聚类数目；二、提出了一个基于进化算法的任意形状数据自适应聚类方法。在该算法中，使用一种基于标签的编码方式对搜索空间编码，并且设计了一种改进的Modularity指标作为目标函数用来自动检测最佳聚类数目；三、由于基于标签的编码会导致染色体的长度过大，提出了一种基于实数的编码方式和一种基于距离矩阵的有效性指标对任意形状数据集进行聚类；四、前面提出的三种算法都是基于单一算法的聚类算法，方法四提出了一种进化多目标算法框架用来将不同的聚类算法有机的结合在一起。最终将提出的算法对大量的任意形状数据做聚类可得，提出的算法得到了比较好的结果。

3) 过分割问题和欠分割问题是图像分割中常出现的问题，在图像分割中，自动的确定最佳图像分割数目也是一个比较重要的问题。本文提出了一种结合Mean-shift算法和自适应进化聚类算法的自适应图像分割算法。首先使用Mean-shift算法对图像进行预处理，得到一幅过分割图像，然后使用自适应进化聚类算法对该过分割图像精确分割，得到最

佳分割数目。该算法不仅可以降低图像分割的时间复杂度，还可以保存图像目标间的连通性。通过与其它算法对比可知，提出的方法可以有效地对图像进行分割。

4) 大量的图像分割算法都是基于单一特征空间的算法，即使有将多特征空间结合在一起的分割算法也是仅仅将多个特征进行直接叠加，所以在叠加时权重的设置非常困难。本文提出了一种基于进化多目标框架的图像分割算法，该算法可以将多维特征空间有机的结合在一起，并且使用颜色特征和纹理特征作为两个目标特征进行测试。最终的结果表明，提出的方法可以有机的将颜色特征和纹理特征融合到图像分割中。

关键词： 聚类，自适应聚类，自适应图像分割，进化算法，多目标进化算法，多目标进化聚类

Abstract

As an unsupervised learning method, clustering has been widely used in many fields such as machine learning, data mining, artificial intelligence and image processing. And, it has been a hotspot research in these fields. With the increasing attention on clustering, several problems need to be solved. Firstly, how to determine the number of clustering automatically. Secondly, how to cluster and get the global optimization solution. Thirdly, how to do clustering for arbitrarily shaped data sets. Fourthly, how to integrate various different clustering methods together. Fifthly, how to take clustering method into application fields such as image processing. In this paper, our main contributions are as follows:

1) We introduce an automatic clustering method based on evolutionary algorithms (EAs). The basic idea is to convert a clustering problem into a global optimization problem and tackle it by an EA. A new validity index, which balances the inter-cluster consistency and the intra-cluster consistency, is proposed to be the objective function. Three adaptive coding schemes, which can deal with variable-length optimization problems by using a fixed-length chromosome, are designed to detect the cluster number automatically. The validity index and adaptive coding schemes are incorporated in an EA for automatic clustering. Our approach is compared with some widely used validity indices and an adaptive coding scheme on some artificial data sets and two real world problems. The experimental results suggest that our method not only successfully detects the correct cluster numbers but also achieve stable results for most of test problems.

2) We propose four methods for automatic detecting the number of clusters in arbitrarily shaped data sets. In the first method, path-based clustering is a well-known method for extracting arbitrarily shaped clusters. However, its high time complexity limits some possible applications. In this scheme, we propose two new algorithms to speed up the original path-based clustering method. A basic method focuses on the path-distance calculation. A modified Floyd algorithm is applied to reduce the time complexity. A preprocessing procedure is used to reduce the number of data points to the path-based clustering algorithm. Moreover, this algorithm can automatic determine the number of clusters by a box clustering. In the second method, an evolutionary arbitrarily shaped clustering (EASC) method is proposed for extracting arbitrarily shaped clusters. In EASC, the path distance is used

to measure the similarity between data points and a modified Modularity index is utilized as the optimization objective. In the third method, we apply real-based coding scheme to represent solutions in search space and establish a new validity index to cluster arbitrarily shaped data sets. In the forth method, we propose a multiobjective evolutionary frame to mix the different clustering algorithms. The new approaches are applied to a variety of test data sets with arbitrarily shapes and the experimental results show that our method is efficient in dealing with the given problems.

3) Undersegmentation or oversegmentation is a challenge faced by image segmentation methods, and it is extreme important to determine the current number of regions of an image in real-world applications. In this paper, we introduce an adaptive strategy to do so. The basic idea is to firstly oversegmentation an image by using the Mean-shift method; and then segment the obtained results by using an evolutionary algorithm. In the second stage, a feature is extracted for each region obtained by the Mean-shift method, and a new fitness function is designed to determine the proper number of clusters. The adaptive approach is applied to a variety of images, and the experimental results show that our method is both efficient and effective for image segmentation.

4) Although various image segmentation methods have been proposed in recent decades, most of these methods are based on only a single feature space. How to combine various features to image segmentation is a challenge problem. To address this problem, we propose to combine different features based on multiobjective evolutionary optimization. Two optimization objectives, which are based on color and texture features respectively, are therefore designed for image segmentation. The experiments show that our method is able to combine multiple features for image segmentation successfully.

Keywords: clustering, adaptive clustering, adaptive image segmentation; evolutionary algorithm, multiobjective evolutionary algorithm, multiobjective evolutionary clustering

目录

表格

插图

参考文献