Annals of Operations Research
https://doi.org/10.1007/s10479-020-03606-8

S.I.: DATA MINING AND DECISION ANALYTICS

An automatic clustering for interval data using the genetic algorithm

Tai Vovan[1] · Dinh Phamtoan[2,3,4] · Le Hoang Tuan[3,5] · Thao Nguyentrang[2,3]

Abstract
This paper proposes an Automatic Clustering algorithm for Interval data using the Genetic algorithm (ACIG). In this algorithm, the overlapped distance between intervals is applied to determining the suitable number of clusters. Moreover, to optimize in clustering, we modify the Davies & Bouldin index, and to improve the crossover, mutation, and selection operators of the original genetic algorithm. The convergence of ACIG is theoretically proved and illustrated by the numerical examples. ACIG can be implemented effectively by the established Matlab procedure. Through the experiments on data sets with different characteristics, the proposed algorithm has shown the outstanding advantages in comparison to the existing ones. Recognizing the images by the proposed algorithm gives the potential in real applications of this research.

Keywords Cluster analysis · DB index · Genetic algorithm · Interval data · Overlap distance

1 Introduction

Clustering is green to partition the data into groups so that similar objects belong to the same group and the dissimilar objects to different groups (Höppner and Böttcher 2007). It is

✉ Thao Nguyentrang
nguyentrangthao@tdtu.edu.vn

Tai Vovan
vvtai@ctu.edu.vn

Dinh Phamtoan
dinh110094@gmail.com; phamtoandinh@vanlanguni.edu.vn

Le Hoang Tuan
tuanlh@uit.edu.vn

1 College of Natural Science, Can Tho University, Can Tho, Vietnam

2 University of Science, Ho Chi Minh City, Vietnam

3 Vietnam National University, Ho Chi Minh City, Vietnam

4 Faculty of Engineering, VanLang University, Ho Chi Minh City, Vietnam

5 Department of Mathematics, University of Information Technology, Ho Chi Minh City, Vietnam

Published online: 18 April 2020