International Journal of Information Technology & Decision Making
Vol. 20 (2021)

DOI: 10.1142/S0219622021500309

World Scientific
www.worldscientific.com

An Ensemble Clusterer Framework based on Valid and Diverse Basic Small Clusters

Tao Sun*, Saeed Mashdour† and Mohammad Reza Mahmoudi‡,§

*Information Technology Center,
Guangzhou Academy of Fine Arts,
Guangzhou, Guangdong Province, P. R. China

†Nourabad Mamasani Branch,
Islamic Azad University, Mamasani, Iran

‡Department of Statistics, Faculty of Science,
Fasa University, Fasa, Iran
§rezastat@yahoo.com

Published

Clustering ensemble is a new problem where it is aimed to extract a clustering out of a pool of base clusterings. The pool of base clusterings is sometimes referred to as ensemble. An ensemble is to be considered to be a suitable one, if its members are diverse and any of them has a minimum quality. The method that maps an ensemble into an output partition (called also as consensus partition) is named consensus function. The consensus function should find a consensus partition that all of the ensemble members agree on it as much as possible. In this paper, a novel clustering ensemble framework that guarantees generation of a pool of the base clusterings with the both conditions (diversity among ensemble members and high-quality members) is introduced. According to its limitations, a novel consensus function is also introduced. We experimentally show that the proposed clustering ensemble framework is scalable, efficient and general. Using different base clustering algorithms, we show that our improved base clustering algorithm is better. Also, among different consensus functions, we show the effectiveness of our consensus function. Finally, comparing with the state of the art, we find that the clustering ensemble framework is comparable or even better in terms of scalability and efficacy.

Keywords: Ensemble clusterer; k-means clusterer algorithm; data validity.

1. Introduction

According to some studies, as a task in the decision support system,[1–6] pattern detection,[7,8] statistical data analysis,[9] machine learning,[10,11] optimization[12–15] and data mining[16] fields, clustering task has been regarded as a prominent issue.[17,18] It is

§Corresponding author.