Particle Swarm Clustering Ensemble

Abbas Ahmadi
Pattern Analysis and Machine Intelligence Lab,
Department of Systems Design Engineering,
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
abbas@pami.uwaterloo.ca

Fakhri Karray
Pattern Analysis and Machine Intelligence Lab,
Department of Electrical and Computer Engineering,
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
karray@uwaterloo.ca

Mohamed Kamel
Pattern Analysis and Machine Intelligence Lab,
Department of Electrical and Computer Engineering,
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
mkamel@pami.uwaterloo.ca

ABSTRACT

Extracting natural groups of the unlabeled data is known as clustering. To improve the stability and robustness of the clustering outputs, clustering ensembles have emerged recently. In this paper, an ensemble of particle swarm clustering algorithms is proposed. That is, the members of the ensemble are based on the cooperative swarms clustering approaches. The performance of the proposed particle swarm clustering ensemble is evaluated using different data sets and is compared to that of other clustering techniques.

Categories and Subject Descriptors: H.3.3 [Information Search and Retrieval]: Clustering

General Terms: Algorithms, Design.

Keywords: Particle swarm optimization(PSO), data clustering, clustering ensemble, multiple cooperative swarms.

1. INTRODUCTION

Particle swarm optimization is a bio-inspired search algorithm mimics the swarming behavior of flocks of birds [5, 6]. PSO has been applied into a number of clustering applications such as document clustering, gene expression and image segmentation all of which use a single swarm to deal with clustering problem. The single swarm clustering is based on search technique which does not rely on the initial solutions and increases the possibility of converging to a near-optimal solution [3]. When the dimensionality of the data is relatively high and the number of clusters is large, the ability of the single swarm clustering is not sufficient to explore all of the solution space thoroughly. Instead, multiple cooperative swarms can be considered to determine clusters' centers [2]. Multiple cooperative swarms clustering approach distributes the search space among several swarms each of which is responsible to search its corresponding portion while cooperating with other swarms [2].

To improve the stability and robustness of the clustering outputs, clustering ensembles have emerged recently [1]. In this paper, a new ensemble of clustering approaches is introduced in which the members of the ensemble belong to the multiple cooperative swarms clustering category. Each member provides its solution for the given data as a vector of labels assumed as a new feature. The new extracted features are used in a aggregation module which is based on multiple cooperative swarms clustering and the final clustering solution is released accordingly.

In the following, we first introduce our proposed particle swarm clustering ensemble. We next present experimental results. We finally provide concluding remarks.

GECCO'08, July 12–16, 2008, Atlanta, Georgia, USA.
ACM 978-1-60558-130-9/08/07.

2. ENSEMBLE OF PSO CLUSTERINGS

Clustering is the process of extracting natural grouping of data based on some measures of similarity. In clustering ensemble, several clustering algorithms provide solution to the given clustering task. These solutions obtained by B different clustering algorithms are combined in aggregation module to obtain a final clustering solution. Hence, two main components should be addressed in order to build a clustering ensemble. These components are the generation of different partitions (diversification) and consensus function (aggregation) as explained next.

2.1 Diversification

First, the model order of each clustering algorithm is selected randomly from a pre-specified range, $k \in [k_{min}, k_{max}]$, where k_{min}, k_{max} are the pre-specified maximum and minimum values for the model order of the given clustering problem. Having the model order selected, the cooperative swarms clustering algorithm [2] is applied on the data and the clustering solution is expressed by a vector of labels. In other words, each individual clustering can be presumed as feature extraction in which the original feature space $\mathbf{X}$ is mapped into a new feature space. Thus, each individual clustering algorithm reduces the dimensionality of the data.

2.2 Aggregation

The solution of different clustering algorithms together provide a new feature space of dimension B represented by $\tilde{X}$. This new feature space, considered as a new representation of the original data, is used in another multiple cooperative swarms clustering with a priori given number of clusters. The schematic presentation of the proposed cooperative swarms clustering ensemble is depicted in Fig. 1.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed ensemble of multiple cooperative swarms clustering algorithms, three

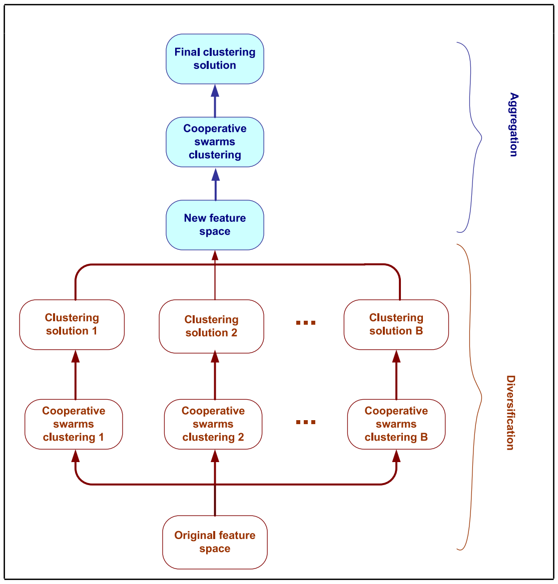

Figure 1: Architecture of the proposed method

data sets iris, zoo and liver-disordered from UCI machine learning repository [4] are selected.

The PSO parameters are considered as $w = 1.2$(decreasing gradually), $c_1 = 1.49$, $c_2 = 1.49$, and $n = 30$ (for all swarms). To keep a balance between compactness and separation measures in the combined measure, w_1 has been fixed to 0.85 [2]. Also, the model order is assumed to be equal to the number of classes for all data sets. Besides, the size of ensemble is $B = 10$ and the maximum and minimum values for the model orders of the clustering algorithms are assumed to be $k_{min} = 2$ and $k_{max} = 10$, respectively.

The proposed approach is compared with the ensemble of other clustering techniques such as k-means, k-harmonic means, fuzzy c-means, hybrid PSO, and single swarm clustering for different data sets in terms of combined measure as presented in Fig. 2. The results have been obtained by averaging over 30 independent runs. According to Fig. 2, the proposed approach outperforms the other techniques using different data sets. Moreover, the ensemble approaches based on k-means, k-harmonic means, and fuzzy c-means converges quickly as compared to the ensemble approaches based on particle swarm optimization. However, their final solution inferiors that of PSO clustering ensembles in terms of combined measure. Also, a sharp improvement in the behavior of clustering ensemble based on hybrid PSO and multiple cooperative swarms is duo to switching from k-means to single swarm in hybrid approach and beginning the cooperation among swarms in the multiple cooperative swarms approach.

4. CONCLUSIONS

In this paper, we introduced a new clustering ensemble scheme based on multiple cooperative swarms approach. The proposed approach encompassed two main components: diversification and aggregation. In diversification component, several clustering algorithms based on multiple cooperative swarms with different model orders have been used to provide solutions for the given data. Next in the aggregation

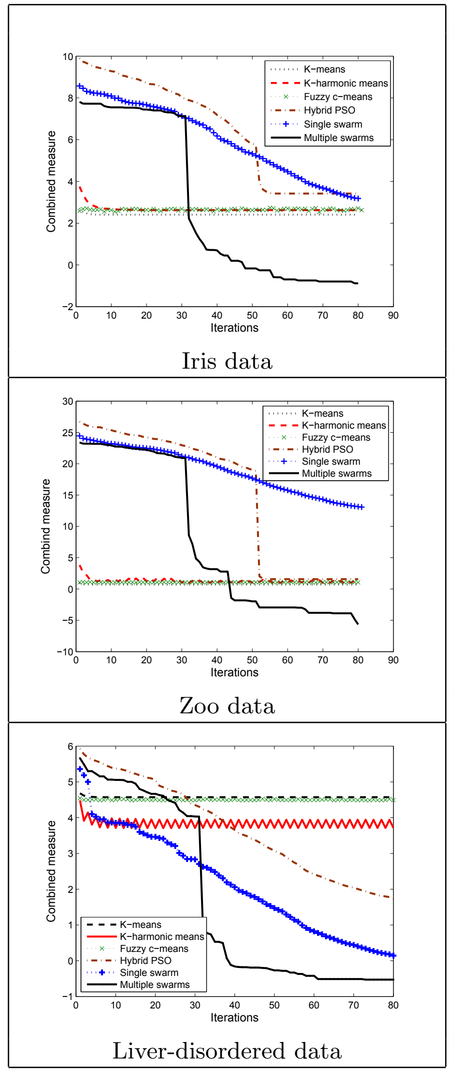

Figure 2: Comparing convergence of the proposed method with others in terms of combined measure

component, the outputs of the clustering algorithms in diversification part have been considered as new feature space to be clustered by a new multiple cooperative swarms approach with known model order. The proposed approach has been evaluated on a number of data sets from UCI machine learning repository.

5. REFERENCES

[1] A. Topchy, A.K. Jain and W. Punch, "Clustering Ensembles: Models of Consensus and Weak Partition," *IEEE Trans. PAMI 2005*, 27(12):1866-1881.

[2] A. Ahmadi, F. Karray and M. Kamel, "Multiple Cooperating Swarms for Data Clustering," *IEEE Proceedings of Swarm Intelligence Symposium 2007*, pages 206-212.

[3] X. Cui, T.E. Potok and P. Palathingal, "Document Clustering Using Particle Swarm Optimization," *IEEE Proceedings of Swarm Intelligence Symposium 2005*, pages 185-191.

[4] C.L. Blake and C.J. Merz, "UCI Repository of Machine Learning Databases," 1998.

[5] J. Kennedy, R.C. Eberhart and Y. Shi, "Swarm Intelligence," *Morgan Kaufman Publishers 2001*.

[6] A.P. Engelbrecht, "Fundamentals of Computational Swarm Intelligence," *John Wiley and Sons 2005*.