Clustering Analysis on Alumni Data Using Abandoned and Reborn Particle Swarm Optimization

Paulus Mudjihartono[1], Thitipong Tanprasert[2], Rachsuda Jiamthapthaksin[2]

[1]Department of Informatics, Universitas Atma Jaya Yogyakarta, Indonesia
[2]Department of Computer Science, Assumption University, Bangkok, Thailand
[1] p5729431@au.edu

Abstract—**Alumni data is one of the most important data that university management uses for developing the learning process decisions. This paper applies the idea of Abandoned and Reborn PSO (AR-PSO) to convert a clustering problem into the optimization form with an objective function to minimize the ugliness of the desired clusters. This algorithm of Clustering using AR-PSO (CAR-PSO) is slightly adapted to the cluster problem domain. The generated clusters need to be examined to decide if they are acceptable. There are three evaluations; the closeness, the separation and the purity. Finally, the experiment results show that the CAR-PSO is comparable with *k*-means in both types of alumni data while leaving the other two clustering algorithms.**

Keywords: clustering, alumni data, CAR-PSO, evaluations.

I. INTRODUCTION

One of data mining task is clustering. It deals with the big amount of data that need to be clustered to further take the information out of it. Many researchers have tried many algorithms to do clustering. Some of them do their own algorithms, and some modify it, and they do comparisons. He et al. [1] applied hierarchical *k*-means clustering algorithm based on ensemble learning that yields the better cluster accuracy. Another clustering algorithm is introduced by Wang and Jing [2] for dealing with unsupervised relation extraction. It introduced co-kmeans based on co-clustering theory, which turned out that it outperforms *k*-means.

Alternatively Particle Swarm Optimization (PSO) as a tool for searching optimized values from the objective function has been also used in many other area, which originally is not an optimization problem. In general, PSO can be applied to many different problems through problem mapping. The optimization problems sometimes cannot be transformed into a derivable objective function. PSO in [3] plays the important role in solving this kind of problems by avoiding the divergence.

Mudjihartono [4] did AR-PSO in University Timetable Problem and compared it to the original-PSO. It showed the better results of AR-PSO in terms of how the solution deals with the customized timetable constraints. The contribution of this paper is introducing the algorithm of Clustering using AR-PSO (CAR-PSO), which adapts AR-PSO in order to address the clustering problem.

The organization of the rest of the paper is as follows. Section II provides related works; Section III provides the idea of CAR-PSO, while Sections IV and V provide experiment results and conclusions, respectively. Experiment Results is the main body of the paper that consists of three aspects of evaluation in detail; the closeness, the separation and the purity.

II. RELATED WORKS

A modified PSO has once been introduced by [5]. In this modified algorithm, the particle movement is calculated not only by personal and global best but also by the position of the nearest neighbor of the personal best. A fuzzy clustering is used to improve the basic PSO. It results better in multi-modal functions than that in basic PSO.

Consensus Clustering based on PSO also was once presented by [6]. One of the clustering algorithms to be contested is PSO and the result shows that PSO is good for some datasets. Sharma and Sisodia [7] offered a solution over the weakness of the interval length of fuzzy time series forecasting. They used modified-turbulent PSO to automate the clustering for partitioning the universe of discourse.

A new approach clustering based on the Artificial Bee Colony (ABC) is proposed in [8] to deal with categorical data. It enhances the k-modes type of algorithm that tends to fall into local optima. The algorithm combines *k*-modes algorithm and ABC optimization procedure. ABC-*k*-modes, the name of the algorithm, outweighs the *k*-modes algorithms: k-modes, Fuzzy-*k*-modes, and Genetic *k*-modes in term of accuracy, precision, recall and Rand index.

The PSO has undergone some modifications several times to get better result or to adapt the problem domain. The AR-PSO is a modification of PSO of how the (some) particles moved. It is a considerably worth to apply it to do clustering with some another modification.

978-1-4673-8139-0/16/$31.00 ©2016 IEEE