

Improved Differential Evolution on Optimizing 2-Level Ant Clustering Using of Variation

Phichete Julrode

Department of Informatics Faculty of Science and Technology,
Phuket Rajabhat University, Phuket 83000, Thailand
Email: phichete@pkru.ac.th

Abstract— Ant based Clustering (ANT) is a very powerful tool for automatic detection of relevant clusters. The extended version of ANT, two-level Ant based Clustering (2LANT) was introduced for improving ANT clustering in explorative manner. However, structural methods for efficiently confirming the competent optimization of 2LANT initialization are lacking. Due to the important advantages over other optimization algorithms belonging to differential evolution (DE) approach, this paper investigates the utilization of the original DE as well as the variations, here called VarDE1 and VarDE2 as tools for optimizing the initial cluster weights of 2LANT. Such investigated approaches are respectively so called DE+2LANT, VarDE1+2LANT and VarDE2+2LANT. With respect to the different choices of mutation process, both variant DEs would get better accuracy than the original one. More elitism on mutation process is involved with VarDE2+2LANT rather than with VarDE1+2LANT; whilst the most random mutation is applied by DE+2LANT. 10-fold cross validation experiments are taken on real-world and artificial data sets with an identified number of clusters. Within the scope of this paper, the investigation results point out the better clustering performance of the variant DEs, VarDE2+2LANT over the related approaches.

Index Terms— Two-level ant based clustering, differential evolution, agglomerative hierarchical clustering

I. INTRODUCTION

Ant Based Clustering (ANT) [1], originally suggested by Dorigo, is one of the efficient methods for clustering problems. It is a multi-agent computational algorithm to find the best solution of cluster provided to the space of initialization through a competitive, nonlinear, unsupervised learning process. To improve ANT on the issue of explorative manner, two-level ANT (2LANT) was introduced [2]. In the first level of 2LANT, ant based clustering (ANT) is utilized as a training procedure; and yields a set of trained clusters. In the second level, a traditional clustering approach such as agglomerative hierarchical clustering (AHC) is applied for grouping the trained clusters, which obtained from the first-level ANT [3]. The grouping criterion is to apply the shortest distance between clusters. The clusters topology is refined in this second level. However, in order to improve the clustering accuracy, an efficient optimization

algorithm should be executed on ANT for achieving the appropriate, evolved set of initial clusters. Differential evolution (DE) [4]-[6] have been widely employed to deal with optimization problems. It is a population-based evolutionary algorithm (EA) [7] that applies the operators: crossover, mutation and selection. DE has been proposed particularly for real number optimization problems. There exist three main advantages of DE over many conventional EAs: finding the true global minimum of a multi-modal search space regardless of the initial parameter values, fast convergence, and using a few control parameters. By these reasons, this paper investigates the utilization of the original DE as well as the variations of it namely VarDE1 and VarDE2 as a tool for optimizing the 2LANT on account of cluster weights initialization. Such investigated approaches are respectively so called, DE+2LANT, VarDE1+2LANT and VarDE2+2LANT. More elitism on mutation process is involved with VarDE2+2LANT rather than with VarDE1+2LANT; whilst the most random mutation is applied by DE+2LANT. Hepatitis, Parkinson, Iris, Wine and 2-D artificial data sets with identified number of clusters are tested in the experiments. The remainder of this paper is organized as follows: Section II, clustering ANT and 2LANT are briefly reviewed. Section III gives descriptions of the original DE, the VarDE1+2LANT and VarDE2+2LANT. Section IV provides the investigation of performance of the related approaches. Finally, the conclusions are given in section V.

II. OVERVIEW OF THE ANT BASED CLUSTERING (ANT) AND TWO-LEVEL ANT (2LANT)

A. Ant based Clustering (ANT)

The ant systems, developed with concept of simple multi-agent principles emphasize distributiveness, flexibility, and robustness. Ant-based algorithm has been developed using swarm intelligence principles that emphasize distributiveness, direct or indirect interactions among relatively simple agents, flexibility, and robustness [8], [9]. By such competent characteristics, ant-based clustering more relieves the fast convergence during searching process than several other evolutionary approaches. Groups of ants cooperate to move cluster centers in feature space to search for optimal clusters partition.

Manuscript received September 8, 2013; revised November 28, 2013.

Initially, the feature values are normalized between 0 and 1. Each ant is assigned to a particular feature of a cluster in a partition. The ants never change the feature, cluster or partition assigned to them. After randomly moving the cluster centers for a fixed number of iterations, called an epoch, the quality of the partition is evaluated. According to the former ant colony optimization, particular clusters partition is selected as the best one based on a certain cascading rules of probability. A certain additional memories are exploited for storing some numbers of good clusters. However, the ant colony optimization used in this paper selection does not need the additional memories to store more than one good cluster centers; instead, it applies the heuristic technique, based on Boltzmann probability [10]. Such technique supports the exploration of the good clusters. Similar to the former ant colony optimization, there are two directions for the random movement of the ant. The positive direction is when the ant is moving in the feature space from 0 to 1, and the negative direction is when the ant is moving in the feature space from 1 to 0. If during the random movement the ant reaches the end of the feature space the ant reverses the direction. After a fixed number of epochs the ants stop. The cluster centers, finally obtained are then used as the clustering approaches.

B. The 2LANT

The 2LANT can be viewed as an extension of ANT. In order to hierarchically grouping the clusters, a set of trained clusters are submitted from ANT in the first level to a traditional clustering approach, i.e. an agglomerative hierarchical clustering approach (AHC) [11] in the second level. The grouping criterion is to apply the shortest distance between clusters. The process of 2LANT begins with ANT. Similar to 2LANT [12], more number of cluster nodes than apriori known are recruited in input layer for executing ANT. This may improve the clustering process in an explorative manner. Then, AHC groups the trained clusters, which are obtained from ANT into a hierarchical structure according to proximity of trained clusters. Until attaining the identified number of clusters, AHC performs a subsequent one-by-one agglomeration stage of merging the most similar clusters with respect to the minimum distance. A refinement of clusters topology is done in the second level where a final set of cluster is achieved.

III. USING DIFFERENTIAL EVOLUTION (DE) AND THE VARIANT DES (VARDE1, VARDE2) ON OPTIMIZING 2-LEVEL ANT (DE+2LANT, VARDE1+2LANT AND VARDE2+2LANT)

Like several EAs, DE has three operations: mutation, crossover and selection. For the purpose of using DE for optimizing 2LANT with respect to initial set of cluster weights, DE starts with W population size. Each individual in the population is represented by $\mathbf{M}_j(t) \subset \Re^{AxK}, j \in \{1, \dots, W\}$, where A and K are respectively the number of attributes and clusters.

$$\begin{aligned}\mathbf{M}_j(t) &= ([clusterWeight_j^1]_{Ax1}, \dots, [clusterWeight_j^K]_{Ax1}) \\ &= (m_{j1}(t), \dots, m_{jk}(t)),\end{aligned}$$

where $m_j(t) \subset \Re^A$. After population is initialized, an iterative process is started. At each iteration, a target matrix, $\mathbf{T}(t) = (t_1(t), \dots, t_K(t))$ which is associated with the best fitness is found. The fitness here is referred by clustering accuracy of 2LANT. Then, a mutation matrix, $\mathbf{V}_j(t)$ is created, which is a combination of randomly chosen individual set of initial clusters $\mathbf{M}_j(t)$. $\mathbf{V}_j(t)$, a mutated solution can be generated by using a strategy like:

$$\mathbf{V}_j(t) = \mathbf{M}_{rand_0}(t) + F(\mathbf{M}_{rand_1}(t) - \mathbf{M}_{rand_2}(t)). \quad (1)$$

where indices $rand_0$, $rand_1$, $rand_2$ are randomly and mutually different integers generated in the range $[1, W]$. $F \in [0, 2]$ is a scale factor, which controls the mutation size. Each mutation matrix, $\mathbf{V}_j(t)$ is crossover with target matrix, $\mathbf{T}(t)$ to generate the trial matrix, $\mathbf{U}_j(t) = (\mathbf{u}_{j1}(t), \dots, \mathbf{u}_{jK}(t))$ by a following rule: $u_{jk}(t+1)$ would be replaced by the corresponding elements in $M_j(t+1)$ if $rand_{1j} \leq CR$ $rand_{1j} \leq CR$ or $J = j_{rand}$, otherwise replaced by $T_j(t)$. Where CR is a crossover constant in the range $[0, 1]$, specified by users, and j_{rand} is a randomly chosen integer in the range $[1, K]$ to ensure that the trial matrix, $\mathbf{U}_j(t+1)$ will get at least one element from the mutation matrix. Finally, the selection operation favors the better one, or $\mathbf{U}_j(t+1)$; is regulated by:

$$\mathbf{M}_{j(t+1)} = \begin{cases} \mathbf{U}_j(t+1) & f(\mathbf{U}_j(t+1)) \geq f(\mathbf{M}_j(t)), \\ \mathbf{M}_j(t) & \text{Otherwise} \end{cases} \quad (2)$$

f refers to the fitness function. $\mathbf{M}_{j(t+1)}$ from (2), for all j represent the new population for the next iteration of DE. Fig. 1 illustrates the utilization of DE on optimizing 2LANT with respect to the initial weights.

Fig. 1 can be explained in the following steps:

Step 1: \mathbf{X}_{AxP} represents P input data with A dimensions. A set of $\mathbf{M}_1(0), \dots, \mathbf{M}_w(0)$ corresponds to DE population; would be a set of initial cluster weights.

Step 2: DE operations are performed; the new DE population $\mathbf{M}_1(t), \dots, \mathbf{M}_w(t)$ is generated; and is one-by-one presented to 2LANT in order to compute the fitness.

Step 3: According to 2LANT, an individual, $\mathbf{M}_j(t) \subset \Re^{AxK}$ assigned to ANT that associated with the first level of 2LANT. ANT generates a set of trained cluster, $\mathbf{M}'_j(t) \subset \Re^{AxK}$ is submitted to AHC in the 2LANT second level for cluster grouping purpose and to DE to represent $\mathbf{M}_j(t+1)$. The accuracy fitness of $\mathbf{M}_j(t+1) = f(\mathbf{M}''_j(t))$ is generated by AHC, where $\mathbf{M}''_j(t) \subset \Re^{AxK'}, K' < K$.

Step 4: Repeat Step 2-4 until maximum iteration is reached.

The DE and a couple of variant DEs, VarDE1+2LANT and VarDE2+2LANT are investigated in this paper. The difference between the variant DEs and the original one locates on the mutation process. In the DE, mutation operation is based on (1); while in

VarDE1+2LANT and VarDE2+2LANT, they are respectively based on (6) and (7).

$$\mathbf{V}_j(t) = \mathbf{T}(t) + F (\mathbf{M}_{rand1}(t) - \mathbf{M}_{rand2}(t)) \quad (3)$$

$$\mathbf{V}_j(t) = \mathbf{T}(t) + F (\mathbf{T}(t) - \mathbf{M}_{rand2}(t)) \quad (4)$$

It is seen from (1), (3) and (4) that the mutant vectors that are produced by VarDE2+2LANT are rather elitism, since they locate around the target matrix, $\mathbf{T}(t)$ with which the best fitness is associated; while those produced by the original DE are most randomly generated.

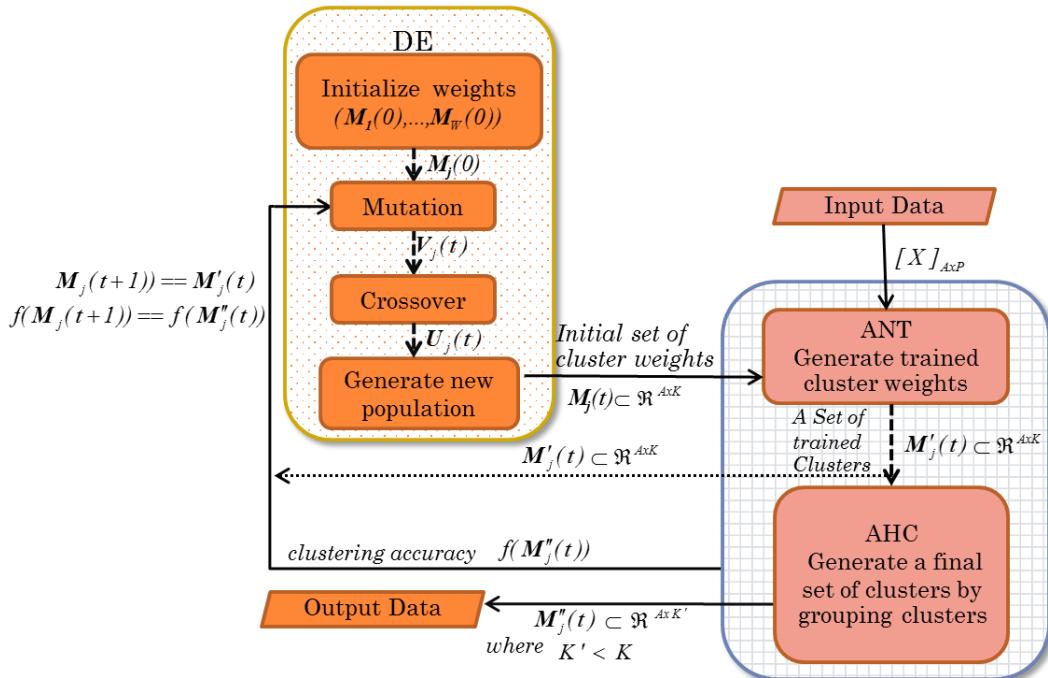


Figure 1. DE+2LANT procedure

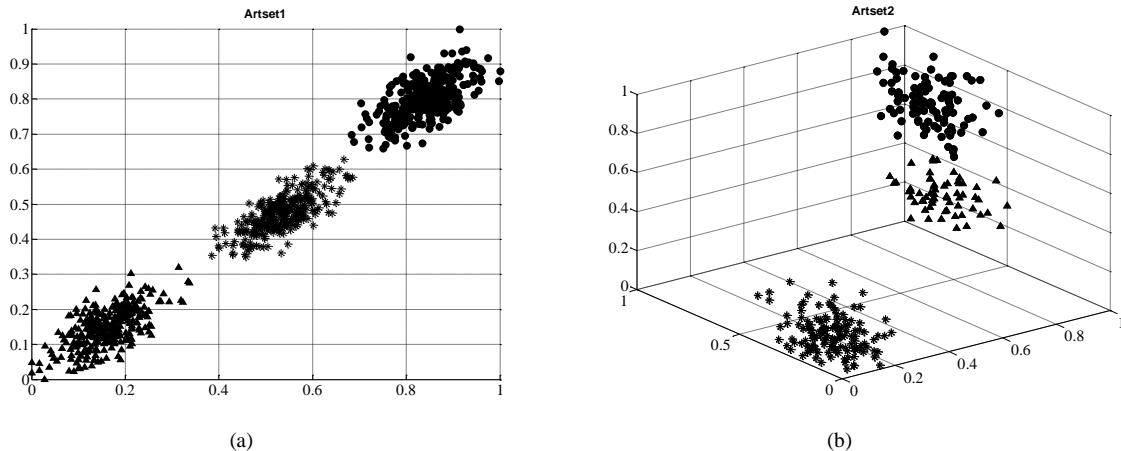


Figure 2. (a) 2-D artificial data sets, (b) 3-D artificial data sets

IV. EXPERIMENTS AND RESULTS

The experiments are performed on the real-world benchmarks, Hepatitis, Parkinson, Iris, Wine from the University of California at Irvine–Machine Learning Repository [13], 2-D and 3-D artificial data sets show in Fig. 2(a) and Fig. 2(b). The first data set, 2-class Hepatitis has 19-D 155 data; 2-class Parkinson has 22-D 195 data; 3-class Iris has 4-D 150 data; 2-class Wine has 13-D 178 data; 2-class, 2-D artificially simulated data set contains 900 artificial data points and 2-class, 3-D artificially simulated data set contains 900 artificial data

points. The scale factor, F and the population size related to DE are consecutively set to 0.6 and 35. The performance measurement of clustering relies on the following two cluster validity indices: One is adjusted rand index (ARI) [14] which is a measure of agreement of clustering results against those of apriori known. The ARI value lies between zero and one. A high value indicates that the clustering and known results are very similar. The other, Minkowski score (MS) [15] index is the normalized distance between two matrices, clustering and apriori known class. Lower MS implies better clustering solution; and a perfect solution will have a zero

score. Table I. and Table II. consecutively indicate the average ARI and MS of the related methods that are obtained by 10-cross validation run on the 5 data sets. The same set of initial cluster weights is applied to all methods for the beginning step.

TABLE I. THE AVERAGE ARI BASED ON 10-CROSS VALIDATION RUNS

| Data Set | <i>VarDE2+2L ANT</i> | <i>VarDE1+2L ANT</i> | <i>DE+2L ANT</i> | <i>ANT</i> |
|----------------|----------------------|----------------------|------------------|------------|
| Hepatitis | 0.78 | 0.76 | 0.62 | 0.62 |
| Parkinson | 0.75 | 0.69 | 0.67 | 0.66 |
| Iris | 0.97 | 0.97 | 0.91 | 0.87 |
| Wine | 0.96 | 0.96 | 0.95 | 0.94 |
| 2-D artificial | 1.00 | 1.00 | 0.98 | 0.96 |
| 3-D artificial | 1.00 | 0.99 | 0.98 | 0.96 |

With respect to Table I. and Table II., both ARI and MS provide the same meaning of the results. It points out

the better clustering performance of VarDE2+2LANT and VarDE1+2LANT over the others. A larger performance difference between VarDE2+2LANT, VarDE1+2LANT, DE+2LANT and ANT is shown when they are applied on the data with larger amount of attributes. The evidence is shown in Fig. 3.

TABLE II. THE AVERAGE MS BASED ON 10-CROSS VALIDATION RUNS

| Data Set | <i>VarDE2+2L ANT</i> | <i>VarDE1+2L ANT</i> | <i>DE+2L ANT</i> | <i>ANT</i> |
|----------------|----------------------|----------------------|------------------|------------|
| Hepatitis | 0.22 | 0.24 | 0.38 | 0.38 |
| Parkinson | 0.25 | 0.31 | 0.33 | 0.34 |
| Iris | 0.03 | 0.03 | 0.09 | 0.13 |
| Wine | 0.04 | 0.04 | 0.05 | 0.06 |
| 2-D artificial | 0.00 | 0.00 | 0.02 | 0.04 |
| 3-D artificial | 0.00 | 0.01 | 0.02 | 0.04 |

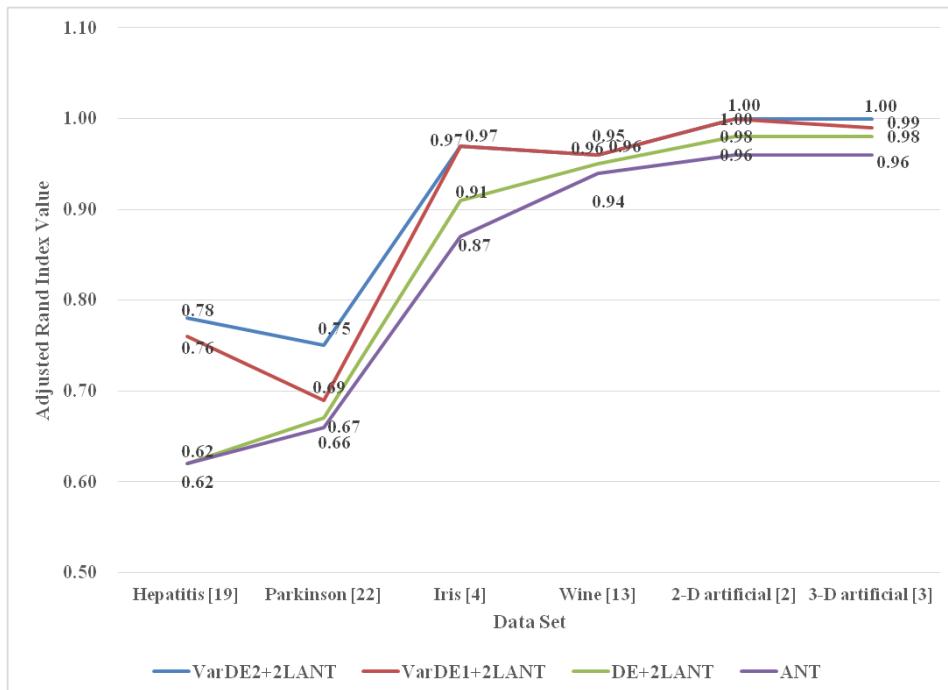


Figure 3. The ARI values as results of applying data with various number of attributes

V. CONCLUSIONS

In this paper, the DE as well as the variations, here called VarDE1 and VarDE2 are investigated with respect to the optimization of the initial cluster weights of 2LANT. The investigation results show the better clustering performance of the variant DEs, VarDE1+2LANT and VarDE2+2LANT over the DE+2LANT and ANT. The better performance of those variant DEs over the others gets more dominant when data sets with a high number of attributes are applied. This may indicate, within a scope of this paper that the more elitism of the DE mutation process, the better clustering performance may be obtained. However, the results on applying the clustering approaches on 19-D

Hepatitis and 22-D Parkinson show the low performance. This is due to the noisy and irrelevant attributes that are hidden in the data set. One of the future works would relate to the reduction of such noise and irrelevant attributes as the clustering preprocessing.

REFERENCES

- [1] M. Dorigo, "Optimization, learning and natural algorithms," Ph.D. Thesis, Dipartimento di Elettronica, Politecnico di Milano, Italy, 1992.
- [2] M. Dorigo and T. Stutzle, "The ant colony optimization metaheuristic: algorithms, applications, and advances," *Handbook of Metaheuristics*, F. Glover and G. Kochenberger, Eds., Kluwer Academic Publishers, vol. 57, 2003, pp. 250-285.
- [3] E. Lumer and B. Faieta, "Diversity and adaptation in populations of clustering ants," in *Proc. Third International Conference on*

- Simulation of Adaptive Behavior: From Animals to Animats3*, MIT Press Cambridge, 1994, pp. 489–508.
- [4] R. Storn and K. Price, “Differential evolution-A simple and efficient adaptive scheme for global optimization over continuous spaces,” in *Proc. ICSI, Berkeley, CA*, Tech. Rep. TR-95-012, 1995.
 - [5] R. Storn and K. Price, “Differential evolution-A simple and efficient heuristic strategy for global optimization over continuous spaces,” *J. Glob. Optim.*, vol. 11, no. 4, pp. 341–359, Dec 1997.
 - [6] K. Price, R. Storn, and J. Lampinen, *Differential Evolution-A Practical Approach to Global Optimization*. Berlin, Germany: Springer-Verlag, 2005, Ch. 2.
 - [7] Z. F. Wu, H. K. Huang, and Y. Zhang, “A modified differential evolution algorithm with self-adaptive control parameters,” in *Proc. 3rd International Conference Intelligent System and Knowledge Engineering*, Nov 17-19, 2008, pp. 524-527.
 - [8] E. Bonabeau, M. Dorigo, and G. Theraulaz, *Swarm Intelligence: From Natural to Artificial Systems*, New York, USA: Oxford University Press, Inc., 1999.
 - [9] M. Dorigo, E. Bonabeau, and G. Theraulaz, “Ant algorithms and stigmergy,” *Future Generation Computer Systems*, vol. 16, no. 8, pp. 851-871, 2000.
 - [10] C. Y. Lee, “Entropy-Boltzmann selection in the genetic algorithms,” *IEEE Transactions on Systems Man and Cybernetics Part B Cybernetics a Publication of the IEEE Systems Man and Cybernetics Society*, vol. 33, no. 1, pp. 138-149, 2003.
 - [11] X. Dai, Y. He, and Y. Sun, “A Two-layer text clustering approach for retrospective news event detection,” in *Proc. Artificial Intelligence and Computational Intelligence International Conference*, Oct 23-24, 2010, pp. 364-368.
 - [12] P. Julrode and S. Supratid, “A novel fuzzy ant-based clustering using a nonparametric balance between exploitation and exploration,” in *Proc. 13th International Conference on Applied Computer Science WSEAS*, Morioka City, Iwate, Japan, April 23-25, 2013, pp. 40-52.
 - [13] UCI repository of machine learning databases. Dept. of Info. And Comp. Sci. UCI. CA. [Online]. Available: <http://www.ics.uci.edu/~mlearn/>
 - [14] X. L. Xie and G. Beni, “A validity measure for fuzzy clustering,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841-847, Aug 1991.
 - [15] N. Jardine and R. Sibson, *Mathematical Taxonomy*, New York: Wiley, 1971.



Phichete Julrode is a lecturer at the Department of Informatics, Faculty of Science and Technology, Phuket Rajabhat University, Thailand. He received his bachelor degree in Computer Science from Phetburi Rajabhat University, Thailand in 1995, master degree in Computer Science, Chiangmai University, Thailand, 2005 and doctoral degree in Information Technology from Rangsit University, Thailand in 2012. His research interests include artificial intelligence, data mining, machine learning, and evolutionary algorithms.