Investigation on the Effects of ACO Parameters for Feature Selection and Classification

Shunmugapriya P.[1], Kanmani S.[2], Devipriya S.[2], Archana J.[2], and Pushpa J.[2]

[1] Dept. of CSE
[2] Dept. of IT
Pondicherry Engineering College, Puducherry, India
pshunmugapriya@gmail.com

Abstract. Ant Colony Optimization (ACO) is the most familiar meta-heuristic search algorithm and has been used in optimization of a number of feature selection (FS) problems. As a meta-heuristic search algorithm, ACO requires a set of parameters to solve the problem of feature selection. Pheromone Evaporation Rate (PER) is the most important among all these parameters. Setting up the values of these parameters is a big deal. Usually, these parameters are set up by experimenting through a number of values and finally selecting a set of values that seems to work well for feature selection. The change in optimal feature selection in accordance to different values of PER and other ACO parameters is discussed in this paper. ACO is applied for feature selection and classification of 10 datasets. From the experimental results, it can be seen that, the optimal value for the evaporation rate (ρ) lies around 0.7 leading to selection of best features and increase in classification accuracy.

Keywords: Optimization, Ant Colony System, Pheromone, Pheromone Evaporation, Classification, Feature Selection, ACO parameters.

1 Introduction

Feature selection is constructively used by a number of Machine Learning algorithms especially Pattern Classification [2, 3, 4, 8, 9 and 10]. The presence of redundant, irrelevant and noisy data may result in poor prediction (classification) performance. Feature selection extracts the relevant and most useful features without affecting the original representation of the dataset. The generic purpose of a Feature Selection Algorithm related to Pattern Classification is the improvement of the classifier or learner, either in terms of learning speed, generalization capacity or simplicity of the representation [10].

Ant Colony Optimization (ACO) is a meta-heuristic search algorithm which has been successfully employed to implement feature selection in numerous applications [5, 7, 21, 22, 23, 24, 25, 26, 27 and 28]. It can be inferred from these works that ACO leads to optimal selection of features and effectively increases the prediction results. ACO employs certain parameters to solve the optimization problems. These parameters are Pheromone Evaporation Rate (PER), Local Pheromone Update (LPU),

V.V. Das and J. Stephen (Eds.): CNC 2012, LNICST 108, pp. 136–145, 2012.

parameter stating relative importance (β), parameter which decides the component selection(τ_o) and the number of ants [1] and [11].

While performing optimization using ACO, its parameters have to be fine tuned and assigned values. These values are assigned after experimenting with an allowed set of numbers. The performance of the ant colony system changes with respect to the change of values of the parameters especially PER. Works have been carried out analyzing the role of ACO parameters in combinatorial optimization problems like Traveling Salesman Problem, online parameter adaptation etc [11, 14, 15, 16, 17, 18, 19 and 20]. Our work searches for the optimal values for ACO parameters in relation to FS problem optimization. 10 standard datasets have been used to check for the behavior of ACO according to different values for the PER, LPU and β. It could be inferred from the results that, ACO leads to better optmization, when the value of PER is assigned between 0.1 and 0.7. For the value of PER from 0.1 to 0.7, classification accuracy keeps increasing and the accuracy takes a transition when PER exceeds around 0.75 and starts to decrease. So, from the experiments conducted, the optimal value to be assigned to PER when optimizing FS by ACO is a value around 0.75.

This paper is organized in 6 sections. Feature Selection and Classification are discussed in section 2. Section 3 gives a brief description of ACO and Pheromone Trial. Section 4 outlines the ACO algorithm and ACO parameters. The computational experiments and results are described in section 5. Section 6 concludes the paper.

2 Classification and Feature Selection

2.1 Classification

A classifier takes a set of features as input and these features have different effect on the performance of classifier. Some features are irrelevant and have no ability to increase the discriminative power of the classifier. Some features are relevant and highly correlated to that specific classification. For classification, sometimes obtaining extra irrelevant features is very unsafe and risky [2]. A reduced feature subset, containing only the relevant features helps in increasing the classification accuracy and reducing the time required for training.

2.2 Feature Selection (FS)

Feature selection is viewed as an important preprocessing step for different tasks of data mining especially pattern classification. When the dimensionality of the feature space is very high, FS is used to extract the relevant and useful data. FS reduces the dimensionality of feature space by removing the noisy, redundant and irrelevant data and thereby makes the feature set more suitable for classification without affecting the accuracy of prediction [4]. It has been proved in the literature that “classifications done with feature subsets given as an output of FS have higher prediction accuracy than classifications carried out without FS” [3].

A number of algorithms have been proposed to implement FS. Apart from the ordinary FS algorithms there are two types of feature selection methods related to pattern classification: Filter Approach and Wrapper Approach. Evolutionary algorithms are used widely for searching the best subset of features through the entire feature space [5, 7, 21, 22, 23, 24, 25, 26, 27 and 28].

3 Ant Colony Optimization and Pheromone Trail

Ant Colony Optimization (ACO) was introduced in early 1990s by M.Dorigo and his colleagues [6]. ACO algorithm is a novel nature inspired meta- heuristic for the solution of hard combinatorial optimization problems. The main inspiration source of ACO is the foraging behavior of real ants. Ants are social insects living in colonies with interesting foraging behavior. An ant can find the shortest path between the food source and a nest. Initially, ants walk in random paths in search of food sources. While walking on the ground between the food source and the nest, ants deposit a chemical substance called the pheromone on the ground and a pheromone trial is formed. This pheromone evaporates with time. So, the shorter paths will have more pheromone than the longer paths. Ants can smell the pheromone and when choosing their paths, they tend to choose the paths with stronger pheromone concentration. This way, as more ants select a particular path, more and more pheromone will be deposited on the path. At a certain point of time, this path (the shortest path) will be selected by all the ants [1].

4 ACO Algorithm and its Parameters

In this work, ACO algorithm is used to optimize the selection of features and the features selected are used to form a training set. The classifier gets trained with the selected features and tests on a validation set. If the accuracy of prediction is better, then ACO is allowed to proceed with the optimal selection of features. So, the classification algorithm together with ACO plays the role of feature selector while simultaneously increasing the accuracy of classification.

4.1 ACO Parameters

ACO employs a set of parameters for optimization and states a mechanism to assign values to these parameters. The parameters employed within ACO are listed and explained as follows [1] and [adapted from 11]:

τ_o - This parameter determines whether the ant uses the greedy or probablistic form of component selection equation at each step of the algorithm.
ϕ - The local pheromone updating factor(LPU).
ρ - The global pheromone updating factor (Evaporation rate) (PER)
β - The relative importance placed on the visiblity heurisic.

All these parameters are important and plays effective role in optimization. ACO requires setting up the values of these parameters, before commencing its selection process. Apart from this, the number of ants to do the search procedure should also be decided. The ACO mechanism allows each ant to maintain its own parameters and in turn these are used to adapt to other paramter values. PER is the significant parameter as it decides where more pheromone is to be accumulated and what is to be selected[1].

4.2 The ACO Algorithm

The implementation of ACO based feature selection is based on the ACO algorithm proposed by Nadia Abd-Alsabour et al [7]. The following equations (1), (2), (3) are used in implementing ACO to select optimal features and increase the classification accuracy [adapted from [7]].

$$P_i = \tau_i . \Delta\tau_i . \quad (1)$$

$$\tau_i = (1-\varphi).\tau_i + \varphi.\tau_0 . \quad (2)$$

$$\tau_i = (1-\rho).\tau_i + \left(\rho.\frac{1}{L_{best}}\right)^{\beta} . \quad (3)$$

At the start of the algorithm, all the parameters are initialized. The features of the dataset are assigned a pheromone value each and are initialized to a small positive number in the range [0, 1]. Each ant selects a feature based on the probability value as given in (1). Accumulation of pheromone is done, when a feature is selected by using equation (2). Because the classifier is also involved, the pheromone accumulation actually encourages the selection of the feature that is more relevant and has a positive effect on the classification accuracy. After all the ants have finished a run, the ant producing the highest classification accuracy is considered as the best ant (L) and then the global pheromone update is done using equation (3). Global pheromone update actually leads to the pheromone evaporation of irrelevant features. After a predetermined number of iterations are over, the algorithm halts yielding the set containing optimal features.

5 Experiments and Discussions

5.1 Datasets

We have arrived at the optimal values for ACO parameters to be used with Feature Selection and classification, based on the experiments conducted using 9 UCI (University of California, Irvine) datasets and the HIV dataset. The datasets are taken from the UCI repository [12] and the datasets used are Cleveland Heart, Hepatitis, Lung Cancer, Dermatology, Pima Indian Diabetes, Liver, Wisconsin, Diabetes and HIV. The description of the data sets is given in Table 1. All the datasets listed in Table 1 are standard datasets and have been used in a number of classification and feature selection problems. In order to arrive at optimal values for the ACO parameters, these datasets are used.

5.2 Experiment

All computations are done using WEKA (Waikato Environment of Knowledge Analysis) [13]. As discussed in section 4.1, the ACO parameters used for optimization of feature selection are ρ, ϕ, β, τ_o, τ_i and the number of ants. The number of ants is usually set equal to the number of features and the pheromone values τ_i of the features are initialized to a small positive number in the range[0,1]. Because the pheromone values assigned are probability distributed, the pheromone values can be randomly assigned to the features. ρ, ϕ, β should also be assigned values in the range [0, 1]. The effect of these parameters over optimization has been discussed in a number of works in the literature [11, 14, 15, 16,17,18,19 and 20].

Table 1. Datasets Description

Dataset	No. of Samples	No. of Features	No. of Classes
Heart-C	303	13	2
Dermatology	366	34	6
Hepatitis	155	19	2
Lung Cancer	32	56	2
Pima Indian Diabetes	768	8	2
Liver	345	6	2
Wisconsin	699	9	2
Lymphography	148	18	4
Diabetes	768	9	2
HIV	500	21	3

The parameter of global update ρ, which is also the parameter indicating the evaporation rate, has significant effect over the selection of features, based on the value set to it. The relative factor β and LPU ϕ, affects the optimal feature selection but compared to the evaporation rate ρ, they have lesser significance. The importance of the pheromone evaporation rate has been revealed in the literature [1, 11, 14, 15, 16,17,18,19 and 20].

Table 2. HEART C - ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy(%)
0.1 TO 0.69	0.0	13	1(2)	65.67
	0.10	13	3(7,11,13)	71.94
	0.2	13	2(7,13)	82.25
	0.3	13	5(2,3,7,11,13)	86.85
0.7 To1	0.0	13	1(7)	58.74
	0.1	13	2(1,7)	59.73
	0.2	13	1(2)	65.67

The following tables 2,3,4,5,6,7,8,9,10 and 11 show how the values of the parameters affect the behavior of ACO in optimal selection of features and in increase of classification accuracy. For Heart dataset, when the evaporation rate ρ is assigned the value from 0 to 0.69, the features are selected and the accuracy keeps increasing. When ρ takes up values in the range of [0.7, 1], the classification accuracy keeps decreasing and is not a favorable situation. The relative factor β has the same effect over optimization for all the values in the interval [0, 1]. The local pheromone update ϕ has yielded better result when it is set to 0.3.

Table 3. HEPATITIS **-** ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy (%)
0.1 TO 0.73	0.0	19	2(4,7)	58.06
	0.1	19	17(all except 13,19)	60
	0.2 to 1	19	14(all except 13,15,16,17,19)	65.16
0.74 To 1	0.0 to 1	19	13(all except 10,11,14,15,17,19)	77.45

Table 4. LUNG CANCER - ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy (%)
0.1 To 0.78	0.0 to 1	56	33(1,2,4,6,7,9,11,12,17,18, 21,22,24,26,27,28,30,31,33,36, 37,38,39,41,42,43,44,45,46,47,48, 52,54)	87.37
0.79 To 0.87	0.0	56	50(all except 1,3,5,10,13,20)	81.25
	0.1 to 1	56	53(all except 3,5,13)	82.37
0.88	0.0	56	52(all except 1,3,5,13)	71.25
	0.1	56	6(1,2,4,6,24,49)	70.87
	0.2 to 1	56	1(1)	65.62
0.89 to 1	0.0	56	50(all except 1,3,5,10,13,20)	80.25
	0.1 to 1	56	2(1,2)	65.62

For Hepatitis dataset, the impact of the parameters is in a reverse order to that of Heart C. ACO gives lesser accuracy for the value of ρ from 0 to 0.73 and the accuracy gets increased for the value of ρ higher than 0.73. β has the same effect for all the values in the range [0, 1]. ϕ affects the optimization process for lower values of ρ and produces the same result for the higher values of ρ.

Table 5. DERMATOLOGY- ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy (%)
0.1 TO 0.77	0.0 to 0.4	34	28(all except 4,7,11,13,20,34)	96.90
	0.5 to 1	34	27(all except 4,7,9,11,13,20,34)	98.35
0.78 To 1	0.0	34	33(all except 13)	95.90
	0.1	34	33(all except 13)	94.5
	0.2 to 1	34	1(1)	35.79

Table 6. PIMA- ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy (%)
0.1 TO 0.77	0.0 to 1	8	6(except 5,7)	81.11
0.78	0.0 To 0.2	8	6(except 5,7)	81.11
	0.3 To 1	8	6(except 3,4)	89.82
0.79 To 0.81	0.0 To 0.2	8	6(except 5,7)	81.11
	0.3 and 0.4	8	6(all except 4,8)	74.73
	0.5 To 1	8	1(1)	67.83
0.82 To 1	0.0	8	8(all)	80.11
	0.1	8	6(all except 4,8)	74.73
	0.2 To 1	8	1(1)	67.83

Table 7. LYMPHOGRAPHY – ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Seleted	Accuracy
0.1 to 0.74	0.0	18	1(2)	70.94
	0.1 To 1	18	16(all except 1,13)	78.35
0.75 to 1	0.0 to 1	18	1(2)	70.94

Table 8. LIVER - ACO based Feature Selection and Classification Accuracy

Evaporation Rate ρ	ϕ	Actual Features	Features Selected	Accuracy(%)
0.0 to 0.78	0.0 to 1	6	4(except 4,6)	88.63
0.79 To 1	0.0 to 0.3	6	4(except 4,5)	84.63
	0.4 to 1	6	1(1)	57.97

Table 9. WISCONSIN **-** ACO based Feature Selection and Classification Accuracy

Evaporation Rate (ρ)	ϕ	Actual Features	Features Selected	Accuracy(%)
0.0 To 0.71	0.0 to 1	9	4(except 4,6)	87.65
0.72 To 1	0.0 to 0.3	9	4(except 4,5)	83.0
	0.4 to 1	9	1(1)	63.29

Table 10. HIV **-** ACO based Feature Selection and Classification Accuracy

Evaporation rate (ρ)	ϕ	Actual features	Features seleted	Accuracy(%)
0.1 To 0.78	0.0 to 0.7	21	all	77.25
	0.8 To 1	21	7(2,4,9,10,11,12,19)	85.65
0.79 To 1	0.0 to 1	21	all	77.25

Table 11. DIABETES - ACO based Feature Selection and Classification Accuracy

Evaporation Rate (ρ)	ϕ	Actual features	Features seleted	Accuracy
0.0 to 0.69	0.0 to 1	8	8(all)	73.82
0.7 To 0.74	0.0 to 0.3	8	6(except 4,8)	84.11
	0.4	8	7(all except 8)	73.95
	0.5 To 1	8	8(all)	73.82
0.75 To 1	0.0 To 0.1	8	8(all)	73.82
	0.2	8	6(all except 4,8)	74.73
	0.3 To 1	8	7(all except 8)	73.95

From the tables 2,3,4,5,6,7,8,9,10 and 11 it can be seen that, except Hepatitis, for all other datasets, the feature selection is optimal and the accuracy is higher, when the evaporation rate is set to 0.75 and below it. The local pheromone update factor has only a little significance and performs its best when set to the values in the range [0.3, 0.5]. The relative factor β has the same effect on optimization for all the values in between [0, 1].

From the data represented in Tables 2 to 11, it can be inferred that,

a. The relative factor β affects the optimization process in the same way for all the values in the range [0, 1].
b. The local pheromone update factor ϕ, shows varied performance based on the values it is assigned to. However, the results suggest ϕ gives best results when it is assigned values in between 0.3 to 0.5.
c. When the PER ρ is assigned values around 0.7, it leads to best optimization and higher classification accuracy.
d. Except for Hepatitis dataset, the accuracy increases for the values of PER from 0.1 to 0.7(roughly) and the accuracy starts decreasing when PER takes values from 0.75 to 1.
e. When ACO is applied for optimization of feature selection, the parameters can be set to the optimal values suggested by this experiment.
f. The optimal values are 0.3 to 0.5 for LPU, 0.7 to 0.75 for PER. β can take any value from 0 to 1.

6 Conclusion

ACO has been widely employed to solve combinatorial optimization problems. Literature has proved FS implemented using ACO has always resulted in optimal feature selection and better classification accuracy. However, setting the values of parameters required in ACO mechanism is usually a time consuming process. These parameters are usually set by running the trial and error on all possible values and finally selecting the numbers that yield best results. In this work, ACO in combination with a classifier is employed for optimal selection of features. We have experimented by assigning all the allowed values to the ACO parameters using 10 different data sets and arrived at optimal values for these parameters within the allowed range of values.

References

1. Dorigo, M., Stutzle, T.: Ant Colony Optimization. The MIT Press, Massachusetts (2004)
2. Rezaee, M.R., Goedhart, B., Lelieveldt, B.P.F., Reiber, J.H.C.: Fuzzy feature selection. Pattern Recognition 32, 2011–2019 (1999)
3. Laura Santana, E.A., Ligia Silva Anne Canuto, M.P., Pintro, F., Vale, K.O.: A Comparative Analysis of Genetic Algorithm and Ant Colony Optimization to Select Attributes for a Heterogeneous Ensemble of Classifiers, pp. 465–472. IEEE (2010)
4. Ahmed, E.F., Yang, W.J.M., Abdullah, M.Y.: Novel method of the combination of forecasts based on rough sets. Journal of Computer Science 5, 440–444 (2009)
5. Sadeghzadeh, M., Teshnehlab, M.: Correlation-based Feature Selection using Ant Colony Optimization. World Academy of Science, Engineering and Technology 64, 497–502 (2010)
6. Dorigo, M., Di, C.G., Gambardella, L.M.: Ant algorithms for discrete optimization. Artificial Life 5, 137–172 (1999)
7. Abd-Alsabour, N., Randall, M.: Feature Selection for Classification Using an Ant Colony System. In: Sixth IEEE International Conference on e–Science Workshops, pp. 86–91 (2010); Kuncheva, L.I.: Combining Pattern Classifiers, Methods and Algorithms. Wiley Interscience (2005)
8. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Recognition, 2nd edn. John Wiley & Sons, Inc. (2001)
9. Molina, L.C., Belanche, L., Nebot, À.: Feature Selection Algorithms: A Survey and Experimental Evaluation. In: Second IEEE International Conference on Data Mining, pp. 155–172 (2002)
10. Randall, M.: Near Parameter Free Ant Colony Optimisation. In: Dorigo, M., Birattari, M., Blum, C., Gambardella, L.M., Mondada, F., Stützle, T. (eds.) ANTS 2004. LNCS, vol. 3172, pp. 374–381. Springer, Heidelberg (2004)
11. Frank, A., Asuncion, A.: UCI Machine Learning Repository. University of California, School of Information and Computer Science, Irvine, CA (2010), `http://archive.ics.uci.edu/ml`
12. WEKA: A Java Machine Learning Package, `http://www.cs.waikato.ac.nz/~ml/weka/`
13. Ridge, E., Kudenko, D.: Screening the Parameters Affecting Heuristic Performance. In: Preceeding of Genetic and Evolutionary Computation, GECCO 2007, p. 180. ACM (2007)
14. Matthews, D.C.: Improved Lower Limits for Pheromone Trails in Ant Colony Optimization. In: Rudolph, G., Jansen, T., Lucas, S., Poloni, C., Beume, N. (eds.) PPSN 2008. LNCS, vol. 5199, pp. 508–517. Springer, Heidelberg (2008)
15. Stutzle, T., López-Ibáñez, M., Pellegrini, P., Maur, M., de Oca, M.M., Birattari, M., Dorigo, M.: Parameter Adaptation in Ant Colony Optimization, IRIDIA Technical Report Series Technical Report No. TR/IRIDIA/2010-002 (2010)
16. Kumar, P.: A Note on the Parameter of Evaporation in the Ant Colony Optimization Algorithm. International Mathematical Forum 6(34), 1655–1659 (2011)
17. Ivković, N.: Investigating MAX-MIN_ Ant System Parameter Space
18. Dobslaw, F.: A Parameter-Tuning Framework for Metaheuristics Based on Design of Experiments and Artificial Neural Networks. World Academy of Science, Engineering and Technology 64, 213–216 (2010)
19. Pellegrini, P., Favaretto, D., Moretti, E.: On MAX-MIN Ant System's parameters

20. Sivagaminathan, R.K., Ramakrishnan, S.: A hybrid approach for feature subset selection using neural networks and ant colony optimization. Expert Systems with Applications 33, 49–60 (2007)
21. Aghdam, M.H., Ghasem-Aghaee, N., Basiri, M.E.: Text feature selection using ant colony optimization. Expert Systems with Applications 36, 6843–6853 (2009)
22. Al-Ani, A.: Feature Subset Selection Using Ant Colony Optimization. International Journal of Computational Intelligence 2(1), 53–58 (2005)
23. Al-Ani, A.: Ant Colony Optimization for Feature Subset Selection. World Academy of Science, Engineering and Technology 4, 35–38 (2005)
24. He, Y., Chen, D., Zhao, W.: Ensemble classifier system based on ant colony algorithm and its application in chemical pattern classification. Chemo Metrics and Intelligent Laboratory Systems, 39–49 (2006)
25. Robbins, K., Zhang, W., Bertrand, J.: The ant colony algorithm for feature selection in high-dimension gene expression data for disease classification. Mathematical Medicine and Biology, 413–426 (2007)
26. Kanan, H., Faez, K.: An improved feature selection method based on ant colony optimization (ACO) evaluated on face recognition system. Applied Mathematics and Computation, 716–725 (2008)
27. Chandra, A., Yao, X.: Ensemble learning using multi-objective evolutionary algorithm. Journal of Mathematical Modeling and Algorithms 5(4), 417–445 (2006)