J.Natn.Sci.Foundation Sri Lanka 2018 46 (2): 143 - 151
DOI: http://dx.doi.org/10.4038/jnsfsr.v46i2.8414

RESEARCH ARTICLE

An optimal feature selection method using a modified wrapper-based ant colony optimisation

Saba Jameel and Saif Ur Rehman*
University Institute of Information Technology, PMAS Arid Agriculture University, Rawalpindi, Pakistan.

Revised: 25 October 2017; Accepted: 16 November 2017

Abstract: In feature selection, applications often require very high-dimensional data. Feature selection algorithms are therefore designed to identify the relevant feature subset from the original features, which can facilitate subsequent analysis such as classification and clustering. Also, feature reduction helps to reduce dataset dimensionality, lessen the running time, and/or improve the prediction accuracy. In this paper, a new wrapper-based feature selection approach is proposed based on ant colony optimisation (ACO). In the proposed approach, an ACO search environment is built and every ant probabilistically selects attributes depending on the pheromone and heuristic values linked with every edge. Furthermore, a heuristic function is used along with the values of pheromone for the selection of the ideal attribute subset. Naïve Bayes classifier is used to compute the fitness of each selected feature subset. The computed classification accuracy from naïve Bayes classifier is used as a fitness function. Different datasets are used for the experimental evaluation of the proposed approach. The experimental results of the proposed technique are very promising. The proposed technique increased accuracy by 5 % on average during experimentation on all datasets used when the subset feature selection is performed. Moreover, in 9 out of 15 datasets, the accuracy is improved when the feature subsets are selected using the proposed technique and the existing genetic search technique.

Keywords: Ant colony optimisation, feature selection, symmetric uncertainty, wrapper method.

INTRODUCTION

In the past few years the rapid growth of technologies resulted in the production of large amount of data. In machine learning it is treated as a tricky issue and also a challenge for classification (Kashef & Nezamabadi-pour, 2015). Ideally, having more features suggests more discriminating influence in classification (Duda *et al*., 2001). This is not generally valid in practical experience, since not all the features present in high dimensional datasets facilitate the prediction of target class. A large portion of the features in the datasets are redundant and irrelevant to the model and may negatively affect the prediction precision (Liu & Yu, 2005). Feature selection technique is commonly used to deal with this problem. Feature subset selection is considered as an important research problem in knowledge discovery process, not only for the insight gained to find relevant modelling variables but also for the enhanced understandability, scalability, and possibly, accuracy of the resulting models. The process of feature selection is to select a subset of features from the whole dataset by removing redundant features, when two or greater than two features hold the same predictive information and also irrelevant features, the features have no impact on class label. In general, the result of feature reduction is for classifiers to come across small and clear information, which helps to save long computational time (Uysal & Gunal, 2012).

To find out the optimal feature subset, feature selection requires a searching technique that explores all possible feature subset space. The whole search space containing all possible subsets of features have size 2n where n is the total number of features. Practically this technique is not feasible even for a small number of features because

* Corresponding author (saif@uaar.edu.pk; https://orcid.org/0000-0002-5810-6479)

of its computational complexity. Therefore, feature selection is a discrete optimisation problem (Markid *et al*., 2015a). To avoid this difficulty heuristic search strategies [i.e. genetic algorithm, simulated annealing, ant colony optimisation (ACO) etc.] are proposed. The feature selection strategies can be divided into four groups: filter, wrapper, embedded and hybrid methods (Saeys *et al*., 2007).

In filter methods statistical measures are used to calculate the relevance of features. A filter method can successfully eliminate irrelevant attributes without using any learning algorithm (Tabakhi & Moradi, 2015). Filter models are computationally inexpensive (Markid *et al*., 2015a).

In wrapper methods, to calculate the worth of a feature subset, a learning algorithm is used besides some search strategy. They are computationally expensive as compared to filter models but regarded as better when concerning classification accuracy (Yu & Liu, 2003).

In embedded methods, feature selection is integrated as a part of the learning process. These strategies utilise the entire dataset and do not split the dataset into test and training sets. They find the optimal subset of features faster as compared to wrapper models because in this technique each selected subset is not evaluated like in the wrapper method (Markid *et al*., 2015a).

Some researchers have combined the filter and wrapper approaches for feature selection process in successive steps (Tabakhi *et al*., 2014). This approach is termed as hybrid feature selection scheme. In the hybrid scheme, initially, the filter method is applied to select a significant subset of features and from this significant subset of features the optimal feature set is filtered out using the wrapper method (Unler *et al*., 2011).

More recently, swarm-based strategies are earning fame due to their superior performance to solve feature selection problem. Among swarm-based approaches to feature selection, ACO is a promising way to deal with combinational optimisation problem and has been extensively used in feature selection (Al-Ani, 2005; Aghdam *et al*., 2009). ACO has some advantages: it follows nature in a parallel way, provides multi agent structure, positive feedback and has good local and global search ability.

Different feature selection methods based on ACO have been reported in recent years (Markid *et al*., 2015b). Among these approaches, binary ACO for feature selection was suggested for the first time (Touhidi *et al*., 2007). In this approach, all of the features are visited by ants and are allowed to either select it or not. In their approach, the ants can only analyse the successive feature and are unable to travel in any desired sequence of attributes. This limitation decreases the search exploration that results from non-optimal solutions.

In 2012, another feature selection scheme was introduced by Abd-Alsabour *et al*. (2012). In their work, the impact of fixing the length of selected subsets of features by using ACO for feature selection was studied (Abd-Alsabour *et al*., 2012). Extensive experiments were performed to check the impact of fixing the number of features with varying size of the features. The results revealed that varying the length of selected subsets of attributes delivered better performance as compared to selected subset of attributes with fixed length with regard to classification accuracy.

Chen *et al*. (2013) presented an ACO based algorithm, which is called ant colony optimisation feature selection (ACOFS). They used a binary strategy. In this algorithm, the features are contained in a sequence with two arcs between every feature (node) and its subsequent one. The well-known F- score is used as a heuristic function and competitive results were obtained in this approach. But, due to the topology of ACOFS any type of statistical information cannot be used and the final subset will not facilitate to eliminate redundant features.

Markid *et al*. (2015a) proposed bidirectional ant colony optimisation feature selection (BDACOFS). BDACOFS was motivated from the earlier work, ACOFS (Chen *et al*., 2013). The adopted binary strategy makes it more flexible, so that the BDACOFS can use statistical information. Markid *et al*. (2015a) designed the topology of the proposed algorithm as a circular graph. Two factors were used to calculate the heuristic desirability F-score and mutual information. From the preliminary experimentations, authors have concluded that their technique BDACOFS performed well and removed redundant features as compared to ACOFS.

Tabakhi *et al*. (2014) proposed a filter strategy using an unsupervised feature selection method based on ACO algorithm, called UFSACO. The UFSACO executes an implicit relevance and explicit redundancy analysis. Although UFSACO performs well in some cases the relevance of attributes in the dataset cannot be decided without redundancy among attributes. Therefore, UFSACO cannot remove irrelevant attributes from the dataset (Tabakhi *et al*., 2014).

In 2015, a new feature subset selection approach was proposed by Khan and Baig (2015). In their technique, for the computation of the heuristic desirability two well-known measures were used. These two measures are minimum redundancy and maximum relevance (mRMR). The investigational results demonstrated that the proposed strategy outperforms particle swarm optimisation (PSO) and genetic algorithm (GA) based feature subset selection. Furthermore, they showed that ACO is a robust method for feature subset selection as compared to PSO or GA based feature selection schemes. The performance of the proposed approach may be enhanced by utilising different measures of relevance (e.g. MI) rather than mRMR.

Sequence based feature selection (SFS) is another feature selection (Markid *et al*., 2015b). This algorithm is based on ACO approach. This method uses a completely linked graph of nodes to represent the problem and puts the pheromone on edges instead of nodes. Each ant should visit all the nodes and at the end of iteration every ant has a sequence of nodes. These sequences were passed to the next step for selecting a subsequence of it and considered as candidate solutions. SFS technique expands the quantity of possible solutions (for assessment) but the quantity also increases the computational time of their algorithm.

Maximum margin feature selection (MMRFS) is a wrapper method for feature subset selection. This method was proposed by Cheng *et al*. (2007). In the method, MMRFS uses information gain to weigh the correlation between each of the dataset features and class labels, and then selects features with less redundancy covering new training samples.

Guyon *et al*. (2002) proposed a support vector machine recursive feature elimination (SVM_RFE). SVM_RFE starts by selecting the dataset features *via* a greedy backward feature elimination. This technique initially builds a linear classifier, then uses the weight vector of the hyperplane constructed by the training samples in order to rank the features. During each iteration, lower ranked features are removed and a new hyperplane is constructed and so on. The limitation of this technique is that it works only with linear kernel (Guyon *et al*., 2002).

In 2016, Gu *et al*. proposed a cat swarm optimisation (CSO) based method for handling high dimensional feature subset selection. In their work, the CSO is modified to be suited for combinatorial optimisation. The modified CSO based variant is embedded in a wrapper feature selection approach. Gu *et al*. (2016) concluded that the number of new solutions found in terms of feature subset by CSO-kNN is considerably larger than that was found by the PSO-based methods. Unlike the PSO-based method where the final optimal results heavily rely on the initialisation in terms of the number of selected features, the purposed CSO method performs consistently well and is less sensitive to initialisation.

Recently in 2016, Wang *et al*. introduced a new rough set model called fuzzy neighbourhood rough set. This model decreased the possibility that an instance belongs to multiple classes. The dependency between fuzzy decision and condition features is defined. Using this dependency, the significance of a candidate feature is evaluated. The experimental results showed that the algorithm can find a small and effective subset of features and obtain high classification accuracy. The authors also found that the two parameters have great impact on the performance of the proposed attribute reduction algorithm. Selection of suitable values of parameters for each dataset according to the numbers of selected features and classification accuracies is an important issue of this technique.

Li and Oh (2016) proposed a new method to improve the quality of feature selection. They disclosed that the information about the interactions between two given features is very helpful for improving the original feature selection algorithms. In their study, they used the classification accuracy as an evaluation measure for interaction but they also added that evaluation measure could be changed if the aim of feature selection is other than classification. The proposed method does not include redundancy among its features (Li & Oh, 2016).

Various methods have been proposed for feature selection based on ACO to eliminate insignificant features from the dataset. However, from the above detailed survey of the different feature selection techniques, it can be summarised that still there is room to search for an optimised feature subset without degrading classification accuracy and reducing the dimension of datasets.

An important issue in feature selection is the development of a method, which can select a high quality feature subset and eliminate both redundant and irrelevant features from the dataset under observation. In this paper, a new wrapper-based feature selection approach using ACO is proposed for the feature subset section. In the proposed algorithm, ACO is adopted to guide the search procedure and to remove the irrelevant as well as redundant features from the datasets. Symmetrical uncertainty is used as a heuristic function. Naïve Bayes classifier performance is used to evaluate the significance of each selected subsets of features.

METHODOLOGY

Feature selection is a discrete optimisation problem (Markid *et al*., 2015a), and therefore the optimisation capability of ACO is utilised for the feature selection. In the proposed technique, ACO is used as a population-based mechanism for the selection of feature subset. The wrapper-based methods primarily used the population-based selection of feature subset, in which a learning algorithm is employed to evaluate the usefulness of selected subsets of features (Ali & Shahzad, 2012).

Ant colony optimisation

In 1996, Dorigo *et al*. embraced this idea and suggested an 'artificial colony of ant's algorithm', which was known as the ant colony optimisation. ACO is a meta-heuristic algorithm that was motivated by the seeking activities of ants (Kashef & Nezamabadi-pour, 2015). In ACO, ants act as agents and each ant constructs its own candidate solution. The solution is constructed by using a probability function based on the pheromone value and desirability value. On the completion of each iteration, all the candidate solutions are evaluated and the pheromones on visited paths are updated. In successive iterations, optimised solution was obtained. The main deliberation for applying ACO to feature selection problem is as follows:

ACO's search space

Search space is one of the main critical factors, when one attempts to obtain improved results from ACO based algorithm. ACO based algorithm mostly utilised the complete graph with *m-nodes* where *m* is the total number of attributes (Markid *et al*., 2015a). The other is binary strategy where attributes are in a sequence with two edges between each node and its successor (Chen *et al*., 2013). However, the complexity of edges among the nodes has been reduced in binary approach, but there is no possibility to apply any sort of mutual statistical information between nodes (Haindl *et al*., 2006) as heuristic function, because of the absence of full association between nodes of graph. Therefore, redundant features cannot be eliminated in this binary approach.

In the proposed approach, the complete graph is used as a search space. It is an M*M graph; M is the total number of features except for target feature. Each feature is represented through a node. The start node is selected by every ant through guided random selection; for this purpose information gain (IG) is used. The search space also consists of a terminal node. The terminal node is utilised to end the search and is associated with every node in the graph. When a terminal node is selected by an ant, it does not include further nodes and its trail is regarded as complete. In the proposed algorithm, the number of features to be selected by an ant is not predefined (limitation of ACO approach), it selects the feature subsets arbitrarily.

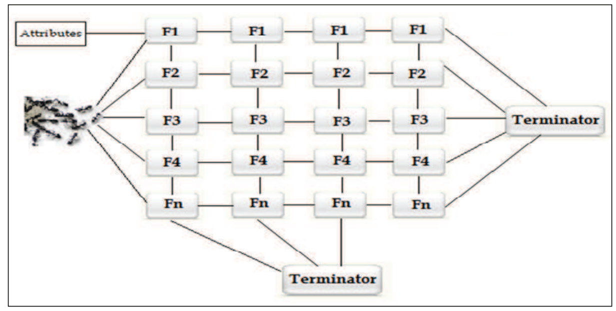

Figure 1: Search space of ACO for ant's traversal

Attribute's selection

After selecting the first attribute/node to move from the present attribute to the next, an ant utilises two factors to compute the selection probability of subsequent attributes. The first factor in computing the selection probability is the pheromone amount that is associated with the edges between the nodes. The second factor is the heuristic function that illustrates the significance of an attribute. If an ant is at node *l,* the probability of selecting the next node as *m* is computed through the probability equation. It is essential that the node *m* has not been visited up till now. This probability is computed by using the following equation (Khan & Baig, 2015):

$$\rho_{ij} = \frac{[\tau_{lm}]^{\alpha}.[\eta_{lm}]^{\beta}}{\sum_{K\in S}[\tau_{lm}]^{\alpha}.[\eta_{lm}]^{\beta}} \quad ...(1)$$

In equation (1) τ_{lm} represents the pheromone amount among node *l* and node *m*, and η_{lm} is the heuristic value. The influencing parameters are α and β, which influence the value of pheromone and heuristic, respectively.

Heuristic function

In the proposed algorithm, to calculate the heuristic function, symmetrical uncertainty (SU) is used. Symmetry is a desired property for measuring correlations among attributes. SU has various advantages i.e. its nature is symmetric hence SU (i, j) is same as that of SU (j, i), consequently it decreases the required number of comparisons. It is not affected by multi valued features as the information gain, and its values are normalised in the range [0, 1]. SU is calculated using equation (2) (Ali & Shahzad, 2012).

$$SU(X,Y) = 2\left[\frac{IG\ (X/Y)}{H(X)+H\ (Y)}\right] \quad ...(2)$$

IG(X/Y) represents information gain of attributes *X* and *Y*. The entropy of attribute *X* is represented by *H(X)* and the entropy of attribute *Y* is *H(Y)*. If the value of *SU(X, Y)* is 1, it indicates that *X* and *Y* are absolutely dependant; a 0 value suggest that A and B are completely independent. The proposed algorithm calculates the SU between all the attributes.

Fitness function

To discover the importance of a particular subset of attributes fitness function is used. In the proposed algorithm, the classification accuracy is used as a fitness function. When a specific subset is constructed by an ant, we retain this particular subset of attributes and to evaluate its worth we run the classifier [naïve Bayes (NB) is used for this purpose]. The fitness of the specific subset of attribute is calculated through equation (3) (Markid *et al*., 2015a).

$$fitness = \frac{\varphi}{M} \quad ...(3)$$

where φ represents the quantity of example correctly classified by the classification algorithm, and M is the aggregate amount of test cases. Fitness is computed for every fold and then averaged.

Pheromone update

When every ant completes their visit, the pheromone values are updated, so for search, future ants can utilise this information. The values of pheromone are updated through equation (4)

$$\tau_{lm}(t+1) = \tau_{lm}(t) * Q * \lambda \quad ...(4)$$

The value of pheromone among node *l* and *m* in the current iteration is represented by $\tau_{lm}(t)$ and Q represents the quality of the solution; quality depends on the value of fitness received from the current iteration. The aim of the pheromone upgrade is to boost the value of pheromone connected with the best solution, and to diminish those that are not connected with the best ones generally. This is accomplished by reducing the pheromone values through evaporation. In this way for the future ants the best attributes subset will become more appealing and have more chances of selection. To avoid premature convergence a moderation factor λ is used in equation (4).

The following equation (5) is used to evaporate pheromone value (Ali & Shahzad, 2012).

$$\tau_{lm}(t) = \tau_{lm}(t)(1-\rho) \quad ...(5)$$

The value of pheromone among node *l* and *m* in the current iteration is represented by $\tau_{lm}(t)$ and ρ is the pheromone evaporation rate.

Proposed algorithm

Input: (**D**: dataset, **C:** maximum number of allowed iterations, **A:** define the number of ants. $\mathbf{S_c}$**:** current subset, $\mathbf{S_g}$**:** global best subset)
Output: Subset // an optimal subset

1. **begin algorithm**
2. Compute the information gain $IG_{i,}$ of each feature with target feature, i= 1.....n.
3. Compute the symmetrical uncertainty (SU) using equation 2, $S_{i,j}$ between features, i, j = 1.....n.
4. Initialise the intensity of pheromone associated with the features, i, j = 1.....n
5. Initialise ACO parameters
6. Generate a population of ants.
7. ***for*** *t=1 to C* ***do***
8.   ***for*** *K= 1 to A* ***do***
9.   *Generate a subset of feature S.*
10.   *Evaluate each feature subset S, using the computed value of SU between features*
11.   ***if*** *($S_c > S_g$)*
12.   *$S_g = S_c$*
13.   ***end if***
14.   ***end for***
15.   *Update pheromone according to the pheromone updating rule from equation (4) and goto Line No. 7*
16. ***end for***
17. Report best feature subset, S_g
18. **end algorithm**

RESULTS AND DISCUSSION

The proposed feature selection approach was evaluated using fifteen different datasets from the UCI machine learning repository (Hettich & Bay, 1996). These datasets included common and benchmark datasets used for feature selection performance evaluation. The primary statistical details about the datasets are given in Table 1.

The proposed technique is implemented in Microsoft Visual C#.Net using 4.0 .Net framework. The experiments were conducted on HP Core i5-2430M CPU @ 2.40 GHz, 6Gb RAM running Windows 10, 64-bit. 10-fold cross-validation is used for all of the experiments. In this cross-validation, the datasets are divided randomly into 10 equal sized, mutually exclusive subsets. Each of the subsets is used once for testing and the rest of 9 are utilised for training. The parameters used in the experiments incorporate pheromone evaporation rate ρ - 0.09; maximum iteration of the proposed algorithm - 100; moderation factor used in pheromone update - 0.1; number of ants in a generation - 15; values of α and β, which influence the value of pheromone and heuristic, respectively - 1.

Table 1: Dataset statistics used

Datasets	No. of attributes	No. of instances	No. of classes
Tic-tac-toe	9	958	2
Iris2	4	150	3
Vehicle	18	848	4
Diabetes	8	768	2
Breast-cancer-wisconsin-MD	10	699	2
Soybean-large_MD	35	307	4
Zoo	17	101	7
Glass	9	214	6
Hepatitis_MD	19	155	2
Lenses_MD	5	24	3
Winequality-red-MD	11	1599	6
SPECT	22	267	2
Heart	13	270	2
German	20	1000	2
Blood-transfusion	4	748	2

Table 2: Accuracy calculated with all features and after selecting optimal feature subset by applying feature selection with proposed optimal feature selection using ACO

	Without attribute subset selection	Feature selection using genetic search	Feature subset selection (ACO)
Datasets		Accuracy	
Tic-tac-toe	69.83	73.06	74.21
Iris2	86.67	92.00	92.00
Vehicle	57.09	58.51	60.17
Diabetes	72.91	75.78	75.78
Breast-cancer-wisconsin-MD	96.57	97.28	97.28
Soybean-large_MD	97.72	98.33	99.02
Zoo	41.58	98.03	97.03
Glass	66.82	64.01	66.82
Hepatitis_MD	65.16	73.71	74.19
Lenses_MD	37.50	88.00	87.50
Winequality-red-MD	56.84	60.97	61.10
SPECT	68.16	70.03	75.65
Heart	80.74	84.81	85.92
German	68.90	75.60	76.40
Blood-transfusion	73.26	77.40	76.07

Based on the above parameter settings, the classification accuracy is computed using naïve Bayes classifier for all the datasets with entire set of features. Furthermore, using the proposed algorithm optimal feature subset is selected and on this reduced feature subset the classification accuracy is calculated. Feature selection using the proposed optimal ACO's accuracy has been compared with the feature selection based on genetic search by using naïve Bayes classifiers. This method is already implemented in WEKA, which is a machine learning open source software package. The proposed optimal ACO based algorithm achieved better performance on nine datasets out of fifteen. The accuracy values of (a) without feature subset selection; (b) with attribute subset selection and (c) feature subset selection using genetic search against each dataset is shown in Table 2.

During the experimentations on the different datasets, the proposed ACO based feature selection technique results in better accuracy as compared to the accuracy computed using without feature subset selection. The maximum accuracy is achieved on Zoo and Lenses_MD datasets, which is 55 % and 50 % improved, respectively. Overall, on average more than 5 % accuracy is improved using the proposed approach. Moreover, when the proposed ACO based feature subset selection technique is compared with the results obtained from feature selection using genetic search, it was observed that the proposed technique produce better accuracy in nine out of 15 datasets under observation.

In addition to the accuracy, the number of features reduced are also demonstrated. For the experimentation, two competing feature subset selection techniques are considered. These two techniques are ACOFS (Chen *et al*., 2013) and UFSACO (Tabakhi *et al*., 2014). The source code of ACOFS was obtained from the authors and the UFSACO code is re-implemented. The datasets discussed in Table 1 are considered for evaluation. During the experimentation, the proposed ACO technique performed better in most of the cases and it has reduced more number of attributes than the competing ACOFS and UFSACO.

In Table 3, the number of features reduced by the proposed ACO based feature selection and the two competing feature subset selection techniques is given. In most of the cases, the proposed ACO based technique performance was good. Best performance of the proposed ACO is on Breast-cancer-wisconsin-MD dataset in which 31 irrelevant features are reduced whereas on the same dataset UFSACO reduced only 19 features and ACOFS reduced only 22 attributes. Only on the Glass dataset, the proposed ACO based technique has not removed any irrelevant feature but ACOFS has reduced two irrelevant features. However, overall the performance of the proposed feature subset selection technique is more accurate as compared to the other competing feature subset selection techniques.

Table 3: Number of reduced features after applying feature selection using the proposed ACO scheme

Datasets	Total features	No. of features reduced		
		Proposed ACO based technique	UFSACO Tabakhi *et al*., 2014	ACOFS Chen *et al*., 2013
Tic-tac-toe	9	4	2	3
Iris2	4	2	0	1
Vehicle	18	9	5	6
Diabetes	8	5	2	1
Breast-cancer-wisconsin-MD	10	5	3	2
Soybean-large_MD	35	31	19	23
Zoo	17	12	7	5
Glass	9	1	0	2
Hepatitis_MD	19	16	10	14
Lenses_MD	5	3	1	2
Winequality-red-MD	11	6	4	3
SPECT	22	18	15	16
Heart	13	7	5	6
German	20	13	9	14

CONCLUSION

In this paper, a new optimal wrapper-based feature selection technique is proposed. The new proposed scheme combines both the accuracy of the wrapper feature selection model and adequate performance of the well-known ACO algorithm. The proposed ACO based feature selection scheme performance reduced the dataset dimensionality without compromising the prediction accuracy of the classifier.

REFERENCES

1. Abd-Alsabour N., Randall M. & Lewis A. (2012). Investigating the effect of fixing the subset length using ant colony optimization algorithms for feature subset selection problems. *2012 13th International Conference on Parallel and Distributed Computing, Applications and Technologies*, Beijing, China, 14 − 16 December, pp. 733 − 738. DOI: https://doi.org/10.1109/PDCAT.2012.84
2. Aghdam M.H., Ghasem-Aghaee N. & Basiri M.E. (2009). Text feature selection using ant colony optimization. *Expert Systems with Applications* **36**(3): 6843 − 6853. DOI: https://doi.org/10.1016/j.eswa.2008.08.022
3. Al-Ani A. (2005). Feature subset selection using ant colony optimization. *International Journal of Computational Intelligence* **2**(1): 53 − 58.
4. Ali S.I. & Shahzad W. (2012). A feature subset selection method based on symmetric uncertainty and ant colony optimization*, 2012 International Conference on Emerging Technologies (ICET)*, Islamabad, Pakistan, 8 − 9 October, pp. 1 − 6. DOI: https://doi.org/10.1109/ICET.2012.6375420
5. Chen B., Chen L. & Chen Y. (2013). Efficient ant colony optimization for image feature selection. *Signal Processing* **93**(6): 1566 − 1576. DOI: https://doi.org/10.1016/j.sigpro.2012.10.022
6. Cheng H., Yan X., Han J. & Hsu C.-W. (2007). Discriminative frequent pattern analysis for effective classification. *Proceedings of the 23rd IEEE International Conference on Data Engineering (ICDE)*, Istanbul, Turkey, 15 − 20 April, pp. 716 − 725. DOI: https://doi.org/10.1109/ICDE.2007.367917
7. Dorigo M., Maniezzo V. & Colorni A. (1996). Ant system: optimization by a colony of cooperating agents. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* **26**(1): 29 − 41. DOI: https://doi.org/10.1109/3477.484436
8. Duda R., Hart P. & Stork D. (2001). *Pattern Classification.* John Wiley & Sons, New Jersey, USA.
9. Gu S., Cheng R. & Jin Y. (2016). Feature selection for high-dimensional classification using a competitive swarm optimizer. *Soft Computing* **22**(3): 811 − 822.
10. Guyon I., Weston J., Barnhill S. & Vapnik V. (2002). Gene selection for cancer classification using support vector machines. *Machine Learning* **46**: 389 − 422. DOI: https://doi.org/10.1023/A:1012487302797
11. Haindl M., Somol P., Ververidis D. & Kotropoulos C. (2006). Feature selection based on mutual correlation. *Progress in Pattern Recognition, Image Analysis and Applications* (eds. J.F. Martinez-Trinidad, J.A. Carrasco Ochoa & J. Kittler) CIARP 2006. Lecture Notes in Computer Science, volume 4225, pp. 569 − 577. Springer, Berlin, Heidelberg, Germany. DOI: https://doi.org/10.1007/11892755_59
12. Hettich S. & Bay S.D. (1996). *The UCI KDD Archive.* Department of Information and Computer Science, University of California, USA.
13. John G.H., Kohavi R. & Pfleger K. (1994). Irrelevant features and the subset selection problem. *Machine Learning: Proceedings of the Eleventh International Conference*, Rutgers University, New Brunswick, USA, 10 − 13 July, pp. 121 − 129. DOI: https://doi.org/10.1016/B978-1-55860-335-6.50023-4
14. Kashef S. & Nezamabadi-pour H. (2015). An advanced ACO algorithm for feature subset selection. *Neurocomputing* **147**: 271 − 279. DOI: https://doi.org/10.1016/j.neucom.2014.06.067
15. Khan A. & Baig A.R. (2015). Multi-objective feature subset selection using mRMR based enhanced ant colony optimization algorithm (mRMR-EACO). *Journal of Experimental and Theoretical Artificial Intelligence* **28**(6): 1 − 13.
16. Li S. & Oh S. (2016). Improving feature selection performance using pairwise pre-evaluation. *BMC Bioinformatics* **17**(1): 312. DOI: https://doi.org/10.1186/s12859-016-1178-3
17. Liu H. & Yu L. (2005). Toward integrating feature selection algorithms for classification and clustering. *IEEE Transactions on Knowledge and Data Engineering* **17**(4): 491 − 502. DOI: https://doi.org/10.1109/TKDE.2005.66
18. Markid H.Y., Dadaneh B.Z. & Moghaddam M.E. (2015a). Bidirectional ant colony optimization for feature selection. *2015 International Symposium on Artificial Intelligence and Signal Processing (AISP),* Mashhad, Iran, 3 − 5 March, pp. 53 − 58. DOI: https://doi.org/10.1109/AISP.2015.7123519
19. Markid H.Y., Dadaneh B.Z. & Moghaddam M.E. (2015b). Sequence based feature selection using ant colony optimization. *2015 5th International Conference on Computer and Knowledge Engineering (ICCKE)*, Mashhad, Iran, 29 October, pp. 100 − 105. DOI: https://doi.org/10.1109/ICCKE.2015.7365867
20. Saeys Y., Inza I. & Larrañaga P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics* **23**(19): 2507 − 2517.
21. Tabakhi S. & Moradi P. (2015). Relevance–redundancy feature selection based on ant colony optimization. *Pattern Recognition* **48**(9): 2798 − 2811. DOI: https://doi.org/10.1016/j.patcog.2015.03.020
22. Tabakhi S., Moradi P. & Akhlaghian F. (2014). An unsupervised feature selection algorithm based on ant colony optimization. *Engineering Applications of Artificial Intelligence* **32**: 112 − 123.

DOI: https://doi.org/10.1016/j.engappai.2014.03.007

23. Touhidi H., Nezamabadi-pour H. & Saryazdi S. (2007). Feature selection using binary ant algorithm. *Frist Joint Congress on Fuzzy and Intelligent Systems*, Mashhad, Iran, 29 − 31 August.
24. Unler A., Murat A. & Chinnam R.B. (2011). mr^2 *PSO*: A maximum relevance minimum redundancy feature selection method based on swarm intelligence for support vector machine classification. *Information Sciences* **181**(20): 4625 − 4641.
DOI: https://doi.org/10.1016/j.ins.2010.05.037
25. Uysal A.K. & Gunal S. (2012). A novel probabilistic feature selection method for text classification. *Knowledge-Based Systems* **36**: 226 − 235.
DOI: https://doi.org/10.1016/j.knosys.2012.06.005
26. Wang C., Shao M., He Q., Qian Y. & Qi Y. (2016). Feature subset selection based on fuzzy neighborhood rough sets. *Knowledge-Based Systems* **111**: 173 − 179.
DOI: https://doi.org/10.1016/j.knosys.2016.08.009
27. Yu L. & Liu H. (2003). Feature selection for high-dimensional data: a fast correlation-based filter solution. *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington DC, pp. 856 − 863.