Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 96 (2016) 403 – 409

Procedia Computer Science

20th International Conference on Knowledge Based and Intelligent Information and Engineering Systems, KES2016, 5-7 September 2016, York, United Kingdom

An Innovative Approach for Attribute Reduction using Rough Sets and Flower Pollination Optimisation

Waleed Yamany[a,*], Eid Emary[b], Aboul Ella Hassanien[b], Gerald Schaefer[c], Shao Ying Zhu[d]

[a] *Faculty of Computers and Information, Fayoum University, Egypt*
[b] *Faculty of Computers & Information, Cairo University, Egypt*
[c] *Department of Computer Science, Loughborough University, U.K.*
[d] *Department of Computing and Mathematics, University of Derby, U.K.*

Abstract

Optimal search is a major challenge for wrapper-based attribute reduction. Rough sets have been used with much success, but current hill-climbing rough set approaches to attribute reduction are insufficient for finding optimal solutions. In this paper, we propose an innovative use of an intelligent optimisation method, namely the flower search algorithm (FSA), with rough sets for attribute reduction. FSA is a relatively recent computational intelligence algorithm, which is inspired by the pollination process of flowers. For many applications, the attribute space, besides being very large, is also rough with many different local minima which makes it difficult to converge towards an optimal solution. FSA can adaptively search the attribute space for optimal attribute combinations that maximise a given fitness function, with the fitness function used in our work being rough set-based classification. Experimental results on various benchmark datasets from the UCI repository confirm our technique to perform well in comparison with competing methods.

Peer-review under responsibility of KES International

Keywords: Pattern recognition; attribute reduction; rough sets; flower pollination optimisation.

1. Introduction

Attribute reduction is one of the essential problems in the fields of data mining, machine learning, and pattern recognition[1]. Attribute reduction is mainly concerned with selecting the smallest subset of attributes for a given problem while preserving a suitably high accuracy in representing the original attributes[2]. In real world problems, attribute reduction is often a necessity due to the presence of noisy, misleading or irrelevant attributes[3], while attribute reduction enables data processing techniques such as machine learning algorithms to yield good performance[2].

* Corresponding author.
E-mail address: wsy00@fayoum.edu.eg

Peer-review under responsibility of KES International
doi:10.1016/j.procs.2016.08.083

One of the well-known approaches for attribute reduction is based on rough set theory[4] which has the inherent ability to deal with vagueness and uncertainty in data analysis. Roughs sets have been extensively used in data mining[5], machine learning[6] and other fields. Here, knowledge is considered as a kind of discriminability, which can also be employed as an instrument to reduce attribute dimensionality and establish data dependencies. While various rough set-based algorithms for attribute reduction have been proposed, the main idea of these methods is to detect minimal reducts by creating every conceivable reduct and subsequently selecting the one with smallest length. This can be performed by developing a kind of detectability capacity from a given dataset and then streamlining it[6]. On the other hand, the number of possible subsets is typically very large and considering all attribute subsets for choosing the optimal one is considered an NP-hard problem. To address this, rough sets can be integrated with optimisation approaches such as genetic algorithms[14], ant colony optimisation[8], or particle swarm optimisation[9].

Wroblewski[7] integrated a genetic algorithm (GA) with a greedy algorithm to produce small reducts, however could not demonstrate that the produced subset is a reduct. ElAlami[13] also made use of GAs to locate ideal relevant attributes. Zhai et al.[14] proposed an incorporated attribute extraction approach that uses both rough sets and GAs. Jensen and Shen[8] applied ant colony optimisation to find rough set reducts, while Wang et al.[9] developed a method for attribute reduction using particle swarm optimisation hybridised with rough sets. Although, stochastic techniques can yield strong solutions for global optimisation, this is accomplished at the expense of computational cost[2].

In this paper, we propose a new attribute reduction technique. In our approach, we make use of the Flower Search Algorithm (FSA) to discover ideal attribute subsets (reducts). FSA is a relatively new evolutionary computation algorithm proposed by Yang[10] and is based on the flower pollination process of flowering plants. FSA can adaptively search the attribute space for optimal attribute combinations that maximise a given fitness function, with the fitness function used in our work being rough set-based classification. Experimental results on various benchmark datasets from the UCI repository confirm our technique to perform well in comparison with competing methods.

The organisation of the remainder of the paper is as follows: Section 2 describes the fundamentals of rough set theory, while Section 3 explains the flower search algorithm. Our FSA algorithm based on rough sets for attribute reduction (FLRSAR) is then presented in Section 4. Experimental results are given in Section 5, while Section 6 concludes the paper.

2. Rough Set Theory

In this section, we present some of the necessary fundamentals for rough set (RS) theory and RS-based feature selection. For more exhaustive descriptions of the theory, we refer to[4] and other publications.

Let $I = (O, S, B, f)$ be an information system, such that O is a finite non-empty set of instances, S is a finite non-empty set of attributes, and B is the set union of attribute scopes such that $B = \bigcup_{s \in S} B_s$ for B_s indicate the value scope of attribute s. $f : O \times S \longrightarrow B$ is an information function that associates a unique magnitude of each attribute with every instance in O such that $f(x, s) \in B_s$ for any $s \in S$ and $x \in O$.

For any $P \subseteq S$ there exists an associated indiscernibility relationship $IND(Z)$

$$IND(Z) = \{(x, y) \in O \times O | \forall s \in p, f(x, s) = f(y, s)\}. \tag{1}$$

The partition of O, deduced by $IND(Z)$ is indicated by $O/IND(Z)$ and can be computed as

$$O/IND(Z) = \otimes\{s \in Z : O/IND(\{s\})\}, \tag{2}$$

where

$$R \otimes K = \{X \bigcap Y : \forall R \in X, \forall K \in Y, X \bigcap Y \neq \emptyset\}. \tag{3}$$

For subset $X \subseteq O$ and equivalence relationship $IND(Z)$, the Z-lower and Z-upper approximations of X are determined as

$$Z_*(X) = \{x \in O : [x]_Z \subseteq X\} \tag{4}$$

and

$$Z^*(X) = \{x \in O : [x]_Z \bigcap X \neq \emptyset\}, \tag{5}$$

respectively.

For two subsets $E, W \subset S$ which give raise to two equivalence relationships $IND(E)$ and $IND(W)$, the E-positive and E-negative regions of W can be determined as

$$POS_E(W) = \bigcup_{X \in O/IND(W)} E_*(X) \tag{6}$$

and

$$NEG_E(W) = O - \bigcup_{X \in O/IND(W)} E^*(X) \tag{7}$$

respectively, while the E-boundary region of W is defined as

$$BND_E W = \bigcup_{X \in O/IND(W)} E^*(X) - \bigcup_{X \in O/IND(W)} E_*(X). \tag{8}$$

A significant problem in data analysis is finding dependencies among attributes. Dependencies can be determined in the following manner. For $E, W \subset S$, E fully depends on W, if and only if $IND(E) \subset IND(W)$. We say that W depends on E to a degree $0 \leq \mu_E(W) \leq 1$, if

$$\mu_E(W) = \frac{\mid POS_E(W) \mid}{\mid O \mid}. \tag{9}$$

If $\mu_E(W) = 1$, then W fully depends on P; if $0 \leq \mu_E(W) \leq 1$, then W depends partially on E; and if $\mu_E(W) = 0$ then W does not depend on E. The dependence degree $\mu_E(W)$ can be used as a heuristic in greedy algorithms for calculating attribute selection.

$I = (O, C \cup D, B, f)$ is called a decision table if $S = C \cup D$ and $C \cap D = 0$ in an information table, where C is the set of condition attributes, and D is the decision attributes set. The dependency degree among condition and decision attributes, $\mu_C(D)$, is named the classification accuracy, induced by the decision attribute set.

The goal of attribute reduction is to discard redundant attributes so that the reduced set has the same classification accuracy as the original set. A reduct is resolved as a subset RED of the condition attribute such that

$$\mu_C(D) = \mu_{RED}(D) \text{ and } \forall P \subset RED, \mu_p(D) \neq \mu_C(D). \tag{10}$$

3. Flower Pollination Search

Flower search is an algorithm based on the flower fertilisation process of flowering plants [10]. Here, the key objective is eventual proliferation by means of fertilisation. Fertilisation is based on the transfer of pollen, which in turn is frequently linked with pollinators, for example insects or other animals [16].

Fertilisation can be caused by self-fertilisation or cross-fertilisation. Cross-fertilisation is caused by pollen from other plants possibly a long distance away as pollinators can cover large areas, consequently yielding global fertilisation. Also, honey bees and birds might follow a Levy flight pattern with flight steps following a Levy distribution [11]. Flower stability can also be used based on the comparability or variety of flowers [16,10].

In the global fertilisation step, pollen are transferred by pollinators and so can travel long distances. This guarantees fertilisation and propagation of the fittest, with the most fittest referred to as g_*. A rule for repositioning can be formulated as

$$X_i^{t+1} = X_i^t + L(X_i^t - g_*), \tag{11}$$

where X_i^t is the i-th solution vector at iteration t and g_* is the current optimal solution. L refers to the quality of the fertilisation with is a step size drawn from a Levy distribution.

Self-fertilisation is fertilisation by pollen from the same flower or flowers of the same plant, typically without a pollinator, and thus leads to local fertilisation. Physical vicinity as well as other factors such as wind, can cause a significant proportion of local fertilisation in the general fertilisation process. Local pollination and flower constancy can be obtained by

$$X_i^{t+1} = X_i^t + \epsilon(X_j^t - g_k^t), \tag{12}$$

where X_j^t and X_k^t are solution vectors drawn randomly from the solution set, while ϵ is drawn from a uniform distribution in the range $[0; 1]$.

4. FSA and RS-based Attribute Reduction (FLRSAR)

In this paper, we employ the idea of flower fertilisation for optimal attribute reduction. An overview of our approach is shown in Fig. 1. Each attribute subset can be interpreted as a position in a search space. For N attributes, there are 2^N different attribute subsets. The optimal solution is the subset that yields the highest classification accuracy and smallest length, for whose identification we utilise the exploration and optimisation characteristics of the flower search algorithm.

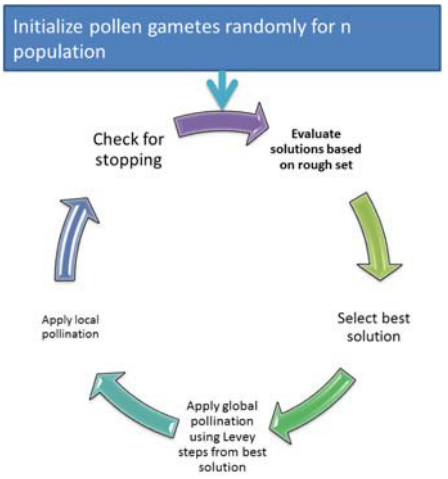

Fig. 1. Overview of the proposed approach.

For our FSA approach, the solution space represents all possible attribute subsets and hence flower positions represent selections of attribute sets. Each attribute is considered as an individual with a value in the range $[-2; 2]$. To decide if a feature will be selected or not, its position value will be thresholded with a constant threshold.

The parameter ϵ in Eq. (12) is important for the performance of the flower algorithm as it controls the diversity of the solutions obtained at a given iteration. In our work, we adaptively tune ϵ throughout the running of the algorithm. At the beginning, we need to allow for larger diversity and hence we use larger values for ϵ, while towards the end, the solutions are near the optimal and thus less diversity is required and smaller values for ϵ are used. Initially, ϵ is set as half the search range. It is then is calculated in each iteration by

$$\epsilon = \epsilon_0 - \frac{t\epsilon_0}{N_{gen}}, \tag{13}$$

where ϵ_0 is the initial value for ϵ, t is the iteration number and N_{gen} is the maximum number of iterations.

As fitness function, we use

$$Fit = \alpha\gamma_R(D) + \beta\frac{|\,C - R\,|}{|\,C\,|}, \tag{14}$$

where $\gamma_R(D)$ is the quality of classification of condition feature set R proportional to decision D, $|\,R\,|$ is the length of the attribute reduction subset, and $|\,C\,|$ is the aggregate number of attributes. The two parameters α and β assign weights to the two parts of the fitness function (i.e., classification quality and subset length respectively) with $\alpha \in [0, 1]$ and $\beta = 1 - \alpha$.

In our experiments, we assume the subset length to be less important compared to classification accuracy and set $\alpha = 0.9$, $\beta = 0.1$. This assures that the optimal solution is at least a real reduct. The quality of every solution is estimated by the fitness function in Eq. (14) with the aim to maximise fitness[9].

Algorithm 1 summarises the main steps of our algorithm in terms of pseudo-code.

Algorithm 1 Proposed FLRSAR algorithm

```
Initialise a population of n flowers/pollen gametes with random solutions.
Select the best solution g*.
t = 0
while t ≤ n do
  for all Flower i in the solution pool do
    if rand < p then
      Draw a d-dimensional step vector L which obeys a Levy distribution.
      Apply global pollination on solution i using Eq. (11).
    else
      Draw ϵ from uniform distribution.
      Randomly choose j, k solutions from the current pool of solutions.
      Apply local pollination on solution i using j, k and Eq. (12).
    end if
    Evaluate the new solution using fitness function from Eq. (14).
    if new solution is better then
      Replace solution i with new solution.
    end if
  end for
  Update the best solution g*.
end while
```

5. Experimental Results

We performed extensive classification experiments on 11 datasets from the UCI Machine Learning Repository[18] listed in Table 1. We use the LEM2 technique[12] to obtain rules from the data and the global strength[15,9] for rule parley in classification. For evaluation, we perform standard ten-fold cross validation. For comparison, we also obtain results for conditional entropy-based attribute reduction (CEAR)[20], discernibility matrix-based attribute reduction (DISMAR)[19] and GA-based attribute reduction (GAAR)[15].

Dataset	Attributes	Instances	CEAR	DISMAR	GAAR	FLRSAR
Breastcancer	9	699	4	5	4	2
M-of-N	13	1000	7	6	6	6
Exactly	13	1000	8	6	6	6
Exactly2	13	1000	11	10*	11	10*
Vote	16	300	11	8*	9	8*
Zoo	16	101	10	5	6	4
Lymphography	18	148	8	7	8	6
Led	24	2000	12	18	8	5*
Soybean-small	35	47	2	2	6	2
Lung	56	32	5	4	6	6
DNA	57	318	6	6	7	5*

Table 1. Reduct sizes found by different attribute reduction methods. * indicates optimal solution.

In Table 1, we list reduct information for the different algorithms. For the evaluated datasets, some may have more than one ideal reduct, while some have only one (exclusive) best reduct. From our results, we can observe that in some cases hill-climbing methods can locate the optimal solution. For example, DISMAR identifies the optimal solution for the Soybean-small and Vote datasets. However, for the other datasets, only sub-optimal solutions are found, leading to redundant attributes. For CEAR, we can see that it often includes even more redundant attributes than DISMAR. As the experimental results confirm, FLRSAR performs better than GAAR. As we can see, FLRSAR

effectively discovers the best reducts on almost all datasets compared to the competing methods. FLRSAR finds the optimal reduct for the Exactly2 and Vote dataset, and is the only method to identify the ideal reduct for the Led and DNA datasets.

FSA has strong exploration capabilities. For some of the datasets with more properties, for instance Lymphography and DNA, the GA-based approach, being influenced by the number of attributes, after having found a sub-optimal solution could not successfully identify a superior one. On the other hand, our FSA-based algorithm is shown to effectively search the attribute domain and to acquire the optimal solution.

Table 2 gives the resulting numbers of decision rules and classification accuracies based on various reducts. As we can see from there, our approach not only successfully yields compact reducts, the classification accuracy is also superior compared to the other methods. Except for three datasets (Breastcancer, Zoo, and Lung), our FSA-based approach gives the best classification performance.

Dataset	CEAR		DISMAR		GAAR		FLRSAR	
	NR	CA	NR	CA	NR	CA	NR	CA
Breastcancer	75	94.20	67	95.94	64	95.65	36	94.64
M-of-N	35	100	35	100	35	100	35	100
Exactly	50	100	50	100	50	100	50	100
Exactly2	178	69.60	230	83	200	80.80	219	85.50
Vote	25	92.33	28	93.67	25	94.00	23	94.67
Zoo	13	94.00	13	94.00	13	92.00	11	92.00
Lymphography	42	72.14	40	74.29	38	70.00	34	78.57
Led	228	83.10	257	78.85	10	100	10	100
Soybean-small	4	100	4	100	4	97.50	5	100
Lung	13	73.33	14	73.30	12	70.00	11	70.00
DNA	192	26.45	191	36.45	191	33.87	182	41.94

Table 2. Obtained classification accuracies for different reducts. NR = number of rules; CA = classification accuracy.

We also conducted a further comparison with the rough set framework RSES [17] and show the results in Table 3. We use exhaustive calculation to acquire reducts in RSES. RSES could not discover all reducts in all datasets (Soybean-small, Lung and DNA) by exhaustive search due memory constraints. Consequently, we utilise a quick genetic algorithm to obtain 10 reducts. We use the decision rules classifier LEM2 algorithm for global rule generation and ten-fold cross-validation evaluation as in our other experiments. As we can see from Table 3, for almost all cases, reducts obtained by FLRSAR are smaller and exhibit higher classification accuracy.

Dataset	TR	MS	NR	CA
Breastcancer	20	4	455	94.40
M-of-N	1	6	19	92.90
Exactly	1	6	41	85.90
Exactly2	1	10	242	76.40
Vote	2	8	67	92.93
Zoo	33	5	109	96.80
Lymphography	424	6	353	85.90
Led	140	5	6636	100
Soybean-small	-	2	31	92.50
Lung	-	4	100	75.00
DNA	-	5	2592	74.30

Table 3. Experimental results by RSES. TR = total reducts; MS = minimaL size of reducts; NR = number of rules; CA = classification accuracy.

6. Conclusions

In this paper, we have proposed an algorithm based on rough sets and flower pollination optimisation for attribute reduction. FSA has robust search capabilities and can effectively find small attribute reducts based on a suitable definition of a fitness function that combines both classification accuracy and attribute set size. Experimental results prove competitive performance for our FSA-based approach showing that FSA combined with rough sets forms a useful technique for the attribute reduction problem.

References

1. Laura, M. C. "A framework for feature selection in high-dimensional domains". Ph.D. Thesis, University of Cagliari,2012.
2. Chen, Y., Miao, D., Wang, R. and Wu, K. "A Rough Set Approach to Feature Selection Based On Power Set Tree". Knowledge-Based Systems 24, 275-281, 2011.
3. Jensen, R. "Combining rough and fuzzy sets for feature selection". Ph.D. Thesis, University of Edinburgh, 2005.
4. Pawlak, Z. "Rough sets", International Journal of Computer and Information Science 11(5), 341-356, 1982.
5. Guo, Q.L and Zhang, M. "Implement web learning environment based on data mining", Knowledge-Based Systems, Vol. 22, No. 6, pp. 439-442, 2009.
6. Swiniarski, R.W and Skowron, A. "Rough set methods in feature selection and recognition", Pattern Recognition Letters 24, 833-849, 2003.
7. Wroblewski, J. "Finding minimal reducts using genetic algorithms". Second Annual Join Conference on Information Sciences, pp. 186-189, 1995.
8. Jensen, R. and Shen, Q., "Semantics-preserve dimensionality reduction: rough and fuzzy-rough-based approaches", IEEE Transactions on Knowledge and Data Engineering 16, 1457-1471, 2004.
9. Wang, X., Yang, J., Teng, X., Xia, W. and Jensen. R, "Feature selection based on rough sets and particle swarm optimization",Pattern Recognition Letters 28, 459-471, 2007.
10. Yang, X.S., "Flower pollination algorithm for global optimization", Unconventional Computation and Natural Computation 2012, Lecture Notes in Computer Science, Vol. 7445, pp. 240-249, 2012.
11. Pavlyukevich, I. "Levy flights, non-local search and simulated annealing", Journal of Computational Physics, 226(2), 1830-1844, 2007.
12. Stefanowski, J. "On rough set based approaches to induction of decision rules". In: Skowron, A., Polkowski, L. (Eds.), "Rough Sets in Knowledge Discovery", Physica Verlag, 500-529, 1998.
13. ElAlami, M. E. "A filter model for feature subset selection based on genetic algorithm". Knowledge-Based Systems 22(5), 356-362.
14. Zhai, L. Y., Khoo, L. P., and Fok, S. C. "Feature extraction using rough set theory and genetic algorithms: an application for the simplification of product quality evaluation". Computers & Industrial Engineering 43(4), 661-676, 2002.
15. Bazan, J., Nguyen, H.S., Nguyen, S.H., Synak, P and Wroblewski, J. "Rough set algorithms in classification problem. In: Polkowski, L., Tsumoto, S., Lin, T.Y. (Eds.), "Rough Set Methods and Applications". Physica-Verlag, pp. 49-88, 2000.
16. Yang, X.S., Karamanoglu, M. and He, X. "Multi-objective Flower Algorithm for Optimization", Procedia Computer Science 18, 861-868, 2013.
17. Skowron, A., Bazan, J., Son, N.H. and Wroblewski, J. "RSES 2.2 User's Guide", Institute of Mathematics, Warsaw University, Warsaw, Poland, 2005.
18. Bache, K. and Lichman, M. "UCI Machine Learning Repository" [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.
19. Hu, K., Lu, Y.C. and Shi, C.Y. "Feature ranking in rough sets". AI Commun 16(1), pp. 41-50, 2003.
20. Wang, G.Y., Yu, H. and Yang, D.C. "Decision table reduction based on conditional information entropy". Chin. J. Comput. 25 (7), 759-766 2002.