A novel hybrid feature selection via Symmetrical Uncertainty ranking based local memetic search algorithm

S. Senthamarai Kannan [a,*], N. Ramaraj [b]

[a] Department of Information Technology, Thiagarajar College of Engineering, Madurai, India
[b] G.K.M. Engineering College, Chennai, India

ARTICLE INFO

ABSTRACT

A novel correlation based memetic framework (MA-C) which is a combination of genetic algorithm (GA) and local search (LS) using correlation based filter ranking is proposed in this paper. The local filter method used here fine-tunes the population of GA solutions by adding or deleting features based on Symmetrical Uncertainty (SU) measure. The focus here is on filter methods that are able to assess the goodness or ranking of the individual features. Empirical study of MA-C on several commonly used datasets from the large-scale Gene expression datasets indicates that it outperforms recent existing methods in the literature in terms of classification accuracy, selected feature size and efficiency. Further, we also investigate the balance between local and genetic search to maximize the search quality and efficiency of MA-C.

1. Introduction

The feature selection problem in terms of supervised inductive learning is: given a set of candidate features, select a subset defined by one of three approaches: (a) the subset with a specified size that optimizes an evaluation measure, (b) the subset of smaller size that satisfies a certain restriction on the evaluation measure and (c) the subset with the best commitment among its size and the value of its evaluation measure [1]. High dimensional data (i.e., data sets with hundreds or thousands of features), can contain high degree of irrelevant and redundant information which greatly degrades the performance of learning algorithms. Therefore, feature selection becomes necessary for machine learning tasks for facing high dimensional data. However, this trend of enormity on both the size and dimensionality poses great challenges to feature selection algorithms. Some of the recent research efforts in feature selection focus on the challenges from handling a huge number of instances [3] to dealing with high dimensional data [2]. This work is concerned about feature selection for high dimensional Gene expression datasets.

Feature selection [20–22] has become the focus of many research areas in recent years. With the rapid advance of computer and database technologies, datasets with thousands of variables and features are now ubiquitous in pattern recognition, data min-ing, and machine learning. Feature selection generally involves a combination of search and attributes utility estimation plus evaluation with respect to specific learning schemes [19].

Feature selection algorithms can broadly fall into the filter model or the wrapper model [4,2]. The filter model relies on general characteristics of the training data to select some features without involving any learning algorithms. It therefore does not inherit any bias of a learning algorithm. They are computationally cheap, as they do not involve the induction algorithm. However, they also take the risk of selecting subsets of features which may not match the chosen induction algorithm. The wrapper model requires one predetermined learning algorithm in feature selection and uses its performance to evaluate and determine the features that are selected. As for each new subset of features, the wrapper model needs to learn a hypothesis (or a classifier). It tends to give superior performance as it finds features that are better suited for the predetermined learning algorithm. But it tends to be more computationally expensive.

In this paper, we propose a novel correlation based memetic framework [5,6,27], i.e., a combination of genetic algorithm (GA) [7,8] and local search (LS) using correlation based filter ranking [9]. Memetic algorithms (MAs) [24] are population-based meta-heuristic search methods inspired by Darwinian's principles of natural evolution and Dawkins' notion of a meme defined as a unit of cultural evolution that is capable of local refinements. Recent studies on MAs have revealed their successes on a wide variety of real world problems. Particularly, they not only converge to high quality solutions, but also search more efficiently than their conventional counterparts.

* Corresponding author.
E-mail addresses: sskit@tce.edu (S. Senthamarai Kannan), ramaraj_gm@yahoo.com (N. Ramaraj).

The goal of MA-C is to improve classification performance and to accelerate the search to identify important feature subsets. In particular, the filter method fine-tunes the population of GA solutions by adding or deleting features based on SU measure. Hence, our focus here is on filter methods that are able to assess the goodness or ranking of the individual features. Empirical study of MA-C on several commonly used datasets from the UCI repository [10] indicates that it outperforms recent existing methods in the literature in terms of classification accuracy, selected feature size and efficiency. Further, we also investigate the balance between local and genetic search to maximize the search quality and efficiency of MA-C.

2. Related work

A central problem in machine learning is identifying a representative set of features from which to construct a classification model for a particular task. The work presented in thesis [11] addresses the problem of feature selection for machine learning through a correlation based approach. The central hypothesis is that good feature sets contain features that are highly correlated with the class, yet uncorrelated with each other.

An integrated approach for simultaneous clustering and feature selection using a niching memetic algorithm [13] makes feature selection an integral part of the global clustering search procedure and attempts to overcome the problem of identifying less promising locally optimal solutions in both clustering and feature selection, without making any a priori assumption about the number of clusters. Within the NMA_CFS procedure, a variable composite representation is devised to encode both feature selection and cluster centers with different numbers of clusters. Further, local search operations are introduced to refine feature selection and cluster centers encoded in the chromosomes. Finally, a niching method is integrated to preserve the population diversity and prevent premature convergence. Results of the computational experiments proposed in [14] clearly show the importance of striking a balance between genetic search and local search. In this work, a multiobjective genetic local search (MOGLS) algorithm is modified by choosing only good individuals as initial solutions for local search and assigning an appropriate local search direction to each initial solution. The recursive least squares algorithm [15] is proposed as an efficient way to generate local models and local cross-validation is used as an economic way to validate different alternatives. As far as model selection is concerned, the winner-takes-all strategy and a local combination of the most promising models are explored. The method proposed is tested on six different datasets and compared with state-of-the-art approaches. A hybrid approach involving genetic algorithms (GA) and bacterial foraging (BF) algorithms for function optimization problems [16] is illustrated using four test functions and the performance of the algorithm is studied with an emphasis on mutation, crossover, variation of step sizes, chemotactic steps, and the lifetime of the bacteria.

ReliefF [17] has proved to be a successful feature selector but when handling a large dataset, it is computationally expensive. An optimization using Supervised Model Construction has been proposed to improve starter selection. Effectiveness has been evaluated using 12 UCI datasets and a clinical diabetes database. Experiments indicate that compared with ReliefF, the proposed method improved computation efficiency whilst maintaining the classification accuracy. In the clinical dataset (20,000 records with 47 features), feature selection via Supervised Model Construction (FSSMC) reduced the processing time by 80%, compared to ReliefF, and maintained accuracy for Naive Bayes, IB1 and C4.5 classifiers.

A Gene ranking method based on Grey Relational Analysis [18] requires less data, does not rely on data distribution and is more applicable to numerical data value. experimentally performed better compared with several traditional methods, including Symmetrical Uncertainty, χ^2-statistic and ReliefF. Especially it is much faster than other methods.

A hybrid genetic rule learning algorithm [25,28] incorporating a local search method embedded in the evolution process to improve the performance of the algorithm. In the local search procedure, the minimum information entropy heuristic is used to specify the importance of features. Irrelevant features are removed and useful features are added. When adding a relevant feature, the corresponding rule condition is also adjusted to improve the rule quality. Experiments show that this hybrid model works well in practice.

A novel feature subset selection algorithm, which utilizes a genetic algorithm (GA) to optimize the output nodes of trained artificial neural network (ANN) has been presented in [29]. The GA is involved to find the optimal relevant features, which maximize the output function for each class. The dominant features in all classes are the features subset to be selected from the input feature group.

A simple filter method for setting attribute weights for use with naive Bayes has been experimented in [30] to show that naive Bayes with attribute weights rarely degrades the quality of the model compared to standard naive Bayes and, in many cases, improves it dramatically. The main advantages of this method compared to other approaches for improving naive Bayes is its run-time complexity and the fact that it maintains the simplicity of the final model.

A new algorithm of data reduction based on a correlation model with data discretization named FCBF$^+$ is proposed in [31] which perform the discretization of continuous attributes in an efficient manner. In this paper, the author aim to solve the current problem that a continuous attribute in a clustering or classification algorithm must be made discrete. Performance evaluation is done on clustering accuracy for all the features, and a reduced set of features is obtained using FCBF$^+$. It is found that the proposed FCBF$^+$ algorithm improves the clustering accuracy of various clustering algorithms.

3. A correlation based memetic algorithm (MA-C)

In this section, we introduce the proposed correlation based memetic feature selection algorithm (MA-C) for classification problems which is depicted in Fig. 1.

In the first step, the GA population is randomly initialized with each chromosome encoding a candidate feature subset. Subsequently, a local search (LS) is performed. The LS is performed on
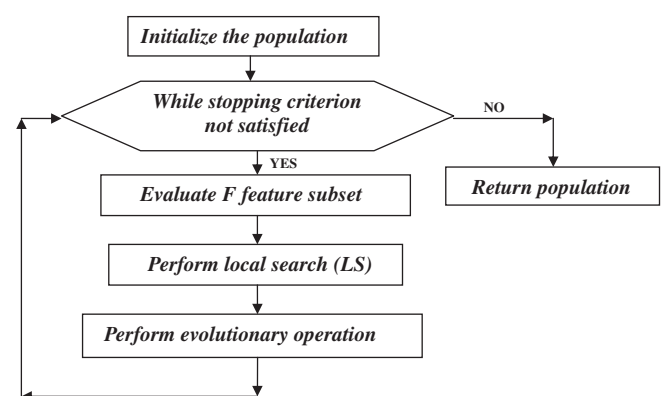

Fig. 1. Flow chart for MA-C.

all or portion of the chromosomes, to reach a local optimal solution or to improve the feature subset. Genetic operators such as crossovers and mutations are performed to generate the next population. This process repeats itself till the stopping conditions are satisfied. Each component is explained as follows.

3.1. Population initialization

In the feature selection problem, a representation for candidate feature subset must be chosen and encoded as a chromosome. A chromosome is a binary string of length equal to the total number of features so that each bit encodes a single feature. A bit of '1' or '0' implies that the corresponding feature has either been selected or rejected. The length of the chromosome is denoted as n. The maximum allowable number of bit '1' in each chromosome is denoted as m. When prior knowledge about the optimal number of features is available, we may limit m to no more than the pre-defined value; otherwise m is equal to n. At the start of the search, a population size of p is randomly initialized.

3.2. Objective function

The objective function is defined in simple words by the classification accuracy:

$$Fitness(c) = Accuracy(Sc), \tag{1}$$

where Sc denotes the corresponding selected feature subset encoded in chromosome c, and the feature selection criterion function Accuracy(Sc) evaluate the significance for the given feature subset Sc, In this paper, Accuracy(Sc) is specified as the classification accuracy for Sc using Naïve Bayes algorithm. Note that when two chromosomes are found to have similar fitness, i.e., the difference between their fitness is less than a small value of ε, then the one with a smaller number of selected features is given higher chances of surviving to the next generation.

3.3. Local search improvement procedure (LS)

Correlation based filter ranking method using Symmetrical Uncertainty proved that it is more efficient to remove redundant features in order to improve the classifier's accuracy. Taking this cue, here we consider the use of correlation based filter ranking method with SU measure as memes or local search heuristics in our MA-C. We show in Section 4 that using filter ranking methods as memes, MA is capable of converging to improved classification accuracy at a lower number of selected features when compared to existing methods. In this section, we discuss how to evaluate the goodness of features for classification using SU based correlation measure. In general, a feature is said to be good if it is relevant to the class concept without being redundant to any of the other relevant features. If we adopt the correlation between two variables as a goodness measure, the above definition can be restructured as a feature that is good if it is highly correlated with the class but not highly correlated with any of the other features. In other words, if the correlation between a feature and the class is high enough to make it relevant to (or predictive of) the class and the correlation between it and any other relevant features does not reach a level so that it can be predicted by any of the other relevant features, it will be regarded as a good feature for the classification task.

SU based correlation measure is based on the information-theoretical concept of entropy, which is a measure of the uncertainty of a random variable. The entropy of a variable X is defined as

$$H(X) = -\sum P(x_i)\log_2(P(x_i)) \tag{2}$$

and the entropy of X after observing values of another variable Y is defined as

$$H(X|Y) = -\sum_i P(y_j) \sum_j P(x_i|y_j)\log_2(P(x_i|y_j)), \tag{3}$$

where $P(x_i)$ is the prior probabilities for all values of X and $P(x_i|y_j)$ is the posterior probabilities of X given the values of Y. The amount by which the entropy of X decreases reflects additional information about X provided by Y and is called information gain (IG), given by

$$IG(X|Y) = H(X) - H(X|Y). \tag{4}$$

According to this measure, a feature Y is regarded to be more correlated to feature X than to feature Z, if $IG(X|Y) > IG(Z|Y)$.

Information gain is symmetrical for two random variables X and Y. Symmetry is a desired property for a measure of correlations between features. However, information gain is biased in favor of features with more values. Furthermore, the values have to be normalized to ensure they are comparable and have the same affect. Therefore, we choose Symmetrical Uncertainty [11], defined as follows:

$$SU(X, Y) - 2[IG(X|Y)/(H(X) + H(Y))]. \tag{5}$$

SU compensates for the information gain's bias toward features with more values and normalizes its values within the range [0, 1] with the value 1 indicating that knowledge of either one of the values completely predicts the value of the other and the value 0 indicating that X and Y are independent. It treats a pair of features symmetrically.

The SU value has two main functions: (1) it can remove the features with SU lesser than threshold and (2) gets every feature's weight that is to be used to guide the initialization of the population for genetic algorithms in memetic framework. The feature having larger SU value gets higher weight. The feature having lesser SU value is removed. All these concepts have been explained in Fig. 2.

Given a data set with N features and a class C, the algorithm finds a set of predominant features subset for the class concept. It consists of two major parts. In the first part, it calculates the $SU_{i,c}$ value for each feature and places them in descending order according to their $SU_{i,c}$ values. The $SU_{i,c}$ value defines the correlation between the feature F_i, the class C. In the second part, it further processes the ordered list to remove the redundant features and keeps only the predominant ones among all the selected relevant features. A feature f_p that has already been determined to be a predominant feature can always be used to filter out other features that are ranked lower than fp and have f_p as one of its redundant peers. The iteration starts from the first element and continues as follows. For all the remaining features (from the one right next to f_p to the last one in the list), if f_p happens to be a redundant peer to a feature f_q, f_q will be removed. Feature f_q is said to be a redundant pair to f_p if the correlation between f_p and f_q is greater than the correlation between f_q and the class C. After completing one round of filtering features based on f_p, the algorithm will take the currently remaining feature right next to f_p as the new reference to repeat the filtering process. The algorithm stops only when there are no more features to be removed. It finally returns the optimal feature subset.

3.4. Evolutionary operators

In the evolution process, standard GA operators such as linear ranking selection, uniform crossover and mutation operators based on elitist strategy may be applied. However, if prior knowledge on the optimum number of features is available, the number of bit '1' in each chromosome may be constrained to a maximum of m in the
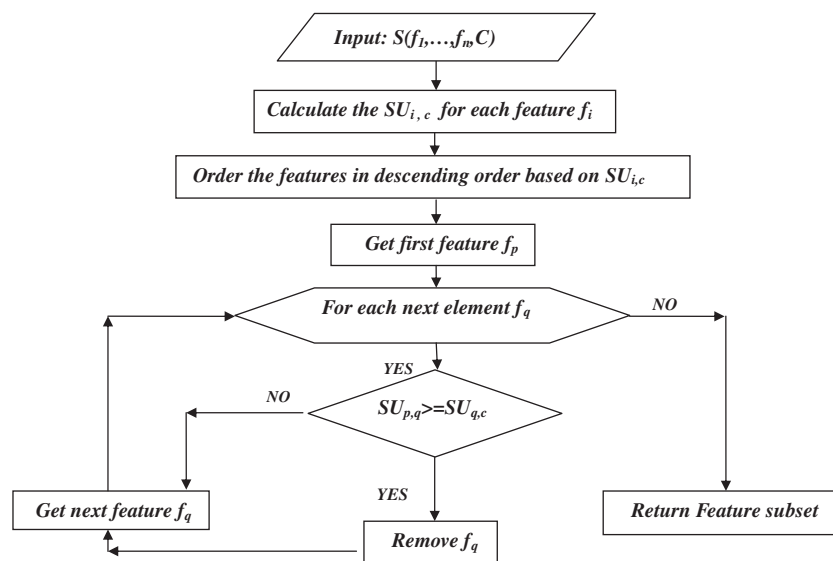
Fig. 2. SU based correlation-filter ranking method – LOCAL SEARCH.

evolution process. Since the standard uniform crossover and mutation operators may violate this constraint, Subset Size-Oriented Common Feature Crossover [12] and mutation are proposed here.

Crossover: We use a Subset Size-Oriented Common Feature Crossover Operator (SSOCF) which keeps useful informative blocks and produces offsprings which have the same distribution than the parents. Offsprings are kept, only if they fit better than the least good individual of the population. Features shared by the two parents are kept by offsprings and the non-shared features are inherited by offsprings corresponding to the ith parent with the probability $(n_i - n_c/n_u)$ where n_i is the number of selected features of the ith parent, n_c is the number of commonly selected features across both mating partners and n_u is the number of non-shared selected features.

Mutation: The mutation is an operator which allows diversity. During the mutation stage, a chromosome has a probability p_{mut} to mutate. If a chromosome is selected to mutate, we choose randomly a number n of bits to be flipped then n bits are chosen randomly and flipped.

4. Experimental results and discussion

In this section, we present an experimental study of MA-C on eight commonly used Gene expression datasets. In the MA-C, we employed a population size that is equal to the number of attributes with the stopping criterion as 6000 fitness function calls for Gene datasets. Further, the maximum number of selected features is constrained to 2500 for the Gene datasets as depicted in Table 1. In our experimental setup, we employ crossover and mutation probabilities of $P_c = 0.6$ and $P_m = 0.1$, respectively. Linear ranking selection with selection pressure of 1.5 is used for selection. The threshold ε to determine fitness similarity between two chromosomes is configured as 0.001 for Gene datasets. The fitness of a chromosome or selected feature subset is evaluated using the Naïve Bayes classifier and the standard 10 folds cross-validation. We use Naïve Bayes in our experiments because the principal conclusions from [23] are that Naïve Bayes offers off-the-shelf solutions to problems with large numbers of samples and attributes, mixed types of variables, and lots of missing values. Here we use the classification accuracy estimated from cross-validation and

Table 1
Datasets and parameters used for experiments.

Dataset	No. of features	No. of instances	No. of classes	Population size	P_c, P_m	Max No. of selected features, m
Breast	24,481	97	2	24,482	0.6, 0.1	2500
CNS	7129	60	2	7130	0.6, 0.1	2500
Leukemia	7129	72	2	7130	0.6, 0.1	2500
Leukemia_3c	7129	72	3	7130	0.6, 0.1	2500
Leukemia_4c	7129	72	4	7130	0.6, 0.1	2500
Ovarian	15,154	253	2	15,155	0.6, 0.1	2500
SRBCT	2308	83	4	2309	0.6, 0.1	2500
MLL	12,582	72	3	12,583	0.6, 0.1	2500

the number of selected features as performance measures. This algorithm was carried out in the WEKA and MAFS environment [26]. The memetic algorithm has been run using MAFS and all other algorithms used in this paper have been run using the WEKA. It is worth noting that the configurations of the parameter used here have been investigated empirically for the datasets considered and are summarized in Table 1.

Table 2 compares the accuracy of various feature selection algorithms with our MA-C. The results of the following algorithms, (a) SU-CFS is a correlation based feature selection method which uses SU measure, (b) genetic algorithm (GA) with Naïve Bayes as subset selection criterion, (c) WFFSA-R [27], which is a wrapper-filter feature selection algorithm and it uses ReliefF as filter ranking method and GA as wrapper method, (d) our proposal MA-C which is a hybrid combining memetic algorithm and correlation has been represented graphically.

In Table 2, Acc denotes the classifier accuracy, in percentage, using Naïve Bayes algorithm and F_s denotes the no of features selected and the column. None specifies the accuracy and the number of selected features to the original dataset without applying any feature selection algorithm. Best results in each row are shown in block letters. From the table, we infer that the redundant attributes are removed efficiently in our algorithm as it is very much reduced when compared to the original algorithm. Also, we can see that as the number of attributes increases, the reduction in attribute and efficiency of resulting attributes increases. From Ta-

Table 2
Performance comparison of proposed MA-C method.

| Datasets	None		SU-CFS		GA		WFFSA-R		MA-C	
	Accuracy (%)	Features selected	Accuracy (%)	Features selected	Accuracy (%)	Features selected	Accuracy (%)	Features selected	Accuracy (%)	Features selected
Breast	90.72	24481	56.7	139	58.76	332	63.91	196	95.26	183
CNS	93.33	7129	75	40	73.33	915	76.66	475	97.78	374
Leukemia	98.61	7129	98.61	76	98.61	710	97.22	384	99.56	387
Leukemia_3c	100	7129	97.22	105	98.61	999	98.61	452	99.53	394
Leukemia_4c	100	7129	94.44	120	94.44	985	97.74	464	98.61	386
Ovarian	98.02	15154	100	32	96.04	313	99.2	292	100	247
SRBCT	100	2308	100	112	100	1044	100	651	100	526
MLL	98.61	12582	98.61	118	100	815	98.61	115	100	108
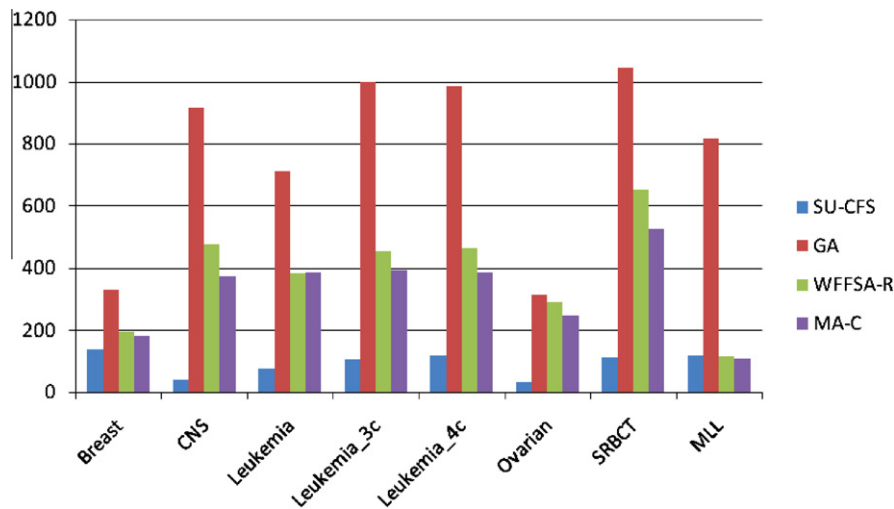
Fig. 3. Comparison of selected features for each feature selection algorithm.
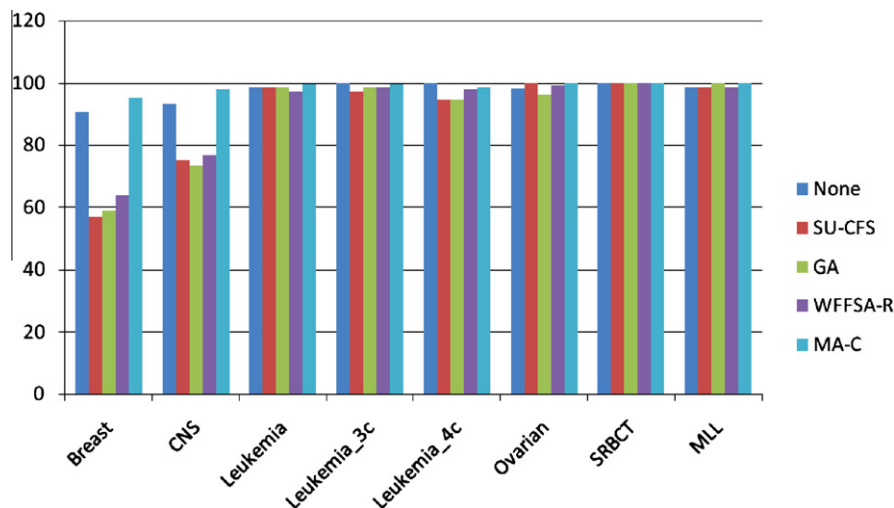

Fig. 4. Comparison of classifier accuracy for each feature selection algorithm.

ble 2, it is obvious that MA-C produce best results except Leukemia_3c, Leukemia_4c datasets.

From Figs. 3 and 4, it is obvious that the proposed MA-C obtains substantial reduction in feature set size maintaining better accuracy compared with other approaches for the chosen Gene datasets of high dimensionality.

5. Computational complexity

In this section, we analyze the computational complexity of the proposed MA-C. The ranking of features based on the filter methods have linear time complexity in terms of feature dimensionality. They are conducted offline and the rank list thus obtained may be

reused for each local search in MA-C. Consequently, the computation for feature ranking is a one-time offline cost and is considered to be negligible compared to that of the fitness evaluation in Eq. (1). Hence, we define the computational cost of a single fitness evaluation as the basic unit of computational cost in our analysis.

The computational complexity for GA can be derived as $\mathrm{O}\,(pg)$, where p is the size of population and g is the number of search generations. The computational complexity for correlation based filter ranking is $\mathrm{O}\,(MN \log N)$ where M is the number of instances and N is the number of features in the dataset. Generally the time complexity of MA-C is high since it combines filter and wrapper technique, but its efficiency in terms of feature selection is high when compared with other algorithms.

6. Conclusion

In this paper, we have proposed a novel hybrid feature selection algorithm (MA-C) based on a memetic framework. We use correlation based filter ranking method as a local search heuristic in the memetic algorithm. The experimental results show that the proposed method has efficient searching strategies and is capable of producing good classification accuracy with small number of features simultaneously. Most importantly, the performance of the proposed approach is better than GA, MA with sequential local search and other existing algorithms cited in the literature. Further, our study on various local search strategies, local search length and intervals allow us to identify a suitable balance tradeoff between genetic and local search in the memetic search. This allows us to maximize the effectiveness and efficiency of the proposed hybrid filter and wrapper feature selection algorithm for classification problem using a memetic framework.

References

[1] Luis Carlos Molina, Lluis Belanche, Angela Nebot, in: IEEE International Conference on Feature Selection Algorithms: A Survey and Experimental Evaluation, 2002, pp. 306–313.

[2] S. Das, Filters, wrappers and a boosting-based hybrid for feature selection, in: Proceedings of the Eighteenth International Conference on Machine Learning, 2001, pp. 74–81.

[3] H. Liu, H. Motoda, L. Yu, Feature selection with selective sampling, in: Proceedings of the Nineteenth International Conference on Machine Learning, 2002, pp. 395–402.

[4] R. Kohavi, G. John, Wrappers for feature subset selection, Artificial Intelligence 97 (1997) 273–324.

[5] Y.S. Ong, A.J. Keane, Meta-Lamarckian in memetic algorithm, IEEE Transactions on Evolutionary Computation 8 (2) (2004) 99–110.

[6] Zexuan Zhu, Y.S. Ong, M. Dash, Markov blanket-embedded genetic algorithm for gene selection, Pattern Recognition 49 (11) (2007) 3236–3248.

[7] Feng Tan, Xuezheng Fu, Yanqing Zhang, Anu G. Bourgeois, A genetic algorithm-based method for feature subset selection, 2007.

[8] Robert R. Biers, Matthew F. Muldoon, Bruce G. Pollock, Steven Manuck, Gwenn Smith, Mark E. Sale, A genetic algorithm-based hybrid machine learning approach to model selection, Journal of Pharmacokinetics and Pharmacodynamics 33 (2) (2006).

[9] L. Yu, H. Liu (Feature selection methods for high-dimensional data: a fast correlation-based filter solution), Springer-Verlag, 2003, pp. 856–863.

[10] P.M. Murphy, D.W. Aha, UCI Repository for Machine Learning Database, Technical Report, Department of Information and Computer Science, University of California, Irvine, California, 1994.

[11] Mark A. Hall, Correlation based feature selection for machine learning, Thesis Report, University of Waikato, April 1999.

[12] C. Emmanouilidis, A. Hunter, J. MacIntyre, A multiobjective evolutionary setting for feature selection and a commonality-based crossover operator, in: Congress on Evolutionary Computing 2000, vol. 2, CEC, 2000, pp. 309–316.

[13] Weiguo Sheng, Xiaohui Liu, Mike Fairhurst, A niching memetic algorithm for simultaneous clustering and feature selection, IEEE Transactions on Knowledge and Data Engineering 20 (7) (2008) 868–879.

[14] H. Ishibuchi, T. Yoshida, T. Murata, Balance between genetic search and local search in memetic algorithm for multiobjective permutation flowshop scheduling, IEEE Transactions on Evolutionary Computation 7 (2) (2003) 204–223.

[15] G. Bontempi, M. Birattari, H. Bersini, A model selection approach for local learning, AI Communications 13 (1) (2000) 41–47.

[16] Dong Hwa Kim, Ajith Abraham, Jae Hoon Cho, A hybrid genetic algorithm and bacterial foraging approach for global optimization, Information Sciences (2007) 3918–3937.

[17] Lijuan Zhang, Zhoujun Li, An optimization of ReliefF for classification in large datasets, Data and Knowledge Engineering 68 (11) (2009) 1348–1356.

[18] Lijuan Zhang, Zhoujun Li, Gene selection for classifying microarray data using grey relation analysis, Discovery Science (2006) 378–382.

[19] Mark A. Hall, Geoffrey Holmes, Benchmarking attribute selection techniques for discrete class data mining, IEEE Transactions on Knowledge and Data Engineering 15 (6) (2003) 1437–1447.

[20] H. Liu, H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, Kluwer Academic Publishers, 1998.

[21] Isabelle Guyon, Andre Elisseeff, An introduction to variable and feature selection, Journal of Machine Learning Research 3 (2003) 157–1182.

[22] I. Guyon, A practical guide to model selection, in: Proceedings of the Machine Learning Summer School Springer Text in Statistics, Springer, 2009.

[23] I. Guyon, V. Lemaire, M. Boullé, Gideon Dror, David Vogel, Analysis of the KDD Cup 2009: Fast scoring on a large orange customer database, in: JMLR Workshop and Conference Proceedings, vol.7, 2009, pp. 1–22.

[24] P. Moscato, Memetic algorithms: a short introduction, in: D. Corne, M. Dorigo, F. Glover (Eds.), New Ideas in Optimization, McGraw-Hill, Maidenhead, UK, 1999, pp. 219–234.

[25] I.S. Oh, J.S. Lee, B.R. Moon, Hybrid genetic algorithms for feature selection, IEEE Transactions on Pattern Analysis and Machine Intelligence l 26 (11) (2004) 1424–1437.

[26] Zexuan Zhu, Yew-Soon Ong, Memetic algorithms for feature selection on micro array data, Advances in Neural Networks – ISNN (2007) 1327–1335.

[27] Z. Zhu, Y.S. Ong, M. Dash, Wrapper–filter feature selection algorithm using a memetic framework, IEEE Transactions on Systems, Man, Cybernetics B 37 (1) (2007) 70–76.

[28] Zhichun Wang, Minqiang Li, A hybrid genetic algorithm for simultaneous feature selection and rule learning, in: Fourth International Conference on Natural Computation, 2008, pp. 8-12.

[29] M.E. ElAlami, A filter model for feature subset selection based on genetic algorithm, Knowledge-Based Systems 22 (5) (2009) 356–362.

[30] Mark Hall, A decision tree-based attribute weighting filter for naive Bayes, Knowledge-Based Systems 20 (2) (2007) 120–126.

[31] S. Senthamarai Kannan, N. Ramaraj, An improved correlation-based algorithm with discretization for attribute reduction in data clustering, Data Science Journal 8 (2009) 125–138.