FEATURE SELECTION USING GENETIC ALGORITHM FOR SONAR IMAGES CLASSIFICATION WITH SUPPORT VECTOR

Hicham LAANAYA (1,2), Arnaud MARTIN (1), Ali KHENCHAF (1), Driss ABOUTAJDINE (2)

[1] ENSIETA, E^3I^2 EA3876
2 rue François Verny 29806 Brest Cedex 9, France
laanayhi, Arnaud.Martin, Ali.Khenchaf@ensieta.fr

[2] Faculté des sciences de Rabat, Morocco
aboutaj@fsr.ac.ma

ABSTRACT

This paper describes an approach being explored to improve the usefulness of machine learning techniques for generating classification rules for complex, real world data. The approach involves the use of genetic algorithms to select the best subset of features to be used by the classification system. This approach has been implemented and tested on sonar images classification. The results are encouraging and indicate significant improvements of the presented approach.
Keywords: sonar image, feature selection, genetic algorithm, classification, support vector machines.

1 INTRODUCTION

Automatic sonar images classification is one of the key areas of interest in the sonar image applications. For example, the images collected by sonar are particularly high-dimensional and difficult to characterize. It can be important to detect a specific kind of sediment, for example the rocks can be used as landmarks for images registration being used for underwater navigation.

Often, the ideal decision border between different classes in such sets is highly non-linear. As a result, training a classifier on such data sets is quite complicated: a large number of parameters have to be estimated using a limited number of samples. An alternative to deal with this problem is to select the appropriate features or attributes given the best classification results. The choice of features or attributes used to represent patterns presented to a classifier, affect:

- the accuracy of the classification function that can be learned using supervised classifier,
- the time needed for learning a classification function,
- the number of examples needed for learning a classification function.

This paper presents a feature subset selection problem in automated design of pattern classifiers. The feature subset selection problem refers the task of identifying and selecting a useful subset of features to be used to represent patterns from a larger set of often mutually redundant, possibly irrelevant.

Fig. 1 shows an automated classification system subdivided into three modules: feature extraction, feature selection, and classification.

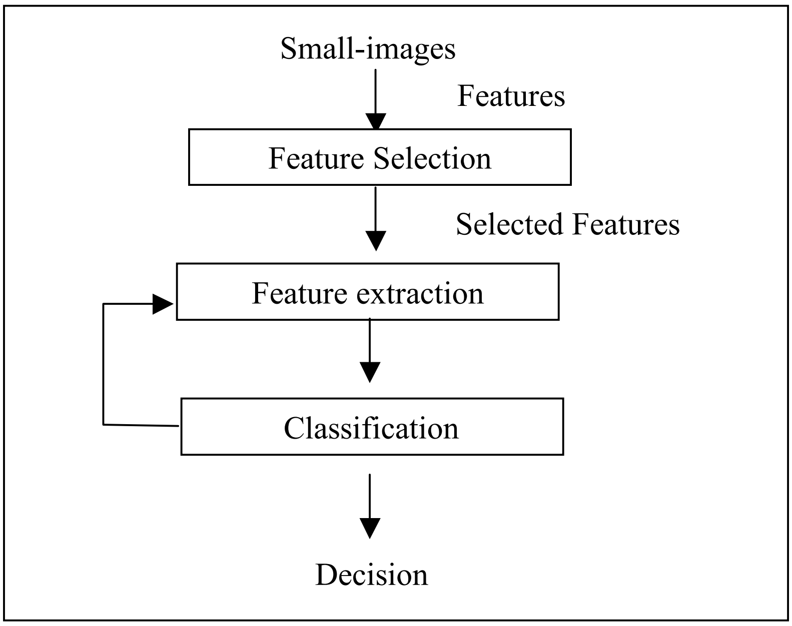


Fig. **1**. Automated classification system for sonar images.

A feature selection stage based on a genetic algorithm is added to the conventional framework to select an appropriate feature subset as inputs to the classifier.

The paper is organized as follow. In the first section we present a brief description of Genetic Algorithms for feature selection, then we present the chosen classifier based on Support Vector Machine (SVM). Finally we give some experimental results of our approach for sonar images classification.

2 GENETIC ALGORITHMS FOR FEATURE SELECTION

Feature subset selection algorithms can be classified into two categories based on whether or not feature selection is performed independently of the learning algorithm used to construct the verifier. If feature selection is done independently of the learning algorithm, the technique is said to follow a filter approach. Otherwise, it is said to follow a wrapper approach [1]. The first one is computationally more efficient but its major drawback is that an optimal selection of features may not be independent of the inductive and representational biases of the learning algorithm that is used to build the classifier. On the other hand, the wrapper approach involves the computational overhead of evaluating a candidate feature subset by executing a selected learning algorithm on the database using each feature subset under consideration.
Feature subset selection in the context of practical applications such as sonar images classification presents a multicriterion optimization function, *e.g.* number of features and accuracy of classification. Genetic algorithms formally introduced in 1970s by John Holland [2] offer a particularly attractive approach for this kind of problems since they are generally quite effective for rapid global search of large and poorly understood spaces. Moreover, genetic algorithms are very effective in solving large-scale problems [3].
So we use them here for subset selection [4,5] in the context of our sonar images classification problem..
Each image is represented as a vector of features. Many features extraction approach can be considered for sonar images classification [6]. The features extraction is calculated after wavelet decomposition presented in section 4.2. By this approach we obtained a features vector of 63 parameters. In our encoding scheme, the chromosome is a bit string whose length

is determined by the number of parameters in the image. Each parameters is associated with one bit in the string. If the i^{th} bit is 1, then the i^{th} parameter is selected, otherwise, that component is ignored (**Fig. 2**). Each chromosome thus represents a different parameter subset.

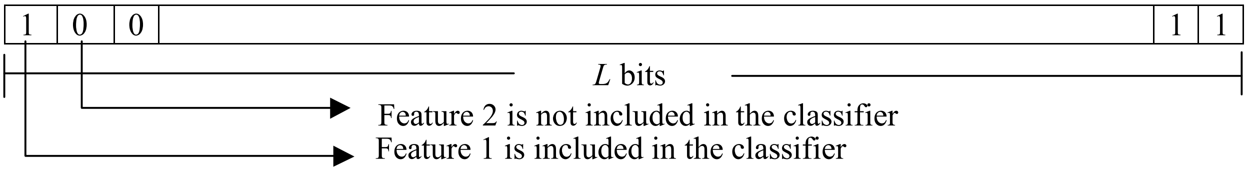


Fig. 2. A *L*-dimensional binary vector

In general the initial population is generated randomly. The goal of feature subset selection is to use fewer features to achieve the same or better performance. Therefore, the fitness evaluation contains two terms:

- accuracy from the validation data and,
- number of features used.

Only the features in the parameter subset encoded by an individual are used in order to train a classifier. The performance of the classifier is estimated using a validation data set and used to guide the genetic algorithm. Each feature subset contains a certain number of parameters. Between accuracy and feature subset size, accuracy is our major concern. Combining these two terms, the fitness function is given by:

$$Fitness = 10^4 \times Accuracy + 100 / d \times Zeros \quad (1)$$

where *Accuracy* is the accuracy rate that an individual achieves, *Zeros* is the number of zeros in the chromosome and *d* the number of features.

Overall, higher accuracy implies higher fitness. Also, fewer features used imply a greater number of zeros, and as a result, the fitness increases. Notice that individuals with higher accuracy would outweigh individuals with lower accuracy, no matter how many features they contain.

Mutation and *crossover* are two of the most commonly used operators with genetic algorithm that represent individuals as binary strings. Mutation operates on a single string and generally changes a bit at random. Crossover operates on two parent strings to produce two offspring.

3 SVM CLASSIFICATION

In the classification task, the images are analyzed in order to be separated. This process uses some features of the images to differentiate every one from the others. This way, the images can be classified in several classes with some characteristic in common.

Then the classification of sediments can be done using anyone of well-known classification techniques. One of them is the SVM given a simple way to obtain good classification results with a reduced knowledge. So, the used classification is based on the SVMs classification algorithm. The principle of SVMs has been developed by Vapnik [7] and has been used in several applications [8]. The classification task is reduced to find a decision border dividing the data into the groups representing the separated classes. The simplest decision case is when the data can be divided into two groups.

Consider the problem where the vectors can be divided into two sets. We must find the optimal decision border that separates these two sets of images. This optimal election will be the one that maximizes the distance from the border to the data. In the two dimensional case, the border will be a line, and in a multidimensional space the border will be an hyperplane. The searched decision function has the form given by:

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i < x_i, x > +b. \qquad (2)$$

The y values of this expression are +1 for positive classification training vectors (representing one class) and -1 for the negative training vectors (representing the other class). Also, the inner product is performed between each training input and the vector that must be classified. Thus, we need a set of training data (x,y) in order to find the classification function. The values α_i are the Lagrange multipliers, b a constant value obtained by the minimization process and the l value will be the number of vectors in the training database. These vectors with a value not equal to zero, are known as support vectors. In our case, the x represents one image from the sonar images training database (in the space of features) and y represents the predicted kind of sediment present on the x image. The (x_i,y_i) represent the images of the training database and there associated kind of sediments. When the data are not linearly separable this scheme cannot be used directly. To avoid this problem, the SVMs can map the input data into a high dimensional features space. The SVM constructs an optimal hyperplane in the high dimensional space and then returns to the original space transforming this hyperplane in a non-linear decision border. The non-linear expression for the classification function is given in the following equation:

$$f(x)=\sum_{i=1}^{l}\alpha_i y_i K(x_i,x)+b, \qquad (3)$$

where K is the kernel that performs the non-linear mapping. The choice of this non-linear mapping function or kernel is very important in order to obtain good classification performance. But, there are no methods to do this choice. One kernel used in our previous work [8] and in this paper; is the radial basis function. This function has the expression given by:

$$K(x,y)=\exp(-\gamma(x-y)^2) \qquad (4)$$

where γ is a parameter that will be tuned by the user. When some data into the sets cannot be separated, the SVM can include a penalty term in the minimization, which makes more or less important the misclassification. The greater is this parameter, the more important is the misclassification error into the minimization procedure.

This approach can be generalized to more then two classes [9, 10] where we can quote different methods:

- *One-vs-one:* we seek to separate each class from the others, and then we fuse the results,
- *One-vs-rest:* we made a classifier for each two classes, and then we fuse the results,
- the direct approach, where we considered directly all the classes.

Report to [9] for more details on this three first methods. Another method can be considered: Direct Acyclic Graph (*DAG*), see [10,11] for details.

4 EXPERIMENTS

4.1 The database

We seek in this article to classify sediments using a sonar images database that we carried out. Database consists of 26 sonar images provided by GESMA (*cf.* **Fig. 3** for an example of one image) cut to 4249 small-images of size 64x64, on which we indicated the kind of sediments (sand, rock, ripple, silt, and cobbles), or the non existence of information when there is a zone in the shade (labelled shadow) (*cf.* **Tab. 1**). On the **Tab. 2** we effective and percentage of each kind of sediment. We notice that the sand sediment is the most represented one. The cobbles sediments are under-represented. A major classification difficulty is due to this difference of effective. Moreover several sediments can appear on a same image (named patch-worked images), see **Tab. 1**. There is 39.7% of patch-worked small-images.

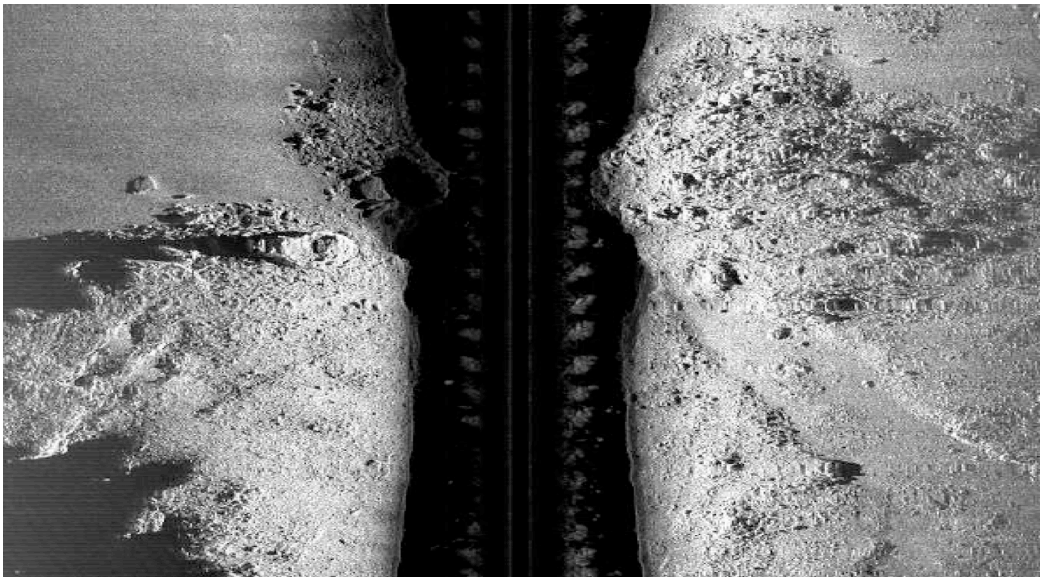
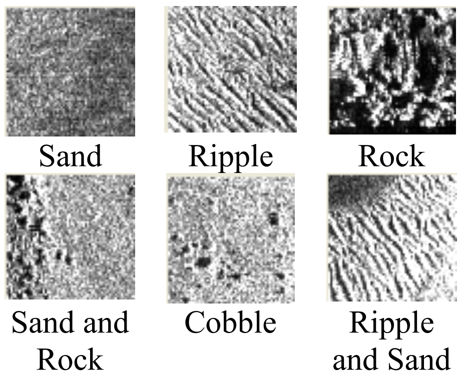


Fig. 3. Example of lateral sonar image (provided by GESMA).



Tab. 1. Sample of small-images with different type of sediment.

Sediment	Effective	%
Sand	2321	21.35
Rock	915	54.62
Ripple	374	8.80
Silt	234	5.50
Cobbles	33	0.77
Shadow	102	2.40
Total	4249	100.00

Tab. 2. Database elements and their effective.

Notice that such a database is quite difficult to realize. Indeed, the expert has a subjective experience, and he can make mistakes on some small-images, even if he has a perception of the global sonar image. So we only have a subjective perception of reality.
SVM classification is a supervised approach. So we need to build a training and test database from our data. **Tab. 3** presents effective of the training (Tr. DB.) and test (Ts. DB.) database obtained randomly in order to get 1/3 of data for training part and 2/3 of data for test part.

	Rock	Cobbles	Sand	Ripple	Silt	Shadow	Total
Tr. DB	319	18	971	147	23	79	1557
Ts. DB	596	15	1350	227	211	293	2692

Tab. 3: Training and Test database effective.

4.2 Feature extraction

In order to extract irrelevant information in the sonar images, many features extraction approach can be considered [6]. Here, we have used the discrete translation invariant wavelet transform [12]. It is based on the choice of the optimal translation for each decomposition level. Each decomposition level d gives four new images. We choose here a decomposition level d=3. For each image I_d^i (the i^{th} image of the decomposition d) we calculate three parameters. The energy is given by:

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} I_d^i(n,m)\,, \tag{5}$$

where N and M are respectively the number of pixels on the rows, and on the columns. The entropy is estimated by:

$$-\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M} I_d^i(n,m)\ln\left(I_d^i(n,m)\right), \tag{6}$$

and the mean is given by:

$$\frac{1}{NM}\sum_{n=1}^{N}\sum_{m=1}^{M}\left|I_d^i(n,m)\right|. \tag{7}$$

So we obtain 63 features Each small-image is then represented in a 63-dimension space.

4.3 Application of SVM on the rough data

The SVM classifier used on our experiments is *libsvm* given in [13]. This algorithm use the *one-vs-one* multi-class approach, the choice of this approach is explained in [10]. We use SVM with Gaussian kernel, and the SVM classifier was trained using the training database after wavelet decomposition with 63 parameters (the wavelet decomposition level used is 3). On **Tab. 4**, we present the obtained effective of each kind of sediment after the tests made on our test database.

We have obtained a global classification rate of 61.74% defined as the number of good classified small-images on the total of small images. Notice that no cobbles small-images are detected. 1099 of 1350 (81.40%) of the sand small-images are detected, 62.41% of the rock small-images are well classified and 58.02% of the shadow small-images are detected. We note a low rate of detection for the two sediments, silt and ripple; indeed, only 6.60% (respectively 2.84%) of the ripple (respectively silt) small-images are detected. The classifier tends to classify all the images in the two classes, sand and rock small-images, both majority classes of the database in terms of effective.

		References					
Tests	Class Name	Sand	Rock	Ripple	Silt	Cobbles	Shadow
	Sand	1099	169	15	8	0	59
	Rock	165	372	46	1	0	12
	Ripple	1099	169	15	8	0	1
	Silt	30	170	5	6	0	0
	Cobbles	144	3	15	6	0	0
	Shadow	73	49	0	0	0	170

Tab. 4. Confusion matrix for the rough sonar data (best classification rate is 61.74%).

4.4 Application of SVM after the application of Genetic Algorithm

Before training the classifier, we apply genetic algorithm for feature selection over the 63 features calculated by wavelet transform. In the genetic algorithm-based feature selection method used in this study, each individual in the genetic algorithm population represents a feature subset as a binary string. A "0" in the i^{th} position indicates that the i^{th} original feature is excluded from the feature subset, and a "1" indicates that the feature is present. To evaluate the fitness of an individual, the selected feature subset is fed into the SVM classifier. The genetic algorithm fitness function combined two optimization criteria:

1) minimization of the error rate of the classifier, defined as the number of good classified small-images on the total of small images,
2) minimization of the number of features.

The following parameters were used: P = 4 (population size); p_c =0.6 (probability of crossover); p_m =0.01 (probability of mutation) G = 10 (number of generations). **Tab. 5** represents the obtained results.

		References					
	Class Name	Sand	Rock	Ripple	Silt	Cobbles	Shadow
Tests	Sand	1095	141	37	29	3	45
	Rock	58	468	27	28	0	15
	Ripple	108	77	38	0	0	4
	Silt	40	111	9	51	0	0
	Cobbles	11	4	0	0	0	0
	Shadow	137	24	4	1	0	127

Tab. 5. Confusion matrix after genetic algorithm (best classification rate is 66.08%).

We have obtained a global classification rate of 66.08%. 81.11% of sand small-images are well classified a rate rather than on the rough data, and 78.52% of rock small-images are detected. Once again, no cobbles small-images are detected as in the rough data. We obtained a classification rate of 16;74% for ripple small-images and of 24.17% for silt small-images. Thus, we have obtained a classification rate better than the classification rate obtained by applying SVM on our rough database.

5 CONCLUSION

In this paper, we have presented and used a new method of feature selection in the context of sediment classification. We have shown that the application of genetic algorithms to feature selection gives better results than without. We have noticed that this approach detects a small part of the three classes (silt and ripple) and cannot detect the cobble small-images that are to few represented. Genetic algorithms are time consuming, but we used them only in the learning stage; in the test stage, we project the data in the space created by the best chromosome. By this way, the classifier gives the class for each individual.

Another approach is actually under study. It consists in the combination of other methods of feature extraction like co-occurrence matrices, and a feature selection method like genetic algorithm: feature extraction to get the relevant parameters of the features (images for example) and feature selection to get the best parameters that gives a best rate of classification.
As cited in the experiments part, we have used the default parameters of the SVM classifier,

another approach is to use genetic algorithm to tune the SVM classifier parameters (C, gamma for gaussian kernel or d the degree of polynomial kernel).
Another problem is the patch-worked small-images. We are working on the realization of a new repartition of the data with a previous manual segmentation of the sediment.

6 REFERENCES

[1] G. John, R. Kohavi, and K. Pfleger. «Irrelevant features and the subset selection problems». In *11th International Conference on Machine Learning*, pages 121–129, 1994.

[2] J.H. Holland «Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control, and Artificial Intelligence», 2nd edition, MIT Press / Bradford Books, 1992.

[3] M. Kudo and J. Sklansky. «Comparison of algorithms that select features for pattern classifiers». *Pattern Recognition*, 33(1):25–41, 2000.

[4] D.E. Goldberg. «*Genetic Algorithms in Search, Optimization & Machine Learning*». Addison-Wesley, 1989.

[5] J. Yang, and V.G. Honavar, «Feature Subset Selection Using a Genetic Algorithm», IEEE Intelligent Systems, 1998, pages 44-49.

[6] A. Martin, G. Sévellec, and I. Leblond, «Characteristics *vs* decision fusion for sea-bottom characterization», Colloque Caractérisation in-situ des fonds marins, Brest, France, 21-22 October 2004.

[7] V.N. Vapnik, «*Statistical Learning Theory*», John Wesley and Sons, 1998.

[8] C. Archaux, H. Laanaya, A. Martin, and A. Khenchaf, «An SVM based churn detector in prepaid mobile telephony». International Conference On Information & Communication Technologies (ICTTA), Damas, Syrie, pp 19-23, 2004.

[9] J. Weston & C. Watkins, Multi-class support vector machines, Technical Report CSD-TR-98-04, Royal Holloway, University of London, Department of Computer Science, 1998.

[10] C.W. Hsu, and C.J. Lin. «A comparison of methods for multi-class support vector machines». Technical report, Department of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, 2001.

[11] J.C. Platt, N. Cristianini, and J. Shawe-Taylor. «Large margin DAGs for multiclass classification». In S. A. Solla, T. K. Leen, and K.-R. Müller, editors.

[12] K.I. Kim, K. Jung, S.H. Park, and H.J. Kim, «Support Vector Machines for Texture Classification», *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 11, November 2002.

[13] C.C. Chang, and C.J. Lin, «*LIBSVM* - A Library for Support Vector Machines» http://www.csie.ntu.edu.tw/~cjlin/libsvm/