Dimensionality reduction of sonar images for sediments classification

[1 2] Hicham LAANAYA, [1] Arnaud MARTIN, [1] Ali KHENCHAF,
[2] Driss ABOUTAJDINE

[1] ENSIETA - E^{3}I^2 EA3876,
2 rue Fran cois Verny, 29806 Brest Cedex 9, France
{laanayhi, arnaud.martin, ali.khenchaf}@ensieta.fr,
[2] Université Mohamed V-Agdal,
Faculté des sciences de Rabat, Maroc
aboutaj@fsr.ac.ma

Abstract *Data in most of the real world applications like sonar images classification are high dimensional and learning algorithms like Support Vector Machines (SVMs) have problems in handling high dimensional data. Here, we show that a non-linear projection method called curvilinear component analysis can effectively reduce the original dimension to a lower dimension. We apply this approach for dimensionality reduction of the sonar images and use SVMs classifiers for sediments classification.*

1 introduction

In many real-world classification problems, high-dimensional data sets are collected, e.g. from sensors. For example, the images collected by a sonar are particularly hight-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. A classifier should therefore have many degrees of freedom, and consequently a large number of parameters. As a result, training a classifier on such data sets is quite complicated: a large number of parameters has to be estimated using a limited number of samples.

One can overcome this problem by first mapping the data to a high-dimensional space in which the classes become (approximately) linearly separable. Kernel-based techniques, such as SVMs, are typical examples of this approach. An alternative is to lower the data dimensionality, rather than increase it. Although it might seem information is lost, the reduction in the number of parameters

one needs to estimate can result in better performance. Many linear methods for performing dimensionality reduction, such as principal component analysis (PCA) and linear discriminant analysis (LDA) are well-established in literature. Here, a non-linear dimensionality reduction method called curvilinear component analysis (CCA, [1]) is considered. The main assumption behind CCA is that the data set is sampled from a (possibly non-linear) manifold, embedded in the high-dimensional space. CCA is an unsupervised method and brings some improvements to Sammon's mapping [2]. Actually, when unfolding a non-linear structure, Sammon's mapping cannot reproduce all distances. One way to face this problem consists in favoring *local topology*: CCA tries to reproduce short distances firstly, long distances being secondary. We present the principle of this method in section 2, SVMs and experimental results are presented, respectively, in section 3 and section 4.

2 General framework of curvilinear component analysis

This section introduce CCA by explaining the basic idea of this method. Although CCA has been already presented in a number of works [1]: CCA takes a set of high-dimensional data and maps them into a low-dimensional space while preserving *local topology* structure of the data. In the CCA, the topology is defined by the distances between all pairs of vectors of original data. Since the topology cannot be entirely reproduced in the projection subspace, witch has a lower dimension than the original subspace, the local topology, the most important, is favored to the detriment of the global topology. The goal of CCA is then to minimize an error function witch characterizes the difference of topology between the original subspace (the space of sonar images x_i) and the projection subspace (y_i):

$$E = \frac{1}{2}\sum_{i}\sum_{j\neq i}(X_{ij} - Y_{ij})^2 F_\lambda(Y_{ij}). \tag{1}$$

with: X_{ij}: represent the euclidian distance between the sonar images x_i and x_j.

Y_{ij}: euclidean distance between the projections y_i and y_j of the sonar images x_i and x_j in the projection subspace $\mathbb{R}^d$ with d its dimension.
$F_\lambda : \mathbb{R}^+ \to [0,1]$ is a decreasing function of its argument, so it is used to favor local topology preservation. For example, F_λ could be a step, exponential or sigmoid function of Y_{ij}. We use the first function because it takes just two values 0 or 1.

The gradient of E is given by the equation 2:

$$\nabla_i E = \sum_{j\neq i} \frac{X_{ij} - Y_{ij}}{Y_{ij}} [2F_\lambda(Y_{ij}) - (X_{ij} - Y_{ij})F'_\lambda(Y_{ij})](y_j - y_i), \tag{2}$$

with $\nabla_i E$ denotes the gradient of E with respect to y_i.

A minimization of E by gradient descent gives the adaptation rule:

$$\Delta y_i = \alpha(t) \sum_{j \neq i} \frac{X_{ij} - Y_{ij}}{Y_{ij}} [2F_\lambda(Y_{ij}) - (X_{ij} - Y_{ij})F'_\lambda(Y_{ij})](y_j - y_i). \quad (3)$$

where $\alpha(t)$ is an adaptation factor that evolves with time.
this rule has several defects:

- for each i, we will calculate a sum over all the j, thus the complexity is $O(N^2)$, with N the dimension of the feature space.
- the process can fall into a local minimum of E.

A simple procedure to avoid these defects consists in choosing a random point y_i, and all the $y_{j \neq i}$ are moved with respect to y_i. The new adaptation rule is given by (method of the modified gradient):

$$\forall j \neq i \quad \Delta y_j = \alpha(t) F_\lambda(Y_{ij})(X_{ij} - Y_{ij}) \frac{y_j - y_i}{Y_{ij}}. \quad (4)$$

the complexity is only in $O(N)$ instead of $O(N^2)$.

We find several applications of the CCA, for example [6] used this method for the reduction of the dimensionality for the detection of the person gender using faces. We find another application in the representation of the phonemes in [1].

3 SVMs classification

In classification task, the different images are separated in order to analyze the information in some way the information that contains. This process uses some characteristics 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. The classification is a task where every images is classified or labeled into several groups.

Then the classification of sediments can be done using anyone of well-known classification techniques. One of theses techniques is the SVMs that give us a simple way to obtain good classification results with a reduced knowledge of the problem. The principles of SVMs have been developed by Vapnik [4] and have been presented in several works like [3]. The classification task is reduced to find a decision border that divide the data into the groups that we want to separate. The simplest decision case is when the data can be divided into two groups. The work presented here is based on the SVMs classification algorithm presented in [5].

Bases of Support Vector Machines In the simplest decision problem we have a number of vectors divided into two sets, and we must find the optimal decision border to divide these sets. 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, in a multidimensional space the border will be an hyperplane. The searched decision function has the next form:

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i < x_i, x > +b. \tag{5}$$

The y values that appear into this expression are +1 for positive classification training vectors and -1 for the negative training vectors. 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 are the Lagrange multipliers obtained in the minimization process and the l value will be the number of vectors that in the training process contribute to form the decision border. These vectors are those with a value not equal to zero and are known as support vectors. On our case, the x represents one image from the sonar images training database 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 kind of sediments.

When the data are not linearly separable this scheme can not be used directly. To avoid this problem, the SVMs can map the input data into a high dimensional feature space. The SVMs 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 (6) where K is the kernel that performs the non-linear mapping.

$$f(x) = \sum_{i=1}^{l} \alpha_i y_i K(x_i, x) + b. \tag{6}$$

The choice of this non-linear mapping function or kernel is very important in the performance of the SVMs. One kernel used in our previous work is the radial basis function. This function has the expression given in (7).

$$K(x, y) = exp(-\gamma(x - y)^2). \tag{7}$$

where γ is a parameter that will be tuned by the user.

When some data into the sets can not be separated, the SVMs can include a penalty term in the minimization, that makes more or less important the misclassification.The greater is this parameter the more important is the misclassification error into the minimization procedure.

4 Experiments

4.1 The database

We seek in this article to classify sediments using a sonar images database which we carried out. It consists of 26 sonar images provided by GESMA (cf. Fig.

1 for an example of that images) cut to 4249 small-images of size 64x64, on which we indicated the kind of sediments (sand, rock, cobbles, ripple, silt and shadow), or the non existence of information when there is a zone in the shade. Moreover several sediments can appear on a same image, which we informed by the existence of a border or not (cf. Fig. 2).

Figure 1: Example of sonar image (provided by GESMA).

Thus, the knowledge discovery using such data is very important for the expert. The approach employed to solve the problem of classification of marines sediments is based on the method of Support Vector Machines (SVM) [3]. On the table 2 we represent some information about our database where B. A. (resp. B. T.) represents training (resp. test) database.

4.2 Application of SVM on the rough data

We have trained our SVM classifier using the training database (B.A.) without any processing, we have used a gaussian kernel. On the table 3 we present some results that we obtained after the tests that we made on our test database (B.T.)

We have obtained a classification rate of 61.74%, we note that no cobbles

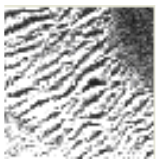
 Sand	 Ripple	 Rock
 Sand & Rock	 Silt	 Ripple with border

Table 1: Example of different kind of small-images.

	Rock	**Cobbles**	**Sand**	**Ripple**	**Silt**	**Shadow**	**Total**
B.A	319	18	971	147	23	79	1557
B.T	596	15	1350	227	211	293	2692
B.A. et Fr	153	16	291	94	15	14	583
B.T. et Fr	266	12	472	136	67	151	1104

Table 2: Manual segmentation of our database.

		References					
	Class name	Rock	Cobbles	Sand	Ripple	Silt	Shadow
Tests	Rock	372	0	165	46	1	12
	Cobbles	3	0	11	1	0	0
	Sand	169	0	1099	15	8	59
	Ripple	67	0	144	15	0	1
	Silt	170	0	30	5	6	0
	Shadow	49	0	73	0	1	170

Table 3: Confusion matrix for the rough sonar data (best classification rate is 61.74%)

image is detected. 1099 of 1350 (81.40%) of the sand images are detected, 62.41% of the rock images are well classified and 58.02% of the shade images are detected, we note a low rate of detection for the two sediments, silt and

ripple, indeed, only 6.60% (resp. 2.84%) of the ripple images (resp. silt) are detected. The classifier tends to classify all the images in the two classes, sand images and rock images. the two majority classes of the database.

4.3 Application of SVM after the application of CCA

Before training the classifier, we reduced the dimension of our data by applying the CCA with a dimension of 5. The table 4 represents the obtained results.

		References					
	Class name	Rock	Cobbless	Sand	Ripple	Silt	Shadow
Tests	Rock	27	8	438	1	115	7
	Cobbles	0	0	15	0	0	0
	Sand	30	2	1180	1	88	49
	Ripple	4	1	221	0	0	1
	Silt	22	1	56	0	132	0
	Shadow	4	0	110	1	34	144

Table 4: Confusion matrix after applying CCA on our database ($d = 5$) (the classification rate is 55.08%)

We obtained a classification rate of 55.08%, a rate lower than the classification rate obtained by applying SVM on our rough database. One gained in computing times but one lost on the classification rate; theses results can be explained by the fact that we haven't obtained a good representation of our data in a space of low dimension.

5 Conclusion

We have used here a new method of dimensionality reduction for sediment classification. We have shown that the CCA applied on our database can't give a best results, this results can explained by the fact that we lost information during projection by the CCA.

An other approach that we can use is the combination between a feature extraction method like wavelet decomposition or co-occurrence matrices and a feature selection method like genetic algorithm for feature selection: 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.

References

[1] P. Demartines, & J. Hérault, *Curvilinear Component Analysis: a self-organizing neural network for non-linear mapping of data sets.* IEEE Trans-

action on Neural Networks 8(1), pp 711-720, 1998.

[2] J.W. Sammon, *A Non-Linear Mapping for Data Structure Analysis*, IEEE Transaction on Computers, C-18(5), 1969.

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

[4] V. N. Vapnik, Statistical Learning Theory, John Wesley and Sons, 1998.

[5] 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.

[6] S. Buchalou, N. Davey, R.J. Frank & T.M. Gale, *Dimensionality Reduction of Face Images for Gender Classification*, Proceedings of IEEE IS, 2004.