

A comparison of PCA and GA selected features for cloud field classification

Miguel Macías-Macías¹, Carlos J. García-Orellana¹, Horacio M. González-Velasco¹, Ramón Gallardo-Caballero¹, Antonio Serrano-Pérez²

¹ Departamento de Electrónica e Ingeniería Electromecánica - Universidad de Extremadura,
Avda. de Elvas s/n, 06071. Badajoz, Spain.
(miguel, carlos, horacio, ramon)@nernet.unex.es

² Departamento de Física - Universidad de Extremadura,
Avda. de Elvas s/n, 06071. Badajoz, Spain.
(asp)@unex.es

Abstract. In this work a back propagation neural network (BPNN) is used for the segmentation of Meteosat images covering the Iberian Peninsula. The images are segmented in the classes land (L), sea (S), fog (F), low clouds (C_L), middle clouds (C_M), high clouds (C_H) and clouds with vertical growth (C_V). The classification is performed from an initial set of several statistical textural features based on the gray level co-occurrence matrix (GLCM) proposed by Welch [1]. This initial set of features is made up of 144 parameters and to reduce its dimensionality two methods for feature selection have been studied and compared. The first one includes genetic algorithms (GA) and the second is based on principal component analysis (PCA). These methods are conceptually very different. While GA interacts with the neural network in the selection process, PCA only depends on the values of the initial set of features.

1 Introduction

In order to understand and model the radiation balance in the climatic system a very accurate information of the cloud cover is needed. Clouds play an important role reflecting the solar radiation and absorbing thermal radiation emitted by the land and the atmosphere, therefore reinforcing the greenhouse effect. The contribution of the clouds to the Earth albedo is very high, controlling the energy entering the climatic system. An increase in the average albedo of the Earth-atmosphere system in only 10 percent could decrease the surface temperature to levels of the last ice age. Therefore, global change in surface temperature is highly sensitive to cloud amount and type.

These reasons make that numerous works about this topic have been published in the last years. Many of them deal with the search of a suitable classifier. Welch [1] used linear discrimination techniques, Lee et al. [2] tested a back-propagation neural network (BPNN), Macías et al. at [3] showed that the classification results obtained with a BPNN were better than those obtained with a SOM+LVQ neural network, and at [4] they used a BPNN to studying the evolution of the cloud cover over Cáceres

(Spain) along 1997. Bankert et al. [5] and Tiam et al. [6][7] used a probabilistic neural network (PNN). In [8] linear discrimination techniques, and PNN and BPNN neural networks were benchmarked and the results showed that BPNN achieves the highest classification accuracy.

Other works are related with the search of an initial feature set that allow to obtain reliable classification results. First works used simple spectral features like albedo and temperature. Later works include textural features too since they are less sensitive to the effects of atmospheric attenuation and detector noise than the first ones [9]. In [1] Welch et al. used statistical measures based on gray level co-occurrence matrix (GLCM) proposed by Haralick et al. in [10]. In [6] several image transformation schemes as singular value decomposition (SVD) and wavelet packets (WP's) were exploited. In [11] Gabor filters and Fourier features are recommended for cloud classification and in [6] authors showed that SVD, WP's and GLCM textural features achieved almost similar results.

In spite of it, the initial set of features and the classifier proposed in each work is very dependent on the origin of the images (season, satellite type, location on the Earth, etc.) that have been used.

In this work we propose a BPNN neural network and GLCM textural features for the segmentation of Meteosat images covering the Iberian Peninsula. The initial GLCM feature set consists of 144 features. Because of the finite size of the prototypes set and in order to remove the redundancy in these features, a selection process has to be used.

In that sense, in [12] Doak identifies three different categories of search algorithms: exponential, sequential and randomised. In [13] Aha et al. use the most common sequential search algorithms for feature selection applied to the clouds classification: the forward sequential selection (FSS) and the backward sequential selection (BSS). In [14], [15] and [16] a genetic algorithm (GA) representative of the randomised category is used for feature selection. They use GA because it is less sensitive than other algorithms to the order of the features that have been selected.

All these algorithms interact with the network in the selection process. Thus, it seems that this process is going to be very dependent of the prototypes selection and classification by the Meteorology experts. This process is particularly problematic in this application, since clouds of different types could overlap on the same pixel of the image. Taking into account this drawback, feature selection algorithms not dependent on the labelled of the prototypes, as principal component analysis (PCA), acquire a notable interest for comparison studies.

Therefore, in this work we want to compare the classification results obtained from the two previously mentioned feature selection methods. In section 2 we show the methodology followed in all the process, namely, the pre-processing stage, a brief of the PCA feature selection algorithm and the characteristics of our GA feature selection algorithm. In section 3 the classification results with both of the feature selection methods are given and finally the conclusion and comments are presented in section 4.

2 Methodology

In this paper images from the geostationary satellite Meteosat are used. This satellite gives multi-spectral data in three wavelength channels. In this work two of them, the visible and infrared channels, are used. The subjective interpretation of these images by Meteorology experts suggested us to consider the following classes: sea (S), land (L), fog (F), low clouds (C_L), middle clouds (C_M), high clouds (C_H) and clouds with vertical growth (C_V).

For the learning step of the neural models, several Meteorology experts selected a large set of prototypes. These are grouped into rectangular zones, of such form that, each of these rectangular zones contains prototypes belonging to the same class. For this selection task a specific plug-in for the image-processing program GIMP was implemented.

In order to compare the classification results obtained by the two feature selection algorithms and to carry out the GA feature selection process, the set of prototypes was divided into a training set, a validation set and a test set. For obtaining an optimal neural network with good generalization, we started from a BPNN with very few neurons in its hidden layer. This network was trained with the training set. The learning process stops when the number of misclassifications obtained on the validation set reaches a minimum. After that, the process was repeated by increasing the network size. The new network is considered optimal if the number of misclassifications over the validation set is lower than the previous one. Finally, we select the optimal feature selection algorithm according to the classification results on the test set.

For the training of the BPNN, the Resilient Backpropagation RProp algorithm described in [17] is used. Basically this algorithm is a local adaptive learning scheme which performs supervised batch learning in multi-layer perceptrons. It differs from other algorithms since it considers only the sign of the summed gradient information over all patterns of the training set to indicate the direction of the weight update. The different simulations were performed by means of the freeware neural networks simulation program SNNS (Stuttgart Neural Network Simulator).

2.1 Preprocessing stage.

Our final aim is the definition of a segmentation system of images corresponding to different times of the day and different days of the year. Therefore, satellite data must be corrected in the preprocessing stage to obtain physical magnitudes which are characteristic of clouds and independent of the measurement process.

From the infrared channel, we obtained brightness temperature information corrected from the aging effects and transfer function of the radiometer. From the visible channel we obtained albedo after correcting it from the radiometer aging effects and considering the viewing and illumination geometry. This correction deals with the Sun-Earth distance and the solar zenith angle at the image acquisition date and time, and the longitude and latitude of the pixel considered. In [7] no data correction is made and an adaptive PNN network is proposed to resolve this issue.

Next, from the albedo and brightness temperature data, which are already characteristic of the cloud, 144 statistical measures based on gray level co-occurrence matrix (GLCM) were computed. These measures constitute the characteristic vector for each pixel in the image. The large dimensionality of this vector and the limited quantity of prototypes available lead us to the case where the sparse data provide a very poor representation of the mapping. This phenomenon has been termed *the curse of dimensionality* [18]. Thus, in many problems, reducing the number of input variables can lead to improved performances for a given data set, even though some information is being discarded. Therefore, this process constitutes one of the fundamentals steps of the preprocessing stage and also one of the most significant factors in determining the performance of the final system.

In the next sections we are going to describe briefly the algorithms used for reducing the dimensionality of the input vector. Two different methods will be applied, GA as representative of the algorithms that select a subset of the inputs and discard the remainder and PCA as representative of the techniques based on the combination of inputs together to make a, generally smaller, set of features.

2.2 PCA feature selection

Principal Components Analysis is one of the most known techniques of multivariate analysis [19]. Due to its versatility, this method has been used for many different purposes related to synthesizing information. This method starts with a large set of variables which are highly intercorrelated and defines new uncorrelated variables, which are linear combination of the initial ones, ordered by the information they account for.

In this study, the 144 mentioned statistical measures were calculated for 4420 pixel extracted from a set of 20 images chosen to be representative of all types of clouds, land and sea. The distance between selected pixels is, at least, five pixels, which means about 35 km for the region of study. This avoids considering too much redundant information.

Next, a PCA was performed with the correlation matrix of the 144 variables and 4420 cases. The correlation matrix was chosen as the dispersion matrix since the variables have different units. Thus, all variables have the same weight irrespective of their original variance. The most representative principal components (PCs) were selected according to the Kaiser's rule [20]. Then, the variable most correlated to each PC was chosen as representative of the information accounted for by the PC.

Since rotating PCs results in a less ambiguous classification of variables, the PCs were also rotated according to Varimax method [21]. This rotation was chosen since it is widely accepted as being the most accurate orthogonal rotation method.

Thus, finally, two sets of variables were selected, one for the case of unrotated PCs (PCANR) and other for the case of rotated PCs (PCAR).

2.3 GA feature selection

The GA algorithm [22] tries to select a subset of features that offer the neural network with the best generalization by using the prototypes selected and labelled by the ex-

perts in Meteorology. That is, the network that, trained with the prototypes of the learning set, achieves the minimum number of misclassifications over the validation set.

For each subset of features the algorithm uses one hidden layer perceptrons where the number of the neurons of the hidden layer changed from 20 till 40. For each topology the training process is repeated 20 times randomizing the weights each time. As fitness we have used the sum of squared error (SSE) over the validation set.

The GA was configured using a cross-over probability of 0.6, a mutation probability of 0.1, a population of 350 individuals, a tournament selection and a steady-state population replacement with a 30% of replacement.

The simulations were done in a Beowulf style cluster with Clustermatic as OS (a patched RedHat 7.2 Linux OS, with *bproc* for cluster management). The cluster is built using on master node, a double Pentium III @ 800 MHz with 1 Gbyte of RAM memory, and 25 nodes, with AMD Athlon @ 900 MHz with 512 Mbytes of memory each. For GA simulations we used the PGAPack [23] simulator with MPI enabled.

3 Results

In order to implement the processes described above, the experts in Meteorology selected 4599 prototypes, 2781 for the training set, 918 for the validation set and 900 for the test set. The prototype selection was made from the Iberian Peninsula Meteosat images corresponding to the years 1995-1998.

In the feature selection process the PCANR algorithm selected 8 variables, the PCAR 17 and the GA gave 13. In table 1 we can observe the set of parameters selected for each algorithm.

Table 1. Parameters selected for each algorithm

Algorithm	Number	Parameters
GA	13	113, 143, 83, 85, 72, 125, 110, 119, 88, 72, 17, 58, 40
PCAR	17	136, 25, 67, 94, 15, 22, 96, 126, 60, 121, 102, 84, 30, 50, 56, 86, 132
PCANR	8	140, 25, 22, 78, 12, 121, 56, 86

Once feature selection is made, we use a BPNN to make the comparison of the algorithms and to make the final classification. In order to select the network with the best generalization for each algorithm we take one hidden layer BPNN with variable number of neurons. We train the neural network with the training set and we calculate the SSE over the validation set each training iteration. The network that reaches a minimum of misclassification over the validation set is chosen as representative for this algorithm.

In the GA case the minimum value for the sum of squared error (SSE) over the validation set was $SSE_v=35$ and this value was reached with 23 neurons in the hidden layer. With the PCAR algorithm $SSE_v=136$ with 24 neurons in the hidden layer and,

finally, with the PCANR algorithm we used 48 neurons in the hidden layer to obtain a $SSE_v=196$.

In tables 2,3 and 4 the percentage of success over the seven classes defined in the learning process and the SSE calculated over the three subsets of prototypes by the network representative of each feature selection algorithm can be observed.

Table 2. Classification results over the learning set reached by the networks with the best generalization over the three sets of features.

Learning set								
Algorithm	F	C _L	C _M	C _H	C _V	L	S	SSE _L
GA	96.4	95.1	94.6	100	96.5	100	100	146
PCAR	87.4	86.1	89.6	98.4	89.2	100	100	347
PCANR	84.2	92.3	93.3	97.8	86.7	94.5	95.0	376

Table 3. Classification results over the validation set reached by the networks with the best generalization over the three sets of features.

Validation set								
Algorithm	F	C _L	C _M	C _H	C _V	L	S	SSE _v
GA	96.4	96	99.4	98.4	98.9	100	100	35
PCAR	83.2	87.9	88.7	98.4	87.8	90.3	100	136
PCANR	74.3	87.9	79.9	98.4	90	88.2	95.2	196

Table 4. Classification results over the test set reached by the networks with the best generalization over the three sets of features.

Test set								
Algorithm	F	C _L	C _M	C _H	C _V	L	S	SSE _T
GA	92.7	86.5	77.6	100	75.9	100	100	167
PCAR	84.6	69.7	78.8	97.3	76.8	97.6	100	194
PCANR	62.6	94.2	94.1	91.8	60.7	69.6	99.3	262

4 Conclusions

Since the feature selection algorithm interacts with the network in the selection process for the GA case, the minimum value for the SSE_v is lower than the minimum obtained with the other algorithms. But it also happens that the value of the SSE_T for the GA algorithm is the lowest obtained. Thus we propose the features selected by the GA algorithm to perform the future automatic segmentation of the Iberian Peninsula Meteosat images.

In Figure 1 an example of an Iberian Peninsula Meteosat image segmentation can be observed. For the final classification a new class, the indecision class (I), has been added. We consider that one pixel belongs to a class when the output of the neuron

representative of this class is bigger than 0.6 and the others outputs are least than 0.4. In other case the pixel is considered to belong to the indecision class.

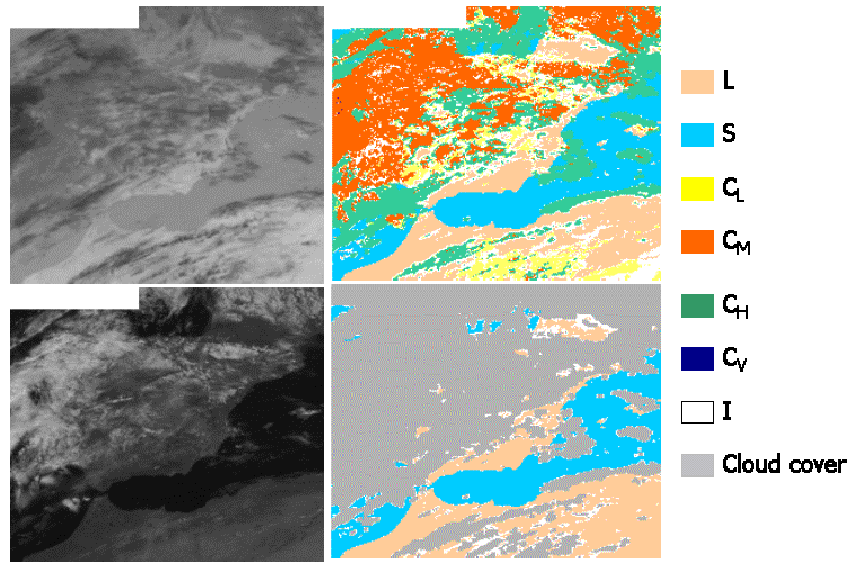


Fig. 1. Example of an Iberian Peninsula Meteosat image segmentation.

Acknowledgements.

Thanks are due to EUMETSAT for kindly providing the Meteosat data and to project 1FD970723 (financed by the FEDER and Ministerio de Educación).

References

1. Welch, R.M., Kuo K. S., Sengupta S. K., and Chen D. W.: Cloud field classification based upon high spatial resolution textural feature (I): Gray-level cooccurrence matrix approach. *J. Geophys. Res.*, vol. 93, (oct. 1988) 12633-81.
2. Lee J., Weger R. C., Sengupta S. K. And Welch R.M.: A Neural Network Approach to Cloud Classification. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 5, pp. 846-855, Sept. 1990.
3. M. Macías, F.J. López, A. Serrano and A. Astillero: "A Comparative Study of two Neural Models for Cloud Screening of Iberian Peninsula Meteosat Images", *Lecture Notes in Computer Science 2085*, Bio-inspired applications of connectionism, pp. 184-191, 2001.
- 4.- A. Astillero, A Serrano, M. Núñez, J.A. García, M. Macías and H.M. González: "A Study of the evolution of the cloud cover over Cáceres (Spain) along 1997, estimated from Meteosat images", *Proceedings of the 2001 EUMETSAT Meteorological Satellite Data Users' Conference*, pp. 353-359, 2001
5. Bankert, R. L et al.,: Cloud Classification of AVHRR Imagery in Maritime Regions Using a Probabilistic Neural Network. *Journal of Applied. Meteorology*, 33, (1994) 909-918.

6. B. Tian, M. A. Shaikh, M. R. Azimi, T. H. Vonder Haar, and D. Reinke, "An study of neural network-based cloud classification using textural and spectral features," *IEE trans. Neural Networks*, vol. 10, pp. 138-151, 1999.
7. B. Tian, M. R. Azimi, T. H. Vonder Haar, and D. Reinke, "Temporal Updating Scheme for Probabilistic Neural Network with Application to Satellite Cloud Classification," *IEEE trans. Neural Networks*, Vol. 11, no. 4, pp. 903-918, Jul. 2000.
8. R. M. Welch et al., "Polar cloud and surface classification using AVHRR imagery: An intercomparison of methods," *J. Appl. Meteorol.*, vol. 31, pp. 405-420, May 1992.
9. N. Lamei et al., "Cloud-type discrimination via multispectral textural analysis," *Opt. Eng.*, vol. 33, pp. 1303-1313, Apr. 1994.
10. R. M. Haralick et al., "Textural features for image classification", *IEEE trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610-621, Mar. 1973.
11. M. F. Aug.eijin, "Performance evaluation of texture measures for ground cover identification in satellite images by means of a neural.network classifier," *IEEE trans. Geosc. Remote Sensing*, vol. 33, pp. 616-625, May 1995.
12. Doak J. An evaluatin of feature selection methods and their application to computer security (Technical Repor CSE-92-18). Davis, CA: University of California, Department of Computer Science.
13. Aha, D. W., and Bankert, R. L.: A Comparative Evaluation of Sequential Feature Selection Algorithms. *Artificial Intelligence and Statistics V.*, D. Fisher and J. H. Lenz, editors. Springer-Verlag, New York, 1996.
14. Eng Hock Tay F. and Li Juan Cao, A comparative study of saliency analysis and genetic algorithm for feature selection in support vector machines", *Intelligent Data Analysis*, vol. 5, no. 3, pp. 191-209, 2001.
15. A. Tettamanzi, M. Tomassini. *Soft Computing. Integrating Evolutionary, Neural and Fuzzy Systems*. Springer, 2001.
16. F.Z. Brill, D.E. Brown and W.N. Martin. Fast genetic selection of features for neural network classifiers. *IEEE Transactions on Neural Networks*, 3(2): 324-328, 1992.
17. M. Riedmiller, M., Braun, L.: A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm. In *Proceedings of the IEEE International Conference on Neural Networks 1993 (ICNN 93)*, 1993.
18. R. Bellman: "Adaptive Control Processes: A Guided Tour". New Jersey, Princeton University Press.
19. I. T. Jolliffe: *Principal Component Analysis*, Springer-Verlag, 271 pp., 1986.
20. H. F. Kaiser: "The application of electronic computer to factor analysis", *Educ. Psychol. Meas.*, vol. 20, pp.141-151, 1960.
21. H. F. Kaiser, "The Varimax criterion for analytic rotation in factor analysis", *Psychometrika*, vol. 23, pp. 187-200, 1958.
22. D.E. Goldberg. *Genetic Algorithms in Search, Optimization & Machine Learning*. Addison Wesley, 1989.
23. D. Levine. *Users Guide to the PGAPack Parallel Genetic Algorithm Library*. Research Report ANL-95/18. Argonne National Laboratory, 1996.