Reduction of Irrelevant Features in Oceanic Satellite Images by means of Bayesian Networks

J. A. Piedra
Department of Languages and Computation
University of Almería, 04120 Almería
jpiedra@ual.es

A. Salmerón
Department of Statistics and Applied Mathematics
University of Almería, 04120 Almería
asalmero@ual.es

F. Guindos
Department of Languages and Computation
University of Almería, 04120 Almería
fguindos@ual.es

M. Cantón
Department of Languages and Computation
University of Almería, 04120 Almería
mcanton@ual.es

Abstract

This paper describes the use of Bayesian networks for the reduction of irrelevant features [1,2] in the recognition of oceanic structures in satellite images. Bayesian networks are used to validate the symbolic knowledge -provided by neuro symbolic or HLKPs (High Level Knowledge Processors) nets- and the numeric knowledge. This provides an automatic interpretation of images. The main objective of this work is the construction of an automatic recognition system for processing AVHRR (Advanced Very High Resolution Radiometer) images from NOAA (National Oceanographic and Atmospheric Administration) satellites to detect and locate oceanic phenomena of interest like upwellings, eddies and island wakes. With this aim, this paper reports on a methodology of knowledge selection and validation. In knowledge selection, filter measures are used. For knowledge validation, Bayesian networks (Naïve Bayes, TAN and KDB) are evaluated.

1. Introduction

A pattern classification system often involves two stages of development: *Extraction and selection of features* that can be used to discriminate between pattern classes, and *classification*, that draws class boundaries in the selected feature space. Feature selection consists of selecting an optimal or suboptimal feature subset from a set of candidate features.

The most common framework for feature selection is to define criteria for measuring the goodness of a set of features [3], and then use a search algorithm to find an optimal or sub-optimal set of features [4]. The objective is to use Bayesian networks for features selection in the recognition of oceanic structures in satellite images [5,6].

The paper is organized as follows. The data set used in this study is described in section 2. The symbolic and numeric data set is explained in section 3. A methodology for knowledge selection and validation by means of Bayesian networks from data is examined in section 4. The experiments carried out with this methodology are reported in section 5, and the paper ends with the conclusions in section 6.

2. Data set

The AVHRR sensor has been a powerful tool in environmental, climatic and geophysical research tasks for more than twenty years. This sensor, on board the Tiros and NOAA satellite series, covers 5 channels in the infrared and visible spectra. Particularly, infrared information has been used in oceanic feature identification.

AVHRR channels 2 and 4 provide visible and infrared information in the ranges 0.725-1.10 μm and 10.50-11.50 μm.

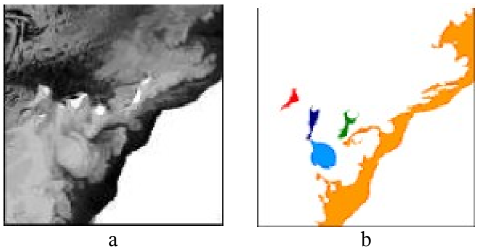

Figura 1. AVHRR scene and feature map

The region of NW Africa, Iberian Atlantic Coast and Mediterranean sea in which the study was carried out is at the easternmost part of the eastern Atlantic. A detailed oceanographic description of this area can be find in [7,8].
In the studied area there are different kinds of important oceanic mesoscalar structures: Upwellings, cold eddies, warm eddies and wakes.

Upwelling is the term used by oceanographers to describe the situation where cool, but nutrient-rich water from the lower layers of the ocean (too dark for plants to grow) raises towards the surface. Brought into the light, such waters become very fertile and rich feeding ground for fish. For an upwelling to occur, there must be a divergence of the surface currents, and this usually happens as a result of the wind field causing surface water drift in the presence of topographic constraints. For example, a longshore wind flowing in the Northern Hemisphere with a coastline to the left of the direction of travel, sets up a geostrophic flow away from the coast, and an upwelling tends to occur. This upwelling is a regular occurrence off the NorthWest African coast [7], and off the Peruvian coast, where wind conditions are suitable. Nonetheless, upwellings may well be intermittent, depending both on the weather and, to a certain extent, on the ocean-wide thermohaline circulation. Because of the importance to commercial fisheries, much oceanographic research has been performed in order to understand and predict upwellings, and remote sensing from satellites is beginning to be used to develop this.

Eddies are structures with a high morphologic and contextual variability difficult to determine. These are different from the surrounding water in temperature and salinity. In addition, an eddy can travel great distances and for a long period of time without mixing itself with the surrounding water. On the other hand, the direction of cold or warm eddies and the degree of symmetry are controlled in our case by trade winds, that being more intense for the first in conditions of calm, whereas with strong winds the seconds are intensified and more symmetrical. In cold eddies [8] the movement is ascending: cold water, rich in nutrients, arises towards the surface. However, warm eddies accumulate and sink warm water, generating a transport of organic matter towards the interior of the ocean.

Wakes [9] are oceanic structures that are characterised by structures of warm type associated with islands. In different studies, wakes have been observed leeward of the Canary Islands. A wake is generated by the obstacle that the islands conform for the pre dominant wind field in this region (trade winds direction NE). This implies a reduction of the intensity of winds in the southwest of the islands and a heating of the surface of the sea in these zones. Wakes are relatively brief and they are identified by a water strip warmer than the rest, this strip is adhered to coast and a filament is extended to open sea.

Figure 1(a) shows the ocean phenomena like upwellings, eddies and island wakes in an AVHRR scene (equalized), and figure 1(b) shows classified ocean phenomena.

3. Data set

The starting point of this study is a symbolic knowledge data base. This data base is obtained from the competitive networks of High Level Knowledge Processors [10] from the satellite images.
On the other hand, a numercic knowledge data base is extracted from the same set of satellite images.

3.1. High-Level Knowledge Processors (HLKP)

HLKP are neural symbolic networks for data classification processes. This classifier builds up a neural network structure from a set of interconnected processing elements (PEs). The processing elements depend on its input, building a network with learning ability. The processing elements have a connected architecture equal to neural classifiers.

HLKP let to use knowledge in non-numerical domains and structurally and operationally similar to some classical neural classifiers. The basic element in this classifier is the High-Level Knowledge Processor (HLKP) [10].
In the classification process, the processing elements save the structure of the knowledge base, the rule depend on this structure (these rules follow ideas of similarity between patterns) and each rule has an associated priority value. The competitive structure of HLKPs selects the output for which the rule that generated the output has the greatest priority.

3.2. Symbolic feature set

The original feature set can be divided into two categories: *morphological features*, described in table 1 and *context features*, which can be seen in table 2.

Label	Feature
1,2,3	Size in pixels above or below threshold
3000	Above-average temperature of the region.
3001	Below-average temperature of the region
3002	Region variability above 4 (see Torres et al, 1997)
3003	Region variability above 6 (see Torres et al, 1997)
3004	Variability below 4 (see Torres et al, 1997)
8000	Presence of a kernel.
8003	Hat or V-shape (for cold or warm eddies).
1,2	Rounded.
8005	Not on the boundary.
8006	The centroids of the subregions are aligned.
8007	Kernel cold with regard to the rest of the region.
8008	Kernel warm with regard to the rest of the region.
8010	Oblique Gaussian shape (for upwelling).
8009	Oblique Gaussian shape
8002	Not on any defined region.

Tabla 1. Morphological features.

Label	Feature (using a GIS database)
	Land zone which the region is on
1024	SE Spain
1025	SW Portugal
1026	Melilla
1027	NW Africa
1028	Sahara
1029	Fuerteventura
1030	Isla de Lobos
1031	Lanzarote
1032	Grand Canary Island
1033	Tenerife
1034	Gomera
1035	La Palma
1036	Hierro
1037	N Cape White
1038	S Cape White
1050	Cantabrian Coast and W of France
	Position of the region with regard to land
2000	North
2001	South
2002	East
2003	West
	Type of land closest to the region
5000	Island
5001	Continental platform
	Position in the ocean in which the region is found
4000	Coast
4001	Transition Zone
4002	High sea
	Hemisphere in which the region is found
6000	North
6001	South
	Sea or ocean in which the region is found
7000	Mediterranean
7001	Atlantic
8001	**The region is not surrounded by clouds.**
8004	**The region is at less than a distance from the islands or near filamentous structures.**

Tabla 2. Context features.

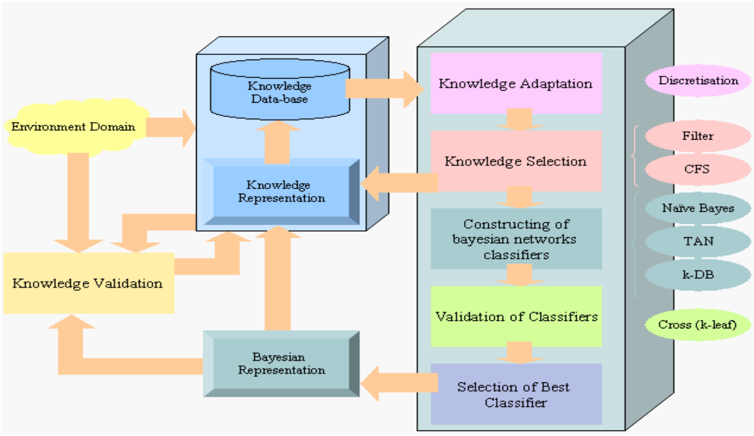

Figura 2. Methodology of knowledge selection and validation

3.3. Numeric feature set

Some of the original feature set can be seen in table 3.

Simple features	*Level of gray*
First Point. Area. Perimeter Density Volume. Volume^2 Equivalent diameter.	Min level of gray Max level of gray Mean level of gray Standard deviation Barycenter level of grays
	Bounding box
	First point Height Width Area Extended
Bounding ellipsoid	
Centroid. Major Axis. Minor Axis. Orientation Excentricity Irradiance	
	Inerce moments
	Moments of Hu Moments of Maitra Moments of Zernike Tensorial moments

Tabla 3. Numeric features

4. Methodology for knowledge selection and validation

A new methodology is proposed in figure 2. The goal is to achieve a better knowledge selection by means of filter metrics, Bayesian classifiers and cross-validation.

Knowledge optimization is shown by means of Bayesian classifiers. This allows several kinds of causal reasoning: inductive, deductive-predictive and intercausal. In figure 2, it is observed a knowledge feedback flow; this makes possible the knowledge validation.

4.1. Knowledge adaptation.

Discretization algorithms [11] can be divided into two categories:

- unsupervised (class-blind) algorithms that discretize attributes without taking into account respective class labels. The representative algorithm is equal-frequency discretizations.
- supervised algorithms discretize attributes by taking into account the class-attribute interdependence. The representative algorithm is the K-means method.

Discretization should significantly reduce the number of possible values of the continuous attribute since large number of possible attribute values contributes to a slow and ineffective process of inductive machine learning. Thus, a supervised discretization algorithm should seek for the minimum number of discrete intervals, and at the same time it should not weaken the interdependency between the attribute values and the class label.

The goal is to search the best discretice aproximation. We applied Wrapper method with bayesian classifier (Naïve Bayes). The best result is shown in experimetal results.

4.2. Knowledge Selection.

Dimension reduction is often used in clustering,classiffication,and many other machine learning and data mining applications. It usually retains the most important dimensions (attributes),removes the noisy dimensions (irrelevant attributes)and reduces computational cost.
We have tried different filter methods [12] for feature selection:

- Mutual information (MI).
- Euclidean distance (ED).
- Matusita distance (MD).
- Kullback-Leibler (KL)
- Shannon entropy (SE)
- Bhattacharyya metric (BM)
- Correlation based Features Seleccion (CFS).

4.3. Constructing of bayesian networks classifiers

Bayesian networks [13] have been successfully used as models for representing uncertainty in knowledge bases. The uncertainty is represented in terms of a probability distribution whose induced independence relations are encoded by the network structure.
Formally, a Bayesian network for a set of variables $X = \{X_1, ..., X_n\}$ consist of a directed acyclic graph where each vertex is labeled with a variable in X, and a set of conditional distributions for each variable X_i given its parents in the graph, which is denoted as $p(x_i|x_{pa(i)})$.

A Bayesian network can be used as a classifier, as long as one of its variables represents the class and the other ones are the features that describe the object that is going to be classified.
Classification is carried out by instantiating the value of the feature variables and then carrying out a *probability propagation* [13] over the class variable, which consists of computing the posterior probability of each class given the observed features. Afterwards, the assigned class is that one with higher posterior probability.

We have tried three simple methods for learning the Bayesian network from the set of numeric data:

1. *Naive-Bayes:* this method is oriented to classification, and is based on the assumption that all the features are conditionally independent when the class is known. This assumption implies that the structure of the network is rather simple, since the only arcs in the network link the class variable with each one of the features, and there are no arcs among the feature variables. The advantage of this naive approach is that the number of parameters to learn from the data is low, improving in this way the accuracy of the estimations [14].
2. *Tree Augmented Naïve Bayes Classifier (TAN)*: [15] TAN models are a restricted family of Bayesian networks in which the class variable has no parents and each attribute has as parents the class variable and at most another attribute.
3. *k-Dependence Bayesian Classifier (KDB):* [16] a k-dependence Bayesian classifier is a Bayesian network which contains the structure of the Naive Bayesian classifier and allows each feature to have a maximum of k feature nodes as parents.

4.4. Validation of Classifiers

Instead of fixed train-test partitions we have performed the experiments with 10-fold cross-validation. This means that the whole dataset is partitioned into 10 sub-datasets, nine are used as training set and the remainder one is reserved initially as test set, and is then used to evaluate the results in the second stage. This is done for the ten possible subdatasets.

4.5. Selection of the best classifier

Selecting classifiers based only on their accuracy, it may often be more effective to attempt to select the classifiers based on their simplicity, for which several measures have been observed: dimension reduction and dependence relationship between features. For classification purposes it may be useful to examine especially the errors.

5. Experimental results

In order to test the performance of the Bayesian classifiers in our system, we have simulated a series of symbolic and numeric data sets as described in section 3, and for each data set the three Bayesian networks have been constructed, using the Elvira tool [17], which is available at *http://leo.ugr.es/~elvira*.

In evaluation process of the symbolic knowledge, a discretization (equal frecuency) has been used for 2 intervals (present or absence of the value). The results of Bayesian classifiers is displayed in table 4, where the accuracy rate and the average number of relevant features can be found. With respect to the accuracy rate, 80% is obtained by HLKPs. This accuracy rate is practycally the maximum that can be reached taking into account the original data, unless new features were included in the system [10]. It can be seen that the classifiers based on Bayesian networks reach the same accuracy level achieved by the neural network (HLKP), but reducing the number of features significantly (12 variables (CFS) instead of 50 (HLKP)).

In the case of numeric data the accuracy rate is shown in table 5. The best discretization is K-Means with k = 100. It is observed, that the methodology improves to HLKPs since its best classification is 89.18 % whereas HLKP is 80 % and dimension reduction is 14 (CFS) instead of 50 (HLKP) or 80 (all numeric features). Examples of the networks obtained in the experiments can be seen in figures 3 and 4.

6. Conclusion

We have explored the use of Bayesian networks as a mechanism for feature selection in a system for automatic recognition of oceanic satellite images. The use of Bayesian networks has provided benefits with respect to HLKPs, not only in the reduction of relevant features, but also in discovering the structure of the knowledge, in terms of the conditional independence relations among the variables.

In future works we plan to improve the accuracy rate of the system including more variables. Furthermore, we expect to use models for avoiding the discretisation of the continuous features when learning Bayesian networks.

More precisely, we will use mixtures of truncated exponentials [17].

Filter	No. of Vars	NB	TAN	KDB
MI	19	78.84	78.85	78.82
ED	19	78.84	78.84	78.84
MD	19	78.83	78.77	78.77
KL	19	78.82	78.87	78.76
SE	48	**79.00**	**79.10**	**78.96**
BM	13	77.60	77.60	77.60
CFS	**12**	77.44	77.36	77.38

Tabla 4. Evaluation with symbolic features (number of variables and accuracy).

Filter	No. of Vars	NB	TAN
IM	60	84.78	84.08
DE	24	88.48	85.78
MA	61	85.28	82.48
KL-1	61	85.08	81.98
KL-2	19	87.68	84.38
SH	1	30.03	31.33
CFS	**14**	**89.18**	**87.08**
All	80	85.88	78.47

Tabla 5. Evaluation with numeric features (number of variables and accuracy)

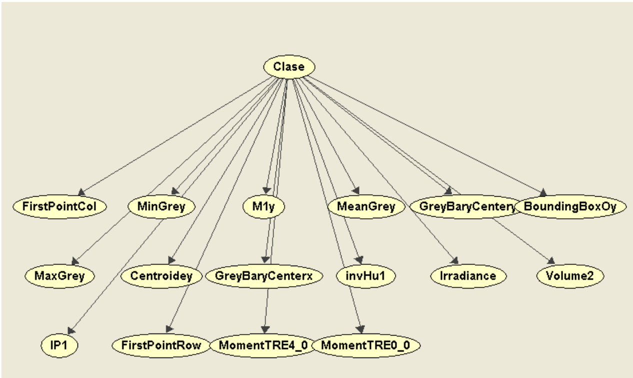

Figura 3. Naive Bayes – CFS, numeric data set.

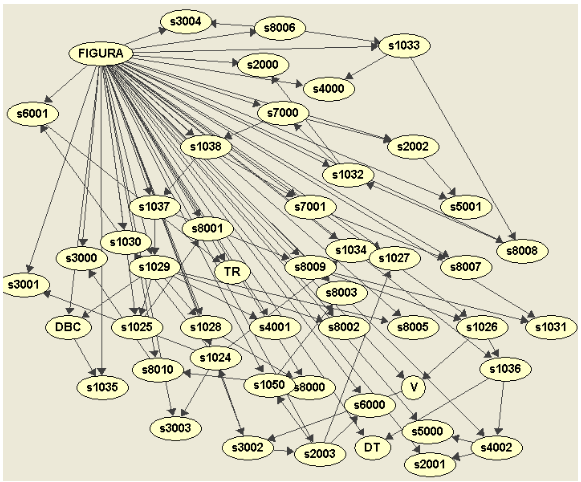

Figura 4. TAN – SH, symbolic data set.

Acknowledgments

This work has been partially supported by the Spanish Ministry of Science and Technology through project TIC2001-2973-C05-02 and FEDER funds.

References

[1] I. Inza, P. Larrañaga, R. Etxeberria, and B. Sierra. 2000. Feature subset selection by bayesian networks based optimization. *Artificial Intelligence*, 123(1-2):157-184.

[2] R.S. Lynch and P.K. Willet. 1998. Bayesian classification and the reduction of irrelevant features from training data. In *Proceedings of the 37th IEEE* Conference *on Decision & Control Tampa*, pages 1591-1592, Florida USA, December.

[3] Y.L. Murphey and H. Guo. 2000. Automatic feature selection - a hybrid statistical approach. *IEEE International Conference on Pattern Recognition*, pages 382-385.

[4] P. Langley and S. Sage. 1994. Induction of selective Bayesian classifiers. *In Proceedings of the Tenth Conference on Uncertainty in Artificial Intelligence*, pages 399-406, Seattle, WA. Morgan Kaufmann.

[5] L. Bruzzone. 1998. Classification of remote sensing images by using the Bayes rule for minimun cost. IEEE Int. Geoscience and Remote Sensing Symposium, pages 1778-1780.

[6] Y. Yamagata and H. Oguma. 1997. Bayesian feature selection for classifying multi-temporal sar and tm data. *IEEE Int. Geoscience and Remote Sensing Symposium*, pages 978-980.

[7] J.L. García. 1998. *Descripción y análisis cuantitativo mediante series de imágenes de satélite de ladinámica de las aguas superficiales del noroeste de África*. Ph.D. thesis, Universidad de las Palmas de Gran Canaria.

[8] A. Tejera. 1996. *Estudio de estructuras oceanográficas mesoescalares en la cuenca canaria mediante los altímetros de los satélites GEOSAT, ERS-1 y TOPEX/POSEIDÓN*. Ph.D. thesis, Universidad de las Palmas de Gran Canaria.

[9] A. Tejera. 1997. Mesoscale vaiability in the canary region from altimetry. *In Proc 3rd ERS Symposium On Space at the service of our enviroment*, volume 3, Florence Italy ESA SP-414.

[10] J.A. Torres, F. Guindos, M. Peralta, and M. Cantón. 2003. Competitive neural-net-based system for the automatic detection of oceanic mesoscalar structures on avhrr scenes. *IEEE Transactions on Geoscience and Remote Sensing*.

[11] Kurgan L. & Cios K.J. 2001. Discretization Algorithm that Uses Class-Attribute Interdependence Maximization, *Proceedings of the 2001 International Conference on Artificial Intelligence (IC-AI 2001)*, pp.980-987, Las Vegas, Nevada.

[12] Ben-Bassat, M. 1982. Use of distance measures, information measures and error bounds in feature evaluation. In Krishnaiah, P. R. and Kanal, L. . N., editors, *Handbook of Statistics*, volume 2, pages 773.791. North-Holland Publishing Company.

[13] F.V. Jensen. 2001. *Bayesian networks and decision graphs*. Springer-Verlag.

[14] R.O. Duda and P.E. Hart. 2001. *Pattern classification*. John Wiley and Sons, New York.

[15] Friedman, N., Geiger, D., and Goldszmidt, M. 1997. Bayesian network classifiers. *Machine Learning*, 29(2):131.164.

[16] Sahami, M. 1996. Learning limited dependence Bayesian classifiers. *In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, pages 335.338.

[17] Elvira Consortium. 2002. Elvira: An environment for creating and using probabilistic graphical models. In J.A. Gámez and A. Salmerón, editors, *Proceedings of the First European Workshop on Probabilistic Graphical Models*, pages 222-230.

[18] S. Moral, R. Rumí, and A. Salmerón. 2001. Mixtures of truncated exponentials in hybrid Bayesian networks. *In Lecture Notes in Arti¯cial* Intelligence, volume 2143, pages 135-143.