Lab session 4:
Land feature selection based on satellite images

Félix Jiménez Monzón and Javier Carnerero Cano
Univesidad Carlos III de Madrid
Avenida de la Universidad 30, 28911, Leganés (Madrid), Spain
{100290742, 100292885}@alumnos.uc3m.es

Abstract

In this work, we evaluate the performance of different feature selection strategies to determine wich ones provides better performance to classify the crop type (16 classes) of a set of pixels in a multiband spectral image. Such spectral image was taken by sensor type Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) on board a satellite and it covers a region of Indiana, United States. We use the Filter, Search, Wrappers and Embedded techniques. Random Forest for Filter methods, and mRMR and RFE for other kind of methods obtain the highest accuracy in this case.

1 Introduction

Airborne Visible / Infrared Imaging Spectrometer (AVIRIS) [1] is an optic sensor, unique in the world, that allows the obtaining of images in the emitted spectral radiation. It is able to measure the molecular absorption and particle scattering signatures in 224 adjacent spectral bands.

The main objective of the AVIRIS project is to identify, measure and monitor the Earth's surface and atmosphere in order to understand, monitor and predict the processes related to the global environment and climate change. In that sense, a spectral image from some crop fields from a region of Indiana (USA) is going to be used. This image was taken with the AVIRIS sensor on board a satellite, and 220 spectral bands have been used.

In the previous part of this work we evaluated different classifiers in order to see the performances of the same. In this case, different types of feature selection techniques will be used to determine which ones provide better performance in this setting.

2 Objective

The objective of this project is to carry out a comparative study of the performance of several feature selection techniques[1], like Filter methods, Search methods, Wrappers and Embedded methods. With these features, we will apply a Linear Support Vector Machine (SVM), which is able to classify and associate the pixels from a spectral picture, taken with a satellite, of various types of crop fields. The input data are spectral images of 145×145 pixels, which represent the molecular absorption and particle scattering signatures in 220 contiguous spectral bands. Each pixel in the image represents an area of 20 m^2. The classifier has been trained so that it can associate the 220 spectral bands of

[1] The results correspond to a fix division of the train and test samples. If we want to see which technique presents the best performance, we have to repeat the process several times in order to calculate the average and the standard deviation of the accuracy of the classifier. We have not done this because of the extremely long computational time needed. So, only with these results we could not assure that a technique is better than the others.

each pixel with a type of crop field (16 classes). Now in this work, we are going to try to reduce the dimensionality of the input data (220 dimensions) minimizing the loss of relevant information. We will take the 75% of the labelled image pixels for training.

First of all, we normalize the pre-processed data to zero mean and unit standard deviation in each of the inputs variables. Then, in order to carry out this project, several methods of feature selection techniques will be used. First, in terms of Filter methods, the performance of the Anova F-Test, Mutual Information, Random Forest and Hilbert-Schmidt Independence Criterion (HSIC) algorithms will be studied. Then, the Minimum Redundancy Maximal Relevance (mRMR) algorithm, which is a Search method, will be studied. After that, a Wrapper feature selection technique will be analyzed, using the Recursive Feature Elimination (RFE) method. And we will also study the performance of two Embedded methods: the L_1-SVM regularization and the L_1-Logistic Regression regularization. Finally, a comparative analysis of the results will be made.

3 Filter methods

Four kinds of Filter methods have been studied. These are: Anova F-Test, Mutual Information, Random Forest and HSIC algorithms. First of all, we would like to note that, after selecting the subset of relevant features, we have analysed their discriminatory capability using a linear SVM as classifier and have used its final test accuracy to evaluate the goodness of the different selection methods. We have used the default parameters for that classifier. They are: $C = 1$ and $\gamma = \frac{1}{n_{features}}$. We are aware that we should have validated those parameters for each new case, since the number of variables is different, but the computational cost for that is very high and it requires a huge amount of time. For this reason, we have used those parameters in all the cases.

3.1 Anova F-Test

Firstly, we carry out the classification applying in first place the Anova F-Test algorithm, which performs a univariate analysis, i.e. it evaluates feature by feature its relevance (in an independently way). This algorithm analyses if the expected values of a feature differ from one class to other. In order to find the optimum number of features, a validation process has been done, validating the performance of the algorithm (from 1 feature to 220 with a step of 5). This evolution can be seen in Figure 1(a). We obtained that the optimum number of features is **201**, and the accuracy obtained using that number is **87.23%**.

3.2 Mutual Information

The next algorithm we have evaluated is the Mutual Information algorithm, whose objective is to measure non-linear relationships in high dimensional spaces. We have also done a validation process in order to determine the optimum number of features. These results can be seen in Figure 1(b). According to the results, the optimum number of features is **181**, and the accuracy obtained is **87.00%**.

3.3 Random Forest

Thirdly, we have studied the performance of the Random Forest algorithm, in which a feature used as a decision node in a tree can be also used to assess the relative importance of that feature. In this case, we have set the number of estimators equal to the number of features (220). Like in the previous cases, we have done a validation process to determine the optimum number of features. The accuracy evolution can be seen in Figure 1(c). The optimum number of features is **176** and the accuracy is **88.66%**.

3.4 Hilbert-Schmidt Independence Criterion

Finally, the performance of the HSIC algorithm has been evaluated. This algorithm considers the covariance in the Hilbert space (by means of kernel functions), which let us measure linear relationships between two variables. In this case, we have used a gaussian kernel. Once again, we have determined the optimum number of features carrying out a validation process, which can be

Table 1: Number of features and test accuracy for each Filter method.

Algorithm	Number of features	Test accuracy (%)
Anova F-Test	201	87.23
Mutual Information	181	87.00
Random Forest	176	88.66
HSIC	196	87.54

seen in Figure 1(d). We determine that the optimum number of features obtained is **196**, and the accuracy is **87.54%**.

3.5 Comparison

We can observe in Table 1 that all Filter methods have a similar performance, both in test accuracy and in number of features. Especially, we can see that the best result is achieved using the Random Forest algorithm, since the test accuracy is the highest and the number of features is the lowest. However, the difference with the performance of the other algorithms is not big enough to affirm which one is the best.

4 Search method: Minimum Redundancy Maximal Relevance

As in the previous section, we would like to note that, after selecting the subset of relevant features, we have analysed their discriminatory capability using a linear SVM as classifier and have used its final test accuracy to assess the performance of the different selection methods. We have used the default parameters for that classifier.

MMMR is an extension of univariate scorings to a multivariate analysis. It based on selecting a relevance and redundance scorings, and computing the features that maximize the difference between the relevance and redundance criteria. Like in the previous cases, we have done a validation process to determine the optimum number of features (from 1 feature to 220 with a step of 5). The accuracy evolution can be seen in Figure 1(e). The optimum number off features and the accurracy are, respectively, **196** and **87.54%**.

5 Wrapper: Recursive Feature Elimination

Now we are going to study a Wrapper method in order to do the feature selection. This method is the RFE algorithm. Once again, we are going to use a linear SVM as classifier, in order to eliminate the useless variables and to check if there is performance degradation.

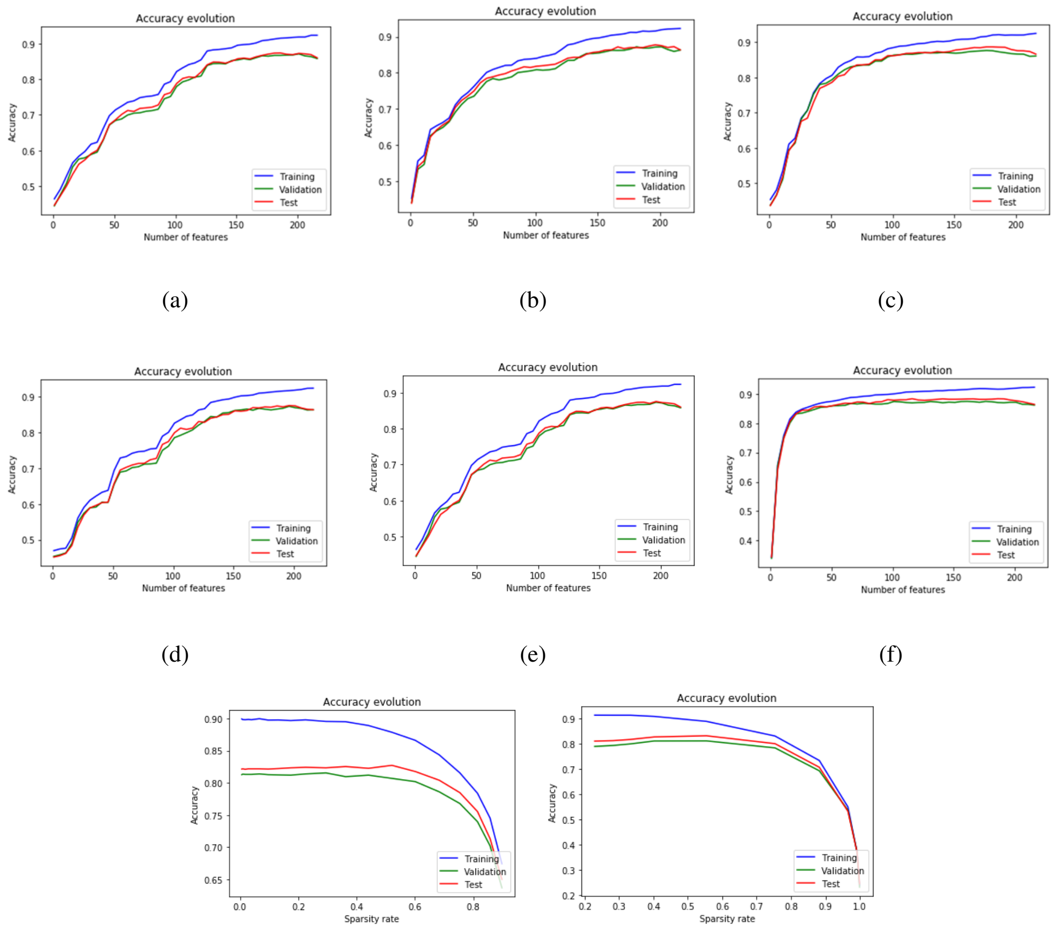
We have set both the number of features to select and the step equal to 10 and we have used a SVM as estimator. We have also done a validation process in order to determine the optimum number of features. The accuracy evolution can be seen in Figure 1(f). The optimum number of features is **161** and the accuracy is **88.23%**.

6 Embedded methods

Two kinds of Embedded methods have been studied. These are: L_1-SVM and L_1-Logistic Regression. Like the previous cases, we have analysed their discriminatory capability using a linear SVM as classifier and have used its final test accuracy to assess the performance of the different selection methods. We have used the default parameters for that classifier.

6.1 L_1-Support Vector Machine

In the L_1-SVM formulation, we modify the regularization term in such a way that the L_1-norm is minimized. Moreover, it is based on the lack of continuity in the origin, which causes most of the coeficients to fall into it, making them to be zero. This provides sparse solutions (automatic feature



(a) (b) (c)

(d) (e) (f)

(g) (h)

Figure 1: Accuracy evolution for the different feature selection techniques. (a) Anova F-Test. (b) Mutual Information. (c) Random Forest. (d) HSIC. (e) MMRR. (f) RFE. (g) L_1-SVM. (h) L_1-Logistic Regression.

selection in linear SVM). The accuracy evolution with respect to the sparsity rate can be seen in Figure 1(g). The optimum number of features is **220**, and the optimum test accuracy is **82.33%**.

6.2 L_1-Logistic Regression

It works in an analogous way than the L_1-SVM method, but in this case it is based on the Logistic Regression algorithm. The accuracy evolution with respect to the sparsity rate can be seen in Figure 1(h). The optimum number of features is **213**, and the optimum test accuracy is **83.14%**.

7 Conclusions

In this project, we have studied the performance of several Filter, Search, Wrapper and Embedded feature selection techniques in order to see which one of them achieves the better result. The problem consisted in cataloguing the different types of crops (16 classes) from a spectral image taken with the AVIRIS sensor on board a satellite.

In this setting, the Random Forest algorithm achieves a better performance than the other Filter methods. On the other hand, mRMR and RFE also obtain higher accuracy than the rest of algorithms.

References

[1] NASA, “AVIRIS (Airborne Visible / Infrared Imaging Spectrometer).” Available: “https://aviris.jpl.nasa.gov/", 2017. [Online]. [Accessed February 28, 2017].

[2] C. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2007.

[3] R. Duda, P. Hart, and D. Stork, *Pattern Classification*. Wiley & Sons, 2nd Edition, 2000.