ESA Round Robin Exercise

Subjects: Meteorology & Atmospheric Sciences | Others

Contributor: Umberto Amato , Anestis Antoniadis , Maria Francesca Carfora

Motivated by the experience acquired in the ESA promoted Round Robin exercise aimed at comparing cloud detection algorithms for PROBA-V sensor, we investigate specific issues related to cloud detection when remotely sensed images comprise only a few spectral bands in the visible and near-infrared by considering a bunch of well-known classification methods. First, we demonstrate the feasibility of using a training dataset semi-automatically obtained from other accurate algorithms. In addition, we investigate the effect of ancillary information, e.g., surface type or climate, on accuracy. Then we compare the different classification methods using the same training dataset under different configurations. We also perform a consensus analysis aimed at estimating the degree of mutual agreement among classification methods in detecting Clear or Cloudy sky conditions. Results are also compared on a high-quality test dataset of 1350 reflectances and Clear/Cloudy corresponding labels prepared by ESA for the mentioned exercise.

Cloud Detection   PROBA-V   statistical learning   machine learning

cumulative discriminant analysis   K-Nearest Neighbor   neural networks

1. Introduction

Despite the large existing literature, cloud detection from images taken by sensors onboard satellites is still an area of very active research. This is essentially due to three main reasons: (i) Cloud detection is an important preliminary step of remotely sensed image processing because clouds affect sensor measurements of radiance emitted by surface up to make data unreliable for a wide range of remote-sensing applications that use optical satellite images; (ii) Cloud detection by itself is a difficult problem (even for experts attempting to visually detect clouds from signatures and/or images) in some conditions as transparent or semi-transparent clouds and in general when the contrast between the cloud and the underlying surface is poor; and (iii) Despite consolidated guidelines for cloud detection algorithms (e.g., use of infrared bands, preliminary removal of noninformative bands), development of new sensors with different hardware capabilities in terms of spatial, spectral and temporal resolution claims for specific algorithms or adaption of existing ones (e.g., re-estimate of new thresholds).In all cases, also considering the recent explosive emerging methods, a problem of validation of methods arises that could help not only in comparing their accuracy but also to understand strengths and weaknesses of the general cloud detection problem.

Classification exercises are sometimes organized where different algorithms are challenged to estimate a cloud mask from radiance detected by a specific sensor. Radiance is endowed with labels on the Clear or Cloudy condition accurately assigned by experts that are blind to the algorithms, so to be used as a validation of the

algorithms themselves. While the main purpose of such exercises is to develop accurate operational algorithms for specific sensors onboard satellites, an important by-side effect is comparison of state-of-art methods on a same, very accurate dataset. In this respect we mention the Landsat comparison exercise and the ESA Round Robin exercise for PROBA-V sensor. Such comparisons are an exceptional way not only to compare algorithms, but especially to discover their weakness in particular climatic/surface conditions and, finally, to progress knowledge on cloud mask detection.The framework proposed by the authors for the Round Robin exercise includes a statistical classification method (Cumulative Discriminant Analysis, CDA), a training set semi-automatically obtained from cloud masks estimated for concurrent sensors, and grouping data in almost homogeneous surface types. In particular our framework was the only one within the exercise that did not use a manual dataset obtained by expert annotation to train the classification. Instead it was relying on a semi-automatic training obtained as the result of consolidated and acknowledged as reliable cloud masks having comparable spatial resolution as the target cloud mask (MODIS and SEVIRI). The only intervention required is spatial and temporal co-registration of the training cloud mask with the target cloud mask.

2. Data

PROBA-V (PRoject for On-Board Autonomy – Vegetation) is a global vegetation monitoring mission, launched in 2013 to assure the succession of the Vegetation instruments onboard the French SPOT-4 and SPOT-5 Earth observation missions. The satellite follows a Sun-synchronous orbit at a height of 820 km, achieving a daily global coverage, except the equatorial region where coverage is guaranteed every two days. The optical instrument onboard provides from 1/3 km to 1 km-resolution data products. It captures a Blue band (centred at 463nm), a Red band (centred at 655nm), a Near Infrared band (centred at 845nm), and a Short wave Infrared band (centred at 1600nm). The data of the traditional Vegetation products, as provided by PROBA-V, are freely accessible for all users. The new, higher resolution products of PROBA-V elder than 1 month share the same full, free and open data policy.

We consider as input data 331 images released by ESA and provided by the organizers of the Round Robin exercise. These images are PROBA-V Level 2A products with Top-of atmosphere reflectance (the four PROBA-V bands radiometrically and geometrically corrected and resampled at spatial resolution 333 m). They conform a complete globe acquisition from four different dates covering the four seasons along the year 2014. PROBA-V scenes are endowed with a sea/land mask and an algorithm for snow/ice detection. The total number of valid scenes available in the 331 files is 7,731,538,861, the remaining ones being off sensor view, sun glint or missing reflectance.

3. Methods

Cloud detection can be formally considered as a binary supervised classification problem. As such, methods for its solution need a representative set of data with labels considered to be "certain" (training dataset). They evaluate patterns in different features and assign data into one of the two classes (Clear or Cloudy). The classification

procedure also involves collection and evaluation of a validation dataset. Once trained, each classifier applies a decision rule to determine if validation data are more likely to have originated from one class or another. This rule partitions the n-dimensional feature space into 2 regions corresponding to the Clear and Cloudy conditions.

Among the different approaches reported in the literature, we consider and compare for the present study seven different supervised classifiers. They fall into the categories usually labelled as Statistical and Machine Learning and are based on different principles, as Discriminant Analysis, Neural Networks, Nearest Neighbour. In the following we briefly describe them.

1. Linear Discriminant Analysis (LDA). It applies the Bayes rule to each scene to select the Clear/Cloudy class so to maximize the posterior probability of the class for a scene given the actual reflectance in that scene. LDA assumes that reflectance follows Gaussian distributions for the Clear and Cloudy classes sharing the same covariance matrix;
2. Quadratic Discriminant Analysis (QDA), that generalizes LDA assuming that also covariance matrix depends on the class (Clear or Cloudy);
3. Principal Component Discriminant Analysis (PCDA): the hypothesis of Gaussian distribution of reflectance is released in favour of a generic distribution estimated by nonparametric regression; in addition the original reflectances are transformed into uncorrelated Principal Components before classification;
4. Independent Component Discriminant Analysis (ICDA): similar to PCDA, but with the original reflectances transformed into Independent Components before nonparametric estimation of the densities; this makes such components independent also for non Gaussian distributions;
5. Cumulative Discriminant Analysis (CDA): the decision rule for classification depends on a single threshold for each feature (spectral band), based on the empirical distribution function, that discriminates scenes belonging to the Clear and Cloudy classes; the threshold is estimated so to minimize at the same time the false positive and false negative rates on the training or on a validation dataset.
6. Artificial Neural Networks (ANN). We use a two-layer feed-forward network, with sigmoid hidden and softmax output neurons for pattern recognition. The network is trained with scaled conjugate gradient backpropagation.
7. k-Nearest Neighbour (KNN) that labels each considered scene based on a voting strategy among the labels assigned to the K closest neighboring scenes belonging to the training dataset.

Methods LDA, QDA, PCDA, ICDA and CDA require estimate of the statistical distribution of radiance. We mention that other methods are available in the literature; results of some of them have not been reported because of poor accuracy on other sensors (Logistic Regression) or unfeasible computational time (Support Vector Machine). All the above methods are pixel-wise, that is they treat pixels separately without taking account of spatial correlations among them or local features that are instead typical of images.