Post classification smoothing using Bayesian techniques in SITS

Rolf Simoes   *National Institute for Space Research (INPE), Brazil*
Gilberto Camara   *National Institute for Space Research (INPE), Brazil*
Alexandre Carvalho   *Institute for Applied Economics Research (IPEA), Brazil*

This vignette describes a Bayesian smoothing method to reclassify the pixels, based on the machine learning probabilities. We consider that the output of the machine learning algorithm provides, for each pixel, the information on the probability of such pixel belonging to each of the target classes. Usually, we label a pixel as being of a given class if the associated class probability is higher than the probability of it belonging to any of the other classes. The observation of the class probabilities of each pixel is taken as our initial belief on what the actual class of the pixel is. We then use Bayes' rule to consider how much the class probabilities of the neighbouring pixels affect our original belief.

Introduction

Image classification post-processing has been defined as "a refinement of the labelling in a classified image in order to enhance its classification accuracy" [Huang et al., 2014]. In remote sensing image analysis, these procedures are used to combine pixel-based classification methods with a spatial post-processing method to remove outliers and misclassified pixels. For pixel-based classifiers, post-processing methods enable the inclusion of spatial information in the final results.

Post-processing is a desirable step in any classification process. Most statistical classifiers use training samples derived from "pure" pixels, that have been selected by users as representative of the desired output classes. However, images contain many mixed pixels irrespective of the resolution. Also, there is a considerable degree of data variability in each class. These effects lead to outliers whose chance of misclassification is significant. To offset these problems, most post-processing methods use the "smoothness assumption" [Schindler, 2012]: nearby pixels tend to have the same label. To put this assumption in practice, smoothing methods use the neighbourhood information to remove outliers and enhance consistency in the resulting product.

Smoothing methods are an important complement to machine learning algorithms for image classification. Since these methods are mostly pixel-based, it is useful to complement them with post-processing smoothing to include spatial information in the result. A traditional choice for smoothing classified images is the majority filter, where the class of the central pixel is replaced by the most frequent class of the neighbourhood. This technique is rather simplistic; more sophisticated methods use class probabilities. For each pixel, machine learning and other statistical algorithms provide the probabil-

ities of that pixel belonging to each of the classes. As a first step in obtaining a result, each pixel is assigned to the class whose probability is higher. After this step, smoothing methods use class probabilities to detect and correct outliers or misclassified pixels.

In this vignette, we introduce a Bayesian smoothing method, which provides the means to incorporate prior knowledge in data analysis. Bayesian inference can be thought of as way of coherently updating our uncertainty in the light of new evidence. It allows the inclusion of expert knowledge on the derivation of probabilities. As stated by Spiegelhalter and Rice [2009]: "In the Bayesian paradigm, degrees of belief in states of nature are specified. Bayesian statistical methods start with existing 'prior' beliefs, and update these using data to give 'posterior' beliefs, which may be used as the basis for inferential decisions". Bayesian inference has now been established as a major method for assessing probability.

Overview of Bayesian estimattion

Most applications of machine learning methods for image classification use only the categorical result of the classifier which is the most probable class. The proposed method uses all class probabilities to compute our confidence in the result. In a Bayesian context, probability is taken as a subjective belief. The observation of the class probabilities of each pixel is taken as our initial belief on what the actual class of the pixel is. We then use Bayes' rule to consider how much the class probabilities of the neighbouring pixels affect our original belief. In the case of continuous probability distributions, Bayesian inference is expressed by the rule:

$$\pi(\theta|x) \propto \pi(x|\theta)\pi(\theta)$$

Bayesian inference involves the estimation of an unknown parameter θ, which is the random variable that describe what we are trying to measure. In the case of smoothing of image classification, θ is the class probability for a given pixel. We model our initial belief about this value by a probability distribution, $\pi(\theta)$, called the *prior* distribution. It represents what we know about θ *before* observing the data. The distribution $\pi(x|\theta)$, called the *likelihood*, is estimated based on the observed data. It represents the added information provided by our observations. The *posterior* distribution $\pi(\theta|x)$ is our improved belief of θ *after* seeing the data. Bayes's rule states that the *posterior* probability is proportional to the product of the *likelihood* and the *prior* probability.

Smmothing using Bayes' rule

Given the general principles of Bayesian inference, smoothing of classified images requires estimating the *likelihood* and the *prior* probability of each pixel belonging to each class. In order to express our problem in a more tractable form, we perform data transformations. More formally, consider a set of K classes that are candidates for labelling each pixel. Let $p_{i,k}$ be the probability of pixel i belonging to class k, $k = 1, \ldots, K$.

We have

$$\sum_{k=1}^{K} p_{i,k} = 1, p_{i,k} > 0$$

We label a pixel p_i as being of class k if

$$p_{i,k} > p_{i,m}, \forall m = 1, \ldots, K, m \neq k$$

For each pixel i, we take the odds of the classification for class k, expressed as

$$O_{i,k} = p_{i,k}/(1 - p_{i,k})$$

where $p_{i,k}$ is the probability of class k. We have more confidence in pixels with higher odds since their class assignment is stronger. There are situations, such as border pixels or mixed ones, where the odds of different classes are similar in magnitude. We take them as cases of low confidence in the classification result. To assess and correct these cases, Bayesian smoothing methods borrow strength from the neighbours and reduced the variance of the estimated class for each pixel.

We further make the transformation

$$x_{i,k} = \log[O_{i,k}]$$

which measures the *logit* (log of the odds) associated to classifying the pixel i as being of class k. The support of $x_{i,k}$ is $\mathbb{R}$. Let V_i be a spatial neighbourhood for pixel i. We use Bayes' rule to update the value $x_{i,k}$ based on the neighbourhood, assuming independence between the classes. In this way, the update is performed for each class k at a time.

For each pixel, the random variable that describes the class probability is denoted by $\theta_{i,k}$. Therefore, we can express Bayes' rule for each combination of pixel and class as

$$\pi(\theta_{i,k}|x_{i,k}) \propto \pi(x_{i,k}|\theta_{i,k})\pi(\theta_{i,k}).$$

We assume the prior distribution $\pi(\theta_{i,k})$ and the likelihood $\pi(x_{i,k}|\theta_{i,k})$ are modelled by Gaussian distributions. In this case, the posterior will also be a Gaussian distribution. To estimate the prior distribution for a pixel, we consider that all pixels in the spatial neighbourhood V_i of pixel i follow the same Gaussian distribution with parameters $m_{i,k}$ and $s_{i,k}^2$. Thus, the prior is expressed as

$$\theta_{i,k} \sim N(m_{i,k}, s_{i,k}^2).$$

In the above equation, the parameter $m_{i,k}$ is the local mean of the probability distribution of values for class k and $s_{i,k}^2$ is the local variance for class k. We estimate the local mean and variance by considering the neighbouring pixels in space. Let $\#(V_i)$ be the number of elements in the spatial neighbourhood V_i. The local mean is calculated by:

$$m_{i,k} = \frac{\sum_{j \in V_i} x_{j,k}}{\#(V_i)}$$

and the local variance by

$$s_{i,k}^2 = \frac{\displaystyle\sum_{j \in V_i}[x_{j,k} - m_{i,k}]^2}{\#(V_i) - 1}.$$

We also consider that the likelihood follows a normal distribution. We take the likelihood as being the distribution of $x_{i,k}$, conditioned by the local variable $\theta_{i,k}$. This conditional distribution is also taken as normal with parameters $\theta_{i,k}$ and σ_k^2, expressed as

$$x_{i,k}|\theta_{i,k} \sim N(\theta_{i,k}, \sigma_k^2)$$

In the likelihood equation above, σ_k^2 is a hyper-parameter that controls the level of smoothness. The Bayesian smoothing estimates the value of $\theta_{i,k}$ conditioned by the data $x_{i,k}$. This is the updated value of the logit of class probability for class k of pixel i. Since both the prior and the likelihood are assumed as Gaussian distribution, based on Bayesian statistics the value of conditional mean for a normal distribution is given by:

$$E[\theta_{i,k}|x_{i,k}] = \frac{m_{i,t} \times \sigma_k^2 + x_{i,k} \times s_{i,k}^2}{\sigma_k^2 + s_{i,k}^2}$$

which can also be expressed as

$$E[\theta_{i,k}|x_{i,k}] = \left[\frac{s_{i,k}^2}{\sigma_k^2 + s_{i,k}^2}\right] \times x_{i,k} + \left[\frac{\sigma_k^2}{\sigma_k^2 + s_{i,k}^2}\right] \times m_{i,k}$$

The updated value for the class probability of the pixel is a weighted average between the original logit value $x_{i,k}$ and the mean of the class logits $m_{i,k}$ for the neighboring pixels. When the local class variance of the neighbors $s_{i,k}^2$ is high relative to the smoothing factor σ_k^2, our confidence on the influence of the neighbors is low, and the smoothing algorithm gives more weight to the original pixel value $x_{i,k}$. When the local class variance $s_{i,k}^2$ decreases relative to the smoothness factor σ_k^2, then our confidence on the influence of the neighborhood increases. The smoothing procedure will be most relevant in situations where the original classification odds ratio is low, showing a low level of separability between classes. In these cases, the updated values of the classes will be influenced by the local class variances.

The hyperparameter σ_k^2 sets the level of smoothness. If σ_k^2 is zero, the smoothed value $E[\mu_{i,,k}|l_{i,k}]$ is equal to the pixel value $l_{i,k}$. Higher values of σ_k^2 will cause the assignment of the local mean to the pixel updated probability. In practice, σ_k^2 is a user-controlled parameter that will be set by users based on their knowledge of the region to be classified. In our case, after some classification tests, we decided to set the parameters V as the Moore neighborhood where each pixel is connected to all those pixels with Chebyshev distance of 1, and $\sigma_k^2 = 20$ for all k. This level of smoothness showed the best performance in the technical validation.

Use of Bayesian smoothing in SITS

Doing post-processing using Bayesian smoothing in SITS is straightforward. The result of the `sits_classify` function applied to a data cube is set of more probability images, one per requested clasification interval. The next step is to apply the `sits_label_- classification` function. By default, this function selects the most likely class for each pixel considering only the probabilities of each class for each pixel. To allow for Bayesian smooting, it suffices to include the `smoothing = bayesian` parameter. If desired, the variance parameter (associated to the hyperparameter σ_k^2 described above) can control the degree of smoothness.

```
# Retrieve the data for the Mato Grosso state

# select the bands "ndvi", "evi"
samples_2bands <- sits_select_bands(samples_mt_4bands, ndvi, evi)

#select a rfor model
rfor_model <- sits_train(samples_2bands, ml_method = sits_rfor())

# Classify a raster file with 23 instances for one year
# select the bands "ndvi", "evi" from the "inSitu" package
evi_file <- system.file("extdata/Sinop", "Sinop_evi_2014.tif", package = "inSitu")
ndvi_file <- system.file("extdata/Sinop", "Sinop_ndvi_2014.tif", package = "inSitu")
files <- c(ndvi_file, evi_file)

# define the timeline for the files
time_file <- system.file("extdata/Sinop", "timeline_2014.txt", package = "inSitu")
timeline_2013_2014 <- scan(time_file, character())

# create a data cube based on the information about the files
raster_cube <- sits_cube(name = "Sinop", timeline = timeline_2013_2014,
                         bands = c("ndvi", "evi"), files = files)

# classify the raster image and generate a probability file
raster_probs <- sits_classify(raster_cube, ml_model = xgb_model, memsize = 4, multicores =

# label the probability (no smoothing applied by default)
raster_class <- sits_label_classification(raster_probs)

# smooth the result with a bayesian filter
raster_class_bayes <- sits_label_classification(raster_probs, smoothing = "bayesian")

# plot the smoothened image
plot(raster_class_bayes, time = 1, title = "Sinop-smooth")
```

The result is shown below.

```
knitr::include_graphics(system.file("extdata/markdown", "sinop_class.png", package = "sits"
```
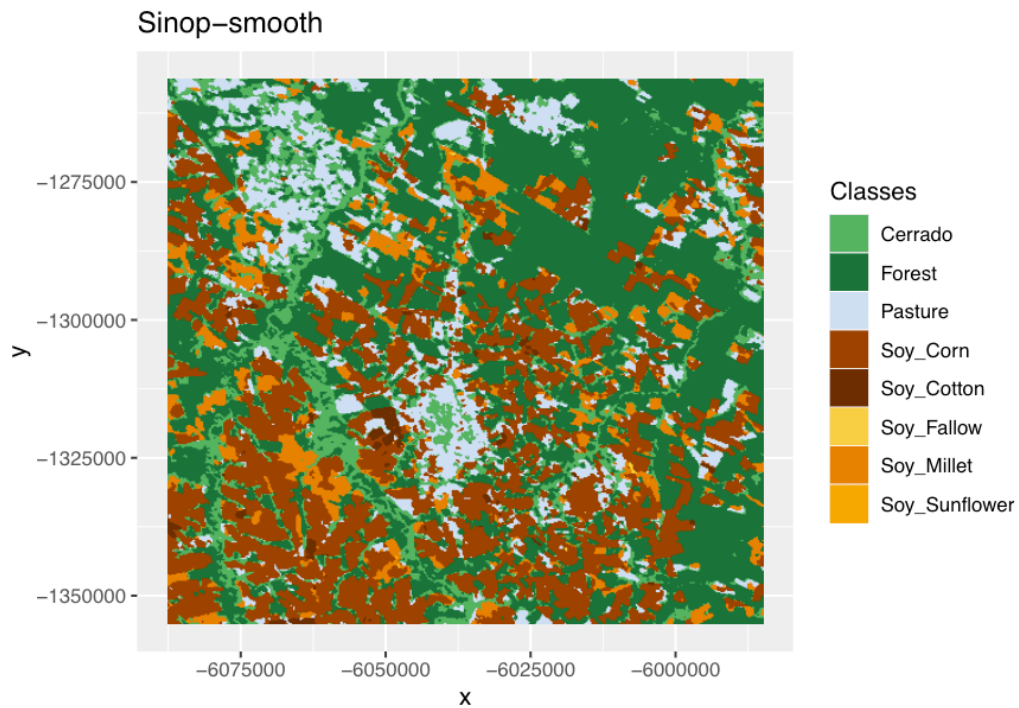

Figure 1: Classified image post-processed with Bayesian smoothing. The image coordinates (*meters*) shown at vertical and horizontal axis are in MODIS sinusoidal projection.

References

Xin Huang, Qikai Lu, Liangpei Zhang, and Antonio Plaza. New postprocessing methods for remote sensing image classification: A systematic study. *IEEE Transactions on Geoscience and Remote Sensing*, 52(11):7140–7159, 2014.

Konrad Schindler. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE transactions on geoscience and remote sensing*, 50(11):4534–4545, 2012.

D. Spiegelhalter and K. Rice. Bayesian statistics. *Scholarpedia*, 4(8):5230, 2009.