

MAPPING OF BURNED AREA USING PRESENCE AND BACKGROUND LEARNING FRAMEWORK ON THE GOOGLE EARTH ENGINE PLATFORM

D.Attaf^{*1,2}, K.Djerriri¹, D. Mansour¹, D. Hamdadou²

1 Department of Earth Observation, Centre des Techniques Spatiales, Arzew, Algeria,
2 Department of Computer Sciences, University of Oran 1 Ahmed Benbella, Laboratory LIO,
Algeria

Commission VI, WG VI/4

KEY WORDS: Burned area, feature extraction, Google Earth Engine platform, one-class classification, SVM Classification, Presence and background learning framework.

ABSTRACT:

Mapping of burned areas caused by forest fires was always a main concern to researchers in the field of remote sensing. Thus, various spectral indices and classification techniques have been proposed in the literature. In such a problem, only one specific class is of real interest and could be referred to as a one-class classification problem. One-class classification methods are highly desirable for quick mapping of classes of interest. A common used solution to deal with One-Class classification problem is based on one-class support vector machine (OC-SVM). This method has proved useful in classification of remote sensing images. However, over-fitting problem and difficulty in tuning parameters have become the major obstacles for this method. The new Presence and Background Learning (PBL) framework does not require complicated model selection and can generate very high accuracy results. On the other hand the Google Earth Engine (GEE) portal provides access to satellite and other ancillary data, cloud computing, and algorithms for processing large amounts of data with relative ease. Therefore, this study mainly aims to investigate the possibility of using the PBL framework within the GEE platform to extract burned areas from freely available Landsat archive in the year 2015. The quality of the results obtained using PBL framework was assessed using ground truth digitized by qualified technicians and compared to other classification techniques: Thresholding burned area spectral Index (BAI) and OC-SVM classifiers. Experimental results demonstrate that PBL framework for mapping the burned areas shows the higher classification accuracy than the other classifiers, and it highlights the suitability for the cases with few positive labelled samples available, which facilitates the tedious work of manual digitizing.

1. INTRODUCTION

Forest Fire is an important environmental issue in forest ecosystems. Each year, thousands of hectares are burnt in the Mediterranean countries, especially during the summer season. Even if forest fires can be beneficial, they endanger the forest biodiversity by changing the biomass stocks and the fauna. Moreover, they damage the soil fertility, the water quality, the hydrological cycles and can also lead to soil erosion (Zammit, et al., 2008).

Assessment of the damage caused by a forest fire plays an important role after the fire extinction. Indeed, an accurate detection provides crucial information for forest offices to plan the restoration and rehabilitation programs. It also helps the fire brigades to better locate the burnt areas, to understand the fire propagation and then, to pre-empt fight against future fires in the area. (Zammit, et al., 2008).

Remote sensing is a valuable tool to assess burnt areas. Several studies have shown the great potential of remote sensing concerning burnt area mapping since it provides spectral information on a large surface, even for remote areas. Many supervised classification methods have been proposed to address this problem. In traditional one, the completeness and representativeness of training sets are of vital importance to the classification accuracy. However, in many applications it is highly labor-intensive and time-consuming to collect sufficient training samples. There are also circumstances that we only concern about one specific class, regardless of other classes. In such cases, it is hard or not necessary to collect training sets for

classes other than the target class, and hence traditional supervised classification methods might be inaccurate due to incomplete training sets. A feasible solution to deal with these circumstances is to use one-class classification methods. In one-class classification, points belonging to a specific class of interest are denoted as positive data (i.e., presence data), while those belonging to other classes are referred to as negative data (i.e., absence data). One-class classifier aims to generate a decision model from an input training dataset that usually consists of positive data only, so as to recognize the positive data in a test dataset. A number of one-class classification methods have been developed. For example, one-class support vector machine (OC-SVM) seeks a hyperplane to separate most of the samples from the origin with the maximum margin in the feature space, and has proved to be useful in the one-class classification of remote sensing images. However, due to the lack of negative training samples, OC-SVM normally suffers from an over-fitting problem. Improvement can be achieved with the aid of samples to be classified, which are defined as unlabeled data. Recently, Li et al. 2011 proposed a new Presence and Background Learning (PBL) algorithm that has great advantage in one-class classification. This algorithm does not require complicated model selection and can generate very high accuracy results in niche modelling. (Ao, et al., 2017). Most recently, the growing availability of large-volume open-access remote sensing data and the development of advanced machine learning tools have been integrated with recent implementations of powerful cloud computing resources. This offers new opportunities for broader sets of applications at new

spatial and temporal scales in the geospatial sciences and addresses the limitation of existing methods and products. Specifically, the advent of powerful cloud computing resources, such as NASA Earth Exchange, Amazon's Web Services, Microsoft's Azure, and Google cloud platform has addressed these Geo Big Data problems. For example, Google Earth Engine (GEE) is an open-access, cloud-based platform for parallel processing of petabyte-scale data. It hosts a vast pool of satellite imagery and geospatial datasets, and allows web-based algorithm development and results visualization in a reasonable processing time. In addition to its computing and storage capacity, a number of well-known machine learning algorithms have been implemented, allowing batch processing using JavaScript on a dedicated application programming interface (API). Therefore, this study mainly aims to investigate the possibility of using the PBL framework within the GEE platform to extract burned areas from open access remote sensing data archive. (Mahdianpari, et al., 2018).

2. USED DATA AND METHODS

The proposed methodology consists in applying three classification schemes: thresholding burned area spectral Index, SVM one-class classification of Landsat-8 spectral bands and Presence and Background Learning framework based classification. OC-SVM classifier only requires positive labelled data, while PBL framework requires positive and unlabeled training samples.

2.1 Study Area and used data

The proposed methodology has been applied to the department Sidi Bel Abbes which is located in the north west of Algeria 56 km south of the Mediterranean Sea. The north of the department is a mountainous area that suffers every year of forest fires which cause large burned areas. The geographic situation of the study area is illustrated in Figure 1.

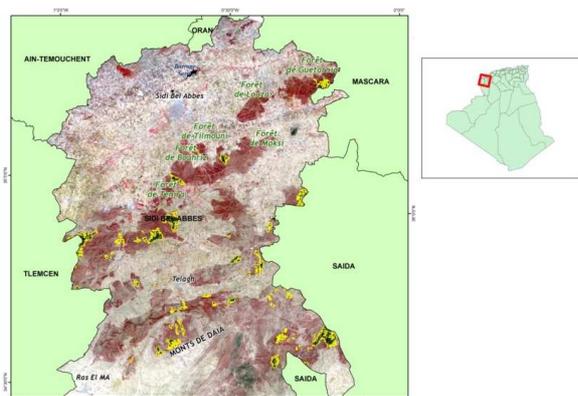
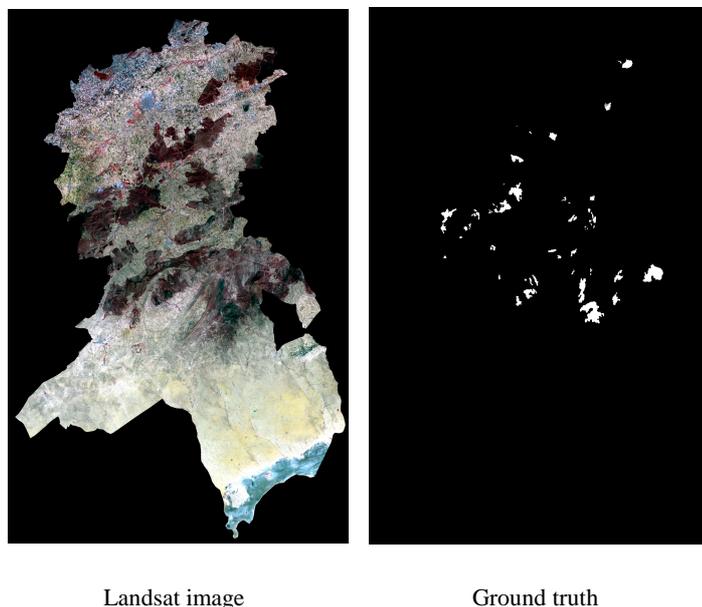


Figure 1. Areas affected by forest fires in the department of Sidi Bel Abbes summer 2015

2.2 Google Earth Engine platform

Google Earth Engine (GEE, <https://earthengine.google.org>) is a cloud-based platform for scientific analysis and visualization of petabyte-scale geospatial datasets (N. Gorelick, et al., 2017). It stores several decades of historical images and scientific datasets (Liu, et al., 2016). We used the Landsat 8 TM

calibrated top-of-atmosphere reflectance data, with the lowest cloud score from the images selection dates between June and August 2015. The quality of the results obtained using PBL framework was assessed using ground truth digitized by qualified technicians.



Landsat image

Ground truth

Figure 2. Landsat image used in our study and its associated ground truth

In this work, the three classification schemes cited bellow were performed on GEE platform.

2.3 Burned Area Index (BAI)

Spectral burn indices supply an optimum measure and beneficial signal for fire effects (Key and Benson, 2005). Hence, the Burn Area Index (BAI) was applied to Landsat 8 satellite image. The Burned Area Index (BAI) was developed by (Chuvieco et al. 2002) to assist in the delineation of burn scars and assessment of burn severity. It is based on the spectral distance to charcoal reflectance. It uses red and near-infrared (NIR) wavelengths.

$$BAI = \frac{1}{(0.1 - Red)^2 + (0.06 - NIR)^2} \quad (1)$$

2.4 One-Class Support Vector Machine (OC-SVM)

OC-SVM is a machine-learning algorithm derived from the standard Support Vector Machine algorithm and designed to tackle single class classification problems. It optimizes separation of the target class by constructing a hyperplane that best represents the multidimensional edge of the feature space. The hyperplane is obtained by minimizing the function:

$$(w, \xi, \rho) = \frac{1}{2} \|w\|^2 - \rho + \frac{1}{mv} \sum_{i=1}^m \xi_i \quad (2)$$

Where v is used to control the volume of the sphere that bounds the portion of the feature space where the training samples are

located. It defines an upper limit to the outliers fraction found in the data, w is the width of the region besides the decision boundary, and m is the number of training instances.

Although developed for linear problems, different kernel functions for nonlinear problems can be used with OC-SVM. Among the kernel functions available for computation, the most widely used are the radial basis function (RBF), the sigmoid function, and the polynomial function.

Kernel choice is based on the data type and on the n -dimensional feature distribution. In the present analysis we used the RBF kernel in the present analysis which has been extensively and successfully used in remote sensing image-processing tasks. According to the boundaries defined by the kernel function of the OC-SVM, the classifier returns positive values for classes that are similar to the classes in the training sample and negative values otherwise. (A. A. Pereira, et al., 2017)

2.5 PBL one-class Classification framework

Unlike the commonly used positive-only one-class methods like OC-SVM. We used PBL, a positive unlabeled framework (Li, et al., 2011) that has been successfully applied in one-class remote sensing classification (Deng, et al., 2011), (Ao, et al., 2017). A brief description of PBL is provided in the following (Ao, et al., 2017).

Let $y = 1$ denote positive data, $y = 0$ denote negative data, and x denote the vector of features associated with a SP. In one-class classification, we aim to model the probability of a pixel being from the positive class conditioned on its features, $\Pr(y = 1 | x)$. Traditional supervised learning methods cannot correctly model $\Pr(y = 1 | x)$ without negative samples $\Pr(y = 0)$. To address this problem, the PBL framework was developed to model $\Pr(y = 1 | x)$ without relying on samples of the negative class (Li, et al., 2011).

In PBL algorithm, a different model is first trained using a dataset that consists of labeled positive (denoted as $s = 1$) and unlabeled samples (denoted as $s = 0$). This model is then used to calibrate the desired model. The trained model is denoted as $\Pr(s = 1 | x, \eta = 1)$, where $\eta = 1$ denotes the positive-unlabeled case-control sampling scenario. The unlabeled data may contain both positive and negative samples, so the trained model $\Pr(s = 1 | x, \eta = 1)$ is not equal to the desired model $\Pr(y = 1 | x)$ but they have the following relationship:

$$\Pr(y = 1/x) = \frac{1-c}{c} \times \frac{\Pr(s=1/x, \eta=1)}{1-\Pr(s=1/x, \eta=1)} \quad (3)$$

where c is a constant (Li, et al., 2011). Thus, the obtaining of the desired model $\Pr(y = 1 | x)$ needs to estimate first the constant c . Let O denote a set of the 'prototypical positive' pixels. The predicted value $\Pr(s = 1 | x, \eta = 1)$ of any prototypical positive pixel is equal to c (Li, et al., 2011). In practice, the predicted probabilities of multiple prototypical positive pixels are averaged to obtain a more reliable estimator of c :

$$c = \frac{1}{n} \sum_{x \in O} \Pr(s = 1/x, \eta = 1) \quad (4)$$

Where n is the cardinality of O .

PBL requires a binary classifier that can estimate posterior probability be trained on the labelled positive and unlabeled data. In this work we used the BPL algorithm since GMO Maximum Entropy classifier (gmoMaxEnt) implemented in

GEE platform (Deng, et al., 2011). Multinomial logistic regression is generalization of linear regression using the softmax transformation function and the main task is to minimize an error function by taking the negative logarithm of the likelihood, which means cross-entropy. The main difference from other models and algorithms is the outcome score that could be considered as a probability value (Haykin, 2008), (C.M. Bishop, 2006),

The BPL trained using labelled positive and unlabeled samples results in the naive model $\Pr(s = 1 | x, \eta = 1)$. The predicted values of the positive samples in the training set were then averaged to estimate the constant c . The naive model $\Pr(s = 1 | x, \eta = 1)$ was used to calibrate the desired model $\Pr(y = 1 | x)$ though Equation (1). Binary classification of a test sample is achieved by applying a decision threshold of 0.5.

3. EXPERIMENTS, RESULTS AND DISCUSSION

We evaluate the classification by employing 02 different criteria: the estimated Cohen's Kappa statistic (k), and the Fscore. The first metric (k) is an overall accuracy metric which compensates for the chance agreement between classes. The second score is the average of the harmonic means between precision and recall for the burned area class.

This measure is sensitive to class accuracy, but additionally takes into account the number of correctly classified pixels over the number of predicted labels (Sokolova, et al., 2009).

Table I illustrates the one-class classification results of PBL, OC-SVM, and binary classification results of thresholding burned area spectral Index for the burned area class in the study area. Accuracy assessment using ground truth classes revealed that the all algorithms successfully extracted the class of interest in the Landsat dataset (Table II). The highest accuracies were obtained with the BPL classifier with F-score of 0.87 for the Ain-Adden area.

We note that the thresholding burned area spectral Index gives the second better F-score for the Taoudmout, as shown in table I, however the F-score and Kappa statistic results given by this classifier are strongly related to the threshold value, in this work the threshold is set to 40.

OC-SVM does not perform well in our experiments. The obtained results show that positive-unlabeled methods generally outperform positive-only methods (Li, et al., 2011). In addition, OC-SVM is sensitive to the input parameters such as gamma and nu coefficient, which can hinder its adoption by non-expert users.

Communes	OC-SVM		OC-PUL		BAI (40)	
	F1	Kappa	F1	Kappa	F1	Kappa
Ain Adden	0.67	0.64	0.87	0.86	0.73	0.70
Bir El Hmam	0.67	0.66	0.78	0.78	0.59	0.58
Taoudmout	0.73	0.69	0.76	0.73	0.79	0.76
Oued Taourira	0.55	0.52	0.73	0.72	0.60	0.58
Tafissour	0.54	0.52	0.65	0.63	0.67	0.66
Tenira	0.31	0.30	0.74	0.74	0.37	0.36

Table 1. Classification results of burned areas class

Table II illustrates the one-class classification results of PBL, OC-SVM, and BAI thresholding for the burned Area class in the study area.

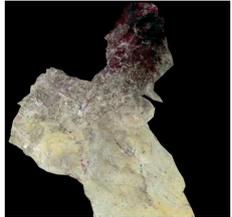
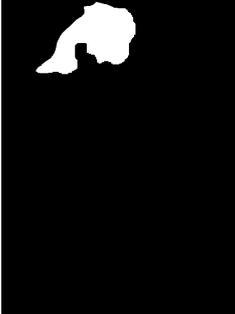
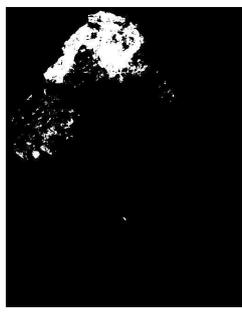
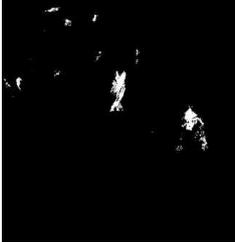
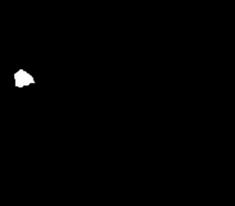
False-Color composite	Ground-truth	OC-BPL Ain-Aden	OC-SVM	BAI Index
				
Bir El Hmam				
				
Taoudmout				
				
Oued Taourira				
				
Tafissour				
				
Tenira				
				

Table 2. Results of extraction of burned areas (from left to right: false color composite image, Ground truth, PBL, OC-SVM and BAI)

4. CONCLUSION

In this work we applied three classification schemes for the extraction of burned areas from Landsat-8 images. The experiments demonstrated high kappa statistic and F-score for the PBL framework based classification with the advantage that it requires only labelled samples of the class of interest with unlabeled samples. The obtained experimental results confirm that PBL based approach is one of the most robust one-class classification methods, consistently outperforming OC-SVM, where both additional labelled and unlabeled data are responsible for the improvement of classification accuracy.

REFERENCES

- O. Zammit, X. Descombes, and J. Zerubia., 2008: "Combining one-class support vector machine learning and hysteresis thresholding: Application to burnt area mapping".
- Z. Ao 1, Y. Su, W. Li, Q. Guo, and J. Zhang., 2017, " One-Class Classification of Airborne LiDAR Data in Urban Areas Using a Presence and Background Learning Algorithm", 27 September 2017.
- M. Mahdianpari, B. Salehi, F. Mohammadimanesh, S. Homayouni and E. Gill., 2018: "The FirstWetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform".
- X. Liua, G. Hub, Y. Chenb, X. Lib, X. Xua, S. Lic, F. Peid, S. Wang, ., 2018: "High-resolution multi-temporal mapping of global urban land using Landsat images based on the Google Earth Engine Platform", 1-848.
- X. Deng, W. Li, X.Liu, Q. Guo, and S. Newsam., 2017: "One-class remote sensing classification: one-class vs. binary classifiers". *International Journal of Remote Sensing*, 39(6), 1890-1910.
- Z. Ao, Y. Su, W. Li, Q. Guo, and J. Zhang., 2017: "One-Class Classification of Airborne LiDAR Data in Urban Areas Using a Presence and Background Learning Algorithm". *Remote Sensing*, 9(10), p.1001.
- M. Sokolova, and G Lapalme., 2009: "A systematic analysis of performance measures for classification tasks". *Information Processing & Management*, 45(4), 427-437.
- S. Haykin., 2008: *Neural Networks and Learning Machines*, 3rd Edn. Upper Saddle River: NJ: Prentice Hall.
- C.M. Bishop., (2006): "Pattern Recognition and Machine Learning". New York, NY: Springer.
- A. A. Pereira, J. M. C. Pereira, R. Libonati , D. Oom , A. W. Setzer, F. Morelli , F. Machado-Silva and L. Marcelo Tavares de Carvalho., 2017. "Burned Area Mapping in the Brazilian Savanna Using a One-Class Support Vector Machine Trained by Active Fires". *Remote Sens*, 9, 1161;
- N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thaub, R. Moore., 2017, "Google Earth Engine: Planetary-Scale Geospatial Analysis For Everyone". *Remote Sensing of Environment*, <https://doi.org/10.1016/j.rse.2017.06.031>
- E. Chuvieco, M. P. Martin, and A. Palacios, 2002. "Assessment of different spectral indices in the red near-infrared spectral domain for burned land discrimination". *International Journal of Remote Sensing*, 23, 51035110.
- W. Li, Q. Guo, and C. Elkan., 2011 "Can we model the probability of presence of species without absence data?". *Ecography*, 34(6), 1096-1105.