Wildfires early alert model *

GO-HUB

October, 2018

1 Introduction

Fires burn millions of hectares of plant surface in the earth every year. This has a great economic, human and ecological cost. Early detection of wildfires is key to minimizing their effects through early intervention of forest services. Also, could prevent possible loss of human life by alerting people in areas at risk.

Currently, each region depends on its own fire detection system. Despite this, there are areas of the Earth populated with vegetation that do not have any fire warning system or that have inefficient fire warning systems.

A global fire warning solution is suitable both for areas with low or no coverage and for strengthening warning systems in areas with a functional warning system. There are fire detection systems such as the *NASA* which warns of fires within three hours of their occurrence (`http://search.earthdata.nasa.gov`). Unfortunately sometimes 3 hours is too long. Fires are hardly controllable once they have begun to spread.

This work proposes a global fire detection system based on combining fire data from *NASA* and other sources with real-time analysis of social networks. In some cases this system could reduce the time of the fire detection system from the *NASA*. The system does not require the active collaboration of people although it can benefit from it.

Generally speaking, the early warning model consists of 4 parts:

1. Fire risk model: the global fire risk is estimated daily using a model based on [1] achieving a spatial resolution of 500m.

2. Selection mask: Excludes those areas that are not of interest to the model.

3. Textual analysis of social networks and APIs: As many text messages as possible are analyzed on social networks and other selected sources in a pseudo-random manner looking for wildfire alarms.

*GO-HUB (2018) Wildfires early alert model. *SOFIA: Fire Prevention Intelligence* `https://github.com/gohubvlc/sofia-project`

1

4. Analysis of images: If text analysis is not enough to decide whether to issue an alert, the decision is reinforced by analyzing social network images and satellite images.

5. Decision-making based on a linear model: Combining the analysis of social networks and the fire risk model a linear model evaluates whether to issue a fire alert.

2 The model

The following section will develop the model step by step.

2.1 Fire risk model

Starting in the last century fire danger rating (FDR) systems have been in operation in many countries around the world, especially in Canada, Australia, Russia and the United States [2]. In all cases, the interest is in the pre-fire environment [3], which refers to the type and condition of fuels as influenced by climate, weather and land management.

Satellite remote sensing contributes greatly to global environmental monitoring, offering accessible, low-cost, worldwide and updated daily information. Taking advantage of this information is key to know the level of fire risk before a fire occurs.

The prediction of fire risk areas has a central role in prevention and timely assistance. In this work we propose the use of different satellite products provided by NASA, based on scientific evidence on the subject [2], [4], to generate a fire risk model.

2.1.1 Data Requirements

It is possible to predict the conditions of fire risk from the temporal analysis of environmental variables built by remote sensing. The selection of the satellite products is based on the relevance of the measured variables, good spatial resolution and high revisit frequency. In turn, the model includes satellite products with processed information such as monthly burned areas or annual land cover.

The included satellite images have been:

1. Daily surface temperature (TS) at 1km spatial resolution (i.e, MOD11A1)

2. Daily precipitately water (PW) at 1km spatial resolution (i.e, MOD05L2)

3. Daily aerosol optical depth model (AOD) at 1km spatial resolution (i.e, MCD19A2)

4. 8-day composite of surface reflectance at 500m spatial resolution (i.e, MOD09A1)

5. Monthly burned-area at 500m spatial resolution (i.e, MCD45A1) [*]

6. Annual land cover map at 500m spatial resolution (i.e, MCD12Q1) [*]

Data can be downloaded daily from `https://ladweb.modaps.eosdis.nasa.gov` and `https://search.earthdata.nasa.gov`.

The images were re-sampled to 500m resolution to reduce between-pixel heterogeneity, mosaicked into a single image and filtered with a 3x3 pixel moving window using a Gaussian averaging kernel. The selected project was WGS84.

The following indices can be constructed from satellite products:

- Normalized Multiband Drought Index (NMDI) is calculate with the surface reflectance by the near infrared (NIR) and shortwave infrared bands centred at 860 nm, 1640 nm, and 2130 nm, as shown below:

$$NMDI = \frac{R_{860nm} - (R_{1640nm} - R_{2130nm})}{R_{860nm} + (R_{1640nm} - R_{2130nm})}$$

  where R represents the apparent reflectance observed by a satellite sensor. NMDI is useful for monitoring soil and vegetation moisture [5].

- Normalized Difference Vegetation Index (NDVI) is carried out with the red and NIR spectral bands centered at 640 nm and 860 nm as seen:

$$NDVI = \frac{R_{NIR} - R_{red}}{R_{NIR} + R_{red}}$$

  where R is the surface reflectance values of MODIS. NDVI is more directly related to chlorophyll activity of the vegetation than to droughtiness conditions [6].

The MCD12Q1 product provides six different classifications [7]. These include the Geosphere-Biosphere Programme land cover classification, the University of Maryland classification scheme, the Biome classification scheme described by [8], the LAI/fPAR Biome scheme described by [9], the Plant Functional Type scheme described by [10], and the FAO-Land Cover Classification System that contains the layers: land cover (LCC1), land use (LCC2) and the surface hydrology (LCC3).

All water bodies are masked using the surface hydrology (MCD12Q1 - LCC3).

The active fire obtained from the products VNP14 (M-band) and VNP14IMG (I-band) is used for model validation. Each active fire location represents the center of a 1 km pixel that is marked by the algorithm as containing a fire within the pixel. The images can be downloaded from `https://ladsweb.modaps.eosdis.nasa.gov`.

2.1.2 Filling the missing data

Remotely sensed imagery time-series often presents gaps that reduce the usefulness of these data sources for modeling and monitoring environmental phenomena. In particular, gaps are problematic in tropical and subtropical areas where

3

persistent cloud cover can obscure portions of the image seasonally or throughout the year. It is possible to find in the bibliography different approaches to fill the gaps, being more convenient for the characteristics of this study, to select a Spatio-temporal gap-filling.

We select the methodology present in [11] that leverages the wealth of empirical information present within large imagery time-series to fill missing pixels. Their method produces highly accurate results while utilizing a conceptually simple, computationally efficient algorithmic framework.

2.1.3 Risk model algorithm

Detailed methodology is based on work published by [4]. The proposed system consisted of eight stages:

1. To maximize the potential of the model, all six land cover categories are used.

2. We relate the input variables (TS, PW, AOD, NMDI and NDVI) for each of the land use categories.

3. We calculated the specific average values by study category for the input variables during the period i (i.e., $\overline{NMDI_i}$, $\overline{NDVI_i}$) and j day (i.e., $\overline{PW_j}$, $\overline{TS_j}$ and $\overline{AOD_j}$) using areas that did not have fires during the last year. At the same time, as these variables present a temporary behavior, the seasonal averages were calculated to decrease the error in the estimation.

4. For each pixel and throughout each period $i + 1$ or day $j + 1$, the fire danger conditions were calculated. For this purpose, the instantaneous values of the input were compared with their respective seasonal mean values calculated in the previous step.

5. As a condition of high fire danger, the prevalence of the following criteria was analyzed:

   - $TS_j \geq \overline{TS_j}$: high temperatures may favour fire
   - $NMDI_i \leq \overline{NMDI_i}$: low vegetation moisture may favour fire
   - $NDVI_i \leq \overline{NDVI_i}$: low greenlees may indicate dry vegetation and therefore a good fuel
   - $PW_j \leq \overline{PW_j}$: The low level of water vapour in the atmosphere could be associated with the flammability of fuels
   - $AOD_j \geq \overline{AOD_j}$ high AOD values may indicate the presence of smoke

6. Five hazard categories were established based on the following conditions:

   - Extremely high: when all five variables belong to the high hazard class

- Very high: when at least three of the five variables belong to the high hazard class
- High: when at least two of the five variables belong to the high hazard class
- Moderate: when at least one of the five variables belongs to the high hazard class
- Low: when all five variables belong to the low hazard class.

7. The FDR was applied of risk according to the risk category most frequently for each land cover.

8. In order to validate the modeled categories, the historical information from 2011 (VIIRS images available) of active fires and monthly burned-area was employed.

2.2 Selection masks

The early warning fire system proposed in this work operates on a global scale. This implies that it will require a large computing capacity. In order to reduce the load of calculations some areas will be excluded using Boolean filters. These areas are:

- Areas with more than 10k inhabitants/km^2 obtained from `https://landscan.ornl.gov`. We assume that these areas are urban areas. It is not useful to alert of fires in these areas since the goal of the model is to generate wildfire alerts. Furthermore, if a fire were to be observed from a densely populated urban nucleus, the authorities would undoubtedly be alerted in time as a large number of people would be able to see it.

- Areas in which no fire has occurred since 2000 according to the *NASA* fire database *Fire Information for Resource Management System (FIRMS).*

- Areas in which there is an active fire according *FIRM*.

.

2.3 Social network and API text analysis

In this stage as much post geolocalized on social networks and information received from the API SoFiA and related with to is downloaded. Those regions covered by the masks produced in the previous stage of the model will be ignored for the search 2.2. Even ignoring these regions there is too much information on social networks to process it all. A selection criteria is established. It prioritizes the post according to the fire risk fr evaluated in its location. In addition, a minimum threshold of fr from which no post is taken into consideration, t, is established. t will depend on the maximum amount of post that can be processed in each batch, c_{max}.

In each batch c_{max} posts are downloaded and sorted by its fr. We set $n = 10$ percentiles in the range $fr > t$. In each of them as many post as $\int_{fr_i}^{fr_{i+1}} g(fr) \times c_{max}/(\int_t^1 g(fr))$ are processed. They are randomly chosen from among the candidates, being $g(fr) = e^{(x-t)/(1-t)ln(2)} - 1$.

The selected posts are processed to be introduced into a dense neural network with a depth to be determined in an optimization process. The following variables are considered for the dense neural network:

- Bag of fire-related keywords contained in each post. The keywords will be determined from among those that can be found in a post that alerts about a fire. For example textfire, textsmoke, textfire, etc.

- Reverse of the population density in the position of the post. It is intended to penalize the areas with more population and therefore with more possibility of fire sighting.

- Fire risk estimated in the first stage of the model.

The network has a single soft-max output that indicates if the post is alerting or not about fire. Positive reports are clustered whit the *Mean-Shift Clustering* algorithm. This is simply a proposal, other clustering algorithms might be more convenient for the model.

The possibility of it being a real alert is evaluated in each report as

$$AlertRisk = \frac{p}{d \times fr}$$

, being p the number of positive reports in the cluster. Thus, in an uninhabited area a single report would be enough to trigger the alert for a high fire risk and on the contrary many reports would be needed in a cluster with low fire risk and densely populated.

2.4 Image analysis

If *AlertRisk* is too high with respect to an established threshold it is considered necessary to issue an alert automatically. If *AlertRisk* is too low with respect to an established threshold the cluster is discarded. In any other case, images are requested to reinforce the area enclosed by the complex hue of the posts forming the cluster or the posts enclosed in an area of 5 km around each point if there are less than four points in a cluster. The images will be obtained from the following sources: post from social networks that have been evaluated and contain images, the geostationary satellite network *GOES* and the SoFia API. These images are evaluated by a ResNet[12] object detection neuronal network. It is important to note that although the architecture of the networks that process the satellite images, the images from the social networks, and the API may match their sets of images are different.

Users who volunteer will also be able to catalogue images of possible fires. They will receive as many images as possible. The judgment given by the user

will prevail over the model if it is the case that a same image has been classified by a human and the object identification network.

2.5   Decision making based on a linear model

Finally, the results are entered into a linear model with two input variables:

- The percentage of images evaluated as fires.

- The average fire risk at the points where warnings have been issued.

This last regressor will decide whether or not to issue a warning.

3   Training

Different parts of the model are machine learning algorithms that require training. This section indicates where the tagged data is obtained for each of them.

It will be necessary to dedicate staff to tagging for the dense network that is used for the detection of alerts in social networks and API. The help of the community of SoFiA users could also be requested. In any case, active learning techniques will be used to determine which samples are to be tagged.

Image recognition networks will be trained with images of places and dates where *FIRMS* indicates that there has been a fire and have also been catalogued as fires by APP users.

4   Conclusions

A fire early warning model based on the use of social networks is proposed, limiting the study area to the maximum in order to make the analysis viable.

The quality of some parts of the model will depend on the collection of sufficient data to make it viable.

This work is a proposal that leaves many points open due to lack of time either to develop them fully or because they require in-depth analysis. Despite this, the proposal is realistic and the model is viable.

References

[1] Ehsan H Chowdhury and Quazi K Hassan. Development of a new daily-scale forest fire danger forecasting system using remote sensing data. *Remote Sensing*, 7(3):2431–2448, 2015.

[2] Ehsan H Chowdhury and Quazi K Hassan. Operational perspective of remote sensing-based forest fire danger forecasting systems. *ISPRS Journal of Photogrammetry and Remote Sensing*, 104:224–236, 2015.

[3] Theresa Jain, Russell Graham, and David Pilliod. Tongue-tied: confused meanings for common fire terminology can lead to fuels mismanagement. pages 22–26, 2004.

[4] Ehsan H Chowdhury and Quazi K Hassan. Use of remote sensing-derived variables in developing a forest fire danger forecasting system. *Natural Hazards*, 67(2):321–334, 2013.

[5] Lingli Wang, John J Qu, and Xianjun Hao. Forest fire detection using the normalized multi-band drought index (nmdi) with satellite measurements. *Agricultural and forest meteorology*, 148(11):1767–1776, 2008.

[6] B Leblon, M Alexander, J Chen, and S White. Monitoring fire danger of northern boreal forests with noaa-avhrr ndvi images. *International Journal of Remote Sensing*, 22(14):2839–2846, 2001.

[7] M Friedl and D Sulla-Menashe. Mcd12q1 modis/terra+ aqua land cover type yearly l3 global 500m sin grid v006 [data set]. *NASA EOSDIS Land Processes DAAC. doi*, 10, 2015.

[8] Steven W Running, Ramakrishna R Nemani, Faith Ann Heinsch, Maosheng Zhao, Matt Reeves, and Hirofumi Hashimoto. A continuous satellite-derived measure of global terrestrial primary production. *AIBS Bulletin*, 54(6):547–560, 2004.

[9] Ranga B Myneni, S Hoffman, Yuri Knyazikhin, JL Privette, J Glassy, Yuhong Tian, Y Wang, X Song, Y Zhang, GR Smith, et al. Global products of vegetation leaf area and fraction absorbed par from year one of modis data. *Remote sensing of environment*, 83(1-2):214–231, 2002.

[10] Gordon B Bonan, Samuel Levis, Laurent Kergoat, and Keith W Oleson. Landscapes as patches of plant functional types: An integrating concept for climate and ecosystem models. *Global Biogeochemical Cycles*, 16(2):5–1, 2002.

[11] Daniel J Weiss, Peter M Atkinson, Samir Bhatt, Bonnie Mappin, Simon I Hay, and Peter W Gething. An effective approach for gap-filling continental scale remotely sensed time-series. *ISPRS Journal of Photogrammetry and Remote Sensing*, 98:106–118, 2014.

[12] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.