Predicting the burned area in forest fires

Project proposal

Pietro Dellino
pietro.dellino@student.ecp.fr

Lucie Donne
lucie.donne@student.ecp.fr

Valentin Gonsolin
valentin.gonsolin@student.ecp.fr

Baptiste Marvaldi
baptiste.marvaldi@student.ecp.fr

MOTIVATION AND PROBLEM DEFINITION

Forest fires are a really important issue because they are dangerous and must be stopped before they are spread too much to be extinguished by human means. Being able to predict them precisely is then very helpful to organize human and logistical resources to counter them in time.

Our aim is not only to predict a forest fire, which is already commonly done by the relevant authorities, but also to predict the burned area in case of fire, which could give an idea of the resources to be mobilised.

Our algorithm could be a support for authorities, who could use it as a complement of the systems they already use to calculate the risk of forest fire.

Our study will be based on a work called “Machine Learning and Dataming Algorithms for Predicting Accidental Small Forest Fires” (see references), especially for the methods of dimensionality reduction, because our aim is different: we don’t want to predict the location of the fire.

Figure 1: A forest fire near Red Lake, Ontario, in May 1986. (Photo by Brian J. Stocks)

METHODOLOGY

Our goal is to establish links between the burned area and various parameters such as humidity, temperature, wind, the period in the year. Since we know the output value, we will use supervised learning and more precisely regression because the output value is a real number. There are other studies on forest fires which use for example clustering algorithms in order to understand where fires are most likely to start.

Yet, we will be more interested in the burned surface which is why we will be using different methods. We will also be using dimensionality reduction methods in order to simplify our dataset. Dimensionality reduction also allows us to apply our method to other datasets where the parameters can differ a bit. This problem is really interesting then because it's linked to many machine learning methods.

Our work will be, in the first place, to completely define the parameters and the model. Then, we will use dimensionality reduction to simplify the model and adapt our data. Finally, we will use regression algorithms depending on the data and for example the least square method.

We hope that this method, which is quite typical in machine learning problems, will allow us to get a relevant solution for our problem that we will check with in the evaluation part.

EVALUATION

For the training of our algorithm we will use an existing dataset of the numerous forest fires that occurred in the northern region of Portugal. This dataset includes geographical coordinates, the date and a lot of meteorological information (like wind, humidity, temperature, rain…). Some of this information is useless to us, for instance the geographical coordinates, because our

aim is to predict the burned area and not the position. Moreover, we want to use algorithm not only in Portugal, but in other countries too.

To assess the efficiency of our program, we will use in a first place the part of the dataset unused in the first step of training. With the input values we can check if the predicted burned area is far from the real one.

We can then use a dataset from another country, for example one from the United States, and then check once again if the algorithm works for another country.

Finally, for the sake of getting the most efficient algorithm, we'll use the different machine learning algorithms seen during the lessons and take on the one with the smallest error in result of the simulations.

REFERENCES

[1] Ian Downard. 2018. *Predicting forest fires with spark machine learning*. https://mapr.com/blog/predicting-forest-fires-with-spark-machine-learning/

[2] Vasanth Iyer, S. Sitharama Iyengar, N. Paramesh, Garmiela Rama Murthy, Mandalika B. Srinivas. 2011. *Machine Learning and Dataming Algorithms for Predicting Accidental Small Forest Fires*. International Institute of Information Technology, Hyderabad, India. Louisiana State University, Baton Rouge, USA. University of New South Wales, Sidney, Australia. Brila Institute of Technology & Science, Hyderabad Campus, India.

[3] D. Dua, E. Karra Taniskidou. 2017. *UCI Machine Learning Repository*. http://archive.ics.uci.edu/ml. Irvine, CA: University of California, School of Information and Computer Science.

[4] Paulo Cortez, Anibal Morais. *A Data Mining Approach to Predict Forest Fires using Meteorological Data.* Department of Information Systems/R&D Algoritmi Center, University of Minho, Guimarães, Portugal.