Using Remote Sensing Data to Understand Fire Ignition during the 2019-2020 Australia Bushfire Season

A thesis submitted for the degree of

Bachelor of Commerce (Honours)

by

Weihao Li

28723740

Department of Econometrics and Business Statistics

Monash University

Australia

May 2020

Contents

Chapter 1

Statement of the topic

Along with the extreme heatwave in Australia 2019-2020, one of the most devastating bushfire season in history had been witnessed. Lighting strikes and arson were been discussed among pubilc as the main cause of this disaster. This research will explore the methods of fire ignition during 2019-2020 Australia bushfires season and provide a model to predict the fire risk of neighbourhoods. Hotspots data from the JAXA's Himawari-8 satellite and weather data from the Australian Bureau of Meteorology will be used in ignition identification and bushfire danger estimation. In addition, an interactive web application embeds with research outcomes will be built for data visualization purpose.

1.1 Motivation

<left blank>

1.2 Research aim and questions

This research aims to answer the following questions:

1. How to find the ignition methods during 2019-2020 Australia bushfire season?
2. How to model fire risk of neighbourhoods?

1.3 Research plan

<left blank>

1.4 Scope of research

<left blank>

Chapter 2

Literature Review

<left blank>

Chapter 3

Data collection and processing

In order to answer the reserach questions, this research will use mutiple geospatial datasets and weather datasets.

3.1 Sources

Two methods were performed to collect necessary datasets, including retrieve files from servers and crawl data from websites. For reproducibility purpose, only public data will be used in this research.

To obtain fire history, we used the hotspots data from JAXA's Himawari-8 satellite. This satellite is positioned in geostationary orbit at 140 degrees east longitude, and the revisit period is 10-minute. Its management system - JAXA's P-Tree system, provides WildFire observation product with 2km spatial resolution. However, it is a beta version product and the data is not quality assured (P-Tree System, 2020).

Weather data were collected from the Australian Bureau of Meteorology, by using an R package - Bomrang. Due to the limitation of APIs provided by the package, we crawled data from BOM's website for extra information.

To characterise the fuel we used forest of Australia (2018) from the Australian Bureau of Agriculture and Resource Economics and Sciences. It is a fuel layer contains the vegetation information across Australia.

Table 3.1: *Data information*

Data set name	Spatial Resolution	Temporal resolution	Time
Hotspots data - JAXA's Himawari-8 satellite	$0.02^{\circ} \approx 2km$	Per 10 minutes	2015-2020
Weather data - Australian Bureau of Meteorology		Daily	2019-2020
Map - OpenStreetMap	2m		2020
Fuel layer - Australian Bureau of Agriculture and Resource Economics and Sciences	100m		2018
Victorian CFA fire stations - Department of Environment, Land, Water & Planning	20m		2020
Victorian Recreation sites - Department of Environment, Land, Water & Planning	10m		2020

Information on the datasets not covered above can be found in Table 3.1.

3.2 Pre-processing

An R package sf (Edzer, 2020) has been used to convert all geospatial data into simple features. We then used Tidyverse (Hadley et al., 2020) as a data wrangling tool to manipulate data.

To preprocess the hotspots data, we selected the observations in Australia from the full disk. Meanwhile, hotspots with irradiance under 100 watt per square metre will be deleted. We restricted our study to hotspots that have significant firepower. An hour id has been assigned to each observation range from 1 to T, represents the relative time the hotspot being observed. On top of the tidy hotspots data, a clustering algorithm was developed to identify fire clusters. Details about the algorithm can be found in table 3.2.

Table 3.2: *A clustering algorithm for hotspots*

Algorithm 1 Hotspots clutering

input: Hotspots dataset H : (Hour_id$^{(n)}$, Coordinates$^{(n)}$), n = 1, 2, ... N
An empty dataset F : (Fire_id$^{(m)}$, Coordinates$^{(m)}$, Active$^{(m)}$), m = 1, 2, ...
An empty vector K $\in \mathbb{N}_1^n$
A distance hyperparameter $r_0 \in \mathbb{R}^+$
A time hyperparameter $t_0 \in \mathbb{N}^+$

output: A vector K $\in \mathbb{N}_1^n$ contains memberships of hotspots
A dataset F contains fire clusters information including memberships, latest centroids and time from last updated

1 : select subset $H_c \in$ H where Hour_id == 1
2 : calculate distance matrix D for Coordinates in H_c
3 : assign 1 to a zero adjacency matrix A for where D $\leq r_0$ // hotspots with relative distance less or equal to r_0 will be considered belong to the same cluster
4 : create undirected unweighted graph G from A
5 : record memberships of G to K
6 : record clusters classes to Fire_id and record clusters centroids to Coordinates of F
7 : set Active in F to t_0 // Active clusters are fire being observed in the last t_0 hour
8 : **for** hour = 2, ... T **do**
9 : let Active -1 and select subset $F_c \in$ F where Active ≥ 0
10 : select subset $H_c \in$ H where Hour_id == hour
11 : append Coordinates from F_c to H_c
12 : repeat step 2 - 4
12 : **for** $h_i =$ each hotspot in H_c **do**
13 : **if** h_i share the same membership as one of active clusers in F_c **then**
14 : copy the corresponding Fire_id of the nearest active cluser to K
15 : **else** copy the membership from G to K
16 : **end if**
17 : **end for**
18 : update F for clusters involed in current timestamp and reset corresponding Active to t_0
19 : **end for**

Chapter 4

Exploratory data analysis

Chapter 5

Modelling

<plan>

5.1 Predicting ignition method

<plan>

5.2 Modelling fire risk

<plan>

Chapter 6

Timeline

The research plan for this semseter can be found in Table 6.1. Future research plan can be found in Table 6.2.

Table 6.1: *Research plan till week 9*

Timeline	Tasks
Week 2	Geographic data background reading
Week 3	Collect Remote sensing data (JAXA himawari-8 satellite) and explore BOM weather data APIs (Bomrang)
Week 4	Collect Road Map (OpenStreetMap) and read articles in SpatioTemporal data visualization and modelling
Week 5	Develop clustering algorithm for remote sensing data
Week 6	Test diferent hyperparameters for clusetring
Week 7	Exploratory data analysis on fire clusters and data integration
Week 8	Feature planning for the shiny app
week 9	Write research proposal and prepare the first presentation

Table 6.2: *Research plan since June*

Timeline	Tasks
June - July	Modelling fire ignition and fire risk
August	Consolidate findings and create mockups of the shiny app
September	Develop the shiny app and perform different levels of testing
October	Write thesis and prepare the second presentation

Chapter 7

Supplementary materials

Chapter 8

Biblography

Chapter 9

Words from last week

9.1 Data

9.2 Methodology

Both supervised and unsupervised learning will be implemented to reach the research aims.

To understand the ignition of bushfires, a customized clustering algorithm will be developed to convert hotpots data into fire history, which will contain the starting time and coordinates of each fire. This algorithm will mainly involve simulating fire growth, deciding fire boundaries controlled by tolerance and assigning hotspots data to the most probable cluster. After the clustering result being obtained, fire history will be visualized to diagnostic the performance of the algorithm. It will be done by comparing the behaviour of the same fire under different sets of hyperparameters.

Exploratory data analysis of fire history and its relative factors, like weather condition, distance to the nearest road and distance to the nearest recreation site will be performed. Prior knowledge and featuring engineering will be needed to fully understand the relationship. We expect to discover relationships between the ignition of fire with these factors, which can help us identify the cause of bushfires later on.

In order to examine the sources of fire ignition, different strategies will be used depending on the outcome in the previous section. If the findings from the analysis are strong and directly related to potential sources of fire ignition, hypothesis tests will be conducted to examine the pattern. If the evidence is weak, we will consider developing another clustering algorithm on fire history. This algorithm will be designed to maximize the distance between bushfire started with different causes in a high dimensional space. A probability model then can be built on top of it, which can provide a probabilistic answer for the source of bushfire ignition during 2019-2020 bushfire season.

Models for predicting fire risk of neighbourhoods will be built using raw hotspots data instead of the fire history because the hotspots data can be considered generated from a partially observable Markov decision process, and the underlying state is the development of the bushfire. From low complexity models like logistic regression to high complexity models like deep neural network will be tested.

For sharing our research outcome, a shiny app will be built and hosted online. In addition, both static and dynamic visualization tools will be considered using. Due to the nature of Spatio-temporal data, which has at least 3-dimensional features, static map view without faceting can only provide limited information. Meanwhile, faceting map view with time will be limited by the size of caravans. Animation based map view is computationally expensive and distracting though it provides more information. Better ways for visualizing Spatio-temporal data will be explored during the development. The potential product will be an interactive map view with triggers to transform data and manipulate the aesthetics specifications.

9.3 Preliminary Results

Appendix A

Additional stuff

You might put some computer output here, or maybe additional tables.

Note that line 5 must appear before your first appendix. But other appendices can just start like any other chapter.