A clustering algorithm to organize satellite hotspots data for the purpose of tracking bushfires remotely

by Weihao Li, Emily Dodwell, and Dianne Cook

Abstract An abstract of less than 150 words.

Introduction

Bushfires are a major problem for Australia, and many other parts of the globe. There is concern that as the climate becomes hotter, and drier, that the impact of fires becomes much more severe and extensive. In Australia, the 2019-2020 fires were the worst on record causing extensive ecological damage, as well as damage to agricultural resources, properties and infrastructure. The Wollemi pine, rare prehistoric trees, required special forces intervention to prevent the last stands in the world, in remote wilderness areas, from being turned into ash.

Contributing to the problem is that many fires started in very remote areas, locations deep into the temperate forests ignited by lightning, that are virtually impossible to access or to monitor. Satellite data provides a possible solution to this, particularly remotely sensed hotspot data, which may be useful in detecting new ignitions and movements of fires. Understanding fires in remote areas using satellite data may provide some help in developing effective strategies for mitigating bushfire impact.

This work addresses this topic. Using hotspot data, can we cluster in space and time, in order to determine (1) points of ignition and (2) track the movement of bushfires.

This paper is organised as follows. The next section provides an introduction to the literature on spatiotemporal clustering and bushfire modeling and dynamics. Section Algorithm describes the clustering algorithm, and section Application illustrates how the resulting data can be used to study bushfire ignition.

Background

Algorithm

Data pre-processing

To illustrate this algorithm, hotspot data during 2019-2020 Australia bushfire season taken from Himawari-8 satellite (P-Tree System, 2020) was used. This hotpot dataset contained records of 1989572 hotspots for 6 months in the full disk of 140 °east longitude. Before applying the algorithm on the data, two necessary pre-processing steps needed to be performed. Given we focused on the bushfires in Victoria, only the hotspots within the boundary of Victoria were kept. Besides, a threshold (irradiance over 100 watts per square metre) for fire power was used to reduce noises from the background. The final dataset contained 75936 hotspots, which is shown in Figure 1.

Steps

This algorithm was consist of 4 steps:

1. Divided hotspots into different segments
2. Clusterd hotspots in each segment individually
3. Updated the clustering results recursively
4. Computed the ignition location for each cluster

1. Divided temporal dimension into different segments

One of the difficulties to perform clustering on spatiotemporal data was to determine the scaling of the temporal dimension. Similarly, a characteristic of hotspot data was cloud cover could lead to missing observations of a bushfire in several hours. One possible solution for both issues was flattening the temporal dimension in a small window. In other words, we predetermined the temporal dependence by introducing a parameter *ActiveTime*. The interpretation of *ActiveTime* was the time a fire can stay smouldering but undetectable by satellite before flaring up again.

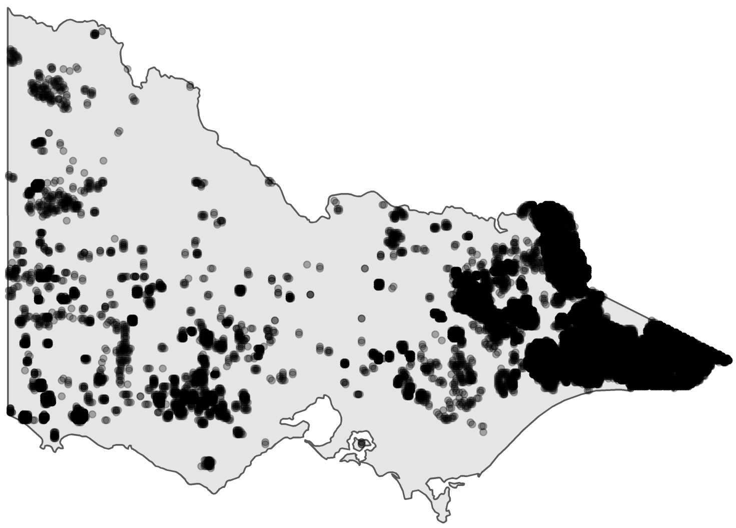


Figure 1: A map shown the distribution of hotspot locations in Victoria during 2019-2020 Australia bushfire season.

Given a certain value of *ActiveTime* and the length of the time frame T, the algorithm would define several segments,

$$S_t = [max(1, t - ActiveTime), t], \quad t = 1, 2, ..., T$$

For example, if the dataset contains 48 hours of hotspots and the $ACtiveTime = 24\ hours$, there will be 48 segments $\boldsymbol{S}_1, \boldsymbol{S}_2, .., \boldsymbol{S}_{48}$, where

$$\begin{aligned}
\boldsymbol{S}_1 &= [1, 1] \\
\boldsymbol{S}_2 &= [1, 2] \\
&... \\
\boldsymbol{S}_{25} &= [1, 25] \\
\boldsymbol{S}_{26} &= [2, 26] \\
&... \\
\boldsymbol{S}_{47} &= [23, 47] \\
\boldsymbol{S}_{48} &= [24, 48]
\end{aligned}$$

2. Cluster hotspots in each segment individually

Since the algorithm broke the temporal dimension in **step 1**, the local clustering for each segment only needed to address the hotspots spatially by introducing the second parameters *AdjDist*. *AdjDist* represented the potential potential distance a fire could spread with respect to the temporal resolution of the data. For example, let $AdjDist = 3000m$ and the temporal resolution of the data is 10-minute, then the potential speed of the bushfire is $3000m/10\ min = 18km/h$.

Given a certain value of *AdjDist* and the segment $\boldsymbol{S}_t$, the algorithm would

(a) Append a randomly selected hotspot h_i to a empty list $\boldsymbol{L}$, where h_i is the ith hotspot in the segment $\boldsymbol{S}_t$, and let pointer $\boldsymbol{P}$ point to the first element of the list $\boldsymbol{L}$.

(b) Visit every h_i where $h_i \notin \boldsymbol{L}$. If $geodesic(h_i, \boldsymbol{P}) \leq AdjDist$, append h_i to list $\boldsymbol{L}$.

(c) Move pointer $\boldsymbol{P}$ to the next element of the list $\boldsymbol{L}$.

(d) Repeat (b) and (c) till the pointer $\boldsymbol{P}$ reach to the end of the list $\boldsymbol{L}$.

(e) For hotspots $h_i \in \boldsymbol{L}$, assign a new membership to them. Repeat (a) to (d) for unassigned hotspots in segment $\boldsymbol{S}_t$.

3. Updated the clustering results recursively

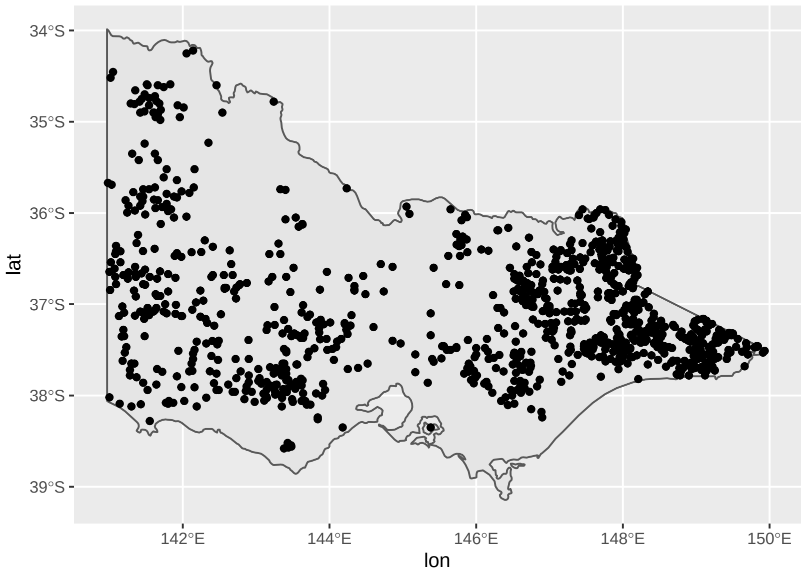
Effects of parameter choices

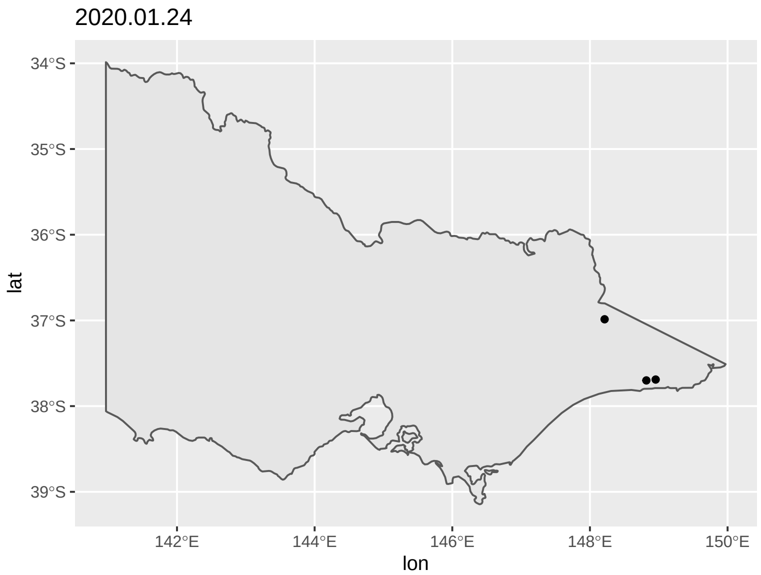
There are two parameters that can be tuned in this algorithm. They are `adj_dist`, which is the density distance and `active_time`, which is the .

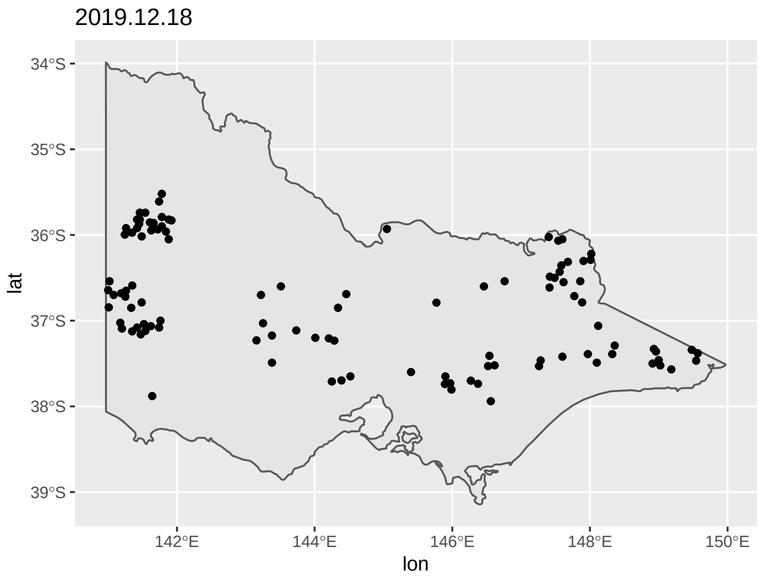
Application

Determining the ignition point and time for individual fires

Show ignition points for a particularly heavy day and another for a particularly light day







Tracking fire movement

Display showing how a fire moves over time, maybe two or more fires

Allocating resources for future fire prevention

Merging data with camp sites, CFA, roads, . . .

Summary

Acknowledgements

- The code and files to reproduce this work are at XXX
- Data on hotspots can be downloaded from XXX

Bibliography

P-Tree System. JAXA Himawari Monitor - User's Guide, 2020. URL https://www.eorc.jaxa.jp/ptree/userguide.html. [p1]

Weihao Li
Monash University
line 1
line 2

wlii0039@student.monash.edu

Emily Dodwell
AT&T
line 1
line 2

emily@research.att.com

Dianne Cook
Monash University
line 1
line 2

dicook@monash.edu