Using remote sensing data to understand the ignition of 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 sources 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 Bureau of Meteorology of Australia 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 Background and literature review

<left blank>

1.3 Research aim

The aim of this research is to use remote sensing data and apply machine learning framework to understand the sources of bushfire ignition and predict fire risk. There are 5 sub-objectives:

1. Data integration of remote sensing data, weather data, map data and vegetation data.
2. Develop customized clustering algorithm to classify fire clusters.
3. Exploratory data analysis of fire clusters.
4. Examine source of bushfire ignitions using weather records, road map and recreation site locations.
5. Explore bushfire risk models based on weather condition, geometry information and fire history.
6. Develop an interactive shiny app to present research outcome.

1.4 Research plan

1.4.1 Data

In order to collect necessary datasets, two methods will be performed, including retrieve files from servers and crawl data from websites. For reproducibility purpose, only public data will be used in this research. Information on the datasets can be found in Table 1.1.

1.4.2 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

Table 1.1: *Data information*

Data set name	Spatial Resolution	Temporal resolution	Time
Hotspots data - JAXA's Himawari-8 satellite	$0.02° \approx 2km$	Per 10 minutes	2015-2020
Weather data - Bureau of Meteorology of Australia		Daily	2019-2020
Map - OpenStreetMap			2020
Fuel layer - Australian Bureau of Agriculture and Resource Economics and Sciences			2018

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.

1.5 Preliminary Results

1.6 irrelevant

1.7 code

Included in this template is a file called `sales.csv`. This contains quarterly data on Sales and Advertising budget for a small company over the period 1981–2005. It also contains the GDP (gross domestic product) over the same period. All series have been adjusted for inflation. We can load in this data set using the following command:

```
sales <- ts(read.csv("data/sales.csv")[,-1], start=1981, frequency=4)
```

Any data you use in your thesis can go into the data directory. The data should be in exactly the format you obtained it. Do no editing or manipulation of the data outside of R. Any data munging should be scripted in R and form part of your thesis files (possibly hidden in the output).

1.8 Figures

Figure 1.1 shows time plots of the data we just loaded. Notice how figure captions and references work. Chunk names can be used as figure labels with `fig:` prefixed. Never manually type figure numbers, as they can change when you add or delete figures. This way, the figure numbering is always correct.

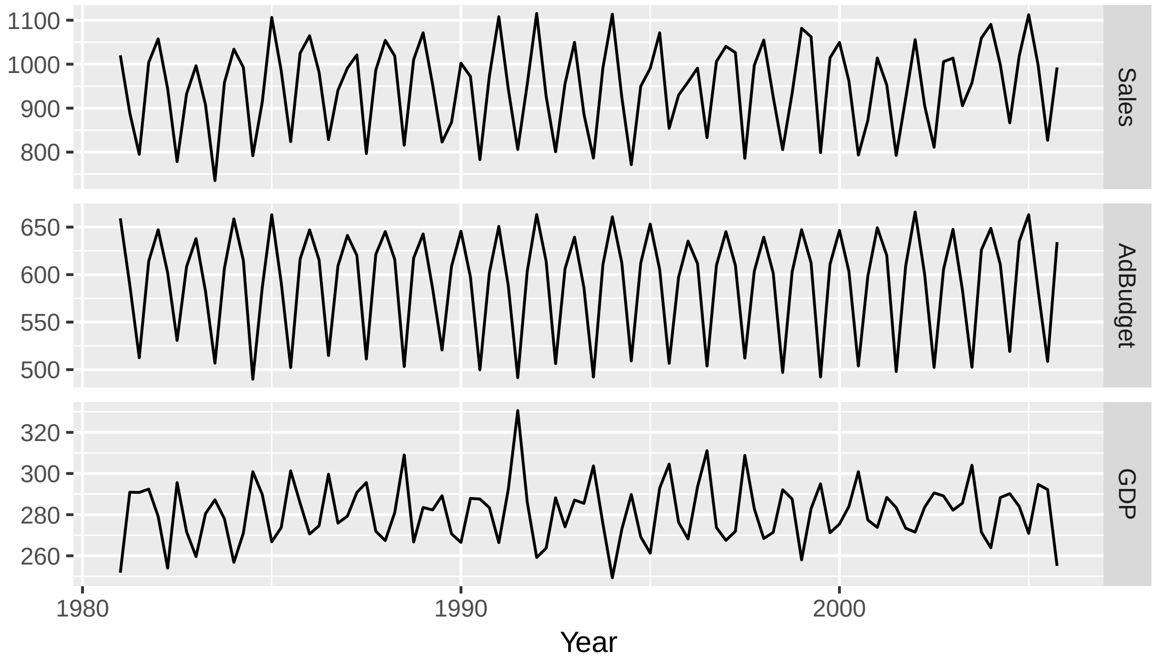

Figure 1.1: *Quarterly sales, advertising and GDP data.*

1.9 Results from analyses

We can fit a dynamic regression model to the sales data.

If y_t denotes the sales in quarter t, x_t denotes the corresponding advertising budget and z_t denotes the GDP, then the resulting model is:

$$y_t - y_{t-4} = \beta(x_t - x_{t-4}) + \gamma(z_t - z_{t-4}) + \theta_1\varepsilon_{t-1} + \Theta_1\varepsilon_{t-4} + \varepsilon_t \tag{1.1}$$

where $\beta = 2.28$, $\gamma = 0.97$, $\theta_1 = NA$, and $\Theta_1 = -0.90$.

1.10 Tables

Let's assume future advertising spend and GDP are at the current levels. Then forecasts for the next year are given in Table 1.2.

Again, notice the use of labels and references to automatically generate table numbers. In this case, we need to generate the label ourselves.

Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
1000.2	947.7	1052.7	919.9	1080.5
1013.1	959.3	1066.8	930.9	1095.3
1076.7	1022.9	1130.6	994.4	1159.0
1003.5	949.7	1057.4	921.2	1085.8

Table 1.2: *Forecasts for the next year assuming Advertising budget and GDP are unchanged.*

The `knitLatex` package is useful for generating tables from R output. Other packages can do similar things including the `kable` function in `knitr` which is somewhat simpler but you have less control over the result. If you use `knitLatex` to generate tables, don't forget to include `results="asis"` in the chunk settings.

Chapter 2

Exponential Smoothing

2.1 Organizing your ideas

Imagine you are writing for your fellow Honours students. Topics that are well-known to them do not have to be included here. But things that they may not know about should be included. Resist the temptation to discuss everything you've read in the last year.

Do not organize your chapter around the papers you have read with one section per paper. Instead, you should organize your chapters around themes, and within each theme provide a story explaining the development of ideas. It is usually helpful to plan out a table of contents first with major section headings.

When you are discussing results from several papers or books, you will need to adopt a common notation to ensure your chapter makes sense. Do not use different notation for the same thing.

2.2 Citations

All citations should be done using markdown notation as shown below. This way, your bibliography will be compiled automatically and correctly.

Exponential smoothing was originally developed in the late 1950s (Brown, 1959, 1963; Holt, 1957; Winters, 1960). Because of their computational simplicity and interpretability, they became widely used in practice.

Empirical studies by Makridakis and Hibon (1979) and Makridakis et al. (1982) found little difference in forecast accuracy between exponential smoothing and ARIMA models. This made the family of exponential smoothing procedures an attractive proposition (see Chatfield et al., 2001).

The methods were less popular in academic circles until Ord, Koehler, and Snyder (1997) introduced a state space formulation of some of the methods, which was extended in Hyndman et al. (2002) to cover the full range of exponential smoothing methods.

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.

Bibliography

Brown, RG (1959). *Statistical forecasting for inventory control*. McGraw-Hill, New York.

Brown, RG (1963). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, New Jersey: Prentice Hall.

Chatfield, C, AB Koehler, JK Ord, and RD Snyder (2001). A new look at models for exponential smoothing. *The Statistician* **50**, 147–159.

Holt, CE (1957). *Forecasting trends and seasonals by exponentially weighted averages*. O.N.R. Memorandum 52/1957. Carnegie Institute of Technology.

Hyndman, RJ, AB Koehler, RD Snyder, and S Grose (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* **18**(3), 439–454.

Makridakis, S, A Anderson, R Carbone, R Fildes, M Hibon, RLJ Newton, E Parzen, and R Winkler (1982). The accuracy of extrapolation (time series) methods: results of a forecasting competition. *Journal of Forecasting* **1**, 111–153.

Makridakis, S and M Hibon (1979). Accuracy of forecasting: an empirical investigation (with discussion). *Journal of Royal Statistical Society (A)* **142**, 97–145.

Ord, JK, AB Koehler, and RD Snyder (1997). Estimation and prediction for a class of dynamic nonlinear statistical models. *Journal of American Statistical Association* **92**, 1621–1629.

Winters, PR (1960). Forecasting sales by exponentially weighted moving averages. *Management Science* **6**, 324–342.