Learning Forest Wildfire Dynamics from Satellite Images Using Reinforcement Learning

Sriram Ganapathi Subramanian
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Canada
s2ganapa@uwaterloo.ca

Mark Crowley
Department of Electrical and Computer Engineering
University of Waterloo
Waterloo, Canada
mcrowley@uwaterloo.ca

Abstract

Forest wildfires are a perennial problem in many parts of the world requiring high financial and social costs to measure, predict and control. One key challenge is modelling the dynamics of fire spread itself which usually relies on computationally expensive, hand crafted physics-based models. The question we ask is can we learn a dynamics model by treating wildfire as an agent spreading across a landscape in response to neighbourhood environmental and landscape parameters. The problem is modelled as a Markov Decision Process where the fire is the agent at any cell in the landscape deciding whether to spread the fire into neighbouring cells. The set of suitable actions the fire can take at any point of time includes moving North, South, East, West or to not spread at this time. Rewards are provided at the end of the epoch based on correctly classifying cells which are on fire or not. We apply two algorithms to this problem: Value Iteration and Asynchronous Advantage Actor-Critic (A3C)[8] which is a recent direct policy search approach that utilizes Deep Learning to perform simultaneous state-space approximation and policy representation.

The data for the start state and rewards come solely from satellite images of a region in northern Alberta, Canada which is prone to large wildfires. Two events are used, the Fort McMurray fire of 2016 which led to the unprecedented evacuation of almost 90,000 people for several months [10] and the Richardson fire of 2011 which was smaller but just as dramatic at the time. Experiments are carried out training a wildfire spread policy for one region on multiple time frames as well as testing the transferability of that policy to the data from a second region. The results obtained indicate that it is useful to think of fire as a learning agent to understand its characteristics in a spatial environment.

Keywords: Value iteration; Asynchronous Advantage Actor-Critic; Satellite images; Forest Wildfires; Spatiotemporal Model; Spatial information; Deep Reinforcement Learning.

1 Introduction

The risk, costs and impacts of forest wildfires are a perennial and unavoidable concern in many parts of the world. A number of factors contribute to the increasing importance and difficulty of this domain in future years including climate change, growing urban sprawl into areas of high wildfire risk and past fire management practices which focused on immediate suppression at the cost of increased future fire risk [9].

There are a wide range of challenging decision and optimization problems in the area of Forest Fire Management [5], many of which would benefit from more responsive fire behaviour models which could be run quickly and updated easily based on new data. For example, the simplest decision problem could be whether to allow a fire to burn or not, which requires a great deal of expensive simulation to evaluate fully [3].

The field of forest and wildland fire behaviour modelling is a very active one involving trials in real forest conditions, lab experiments, physics-based fire modelling and more [1]. A challenge in the field is balancing the detail and computational cost of a physics-based model with the high level of accuracy needed for important decision making and planning problems. In this work we are contributing to that effort by investigating how RL could be used to model wildfire spread as an agent on the landscape taking spatial actions in reaction to its environment.

Previous work on the application of reinforcement learning to spatial processes [2] has investigated modelling the state variable to represent the land cover and environmental characteristics and the action variable to represent interaction between characteristics of the neighbourhood geographic space. We apply these general principles in this work to wildfire prediction.

This work has similarities to the use of intelligent systems for predicting burned areas as suggested in [7]. However, their work focused on burned area alone where we look at the more specific problem of prediction of actual fire spread location over the short term. Their work also relies on high resolution field data which may not be possible to obtain for all fires. Our study uses easily accessible satellite data from government agencies. A summary [4] of the merits of various software packages and tools (such as GeoMISC, GeoKNN, GeoMLP etc) in relation to Machine Learning on Geospatial Data was useful in developing this work.

2 Design and Experiments

The problem is formulated a Markov Decision Process (MDP) $< S, A, P, R >$ where the state is each location on the landscape and the 'agent' taking actions is a fire spreading across the landscape. A state $s \in S$ corresponds to the state of a cell in the landscape (x, y, t, l, w, d, b) where x, y are the location of the cell, t is the temperature and l is the land cover value of the cell (derived from satellite images), and w and d are wind speed and direction (both of which are the assumed same for all cells in an image). The state element b records whether the cell has been burned by fire ($b = 1$) or not ($b = 0$) or whether this is uncertain from the satellite images ($b = 2$). The initial state will come from satellite images and where certain cells are set to have just ignited fire.

The reward function R is formulated in which the cells being damaged by the wild fire were assigned a reward value of one and the cells clearly undamaged by the fire got a value of -1. All the other cells were initialised with a reward value of 0. A reward value is set up for every cell. A certain location of fire is given a value of 1 and a certain location having no fire is given a reward value of 0. The action is then, at each instant, for the fire to 'choose' one of the North, South, East or West or stay actions to take from the current cell. The goal of the agent is to choose its actions so as to maximize the rewards gained within a time constraint. Wind Speeds and Wind Direction are taken into consideration before an action is taken. These values are assumed to be a constant for the entire region of consideration. The dynamics for any particular cell $P(s'|s, a)$ is a function from one state s to the next s' given an action a. In this formulation the dynamics are actually quite simple, other aspect of the cell state do not change quickly, or all, so the action of spreading a fire into a cell directly alter the chance that neighbouring cell will move to a burn state. After this the fire can take any one of possible actions from that cell. The fire is constrained to not cross the boundary of the domain of study. Figure 1 shows a representation of the model. The light/green cells are unburned and dark/red are those cells affected by fire. The cirlce represents an ignition point while at each action choice the current cell, decides to spread to a neighbouring cell or not.

2.1 Using Satellite Data

The study area under consideration is a northern part of the province Alberta, Canada (53.9333N, 116.5765 W). The satellite images are downloaded from the USGS Earth Explorer data portal (https://earthexplorer.usgs.gov/) for the Alberta. Figure 2(b) shows an example where the smoke over burning areas can be seen. The time of image capture is during the occurrence of Fort McMurray (May 1 2016 to August 1 2016) and Richardson fires (May 15, 2011 September 2011). All the images were corrected for missing values and outliers. Additional pre-processing steps were carried out as outlined in [6]. For the region under consideration the original 7611×8401 pixel images are divided into distinct cells of 100×100 pixels which are used to represent a single cell state.

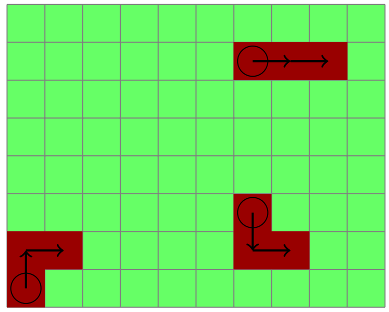
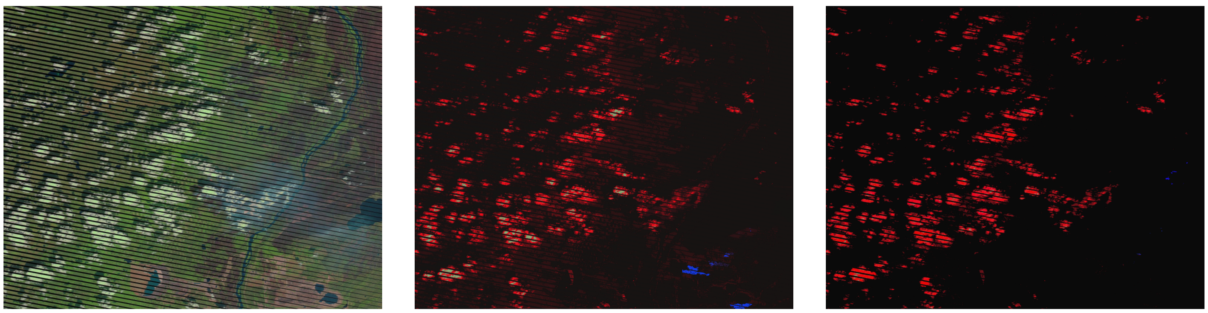


Figure 1: A schematic of the wildfire motion domain.



(a) The satellite image of june 8, 2011 in the region of Alberta

(b) Output obtained from value iteration

(c) Output obtained from deep learning

Figure 2: The wildfire spread domain as a grid model representing input satellite data and output wildfire prediction

The land cover value is obtained by processing the satellite images in a open source geo-processing software (Multispec). The affected state of a cell is also obtained from the satellite image. Temperature is obtained from processing thermal images from the same data source. Wind speed and wind direction are obtained from NOAA satellite products for the region of study. They are assumed to be fairly constant over the region of the study area. (http://www.ospo.noaa.gov/Products/atmosphere/wind.html)

2.2 Solution 1 : Value Iteration Method

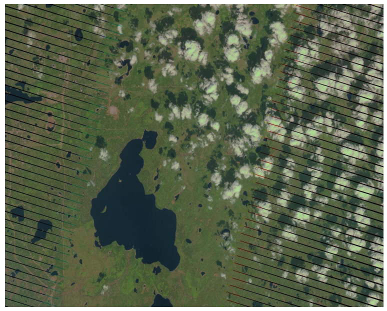
For the first attempt at this problem we use asynchronous value iteration where the optimal value of the state $V^*(s)$ under the greedy fire spread policy is give by the Bellman Equation:

$$V^*(s) = R(s) + max_a \gamma \sum_{s'} P(s'|s,a) V^* s' \tag{1}$$

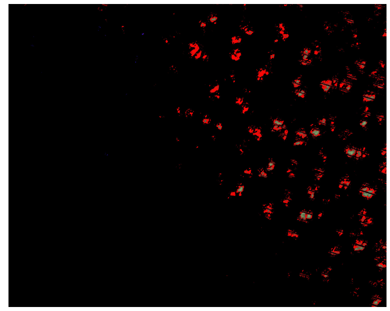
where s' is the successor state and γ denotes the discount factor which we set to 0.99.

2.3 Solution 2: Deep Learning Method

The second algorithm we tried on this domain was the **Asynchronous Advantage Actor Critic (A3C)** algorithm [8]. In this algorithm a global network is defined in addition to multiple worker agents with individual sets of parameters. The input to this global network was formalised as a grid of 100 X 100 cells with each cell having state values which is an average of the state values of several pixels from the previous setup derived from the same set of satellite images. Here the reinforcement learning problem is formulated as the fire being the agent and deciding to spread from one cell to another. In $A3C$ we have the advantage of defining multiple worker agents. Each separate instance of a fire (unconnected to other fires) in a neighbourhood is given its environment as an input and the fire is defined as an individual worker. Each worker (fire) would then update the global environment and we have plotted the result obtained. This method was also run on the same computer as the value iteration and it took an average of 40 minutes to run. This is much better compared to the Value iteration method. The deep network used is based on that given in [8] which uses an input layers of 100×100 pixel windows from the satellite image for the start date. Then there is an convolution layer of 16 88 filters with a stride of 4 follow by a rectifier nonlinear unit (ReLU) layer. The second hidden layer convolves 32 4×4 filters usin



(a) The satellite image of june 14, 2016 in the region of Alberta



(b) Output obtained from value iteration

Figure 3: Comparing the output from value iteration to the actual scenario

Method	(A) Richardson Fire(1st Month)	(B) Richardson Fire(2nd Month)	(C) Fort McMurray Fire
Value Iteration	79.6%	77.5%	70%
Deep Learning	87.3%	83.2%	81.8%

Table 1: Accuracy of each algorithm on the three different test scenarios.

a stride of 2 also followed by a ReLU layer. The final fully-connected layer uses 256 ReLUs which is output to a layer of thresholded functions for the output of each of the five possible actions.

2.4 Experimental Setup

In the first two experiments, (A) and (B), satellite images from May, June and July 2011 of the Richardson wildfire were chosen. The May images were taken right from before the ignition of the fire and used as the start state for RL. The satellite images in June (A) and July (B) were used to determine the target burned area and so inform the reward function. During training the predicted final burned areas at the end of the period were used to compute the reward for the training epoch. The values obtained as a result were used to obtain a value function that predicts the growth and characteristics of fire after fixed time durations. For testing the output from the value iteration was compared to an actual satellite image from halfway through the training period for the same.

In the third experiment (c), we applied the policy learned on the Richardson fire to satellite images of a different fire in Fort McMurray five years leater. This is the same region of Northern Alberta but exactly the same area so general properties should be transferable. The ignition location and initial spread are taken from the satellite images of the month of May As before, we tested the predicted burned areas a month after ignition agains the true satellite and data.

3 Results

All experiments were run on an intel core i7-6700 cpu with 32Gb ram. The value iteration approach took an average of 2.5 hours to converge while the A3C solution took aroud 4 hours. Figure 2(c) shows the results obtained from the experiment (A). The red in the output image corresponds to the pixels which were on fire and were classified correctly as fire by the algorithm (ie. true positives). The blue pixels represent flase positives, those there were classified as burning but were not yet at that point. White pixels represent false negatives where the policy predicted no fire but fire was indeed present. It can be seen visually that the value iteration algorithm was able to capture most of the large fire zones of the Richardson fire on June 8. Still some of the smaller ignitions were not captured by the value iteration.

Figure 3 shows the results from the experiment (B). The day of comparison was on June 14, 2016. Again we can see that the value iteration algorithm did fairly well capturing the major fire spots.

Figure **??** shows the results obtained from the Deep Learning method. The deep learning method performs slightly better than a simple value iteration as some small fire instances are also being identified by $A3C$ in addition to the ones identified by value iteration technique.

The ROC curve in Figure 4 shows the effect of varying the threshold used for classifying if a pixel is affected by fire or not based on the value. The value was chosen to be 0.83 from the curve as it has high degree of true positives with only a small number of false positives.

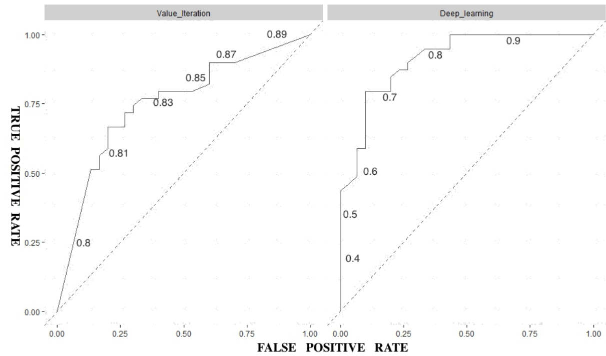


Figure 4: ROC curve for analysing sensitivity of policy to classification threshold parameter.

Table 1 presents the accuracy (as a percentage) obtained in both the methods across different experiments. Columns (A) and (B) test the policy's prediction on the same fire event where the test point is 1 month or 2 months after ignition where the reward comes from a point a month later. Column (C) is testing the policy's prediction on a completely different fire five years later, in the same region, predicting a fire one month after ignition.

4 Conclusion

In this work, we presented a novel approach for utilizing RL to augment physics-based forest wildfire simulations by learning the spread dynamics directly from readily available satellite image data. We demonstrated that a simple agent based model of fire can obtain reasonable fire spread predictions using value iteration and that this can be significantly improved using a Deep RL representation. In future work we plan to look at enriching the model by including more land characteristics such as moisture, slope and directional aspect as state variables in individual cells. We will also perform a wider comparison against different algorithms such as those in [7]. We plan to investigate improvements to the structure of the Deep Neural Network policy representation tailored to this kind of spatially spreading problem.

References

[1] Mark A. Finney, Jack D. Cohen, Sara S. McAllister, and W. Matt Jolly. On the need for a theory of wildland fire spread. *International Journal of Wildland Fire*, 22(1):25–36, 2013.

[2] Garcia F Forsell N and Sabbadin R. Reinforcement learning for spatial processes. *ResearchGate*, 2009.

[3] Rachel M. Houtman, Claire A. Montgomery, Aaron R. Gagnon, David E. Calkin, Thomas G. Dietterich, Sean McGregor, and Mark Crowley. Allowing a wildfire to burn: Estimating the effect on future fire suppression costs. *International Journal of Wildland Fire*, 22(7):871–882, 2013.

[4] Pozdnoukhov A Kanevski M and Timonin V. Machine learning algorithms for geospatial data applications and software tools. *iEMSs*, 2008.

[5] David L Martell. A Review of Recent Forest and Wildland Fire Management Decision Support Systems Research. *Current Forestry Reports*, 1(2):128–137, 2015.

[6] Anya M. Reading Matthew J. Cracknell n. Geological mapping using remote sensing data: A comparison of five machine learning algorithms,their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers & Geosciences*, 63, 2014.

[7] Leonardo Vanneschi Mauro Castelli and Ale Popovi. Predicting burned areas of forest fires - an artificial intelligence based approach. *Fire Ecology*, 11, 2015.

[8] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Alex Graves, Ioannis Antonoglou, Daan Wierstra, and Martin A. Riedmiller. Playing atari with deep reinforcement learning. *CoRR*, abs/1312.5602, 2013.

[9] Claire Montgomery. Fire: An Agent and a Consequence of Land Use Change. In Joshua M. Duke and JunJie Wu, editors, *The Oxford Handbook of Land Economics*, chapter 13, pages 281–301. Oxford University Press, 2014.

[10] Andrea Woo and Carrie Tait. Up to 90,000 evacuated from Fort McMurray; some ordered to move again as fire spreads, may 2016.