

Xây dựng mô hình Machine Learning dự báo cháy rừng ở các tỉnh Tây Nguyên dựa vào dữ liệu lịch sử thời tiết.

Nguyễn Đại Kỳ
19521731

Văn Viết Hiếu Anh
19521225

Lê Văn Phước
19522054

Ngày 1 tháng 8 năm 2021

Môn học: CS114 - Máy học
Giảng viên hướng dẫn: Lê Đình Duy
Phạm Nguyễn Trường An

Mục lục

1	Abstract	2
2	Introduction	2
3	Dataset	2
3.1	Source of Data	2
3.1.1	Weather Data	2
3.1.2	Fire Data	2
3.2	Imputation of Data	3
3.3	Creation of Dataset	3
4	Methods	3
4.1	Convolutional neural network	3
4.2	Fully-connected neural network	3
5	Conclusion	3

1 Abstract

Bài viết là về quá trình thực nghiệm nghiên cứu các model Machine Learning với mục đích chọn ra mô hình tối ưu để dự đoán mức độ cháy rừng dựa vào dữ liệu thời tiết trong lịch sử của từng địa phương. Với mục đích hỗ trợ trong việc dự đoán để phục vụ trong công tác phòng chống cháy rừng ở nước ta. Vì bài viết là ghi chép của quá trình thực nghiệm nên sẽ có nhiều phương pháp được đưa ra sử dụng.

2 Introduction

Ở nước ta có 3 thảm họa lớn nhất, gây thiệt hại lớn hàng năm về cả người và của. Cùng với lũ lụt và hạn hán, cháy rừng là một thảm họa gây thiệt hại không chỉ về kinh tế mà còn cả con người và hệ sinh thái. Theo thống kê của Cục Kiểm lâm từ năm 1992 đến 2006, trung bình mỗi năm xảy ra 1254 vụ cháy rừng gây thiệt hại khoảng 6646 ha rừng, trong đó có 2854 ha là rừng tự nhiên và 3791 ha là rừng trồng. Bên cạnh việc nâng cao năng lực phòng cháy chữa cháy rừng (PCCCR) cho lực lượng kiểm lâm như đầu tư trang thiết bị, cơ sở vật chất, xây dựng cơ chế điều hành phối hợp và tuyên truyền nâng cao nhận thức trách nhiệm của chủ rừng và người dân, công tác cảnh báo nguy cơ cháy rừng cũng như tổ chức phát hiện sớm và thông báo kịp thời điểm cháy rừng là rất cần thiết.

Từ đầu năm 2007, Cục Kiểm lâm (Bộ Nông nghiệp và Phát triển Nông thôn) đã lắp đặt và vận hành trạm thu ảnh viễn thám MODIS tại Hà Nội với mục đích chính là phát hiện sớm các điểm cháy rừng (hotspots) trên toàn lãnh thổ Việt Nam. Hệ thống trạm thu của TeraScan đã tự động thu nhận, xử lý và sao lưu dữ liệu ảnh MODIS hàng ngày từ 2 vệ tinh TERRA và AQUA với mô-đun Vulcan tự động xử lý và tạo ra dữ liệu các điểm cháy sử dụng thuật toán ATBD-MOD14.

Hệ thống này cung cấp dữ liệu về điểm cháy ghi nhận được từ vệ tinh và lưu lại thời gian và tọa độ cháy. Từ khi bắt đầu lắp đặt đến nay hệ thống dữ liệu cháy của cục kiểm lâm được ghi lại được gần 1 triệu điểm cháy. Nhờ lượng dữ liệu này việc xây dựng một hệ thống tự động phân tích mức độ cháy rừng dựa vào các đặc trưng cơ bản của dữ liệu khí tượng thủy văn là hoàn toàn có cơ sở và khả quan.

3 Dataset

Để xây dựng được mô hình, việc đầu tiên sau khi khai thác dữ liệu là phải thực hiện phân tích và làm sạch dữ liệu. Từ đó chọn ra những đặc trưng có ảnh hưởng nhiều đến output của bài toán để tạo ra được dataset phù hợp.

3.1 Source of Data

Cả 3 nguồn dữ liệu gồm 2 nguồn dữ liệu thời tiết và nguồn dữ liệu về điểm cháy đều không có API cung cấp một cách đại chúng cho việc khai thác. Tuy nhiên vì các website này được xây dựng dựa trên cấu trúc Asynchronous (ASP.NET, JQuery hoặc ReactJS) nên sau khi phân tích và ghi lại các request, ta hoàn toàn có thể tìm được các cổng API và phương thức giao tiếp với server. Từ đó dùng vào việc khai thác dữ liệu.

3.1.1 Weather Data

weather.com

worldweatheronline.com

3.1.2 Fire Data

firewatchvn.kiemlam.org.vn

3.2 Imputation of Data

3.3 Creation of Dataset

4 Methods

4.1 Convolutional neural network

4.2 Fully-connected neural network

5 Conclusion

Tài liệu