PREDICTING FOREST FIRES WITH DIFFERENT DATA MINING TECHNIQUES

[1]Madhurima De, [2]Linika Labdhi, [3]Bindu Garg

[1]Student, [2]Student, [3]Associate Professor
[1,2,3]Department of Computer Engineering,
[1,2,3]Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, Maharashtra, India

***Abstract*: Forest fires are one of the most frequently occurring disasters in recent years. The behaviour of forest fire and its severity result from a combination of factors such as available fuels, physical setting, and weather. Analysis of historical meteorological data and national fire records in western North America show the primacy of climate in driving large regional fires via wet periods that create substantial fuels, or drought and warming that extend conducive fire weather.**
The effects of forest fires creates a very lasting impact on the environment as it leads to deforestation and global warming, which is also one of its major cause of occurrence. Forest fires are dealt by collecting the satellite images of forest and if there is any emergency caused by the fires then the authorities are notified to mitigate its effects. In this work, we will be exploring various Data Mining (DM) approaches to predict the burnt area of forest fires. Five different DM techniques, e.g. Support Vector Machines (SVM) and Random Forests, and four distinct feature selection setups (using spatial, temporal, FWI components and weather attributes), were tested on recent real-world data.

Index Terms: Forest Fires, Support Vector Machine, Supervised Learning Algorithms, Data Mining Application.

I. INTRODUCTION

Forests fires are as old as the forests themselves. They pose a threat not only to the forest wealth but also to the entire regime to fauna and flora disturbing the bio-diversity and the ecology and environment of a region. During summer, when there is no rain for months, the forests become littered with dry senescent leaves and twinges, which could burst into flames ignited by the slightest spark. The Himalayan forests, particularly, Garhwal Himalayas have been burning regularly during the last few summers, with colossal loss of vegetation cover of that region.

1.1 Causes of Forest Fire

1. Natural causes - Many forest fires start from natural causes such as lightning which set trees on fire. However, rain extinguishes such fires without causing much damage. High atmospheric temperatures and dryness (low humidity) offer favourable circumstance for a fire to start.

2. Man-made causes - Fire is caused when a source of fire like naked flame, cigarette or bidi, electric spark or any source of ignition comes into contact with inflammable material.

1.2 Types of Forest Fire

1. Surface Fire - A forest fire which primarily burn as a surface fire, spreads along the ground as the surface litter (senescent leaves and twigs and dry grasses etc) on the forest floor and is engulfed by the spreading flames.

2. Underground Fire - The fires of low intensity, consuming the organic matter beneath and the surface litter of forest floor are sub-grouped as underground fire. In most of the dense forests a thick mantle of organic matter is find on top of the mineral soil. This fire spreads in by consuming such materials. The other terminology for this type of fire is Muck fires.

3. Firestorms - Among the forest fires, the fire spreading most rapidly is the firestorm, which is an intense fire over a large area. As the fire burns, heat rises and air rushes in, causing the fire to grow. More air makes the fire spin violently like a storm.

Data mining is the process of finding anomalies, patterns and correlations within large data sets to predict outcomes. Various techniques can be used in this information to increase revenues, cut costs, improve customer relationships, and reduce risks and more. Its foundation comprises three intertwined scientific disciplines: statistics (the numeric study of data relationships), artificial intelligence (human-like intelligence displayed by software and/or machines) and machine learning (algorithms that can learn from data to make predictions). Data mining – as well as predictive modelling and real-time analytics – are used in oil and gas operations. Indeed, the fire detection domain uses several DM techniques.

II. FOREST FIRE DATA

Attribute Information: The forest fire dataset is a multivariate dataset which is a data set consisting of two or more than two variables is referred to as multivariate dataset. It has 13 attributes having 517 instances. The attributes are explained as follows:

1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9 where the fire occurred.

2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9 where the fire occurred.

3. Month - month: 'Jan' to 'Dec' of the year

4. Day - day: 'mon' to 'sun' of the week

5. FFMC - FFMC index from the FWI system: 18.7 to 96.20. It is the Fine Fuel Moisture Code which denotes the moisture content surface litter and influences ignition and fire spread.

6. DMC - DMC index from the FWI system: 1.1 to 291.3. It is the Duff Moisture Code which represents the moisture content of shallow and deep organic layers, which affect fire intensity.

7. DC - DC index from the FWI system: 7.9 to 860.6. It is the Drought Code which represents the moisture content of shallow and deep organic layers, which affect fire intensity.

8. ISI - ISI index from the FWI system: 0.0 to 56.10. It is the Initial Spread Index which determines and correlates it with fire velocity spread.

9. Temp - temperature in Celsius degrees: 2.2 to 33.30 of the area.

10. RH - relative humidity in %: 15.0 to 100 in the air.

11. Wind - wind speed in km/h: 0.40 to 9.40 at the time of fire.

12. Rain - outside rain in mm/m2: 0.0 to 6.4.

13. Area - The forest area (in ha): 0.00 to 1090.84 that burned during the forest fire.

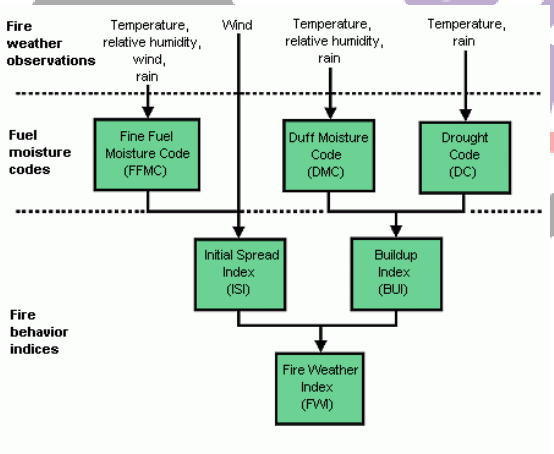
Fire weather observations
Temperature, relative humidity, wind, rain
Wind
Temperature, relative humidity, rain
Temperature, rain
Fuel moisture codes
Fine Fuel Moisture Code (FFMC)
Duff Moisture Code (DMC)
Drought Code (DC)
Initial Spread Index (ISI)
Buildup Index (BUI)
Fire behavior indices
Fire Weather Index (FWI)

Fig. 2.1: Fire weather index structure

This park contains a high flora and fauna diversity. From January 2000 to December 2003, the data used in the experiments was collected and it was built using two sources. The inspector that was responsible for the Montesinho fire occurrences collected the first database. At a daily basis, several features were registered, such as the time, date, spatial location within a 9×9 grid (*x* and *y* axis), the type of vegetation involved, the six components of the FWI system and the total burned area every time a forest fire occurred.

III. MACHINE LEARNING TECHNIQUES

3.1 Logistic Regression

Logistic regression is a statistical method for predicting binary classes. The outcome or target variable is binary in nature. For example, it can be used for cancer detection problems. It computes the probability of an event occurrence. It is a form of binomial regression. The relationship between the dependent variable and the independent variable helps it to predict the target variable. To determine their probability and map them to some discrete values, the logistic regression uses sigmoid function. The sigmoid function is as follows:-

$$\phi(z) = \frac{1}{1+e^{-z}}$$

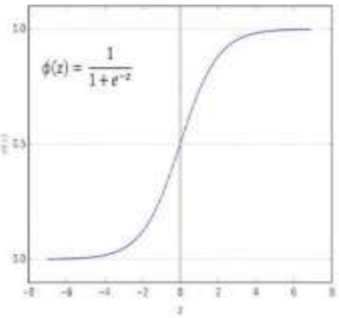

Fig. 3.1.1: Graphical representation of sigmoid function

The value of is within the range of [0, 1]. Hence it can be said that represents the probability of the occurrence of ‘z’.

3.2 Support Vector Machine (SVM)

SVM or Support Vector Machine is a linear model for classification and regression problems. It can solve linear and non-linear problems and work well for many practical problems. The data is plotted in N-Dimensional space where the coordinates in the plot corresponds to its value .The algorithm then creates a line or a hyperplane which separates the data into classes. According to the SVM algorithm we find the points closest to the line from both the classes. These points are called support vectors. Now, we compute the distance between the line and the support vectors. This distance is called the margin. Our goal is to maximize the margin. The hyperplane for which the margin is maximum is the optimal hyperplane.

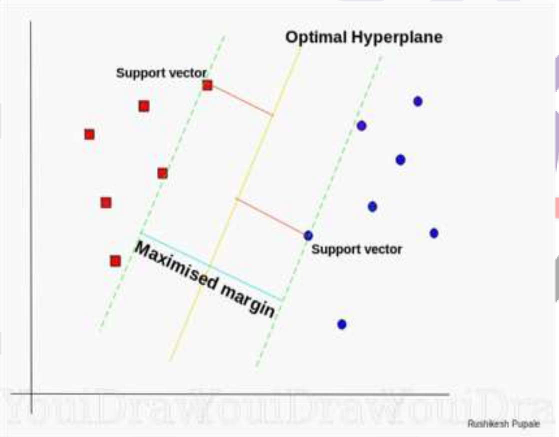

Fig. 3.2.1: Optimal Hyperplane using the SVM algorithm

3.3 Decision Trees

A decision tree is a flowchart-like structure in which each internal node represents a test on a feature (e.g. whether a coin flip comes up heads or tails) , each leaf node represents a class label (decision taken after computing all features) and branches represent conjunctions of features that lead to those class labels. The paths from root to leaf represent classification rules. Decision trees are constructed via an algorithmic approach that identifies ways to split a data set based on different conditions. It is one of the most widely used and practical methods for supervised learning. Decision Trees are a non-parametric supervised learning method used for both classification and regression task.

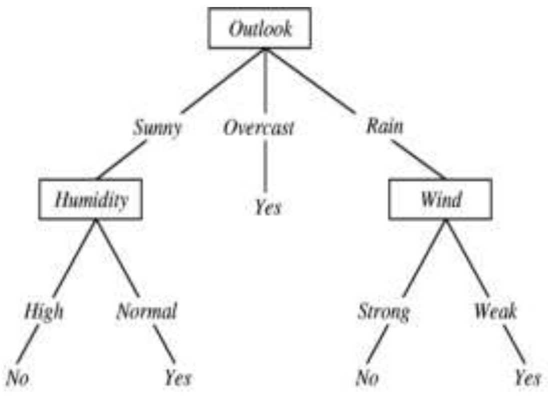

Fig. 3.3.1: Example of a decision tree for rain forecasting

3.4 Random Forest

Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random forest spits out a class prediction and the class with the most votes becomes our model's prediction. The advantage of random forest over decision tree is that they correct the over-fitting nature of the decision trees.

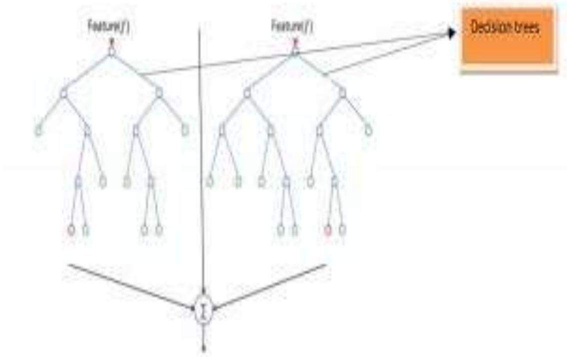

Fig. 3.4.1: Overview of the two decision trees that constitute a Random forest

IV. CONCLUSION AND FUTURE WORK

Forest fires cause a significant environmental damage while threatening human lives. A substantial effort was made in the last two decades to build automatic detection tools that could assist Fire Management Systems (FMS). The three major trends are the use of satellite data, infrared/smoke scanners and local sensors (e.g. meteorological). In this work, we propose a Data Mining (DM) approach that uses the forest fire dataset from the UCI machine learning repository. The database included spatial, temporal, components from the Canadian Fire Weather Index (FWI) and four weather conditions. Five different DM algorithms, including Support Vector Machines (SVM), and four feature selections (using distinct combinations of spatial, temporal, FWI elements and meteorological variables) were tested. This problem was modeled as a regression task, where the aim was to predict the occurance of future forest fires. However, this work opens room for the development of automatic tools for fire management support. Since the FMS system is widely used around the world, further research is need to confirm if direct weather conditions are preferable than accumulated values, as suggested by this study.

REFERENCES

1. Soo Chin Liew, "Satellite detection of forest fires and burn scars", 2001.
2. P.W. Adriaans, D. Zantinge, Data Mining, Addison-Wesley, 1996.
3.Tukey. J. W. (1962)," The future of data analysis", Ann. Statist.33, 1-67.
4. S. Agarwal, "Data Mining: Data Mining Concepts and Techniques," 2013 International Conference on Machine Intelligence and Research Advancement, Katra, 2013, pp. 203-207.
5. P. Cortez and A. Morais, "A Data Mining Approach to Predict Forest Fires using Meteorological Data.", In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence, Proceedings of the 13th EPIA 2007 – Portuguese Conference on Artificial Intelligence, December, Guimarães, Portugal, pp. 512-523, 2007, APPIA, ISBN-13 978-989-95618-0-9.
6. G. E. Sakr, I. H. Elhajj, G. Mitri and U. C. Wejinya, "Artificial intelligence for forest fire prediction," 2010 IEEE/ASME International Conference on Advanced Intelligent Mechatronics, Montreal, ON, 2010, pp. 1311-1316.
7. Bühlmann, Peter, "Bagging, Boosting and Ensemble Methods", Handbook of Computational Statistics. 10.1007/978-3-642-21551-3_33, 2012.

8. T. M. Khoshgoftaar, J. Van Hulse and A. Napolitano, "Comparing Boosting and Bagging Techniques With Noisy and Imbalanced Data," in IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 41, no. 3, pp. 552-568, May 2011.

9. L. Breiman, "Bagging predictors", Machine Learning, vol. 24, no. 2, pp. 123-140, 1996.

10. Goutte, Cyril & Gaussier, Eric, “A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation”, Lecture Notes in Computer Science. 3408. 345-359. 10.1007/978-3-540-31865-1_25, 2012

11. D. Zhang, J. Wang, X. Zhao and X. Wang, “A Bayesian Hierarchical Model For Comparing Average F1 Scores,” 2015 IEEE International Conference on Data Mining Atlantic City, NJ, 2015, pp. 589-598. Doi:10.1109/ICDM.2015.44