Proceedings for the 5th International Fire Behavior and Fuels Conference
April 11-15, 2016, Portland, Oregon, USA
Published by the International Association of Wildland Fire, Missoula, Montana, USA

A NOVEL FOREST FIRE PREDICTION TOOL UTILIZING FIRE WEATHER AND MACHINE LEARNING METHODS

Leo Deng*
Portland State University, Portland, Oregon, USA, leodeng5@gmail.com
Marek Perkowski
Portland State University, Portland, Oregon, USA, h8mp@pdx.edu
John Saltenberger
Northwest Interagency Coordination Center, Portland, Oregon, USA,
jsaltenb@blm.gov

INTRODUCTION

Wildfires are an essential and natural part of ecosystems that help restore them. Many species of plants rely on such fires to cleanse the environment for better regeneration and growth. Additionally, they have the potential to support the growth of thriving native species while eliminating invasive species. The result is most often a newly created ground perfect for the future plants that still live. It is a catalyst for promoting biological diversity and keeping ecosystems healthy. Despite their integral contributions to the environment, wildfires also threaten people and property, especially when they are unable to be contained (Fried *et al.* 2004). We have been able to keep the advantages of wildfires much greater than its disadvantages, but due to climate change, fire seasons are becoming uncontrollable, and there is a need for a more efficient management system. Better fire forecasting becomes more and more crucial to retain a balanced relationship between wildfires, humans, and the environment. The purpose of this work is to aid these management agencies on planning and strategy to efficiently manage wildfires and being prepared to contain hazardous, unwanted fires.

There has been a spark of interest in the use of data mining in the field of wildfire management. Many techniques have been developed in attempt to increase fire awareness (Lee *et al.* 2002; Cruz *et al.* 2005; Alonso-Betanzos *et al.* 2003; Vega-Garcia *et al.* 1996; Hsu *et al.* 2002; Stojanova *et al* 2006; Sitanggang *et al.* 2013). In Portugal, an attempt was made to predict the number of acres future wildfires would burn using machine learning methods combined with regression techniques, based on weather attributes and the Fire Weather Index (FWI) for wildfires. The method is unique in that it takes advantage of easily obtainable fire and weather information from existing local sensors. However, this model used a continuous method which resulted in relatively poor prediction accuracy (Cortez *et al.* 2007). In addition, the database used was limited to the Fire Weather Index and basic meteorological variables, and a limited range of time. The machine learning methods used included Decision Trees, Random Forests, SVM, Neural Networks and Naïve Bayes. The best configuration developed utilized the SVM method. In contrast to these previous works, this work introduces improved novel Machine Learning (ML) methods, where the emphasis is on predicting future forest fire intensities with the use of real-time and easily-obtained meteorological data from existing local sensors. This work is demonstrated on a complete database with an optimized training set of historical weather attributes, and optimized machine learning methods. The result of the prediction will be discretized in the terms of the magnitude of the fire, as needed by fire management agencies.

*1924 NW Norfolk Ct. Portland, OR 97229
leodeng5@gmail.com,

MATERIALS AND METHODS

Fire weather database

The data collected from local sensors by the Northwest Interagency Coordination Center was used for testing and developing the tools in this work. The data includes the monthly averages of ten different relevant fire weather attributes including National Fire Danger Rating System (NFDRS) indices.

The attributes are: 100 hour dead fuel moisture (F100), 1000 hour dead fuel moisture (F1000), Live Fuel Index 1-100 (LFI), Sum of Rain Duration in hours (RainDur), Sum of Rain Amount in inches (RainAmt), Average Temperature in °F (Temp), Maximum Temperature in °F (Max Temp), Minimum Relative Humidity % (MinRH), Wind Speed in mph (Wind), and Duff Moisture Code (DuffMC)

Each of these attributes of the fires in the database is multiple-valued, and this data was integrated from the 12 predictive service areas in Oregon and Washington States over a time period of 32 years. Each of the 1443 instances includes the number of acres burnt by the fire.

Machine learning methods

Six different machine learning methods were selected and used in the Orange Machine Learning software suite on the fire weather data. Three are based on multiple-valued logic: a Disjunctive Normal Form (DNF) rule based method, Decision Trees, and Naïve Bayes (Barber 2012; Alpaydin 2014). The other three are based on continuous representation: the Support Vector Machine (SVM) along with the radial basis and polynomial kernel functions. As a result of the varying concepts these methods are based on, one cannot be absolutely named better than another; their ability to optimize with precision is dependent on the type of data that is being tested. We intentionally selected different types of methods and different representations, with the intent to find the best method with the data (Zupan *et al.* 2007).

Strategically testing and selecting the attributes

For each of the 7 intensity levels, the machine classifies the fire into one of two categories – less than or greater than a specific number of acres burnt. This is done multiple times as shown in Fig. 1 resulting in the discovery of the final intensity of the fire. The 10 fire weather attributes utilized in this work were tested in different combinations to confirm their relevance to the intensity of a wildfire. The Support Vector Machine and rbf kernel were used on each attribute individually to identify their individual potential for predicting the intensity of a fire. The 10 attributes are ranked from the highest to the lowest average accuracy.

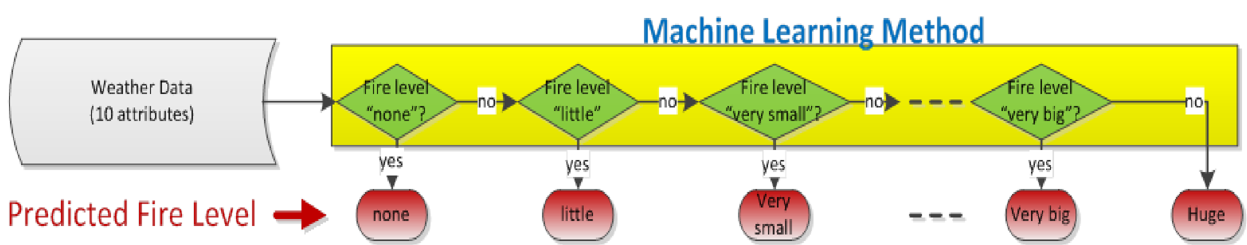

Fig. 1. Final FFPT illustration

Fig. 2 illustrates how the best set of input data is determined using the two support vector machine methods. The "4 best attributes" and "7 best attributes" are determined from the previous step; individual accuracies are used to rank them by importance. The "Medians of attributes" option takes the median value of each of the 10 attributes and uses those single values to train on the data. These four training set options are used to train a linear SVM as well as the radial basis function (RBF) kernel SVM using a 97%/3% training/testing set on randomly selected data. This is repeated 5 times for each method to see which preprocessing option gives the highest accuracy and ultimately the optimal training set that will be used to test each of the six different machine learning methods later.

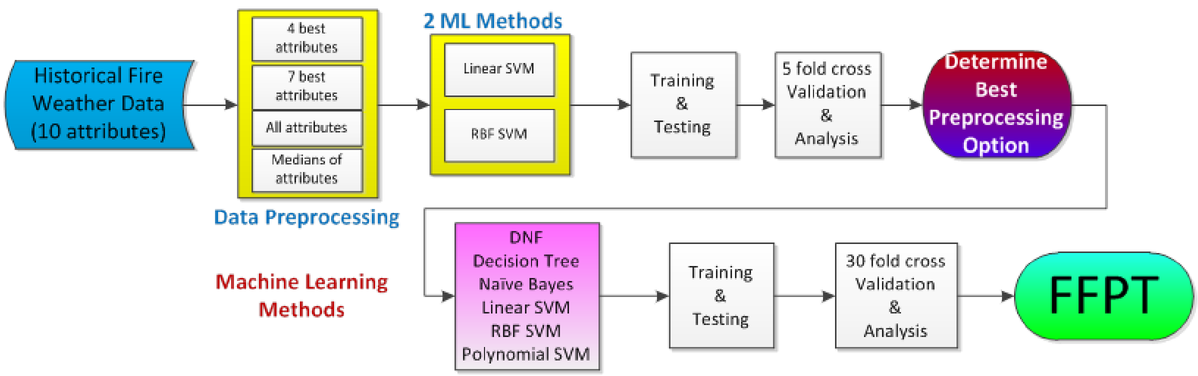

Fig. 2. Determining best set of input data and applying to 6 ML Methods to develop FFPT

Applying six methods on selected attributes to optimize FFPT

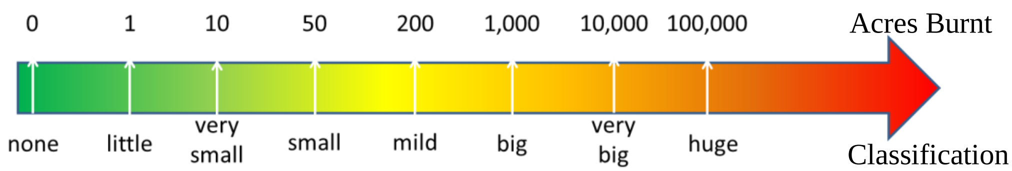

Fig. 3. Classification levels used by FFPT

Once the most optimal training set was determined, this set of data was used to train on all 6 different machine learning methods using the same 97%/3% training/testing method repeating on each intensity/method combination 36 times using randomly selected data each time. The tool was made to classify fires into one of 6 specific intensity levels as shown in Fig. 3 where intensity levels 0 and 1 – none and very small – were combined together in the testing. For each of the 1443 instances used in this work, the machine must perform as many as 6 different separations between the two adjacent intensity levels, varying depending on its specific fire intensity level. For example, for the "huge" fire, all six separations must be done, as the tool will keep asking if the fire is smaller than a certain number of acres until it reaches the final stage of classification. However, for fire level "none", only the first separation is necessary, as the tool simply asks if there is a fire or not, and if not, then the tool reaches the conclusion "none". This process is illustrated in Fig. 1. Six different machine learning methods and six different intensity separations result in a total of 36 accuracies. These values are the result of the mean of the accuracies from the 30 random trials in previous step. Finally, each of the six machine learning methods was given an overall average accuracy.

RESULTS AND DISCUSSION

Fig. 5-7 show the accuracy results of each of the 10 attributes tested individually with the support vector machine, the 6 ML methods accuracy for each of the 7 intensity levels, and the final master average accuracies of each version of the FFPT respectively.

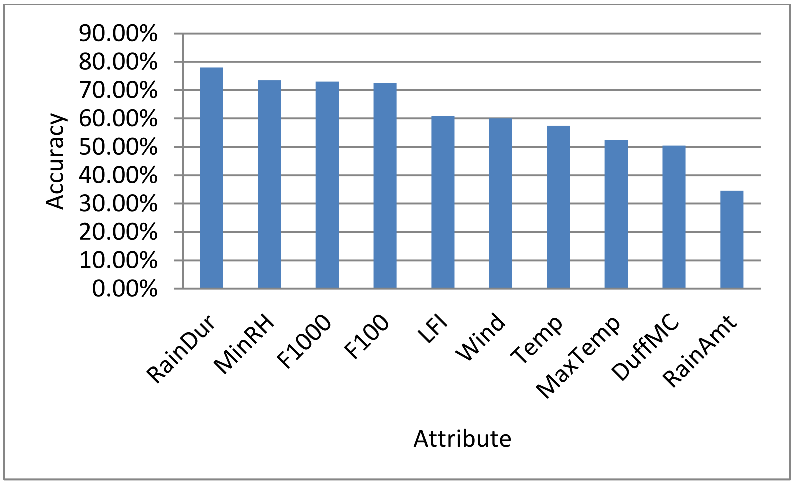

Fig. 5. Accuracy of each individual attribute using SVM

The results in Fig. 5 show each of the 10 attributes' individual accuracies when tested with a 97%/3% training/testing set. Clearly the top 4 attributes are Rain Duration, Minimum Relative Humidity, F1000, and F100. The top 7 attributes additionally include the LFI, Wind, and Temperature. These two sets of attributes, along with the set of all attributes and the set of all attributes' median values, are tested with 2 machine learning methods in the next step to determine which of these sets of attributes is the optimal training set.

Using 7 attributes with the rbf Kernel SVM gave a much higher accuracy than when using 4 attributes. The least optimal training set was the one that was using median values. Rbf Kernel SVM is almost always better than the Linear SVM. These results determine that in all future steps, all attributes – as opposed to a select few – are used to train the 6 machine learning methods, as they all have a significant role to play in the intensity of a fire.

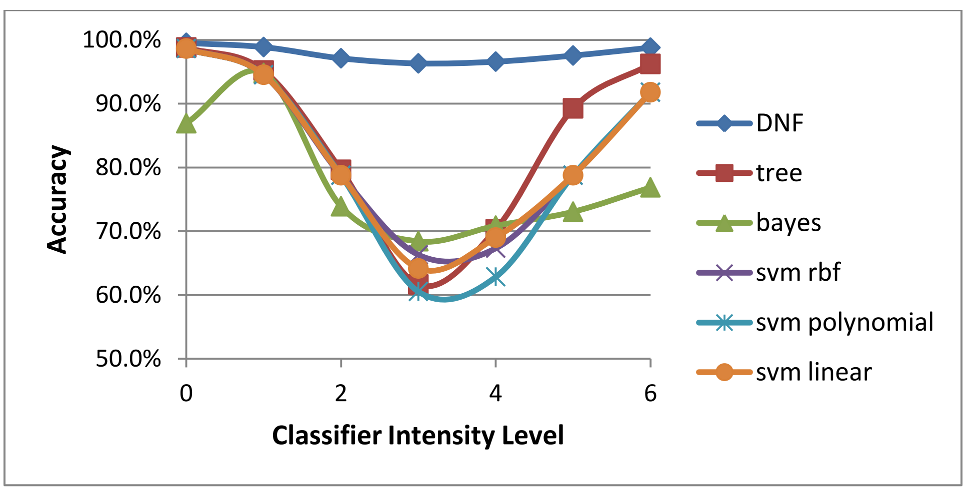

Fig. 6. Accuracy of each machine learning method for predicting each of the 7 intensity levels of wildfires

The trend seen in Fig. 6 for all 6 methods shows that this tool is most accurate at predicting very small fires or very huge fires, and the accuracy drops by some margin as the intensity level nears the center of the pool of data. This can be explained, because for classifying mild fires, the machine learning process has to deal with an equal amount of data on both sides of the separation. Looking at whether a fire is huge or not, on the other hand, is much easier with almost all of the data being less intense than "huge" and only a small portion of months with extreme weather conditions that resulted in the burning of more than 100000 acres of land.

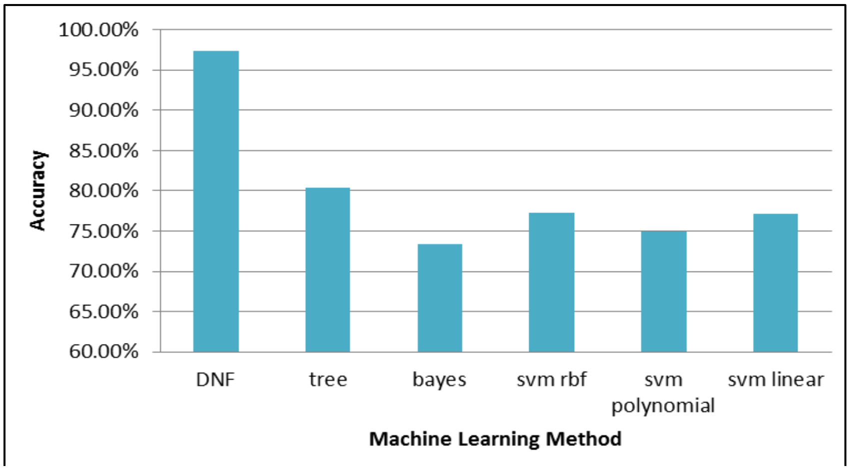

Fig. 7. Average accuracy for each machine learning method

Fig. 7 gives the final average accuracies of each machine learning method when tested on all attributes with a 97%/3% training/testing set. The DNF rule based method gives the highest average accuracy at 97.8%, and a maximum accuracy of 98.8% when predicting huge fires. With so much data from the past 32 years – 1443 months of fires – the DNF method shows the best ability to optimize the large data set with precision. Its lowest accuracy when classifying fires into very small vs. mild fires stayed very high at 96.3% unlike the other 5 methods whose accuracies all dropped significantly by at least 25%.

In this work a novel tool for forecasting wildfires was developed, providing a specific intensity level for a given fire based on the amount of land it would burn. Forecasting of wildfire intensity levels is dependent on the accuracy of weather attribute forecasts. An important factor to consider is that despite the rain duration's high impact on the intensity of a fire, methods to predict this variable are much less accurate than others such as rain amount. The importance of this prediction tool lies with the wildfire management agencies' need to increase awareness of burning wildfires in order to make an educated decision as to which events require the most or least attention.

Conclusion

Our results support the expectation that the newly developed tool will perform with highly accurate information, ultimately benefitting fire managers in their preparations, resource allocation, and minimizing additional assistance for unexpected intense wildfires. A novel Forest Fire Prediction Tool (FFPT) utilizing a disjunctive normal form (DNF) based method was developed and used for wildfire prediction for the Pacific Northwest United States. This method was the best chosen out of six different machine learning methods all tested with random selected data from the historical fire weather database of the past 32 years. In contrast to previous methods of fast fire detection, this tool makes it possible to enable proactive resource management for firefighting response teams, promoting the conservation of valuable resources. This will inevitably result in a significantly higher control and balance of large fires as well as lowered costs for land restoration.

Results showed that of the six machine learning methods used, given a 97%/3% training/testing combination, DNF rule based method, used with all 10 fire weather attributes gave the highest average accuracy of 97.8%, the highest accuracy reported for forest fire intensity prediction in literature. The prediction accuracy is higher for small and large scale fires. This tool will result in a much greater awareness for wildfire management, allowing response teams to conduct accurate planning and decision making.

REFERENCES

Alpaydin, Ethem. Introduction To Machine Learning. Cambridge, MA: MIT Press, 2014.

Alonso-Betanzos, Amparo, et al. "An intelligent system for forest fire risk prediction and fire fighting management in Galicia." *Expert Systems with Applications* 25.4 (2003): 545-554.

Barber David. *Bayesian Reasoning and Machine Learning*. Cambridge University Press, 2012. 203-213.

Cortez Paulo and Anibal de Jesus Raimundo Morais. "A Data Mining Approach to Predict Forest Fires Using Meteorological Data." (2007).

Cruz, Miguel G., Martin E. Alexander, and Ronald H. Wakimoto. "Development and testing of models for predicting crown fire rate of spread in conifer forest stands." *Canadian Journal of Forest Research* 35.7 (2005): 1626-1639.

Fried, Jeremy S., Margaret S. Torn, and Evan Mills. "The impact of climate change on wildfire severity: a regional forecast for northern California." *Climatic change* 64.1-2 (2004): 169-191.

Hsu, Wynne, Mong Li Lee, and Ji Zhang. "Image mining: Trends and developments." *Journal of intelligent information systems* 19.1 (2002): 7-23.

Lee, B. S., et al. "Information systems in support of wildland fire management decision making in Canada." *Computers and Electronics in Agriculture* 37.1 (2002): 185-198.

Lee, B. S., P. M. Woodard, and S. J. Titus. "Applying neural network technology to human-caused wildfire occurrence prediction." *AI applications* (1996).

Sitanggang, Imas Sukaesih, et al. "Predictive Models for Hotspots Occurrence using Decision Tree Algorithms and Logistic Regression." Journal of Applied Sciences, 13.2, 252-261.

Stojanova Daniela, et al. "Learning to Predict Forest Fires with Different Data Mining Techniques." *Conference on Data Mining and Data Warehouses (SiKDD 2006), Ljubljana, Slovenia*. 2006.

Zupan, Blaz, et al. "Orange and decisions-at-hand: Bridging predictive data mining and decision support." *Proceedings of the ECML/PKDD Workshop on Integrating Aspects of Data Mining, Decision Support and Meta-Learning, Freiburg, Germany*. 2001.