Paper 9361-2016

Predicting forest fire occurrence and incremental fire rate using SAS® 9.4 and SAS® Enterprise Miner ™ 14.1

Quyen Nguyen, Dr. Goutam Chakraborty, Oklahoma State University

ABSTRACT

Fast detection of forest fires is a great concern among environmental experts and national park managers because forest fires create economic and ecological damages and endanger human lives. For effective fire control and resource preparation, it is necessary to predict fire occurrences in advance and estimate the possible losses caused by fires. For this purpose, real-time sensor data of weather conditions and fire occurrence are highly recommended to use in order to support the predicting mechanism. The objective of this study is to use SAS® 9.4 and SAS® Enterprise Miner™ 14.1 to predict the probability of fires and to figure out special weather conditions resulting in incremental burned areas in Montesinho Park forest (Portugal). The data set was obtained from the Center for Machine Learning and Intelligent Systems at University of California, Irvine and contains 517 observations and 13 variables from January 2000 to December 2003. Support Vector Machine analyses with variable selection were performed on this data set for fire occurrence prediction with a validation accuracy of approximately 60%. The study also incorporates Incremental Response technique and Hypothesis testing to estimate the increased probability of fire as well as the extra burned area under various conditions. For example, when there is no rain, a 27% higher chance of fires and 4.8 hectares of extra burned area are recorded, compared to when there is rain.

INTRODUCTION

The impacts of forest fires in forest preservation and in human lives have raised a great concern about forest fire controlling. Despite the increasing expenses to control this disaster, millions of forest hectares over the world are destroyed every year. In Portugal, a highly affected country by forest fire, 2.7 million hectares of forest have been destroyed from 1980 to 2005 [2]. Especially in fire season of 2003 and 2005, affected area in Portugal accounted for 4.6% and 3.1% of the territory respectively with about 20 human deaths each year. For forest fire controlling purpose, many approaches have been considered. These approaches are grouped into three major categories: satellite-based, infrared/smoke scanners, and local sensors (meteorological sensors) [1]. In Portugal forest situation, meteorological is the most appropriate approach for fire detection and tracking with a low cost and real-time data since automatic meteorological stations are often available in Portugal (with 162 official stations) [2]. Meteorological sensors track information such as temperature and relative humidity to estimate the possibility of fires [1]. Of the indexes measured in the meteorological sensor system, the Canadian forest Fire Weather Indexes (FWI), which were correlated with fire activity in Portugal and other southern Europe countries, are also incorporated [4].

The main areas of research in this paper are as follows:

- Use traditional hypothesis testing to compare fire rate and burned area in different weather conditions
- Verify the hypothesis testing with Incremental Response models using four treatment variables interchangeably, including: x-y coordinate, rain, wind and temperature

LITERATURE REVIEW

Forest fire prediction is not a new topic explored in data mining and analytics. Cortez and Morais (2007) suggested multiple data mining models to predict the possible burned area by forest fires in a national park in Portugal. When a fire happened, the burned area figure was collected with other meteorological data of weather conditions correlating to that specific fire. In the study by Cortez and Morais, models used include linear regression, decision tree, random forest, support vector machine and so on. And the target variable is numeric denoting the burned areas caused by fires [2]. The study did a good job in predicting how big a fire could be in a specific combination of weather conditions. These model results were good enough in predicting forest fires for fire control purposes. For this reason, forest burned areas and fire rate predictions will not be covered in this paper. This paper, which uses the same data set as that of Cortez and Morais, will focus mainly on incremental response analysis.

Incremental response has been also applied for a few years in business and marketing analysis since introduced. Incremental response models that use a pair of training data sets (treatment and control) to measure the incremental effectiveness of a direct marketing program [3]. Applying the same concept, the incremental seriousness of fire occurrences is measured in this study. The incremental seriousness of fire when a treatment event happens or in case of a special event is measured in two approaches: by fire probability and by burned area by fires.

DATA

The data set in this study was obtained from Center for Machine Learning and Intelligent Systems at University of California, Irvine. The original data set has 517 observations, and 13 variables, collected from January 2000 to December 2003. Target variable is "area", which is the burned area of the forest (in hectares). Input variables include: indexes from the Canadian danger rating system of Fire Weather Indexes (FWI), month of year, day of week of the observed records, temperature, relative humidity, and outside rain. Following are variables used for modeling:

Variable Name	Role	Level	Description	Range of Values
X	Input	Nominal	x-axis spatial coordinate within the park map	1 to 9
Y	Input	Nominal	y-axis spatial coordinate within the park map	2 to 9
month	Input	Nominal	Month of the year	Jan to Dec
day	Input	Nominal	Day of the week	Mon to Sun
FFMC	Input	Interval	Fine Fuel Moisture Code index (FWI system)	18.7 to 96.2
DMC	Input	Interval	Duff Moisture Code index (FWI system)	1.1 to 291.3
DC	Input	Interval	Drought Code index (FWI system)	7.9 to 860.6
ISI	Input	Interval	Initial Spread Index (FWI system)	0.0 to 56.10
temp	Input	Interval	Temperature in Celsius degrees	2.2 to 33.30
RH	Input	Interval	Relative humidity in %	15.0 to 100
wind	Input	Interval	Wind speed in km/h	0.40 to 9.40
rain	Input	Interval	Outside rain in mm/m2	0.0 to 6.4
area	Target	Interval	The burned area of the forest (in ha)	0.00 to 1090.84

Table 1. Original variables - name, role, measurement level, description and range of values

DATA PREPARATION

The original data set has 13 attributes and 517 observations describing forest fire occurrence associating with different weather conditions. For modeling purpose, the data set is subjected to cleaning methods including filtering extreme values and transforming variables (binary variable creation). The final data set has 506 observations (containing only observations with burned area less than 100 hectares, or 98% of the total observations) and 18 attributes, of which 5 new binary variables are created in data preparation process. For implementing incremental response analysis using SAS® Enterprise Miner ™, a new binary target is created to indicate whether or not a fire happens. Four treatment variables are also created and used interchangeably to predict the increased fire rate and burned area regarding each of these four risk indicator variables (Figure 1).

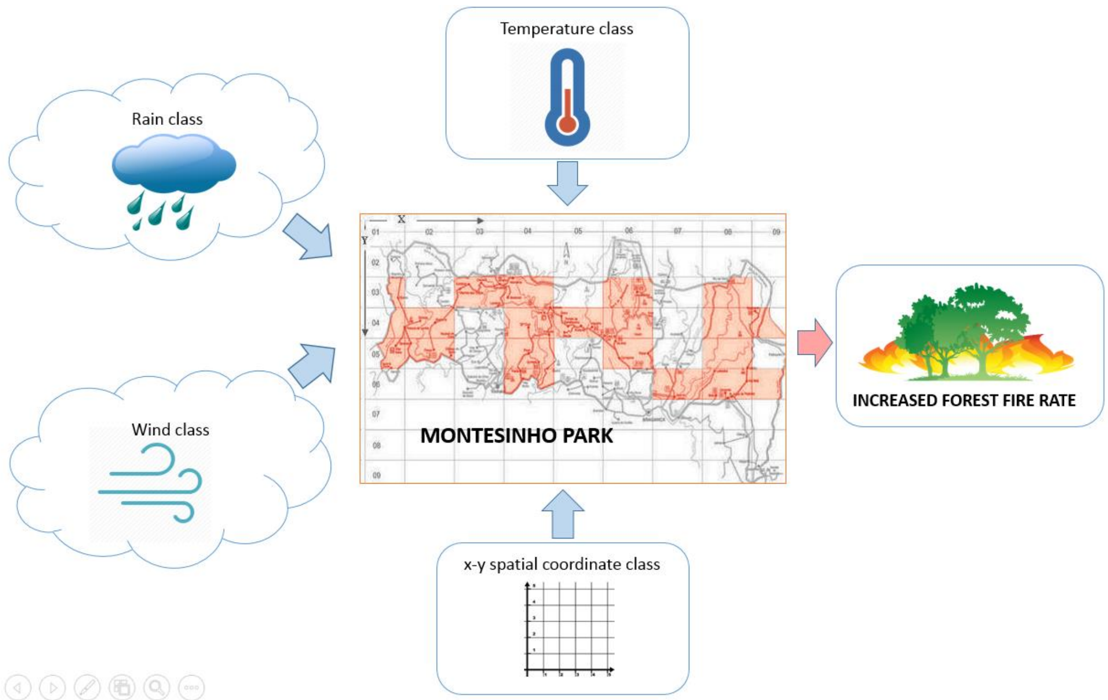

Figure 1. Created treatment variables for Incremental Response Analyses

Details of the five new created variables as follows:

Variable Name	Role	Level	Description
fire	Target	Binary	Fire occurrence: 1 = Yes; 0 = No
xy_class	Treatment	Binary	x-y spatial coordinate in Montesinho Park with high risk of fire: 1 = high risk; 0 = not high risk
temp_class	Treatment	Binary	Temperature classification: 1 = high temperature (temperature ≥ 15ºC); 0 = normal temperature.
rain_class	Treatment	Binary	Rain classification: 1 = no rain; 0 = rain
wind_class	Treatment	Binary	Wind classification: 1 = high wind speed (wind speed ≥ 8km/h); 0 = low wind speed.

Table 2. Created variables' name, role, measurement level, description and range of values

Red parts in the map bellow (Figure 2) are regions with high risk of fire in Montesinho Park (xy_class = 1) which are classified using the combination of a high fire rate and/or a large burned area by forest fires from January 2000 to December 2003.

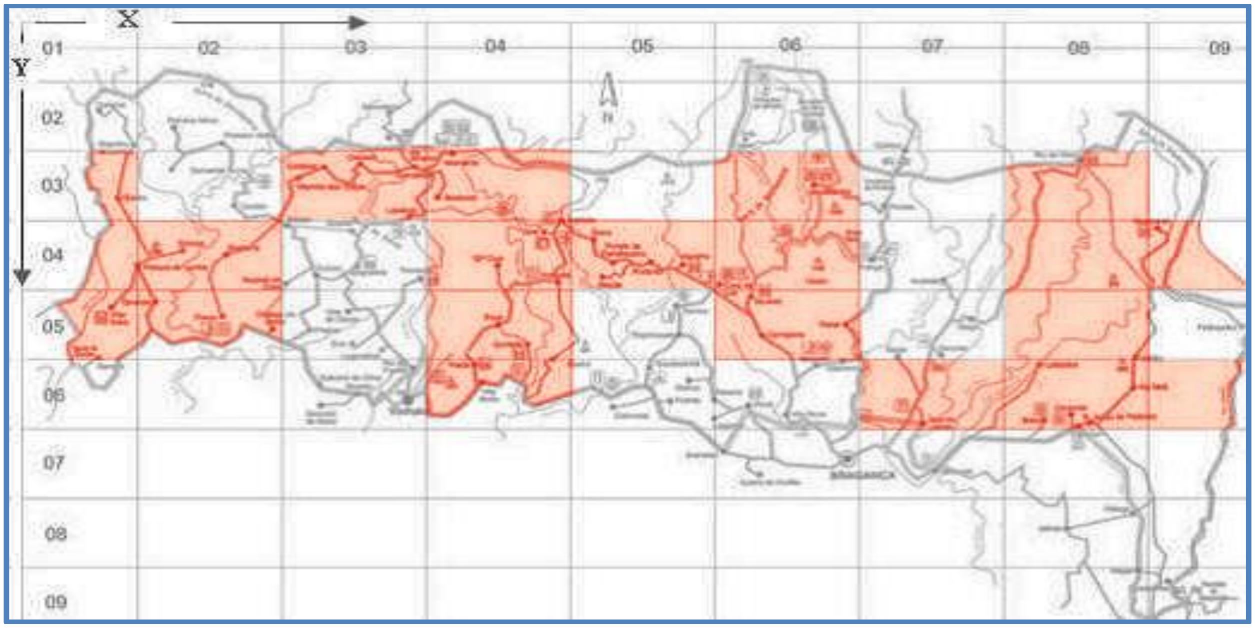

Figure 2. Montesinho Park (Portugal) map with details of high risk x-y spatial coordinate

Role	Measurement Level	Frequency Count
INPUT	INTERVAL	8
INPUT	NOMINAL	4
TARGET	BINARY	1
TARGET	INTERVAL	1
TREATMENT	BINARY	4

Figure 3. Final data set variable summary

Variable	Label	Missing	N	Minimum	Maximum	Mean	Standard Deviation	Skewness	Kurtosis
DC	Drought Code index from the FWI system: 7.9 to 860.6	0	506	7.9	860.60	545.696	250.032	-1.0739	-0.314
DMC	Duff Moisture Code index from the FWI system: 1.1 to 291.3	0	506	1.1	291.30	110.146	64.170	0.5629	0.227
FFMC	Fine Fuel Moisture Code index from the FWI system: 18.7 to 96.20	0	506	18.7	96.20	90.630	5.566	-6.5461	66.218
ISI	Initial Spread Index from the FWI system: 0.0 to 56.10	0	506	0.0	56.10	9.040	4.592	2.5274	21.245
RH	% relative humidity	0	506	15.0	100.00	44.421	16.415	0.8493	0.393
area	burned area of the forest (in ha)	0	506	0.0	95.18	6.308	13.822	3.4961	13.881
rain	outside rain in mm/m2	0	506	0.0	6.40	0.022	0.299	19.6043	412.328
temp	temperature in Celsius degrees	0	506	2.2	33.30	18.824	5.832	-0.3156	0.116
wind	wind speed in km/h	0	506	0.4	9.40	4.027	1.800	0.5655	0.039

Figure 4. Interval Variable Summary Statistics

Variable	Label	Type	Number of Levels	Missing
X	x-axis spatial coordinate within the Montesinho park map: 1 to 9	N	9	0
Y	y-axis spatial coordinate within the Montesinho park map: 2 to 9	N	6	0
day		C	7	0
fire		N	2	0
month		C	12	0

Figure 5. Class variable summary statistics

MODEL BUILDING

The data set is partitioned into train (60%) and validation (40%) subsets to provide more precise assessments. In this study, there is no need to set prior probabilities since the percentages of target events (fire =1 and fire =0) are pretty equal in both data sets.

In this paper, classification models to predict whether or not a fire will occur are not the main focus in model building. Even though some classification models were built, these models did not work out as anticipated. The best classification model selected is Support Vector Machine with Polynomial kernel having training accuracy of 90%, and validation accuracy of about 60%. The high fluctuation between training and validation accuracy perhaps due to the small sample size of only about 500 observations. Selected variables in SVM models include: x and y coordinate, month of year, drought code index, relative humidity percentage, fine fuel moisture code index and outside rain. Of these variables, rain and x-y coordinates are also used to create treatment variables for incremental response analyses.

HYPOTHESIS TESTING USING PROC GLM

Before building any incremental response models using four treatment variables, traditional hypothesis testing is applied to explore the impact of each treatment on the target variables. PROC GLM is used to compare the means of fire probability between control group (treatment event = 0) and treatment group (treatment event = 1). Under PROC GLM, Levene's Test for Homogeneity of variance assumption, Welch and Dunnett post-test are also incorporated.

PROC GLM code for hypothesis testing with the predictor variable of x-y coordinate as follows:

```
PROC GLM DATA=create.fire_final;
    CLASS xy_class;
    MODEL Fire Area= xy_class;
    MEANS xy_class / WELCH   HOVTEST=LEVENE ;
    MEANS xy_class / DUNNETT ALPHA=0.05;
```

Using x-y coordinate as predictor variable in PROC GLM (Figure 6), we can see that both fire occurrence and burned area models are significant at 95% confident interval. In comparison with the control group, the treatment group shows a 21.7% of higher fire rate and an extra burned area of 4.56 hectares.

x-y class treatment

Dependent Variable: fire

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	5.3140069	5.3140069	22.11	<.0001
Error	504	121.1148468	0.2403072		
Corrected Total	505	126.4288538			

Comparisons significant at the 0.05 level are indicated by ***.

xy_class Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 0	0.21665	0.12614	0.30717	***

Dependent Variable: area burned area of the forest (in ha)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	2356.90877	2356.90877	12.62	0.0004
Error	504	94119.12233	186.74429		
Corrected Total	505	96476.03109			

Comparisons significant at the 0.05 level are indicated by ***.

xy_class Comparison	Difference Between Means	Simultaneous 95% Confidence Limits		
1 - 0	4.563	2.039	7.086	***

Figure 6. Hypothesis test result for x-y class treatment

Similarly using predictor variable of rain class with the same code as above, figure 7 reveals that fire rate model is significant at 85% confidence interval and burned area model is significant at 65% confidence interval. At a common confidence interval of 65%, rain class event is responsible for 26.6% of increased fire rate and 4.75 hectares of increased burned area.

rain class treatment

Dependent Variable: fire

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.5573678	0.5573678	2.23	0.1358
Error	504	125.8714859	0.2497450		
Corrected Total	505	126.4288538			

Comparisons significant at the 0.35 level are indicated by ***.

rain_class Comparison	Difference Between Means	Simultaneous 65% Confidence Limits		
1 - 0	0.26606	0.09946	0.43267	***

Dependent Variable: area burned area of the forest (in ha)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	178.34723	178.34723	0.93	0.3344
Error	504	96297.68386	191.06683		
Corrected Total	505	96476.03109			

Comparisons significant at the 0.35 level are indicated by ***.

rain_class Comparison	Difference Between Means	Simultaneous 65% Confidence Limits		
1 - 0	4.7594	0.1511	9.3676	***

Figure 7. Hypothesis test result for rain class treatment

For wind class predictor variable, both fire rate and burned area models are significant at 85% confidence interval (Figure 8). Wind class event is responsible for 33.3% of increased fire rate and 5.3 hectares of increased burned area.

wind class treatment

Dependent Variable: fire

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	1.9288538	1.9288538	7.81	0.0054
Error	504	124.5000000	0.2470238		
Corrected Total	505	126.4288538			

Comparisons significant at the 0.15 level are indicated by ***.

wind_class Comparison	Difference Between Means	Simultaneous 85% Confidence Limits		
1 - 0	0.33333	0.16135	0.50531	***

Dependent Variable: area burned area of the forest (in ha)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	487.33777	487.33777	2.56	0.1103
Error	504	95988.69333	190.45376		
Corrected Total	505	96476.03109			

Comparisons significant at the 0.15 level are indicated by ***.

wind_class Comparison	Difference Between Means	Simultaneous 85% Confidence Limits		
1 - 0	5.2984	0.5230	10.0738	***

Figure 8. Hypothesis test result for wind class treatment

For temperature predictor variable, both fire occurrence and burned area models could only be significant at a very low confidence interval of about 50% (Figure 9). This indicates that temperature classification does not have a significant impact on the fluctuation of fire rate nor burned area.

temperature class treatment

Dependent Variable: fire

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.1680057	0.1680057	0.67	0.4132
Error	504	126.2608480	0.2505176		
Corrected Total	505	126.4288538			

Dependent Variable: area burned area of the forest (in ha)

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	70.42592	70.42592	0.37	0.5443
Error	504	96405.60518	191.28096		
Corrected Total	505	96476.03109			

Figure 9. Hypothesis test result for temperature class treatment

INCREMENTAL RESPONSE MODELS IN SAS® ENTERPRISE MINER ™

The primary purpose of incremental response algorithm in SAS® Enterprise Miner ™ is to find the increased response rate as well as the increased revenue when a promotion (treatment) is applied to a marketing campaign. This paper applies the same technique to the situation of forest fire in which the response rate emphasizes the fire rate and the revenue emphasizes the burned area in hectares.

Incremental response or Net Lift works based on the approach of weight of evidence (WOE) and information value (IV) to avoid a situation called curse of dimensionality and the overfitting of data. Doing so, all levels of the dummy variable are enumerated with the pattern of response variable within each level.

Denote Y = 1 for the situation when the target event occurs (forest fire), and Y = 0 when the target event does not occur (no fire). The predictor variables X is grouped into I mutually exclusive bins (levels). Then the weight of evidence (WOE) for each level of treatment variable is calculated as:

$$WOE = \log \frac{P(X = x_i | \, Y = 1)}{P(X = x_i | \, Y = 0)} \quad for \; i = 1, 2, \dots, I$$

Where P (X|Y) denotes the conditional probability of a certain X when Y event occurs.

Information value of a variable is calculated as:

$$IV = \sum [P(X = x_i | \, Y = 1) - P(X = x_i | \, Y = 0)] \, . WOE_i$$

In assessing the Information Value of a variable, a value less than 0.02 indicates that the variable is not predictive, whereas a value greater than 0.3 shows that the variable has a strong predictive power [3].

For incremental response model with the Treatment group (T) and Control group (C), the WOE and IV concepts are modified. And the use of Net Weight of Evidence (NWOE) and Net Information Value (NIV) is suggested.

$$NWOE = log \frac{P(X = x_i | \, Y = 1)_T / P(X = x_i | \, Y = 0)_T}{P(X = x_i | \, Y = 1)_C / P(X = x_i | \, Y = 0)_C}$$

$$NIV = \sum [P(X = x_i | \, Y = 1)_T P(X = x_i | \, Y = 0)_C - P(X = x_i | \, Y = 1)_C P(X = x_i | \, Y = 0)_T] \, . NWOE_i$$

When a model is used in both training and validation data in SAS® Enterprise Miner ™, Penalized or Adjusted Net Information Value (PNIV) is automatically generated to reflect the difference between the NWOE of training data and the NWOE of validation data.

$$PNIV = NIV - Penalty$$

$$Penalty = \sum [P(X = x_i | \, Y = 1)_T P(X = x_i | \, Y = 0)_C - P(X = x_i | \, Y = 1)_C P(X = x_i | \, Y = 0)_T] \, . \omega_i$$

Where $\omega = NWOE_{Train} - NWOE_{Valid}$

Incremental response models in this paper will use cut off value of 0.3 for PNIV to select variables predicting increased fire rate and burned area.

In this modeling part, there are four models built using four different treatment variables, including x-y class, temperature class, rain class, and wind class (Figure 10). The purpose of these models is to measure the incremental probability of fire and incremental burned area when the treatment event occurs and to confirm the results of hypothesis testing earlier.

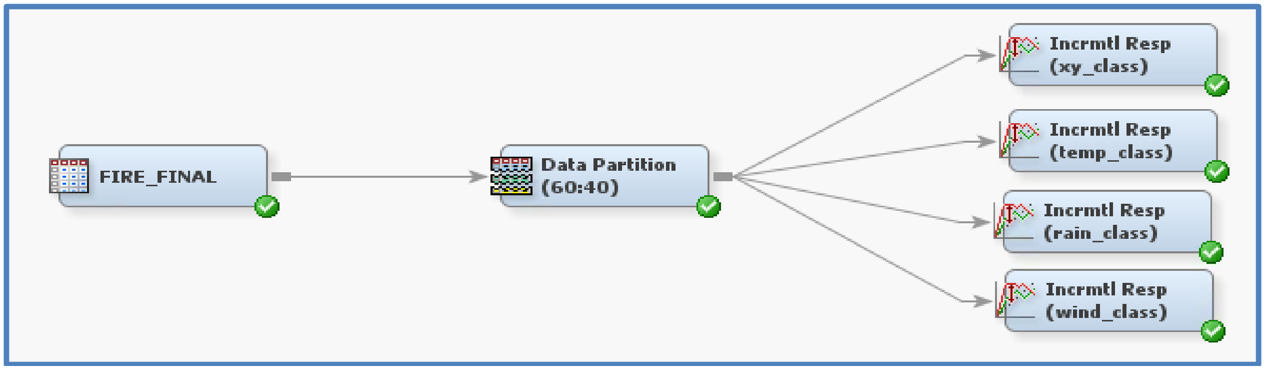

Figure 10. Incremental Response model process flow diagram

From figure 11, we can see that for all four different treatment variables, incremental rate of fire is recorded in both training and validation data sets. This means that probability of fire occurrence will increase when treatment event is equal to 1 (high risk spatial coordinates, high temperature condition, condition without rain, and high wind speed to spread fire). However, the same result is not recorded in the incremental average burned area. In figure 12, incremental area in treatment event of both training and validation data sets is only found in x-y class event and rain class event. These two treatment variables will be used for further analyses and fire control recommendations.

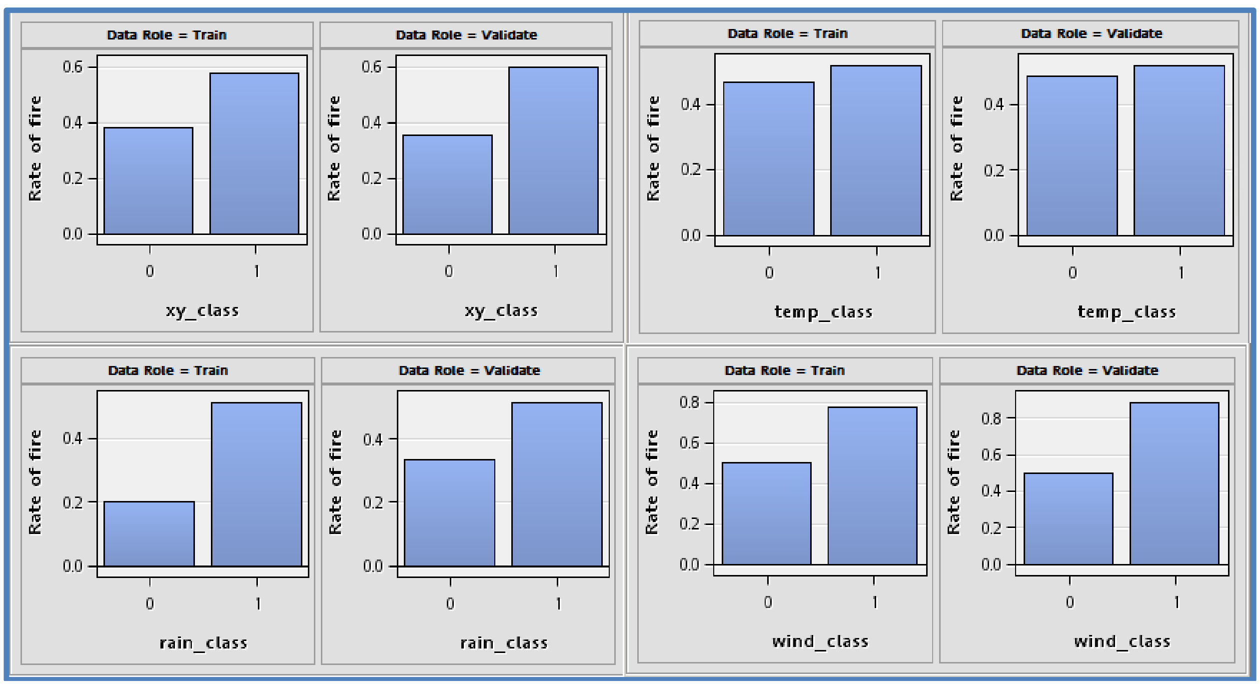

Figure 11. Rate of fire outcome across different treatment events.

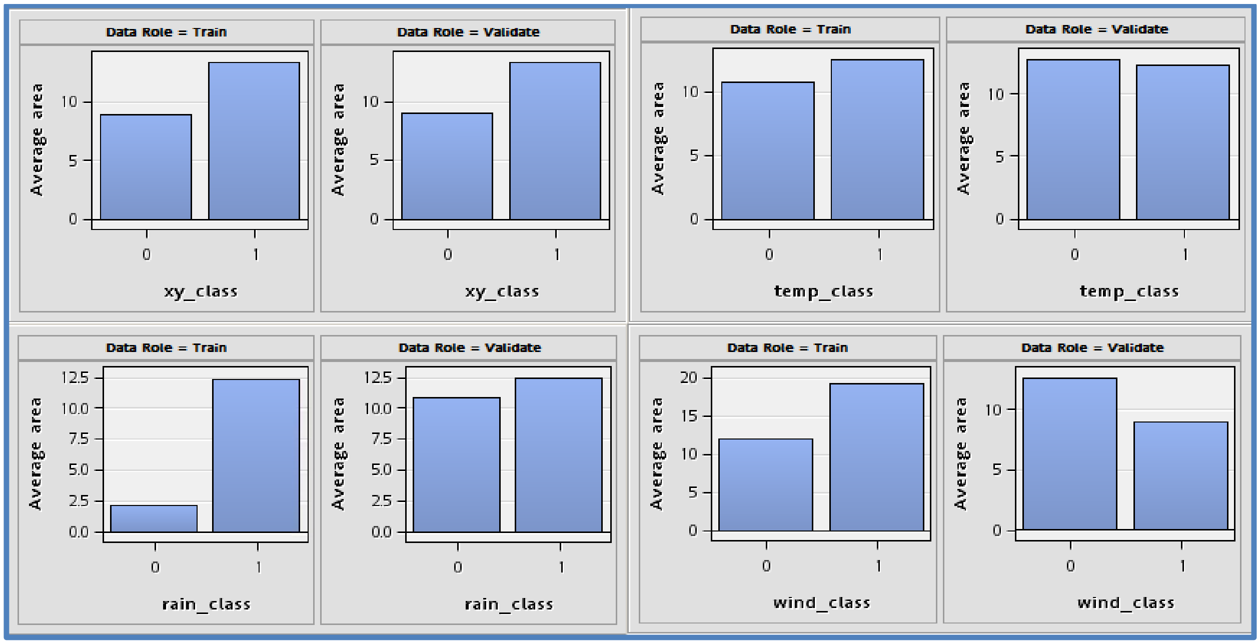

Figure 12. Average burned area across different treatment events.

Incremental response with the treatment of x-y class: Figure 11 and 12 revel that when the event of x-y class occurs (in forest regions with a high risk of fire), the probability of fire increases by more than 20%, and the average burned area increases by 4.5 hectares. This verifies the result of hypothesis testing using PROC GLM earlier.

Incremental response with the treatment of rain class: When the event of rain class occurs (no rain in the area), the probability of fire increases by 20% and 30% in the validation and the training data respectively. And the average burned area increases by 2 hectares and 10 hectares in the validation and the training data respectively. The difference increment between training and validation data could be explained by a rather low confidence interval of hypothesis testing with rain class treatment where the increments in fire rate and in burned area are calculated at 26.6% and 4.75 hectares respectively.

EXPLAINING BEST MODELS

Among four incremental response models, only models using x-y coordinate treatment and rain class treatment are stable enough to display an increment in both fire rate and burned area using a split of 60:40 for the training and validation data.

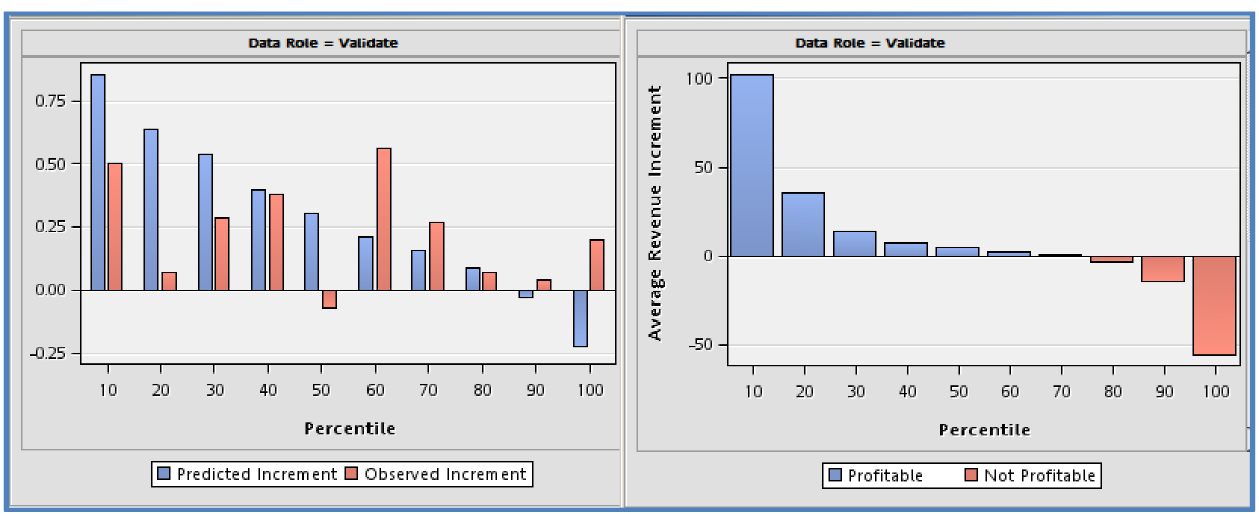

Figure 13. x-y class treatment – Validation incremental fire rate (left) and incremental average burned area (right)

In figure 13, the predicted increment (blue bar) represents the model-generated increased rate of fire when the treatment event occurs at difference deciles of the data set. Similarly, the observed increment (red bar) represents the actual increment recorded from the data set. A good incremental response model normally has predicted and observed increments pretty close to each other. And this is also observed in the model using x-y class treatment. The right part of figure 13 is called average revenue increment. This indicates the respective increment in revenue for each decile in a marketing campaign when there is a promotion (treatment event). All the blue deciles (profitable) are those having increased revenue outweighing its cost. In the context of this study, the average revenue increment chart represents the observations with an increment in burned area when the treatment event occurs. On the other hand, the red deciles represents those observations with a decrease in burned area when the treatment event occurs.

In the top decile of incremental model using x-y class as treatment variable (Figure 13), validation observed incremental fire rate is about 50%. This means the top decile of the data set is responsible for about 50% higher likelihood of fire occurrence. The predicted incremental fire rate is about 85%. The difference between predicted and observed increments measures how good the performance of this incremental response model is. Similarly, the incremental average burned area in the first decile is about 100 hectares. This can be interpreted as that the top decile of the data set would result in an increased burned area of 100 hectares in the forest.

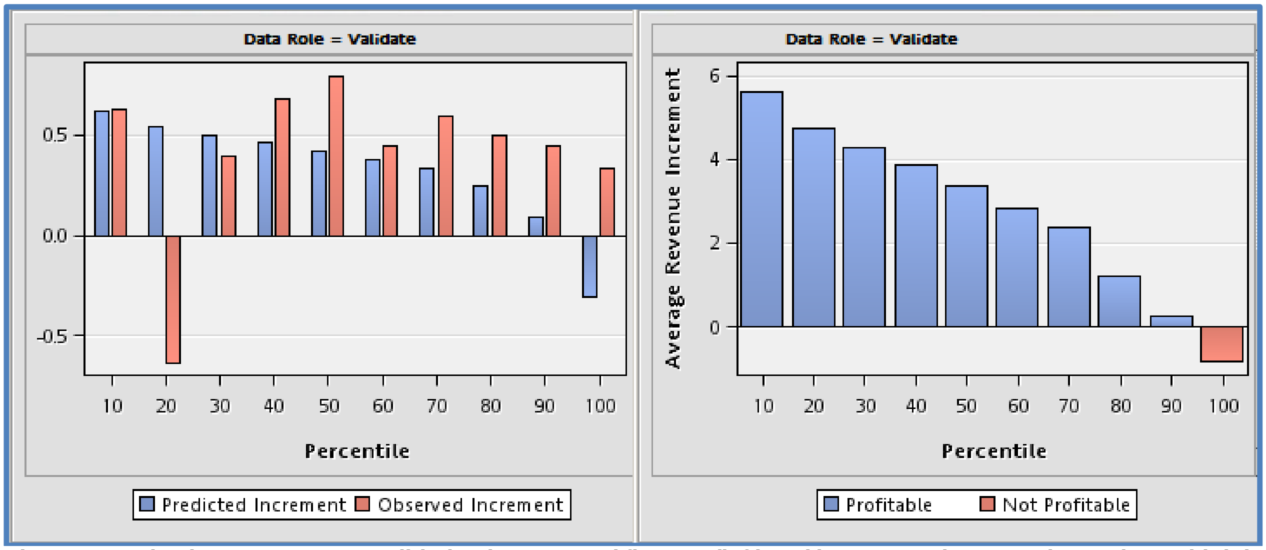

Figure 14. Rain class treatment – Validation incremental fire rate (left) and incremental average burned area (right)

For incremental model using rain class as treatment variable (Figure 14), there is a high fluctuation of validation observed incremental fire rate with a negative observed increment in the second decile. This complies with the hypothesis testing result of rain class treatment model as the model is significant at a rather low confidence interval. Validation incremental observed fire rate in the first decile is about 63% or 63% higher likelihood of fire occurrence. And the validation average burned area has its top decile incremental area of 5.6 hectares. This top decile's increased burned area is pretty low compared to that of model using xy-class treatment. We can see that overall, the incremental response model using x-y class has the best quality among all four models.

Selected variables using Adjusted NIV for x-y class incremental response model includes: x and y coordinate, outside rain, day of week, month of year, and outside temperature (Figure 15).

Variable Name	Adjusted Net Information Value	Net Information Value	Rank Percentile	Selection	Label
X	73.04266	1364.88	8.333333	Yes	x-axis spatial coordinate within the Montesinho park map: 1 to 9
rain	0	0	16.66667	Yes	outside rain in mm/m2
day	-61.8118	37.23674	25	Yes	
Y	-467.381	256.0162	33.33333	Yes	y-axis spatial coordinate within the Montesinho park map: 2 to 9
temp	-1049.71	1986.24	41.66667	Yes	temperature in Celsius degrees
month	-1302.89	923.643	50	Yes	
ISI	-1860.54	1064.582	58.33333	No	Initial Spread Index from the FWI system: 0.0 to 56.10
DMC	-1874.56	612.7359	66.66667	No	Duff Moisture Code index from the FWI system: 1.1 to 291.3
wind	-2381.72	242.8717	75	No	wind speed in km/h
RH	-2437.95	906.819	83.33333	No	% relative humidity
DC	-2458.76	2479.621	91.66667	No	Drought Code index from the FWI system: 7.9 to 860.6
FFMC	-3071.56	686.0202	100	No	Fine Fuel Moisture Code index from the FWI system: 18.7 to 96.20

Figure 15. x-y class treatment – Adjusted Net Information Value of selected variables

Selected variables using Adjusted NIV for rain class incremental response model includes: x and y coordinate, fine fuel moisture code index, outside rain, drought code, and initial spread index (Figure 16).

Variable Name	Adjusted Net Information Value	Net Information Value	Rank Percentile	Selection	Label
X	5108.269	33934.84	8.333333	Yes	x-axis spatial coordinate within the Montesinho park map: 1 to 9
FFMC	119.4143	21189.9	16.66667	Yes	Fine Fuel Moisture Code index from the FWI system: 18.7 to 96.20
rain	0	0	25	Yes	outside rain in mm/m2
DMC	-49.7613	1088.95	33.33333	Yes	Duff Moisture Code index from the FWI system: 1.1 to 291.3
DC	-205.044	16057.09	41.66667	Yes	Drought Code index from the FWI system: 7.9 to 860.6
ISI	-221.173	16520.38	50	Yes	Initial Spread Index from the FWI system: 0.0 to 56.10
month	-238.438	3418.909	58.33333	No	
day	-886.488	21233.44	66.66667	No	
RH	-1538.14	5018.63	75	No	% relative humidity
wind	-3435.37	14517.95	83.33333	No	wind speed in km/h
Y	-5790.22	16266	91.66667	No	y-axis spatial coordinate within the Montesinho park map: 2 to 9
temp	-6940.2	15504.35	100	No	temperature in Celsius degrees

Figure 16. Rain class treatment – Adjusted Net Information Value of selected variables

The common variables selected by both incremental response models are: x and y coordinate, and outside rain. These variables are also the important variables selected to create treatment variables

CONCLUSIONS

- Incremental response analyses confirm that high threat areas (x and y coordinate class of 1) account for a higher rate of fire, about 22% higher fire rate and 4.6 hectares of extra burned area on average. This suggests that we could use artificial rain or watering system to control the seriousness of forest fire in some high-threat areas, for example the top decile region with an increased burned area of 100 hectares and increase fire rate of 50%.
- Rain and no rain conditions also have an effect on the incremental rate of fire as well as average burned area. Top decile observations using rain class treatment models account for 63% higher likelihood of fire and 5.6 hectares of extra burned area.

FUTURE SCOPE

The real time system using meteorological data is a promising factor for a continuous fire forecast system in Montesinho Park forest. This could help to create interactive daily and weekly dashboards showing high risk region in the park. By locating high risk areas with real-time updates of fire threat indicators, the park management board can arrange resources and workforce to control over the forest fire situations. Also, the application of incremental response modeling technique could provide a good reference for fire control preparation regarding difference deciles of fire threat. For example, in the forest regions that are classified as top decile by x-y class treatment incremental response model, it is expected that burned area could go up by 100 hectares. Park managers should take a close look at this part of the forest to have sufficient fire control resources. For x-y class treatment model, month of the year is selected to be a predictor model using Penalized net information value criterion. By looking at the weight of evidence chart of the month variable, park managers could figure out which months of the year that have a higher threat of fire.

REFERENCES

[1] Arrue, B. C., Ollero, A., & Matinez de Dios, J. R. (2000). An intelligent system for false alarm reduction in infrared forest-fire detection. *IEEE Intelligent Systems and Their Applications*, 15(3), 64-73.

[2] Cortez, P., and Morais, A. (2007). A Data Mining Approach to Predict Forest Fires using Meteorological Data, *Proceedings of the 13th EPIA 2007*, Portugal, 512-523.

[3] Lee, T., Zhang, R., Meng, Z. and Ryan, L. (2013). Incremental Response Modeling Using SAS® Enterprise Miner ™, *Proceedings of the SAS Global Forum 2013 Conference*, Paper 096-2013, Cary, NC: SAS Institute Inc.

[4] Viegas, D., Biovio, G., Ferreira, A., Nosenzo, A. & Sol, B. (1999). Comparative Study of various methods of fire danger evalutation in southern Europe. *International Journal of Wildland Fire*, 9, 235–246.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Quyen Nguyen

Oklahoma State University

Phone: 405-612-2465

E-mail: quyen.t.nguyen@okstate.edu

Quyen Nguyen is a graduate student in Management Information Systems with SAS® and OSU Predictive Analytics Certificate at Spears School of Business, Oklahoma State University. She has 5 years of work experience in business analysis with Fortune 500 companies. She holds following credentials: SAS® Statistical Business Analyst, SAS® Advanced, Base programmer and SAS® Certified Predictive Modeler.

Dr. Goutam Chakraborty

Oklahoma State University

Stillwater, OK, 74078

E-mail: goutam.chakraborty@okstate.edu

Dr. Goutam Chakraborty is Ralph A. and Peggy A. Brenneman professor of marketing and founder of SAS and OSU data mining certificate and SAS and OSU marketing analytics certificate at Oklahoma State University. He has published in many journals such as Journal of Interactive Marketing, Journal of Advertising Research, Journal of Advertising, Journal of Business Research, etc. He has chaired the national conference for direct marketing educators for 2004 and 2005 and co-chaired M2007 data mining conference. He has over 25 years of experience in using SAS® for data analysis. He is also a Business Knowledge Series instructor for SAS®.