Available online at www.sciencedirect.com

SciVerse ScienceDirect

Procedia Computer Science 12 (2012) 282 – 287

Procedia
Computer Science

Complex Adaptive Systems, Publication 2
Cihan H. Dagli, Editor in Chief
Conference Organized by Missouri University of Science and Technology
2012- Washington D.C.

Estimation of the burned area in forest fires using computational intelligence techniques

A. Murat Özbayoğlu*, Recep Bozer

TOBB University of Economics and Technology, Department of Computer Engineering, Ankara, 06560, Turkey

Abstract

Forest fires have environmental impacts that create economic problems as well as ecological damage. Developing a means to predict the possible size of a fire shortly after it first breaks out has the potential to guide proper resource allocation for improved fire control and was the main motivation of this research. In this study, the burned areas resulting from possible forest fires were estimated using historical forest fire records which contained parameters like geographical conditions of the existing environment, date and time when the fire broke out, meteorological data such as temperature, humidity and wind speed, and the type and number of trees in a unit area. The data was from the Department of Forestry in Turkey and contained 7,920 forest fire records from 2000 and 2009. The output from the estimation methods implemented in this work predicted the size of the area lost due to the fire and the corresponding fire size, i.e. big, medium, or small fire. Some of the estimation methods investigated were Multilayer Perceptron (MLP), Radial Basis Function Networks (RBFN), Support Vector Machines (SVM) and fuzzy logic. The results of these estimates are presented and compared to similar studies in literature.

Keywords: Forest fire loss estimation; forest fire burned area; computational intelligence; artificial neural networks; Radial Basis Function; Multilayer Perceptron; Support Vector Machines

1. Introduction

Forest fires are one of the environmental problems that threaten forests and natural life in the world. In general expression, a fire which tends to spread freely and can burn all living and natural components within the forest is called forest fire. Every year, an average of 4 million hectares of forest in the world and an average of 550 thousand hectares of forest in the Mediterranean belt burns. In particular, Turkey is located in a very sensitive area; approximately 60% of its forests are in first degree sensitive areas. For the year 2008, 2135 forest fires occurred; as a result 29749 hectares of forest were burned. The economic damage caused by these fires was approximately 100

* Corresponding author. Tel.: +90-312-292-4073; fax: +90-312-292-4081.
E-mail address: mozbayoglu@etu.edu.tr.

 doi:10.1016/j.procs.2012.09.070

million USD [1]. This indicates the necessity of trying to prevent and/or minimize the damage caused by forest fires.

The aim of this study was to estimate the size of a burned area due to a forest fire. However, the models were implemented using only meteorological, climate and topography data which are readily available at real-time. Even if the fire size were to be estimated roughly, this would provide a tremendous benefit for firefighting units to allocate resources, especially in peak seasons where there may be multiple fires happening at the same time.

2. Literature Review

Forest fire prevention and loss estimation systems depend on real time data retrieval from various sources. In order to implement these systems, a lot of countries have formed forest fire database and/or decision support systems. Among the most notable systems, European Forest Fire Information Systems (EFFIS) [2], Canadian Forest Fires Danger Rating System-CFFSDRS [3] and National Fire Danger Rating System-NFDRS in USA [4] can be counted. These systems use landscape information, topography, real-time weather data, etc. to identify forest fire possibility and risk at any given time. CFFSDRS uses subsystems such as Fire Weather Index (FWI) and Fire Behavior Prediction (FBP).

There are several studies performed on forest fire prediction systems using forest fire databases based on computational intelligence models. Brillinger et. al. [5] created a statistical model using historical data in order to predict the possibility of a forest fire for a given particular topography and elevation. The system was based on previous seasonal fire history in that region without using weather conditions. Jaiswal et. al. [6] created a Geographical Information System (GIS) based model calculating forest fire risk using climate, landscape and topographic data. Iliadis [7] created a decision support system using fuzzy logic and algebra. He predicted the risky forest areas in Greece and out of 20, he correctly spotted 12 of them. In one application, association rule mining along with clustering was used for identifying the risky regions for forest fires and burned area estimation for existing forest fires [8]. Cheng and Wang used spatio-temporal data mining techniques and recurrent neural networks to estimate the burned area in forest fires [9]. Cortez and Morais used multivariate regression, decision trees, random forest, artificial neural networks and SVM to estimate forest fire loss and obtained the best results through SVM [10]. The success rate was %46 for estimation errors of less than 1 hectare. In another study [11], forest fire prediction was implemented with neural networks and SVM using only humidity and precipitation. In this study, we used a different novel approach, using output clusters for different burned area sizes.

3. Forest Fire Data

Forest fires are caused by several factors. Weather conditions such as temperature, relative humidity, wind are important elements on the spread property of a forest fire. Relative humidity, which is related to combustible content's moisture, is at its highest degree at the early morning, but by the midday it gets its minimum value. When it is less than 10%, it becomes very dangerous. Evaporation increases with increasing temperature, and increased evaporation reduces humidity. In general, fire risk starts at 77 °F, and it increases with increasing temperature. Another weather factor is wind, which has destructive effects on the branches of flammable materials, makes the fire spread quickly and negatively affects firefighting. In terms of forest fires, risk starts at a wind speed of 15 km/hr.

Land conditions are also influential on forest fires. Aspect which shows the degree of sunlight on hills is important. Also, high tendency prevents the spread of fire. In addition, fire risk reduces with increasing altitude. 80% of forest fires in Turkey have been at an altitude of 0-400 meters.

Season and hour, when the forest fires occur in a day, have considerable effects on the behavior of fires. Most forest fires occur especially from spring to autumn. Moisture content of air is related to the hour of the day, since sun light and evaporation directly influences it. Furthermore, type and number of trees in unit area determine sensitivity of the forest in a possible fire. Some trees tend to burn very quickly when a fire occurs, in contrast some are resistant. Generally the trees which have wide leaves cause the fire becoming bigger, whereas pin leaves make the fire spread slowly. If the tree distribution in an area is condensed, enlargement is easy, fire can spread very quickly and this makes prevention of fires difficult, so the number of trees in a unit area is also an important factor.

Our proposed solution includes nine input parameters, relative humidity, wind speed, temperature, aspect, tendency, season, hour, tree type, number of trees in unit area, and one output parameter, burned area due to the fire.

The data used in the experiments was collected from 2000 to 2009 and included 7920 forest fires. A lot of preprocessing was performed on the data due to inconsistencies, different metrics used by different stations, manual data entry errors, etc. Nonnumeric values were enumerated; values in each column were normalized to [0-1]. Euclidean distance was used in distance calculations between different points in the data set. Figure 1 illustrates the distribution of the forest fires based on their amount of burned areas. Figure 1 (note that the x-axis is not linear, and y- axis values are log transformed) indicates most of the fires (78%) were less than 1 hectares in size.

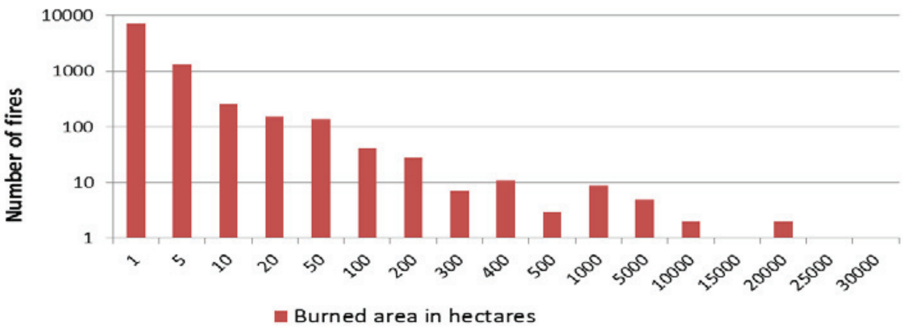

Fig. 1. The distribution of the burned area values in the forest fire data set

4. Methodology

MLP [12], RBFN [12] and SVM [13] were implemented in this study for forest fire burned area estimation. Several different input-output configurations were implemented. Burned area values were clustered using k-means and different numbers of output clusters were tested. Using k-means, cluster centroids were also calculated. For each network model, the numbers of neurons in the output layer were determined by the number of clusters. For the training and the cross validation (CV) set, the output was set as a fuzzy data vector in such a way that one data could be part of 2 clusters with a certain membership degree. The fuzzification of the output was represented in Figure 2. For example, if the burned area of a particular forest fire was 150 hectares, then that fire would be represented as a small fire with a membership degree of (150–63.86) / (273.05–63.86) = 0.41, and as a medium fire with a membership degree of (273.05–150) / (273.05–63.86) = 0.59, hence the output vector would be [0 0.41 0.59 0 0].

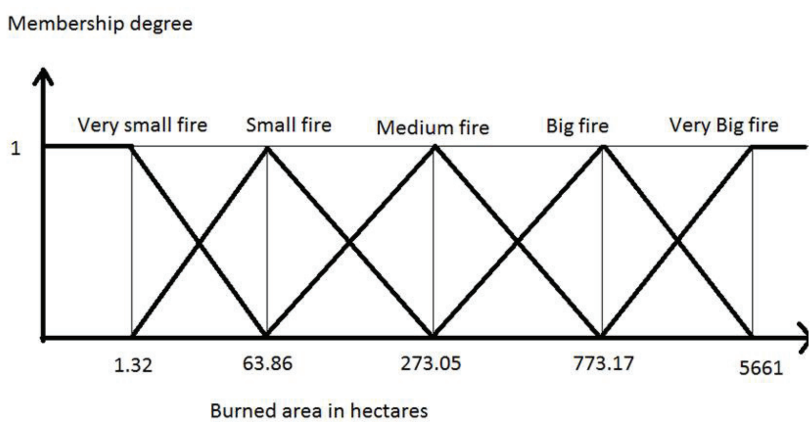

Fig. 2. Output cluster values for 5 cluster case

During the experiments, it was observed that due to the uneven distribution of burned areas of the forest fires, it was very difficult to predict the burned area in hectares, since almost %80 of the data represented fires that were less than 1 hectares, but there were also a handful of forest fires that had more than 10000 hectares in burned area. As a result, the output values were fuzzified and different number of clusters (2,3 and 5) were used in the output. The results indicate that trying to identify the clusters (small fire, medium fire, big fire, etc.) were a better approach than trying to predict the actual burned area. The output burned area clusters were tabulated in Table 1 for 5 cluster case.

Table 1. Output cluster values for 5 different sizes of forest fires

Cluster No	Cluster Description	Cluster center (in hectares)	Number of data points in cluster	Number of test data in cluster
1	Very small fire	1.32	7825	1565
2	Small fire	63.86	76	15
3	Medium fire	273.05	14	3
4	Big fire	773.17	3	1
5	Very big fire	5661	2	1

The data set was divided into 3 sections. 60% of the data was used for training purposes, 20% was used for CV and the remaining 20% of the data was reserved for testing. Table 1 indicates the numbers of data points in the clusters were not evenly distributed. Training with such a distribution would prevent the models to learn all clusters properly. So, the amount of data points in each cluster were normalized by duplicating the data in the misrepresented clusters, hence each cluster had similar amount of data. The same data distribution was used for training, CV and testing. But, for the 5 cluster case, "Big Fire" and "Very Big Fire" cases only had 3 and 2 data points. As a result, for "Big Fire", 1 data point (before duplication) was assigned for training, 1 data point was spared for CV and 1 data point was used in testing; in a similar fashion for "Very Big Fire", 1 data point was assigned for training and 1 data point was used in testing. So, the data from "Very Big Data" was not represented in the CV data set.

In a separate model, the training input data was clustered into 350 different clusters using k-means algorithm. Then, for each cluster, the mean output value, which is the burned area in hectares, was calculated. Then, each test data vector was matched against the clusters to identify which cluster the data belonged in. Finally, the test data output was compared to the cluster center output (mean output value of the cluster) and performance was measured.

5. Results and Discussions

The results from four of several different fire estimation models are described herein. For the models whose performances were had less than 50 % True, the results are not explicitly reported. Direct estimation of the burned area in hectares, log estimation of the burned area, log estimation of 5 clusters and 3 clusters cases were among the unsuccessful models investigated. Though natural logarithm transformed 2 clusters output case were relatively successful, those results are not explicitly stated here. In this paper, the best results obtained from 5 clusters, 3 clusters and 2 clusters outputs are presented.

In Table 2 the performance results of MLP for 5 clusters at the output are tabulated. RBFN results are not presented here due to their poor performance. MLP success was 53.02% when all the clusters were treated equally.

Table 2. MLP performance for 5 clusters at the output

Cluster No	True	False	Total Data points in Cluster	True %	False %
1	1228	337	1565	78.45	21.55
2	8	7	15	53.33	46.67
3	1	2	3	33.33	66.67
4	0	1	1	0.00	100.00
5	1	0	1	100.00	0.00
Total data points Percent success (%)	1238 **78.11**	347 21.89	Average Percent Success Rate (%) when all clusters have equal weight	**53.02**	46.98

In another model, 3 output clusters were used. The results are tabulated in Table 3. The MLP overall performance was improved to 62.89% compared to 53.02% for the 5 cluster case.

Table 3. MLP performance for 3 clusters at the output

Cluster No	True	False	Total Data points in Cluster	True %	False %
1	1402	179	1581	88.68	11.32
2	2	0	2	100.00	0.00
3	0	1	1	0.00	100.00
Total data points Percent success (%)	1404 **88.63**	180 11.37	Average Percent Success Rate (%) when all clusters have equal weight	**62.89**	37.11

In one of the previous studies [11], models only using 2 parameters (humidity and wind speed) were developed and the results were provided. In order to compare the results obtained in this study, same parameters were used in another developed model. The results from the reduced parameter MLP model are tabulated in Table 4.

Table 4. MLP performance for the reduced input (2 inputs) and 3 clusters at the output model

Cluster No	True	False	Total Data points in Cluster	True %	False %
1	1532	49	1581	96.90	3.10
2	0	2	2	0.00	100.00
3	1	0	1	100.00	0.00
Total data points	1533	51	Average Percent Success Rate (%) when all clusters have equal weight	**65.63**	34.37
Percent success (%)	**96.78**	3.22			

The results from Table 4 indicate better overall outcome (65.63%) was observed by using the reduced parameter model, however again, no improvement was visible with the RBFN model (not presented here).

In another developed model the data was clustered to only 2 clusters at the output (small and big fire). This time, SVM was also included in the developed models. Instead of directly using the burned area in hectares, natural logarithm (ln) of the output was used by applying ln(x+1) to the output. These new 2 cluster models were developed with MLP, RBFN and SVM and the best results were obtained with SVM. These results are tabulated in Table 5.

Table 5. SVM performance for the (ln x + 1) transformed output with 2 cluster model

Cluster No	True	False	Total Data points in Cluster	True %	False %
1	1077	203	1280	84.14	15.86
2	130	174	304	42.76	57.24
Total data points	1207	377	Average Percent Success Rate (%) when all clusters have equal weight	**63.45**	36.55
Percent success (%)	**76.20**	23.80			

Applying natural logarithm transformation to the forest fire data did not improve the performance of the estimation process. However, the distributions of the data to different clusters were positively affected. The models were also analyzed using the error metrics based on the deviation from target for the actual burned area. For this purpose, RMSE (Root Mean Square Error), MAE (Mean Absolute Error), MAPE (Mean Absolute Percent Error) values are calculated and the model results are presented in Table 6.

Table 6. Performance comparison of models used in the study

Estimation model	Model details	RMSE	MAE	MAPE (%)
MLP	3 cluster, 2 input	15.85	4.11	51
RBFN	5 cluster	18.35	4.05	54
SVM	2 cluster with ln transformed output	7.33	3.36	69

Table 6 indicates that the burned area estimation had a high deviation from target, (more than 50% MAPE). One explanation for this is the large number of very small fires and a handful of very big fires within the data set resulting in a very high overall MAPE, since almost %80 of the fires resulted in less than 1 hectare of burned area; for such fires, even a 1 hectare deviation would result in more than %100 error. At the same time, RMSE was also recorded relatively high compared to the MAE due to the huge error deviations in big fires. The contribution of the squared error from big fires resulted in high RMSE, since RMSE is more sensitive to outlier data and large errors.

When these results were compared to the previous studies performed in the literature, it was observed that the current study fared well compared with the other models in literature, even though there were not many studies performed for burned area estimation. Also it should be noted that the estimation results in these types of studies are very dependent on the quality of the data set. In [7] 60% was achieved for risky area identification and in [10] good overall estimation was achieved with the best MAE performance of 2.85 hectares and best RMSE was 12.71. In [11], a different analysis was implemented with the detection of a fire or no fire day with over 90% accuracy. However, they did not perform burned area estimation. The best overall estimation results obtained in this study

(66%) were obtained by the implementation of MLP model to 3 output clusters (small, medium and big fire) and MAE of 3.36 hectares and RMSE of 7.33 hectares using SVM model with 2 clusters at the output (small or big fire).

The results of the last model where the input vectors were clustered into 350 different clusters and the corresponding mean output values were calculated for each cluster were also satisfactory. In [10], the percent of test data points within 1 hectare of error was 46%, in this study 53% was achieved for less than 1 hectares of error. When the tolerance was increased to 2 hectares, the model in [10] achieved 61% success, whereas 72% of the data points in this study were within 2 hectares of the output mean value of the clusters.

6. Conclusions

In this study, the burned forest area during a forest fire was estimated using different models. Trying to forecast the size of a staring fire is especially important for firefighters in order to figure out how much manpower and resources should be allocated in each fire. Since the environmental damage, ecological effects and financial loss in these forest fires might have a considerable impact on a country or region, any level of success in a rough estimate for the forest fire can be crucial in proper resource allocation and scheduling for extinguishing these fires. Several different estimation models were prepared and the most successful ones are presented in this study. The results indicate that in some models the performances in the estimation process were above %60, beating most of the studies implemented in literature. The best model turned out to be an MLP model using only two inputs (humidity and wind speed) and three clusters (small, medium and big fire) in the output with more than %65 success rate.

Future works may include probabilistic models that can detect the contingency of the start of a possible forest fire under certain conditions. Such a model can be combined with the models developed in this study to provide a complete system where risky conditions for a big fire can be closely monitored, also live satellite or camera views can be included in order to provide visual features that can increase the performance of the overall model.

References

1. N. Dogan N., Forest Fire Management, Deparment of Forestry Publications, (2009)
2. EFFIS European Forest Fire Information System, http://effis.jrc.ec.europa.eu/index.php. (May 17, 2012).
3. B. Stocks, T. Lynham, B. Lawson, M. Alexander, C.V. Wagner, R. McAlpine, D. Dube, The Canadian forest fire danger rating system: an overview, The Forestry Chronicle, 65, 450-457, (1989).
4. J. Deeming, R. Burgan, J. Cohen, The national fire danger rating system-1978, US Department of Agriculture Forest Service General Technical Report, INT-39, 63, (1977).
5. D.R. Brillinger, H.K. Preisler, J.W. Benoit, Risk assessment: a forest fire example. In D. Goldstein, T. Speed, (Eds.), Statistics and Science: a Festschrift for Terry Speed, IMS Lecture Notes Monograph Series, vol. 40, Institute of Mathematical Statistics, Beachwood, OH, 177-196, (2003).
6. R. Jaiswal, M. Saumitra, D. Kumaran, S. Rajesh, Forest fire risk zone mapping from satellite imagery and GIS, International Journal of Applied Earth Observation and Geo-information, 4, 1-10, (2002).
7. L. Iliadis, A decision support system applying an integrated fuzzy model for long term forest fire risk estimation, Environmental Modeling and Software, 20, 613-621, (2005).
8. T. Cheng, J. Wang, Applications of spatio-temporal data mining and knowledge for forest fire, In Proceedings of the ISPRS Technical Commission VII Mid Term Symposium, Enschede, The Netherlands, 148-153, (2006).
9. T. Cheng, J. Wang, Integrated spatio-temporal data mining for forest fire prediction, Transactions in GIS, 12, 591-611, (2008).
10. P. Cortez, A. Morais, A data mining approach to predict forest fires using meteorological data, In Proceedings of the 13th Portugese Conference on Artificial Intelligence, 512-523, (2007).
11. G.E. Sakr, H.I. Elhajj, M. George, Efficient forest fire occurrence prediction for developing countries using two weather parameters, Engineering Applications of Artificial Intelligence,24, 888-894, (2011).
12. S. Haykin, Neural Networks: A Comprehensive Foundation, Second Edition, Prentice Hall, Upper Saddle River, NJ (1999).
13. V.N. Vapnik, Statistical Learning Theory, John Wiley & Sons, (1998).