Journal of Theoretical and Applied Information Technology
15th December 2011. Vol. 34 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

AN EVALUATION OF K-MEANS BASED ANN USING FOREST FIRE DATA IN SPATIAL DATA MINING

[1]N.NAGA SARANYA, [2]G.PADMAPRIYA, [3]S.HEMALATHA, [4]M.HEMALATHA
[1,2,3] Research Scholar (C.S), Karpagam University, Coimbatore, India- 641021.
[4]Head, Department of Software Systems, Karpagam University, Coimbatore, India -641021.
E-mail: nachisaranya01@gmail.com

ABSTRACT

The explosive growth of spatial data and extensive utilization of spatial databases [1] emphasize the necessity for the automated discovery of spatial knowledge. In modern times, spatial data mining has emerged as an area of huge research. Forest fires are a chief green concern, causing inexpensive and environmental damage while endangering human lives across the world. The fast or early detection of forest fires [2] is a essential element for controlling such phenomenon. The application of remote sensing is at present a significant method for forest fires monitoring [3], particularly in vast and remote areas. This paper presents an intelligent system to detect the presence of forest fires in the forest spatial data using SMO and Artificial Neural Networks. Extensive experimental assessments on publicly available spatial data illustrated the efficiency of the proposed system in effectively detecting forest fires. Finally, since large fires are rare dealings, outlier detection techniques will also be addressed.

Keywords: *Spatial Data Mining, Forest Fire Data, SMO, ANN*

1. INTRODUCTION

India, with a forest cover of 76.4 million hectares, contains a range of climate zones, including the steamy south, northwestern deserts, Himalayan Mountains, and the wet north-east. Forests are widely distributed in the country. India's forests are artistic with a mixture of biomes and biological communities. The forest vegetation [4] in the country varies from tropical evergreen forests in the West Coast and in the Northeast to alpine forests in the Himalayas in the North. In between the two edges, there are semi-evergreen forests, deciduous forests, sub-tropical broad-leaved hill forests, sub-tropical pine forests, and sub-tropical montane temperate forests. With increasing population pressure, the forest cover of the country is weakening at an upsetting rate. Along with various factors, forest fires are a major reason of poverty of Indian forests. According to a Forest Survey of India Report, about 50 percent of forest areas in the country are fire prone (ranging from 50 percent in some states to 90 percent in the others). About 6 percent of the forests are prone to severe fire damage.

2. TRAINING OF ANN PARAMETERS

2.1 Partitioning Algorithm

K-Means algorithm is very popular for data clustering [5]. The Algorithm goes like this

Step1: Select k Center in the problem space (it can be random).

Step2: Partition the data into k clusters by grouping points that are closest to those k centers.

Step3: Use the mean of these k clusters to find new centers.

Step4: Repeat steps 2 and 3 until centers do not change.

This algorithm normally converges in short iterations.

A set of observations $(\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n)$ [6], where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into k sets $(k \le n)$ $\mathbf{S} = \{S_1, S_2, \ldots, S_k\}$ so as to minimize the within-cluster sum of squares (WCSS):

$$\underset{\mathbf{S}}{\arg\min} \sum_{i=1}^{k} \sum_{\mathbf{x}_j \in S_i} \|\mathbf{x}_j - \boldsymbol{\mu}_i\|^2$$

where $\boldsymbol{\mu}_i$ is the mean of points in S_i.

Journal of Theoretical and Applied Information Technology
15th December 2011. Vol. 34 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

2.2 Proposed Multi Layer Neural Network Architecture

The following diagram illustrates the proposed multi-layer neural network design which is going to be used in this project [7],[8].

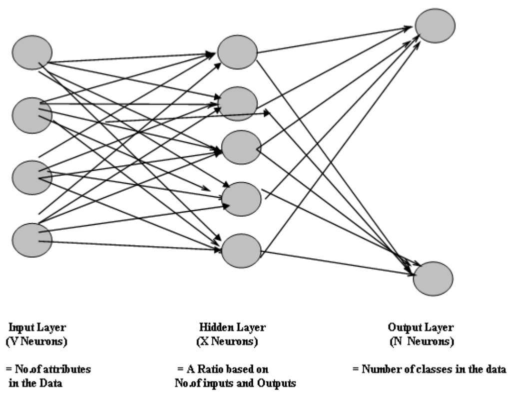

Figure 1: Multi Layer Neural Network Architecture

MLNN Algorithm:

1.Initialize the weights in the network (often randomly)
2. repeat
* for each example e in the training set do
1. O = neural-net-output (network, e);
forward pass
2. T = teacher output for e
3. Calculate error (T - O) at the output units
4. Compute delta_wi for all wt from hidden layer to output layer ;
backward pass
5. Compute delta_wi for all wt from input layer to hidden layer ;
backward pass continued
6. Update the weights in the network
* end
3. until all examples classified correctly or stopping criterion satisfied
4. return (network)

3. VULNARABILITY OF THE INDIAN FOREST

The ecological and socio-economic consequences of wild land fires in India include:

Loss of timber, loss of bio-diversity, loss of wildlife habitat, global warming, soil erosion, loss of fuel wood and fodder, damage to water and other natural resources, loss of natural regeneration. Estimated average tangible annual loss due to forest fires in country is Rs.440 crore

The vulnerability of the Indian forests to fire varies from place to place depending upon the type of vegetation and the climate. Every year there are one or two major incidences of forest fire in this region. The other parts of the country dominated by deciduous forests are also damaged by fire is shown in table1.

Table 1. Susceptibility of Indian forests to wildfire

S.No	Type of Forests	Fire frequent(%)	Fire Occasional (%)
1	Coniferous	8	40
2	Moist Deciduous	15	60
3	Dry Deciduous	5	35
4	Wet/Semi-Evergreen	9	40
5	North-eastern Region	50	45

3.1 Forest Fire Statistics

In India there are no comprehensive data to indicate the loss to forests in terms of area burned, values, and volume and regeneration damaged by fire [10]. The available forest fire statistics are not reliable because they under estimate fire numbers and area burned. The reason behind this is attributed to the fear of accountability. However, Forest Survey of India in a country-wide study in 1995 estimated that about 1.45 million hectares of forest are affected by fire annually. According to an assessment of the Forest Protection Division of the Ministry of Environment and Forests, Government of India, 3.73 million hectares of forests are affected by fires; annually in India is shown in fig 2 and fig 3.

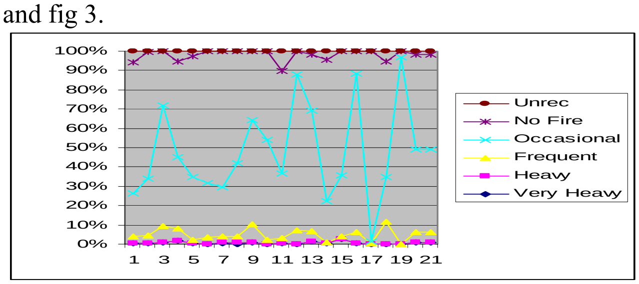

Figure 2: Extent of fire incidents (ha)

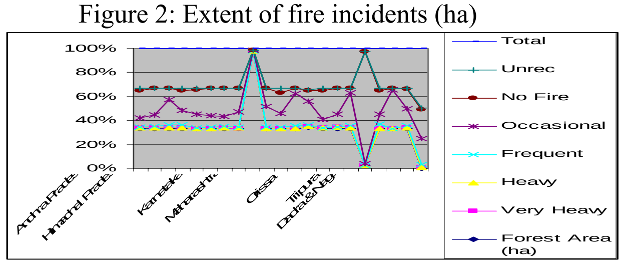

Figure 3: Degree of fire incidence in forest areas of India.

Journal of Theoretical and Applied Information Technology
15th December 2011. Vol. 34 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

3.2 Research Issues

In India, there is an urgent need to initiate research in the fields of fire detection, suppression, and fire ecology for better management of forest fires. The research and technology developed in western countries always suitable for the Indian environment. Thus, it is essential that original research specific for Indian conditions be conducted. The Government is considering setting up a National Institute of Forest Fire Management with satellite centres in different parts of the country to bring the latest forest fire fighting technologies to India through proper research, training of personnel, and technology transfer on a long term basis.

4. EXPERIMENTAL RESULTS

All experiments reported in this study were conducted using the rapid miner, an open source library for the statistical environment [11] that facilitates the use of DM techniques in classification and regression tasks. In particular, the rapid miner uses the random forest (RF algorithm by L. Breiman and A. Cutler), nnet (for the NN) and kernlab [12] packages. Before fitting the models, some preprocessing was required by the K-Means, NN models. Also, for the NN and SVM methods, all attributes were standardized to a zero mean and one standard deviation. Next, the regression models were fitted. The MR parameters were optimized using a least squares algorithm, while the DT node split was adjusted for the reduction of the sum of squares. Regarding the remaining methods, the default parameters were adopted for the RF (e.g. T = 1500), the NN were adjusted using NR = 3 trainings and E = 100 epochs of the BFGS algorithm and the Sequential Minimal Optimization algorithm was used to fit the SVM. After fitting the DM models, the outputs were post processed using the inverse of the logarithm transform. In few cases, this transformation may lead to negative numbers and such negative outputs were set to zero.

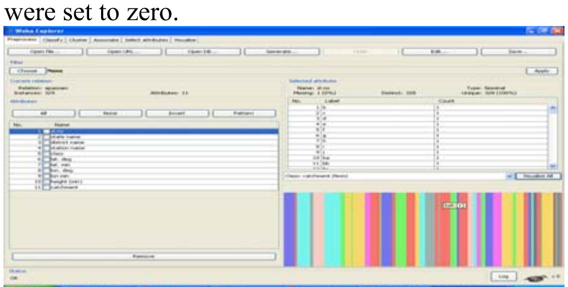

Figure 4: Overall Instances of Forest Fire Data

Test mode: evaluate on training data

Classifier model (full training set)

SMO: Kernel used: Linear Kernel: K(x, y) = <x, y>

Classifier for classes: dbd, dbe

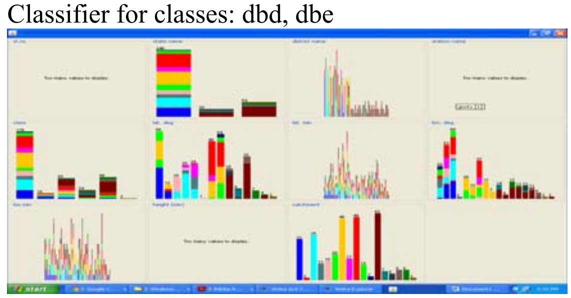

Figure 5: Clustering of Forest Fire Data

BinarySMO

Machine linear: showing attribute weights not support vectors.

-0.0043 * (normalized) sl.no=b

+ -0.0147 * (normalized) sl.no=c……….

Number of kernel evaluations: 447 (82.339% cached)

Classifier for classes: dbd, dae

Table 2: Evaluation on training set

Correctly Classified Instances	328 (100 %)
Incorrectly Classified Instances	0 (0 %)
Kappa statistic	1
Mean absolute error	0.1038
Root mean squared error	0.223
Relative absolute error	98.151 %
Root relative squared error	97.0934 %
Total Number of Instances	328
Ignored Class Unknown Instances	1

Accuracy by Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	0.998	dbd
	1	0	1	1	1	1	dbe…………
Wt Avg.	1	0	1	1	1	1	

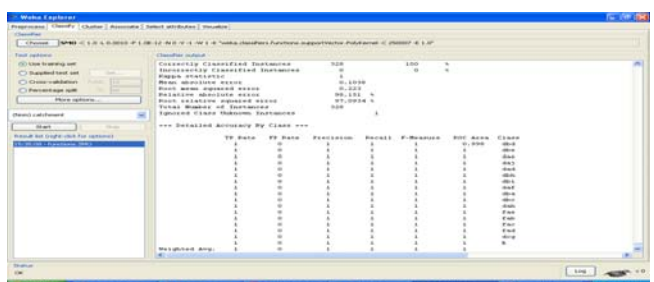

Figure 6: Classified Instances of Forest Fire Data

Confusion Matrix

a b c d e f g h i j k l m n o p q<-- classified as

32 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 |a = dbd

0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = dbe……….

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 2 | q = k

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

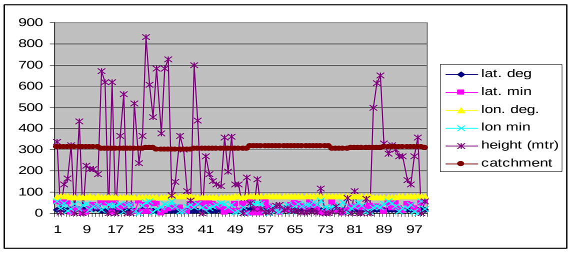

Figure 7: Plot for forest fire in India

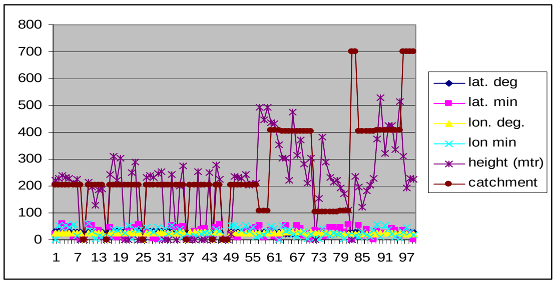

Figure 8: Plot for Frequent forest fire in India

4.1 Outlier Prediction

Time taken to build model: 6.92 seconds
Predictions on test split
inst#, actual, predicted, error, probability distribution (class)
1 6:dbh 7:dbi + 0.066 0.007 0.096 0.059 0.051 0.11 *0.118 0.074 0.103 0.081 0.044 0.088 0.015 0.029 0.037 0.022 0 (obsy)
2 1:dbd 1:dbd *0.118 0.007 0.081 0.051 0.066 0.096 0.103 0.074

Detailed Accuracy by Class

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.933	0.021	0.875	0.933	0.903	0.983	dbd
	0	0	0	0	0	0.928	dbe
	0.8	0.029	0.727	0.8	0.762	0.965	dae
	0	0	0	0	0	?	k
Wt Avg.	0.795	0.027	0.767	0.795	0.774	0.974	

Table 3: Evaluation on test split

Correctly Classified Instances	89 (79.4643 %)
Incorrectly Classified Instances	23 (20.5357 %)
Kappa statistic	0.767
Mean absolute error	0.1041
Root mean squared error	0.2237
Relative absolute error	98.2783 %
Root relative squared error	97.3805 %
Total Number of Instances	112

Confusion Matrix

a b c d e f g h i j k l m n o p q <-- classified as
14 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 | a = dbd
1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = dbe
0 0 8 1 1 0 0 0 0 0 0 0 0 0 0 0 0 | c = dae

5. CONCLUSION

Forest fires cause a major ecological damage while threatening human lives. Here two different DM algorithms, including K-Means, SMO and Neural Network (NN) were tested. The proposed solution, which is based in a ANN, is capable of predicting small fires, which form the majority of the fire occurrences. The drawback is the lower predictive accuracy for large fires. To our facts, this is the first time the burn area is predicted using only meteorological based data and further tentative research is required. As argued, predicting the size of forest fires is a challenging task. To improve it, we believe that additional information required, such as the type of vegetation and firefighting intervention. Nevertheless, the proposed model is still useful to recover firefighting resource management. Finally, since large fires are rare dealings, outlier detection techniques [13] will also be addressed.

ACKNOWLEDGEMENT:

I thank the management of Karpagam University for their kind support to do this article successfully.

REFRENCES:

[1] D. Hand, H. Mannila, and P. Smyth. Principles of Data Mining. MIT Press, Cambridge, A, 2001.

[2] B. Arrue, A. Ollero, and J. Matinez de Dios. An Intelligent System for False Alarm Reduction in Infrared Forest-Fire Detection. IEEE Intelligent Systems, 15(3):64–73, 2000.

[3] D. Stojanova, P. Panov, A. Kobler, S. Dzeroski, and K. Taskova. Learning to Predict Forest Fires with Different Data Mining Techniques. In D. Mladenic and M. Grobelnik, editors, 9th International multiconference Information Society (IS 2006), Ljubljana, Slovenia, 2006.

[4] K. Malarz, S. Kaczanowska, and K. Kulakowski. Are forest fires predictable? International Journal of Modern Physics, 13(8):1017–1031, 2002.

[5] Hamerly, Greg, and Elkan, Charles. "Learning the k in k-means." Retrieved from the World Wide Web at http://books.nips.cc/papers/files/nips16/NIPS2003_AA36.pdf through Google Scholar on November 25, 2006. Found.

[6] Zhao, Tong, Nehorai, Arye, and Porat, Boaz. "K-Means Clustering-Based Data Detection and Symbol-Timing Recovery for Burst-Mode Optical Receiver." IEEE Transactions on Communications.Vol. 54. No 8. August 2006. 1492-1501.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[7] Flexer. Statistical evaluation of neural networks experiments: Minimum requirements and current practice. In Proceedings of the 13th European Meeting on Cybernetics and Systems research, volume 2, pages 1005–1008, Vienna, Austria, 1996.
[8] Azimi-sadjadi, M.R. Sheedvash, S. Trujillo, F.O. Recursive dynamic node creation in multilayer neural networks, IEEE transactions on Neural Networks issue date: mar 1993 Vol 4(2) on page(s): 242 - 256 ISSN: 1045-9227 cited by: 5 inspec accession number: 4415954 doi: 10.1109/72.207612.
[9] Dundar, G. Rose, K. Dept. of Electr. & Electron. Eng., Bogazici Univ., Istanbul, "The effects of quantization on multilayer neural networks", IEEE Transactions on Neural Networks Issue Date: Nov 1995 Vol6(6), On page(s): 1446 - 1451 ISSN: 1045-9227 Cited by: 17 INSPEC Accession Number: 5110564 Digital Object Identifier: 10.1109/72.471364.
[10] Meisner, B. N. Correlation of National Fire Danger Rating System Indices and Weather Data with Fire Reports. Forest Fire Laboratory, Pacific Southwest Research Station, Riverside: USDA Forest Service; 1993; Final Report, Interagency Agreement No. R500A.2002.
[11] P. Cortez. Rapid Miner: Data Mining with Neural Networks and Support Vector Machines using Rapid Miner. In R. Rajesh (Ed.), Introduction to Advanced Scientific Soft wares and Toolboxes, In Press.
[12] Kivinen J, Smola A, Williamson R (2004). \Online Learning with Kernels." IEEE Transactions on Signal Processing, 52.URL http://mlg.anu.edu.au/~smola/ papers/ KivSmoWil03.pdf.
[13] J. Zhao, C. Lu, and Y. Kou. Detecting Region Outliers in Meteorological Data. In Proceedings of the 11th ACM International Symposium on Advances in Geographic Information Systems, pages 49–55, New Orleans, USA, 2003.

AUTHOR PROFILES:

N.Naga Saranya, currently pursuing her Ph.D. degree Under the guidance of Dr. M.Hemalatha, Head, Dept of Software Systems, Karpagam University, Tamilnadu, India. One year of research experience and she published four international journals and also presented four papers in various national and international conferences. Area of interest is Data Mining.

Dr..M.Hemalatha completed MCA M.Phil., PhD in Computer Science and currently working as a Asst Professor and Head, Dept of Software Systems in Karpagam University. Eleven years of Experience in teaching and published sixty papers in International Journals and also presented seventy papers in various National conferences and international conferences. Area of research is Data Mining, Software Engineering, Bioinformatics, and Neural Network. Also reviewer in several National and International journals.