Media Statistika

p-ISSN 1979 – 3693 e-ISSN 2477 – 0647
MEDIA STATISTIKA 14(2) 2021: 118-124
http://ejournal.undip.ac.id/index.php/media_statistika

PREDICTION OF FOREST FIRE USING NEURAL NETWORKS WITH BACKPROPAGATION LEARNING AND EXREME LEARNING MACHINE APPROACH USING METEOROLOGICAL AND WEATHER INDEX VARIABLES

Dedi Rosadi[1*], Deasy Arisanty[2], Dina Agustina[3]
[1]Department of Mathematics, Gadjah Mada University, Indonesia
[2]Department of Geography Education, Lambung Mangkurat University, Indonesia
[3]Department of Mathematics, Padang State University, Indonesia

E-mail: dedirosadi@gadjahmada.edu

DOI: 10.14710/medstat.14.2.118-124

Article Info:
Received: 10 December 2020
Accepted: 28 December 2021
Available Online: 11 Januari 2022

Keywords:
Forest fire prediction, neural networks, backpropagation, extreme learning machine
.

Abstract: Forest fire is one of important catastrophic events and have great impact on environment, infrastructure and human life. In this study, we discuss the method for prediction of the size of the forest fire using the hybrid approach between Fuzzy-C-Means clustering (FCM) and Neural Networks (NN) classification with backpropagation learning and extreme learning machine approach. For comparison purpose, we consider a similar hybrid approach, i.e., FCM with the classical Support Vector Machine (SVM) classification approach. In the empirical study, we apply the considered methods using several meteorological and Forest Weather Index (FWI) variables. We found that the best approach will be obtained using hybrid FCM-SVM for data training, where the best performance obtains for hybrid FCM-NN-backpropagation for data testing.

1. INTRODUCTION

Forest fire is one of important catastrophic events and have great impact on environment, infrastructure and human life. In literature, for forest fire modelling, in particular for early warning detection system, there are various methods available, which can be categorized into physics-based mode, statistical model and machine learning model, see e.g., Buia et.al, (2017). For empirical modelling, there are various types of data can be used, for instance, using satellite data, using infra-red/fog or fire detector, using data from various ground sensors, such as the weather and meteorological data, see e.g. Cortez and Morais (2007).

Among all, Machine Learning (ML) approach is one of the most popular methods used in the literature. For instance, Cortez and Morais (2017) applied the classification analysis based on several machine learning approach, i.e. Multiple Regression (MR), Decision trees (DT), Random Forests (RF), Neural Networks (NN) and Support Vector Machines (SVM) for forest fire prediction using several meteorological and Forest Weather Index (FWI) variables. Using the same data, Shidiq and Mustofa (2014) applied hybrid approach between the clustering and the classification methods, i.e., applied cosine-distance Fuzzy C-Means (FCM) clustering together with NN with backpropagation learning classification. Using this hybrid approach, it shows that it will provide more accurate classification than using classical approach (SVM, K-Nearest Neighbor/KNN, DT, and Naive Bayes/NB) either for data with or without classification. In Rosadi and Andriyani (2020), it was applied hybrid FCM and ensemble Adaptive Boosting

(AdaBoost) classification and using the same data, it was found better performance than using methods considered in Cortez and Morais (2007) and Shidiq and Mustofa (2014). In this study, we consider the modelling of the same data as Rosadi and Andriyani (2020), however we consider hybrid approach between FCM clustering and NN classification, where we consider method *backpropagation learning* and *extreme learning machine* for NN optimization algorithm. For comparison purpose, we apply the same FCM clustering with SVM classification.

The rest of this paper organized as follows. In Section 2 we provide short summary of some methods considered in the study. In Section 3, we provide the algorithm used in the study, and in Section 4, we discuss the empirical results. In the last section, we conclude our study discussed in this paper.

2. LITERATURE REVIEW

2.1. Clustering Method: Fuzzy C- Means Clustering

The Fuzzy C-Means (FCM) clustering (see e.g., Bezdek, 1981) is an extension of the classical k-mean clustering approach. The FCM algorithm will cluster *n* elements, denotes as $X = \{\mathbf{x}_1, \cdots, \mathbf{x}_n\}$ into *m* fuzzy clusters. Given the data, using FCM algorithm, we obtain *m* centres of the clusters, denotes as $C = \{\mathbf{c}_1, \cdots, \mathbf{c}_m\}$ and the matrix $W = w_{ij} \in [0,1], i = 1, \cdots, n, j = 1, \cdots, m$, where each w_{ij} shows degree of memberships for each of element $\mathbf{x}_i$ on each of cluster $\mathbf{c}_j$. Detail algorithm can be found in, e.g., Chattopadhyay, Pratihar and Sarkar (2011).

2.2. Classical Classification Approach: Support Vector Machine

Support Vector Machine (SVM) firstly introduced by Boser, Guyom and Vapnik (1992). SVM is one of the supervised learning methods and can be used either for classification of for regression analysis. The SVM has been applied in various application such as hand writing analysis, face recognition, pattern classification, etc. In the SVM analysis, to classify the data, it will be defined the hyperplane in the data space of dimension N (where N denotes the number of variables or features) which will classify the data. In the SVM algorithm, the goal is to obtain the area which will have the maximum margin, i.e., the maximum distance between point within the classes. This algorithm can be used for data which is linearly separable or not. We provide a short summary of the method for the data which is linearly separable below.

The hyperplane for linearly separable data can be written as

$$g(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \mathbf{w}_0$$

Assume

$$g(\mathbf{x}) \geq 1, \forall \mathbf{x} \in class\ 1$$

and $g(\mathbf{x}) \leq -1, \forall \mathbf{x} \in class\ 2$

Next, we calculate the margin z for each support vector

$$z_+ = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}, \forall \mathbf{x} \in class\ 1, z_- = \frac{|g(\mathbf{x})|}{\|\mathbf{w}\|} = \frac{1}{\|\mathbf{w}\|}, \forall \mathbf{x} \in class\ 2$$

Therefore, we obtain the total margin *z* as

$$z = \frac{1}{\|\mathbf{w}\|} + \frac{1}{\|\mathbf{w}\|} = \frac{2}{\|\mathbf{w}\|}$$

where $\mathbf{w}$ denote the weight vector.
The weight $\mathbf{w}$ can be obtained by using constrained nonlinear optimization approach, e.g., using KKT (Karush-Kuhn-Tucker) approach. Define the weight vector as

$$\mathbf{w} = \sum_{i=0}^{N} \lambda_i y_i \mathbf{x}_i$$

Minimization of the weight is applied using the KKT condition

$$\sum_{i=0}^{N} \lambda_i y_i = 0$$

See e.g., Christmann and Steinwart (2008) for further detail of the SVM approach.

2.3. Extreme Learning Machine approach

Extreme Learning Machine (ELM) is the learning algorithm for a Single-Layer Feed-Forward Neural Network (SLFN) introduced in Huang, Zhu and Siew (2004). In ELM, the optimal tuning parameter will be obtained using one step approach. It is known that this algorithm has very fast learning process and will have a good performance, as long as the number of neurons in the hidden layer is sufficiently large and the number of available data for tuning the parameter is sufficient.

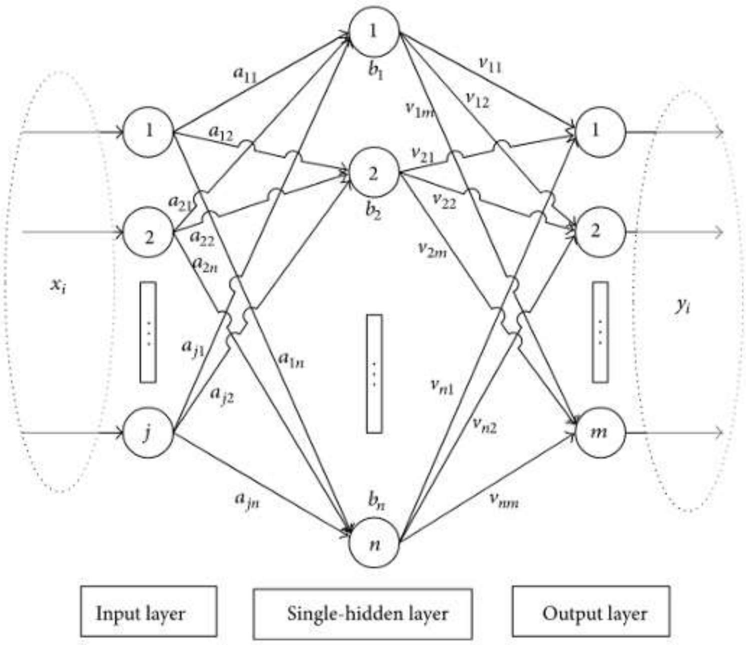

Figure 1. SLFN (Single Hidden Layer Feedforward Neural Network)

ELM model can be written as

$$F_L(x) = \Sigma_{i=1}^{\mathrm{L}} \beta_i g_i(x) = \Sigma_{i=1}^{\mathrm{L}} \beta_i g\left(w_i * x_j + b_i\right), j = 1, \ldots, N$$

where:

L denotes the number of neurons in hidden layer
N denotes the number of training data

β denotes weight vector between the hidden layer and the output
w denotes weight vector between the input and the hidden layer
g denotes the activation function
b denotes the bias, and
$\boldsymbol{x}$ denotes the vector of input data

This equation can be written as

$$\boldsymbol{T} = \boldsymbol{H\beta}$$

Where

$$\boldsymbol{H} = \begin{bmatrix} g(w_1 * x_1 + b_1) & \cdots & g(w_L * x_1 + b_L) \\ \vdots & \ddots & \vdots \\ g(w_1 * x_N + b_1) & \cdots & g(w_L * x_N + b_L) \end{bmatrix}_{N\times L}$$

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_L^T \end{bmatrix}_{L\times m} \quad \boldsymbol{T} = \begin{bmatrix} t_1^T \\ \vdots \\ T_N^T \end{bmatrix}_{N\times m}$$

m denotes the number of outputs
$\boldsymbol{H}$ denotes the *Hidden Layer Output Matrix*
$\boldsymbol{T}$ denotes the matrix of training data

ELM algorithm works as follows

1. First, generate randomly the weight w_i and the bias b_i, $i = 1, \dots, L$
2. Calculate the output from hidden layer
3. Multiply the data training T with the transpose of matrix random weight

$$\boldsymbol{H} = \boldsymbol{T}.\boldsymbol{W^T}$$

4. Choose the activation function. For our study, we apply the Relu function, i.e., g(x)=max(0,x)
5. Calculate the weight vector

$$\widehat{\boldsymbol{\beta}} = \boldsymbol{H}^+.\boldsymbol{T}$$

6. Here we may apply the Moore-Penrose pseudo inverse, i.e.

$$\boldsymbol{H}^+ = (\boldsymbol{H^T}.\boldsymbol{H})^{-\boldsymbol{1}}.\boldsymbol{H^T}$$

7. Use $\hat{\beta}$ to obtain the output

$$\boldsymbol{T} = \boldsymbol{H}.\widehat{\boldsymbol{\beta}}$$

2.4. Backpropagation Learning algorithm

Backpropagation learning is the learning algorithm for obtaining the optimum weight of Multi-*layer Perceptron neural networks*. The steps of *backpropagation* are given as follows

1. Error calculation. Calculate the difference between the output with the real data
2. Minimum Error. Check whether the error has been minimum or not
3. Updating parameter. If the error is large, update the parameter iteratively until the minimum error is obtained. In this step, we may apply the *gradient descent* approach or delta method.
4. Prediction using the model. When the minimum error is reached, it will be used for prediction or other modelling aspects.

See for instance, Rumelhart and Hinton (1986) for further detail.

3. RESEARCH METHOD

3.1. Description of Data

For empirical study we use the same data as Cortez and Morais (2007), and it is available in the UCI *machine learning repository*. The number of cases is 517, from year 2000 until 2007. The original data contains 12 variables, which are meteorological and *forest weather index* (FWI) variables. As Shidiq and Mustofa (2014) and Rosadi and Andriyani (2020), here we use 8 variables only in the study, namely *FFMC, DMC, DC, ISI, Temperature, RH, Wind* and *Rain*.

3.2. The Algorithm

The algorithm for modeling can be described as follows

Data Processing

1. Firstly, the data is classified into two categories, namely the data with the variable Area has the value 0 (it is labelled as "non burned area") and the data with the variable Area has the value larger than 0 (it is labelled as "burned area")
2. Apply the min-max normalization to the data

$$v_i' = \frac{v_i - \min A}{\max A}(\mathrm{newm}\,\mathrm{ax}\, A - \mathrm{new}\min A) + \mathrm{new}\min A$$

   where the interval [0,1] is used for the value of [new min *A*, new max *A*].

Clustering step

3. The "burned area" then clustered into two clusters by applying the fuzzy c-means algorithm. The cluster names are "not heavily burned" and "heavily burned". Here we only used the cosine similarity measure, which already showed will give optimal distance for clustering this data, see e.g. Shidiq and Mustofa (2014). For the computation, we apply the function *fcm* in the package *ppclust* (see Cebezi, 2019) of R (R Core Team, 2021)
4. Combine the results of clustered data from step 3 with the data from "non burned area".

Classification step

5. Data is randomly splitted into training (70 and 80 percent) and testing data (30 and 20 percent)
6. We apply the considered classification method to the training and testing data. Here we apply the SVM, NN-backpropagation algorithm and NN-ELM. Here we use the function svm in the package *e1071* (Meyer, 2020), the function *elm_train* in the package *elmNNRcpp* (Gosso, 2012; Huang, Zhu, Ding and Zhang, 2011; Mouselimis, 2021) and the function *neuralnet* in the package *neuralnet* (Fritsch, et.al., 2019) ,
7. To check the performance of the methods, we calculate the accuracy measure, which is define as

$$\text{Accuracy}=\frac{\text{TP + TN}}{\text{TP + TN + FP + FN}}$$

   where TP, TN, FP, FN denotes the number of cases *true positive, true negative*, *false positive* and *false negative* in the classification of the categorical data.

4. RESULTS AND DISCUSSION

The summary of the empirical study is given in Table 1. In the study we provide the results for the data training and data testing size (80%:20%) and (70%;30%). For classification we consider the SVM and NN with backpropagation and ELM method. From Table 1, it can be seen that for training data, SVM will provide the best accuracy. For data testing, we obtain the best performance is obtained for NN with backpropagation with the accuracy more than 82%. From this empirical study, we in general can see that the hybrid approach between fuzzy c-means clustering and NN with backpropagation learning will provide a good method for prediction the size of the forest fire.

Table 1. Accuracy of the Method

Algorithm	The Ratio Between Data Training and Data Testing	Accuracy for Data Training	Accuracy for Data Testing
NN-ELM	8:2	47.82%	47.57%
	7:3	47.51%	48.38%
NN-Backpropagation	8:2	83.81%	82.52%
	7:3	83.70%	83.22%
SVM	8:2	86.47%	82.52%
	7:3	86.74%	-

5. CONCLUSION

One of the key success for controling the forest fire is the early detection of the fire. In this paper we consider the hybrid algorithm between the clustering and classification for forest fire prediction where we apply the method using several meteorological and *forest weather index* (FWI) variable. From empirical study we found that NN with backpropagation learning will provide a good method for prediction the size of the forest fire. To obtain the best method for the particular data, we suggest to study various combination of the hybrid classification and clustering.

ACKNOWLEDGEMENT

The authors acknowledge the research funding from PD DIKTI 2021.

REFERENCE

Bezdek, J. C.1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, NY.

Boser, B., Guyon, I. and Vapnik, V. 1992. A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, Pittsburgh., pp. 144-152

Buia, D.T., Buib, Q.-T., Nguyenc, Q.-P., Pradhand, B., Nampak, H. and Trinh, P. 2017. A Hybrid Artificial Intelligence Approach Using GIS-Based Neural-Fuzzy Inference System and Particle Swarm Optimization for Forest Fire Susceptibility Modeling at A Tropical Area. *Agricultural and Forest Meteorology*, Vol. 233, pp. 32-44

Cebeci, Z. 2019. Comparison of Internal Validity Indices for Fuzzy Clustering. *Journal of Agricultural Informatics*, Vol. 10, pp. 1-14.

Chattopadhyay, S., Pratihar, D. K. and Sarkar, S. C, D, 2011. A Comparative Study of Fuzzy C-Means Algorithm and Entropy-Based Fuzzy Clustering Algorithms. *Computing and Informatics*. Vol. 30 , pp. 701

Christmann, A. and Steinwart, I., 2008, *Support Vector Machines*, Springer-Verlag, New York

Cortez, P. and Morais, A. 2007, A Data Mining Approach to Predict Forest Fires usingMeteorological Data. In: Neves J, Santos M F, Machado J (eds.) *EPIA 2007*, pp.512-523.

Fritsch, S., Guenther, F., Wright, M.N., Suling, M. and Mueller, S.M.. *Neuralnet: Training of Neural Networks,* R package version 1.44.2. CRAN , Vienna, Austria.

Gosso, A. 2012. *elmNN: Implementation of ELM (Extreme Learning Machine) algorithm for SLFN (Single Hidden Layer Feedforward Neural Networks)*. CRAN , Vienna, Austria

Huang, G-B., Zhu, Q-Y. and Siew, C. 2004. Extreme Learning Machine: A New Learning Scheme of Feedforward Neural Networks. *IEEE International Joint Conference on Neural Networks* 2004 (IEEE Cat. No.04CH37541)

Huang G-B., Zhou, H., Ding, X., Zhang, R. 2011. Extreme Learning Machine for Regression and Multiclass Classification. *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 42, pp. 513–529.

Meyer, D., 2020, Support Vector Machines* The Interface tolibsvmin package e107 FH Technikum Wien, Vienna, Austria

Mouselimis, L. 2021. *elmNNRcpp: The Extreme Learning Machine Algorithm*. R package version 1.0.3, CRAN , Vienna, Austria

R Core Team, 2021. R: A Language and Environment for Statistical Computing, R Foundation for Statistical, Vienna.

Rosadi D. and Andriyani, W. 2020. Prediction of Forest Fire using Ensemble Method, Presented in ICMSE 2020, Semarang, Indonesia, 6 Oktober 2020

Rumelhart, D.E. Hinton, G.E. and Williams, R. J. 1986. Learning Representations by Back-Propagating Errors. *Nature*, Vol. 323, pp. 533-536.

Shidik, G. F. and Mustofa, K. 2014. Predicting Size o Forest Fire Using Hybrid Model. *Proceeding ICT EurAsia Conference 2014*, pp 316-327