An Integrated Approach to the Regression Problem in Forest Fires Detection

I. S. Sineva, *Member, IEEE*
Moscow Technical University of Communications and Informatics
Moscow, Russia
iss@mtuci.ru

M. D. Molovtsev
Moscow Technical University of Communications and Informatics
Moscow, Russia
mishamolovtsev@gmail.com

***Abstract*-Machine learning procedures are used in almost humans' life areas: recommendations systems for online-services, self-driving cars, machine translation systems and so on. But machine learning tasks complexity and importance are grown up and solutions based on a single algorithm can have bad quality. The complex approach was developed for burned area estimating on the forest fires detection problem. The developed approach was tested on the data of Montensinho park. In this paper developed approach is compared with classic regression algorithms and its advantages are shown. The significantly better quality of the proposed solution is demonstrated.**

Keywords—machine learning; regression; forest fires; linear regression; random forest

I. Introduction

With the active development of information technology and mathematical application, the machine learning methods usage is becoming more and more widespread: recommendation systems for online-services [1, 12, 13], medical applications [2], self-driving cars, security systems and so on.

With the high demand for solving problems by machine learning methods, software and hardware capabilities for this are also growing. For example, a large number of libraries with machine learning methods have been developed for almost every programming language. These libraries implement almost all the most popular methods of machine learning, as well as additional procedures for data preprocessing, metric evaluation, visualization and so on.

Also implemented algorithms are becoming more and more optimized, which allows you to run them in the absence of large computing resources such as mobile phones.

All this allows machine learning specialists to solve and automate a variety of issues, the solution of which was previously impossible for various reasons.

One of the most actual issues where machine learning could do well is forest fires prediction. Forest fires are a serious environmental problem that causes serious economic damage and poses a great danger to human life. It is possible to put out big fires in 2010 in the European territory of the Russian Federation and in 2012 in territories of the Siberian and Far East Federal district. Such natural disasters have contributed to the scale of the national disaster. Thus, there is a need to detect forest fires in the early stages.

II. Source Data

Data from the national Park of Montesino, located in the North of Portugal, were taken as source data [3].

Historical data relating to the period from January 2000 to December 2003 were collected for the study. The data contains the coordinates and the fire source, the date, the meteorological sensor readings: relative humidity, wind speed, air temperature, the amount of precipitation, as well as the indices of the Canadian forest fire hazard assessment system and the burned area size.

The [4] presents the first solution to the problem forest fires prediction problem. The support vector machine was chosen as the final method.

In [5] the classification problem was solved for this data and it can be used as a part of integrated approach for burned area size prediction.

III. Regression Problem

Regression problem is one of the basic machine learning problems and its' solutions can be used in different ways:

1. Prediction of some variable using other variables. A special case is the problem of time series prediction.

2. Construction of function $y = f(x)$. This will determine the type of relationship between the dependent variable (predictor) and the independent (factors).

3. Determining the contribution of individual independent variables to predictor variation

In the forest fires prediction problem regression algorithms are uses for burned area size prediction. it is necessary to learn not only that there was a fire, but also how big the fire center is. This will help emergency responders assess their next actions and the resources needed to minimize damage.

In [6, 7] the possibility of using fuzzy logic methods is considered. The main idea of the work is to build a fuzzy rules system which help to predict the buried are size. These rules are generated using genetic algorithms. But it doesn't prediction burned area directly. It predicts some fuzzy vector of membership to 3 classes: low, medium and high. Using this

978-1-7281-4772-7/20/$31.00 ©2020 IEEE

fuzzy vector, burned area is calculated. The results quality is evaluated using MAE and RMSE. It provides to compare this solution with developed approach.

In this paper new approach for regression problem is presented.

This approach consists of 2 algorithms used sequentially:

1. Random forest regressor
2. Elastic net

The quality of this sequence was checked using MAE and RMSE metrics and was compared with quality of most popular regression algorithms

A. Random Forest

In order to avoid confusion between the concepts of forest and trees from the context of the problem and the terminology of the algorithm, the latter will be written in quotation marks.

This algorithm is one of the most popular algorithms in classification and regression problems. It based on ensemble method called bagging (bootstrap aggregation).

The main bagging idea is as follows: N subsamples with repetition are constructed and basic algorithms are independently trained on each of them.

Random forest basic algorithm is decision "tree".

The structure of the tree is "leaves" and "branches". The edges ("branches") of the decision tree contain the attributes on which the target function depends, the "leaves" contain the values of the target function, and the other nodes contain the attributes, which differ in cases. To predict a new case, you need to go down the tree to the sheet and give the appropriate value [8, 9].

The main method idea is to divide the set of training objects into areas with minimal MAE(mean absolute error).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (1)$$

Where n – observation count,

y_i - actual burned area,

$\hat{y}_i$ - predicted burned area.

There are several specificities of random forest:

1. It constructs N subsamples with repetitions. Size of subsample can be determinate by the expression:

$$N(1-1/N)^{\frac{1}{N}}, \quad (2)$$

where N - training set size.

2. For every "tree" select m random features. The number of selected features can be determinate by expression:

$$m \approx \sqrt{M}, \quad (3)$$

where M - features count.

The final result of random forest prediction can be determinate by expression:

$$a(x) = \frac{1}{n}\sum_{i=1}^{N} a_i(x), \quad (4)$$

where $a_i(x)$ - i-th "tree" solution.

B. Elastic Net

In regression problems random forest has only one disadvantage: it works badly when there is any linear area (figure 2). This happens because random forest is a piecewise constant function [10].

The linear regression model determinates in common case by expression:

$$y = \sum_{i=1}^{n} x_i\beta_i + \beta_0 + \varepsilon \quad (5)$$

where y - dependent variable(in forest fires detection problem it is residuals of random forest),

x_i - i-th independent variable,

β - models' parameters,

ε - error.

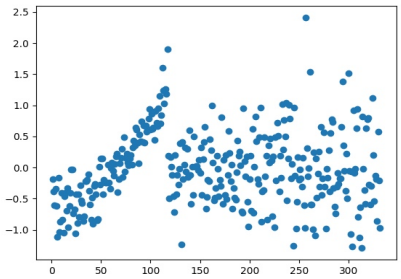

Fig. 1. Residuals of random forest

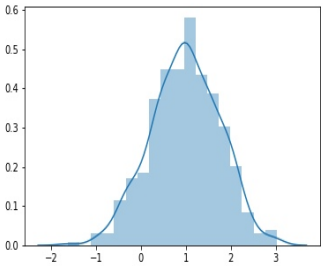

Fig. 2. Random forest residuals' distribution

Model can be learned using gradient descent or its' modifications: stochastic gradient descent or mini batched gradient descent.

Also linear regression mode can be learned using least squares method in matrix form:

$$B = (X^T X)^{-1} X^T Y \quad (6)$$

where Y - burning area vector,

X - observations' matrix,

B - models parameters vector.

Sometimes modified linear regression algorithms are used to combat retraining and selection of the most important features. These modifications are Ridge and Lasso regression.

Ridge regression uses L2 regularization to prevent overfitting. And it can be learned in matrix form:

$$B = (X^T X + \lambda I)^{-1} X^T Y \quad (7)$$

where λ - regularization coefficient,

I - identity matrix with same size as observations' matrix.

But elastic net uses also L1 regularization that doesn't have solution in matrix form.

In elastic net the follow cost function should be minimized[11]:

$$J(\beta_0, \beta_1, ..., \beta_k) = \frac{1}{2}\sum_{i=1}^{n}(y_i - \sum_{j=1}^{k} x_{i,j}\beta_j - \beta_0)^2 + \lambda_1 \sum_{j=1}^{k}|\beta_j| + \lambda_2 \sum_{j=1}^{k}\beta_j^2 \quad (7)$$

where y_i - i-th observations' burned area,

$x_{i,j}$ - j-th parameter of i-th observation,

λ_1 - regularization coefficient for L1 regularization,

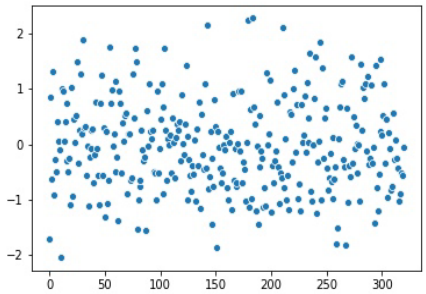

Fig. 3. Residuals of elastic net

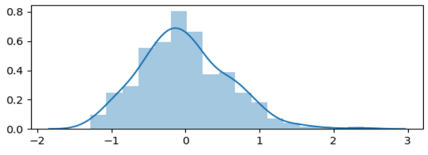

Fig. 4. Random forest residuals' distribution

λ_2 - regularization coefficient for L2 regularization.

If λ_1 equals to 0, elastic net became a ridge regression. If λ_2 equals to zero, elastic net become a lasso regression.

The main advantage of elastic net model is that it can use all advantages of ridge and lasso regression:

1. L2 regularization prevents over-training of the model by prohibiting disproportionately large weights.

2. L1 implements this by selecting the most important factors that have the strongest impact on the result.

After this function minimization is needed to check distribution of ε. Values of ε should be random. If their distribution is normal with mean equal 0, everything is good and it is possible to use a lot of statistic tests for model analysis.

IV. Results

For algorithms comparison were chosen following metrics:

1. MAE – mean absolute error

2. RMSE – root mean-squared error

$$RMSE = \frac{1}{n}\sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (8)$$

Developed approach is compared with the most popular regression algorithms:

1. Linear regression (LR)

2. Support vector machine (SVM)

3. Decision tree (DT)

4. Random forest (RF)

Also cause of data has categorical value (month of observation) there will used different encoding methods:

1. One hot encoding. It transforms a categorical feature with n values into n-dimensional binary vector. If categorical feature equals k, after transformation this observation will have 1 on k-th position and 0 on others.

2. Mean target encoding. It replaces every value of categorical feature with mean target value for this value.

All encodings are trained on train set and if they will be used on value they never seen they will ignore it. It means that one hot encoding will return zeros vector for this value, and mean target encoding will return zero or some default value. For mean target encoding it is better to return some default value (for example -999) because it is possible to get zero as mean target for some categorical feature value.

The resulting metric values can be interpreted as noise parameters, where MAE is the noise Medina and RMSE is the standard deviation.

Based on the results of the analysis of algorithms, it can be

concluded that the use of mean target encoding allows you to slightly reduce the RMSE. This may be due to the fact that the categorical variable has 12 values and turning it into a 12-dimensional binary vector is not the best solution. However, changing the data preprocessing methods does not reduce the value of the metrics to match the preprocessing approach.

TABLE I. COMPARISON OF ALGORITHMS

Algorithm	MAE	RMSE
Development approach	12.3	28.1
LR with one hot encoding	23	63
LR with mean target	22	62
SVM with one hot encoding	13	63
SVM with mean target	13	61
Decision tree with one hot encoding	25	70
Decision tree with mean target	19	69
Random forest with one hot encoding	20	66
Random forest with mean target	21	65
Algorithm from original work	12.71	58.3
Fuzzy rules system	15.13	50.3

V. CONCLUSION

Based on previous paper full approach for forest fires detection problem were developed. This approach based in several sequentially used machine learning methods that solve different tasks. It provides better quality compared to solution based on only one algorithm.

The approach consists of two steps:

1. Classification. At this step algorithm decides is there fire or not. It there is no fire the observation is skipped and its' burned area size marked as 0.

2. If there is a fire algorithm predicts its' size using regression algorithms sequel that give us better prediction quality.

Every step was validated and compared with the most popular and appropriate for this data machine learning methods.

The developed approach can be used not only to solve the problem of forest fire detection and assessment of the fire source. One example of another approach application can be floods. As another frequent and dangerous natural disaster, building a system to predict them can work in a similar way:

1. Flood location prediction.

2. Destruction area estimation.

However, it is important to understand that a system trained on fire data cannot be applied to data for floods (or any other natural disasters). There are several reasons for this:

1. It is necessary to distinguish different patterns. Because in fires, climate indicators behave in one way, and in floods in another. Thus, when applying a system trained on forest fire data to flood data, the system will recognize fire patterns rather than flood patterns.

2. Perhaps need different data. For the forest fire prediction was enough climatic data: temperature, precipitation, humidity, wind speed, and so on. These data may not be sufficient to predict flooding and may need data on terrain and nearby water sources. And maybe the existing data will be completely useless and need fundamentally different data.

But here the approach itself is important, which does not depend on the data and can show good results in other tasks.

REFERENCES

[1] Segaran T. *Programming Collective Intelligence: Building Smart Web 2.0 Applications.* Paperback – August 26, 2007.

[2] Molovtsev M.D., Sineva I.S. Usage of data mining techniques for isolation of trembling hyperkenesis in the diagnosis of Parkinson's disease. *Telecommunications and information technology*, 2018, No.1, pp. 89-94. (In Russian)

[3] Forest Fires Data Set, available at: http://archive.ics.uci.edu/ml/datasets/Forest+Fires. (Accessed 18 February 2019).

[4] P. Cortez, A. Morais. A Data Mining Approach to Predict Forest Fires using Meteorological Data. In J. Neves, M. F. Santos and J. Machado Eds., New Trends in Artificial Intelligence. *Proceedings of the 13th EPIA 2007 – Portuguese Conference on Artificial Intelligence,* December, Guimarães, Portugal, pp. 512-523, 2007.

[5] Molovtsev Mikhail D., Sineva Irina S. Classification Algorithms Analysis in the Forrest Fires Problem. *International Scientific and Practical Conference "Quality Management, Transport and Information Security, Information Technologies",* Sochi, Russia, September 23-27, 2019, pp. 548-553.

[6] Fowler, A., Teredesai, A. M., & De Cock, M. The 28th North American Fuzzy Information Processing Society Annual Conference (NAFIPS2009) Cincinnati, Ohio, USA - June 14-17, 2009.

[7] Pourghasemi, H. reza, Beheshtirad, M., & Pradhan, B. A comparative assessment of prediction capabilities of modified analytical hierarchy process (M-AHP) and Mamdani fuzzy logic models using Netcad-GIS for forest fire susceptibility mapping. Geomatics, Natural Hazards and Risk, 7(2), 861-885, 2014

[8] V'yugin V.V. *Matematicheskie osnovy teorii- mashinnogo obucheniya I prognozirovaniya* [Mathematical foundations of the theory of machine learning and forecasting]. *MCCME E*, 2018, 384 p.

[9] Flach P. Machine Learning: *The Art and Science of Algorithms That Make Sense of Data.* Cambridge University Press; 1 edition, 2012, 409 p.

[10] Chistyakov S. Proceedings of RAS Kerelian scientific center. *Sluchajnylesa:obzor* [Random forests: review], 2013, no 1, pp. 117-136.

[11] Nikolenko S., Kadurin A., Archangelskaya E. *Glubkoeobuchenie. Pogruzhenievmirneironnyhsetey*[Deep learning. Dive into the world of neural networks] SPB, 2018. 480 p.

[12] A.S. Adzhemov, I.S. Sineva, "Efficiency of genetic-like coding algorithm for metric space sources," *Synchroinfo Journal*, vol. 3, no. 1, pp. 2-5, 2017.

[13] D.A. Yakovlev, I.S. Sineva, "Network anomaly detection based on flow intensity measuring in the decay model for DDOS protection," *T-Comm*, vol. 13, no.1, pp. 41-44, 2019. (in Russian)