Machine Learning Methods Performance Evaluation*

Zakoldaev D. A.[1], Vorobeva A. A.[2]

[1]PhD, associate Professor, Department Computer Systems Design and Security, St. Petersburg National Research University of Information Technologies, Mechanics and Optics, Russia, St. Petersburg
[2]PhD, associate Professor, Department Computer Systems Design and Security, St. Petersburg National Research University of Information Technologies, Mechanics and Optics, Russia, St. Petersburg
[1]d.zakoldaev@itmo.ru

Abstract. In this paper, we describe an approach for air pollution modeling in the data incompleteness scenarios, when the sensors cover the monitoring area only partially. The fundamental calculus and metrics of using machine learning modeling algorithms are presented. Moreover, the assessing indicators and metrics for machine learning methods performance evaluation are described. Based on the conducted analysis, conclusions on the most appropriate evaluation approaches are made.

Keywords: machine learning; air pollution modeling; environmental modeling

1. Introduction

Modern society is currently at an active stage of globalization, and the needs of the economic, political,cultural, and other areas are always growing. Growing needsalso entails an increase in the scale of production activities. However,large production processes that are an integral partthe world economy are capable of being sources of environmental pollutionand affect both the health of thenearby territory's population andon the flora and fauna state.

Research in the field of environmental monitoringconducted for many years by scientific communities from different countries.Of particular interest in this area is the developmentmethods for making short-term forecasts of air pollution,the results of which can be used for operationalresponse to industrial emissions and the formation of a managerexposure to prevent the spread of pollution beyonda certain controlled area.

In this paper, we solve the task of controltheharmful substances concentration emitted into the air duringvital activity of a critical facility and organizationtimely response to such emissions. The considered object activityis coaltransportation.

The monitoring module is based on the concept of distributedself-organizing cyber-physical systems, representedas a set of various elements: sensors, data transmission means, computing devices, etc. The main rolehereare played by the system's sensors responsible for collecting meteorological data anddata on individualpollutant concentrations in the atmosphereover thesanitary zone territory.

2. Approach for Machine Learning Methods Performance Evaluation

Below we present and discuss metrics that can be used to assess the effectiveness of various artificial intelligence methods in solving the problem of environmental pollutionpredicting [1-3].

MSE (Mean Square Error). The essence of the MSE estimation method is to calculate the square sum ofactual values deviationsfrom the calculated initial values. However, squaring the magnitude of the deviation significantly increases those values that lie far from all others, or decreases the values of deviations that are between 0 and 1.

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2$$

RMSE (Root Mean Square Error).The advantage of RMSE over MSE is that the order of the estimated values coincides with the magnitude of the error or deviation, however, it is much easier to evaluate the effectiveness of a predictive model based on MSE.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)^2}$$

MAE (Mean Absolute Error) is used to estimate the absolute error N of the prediction results. The undoubted advantage of MAE is that the modules of the deviation magnitude do not multiply the deviations that are considered outliers. Therefore, this estimate is more robust than MSE and corresponds to the median.

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|y_t - \hat{y}_t|$$

The determination coefficient R reflects the percentage of variance. Thedetermination coefficient is used in regression analysis more than in other forecasting methods, therefore it can be used when evaluating extrapolation models. Moreover, it is scale-free. If the model fits the data series perfectly, the R^2 value is 1. If the model does not describe the series at all, but is just a straight line, then the coefficient of determination becomes equal to 0. In cases with nonlinear models, the coefficient can also become negative, but at the same time it is uninterpretable.

$$R^2 = 1 - \frac{MSE(model)}{MSE(baseline)}, MSE(baseline) = \frac{1}{T}\sum_{t=1}^{T}|y_t - \bar{y}_t|$$

However, it is calculated from the training part of the sample, which means it simply shows how well the data is described. However, the accuracy of the description does not guarantee the accuracy of the forecasts. Therefore, this coefficient can be used to assess the adequacy of the model.

MAPE (Mean Absolute Percentage Error) is the mean absolute percentage error. This ratio can be measured in fractions or percentages and be interpreted as a percentage of deviation from the actual values.

$$MAPE = \frac{1}{h}\sum_{j=1}^{h}\frac{|e_{T+j}|}{y_{T+j}},$$

where e_{T+j} is a forecasting step error.

To describe situations ofthe null hypothesis acceptance or rejection within the forecasting tasks, statistical methods are used.Such methods describe first and second type errors, acceptance of a correct null hypothesis and rejection of an incorrect null hypothesis.

The statistical hypothesis is tested and correctly accepted as True Positive if the experimental result is consistent with the null hypothesis. If the null hypothesis is rejected correctly, it is the False Negative hypothesis.

During the hypotheses statistical testing, errors of the 1st and 2nd type may appear. False Positive Error (type I error) means that the null hypothesis was rejected incorrectly, and in the case of False Negative Error (type II errors), the null hypothesis is incorrectly accepted.

In the forecasting module being developed, the level of coal dust concentration is determined in two successive stages:
- classification;
- regression.

The classification defines two value categories - zero and non-zero coal dust concentrations. A zero value means a value significantly less than the resolution of the sensor. In the case of classifying a given observation as non-zero, the second stage - regression - is carried out to determine a specific concentration value.

The following values are used to assess the quality of the classification:
- True Positive - the number of correctly classified objects of the relevant class;
- True Negative - number of correctly classified objects of irrelevant class;
- False Positive - the number of objects of an irrelevant class defined as objects of a relevant class;
- False Negative - the number of objects of the relevant class, defined as objects of the irrelevant class.

Based on these values, the following metrics for assessing the effectiveness of predicting environmental pollution based on artificial intelligence are calculated:
- Accuracy - percentage of correctly recognized relevant and irrelevant objects out of the total number of instances;

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + False\ Positive + True\ Negative + False\ Negative}$$

- Precision - the proportion of relevant objects among recognized objects;

$$Precision = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

- Recall - the proportion of relevant instances that were retrieved from the total number of instances;

$$Recall = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

- F1 is the harmonic mean combining positive predictive value and completeness.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

There are main varieties for Precision, Recall, F1 metrics:

- micro - the absolute values True Positive, True Negative, False Positive, False Negative are used when calculating the specified metric in the micro version;
- macro - when calculating the specified metric in the macro variety, the fractional values True Positive, True Negative, False Positive, False Negative are used in relation to the sample size;
- weighted - when calculating a given metric in the weighted variety, the fractional values True Positive, True Negative, False Positive, False Negative are used in relation to the sample size, taking into account the weight of the given metric by the relative number of examples available for it.

When building classification models for the forecasting module, the F1 macro metric was chosen as a priority due to the following reasons:

- the data sample is not balanced;
- F1 is a metric that combines Precision and Recall metrics.

To assess the quality of the regression, the metrics MSE and MAE were taken into account, however, preference was given to the MAE metric, since the work is carried out mainly with numbers less than one.

3. Conclusion

In this paper we described and discussed machine learning performance evaluation methods related the task of environmental pollution modeling and forecasting. This study is a part of the environmental monitoring module development, and here we presented methods and calculus for forecasting performance evaluation. Such metrics as Mean Square Error, Root Mean Square Error, Mean Absolute Error, R determination coefficient, Mean Absolute Percentage Error were described. To assess the classification performance, provided by the machine learning models, such indicators as Accuracy, Recall, F1, which depend on the number of True Positive, True Negative, False Positive and False Negative classification cases. According to the conducted analysis, Mean Absolute Error were chosen as the most relevant indicator in our case, as numbers less that 1 are prevalent in our project.

4. Acknowledgement

References

1. Bai L. et al. Air pollution forecasts: An overview //International journal of environmental research and public health. – 2018. – Т. 15. – №. 4. – С. 780.
2. Mishra D., Goyal P., Upadhyay A. Artificial intelligence based approach to forecast PM2. 5 during haze episodes: A case study of Delhi, India //Atmospheric Environment. – 2015. – Т. 102. – С. 239-248.
3. Bai L. et al. Air pollution forecasts: An overview //International journal of environmental research and public health. – 2018. – Т. 15. – №. 4. – С. 780.