UDC 681.3

R.M. Alguliyev, F.J. Abdullayeva
e-mail: r.alguliev@gmail.com, a_farqana@mail.ru

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

WEB APPLICATION ANOMALY DETECTION BASED ON LOGISTIC REGRESSION

In this paper, the detection method of anomalies in cloud computing web applications, based on logistic regression is proposed. In this study for the detection of anomalies URLs in "HTTP GET requests" are considered. A URL also termed a web address as well, usually is composed of one absolute path and several query parameters. The values of these parameters typically describe the URLs as a time series.

Logistic regression is one of the most popular classification methods. In this study it modeled as follows: If there are n independent features $y_1, y_2, ..., y_n$ then the conditional probability of web application anomaly detection for logistic regression is given by

$$P = p\left(z = 1 \mid y_1, y_2, ..., y_n\right), \tag{1}$$

$$p = e^z / \left(1 + e^z\right) \; z = \gamma_0 + \gamma_1 y_1 + \gamma_2 y_2 + ... + \gamma_n y_n \tag{2}$$

where γ_i is the coefficient, and $y_1, y_2, ..., y_n$ are features.

The implementation of the method is conducted in the Tensorflow library of Python program package on the "HTTP dataset CSIC 2010". To the input of the model, are given the URL templates of the "HTTP dataset CSIC 2010" dataset. A URLs in the dataset are encoded and incorporate both normal and anomalous queries. 'reqlen', 'pathdepth', 'arglen', 'narg', 'nletterarg', 'ndigitarg', 'notherarg', 'nspecarg', 'nletterpath', 'ndigitpath', 'notherpath', 'sqlpl', 'loginpl', 'typeimg', 'typejs', 'typeother', 'getflag', 'postflag', 'putflag', 'port', 'pllen', 'label' are features of the dataset.

The detection accuracy and test results for various metrics of the proposed logistic regression method for web application anomaly detection are shown in Table 1.

Table 1.

The web application anomaly detection accuracy of logistic regression method

Metod	Precision	Recall	F-measure
Logistic regression	0.88	0.60	0.72

As shown by the results of experiments in Table 1, the logistic regression has produced well results. The confusion matrix of the logistic regression method in web application anomaly detection issue is shown in Table 2.

Table 2.

Confusion matrix of the logistic regression method

True	0	1	Total
0	9365	6120	15485
1	1279	54721	56000

The logistic regression model has detected 9365 points from the 15485 points marked as an anomaly (0) in the dataset as anomaly correctly, in the detection of 6120 points it made a mistake. This model has detected 54721 points from the 56000 points correctly as normal (1), in this class in the detection of 1279 points it made a mistake. This situation can be considered a good result in the classification issue. In addition, good results have been achieved in the detection accuracy of the logistic regression model on the various metrics listed in Table 2. Thus, *Precision, Recall, F-measure* metrics achieved 0.88, 0.60, 0.72 values, respectively.

The ROC curve constructed on the true positive and true negative parameters visually represents the effectiveness of the model better (see figure). It is seen from the figure 1, the ROC curve gradually approximates to 1.

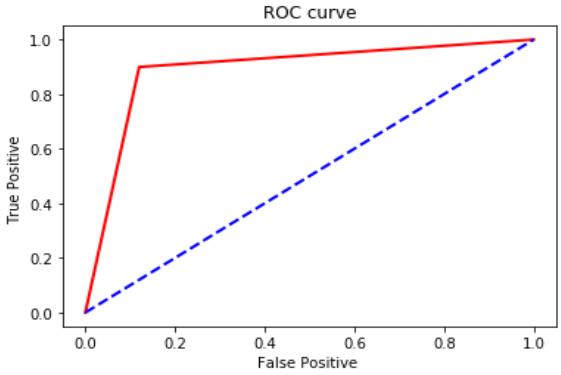

Fig. ROC curve of logistic regression model on “HTTP dataset CSIC 2010” dataset

This work was supported by the Science Development Foundation under the President of the Republic of Azerbaijan – ***Grant No. EIF-KETPL-2-2015-1(25)-56/05/1.***

---

1. Yu J., Tao D., Lin Z. A hybrid web log based intrusion detection model, Proc. of the IEEE 4th international conference on cloud computing and intelligence systems (CCIS), 2016, Beijing, China, pp. 356-360.

2. Zolotukhin M., Hämäläinen T., Analysis of HTTP requests for anomaly detection of web attacks, IEEE 12th international conference on dependable, autonomic and secure computing, 2014, Dalian, China, pp. 406-411.

UDC 004.056.5

R.M. Alguliyev, R.M. Aliguliyev, L.V. Sukhostat
e-mail: r.alguliev@gmail.com, r.aliguliyev@gmail.com, lsuhostat@hotmail.com

Institute of Information Technology, Azerbaijan National Academy of Sciences, Baku, Azerbaijan

PURITY-BASED CONSENSUS CLUSTERING FOR ANOMALY DETECTION IN BIG DATA

The paper proposes a weighted consensus clustering for efficient integration of single clustering methods.