COMPARATIVE STUDY OF FEATURE SELECTION TECHNIQUES RESPECTING NOVELTY DETECTION IN THE INDUSTRIAL CONTROL SYSTEM ENVIRONMENT

Jan Vavra, Martin Hromada

This Publication has to be referred as: Vavra, J[an] & Hromada, M[artin] (2018). Comparative Study of Feature Selection Techniques Respecting Novelty Detection in the Industrial Control System Environment, Proceedings of the 29th DAAAM International Symposium, pp.1084-1091, B. Katalinic (Ed.), Published by DAAAM International, ISBN 978-3-902734-20-4, ISSN 1726-9679, Vienna, Austria
DOI: 10.2507/29th.daaam.proceedings.155

Abstract

The emerging trend of interconnection between business processes and industrial processes resulted in a considerable number of cyber security incidents that show us how vulnerable Industrial Control Systems (ICS) are. These usually legacy systems were not designed with cyber security in mind. Therefore, there is a necessity for the reliable cyber security system. The anomaly detection based on machine learning techniques is the one potential way how to protect the system against cyber-attacks effectively. However, the ICS has become more sophisticated; therefore, produce high-dimensional datasets. Hence, the dimensionality reduction for the dataset is required due to high computational complexity. We introduce the comprehensive study on dimensionality reduction techniques which are applied to ICS network cyber security. Moreover, obtained results are evaluated by novelty detection algorithm where One-Class Support Vector Machine algorithm is used

Keywords: industrial control system; cyber security; anomaly detection; feature selection; support vector machine

1. Introduction

Our contemporary society depends on highly sophisticated Information and Communication Technology (ICT) and Industrial Control Systems (ICS). The interconnection between ICT and ICS resulted in the opening of ICS system to new threats. According to Knapp [1], industrial networks are becoming more attacked. Moreover, these cyber-attacks are more sophisticated, and therefore more fatal.

The ICS is often confused with Supervisory Control and Data Acquisition (SCADA). However, SCADA and Distributed Control Systems (DCS) are main subgroups of ICS. We decided to adopt designation ICS for cyber-physical [2] systems in the article which are commonly implemented in critical information infrastructure (CII). The compromising of integrity, confidentiality and availability of ICS can have serious implication on an economy, human life, and therefore severe impact on the state itself. Ferrag et al. [3] cited that SCADA systems will become more interconnected due to the Internet of Things (IoT). Furthermore, they provide a comprehensive study of significant threats to Smart Grids. They also highlight new research challenges as detecting and avoiding further attacks or IoT-driven Smart Grids. Moreover, Cvitić et al. [4] have done mapping of vulnerabilities and threats to IoT in connection to architecture layers.

They concluded the increasing number of IoT devices will become a challenging task for maintenance and security [4]. Maglaras et al. [5] concluded emerging of new challenges due to the synergy between the ICS and the IoT. They identify main deficiencies in the implementation of cyber security solutions in ICS environments.

The aim of the article is primarily connected with machine learning techniques, especially anomaly-based detection. The main procedure of the research is based on feature selection techniques which are evaluated by the classification algorithm according to commonly used criteria. Additionally, the best result for each feature selected technique is selected according to multi-criteria evaluation. Thus, it is one of the possible ways how to ensure cyber security of ICS systems. Chandola et al. [6] described trends and applications of anomaly detection systems in a considerable number of fields. Moreover, this highly cited survey provides a complex review of anomaly detection techniques and identify their advantages and disadvantages. In addition, we decided to adapt one-class classification technique more specifically One-Class Support Vector Machine (OCSVM). This algorithm can be classified as novelty anomaly detection also known as outlier anomaly detection or semi-supervised anomaly detection. Maglaras and Jiang [7] discussed the possible solution of machine learning techniques where OCSVM was selected as the best-suited choice. Furthermore, Raczko and Zagajewski [8] identify Support Vector Machine (SVM) as a classifier algorithm which is best for complex classification problems and conclude SVM more suitable to implement in large systems in comparison to Artificial Neural Network (ANN) and Random Forest (RF). Omer et al. [9] demonstrated high classification capabilities of SVM even better than ANN.

The feature selection is one of the main challenges for classifiers. Moreover, high dimensional datasets have become the serious problem especially for the highly complex system like SCADA. Thus, there is a significant demand to reduce the dimensionality of the dataset. Dash and Liu [10] discusses a considerable number of feature selection techniques which are commonly used in real-world classification tasks. Moreover, they also described basic concept of elimination or selection of irrelevant features. The highly cited study [11] created by Guyon and Elisseeff pointed out strengths and weaknesses of different feature selection techniques. The techniques were tested on varied datasets with the different number of variables.

The rest of the article is organized as follows. In the section, II has described principles of SVM algorithm. Section III is dedicated to feature selection techniques. The experimental setup is specified in section IV. The section V shows the results of the experiment and section VI presents a conclusion of the article.

2. Support Vector Machine

This section is dedicated to the definition of SVM and OCSVM. SVM was created by Vladimir Vapnik and published in the publication the nature of statistical learning theory [12]. Moreover, the SVM can be included into supervised classification algorithms. The OCSVM is a specific example of SVM which is commonly used for the binary classification task. The basic idea of SVM is to create the widest margin near the boundary between two sets of data. Additionally, the separation vectors between two groups of data are usually called as hyperplane where is essential to maximize the margins. The real example of separation by hyperplane can be shown in fig. 1, where the hyperplane is represented by dashed and solid lines and the data for the different class are represented by asterisks and circles. The supervised learning algorithms is based on the dataset of examples $x_i \in X$ and the labels $y_i \in Y$. However, there are two states in novelty anomaly detection. The system can distinguish between normal operation of the system and the anomalies within the system.

Fig. 1. The representative example linear hyperplane in SVM algorithm [13]

The linear hyperplane is calculated with the intention of the maximization of the margin space between two different datasets. However, the paper is based on a non-separable dataset where the “slack variable” represented by ξ is [14]. The equation for separation of non-separable dataset is presented in (1).

$$f(\bar{x}) - \xi = \bar{w}\bar{x} + b \quad (1)$$

The main boundary is also known as hyperplane which is defined as $f(\bar{x}) = \{0\}$ with the boundary for positive examples $f(\bar{x}) = \{1\}$ and the boundary for negative examples $f(\bar{x}) = \{-1\}$. To optimize the SVM classification capabilities, we need to maximize the width of hyperplane defined as $\max \frac{2}{\|w\|}$. However, the paper is based on a nonlinear separation which is applied to the collected dataset. Therefore, it is appropriate to transform data into higher dimensional space where the data are separable (2) [13]. Thus, we use kernel function shown in equation (3).

$$\Phi: R^d \rightarrow \mathcal{H} \quad (2)$$

$$K(x_i, x_j) = (\Phi(x_i) \cdot \Phi(x_j)) \quad (3)$$

In the case of OCSVM, the data are separated from the origin in feature space by a hyperplane according to equation (4).

$$f(x) = \sum_{i=1}^m \alpha_i k(x_i, x) - \rho \quad (4)$$

The decision function (4) is used to separate the data from anomalies. This separation is implemented by kernel function $k(x_i, x)$. Furthermore, we choose radial basis function (RBF) in order to solve nonlinear separation problem. [13] The RBF kernel function is represented by equation (5).

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2), \quad \gamma > 0 \quad (5)$$

Where x_i represents data points, x represents landmark and γ is a gamma parameter for SVM. Gamma is the parameter of the nonlinear classification due to RBF kernel. Moreover, this parameter is a trade-off between error due to bias and variance of the predictive model. Therefore, there are two main problems, a problem of overfitting of the model and the boundary does not correspond with the complexity of data [13].

3. Feature Selection

Feature selection is one of the hot topics in a machine learning area. The increasing complexity of the contemporary problems in the industry or society resulted in high dimensional data. Moreover, the processing of the data can be a highly computational operation which demands unacceptable quantity of time. Therefore, there is the immense interest in dimension reduction of high dimensional data. Hence, the feature reduction techniques are broadly applied. The feature selection techniques reduce the original number of dimensions of the data due to the importance of each feature. Moreover, each subset will be evaluated according to OCSVM algorithm.

The correlation techniques are selected as the first group for the examination of feature reduction. The correlation techniques calculate the relationship between variables. The correlation techniques are based on a simple assumption that highly similar features are redundant and unnecessarily increase the dimensionality of the dataset. Moreover, algorithms based on Pearson, Kendall and Spearman correlations were detailed described in the article [15]. The approach is based on searching of the correlation matrix in order to find the most correlated features which should be excluded.

The feature selection techniques based on c were adopted for dimensionality reduction of the dataset which were discussed by Guyon and Elisseeff [11]. Furthermore, the ROC curve is calculated for each feature in the dataset in connection to a specific class. The area under the ROC curve is used as the main parameter to distinguish between important and unimportant feature.

Recursive Feature Elimination (RFE) was chosen as the last feature selection method. This technique creates subsets of features which are consequently evaluated. Moreover, in each iteration, the feature is included or excluded according to its perforation. This process is named recursive elimination. The RFE is a wrapper method based on greedy optimization. The RFE method was examined in the paper [16].

4. Experimental Setup

In order to investigate the possibility of deployment of semi-supervised based anomaly detection we exploit dataset which was developed by Lemay and Fernandez [17]. This fully labeled ICS dataset is based on Modbus communication protocol where six datasets are classify as normal. Furthermore, in the five datasets, the malicious activities are presented. We decided to use three of them ("CnC uploading exe", "6RTU with operate", "moving two files"). Moreover, the architecture of the testbed is shown in fig. 2.

Fig. 1. Testbed used for generation of ICS dataset [17]

We present the evaluation of feature selection techniques for anomaly detection systems in ICS networks. Moreover, a considerable number of network-based features from pcap files were collected and evaluated. The obtained data were preprocessed. Consequently, we created twenty-one datasets which corresponding to three cyber-attacks and eight feature selection technique. The features in datasets were selected according to Pearson, Kendall and Spearman correlations in the first phase. The RF, SVM, and ANN used to distinguish between an important and unimportant feature in the second phase. Lastly, we used the RFE method in order to select the most important features. Thus, each dataset was verified according to the multistep procedure presented in the paper [13]. A considerable number of OCSVM classification models were created according to gamma parameter. The best-suited result was established via multi-criteria evaluation based on multiple criteria (Accuracy, Sensitivity, Specificity, Precision, False Positive Rate (FPR) and Time).

- **Accuracy** - It represents the correct classification of the model. Moreover, accuracy is calculated as correct classify observation divided by the total observations.
- **Sensitivity** - Sensitivity is also known as recall or true positive rate. Moreover, it is based on true positive condition and predicted positive condition. The criterion expresses how much relevant results are retrieved by the predictive model.
- **Specificity** - Specificity is also known as True negative rate. This criterion represents the measure of how correctly the negatives examples are classified.
- **Precision** - The criterion is also known as positive predictive value. It takes into account true positive value and false positive value. The precision gives us information about how many relevant and irrelevant results give us the predictive model.
- **FPR** - This criterion is commonly known as false alarm rate. The predictive model improperly identifies normal harmless behavior as an anomaly which may lead to disruption of ICS. Therefore, FPR is highly important for critical infrastructure because the availability of the services is the most important criterion for ICS.
- **Time** - Time represents necessary time period for creation and evaluation of the predictive model. [13]

5. Results

The three cyber-attacks were chosen (“CnC uploading exe”, “6RTU with operate”, “moving two files”) to test feature selection methods. Each cyber-attack was represented by pcap file. Additionally, we extracted two hundred and ninety-six feature from each pcap file; and consequently created three datasets. The datasets were preprocessed and cleaned from zero variance features or “empty” features. Furthermore, numerical transformations were applied on all datasets. The correlation techniques were applied on clean datasets without cyber-attacks which correspond with novelty detection ideology. The correlation coefficients (Pearson, Kendall, and Spearman) were calculated for each dataset. Furthermore, all features with variance higher than 0.8 were excluded from datasets. Thus, nine subsets are created according to correlation technique and cyber-attack. The example of selected features by Pearson for cyber-attack “CnC uploading exe” is presented in Tab. 1.

tcp.stream	ip.src	ip.dst
tcp.dstport	tcp.window_size	tcp.analysis.push_bytes_sent
tcp.analysis.bytes_in_flight	tcp.analysis.ack_rtt	tcp.nextseq
tcp.ack	tcp.flags.push	tcp.flags.ack
tcp.flags.fin	tcp.analysis.initial_rtt	frame.time_delta
modbus.bitnum	modbus.data	modbus.func_code
modbus.reference_num	modbus.request_frame	tcp.checksum
tcp.options.sack_perm	tcp.payload	tcp.pdu.size
tcp.port	tcp.reassembled.data	tcp.reassembled.length
tcp.segment	tcp.segment_data	tcp.stream
tcp.time_delta	tcp.window_size_scalefactor	

Table 1. Selected features according to Pearson correlation for cyber-attack “CnC uploading exe”

The feature selection based on classification algorithms were performed by RF, SVM, and ANN algorithms. Hence, the nine subsets were created. The illustrative example of results according to classification algorithm RF for cyber-attack "CnC uploading exe" is as follow: tcp.srcport, tcp.dstport, tcp.window_size_value, tcp.window_size, tcp.seq, tcp.ack, tcp.port, tcp.time_relative.

Lastly, the RFE method was used to choose the most appropriate features in relation to the chosen cyber-attacks. Thus, three subsets were created for each cyber-attack. The illustrative result for RFE based on "CnC uploading exe" goes as follows: tcp.srcport, tcp.dstport, tcp.window_size_value, tcp.window_size, tcp.seq, tcp.ack, tcp.port.

Subsequently, the five hundred classification models were generated for each subset; therefore, ten thousand and five hundred predictive models were created in total. Each classification model was built on the normal behavior of the system and evaluated on the test dataset which consisted of normal for different values of the gamma parameter (from 0.002 to 1 after 0.002) and abnormal records. Moreover, Accuracy, Sensitivity, Specificity, Precision, False Positive Rate (FPR) and Time were calculated as an assessment factor for each classification model. Thus, the descriptive statistics for all results can be seen in Tab. 2.

		CnC uploading exe						6RTU with operate						moving two files					
		Acc.	Sens.	Spec.	Prec.	FPR	Time	Acc.	Sens.	Spec.	Prec.	FPR	Time	Acc.	Sens.	Spec.	Prec.	FPR	Time
Pearson	Min.	0.420	0.991	0.245	0.285	0.357	6.221	0.854	0.610	0.849	0.184	0.023	7.586	0.389	0.998	0.082	0.353	0.609	24.370
	1st Qu.	0.485	0.997	0.268	0.365	0.580	15.278	0.869	0.993	0.863	0.268	0.078	21.225	0.491	1.000	0.097	0.461	0.772	45.450
	Median	0.612	1.000	0.327	0.521	0.673	23.555	0.894	1.000	0.886	0.408	0.114	31.646	0.635	1.000	0.130	0.614	0.870	107.750
	Mean	0.621	0.998	0.359	0.534	0.641	25.714	0.899	0.977	0.896	0.458	0.104	31.489	0.652	1.000	0.172	0.632	0.828	96.270
	3rd Qu.	0.740	1.000	0.420	0.681	0.732	38.500	0.924	1.000	0.922	0.613	0.137	43.455	0.815	1.000	0.228	0.804	0.903	130.510
	Max.	0.892	1.000	0.643	0.875	0.755	43.496	0.956	1.000	0.977	0.904	0.151	55.626	0.915	1.000	0.392	0.911	0.918	181.670
Kendall	Min.	0.345	0.995	0.224	0.192	0.292	6.470	0.825	0.734	0.825	0.023	0.017	9.100	0.196	0.998	0.064	0.149	0.618	26.750
	1st Qu.	0.396	0.997	0.238	0.256	0.622	17.330	0.832	1.000	0.830	0.057	0.108	31.650	0.301	1.000	0.073	0.261	0.821	66.600
	Median	0.509	1.000	0.278	0.394	0.723	26.350	0.861	1.000	0.855	0.220	0.145	57.530	0.483	1.000	0.096	0.453	0.904	121.990
	Mean	0.558	0.999	0.333	0.456	0.667	24.340	0.870	0.986	0.867	0.285	0.133	47.610	0.521	1.000	0.139	0.494	0.861	114.030
	3rd Qu.	0.690	1.000	0.378	0.620	0.762	32.360	0.900	1.000	0.892	0.443	0.170	62.540	0.749	1.000	0.179	0.734	0.927	162.840
	Max.	0.920	1.000	0.708	0.906	0.776	36.560	0.950	1.000	0.983	0.927	0.175	74.010	0.913	1.000	0.382	0.909	0.936	186.280
Spearman	Min.	0.340	0.993	0.222	0.187	0.292	6.199	0.830	0.737	0.829	0.050	0.024	9.088	0.281	0.998	0.071	0.239	0.605	23.810
	1st Qu.	0.418	0.995	0.244	0.284	0.608	16.598	0.838	1.000	0.836	0.096	0.099	29.657	0.389	1.000	0.082	0.354	0.784	47.560
	Median	0.530	0.996	0.286	0.423	0.714	25.273	0.859	1.000	0.853	0.209	0.147	52.904	0.567	1.000	0.112	0.542	0.888	107.100
	Mean	0.568	0.996	0.338	0.470	0.663	23.184	0.875	0.987	0.872	0.315	0.128	44.132	0.589	1.000	0.156	0.566	0.844	118.890
	3rd Qu.	0.709	0.997	0.393	0.643	0.756	30.925	0.910	1.000	0.901	0.495	0.164	58.926	0.802	1.000	0.216	0.791	0.918	170.090
	Max.	0.920	1.000	0.708	0.906	0.778	34.401	0.953	1.000	0.976	0.893	0.171	64.162	0.917	1.000	0.395	0.914	0.929	270.120
RF	Min.	0.881	1.000	0.614	0.854	0.248	3.785	0.939	0.787	0.955	0.782	0.017	4.414	0.851	0.982	0.260	0.843	0.619	16.550
	1st Qu.	0.886	1.000	0.624	0.860	0.297	3.962	0.963	0.966	0.958	0.801	0.025	4.846	0.885	0.985	0.297	0.889	0.678	16.910
	Median	0.904	1.000	0.663	0.882	0.337	4.027	0.964	0.991	0.961	0.812	0.039	5.421	0.892	0.985	0.308	0.898	0.692	17.090
	Mean	0.905	1.000	0.667	0.882	0.333	4.057	0.966	0.971	0.965	0.836	0.035	5.238	0.891	0.989	0.311	0.895	0.689	18.320
	3rd Qu.	0.920	1.000	0.704	0.902	0.376	4.117	0.969	0.991	0.975	0.881	0.042	5.548	0.897	0.994	0.322	0.904	0.703	17.380
	Max.	0.938	1.000	0.752	0.923	0.386	5.004	0.979	0.991	0.983	0.920	0.045	6.595	0.915	1.000	0.382	0.921	0.740	32.370
ANN	Min.	0.459	0.933	0.259	0.333	0.241	3.566	0.908	0.675	0.955	0.785	0.007	5.413	0.433	0.994	0.087	0.401	0.014	10.330
	1st Qu.	0.641	0.953	0.334	0.581	0.553	3.715	0.970	0.939	0.970	0.858	0.017	5.668	0.785	0.995	0.193	0.777	0.583	11.810
	Median	0.710	0.958	0.383	0.671	0.617	3.786	0.974	0.982	0.976	0.889	0.024	5.846	0.856	0.995	0.267	0.852	0.733	12.220
	Mean	0.708	0.965	0.405	0.665	0.596	4.179	0.973	0.962	0.976	0.887	0.024	5.900	0.807	0.996	0.375	0.799	0.625	13.690
	3rd Qu.	0.773	0.965	0.447	0.749	0.666	4.866	0.978	0.984	0.983	0.920	0.030	6.097	0.925	0.996	0.417	0.924	0.807	14.310
	Max.	0.919	1.000	0.759	0.939	0.742	7.220	0.989	0.992	0.993	0.969	0.045	6.878	0.995	1.000	0.986	0.999	0.913	22.290

SVM	Min.	0.886	1.000	0.624	0.860	0.297	3.962	0.951	0.877	0.946	0.740	0.018	4.514	0.885	0.998	0.319	0.880	0.599	17.480
	1st Qu.	0.899	1.000	0.651	0.875	0.328	4.129	0.959	0.964	0.955	0.782	0.029	4.752	0.894	0.998	0.341	0.889	0.633	18.070
	Median	0.905	1.000	0.665	0.883	0.335	4.186	0.962	0.966	0.961	0.816	0.039	4.911	0.902	0.998	0.354	0.897	0.646	18.410
	Mean	0.903	1.000	0.662	0.881	0.338	4.200	0.963	0.964	0.963	0.823	0.037	5.002	0.902	0.999	0.356	0.897	0.644	19.710
	3rd Qu.	0.908	1.000	0.672	0.887	0.350	4.246	0.966	0.985	0.971	0.862	0.045	5.246	0.907	0.998	0.367	0.903	0.659	18.920
	Max.	0.920	1.000	0.704	0.902	0.376	4.992	0.974	0.986	0.982	0.916	0.054	9.476	0.919	1.000	0.401	0.916	0.681	35.140
RFE	Min.	0.892	1.000	0.637	0.867	0.292	3.734	0.838	0.526	0.973	0.874	0.015	4.892	0.874	0.998	0.301	0.867	0.622	16.530
	1st Qu.	0.905	1.000	0.665	0.883	0.305	3.884	0.969	0.920	0.979	0.900	0.018	5.183	0.893	0.999	0.336	0.887	0.646	17.060
	Median	0.911	1.000	0.680	0.890	0.320	3.943	0.977	0.971	0.980	0.908	0.020	5.294	0.896	0.999	0.344	0.891	0.656	17.500
	Mean	0.910	1.000	0.678	0.889	0.322	3.975	0.968	0.923	0.980	0.909	0.020	5.335	0.896	0.999	0.345	0.891	0.656	18.120
	3rd Qu.	0.917	1.000	0.695	0.898	0.335	4.014	0.980	0.987	0.982	0.916	0.021	5.457	0.901	0.999	0.354	0.896	0.664	18.260
	Max.	0.922	1.000	0.708	0.904	0.363	4.856	0.986	0.988	0.985	0.931	0.027	6.333	0.910	1.000	0.378	0.906	0.699	24.240

Table 2. The results for all feature selection technique per cyber-attacks.

Tab. 2 shows us overall results for classification models. As can be seen, the ANN outperform other techniques in the most important criteria FPR and Time which is needed to be minimized. Otherwise, the rest of the criteria should be maximized. Furthermore, the ANN shows promising results in Accuracy, Sensitivity, Specificity, Precision in all cases.

Moreover, the selected features for "CnC uploading exe" goes as follow: X_ws.col.Protocol, tcp.window_size_value, tcp.window_size, tcp.flags.push, modbus.func_code, tcp.flags.str, tcp.segment. The selected subset of feature for 6RTU with operate cyber-attack goes as follow: ip.src, ip.dst, tcp.dstport, tcp.window_size_value, tcp.flags, tcp.window_size, tcp.flags.ack, tcp.segment, tcp.time_relative. Lastly, the features selected by ANN for "moving two files" goes as follow tcp.window_size_value, tcp.window_size, udp.dstport, udp.srcport.

Fig. 3. The results for all OCSVM classification models in consideration of "6RTU with operate" based on features selected by ANN.

The best-performing feature selection technique based on ANN is evaluated according to multiple criteria (Accuracy, Sensitivity, Specificity, Precision, FPR and Time). Therefore, we decided to present results in detail. The overall results for ANN are represented in Fig. 3 where Index represents different values of gamma parameter (from 0.002 to 1 after 0.002) and the axis Y is represented by an actual value of Accuracy, Sensitivity, Specificity, Precision, FPR (from 0 to 1) and Time (from 0 to 7 ms).

As we mentioned before, we need to minimize FPR and Time and maximize Accuracy, Sensitivity, Specificity, Precision criterions. Moreover, figure 3 shows us that the best performance of OCSVM is connected with the small value of the gamma parameter. Lastly, we use multi-criteria evaluation in order to calculate and choose a suitable gamma parameter to increase classification capabilities of OCSVM based on selected criteria. The Tab. 3 represents best-suited classification models based on gamma parameter for different feature selection techniques. As can be seen in the Tab. 3 the selected classification model for ANN by multi-criteria evaluation is the robust solution in comparison with other feature selection techniques. The rest of the feature selection technique indicate similar results in all criteria.

	CnC uploading exe						6RTU with operate						moving two files					
	Acc.	Sens.	Spec.	Prec.	FPR	Time	Acc.	Sens.	Spec.	Prec.	FPR	Time	Acc.	Sens.	Spec.	Prec.	FPR	Time
Pearson	0.891	0.991	0.639	0.873	0.361	6.827	0.932	0.771	0.972	0.877	0.028	8.524	0.915	0.999	0.389	0.910	0.611	25.668
Kendall	0.920	0.996	0.708	0.906	0.292	6.470	0.945	0.808	0.979	0.904	0.021	9.442	0.912	0.999	0.381	0.907	0.619	27.566
Spearman	0.920	0.996	0.708	0.906	0.292	6.199	0.945	0.824	0.973	0.877	0.027	9.993	0.917	0.998	0.395	0.914	0.605	24.064
RF	0.936	1.000	0.747	0.921	0.253	3.853	0.978	0.956	0.983	0.920	0.017	4.475	0.913	0.992	0.375	0.915	0.625	17.119
ANN	0.919	0.961	0.759	0.938	0.241	3.813	0.937	0.752	0.992	0.966	0.008	5.835	0.995	0.995	0.986	0.999	0.014	11.105
SVM	0.920	1.000	0.703	0.902	0.297	4.046	0.971	0.923	0.982	0.916	0.018	4.595	0.916	0.998	0.392	0.913	0.608	17.635
RFE	0.920	1.000	0.703	0.902	0.297	3.871	0.984	0.980	0.984	0.927	0.016	5.035	0.909	0.999	0.374	0.904	0.626	16.693

Table 3. Results for the best-selected gamma parameter according to multi- criteria evaluation.

6. Conclusion

The presented research is focused on improving detection capabilities of the anomaly detection system in the SCADA environment. Therefore, the multiple feature selection techniques were selected in order to achieve the goal. We decided to use three groups of feature selection techniques (correlation techniques, techniques based on classification and RFE technique). A considerable amount of OCSVM models were created to evaluate subsets of features for different values of gamma parameter which has the high influence on the performance of OCSVM classification. At last, the multi-criteria evaluation was used to select the best-suited classification model.

The overall results show promising performance of feature reduction techniques based on ANN. The value of the most important criteria (FPR and Time) were relatively low in comparison with other solutions for all cyber-attacks. However, the decomposition of the results for specific criterions is much wider than other feature selection techniques which can lead to worse results due to the chosen gamma parameter.

The worst performing group of feature selection techniques is correlations methods. The FPR and Time criterions are proportionally higher than others solutions because of the higher dimensionality of the selected dataset. However, the main advantage of the feature selection based on correlation techniques is that these techniques do not need label information; therefore, the information about cyber-attacks are not needed. Otherwise, the other feature selection technique computes important of each feature according to the label which can be tricky to use in novelty detection system.

Even the most of the results for all feature selection techniques gives promising results, the high value of FPR criterion is unacceptable in all cases. We can describe FPR as one of the most important criteria for SCADA systems due to its importance to the availability of the services; therefore every wrong classified record can result in high impact at SCADA system especially in case of big data. Therefore, there is a necessity for application of techniques and methods for FPR reduction. The one of the possible way how to achieve this goal is to apply deep learning method. The direction of our future research will include investigation of deployment recurrent neural network in SCADA environment.

7. Acknowledgments

This work was funded by the Internal Grant Agency (IGA/FAI/2018/003) and supported by the project ev. no. VI20152019049 "RESILIENCE 2015: Dynamic Resilience Evaluation of Interrelated Critical Infrastructure Subsystems", supported by the Ministry of the Interior of the Czech Republic in the years 2015-2019 and also supported by the research project VI20172019054 "An analytical software module for the real-time resilience evaluation from point of the converged security ", supported by the Ministry of the Interior of the Czech Republic in the years 2017-2019. Moreover, this work was supported by the Ministry of Education, Youth and Sports of the Czech Republic within the National Sustainability Programme project No. LO1303 (MSMT-7778/2014) and also by the European Regional Development Fund under the project CEBIA-Tech No. CZ.1.05/2.1.00/03.0089. Finally, we thank Lemay and Fernandez [17] who provides ICS datasets.

8. References

- [1] Knapp, E. Industrial network security: securing critical infrastructure networks for Smart Grid, SCADA, and other industrial control systems. Waltham, MA: Syngress, c2011, xvii, 341 p. ISBN 978-1-59749-645-2.
- [2] Klimeš, J. (2014). Using formal concept analysis for control in cyber-physical systems. 24th DAAAM International Symposium on Intelligent Manufacturing and Automation. Procedia Engineering (pp. 1518-1522).
- [3] Ferrag, M. A., Maglaras, L. A., Janicke, H., & Jiang, J. (2016). A survey on privacy-preserving schemes for smart grid communications. arXiv preprint arXiv:1611.07722.
- [4] Cvitić, I., Vujić, M., & Husnjak, S. (2016, January). Classification of security risks in the IoT environment. In 26th DAAAM International Symposium on Intelligent Manufacturing and Automation (pp. 0731-0740).
- [5] Maglaras, L. A., Kim, K. H., Janicke, H., Ferrag, M. A., Rallis, S., Fragkou, P., Maglaras, A., & Cruz, T. J. (2018). Cyber security of critical infrastructures. ICT Express
- [6] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. ACM computing surveys (CSUR), 41(3), 15.
- [7] Maglaras, L. A., & Jiang, J. (2014, August). Intrusion detection in scada systems using machine learning techniques. In Science and Information Conference (SAI), 2014 (pp. 626-631). IEEE.
- [8] Raczko, E., & Zagajewski, B. (2017) Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images. European Journal of Remote Sensing 50:1, pages 144-154.
- [9] Omer, G., Mutanga, O., Abdel-Rahman, E. M., & Adam, E. (2015). Performance of support vector machines and artificial neural network for mapping endangered tree species using WorldView-2 data in Dukuduku forest, South Africa. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 8(10), 4825-4840.
- [10] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(3), 131-156.
- [11] Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.
- [12] Vapnik, V. (2013). The nature of statistical learning theory. Springer science & business media.
- [13] Vávra, J., & Hromada, M. (2018, April). Novelty Detection System Based on Multi-criteria Evaluation in Respect of Industrial Control System. In Computer Science On-line Conference (pp. 280-289). Springer, Cham.
- [14] Burges, C. J. (1998). A tutorial on support vector machines for pattern recognition. Data mining and knowledge discovery, 2(2), 121-167.
- [15] Croux, C., & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. Statistical methods & applications, 19(4), 497-515.
- [16] Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46(1-3), 389-422.
- [17] Lemay, A., & Fernandez, J. M. (2016, August). Providing SCADA Network Data Sets for Intrusion Detection Research. In CSET@ USENIX Security Symposium.