ISSN 0146-4116, Automatic Control and Computer Sciences, 2016, Vol. 50, No. 8, pp. 726–733. © Allerton Press, Inc., 2016.
Original Russian Text © L.V. Utkin, V.S. Zaborovskii, S.G. Popov, 2016, published in Problemy Informatsionnoi Bezopasnosti, Komp'yuternye Sistemy.

Detection of Anomalous Behavior in a Robot System Based on Deep Learning Elements

L. V. Utkin*, V. S. Zaborovskii, and S. G. Popov

Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

**e-mail: lev.utkin@gmail.com*

Received June 11, 2016

Abstract—The preprocessing procedure for anomalous behavior of robot system elements is proposed in the paper. It uses a special kind of a neural network called an autoencoder to solve two problems. The first problem is to decrease the dimensionality of the training data using the autoencoder to calculate the Mahalanobis distance, which can be viewed as one of the best metrics to detect the anomalous behavior of robots or sensors in the robot systems. The second problem is to apply the autoencoder to transfer learning. The autoencoder is trained by means of the target data which corresponds to the extreme operational conditions of the robot system. The source data containing the normal and anomalous observations derived from the normal operation conditions is reconstructed to the target data using the trained autoencoder. The reconstructed source data is used to define a optimal threshold for making decision on the anomaly of the observation based on the Mahalanobis distance.

Keywords: robot systems, detection of anomalous behavior, autoencoder, Mahalanobis distance, transfer learning, machine learning

DOI: 10.3103/S0146411616080319

1. INTRODUCTION

Recently, robot systems (RS) have ever-increasing value due to certain factors including the capability to solve more efficiently many more complex problems than single robots [1, 2]. In the RS the robots are usually equipped with a system of sensors used to obtain as much information as possible about the external "world". As pointed out in [3], each sensor can cover different aspects of the environment. One of the important problems of RS learning is efficient integration of the numerous distributed sensors which can actuate the GPS devices, temperature sensors, altimeters, scanners, etc. [4].

At the same time, efficient integration of the measurements and joint functioning of all robot subsystems in the RS significantly depends on the reliability or fault-tolerance of the RS elements, as well as on quick detection of failures or anomalous behavior of the RS elements. It follows that the systems of control of the robots should be complemented with the anomalous behavior detection systems, whose purpose consists in detecting the anomalous behavior and sending their signals, for example, to the operator [5]. As pointed out in [5], the failures of robots and their subsystems are not restricted only to the wear and aging of the system hardware. The nature of the anomalous behavior can be very different. The authors of the work [6], for example, take into account so-called contextual failures [7], which occur when the failed sensor reports the correct information, which, in fact, is not so under the operating conditions of the system. For instance, a sensor can be partially destroyed, but physically continue operating, transferring the false information which is nevertheless within the preset limits from the point of view of accuracy of its perception.

At present, there are many publications devoted to anomalous behavior detection using machine learning methods [8–11]. However, we note the work [10], where the authors study the online methods of detection of anomalies for the robots and propose using the Mahalanobis distance to detect the failures. The main idea on which the approach offered in [10] is that a sliding window method is used to analyze the matrix of past data where the matrix columns correspond to the measured attributes, and the matrix rows maintain the values of these attributes within certain time instants or time intervals. The degree or value of the "anomaly" of the current observation (last line of the observation matrix) is estimated based on the previous observations using the Mahalanobis distance, which calculates the distance between the vector of the current observations and numerous observations represented in the matrix of the preceding data. Thus, the Mahalanobis distance can indicate whether the current multidimensional observation is anomalous as compared to the previous observations. This is an excellent idea proposed in the work [10], as in contrast to the generally accepted Euclidean distance, the Mahalanobis distance permits us to consider the distribution of observations and measurements or "weigh" the data by calculating the variance of each random value and covariance between the random values. Thus, the point located within the data aggregation area has a smaller Mahalanobis distance. At the same time, the emissions and anomalous measurements have a longer Mahalanobis distance. The anomalous behavior detection method proposed in the work [10] uses the threshold for the distance to adopt the decision whether the current values are normal or anomalous.

The main drawback of the anomaly detection method [10] is that the Mahalanobis distance cannot be calculated when the dimensionality of the variables exceeds the number of observations. This is because it is necessary to calculate the inverse matrix of covariances to define the Mahalanobis distance. Some approaches to solve partially this problem can be found in the literature (see, for example, [12]). However, these approaches require significant computational resources and often are barely interpretable. It should be noted that robot systems can consist of hundreds of robots equipped with numerous sensors. Thus, the number of features or attributes becomes significant as compared to the number of the previous observations and measurements. Therefore, it is necessary to develop simple, from the computational point of view, methods of anomaly detection in the robot systems, taking into account the larger dimensionality of the data.

Another problem which can be met while detecting anomalous behavior in the robot systems is the lack of information about what the anomalous behavior is. The authors of the work [10] introduce a threshold for the Mahalanobis distance such that the anomalous behavior takes place when the Mahalanobis distance is larger than the threshold. However, despite the interesting approach for selecting the optimal threshold represented in the work [10], its definition causes difficulties when the training sample is small.

It is necessary to specify an important characteristic of the robot system that the robot may learn in different operational conditions and environments, and it is important to use this information, i.e., consider the difference of the data obtained in different conditions. Despite the importance of some conditions, it is difficult to obtain a large amount of data corresponding to these conditions. For example, the robots can operate in extreme conditions (high or low temperatures, corrosive media, etc.) and in normal conditions. Learning in extreme conditions is difficult, as it leads, for example, to damage of the equipment. As a result, we have a large volume of training data in normal conditions and a small volume in extreme conditions. Another example is the analysis of images during the day and in twilight.

One of the interesting and powerful approaches improving the learning quality in the above cases and considering the different learning conditions is transfer learning which can be treated as the method to transfer the useful information from one or some "initial" data sources (normal conditions) to the target set (extreme conditions) [13]. The main purpose of learning consists in solving the problem of machine learning when the learning data obtained in normal operation conditions (source data) and test data obtained in extreme conditions (target data) have different distributions of probabilities or are represented in different feature spaces. The main idea on which transfer learning is based consists in defining the total source and target data despite the difference of their distributions [14].

One of the most efficient and popular methods of the RS machine learning is the deep learning method. That is why later the authors will use an important element for deep learning—a sparse autoencoder which is a special type of the neural network to solve the transfer learning problem. The autoencoder implements the algorithm without a teacher using the back propagation of error so that the input and output of the network would be possibly identical. An interesting application of the autoencoder to solve the transfer learning problem in detection systems for speech emotions was proposed in the work [15]. At the first stage, the autoencoder was trained on the elements of the target training sample. At the second stage, the data from the source training sample was reconstructed using the previously trained autoencoder. Then, the target and reconstructed source data were integrated to construct the classifier.

It should be noted that the autoencoders have already been used in problems of anomalous behavior detection [16, 17]. Particularly, in the work [16] the autoencoder was used to reduce the dimensionality of the data. In fact, the authors [16] replaced the linear method of the main components by the autoencoder. In the work [17] the structural noisy autoencoder was proposed where the data conversion error was considered as the failure index.

We propose to use the autoencoder for two aims. First, we try to solve the problem of calculation of the Mahalanobis distance in the case of a small sample and high dimensionality of the data using the autoencoder to reduce the data dimensionality. But, unlike the generally accepted use of the autoencoder to reduce the dimensionality, we apply it also for transfer learning. That is why, we, second, generalize the approach to the transfer learning using the autoencoder proposed in [15] to define the threshold for decision making for the Mahalanobis distance. The source data set is reconstructed in the target data using the autoencoder. We study the case when the target data obtained in the extreme conditions do not include the examples corresponding to the anomalous behavior of the robot systems, i.e., the target data consists only of the "normal" examples" (matrix of the past observations). At the same time, it is assumed that the source data includes both normal and anomalous examples classified by the operator. In order to improve the method of detection of anomalous behavior of the robots using the Mahalanobis distance, we propose to consider the distance between the reconstructed normal and abnormal source data.

The paper is arranged as follows. The formal statement of the problem of detection of anomalous behavior considering the source and target data is given in Section 2. The detailed description of the anomalous behavior detection method using the Mahalanobis distance proposed in the work [10] is given in Section 3. A brief description of the autoencoder and methods of its learning is represented in Section 4. The transfer learning algorithm using the autoencoder offered in the work [15], is considered in Section 5. A new procedure for preliminary processing of data to allow some problems related to the anomalous behavior detection method [10] to be solved is proposed in Section 6. Section 7 includes the final remarks on the work.

2. FORMAL PROBLEM STATEMENT

It is assumed that every moment k we record one observation in the form of the vector consisting of m characteristics or attributes obtained from all sensors T of the RS robots. After the time moment r, a training set of observations is available obtained in the normal operation conditions of the robot systems consisting of r examples $Q_s = \{(\mathbf{x}_1^s, y_1^s), (\mathbf{x}_2^s, y_2^s), \dots, (\mathbf{x}_r^s, y_r^s)\}$, having some probability distribution D_s. Here $\mathbf{x}_i^s \in \mathbf{R}^m$ is a vector of m features obtained after the time i; $y_i^t \in \{-1, 1\}$ is a label of the class corresponding to the normal behavior of the robot system at the time moment i ($y_i^t = 1$) and anomalous behavior at the same moment ($y_i^t = -1$). Let us call this set the source.

There is also a training set obtained in the extreme operation conditions consisting of n examples $Q_t = \{\mathbf{x}_1^t, \mathbf{x}_2^t, \dots, \mathbf{x}_n^t\}$, having another probability distribution D_t. It is assumed that all examples in Q_t correspond to the normal behavior of the robot system. In other words, we do not have any examples corresponding to the anomalous behavior among the sample elements obtained in the extreme conditions. Let us call this set the target.

Let us also assume that the source and target training samples have similar features. The purpose of the authors is to develop an algorithm for the anomalous behavior detection of the robot system considering the above features of the source and target training samples using the Mahalanobis distance.

3. DETECTION OF ANOMALIES USING THE MAHALANOBIS DISTANCE

Let us consider the online method of the anomalous behavior detection proposed in the work [10] in this section. According to this method the matrix of the past data H, consisting of n rows, is created so that each row contains the values m of the characteristics or attributes obtained from the sensors within certain moments of time, for example, GPSx, GPSy, odometry, telemetry, etc. The sliding window is defined containing q rows of the matrix H. At each moment of time n after obtaining some vector $\mathbf{x}_n^t$, the matrix H is modified. Using H, the "anomaly" index for the current vector $\mathbf{x}_n^t$ is estimated using the Mahalanobis distance. It is calculated between the vector of the dimensionality m and a group of other vectors in mean-square deviation units.

Let us assume that the center of the set of vectors is the vector $\mu = (\mu_1, \dots, \mu_m)$ and let us designate the covariance matrix S. Here, the center means a sample average value or sample median of the set of vectors limited by the sliding window. Then, the Mahalanobis distance between the vector of characteristics $\mathbf{x}_n^t$ and the center of the set is defined as

$$M(\mathbf{x}_n^t, \mu) = \sqrt{(\mathbf{x}_n^t - \mu)^{\mathrm{T}} S^{-1} (\mathbf{x}_n^t - \mu)}.$$

Here S^{-1} is the inverse covariance matrix. Thus, the Mahalanobis distance is used to define whether the vector $\mathbf{x}_n^t$ is anomalous according to the matrix H. Particularly, if the vector $\mathbf{x}_n^t$ with the values similar or close to the observed vectors in the matrix H, is located in the m-dimensional space in the area where observations are concentrated, the Mahalanobis distance will be small. If the vector $\mathbf{x}_n^t$ is located outside the area of

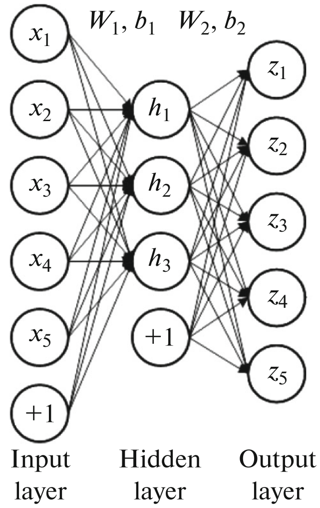

Fig. 1. Autoencoder structure.

concentration of other vectors, the Mahalanobis distance becomes longer, which corresponds to the anomalous behavior. The authors of the work [10] specify that many general types of anomalous behavior can be detected using the Mahalanobis distance. It should be noted that this distance represents a more powerful multidimensional method of detection of the emissions or anomalies as compared to the Euclidean distance as the Mahalanobis distance considers correlation between the random values and their range [18].

To increase the efficiency of the method the authors of the work [10] also suggested finding the subset of the correlated features based on the calculation of the standard Pearson correlation coefficient. With the subsets of the correlated features the authors of [10] define the thresholds of the longest Mahalanobis distance to adopt the decision on each subset whether the current vector is anomalous.

This method is very interesting and efficient. Besides, the numerous experiments represented in [10] confirm its quality. But, first, it does not consider the use of the source training data obtained in normal operational conditions. Second, the Mahalanobis distance cannot be calculated if the number of the features m exceeds the number q of the rows in the sliding window. Third, the threshold of adoption of the decisions is selected rather randomly. The selection of the optimal thresholds requires a large volume of training data. That is why we suggest using the autoencoder in order to pass over all specified disadvantages of the method and include the source training data in the anomalous behavior detection procedure.

4. AUTOENCODER

The autoencoder is the type of the neural network where the output data coincides or is similar to the input data. The autoencoder has two operational modes: coding (conversion of the data into new data defined by the hidden layer) and decoding (data reconstructing). During the coding process, the input data is transformed into the data at the hidden layer output. In the decoding process, the data at the hidden layer output is reconstructed in the output data. Let us assume that the autoencoder hidden layer consists of p elements (see Fig. 1). If the input example from the target learning set is designated as $\mathbf{x}$, the activation of each neuron of the hidden layer designated as h_j, $j = 1,\ldots,p$, is calculated as

$$h(\mathbf{x}) = \sigma(W_1\mathbf{x} + b_1).$$

Here $\sigma(z)$ is an activation or threshold function, for example, $\sigma(z) = 1/(1 + \exp(-z))$; W_1 is such a weight matrix that its element $w_{ij}^{(1)}$ is the weight of connection between the weight between the j-element of the input layer and i-neuron of the hidden layer; here b_1 is the vector of the absolute terms; $h(\mathbf{x}) = (h_1(\mathbf{x}),\ldots,h_p(\mathbf{x}))$, the vector of activations of the hidden-layer neurons.

The neural network output is defined as

$$\mathbf{z} = \sigma(W_2 h(\mathbf{x}) + b_2).$$

Here W_2 is such a weight matrix that its element $w_{ij}^{(2)}$ is the weight of connection between the weight between the j-neuron of the hidden layer and i-element of the output layer, where b_2 is the vector of the absolute terms. Each training example $\mathbf{x}$ is displayed in its code $h(\mathbf{x})$ and its reconstruction $\mathbf{z}$.

If there is a set of n input examples $\mathbf{x}$, the weight matrices W_1 and λ, the absolute members b_1 and b_2 are calculated using the known back-propagation algorithm which minimizes the reconstruction error

$$J(W_1, b_1, W_2, b_2) = \sum_{i=1}^{n} \|\mathbf{x} - \mathbf{z}\|^2.$$

Here the reconstruction error $L(\mathbf{x}, \mathbf{z}) = \|\mathbf{x} - \mathbf{z}\|^2$ is considered as the function of losses, where $\|\cdot\|^2$ is the quadratic standard. Thus, each example is displayed in its code and then reconstructed into $\mathbf{z}$ so that $\mathbf{x} \approx \mathbf{z}$.

The simplest form of regularization

$$J(W_1, b_1, W_2, b_2) = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{x} - \mathbf{z}\|^2 + \lambda W,$$

where

$$W = \sum_{i=1}^{m}\sum_{j=1}^{p} (w_{ij}^{(2)})^2 + \sum_{i=1}^{p}\sum_{j=1}^{m} (w_{ij}^{(2)})^2$$

is the summand limiting the weights and added to the function of losses, and λ, the parameter controlling the influence of regularization.

Another method of regularization is adding the Kullback-Leibler divergence $\mathrm{KL}(\rho\|\rho_j^*)$ as the penalty summand to provide the autoencoder sparsity defined as

$$\mathrm{KL}(\rho\|\rho_j^*) = \rho \log \frac{\rho}{\rho_j^*} + (1 - \rho) \log \frac{1 - \rho}{1 - \rho_j^*}.$$

Here ρ is the sparsity level; ρ_j^*, the average activation of the hidden neuron with the number j (average by all examples of the training set), defined as

$$\rho_j^* = \frac{1}{n_t}\sum_{i=1}^{n_t} h_j(\mathbf{x}_i).$$

Adding the penalty summand based on the Kullback-Leibler divergence to $J(W_1, b_1, W_2, b_2)$, the target function is obtained

$$J_{\text{sparse}}(W_1, b_1, W_2, b_2) = \sum_{i=1}^{n_t} \|\mathbf{x} - \mathbf{z}\|^2 + \lambda W + \beta \sum_{j=1}^{p} \mathrm{KL}(\rho\|\rho_j^*).$$

Here β controls the penalty summand weight. The idea behind the penalty summand consists in the fact that those values of activation are significantly penalized which significantly differ from the parameter ρ.

One of the most important purposes of the autoencoder is reduction of the data dimensionality. Besides, if the features of the observations are correlated, the reduction of the dimensionality can be significant, i.e., we can obtain $p \ll m$.

5. TRANSFER LEARNING USING THE AUTOENCODER

An interesting algorithm of transfer learning using the autoencoder is suggested in the work [15]. According to this algorithm, the autoencoder with one hidden layer is learnt using the target examples $\mathbf{x}_k^t$, $k = 1, \ldots, n$. Applying this procedure, the autoencoder is learnt by calculating the matrices W_1, W_2 and parameters b_1, b_2. To transfer learning each example, $\mathbf{x}_k^s$ from the source training set is calculated as a new reconstructed example $\mathbf{x}_k^{s\to t}$, obtained using the learnt set of the parameters W_1, W_2, b_1, b_2, i.e.,

$$\mathbf{x}_k^{s\to t} = SA_{\text{Recon}}(\mathbf{x}_k^s).$$

Here

$$SA_{\text{Recon}}(\mathbf{x}) = \sigma(W_2\sigma(W_1\mathbf{x} + b_1) + b_2)$$

is the autoencoder output. We use the designation $s \to t$ for new reconstructed examples in order to show that they are obtained by converting the source data into the target data. Using the given examples, all observations $\mathbf{x}_k^s$ are reconstructed from the source data. As is specified in the work [15], the reconstruction procedure, in its turn, reduces the difference between the source and target data. The reconstructed source data and target data can be integrated to be used in the classification using one classifier.

6. PROPOSED PROCEDURE OF PRELIMINARY PROCESSING

We suggest using the autoencoder for solving the following three tasks: to consider the source training sample, the high dimensionality of the training data, to transfer learning and define the optimal threshold for adoption of the decision while applying the Mahalanobis distance.

The first task. The normal and abnormal data of the source training sample is reconstructed in the target data. This problem can be solved after training the autoencoder using the target sample data. As a result of this procedure two sets of normal and abnormal reconstructed data are obtained. Under certain conditions, the normal data of the reconstructed source sample can be integrated with the target sample data to increase the dimension of the matrix of the previous observations.

The second task. One of the main purposes of use of the autoencoder is to reduce the data dimensionality. Moreover, if the features of the observations are correlated, the reduction becomes significant, i.e., it is possible to achieve the ratio $p \ll m$. In fact, the robot systems mainly consist of identical robots with similar sets of sensors. Although each sector can cover different aspects of the environment and its measurements can significantly differ from the measurements of the similar sensors belonging to other robots, there is often a point of correlation between the observations of these sensors. That is why the dimensionality of all data sets including the reconstructed normal and abnormal data of the source sample, the target data can be reduced by using the autoencoder. As a result, the data set is obtained at the hidden layer output $\{h(\mathbf{x}_k^{s\to t})\}$ and $\{h(\mathbf{x}_k^t)\}$, having the dimensionality p, such that $p < n$. Here $h_i(\cdot)$ is activation of the relevant i-neuron of the hidden layer whose values are defined by the learnt autoencoder. Thus, the result of solving of the second problem of preliminary data processing is the data set of reduced dimensionality. For the new vectors of small dimensionality, it is possible to find the reverse covariance matrix S^{-1} and calculate the Mahalanobis distance.

The third task. Using the reconstructed data of the source training sample the threshold can be obtained to adopt the decision on the RS anomalous behavior. For each reconstructed vector $h(\mathbf{x}_k^{s\to t})$ at $y_k^s = -1$ (abnormal data) the Mahalanobis distance is calculated designated as $M(h(\mathbf{x}_k^{s\to t}), \mu)$, between the vectors $h(\mathbf{x}_k^{s\to t})$ from the set $\{h(\mathbf{x}_k^{s\to t})\colon\ y_k^s = -1\}$ and the compatible set of the normal reconstructed data of the source training sample and target data

$$\{h(\mathbf{x}_k^{s\to t}):\ \ y_k^s = 1\} \cup \{h(\mathbf{x}_k^t)\}.$$

The choice of the threshold depends now on the decision strategy specified in the problem. Particularly, it is possible to apply the pessimistic strategy to select the threshold according to which the smallest Mahalanobis distance is selected from the set of all distances defined by the set $\{h(\mathbf{x}_k^{s\to t})\colon\ y_k^s = -1\}$, i.e.,

$$\text{threshold}_{\text{pessim}} = \min\left\{M(h(\mathbf{x}_k^{s\to t}), \mu)\colon\ y_k^s = -1\right\}.$$

Another strategy (optimistic) is to select the longest Mahalanobis distance

$$\text{threshold}_{\text{optim}} = \max\left\{M(h(\mathbf{x}_k^{s\to t}), \mu)\colon\ y_k^s = -1\right\}.$$

These extreme strategies are partial cases of a more general cautious strategy defined as the linear combination of the pessimistic and optimistic strategies with some caution parameter α:

$$\text{threshold} = \alpha \cdot \text{threshold}_{\text{pessim}} + (1 - \alpha) \cdot \text{threshold}_{\text{optim}}.$$

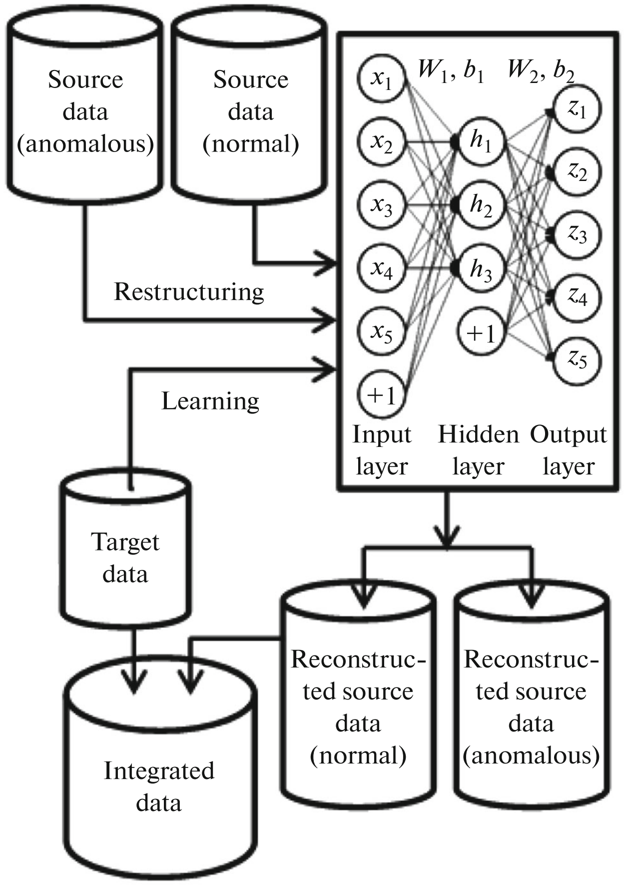

Fig. 2. Preliminary data processing.

The average Mahalanobis distance can be also used. The choice of some or other strategy depends on the application problem where the robot system is used.

The three above-mentioned problems can be considered as the stage of preliminary data processing for the anomalous behavior detection procedure using the Mahalanobis distance and sliding window suggested in the work [10]. This stages are schematically represented in Fig. 2.

7. CONCLUSIONS

The procedure of the preliminary data processing to detect the anomalous behavior in the robot systems consisting of the main related problems has been suggested in the article. It is a preliminary stage to implement the efficient method proposed in the work [10]. But the efficiency and functionality of the method represented in [10] is limited by some conditions which should be taken into account. The main limitation is the necessity to consider the significant dimensionality of the training data features to the robot systems with many robots and relevant sensors. The second limitation is related with the need to consider a small volume of data in the target training sample due to the difficulties appearing while training the robots in the extreme operation functions. That is why the suggested procedure is aimed at passing over the given limitations.

The procedure of preliminary data processing is based on the use of the autoencoder, which is a powerful instrument to solve different problems encountered in machine learning. The autoencoder is, first of all, used in the work to reduce the dimensionality of the training data to calculate the Mahalanobis distance, which is one of the best metrics to detect anomalous behavior of robots or sensors in robot systems [10]. Besides, the autoencoder was also applied to solve the transfer learning problems. The source training sample with the normal and abnormal data obtained in the normal operational conditions was reconstructed in the target sample corresponding to the extreme operation conditions using the learnt autoencoder. The reconstructed data was used to search for the optimal threshold to adopt the decisions on the anomalous behavior based on the Mahalanobis distance. Some decision strategies were suggested to select the optimal threshold.

The proposed procedure of preliminary data processing allows us to solve several problems. However, it should be noted that calculation of the Mahalanobis distance is a sufficiently complex procedure from the computational point of view, which is complicated with the increase of the number of robots and sensors in the robot system. Despite the decrease in the dimensionality using the autoencoder, the calculation complexity can limit the use of the anomaly detection method. That is why the problem of the search for other methods of efficient RS anomalous behavior detection is urgent. One of the methods is to use the k-sparse autoencoder [19], which has very interesting and useful properties. But use of the k-sparse autoencoder for RS monitoring is the work of further studies.

It should be also noted that the suggested use of the autoencoder can be extended to other anomalous behavior detection methods different from the method proposed in the work [10], as the autoencoder solves important problems of learning and dimensionality reduction. These problems can be met in a variety of different schemes for detecting anomalies and failures in systems.

ACKNOWLEDGMENTS

The research was partially supported by the Russian Foundation for Basic Research, grant no. 16-29-04319.

REFERENCES

1. Navarro, I. and Matia, F., An introduction to swarm robotics, *ISRN Rob.,* 2013, vol. 2013, no. 608164, pp. 1–10.
2. Tan, Y. and Zheng, Z.-Y., Research advance in swarm robotics, *Def. Technol.,* 2013, vol. 9, pp. 18–39.
3. Pronobis, A., Mozos, O., and Caputo, B., SVM-based discriminative accumulation scheme for place recognition, *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA 2008,* 2008, pp. 522–529.
4. Cowley, A., Hsu, H.-C., and Taylor, C., Distributed sensor databases for multi-robot teams, *Proceedings of the 2004 IEEE International Conference on Robotics and Automation, ICRA '04,* 2004, vol. 1, pp. 691–696.
5. Khalastchi, E., Kaminka, G., Kalech, M., and Lin, R., Online anomaly detection in unmanned vehicles, *Proc. of 10th Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS 2011),* 2011, vol. 1, pp. 115–122.
6. Khalastchi, E., Kalech, M., and Rokach, L., Sensor fault detection and diagnosis for autonomous systems, *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2013),* 2013, pp. 15–22.
7. Chandola, V., Banerjee, A., and Kumar, V., Anomaly detection: A survey, *ACM Comput. Surv.,* 2009, vol. 41, no. 3, pp. 15:1–15:58.
8. Bartkowiak, A., Anomaly, novelty, one-class classification: A comprehensive introduction, *Int. J. Comput. Inf. Syst. Ind. Manage. Appl.,* 2011, vol. 3, pp. 61–71.
9. Erfani, S., Rajasegarar, S., Karunasekera, S., and Leckie, C., High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning, *Pattern Recognit.,* 2016, vol. 58, pp. 121–134.
10. Khalastchi, E., Kalech, M., Kaminka, G., and Lin, R., Online data-driven anomaly detection in autonomous robots, *Knowl. Inf. Syst.,* 2015, vol. 43, no. 3, pp. 657–688.
11. Pimentel, M., Clifton, D., Clifton, L., and Tarassenko, L., A review of novelty detection, *Signal Process.,* 2014, vol. 99, pp. 215–249.
12. Thomaz, C., Gillies, D., and Feitosa, R., A new covariance estimate for Bayesian classifiers in biometric recognition, *IEEE Trans. Circuits Syst. Video Technol.,* 2004, vol. 14, no. 2, pp. 214–223.
13. Pan, S. and Yang, Q., A survey on transfer learning, *IEEE Trans. Knowl. Data Eng.,* 2010, vol. 22, no. 10, pp. 1345–1359.
14. Yang, S., Lin, M., Hou, C., Zhang, C., and Wu, Y., A general framework for transfer sparse subspace learning, *Neural Comput. Appl.,* 2012, vol. 21, no. 10, pp. 1801–1817.
15. Deng, J., Zhang, Z., Marchi, E., and Schuller, B., Sparse autoencoder-based feature transfer learning for speech emotion recognition, *Humaine Association Conference on Affective Computing and Intelligent Interaction, ACII-2013,* Geneva, 2013, pp. 511–516.
16. Sakurada, M. and Yairi, T., Anomaly detection using autoencoders with nonlinear dimensionality reduction, *Proceedings of the MLSDA 2014 2nd Workshop on Machine Learning for Sensory Data Analysis,* Gold Coast, 2014, pp. 4–11.
17. Tagawa, T., Tadokoro, Y., and Yairi, T., Structured denoising autoencoder for fault detection and analysis, *JMLR: Workshop and Conference Proceedings,* 2014, vol. 39, pp. 96–111.
18. Brereton, R., The Mahalanobis distance and its relationship to principal component scores, *J. Chemom.,* 2015, vol. 29, pp. 143–145.
19. Makhzani, A. and Frey, B., k-Sparse autoencoders, *arXiv preprint: 1312.5663,* 2013.

Translated by Yu. Bezlepkina