Structural health monitoring system of construction facilities: enhanced training approach

Kats V
Moscow State University of Civil Engineering, Yaroslavskoe shosse, 26, Moscow, 129337, Russia vladk.94@mail.ru

Abstract. Structural health monitoring (SHM) of constructions under operating conditions becomes a relevant problem. When dealing with this problem the specific nature and environment conditions of the controlled facility should be considered. Therefore it is very important to continuously train the monitoring system with new data acquired from sensors. Current paper presents an approach of considering training the SHM system as training of defect evolution detection model. This method consists of four subsequent stages which allows to build and train defect evolution detection model on the acoustic emission sensor data. It includes data preparation stages (feature extraction, feature selection), outlier detection and training stage via proposed modification of the One-Class SVM anomaly detection method. Proposed approach was verified on real fuel and energy infrastructure facility. Obtained results give good grounds for utilizing the suggested approach in SHM systems with acoustic emission sensors allowing detection evolutionary defects in controlled facility which contributes to prevention of emergency situations that may have economic, social and ecological consequences.

1. Introduction
Structural health monitoring (SHM) of constructions during their lifecycle is an important condition for providing their sustainable functioning. The purpose of conducting the monitoring consists in early detection of operating defects, prevention emergency situations and minimization of the damage due to failure of the construction [1]. In such a formulation a problem of conducting monitoring on the constructions becomes very urgent. It especially concerns facilities forming fuel and energy infrastructure, e.g. reservoirs, oil pump and compressor stations. Recent emergency in Norilsk caused by the pile foundation under the oil reservoir once again showed the necessity of conducting continuous technical condition estimation of this kind of facilities [2]. In this regard, the problem of development intellectual structural health monitoring systems that allow to detect operating defects in constructions during their evolution become very important.

Behind the most of modern SHM systems lie the collection and processing of the sensors data obtained by acoustic emission (AE) method [3]. It is essential that AE signals parameters emitted by the defect in the construction undergoes alterations depending on various factors, e.g. defect type, its nature, noise during data collection process [4]. In this regard, the implementation of the monitoring system training method that allows to consider special properties of controlled facility.

Current AE diagnostics systems generally use a system settings module that register signals only if the signal amplitude exceeds the user defined amplitude threshold [5]. Such a thresholding approach leads to significant loss of acoustic signal information content and consequently doesn’t allow to consider the entire characteristics of the controlled facility.

Complex approach in training of the SHM systems also supposes building the system of pattern recognition and classification that lies in the foundation of decision making regarding the hazard class
of defects. Such an approach supposes a training mode that utilizes multivariate data analysis. Authors of paper [6] propose a similar approach that uses preliminary training. However, the extracted features given in that paper do not have enough informative content and may not be useful during classification. Several features do not have sensitivity to defect evolution process, others may distort the classification results that lead to incorrect technical condition estimation and subsequently lead to making inappropriate decisions.

In paper [7] authors considered a classifier training method based on K-nearest neighbours algorithm. Nevertheless, proposed method utilizes features extraction only on short time domain that results in the impossibility of considering long-term trends of AE time series and exclude the possibility of forecasting the defect evolution. Moreover, the diagnostic feature extraction is conducted on noisy signal despite the preliminary filtration of high-amplitude interference is absent.

During the training data collection process AE systems can register outliers that are important to take into account, so they can lead to incorrect classification regarding the hazard class of defect [8].

In current paper as a part of providing a complex approach to technical condition estimation an original approach of training the SHM system of construction facilities is proposed. This approach is implemented through a subsequent application of methods that include feature extraction technique, feature selection, outlier detection in multidimensional feature set and anomaly detection methods. Their joint application allows to build a model of defect evolution recognition upon which proposed conducting the identification of the hazard class of defects.

2. Methods

Figure 1 demonstrates a diagram of the training process of decision support system for defect evolution detection. An important peculiarity of proposed approach is an opportunity of effective processing and analysis of AE time series acquired during the data collection process and building defect evolution recognition model on its basis. The proposed approach consists of four general stages considered below in detail.

![Figure 1. Sequential method of training the acoustic emission SHM system.](image)

2.1. Feature Extraction on short-term and mid-term domain

An important component of training set collection process in proposed approach is a method of transforming the source AE signal time series into diagnostic features by utilizing which the defect evolution recognition model is trained. At the heart of it lies a method of feature extraction proposed by us in paper [9] that analyzes properties of AE signal time series in two time domains within sliding windows with overlays. Frequency and time-frequency features are calculated in short-term windows which are then averaged in mid-term domain by means of calculating statistical parameters of probability distribution and aggregated in diagnostic features matrix [10]. The approach described above allows to identify long-term trends of alteration of the AE signal flux that is connected to degradation processes of operation defects in the controlled facility.

However, this method have constrained application: large dimension of the diagnostic feature matrix may lead to the increase of overfitting probability of classification model. The last results in dramatic decrease of decision making quality and significantly complicates the classification results interpretation process.
2.2. Optimal Feature Set selection
The solution of the problem noted above requires feature selection approach. Modern methods of dimensionality reduction divide on feature selection and method of mapping the source feature space to lower-dimensional by the means of constructing linear and non-linear combination of source features. It should be noted that the most effective feature selection methods are those which consider the extent of correlation impact on classification results quality. In this regard, feature selection method should utilize a metric that would allow to parametrize separability of the joint distribution of original feature space optimal projection.

Current paper proposes to utilize a greedy feature selection method together with the maximum mutual information criterion. The value of mutual information determines how close the joint distribution of the feature subset to the product of their marginal distributions, in particular, if distributions X and Y are independent then the mutual information coefficient is equal to zero.

So the problem of feature selection with criterion of maximum mutual information is formulated as follows:

\[
 f = \arg \max_f I(X_f, y)
\]

where \(|X_f| = k\) - size of optimal feature subset,
\(X_f\) - subset of AE signals selected features,
\(y\) - vector of defect hazard classes values

\[
 I(X_f, Y) = \sum_{i=1}^{|X_f|} \sum_{j=1}^{|Y|} P(i, j) \log \left( \frac{P(i, j)}{P(i)P(j)} \right)
\]

On step \(i\) the proposed algorithm increases the dimension of features by joining the feature which in the aggregate with features subset on step \(i - 1\) gives the maximal value of mutual information coefficient. The algorithm stops when \(|X_f| = k\).

2.3. Outliers removal
The problem of outlier removal occurs when forming the training set of the decision support system of defect evolution detection. As noted above, it appears due to individual specialties of the controlled facility and also under the influence of external factors. Those outliers which get into training set during data collection process may lead to classification errors of the monitoring system and respectively to incorrect estimation of defect hazard class.

Current paper proposes an approach of outlier removal based on modified method of calculating the deviation in probability distribution density of the potential outlier in relation to neighbor AE events. This method utilizes the local density level coefficient:

\[
 LOF_k(A) = \sum_{B \in N_k(A)} \frac{ld_k(B)}{ld_k(A)}
\]

where \(ld_k(A) = \left( \frac{\sum_{B \in N_k(A)} d(A, B)}{|N_k(A)|} \right)^{-1}\) - local density of AE event A, \(N_k(A)\) - set of k-neighbor points in relation to A, \(LOF_k(A)\) - local density level factor for event A.

AE events with local density level factor that exceed threshold are considered as outliers. The important problem that constrains practical application of the local outlier factor method is selection of the threshold value since incorrect threshold selection may lead to excessive AE data filtration and as a result may influence the correctness of the hazard class recognition. Authors of paper [11] recommend to set the threshold value equal or greater to one.
Current paper presents an original algorithm of the defect hazard class recognition based on probability distribution characteristics of calculated local outlier factor values.

\[
\begin{align*}
T_{i+1} &= \mu(LOF_i) + \text{step} \ast \sigma(LOF_i) \\
\left| T_{i+1} - T_i \right| &\neq 0
\end{align*}
\]

where \( T_i \) – threshold value on \( i \)-th step, \( LOF_i \) – corresponding local outlier factor value, \( \mu \) – mean, \( \sigma \) – standard deviation.

Each \( i \)-th step calculate new threshold value then algorithm filters out outliers that exceed this threshold value. After that new local outlier factor values are calculated from the rest of the feature set. The convergence of this optimization procedure is provided by the comparison of the threshold values between \( i \)-th and on \( i-1 \)-th steps.

2.4. Training the defect evolution detection model

Training supervised pattern recognition model is usually limited to application of binary or multiclass classification methods [12]. The problem of defect evolution detection on real controlled facility is complicated by the fact that prior information about defects of higher hazard class is generally absent. Sustainably functioning facility has only defects with hazard class not higher than one against noises of various origins [10]. Consequently, training methods based on binary and multiclass classification when a prior information of hazard class distributions exist are not applicable to this problem. Therefore it requires an additional training of algorithm that are used in defect evolution recognition system. There are indications that the approach that allows to solve this problem refers to anomaly detection methods [13]. AE events that forms the training set are usually conforms to the first hazard class as mentioned above. Events that appear during monitoring and do not belong to a prior known first hazard class events are anomalies in comparison to events from training set.

Anomaly detection represents a subclass of problems of pattern recognition that offer the opportunity of identifying the AE events that are different from general part of events in training set by statistical and metrical properties. Several methods of anomaly detection are currently used:

- Methods based on analysis of distance distribution density. This group of method expects that distances between normal observations in feature space are small enough while long distances are common to anomaly events [14]. These include K-nearest neighbor method and local outlier factor that utilizes reachability distance metric.
- Method that based on clustering, e.g. well-known K-means method. However it refers to unsupervised method since it may lead to significant mistakes in anomaly detection.

Current paper propose a modification of One-Class Support Vector Machine (One-class SVM) anomaly detection method. It represents one of the form of the classifier which requires only information about the events that belong to one class [15]. The main benefit of SVM-algorithm is the opportunity of the classifier to build non-linear separating boundaries that allows to significantly improve the results of classification.

However the results of original algorithm One-Class SVM depend considerably on the configuration parameters \( \nu \) and \( \gamma \) [15]. Proposed modification of One-Class SVM allows to find optimal values of configuration coefficients \( \nu \) and \( \gamma \), and thereby improves the quality of defects evolution recognition. The solution of corresponding optimization problem has been conducted by genetic algorithm.

\[
\begin{align*}
\arg \min_{w_{\nu,\gamma}} E(w_{\nu,\gamma}, \Phi(X), \bar{y}) \\
0 < \nu < 1 \\
\gamma \in \mathbb{R}
\end{align*}
\]

where \( E(\cdot) \) - error function, \( \bar{y} \) – true class value,

\( w_{\nu,\gamma} \) – weight vector under selected values of \( \nu, \gamma \),

\( \Phi(X) = \exp(\gamma \|X\|^2) \).
3. Results and discussion

Figure 2 demonstrates experimental acoustical time series obtained via technical condition diagnostics of constructions of the controlled facility after preliminary filtration and stationary noise removal. Time series on Fig. 2a, 2b and 2c correspond to three defects of different hazard classes. The process collection of acoustic emission data and controlled facility characteristics were described in previous paper [10]. These time series were utilized for training the defect evolution model via methods described above. Fig 2c demonstrates the process of transition between II and III hazard classes.

![Figure 2](image1)

**Figure 2.** Fragments of AE time series. Sampling frequency – 2.5MHz: a) defect of I hazard class b) defect of II hazard class c) transient defect from II to III hazard class

Figure 3 presents several of diagnostic features obtained after first stage of proposed method. They describe evolution of AE signal during the transfer from one hazard class to the higher one. Following set of spectral and time features were calculated on short term domain. These included energy of signal, correlation interval, entropy, kurtosis, skewness, spectral centroid. For the purpose of averaging following features were after calculated over short-term features in mid-term windows: maximum and minimum value, standard deviation, mean value, variation coefficient. The feature extraction stage resulted in 48 features.

It follows from fig. 3 that calculated statistical features on short and mid-term domains demonstrate defect evolution to higher hazard classes. It should be noticed that defect evolution recognition was conducted by means of long-term trends analysis of AE signals. This fact distinguishes proposed method of other approaches utilized in current AE diagnostic systems that are based on separate AE events analysis [1].

![Figure 3](image2)

**Figure 3.** Diagnostic feature values calculated in short and mid-term windows: mean value of spectral centroid, maximum of kurtosis, maximum energy
Within the stage of feature selection the original set of features was mapped to four-dimensional space via proposed feature selection method. It follows from the Fig. 4 that bigger values of mutual information coefficient have four features: variation coefficient of energy, kurtosis of correlation intervals, the variation of entropy and standard deviation of spectral centroid. It can be seen from the Fig. 5 that the standard deviation of spectral centroid has high correlation between other features. Each feature on Fig. 5 correspond to fragments of original time series fragments from Fig. 2. An additional adjustment on correlation values was conducted after the main phase of features selection stage. Thus, three features of the original feature set was utilized for training the defect evolution model.

![Figure 4](image1.png)
**Figure 4.** Feature weights obtained via proposed feature selection method

![Figure 5](image2.png)
**Figure 5.** Pair-wise features distribution plot

Figure 6 presents results of the stage III – outlier detection and removal from training set under the given confidence level of 0.0075. Blue color marks the outliers found during the algorithm execution, green color marks normal events corresponding to AE signals emitted from the defect.

![Figure 6](image3.png)
**Figure 6.** Outliers detection in optimal features space

It can be noticed that proposed modification of local outlier factor method allows to detect outliers in diagnostic features matrix after feature selection. Meanwhile, the mean value of local density factor was closer to 1.8477, while normal events had values closer to 1.062.
Fig. 7 represents the training results of defect evolution detection model (stage IV). The model described here as contour which is built on basis of training set via described above modified One-Class SVM method. AE events that corresponding to satisfying technical condition of constructions are located inside the contour while events that correspond to processes of defect forming and evolution are located outside. Anomaly events are marked by pink colour. Meanwhile, the classification error which was estimated by 10-fold cross-validation with “Accuracy” metric was 1.4%.

Conclusions

Current study presented a modified method of training the acoustic emission monitoring system of construction facilities based on step-by-step transformation of AE signals to representative training set by means of which the defect evolution detection model can be built.

Proposed method consist of four subsequent stages: features extraction on two time resolution domains, feature selection, outlier detection and learning approach based on anomaly detection method.

The practical approbation of proposed method was conducted on real fuel and energy infrastructure facility. It was verified by providing cross-validation results of training process with “Accuracy” metric.

This method can be utilized for building effective intellectual diagnostic monitoring system of facility constructions. The monitoring system enriched by proposed method allows to detect evolution of operating defects and prevent emergency situations that may have economic, social and ecological consequences.

References

[1] State Standard 31937-2011. Buildings and constructions. Rules of technical condition monitoring (in Russian).
[2] Consequences of oil spills in Norilsk rbc.ru: URL: https://www.rbc.ru/photoreport/03/06/2020/5ed7b4ac9a794710786cc0d (in Russian).
[3] Gholizadeh S and Baharudin, B 2015 A review of the application of acoustic emission technique in engineering Structural Engineering & Mechanics 54 pp. 1075-1095.
[4] Aggelis D 2016 Acoustic Emission Analysis for NDE in Concrete Innovative AE and NDT Techniques for On-Site Measurement of Concrete and Masonry Structures 20 pp. 69-88.
[5] Agletdinov E, Vinogradov A and Merson D 2019 Proc. of conf. Perspective materials and technologies pp. 561-562 (in Russian)
Acknowledgments
The reported study was funded by RFBR according to the research project № 20-37-90089.
Author thanks companies Strategia NK Ltd. and Diaform Ltd. for the help in conducting acoustic emission diagnostics on the controlled facility.