UDC 517.444 Hrynchenko P.V.[1], Bakurova A.V.[2]

APPLICATION OF WAVELET ANALYSIS FOR DATA VALIDATION IN RISK ASSESSMENT SYSTEMS

The object of the study is two information systems. The first system (S.1) is a functioning system, the state of which can be normal or abnormal, described by a set of system parameters at a certain point in time. That is, there are time series describing the normal behavior of this system and deviations from it. The second system (S.2) is designed to detect abnormal activity in the first system. The outputs of this anomaly detection system are presented by a time series, each of the values of which characterizes the state of the system at a certain point in time by two parameters – the confidence of the second anomaly detection system in assessing the state of the first system and the severity of the anomaly that occurred in the first system.

The aim of the study is to develop a method for assessing the stability of an anomaly detection system based on the analysis of changes system confidence and severity of the detected anomalies using wavelet analysis for data validation in risk assessment system (third system – S.3).

The method is developed using the Daubechies wavelet transform of the 4th order (DB.4), which allows:

– separate normal system behavior and abnormal states through the analysis of various frequency components of the signal;

– decompose the signal into approximating and detailing coefficients, which helps to identify system instability.

The use of DB.4 wavelets in this study is due to their unique properties:

– compact carrier – allows for effective analysis of local signal features;

– good frequency localization – wavelets are able to decompose a signal into frequency components without losing information;

– orthogonality – ensures the possibility of accurate signal reconstruction;

– smoothness – DB.4 is smooth enough for time series analysis [1].

Fourth-order wavelets provide a balance between time localization, frequency resolution, and computational complexity. They are particularly suitable for analyzing data associated with abrupt changes, which is typical for anomaly detection systems.

[1] Postgraduate Student of the Department of System Analysis and Computational Mathematics, National University «Zaporizhzhia Polytechnic»

[2] Professor of the Department of System Analysis and Computational Mathematics, National University «Zaporizhzhia Polytechnic»

The developed method for assessing the stability of the anomaly detection system includes several key stages. First of all, the DB.4 wavelet transform is applied, which allows separating long-term trends and short-term fluctuations in the system's confidence. This makes it possible to identify uncharacteristic changes in its operation, which is especially important when analyzing the behavior of the anomaly detection system.

The next step is statistical analysis, which determines the correlation between the confidence of the system and the severity of the detected anomalies. In addition, the energy of the components is calculated, which characterizes the intensity of high-frequency changes in the signal, which allows for a more accurate assessment of the level of instability of the system.

Stability criteria are used for a comprehensive assessment of the system's performance (S.2):

1) correlation coefficients between system confidence and anomaly severity in (S.2);

2) the results of the energy analysis of the signal components, as well as the assessment of the level of instability S.2 through the threshold energy values.

The behavior of the system (S.2) is analyzed during transitions between the state of normal system operation and its abnormal behavior, which allows for the analysis of system behavior, such as the relationship between the level of confidence of the system in detecting anomalies and the severity of anomalies.

Determination of instability levels (S.2) is carried out based on the analysis of the energy of the components, which makes it possible to quantitatively assess the stability of the system (S.2) to external disturbances. In addition, a correlation analysis of the confidence and severity of anomalies is performed, which helps to determine the degree of consistency of the system's conclusions.

The developed data validation method has been tested for integration with the ACRAM risk assessment system (S.3) [2]. ACRAM is a risk assessment methodology being developed within the Telemetry project, aimed at adaptive threat management and predictive risk analysis in complex information systems.

Main areas of application of the developed method in ACRAM:

– preliminary data validation, which allows determining the reliability of input data and identifying periods of instability (S.2);

– data filtering will allow excluding periods with high instability from risk assessment analysis;

– Integrated assessment can be used in combination with ACRAM for two-tier risk analysis;

– validation of risk assessment results will improve the accuracy of conclusions (S.3).

The developed method was tested by analyzing the test data obtained as a result of work (S.2). The study revealed a significant negative correlation (-0.853) between the confidence of the system (confidence) and the severity of the detected anomalies (severity). This means that as the severity of the detected anomaly increases, the confidence of the system (S.2) decreases. This result confirms that the system adequately responds to anomalies, reducing confidence when moving to more critical states.

In the context of the wavelet transform, we can say that the signal is a sum of different components, each of which reflects certain features of the signal at different scale levels (or frequencies). These components contain information about high-frequency and low-frequency variations of the signal.

The sum of the energy of these components is equal to the total energy of the original signal, which means that the wavelet transform is an energy-saving process.

Wavelet analysis allows us to calculate the energy of components – a metric that reflects the level of instability of the system [3] using the formula (1).

$$E_x = \sum_{n=-\infty}^{\infty} x^2[n], \tag{1}$$

where x is a real number.

For different time windows, the mean energy value (2.968) and standard deviation (0.0894) [4] were calculated using the formula (2).

$$STD = \sqrt{\frac{\sum(X - \bar{X})^2}{n}}, \tag{2}$$

where:

X – individual values of the energy window calculated as the sum of squares based on the time series of the component energies,

$\bar{X}$ – average value of window energy,

n – number of windows.

Using formulas (1) and (2), threshold values were determined instability S.2:

– lower threshold: 2.879;

– average threshold: 3.057;

– high threshold: 3.147.

The average energy per window (2.968) was between the lower and average thresholds, indicating an average level of system stability (S.2). During the study, it was possible to determine the time intervals in which the system demonstrated instability. This is important for subsequent improvement of the anomaly detection method. (S.2).

Conclusions. Thus, the developed method for assessing the stability of the anomaly detection system has proven its effectiveness, allowing to assess the stability of the anomaly detection system (S.2). During the analysis, it was confirmed that the use of the DB.4 wavelet transform effectively solves the problem of separating normal and abnormal behavior of the system, ensuring accurate detection of unstable states. One of the key results of the study was the identification of a significant negative correlation (-0.853) between the confidence of the system and the severity of the detected anomalies. This means that when moving to more critical states, the level of confidence of the system decreases. An additional calculation of the energy of the components showed that (S.2) demonstrates an average level of stability. The calculated threshold values made it possible to classify the time intervals of instability, which makes it possible to further take them into account when assessing the reliability of the system. The goal of further research is to integrate the proposed method for assessing the stability of the anomaly detection system with the ACRAM system (S.3).

This work has been supported by the Horizon Europe project TELEMETRY.

REFERENCES

1. Image Reconstruction Using Wavelet Transform with Extended Fractional Fourier Transform. (2014). Blekinge Institute of Technology. Rajesh T.

2. Fuzzy logic-based methodology for building access control systems based on fuzzy logic. (2024). 6th International Workshop on Modern Data Science Technologies. Lviv-Shatsk, Ukraine.: Lytvyn, V., Bakurova, A., Zaritsky, O., Gritskevich., A., Hrynchenko, P., Tereschenko, E., Shyrokorad, D.

3. Energy of a signal. [Online]. Available: https://www.gaussianwaves.com/2013/12/power-and-energy-of-a-signal/.

4. Standard deviation. [Online]. Available: https://berg.com.ua/indicators-overlays/stdev