MACHINE LEARNING METHODS FOR PROBABILISTIC ASSESSMENT OF SOLID DAMAGE INDICATORS

D. Marmysh

Belarusian State University, Minsk, Belarus

Joint Institute of Dalian University of Technology and Belarusian State University

In the exploring of multicomponent systems in which a complex stress-strain state arises due to the interaction of various loading conditions, the concept of a body with a dangerous volume is a convenient and effective mathematical tool for the researcher. Damage analysis of a multicomponent system allows optimization of loading conditions.

Damage analysis is based on knowledge about the distribution of stress and strain fields in a solid $\sigma_{ij} = \sigma_{ij}(x, y, z)$, $\varepsilon_{ij} = \varepsilon_{ij}(x, y, z)$ where $i, j = 1, 2, 3$, $x, y, z \in \square$. Further, based on the quantitative assessment of the stress-strain state, it is possible to carry out a quantitative analysis of damage according to the formulas [1]

$$V = \int_{\Omega} dV \text{, где } dV = \begin{cases} dxdydz, & \sigma_{ij}(x, y, z) \geq \sigma_{\mathrm{lim}}, \\ 0, & \sigma_{ij}(x, y, z) < \sigma_{\mathrm{lim}}, \end{cases}$$

$$\Psi = \int_{\Omega} \varphi dV \text{, where } \varphi = \frac{\sigma_{ij}}{\sigma_{\mathrm{lim}}}.$$

In the mathematical model of damage to a continuous medium, the limiting stress σ_{lim} can be chosen or calculated based on the selected strength theory.

The representation of a solid continua in the form of a discrete model allows you to apply various approaches and apply High-performance computing (HPC) methods, e.g. parallel computing methods. HPC approaches can also be implemented algorithmically using special CUDA technology, which allows to speed up computing processes many times and reduce computer cost. For example, figure 1 shows the acceleration coefficient in potential distribution calculation for the half-space. This problem arises when discretizing a continuous distribution and approximating potential distribution by boundary elements. Figure 1 shows that with the use of CUDA technology, the computing speed can be increased by more than 100 times.

Parallelization of calculations can be applied due to the fact that when sampling a continuum model, quantitative indicators of damage to a single discrete element can be calculated independently of others. This simple fact allows you to use multithreading when performing computational procedures.

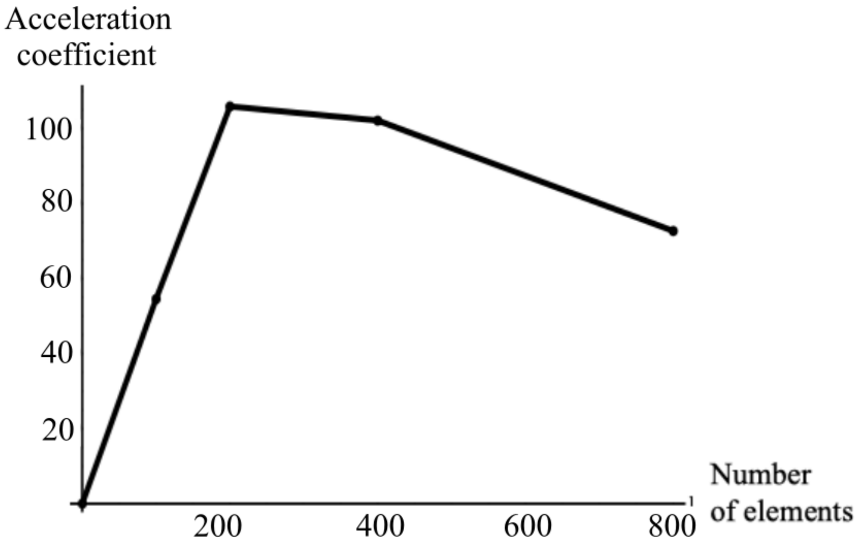

Fig. 1. Acceleration coefficient in potential distribution calculation for the half-space

The above formulas for damage indicators are an analytical model that allows you to accurately determine the quantitative characteristics and configuration of the damaged area. However, in real engineering applications, damage can be calculated with a certain degree of probability, so statistical methods and machine learning methods can be used to quantify it.

Monte Carlo methods are effectively used to calculate spatial integrals in complex geometry. Monte Carlo methods ensure the accuracy of the order $C/\sqrt{N}$, where *N* is the number of calculated nodes and, according to the law of large numbers, the numerical values of the damage integrals obtained by the Monte Carlo method converge to the exact value with an increase in the number of *N* [2].

Numerical approaches to assessing the damage to the medium are based on a preliminary calculation of the stress state in the selected node. If the operating voltage exceeds the limit, then the cumulative value of the dangerous volume increases by the value of the volume of the discrete cell according to simple summation formulas

$$V = K \cdot dV\,,\ \Psi = dV \cdot \sum_{i=1}^{K} \frac{\sigma_{ij}}{\sigma_{\lim}},$$

where K is the number of points at which the stress at point σ_{ij} exceed the limit $\sigma_{\lim}$ ($K < N$).

From a geometric point of view, the function $\sigma_{\lim} = f_{\lim}(\mathbf{x})$ sets some surface bounding the area $D \subset \mathbf{R}^m$. From a physical point of view, the area of Ω should be limited. Thus, when calculating the dangerous volume and integral damage, it is initially necessary to determine whether the point in question belongs to the area of Ω or not, i.e. to conduct a binary classification with one of two possible answers

$$\delta(A)=\begin{cases}0, & A\notin\Omega,\\ 1, & A\in\Omega.\end{cases}$$

The problem of binary classification can be solved by various methods. One of the most effective methods of machine learning is the method based on logistic regression. The logistic regression model allows you to calculate the probability $P(\mathbf{x},\boldsymbol{\alpha})$ from 0 to 1 belonging to the damaged area. Probability $P(\mathbf{x},\boldsymbol{\alpha})$ is a continuous function, therefore, 0.5 is usually chosen as the threshold value, i.e. $A\in\Omega$ if $P(\mathbf{x},\boldsymbol{\alpha})\geq 0.5$ [3, 4].

An approach based on the calculation of probability using the logit-model was applied to the calculation of the damage of an elastic half-plane under the action of two types of forces distributed over the surface $q(x)=q_0\sqrt{1-x^2/a^2}$ (model of a non-conformal contact case) and $q(x)=q_0\Big/\sqrt{1-x^2/a^2}$ (model for rigid stamp interaction). In both cases, the logit-model looks like

$$P(x_i,y_i,\boldsymbol{\alpha})=\left[1+\exp(-\alpha_0-\alpha_1x_i-\alpha_2y_i)\right]^{-1},\ i=1,2\,.$$

The quality of the logit model can be assessed based on ROC analysis [5]. In this case, a ROC curve is constructed in the axes of specificity (S_p) and sensitivity (S_e), which determine how the proportion of truly positive and truly negative cases identified by the model, respectively.

$$S_p=\frac{TN}{TN+FP},\ S_e=\frac{TP}{TP+FN},$$

where TN is the number of correctly classified negative responses;
TP is the number of correctly classified positive responses;
FP is the number of negative responses classified as positive;
FN is the number of negative responses classified as negative.

Figures 2 and 3 show the ROC curves for both cases of the normal pressure distribution along the boundary of the half-plane.

To quantify the quality of the classifier, the size of the AUC area located between the ROC curve and the S_p specificity axis is determined.

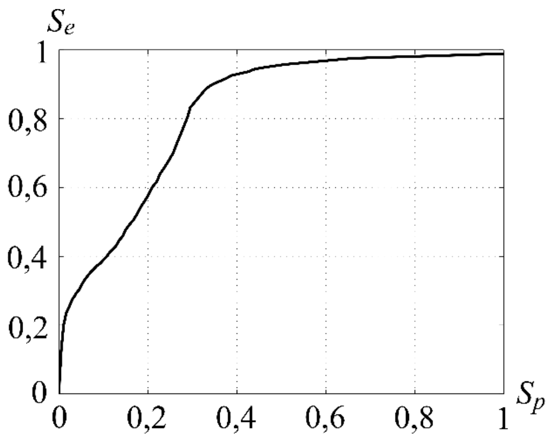

Figure 2 – ROC-curve of damageability for non-conformal contact case

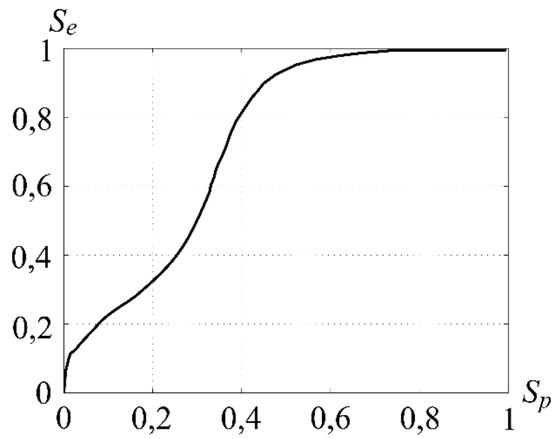

Figure 3 – ROC-curve of damageability for rigid stamp interaction cases

A set of training precedents for building a logistic regression model for assessing the damage to the environment was randomly generated from a uniform distribution over a two-dimensional area containing a dangerous volume. For a random sample of use cases, it is important to understand the stability of numerical algorithms in estimating the parameters of the logit-model and, accordingly, in terms of classification quality. Figures 4 and 5 show box diagrams for the accuracy of classification and the area under the ROC curve with a different number of training precedents, from which it is possible to determine the degree of dispersion and asymmetry in the assessment of these indicators.

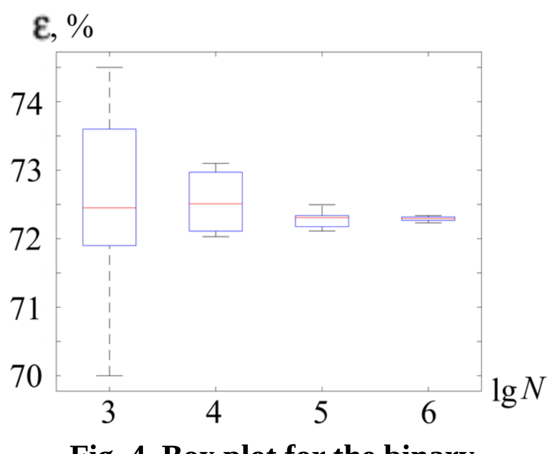

Fig. 4. Box plot for the binary classification accuracy

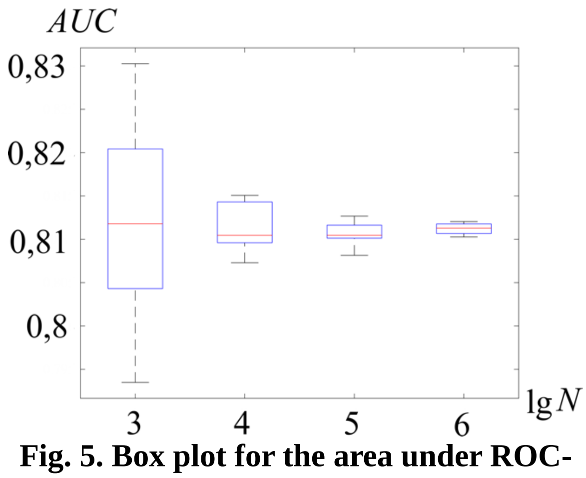

Fig. 5. Box plot for the area under ROC-curve AUC

It can be seen from diagrams 4 and 5 that with an increase in the number of use cases for training the model, the spread in the assessment of classification accuracy (ε) and in the quality of the regression model (*AUC*) is decreasing. For classification accuracy with a training sample size of 103, the difference between the first and third quartile is 1.7%, for a sample size of 106 – 0.05%. For the AUC indicator, similar indicators are 0.017 and 0.0011, respectively. The decrease in the spread of indicators allows us to get the conclusion about the stability of numerical algorithms for estimating the parameters of the logistic regression model and the quality of binary classifiers.

References

1. Sosnovskiy, L. A. Tribo-Fatigue. Wear-Fatigue Damage and its Prediction / L. A. Sosnovskiy // Series: Foundations of Engineering Mechanics, Springer, 2005. — 424 p.

2. Fishman G. S. Monte Carlo Concept, Algorithms and applications. Springer, 1996. 698 p

3. Hosmer, D. W. Applied logistic regression / D.W. Hosmer, S. Lemeshow. – 2nd edition. – Wiley: NY, 2000. – 375 p.

4. Niu, L. A review of the application of logistic regression in educational research: common issues, implications, and suggestions / L. Niu // Educational review. – 2020. – Vol. 72, iss. 1. – Pp. 41–67.

5. Zweig, M.H. ROC plots: a fundamental evaluation tool in clinical medicine / M.H. Zweig, G. Campbell // Clinical chemistry. – 1993. – Vol. 39, iss. 4. – Pp. 561–577.