THE USE OF FUZZY CLUSTERING FOR BUILDING A KNOWLEDGE BASE IN THE PROBLEMS OF PREDICTING THE DURABILITY OF CORRODING STRUCTURES

Korotka L., Candidate of technical sciences, Associate Professor; Karasov O.
Ukrainian State University of Chemical Technology, Ukraine, Dnipro

Quite a lot of works are devoted to the study of methods for processing multidimensional data arrays in various subject areas, and these studies are far from being completed [1, 2]. Before starting to work with such data, if possible, preliminary processing should be carried out, for example, removal of abnormal or noisy elements, etc.

One of the directions of processing multidimensional arrays of various structures is cluster data analysis, which allows you to space heterogeneous data into subsets of a more or less homogeneous structure.

There are a sufficient number of clustering methods with different metrics that can be grouped as clear and fuzzy. The use of clear clustering methods allows you to split the original set of objects into several, as a rule, non-overlapping subsets. In this case, it is assumed that any object from the resulting partition into subsets belongs to a single cluster. Fuzzy clustering methods allow you to divide the source data into subsets in such a way that an element can belong simultaneously to several clusters with different degrees of membership. It is obvious that the use of clear or fuzzy clustering methods entirely depends on the problem being solved. The advantages and disadvantages of clustering methods are well known.

Solving practical problems using one or another clustering algorithm, it is necessary to determine the metric. It is known that the simplest is the Euclidean metric, but at the same time, it allows one to obtain a partition in the form of hyperspheres, which is not always practical. Algorithms are known that use a norm to obtain clusters that allows one to obtain subsets of the partition of various shapes. In the absence of information about the best number of subsets of the partition, it is advisable to use a criterion for assessing the quality of clustering, which can be represented by an evaluation function.

One of the most common fuzzy clustering algorithms is the c-means fuzzy clustering method (fuzzy clustering, soft k-means, c-means, fcm algorithm), which is

a generalization of R. Yager's and is also well known. In fact, it consists of the following two stages:

- the data point with the maximum potential to represent the first cluster is selected;

- to determine the next fuzzy cluster and the coordinates of its center in the neighborhood, which is specified by the parameter γ, all data points in the vicinity of the center of the first cluster are deleted.

These two procedures are repeated until all data points are inside the vicinity of the radius of the parameter γ of the sought cluster centers. In the general case, small values of this parameter γ lead to the finding of a small number of clusters larger in terms of the number of points (usually γ = 0.5).

The paper discusses corrosive constructs that operate in a corrosive environment [3, 4]. The input variables are: initial σ_0 and ultimate stresses in the element [σ], corrosion rate v_0, square A_0 and perimeter P_0 bar structural elements, error [ε] numerical solution of a system of differential equations (SDE), which describes the process of accumulation of geometric damage to the structure. The output variable is the integration step h_t to obtain an estimated predicted durability value.

In the numerical solution of SDE, the problematic aspect is the error of the obtained solution [3, 4], which, in turn, is a function of many variables $\varepsilon(v_0, A_0, P_0, \sigma_0, [\sigma], h_t)$. Each variable can be written as a linguistic variable, term-sets, which are defined using fuzzy clustering.

The center of the corresponding cluster is considered as the core of the term set. The carrier of the set is the cluster boundaries. The resulting clusters are considered as a basic knowledge base [4], which requires additional tuning [4, 5]. The α-level principle of generalization is used for the construction of membership functions.

The rule base should provide the ability to achieve the required accuracy of the fuzzy model, in particular, after the parameters of the latter have been determined. At the same time, in order to reduce computational costs and present the model in a more intuitive way: the number of rules contained in the knowledge base should be

as small as possible. In addition, reducing the number of rules in a multi-input model may be a prerequisite for tuning its parameters.

References

1. Solodka N.A., Shapovalov D.D. Analiz i obrabotka dannyh kompanii kak sposob povyshenija ee pribyl'nosti [Analysis and processing of company data as a method of profitability improving]. Elektronnij naukovo-praktichnij zhurnal «Shidna Єvropa: ekonomika, biznes ta upravlinnja. 2019. № 6 (23) S. 756–760. (in Russian) (in Scholar)

2. Us S.A., Korjashkina L.S. Modeli ta metody pryjnjattja rishen'(druge vydannja) [Models and methods of decision making] // Dnipropetrovs'k, NGU, 2018, 270 s. (in Ukranian) (in Scholar)

3. Korotka L.I., Korotka Y.A. The use of elements of computational intelligence in problems of forecasting of corroding constructions durability. Mathematical and computer modelling. Series: Technical sciences – 2017. Issue 16. – P 64-71. (in Scholar)

4. Korotka L. I. The use of fuzzy clustering in solving problem in predicting the durability of corrosive structures. Mathematical modeling. №2(43)2020, С. 44-54. (DOI: https://doi.org/10.31319/2519-8106.2(43)2020.219266) (in Scholar)

5. Pegat A. Fuzzy modeling and control. Springer Studies in Fuzziness and Soft Computing, 2001. 728 p.