UDC 681.142

Metrological conformity of classification of objects according to properties represented by ordinal and nominal scales

N. Yaremchuk, E. Volodarsky, O. Hoda

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Ave., 37, 03056, Kyiv, Ukraine
yaremchukna@i.ua; vet-1@ukr.net; R_olia@i.ua

Abstract

To achieve metrological conformity, the procedure of classification of objects, based on observations of ordinal and nominal properties, provides for verification of the classified data performing reference measurements followed by further consideration of their uncertainties. The result of the verification is the compliance of the categories of nominal and ordinal properties, to which their manifestations are attributed according to observations, with the categories established by the reference procedure. To account for the uncertainty of the reference procedure, a matrix of the reference relation is composed, which characterizes the correspondence of the categories established by the reference procedure to the true categories according to their definition. To obtain the classification reliability indicators, it is suggested to combine the ordinal variances of the verification matrices and the reference ratio followed by the estimation of conditional probabilities characterizing the classification reliability.

Keywords: metrological conformity; ordinal properties; nominal properties; classification reliability indicators; uncertainty.

Received: 08.12.2022 Edited: 23.12.2022 Approved for publication: 28.12.2022

1. Introduction

Modern methods of data processing include information that is presented not only in a numerical way, but, to a large extent, using categorized variables, which is according to their individual properties have been attributed to a specific group or category in ordinal or nominal scales [1].

If a relation of order is established between the categories of the scale, the resulting scale is a quasi-order (or pre-order) scale. Such a scale is an intermediate scale between nominal and ordinal scales. Although nominal scales are less informative than higher-order ones, they can provide a quick assessment of the property under consideration and all the necessary interpretation for decision-making. In addition, when conducting research to make decisions about the state of the object according to its properties; when determining the quality of products according to the established categories; when constructing control charts of technological processes based on qualitative data, there is a need to process the obtained data on quasi-order scales, if they are consistent in a metrological way [2—5].

If the goal of the used procedure of experimental informatics is to attribute an object to a certain category according to its properties, then such a procedure is called classification [6].

To assess the reliability of the classification, it is necessary to have physical measurement standards that would reproduce the categories of properties according to their definition. Metrological conformity of classification results in the presence of such standards can be achieved by calibration or verification. In most cases, calibration is used as a training procedure for operators. It is carried out before the classification of technical systems, which *a priori* allows obtaining classification reliability indicators that account for the uncertainty of measurement standards. Verification is used to check the obtained classification results. It is necessary to combine the classification reliability indicators, obtained during verification, with the uncertainty characteristics of measurement standards.

The paper considers the verification of classification according to observation of manifestations of nominal properties, for which a reference procedure is used based on the measurement of these manifestations followed by using the established classification scale. The result of the verification is the compliance of the categories of nominal properties, to which their observational manifestations are attributed, with the categories established by the reference procedure.

Table 1

Results of verification of the procedure for visual observation of the roughness level of treated surfaces

Categories to which the level of surface is assigned according to visual observation, (j =1.2) / Categories to which the level of surface treatment is assigned according to the reference instrument and classification scale (i =1.2)	Semi-clear (j =1)	Rough (j =2)	Sum
Semi-clear (i =1)	80(P_{11} = 0.8)	20(P_{12} = 0.2)	100
Rough (i =2)	10(P_{21} = 0.1)	90(P_{22} = 0.9)	100
Sum	90	110	200

2. Verification of classification by observing the manifestations of nominal properties using the reference procedure

Visual and tactile observation of the roughness level of treated surfaces implies a selective check when the same points of the treated surfaces are subject to repeated measurements and classification, with the categories being determined by the indicators of the reference instrument and the classification scale (Table 1).

The stochastic matrix in Table 1 allows calculating the classification quality indicators.

The variance of the stochastic matrix is estimated by rows using the formula of the variance measure for the ordinal scale [7, 8]:

$$D = \frac{4}{m-1}\sum_{k=1}^{m-1} F_k (1 - F_k), \qquad (1)$$

where F_k is the cumulative relative frequency for the k-th level. The value of D is normalized in the range from 0 to 1.

According to (1) and Table 1 we obtain $D_1 = 0.64$, $D_2 = 0.36$.

To assess the quality of the classification procedure, the Frobenius norm is used, which is equal to 0 for a correct classification:

$$G = \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{m}(P_{ij} - I_{ij})^2, \qquad (2)$$

where I_{ij} corresponds to the unit matrix of an ideal classification: $I = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix}$.

In accordance with (2) and Table 1, we obtain $G = \frac{1}{2}(P_{12}^2 + P_{21}^2) = 0.10$.

Verification of the classification procedure was carried out, only if the categories, defined by the reference instrument, corresponded to the true ones according to their definition.

However, if we analyse a metric classification based on the results of measuring the level of roughness [5, 8], we can obtain a real matrix of the reference relation, which becomes ambiguous due to the measurement uncertainty of the reference instrument (Table 2).

The variance of the reference matrix by rows is $D_1^{em} = 0.28$, $D_2^{em} = 0.059$. The Frobenius norm of the reference matrix is $G^{em} = 0.0029$.

Thus, the variance of the matrix of reference data is much smaller than the variance of the matrix of experimental data to be verified, which should be so.

To determine the reliability of the classification procedure under metrological conformity, one can use the composition of ambiguous relations obtained from the data in Table 1 and Table 2, namely $R_3 = R_1 \circ R_2$.

However, it should be noted that the convolution of the matrices provides only approximate accounting for the uncertainty of the measurement standards used in verification. Therefore, a method to account for the uncertainty of measurement standards using the total variance of verification and reference data by rows of a stochastic matrix is suggested, namely $D_1^{в} = D_1 + D_1^{em}$, $D_2^{в} = D_2 + D_2^{em}$.

Then, considering the formula for the ordinal variance (1), we obtain the probabilities of cor-

Table 2

Reference ratio, accounting for the uncertainty of the reference instrument

Categories according to reference measurement instrument, (i =1.2) / True categories according to their definition, (l =1.2)	i =1	i =2
l =1	0.925	0.075
l =2	0.015	0.985

Table 3

Matrix of stochastic classification under metrological traceability

True categories according to their definition (i =1.2) \ Categories, to which the level of surface treatment is assigned according to visual observation (j =1.2)	Semi-clear (j =1)	Rough (j =2)
Semi-clear (i =1)	P_{11} = 0.64	P_{12} = 0.36
Rough (i =2)	P_{21} = 0.12	P_{22} = 0.88

rectly assigning the roughness level of the treated surfaces to certain categories by verifying the classification procedure based on visual observation: $P^{в}_{11(12)} = 0.5 \pm \sqrt{0.25 - D^{в}_{1}/4}$; $P^{в}_{22(21)} = 0.5 \pm \sqrt{0.25 - D^{в}_{2}/4}$.

According to the data in Table 1 and Table 2, we obtain $P_{11} = 0.64$; $P_{12} = 0.36$; $P_{22} = 0.88$; $P_{21} = 0.12$.

Then the stochastic matrix of the verification results accounting for the uncertainty of measurement standards takes the form of Table 3.

The probabilities of a correct classification of P_{11} and P_{22} have decreased in comparison with Table 1, provided that the uncertainty of the measurement standards used in verification is accounted for. Thus, a method for assessing the reliability of the classification procedure under metrological conformity is suggested.

3. Conclusion

If the result of applying a nominal scale is to obtain one or more categories of the scale, to which the nominal property under consideration can be attributed, then such a procedure is called a classification procedure. An indicator of the reliability of the classification may be the probability of correct assigning to a certain category, which can be determined by verifying the classification procedure using measurement standards corresponding to the categories of the scale or by using the reference procedure.

The paper considers the verification of classification by observing the manifestations of nominal properties, using the reference procedure based on the measurement of these manifestations followed by using the established classification scale that ensures metrological conformity of the classification procedure. The result of the verification is the compliance of the categories of nominal properties, to which their observational manifestations are attributed, with the categories established by the reference procedure.

To obtain the characteristics of the reliability of the classification, a stochastic classification matrix is composed, the components of which are the conditional probabilities of assigning the manifestations of the nominal property to a certain category, while these manifestations were assigned to the same or another category according to the reference procedure. The diagonals of the matrix correspond to the probability of correct assignment of the nominal property to a certain category.

To establish metrological conformity, it is necessary to compose a matrix of correspondence between the categories established by the reference procedure and true categories according to their definition. For an ideal reference procedure, a full correspondence is established, which is represented by the units in the diagonal of the correspondence matrix.

In case of imperfections of the reference procedure, to account for the uncertainty of the classification results according to the reference procedure, a matrix of the reference relation is composed, the components of which are the conditional probabilities of conformity of the categories established by the reference procedure to the true categories according to their definition.

The paper suggests a method that allows accounting for imperfections of the reference procedure by combining the ordinal variances of the verification matrices and reference ratio by rows, followed by the estimation of conditional probabilities characterizing the reliability of the classification under metrological conformity.

Метрологічна узгодженість класифікації об'єктів за властивостями, що відображаються порядковими та номінальними шкалами

Н.А. Яремчук, Є.Т. Володарський, О.Ю. Года

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", пр. Перемоги, 37, 03056, Київ, Україна
yaremchukna@i.ua; vet-1@ukr.net; R_olia@i.ua

Анотація

Номінальні властивості відображаються за номінальними шкалами, які є прийнятим за домовленістю рядом категорій номінальної властивості, який можна подати назвами категорій, числами або символами, що є ідентифікаторами положення категорій у ряду, з яким порівнюються конкретні прояви номінальної властивості.

Хоча й номінальні шкали є менш інформативними, ніж шкали вищого порядку, вони можуть забезпечити швидке оцінювання досліджуваної властивості та необхідну інтерпретацію для прийняття рішень. Крім того, при проведенні досліджень із метою прийняття рішень про стан об'єкта за його властивостями виникає необхідність в опрацюванні даних, отриманих за шкалами квазіпорядку, за умов їхньої метрологічної простежуваності.

Якщо кінцевим результатом застосування номінальної шкали є отримання однієї або декількох категорій шкали, до яких можна віднести досліджувану номінальну властивість, то така процедура є процедурою класифікації. Показником достовірності класифікації може бути ймовірність правильного віднесення до певного класу еквівалентності, яку можна визначити за верифікацією процедури класифікації, використовуючи еталони, що відповідають категоріям шкали, або за еталонною процедурою.

Розглянуто верифікацію класифікації за спостереженням проявів номінальної властивості, у рамках якої використовується еталонна процедура, заснована на вимірюванні цих проявів із подальшим використанням встановленої шкали класифікації. Результатом верифікації є відповідність категорій номінальних властивостей, до яких віднесені їхні прояви за спостереженням, категоріям, встановленим за еталонною процедурою.

Для встановлення метрологічної простежуваності необхідно скласти матрицю відповідності категорій, закріплених за еталонною процедурою, істинним категоріям за їхнім визначенням. У роботі запропоновано метод урахування неідеальності еталонної процедури, застосовуючи об'єднання порядкових дисперсій матриць верифікації, та еталонного відношення за рядками з подальшою оцінкою умовних ймовірностей, що характеризують достовірність класифікації.

Ключові слова: метрологічна узгодженість; порядкові властивості; номінальні властивості; невизначеність; показники достовірності.

References

1. Dybkaer R. Concept system on "quantity": formation and terminology. *Accreditation and Quality Assurance*, 2013, vol. 18, pp. 253–260. doi: 10.1007/s00769-013-0977-8
2. Torra V. et al. Regression for ordinal variables without underlying continuous variables. *Information Sciences*, 2006, vol. 176, issue 4, pp. 465–474. doi: 10.1016/j.ins.2005.07.007
3. Bashkansky E., Gadrich T. Some metrological aspects of ordinal measurements. *Accreditation and Quality Assurance*, 2010, vol. 15, issue 6, pp. 331–336.
4. Franceschini F., Galetto M. Ordered Samples Control Charts for Ordinal Variables. *Quality and Reliability Engineering International*, 2005, vol. 21, pp. 177–195. doi: 10.1002/qre.614
5. Bashkansky E., Gadrich T. Evaluating quality measured on a ternary ordinal scale. *Quality and Reliability Engineering International*, 2008, vol. 24, pp. 957–971. doi: https://doi.org/10.1002/qre.946
6. Benoit E. New scale classification within the representational theory of measurement. *Journal of Physics: Conference Series*, 2013, 459(1):012004. doi: 10.1088/1742-6596/459/1/012004
7. Gadrich T., Bashkansky E. ORDANOVA: Analysis of Ordinal Variation. *Journal of Statistical Planning and Inference*, 2012, vol. 142, pp. 3174–3188. doi: 10.1016/j.jspi.2012.06.004
8. Blair I., Lacy M. Statistics of Ordinal Variation. *Sociological Methods & Research*, 2000, vol. 28, issue 3, pp. 251–280. doi: 10.1177/0049124100028003001