UDC 681.142 DOI:

The methods of data processing according to the measurement/classification procedure for quality indicators of objects

O. Hoda, R. Dovhaliuk, N. Yaremchuk

National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Peremohy Ave., 37, 03056, Kyiv, Ukraine
R_olia@i.ua; ruslanadovhaliuk@gmail.com; yaremchukna@i.ua

Abstract

The measurement/classification procedure of object properties, such as quality indicators, is used for determination of object quality category. According to measurement uncertainty, the fuzzy classification scale is constructed. In accordance with this scale, fuzzy classification results are obtained. The use of fuzzy averaging for multiple measurements of individual quality indicators are proposed. The fuzzy logic operators that can be used in the construction of group quality indicators are considered and recommendations for their application are given. As an example of application of the proposed method, the definition of water quality categories on its biochemical properties is used.

Keywords: measurement/classification procedure; measurement uncertainty of a fuzzy scale; fuzzy quality indicators; group quality indicators.

Received: 24.12.2021 Edited: 16.03.2022 Approved for publication: 28.03.2022

1. Introduction

The paper is dedicated to solving the problem of assessing the quality indicators of objects on scales defined by several ordered categories [1] taking into account measurement uncertainties [2]. As an example, the procedure of determining a category of water quality by individual or group indicators was chosen. Thus, according to the content of harmful substances, for example, "nitrate content NH_4^+ in surface waters mg/dm³", the following categories of water quality are defined [3]: T_1 – "Class I – very clean" (less than 0.1), T_2 – "Class II – clean" (0.10–0.30), T_3 – "Class III – contaminated" (0.31–1.0), T_4 – "Class IV – dirty" (1.01–2.50). In accordance with these classes and categories, the application of individual quality indicators solves the problem of assigning water quality to the appropriate verbal category for one component that affects water quality. At the same time the defined scale with a certain linguistic variable is used. When using group indicators of water quality, the problem of assigning water quality to the appropriate verbal category by several components is solved. However, the components of measurement uncertainty that underlie the classification and the ambiguity of the expert information that may be used in the classification are not taken into account. Thus, the following components of uncertainty can be identified: instrumental uncertainty due to the uncertainty of measuring the property reflected by the quality indicator; semantic uncertainty arising from discrepancies in the boundaries of the categories given in different sources.

When measuring the properties that characterize the biochemical composition of water, the measurement uncertainty is significant and averages 20–30%. Then the decision to assign water quality to a certain category is made on a scale with a fuzzy linguistic variable, and the uncertainty of the measurement/classification procedure is reflected in the scope of degrees of belonging of measurement results to different categories of the classification scale [2, 3].

2. The method of data processing according to measurement/classification of individual quality indicators

In accordance with the method [4], taking into account the measurement uncertainty, the following term set of the linguistic variable "content NH_4^+ in surface waters mg/dm³" is formed:

$$\mu_{T_1} NH_4^+ = \begin{cases} 1, \textit{ for } 0 \le x \le 0.06 \\ (0.14 - x)/0.08 \textit{ for } 0.06 < x \le 0.14 \\ 0 \textit{ in other cases} \end{cases}$$

$$\mu_{T_2} NH_4^+ = \begin{cases} (x - 0.06)/0.08, \textit{ for } 0.06 < x \le 0.14 \\ 1, \textit{ for } 0.14 < x \le 0.18 \\ (0.42 - x)/0.24 \textit{ for } 0.18 < x \le 0.42 \\ 0 \textit{ in other cases} \end{cases} \quad (1)$$

$$\mu_{T_3} NH_4^+ = \begin{cases} (x-0.18)/0.24, \textit{ for } 0.18 < x \le 0.42 \\ 1, \textit{ for } 0.42 < x \le 0.6 \\ (1.4-x)/0.8 \textit{ for } 0.6 < x \le 1.4 \\ 0 \textit{ in other cases} \end{cases}$$

$$\mu_{T_4} NH_4^+ = \begin{cases} (x-0.06)/0.8, \textit{ for } 0.6 < x \le 1.4 \\ 1, \textit{ for } 1.4 < x \le 3.5 \\ 0 \textit{ in other cases} \end{cases}$$

The result of the measurement/classification procedure is determined by the cross section of the measurement result with the term set (1). Thus, if the content is 12 mg/dm^3, the fuzzy result of the measurement/classification procedure is determined by the cross section with the terms T_1 and T_2, and is respectively T_1 | 0.25; T_2 | 0.75.

When determining water quality classes and categories for individual indicators, multiple measurements are performed, arithmetic mean values for each indicator are separately compared with the classification scale, the worst values (maximum or minimum) are also compared with the scale, and then the final decision on a certain water quality indicator is made. When using scales with fuzzy linguistic variable, a more universal approach is proposed, which provides information on the average values of quality indicators. To do this, individual measurement results are classified on a scale with fuzzy linguistic variable, and then combined using the fuzzy averaging operator [5]. Arithmetic mean (AM) for each category T_i of measurement results is:

$$\mu_{AM}(T_i) = \frac{1}{n}\sum_{i=1}^{n} \mu(T_i) \qquad \forall T_i \in M, \tag{2}$$

where n is the number of measurements, $\mu(T_i)$ is the degree of belonging of a single measurement result to the category T_i, $\mu_{AM}(T_1)$ is the degree of belonging of the result of multiple measurements to the category T_i, M is the set of categories of the scale with fuzzy linguistic variable.

For example, we can use the results of NH_4^+ measurements obtained for the Southern Bug River (Table 1).

According to formula (2) for the data of Table 1 we obtain:

$\mu_{AM}(T_2) = 0.44$, $\mu_{AM}(T_3) = 0.56$.

While the outcome of the measurement/classification procedure is as follows: T_2 | 0.44; T_3 | 0.56.

3. The methods of data processing according to the measurement/classification of group quality indicators

When constructing group indicators of water quality, indicator scales of the various components in water content is used, which are related to the procedure of their measurement. Thus, the measurement of biochemical oxygen demand (*BOD*) is performed according to the method [6]. According to this method, the linguistic characteristics of the term sets of the scale "*BOD* in surface waters mg/dm^3" are as follows: T_1 – "Class I – very clean" (less than 1.0), T_2 – "Class II – clean" (1.0–2.1), T_3 – "Class III – contaminated" (2.2–7.0), T_4 – "Class IV – dirty" (7.1–12.0). According to the relative measurement error of *BOD* set as the confidence interval with a probability P = 0.95 of 30% (1–10 mg/dm^3) and 25% (10–100 mg/dm^3), a scale for fuzzy linguistic variable was constructed:

$$\mu_{T_1} BOD = \begin{cases} 1, \textit{ for } 0 \le x \le 0.7 \\ (1.3-x)/0.6 \textit{ for } 0.7 < x \le 1.3 \\ 0 \textit{ in other cases} \end{cases}$$

$$\mu_{T_2} BOD = \begin{cases} (x-0.7)/0.6, \textit{ for } 0.7 < x \le 1.3 \\ 1, \textit{ for } 1.3 < x \le 1.47 \\ (2.73-x)/1.26 \textit{ for } 1.47 < x \le 2.73 \\ 0 \textit{ in other cases} \end{cases} \tag{3}$$

$$\mu_{T_3} BOD = \begin{cases} (x-1.47)/1.26, \textit{ for } 1.47 < x \le 2.73 \\ 1, \textit{ for } 2.73 < x \le 4.9 \\ (8.9-x)/4.0 \textit{ for } 4.9 < x \le 8.9 \\ 0 \textit{ in other cases} \end{cases}$$

$$\mu_{T_4} BOD = \begin{cases} (x-4.9)/4.0, \textit{ for } 4.9 < x \le 8.9 \\ 1, \textit{ for } 8.9 < x \le 15.0 \\ 0 \textit{ in other cases} \end{cases}$$

Table 1

The results of measurements and classification of NH_4^+ content for the Southern Bug River

NH_4^+ $\frac{mg}{dm^3}$	Measurement result	0.27	0.40	0.33	0.38	0.20	0.31
	Classification result	T_2 | 0.63; T_3 | 0.37	T_2 | 0.08; T_3 | 0.92	T_2 | 0.38; T_3 | 0.62	T_2 | 0.17; T_3 | 0.83	T_2 | 0.92; T_3 | 0.08	T_2 | 0.46; T_3 | 0.54

Table 2

Results of measurements and classification of *BOD*

BOD $\frac{mg}{dm^3}$	Measurement result	3.4	1.52	3.92	3.2	3.5	2.8
	Classification result	T_3 | 1	T_2 | 0.96; T_3 | 0.04	T_3 | 1	T_3 | 1	T_3 | 1	T_3 | 1

The results of the measurement/classification procedure for the indicator of *BOD* are determined by the cross section of the measurement results of Table 2 with the term set (3).

Using fuzzy averaging operator, the outcome of the procedure of repeated measurement/classification of the *BOD* indicator is obtained: T_2(*clean*)|0.16; T_3 (*contaminated*)|0.84.

That is, when submitting a measurement classified by individual results, information about the availability of the result corresponding to the lower category is retained.

When constructing group indicators of water quality, it is necessary to combine individual indicators. According to the theory of fuzzy decision-making, associations are performed on the basis of unified data using a common scale of decision-making [7, 8]. For this case, these are the conditions of verbal and metric unification. When constructing a unified scale, the boundaries of individual terms are divided into the normalizing value – the end of the range. Then, the unified scales for the above two indicators are as follows: for NH_4^+ T_1 – "Class I – very clean" (less than 0.04), T_2 – "Class II – clean" (0.12), T_3– "Class III – contaminated" (0.4), T_4 – "IV class – dirty" (1); for *BOD* T_1 – "Class I – very clean" (less than 0.083), T_2 – "Class II – clean" (0.183), T_3 – "Class III – contaminated" (0.58), T_4 – "Class IV – dirty" (1). It is obvious that metric unification is not carried out.

Since only the conditions of verbal unification are met, the paper proposes the use of fuzzy union operators, based only on verbal unification with the presentation of the original data in a classified form.

The analysis of literature sources has demonstrated that, when combining two or more indicators by category T_i $(i=\overline{1,M})$, the following operators can be used:

– S-norm "algebraic sum", which for indicators A and B is equal to

$$\mu_{A+B}(T_i)=\mu_A(T_i)+\mu_B(T_i)-\mu_A(T_i)\cdot\mu_B(T_i);$$

– averaging operator $\mu_{AM}(T_i)=\frac{1}{n}\sum_{j=1}^{n}\mu_j(T_i),$

where n is the number of indicators that combine.

– Fuzzy – logical sum, which for indicators A and B is equal to

$$\mu_{F(A+B)}(T_i)=\mu_{A+B}(T_i)/(1-\mu_A(T_i)\cdot\mu_B(T_i)).$$

For the example shown in Table 1, 2 the results of the association are:

– S-norm "algebraic sum":

T_2(*clean*)|0.36; T_3 (*contaminated*)|0.64;

– averaging: T_2(*clean*)|0.3; T_3 (*contaminated*)|0.7;

– Fuzzy – logical sum:

T_2(*clean*)|0.25; T_3 (*contaminated*)|0.75.

According to the analysis of the results of modeling the combination of indicators using the above operators, the following conclusions and recommendations were obtained:

– the S-norm "algebraic sum" operator has the greatest variance and the greatest sensitivity to the upper or lower categories;

– the Fuzzy – logical sum operator has the lowest variance and the lowest sensitivity to the upper or lower categories;

– the results of aggregating by the averaging operator are between the S-norm "algebraic sum" and Fuzzy – logical sum;

– the S-norm "algebraic sum" operator can be recommended for the construction of group indicators if it is important to take into account the dispersal of the result of the classification by individual indicators;

– Fuzzy – logical sum operator can be recommended for the construction of group indicators if the purpose is to assign the result of the classification to certain categories of individual indicators;

– the averaging operator can be used for any number of indicators and retains the property of unit separation and can therefore be recommended for group indicators as the most universal.

4. Conclusions

Given the measurement uncertainty, the scale of the measurement/classification procedure is a scale with fuzzy linguistic variable, the categories of which correspond to the term set of fuzzy linguistic variable, and the membership functions of individual terms correspond to the characteristics of measurement uncertainty at the boundaries between the terms.

When processing the results of multiple measurements, individual measurement results are classified and then combined by fuzzy averaging. In this case, the combined function of belonging to the classification result retains data on the degree of belonging to the upper or lower categories.

The operators of fuzzy logic that can be used in the construction of group quality indicators are considered and recommendations for their application are given.

Методи опрацювання даних, отриманих за процедурою вимірювання/класифікації показників якості об'єктів

О.Ю. Года, Р.Ю. Довгалюк, Н.А. Яремчук

Національний технічний університет України "Київський політехнічний інститут імені Ігоря Сікорського", пр. Перемоги, 37, 03056, Київ, Україна
R_olia@i.ua; ruslanadovhaliuk@gmail.com; yaremchukna@i.ua

Анотація

У роботі розглянуто процедуру вимірювання/класифікації властивостей об'єктів, що є показниками якості, яка використовується для визначення категорії якості об'єкта. Відповідно до розглянутої процедури будується шкала з метричним носієм за діапазоном вимірювальної властивості, на якій нанесено границі, що відповідають категоріям якості об'єкта за цією властивістю. Відповідно до цих класів і категорій при застосуванні одиничних показників якості вирішується задача віднесення якості води до відповідної вербальної категорії за одним компонентом. При урахуванні невизначеності вимірювання шкала стає нечіткою і загальна терм-множина шкали відповідає категоріям класифікації. Показано, що результат класифікації включає найменування категорій з позначенням ступеня приналежності до цієї категорії. Через те, що внаслідок урахування невизначеності функції приналежності окремих категорій перерізаються, то характеристикою невизначеності процедури вимірювання/класифікації стає розмах класифікованих даних за сусідніми категоріями.

При побудові групових показників необхідне об'єднання одиничних показників, що підлягали багаторазовим вимірюванням. Згідно з теорією прийняття нечітких рішень об'єднання проводять на основі уніфікованих даних із використанням загальної шкали прийняття рішень. Для опрацювання багаторазових вимірювань за процедурою вимірювання/класифікації запропоновано метод із класифікацією кожного результату вимірювання і подальшим використанням оператора нечіткого усереднення. Як приклад застосування запропонованого методу використано визначення категорій якості води за її біохімічними властивостями. Проведено моделювання побудови групових показників якості за умов вербальної уніфікації з використанням наступних операторів нечіткої логіки: S-норма "алгебраїчна сума", оператор нечіткого усереднення, Fuzzy – логічна сума. Проведено порівняння та розроблені рекомендації з їх використання.

Ключові слова: процедура вимірювання/класифікації; невизначеність вимірювання нечіткої шкали; нечіткі показники якості; групові показники якості.

Методы обработки данных, полученных по процедуре измерения/классификации показателей качества объектов

О.Ю. Года, Р.Ю. Довгалюк, Н.А. Яремчук

Национальный технический университет Украины "Киевский политехнический институт имени Игоря Сикорского", пр. Победы, 37, 03056, Киев, Украина
R_olia@i.ua; ruslanadovhaliuk@gmail.com; yaremchukna@i.ua

Аннотация

В работе рассмотрена процедура измерения/классификации свойств объектов, являющихся показателями качества, используемая для определения категории качества объекта. Согласно процедуре строится шкала с метрическим носителем по диапазону измерительного свойства, на котором нанесены границы, соответствующие категориям качества объекта по этому свойству. При учете неопределенности измерения шкала становится нечеткой и общее терм-множество шкалы соответствует категориям классификации. Показано, что результат классификации включает в себя наименование категорий с указанием степени принадлежности к этой категории. Показано, что характеристикой неопределенности процедуры измерения/классификации становится размах классифицированных данных по соседним категориям.

Для обработки многократных измерений по процедуре измерения/классификации предложен метод классификации каждого результата измерения с последующим использованием операторов нечеткого усреднения. В качестве примера применения предлагаемого метода использовано определение категорий качества воды по ее биохимическим свойствам.

Ключевые слова: процедура измерения/классификации; неопределенность измерения нечеткой шкалы; нечеткие показатели качества; групповые показатели качества.

References

1. 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
2. Yaremchuk N.A., Semenyuk R.S. Sposoby urakhuvannia nevyznachenosti pry pobudovi linhvistychnykh shkal [Methods of taking into account uncertainty in construction of linguistic scales]. Spetsvypusk: Dopovidi KhII Mizhnarodnoi naukovo-tekhnichnoi konferentsii "Metrolohiia ta vymiriuvalna tekhnika" [The special issue: Reports of the *XII International Scientific and Technical Conference “Metrology and Measurement Techniques”*], 2020, no. 3A, pp. 77–81 (in Ukrainian). Available at: http://umj.metrology.kharkov.ua/article/view/217969
3. Semeniuk R., Yaremchuk N. The methods of obtaining and elaborating classified data in intellectual measurement systems. *Central European Researchers Journal*, 2020, vol. 6, issue 1, pp. 27–36.
4. MVV №081/12-0106-03. Metodyka vykonannia vymiriuvan masovoi kontsentratsii amonii-ioniv fotokolorymetrychnym metodom z reaktyvom Neslera [Method of measuring the mass concentration of ammonium ions by photocolorimetric method with Nesler’s reagent]. Kyiv, 2003 (in Ukrainian).
5. Zak Yu.I. Prinyatie reshenij v usloviyah nechetkih i razmytyh dannyh: Fuzzy – tekhnologii [Decision making in conditions of fuzzy and blurred data: Fuzzy technologies]. Moscow, Librokom, 2013. 352 p. (in Russian).
6. MVV №081/12-0014-01. Metodyka vykonannia vymiriuvan biokhimichnoho spozhyvannia kysniu *BOD* [Methods of measuring biochemical oxygen demand *BOD*]. Kyiv, 2002. 41 p. (in Ukrainian).
7. Piegat A. Fuzzy modeling and control [Nechetkoe modelirovanie i upravlenie. 2nd ed. Translation from English]. Moscow, Binom, 2013. 798 p. (in Russian).
8. Saaty T.L. Decision Making with Dependence and Feedback: The Analytic Network Process [Prinyatie reshenij pri zavisimostyah i obratnyh svyazyah. Analiticheskie seti. 3nd ed. Translation from English]. Moscow, Librokom, 2011. 360 p. (in Russian).