УДК 681.142

Н. А. Яремчук, О. Ю. Редьога, О. М. Сікоза

ОСОБЛИВОСТІ АРИФМЕТИЗАЦІЇ ДИСКРЕТНИХ ВЕРБАЛЬНИХ ШКАЛ

Вступ

Вербальні ординальні шкали якості характеризують зв'язок між станом об'єкта або його властивостями і словесним відображенням цього стану (властивості). В цьому випадку дискретна вербальна шкала строгого лінійного порядку може бути позначена як $S=(X;R_{>})$, де $X=\{x_0,\dots x_i\dots,x_{m+1}\}$ – сукупність точок шкали, а $R_{>}$ – відношення строгого лінійного порядку. Наприклад, вербальна шкала якості з наступними найменуваннями точок шкали: x_0 – «пристрій зовсім не відповідає вимогам до його якості», x_1 – «пристрій частково відповідає вимогам до його якості», x_2 – «пристрій відповідає основним вимогам до його якості», x_3 – «пристрій майже повністю відповідає вимогам до його якості», x_4 – «пристрій відповідає всім вимогам до його якості». Така шкала є вербальною або «не числовою» шкалою вимірювання якості. Але використання «не числових» шкал для вимірювання якості робить неможливою подальшу обробку результатів вимірювання за такими шкалами з використанням методів, які орієнтовані на числові вихідні дані. Крім того, збільшилось число випадків взаємодії з областями математичної кібернетики при вирішенні задач упорядкування об'єктів або їх властивостей за ступенем «надання переваги», «цінності», «корисності» тощо, які теж вимагають числових значень.

Тому ставиться задача арифметизації дискретної вербальної шкали вимірювання якості [1], тобто задача присвоєння визначених дійсних чисел точкам шкали $S=(X;R_{>})$ зі збереженням заданих відношень, що входять в структуру $R_{>}$. Іншими словами, знаходять відповідність φ між множиною вербальних точок X шкали S і множиною дійсних чисел, що задає арифметизацію шкали вимірювання якості ПЗ і є гомоморфізмом емпіричної системи в числову.

В залежності від наявності або відсутності апріорної інформації розглядають арифметизацію з врахуванням апріорної інформації і стохастичну арифметизацію відповідно.

В роботі [2] розглянуто принцип стохастичної арифметизації, за результатами якої отримують числа, що відповідають вербальним точкам шкали. Але ця відповідність характеризує певною невизначеністю, і чим ця невизначеність менша, тим більша точність встановлення арифметизованої (числової) шкали. Такий аналіз не проводився і тому задачею статті є дослідження способів оцінювання неточності арифметизації вербальних шкал, або невизначеності встановлення шкали, та подання рекомендацій зі зменшення даної невизначеності.

Постановка задачі: дослідження способів оцінювання неточності переходу від множини вербальних точок ординальної шкали до множини чисел при арифметизації шкали з розробкою рекомендацій щодо зменшення невизначеності арифметизованої шкали.

1. Аналіз способів подання арифметизованої шкали

Вербальна шкала встановлюється за кортежем $\langle Q, X, M, F \rangle$, де Q – сукупність проявів властивості; X – сукупність вербальних точок шкали, що відповідають проявам властивості за операцією $M : Q \to X$; операція F – один до одного перенесення відношень з емпіричної системи $\boldsymbol{Q}$ до вербальної системи $\boldsymbol{X}$.

Числова шкала встановлюється за кортежем $\langle Q, N, M, F \rangle$, де N – сукупність числових точок шкали, $N = (n_1, \ldots)$, $n_i \in R^+$. Таким чином при арифметизації визначається перехід $X \to N$.

При стохастичній арифметизації такий перехід характеризується невизначеністю, яку потрібно оцінювати, тому що вона характеризує невизначеність встановлення шкали.

При стохастичній арифметизації приймається рівномірний розподіл на множині траєкторій арифметизації. Виходячи з цього, в роботі [2] запропоновано наступну модель розподілу щільності ймовірності значень, що відповідають вербальним точкам шкали:

$$p(j;i) = \frac{m!}{(i-1)!(m-i)!} \prod_{k=1}^{i-1} \left(x + \frac{k}{n} \right) \prod_{l=1}^{m-i} \left((1-x) + \frac{l}{n} \right) \frac{n^{m-1} \cdot n!}{(m+n)!}, \tag{1}$$

де $x = 0, \frac{1}{n}, \ldots, \frac{n-1}{n}$, 1 може приймати n дискретних значень,

де n – максимальне значення шкали, за допомогою якої проводиться арифметизація,

i – порядковий номер точки вербальної шкали, $i = \left\{ \overline{0, m+1} \right\}$,

j – номер точки на арифметизованій шкалі, $j = \left\{ \overline{0, m+1} \right\}$, що відповідає номеру i вербальної шкали,

$m+2$ – кількість вербальних і числових точок.

Математичне очікування розподілу використовують для визначення числових точок шкали:

$$\varphi_0\left(i;m,n\right)=M\left(i;m,n\right)=i\cdot n\cdot\left(m+1\right)^{-1}. \qquad (2)$$

Середнє квадратичне відхилення (СКВ) $\sigma(i;m,n)$, що характеризує розсіювання значень при арифметизації, дорівнює:

$$\sigma\left(i;m,n\right)=\sqrt{D\left(i;m,n\right)}=\sqrt{n^2\cdot\frac{i\cdot\left(m-i+1\right)}{\left(m+1\right)^2\cdot\left(m+2\right)}+n\cdot\frac{i\cdot\left(m-i+1\right)}{\left(m+1\right)\cdot\left(m+2\right)}}\ . \qquad (3)$$

Аналіз наведеного виразу для СКВ показує, що значення n використовується як множник діапазону, тому подальші розрахунки проводились для нормалізованого діапазону від 0 до 1. Друга складова СКВ зменшується зі збільшенням n, тому при розрахунках виходили з умови, що $n>>m$, і СКВ визначається як:

$$\sigma\left(i;m,n\right)=n\cdot\frac{i}{m+1}\cdot\sqrt{\frac{m-i+1}{i\cdot\left(m+2\right)}}\ ,\ i=\overline{\left\{1,m+1\right\}}. \qquad (4)$$

При великій кількості n модель (1) замінюють неперервною моделлю β – розподілу:

$$f\left[x;i;\left(m-i+1\right)\right]=x^{i-1}\cdot\left(1-x\right)^{m-i}\cdot\frac{\Gamma\left(m+1\right)}{\Gamma(i)\cdot\Gamma\left(m-i+1\right)}. \qquad (5)$$

Тоді маємо математичне сподівання: $\varphi_0(i;m)=i\cdot(m+1)^{-1}$.

Середнє квадратичне відхилення: $\sigma(i;m)=\dfrac{i}{m+1}\cdot\sqrt{\dfrac{m-i+1}{i\cdot(m+2)}}$

Характеристикою переходу $X\rightarrow N$ є матриця відповідності між вербальними точками i та числами за номерами j. Якщо похибка встановлення ординальної шкали відсутня, то по діагоналі матриці розміщуються одиниці (6):

$P'_{ji}=$

$j \backslash x_i$	x_0	x_1	…	x_{m+1}
0	**1**	0	…	0
1	0	**1**	…	0
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$m+1$	0	0	…	**1**

При цьому

$$0\le i\le m+1, \qquad 0\le j\le m+1. \qquad (6)$$

В реальній матриці (для шкали з похибками встановлення) по рядкам розташовані умовні ймовірності того, то числовій точці під номером j відповідає ряд вербальних точок з різними номерами.

Стохастична матриця, отримана при арифметизації, відрізняється від ідеальної наявністю розсіювання вздовж діагоналі матриці. Для розрахунків умовних ймовірностей в матриці авторами роботи використано поняття класу еквівалентності шкали. За РГМ 2008 [3] клас еквівалентності – підмножина проявів вимірюваної властивості, що приймається як умовно-нерозрізнена в шкалі вимірювань цієї властивості. Для числової шкали клас еквівалентності це частина шкали, що лежить в межах ± 0.5 поділки шкали від значення певної точки шкали. Ймовірність віднесення до певного класу еквівалентності відповідає площі під кривою щільності β– розподілу. Матриця відповідності отримана для $m=2$ з ймовірностями віднесення до певного класу еквівалентності або з ймовірностями відповідності i до j наведена нижче:

$P'_{ji} =$

j \ x_i	x_0	x_1	x_2	x_3
0	**0,67**	0,3	0,03	0
1	0,22	**0,45**	0,22	0,11
2	0,11	0,22	**0,45**	0,22
3	0	0,03	0,3	**0,67**

Моди умовних розподілень за рядками знаходяться на діагоналі матриці. Якщо керуватись умовним одномірним розподілом ймовірностей (кожен рядок матриці), то точка відповідності знаходиться за модою розподілу (тобто точка $j=1$ відповідає x_1, точка $j=2$ відповідає x_2, тощо). Таким чином перевищення ймовірності моди над ймовірностями сусідніх точок свідчить про правильну відповідність вербальної і числової точок.

Значення ймовірностей навколо діагоналі свідчать про розсіювання матриці. Критерії для оцінки розсіювання наведено в наступному розділі.

2. Оцінювання розсіювання матриці відповідності арифметизованої шкали якості

Відмінність реальної матриці від ідеальної характеризує похибки реальної матриці.

Для оцінювання неідеальності матриці P'_{ji} у порівнянні з матрицею P_{ji} (1.6) можна використати норму Фробеніуса [4]:

$$G = \frac{\sum_{i=1}^{m+2}\sum_{j=1}^{m+2}\left(P'_{ji} - P_{ji}\right)^2}{2(m+2)} = 0.15\,,\ \left(0 \le G \le 1\right).$$

Таке значення норми Фробеніуса свідчить про велике розсіювання ймовірностей вздовж діагоналі матриці відповідності.

Дослідження показали, що зі збільшенням точок вербальної шкали збільшується і розсіювання матриці відповідності. Так, при кількості точок 7 матриця відповідності має вигляд:

$P'_{ji}(7) =$

x_i / j	x_0	x_1	x_2	x_3	x_4	x_5	x_6
0	**0,58**	0.35	0.06	0,01	0	0	0
1	0,18	**0.41**	0.31	0.1	0.02	0	0
2	0,17	0.17	**0.32**	0.24	0.09	0.01	0
3	0.03	0.1	0.21	**0.31**	0.21	0.11	0.03
4	0	0.01	0.09	0.24	**0.32**	0.17	0.17
5	0	0	0.02	0.1	0.31	**0.41**	0.17
6	0	0	0	0,01	0.06	0.35	**0.58**

З матриці $P'_{ji}(7)$ видно, що розсіювання умовних ймовірностей за точками шкали збільшилось і для центральних точок шкали умовні ймовірності не дорівнюють нулю у всьому рядку. Для порівняння при аналізі було визначено такі критерії якості матриці відповідності: норма Фробеніуса (характеризує розсіювання), мінімальне перевищення моди (характеризує якість класифікації окремих класів еквівалентності), розмах умовного одномірного розподілу за точками шкали (характеристика невизначеності встановлення арифметизованої шкали). Слід відмітити, що при наявності в розподілі точок з малою ймовірністю характеристика розмаху може бути завищена. Тому як характеристику невизначеності обрано розмах за загальною ймовірністю P=0,95, тобто з вилученням точок з малою ймовірністю. В табл. 1 наведено результати обчислення критеріїв розсіювання і розмитості арифметизованої шкали.

За порівнянням даних, наведених в табл. 1, можна зробити наступні висновки: зі збільшенням кількості точок вербальної шкали збільшується розсіювання встановлення арифметизованої шкали. Норма Фробеніуса значно збільшується (від 0,15 при $m+2=4$ до 0,24 при $m+2=7$). Перевищення моди стає незначним, що ускладнює правильне визначення точок шкали. Тому без наявності апріорної інформації рекомендується встановлювати шкали з меншою кількістю точок шкали. А потім після вимірювань за цією шкалою або калібрування за артефактами поліпшувати характеристики шляхом використання апостеріорної інформації і тоді збільшувати кількість точок шкали.

***Таблиця* 1.**

Значення критеріїв розсіювання і розмитості арифметизованої вербальної шкали якості

Кількість точок вербальної шкали $(m+2)$	Норма Фробеніуса, G	Мінімальне перевищення ймовірності моди над сусідніми значеннями	Максимальний розмах за точками шкали $P=0.95$
4	0,1528	0.23	3
5	0,1867	0.2	5
6	0,2183	0.11	5
7	0,2367	0.098	7

Висновки

В статті розглянуто задачу арифметизації дискретних ординальних шкал якості, вирішення якої полягає в тому, що вербальним точкам шкали ставлять у відповідність числа. Задача арифметизації може вирішуватись за наявністю або відсутністю апріорної інформації (остання називається стохастичною арифметизацією). Показано, що стохастична арифметизація супроводжується певною невизначеністю. Для зменшення невизначеності рекомендовано збільшення кількості дискретних значень числової шкали.

В статті визначено декілька критеріїв для характеристики якості арифметизації вербальної шкали: норма Фробеніуса (характеризує розсіювання), мінімальне перевищення моди умовного розподілення рядка матриці відповідності «вербальні точки» – «числа» (характеризує якість класифікації окремих класів еквівалентності), розмах умовного одномірного розподілу за точками шкали (характеристика невизначеності встановлення арифметизованої шкали).

Наведено рекомендації з вибору кількості точок шкали з врахуванням норми Фробеніуса та значення мінімального перевищення моди. В результаті аналізу встановлено, що за умови стохастичної арифметизації при збільшенні кількості точок вербальної шкали зменшується перевищення моди між сусідніми класами еквівалентності та збільшується норма Фробеніуса, що вказує на збільшення невизначеності. Тому при стохастичній арифметизації число дискретних градацій повинно бути обмежене і процедуру арифметизації необхідно проводити в декілька етапів. Спочатку проводиться стохастична арифметизація і встановлюється шкала з мінімальною кількістю точок, наприклад, «Низький», «Середній», «Високий», потім за встановленою шкалою збирається апостеріорна інформація за допомогою якої ординальна шкала уточнюється.

Список використаної літератури

1. *Хованов Н. В.* Математические основы теории шкал измерения качества [Текст]// Н. В. Хованов/– Ленинград.: 1982. – 169 с.
2. *Яремчук Н.А.* Арифметизація ординальних шкал вимірювання якості програмних засобів[Текст]// Н. А. Яремчук, О. Ю. Редьога / Інформаційні системи, механіка та керування. – 2011. – №11. – С. 5-–15.
3. РМГ 83–2008 Шкалы измерений. Термины и определения. Москва.– 2008.– 20 с.
4. Emil Bashkansky, Tamar Gadrich Some metrological aspects of original measurements/ Accred. – Qual. – Assur. – 2010. – p. 331–336.