

УДК 519.25:303.2

П. М. Григоруk, к.т.н., доцент, Т. П. Завгородня, д.е.н., професор.

Хмельницький національний університет

ВИКОРИСТАННЯ АПАРАТУ НЕЧІТКИХ МНОЖИН ДЛЯ ШКАЛЮВАННЯ ДАНИХ

Анотація. Розглянуто напрямки використання шкали Харрінгтона для визначення відповідності між значеннями деякого вимірюваного показника і суб'єктивній оцінці бажаності. Наведено правила переведення діапазону значень показника до діапазону значень кодової лінгвістичної шкали.

Ключові слова. М'які вимірювання, нечітка множина, шкала бажаності, лінгвістична шкала.

При зборі первинних даних дослідник має справу з великою кількістю атрибутів, що характеризують об'єкт дослідження. Опис вигляду даних і, за необхідності, механізму їх породження – початок будь-якого статистичного дослідження.

Ознака – це деяка загальна для всіх об'єктів якість, конкретні прояви якої взагалі кажучи, можуть змінюватися від одного об'єкту до іншого [1]. Вона не існує сама по собі, і є результатом абстрагування дійсності, її ідеальною конструкцією. В процесі абстрагування можна виділити три рівні:

- змістовний, на якому здійснюється відбір сукупності тих об'єктів, які відображають досліджувану реальність;
- концептуальний, спрямований на виявлення теоретичних понять в якості змінних і переведення їх на мову досліджуваних ознак;
- формалізаційний, на якому здійснюється операціоналізація понять, в ході якої здійснюється зв'язування теоретичних понять з емпіричними даними, тобто виділення індикаторів, що відображають властивості понять, і вибір допустимого набору значень кожного індикатора.

Результатом такого абстрагування є ознаковий простір, який моделює досліджувану реальність. В ньому зосереджені лише найбільш істотні з погляду

дослідження властивості явища або процесу, що вивчається. Формування простору має апріорний характер і в значній мірі визначається суб'єктивним баченням дійсності дослідником. Найбільший вплив воно має на етап формалізації, в результаті чого одні характеристики явища визначаються найбільш істотними, а інші – не значущими. Різні підходи до формалізації можуть призводити до різних просторів.

Наповнення ознакового простору даними здійснюється шляхом вимірювання, під яким розуміють приписування чисел властивостям або подіям відповідно до певних правил, так що відношення між числами репрезентують відношення між цими об'єктами [2]. Як зазначає І. Пфанцагль, об'єктами вимірювання є властивості, тобто, вимірюється не сам по собі об'єкт, а лише його окремі характеристики, параметри. При цьому, вимірюючи одну властивість, ми ігноруємо інші. З цієї точки зору різні за природою об'єкти стають подібними з позиції однакових значень властивості [3, с.11]. Операція встановлення відповідності між вимірними емпіричними даними і певними символічними характеристиками деякої штучно створеної системи називається шкалюванням. Питання, пов'язані з формуванням теорії вимірювань та вибору відповідних шкал в досліджуванні соціально-економічних систем розглядалися С. Стівенсом, П. Суппесом, Дж. Зінесом, Д. Скотом, Р. Льюїсом, І. Пфанцаглем, С. А. Айвазяном, Ф. М. Бородкіним, В. В. Вітлінським, А. Г. Гранбергом, Л. Заде, Т. С. Клебановою, О. О. Недосекіним, Ю. Н. Тостовою, Н. В. Ховановим та багатьма іншими.

Метою даної статті є дослідження питань, пов'язаних із шкалюванням даних в умовах невизначеності.

Невизначеність або навіть суперечність прояву ознак є характерною особливістю ринкового середовища, неусувною його якістю, пов'язаною з тим, що на ринкові умови здійснює свою одночасну дію велика кількість чинників різної природи і спрямованості, що не підлягають сукупній оцінці. Але і навіть якщо б всі ринкові чинники були в моделі враховані (що на практиці є

неймовірним), збереглася б неусувна невизначеність щодо характеру реакцій ринку на ті або інші дії [4].

Однією з форм прояву невизначеності при вимірюваннях є нечіткість, яка характеризує їх неточність, наближеність дійсному значенню. Обробка таких значень здійснюється в рамках теорії м'яких вимірювань, заснованих на використанні апарату нечітких множин [5].

Фундаментальним поняттям м'яких вимірювань універсальна шкала, яка розширює поняття традиційних шкал репрезентативної теорії вимірювань. Основна ідея, що дозволяє сформулювати нечіткі вимірювання, полягає в сполученні деякої лінгвістичної шкали з основною шкалою за рахунок використання загального метричного носія.

Неточності вимірювань відповідає носій нечіткої множини – універсальна множина X , до якої відносяться всі результати спостережень ознаки в рамках оцінюваної статистики, а невизначеності – відома функція належності. Нечітка множина A є множиною значень носія, причому кожному значенню носія зіставлений ступінь належності цього значення даній множині.

Отже, функція належності $\mu_A(x)$ – це функція, областю визначення якої є носій X , $x \in X$, а областю значень – одиничний інтервал $[0; 1]$. Чим вище значення μ_A , тим вище оцінюється ступінь приналежності елементу носія x нечіткій множині A .

Схематично вигляд функції належності представлений на рис. 1. Відмітимо, що далеко не завжди вона має симетричний щодо деякого значення x вигляд.

Безперечною перевагою використання функції належності є те, що всі ознаки приводяться до єдиного узагальненого безрозмірного показника, який приймає значення з відрізка $[0;1]$. В результаті цього ознаки набувають нової корисної властивості порівнянності між собою.

Таким чином, нечітка множина характеризує модельоване нечітке поняття і представляється у вигляді функції належності, яка задає для кожного значення A ступінь упевненості у приналежності його до деякого класу зна-

чень, який вимірюється в деякій шкалі оцінок. В ролі нечіткого поняття може виступати або лінгвістичне поняття, або нечітке число.

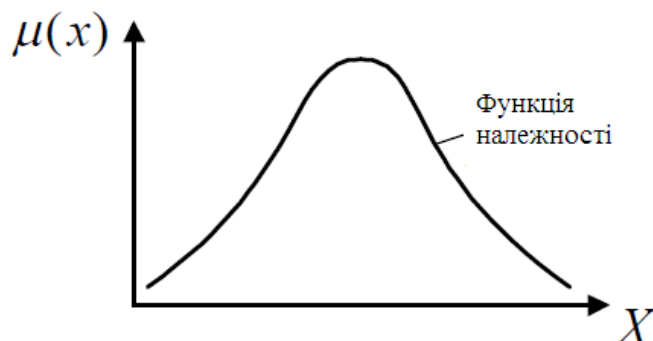


Рис. 1. Схематичний вигляд функції належності

Результатом м'яких вимірювань є вибрані на заданій шкалі лінгвістичні значення оцінюваних в процесі спостереження змінних.

Реалізація м'яких вимірювань полягає в проектуванні всіх можливих вимірювань на універсальну множину X , представлену у вигляді деякої шкали. Ця шкала, з іншого боку, визначається на відомих елементах терм-множини лінгвістичних змінних і поставлена у відповідність конкретному вимірюванню.

Основними питаннями шкалювання вхідної інформації є по-перше, вибір прийнятної шкали X , по-друге, вибір функції належності.

При виборі шкали необхідно враховувати ряд вимог, серед найважливіших з яких варто відзначити наступні:

- шкала повинна бути числовою і безрозмірною для забезпечення зіставлення ознак різної природи;
- шкала повинна бути універсальною, тобто, застосовною як до параметричної, так і до непараметричної вхідної інформації;
- шкала повинна мати область визначень для всіх значень всіх ознак.

Оцінкам інтенсивності прояву ознак, які зазвичай використовуються людиною, властиві наступні основні риси [4]:

⇒ якісність оцінок – словесних формулювань для вимірювання властивостей, для яких не розроблені кількісні шкали;

⇒ приблизність оцінок – їх розпливчатість навіть у тому випадку, коли властивість може бути зміряна за кількісною шкалою.

⇒ використання протилежних, біполярних оцінок, що виражають крайні рівні шкали;

⇒ наявність нейтральної оцінки;

⇒ використання модифікаторів оцінок типу «дуже», «надзвичайно», «майже»;

⇒ симетричність градацій протилежних оцінок;

⇒ використання від п'яти до семи градацій при оцінці властивостей.

Для адекватного вибору функції належності необхідно сформулювати вимоги, що пред'являються до таких функцій:

- функція повинна бути безперервною, гладкою і монотонною;
- вигляд функції повинен відповідати реальним вирішенням дослідника за оцінкою стану ознак;
- чутливість функції в областях, близьких до 0 або 1 повинна бути істотно нижче, ніж в середній зоні. Під чутливістю функції належності розуміють відношення викликаного приросту за шкалою переваг до приросту за лінгвістичною шкалою (шкалою поточних значень ознаки).

Вказаним вимогам відповідає функція належності Харрінгтона [6]. Сформована на її основі шкала Харрінгтона відноситься до вербально-числових шкал, які дозволяють виразити ступінь інтенсивності критерійної властивості, що має суб'єктивний характер. До складу таких шкал входять, як правило, змістовний опис градацій шкали і числові значення, що відповідають кожній з градацій шкали. Шкала значень функції Харрінгтона називається ще шкалою бажаності (переваг), оскільки вона дозволяє визначати переваги рівня вимірюваної ознаки на основі його суб'єктивного сприйняття. Щоб отримати шкалу бажаності, зручно користуватися готовою розробленою таблицею відповідності між відносинами переваги в емпіричній і числовій системах

(табл. 1), причому значення $d \approx 0$ відповідає абсолютно неприйнятній величині ознаки, $d \approx 1$ – ідеальній його величині.

Таблиця 1

Зв'язок між кількісними значеннями безрозмірної шкали бажаності і лінгвістичним виразом бажаності (сприйняттям дослідника)

Бажаність	Відмітки на шкалі бажаності
Дуже добре	0,80..1,00
Добре	0,63..0,80
Задовільно	0,37..0,63
Погано	0,20..0,37
Дуже погано	0,00..0,20

Числові значення градацій шкали Харрінгтона отримані експертним шляхом на основі аналізу і обробки великого масиву статистичних даних. Граничні значення шкали переваг можуть бути віднесені до будь-якого рівня бажаності, на розсуд дослідника. Проте на практиці така ситуація з'являється у край рідко.

У загальному вигляді функцію Харрінгтона можна записати у вигляді

$$d = \exp(-\exp(-z)) \quad (1)$$

де d - значення шкали переваг;

z – значення кодової лінгвістичної шкали Z , розташовані на деякому відрізку $[-Z^* ; Z^*]$. Звичайне значення Z^* обирається в межах від 3 до 6.

Для випадку лінгвістичної шкали $Z = [-6 ; 6]$ функція Харрінгтона графічно показана на рис.2. Як видно з рисунка, цього діапазону вистачає, щоб значення функції в крайніх його точках достатньо близько наближалися до нуля і одиниці відповідно.

Характерними точками кривої є точки перегину $d(0) = \exp(-1) = 1/e \approx 0,37$; $d(0,77) = 1 - 1/e \approx 0,63$. Саме ці міркування, а також зручності обчислень вплинули на вибір цих точок як граничних при визначенні рівня бажаності. Зна-

чення $d=0,2$ і $d=0,8$ та відповідні ним значення $z= - 0,48$ і $z=1,50$ вибрані, як вже наголошувалося, експертним шляхом.

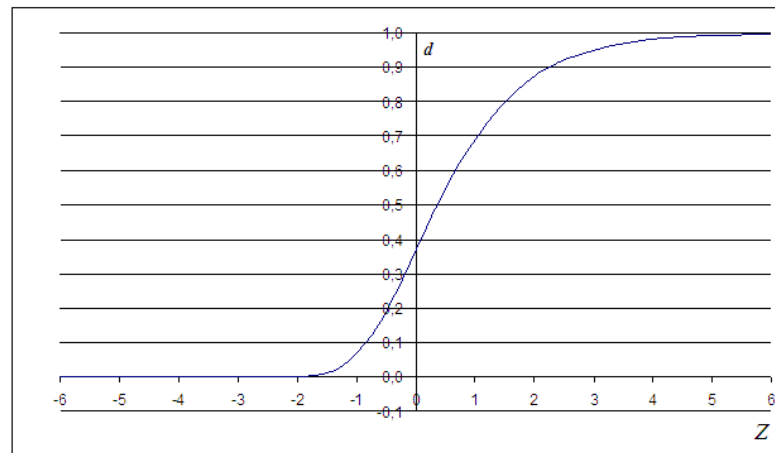


Рис. 2 – Графік функції Харрінгтона при $z \in [-6 ; 6]$

При використанні шкали переваг виникає питання переведення поточних значень вимірюваної ознаки в значення кодової шкали. Якщо діапазон значень множини X симетричний щодо нуля, а межа виражається скінченим числом, то переведення може здійснюватися за формулою:

$$z = \frac{x}{X^*} Z^*, \quad (2)$$

де x – поточне значення ознаки;

X^* – максимальне значення ознаки;

Z^* – обране максимальне значення кодової лінгвістичної шкали.

Його, як вже наголошувалося, рекомендується обирати в інтервалі від 3 до 6, на розсуд дослідника і виходячи із змістовного аналізу ознаки і необхідної точності розрахунків.

Якщо діапазон значень довільний і має вигляд $[X_1, X_2]$, то переведення здійснюється за формулою:

$$z = 2 \frac{x - X_c}{X_2 - X_1} Z^*, \quad (3)$$

де x – поточне значення ознаки;

Z^* – обране максимальне значення кодової лінгвістичної шкали;

X_c – значення середини інтервалу $[X_1, X_2]$.

Якщо перетворення проводиться до несиметричного проміжку кодової шкали $[Z_1, Z_2]$, то можна скористатись формулою:

$$z = Z_1 + \frac{x - X_1}{X_2 - X_1} (Z_2 - Z_1). \quad (4)$$

Якщо задовільною є точність розрахунків до сотих можна обрати $Z_1 = -2$, $Z_2 = 5$. В інших випадках граничні значення діапазону кодової шкали можна обрати, виходячи з табульованих значень функції Харрінгтона на деякому, достатньо великому проміжку.

В тому випадку, якщо відома деяка апріорна інформація про рівень задовільної бажаності $[X_L, X_H]$, то переведення вихідних значень до кодової лінгвістичної шкали можна виконати з використанням формули:

$$z = \frac{x - X_L}{X_H - X_L} \cdot 0,77. \quad (5)$$

Значення $z = 0,77$ відповідає верхньому значенню задовільного рівня бажаності $d = 0,63$. Перетворивши формулу (4), можна отримати граничні значення точок початкової шкали для інших рівнів бажаності.

В деяких випадках необхідно, щоб в кодованій шкалі діапазон задовільної бажаності був не $[0; 0,77]$, а дещо інший, що визначається деяким відрізком $[Z_1, Z_2]$. Цього можна добитися шляхом стиснення (розтягування) графіка

функції бажаності і зсуву його уздовж осі Z . При цьому функція Харрінгтона матиме вигляд:

$$d = \exp\left(-\exp\left(-\frac{0,77}{(Z_2 - Z_1)} \cdot (z - Z_1)\right)\right). \quad (6)$$

При цьому слід відмітити, що вигляд графіка самої функції також змінюється, і може спричинити зміну діапазону значень нормованої шкали. Сказане проілюстроване на рис. 3, на якому представлений графік функції Харрінгтона на відрізку $[-6 ; 6]$ з діапазоном значень задовільного рівня бажаності $[1 ; 3]$.

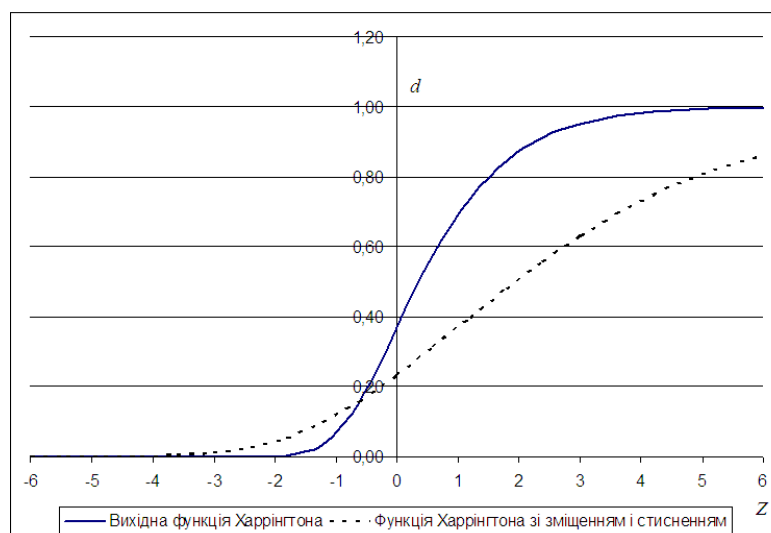


Рис. 3. Графік функції Харрінгтона зі зміщенням і стисненням

Слід зазначити, що всі наведені перетворення можна проводити комплексно, однак в кожному конкретному випадку потрібно надавати їх чітке обґрунтування.

Таким чином, нами розглянуто способи визначення відповідності між значеннями деякої ознаки і суб'єктивною оцінкою бажаності. Отримані результати можна використати для визначення структури досліджуваних об'єктів стосовно побудованого інтегрального показника, для якого зазвичай початково відсутні межі, які дозволяють однозначно охарактеризувати ступінь їх бажаної якості.

Список використаних джерел

1. Толстова Ю. Н. Анализ социологических данных: методология, дескриптивная статистика, изучение связей между номинальными признаками / Ю. Н. Толстова. – М. : Научный мир, 2000. – 352с. – ISBN 5-89176-086-X10.
2. Стивенс С. Математика, измерение и психофизика / С. Стивенс // Экспериментальная психология / Под ред. С. С. Стивенса ; пер. с англ под ред. действ. чл. АМН СССР П. К. Анохина, докт. пед. наук В. А. Артемова. – М. : Иностранная литература, 1960. – т. 1. – С. 11-92.
3. Пфанцагль И. Теория измерений / И. Пфанцагль ; пер. с англ. В. Б. Кузьмина . – М. : Мир, 1976. – 248 с.
4. Недосекин А. О. Нечетко-множественный анализ риска фондовых инвестиций : [монография] / А. О. Недосекин. –СПб. : Сезам, 2002. – 181 с.
5. Заде Л. Понятие лингвистической переменной и ее применение к принятию приближенных решений / Л. Заде ; [пер. с англ. Н. И. Ринго] / [Под ред. Н. Н. Моисеева и С. А. Орловского]. – М. : Мир, 1976. – 176 с. – (Серия: Математика. Новое в зарубежной науке. Вып.3).
6. Адлер Ю. П. Планирование эксперимента при поиске оптимальных условий / Ю. П. Адлер, Е. В. Маркова, Ю. В. Грановский. – М : Наука, 1976. – 139 с.

Abstract. The article describes directions for using the Harrington scale to determine the correspondence between the measured values of some indicators and the subjective assessment of desirability. Transfer rules a range of values of indicator values to a range of encoded linguistic scale are given.

Keywords. Soft measurement, fuzzy set, the desirability scale, the linguistic scale