

ДЕЯКІ АСПЕКТИ НЕЧІТКОЇ КЛАСТЕРИЗАЦІЇ ОБ'ЄКТІВ МИТНОГО КОНТРОЛЮ В РАМКАХ АВТОМАТИЗОВАНОЇ СИСТЕМИ АНАЛІЗУ ТА УПРАВЛІННЯ РИЗИКАМИ

У статті розглядаються питання ідентифікації об'єктів митного контролю, автоматизації процесу класифікації товарів на рівні митних органів. Обґрунтовано можливість застосування алгоритму нечіткої кластеризації для створення методу ідентифікації.

The article deals with objects of customs control identification, classification goods process automation at the customs. It had been proven the possibility to use fuzzy clustering algorithm to create a method of identification.

Ключові слова: нечітка кластеризація, алгоритм fuzzy C-means, ідентифікація об'єктів митного контролю.

На даний час перед митними органами поставлене подвійне завдання: з одного боку, – максимальне скорочення часу здійснення митного оформлення, спрощення митних процедур, сприяння розвитку зовнішньоекономічної діяльності; з іншого боку, – необхідність здійснення дієвого контролю над товарами, що переміщуються через митний кордон України, підвищення рівня збору митних платежів, забезпечення дотримання законодавства суб'єктами зовнішньоекономічної діяльності.

З метою практичної реалізації положень стандартних правил 6.3 та 6.4 Міжнародної конвенції про спрощення та гармонізацію митних процедур (зі змінами) [1], та як один із шляхів посилення контролю за митним оформленням, було розроблено та впроваджено автоматизовану систему аналізу та управління ризиками під час митного контролю та митного оформлення товарів із застосуванням вантажної митної декларації (далі – АСАУР).

На даному етапі розвитку АСАУР існує необхідність створення алгоритму ідентифікації товарів, що переміщуються через митний кордон та заявляються до митного оформлення, для контролю за правильністю класифікації товарів на рівні 14 знаків товарного коду. Існує широкий спектр товарів, які не деталізовані на рівні десятизначного коду Української класифікації товарів зовнішньоекономічної діяльності (далі – УКТЗЕД) [2], але мають відмінні характеристики, що впливають на рівень митної вартості, а як наслідок і на рівень оподаткування, що обумовлює використання додаткового чотиризначного коду товару.

Дієвого автоматизованого механізму контролю за відповідністю встановленого додаткового коду характеристикам заявленим в графі 31 вантажної митної декларації (далі – ВМД) на даний час не існує.

Метою статті є дослідження алгоритму нечіткої кластеризації fuzzy C-means для аналізу об'єктів митного контролю з метою впровадження експертного модуля ідентифікації товарів при митному контролі та митному оформленні з використанням ВМД.

Для розв'язання вище зазначеної задачі розглянемо алгоритми кластеризації на основі міри близькості, що дозволяють віднести той чи інший об'єкт ідентифікації до певного класу (кластеру). Кластерний аналіз передбачає таке розбиття множини об'єктів на підмножини, при якому кожен об'єкт може потрапити лише в один кластер, тобто – чітке розбиття. Для розв'язання задачі ідентифікації об'єктів митного контролю таке обмеження не завжди актуальне. Тому існує необхідність використання нечіткого підходу в задачах кластеризації. Це дозволить провести розбиття об'єктів ідентифікації таким чином, щоб визначити ступень належності кожного об'єкта до кожної підмножини (кластера), та в разі невизначеності, в рамках АСАУР, ініціювати візуальне інформування посадової особи митних органів, що здійснює митний контроль та митне оформлення в необхідності проведення додаткових форм контролю.

Відомо, що кластеризація – це об'єднання об'єктів у групи (кластери) на основі схожості ознак для об'єктів однієї групи й відмінностей між групами. Алгоритми кластеризації не опираються на традиційні статистичні методи; вони можуть використовуватися в умовах майже повної відсутності інформації про закони розподілу даних. Кластеризацію проводять для об'єктів з кількісними (числовими), якісними або змішаними ознаками.

Провідними дослідниками, що зробили найбільший внесок у розв'язок задач кластеризації, по праву вважаються Фішер, Макнаотон, Кауфман, Роузеув, Бездек. На даний час завдання кластеризації має велику кількість розв'язків. Різносторонні випадки й підходи до її розв'язку розглянуті в [3].

Методи кластеризації умовно розділяють на ієрархічні й неієрархічні, серед яких найбільше поширення одержали методи розбиття. Спільними для методів розбиття є відстань як метод порівняння даних, множина центрів кластерів і матриця належності як основні результати роботи алгоритму, функціонал, мінімізація якого визначає розв'язок, а також набір обмежень. Серед методів розбиття найбільш відомі методи k-середніх, Fuzzy C-means і кластеризація по Гюстафсону-Кесселю. Перший відрізняється тим, що одержувані з його допомогою значення матриці приналежності є чіткими і приймають значення з множини {0,1}. Fuzzy C-means і кластеризація по Гюстафсону-Кесселю – нечіткі методи кластеризації, у них матриця приналежності складається з відповідних значень функцій приналежності нечітким кластерам. Відмінність цих методів проявляється у формі кластерів, якими вони оперують. Fuzzy C-means породжує

кластери сферичної форми, а кластеризація по Гюстафсону-Кесселю – еліпсоїди.

Недоліками методів, що не використовують математичний апарат нечіткої логіки, є:

- використання для аналізу центрів кластерів, а не всіх даних;
- віднесення об'єктів тільки до одного із кластерів, а не їх частковий розподіл по класах.

Розглянемо задачу кластеризації для товарної підкатегорії УКТЗЕД **8471609000** «Пристрої введення або пристрої виведення, що містять або не містять в одному корпусі запам'ятовувальні пристрої» [2].

Постановка задачі:

- $X = (X_1, X_2, \dots, X_n)^T$ – об'єкти митного контролю (далі – об'єкти), що підлягають кластеризації (n – кількість об'єктів). Кожний об'єкт $X_k = (x_{k_1}, x_{k_2}, \dots, x_{k_p})$ являє собою точку в p -мірному просторі ознак $k = \overline{1, n}$.

- c – кількість кластерів.

Для кожного елемента множини X необхідно визначити ступінь належності до класів.

Елементи одного кластера повинні бути так близькі кожний кожному, як це тільки можливо, і, одночасно, кластери повинні бути на найбільшому видаленні один від одного. Для забезпечення процесу кластеризації використовують міру близькості, у якості якої звичайно визначають відстань між двома об'єктами (точками в p -мірному просторі) X_k і X_l у вигляді функції:

$d : X \times X \rightarrow R$, яка має наступні властивості:

$d(X_k, X_l) = d_{kl} \geq 0$ – відстань між двома об'єктами має не від'ємне значення;

$d_{kl} = 0 \Leftrightarrow X_k = X_l$ – відстань між двома об'єктами дорівнює нулю, якщо об'єкти однакові;

$d_{kl} = d_{lk}$ – відстань між об'єктом X_k і X_l дорівнює відстані між X_l і X_k .

Функція належності $m_{s_i} : X \rightarrow [0,1]$ описує розбиття множини $X = (X_1, X_2, \dots, X_n)^T$ на нечіткі підмножини S_i ($i = \overline{1, c}$).

Позначимо ступінь приналежності об'єкта $X_k = (x_{k_1}, x_{k_2}, \dots, x_{k_p})$ до підмножини S_i як m_{ik} , тобто $m_{ik} \equiv m_{s_i}(X_k)$. V_{cn} – множина усіх дійсних матриць розміром $c \times n$. Тоді матрицею ступенів належності називається матриця $M = [m_{ik}] \in V_{cn}$, для якої мають виконуватись наступні умови:

1. $m_{ik} \in [0,1]$ $i = \overline{1, c}$ $k = \overline{1, n}$ (на відміну від чіткого, при нечіткому c – розбитті будь-який об'єкт одночасно належить до різних кластерів, але з різним ступенем);

2. $\sum_{i=1}^c m_{ik} = 1$, де $k = \overline{1, n}$ (сума ступенів належності до нечітких кластерів одного об'єкту має бути нормалізована до 1);

3. $\sum_{k=1}^n m_{ik} \in (0, n)$, де $i = \overline{1, c}$ (кількість кластерів має бути більше нуля та менше кількості об'єктів).

$V_i = (v_{i_1}, v_{i_2}, \dots, v_{i_p})$, де $i = \overline{1, c}$ – центри кластерів, тобто точки в p -мірному просторі, навколо яких сконцентровані відповідні об'єкти.

При використанні евклідової відстані завдання нечіткої кластеризації полягає в знаходженні такої матриці ступенів приналежності M й таких координат центрів кластерів $V = (V_1, V_2, \dots, V_c)$, які б забезпечили мінімум наступного критерію [4]:

$$\sum_{i=1}^c \sum_{k=1}^n (m_{ik})^m \cdot \|X_k - V_i\|^2 \rightarrow \min, \quad (1)$$

де $V_i = \frac{1}{\sum_{k=1}^n m_{ik}} \sum_{k=1}^n (m_{ik})^m \cdot X_k$ – центр i -го кластеру, $i = \overline{1, c}$,

m – експоненціальна вага ($m \geq 1$).

Зазначимо, що значення експонентної ваги встановлюється до початку кластеризації. Експонентна вага впливає на матрицю ступенів приналежності: із збільшенням значення m кінцева матриця c -розбиття стає більш "розпливчатою", і при $m \rightarrow \infty$ вона прийме вид $M = [1/c]$, що не є адекватним розв'язком,

оскільки однаковий ступінь належності до кластерів буде визначений для всіх об'єктів кластеризації. На сьогодні не існує теоретично обґрунтованого правила вибору значення. Звичайно встановлюють $m = 2$.

Дані представлені наступною таблицею:

Таблиця 1

Об'єкти ідентифікації та їх характеристики

X_k	x_{k_1} (вага товару, кг)	x_{k_2} (митна вартість товару, \$)	Скорочена назва об'єкту ідентифікації
1	4	132	Монітор ЖК широкоформатний
2	4.2	129	Монітор ЖК широкоформатний
3	5.41	130	Монітор ЖК широкоформатний
4	5.61	135	Монітор ЖК широкоформатний
5	1.6	25.5	Сканер А4
6	4.57	25.5	Сканер А4
7	3.08	25.5	Сканер А4
8	2.47	25.5	Сканер А4
9	2.72	25.5	Сканер А4
10	3.6	28	Сканер А4
11	3	30	Сканер А4

Використавши зазначений алгоритм [5], встановимо:

1. Центри нечітких кластерів (рис. 1): V_1 з координатами [4.8036; 131.4382];
 V_2 з координатами [3.0061 30.5018].
2. Ступені належності об'єктів до нечітких кластерів (табл. 2)

Таблиця 2

Ступені належності об'єктів ідентифікації до нечітких кластерів

X_k	Ступінь належності кластеру V_2	Ступінь належності кластеру V_1
1	0,0001	0,9999
2	0,0006	0,9994
3	0,0002	0,9998
4	0,0012	0,9988
5	0,9976	0,0024
6	0,9976	0,0022
7	0,9978	0,0027
8	0,9973	0,0022
9	0,9978	0,0006
10	0,9994	0,0542
11	0,9458	0,9458

Результати нечіткої кластеризації об'єктів ідентифікації у графічному вигляді наведені на рис. 1.

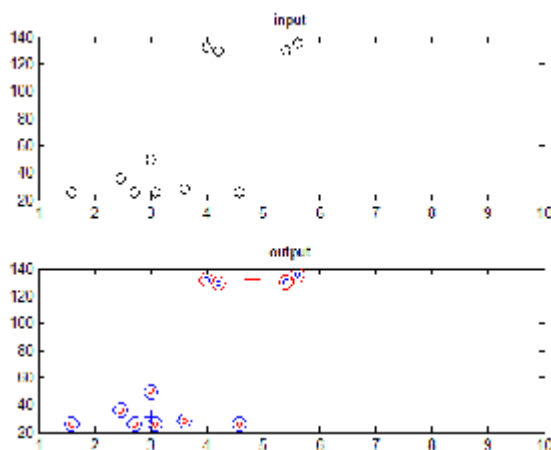


Рис. 1. Результати нечіткої кластеризації

Нечіткі правила можна синтезувати за результатами нечіткої кластеризації. Якщо об'єкти

кластеризації мають дві ознаки ($n=2$), то результати нечіткої розбиття можна представити тривимірним графіком: для кожного об'єкта відкласти по осях абсцис і ординат значення ознак, а по осі аплікату – ступінь приналежності об'єкта нечіткому кластеру. Кількість таких графіків буде дорівнює числу кластерів (n). Зазначений підхід дозволить екстрагувати нечіткі правила для ідентифікації об'єктів митного контролю.

Висновки. Таким чином проаналізовано можливість застосування методу кластеризації fuzzy c-means для створення алгоритму ідентифікації об'єктів митного контролю при визначенні додаткового коду товару. Отримані результати досліджень запропонованим методом дозволяють стверджувати про можливість його агрегування до модулю програмно-інформаційного комплексу «Інспектор 2006», що використовується митними органами в процесі митного оформлення та митного контролю, для розв'язання задач ідентифікації, та в рамках АСАУР, ініціювання візуального інформування посадової особи митних органів, в необхідності проведення додаткових форм контролю у випадках невизначеності.

Література

1. International Convention on the simplification and harmonization of customs procedures // Official Journal (EC) L 100, 21.04.1975, p. 0002 – 0017.
2. Про Митний тариф України (із змінами) [Електронний ресурс]: закон України: [прийнято ВР 05.04.2001 № 2371-III]. – Режим доступу: www.rada.gov.ua.
3. Методы и модели анализа данных: OLAP и Data Mining / [Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И.]. – СПб.: Петербург, 2004. – 336 с.
4. Zimmermann H. – J. Fuzzy Set Theory – and Its Applications. 3rd ed. – Kluwer Academic Publishers, 1996. – 435p.
5. Леоненков А.В. Нечеткое моделирование в среде MATLAB и fuzzyTECH / А.В. Леоненков. – СПб.: БХВ-Петербург, 2003. – 276 с.

Надійшла 12.5.2010 р.

УДК 621.396.6.019.3

О.М. ШИНКАРУК, І.І. ЧЕСАНОВСЬКИЙ
Хмельницький національний університет

ОКРЕМІ МЕТОДИЧНІ АСПЕКТИ ВИКОРИСТАННЯ НАДМІРНОСТЕЙ ПРИ ВИРІШЕННІ ПРОБЛЕМИ ЗАБЕЗПЕЧЕННЯ НАДІЙНОСТІ РАДІОТЕХНІЧНИХ СИСТЕМ

В статті представлено у формалізованому вигляді основні напрямки забезпечення високого рівня надійності багатофункціональних радіотехнічних систем із використанням наявних (потенційних) надмірностей. При цьому розглянута можливість адаптації профілактико-відновлювальних заходів багатофункціональних радіотехнічних систем до особливостей функціонування їх структурних підсистем, характерних тенденцій зміни технічного стану з урахуванням дії чинників зовнішнього впливу. Одержані вирази для окремих комплексних показників надійності з урахуванням наявних резервів як в структурі так і в системі експлуатації радіотехнічних систем. Представлені вирази для оцінки ефективності використання резервів на основі показників безвідмовності.

In the article basic directions of providing of high level of reliability of radio of the technical systems are presented in the formalized kind with the use of potential surpluses. Possibility of adaptation of prophylactic and repair works of radio of the technical systems is thus considered to the features of functioning of them structural subsystems, characteristic tendencies of change of the technical state taking into account the action of factors of external influence. Expressions are got for separate complex reliability indexes taking into account present backlogs as in a structure so in the system of exploitation of radio of the technical systems. Expressions are presented for the estimation of efficiency of drawing on reserves on the basis of indexes of faultlessness.

Ключові слова: адаптація, радіотехнічні системи, надмірність, температура, алгоритм, резервування.

Вступ. Аналіз сучасних тенденцій розвитку радіотехнічних систем (РТС) свідчить про стрімке підвищення їх функціональності, що призводить до значного ускладнення їх структури, а відповідно до зниження показників надійності. Незважаючи на поліпшення якості елементної бази, підвищення надійності таких систем може бути забезпечене тільки на основі комплексного підходу. Сутність даного підходу полягає в системному урахуванні складної структури РТС, сумісному використанні методів резервування (зокрема, почасового і структурного) і вдосконаленні методів організації і проведення профілактико-відновлювальних заходів (ПВЗ) з максимальним урахуванням чинників і умов реального функціонування систем [1,2].

Як на думку авторів, одним із шляхів реалізації такого підходу є синтез математичної моделі процесу функціонування, що забезпечує урахування структури та тенденцій зміни значень показників надійності окремих підсистем, а також визначення оптимальної періодичності їх ПВЗ. На підставі цього та з урахуванням прийнятої стратегії експлуатації (максимальна надійність при допустимих затратах чи мінімальні затрати при допустимому рівні надійності) передбачається суміщення операцій ПВЗ в певні