

УДК 338.26:351.82(477.83)

Р. Селіверстов

НЕЧІТКИЙ КЛАСТЕРНИЙ АНАЛІЗ ОКРЕМИХ ПОКАЗНИКІВ СОЦІАЛЬНОГО РОЗВИТКУ РАЙОНІВ ЛЬВІВСЬКОЇ ОБЛАСТІ

На прикладі аналізу окремих показників соціального розвитку наведено схему нечіткої кластеризації адміністративно-територіальних одиниць Львівської області. Отримано результати, що можуть бути використані під час побудови ефективних економіко-математичних моделей розвитку територій.

Ключові слова: нечітка логіка, кластеризація, класифікація, соціальний розвиток.

Управлінські рішення в галузі державного управління та місцевого самоврядування повинні бути науково обґрунтовані. За сучасних величезних обсягів інформації прийняття ефективного управлінського рішення у багатьох випадках неможливе без аналізу об'єктів управління на основі економіко-математичних моделей. Для забезпечення високої якості отриманих результатів доцільно для кожної групи однорідних об'єктів розробляти свою модель. Внаслідок цього постає проблема формування груп однотипних (однорідних) об'єктів дослідження.

Класичні статистичні методи визначення однорідності сукупності добре зарекомендували себе у випадку, коли аналіз проводиться за однією характеристикою. Коли ж об'єкти дослідження характеризуються низкою показників, доцільно здійснювати багатовимірне ранжування або кластеризацію. У результаті багатовимірного ранжування кожному об'єктові присвоюється інтегрований показник, який визначає його місце у рейтингу. Це може бути підґрунтям для подальшої класифікації (групування) об'єктів за принципом “кращі – середні – гірші”. Одна із методик багатовимірного ранжування з використанням нечіткої логіки апробована у дослідженні [1]. Результатом кластеризації є лише сформовані групи “схожих” об'єктів, кожна з яких (груп) якомога суттєвіше відрізняється від іншої (досягається за рахунок вирішення задачі оптимізації). З поверхневим оглядом алгоритмів та використанням нечітких методів кластеризації даних можна ознайомитися, наприклад, у працях деяких науковців [2 – 4].

Мета статті – здійснення нечіткої кластеризації адміністративно-територіальних одиниць (районів) Львівської області на основі аналізу окремих показників соціального розвитку за 2011 рік, за який на момент проведення дослідження вдалося отримати повні та вичерпні статистичні дані з відкритих інформаційних джерел. Подібні дослідження (без використання нечіткої логіки на основі порівняння індексу розвитку з еталонним значенням) розвитку малого підприємництва у адміністративно-територіальних одиницях України на національному рівні вже здійснені [5]. Особливістю нечіткої кластеризації у даному випадку є те, що кожна адміністративно-територіальна одиниця може бути віднесена до будь-якого з кластерів (характеристики яких наперед невідомі) з відповідним рівнем приналежності.

За вихідну інформацію для аналізу з інформаційного банку даних статистичних показників “Регіональна статистика Львівщини” (“Регіостат”) [6] для

двадцяти районів Львівської області було взято значення дев'яти показників рівня соціального розвитку (табл. 1):

1. Середньомісячна заробітна плата, грн.
2. Кількість зареєстрованих злочинів.
3. Забезпеченість житлом на душу населення, м².
4. Забезпеченість лікарями на 1000 населення.
5. Забезпеченість лікарняними ліжками на 1000 населення.
6. Забезпеченість дошкільними закладами, дітей на 100 місць.
7. Природний приріст (скорочення) населення.
8. Рівень офіційного безробіття, %.
9. Навантаження на одне робоче місце.

На основі цієї інформації формується матриця спостережень із нормованих (приведених до інтервалу від 0 до 1) значень показників (формула 1):

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{19} \\ x_{21} & x_{22} & \dots & x_{29} \\ \dots & \dots & \dots & \dots \\ x_{20,1} & x_{20,2} & \dots & x_{20,9} \end{pmatrix}. \quad (1)$$

Таблиця 1

Вихідна інформація для аналізу

Райони		Показники								
		1	2	3	4	5	6	7	8	9
Бродівський	1	2 046,35	419	21,9	29	6,7	162	-381	3,9	109
Буський	2	1 851,70	246	23,8	26	5,9	123	-200	1,3	52
Городоцький	3	1 836,79	299	24,8	37	6,9	137	-285	1	15
Дрогобицький	4	1 709,27	272	24,4	19	4,2	92	-235	1,8	69
Жидачівський	5	1 924,71	430	23,4	27	6,6	134	-698	2,1	80
Золочівський	6	1 923,37	350	21,4	32	7,3	130	-380	4,2	100
Кам'янка-Бузький	7	2 657,29	271	23,6	26	5,9	127	-200	0,8	27
Мостиський	8	1 765,73	266	23,9	26	5,1	115	-189	1,3	15
Жовківський	9	1 782,20	505	19,6	29	5,4	120	-139	1,2	47
Миколаївський	10	2 216,23	252	23,1	24	12,2	112	-83	3	46
Перемишлянський	11	2 112,64	165	23,6	29	7,2	121	-328	3,4	110
Пустомитівський	12	1 990,20	608	26,9	22	2,8	142	-154	0,9	14
Радехівський	13	1 806,33	276	23,2	26	5,4	113	-189	1,4	213
Самбірський	14	1 699,94	496	22,2	16	4,2	181	-141	2	429
Сколівський	15	2 193,61	275	20,4	29	4,6	128	-52	1,5	17
Сокальський	16	2 943,35	400	20,5	23	6,2	116	-363	1	15
Старосамбірський	17	1 758,69	252	21,6	22	6,7	166	-302	2	40
Стрийський	18	2 793,49	323	22,6	23	4,7	143	-107	1,1	44
Турківський	19	1 883,62	184	21,8	22	6,9	132	63	2,1	49

Яворівський	20	1 908,55	514	18	26	5,7	156	345	1,7	35
-------------	----	----------	-----	----	----	-----	-----	-----	-----	----

Завдання полягає у розподілі районів на наперед задану кількість груп (кластерів) так, щоб в межах однієї групи були максимально подібні за набором досліджуваних показників райони, а самі групи якомога суттєвіше відрізнялися одна від іншої. За міру подібності вибрано класичну для багатовимірного (у нашому випадку дев'ятивимірного) простору відстань за Евклідом (формула 2):

$$\|\vec{a} - \vec{b}\| = \sqrt{\sum_{k=1}^9 (a_k - b_k)^2}, \quad (2)$$

де a_k і b_k ($k = 1, 2, \dots, 9$) – координати векторів \vec{a} і \vec{b} відповідно.

Для розв'язування задачі нечіткої кластеризації обрано метод FCM (FuzzyC-Means) з експоненціальною вагою 2 [7]. Результат шукається у формі матриці приналежності $M = \{m_{ij}\}$ ($i = 1, 2, \dots, n$; $j = 1, 2, \dots, 20$), елементи якої виражають ступінь приналежності j -го району до i -го кластера, координати центру якого $c_i = (c_{i1}, c_{i2}, \dots, c_{i9})$ визначаються за формулою (n – кількість кластерів) (формула 3):

$$c_{ik} = \frac{\sum_{j=1}^{20} (m_{ij})^2 x_{jk}}{\sum_{j=1}^{20} (m_{ij})^2}. \quad (3)$$

Результуюча матриця приналежності повинна мінізувати суму відстаней від районів до центрів кластерів (формула 4):

$$\sum_{i=1}^n \sum_{j=1}^{20} (m_{ij})^2 \|c_i - x_j\| \rightarrow \min \quad (4)$$

за умов розподіленості кожного району між усіма кластерами та відсутності порожніх кластерів і кластерів, які містили б усі райони.

Задача оптимізації розв'язана ітераційним методом у середовищі TurboPascal. На кожному кроці елементи матриці приналежностей обчислювалися за формулою 5 (на першому кроці матриця формувалася випадковим методом):

$$m_{ij} = \frac{1}{\|c_i - x_j\|^2 \sum_{k=1}^n \frac{1}{\|c_k - x_j\|^2}}. \quad (5)$$

Процес обчислення припинявся, коли усі відповідні елементи матриць приналежностей, отриманих на двох сусідніх кроках ітерації, відрізнялися не більше, ніж на 0,1 (формула 6):

$$\max_{1 \leq i \leq n; 1 \leq j \leq 20} |m_{ij}^{(k)} - m_{ij}^{(k-1)}| < 0,1. \quad (6)$$

Результати обчислень подано у табл. 2. У комірці на перетині рядка (району) та стовпця (наперед заданій кількості кластерів) вказано номер кластера, у який потрапив відповідний район у результаті процесу кластеризації, згідно з найбільшим коефіцієнтом приналежності.

У випадку двох кластерів райони розподілились між ними так: 12 і 8 районів відповідно у першому та другому кластерах. Такий поділ районів на дві групи можна

було б вважати доволі умовним, але цікавим є факт, що збільшення кількості кластерів до трьох перенесло до новоутвореного кластера чотири райони з першої групи і лише один (Самбірський) з другої, що може свідчити про стійкість суттєвої подібності Кам'янка-Бузького, Жовківського, Пустомитівського, Сколівського, Сокальського, Стрийського та Яворівського районів. Із аналогічних міркувань стосовно поділу районів на більшу кількість кластерів можна зробити висновок про існування ще двох стійких груп – Бродівський, Жидачівський, Золочівський, Самбірський, Старосамбірський райони (ця група не змінюється для випадків поділу на 2 і 3 кластери) та Буський, Городоцький, Дрогобицький, Мостиський, Радохівський, Турківський (не змінюється для випадків поділу на 4 і 5 кластерів). Залишилися без розгляду два райони – Миколаївський та Перемишлянський, які у випадку поділу на 4 кластери утворили найменшу за обсягом групу, що може бути цікавим фактом для експертів-аналітиків.

Таблиця 2

Розподіл районів між кластерами

Райони	Кількість кластерів			
	2	3	4	5
Бродівський	1	2	1	1
Буський	1	1	3	3
Городоцький	1	1	3	3
Дрогобицький	1	1	3	3
Жидачівський	1	2	1	4
Золочівський	1	2	1	1
Кам'янка-Бузький	2	3	4	2
Мостиський	1	1	3	3
Жовківський	2	3	4	5
Миколаївський	1	1	2	1
Перемишлянський	1	1	2	1
Пустомитівський	2	3	4	5
Радохівський	1	1	3	3
Самбірський	2	2	1	4
Сколівський	2	3	4	5
Сокальський	2	3	4	2
Старосамбірський	1	2	1	4
Стрийський	2	3	4	2
Турківський	1	1	3	3
Яворівський	2	3	4	5

Необхідно зазначити, що отримані результати не можна вважати єдиним вирішенням задачі кластеризації, а є лише локальним по відношенню до сформованої доволно початкової матриці приналежності, що проте ніяк не применшує його значення. Знаходження глобального мінімуму у задачі оптимізації вважається доволі складним завданням у зв'язку з складністю алгоритму та великим об'ємом обчислень [8].

Отже, на підставі результатів кластерного аналізу райони Львівської області за подібністю показників рівня соціального розвитку у 2011 р. можна поділити на чотири групи (табл. 3). За цього, на відміну від багатовимірного ранжування і класифікації, неможливо однозначно розташувати ці групи за зростанням чи спаданням усередненого рівня соціального розвитку районів. Для встановлення особливостей кожної групи потрібно проводити додаткові дослідження на основі економіко-математичних моделей або залучаючи експертне середовище.

Таблиця 3

Узагальнення результатів кластерного аналізу

Група 1	Група 2	Група 3	Група 4
Кам'янка-Бузький	Буський	Бродівський	Миколаївський
Жовківський	Городоцький	Жидачівський	
Пустомитівський	Дрогобицький	Золочівський	
Сколівський	Мостиський	Самбірський	Перемишлянський
Сокальський	Радехівський	Старосамбірський	
Стрийський	Турківський		
Яворівський			

Отримані результати є підґрунтям для ґрунтового аналізу кожної групи на рахунок притаманних їй особливостей. Вони можуть бути використані під час побудови економіко-математичних моделей розвитку територій, а також під час оцінювання надійності функціонування соціальних систем.

Література

1. Селіверстов Р. Г. Нечіткий кваліметричний аналіз соціально-економічного розвитку міст обласного підпорядкування Львівської області / Р. Г. Селіверстов // Ефективність державного управління [Текст] : зб. наук. пр. — Вип. 28 / за заг. ред. проф. В. С. Загорського, доц. А. В. Ліпенцева. — Львів : ЛРІДУ НАДУ, 2011. — С. 158—164.
2. Нейский И. М. Классификация и сравнение методов кластеризации [Электронный ресурс] / И. М. Нейский. — Режим доступа : http://it-claim.ru/Persons/Neyskiy/Article2_Neyskiy.pdf.
3. Нечеткий кластерный анализ на примере социально-экономических показателей крупных городов России [Электронный ресурс]. — Режим доступа : <http://r2.pcnews.ru/articles/data-mining-means-17-29669.html>.
4. Тараскина А. С. Нечеткая кластеризация по модифицированному методу с-средних и ее применение для обработки микрочиповых данных / А. С. Тараскина // Проблемы интеллектуализации и качества систем информатики [Текст]. — 2013. — № 13. — С. 217—228.
5. Лукань Л. Застосування кластерного аналізу для оцінки розвитку малого підприємництва в регіонах України / Л. Лукань, Г. Цегелик // Формування ринкової

економіки в Україні [Текст]. — 2009. — Вип. 19. — С. 73—80.

6. Регіональна статистика Львівщини [Електронний ресурс]. — Режим доступу : http://database.ukrcensus.gov.ua/PXWEB2006p/dialog/statfile1_r.asp.

7. Fuzzy C-Means Clustering [Electronic resource]. — Access mode : http://home.deib.polimi.it/matteucc/Clustering/tutorial_html/cmeans.html.

8. Тараскина А. С. Нечеткая кластеризация...

R. Seliverstov

FUZZY CLUSTER ANALYSIS OF SOME SOCIAL DEVELOPMENT INDICATORS OF LVIV REGION DISTRICTS

On the example of analysis of some social development indicators of Lviv region districts the fuzzy clustering scheme is given. The received results can be used for the construction of efficient mathematical economic models of regional development.

Key words: fuzzy logic, clustering, classification, social development.