

УДК 519.2

Г.Б. Варшавьяк¹, А.И. Ходырев²¹Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков²Харьковский институт банковского дела Университета банковского дела НБУ, Харьков

ОЦЕНКА БЛИЗОСТИ МНОГОМЕРНОЙ КЛАССИФИКАЦИИ, ВЫПОЛНЕННОЙ ЭКСПЕРТНЫМИ МЕТОДАМИ

Показано применение меры близости между различными вариантами разбиения конечного множества объектов на пересекающиеся подмножества для оценки совпадения вариантов территориальной организации размещения производительных сил.

Ключевые слова: кластерный анализ, алгоритмы кластерного анализа, оценка устойчивости алгоритмов кластерного анализа, размещение производительных сил, региональная экономика.

Введение

В настоящее время разбиение конечного множества объектов, каждое из которых характеризуется совокупностью количественных признаков, на непересекающиеся подмножества выполняются двумя принципиально различными способами.

Первый способ – это формальная классификация (разделение объектов, группировка) – выполняемая методами кластерного анализа [1].

Второй способ – это классификация, выполняемая экспертными методами. В этом случае решение задачи неизбежно носит субъективный характер. Эксперты, выполняя классификацию, неизбежно выходят за рамки формальных ограничений, дополняя систему признаков интуитивными представлениями об особенностях изучаемой системы.

Последнее утверждение верно только для случая научной классификации. Если же эту процедуру проводят в интересах аудита или судебной власти, то эксперт обязан быть связан необходимыми процессуальными ограничениями.

Цель работы. Определить количественную меру различия между классификациями, выполненными экспертным путем, на примере территориальной организации размещения производительных сил (РПС).

Анализ литературы. Схемы принятого в настоящее время структурирования РПС Украины приведены в работах [2, 3]. Авторы данного сообщения не считают возможным, в силу своей компетенции, обсуждать корректность этих схем с позиции экономической географии и используют их только в качестве объекта приложения описываемой методики. Оценки различия применения разнообразных методов кластерного анализа к одному и тому же массиву наблюдений приведены в работах [4 – 6].

Изложение результатов

Пусть $I = \{I_1, I_2, \dots, I_m\}$ конечное множество из m объектов, характеризуемых n количественными

признаками. Пусть R и S – два различных разбиения конечного множества I . Оценкой различия этих разбиений, согласно работе [5] назовем величину

$$U(R, S) = \frac{d(R, S)}{\frac{1}{2} \left(\sum_{g=1}^{I_1} |R_g|^2 + \sum_{k=1}^{I_2} |S_k|^2 \right)}, \quad (1)$$

где

$$d(R, S) = \frac{1}{2} \left(\sum_{g=1}^{I_1} |R_g|^2 + \sum_{k=1}^{I_2} |S_k|^2 - \sum_{g=1}^{I_1} \sum_{k=1}^{I_2} |R \cap S_k|^2 \right). \quad (2)$$

В выражениях (1), (2) принято, что I_1 – количество подмножеств кластеров исходного множества m объектов, выделенных одним экспертом, I_2 – то же вторым экспертом $|R_g|$, $|S_k|$ – мощности соответствующих подмножеств. В работе [5] доказано, что величина $U(R, S)$ будет равна нулю при полностью совпадающих классификациях и единице при полностью несовпадающих.

Представим существующее административно-территориальное деление Украины в виде табл. 1.

Таблица 1

Административно-территориальное деление Украины

Условное обозначение административно-территориальной единицы	Наименование административно-территориальной единицы (АТЕ)
1.	Автономная республика Крым
2.	Винницкая область
3.	Волинская область
4.	Днепропетровская область
5.	Донецкая область
6.	Житомирская область
7.	Закарпатская область
8.	Запорожская область
9.	Ивано-Франковская область
10.	Киевская область
11.	Кировоградская область
12.	Луганская область

Окончание табл. 1

13.	Львовская область
14.	Николаевская область
15.	Одесская область
16.	Полтавская область
17.	Ривненская область
18.	Сумская область
19.	Тернопольская область
20.	Харьковская область
21.	Херсонская область
22.	Хмельницкая область
23.	Черкасская область
24.	Черновицкая область
25.	Черниговская область
26.	г.Киев
27.	Г.Севастополь

Различия между приведенными в работах [2, 3] классификациями определяли, используя условия (2), (3).

Для упрощения дальнейшего изложения примем, что классификация РПС, приведенная в [2, с. 206, 207] получила код "ЖУК". Классификация, приведенная в [3, с. 21], получила код "РПСА". Классификация, приведенная [3, с. 22], получила код "РПСБ". Классификация, приведенная [3, с. 24], получила код "Заставний". Соответствие между единицами административно-территориального деления и кластерами показано на рис. 1. Последовательность определения близости между различными кластерами (вариантами РПС) следующая.

Шаг 1. Подготовка исходных данных.

Вектор исходных данных для каждого разбиения областей по экономическим районам Украины формируется в виде столбца из 27 элементов, в котором номер строки – номер области Украины в табл.1 а значение ячейки – номер кластера, к которому отнесена область в данном разбиении.

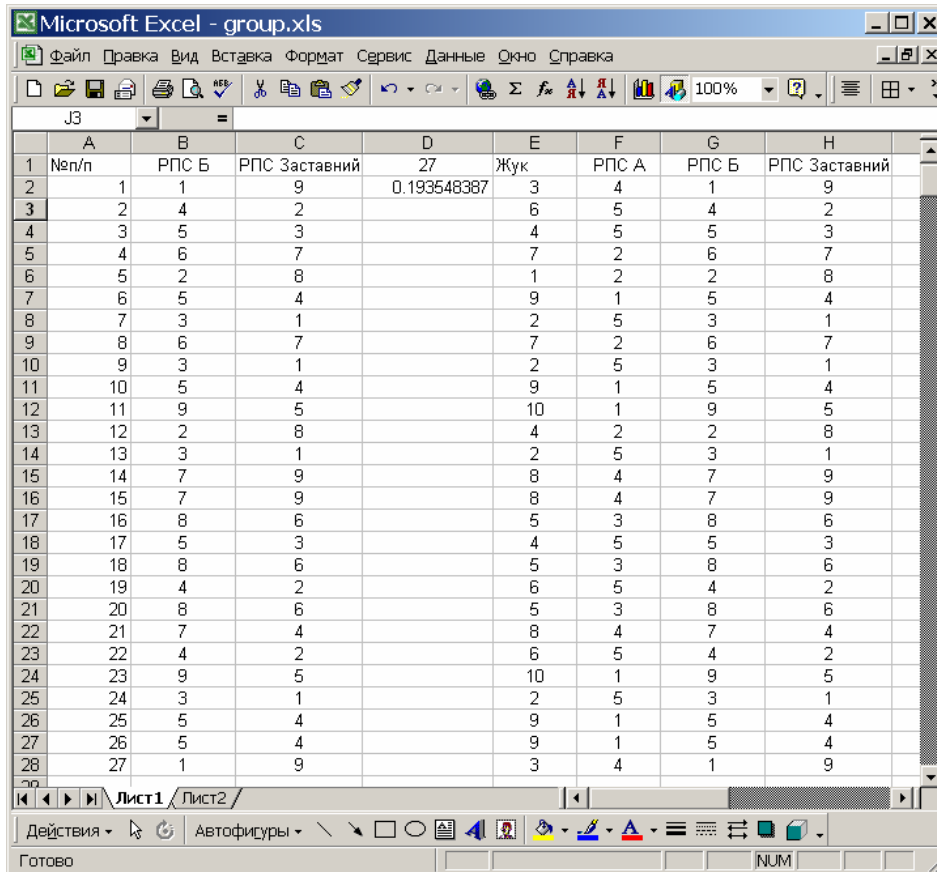


Рис. 1. Размещение исходных данных на Листе 1.

Шаг 2. Расчёт алгоритма вычисления меры сходства реализован на языке WordBasic в качестве подпрограммы, вызываемой горячей клавишей и использующей исходные данные, помещённые в документ Excel.

В столбцы В и С необходимо поместить сравниваемые выборки. В ячейке D1 должно храниться количество элементов выборки (число регионов Украины).

Результат вычисления (значение меры сходства разбиений, помещённых в столбцы В и С), автоматически будет помещён в ячейку D2 (см. рис. 1).

После заполнения столбцов В и С на Лист 1 произойдёт автоматический подсчёт числа элементов кластеров разбиения на Лист 2 (рис. 2).

Шаг 3. Повторное вычисление меры сходства для новых пар векторов и заполнение таблицы результатов.

Для заповнення таблиці результатів (рис. 3) необхідно по череді вычислити меру схожості для всіх пар векторів разбиєний, скопіював по череді вектора в стовпці В і С і вызвав процедуру вычисления значення меры схожості.

	A	B	C	D	E
1	№ кластер	R	S		
2	1	2	4		
3	2	2	3		
4	3	4	2		
5	4	3	5		
6	5	6	2		
7	6	2	3		
8	7	3	2		
9	8	3	2		
10	9	2	4		
11	10	0	0		

Рис. 2. Промежуточні розрахунки на Лист 2

	J	K	L	M	N
3					
4		Жук	РПС А	РПС Б	РПС Заставний
5	Жук	-	0.379032258	0.125	0.151162791
6	РПС А	0.379032	-	0.396946665	0.356589147
7	РПС Б	0.125	0.396946665	-	0.193548387
8	РПС Заставний	0.151163	0.356589147	0.193548387	-
9					
10					
11					

Рис. 3. Таблиця результатів

Аналізуючи приведені на цьому рисунку варіанти, можна прийти к наступному. В цілому варіанти класифікації сильно различаються між собою: від 0,125 до 0,396. Це свідчить про те, що їх автори інтуїтивно використовували різні методическі підходи при проведенні групувань.

Выводы

1. Показана применимость методов анализа формальных классификаций с классификациями, выполненными экспертными методами.

2. Определены меры схожості различных вариантов размещения производительных сил Украины.

Список литературы

1. Дюран Б., Оддел П. Методы кластерного анализа. – М.: Статистика, 1977 – 186 с.
2. Жук М.В.; Круль В.П. Розміщення продуктивних сил і економіка регіонів України. – К.: Кондор, 2006. – 296 с.
3. Розміщення продуктивних сил України. – К.: ВД "Книга", 2001, – 552 с.
4. Миркин Б.Г., Черный Л.Б. Об измерении близости между разбиениями конечного множества объектов. // Автоматика и телемеханика. 1970. – № 5. – С. 120-127.
5. Дубницький В.Ю. Оцінка устойчивости алгоритмов кластерного анализа // Информационные системы. – Х.: ХВУ, 1997. – Вып. 1 (5). – С. 129-134.
6. Андрейчиков А.В., Андрейчикова О.Н. Анализ, синтез, планирование решений в экономике – М.: Финансы и статистика, 2000. – 368 с.

Поступила в редколлегию 4.07.2008

Рецензент: канд. техн. наук, доцент В.Ю. Дубницький, Харьковський інститут банківського дела Університета банківського дела НБУ, Харків.

ОЦІНКА БЛИЗЬКОСТІ БАГАТОВИМІРНОЇ КЛАСИФІКАЦІЇ, ВИКОНАНОЇ ЕКСПЕРТНИМИ МЕТОДАМИ

Г.Б. Варшав'як, О.І. Ходирев

Показано застосування міри близькості між різними варіантами розбиття скінченної множини на непересічні підмножини для оцінки співпадіння варіантів територіальної організації розміщення продуктивних сил.

Ключові слова: кластерний аналіз, алгоритми кластерного аналізу, оцінка стійкості алгоритмів кластерного аналізу, розміщення продуктивних сил, регіональна економіка.

VALUATION OF PROXIMITY OF MULTIDIMENSIONAL CLASSIFICATION AS PERFORMED BY EXPERT METHOD

G.B. Varschaviak, A.I. Khodyrev

Application of a proximity measure shown between various version of division of a finite set of objects into non-intersecting subsets in order to evaluate coincidence of variants for territorial organization of productive force location.

Keywords: cluster analysis, algorithms of cluster analysis, estimation of stability of algorithms of cluster analysis, placing of productive forces, regional economy.