

Романов Вадим Николаевич

КЛАСТЕРНЫЙ АНАЛИЗ НА ОСНОВЕ НЕЧЕТКИХ МОДЕЛЕЙ

В статье дано обобщение кластерного анализа для случая неопределенной информационной среды с использованием нечетких моделей. Показаны преимущества предлагаемого подхода, позволяющего уменьшить неоднозначность распределения объектов по кластерам и уровням порядка, обосновать выбор меры сходства между объектами, нивелировать ошибки, связанные с несогласованностью данных.

Адрес статьи: www.gramota.net/materials/1/2013/10/46.html

Статья опубликована в авторской редакции и отражает точку зрения автора(ов) по рассматриваемому вопросу.

Источник

Альманах современной науки и образования

Тамбов: Грамота, 2013. № 10 (77). С. 147-151. ISSN 1993-5552.

Адрес журнала: www.gramota.net/editions/1.html

Содержание данного номера журнала: www.gramota.net/materials/1/2013/10/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: almanac@gramota.net

геологических пород, но и, двигаясь по уложенным человеком ходам лабиринта, задуматься над историей вообще и смыслом жизни каждого из нас. Узкие и более широкие проходы лабиринта, повороты и, наконец, поиск выхода – не это ли наша нелегкая и, порой, запутанная жизнь?

Нужно сказать, что в Ломовском природно-ландшафтном парке-музее все природные комплексы и объекты, созданные человеком, тесно связаны между собой и несут особую смысловую нагрузку. Историческая данность того или иного места в Ломах дополняется ее сегодняшним восприятием через выражение всех чувств с помощью творений рук человеческих.

За прошедшее время в парке-музее посажено около 20 тысяч деревьев и декоративных кустарников, и работа эта продолжится.

Традиционно в течение многих лет на территории парка-музея проходит Всероссийский фестиваль фольклора и ремесел «Русь песенная, Русь мастеровая». В 2012 году он был юбилейным, десятым по счету. И, конечно же, в связи с этим парк пополнится новыми объектами, чьи смысловые акценты дополнят имеющиеся.

Ломовский природно-ландшафтный парк-музей известен и в международном масштабе, сведения о нем размещены в программе «Вся планета» *Google*.

Развитие экологических потребностей является элементом гармонической системы человеческих потребностей в целом. Каждый посетитель таких музеев может найти в них что-либо интересное, отражающее его мировосприятие, осознание самого себя в этом мире. Музейные средства понимаются нами как инструмент становления и развития экологической культуры личности.

Список литературы

1. **Вершинина Т. П.** Эколого-просветительная работа в Хоперском заповеднике // Сборник научных статей, посвященный 65-летию Хоперского заповедника. Воронеж, 2000. С. 270-272.
2. **Захлебный А. Н., Дзятковская Е. Н.** Проблемы проектирования образования в интересах устойчивого развития // Экологическое образование. 2007. № 1. С. 3-6.
3. **Кужанова Н. И.** Экологическое сознание и культура в экологическом образовании: что изменилось? // Наука и школа. М.: Изд-во МГПУ, 2004. № 1.
4. **Лихачев Д. С.** Экология культуры // Памятники Отечества. М., 1980. № 2 (2).
5. **Мамедов Н. М.** Основание экологического образования // Философия экологического образования / под ред. Н. К. Лисеева. М.: Прогресс-Традиция, 2001. С. 72-78.

MUSEUM EXPOSITIONS OF VORONEZH DISTRICT AS ECO-CULTURAL SPACE

Polyanskaya Evgeniya Ivanovna, Ph. D. in Pedagogy
Borisoglebsk State Teachers' Training Institute
polanska70@mail.ru

In the article the problems of population's ecological culture formation using museum expositions are considered. The author gives a number of the examples of the ecological-enlightening activity of Voronezh region museums. These museums employees' work contributes to the development of the ecological demands of Voronezh district residents.

Key words and phrases: ecological education; ecological culture; museum exposition; environmental park-museum; eco-cultural space.

УДК 519.8

Физико-математические науки

В статье дано обобщение кластерного анализа для случая неопределенной информационной среды с использованием нечетких моделей. Показаны преимущества предлагаемого подхода, позволяющего уменьшить неоднозначность распределения объектов по кластерам и уровням порядка, обосновать выбор меры сходства между объектами, нивелировать ошибки, связанные с несогласованностью данных.

Ключевые слова и фразы: классификация; кластерный анализ; кластер; нечеткая модель; нечеткая мера сходства.

Романов Вадим Николаевич, д.т.н., профессор
Национальный минерально-сырьевой университет «Горный»
vromanvri@mail.ru

КЛАСТЕРНЫЙ АНАЛИЗ НА ОСНОВЕ НЕЧЕТКИХ МОДЕЛЕЙ[©]

Кластерный анализ является разновидностью задачи классификации, когда отсутствует множество представительства (эталонов). Он состоит в объединении объектов в группы (кластеры) в зависимости от степени сходства, определяемой по ряду критериев (признаков, свойств) [1, с. 267]. Различные варианты кластерного анализа имеют следующие недостатки и ограничения:

- использование только количественных значений критериев и статистических методов обработки данных;

- сильная зависимость результатов анализа от выбора меры сходства между объектами;
- использование в качестве меры сходства между объектами среднего или среднеквадратичного расстояния и приближения «ближайших соседей» с последующим усреднением по объединяемым объектам, что приводит к потере полезной информации и необоснованным заключениям о числе и составе кластеров и уровнях порядка;
- ошибки анализа, обусловленные использованием нетранзитивного отношения сходства вместо транзитивного отношения подобия.

Во многих случаях информация об объектах и критериях является неполной, разнородной, и имеется неопределенность в оценке объектов по критериям. Поэтому статистические методы непригодны, и использование точной числовой меры сходства может приводить к ошибкам. Целью статьи является исследование применимости нечетких моделей в кластерном анализе, что позволяет обобщить схему анализа на случай неопределенности и устранить отмеченные недостатки и ограничения. Задача формулируется в следующем виде. Дано множество объектов $X = \{x_1, \dots, x_m\}$, оцениваемых по множеству критериев (признаков, свойств) K_1, \dots, K_n , представленных нечеткими градациями в диапазоне [ОН, ОВ], где ОН – очень низкое значение, ОВ – очень высокое. Требуется распределить объекты по кластерам и уровням порядка. Алгоритм решения задачи включает следующие шаги:

1. Исходная количественная и качественная информация об объектах и критериях, полученная измерениями и экспертными методами, преобразуется в нечеткие градации способом, изложенным в [4, с. 123].

2. Определяется мера близости α_{ij}^l между объектами x_i и x_j по критерию K_l следующим образом: если объекты x_i и x_j имеют одинаковые значения критерия K_l , то полагаем $\alpha_{ij}^l = \text{ОВ}$; если x_i и x_j отличаются на одну градацию, то $\alpha_{ij}^l = \text{В}$ (высокое значение) и т.д. [3, с. 125]. Более точно значения меры близости в зависимости от значений критериев, представленных в виде нечетких градаций, определяются по матрице соответствий (Табл. 1) и дают степень близости объектов по критериям. Расстояние, определяющее степень различия между объектами x_i , x_j по критерию K_l , задается величиной, обратной (противоположной) мере близости $d_{ij}^l = \bar{\alpha}_{ij}^l$, а именно, если $\alpha_{ij}^l = \text{ОВ}$, то $d_{ij}^l = \text{ОН}$ и т.д.

Таблица 1. Матрица соответствий

Нечеткие градации x, y	ОН	ОН-Н	Н	Н-С	С	С-В	В	В-ОВ	ОВ
	$\alpha(x, y)$								
ОН	ОВ	В-ОВ	В	С-В	С	Н-С	Н	ОН-Н	ОН
ОН-Н		ОВ	В-ОВ	В	С-В	С	Н-С	Н	ОН-Н
Н			ОВ	В-ОВ	В	С-В	С	Н-С	Н
Н-С				ОВ	В-ОВ	В	С-В	С	Н-С
С					ОВ	В-ОВ	В	С-В	С
С-В						ОВ	В-ОВ	В	С-В
В							ОВ	В-ОВ	В
В-ОВ								ОВ	В-ОВ
ОВ									ОВ

Примечание. Так как $\alpha(x, y) = \alpha(y, x)$, то в таблице даны только значения выше главной диагонали. ОН – очень низкое значение, Н – низкое, С – среднее, В – высокое, ОВ – очень высокое, (ОН-Н) – между очень низким и низким значениями, (Н-С) – между низким и средним, (С-В) – между средним и высоким, (В-ОВ) – между высоким и очень высоким.

3. По матрице соответствий определяется степень близости объектов по каждому критерию.

4. Определяется степень близости объектов по всем критериям. Для этого необходимо агрегировать результаты по каждому критерию, используя подходящую модель принятия решений, которая должна выбираться так, чтобы учесть малые значения степеней близости. Использование средних (функции Хемминга, Евклида), которые, вообще говоря, не удовлетворяют этому требованию, приводит к потере полезной информации и неоднозначности объединения объектов в кластеры. Чтобы сохранить всю полезную информацию и максимально уменьшить неоднозначность объединения объектов в кластеры, используем меру близости по наибольшему различию

$$\alpha_{ij} = \min_l \alpha_{ij}^l. \quad (1)$$

Из (1) следует, что степень близости α_{ij} объектов x_i , x_j по всем критериям определяется по наименьшему значению меры близости для них по отдельным критериям. Такая оценка является надежной и позволяет уменьшить влияние неадекватности модели принятия решений на результаты анализа. Иногда удобно перейти к расстояниям, тогда (1) запишется в виде

$$d_{ij} = \max_l d_{ij}^l. \quad (2)$$

5. Итогом анализа является объединение объектов в кластеры, распределенные по уровням порядка, в соответствии со значениями нечеткой меры близости.

6. Если для группы однородных объектов результаты анализа оказываются одинаковыми, то для разделения объектов можно использовать функцию Хемминга, как наиболее удобную для вычислений. Ее применение является корректным для объектов одного уровня.

$$\alpha_{ij} = 1/n \sum_l \alpha_{ij}^l. \quad (3)$$

Рассмотрим пример. Исходные данные для примера представлены в Табл. 2.

Таблица 2. Исходные данные для примера

Объекты	Критерии				
	K_1	K_2	K_3	K_4	K_5
x_1	С	Н-С	С	Н	Н-С
x_2	С	Н	В-ОВ	ОН	Н-С
x_3	С-В	Н-С	ОН	Н-С	Н-С
x_4	Н	С-В	ОВ	Н-С	ОН
x_5	В	ОН	В	Н	Н-С
x_6	ОН-Н	С-В	В	Н-С	Н
x_7	ОН	В	ОН	ОН-Н	В
x_8	ОН	ОВ	ОН	Н-С	ОН
x_9	ОН	В-ОВ	ОН	ОН-Н	Н
x_{10}	ОВ	ОН	В	В	ОН-Н
x_{11}	В-ОВ	ОН	С-В	ОВ	Н-С
x_{12}	ОВ	ОН	В	Н-С	Н-С
x_{13}	С	ОН	В	Н-С	ОВ
x_{14}	ОН-Н	С-В	В	С-В	Н

Как видно из Табл. 2, исходные данные являются разнородными, и их оценки по критериям не согласованы, что соответствует реальной ситуации. Определим значение меры близости объектов по каждому критерию, используя Табл. 1. Для критерия K_1 результаты даны в Табл. 3.

Таблица 3. Степень близости объектов по критерию K_1

Объекты	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_1	ОВ	ОВ	В-ОВ	В	В	С-В	С	С	С	С	С-В	С	ОВ	С-В
x_2		ОВ	В-ОВ	В	В	С-В	С	С	С	С	С-В	С	ОВ	С-В
x_3			ОВ	С-В	В-ОВ	С	Н-С	Н-С	Н-С	С-В	В	С-В	В-ОВ	С
x_4				ОВ	С	В-ОВ	В	В	В	Н	Н-С	Н	В	В-ОВ
x_5					ОВ	Н-С	Н	Н	Н	В	В-ОВ	В	В	Н-С
x_6						ОВ	В-ОВ	В-ОВ	В-ОВ	ОН-Н	Н	ОН-Н	С-В	ОВ
x_7							ОВ	ОВ	ОВ	ОН	ОН-Н	ОН	С	В-ОВ
x_8								ОВ	ОВ	ОН	ОН-Н	ОН	С	В-ОВ
x_9									ОВ	ОН	ОН-Н	ОН	С	В-ОВ
x_{10}										ОВ	В-ОВ	ОВ	С	ОН-Н
x_{11}											ОВ	В-ОВ	С-В	Н
x_{12}												ОВ	С	ОН-Н
x_{13}													ОВ	С
x_{14}														ОВ

Аналогично определяется степень близости по критериям K_2, \dots, K_5 . Результаты не приводятся, чтобы не загромождать изложение, так как соответствующие расчеты можно легко проделать, используя данные Табл. 1 и Табл. 2. Теперь определим степень близости между объектами для нашего примера по всем критериям из соотношения (1). Результаты представлены в Табл. 4.

Таблица 4. Степень близости объектов по всем критериям, определенная по (1)

Объекты	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}
x_1	ОВ	С-В	С	С	С-В	С-В	С	Н-С	С	С	Н	С	Н-С	С-В
x_2		ОВ	ОН-Н	С-В	В	С-В	ОН-Н	ОН-Н	ОН-Н	Н	ОН	С	Н-С	Н-С
x_3			ОВ	ОН	Н	Н	Н-С	Н-С	Н-С	Н	Н-С	Н	Н	Н
x_4				ОВ	Н-С	В	ОН	ОН	ОН	Н	Н-С	Н	ОН	В
x_5					ОВ	Н-С	Н	ОН	ОН-Н	С	Н	В	Н-С	Н-С
x_6						ОВ	Н	Н	Н	ОН-Н	Н	ОН-Н	Н	В
x_7							ОВ	Н	С	ОН	ОН-Н	ОН	Н	Н
x_8								ОВ	В	ОН	ОН	ОН	ОН	Н
x_9									ОВ	ОН	ОН-Н	ОН	ОН-Н	Н
x_{10}										ОВ	В	С-В	ОН-Н	ОН-Н
x_{11}											ОВ	Н-С	Н-С	Н
x_{12}												ОВ	Н-С	ОН-Н
x_{13}													ОВ	Н
x_{14}														ОВ

Из данных Табл. 4 определяем последовательность объединения объектов в кластеры и их распределение по уровням порядка. Результаты анализа представлены в Табл. 5. Результаты анализа на уровнях ОВ и В-ОВ можно уточнить, используя функцию Хемминга (3). Такая процедура корректна, так как применяется к объектам одного уровня $\{x_5, x_{12}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_8, x_9\}$, $\{x_{10}, x_{11}\}$ с одинаковой наименьшей степенью близости В (высокая) в каждом кластере. Степень близости этих объектов по критериям, определенная из Табл. 1 и Табл. 2., дана в Табл. 6.

Таблица 5. Результаты кластерного анализа

Степень близости объектов	Число кластеров	Состав кластеров
ОВ	0	Все объекты разделены
В-ОВ	0	Все объекты разделены
В	4	$\{x_5, x_{12}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_8, x_9\}$, $\{x_{10}, x_{11}\}$
С-В	5	$\{x_1, x_2\}$, $\{x_5, x_{12}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_8, x_9\}$, $\{x_{10}, x_{11}\}$
С	4	$\{x_1, x_2, x_5, x_{12}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_8, x_9\}$, $\{x_{10}, x_{11}\}$
Н-С	4	$\{x_1, x_2, x_5, x_{12}, x_{13}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_3, x_8, x_9\}$, $\{x_{10}, x_{11}\}$
Н	4	$\{x_1, x_2, x_5, x_{12}, x_{13}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_3, x_8, x_9, x_7\}$, $\{x_{10}, x_{11}\}$
ОН-Н	3	$\{x_1, x_2, x_5, x_{12}, x_{13}\}$, $\{x_4, x_6, x_{14}, x_{10}, x_{11}\}$, $\{x_3, x_8, x_9, x_7\}$
ОН	1	Все объекты объединяются в один кластер

Примечание. Под кластером понимается подмножество, состоящее не менее чем из двух объектов. Кластеры отделяются фигурными скобками.

Таблица 6. Степень близости α_{ij}^l объектов $\{x_5, x_{12}\}$, $\{x_4, x_6, x_{14}\}$, $\{x_8, x_9\}$, $\{x_{10}, x_{11}\}$

i	j	$l=1$	$l=2$	$l=3$	$l=4$	$l=5$
6	14	OB	OB	OB	B	OB
5	12	B	OB	OB	B-OB	OB
8	9	OB	B-OB	OB	B	B
10	11	B-OB	OB	B-OB	B	B
4	6	B-OB	OB	B	OB	B
4	14	B-OB	OB	B	B	B

По данным Табл. 6 определим степень близости объектов по всем критериям из соотношения (3), используя правила нечеткой арифметики [2, с. 145]. Результаты для уровней OB и B-OB приведены в Табл. 7. Из расчетов следует, что объект x_4 на уровнях OB и B-OB не может быть объединен в один кластер с объектами x_6, x_{14} , так как его степень близости с объектом x_{14} , рассчитанная по (3), составляет B.

Таблица 7. Результаты кластерного анализа на уровнях OB и B-OB по (3)

Степень близости объектов	Число кластеров	Состав кластеров
OB	1	$\{x_6, x_{14}\}$
B-OB	4	$\{x_6, x_{14}\}, \{x_5, x_{12}\}, \{x_8, x_9\}, \{x_{10}, x_{11}\}$

Полученные результаты показывают определенную аналогию между кластерным анализом и анализом связности (топологическим анализом), применение которого к комплексам разностей объектов может быть полезным, хотя и требует дополнительных усилий. Расчеты, проведенные с использованием числовой меры близости, определяемой функцией Евклида и приближением ближайших соседей, показывают, что результаты являются похожими, но не идентичными. В частности, наблюдается увеличение числа уровней порядка до 13 (вместо 9 при нашем анализе), причем разность между этими дополнительными уровнями составляет менее 0,01 (относительная разность колеблется от 0,7% до 1,5%), т.е. не превышает ошибки данных. Появление таких дополнительных уровней трудно обосновать и интерпретировать.

Таким образом, предложенный подход позволяет проводить кластерный анализ в условиях неопределенности исходных данных, уменьшить ошибки и неоднозначность объединения объектов в кластеры, делает более обоснованным определение уровней порядка и числа кластеров на каждом уровне по сравнению с анализом, использующим точную числовую меру сходства. Кроме того, значительно уменьшается трудоемкость анализа.

Список литературы

1. Айвазян С. А., Бухштабер В. М., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Классификация и снижение размерности. М.: Финансы и статистика, 1989. 607 с.
2. Романов В. Н. Нечеткие модели принятия решений // Альманах современной науки и образования. Тамбов: Грамота, 2013. № 5. С. 144-147.
3. Романов В. Н. Нечеткие системы. СПб.: ЛЕМА, 2009. 183 с.
4. Романов В. Н. Определение существенных признаков в задачах идентификации топологическими методами // Альманах современной науки и образования. Тамбов: Грамота, 2013. № 7. С. 122-129.

CLUSTER ANALYSIS BASING ON FUZZY MODELS

Romanov Vadim Nikolaevich, Doctor in Technical Sciences, Professor
National Mineral Resources University
vromanvpi@mail.ru

In the article the generalization of cluster analysis for indefinite information environment case using fuzzy models is given. The advantages of the suggested approach, which allows reducing the ambiguity of objects distribution according to clusters and order levels, substantiating the choice of affinity measure between objects, and leveling mistakes connected with data disagreement are shown.

Key words and phrases: classification; cluster analysis; cluster; fuzzy model; fuzzy measure of affinity.