

Интернет-журнал «Наукоедение» ISSN 2223-5167 <http://naukovedenie.ru/>

Том 9, №1 (2017) <http://naukovedenie.ru/vol9-1.php>

URL статьи: <http://naukovedenie.ru/PDF/09TVN117.pdf>

Статья опубликована 17.02.2017

Ссылка для цитирования этой статьи:

Миронова Ю.Н. Трудности классификации геоинформационных объектов // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 9, №1 (2017) <http://naukovedenie.ru/PDF/09TVN117.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.

УДК 910(075.8)+51(71+74)

Миронова Юлия Николаевна

ФГАОУ ВО «Казанский (Приволжский) федеральный университет»

Елабужский институт (филиал), Россия, Елабуга¹

Доцент кафедры «Математики и прикладной информатики»

Кандидат физико-математических наук

Профессор РАЕ

E-mail: mironovajn@mail.ru

РИНЦ: http://elibrary.ru/author_profile.asp?id=384800

SCOPUS: <http://www.scopus.com/authid/detail.url?authorId=6603018879>

Трудности классификации геоинформационных объектов

Аннотация. В связи с тем значением, которое имеют классификации, возможность автоматизации процесса их получения служит важным средством в геоинформатике (и в других науках).

При классификации геоинформационных понятий приходится сталкиваться с различными трудностями. Часть их связана с проблемой соответствия применяемых методов существу и уровню строгости поставленных задач. Это вызывает необходимость дополнительного учета геоинформационных аспектов моделируемых явлений.

При многих классификациях возникает проблема оптимального выбора системы исходных показателей. Большинство классификационных задач в геоинформатике имеют показатели различной природы: количественные и качественные. Это накладывает определенные ограничения на возможность использования различных методов классификации в геоинформатике.

Целью классификаций является получение некоторого заранее заданного (или не заданного) количества групп геоинформационных объектов (классов геоинформационных объектов). Объекты одного класса должны быть максимально однородны, а объекты разных классов - максимально различаться.

Исследуемая совокупность геоинформационных объектов описывается различными показателями и метриками, кроме того, она фиксирована в пространстве M атрибутивных признаков. В этом пространстве геоинформационные объекты теряют свою первоначальную природу и независимо от нее становятся M -мерными точками.

Изучая данные, полученные с помощью сопоставления объектов, мы можем разделить объекты на некоторые классы.

¹ 423604, Россия, Республика Татарстан, г. Елабуга, ул. Казанская, д. 89, Главный корпус, к. 85

В данной работе мы вкладываем наше множество объектов в топологическое пространство (в данном случае - в лексикографически упорядоченный квадрат), и можем исследовать различные случаи, используя ранее полученные свойства нашего топологического пространства. Важно представление нашего множества точек $M(x, y)$, как подмножества лексикографически упорядоченного квадрата. При добавлении новых точек мы можем не изменять общую структуру множества, порядок следования точек определяется их весами.

Ключевые слова: геоинформационная система; классификация; геоинформатика; класс объектов; матрица; метрика; топология; топологическое пространство; лексикографически упорядоченный квадрат; подмножество

Введение

Классификация изучаемых явлений и процессов в присуща всем наукам, в том числе и геоинформатике, в которой классификация часто выступает не только методом, но и целью научного исследования.

В связи с тем значением, которое имеют классификации, возможность автоматизации процесса их получения служит важным средством в геоинформатике (и в других науках).

Большое количество методов классификации допускает построение соответствующих им формальных алгоритмов - алгоритмов автоматической классификации. Обычно эти алгоритмы создаются без учета специфики тех наук, в которых они могут использоваться. Такая их универсальность имеет положительное значение, позволяя применять в геоинформатике (см. [3], [4], [13], [15]) алгоритмы, уже разработанные математиками-прикладниками или специалистами в области естественных наук.

При использовании универсальных алгоритмов могут возникнуть сложности, связанные с недоучетом специфики геоинформационных задач и подходов, поэтому требуется анализ современного состояния классификаций в геоинформатике, специфических черт и перспектив их развития. Рассмотрим эти проблемы подробнее.

При классификации геоинформационных понятий приходится сталкиваться с различными трудностями. Часть их связана с проблемой соответствия применяемых методов существу и уровню строгости поставленных задач. Это вызывает необходимость дополнительного учета геоинформационных аспектов моделируемых явлений.

Формулировки геоинформационных задач, описания явлений допускают некоторый разброс характеристик явлений, по крайней мере на современном этапе исследований. То есть алгоритмы многомерной классификации имеют более высокий уровень строгости, чем уровень строгости и точности самих задач. Это иногда приводит к результатам, не отвечающим существу и содержательному смыслу геоинформационных задач. Поэтому возникла и стала применяться в исследованиях теория нечетких множеств (см. [12], [2]), а также попытки разработки на ее основе методов классификации.

При многих классификациях возникает проблема оптимального выбора системы исходных показателей. С одной стороны, эта система должна всесторонне описывать изучаемое явление, с другой - включение в рассмотрение всех доступных данных может привести к их избыточности. Данные не должны дублировать друг друга, зависеть друг от друга и т.д. В противном случае они могут привести к искажению конечного результата.

Трудно найти тот критерий, который позволит оценить необходимость включения того или иного показателя как индикатора характеристики геоинформационного объекта. Большую помощь здесь может оказать предварительное всестороннее изучение геоинформационного

объекта, что позволит установить круг показателей, его отображающих. Другой путь решения этой проблемы - экспериментальная проверка степени влияния данных на конечный результат.

Также мы должны учесть различную степень значимости используемых показателей для характеристики объектов. Некоторые из них столь важны, что их исключение не позволяет моделировать объект, другие же лишь дополняют основную систему.

Это требует «взвешивания» показателей, что ведет к изменению степени их влияния на конечный результат. Однако определение «весов» - самостоятельная, сложная, во многом не решенная задача. Она может решаться, например, экспертным опросом специалистов по теме исследования.

Большинство классификационных задач в геоинформатике имеют показатели различной природы: количественные и качественные. Это накладывает определенные ограничения на возможность использования различных методов классификации в геоинформатике.

Так как большая часть данных, которые учитываются при геоинформационных классификациях, имеет качественный характер, то соответствующие алгоритмы должны уметь работать с характеристиками нечисловой природы. Создание таких алгоритмов - это развивающаяся область исследований.

Показатели, применяемые для описания какого-либо признака у различных объектов, часто бывают несопоставимы. Это связано как с различием подходов к оценке показателя для различных объектов, так и с отсутствием объективных методов их определения.

Существенны трудности при выборе наиболее подходящей математической модели для конкретного объекта. В настоящее время имеется большое количество моделей, пригодных для решения классификационных задач, но не отвечающих в полной мере специфике отображаемых геоинформационных объектов.

Создание математической модели геоинформационного объекта начинается с нормировки исходных показателей. Например, при оценках геоинформационных объектов и их систем необходимо привести систему исходных данных к такому виду, когда используемые показатели описывают отклонения характеристик геоинформационного объекта от оптимальных оценок. Это позволяет правильно задать ориентацию показателей между наихудшим и наилучшим условием для каждого из них, и тем самым правильно соизмерить их между собой [3].

Для классификаций геоинформационных объектов важны временные характеристики изучаемых объектов. Геоинформационные объекты и их системы должны рассматриваться как пространственно-временные образования.

Важными понятиями является районирование, определяемое как процедура вычленения целостных территориальных геоинформационных систем, а также привязка объектов к земной поверхности.

Классификации геоинформационных объектов можно разделить на следующие классы:

- по сходству признаков;
- по связям.

При классификациях могут решаться задачи как разделения исследуемой совокупности объектов и явлений на классы, так и отнесение одного или нескольких объектов (или явлений) к уже существующим классам. Классы могут быть описаны одним из способов:

- а. перечислением списка объектов и явлений, ими охватываемых;

- b. указанием общих свойств объектов и явлений, включенных в них;
- c. характеристикой отдельных их представителей, рассматриваемых как типичные [3].

Методы классификации можно разделить на следующие группы:

- 1. по цели: оценочные и типологические классификации;
- 2. по определению класса: обычные и нечеткие [12];
- 3. по наличию известной заранее информации о статистических свойствах классов: параметрические и непараметрические.

Наконец, кроме классификации самих геоинформационных объектов, возникает задача классификации их взаимодействий. Она гораздо труднее из-за увеличения количества параметров для описания процессов этих взаимодействий.

Основная часть

Рассмотрим процесс классификации геоинформационных объектов

Объектом классификации, как правило, является некоторый геоинформационный объект [3].

Обозначим всё множество геоинформационных объектов символом

$$O = \{o_1, o_2, \dots, o_N\} \quad (1)$$

где: o_i - i -й геоинформационный объект, N - количество геоинформационных объектов.

Целью классификаций является получение некоторого количества групп геоинформационных объектов (классов геоинформационных объектов). Объекты, принадлежащие одному классу геоинформационных объектов, должны быть максимально однородны, а объекты разных классов - максимально различаться.

Группа (класс) геоинформационных объектов иногда называется кластером или таксоном, а методы получения классов называют методами классификации, кластерного анализа, распознавания образов или числовой таксономии.

Исследуемая совокупность геоинформационных объектов описывается различными показателями и метриками, кроме того, она фиксирована в пространстве M атрибутивных признаков. В этом пространстве геоинформационные объекты теряют свою первоначальную природу и независимо от нее становятся M -мерными точками.

Результатом представления геоинформационного объекта в признаковом пространстве является некоторая матрица. Если мы имеем N объектов, обладающих каждый M признаками, то

$$X = \begin{pmatrix} o_1 \\ \dots \\ o_N \end{pmatrix} = (o^{(1)}, \dots, o^{(M)}) = \begin{pmatrix} o_1^{(1)} & \dots & o_1^{(j)} & \dots & o_1^{(M)} \\ o_2^{(1)} & \dots & o_2^{(j)} & \dots & o_2^{(M)} \\ \dots & \dots & \dots & \dots & \dots \\ o_{N-1}^{(1)} & \dots & o_{N-1}^{(j)} & \dots & o_{N-1}^{(M)} \\ o_N^{(1)} & \dots & o_N^{(j)} & \dots & o_N^{(M)} \end{pmatrix}$$

Здесь $o_i^{(j)}$ - значение j -го признака для i -го элемента нашей системы геоинформационных объектов.

Можно сопоставить две таких матрицы, и получить матрицу различий наших систем:

$$A = \begin{pmatrix} a_{11} & \dots & a_{ij} & \dots & a_{1M} \\ a_{21} & \dots & a_{2j} & \dots & a_{2M} \\ \dots & \dots & \dots & \dots & \dots \\ a_{i1} & \dots & a_{ij} & \dots & a_{iM} \\ \dots & \dots & \dots & \dots & \dots \\ a_{N1} & \dots & a_{Nj} & \dots & a_{NM} \end{pmatrix}$$

Здесь a_{ij} - результат сопоставления i -й и j -й системы геоинформационных объектов. Обычно a_{ij} означает меру различия (или сходства) объектов.

Переход к матрице A осуществляется с помощью задания метрики d (или расстояния между геоинформационными объектами [8]).

Изучая данные, полученные с помощью сопоставления объектов, мы можем разделить объекты на некоторые классы.

Пусть некоторый объект z имеет n признаков x_1, x_2, \dots, x_n . В общем случае они могут иметь различную важность (или вес), но мы расположим их в порядке уменьшения важности признаков, т.е. так, чтобы $x_1 < x_2 < \dots < x_n$, здесь, чем больше номер признака, тем меньше его вес. Причем для всех $i=1, n$ имеем $x_i \in [0,1]$. Этого можно добиться, разделив соответствующий вес на определенную величину.

Далее, каждый из признаков x_1, x_2, \dots, x_n имеет n характеристик y_1, y_2, \dots, y_n , которые таким же образом распределены в порядке убывания их важности, для всех $j=1, n$ имеем $y_j \in [0,1]$.

Таким образом, множество точек вида $M = (x_1(y_1, y_2, \dots, y_n), x_2(y_1, y_2, \dots, y_n), \dots, x_n(y_1, y_2, \dots, y_n))$ вкладывается в квадрат Q , описанный ниже. То есть имеем множество точек $M(x, y)$, где вводится следующий порядок:

Пусть $M_1(x_1, y_1), M_2(x_2, y_2)$. Тогда $M_1 < M_2$, если:

- 1) $x_1 < x_2$;
- 2) $x_1 = x_2, y_1 < y_2$.

То есть порядок определяется, как в лексикографически упорядоченном квадрате Q . Таким образом, мы вкладываем наше множество объектов в топологическое пространство (см., например, [8]), и можем исследовать различные случаи, используя ранее полученные свойства нашего топологического пространства.

Лексикографически упорядоченный квадрат. Описание пространства

Напомним определение нашего пространства [10].

Рассмотрим на плоскости OXY замкнутый квадрат со сторонами, параллельными осям координат и вершинами $(0,0), (0,1), (1,0), (1,1)$ (см. рис. 1) и упорядочим множество всех точек $z = (x, y), 0 \leq x \leq 1, 0 \leq y \leq 1$ этого квадрата в лексикографическом порядке, то есть:

$(x, y) < (x', y')$, если $x < x'$ или
если $x = x'$ и $y < y'$.

Полученные в результате такого упорядочения порядковые интервалы и полуинтервалы $[0, \alpha [$ и $] \beta, 1]$ образуют базу нашего пространства Q .

Эти интервалы имеют следующий вид: пусть даны $z_1 < z_2$, $z_1 = (x_1, y_1)$, $z_2 = (x_2, y_2)$, причем $x_1 < x_2$, тогда для любой точки z , лежащей в полосе $0 \leq y \leq 1, x_1 < x < x_2$, мы получим, что $z_1 < z < z_2$.

Полуинтервалы $x = x_1, y_1 < y \leq 1$ и $x = x_2, 0 \leq y \leq y_2$ также содержатся в порядковом интервале $[z_1, z_2]$, если $y_1 \neq 1, y_2 \neq 0$.

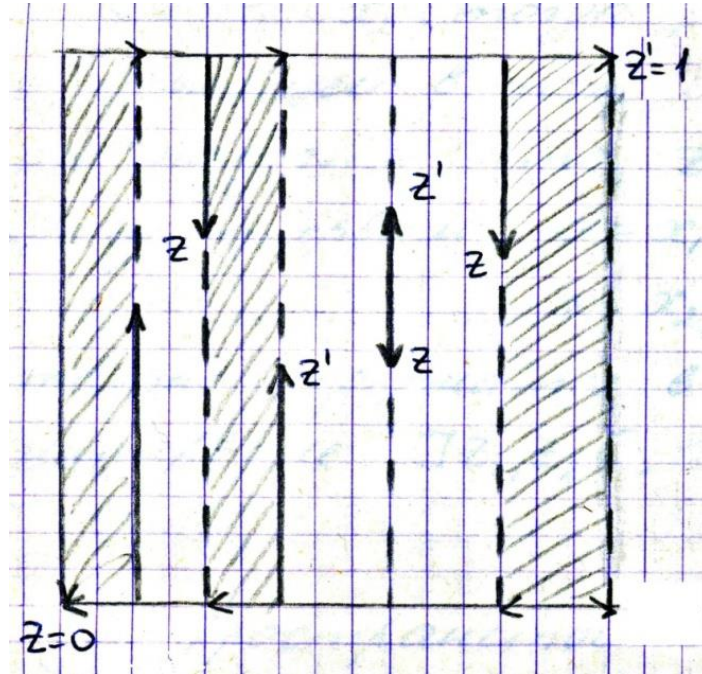


Рисунок 1. Лексикографически упорядоченный квадрат

Итак, множество точек вида $M = (x_1(y_1, y_2, \dots, y_n), x_2(y_1, y_2, \dots, y_n), \dots, x_n(y_1, y_2, \dots, y_n))$ вкладывается в квадрат Q (см. рис. 2). Мы вкладываем наше множество объектов в топологическое пространство, и можем исследовать различные случаи, используя ранее полученные свойства нашего топологического пространства.

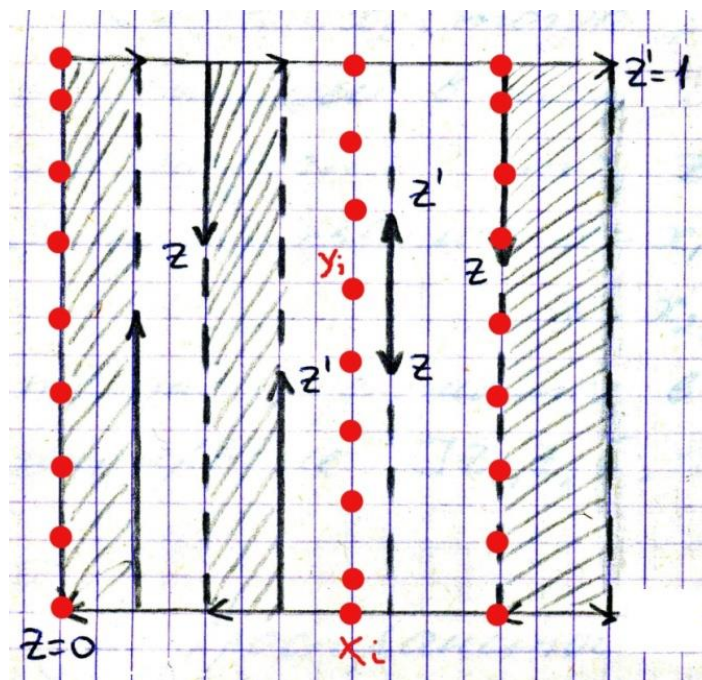


Рисунок 2. Лексикографически упорядоченный квадрат с вложенной в него системой точек

В работе [10] доказаны следующие свойства лексикографического квадрата:

1. Пространство является линейно упорядоченным и содержит наибольший и наименьший элемент.
2. Интервалы и полуинтервалы $[0, \infty [$ и $] \beta, 1]$ образуют базу некоторой топологии на Q .
3. Существование системы мощности c попарно не пересекающихся интервалов.
4. Несепарабельность.
5. Хаусдорфовость.

Также можно доказать и другие топологические свойства лексикографически упорядоченного квадрата.

Немалую роль здесь играет и наглядное представление нашего множества точек $M(x, y)$, как подмножества лексикографически упорядоченного квадрата. При добавлении новых точек мы можем не изменять общую структуру множества, порядок следования точек определяется их весами.

Заключение

Таким образом, мы вкладываем наше множество объектов в топологическое пространство, являющееся лексикографическим пространством, и можем классифицировать наши объекты, добавлять в нашу систему объектов новые объекты, исследовать появляющиеся топологические свойства наших объектов, используя известные свойства лексикографически упорядоченного квадрата.

ЛИТЕРАТУРА

1. Бабенко Л.К. Защита данных геоинформационных систем: учеб. пособие для студентов вузов / Бабенко Л.К., Басан А.С, Журкин И.Г. и др. Под ред. И.Г. Журкина. - М.: Гелиос АРВ, 2010. - 336 с. - 400 экз. - ISBN 978-5-85438-198-7.
2. Боженюк А.В., Розенберг И., Ястребинская Д.Н. Нахождение живучести нечетких транспортных сетей с применением геоинформационных систем. - М.: Научный мир, 2012. - 176 с.
3. Капралов Е.Г. Геоинформатика: в 2 кн. Кн. 1: учебник для студ. высш. учеб. заведений / Е.Г. Капралов, А.В. Кошкарёв, В.С. Тикунов и др.; под ред. В.С. Тикунова. - 3-е изд., перераб. и доп. - М.: Издательский центр «Академия», 2010. - 400 с. - 2500 экз. - ISBN 978-5-7695-6468-0.
4. Капралов Е.Г. Геоинформатика: в 2 кн. Кн. 2: учебник для студ. высш. учеб. заведений / Е.Г. Капралов, А.В. Кошкарёв, В.С. Тикунов и др.; под ред. В.С. Тикунова. - 3-е изд., перераб. и доп. - М.: Издательский центр «Академия», 2010. - 432 с. - 2500 экз. - ISBN 978-5-7695-6820-6.
5. Миронова Ю.Н. Состав геоинформационной системы // Актуальные проблемы гуманитарных и естественных наук. №4 (75) 2015 г. Ч.1., с. 88-90.
6. Миронова Ю.Н. Геоинформационные системы // Актуальные проблемы гуманитарных и естественных наук №03 (62) 2014 Ч.1., Москва, с. 63-65.
7. Миронова Ю.Н. Применение систем глобального позиционирования в геоинформационных системах // Теоретические и прикладные проблемы географии: Материалы международной научно-практической конференции (Астана, 9-10 июня 2014 г.). - Астана, 2014, часть II, с. 307-309.
8. Миронова Ю.Н. Математические аспекты геоинформатики // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 7, №5 (2015) <http://naukovedenie.ru/PDF/93TVN515.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ. DOI: 10.15862/93TVN515.
9. Миронова Ю.Н. Новые методы виртуального моделирования в геоинформационных технологиях // Интернет-журнал «НАУКОВЕДЕНИЕ» Том 8, №5 (2016) <http://naukovedenie.ru/PDF/03TVN516.pdf> (доступ свободный). Загл. с экрана. Яз. рус., англ.
10. Миронова Ю.Н. Некоторые топологические свойства лексикографически упорядоченного квадрата // Международный журнал прикладных и фундаментальных исследований. - 2015. - №12-10. - С. 1908-1909. URL: <http://www.applied-research.ru/ru/article/view?id=8408>.
11. Скворцов А.В., Сарычев Д.С. Технология построения и анализа топологических структур для геоинформационных систем и систем автоматизированного проектирования // Вестник Томского государственного университета. 2002. №275. с. 60-63.
12. Яхьяева Г.Э. Нечеткие множества и нейронные сети: Учебное пособие. - М.: Интернет-Университет Информационных технологий; БИНОМ, 2006. - 316 с.
13. Mironova Yu.N. The classification of geoinformation objects // International Journal Of Applied And Fundamental Research. - 2016. - №2 - URL: www.science-sd.com/464-25151 (29.09.2016).
14. Интернет-ресурс КБ «Панорама» <http://www.gisinfo.ru>.
15. Интернет ресурс «Геоинформатика. Под редакцией проф. В.С. Тикунова. М.: «Академия», 2005» <http://www.studfiles.ru/preview/1817795/>.

Mironova Yuliya Nikolaevna

Kazan (Volga region) federal university
Elabuga institute (branch), Russia, Elabuga
E-mail: mirnovajn@mail.ru

The difficulties of classification of geoinformation objects

Abstract. In connection with the importance of classification, possibility of automation of their production process serves as an important tool in Geoinformatics (and in other Sciences).

The classification of the geographic information concepts has to face various difficulties. Of them are related to the problem of matching methods applied to the substance and the level of rigor of the tasks. This causes the need for additional accounting GIS aspects of the modeled phenomena.

In many classifications, there arises the problem of optimal choice of the system baseline. Most classification tasks in Geoinformatics indicators are of different nature: quantitative and qualitative. This imposes certain restrictions on the use of different classification methods in Geoinformatics.

The purpose of classifications is to obtain some pre-specified (or not specified) the number of groups of GIS objects (of classes of geospatial objects). The objects of the same class should be as homogeneous as possible, whereas objects of different classes as much as possible to vary.

The population under study geographic information describes the various indices and metrics, moreover, is fixed in space M attribute characteristics. In this space of GIS objects lose their original nature and independent of it become M -dimensional points.

By studying the data obtained by mapping objects, we can divide the objects in some classes.

In this paper, we put our a set of objects in a topological space (in this case - in the lexicographically ordered square), and can explore the different cases, using the previously obtained properties of our topological space. It is important to the performance of our set of points $M(x,y)$, as a subset of the lexicographically ordered square. Adding new points we can not change the overall structure of the set, the order of points is determined by their weights.

Keywords: geographic information system; classification; Geoinformatics; class of objects; the matrix; metrics; topology; topological space; lexicographically ordered square; subset