

Идентификация пространственных группировок почвенной мезо-фауны на уровне микрорельефа

У роботі проведено аналіз класифікаційних дендрограм з метою ідентифікації просторових угруповань ґрунтової фауни. Такий підхід ще не дуже поширений в індикаційних дослідженнях, але є досить перспективним. Показні переваги та недоліки цього методу в порівнянні з іншими класифікаційними техніками.

Информацию о характере взаимодействия почвенных животных с условиями окружающей среды возможно получить в полевых и лабораторных условиях. Лабораторные эксперименты отличаются гораздо более высокой точностью и возможностью проверки результатов в повторных исследованиях. Однако адекватность полученных результатов реальным условиям остается весьма неопределенной.

В условиях полевого эксперимента очень сложно получить точные и воспроизводимые результаты, учесть множество действующих факторов и их взаимодействие. Эффективным методическим приемом изучения закономерностей формирования структуры комплексов почвенных животных является комбинация пространственно упорядоченного способа получения почвенно-зоологических проб и статистического анализа данных.

Экологические системы можно отнести к разряду сложных систем, которые далеко не всегда могут быть описаны достаточно точно на языке статистических или иных аналитических математических моделей. Геометрический подход переводит задачу формирования знаний на язык геометрических отношений между эмпирическими фактами, выступающими целостными информационными единицами и отображаемыми точками в пространстве признаков. Это, с одной стороны, делает более понятными критерии и принципы построения правил вхождения объектов в определенные классы эквивалентности, которые основываются на сравнении объектов с помощью мер, имеющих интерпретацию расстояний. С другой стороны, следует иметь в виду, что использование геометрического подхода при неограниченном расширении множества эмпирических фактов автоматически приводит к минимальным, теоретически достижимым ошибкам при принятии решений. Кроме этого, все операции без затруднений распараллеливаются, а визуализация геометрической структуры множества точек позволяет организовать исследование закономерностей в совокупности эмпирических фактов средствами интерактивной когнитивной графики. Более того можно получать наглядные визуальные представления о логических закономерностях в структуре данных — для этого применяется специальная локальная геометрия. В отличие от символьных логических методов, реализующих операции над признаками, в геометрическом подходе главными элементами выступают объекты, а основным видом операций является определение расстояния между объектами в многомерном пространстве признаков.

Методы формирования знаний имеют много общего с решением задач классификации, диагностики и распознавания образов. Но одной из главных отличительных черт методов формирования знаний является функция интерпретации закономерностей, кладущихся в основу правил нахождения объектов в классы эквивалентности.

Коренная проблема геометрического подхода состоит в поиске ответа на вопрос: какие признаки и какую меру следует выбрать для определения расстояний между объектами? В известных методах анализа данных эта задача формулируется как подбор взвешенной метрики с использованием обучающей и частично обучающей информации [1] или как оцифровка переменных, основанная на максимизации статистического критерия [3]. Удачное решение указанной задачи геометрически выражается расширением «сферы действия» объектов, выступающих в роли представителей своих классов эквивалентности.

Традиционные методы анализа многомерных данных, опирающиеся на геометрическую метафору, используют представление об общем пространстве признаков для всех объектов и об одинаковой мере, применяемой для оценки их сходства или различия. Такое представление уместно, например, при изучении однородных физических феноменов на статистическом уровне системной организации, в которых объект можно рассматривать как реализацию многомерной случайной величины с ясным физическим смыслом, когда есть все основания интерпретировать зафиксированные особенности объектов как случайные флуктуации, обусловленные воздействием шумов, погрешностями измерительных приборов и т. п. [2].

В задачах формирования знаний, когда мы имеем дело с системами надкибернетического уровня сложности, каждый объект следует рассматривать как самостоятельный информационный факт (совокупность зафиксированных значений признаков), имеющий важные уникальные особенности. Последние раскрываются путем конструирования для любого объекта собственного пространства признаков и нахождения индивидуальной меры, определяющих иерархию его сходства с другими объектами, релевантную заданному контексту. Без такого раскрытия описания объектов они нивелированы, могут содержать много ненужных, шумящих, отвлекающих и даже вредных деталей, и «сферы действия» объектов как представителей своих классов эквивалентности сужаются [2].

Конструирование собственного пространства признаков и нахождение индивидуальной меры называется локальным преобразованием пространства признаков.

Задача определения контекстно-зависимой локальной метрики заключается в нахождении линейного преобразования новой векторной переменной. Для определения такого преобразования подходит хорошо разработанный аппарат методов многомерного линейного анализа данных.

Критерий качества локальной метрики определяется контекстом, а его конкретная форма задается исследователем. Например, с учетом информации о принадлежности объектов к тем или иным классам эквивалентности это может быть стандартный для линейного дискриминантного анализа критерий, построенный на отношении разброса между классами к внутриклассовому разбросу. Или, имея в виду сложную неоднородную структуру классов, целесообразнее

строить критерий качества на оценке первых ближайших к анализируемому объектам, то есть фактически на локальной оценке отношения правдоподобия.

Формирование знаний методами локальной геометрии может быть успешно применено для решения задач диагностики в экологических исследованиях. Область проблем диагностики весьма широка, но в данной работе мы рассмотрим анализ конкретного случая.

Для изучения пространственного распределения почвенной мезофауны в чернокленовой дубраве в средней трети склона правого берега р. Самара было заложено три параллельных трансекты, вдоль которых располагались почвенно-зоологические пробы размером 50x50 см с интервалом 2,5 метра. Трансекты находились на расстоянии 2,5 между собой. Всего было сделано 75 проб.

Изучение пространственного распределения почвенных беспозвоночных с помощью кластерного анализа позволило выделить три группировки почвенных беспозвоночных. Каждый из кластеров представляет собой определенную группировку почвенных беспозвоночных, которую можно выделить в пределах склона правого берега р. Самара. Количественные характеристики группировок почвенной мезофауны, характерных для выделенных кластеров, представлены в таблице 1. В задачу исследования входит выявление видов почвенных животных, которые могут диагностировать соответствующие группировки животного населения. Эта задача будет решена методами локальной геометрии.

Задача классификации выделенных группировок и отыскание наиболее информативных признаков может быть решена методами дискриминантного анализа. Дискриминантный анализ — достаточно часто используемая процедура для решения задач распознавания образов с обучением. Для выявления положительных особенностей альтернативных процедур приведем его в качестве отправной точки в процессе идентификации экологических объектов.

На рисунке 1 в пространстве дискриминантных функций 1 и 2 показаны кластеры. Как мы видим, этот подход позволяет весьма четко разделить группировки. Таким образом, применяя дискриминантные функции, мы достаточно точно можем идентифицировать принадлежность интересующей нас пробы к соответствующей группировке. Однако, формальное решение задачи и выявление количественных зависимостей, позволяющих провести диагностику не дает качественной интерпретации зависимостей, лежащих в основе различий между группами. Достоверность интерпретации дискриминантных функций и определение их экологического смысла являются субъективными при всей формальности процедуры и зависят от квалификации исследователя и его точки зрения на рассматриваемый объект. Для определения экологического смысла дискриминантных функций может быть использован анализ стандартизированных коэффициентов этих функций, которые приведены в табл. 1. Мы видим, что значения дискриминантной функции 1 в основном определяются изменением плотности дождевого червя *Aporrectodea rosea*, и отчасти — других степных видов почвенных беспозвоночных. Различия вдоль оси, которая задается функцией 2, зависят от изменения плотности влаголюбивых видов, таких как *Octolasion lacteum*, *Enchitreidae* sp. sp., *Schisoturanus dmitriewi* и некоторых других. Таким образом, можно заключить, что изменения структуры комплекса почвенных беспозвоночных чернокленовой дубравы в средней части склона зависят от процессов сивлатизации и обеспеченности почвенного покрова влагой.

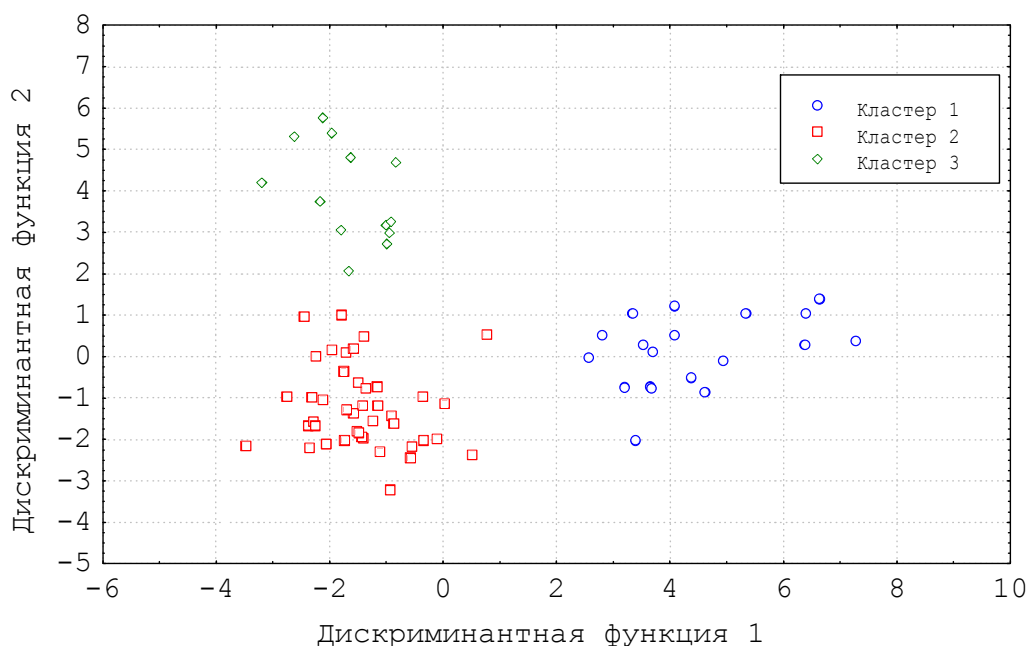


Рис. 1. Распределение пробных площадей в пространстве дискриминантных функций

Однако, полученные результаты допускают и другую интерпретацию. Например, можно допустить, что между указанными группами существуют антагонистические взаимоотношения различной природы — трофической, химической и т.д.

Простые и ясные вербальные правила диагностики дискретных группировок можно получить методами локальной геометрии. Идея такого подхода состоит в следующем. При анализе явления или процесса мы переходим от пространства с началом координат в некоторой абстрактной точке в другое — локальное пространство. Локальное пространство образуется путём перехода к новой векторной переменной, например, $\Delta = |x - x_i|$, где x_i — выбранный объект, относящийся к одной из анализируемых группировок. Он называется центральным объектом. Локальное пространство может быть построено либо для одного объекта каждого класса, либо можно построить несколько множеств для нескольких объектов. Логические закономерности в локальном пространстве отыскиваются с помощью комбинации применения методов линейной алгебры и интерактивной графики.

Алгебраическими методами ищется новая ось в локальном пространстве (вектор), на которой распределение проекций объектов удовлетворяет заданному критерию (например, выражающему стремление сгруппировать около нулевой отметки объекты того же класса, что и центрального объекта). Так как интерес представляют только сравнительно небольшая область около нулевой отметки на новой оси, то удаленные от данной отметки объекты подвергаются исключению с использованием средств интерактивной графики. После

каждого исключения параметры новой оси перерасчитываются и визуальный анализ полученного распределения дает основание для произведения ещё одного акта исключения объектов, либо для остановки процедуры поиска логической закономерности.

Таблица 1. Плотность населения кластеров (экз./пробу) и стандартизованные коэффициенты дискриминантных функций

Вид	№ кластера			Всего по площади	№ функции	
	1	2	3		1	2
Octolasion lacteum	2,37	3,16	11,62	4,43	0,23	1,16
Aporrectodea rosea	6,68	0,79	0,69	2,27	-1,08	0,23
Enchytraeidae sp. sp.	1,74	1,42	2,62	1,71	-0,08	0,53
Megaphyllum r. rossicus	1,68	0,93	1,77	1,27	0,01	-0,26
Curculionidae sp.sp.	0,84	1,02	1,31	1,03	0,01	-0,08
Aranea sp. sp.	0,68	0,56	0,85	0,64	0,08	0,02
Schisoturanus dmitriewi	0,63	0,33	1,31	0,57	-0,07	0,35
Rhisotrogus aestivus	0,26	0,51	0,31	0,41	0,13	0,02
Athous haemorrhoidalis	0,42	0,28	0,77	0,40	-0,20	0,27
Staphilinidae sp. sp.	0,32	0,30	0,38	0,32	-0,14	0,27
Aporrectodea calliginosa	0,21	0,40	0,08	0,29	0,24	-0,35
Geophilus proximus	0,21	0,33	0,08	0,25	0,17	-0,11
Cardiophorus rufipes	0,32	0,23	0,15	0,24	-0,12	-0,14
Carabidae sp. sp.	0,16	0,30	0,08	0,23	-0,07	0,02
Megaphyllum sjaelandicum	0,16	0,28	0,08	0,21	0,22	-0,06
Helodrilus antipae	0,16	0,19	0,31	0,20	0,21	-0,39
Cylindronotus brevicollis	0,26	0,14	0,08	0,16	-0,03	-0,16
Mollusca sp. sp.	0,05	0,21	0,08	0,15	-0,03	-0,35
Escarius retusidens	0,05	0,12	0,23	0,12	0,22	0,07
Stratiomyidae	0,05	0,16	0,00	0,11	0,03	-0,13
Прочие	0,53	0,37	0,31	0,40		
Всего	17,79	12,02	23,08	15,40		

В качестве математической процедуры построения новой координатной оси в локальном пространстве может быть применен инструментарий множественной регрессии, которая осуществляется по методу наименьших квадратов. При этом классификационная переменная у объектов анализируемого класса принимает значение «0» и значение «1» у объектов других классов. Более подробно с алгоритмом этой процедуры можно познакомиться в работе В.Дюка [2].

В программе Statistica 5.5 реализована процедура под названием «Классификационные дендрограммы» («Classification Trees»), которая позволяет значительно упростить генерацию вербальных классификационных правил. Классификационные дендрограммы — метод, позволяющий предсказывать принадлежность наблюдений или объектов к тому или иному классу категориальной зависимой переменной в зависимости от соответствующих значений одной или нескольких предикторных переменных. В ряде случаев классический дискри-

минантный анализ и анализ классификационных дендрограмм дают сходные результаты. В случае соблюдения условий корректности проведения дискриминантного анализа, предпочтительным следует считать именно его. Однако классификационные дендрограммы имеют гораздо меньшее число ограничений, могут обрабатывать не только континуальные данные, но дискретные. Точность идентификации объектов посредством анализа классификационных дендрограмм могут быть гораздо выше.

На рисунке 2 представлена классификационная дендрограмма, позволяющая сформулировать индикативные правила для идентификации пространственных группировок почвенной мезофауны.

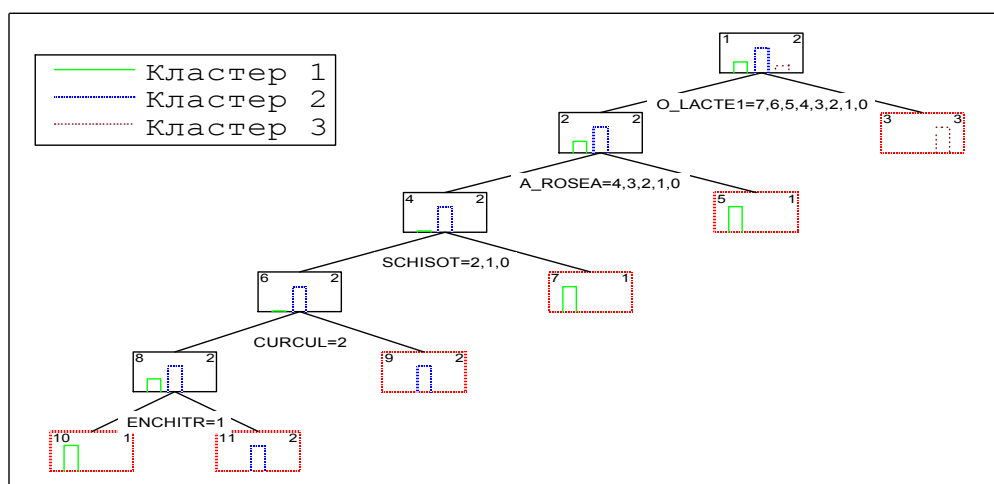


Рис. 2. Классификационная дендрограмма индикации пространственных группировок почвенной мезофауны

На схеме в прямоугольниках находятся гистограммы частот встречаемости трёх анализируемых группировок на каждом этапе их распознавания. В месте каждого дихотомического ветвления указано правило отнесения пробы к одному из двух последующих узлов более низкого уровня. Так, наиболее простым является вербальное правило, которое позволяет идентифицировать кластер 3 и оно гласит: «Если в пробе встречается не меньше 8 экземпляров дождевого червя *Octolasion lacteum*, то такую пробу следует отнести к кластеру 3». К кластеру 1 пробу можно отнести в том случае, если в ней встречено не менее 5 экземпляров дождевого червя *Aporrectodea rosea*, либо не менее 3 экземпляров многосвяза *Schisoturanus dmitriewi*, либо плотность личинок долгоносиков не меньше 2 экземпляров на одну пробу, либо встречается в пробе не менее одного экземпляра червей энхитреид. В противных случаях пробу следует отнести к кластеру 2.

В конечных звеньях классификационной дендрограммы находятся столбцы, соответствующие только одному типу кластеров, что свидетельствует об их полной идентификации с помощью указанных индикативных правил. В отличие от дискриминантного анализа, для идентификации конкретного экологического объекта при применении классификационной дендрограммы не тре-

буется вычисление достаточно сложных аналитических функций, а достаточно применить простые вербальные правила. Точность такой процедуры не ниже, чем при дискриминантном анализе, а может быть даже выше. Так, при использовании для дискриминантного анализа только 5 видов, указанных на рис. 2, точность индикации составляет 97%, тогда как при анализе классификационной дендрограммы точность составляет 100%. Кроме того, точность результатов индикации гораздо меньше зависит от точности определения плотности почвенных беспозвоночных, так в формулировках индикативных правил фигурирует указание порядка величины (больше или меньше некоторого порогового уровня), а не точное его значение.

Библиографические ссылки и примечания

- 1 Айвазян С. А., Бежаева З. И., Староверов О. В. Классификация многомерных наблюдений. М.: Статистика, 1974
- 2 Дюк В. Обработка данных на ПК в примерах. С.-Пб. 1997. 240 с.
- 3 Миркин Б. Г. Анализ качественных признаков и структур. М.: Статистика, 1980