

УДК 681.3.01+519.22+581.55
DOI: 10.36305/2712-7788-2020-1-154-45-55

КОЛИЧЕСТВЕННЫЕ МЕТОДЫ КЛАССИФИКАЦИИ РАСТИТЕЛЬНОСТИ: СОСТОЯНИЕ, ПРОБЛЕМЫ, ОПТИМИЗАЦИЯ

Геннадий Самуилович Розенберг

Институт экологии Волжского бассейна РАН
445003, Самарская область, г. Тольятти, ул. Комзина, 10
E-mail: genarosenberg@yandex.ru

Обсуждаются основные этапы создания автоматических процедур классификации растительности (индексы сопряженности, коэффициенты сходства, простейшие алгоритмы автоматической классификации – построение дендрограмм, дендритов, корреляционных плеяд и пр.). Редукция числа признаков (видов) рассматривается как первое условие оптимизации классификационной процедуры. Обсуждаются результаты эксперимента по редукции видов при автоматической классификации 50 описаний чиевников (формация *Achnatherum splendens* (Trin.) Nevski) в пойме р. Туул в Монголии. Сделан вывод о том, что новые успехи в автоматической классификации растительности следует ожидать не в направлении разработки каких-то новых методов, а в выдвижении новых представлений о структуре и характере динамики растительных сообществ (смена парадигм).

Ключевые слова: растительность; методы классификации; индексы сопряженности; коэффициенты сходства; редукция видов; оптимизация

Его (Аристотеля. – Г.Р.) величайшим и в то же время чреватым наиболее опасными последствиями вкладом в науку была идея классификации, которая проходит через все его работы <...> Аристотель ввел или, по крайней мере, кодифицировал способ классификации предметов, основанный на сходстве и различии...

Джон Десмонд Бернал [1, с. 117].

Введение

Любая отрасль знаний не может обойтись без упорядочения изучаемых объектов, их сортировки по тем или иным признакам, т. е. их классификации. Классификация является необходимым этапом любого исследования, так как без классификации, хотя бы предварительной, с одной стороны, трудно ориентироваться в материале и правильно планировать работу, а с другой – классификация очень важна при завершении исследования, когда подводятся итоги сделанного. Классифицирование – одна из первых ступеней познавательной деятельности; именно этот факт накладывает свой отпечаток на весь дальнейших ход и результативность предпринимаемых исследований.

Классификация (от лат. *classis* – разряд, класс и *facio* – делаю, раскладываю) – разбиение множества (класса) объектов на подмножества (подклассы) по определенным признакам. В научной классификации свойства объекта поставлены в функциональную связь с его положением в определенной системе. Различают искусственную и естественную классификации: в отличие от искусственной (в её основе, как правило, лежат не существенные для систематизации предметов признаки; например, расположение книг по алфавиту или по цвету корешков...), в естественной классификации по максимальному количеству существенных признаков объекта определяется его положение в системе (например, периодическая система элементов).

В качестве оригинального примера искусственной классификации аргентинский писатель Хорхе Луис Борхес (Borges, 1952) предлагал такую «классификацию животных», приписываемую им «одной китайской энциклопедии под названием "Небесная империя благодетельных знаний" (исп. "Emporio celestial de conocimientos benévolos". – Г.Р.). На её древних страницах написано, что животные делятся на:

- a) принадлежащих Императору (исп. *pertenecientes al Emperador*),
- b) набальзамированных (*embalsamados*),
- c) прирученных (*amaestrados*),
- d) сосунков (*lechones*),
- e) сирен (*sirenas*),
- f) сказочных (*fabulosos*),
- g) отдельных собак (*perros sueltos*),
- h) включенных в эту классификацию (*incluidos en esta calcificación*),
- i) бегающих как сумасшедшие (*que se agitan como locos*),
- j) бесчисленных (*innumerables*),
- k) нарисованных тончайшей кистью из верблюжьей шерсти (*dibujados con un pincel finísimo de pelo de camello*),
- l) прочих (*etcétera*),
- m) разбивших цветочную вазу (*que acaban de romper el jarrón*),
- n) похожих издали на мух (*que de lejos parecen moscas*)».

Составленный 40 лет тому назад обзор количественных методов классификации растительности (Миркин, Розенберг, 1979) содержал чуть более 20 алгоритмов. За прошедшие годы в этом направлении наметился заметный прогресс, что связано с «переходом на идеологию» обработки баз фитоценотических данных, таких как «Turbo (Veg)» (Hennekens, 1996); в 1960-80-х гг. под эгидой Р. Уиттекера (Уиттекер, 1980) стало развиваться математическое обеспечение количественных ординационно-классификационных методов анализа растительности. Прежде всего это пакеты программ М. Хилла – TWINSPAN (TWO-way INdicator SPecies ANalysis (Hill, 1979)) и DECORANA (DEtrended CORrespondence ANAlysis); X. Гоча (ORDIFLEX [a FLEXible computer program for four ORDInation techniques: weighted averages, polar ordination, principal component analysis, and reciprocal averaging], COMPCLUS [COMPosite CLUStering] (Gauch, 1977; Gauch, 1982) и др.; имеются и отечественные разработки – «Фитоценолог» (Голуб и др., 1995) и др. Некоторые новые методы количественной классификации можно найти в наших обзора (Уиттекер, 1980; Розенберг, 2007; Розенберг, 2018).

Показатели сопряженности и сходства

Прежде чем говорить о методах количественной классификации, следует несколько слов сказать об определении сходства. «Растительные сообщества (как, впрочем, и экосистемы в целом. – Г.Р.) относятся к числу систем, характеризующихся довольно слабой целостностью <...> В растительном сообществе мы, как правило, встречаемся не с функциональными зависимостями, а со стохастическими» (Васильевич, 1969). А, как известно, стохастическая (статистическая, вероятностная, корреляционная) зависимость между переменными величинами имеет место тогда, когда каждому значению зависимой переменной соответствует не какое-то определенное, а множество значений другой переменной, причем сказать заранее, какое именно значение примет зависимая величина, нельзя. Чаще всего, это объясняется тем, что все факторы, от которых зависит интересующая нас величина, действуют взаимосвязано. В зависимости от того, насколько оптимально сочетаются эти факторы, будет неодинаковой степень воздействия каждого из них на величину результативного показателя. Взаимосвязь между исследуемыми факторами и

результативным показателем проявится, если взять для исследования большое количество наблюдений (объектов) и сравнить их значения. Тогда в соответствии с законом больших чисел влияние других факторов на результативный показатель сглаживается, нейтрализуется, что позволяет установить связь между изучаемыми явлениями. Таким образом, стохастическая связь – это неполная, вероятностная зависимость между показателями, которая проявляется только в массе наблюдений.

Число индексов (коэффициентов) взаимосвязи видов и показателей сходства сообществ весьма велико, что свидетельствует об успешном развитии в последние 40-50 лет «экологической индексологии». Правда, большинство из этих показателей так и остаются на уровне «хорошо что-нибудь с чем-нибудь сложить и на что-то поделить...».

Ю.А. Шрейдер строго и подробно определил и понятие «расстояние» (Шрейдер, 1963), и понятие «сходство» (Шрейдер, 1971). Но вот, что интересно. В книге "Равенство, сходство, порядок" (Шрейдер, 1971) он подробно обсуждает логико-математические аспекты таких понятий, как «отношение», «одинаковость», «эквивалентность», «сходство», «толерантность», «упорядоченность», но нигде не берет на себя смелость предложить, например, конкретную меру сходства. Это демонстрирует, своего рода, «трезвость» при выборе того или иного показателя – каждый из них обладает своей областью применимости и не существует единого и пригодного для всех случаев индекса.

Методы автоматической классификации

«Классификация состоит в объединении насаждений в классы, члены каждого из которых имеют некоторое количество общих характеристик, благодаря чему они оказываются ограниченными от членов других классов. Классификация растительности – не новая область исследований, хотя до последнего времени она базировалась на данных, которые были в значительной степени субъективными» (Грейг-Смит, 1967). «После того как тем или иным методом определено сходство каждой пары площадок, можно приступить к выполнению следующего этапа работы: выделению групп сходных описаний» (Васильевич, 1969). Но не будем забывать, что «сообщество – это условности, объединенные в многомерный континуум <...> При создании классификаций, используют *прагматический подход*. Любая классификация рассматривается как рабочее приближение, и если её уровень достаточен для практического использования, то дальнейшее совершенствование классификации нецелесообразно» (Миркин и др., 2000). Эти три цитаты классиков отечественной количественной (статистической) фитоценологии, фактически, задают границы теоретического базиса для методов автоматической классификации.

Один из основных способов первичного анализа и визуализации корреляционной матрицы (по коэффициентам сопряженности между видами или сходства между описаниями) сводится к построению специальных графиков – *дендрограмм* или *дендритов* (графов «максимального корреляционного пути»). Наиболее простыми способами построения таких графиков являются *метод «корреляционных плеяд»* П.В. Терентьева (Терентьев, 1959) и, так называемая, *«кроцлавская таксономия»* (Florek et al., 1951). И хотя этим методам 60-70 лет они востребованы в фитоценологических исследованиях (простота алгоритмов и хорошая интерпретируемость результатов).

Алгоритм Терентьева осуществляет выделение сильно связанных групп признаков («корреляционных плеяд») и сводится к следующему. Задается пороговое значение коэффициента корреляции r_0 , с помощью которого производится построение

срезов корреляционного цилиндра, из которых формируется последовательность подграфов, принимаемых в качестве «плеяд». Узлами этих подграфов являются все рассматриваемые признаки, а ребрами – корреляционные связи по абсолютной величине больше r_0 . При последовательном уменьшении критического уровня, количество ребер увеличивается, плеяды становятся крупнее и начинают сливаться друг с другом. Окончательно выбирается порог r_0 , скорее отвечающий эстетическим вкусам исследователя, чем каким-то формальным правилам.

Результатом вроцлавской таксономии является полный незамкнутый корреляционный путь, который можно отобразить в виде оптимального дерева – дендрита. Он представляет собой графическую структуру, состоящую из m вершин, соединенных между собой ($m - 1$) ребрами так, что каждая вершина соединена хотя бы с одной другой. Если длину каждого ребра ассоциировать с величиной коэффициента корреляции r_{ij} между вершинами i и j , то оптимальный дендрит имеет максимальную сумму длин соединяющих отрезков из всех возможных. Если принять во внимание, что корреляционная мера по своему смыслу обратна мере дистанции, то граф максимального корреляционного пути идентичен «минимальному дендриту», т. е. дереву минимальной протяженности (minimum spanning tree). При желании, задавшись пороговым значением r_0 , можно полученный полный граф максимального корреляционного пути разбить на подграфы (плеяды), проводя разрыв между теми признаками, которые вошли в первоначальный граф со значением сопряженности меньшим r_0 . Алгоритм вроцлавской таксономии полностью соответствует известному в кластерном анализе методу ближайшей связи (он же – метод «одиночного сцепления»).

Эти два простейших алгоритма автоматической классификации, как бы предваряют, обширный класс методов, объединенных в понятие «*кластерный анализ*».

Пусть имеется матрица наблюдений размерностью $n \times m$, строки i которой соответствуют геоботаническим описаниям ($i = 1, 2, \dots, n$), а столбцы j содержат конкретные показатели встречаемости или проективного покрытия видов ($j = 1, 2, \dots, m$), полученные в точке наблюдения i и выраженные в шкалах измерений произвольного характера (качественных или количественных). Если эти данные понимать, как точки в признаковом пространстве, то задача кластерного анализа формулируется как выделение «сгущений точек (описаний)» и разбиение исходной совокупности на однородные подмножества объектов. Кластерный анализ можно рассматривать также как метод редукции (сжатия) некоторого множества данных в более компактную классификацию объектов.

Кластер определяется, как совокупность точек, лежащих на расстоянии не больше, чем r от некоторого «центра тяжести» в m -мерном пространстве видов (внутри гиперсферы радиуса r или гиперкуба со сторонами $2r$). При этом следует помнить, что введение меры расстояния в m -мерном пространстве является естественным приемом квантификации свойства схожести объектов: чем ближе между собой объекты в данной метрике, тем они более сходны и наоборот. Без этого само понятие «кластер» во многом теряет смысл, поэтому алгоритмы кластерного анализа часто формулируют в терминах дистанций.

Был предпринят ряд попыток разработать аксиоматический подход к введению метрических мер, согласно которым, например, расстоянием D называется двухместная действительная функция $D(x_1, x_2)$, обладающая следующими свойствами:

- $D(x_1, x_2) \geq 0$ – неотрицательная определенность расстояния (хотя тот же коэффициент корреляции Пирсона принимает и отрицательные значения);
- $D(x_1, x_2) = 0$ тогда и только тогда, когда $x_1 = x_2$ – неразличимость тождественных объектов;

- $D(x_1, x_2) = D(x_2, x_1)$ – симметричность расстояния (хотя есть примеры и несимметричных мер);
- $D(x_1, x_2) + D(x_2, x_3) \geq D(x_1, x_3)$ – неравенство треугольника (длина любой стороны треугольника не больше суммы длин двух оставшихся).

Более конкретная математическая формулировка не имеет однозначного смысла, поскольку разные субъекты вкладывают в эту аксиоматику неодинаковое содержание.

В литературе описывается множество различных методов кластеризации, основанных на использовании матриц сходства, оценивании функций плотности статистического распределения, эвристических алгоритмах перебора, идеях математического программирования и др. (Грейг-Смит, 1967; Василевич, 1969; Миркин, Розенберг, 1979; Шитиков идр., 2005; Розенберг, 2007; Розенберг, 2013; Розенберг, 2018). Большая часть этих алгоритмов, при всей их несходстве, методически основаны на одной предпосылке – *гипотезе компактности* (в используемом пространстве признаков измерения (видов), принадлежащие одному и тому же классу объекты (описания), близки между собой, а измерения, принадлежащие разным классам хорошо разделимы друг от друга).

Ограниченный объем статьи не позволяет подробно описать те или иные алгоритмы кластеризации; в этой ситуации ограничусь лишь перечислением основных наиболее распространенных групп эвристических методов кластеризации (Шитиков и др., 2005):

- методы, основанные на *иерархической агломеративной процедуре* (от лат. *agglomerero* – присоединяю, накапливаю);
- *параметрические методы анализа многомерных наблюдений*, принадлежащих к разным классам (два множества будут разделяться тем лучше, чем больше расстояние между их центрами; часто используется выборочное *расстояние Махалонобиса*; к основным преимуществам этой метрики следует отнести учет коррелированности признаков между собой и инвариантность относительно невырожденных линейных преобразований, что избавляет от необходимости нормировки исходной матрицы наблюдений);
- разграничение двух совокупностей, основанное на некоторых *эвристических принципах* (примером может служить метод, предложенный В.Н. Котовым и Н.Г. Терентьевой (Котов, Терентьева, 1989) и использующий понятие «*биоквант*»¹).
- снижение размерности классифицируемой матрицы «описания – виды» путем *факторного анализа*;
- *методы многомерного шкалирования* (можно рассматривать как альтернативу факторному анализу, когда, кроме корреляционных матриц, в качестве исходных данных можно использовать произвольный тип матрицы сходства объектов; многомерное шкалирование – это не сколько определенная математическая процедура, а скорее способ наиболее эффективного размещения объектов, приближенно сохраняющий расстояния между ними в новом пространстве признаков, размерность которого существенно меньше исходного).

Таким образом, можно предложить следующую методику классификации групп многомерных геоботанических описаний (Шитиков и др., 2005):

- формирование матрицы исходных данных;
- снижение признакового пространства путем выделения главных компонент;
- дисперсионный анализ факторов;

¹ Биоквантами называются подмножества объектов (описаний), для которых в заданном метрическом пространстве (видов) все внутргрупповые расстояния меньше всех межгрупповых.

- формирование матрицы расстояний;
- многомерное шкалирование.

Оптимизация процесса автоматической классификации растительности

Редукция числа признаков (видов) является первым условием оптимизации классификационной процедуры. Этот вопрос достаточно подробно обсуждался ранее (Василевич, 1969; Розенберг, 1975; Розенберг, Гареева, 1978; Westhoff, van der Maarel, 1978). Здесь, в качестве примера, приведу результат эксперимента по редукции видов при автоматической классификации 50 описаний чиевников в пойме р. Туул (Тола) в Монголии (формация *Achnatherum splendens* (Trin.) Nevski)². Этот эксперимент направлен на обоснование эмпирически предлагаемых границ редукции видов для оптимизации процедуры классификации растительности по Браун-Бланке (Ellenberg, 1956): выше 10% и ниже 70% встречаемости. Количественная проверка обоснованности этих порогов осложняется тем, что выполнение ряда параллельных классификаций в пространстве видов разной степени редукции, – задача трудоемкая; в данной работе анализ фитоценологических таблиц был заменен автоматической классификацией (метод дендрограммы) с использованием индекса сходства Съёренсена (Грейг Смит, 1967; Василевич, 1969; Миркин, Розенберг, 1979; Шитиков идр., 2005; Розенберг, 2007; Розенберг, 2013; Розенберг, 2018; Florek et al., 1951). Полным перебором всевозможных сочетаний «редукции сверху» (сокращение видов максимальной встречаемости) и «редукции снизу» (исключение из анализа редких видов) с шагом постоянства 0,1, доводимым до абсурда (классификация по одному недоминантному виду *Potentilla dealbata* Bge.), было выполнено 36 классификаций (за основу было взято 50 описаний). Таким образом, задача свелась к сравнению результатов классификаций одних и тех же описаний, но в разных «пространствах» видов, и выбора, фактически, «лучшей классификации».

Результаты выполненных классификаций сравнивались с использованием «меры диссонанса» (Розенберг, 1975; Розенберг, 1978; Розенберг, 2013):

$$D(R, S) = \frac{1}{N(N-1)} \left[\sum_{i=1}^n |R_i|^2 + \sum_{j=1}^m |S_j|^2 - 2 \sum_{i=1}^n \sum_{j=1}^m |R_i \cap S_j|^2 \right],$$

где $|R_i|$ – количество объектов в группе i ($i = 1, 2, \dots, n$) классификации R , $|S_j|$ – количество объектов в группе j ($j = 1, 2, \dots, m$) классификации S , $|R_i \cap S_j|$ – количество общих объектов для i -ой и j -ой групп этих классификаций; N – общее число объектов-описаний (в нашем случае, $N = 50$). Для $D(R, S)$ определена величина «ошибки» (Маамяги, 1977), что позволяет оценить достоверность различия мер диссонанса. Замечу, что $D(R, S)$ меняется от 0 до 1, причем $D(R, R) = 0$.

В качестве «эталонной» классификации (Розенберг, 1975), с которой первоначально проводилось сравнение всех 36 синтезированных классификаций, была принята классификация по полному списку видов. Результаты сравнения приведены в табл. 1. Самой близкой к «эталонной» классификации является группировка описаний при сокращении видов с постоянством выше 0,7 и ниже 0,1 (величина $D(R, S)$ в этом случае достоверно отличается от 1), что подтверждает эмпирическую оценку Х. Элленберга (Ellenberg, 1956). Для подтверждения правильности выбора данных

² Описания выполнены геоботаниками под руководством Б.М. Миркина в 1973 г.; достаточно подробная характеристика объекта дана в работах 7, 8]. Я благодарен коллегам из лаборатории геоботаники и охраны растительности Уфимского института биологии УФИЦ РАН за предоставленную возможность вернуться к этим материалам и провести их обработку.

уровней редукции было проведено сравнение тех же результатов классификации с новым «эталоном» ($x=0,1$ и $y=0,7$); эти результаты приведены в табл. 2.

Результаты этой эталонной классификации достоверно отличаются от всех остальных (кроме классификации по полному списку). Это позволяет сделать вывод о том, что границы редукции видов $x=0,1$ и $y=0,7$ являются приемлемыми оптимизационными порогами. Сходство меры диссонанса $D(R,S)$ при $y=0,7$ и $y=0,6$ свидетельствует о том, что в списке видов (пространстве видов) отсутствовали виды с постоянством от 0,6 до 0,7.

Таблица 1
Значение величины $1 - D(R,S)$ при «редукции снизу» (x) и «редукции сверху» (y) для выделенных групп («эталонная» классификация $y=1, x=0$)

Table 1
The value of $1 - D(R,S)$ in the case of «reduction from below» (x) and «reduction from above» (y) for selected groups («reference» classification: $y = 1, x = 0$)

y	x					
	0	0,1	0,2	0,3	0,4	0,5
1,0	1,00	0,94	0,83	0,85	0,37	0,13
0,9	0,91	0,92	0,87	0,73	0,26	0
0,8	0,91	0,92	0,87	0,73	0,26	0
0,7	0,89	0,98	0,71	0,44	0	0
0,6	0,89	0,98	0,71	0,44	0	0
0,5	0,85	0,80	0,42	0,09	0	0

Примечание. Нули в таблице означают, что классификации не были получены, т. е. все описания объединились в один класс.

Таблица 2
Значение величины $1 - D(R,S)$ при «редукции снизу» (x) и «редукции сверху» (y) для выделенных групп («эталонная» классификация $y=0,7, x=0,1$)

Table 2
The value of $1 - D(R,S)$ in the case of «reduction from below» (x) and «reduction from above» (y) for selected groups («reference» classification: $y = 0,7, x = 0,1$)

y	x					
	0	0,1	0,2	0,3	0,4	0,5
1,0	0,98	0,90	0,92	0,86	0,37	0,13
0,9	0,91	0,93	0,82	0,80	0,23	0
0,8	0,91	0,93	0,82	0,80	0,23	0
0,7	0,89	1,00	0,87	0,50	0	0
0,6	0,89	1,00	0,87	0,50	0	0
0,5	0,87	0,89	0,60	0,16	0	0

В табл. 3 приведена характеристика классификационных групп, полученных из 50 описаний чиевников по оптимальным уровням редукции видов. В группы второго ранга (метод построения дендрограмм) объединилось 48 (96%) описаний. Числовые значения в табл. 3 представляют собой *фитоценотический коэффициент* (произведение постоянства вид в группе на его среднее проективное покрытие, выраженное в 5-балльной шкале). Ранг выделенных единиц, как мне представляется, различен. Правда, это не противоречит формальности самой процедуры: при синтаксономическом анализе (подход Браун-Бланке) сходство оценивается уже не чисто количественно, а учитывается существенность экологической информации о различиях фитоценонов по характеру дифференцирующих видов. Первая единица – чиевники с луком *Allium polyrrhizum*, т. е. наиболее сухой вариант солончаковых сообществ формации. Вторая единица – чиевники седовые (*Suaeda corniculata*) – крайняя степень засоления в условиях лугового увлажнения. Три последующие единицы могут быть объединены в одну ассоциацию чиевников вострецовых (*Elymus chinensis*), причем, единицы 3–4–5 представляют собой ряд уменьшения засоления и увлажнения. В целом, три выделенных ассоциации вполне соответствуют типам чиевников, которые 65 лет тому назад описал А.А. Юннатов (Юннатов, 1954).

Фитоценотические коэффициенты в выделенных группах описаний
(метод дендрограммы, второй ранг)

Таблица 3

Phytoecnotic coefficients of the selected groups of descriptions
(dendrogram method, second rank)

Table 3

Виды Species	Номера групп второго ранга Numbers of second rank groups				
	1	2	3	4	5
<i>Achnatherum splendens</i> (Trin.) Nevski	2,571	5,000	2,913	4,000	3,147
<i>Elymus chinensis</i> (Trin.) Keng	0,571		1,097	1,167	1,286
<i>Taraxacum</i> sp.		1,000	0,371	0,449	0,510
<i>Potentilla dealbata</i> Bge.			0,371	0,250	1,000
<i>Chenopodium album</i> L.	0,326		0,008	1,000	1,571
<i>Potentilla bifurca</i> L.	0,326		0,004		0,408
<i>Carex duriuscula</i> C. A. M.	1,143		0,023		0,123
<i>Puccinellia macranthera</i> Krecz.			0,454	0,250	0,082
<i>Hordeum brevisubulatum</i> (Trin.) Link			0,579		
<i>Polygonum nodosum</i> Pers.		1,000	0,425		0,020
<i>Iris lactea</i> Pall.	0,082		0,768	0,167	
<i>Saussurea amara</i> DC.			0,397	1,389	0,184
<i>Carex enervis</i> C. A. M.			0,212		
<i>Halerpestes salsuginosa</i> (Pall.) Greene			0,765		
<i>Artemisia laciniata</i> Willd.			0,030	0,028	0,082
<i>Potentilla anserine</i> L.			0,208		
<i>Poa subfastigiata</i> Trin.			0,047		
<i>Suaeda corniculata</i> (C. A. M.) Bge.		2,400	0,008		
<i>Puccinellia tenuiflora</i> (Griseb.) Scribn. et Merr.		2,400			
<i>Allium polystachys</i> Turcz.	2,000		0,002		
<i>Artemisia adamsii</i> Bess.	0,020		0,047		
<i>Limonium flexuosum</i> (L.) Ktze.			0,008	0,694	
<i>Chenopodium acuminatum</i> Willd.	0,020			0,111	0,082
<i>Allium odoratum</i> L.	0,184			0,250	
<i>Plantago depressa</i> Willd.			0,008	0,694	0,020
<i>Astragalus adsurgens</i> Pall.			0,008	0,111	0,082
Количество описаний в каждой группе	7	5	23	6	7

Выводы

Завершая это эссе, замечу, что при выполнении кластерного анализа исследователь находится в достаточно тяжелом положении, будучи поставлен перед необходимостью выбора не только комплекта исходных данных, но также метрики расстояния и алгоритма объединения. Например, если для тех же *n* геоботанических описаний можно использовать не менее 5 общепотребимых формул для построения матрицы сходства и не менее 5 широко распространенных методов построения иерархической классификации, то исследователь получит 25 возможных вариантов разбиений, т.е. «деревьев», в разной степени отличающихся друг от друга. В итоге неопределенность исходных данных подменяется другой, еще более туманной, – неопределенностью результатов классификаций. Формализации процедуры выбора того или иного показателя сходства или метода автоматической классификации *не существует* – имеются лишь некоторые, в разной степени логически размытые рекомендации по выбору технологии расчетов.

Приходится констатировать, что за последние полвека развития статистических методов в фитоценологии успехов набирается значительно меньше, чем за весь более чем столетний период становления этих методов. Наверное, это следует объяснить тем, что развитие количественных методов пошло в направлении разработки математических

моделей и создания разного рода баз знаний и данных. Этот процесс не простой и ожидать здесь быстрых успехов не приходится. И здесь, как нельзя, кстати, подходят слова В.И. Василевича (Василевич, 1969), которыми он 50 лет тому назад заканчивал свою монографию и которые не потеряли своей актуальности и сегодня: «Не следует думать, что тот математический аппарат, который в настоящее время используют геоботаники, дает наилучшее решение. С одной стороны, его необходимо углублять и дополнять, совершенствовать уже существующие методы, а с другой стороны, необходимо искать принципиально новые пути решения проблем. Авторитет старых методов так же опасен, как и авторитет старых теорий».

И еще один момент, на котором хотелось бы заострить внимание. Новые успехи фитоценологии следует ожидать *не в направлении разработки каких-то новых методов анализа растительности, а в выдвижении новых представлений о структуре и характере динамики растительных сообществ*; иными словами, должна произойти очередная смена парадигм (Розенберг, 1997).

Литература / References

- Бернал Дж.* Наука в истории общества. М.: Изд-во ин. лит-ры, 1956. 735 с.
 [Bernal J. Science in the history of society. Moscow: Publ. House Foreign Lit., 1956. 735 p.]
- Василевич В.И.* Статистические методы в геоботанике. Л.: Наука, 1969, 232 с.
 [Vasilevich V.I. Statistical methods in geobotany. Leningrad: Nauka, 1969. 232 p.]
- Голуб В.Б., Халеев Е.А., Рухленко И.А.* Пакет программ для обработки геоботанических данных «Фитоценолог» // II совещание «Компьютерные базы данных в ботанических исследованиях». Тез. докл. СПб.: Ботанический институт им. В.Л. Комарова РАН, 1995, с. 13-14.
 [Golub V.B., Khaleev E.A., Ruhlenko I.A. Software package for processing geobotanical data "Phytocenologist" // II Meeting "Computer Databases in Botanical Research": thesis report SPb.: Botanical Institute V.L. Komarov RAS, 1995. P. 13-14.]
- Грейг-Смит П.* Количественная экология растений. М.: Мир, 1967. 359 с.
 [Greig-Smith P. Quantitative ecology of plants. Moscow: Mir, 1967. 359 p.]
- Котов В.Н., Терентьева Н.Г.* Разграничение двух совокупностей биологических объектов на примере планкtonных видов рода *Anabaena* Bory. Препринт. Киев: Институт ботаники им. Н.Г. Холодного АН УкрССР, 1989. 52 с.
 [Kotov V.N., Terentyeva N.G. Differentiation of two sets of biological objects on the example of plankton species of the genus *Anabaena* Bory. Preprint. Kiev: Institute of Botany Sciences Academy of Ukrainian SSR, 1989. 52 p.]
- Маамяги А.В.* Экспертная классификация объектов. Распределение расстояния между классификациями некомпетентных экспертов. Известия АН ЭССР. 1977. Т. 26. Общественные науки, № 2. С. 122-131.
 [Maamagi A.V. Expert classification of objects. The distribution of the distance between the classifications of incompetent experts // News of the Academy of Sciences of the ESSR. 1977. V. 26. Social Sciences, № 2. P. 122-131.]
- Миркин Б.М.* О принципах типизации хорологических единиц растительного покрова речных пойм // Ботанический журнал. 1975. Т. 60, № 3. С. 313-321.
 [Mirkin B.M. On the principles of typing chorological units of plant cover of river floodplaces // Botanical Journal. 1975. V. 60, № 3. P. 313-321.]
- Миркин Б.М., Манибазар Н., Гареева Л.М., Карапов Р.Ш., Максимович С.В., Мухаметшина В.С., Наумова Л.Г.* Растительность речных пойм Монгольской Народной Республики: Опыт синтаксономического анализа с использованием флористических критериев. Л.: Наука, 1980. 280 с.

[*Mirkin B.M., Manibazar N., Gareeva L.M., Kashapov R.Sh., Maksimovich S.V., Mukhametshina V.S., Naumova L.G.* Vegetation of river floodplains of the Mongolian People's Republic: experience of syntaxonomic analysis using floristic criteria. Leningrad: Nauka, 1980. 280 p.]

Миркин Б.М., Наумова Л.Г., Соломец А.И. Современная наука о растительности. Уфа: Логос, 2000. 264 с.

[*Mirkin B.M., Naumova L.G., Solomeshch A.I.* Modern science of vegetation. Ufa: Logos, 2000. 264 p.]

Миркин Б.М., Розенберг Г.С. Количествоные методы классификации, ординации и геоботанической индикации // Итоги науки и техники. Ботаника. Т. 3. М.: ВИНИТИ, 1979. С. 71-137.

[*Mirkin B.M., Rozenberg G.S.* Quantitative methods of classification, ordination and geobotanical indication // Itogi nauki i tekhniki. Botany. V. 3. M.: VINITI, 1979. P. 71-137.]

Розенберг Г.С. О сравнении различных методов автоматической классификации // Автоматика и телемеханика. 1975. № 9. С. 145-148.

[*Rozenberg G.S.* On the comparison of various methods of automatic classification // Automation and Remote Control. 1975. № 9. P. 145-148.]

Розенберг Г.С. Статистические методы в фитоценологии на рубеже тысячелетий (к 50-летию выхода монографии П. Грейг-Смита) // Актуальные проблемы геоботаники: III-я Всероссийская школа-конференция. Лекции. Петрозаводск: Карел. НЦ РАН, 2007. С. 72-116.

[*Rozenberg G.S.* Statistical methods in phytocenology at the turn of the millenniums (on the 50th anniversary of the publication of the monograph by P. Greig-Smith) // Actual Problems of Geobotany: III-rd All-Russian Conference School. Lectures. Petrozavodsk: Karel. SC of RAS, 2007. P. 72-116.]

Розенберг Г.С. Введение в теоретическую экологию / В 2-х т.; Изд. 2-е, исправленное и дополненное. Тольятти: Кассандра, 2013. Т. 1. 565 с. Т. 2. 445 с. URL:<http://www.ievbras.ru/books/books.html> (дата обращения: 03.03.2020)

[*Rozenberg G.S.* Introduction to theoretical ecology / Ed. 2nd, revised and enlarged. Togliatti: Cassandra, 2013. V. 1. 565 p. V. 2. 445 p. [In Russian] Available at: <http://www.ievbras.ru/books/books.html>. (access date: 03.03.2020)]

Розенберг Г.С. Количествоные методы фитоценологии в работах Б.М. Миркина и их современное состояние // Фиторазнообразие Восточной Европы. 2018. Т. 12, № 3. С. 31-54. DOI: 10.24411/2072-8816-2018-10024.

[*Rozenberg G.S.* Quantitative methods of phytocenology in the works of B.M. Mirkin and their current state // Phytodiversity of Eastern Europe. 2018. V. 12, No. 3. P. 31-54. DOI: 10.24411 / 2072-8816-2018-10024.]

Розенберг Г.С., Гареева Л.М. О количественном обосновании допустимых пределов редукции видов при классификации с использованием флористических критериев // Ботанический журнал. 1978. Т. 63, № 6. С. 864-867.

[*Rozenberg G.S., Gareeva L.M.* On the quantitative substantiation of the allowable limits of species reduction in the classification using floristic criteria // Botanical Journal. 1978. V. 63, No. 6. P. 864-867.]

Розенберг Г.С., Смелянский И.Э. Экологический маятник (Смена парадигм в современной экологии) // Журнал общей биологии. 1997. Т. 58, № 4. С. 5-19.

[*Rozenberg G.S., Smelyansky I.E.* Ecological pendulum (Change of paradigms in modern ecology) // Journal of General Biology. 1997. V. 58, No. 4. P. 5-19.]

Терентьев П.В. Метод корреляционных плеяд // Вестник Ленинградского государственного университета. 1959. Т. 9, вып. 2. С. 137-144.

[Terentyev P.V. Method of correlation pleiad // Bulletin of the Leningrad State University. 1959. V. 9, No. 2. P. 137-144.]

Уиттакер Р. Сообщества и экосистемы. М.: Прогресс, 1980. 328 с.

[Whittaker R. Communities and ecosystems. Moscow: Progress, 1980. 328 p.]

Шитиков В.К., Розенберг Г.С., Зинченко Т.Д. Количественная гидроэкология: методы, критерии, решения: в 2-х кн. М.: Наука, 2005. Кн. 1. 281 с.; Кн. 2. 337 с.

[Shitikov V.K., Rozenberg G.S., Zinchenko T.D. Quantitative hydroecology: methods, criteria, solutions: in 2 books. Moscow: Nauka, 2005. Book 1. 281 p.; Book 2. 337 p.]

Шрейдер Ю.А. Что такое расстояние? М.: Физматгиз, 1963. 76 с.

[Schrader Yu.A. What is the distance? M.: Fizmatgiz, 1963. 76 p.]

Шрейдер Ю.А. Равенство, сходство, порядок. М.: Наука, 1971. 256 с.

[Schrader Yu.A. Equality, similarity, order. Moscow: Nauka, 1971. 256 p.]

Юнатов А.А. Кормовые растения пастбищ и сенокосов Монгольской Народной Республики. М.; Л.: Изд-во АН СССР, 1954. 352 с.

[Yunatov A.A. Fodder plants of pastures and hayfields of the Mongolian People's Republic. Moscow; Leningrad: Publ. Academy of Sciences of the USSR, 1954. 352 p.]

Borges J.-L. El idioma analítico de John Wilkins // Otras Inquisiciones. Buenos-Aires: Sur, 1952. P. 121-125.

Ellenberg H. Aufgaben und Methoden der Vegetationskunde // Einführung in die Phytologie. Ulmer; Stuttgart: Kluwer Acad. Publ., 1956. Bd. IV: Grundlagen der Vegetationsgliederung. T. 1. P. 136.

Florek K., Lukaszewicz J., Perkal J., Zubrzycki S. Taksonomia Wroclawska // Przeglad Antropoloji. 1951. V. 17. P. 193-211.

Gauch H.G. ORDIFLEX a flexible computer program for four ordination techniques: weighted averages, polar ordination, principal component analysis and reciprocal averaging: Release B. Ithaca: Cornell Univ. Press, 1977. 195 p.

Gauch H.G. Multivariate Analysis in Community Ecology. Cambridge: Cambridge Univ. Press, 1982. 314 p.

Hennekens S.M. TURBO (VEG). Software package for input, processing, and presentation of phytosociological data. Version 1.0. IBNDLO. Lancaster: Lancaster Univ. Press, 1996. 52 p.

Hill M.O. TWINSPAN a FORTRAN program for arranging multivariate data in an ordered two way table by classification of the individuals and the attributes. Ecology and Systematic. Ithaca: Cornell Univ. Press, 1979. 48 p.

Westhoff V., van der Maarel E. The Braun-Blanquet approach // Classification of Plant Communities. The Hague: Dr. W. Junk, 1978. P. 287-399.

Статья поступила в редакцию 07.04.2020 г.

Rozenberg G.S. Quantitative methods of classification of vegetation: condition, problems, optimization // Plant Biology and Horticulture: theory, innovation. 2020. № 1(154). P. 45-55.

The main stages of creating automatic vegetation classification procedures (contingency indices, similarity coefficients, simple algorithms for automatic classification — construction of dendograms, dendrites, correlation pleiades, etc.) are discussed. The reduction of the number of features (types) is considered as the first condition for the optimization of the classification procedure. The results of the experiment on the reduction of species with the automatic classification of 50 descriptions of *Achnatherum splendens* (Trin.) Nevski formation in the floodplain of the Tuul River in Mongolia are discussed. It is concluded that new successes in the automatic classification of vegetation should be expected not in the direction of developing some new methods, but in advancing new ideas about the structure and character of the dynamics of plant communities (paradigm shift).

Keywords: vegetation; classification methods; contingency indices; similarity coefficients; species reduction; optimization