УДК 510; 519.7; 519.237 ISSN 1995-5499
DOI: https://doi.org/10.17308/sait/1995-5499/2022/3/127-138
Поступила в редакцию 24.06.2022
Подписана в печать 30.09.2022

ФРАКТАЛЬНЫЕ АСПЕКТЫ МОДЕЛИРОВАНИЯ КЛАССИФИКАЦИОННОЙ ЗАДАЧИ

© 2022 Буховец А. Г.✉, Семин Е. А.

*Воронежский государственный аграрный университет имени Императора Петра I
ул. Мичурина, 1, 394087 Воронеж, Российская Федерация*

Аннотация. В центре внимание данной работы математическая модель классификационной задачи, в основе которой понятия теории фрактальных множеств и особенности моделирования ментальных процессов. В рамках такого подхода классификационную деятельность можно рассматривать как бы с двух точек зрения. С одной стороны — это работа с эмпирическим представлением объектов классификации, с другой — ментальное воспроизведение задачи построения классификационных разбиений. Реализация этих двух видов деятельности происходит в процессе построения некоторого фрактального множества, генерируемого посредством рандомизированной системы итерированных функций. Математическая модель классификационной задачи представляется в виде двух пространств: метрического, связанного с феноменологической составляющей, и ультраметрического, отражающего когнитивную сторону решаемой задачи. Как показывает анализ результатов решения задач классификации, фрактальные модели хорошо соответствуют требованиям, предъявляемым к постановке задачи и алгоритмическим особенностям её решения. Сама специфика постановки задачи отражается в необходимости учитывать такие характеристики как изолированность отдельных объектов, возможность устанавливать сходство/различие объектов, компактность пространства признаков и др. Эти свойства характерны для создаваемого в ходе решения ультраметрического пространства. Взаимосвязь этих двух пространств осуществляется посредством моделирования фрактальной структуры. В работе показано, как именно использование фрактального подхода в решении классификационных задач связано с построением ультраметрических пространств. Характерно, что эти ультраметрические пространства являются составной частью алгоритма решения самой задачи. Эта алгоритмическая составляющая часть решения классификационной задачи напрямую связывается с когнитивными процессами и интерпретируется как модель процессов, присущих умственной деятельности.
Ключевые слова: классификационная задача, рандомизированные системы итерированных функций, ультраметрические пространства, фрактальные множества, моделирование ментальных процессов.

ВВЕДЕНИЕ. О ФОРМИРОВАНИИ КЛАССИФИКАЦИОННЫХ СТРУКТУР

Моделирование классификационной задачи уже имеет свою историю. По-видимому, можно считать разумным сделанное предположение о том, «если существует некоторое реальное структурное членение мира, то ему в определённом приближении соответствуют и структура нашего знания» [1, С. 18] и окружающая нас действительность. Вследствие этого возникает потребность в создании таких математических структур, в которых воспроизводились бы компактные множества и при этом отсутствовала связность различных элементов системы. Подходы к построению классификационных схем [2, 3] традиционно опирались на многомерные связные метрические пространства, такие, например, как R^n с заданными в них мерами близости $\rho(x, y)$; $x, y \in R^n$. Эти меры, представленные в виде

✉ Буховец Алексей Георгиевич
e-mail: abuhovets@mail.ru

числового значения, характеризуют степень близости/сходства между произвольными точками этого пространства, которые выполняли роль классифицируемых объектов.

В философских экспликациях постановки классификационных задач можно на наш взгляд, различать два подхода, которые условно будем называть линиями Декарта и Ньютона. С точки зрения Декарта объяснять какие-либо эффекты следует исходя из взаимного расположения объектов: «…физика Декарта есть физика взаимных расположений, а не предположений относительно существований каких-то внутренних сил» [4, С. 162]. Отсюда его стремление создать метод, позволяющий фиксировать положения тел, который в окончательном виде привел к созданию координатной системы. Классификационная задача с этой точки зрения сводится к изучению структур многомерных данных, исследованию взаимного расположения объектов в признаковом пространстве [3, 5, С. 148]. Подход Ньютона был нацелен на описание изменений в (рас)положении тел, исследование их динамики, определение причин, вызывающих эти изменения, и, как следствие, расположение тел относительно друг друга. Пространство, как и время, Ньютон считал абсолютным понятием [6]. Образец такого подхода проявился в открытии закона всемирного тяготения, (правда, приписываемого некоторыми исследователями Гуку). Использование динамических систем — это продолжение и развитие этого направления, имеющего целью установление тех основ, которые формируют структурные особенности.

О необходимости создания подхода, позволяющего автоматически учитывать, как свойства алгоритмов классификации, так и свойства получаемых в результате их выполнения классов, давно уже было отмечено в [2, 5]. Тогда же было высказано предположение о необходимости разработки нового математического аппарата, в большей степени соответствующего отмеченным ниже особенностям классификационной задачи. В настоящей работе будет представлен подход к моделированию классификационной задачи, базирующийся на фрактальной теории, который соединяет в себе как динамические характеристики генерирования, так и геометрические аспекты данных. Этот подход, опирающийся на исследование свойств аттрактора РСИФ [7], можно рассматривать как синтез двух этих подходов ранее существовавших порознь. В его основу была положена идея перехода к моделированию задачи классификации в ультраметрических пространствах.

Мы считаем, что процессы построения классификаций, которые раньше неявно присутствовали в виде описания процедур и представлялись моделями некоторых когнитивных процессов, играют важную роль не только в узкопрактических прикладных вопросах. Изучение и моделирование таких процессов может служить составной частью общей теории ИИ.

1. ОСОБЕННОСТИ ПРЕДСТАВЛЕНИЯ ДАННЫХ И РЕЗУЛЬТАТОВ КЛАССИФИКАЦИОННЫХ ЗАДАЧ

Классификационная задача, как известно, в самом общем виде заключается в следующем. В имеющейся совокупности объектов выделить классы однородных в некотором смысле объектов, т.е. разделить заданное множество таким образом, чтобы объекты, отнесённые к одному классу, были более схожи между собой, чем объекты, принадлежащие различным классам (см., например, [2, С. 18, 5]. В рамках такой постановки задачи одним из основных подходов к построению классификации являются алгоритмы кластерного анализ, использующие геометрическое представление данных.

К особенностям классификационной задачи и получаемых в ходе её решения результатов, на наш взгляд, следует отнести следующие аспекты.

• Объекты в случае геометрического подхода, точнее — топологического, представлены точками (или многомерными векторами), имеющими конечное число координат. Очевидно, что при таком подходе классифицируемые объекты не имеют общей границы.

Такой подход позволяет сохранить индивидуальность каждого объекта и вместе с тем разграничить объекты.

• Интуитивное понятие сходства объектов формализуется такой математической конструкцией как метрика (расстояние, мера близости). Это позволяет рассматривать совокупность объектов как точки некоторого метрического пространства. Проблемы могут возникнуть в том случае, когда сходство выражается псевдометрикой [8] или ультраметрикой [9]. Но эти вопросы, как правило, в работах практического характера не оговариваются. В любом случае основное утверждение, связанное с метрикой, которое обязательно формулируется в классификационной задаче: объекты одного класса должны быть более близки друг к другу, чем объекты, принадлежащие разным классам. Это свойство часто неявно подразумевает, что классы точек образуют в пространстве признаков некоторые скопления, возможно различной плотности. Данное свойство послужило основой для использования в англоязычной литературе термина кластер (cluster), который затем и вытеснил в отечественной литературе другие термины. Алгоритмы, выделяющие кластеры, стали называть алгоритмами кластерного анализа.

• Отдельно сделаем замечание относительно самого пространства и представления результатов выполнения классификационной процедуры. Пространство, образованное совокупностью общих для всех объектов признаков, принято называть признаковым пространством. Под термином классификация подразумевается, как сам результат, так и процедура, с помощью которой он был получен. Ранее этот вопрос активно обсуждался в литературе [1, 10], но в настоящее время утратил свою актуальность и, как правило, каждый автор довольствуется контекстом и/или здравым смыслом.

• Свойства признакового пространства, особенно если за основу берётся евклидово пространство, как правило, не оговариваются специальным образом. По умолчанию полагают, что это пространство является связным или непрерывным — не оговаривается ни то, ни другое, и эти термины не определяются и не фигурируют среди перечисляемых свойств. Кроме этого, модель евклидова пространства предполагает его неограниченность. Эти два свойства в некотором смысле противоречат конкретным особенностям задачи: каждый, кто имел дело с практическими задачами, знает, что значения признаков не могут быть как угодно большими, а в признаковом пространстве имеются области, в которых не могут находиться объекты. На важность существования таких промежутков для признания классификации естественной обращал внимание ещё Любищев А. А. [11].

• Отсутствие упорядоченности. Как правило, классификационный результат рассматривается как измерение, выполненное на номинальном (классификационном) уровне. При этом предполагается, что все объекты одного класса являются эквивалентными, т.е. классификационное разбиение задаёт на множестве объектов отношение эквивалентности. В топологии (см., например, [12]), классы эквивалентности задают на исходном множестве т.н. фактор-множество, а сам процесс построения такого дизъюнктивного разбиения (покрытия) принято называть факторизацией множества.

Указанные выше особенности классификационных задач позволили нам сначала отказаться от рассмотрения классификации как некоторой оптимизационной задачи и переформулировать проблему как задачу исследования структуры многомерных данных в признаковом пространстве [3, 5], а затем перейти к исследованию механизмов генерирования таких структур [7]. Дальнейшее развитие этих идей позволило изменить взгляды как на саму постановку классификационной задачи, так и на скрытые за внешними признаками механизмы формирования самих данных.

Заметим, что отмеченные выше особенности позволяют, на наш взгляд, объяснить дальнейший выбор алгоритмов моделирования фрактальных структур и избежать того, что В. И. Арнольд называл «немотивированной аксиоматикой» и против чего он активно выступал.

2. ПЕРЕХОД К ФРАКТАЛЬНЫМ СТРУКТУРАМ В РЕШЕНИИ КЛАССИФИКАЦИОННЫХ ЗАДАЧАХ

Изменив формулировку задачи классификации, мы тем самым перешли к аппроксимации многомерных данных путём их моделирования посредством фрактальных множеств. При таком подходе к анализу классификационных структур необходимы не столько методы математической статистики, сколько методы теории динамических систем. Фактически это будет означать, что возникающие неопределённости будут рассматриваться не как статистические, а как хаотические [13].

В качестве основного метода решения поставленной задачи была выбрана реализация алгоритма рандомизированных систем итерированных функций (РСИФ) [14], позволяющая генерировать множества, обладающие фрактальными свойствами. Приведём очень краткое описание, чтобы не отсылать читателя к неуказанным в списке литературы источникам, содержащим более подробное описание этих процедур.

Поскольку классификационная задача формализуется в рамках геометрического подхода, то и описание алгоритма РСИФ также приводится в соответствующих терминах. Пусть имеется некоторое множество точек $Z = \{Z_j,\, j = 1, \ldots, K\}$, $Z_j \in R^n$ в метрическом пространстве (R^n, ρ).

В наиболее простом итеративном варианте зависимость $X_t \in R^n$ будет определяться следующей формулой

$$X_t = \xi X_{t-1} + (1-\xi) Z_j^{(t-1)}, \qquad (1)$$

где $0 < \xi < 1$ — значение параметра, $t = 1, 2, \ldots, N$. Впрочем, иногда бывает удобнее использовать величину $\mu = (1-\xi)/\xi$, непосредственно связанную с параметром ξ.

Уравнение (1) можно рассматривать как разностное уравнение, представляющее числовую схему приближенного решения дифференциального уравнения $\frac{\partial X}{\partial t} = (1-\xi)(Z - X)$, являющегося одной из форм уравнения Ланжевена [15, С. 77] для случайного блуждания (винеровского процесса), где вместо традиционно определяемой нормально распределённой случайной величины взята дискретная величина Z.

Выполнение процедуры РСИФ в дискретном варианте F1 начинается с произвольной точки $X_0 \in R^n$ и сводится к следующим преобразованиям:

1. Вычисляются координаты точки X_1 по формуле (1), где в соответствии с распределением вероятностной меры точка $Z_j^{(0)} \in Z$ выбирается случайным образом.

2. Затем точка X_1 принимается за исходную и процесс итеративно повторяется в соответствии с формулой (1) столько раз, сколько точек множества X требуется получить.

В [7] показано, что получить точки множества X можно и другим способом, который будем обозначать F2. Для этого необходимо рассмотреть абсолютно сходящийся ряд $\left(\frac{1-\xi}{\xi}\right)\sum_{n=1}^{\infty} \xi^n = 1, \quad (0 < \xi < 1)$ и в соответствие с распределением вероятностной меры $\{P_1, P_2, \ldots, P_K,\}$ на множестве $Z = \{Z_1, Z_2, \ldots, Z_K,\}$ разнести члены ряда по K различным слагаемым (ячейкам). Каждое полученное таким образом распределение суммы ряда будет в итоге представлять строку $A_i = (a_{i1}, a_{i2}, \ldots, a_{iK})$ некоторой матрицы A. Повторяя эту операцию N раз, получим матрицу A размером $N \times K$. Совокупность строк матрицы $A_\xi^K = \{A_i : i = 1, \ldots, N\}$ образует пространство случайных разбиений. Если из координат точек Z составить матрицу размера $K \times P$, то для того чтобы получить координаты точек множества X следует выполнить умножение $X = A \times Z$. В этом случае строки матрицы X будут представлять координаты точек X_i в пространстве R^n. Графические результаты выполнения процедур F1 и F2 представлены в наших работах [3, 7, 14].

Полученные в результате выполнения процедуры множества точек обладают следующими свойствами [14, 16].

- Каждая точка полученного множества X является выпуклой комбинацией точек $Z = \{Z_1, Z_2, \ldots, Z_K,\}$.
- Эти множества имеют нулевую лебеговскую меру, т. е. представляют собой объединение одноточечных множеств.

• Получаемые множества является совершенными [17], т. е. они замкнуты и не содержат изолированных точек.

• Множества вполне несвязны (вполне разрывны).

• Множества X могут быть гомеоморфно отображены на канторово множество, т. е. имеется возможность построить непрерывные взаимно однозначные отображения — прямое и обратное — между канторовым множеством (или его подмножеством) и множеством, полученным при выполнении указанной ранее процедуры.

• Процедуры F1 и F2 при определенных условиях устойчивы по Лагранжу [18, С. 144].

• Построенные множества имеют сингулярную функцию распределения.

• Построенные множества являются самоподобными, т. е. существуют части множества (реплики, связанные с Z_j), каждая из которых получается из целой части посредством преобразования подобия [7].

Практически это означает, что полученное посредством выполнения РСИФ процедуры множество X является фракталом [19], а точнее, с учетом конечного числа шагов, выполненных при его генерации, предфракталом.

Фактически мы имеем две схемы функционирования РСИФ и построения фрактальных множеств. Можно провести некоторую аналогию с подходами Шрёдингера и Гейзенберга в квантовой механике. Если в первом случае F1 используется аналог численной схемы решения дифференциального уравнения (1), то во втором случае моделирование структуры производится матричным способом, задающим возможные состояния объектов классифицируемой системы.

Процедуры F1 и F2 используются для моделирования структур многомерных данных в классификационных задачах.

3. УЛЬТРАМЕТРИЧЕСКИЕ ПРОСТРАНСТВА В КЛАССИФИКАЦИОННЫХ ПОСТРОЕНИЯХ КАК МОДЕЛИ КОГНИТИВНЫЗ ПРОЦЕССОВ

Противопоставление материального мира духовному можно обнаружить практически в любой философской концепции [20]. Однако это противопоставление, по мнению современных философов условны, в то время как единство мира абсолютно. Но интересно отметить, что это противопоставление находит своё отражение и в моделировании когнитивных процессов. В частности, как полагают некоторые современные учёные, для описания физических процессов адекватным средством являются действительные, а точнее — поле рациональных чисел, в то время как для моделирования ментальных процессов в большей степени соответствуют ультраметрические числовые пространства [21, 22]. Такое противопоставление можно в какой-то мере объяснить особенностями ультраметрических пространств, в основе которых лежат нарушение аксиомы Архимеда и проявление иерархических свойств пространства [23]. Предъявить окончательный ответ в пользу высказанных предположений или отвергнуть их окончательно пока ещё, пожалуй, будет преждевременным. В любом случае рассмотреть подходы к моделированию классификационных схем с учётом указанной выше специфики построения математических моделей представляется нам интересным и поучительным.

В процессе построения классификации с помощью алгоритмов кластерного анализа, мы практически имеем дело, с наборами характеристик объектов — признаками. При этом работа кластерных алгоритмов протекает в признаковом пространстве, которое является прямым отображением измерения реальных феноменологических процессов. Но сам процесс классификации в целом рассматривается как процесс ментальный, некоторые авторы его относят к разделу искусственного интеллекта, где он проходит под названием методов распознавания образов. Фактически, и на это делали упор авторы первых устройств для распознавания образов и нейросетевых моделей, мы стремимся воссоздать не работу с эмпирическими объектами, а пытаемся имитировать некоторые мыслительные процессы. Эти процессы, как показано в работах А. Ю. Хренникова, могут быть представлены с помощью случайных динамических систем [22].

Возможность аппроксимации классифицируемых множеств с помощью двух различных подходов F1 и F2, позволяет представить процесс построения классификации как процесс, который осуществляется в двух различных пространствах. Процедура F1 имеет дело исключительно с объектами признакового пространства $X \subset R^K$, где (R^K, ρ) — метрическое пространство наблюдаемых эмпирических признаков. Этот подход является основным при разработке алгоритмов кластерного анализа и представлении результатов классифицирования. Процедура F2, напротив, переводит рассмотрение классификационной задачи в пространство A_ξ^K, которое, как будет показано далее, можно рассматривать как ультраметрическое (неархимедово) пространство [9]. Такие пространства хорошо представляют модели когнитивных процессов. При этом свойства ультраметрического пространства разительно отличаются от свойств метрических пространств, но при этом близки к свойствам результирующих классификационных разбиений.

В качестве основного математического представления классификационного пространства мы предлагаем использовать пространство A_ξ^K [7], представляющее собой совокупность строк-векторов, построенных в соответствии алгоритмической процедурой F2. Для этого рассмотрим совокупность $A_\xi^K = \{A_i : i = 1, \ldots, N\}$, которая, как было отмечено выше, образуют пространство случайных разбиений. На этом множестве введём норму

$$\|A_i\|_\xi = \max_j |a_{ij}|_\xi, \qquad (2)$$

где $a_{ij} \in A_\xi$, $A_i \in A_\xi^K$. Определенная таким образом норма индуцирует на указанном множестве $A_\xi^K = A_\xi \times A_\xi \times \ldots \times A_\xi$ естественным образом метрику $d_\xi(A_s, A_t) = \|A_s - A_t\|_\xi$, которая удовлетворяет усиленному неравенству треугольника, т. е. является ультраметрикой. Всё это превращает пространство (A_ξ^K, d_ξ) в ультраметрическое [26].

Ультраметрические пространства, как известно, является объектом многочисленных исследований, как в математике, так и в других приложениях науки. Они нашли широкое применение в математической физике, где методы и средства p-адического анализа используются в различных моделях квантовой механики [21], а также теории всплесков (вейвлет-анализа) и криптографии [23]. В биофизике ультраметрические модели позволяют исследовать структуры белка и ДНК [23]. Широкое применение нашли ультраметрические пространства в моделировании когнитивных процессов [22, 25].

Введённое пространство (A_ξ^K, d_ξ), как и все ультраметрические пространства, обладает следующими свойствами [22, 26]:

- ультраметрическое пространство является вполне несвязным;
- любое ультраметрическое пространство имеет нулевую топологическую размерность;
- любой шар с в ультраметрическом пространстве является одновременно как открытым, так и замкнутым;
- любая точка шара может служить центром, радиус шара совпадает с его диаметром;
- множество значений радиуса шара счётное множество;
- если в ультраметрическом пространстве заданы два шара, то взаимное расположение имеет только два варианта: a) шары не пересекаются; b) один из шаров вложен в другой.
- всякий компакт можно покрыть конечным числом шаров фиксированного радиуса без общих точек.

Множество X, как уже было отмечено выше, может быть получено с помощью линейного преобразования $X = A \times Z$, которое переносит свойства A_ξ^K на множество X. При этом следует отметить, что переносимые свойства очень тесно связаны со свойствами классификационных разбиений (см. п. 2). Отметим, что установление наличия подобных свойств в признаковом пространстве является не совсем простой задачей.

4. ОБСУЖДЕНИЕ РЕЗУЛЬТАТОВ

Итак, решение классификационной задачи в рамках предлагаемого подхода распадается на два этапа. На первом этапе исходные данные, представленные в виде точек признакового пространства (R^n, ρ), исследуются с помощью алгоритмов кластерного анализа. Цель исследования заключается в формировании представления о структуре данных в признаковом пространстве и получении статистических оценок параметров РСИФ.

На втором этапе в классификационном пространстве (A_ξ^K, d_ξ) моделируется фрактальная структура с учетом ранее полученных оценок классификационных разбиений. Затем разбиение A_ξ^K отображается в признаковое пространство R^n. Цель этого этапа — убедиться в адекватности классификационной модели и, в случае положительного результата, перейти к прогнозным построениям. В случае недостаточно хорошего совпадения можно перейти к корректировке оценок параметров и с помощью повторения итерационной процедуры построения аппроксимирующих множеств добиться приемлемого результата. Представленная схема в значительной степени учитывает полученную в классификационном пространстве информацию и позволяет существенно улучшить интерпретацию результатов.

Таким образом, предлагаемый нами подход к решению классификационной задачи, т. е. построению классификационного разбиения на основе имеющихся эмпирических данных, вполне укладывается в модели когнитивных процессов. Основное различие заключается в том, что при моделировании в работах [22] используется так называемое символическое (мономиальные) пространство, которое является ультраметрическим. В этом случае процесс описывается данными, записанными в виде последовательности цифр. Например: если $x = (a_n, a_{n-1}, a_{n-2}, \dots, a_0)$, $y = (b_n, b_{n-1}, b_{n-2}, \dots, b_0)$, где a_i, b_i могут принимать значения $0, 1, \dots, p-1$. На этом множестве вводится метрика $d_p(x, y) = \dfrac{1}{p^k}$, если $a_i = b_i$, $i = 0, 1, \dots, k-1$ и при этом $a_k \neq b_k$. В нашем случае в роли цифр будут выступать символы протофрактала $Z = \{Z_j, j = 1, \dots, K\}$ безотносительно к их числовым значениям. Причём для точек, полученных с помощью процедуры F1 по формуле (1), символ Z_j можно определить разрешив соответствующее соотношение относительно символа протофрактала. А в случае точек, полученных по F2, значение соответствующее символу Z_j определяется по положению доминирующего элемента $\hat{a}_m$ в списке $A_i = (a_{i1}, a_{i2}, \dots, a_{iK})$ [7]. При этом следует учитывать, что $p\{Z_j / Z_{j-1}, Z_{j-2}, \dots, Z_0\} = p\{Z_j\}$, т. е. вероятность появление Z_j не зависит от появления символов на предыдущих шагах итераций.

Следует отметить, что попытки построения решения классификационной задачи с привлечением дополнительных пространств предпринимались и раньше. Так, в работах авторов французского коллектива под руководством Э. Диде [27], решение классификационной задачи получается с использованием двух различных пространств: S — пространства покрытий (разбиений) и L — пространства представительств. При этом в ходе решения происходит минимизация классификационного критерия, связывающего характеристики обеих пространств.

В качестве другого примера можно привести подход, предложенный Шрейдером Ю. и Мейном С. [28]. Выделение в моделировании классификационной задачи двух составляющих — мерономии и таксономии, приближает нас к понимаю о модельном представлении двух взаимодополнительных (в смысле Н. Бора) аспектах этой задачи. Но при этом авторы делают следующее замечание: «Таким образом, в конкретном классификационном исследовании с логической неизбежностью взаимо переплетаются таксономические и мерономические процедуры. Это и должно найти отражение в алгоритме классификации, если таковой претендует на продуктивность. Насколько нам известно, алгоритмы, в полной мере отражающие эту связь, пока не найдены» (относится к моменту публикации 1976 г. В некотором смысле содержанием данной статьи мы стремимся восполнить этот пробел)». В целом логическая схема подхода в [29] довольно близка к предлагаемой нами модели. В качестве мерона может быть представлен протофрактал, структура которого и будет определять структуру множества X при гомеоморфных преобразованиях.

ЗАКЛЮЧЕНИЕ

Решение классификационной задачи рассматривается нами как бы в двух взаимно дополнительных пространствах. Фактически в работе представлены математические модели этих двух пространств, моделируемых РСИФ процедурами F1 и F2. При этом классификационное пространство соответствует когнитивным аспектам задачи, а признаковое пространство является информационной базой (феноменологической основой) классификационной задачи. Построения, выполняемые с помощью алгоритмов кластерного анализа с эмпирическими данными, позволяют определить оценки параметров процедур РСИФ и перейти к модели классификации в ультраметрическом пространстве. При этом, как было указано выше, свойства ультраметрических пространств, вполне отвечающие результирующим классификационным разбиениям, позволяют рассматривать прообразы исходных данных и соотносить их со свойствами классификационных разбиений, получаемых в результате выполнения процедур кластерного анализа. Так, например, сферы в ультраметрических пространствах не пересекаются или находятся внутри друг друга. Этот факт важен в работе итеративных алгоритмов [3, С. 45]. Ультраметрическое пространство является по своей природе иерархическим, что создаёт некоторые удобства при выполнении иерархических алгоритмов [3, С. 48]. Области, соответствующие различным кластерам, не пересекаются — это необходимое свойство всех классификационных разбиений. Кроме того, всякий компакт можно покрыть конечным числом шаров фиксированного радиуса без общих точек. Другими словами, многие свойства классификационных построений, отвечающие различным свойствам допустимости [5, С. 163], будут автоматически выполняются. Переход же от ультраметрического пространства к метрическому выполняется с помощью линейного преобразования.

Предложенный подход к решению классификационных задач был в той или иной степени использован нами при определении типичных объектов выборки для социологического опроса, для районирования урожайности зерновых культур и прогнозирования [30], при прогнозировании временных рядов экономических показателей. Эти и другие работы читатель может легко найти в открытом доступе в eLIBRARY.

КОНФЛИКТ ИНТЕРЕСОВ

Авторы декларируют отсутствие явных и потенциальных конфликтов интересов, связанных с публикацией настоящей статьи.

СПИСОК ЛИТЕРАТУРЫ

1. *Субботин, А. Л.* Классификация / А. Л. Субботин. – ИФ РАН, М. – 2001. – 96 с.

2. *Айвазян, С. А.* Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М. : Статистика, 1974. – 240 с.

3. *Буховец, А. Г.* Моделирование структур данных в задачах классификации: математические модели классификационных задач – теория и практика / А. Г. Буховец. – Saarbrücken : Palmarium Academic Publishing, 2012. – 247 с.

4. *Мамардашвили, М. К.* Картезианские размышления / М. К. Мамардашвили. – Москва : Изд. Группа «Прогресс», 1993. – 352 с.

5. Типология и классификация в социологических исследованиях / А. А. Мирзоев, Н. И. Ростегаева, Н. В. Григорьева [и др.]; АН СССР, Институт социологических исследований. – Москва : Наука, 1982. – 296 с.

6. *Ньютон, И.* Математические начала натуральной философии [пер. с лат.] / Исаак Ньютон; ред. и предисл. Л. С. Полака. – М. : Наука, 1989. – 688 с.

7. *Буховец, А. Г.* Структура аттрактора рандомизированных систем итерированных линейных функций / А. Г. Буховец, Т. Я. Бирючинская // Вестник ВГУ, Серия: Системный анализ и информационные технологии. – 2016. – № 2. – С. 5–10.

8. *Barnsley, M. F.* Superfractals / M. F. Barnsley. – Cambridge: Cambridge University Press, 2006. – 464 p.

9. *Коблиц, Н.* p-адические числа, p-адичесциц анализ и дзета функции/ Н. Коблий – М. : Мир, 1982, 192 с.

10. *Cormak, R.V.* A Review of Classification / R. V Cormak // J. of the Royal Statistical society. – 1971. – Vol. 134. – P. 125–162.

11. *Любищев, А. А.* О критериях реальности в таксономии / А.А. Любищев // Информационные вопросы семиотики, лингвистики и автоматического перевода. М. : ВИНИТИ, 1971, вып. 1. – С. 67–82.

12. Введение в топологию: учебное пособие для студентов вузов, обучающихся по специальности «Математика» / Ю. Г. Борисович, Н. М. Близняков, Я. А. Израилевич, Т. Н. Фоменко. – 2-е издание, дополненное. – Москва : Наука, 1995. – 414 с.

13. *Малинецкий, Г. Г.* Нелинейная динамика и хаос: основные понятия / Г. Г. Малинецкий, А. Б. Потапов. – Изд. стереотип. URSS, 2018. – 240 с.

14. *Bukhovets, A. G.* Modeling of fractal data structures / A. G. Bukhovets, E. A. Bukhovets // Automation and Remote Control. – 2012. – Vol. 73, No 2. – P. 381–385.

15. *Романовский, М. Ю.* Введение в эконофизику. Статистические и динамические модели / М. Ю. Романовский, Ю. М. Романовский. – М.-Ижевск, НИЦ «Регулярная и хаотическая динамика», 2007. – 280 с.

16. *Mandelbrot, B.* The Fractal Geometry of Nature / B. Mandelbrot. – W. H. Freeman & Co., 1982. – 460 p.

17. *Хаусдорф, Ф.* Теория множеств / Ф. Хаусдорф. – Москва : ОНТИ, 1937. – 306 с.

18. *Кузнецов, С. П.* Динамический хаос / С. П. Кузнецов. – М. : ФИЗМАТЛИТ, 2006. – 356 с.

19. *Кроновер, Р. М.* Фракталы и хаос в динамических системах / Р. М. Кроновер. – М. : ТЕХНОСФЕРА, 2006. – 488 с.

20. *Капра, Ф.* Дао физики: общие корни современной физики и восточного митицизма / Ф. Капра. – М. : София, 2008. – 416 с.

21. *Владимиров, В. С.* p-адический анализ и математическая физика / В. С. Владимиров, И. В. Волович, Е. И. Зеленов. – М. : Наука, 1994. – 352 с.

22. *Хренников, А. Ю.* Моделирование процессов мышления в p-адических системах координат / А. Ю. Хренников. – М. : ФИЗМАТЛИТ, 2004. – 296 с.

23. *Dragovich B., Khrennikov A. Yu., Kozyrev S. V., Volovich I. V., Zelenov E. I.* p-adic mathematical physics: the first 30 years. p-Adic Numbers Ultrametric Anal. Appl. 9 (2017), no. 2. – P. 87–121.

24. Advances in Non-Archimedean Analysis and Applications. The p-adic Methodology in STEAM-H/ Editors W. A. Zúñiga-Galindo, Bourama Toni – Springer, 2022. – 318 p.

25. A p-adic model for the process of thinking disturbed by physiological and information noise / D. Dubischar, V. M. Gundlach, O. Steinkamp, A. Khrennikov // Journal of Theoretical Biology. – 1999. – Vol. 197, No 4. – P. 451–467. – DOI 10.1006/jtbi.1998. 0887. – EDN YDPXBW.

26. *Bukhovets, A. G.* Ultrametric properties of the attractor spaces for random iterated linear function systems / A. G. Bukhovets, P. V. Moskalev // Journal of Physics: Conference Series, Voronezh, 18–20 декабря 2017 года. – Voronezh: Institute of Physics Publishing, 2018. – P. 012028. – DOI 10.1088/1742-6596/973/1/012028.

27. Методы анализа данных: подход, основанный на методе динамических сгущений: пер. с фр. / кол. авт. под рук. Э. Диде. – М. : ФиС, 1985. – 357 с.

28. *Мейен, С. В.* Методологические аспекты теории классификации / С. В. Мейен, Ю. А. Шрейдер // Вопросы философии. – 1976. – № 12. – С. 67–79.

29. *Шрейдер, Ю. А.* Системы и модели / Ю. А. Шрейдер, А. А. Шаров. – М. : Радио и связь, 1982. – 152 с.

30. *Буховец, А. Г.* Современные подходы и методы в прогнозировании урожайности отдельных видов зерновых культур / А. Г. Буховец, Е. А. Семин, Т. Я. Бирючинская; Воронежский государственный аграрный университет им. Императора Петра I. – Воронеж : ВГАУ им. Императора Петра I, 2016. – 214 с.

Буховец Алексей Георгиевич — д-р техн. наук, профессор, профессор кафедры экономического анализа, статистики и прикладной математики Воронежского государственного аграрного университета имени императора Петра I».
E-mail: abuhovets@mail.ru
ORCID iD: https://orcid.org/0000-0002-6614-044X

Семин Евгений Александрович — канд. экон. наук, доцент кафедры экономического анализа, статистики и прикладной математики Воронежского государственного аграрного университета имени императора Петра I.
E-mail: 113ghz@mail.ru
ORCID iD: https://orcid.org/0000-0002-4674-6532

DOI: https://doi.org/10.17308/sait/1995-5499/2022/3/127-138 ISSN 1995-5499
Received 24.06.2022
Accepted 30.09.2022

FRACTAL ASPECTS OF CLASSIFICATION PROBLEM MODELING

© 2022 A. G. Bukhovets✉, E. A. Semin

Voronezh state agricultural University named after Emperor Peter the Great
1, Michurina Street, 394087 Voronezh, Russian Federation

Annotation. The focus of this work is a mathematical model of the classification problem, which is based on the concepts of the theory of fractal sets and features of modeling mental processes. Within the framework of this approach, classification activities can be viewed from two points of view. On the one hand, it is working with the empirical representation of classification objects, on the other hand, it is a mental reproduction of the task of constructing classification partitions. The realization of these two types of activity occurs in the process of constructing a fractal set generated by a randomized system of iterated functions. The mathematical model of the classification problem is presented in the form of two spaces – metric, associated with the phenomenological component, and ultrametric, reflecting the cognitive side of the problem solution. As the analysis of the results of solving classification problems shows, fractal models well meet the requirements for the formulation of the problem and the algorithmic features of its solution. The very specificity of the problem statement is reflected in the need to take into account such characteristics as the isolation of individual objects, the ability to establish the similarity /difference of objects, the compactness of the feature space, etc. These properties are characteristic of the ultrametric space created during the solution. The relationship of these two spaces is carried out by modeling the fractal structure. The paper shows exactly how the use of the fractal approach in solving classification problems is associated with the construction of ultrametric spaces. It is characteristic that these ultrametric spaces are an integral part of the algorithm for solving the problem itself. This algorithmic component of solving the classification problem is directly related to cognitive processes and interpreted as a model of processes inherent in mental activity.
Keywords: classification problem, randomized systems of iterated functions, ultrametric spaces, fractal sets, modeling of mental processes.

✉ Bukhovets Aleksey G.
e-mail: abuhovets@mail.ru

CONFLICT OF INTEREST

The authors declare the absence of obvious and potential conflicts of interest related to the publication of this article.

REFERENCES

1. *Subbotin A. L.* (2001) Classification: Monograph. *Moscow: IF RAS*. 103 p.

2. *Aivazyan S. A., Bezhaeva Z. I. and Staroverov O. V.* (1974) Classification of multidimensional observations. Moscow : Statistics. 240 p.

3. *Bukhovets A. G.* (2012) Modeling of data structures in classification problems: mathematical models of classification problems – theory and practice. *Saarbrücken : Palmarium Academic Publishing*. 247 p.

4. *Mamardashvili M. K.* (1993) Cartesian reflections. *M. Ed. Progress Group*. 352 p.

5. *Mirzoev A. A., Rostegaeva N. I. and Grigorieva N. V. [et al.]* (1982) Typology and classification in sociological research. *USSR Academy of Sciences, Institute of Sociological Research. Moscow : Nauka*. 296 p.

6. *Newton I.* (1989) Mathematical principles of natural philosophy [trans. from lat.]; ed. and foreword by L. S. Polak. *Moscow : Nauka*. 688 p.

7. *Bukhovets A. G. and Biryuchinskaya T. Ya.* (2016) The structure of the attractor of randomized systems of iterated linear functions. *VSU Bulletin, Series: System Analysis and Information Technologies*. No 2. P. 5–10.

8. *Barnsley M. F.* (2006) Superfractals. *Cambridge : Cambridge University Press*. 464 p.

9. *Koblitz N.* (1982) p-adic numbers, p-adic analysis and zeta functions. *Moscow : Mir*. 192 p.

10. *Cormak R. V.* (1971) A Review of Classification. *J. of the Royal Statistical society*. Vol. 134. P. 125–162.

11. *Lyubishchev A. A.* (1971) About criteria of reality in taxonomy. *Informational issues of semiotics, linguistics and automatic translation. Moscow : VINITI. Iss.* 1. P. 67–82.

12. *Borisovich Yu. G., Bliznyakov N. M., Izrailevich Ya. A., Fomenko T. N.* (1995) Introduction in topology: a textbook for university students studying in the specialty "Mathematics". 2nd edition, supplemented. *Moscow : Nauka*. 414 p.

13. *Malinetsky G. G. and Potapov A. B.* (2018) Nonlinear dynamics and chaos: basic concepts. *Stereotype Publishing House. URSS*. 240 p.

14. *Bukhovets A. G. and Bukhovets E. A.* (2012) Modeling of fractal data structures. *Automation and Remote Control*. Vol. 73, No 2. P. 381–385. DOI 10.1134/S0005117912020154.

15. *Romanovsky M. Yu. and Romanovsky Yu. M.* (2007) Introduction to econophysics. Statistical and dynamic models. *M.-Izhevsk, SIC "Regular and chaotic dynamics"*. 280 p.

16. *Mandelbrot B.* (1982) The Fractal Geometry of Nature. *W. H. Freeman & Co*. 460 p.

17. *Hausdorff F.* (1937) Set theory. *Moscow : ONTI*. 306 p.

18. *Kuznetsov S. P.* (2006) Dynamic chaos. *Moscow : FIZMATLIT*. 356 p.

19. *Kronover R. M.* (2006) Fractals and chaos in dynamic systems. *Moscow : TECHNOSPHERE*. 488 p.

20. *Capra F.* (2008) The Tao of Physics: common roots of modern physics and Eastern Mysticism. *Moscow : Sofia*. 416 p.

21. *Vladimirov V. S., Volovich I. V. and Zelenov E. I.* (1994) p-Adic analysis and mathematical physics. *Moscow : Nauka*. 352 p.

22. *Khrennikov A. Y.* (2004) Modeling of thinking processes in p-adic coordinate systems. *Moscow : FIZMATLIT*. 296 p.

23. *Dragovich B., Khrennikov A. Yu., Kozyrev S. V., Volovich I. V. and Zelenov E. I.* (2017) p-adic mathematical physics: the first 30 years. *p-Adic Numbers Ultrametric Anal*. Appl. 9. No 2. P. 87–121.

24. Advances in Non-Archimedean Analysis and Applications. The p-adic Methodology in STEAM-H. Editors W. A. Zúñiga-Galindo, Bourama Toni. *Springer*. 2022. 318 p.

25. *Dubischar D., Gundlach V. M., Steinkamp O. and Khrennikov A.* (1999) A p-adic model for the process of thinking disturbed by physiological and information noise. *Journal of Theoretical Biology*. Vol. 197, No 4. P. 451–467.

26. *Bukhovets A. G. and Moskalev P. V.* (2018) Ultrametric properties of the attractor spaces for random iterated linear function systems // Journal of Physics: Conference Series, Voronezh, 18–20 December 2017. *Voronezh : Institute of*

Physics Publishing. P. 012028. DOI 10.1088/1742-6596/973/1/012028.

27. Methods of data analysis: an approach based on the method of dynamic thickening: trans. from the French / Col. auth. under the hands of E. Dide. *Moscow : FiS*. 1985. 357 p.

28. *Meyen S. V. and Schrader Yu. A.* (1976) Methodological aspects of the theory of classification. *Questions of Philosophy*. No 12. P. 67–79.

29. *Schrader Yu. A. and Sharov A. A.* (1982) Systems and models. *Moscow : Radio and Communications*. 152 p.

30. *Bukhovets A. G., Semin E. A. and Biryuchinskaya T. Ya.* (2016) Modern approaches and methods in forecasting the yield of certain types of grain crops. Voronezh State Agrarian University. Emperor Peter I. *Voronezh : VGAU named after Emperor Peter I*. 214 p.

Bukhovets Aleksey G. — Dr. Sc. (Eng.), Full Professor, the Department of Economic Analysis, Statistics and Applied Mathematics, Voronezh state agricultural University named after Emperor Peter the Great, Russia.
E-mail: abuhovets@mail.ru
ORCID iD: https://orcid.org/0000-0002-6614-044X

Semin Evgeniy A. — Candidate of Economic Sciences, Associate Professor at the Department of Economic Analysis, Statistics and Applied Mathematics, Voronezh State Agricultural University named after Emperor Peter the Great.
E-mail: 113ghz@mail.ru
ORCID iD: https://orcid.org/0000-0002-4674-6532