Math-Net.Ru

Общероссийский математический портал

С. В. Козырев, Ультраметрика в теории сложных систем, *ТМФ*, 2015, том 185, номер 2, 346–360

DOI: <https://doi.org/10.4213/tmf8955>

Использование Общероссийского математического портала Math-Net.Ru подразумевает, что вы прочитали и согласны с пользовательским соглашением
<http://www.mathnet.ru/rus/agreement>

Параметры загрузки:

IP: 54.198.67.100

26 апреля 2023 г., 17:30:26

© 2015 г.

С. В. Козырев*

УЛЬТРАМЕТРИКА В ТЕОРИИ СЛОЖНЫХ СИСТЕМ

Дается обзор приложений p -адических и ультраметрических методов в теории сложных систем. Рассматриваются следующие примеры: p -адическая параметризация матрицы Паризи в методе реплик; метод иерархической (межбассейновой) кинетики, при помощи которого динамика макромолекул описывается моделями ультраметрической диффузии; двумерная 2-адическая параметризация генетического кода, показывающая, что вырождение генетического кода описывается локальным постоянством отображения в 2-адической метрике. Обсуждаются методы кластеризации для семейства метрик и показывается связь подхода мультикластеризации (ансамблевой кластеризации) с теорией билдингов Брюа–Титса.

Ключевые слова: ультраметрика, сложные системы, кластеризация.

DOI: 10.4213/tmf8955

1. ВВЕДЕНИЕ

В настоящей работе обсуждаются приложения p -адических и ультраметрических методов в теории сложных систем. Для обзора p -адической математической физики и приложений см. труды [1]–[5]. Применение p -адического анализа в математической физике началось с работы [6]. Отправной точкой для приложения ультраметрических методов к сложным системам стал метод нарушения репличной симметрии [7], в котором ультраметрические пространства были использованы для описания пространства состояний спиновых стекол. В работах по методу реплик ультраметрика возникала в результате процесса ветвления в пространстве высокой размерности. Связь нарушения репличной симметрии с p -адическим анализом (p -адическая параметризация матрицы Паризи) была обнаружена в работах [8], [9], в частности, в статье [9] корреляционные функции метода реплик были представлены в виде p -адических интегралов. Обобщение таких результатов на более общие ультраметрические пространства рассматривалось в работах [10]–[12]. В работах [13], [14] рассматривалось применение преобразования Фурье на группах к диагонализации

Исследование выполнено за счет гранта Российского научного фонда (проект № 14-50-00005).

*Математический институт им. В.А. Стеклова Российской академии наук, Москва, Россия.
E-mail: kozyrev@mi.ras.ru

матрицы Паризи (что близко к p -адическому подходу). Настоящая работа содержит в основном обзор результатов, полученных при участии автора.

Понятие “сложной системы” в физике и других областях широко обсуждалось в литературе (см., например, монографии [7], [15]). Наиболее абстрактным образом сложную систему можно обсуждать как систему с высокой колмогоровской сложностью, но реально в физических моделях сложных систем (например, спиновых стекол) всегда применяются соображения общего положения и высокой размерности.

Статистическая механика сложных (неупорядоченных) систем связана также с работами по динамике на ландшафтах энергии. В частности, применительно к динамике макромолекул рассматривалось приближение межбассейновой кинетики, в которой динамика молекулы приближалась переходами между локальными минимумами энергии [16]. При этом локальные минимумы и переходные состояния между ними образуют дерево (disconnectivity graph), и аррениусова кинетика переходов между локальными минимумами принимает вид системы уравнений с матрицей скоростей переходов, являющейся обобщением матрицы Паризи [17]. В простейшем случае динамика такой системы описывается уравнением p -адической диффузии с генератором в виде оператора Владимирова p -адического дробного дифференцирования [8]. В работе [18] было показано, что такая модель дает согласующееся с экспериментами описание динамики белка (связывания миоглобина-СО), в работе [19] это наблюдение было уточнено в области низких температур. В статьях [20], [21] рассматривались различные ультраметрические модели динамики белка. Общая межбассейновая кинетика связана с ультраметрической диффузией со сносом [22]. В работе [23] изучалось p -адическое случайное блуждание. В работе [24] рассматривалось случайное блуждание по границе дерева на комплексной плоскости (на ландшафте энергии, отвечающем функции Дедекинда). В статьях [25], [26] иерархические подходы применялись к обсуждению функционирования молекулярной машины.

Наиболее общий подход к приложению ультраметрики к сложным системам состоит в методах кластеризации, т. е. иерархической классификации объектов на основе их сходства (например, близости расстояний в метрических пространствах) [27]. При этом возникают классификационные деревья кластеров с частичным порядком по вложению, граница такого дерева является ультраметрическим пространством. О применении ультраметрики к процедуре кластеризации и другим приложениям анализа данных см. работы [28], [29]. О двойственности между деревьями и ультраметрическими пространствами см., например, монографию [30]. Иерархическую классификацию можно рассматривать как приближение сложной системы ультраметрической. Метод межбассейновой кинетики можно рассматривать как частный случай такой иерархической кластеризации (по энергиям переходных состояний).

Иерархические модели квантовой статистической механики, связанные с деревьями, изучались в работах [31]–[33].

Обычно в приложениях метрика кластеризации определена неоднозначно, в результате возникает набор деревьев кластеров. Подобный подход (построение ансамбля деревьев кластеров) обычно называется мультикластеризацией (multi-clustering, ensemble clustering) [34]. Мы рассматриваем обобщение такого подхода, в котором деревья кластеров объединяются в один симплициальный комплекс. В p -адическом случае этот комплекс связан с аффинными билдингами Брюа–Титса, в более об-

щем случае мы получаем обобщение билдинга для произвольного случая анализа данных [35]–[37].

Еще один разбираемый в настоящей работе пример – это p -адическая параметризация генетического кода (способа кодирования аминокислот тройками нуклеотидов). Одномерный вариант такой конструкции был предложен в [38], [39]. Мы обсудим двумерную параметризацию генетического кода, основанную на следующем наблюдении: при естественной параметризации нуклеотидов на основе их химических свойств получаемая при этом двумерная 2-адическая параметризация кодонов описывает вырождение генетического кода (вырождение становится локальным по-стоянством отображения на малых расстояниях относительно двумерной 2-адической метрики) [40]–[44].

Для более подробного обзора части обсуждаемых в настоящей работе результатов см. статью [45].

Настоящая работа имеет следующую структуру. В разделе 2 обсуждаются методы кластеризации и двойственность между деревьями и ультраметрическими пространствами. В разделе 3 рассматривается многомерная кластеризация по семейству метрик и соответствующие симплициальные комплексы. В разделе 4 обсуждается p -адическая параметризация матрицы Паризи теории спиновых стекол. В разделе 5 обсуждается связь моделей межбассейновой кинетики и общих уравнений ультраметрической диффузии со сносом. В разделе 6 мы рассматриваем двумерную 2-адическую модель генетического кода. Раздел 7 содержит заключение по обсуждавшимся результатам.

2. КЛАСТЕРИЗАЦИЯ, ДЕРЕВЬЯ И УЛЬТРАМЕТРИЧЕСКИЕ ПРОСТРАНСТВА

Кластеризация и ультраметрика. Кластеризация есть некоторый способ иерархической классификации объектов при помощи деревьев. Здесь мы рассмотрим методы кластеризации наборов точек в метрических пространствах (рассмотренный способ кластеризации называется кластеризацией по ближайшему соседу, можно рассмотреть также другие методы кластеризации).

Пусть (M, ρ) – некоторое метрическое пространство (например, набор точек в $\mathbb{R}^n$ со стандартной евклидовой метрикой). Последовательность точек $\{x_i\}$, $i = 1, \dots, N$, в (M, ρ) называется ε -цепью, если для двух последовательных точек из этой последовательности $\rho(x_i, x_{i+1}) \leq \varepsilon$. Точки $a, b \in (M, \rho)$ назовем ε -связанными, если существует ε -цепь $\{x_i\}$, $i = 1, \dots, N$, соединяющая эти точки, т. е. для такой цепи $a = x_1$, $b = x_N$.

Цепное расстояние $d(a, b)$ между a и b определим как нижнюю грань таких ε , что a и b ε -связаны. Цепное расстояние есть псевдоультраметрика (т. е. удовлетворяет аксиомам ультраметрики, кроме аксиомы невырожденности). Поэтому цепное расстояние определяет ультраметрику на классах эквивалентности точек из (M, ρ) : две точки эквивалентны, если цепное расстояние между ними равно нулю.

Так, например, для ультраметрического пространства цепное расстояние совпадает с ультраметрикой; для связной области в $\mathbb{R}^n$ цепное расстояние между любыми точками равно нулю.

Шар в (M, ρ) относительно цепного расстояния $d(\cdot, \cdot)$ назовем кластером. Кластеризацией (M, ρ) назовем некоторое покрытие пространства M кластерами, причем:

- 1) каждая точка должна лежать хотя бы в одном кластере;
- 2) для каждой пары точек a, b должен существовать минимальный содержащий их кластер $\sup(a, b)$;
- 3) для каждой пары вложенных кластеров $A \subset C$ существует максимальный кластер B такой, что $A \subset B \subset C$, и множество промежуточных кластеров между двумя вложенными кластерами всегда должно быть конечным.

Процедура кластеризации порождает частично упорядоченное дерево кластеров (дендрограмму), в котором:

- 1) вершины дерева суть кластеры;
- 2) частичный порядок задан вложением множеств;
- 3) ребра соединяют кластеры, вложенные без промежуточных.

Рассмотрим функцию на дереве кластеров, ставящую в соответствие кластеру его (цепной) диаметр. Такая функция принимает неотрицательные значения и является монотонно возрастающей относительно частичного порядка. Более того, в каждом конечном пути в дереве (путь есть последовательность вершин дерева, где вершины с соседними номерами в пути являются соседними вершинами дерева) существует ровно одна максимальная вершина.

ПРИМЕР 1. Пусть $\{d_k\}$ есть возрастающая до бесконечности последовательность положительных чисел без точек сгущения, кроме точки ноль, например $\{q^{-k}\}$, где $0 < q < 1$, $k \in \mathbb{Z}$. Тогда в качестве кластеризации в (M, ρ) можно рассмотреть набор всевозможных кластеров с цепными диаметрами q^{-k} .

Двойственность между деревьями и ультраметрическими пространствами. Пусть X – полное локально компактное ультраметрическое пространство. Рассмотрим множество $\mathcal{T}(X)$ шаров в X , содержащее все шары ненулевого диаметра, и шары нулевого диаметра, являющиеся максимальными подшарами в шарах ненулевого диаметра (изолированные точки в X). Тогда $\mathcal{T}(X)$ является частично упорядоченным деревом (деревом шаров в ультраметрическом пространстве X).

Для этого дерева имеют место следующие утверждения [45]:

- 1) множество всех вершин не более чем счетно;
- 2) для каждой вершины I , отвечающей шару ненулевого диаметра, существует конечное множество вершин, соединенных с такой вершиной I ребром (при этом вершины из такого множества, меньшие I относительно частичного порядка, будут отвечать максимальным подшарам в I , таких вершин будет не менее двух);
- 3) каждый (не содержащий самопересечений) путь в дереве $\mathcal{T}(X)$, соединяющий две вершины, конечен, и содержит единственную максимальную вершину (могут существовать и бесконечно убывающие либо возрастающие пути);
- 4) диаметр $d(I)$ шара есть функция на дереве шаров, принимающая неотрицательные значения и монотонно возрастающая относительно частичного порядка, при этом функция $d(I)$ стремится к нулю для убывающих последовательностей вложенных шаров $\{I\}$ и обращается в нуль на вершинах, принадлежащих границе дерева $\mathcal{T}(X)$.

Обратно, пусть у нас есть дерево $\mathcal{T}$ с частичным порядком и функцией $d(\cdot)$ на множестве вершин дерева $\mathcal{T}$, причем выполнены перечисленные выше ограничения на дерево, частичный порядок и функцию $d(\cdot)$.

Определим метрику на $\mathcal{T}$ следующим образом. Между двумя вершинами I, J в дереве существует единственный конечный путь без самопересечений. Обозначим

(единственную) максимальную на этом пути вершину как $\sup(I, J)$, и рассмотрим функцию $d(I, J) = d(\sup(I, J))$. Тогда эта функция будет ультраметрикой на множестве вершин дерева.

Рассмотрим пополнение дерева $\mathcal{T}$ по метрике $d(\cdot, \cdot)$. Такое пополнение содержит само дерево $\mathcal{T}$. Выбросим из пополнения все внутренние вершины дерева $\mathcal{T}$, результат такой операции обозначим как $X(\mathcal{T})$ (внутренними называются вершины дерева, инцидентные более чем одному ребру). Тогда $X(\mathcal{T})$ будет локально компактным ультраметрическим пространством, метрику на котором мы также будем обозначать как $d(\cdot, \cdot)$. Пространство $X(\mathcal{T})$ будем называть границей дерева $\mathcal{T}$.

При фиксированном частичном порядке в $\mathcal{T}$ и выборе различных функций $d(\cdot)$ диаметров шаров (возрастающих относительно фиксированного частичного порядка) соответствующие ультраметрические пространства $X(\mathcal{T})$ будут эквивалентны (будут гомеоморфны и метрики на $X(\mathcal{T})$ для разных функций $d(\cdot)$ будут порождать тот же набор шаров).

3. МНОГОМЕРНАЯ КЛАСТЕРИЗАЦИЯ ПО СЕМЕЙСТВУ МЕТРИК

Описанная в предыдущем разделе процедура кластеризации обобщается на случай, когда вместо одной метрики кластеризации используется набор метрик. При этом получается набор деревьев кластеров, который естественно рассматривать как симплициальный комплекс.

Основная идея выглядит следующим образом. Пусть мы имеем множество X , на котором задано некоторое семейство метрик S . Обозначим через $\mathcal{T}(X, s)$ дерево кластеров в X для некоторой метрики $s \in S$. Может случиться, что для двух разных метрик $r, s \in S$ некоторые кластеры (вершины в деревьях $\mathcal{T}(X, s)$, $\mathcal{T}(X, r)$) совпадают как подмножества в X . Тогда мы можем отождествить такие вершины в этих двух деревьях (склеить эти деревья по таким вершинам). Мы получим граф с частичным порядком.

Результат данной процедуры для набора метрик S обозначим как $\mathcal{T}(X, S)$. Для такого графа фиксируем пару S -кластеров I, J (кластеров относительно всех метрик $s \in S$), один из которых вложен в другой без промежуточных S -кластеров (т.е. промежуточные между двумя этими кластерами вершины в $\mathcal{T}(X, S)$ могут быть кластерами относительно некоторых метрик из S , но не являются кластерами относительно всех метрик из S).

Выберем некоторую метрику $s \in S$ и рассмотрим набор (всевозможных) промежуточных S -кластеров между I, J , т.е. набор $I = J_0 \subset J_1 \subset \dots \subset J_k = J$ (такой набор должен быть конечен по определению). Назовем данный набор симплексом, также назовем симплексами всевозможные подмножества такого набора. Это определяет структуру симплициального комплекса на $\mathcal{T}(X, S)$.

ПРИМЕР 2. На рис. 1, 2 изображены два способа кластеризации на множестве из трех точек и соответствующие деревья $\mathcal{A}, \mathcal{B}$ кластеров. На рис. 3 изображено объединение деревьев кластеров в сеть кластеров $\mathcal{C}$, отвечающую двум метрикам. Симплексы в $\mathcal{C}$ отвечают путям между минимальной и максимальной вершинами цикла в $\mathcal{C}$ и подмножествам таких путей. Рис. 1–3 взяты из работы автора [37].

Рис. 1. Дерево кластеров $\mathcal{A}$. Рис. 2. Дерево кластеров $\mathcal{B}$. Рис. 3. Сеть кластеров $\mathcal{C}$.

ПРИМЕР 3. Рассмотрим многомерные p -адические метрики. Пусть $X = \mathbb{Q}_p^d$ и метрика определена нормой для $z = (z_1, \dots, z_d)$ вида

$$N(z) = \max_{i=1, \dots, d} (q_i |z_i|_p), \quad q_i > 0, \quad (1)$$

т. е. метрика имеет вид $s(x, y) = N(x - y)$. Более общие метрики можно определить с помощью норм $N(Az)$, где A – матрица из $Gl(\mathbb{Q}_p, d)$.

Если для нормы N вида (1) принять условие $p^{-1} < q_i \leq 1$ для любых $i = 1, \dots, d$, то все множества вида $p^k \mathbb{Z}_p^d$, $k \in \mathbb{Z}$, будут шарами относительно всех таких норм N . В частности, для случая

$$p^{-1} < q_1 < \dots < q_d \leq 1 \quad (2)$$

набор промежуточных шаров между $p\mathbb{Z}_p^d$ и $\mathbb{Z}_p^d$ имеет вид $\{B_j\}$, где

$$B_j = \mathbb{Z}_p \times \dots \times \mathbb{Z}_p \times p\mathbb{Z}_p \times \dots \times p\mathbb{Z}_p \quad (3)$$

есть произведение j компонент $\mathbb{Z}_p$ и $d - j$ компонент $p\mathbb{Z}_p$ для каждого $j = 0, \dots, d$.

Если мы изменим порядок коэффициентов в условии (2), то получим другой набор промежуточных шаров между $p\mathbb{Z}_p^d$ и $\mathbb{Z}_p^d$. Поэтому при кластеризации по набору метрик, отвечающему различным перестановкам коэффициентов в (2), мы получим граф кластеров. Симплексы в таком графе (рассматриваемом как симплициальный комплекс) будут отвечать набору (3) и его подмножествам, образам таких наборов относительно перестановок координат в $\mathbb{Q}_p^d$, а также сдвигам таких наборов как наборов множеств в $\mathbb{Q}_p^d$.

Можно также рассмотреть кластеризацию по более общему семейству метрик (отвечающему нормам $N(Az)$, $A \in Gl(\mathbb{Q}_p, d)$), но для таких метрик не обязательно множества $p^k\mathbb{Z}_p^d$, $k \in \mathbb{Z}$, будут шарами (и вообще общих шаров ненулевого диаметра для всего семейства метрик может не существовать). Поэтому в общем случае симплексы в симплициальном комплексе кластеризации определяются как набор промежуточных шаров между двумя шарами для некоторого подсемейства всего семейства метрик кластеризации. Тогда для образа относительно линейного отображения A метрики, удовлетворяющей условию (2) (и метрик для перестановок координат в таком условии), симплексы будут иметь вид образов подмножеств набора (3) относительно линейных отображений $A \in Gl(\mathbb{Q}_p, d)$, а также сдвигов таких симплексов в $\mathbb{Q}_p^d$.

Поскольку множества B_j вида (3) и их образы относительно линейных отображений являются решетками (открытыми компактными $\mathbb{Z}_p$ -модулями в $\mathbb{Q}_p^d$), описанная выше конструкция симплициального комплекса связана с конструкцией аффинного билдинга Брюа–Титса. Для получения аффинного билдинга из симплициального комплекса кластеров для семейства метрик следует взять кластеры, содержащие ноль (т. е. решетки), и профакторизовать подкомплекс таких кластеров по действию группы растяжений на p^k , $k \in \mathbb{Z}$. Симплициальный комплекс относительно семейства метрик, отвечающих условиям (1) и (2), и метрик, заданных путем перестановок координат в этих условиях, соответствует апартаменту в аффинном билдинге. Поэтому для случая общих семейств метрик получаемые сети кластеров являются обобщением аффинных билдингов. Неединственность деревьев кластеров при неоднозначном задании метрики обсуждается, например, в биоинформатике (применительно к вопросам таксономии) и рассматривается как проявление неоднородности процесса биологической эволюции.

4. p -АДИЧЕСКАЯ ПАРАМЕТРИЗАЦИЯ МАТРИЦЫ ПАРИЗИ

В настоящем разделе мы опишем p -адическую параметризацию матрицы Паризи из теории спиновых стекол. Подобная параметризация позволяет представить корреляционные функции модели Шеррингтона–Киркпатрика спинового стекла в виде интеграла по p -адическому аргументу.

Матрица Паризи есть $(n \times n)$ -матрица $Q = (Q_{ab})$, матричные элементы которой равны

$$Q_{aa} = 0, \quad Q_{ab} = q_i, \quad \text{где} \quad \left[\frac{a}{m_i} \right] \neq \left[\frac{b}{m_i} \right], \quad \left[\frac{a}{m_{i+1}} \right] = \left[\frac{b}{m_{i+1}} \right]. \quad (4)$$

Здесь натуральные числа m_i определены так, что их отношения $p_i = m_{i+1}/m_i$ натуральные, и

$$1 = m_0 < m_1 < \dots < m_{k-1} < m_k = n.$$

Значение функции $[x]$ есть наименьшее целое число, большее или равное x (целая часть справа), q_i суть неотрицательные вещественные параметры.

Пример матрицы Паризи для $p_i = 2$ для всех i имеет вид

$$Q = \begin{pmatrix} 0 & q_1 & q_2 & q_2 & q_3 & q_3 & q_3 & q_3 & \dots \\ q_1 & 0 & q_2 & q_2 & q_3 & q_3 & q_3 & q_3 & \dots \\ q_2 & q_2 & 0 & q_1 & q_3 & q_3 & q_3 & q_3 & \dots \\ q_2 & q_2 & q_1 & 0 & q_3 & q_3 & q_3 & q_3 & \dots \\ q_3 & q_3 & q_3 & q_3 & 0 & q_1 & q_2 & q_2 & \dots \\ q_3 & q_3 & q_3 & q_3 & q_1 & 0 & q_2 & q_2 & \dots \\ q_3 & q_3 & q_3 & q_3 & q_2 & q_2 & 0 & q_1 & \dots \\ q_3 & q_3 & q_3 & q_3 & q_2 & q_2 & q_1 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \quad (5)$$

В работах [8], [9] было показано, что для случая $m_i = p^i$ (т.е. когда все $p_i = p$) формула (4) эквивалентна следующей p -адической параметризации: после соответствующей перенумеровки индексов матрицы Паризи матричные элементы этой матрицы имеют вид

$$Q_{ab} = q(|l(a) - l(b)|_p), \quad (6)$$

где $|\cdot|_p$ есть p -адическая норма и функция $q(x)$ определяется как $q(p^i) = q_i$, $q(0) = 0$. Такая перенумеровка задана отображением

$$l: \{1, 2, \dots, p^k\} \mapsto p^{-k}\mathbb{Z}/\mathbb{Z},$$

$$l^{-1}: x = \sum_{i=-k}^{-1} x_i p^i \mapsto 1 + \sum_{i=1}^k x_i p^{i-1}, \quad 0 \leq x_i \leq p-1.$$

Это отображение есть по существу отображение Монна (p -адическая замена переменной) в дискретном случае [46].

Рассмотрим теперь систему кинетических уравнений детального баланса с матрицей переходов, равной матрице Паризи (модель иерархической кинетики, см. далее)

$$\frac{d}{dt} f(a, t) = - \sum_{b=1}^{p^k} Q_{ab} [f(a, t) - f(b, t)].$$

Применяя p -адическую параметризацию матрицы Паризи, получим для этой системы выражение

$$\frac{\partial}{\partial t} f(a, t) = - \int_{p^{-M}\mathbb{Z}/p^N\mathbb{Z}} q(|a - b|_p) [f(a, t) - f(b, t)] d\mu(b),$$

где интегрирование идет по мере Хаара на группе $p^{-M}\mathbb{Z}/p^N\mathbb{Z}$. Взятие совместного предела $M, N \rightarrow \infty$ для такого уравнения дает выражение

$$\frac{\partial}{\partial t} f(x, t) = - \int_{\mathbb{Q}_p} q(|x - y|_p) [f(x, t) - f(y, t)] d\mu(y),$$

которое при $q(|x|_p) = \Gamma_p^{-1}(-\alpha)|x|_p^{-1-\alpha}$ принимает вид p -адического уравнения диффузии, генерируемого дробным оператором Владимирова, изученного в книге [1].

5. ДЕРЕВО БАССЕЙНОВ И МЕЖБАССЕЙНОВАЯ КИНЕТИКА

Обсудим на физическом уровне строгости иерархическую аррениусову кинетику на ландшафтах энергии и покажем, что такая кинетика естественным образом описывается моделью ультраметрической диффузии.

Пусть h есть принимающая положительные значения гладкая функция общего положения (функция Морса) на связной области в $\mathbb{R}^n$. Функцию h будем называть энергией (или ландшафтом энергии). Пусть $S(x, y)$ – пространство путей (гладких кривых) между x и y , $s \in S(x, y)$ есть один из таких путей, а $t \in s$ есть точка на пути s .

Определим функцию двух переменных

$$d(x, y) = \inf_{s \in S(x, y)} \sup_{t \in s} h(t), \quad x \neq y.$$

Для $x = y$ доопределим $d(x, x) = 0$. Значение функции $d(x, y)$ равно максимальному барьеру энергии, который необходимо преодолеть, чтобы переместиться между x и y . Если функция h немонотонна на любом пути между x и y , то $d(x, y)$ равно значению функции h в некоторой критической точке (точке, в которой градиент функции обращается в нуль). Функция $d(x, y)$ является ультраметрикой. Нас будет интересовать такая ультраметрика на множестве X локальных минимумов функции h (в этом случае h немонотонна на любом пути между локальными минимумами из X).

Кластеризация множества X локальных минимумов h по ультраметрике $d(\cdot, \cdot)$ порождает дерево кластеров, которое мы будем называть деревом бассейнов (disconnectivity graph). Такое дерево кластеров описывает набор локальных минимумов функции h и критических точек (перевалов) между локальными минимумами. Диаметры шаров будут отвечать энергетическим барьерам между локальными минимумами для переходов в соответствующих шарах.

ПРИМЕР 4. Пусть функция h имеет три локальных минимума A, B, C , барьер (седловую точку) высотой h_1 между A и B и барьер высотой h_2 между парой A, B и C , при этом $h_1 < h_2$. Составим матрицу расстояний между локальными минимумами, строки и столбцы которой нумеруются минимумами A, B, C :

$$\left(\begin{array}{cc|c} 0 & h_1 & h_2 \\ h_1 & 0 & h_2 \\ \hline h_2 & h_2 & 0 \end{array} \right).$$

ПРИМЕР 5. Рассмотрим случай четырех локальных минимумов A, B, C, D . Пусть барьеры между минимумами имеют следующий вид: барьер высотой h_1 между A и B , а также между C и D и барьер высотой $h_2 > h_1$ между парами A, B и C, D . Матрица расстояний между минимумами, строки и столбцы которой нумеруются в порядке A, B, C, D , имеет вид

$$\left(\begin{array}{cc|cc} 0 & h_1 & h_2 & h_2 \\ h_1 & 0 & h_2 & h_2 \\ \hline h_2 & h_2 & 0 & h_1 \\ h_2 & h_2 & h_1 & 0 \end{array} \right).$$

Матрицы барьеров имеют иерархический вид и являются аналогами матрицы Паризи теории спиновых стекол.

Для ландшафта энергии построим систему уравнений межбассейновой кинетики, пользуясь формулой Эйринга–Полани. Эта формула говорит, что константа скорости реакции (скорость перехода между локальными минимумами для диффузии на энергетическом ландшафте) приблизительно пропорциональна $e^{-\beta\Delta F}$, где β – обратная температура, $\Delta F = F_1 - F_0$ есть разность свободных энергий переходного состояния и начального состояния реакции. Начальным состоянием реакции является один из локальных минимумов энергии, а переходное состояние представляет собой перевал (седловую точку энергии) между двумя локальными минимумами энергии. Аналогичная формула без учета энтропии называется формулой Аррениуса. Свободная энергия равна $F = E - TS$, где E – энергия, $T = \beta^{-1}$ – температура, S – энтропия (в рассматриваемом случае логарифм объема окрестности локального минимума либо седловой точки).

Для дерева бассейнов энергетического ландшафта рассмотрим систему кинетических уравнений Эйринга–Полани (обобщения кинетических уравнений Аррениуса), удовлетворяющую условиям детального баланса,

$$\frac{d}{dt}g(i, t) = - \sum_{j \neq i} C(i, j) [e^{\beta(F(i) - F(\text{sup}(i, j)))} g(i, t) - e^{\beta(F(j) - F(\text{sup}(i, j)))} g(j, t)].$$

Здесь $g(i, t)$ – заселенность i -го локального минимума, $F(i)$ – свободная энергия i -го локального минимума, $\text{sup}(i, j)$ – переходное состояние между локальными минимумами i и j (отвечающее максимальной вершине в пути между i и j в дереве кластеров локальных минимумов для ландшафта энергии). Величина $F(\text{sup}(i, j))$ есть свободная энергия переходного состояния.

Величины $C(i, j)$ образуют набор положительных симметричных по i и j коэффициентов, выбор которых фиксирует модель кинетических уравнений, описывающих аррениусову динамику на ландшафте энергии. Мы используем

$$C(i, j) = \frac{\#(i) \#(j)}{\#^2(\text{sup}(i, j))},$$

где $\#(i)$, $\#(j)$ – объемы окрестностей локальных минимумов, $\#(\text{sup}(i, j))$ – объем окрестности переходного состояния (такой выбор является масштабно-инвариантным).

Пусть $f(i, t) = g(i, t)\#(i)$ есть плотность заселенности i -го локального минимума. Тогда система кинетических уравнений примет вид

$$\frac{d}{dt}f(i, t) = - \sum_{j \neq i} \frac{e^{-\beta E(\text{sup}(i, j))}}{\#(\text{sup}(i, j))} [e^{\beta E(i)} f(i, t) - e^{\beta E(j)} f(j, t)] \#(j).$$

Здесь $E(i)$ и $E(j)$ – энергии локальных минимумов, $E(\text{sup}(i, j))$ – энергия переходного состояния.

В пределе большого числа локальных минимумов система кинетических уравнений принимает вид уравнения ультраметрической диффузии со сносом на ультраметрическом пространстве X локальных минимумов ландшафта энергии:

$$\frac{\partial}{\partial t}f(x, t) + \int_X \frac{e^{-\beta E(\text{sup}(x, y))}}{\nu(\text{sup}(x, y))} [e^{\beta E(x)} f(x, t) - e^{\beta E(y)} f(y, t)] d\nu(y) = 0,$$

где ν – некоторая (борелевская) мера на X , $E(x)$ – вещественная функция на X (глубина потенциальных ям), $E(\sup(x, y))$ – распределение энергий переходных состояний (вещественнозначная монотонно возрастающая относительно частичного порядка функция на дереве шаров $T(X)$ и $E(\sup(x, y)) > E(x)$, $E(\sup(x, y)) > E(y)$).

Для случая $X = \mathbb{Q}_p$ мера ν есть мера Хаара, глубины локальных минимумов одинаковы, $E(x) = \text{const}$, энергии переходных состояний имеют вид $E(\sup(x, y)) = \ln |x - y|_p^\kappa$, а описанное выше уравнение ультраметрической диффузии записывается как

$$\frac{\partial}{\partial t} f(x, t) + D^\alpha f(x, t) = 0;$$

где D^α – оператор Владимирова p -адического дробного дифференцирования, $\alpha = \beta\kappa$. Подобное p -адическое уравнение теплопроводности было рассмотрено в книге [1], где был изучен соответствующий случайный процесс. Диффузии в неархимедовых пространствах изучались в [3], [23].

6. 2-АДИЧЕСКАЯ МОДЕЛЬ ГЕНЕТИЧЕСКОГО КОДА

Белки и нуклеиновые кислоты (ДНК и РНК) являются линейными полимерами, т. е. конечными последовательностями оснований, или словами с буквами из конечного алфавита. Для нуклеиновых кислот имеются четыре сорта оснований, белки состоят из аминокислот двадцати различных сортов. Генетический (аминокислотный) код есть способ кодирования белков на нуклеиновых кислотах, при этом аминокислота кодируется тройкой стоящих рядом нуклеотидов, такая тройка называется кодоном. Имеются 64 различных кодона, при этом генетический код принимает 21 возможное значение (двадцать аминокислот и стоп-кодон, символ остановки трансляции). Таким образом, генетический код является вырожденным. Кроме того, представляет интерес вопрос о связи химических свойств аминокислоты и способа ее кодирования. В работах [40]–[44] была предложена и исследована модель генетического кода, основанная на двумерной 2-адической параметризации пространства кодонов. Одномерная параметризация генетического кода была рассмотрена в работах [38], [39].

Двумерная 2-адическая параметризация генетического кода основана на следующих наблюдениях.

1. Нуклеотиды могут быть охарактеризованы парами 0, 1 на основе двух признаков: нуклеотиды являются либо пуринами, либо пиримидинами; нуклеотиды образуют либо две, либо три водородные связи при спаривании с комплементарными в двойной спирали. Для РНК мы получаем двумерную параметризацию пространства нуклеотидов вида

$$\begin{array}{|c|c|} \hline A & G \\ \hline U & C \\ \hline \end{array} = \begin{array}{|c|c|} \hline 00 & 01 \\ \hline 10 & 11 \\ \hline \end{array}. \quad (7)$$

2. На основе двумерной параметризации нуклеотидов можно задать двумерную параметризацию кодонов (троек нуклеотидов). При этом каждый нуклеотид будет задавать параметризацию на своем масштабе относительно двумерной 2-адической метрики. А именно, рассмотрим 2-адическую плоскость с координатами (x, y) :

$$\begin{aligned} x &= (x_0 x_1 x_2) = x_0 + 2x_1 + 4x_2, & x_i &= 0, 1, \\ y &= (y_0 y_1 y_2) = y_0 + 2y_1 + 4y_2, & y_i &= 0, 1. \end{aligned}$$

Рассмотрим отображение

$$\rho: C_1 C_2 C_3 \mapsto (x, y) = (x_0 x_1 x_2, y_0 y_1 y_2), \quad (8)$$

переводящее кодон в точку 2-адической плоскости по следующим правилам: C_2 определяет пару (x_0, y_0) , C_1 определяет пару (x_1, y_1) , C_3 определяет пару (x_2, y_2) , т. е. нуклеотиды определяют пары (x_i, y_i) по правилу (7), при этом порядок нуклеотидов в кодоне задается правилом $2 > 1 > 3$, другими словами, более важные нуклеотиды в кодоне задают большие по 2-адической норме слагаемые в (x, y) .

На полученной 2-адической двумерной плоскости можно различными способами задать 2-адическую двумерную метрику. Нам будет удобен следующий выбор метрики:

$$d(U, V) = \max(|x - z|_2, q|y - t|_2), \quad 1/2 < q < 1. \quad (9)$$

Здесь $U = (x, y)$, $V = (z, t)$ суть точки на 2-адической плоскости.

Для визуализации полученной параметризации пространства кодонов естественно использовать отображение типа Монна вида

$$\begin{aligned} \eta: x_0 + 2x_1 + 4x_2 &\mapsto 1 + 4x_0 + 2x_1 + x_2, \\ \eta: y_0 + 2y_1 + 4y_2 &\mapsto 1 + 4y_0 + 2y_1 + y_2. \end{aligned}$$

После такого отображения кодоны будут расположены в квадрате 8×8 (где строки и столбцы нумеруются образом отображения η для координат x и y) следующим образом:

AAA	AAG	GAA	GAG	AGA	AGG	GGA	GGG
AAU	AAC	GAU	GAC	AGU	AGC	GGU	GGC
UAA	UAG	CAA	CAG	UGA	UGG	CGA	CGG
UAU	UAC	CAU	CAC	UGU	UGC	CGU	CGC
AUA	AUG	GUA	GUG	ACA	ACG	GCA	GCG
AUU	AUC	GUU	GUC	ACU	ACC	GCU	GCC
UUA	UUG	CUA	CUG	UCA	UCG	CCA	CCG
UUU	UUC	CUU	CUC	UCU	UCC	CCU	CCC

Применяя к этой таблице генетический код, т. е. сопоставляя кодону отвечающую ему аминокислоту, мы получим таблицу аминокислот а 2-адической плоскости [40]

$\frac{\text{Lys}}{\text{Asn}}$	$\frac{\text{Glu}}{\text{Asp}}$	$\frac{\text{Ter}}{\text{Ser}}$	Gly
$\frac{\text{Ter}}{\text{Tyr}}$	$\frac{\text{Gln}}{\text{His}}$	$\frac{\text{Trp}}{\text{Cys}}$	Arg
$\frac{\text{Met}}{\text{Ile}}$	Val	Thr	Ala
$\frac{\text{Leu}}{\text{Phe}}$	Leu	Ser	Pro

В частности, кодоны AAA, AAG отображаются в Lys, кодоны AAU, AAC отображаются в Asp, кодоны CCA, CCG, CCU, CCC отображаются в Pro, и прообразы этих аминокислот являются шарами относительно метрики (9).

Здесь был использован митохондриальный генетический код (версий генетического кода несколько, они отличаются значениями на нескольких кодонах). Из этой таблицы видно, что большая часть вырождения генетического кода отвечает локальному постоянству отображения (8) 2-адического аргумента на малых расстояниях (по одной либо по двум координатам). Такие области постоянства отображения (8) генетического кода (из двух или четырех кодонов) являются шарами относительно метрики (9) диаметров соответственно $q/4$ или $1/4$. Имеются три исключения из этого правила — прообразы аминокислот Leu, Ser и стоп-кодона Ter представляют собой пары шаров.

Также можно на 2-адической плоскости выделить гидрофобные аминокислоты

—	—	—	
—	—	$\frac{\text{Trp}}{\text{Cys}}$	
$\frac{\text{Met}}{\text{Ile}}$	Val		
$\frac{\text{Leu}}{\text{Phe}}$	Leu		

В данной таблице по сравнению с предыдущей удалены символы аминокислот, не являющихся гидрофобными. Мы видим, что гидрофобные аминокислоты кластеризуются относительно двумерной 2-адической метрики (9) (попадают в два шара диаметров $1/2$ и $1/4$), т. е. такая параметризация связана с химическими свойствами аминокислот.

7. ЗАКЛЮЧЕНИЕ

В настоящей работе дан краткий обзор некоторых применений p -адических и более общих ультраметрических методов в анализе данных, статистической физике неупорядоченных систем, динамике макромолекул и биологии. Кроме того, эти методы использовались в теории струн, физике высоких энергий, космологии и других областях [6], [1], [47]–[52] (см. также обзор [2]). Поскольку ультраметрические методы уже нашли различные применения, это позволяет надеяться на дальнейшее расширение области приложений. В частности, актуальные в настоящее время методы мультикластеризации (ансамблевой кластеризации) [34] могут рассматриваться как проявление многомерной иерархии данных, которая может применяться для математического моделирования в области анализа данных и искусственного интеллекта.

Список литературы

[1] В. С. Владимиров, И. В. Волович, Е. И. Зеленов, *p-Адический анализ и математическая физика*, Наука, М., 1994.
[2] B. Dragovich, A. Yu. Khrennikov, S. V. Kozyrev, I. V. Volovich, *p-Adic Numbers Ultrametric Anal. Appl.*, 1:1 (2009), 1–17, arXiv:0904.4205.

- [3] A. N. Kochubei, *Pseudo-Differential Equations and Stochastics Over Non-Archimedean Fields*, Marcel Dekker, New York, 2001.
- [4] V. Anashin, A. Khrennikov, *Applied Algebraic Dynamics*, de Gruyter Expositions in Mathematics, **49**, De Gruyter, Berlin, 2009.
- [5] S. Albeverio, A. Yu. Khrennikov, V. M. Shelkovich, *Theory of p -Adic Distributions*, London Mathematical Society Lecture Note Series, **370**, Cambridge Univ. Press, Cambridge, 2010.
- [6] I. V. Volovich, *Class. Quantum Grav.*, **4**:4 (1987), L83–L87.
- [7] M. Mezard, G. Parisi, M. Virasoro, *Spin-Glass Theory and Beyond*, World Scientific Lecture Notes in Physics, **9**, World Sci., Singapore, 1987.
- [8] V. A. Avetisov, A. H. Bikulov, S. V. Kozyrev, *J. Phys. A: Math. Gen.*, **32**:50 (1999), 8785–8791, arXiv:cond-mat/9904360.
- [9] G. Parisi, N. Sourlas, *Eur. Phys. J. B*, **14**:3 (2000), 535–542, arXiv:cond-mat/9906095.
- [10] A. Yu. Khrennikov, S. V. Kozyrev, *Physica A: Stat. Mech. Appl.*, **359** (2006), 222–240, arXiv:cond-mat/0603685.
- [11] A. Yu. Khrennikov, S. V. Kozyrev, *Physica A: Stat. Mech. Appl.*, **359** (2006), 241–266, arXiv:cond-mat/0603687.
- [12] A. Yu. Khrennikov, S. V. Kozyrev, *Physica A: Stat. Mech. Appl.*, **378**:2 (2007), 283–298, arXiv:cond-mat/0603694.
- [13] D. M. Carlucci, C. De Dominicis, *C. R. Acad. Sci., Ser. IIB: Mech. Phys. Chem. Astr.*, **325**:9 (1997), 527–530, arXiv:cond-mat/9709200.
- [14] C. De Dominicis, D. M. Carlucci, T. Temesvári, *Journal de Physique I (France)*, **7**:1 (1997), 105–115, arXiv:cond-mat/9703132.
- [15] M. Talagrand, *Spin glasses, a Challenge for Mathematicians. Cavity and Mean Field Models*, A Series of Modern Surveys in Mathematics, **46**, Springer, Berlin, 2003.
- [16] O. M. Becker, M. Karplus, *J. Chem. Phys.*, **106**:4 (1997), 1495–1517.
- [17] A. T. Ogielski, D. L. Stein, *Phys. Rev. Lett.*, **55**:15 (1985), 1634–1637.
- [18] V. A. Avetisov, A. H. Bikulov, S. V. Kozyrev, V. A. Osipov, *J. Phys. A: Math. Gen.*, **35**:2 (2002), 177–189, arXiv:cond-mat/0106506.
- [19] В. А. Аветисов, А. Х. Биколов, А. П. Зубарев, *Тр. МИАН*, **285** (2014), 9–32.
- [20] V. A. Avetisov, A. Kh. Bikulov, V. Al. Osipov, *J. Phys. A: Math. Gen.*, **36**:15 (2003), 4239–4246.
- [21] V. A. Avetisov, A. Kh. Bikulov, A. P. Zubarev, *J. Phys. A: Math. Theor.*, **42**:8 (2009), 085003, 18 pp.
- [22] S. V. Kozyrev, *p -Adic Ultrametric Anal. Appl.*, **2**:2 (2010), 122–132.
- [23] S. Albeverio, W. Karwowski, *Stochastic Processes Appl.*, **53**:1 (1994), 1–22.
- [24] S. K. Nechaev, O. A. Vasiliev, *Тр. МИАН*, **245** (2004), 182–201, arXiv:cond-mat/0310079.
- [25] В. А. Аветисов, А. Х. Биколов, А. П. Зубарев, *Вестн. Сам. гос. техн. ун-та. Сер. Физ.-матем. науки*, **1(22)** (2011), 9–15.
- [26] В. А. Аветисов, В. А. Иванов, Д. А. Мешков, С. К. Нечаев, *Письма в ЖЭТФ*, **98**:4 (2013), 270–274.
- [27] F. Murtagh, *Multidimensional Clustering Algorithms*, Compstat Lectures, **4**, Physica, Heidelberg, 1985.
- [28] F. Murtagh, P. Contreras, *p -Adic Numbers, Ultrametric Anal. Appl.*, **4**:1 (2012), 46–56.
- [29] F. Murtagh, *p -Adic Numbers, Ultrametric Anal. Appl.*, **5**:4 (2013), 326–337.
- [30] V. A. Lemin, “Finite ultrametric spaces and computer science”, *Categorical Perspectives* (Kent State University, Kent, OH, USA, August 1998), eds. J. Koslowski, A. Melton, Birkhäuser, Boston, MA, 2001, 219–241.
- [31] M. Greenfield, M. Marcolli, K. Teh, *p -Adic Numbers, Ultrametric Anal. Appl.*, **6**:2 (2014), 81–104.
- [32] M. Marcolli, N. Tedeschi, *p -Adic Numbers, Ultrametric Anal. Appl.*, **6**:2 (2014), 135–154.

- [33] E. Yu. Lerner, M. D. Missarov, *Lett. Math. Phys.*, **22**:2 (1991), 123–129.
- [34] A. Strehl, J. Ghosh, *J. Mach. Learn. Res.*, **3**:3 (2002), 583–617.
- [35] С. В. Козырев, *ТМФ*, **164**:3 (2010), 394–400.
- [36] S. Albeverio, S. V. Kozyrev, *p-Adic Numbers, Ultrametric Anal. Appl.*, **4**:3 (2012), 167–178, arXiv: 1204.5952.
- [37] С. В. Козырев, *ТМФ*, **180**:2 (2014), 234–244, arXiv: 1404.6960.
- [38] B. Dragovich, A. Yu. Dragovich, *p-Adic Numbers, Ultrametric Anal. Appl.*, **1**:1 (2009), 34–41, arXiv: q-bio/0607018.
- [39] B. Dragovich, A. Dragovich, *Computer J.*, **53**:4 (2010), 432–442, arXiv: 0707.3043.
- [40] A. Yu. Khrennikov, S. V. Kozyrev, *Physica A: Stat. Mech. Appl.*, **381** (2007), 265–272, arXiv: q-bio.QM/0701007.
- [41] A. Yu. Khrennikov, S. V. Kozyrev, *J. Theoret. Biol.*, **261**:3 (2009), 396–406, arXiv: 0903.0137.
- [42] С. В. Козырев, А. Ю. Хренников, *Докл. РАН*, **430**:1 (2010), 32–34.
- [43] A. Yu. Khrennikov, S. V. Kozyrev, *p-Adic Numbers, Ultrametric Anal. Appl.*, **3**:2 (2011), 165–168.
- [44] A. Yu. Khrennikov, S. V. Kozyrev, “2-Adic degeneration of the genetic code and energy of binding of codons”, *Quantum Bio-Informatics III* (Tokyo University of Science, Japan, 11–14 March, 2009), QP-PQ: Quantum Probability and White Noise Analysis, **26**, eds. L. Accardi, W. Freudenberg, M. Ohya, World Sci., Signapore, 2010, 193–204.
- [45] С. В. Козырев, *Методы и приложения ультраметрического и p -адического анализа: от теории всплесков до биофизики*, Совр. пробл. матем., **12**, МИАН, М., 2008.
- [46] A. Monna, *Analyse non-Archimédienne*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **56**, Springer, New York, 1970.
- [47] Yu. I. Manin, “Reflections on arithmetical physics”, *Conformal Invariance and String Theory*, eds. P. Dita, V. Gergescu, Academic Press, Boston, MA, 1989, 293–303.
- [48] I. Ya. Aref’eva, B. Dragovich, P. H. Frampton, I. V. Volovich, *Internat. J. Modern Phys. A*, **6**:24 (1991), 4341–4358.
- [49] L. Brekke, P. G. O. Freund, *Phys. Rep.*, **233**:1 (1993), 1–66.
- [50] В. С. Владимиров, “Адельные формулы для 4-частичных древесных струнных и суперструнных амплитуд в одноклассных квадратичных полях”, *Избранные вопросы p -адической математической физики и анализа*, Сборник статей. К 80-летию со дня рождения академика Василия Сергеевича Владимирова, Тр. МИАН, **245**, Наука, М., 2004, 9–28.
- [51] V. S. Varadarajan, “Arithmetic Quantum Physics: Why, What, and Whither”, *Избранные вопросы p -адической математической физики и анализа*, Сборник статей. К 80-летию со дня рождения академика Василия Сергеевича Владимирова, Тр. МИАН, **245**, Наука, М., 2004, 273–280.
- [52] A. Yu. Khrennikov, *Non-Archimedean Analysis: Quantum Paradoxes, Dynamical Systems and Biological Models*, Mathematics and Its Applications, **427**, Kluwer, Dordrecht, 1997.

Поступила в редакцию 24.04.2015,
после доработки 6.05.2015