С.В. Козырев

p-Адические числа и сложные системы

Конференция профессоров РАН по ОМН РАН
14, 15, 16 июня 2016, МИАН

Иерархические методы для сложных систем
Иерархия — деревья, билдинги, всплески (вейвлеты), ультраметрический анализ, p-адические числа.

Многомерные иерархические методы в анализе данных:

Кластеризация — деревья,
Мультикластеризация — сети, билдинги Брюа-Титса и аналоги.

Генетический код — 2-адическая плоскость.

Другие приложения:
Спиновые стёкла, динамика белка, упаковка ДНК.

p-Адические числа

Поле p-адических чисел $\mathbb{Q}_p$ — пополнение поля рациональных чисел по норме

$$|x|_p = p^{-j}, \quad x = p^j \frac{a}{b}, \quad |0|_p = 0.$$

p-Адические числа — ряды вида

$$x = \sum_{l=a}^{\infty} x_l p^l, \quad x_l \in \{0, 1, \ldots, p-1\}.$$

З.И.Боревич, И.Р.Шафаревич, Теория чисел, 3-е изд., М.: Наука, 1985.

В.С.Владимиров, И. В.Волович, Е.И.Зеленов, p-Адический анализ и математическая физика, Наука, М., 1994.

Ультраметрическое пространство: сильное неравенство треугольника для ультраметрики $d(\cdot,\cdot)$

$$d(x,y) \le \max(d(x,z), d(z,y)), \qquad \forall x,y,z \in X.$$

Свойства неархимедовой геометрии:
все треугольники равнобедренны;
шары не пересекаются либо один содержит другой;
любая точка шара есть его центр.
Дерево шаров в (локально компактном) ультраметрическом пространстве —
шары (ненулевого диаметра либо изолированные точки) являются вершинами,
рёбра — пары шаров вида (шар, максимальный подшар).

Можно обсуждать приложения ультраметрических пространств для описании иерархии в различных задачах.

С. В. Козырев, Ультраметрика в теории сложных систем, ТМФ, 185:2 (2015), 346–360

Шары в $\mathbb{Q}_p$ имеют вид

$$p^j\left(n+\mathbb{Z}_p\right), \qquad n=\sum_{l=a}^{-1} n_l p^l, \quad n_l \in \{0,1,\ldots,p-1\},$$

j целое,
a целое отрицательное,
$\mathbb{Z}_p$ есть кольцо целых p-адических чисел (шар $|x|_p \le 1$),
n — параметр на $\mathbb{Q}_p/\mathbb{Z}_p$ – нумерует единичные шары.
$p^j n$ – нумерует шары ненулевого диаметра.

Дерево шаров в $\mathbb{Q}_p$: в каждом шаре p максимальных подшаров в соответствии с разложением

$$\mathbb{Z}_p=\bigcup_{m=0}^{p-1}\left(m+p\mathbb{Z}_p\right).$$

Базис вещественных всплесков в $L^2(\mathbb{R})$.

Базис всплесков имеет параметризацию сдвигами и растяжениями (фиксированной функции либо конечного набора функций)

$$\psi_{jn}(x) = 2^{j/2}\psi\left(2^j x - n\right), \qquad x \in \mathbb{R}, \quad j, n \in \mathbb{Z}.$$

Функция $\psi(x)$ – всплеск (вейвлет). Пример – всплеск Хаара (1909) (разность двух характеристических функций)

$$\psi(x) = \chi_{[0,1/2)}(x) - \chi_{[1/2,1]}(x).$$

Более общие базисы всплесков, кратномасштабный анализ, S.Mallat, Y.Meyer, I.Daubechies, см.

I. Daubechies, Ten Lectures on Wavelets, SIAM, 1992.

Теорема: 1) **Базис p-адических всплесков** в $L^2(\mathbb{Q}_p)$:

$$\psi_{k;jn}(x) = p^{j/2}\psi_k\left(p^{-j}x - n\right),$$

$$x \in \mathbb{Q}_p, \quad j \in \mathbb{Z}, \quad n \in \mathbb{Q}_p/\mathbb{Z}_p, \quad k \in \{1, \ldots, p-1\}.$$

$$\psi_k(x) = \psi(kx), \qquad \psi(x) = \chi(p^{-1}x)\Omega(|x|_p),$$

где $\Omega(x)$ есть характеристическая функция $[0,1]$
(то есть $\Omega(|x|_p)$ есть характеристическая функция $\mathbb{Z}_p$),
χ есть характер

$$\chi(x) = e^{2\pi i\{x\}}, \qquad \{x\} = \sum_{l=a}^{-1} x_l p^l, \quad x = \sum_{l=a}^{\infty} x_l p^l.$$

2) p-Адические всплески — собственные вектора оператора Владимирова дробного дифференцирования

$$D^{\alpha}\psi_{k;\,jn}=p^{\alpha(1-j)}\psi_{k;\,jn}.$$

$$D^{\alpha}f(x)=\Gamma_p^{-1}(-\alpha)\int_{\mathbb{Q}_p}\frac{f(x)-f(y)}{|x-y|_p^{1+\alpha}}d\mu(y),\quad \alpha>0,$$

$$\Gamma_p(-\alpha)=\frac{p^{\alpha}-1}{1-p^{-1-\alpha}}.$$

С.В.Козырев, Теория всплесков как p-адический спектральный анализ. Известия РАН Серия Мат. 2002. Т.66. N.2. С.149–158, arXiv:math-ph/0012019

Спектры более общих операторов. Оператор на локально компактном ультраметрическом пространстве X с борелевской мерой μ

$$Tf(x)=\int_X t(\sup(x,y))(f(x)-f(y))d\mu(y).$$

$\sup(x,y)$ есть минимальный шар, содержащий точки x, y.

Собственные вектора – ультраметрические всплески (базис ψ_{Ij} из локально постоянных функций с нулевым средним с носителями во всевозможных шарах I), собственные значения

$$T\psi_{Ij}=\lambda_I\psi_{Ij},$$

$$\lambda_I=t(I)\mu(I)+\sum_{J>I}t(J)(\mu(J)-\mu(J,I)),$$

(J,I) – максимальный подшар в J, содержащий I.

Отображение Монна (связь вещественных и p-адических всплесков Хаара)

$$\mathbb{Q}_p \to \mathbb{R}_+,$$

$$\mathbb{Q}_p/\mathbb{Z}_p \to \mathbb{Z}_+,$$

$$\sum_{l=a}^{\infty} x_l p^l \mapsto \sum_{l=a}^{\infty} x_l p^{-l-1}.$$

Малые p-адические расстояния отображаются на малые вещественные. Взаимно однозначно почти всюду и сохраняет меру.

$p = 2$: Базис Хаара на положительной полуоси переходит в 2-адический базис всплесков почти всюду.
Вещественный базис всплесков имеет иерархическую структуру.

Другое применение отображения Монна – p-адическая параметризация матрицы Паризи в теории спиновых стёкол.

Фреймы всплесков как орбиты

Действие аффинной группы

$$f(x) \mapsto |a|_p^{-\frac{1}{2}} f\left(\frac{x-b}{a}\right).$$

Орбита всплеска $\psi(x) = \chi(p^{-1}x)\Omega(|x|_p)$ совпадает с набором произведений векторов из базиса $\psi_{k;jn}$ на корни степени p из единицы.

Параметризация базиса всплесков возникает автоматически. Обобщение — фреймы всплесков как орбиты основных функций (локально постоянных с компактным носителем) с нулевым средним.

Фреймом $\{e_n\}$ в гильбертовом пространстве $\mathcal{H}$ называется набор векторов в $\mathcal{H}$, такой, что: $\exists A, B > 0$: $\forall g \in \mathcal{H}$

$$A\|g\|^2 \leq \sum_n |\langle g, e_n\rangle|^2 \leq B\|g\|^2.$$

Метрики в многомерных p-адических пространствах

Норма в $\mathbb{Q}_p^d$

$$N_{q_1,\ldots,q_d}(z) = \max_{i=1,\ldots,d}(q_i|z_i|_p), \qquad q_i > 0.$$

Норма (A-вращение $N_{q_1,\ldots,q_d}$, A есть матрица из $\mathrm{Gl}_d(\mathbb{Q}_p)$)

$$N^A_{q_1,\ldots,q_d}(z) = N_{q_1,\ldots,q_d}(Az).$$

Метрика определена нормой

$$s(x,y) = N(x-y).$$

Пример: $N_{q_1,\ldots,q_d}$, $p^{-1} < q_i \leq 1$, $\mathbb{Z}_p^d$ и $p\mathbb{Z}_p^d$ – шары, при не равных друг другу q_i появляются промежуточные шары между $\mathbb{Z}_p^d$ и $p\mathbb{Z}_p^d$.

Многомерные базисы всплесков как орбиты. Общая схема:
1) Метрика в $\mathbb{Q}_p^d$.
2) Группа автоморфизмов дерева шаров относительно метрики (линейные преобразования плюс сдвиги).
3) Базисы и фреймы всплесков как орбиты такой группы.

Пример: норма

$$\|z\|_p = \max_{i=1,\dots,d}(|z_i|_p).$$

Группа автоморфизмов дерева шаров — сдвиги; однородные растяжения; линейные преобразования, сохраняющие норму (матрицы с матричными элементами из $\mathbb{Z}_p$ и $|\det(\cdot)|_p = 1$.)

Всплески — произведения характеров $\mathbb{Q}_p^d$ на характеристические функции шаров, причём характер на шаре делает одно колебание. Таким образом, всплески нумеруются шарами и волновыми векторами (максимальными подшарами в единичной сфере).

Теорема Набор функций $\{\psi_{k;jn}\}$ является ортонормированным базисом в $L^2(\mathbb{Q}_p^d)$:

$$\psi_k(x) = \chi(p^{-1}k \cdot x)\Omega(\|x\|_p), \qquad x \in \mathbb{Q}_p^d, \quad k \cdot x = \sum_{l=1}^{d} k_l x_l,$$

$$k = (k_1, \ldots, k_d), \qquad k_l = 0, \ldots, p-1,$$

где хотя бы один из k_l не равен нулю;

$$\psi_{k;jn}(x) = p^{-\frac{dj}{2}}\psi_k(p^j x - n), \qquad x \in \mathbb{Q}_p^d, \quad j \in \mathbb{Z}, \quad n \in \mathbb{Q}_p^d/\mathbb{Z}_p^d,$$

$$n = \left(n^{(1)}, \ldots, n^{(d)}\right), \quad n^{(l)} = \sum_{i=\beta_l}^{-1} n_i^{(l)} p^i,$$

$$n_i^{(l)} = 0, \ldots, p-1, \quad \beta_l \in \mathbb{Z}_-.$$

Альтернативные выборы метрики
(q_i, не равные друг другу, $p^{-1} < q_i \leq 1$) —
появляются промежуточные шары между $\mathbb{Z}_p^d$ и $p\mathbb{Z}_p^d$ —
базисы всплесков с матричными растяжениями

$$A = \left(\begin{array}{ccccc} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \vdots & & & & \vdots \\ 0 & 0 & \dots & 0 & 1 \\ p & 0 & \dots & 0 & 0 \end{array} \right),$$

Quincunx basis ($p = 2$)

$$Q = \left(\begin{array}{cc} 1 & -1 \\ 1 & 1 \end{array} \right).$$

Пример применения 2-мерной 2-адической метрики: генетический код на 2-адической плоскости.

Кодоны (тройки нуклеотидов) кодируют аминокислоты.
64 кодона, 20 аминокислот – код вырожден.
Набор всех кодонов (64 элемента) естественным образом нумеруется 2-адической плоскостью 8×8 с 2-мерной 2-адической метрикой

$$d_{1,q}((x,y),(x',y')) = \max(|x-x'|_2, q|y-y'|_2), \quad 1/2 < q < 1,$$

$$x = (x_0 x_1 x_2) = x_0 + 2x_1 + 4x_2, \quad x_i = 0, 1,$$

$$y = (y_0 y_1 y_2) = y_0 + 2y_1 + 4y_2, \quad y_i = 0, 1.$$

Применение к такой таблице кодонов 8×8 митохондриального генетического кода даёт параметризацию кода, относительно которой близкие в метрике $d_{1,q}$ кодоны отображаются в одинаковые аминокислоты, а немного менее близкие – в похожие (гидрофобные аминокислоты кластеризуются в этой метрике):

$\frac{\text{Lys}}{\text{Asn}}$	$\frac{\text{Glu}}{\text{Asp}}$	$\frac{\text{Ter}}{\text{Ser}}$	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

Кластеризация – иерархическая классификация данных.

Данные размечаются иерархической системой (частично упорядоченным деревом) кластеров. В типичном случае кластеры строятся по метрике на данных.

Мультикластеризация — несколько систем кластеров на тех же данных. Несколько метрик, различные метрики порождают различные деревья кластеров.
Сети кластеров с циклами — циклы возникают, когда кластеры относительно разных метрик совпадают как множества.

A.Strehl, J.Ghosh, C.Cardie, Cluster ensembles — a knowledge reuse framework for combining multiple partitions. Journal of Machine Learning Research, 2002. 3. P.583–617.

p-Адический случай: сети кластеров связаны с аффинными билдингами Брюа–Титса.

Кластеризация по ближайшему соседу:
(M, ρ) – метрическое пространство.

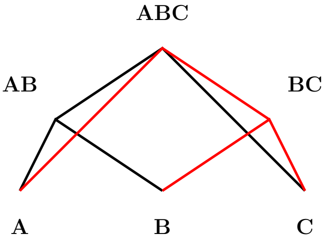
$a = x_0, x_1, \ldots, x_{n-1}, x_n = b$ — ε-цепь в пространстве (M, ρ).

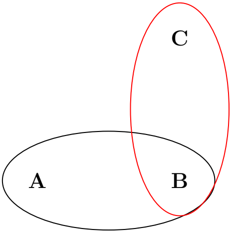
Цепное расстояние между a и b:
$d(a, b) = \inf(\varepsilon$: существует ε-цепь между a, $b)$.

все свойства ультраметрики кроме невырожденности (несовпадающие точки могут иметь нулевое цепное расстояние).
Цепное расстояние определяет ультраметрику на множестве классов эквивалентности точек в M, где a, b лежат в одном классе эквивалентности, если $d(a, b) = 0$.

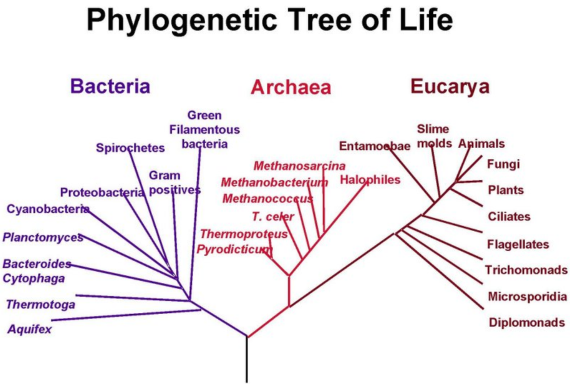
Кластер — шар относительно цепного расстояния.
Кластеризация — покрытие пространства (M, ρ) кластерами.

Мультикластеризация — несколько метрик на M, несколько деревьев кластеров. Деревья склеиваются в сеть.


ABC
AB
BC
A
B
C

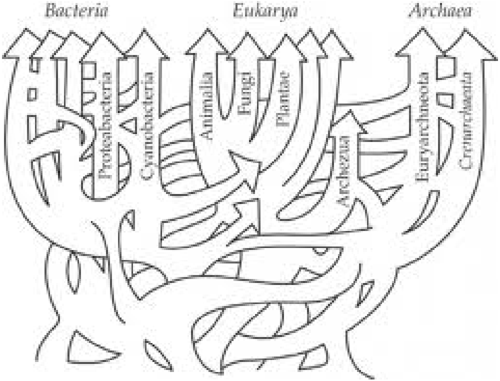

C
A
B

Кластеризация в науках о жизни: дерево жизни
Карл Линней, Systema Naturae, 1735



Рибосомное дерево жизни
Карл Вёзе, 1977, 1985

Филогенетическая сеть



Е.В. Кунин, Логика случая. О природе и происхождении биологической эволюции, Центрполиграф, 2014

Циклы — связь с мультикластеризацией?

Мультикластеризация в p-адических пространствах $\mathbb{Q}_p^d$: связь с аффинными билдингами Брюа–Титса (симплициальными комплексами классов эквивалентности решёток).

Шары в $\mathbb{Q}_p^d$, содержащие нуль — решётки (открытые компактные $\mathbb{Z}_p$-модули). Симплициальный комплекс шаров относительно мультикластеризации по семейству метрик — шары, не обязательно содержащие нуль. Ограничение на шары с центром в нуле и факторизация по растяжениям на степени p даёт аффинный билдинг.

Общий случай мультикластеризации по семейству метрик: можно определить структуру симплициального комплекса и соответствующее понятие размерности.
Таким образом, можно рассмотреть аналоги билдингов и многомерной иерархии в общих конструкциях анализа данных.

Заключение

p-Адические всплески — деревья шаров, группы автоморфизмов таких деревьев, базисы всплесков — системы когерентных состояний. Всплески как собственные вектора псевдодифференциальных операторов.

Многомерные иерархические системы.

Пример многомерной иерархии — 2-адическая плоскость генетического кода.

Анализ данных — кластеризация, мультикластеризация, сети кластеров, аналогия с аффинными билдингами, размерность общих сетей кластеров.