МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФЕДЕРАЛЬНОЕ АГЕНТСТВО ПО АТОМНОЙ ЭНЕРГИИ

РОССИЙСКАЯ АКАДЕМИЯ НАУК

РОССИЙСКАЯ АССОЦИАЦИЯ НЕЙРОИНФОРМАТИКИ

МОСКОВСКИЙ ИНЖЕНЕРНО-ФИЗИЧЕСКИЙ ИНСТИТУТ
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ)

ИНСТИТУТ ОПТИКО-НЕЙРОННЫХ ТЕХНОЛОГИЙ РАН

НАУЧНАЯ СЕССИЯ МИФИ–2007

НЕЙРОИНФОРМАТИКА–2007

**IX ВСЕРОССИЙСКАЯ
НАУЧНО-ТЕХНИЧЕСКАЯ
КОНФЕРЕНЦИЯ**

ЛЕКЦИИ
ПО НЕЙРОИНФОРМАТИКЕ

Часть 1

По материалам Школы-семинара
«Современные проблемы нейроинформатики»

Москва 2007

УДК 001(06)+004.032.26 (06) Нейронные сети
ББК 72я5+32.818я5
М82

НАУЧНАЯ СЕССИЯ МИФИ–2007. IX ВСЕРОССИЙСКАЯ НАУЧНО-ТЕХНИЧЕСКАЯ КОНФЕРЕНЦИЯ «НЕЙРОИНФОРМАТИКА–2007»: ЛЕКЦИИ ПО НЕЙРОИНФОРМАТИКЕ. Часть 1. – М.: МИФИ, 2007. – 178 с.

В книге публикуются тексты лекций, прочитанных на Школе-семинаре «Современные проблемы нейроинформатики», проходившей 24–26 января 2007 года в МИФИ в рамках IX Всероссийской конференции «Нейроинформатика–2007».

Материалы лекций связаны с рядом проблем, актуальных для современного этапа развития нейроинформатики, включая ее взаимодействие с другими научно-техническими областями.

Ответственный редактор
Ю. В. Тюменцев, кандидат технических наук

ISBN 5–7262–0708–4

© *Московский инженерно-физический институт (государственный университет), 2007*

Содержание

***Н. Г. Макаренко*. Стохастическая динамика, марковские модели и прогноз** **52**
Введение 53
Метрика и метрические пространства 58
Пространство компактов и IFS 61
Пространство кодов 66
Меры 70
Пространство мер: метрика Монжа-Канторовича 77
Обратная задача и IFS 81
Предсказание магнитных бурь 84
Заключение 90
Литература 93

Н. Г. МАКАРЕНКО
Главная астрономическая обсерватория РАН, Санкт-Петербург;
Институт математики, Алма-Ата, Казахстан
E-mail: ng-makar@mail.ru, makarenko@math.kz

СТОХАСТИЧЕСКАЯ ДИНАМИКА, МАРКОВСКИЕ МОДЕЛИ И ПРОГНОЗ

Аннотация

Лекция представляет новый метод марковского предсказания временных рядов. Он основан на инвариантной мере случайной динамики, которая реализуется Системой Итеративных Функций (IFS) — сжимающих отображений, снабженных вероятностями. Параметры этой системы получаются в результате решения обратной задачи на основе оценки эмпирической меры из временных рядов.

N. G. MAKARENKO
Pulkovo Astronomical Observatory, St-Petersburg;
Institute of Mathematics, Kazakhstan, Alma-Ata
E-mail: ng-makar@mail.ru, makarenko@math.kz

STOCHASTIC DYNAMICS, MARKOVIAN MODELS AND FORECAST

Abstract

The lecture gives a new method of Markovian prediction of time series. It is based on invariant measure of stochastic dynamics, which is implemented as Probabilistic Iterated Function System (PIFS), i. e. a set of contractive maps with probabilities. Parameters of the system are obtained as a result of the inverse problem solution using the time series empirical measure.

УДК 001(06)+004.032.26 (06) Нейронные сети

Введение

> Сто тысяч миллионов до неба,
> умноженные на отсюда и до фига
> дают представление об идее,
> которую мы попытались до вас
> донести.
>
> *Дуглас Адамс*
> *«Автостопом по Галактике»*

В силу сложившихся традиций мы обычно пытаемся упорядочить свои впечатления о наблюдаемом Мире в форме моделей. Самые простые из них позволяют интерпретировать эксперимент в форме схемы, логической или математической, редуцируя наблюдаемое многообразие Фактов к возможно меньшему числу параметров порядка. Такие модели-интерпретаторы отвечают, или точнее, формализуют вопрос: *«Как это происходит?»* Более продвинутые модели пытаются ответить на вопрос: *«Почему это происходит именно так, а не иначе?»* В любом варианте мы хотим, чтобы модель была не только «объяснительной», но и позволяла бы увидеть чуть дальше «кончика носа» — интервала времени, по которому она строилась. Иными словами, модель должна предсказывать.

Предсказывать поведение динамической системы, вообще говоря, интересно как вперед (Prediction), так и назад во времени (Postdiction). Однако, в присутствии диссипации и конечной точности начальных данных реализовать Postdiction, по-видимому, гораздо сложнее, если это вообще можно сделать. Во-первых, существует слишком много возможных сценариев *Прошлого*, которые могли в принципе закончится существующим *Настоящим*, и все они для нас *равновозможны*. Во-вторых, инверсия даже линейного предиктора приводит к NP-сложным алгоритмам [1].

Современная техника детерминированного нелинейного Prediction временных рядов, о которой я подробно рассказывал в одной из своих прошлых лекций [2], основана на топологических моделях, которые получаются из наблюдений. Реконструкция представляет собой дифференцируемое вложение временного ряда в евклидово пространство подходящей размерности. При этом сам временной ряд рассматривается как *платоновская тень* фазовой траектории аттрактора системы на произвольное направление. Символы веры, необходимые для корректной реализации такой процедуры, сводятся к предположению о существовании абстрактной гладкой

динамической системы с компактным низкоразмерным аттрактором. Предиктор строится по полученной многомерной копии аттрактора.

Его локальный вариант называют *методом аналогов*. Этот метод формализует библейский тезис *«Что было то и будет»*[1] [2]. Другими словами, мы находим в *Прошлом* нашей реконструкции фрагменты траекторий, которые похожи (аналогичны) последнему известному нам фрагменту. После этого выписываем их динамические продолжения, т. е. их «будущее в прошлом». Локальное предсказание получается усреднением вариантов *déjà vu*[3].

Глобальный предиктор основан на всей непрерывной и известной истории ряда и представляет собой нелинейную авторегрессионную модель с памятью, равной размерности вложения. Фактически, мы здесь имеем дело с нелинейным отображением векторов реконструкции, известных из истории ряда, в их последующие по времени значения. Аппроксимация неизвестной нелинейной функции многих переменных может быть получена разными методами.

Искусственные нейронные сети, возможно, являются лучшим аппроксиматором, обученным такому отображению. Однако, в общем случае квадратичного функционала ошибок и конечной обучающей выборки задача аппроксимации не является корректной. Поэтому выбор варианта, претендующего на роль *реального* из множества альтернативных равновозможных решений, всегда является делом весьма интимным даже при наличии хорошего регуляризатора. Для выбора варианта нередко используют даже эстетический принцип: *«Возможно, это хорошо выглядит»*.

Совсем иная ситуация возникает, когда у нас нет оснований полагать, что динамика исходной системы является детерминированной. По разным причинам, возникают большие сомнения в выполнимости *Кредо идеального экспериментатора* [2], необходимого для топологической реконструкции. В этом случае приходится использовать стохастические модели и вероятностный прогноз.

Я позволю себе привести здесь цитату из замечательного эссе *Станислава Лема*, которая очень точно описывает двусмысленность недетерми-

[1]Если ссылка на примечание представляет собой число, заключенное в круглые скобки, например, (3), то она обозначает номер примечания, помещенного в конце данного текущего раздела. Ссылка в виде числа, не заключенного в скобки — обычное подстраничное примечание.

[2]Екл. 1:9.

[3]Уже виденное.

нированной ситуации[4]: *«К понятию вероятности мы прибегаем, когда не знаем чего-либо с полной уверенностью. Но неуверенность эта носит либо чисто субъективный характер (я не знаю, что произойдет, но кто-то другой, возможно, и знает), либо объективный (никто не знает и знать не может). Субъективная вероятность — это протез при информационном увечье; не зная, какая лошадь возьмет приз, и угадывая лишь по числу лошадей (если их четыре, то у каждой один шанс из четырех на победу), мы поступаем как слепой в комнате, заставленной мебелью. Вероятность подобна трости слепца, которой он нащупывает дорогу. Если бы он видел, то не нуждался бы в палке, а если бы я знал, какая лошадь резвее всех, то не нуждался бы в теории вероятностей»*.

К счастью, стохастические модели, основанные на *марковских* процессах, названных в честь А. А. Маркова (1), не зависят от того, что именно характеризует непредсказуемость будущих событий: наши знания о них или сами эти события. Марковский процесс строится на принципе: *будущее зависит от прошлого только через настоящее*.

На формальном языке условная вероятность наблюдать «завтра» некоторое значение x_{n+1}, при условии что ему предшествовала последовательность значений $x_n, x_{n-1}, \ldots, x_0$, редуцируется к условной вероятности, зависящей лишь от наблюдаемого «сегодняшнего» значения x_n:

$$p\big(x_{n+1}|x_n, x_{n-1}, \ldots, x_0\big) = p\big(x_{n+1}|x_n\big).$$

Простейший вариант марковского процесса реализуется AR-моделью, описывающей случайное смещение точки $Tx_n = (ax_n + \xi_n) \to x_{n+1}$, где случайная переменная ξ_n подчиняется гауссовскому распределению $\mathcal{N}(\xi, \sigma)$. Легко показать, что ее эволюция $x_n \to x_{n+1}$ полностью описывается условной (переходной) вероятностью

$$p\big(x_{n+1}|x_n\big) = p(x_{n+1}, x_n)/p(x_n),$$

где $p(x_{n+1}, x_n) = \mathcal{N}(x_{n-1} - ax_n, \sigma)$.

Рассмотренный пример иллюстрирует идею *случайной динамической системы*. Она получается, если вместо одного оператора T использовать случайно выбранные отображения из набора $T \equiv \{T_k\},\ k = 1, 2, \ldots, N$. Каждое $T_i : X \to X$ является непрерывным преобразованием на *компактном* подмножестве $X \in R^n$.

[4] *Ст. Лем.* О невозможности жизни. О невозможности прогнозирования / Вторжение с Альдебарана. – М.: Ин. Лит. 1960.

Случайная орбита такой системы — это просто последовательность точек

$$\{x_m\}_{m=0}^{\infty}: \quad x_m = T_{k_{m-1}} \circ \ldots \circ T_{k_1} \circ T_{k_0} x_0 \equiv T x_0,$$

полученных итеративным применением операторов из набора T. Каждый номер $k_i \in \{1, 2, \ldots, N\}$ может быть выбран независимо, либо с вероятностью, зависящей от номера k_{i-1} предшествующего отображения. В последнем случае мы получим как раз марковский процесс.

Разобьем фазовое пространство X для нашей случайной системы на конечное число клеток $\{A_1, \ldots, A_n\}$. Марковская модель задается $n \times n$ матрицей вероятностей

$$P_{ij} = \mathrm{Prob}_{\mu}\{Tx \in A_j | x \in A_i\},$$

описывающих переходы из клетки A_i в клетку A_j под действием итераций T. Основной проблемой является нахождение этих переходных вероятностей.

Вспомним, что вероятность имеет смысл только в контексте заданной вероятностной меры, которая помечена выше символом μ. К счастью оказывается, что для большинства орбит нашей случайной системы такая инвариантная мера существует! В частности, для каждой из клеток A_i она пропорциональна доле времени, которое проводит в клетке орбита, или, что то же самое, относительному числу точек орбиты, которые в ней «застряли».

Остается проблема поиска подходящих отображений T. Проще всего искать под фонарем, хотя иногда искомое может находиться под крышкой в кастрюле с супом! Такой кастрюлей в нашем случае является геометрия фракталов, о которой я рассказывал в своей лекции [3].

Вспомним, что фрактал является предельным образом рекуррентной динамики *Системы Итеративных Функций (`IFS`)* в пространстве компактов. Последовательные итерации сжатий приводят к единственной неподвижной точке — *фракталу* или *аттрактору `IFS`*. Если снабдить каждое из отображений вероятностями, случайная динамика `IFS` продуцирует единственную инвариантную (мультифрактальную) меру на аттракторе. Случайный алгоритм построения фракталов эквивалентен функционированию искусственной нейронной сети [4].

Предположим теперь, что нам задан аттрактор и мера на нем. Поиск `IFS` и их вероятностей составляет содержание *Обратной задачи* в теории фракталов. Она корректно поставлена и решается как оптимизационная

задача с ограничениями. Известные методы ее решения включают и генетический алгоритм.

Вот к такому необычному контексту приводит современный вариант марковского прогноза. Почти библейский случай: пошел *Саул, Кисов* сын, искать ослиц, а нашел *Царство*[5]. Попыткой понятного описания этого *Фрактального Царства* и является предлагаемая Лекция.

Я следовал форме моих прошлых лекций: каждый раздел сопровождается Примечаниями и Путеводителем по литературе. Последний раздел содержит, в качестве иллюстрации, практическое применение теории к предсказанию магнитных бурь. Эти результаты были получены совсем недавно моими коллегами *Л. М. Каримовой*, *О. А. Круглун* и *С. А. Мухамеджановой* и не были озвучены в устном варианте Лекции.

Я благодарен своей дочери и сотруднице *Ирине Макаренко*, без помощи которой текст вряд ли был бы когда-нибудь закончен.

Примечания

1. *Андрей Андреевич Марков-старший* (14.6.1856–20.7.1922) — выдающийся русский математик, отец математика *А. А. Маркова-младшего*. Родился в Рязани. В 1874 году поступил в Петербургский университет. В 1880 году защитил свою знаменитую магистерскую диссертацию «О бинарных квадратичных формах положительного определителя», а 1881 году — докторскую диссертацию. С 1880 года *А. А. Марков* — приват-доцент Петербургского университета. По предложению *Пафнутия Львовича Чебышева* избран адъюнктом (1886 г.), затем экстраординарным академиком (1890 г.)и ординарным академиком (1896 г.) В 1886 году А. А. Марков — экстраординарный профессор Петербургского университета, а в 1893 году — ординарный профессор. В 1905 году ушел в отставку в звании заслуженного профессора. В 1912 году по собственной просьбе был отлучен от церкви. В письме к Синоду он писал: «Я не усматриваю существенной разницы между иконами и идолами, которые, конечно, не боги, а их изображения».

[5]Книга царств 1:9.

Метрика и метрические пространства

МУ — единица расстояния в
Индии. Означает предел
слышимости мычания коровы.

Невероятно? Но факт!
Еш гвардия, Ташкент, 1971.

В этом разделе мы приведем некоторые общие определения, необходимые для работы в пространствах, которые мы будем использовать.

Метрикой в пространстве X называют функцию $d : X \times X \to R$. Эта запись означает, что каждой паре точек из $X \times X$ ставится в соответствие вещественное число d. Для всех $x, y \in X$ функция d удовлетворяет следующим требованиям:

1. $d(x, y) \geqslant 0$;
2. $d(x, y) = 0$ тогда и только тогда, когда $x = y$;
3. $d(x, y) = d(y, x)$ — свойство симметрии;
4. $d(x, z) \leqslant d(x, y) + d(y, z)$ — неравенство треугольника.

Если отменить условие (2), то мы получим *псевдометрику*, а если заменить свойство (4) на условие «остроугольности»

$$d(x, z) \leqslant \max\bigl(d(x, y), d(y, z)\bigr),$$

то получится *ультраметрика*. Она используется в некоторых задачах распознавания образов [5], когда пространства признаков имеют высокую размерность и приобретают необычные свойства (1).

Пара (X, d) называется *метрическим пространством*. Последовательность точек $x_1, x_2, \ldots, x_k, \ldots$ в (X, d) называют *фундаментальной*, если расстояния между точками уменьшаются, т. е. $d(x_i, x_j) \to 0$, когда $i, j \to \infty$. Если любая такая последовательность сходится к некоторой точке $x \in X$, т. е. $d(x_k, x) \to 0$, когда $k \to \infty$, то пространство (X, d) называют *полным*.

Полные метрические пространства имеют очень важное для нас свойство, которое выражается *теоремой Банаха* (2) или *Принципом сжимающих*

отображений: в полном метрическом пространстве любое сжимающее отображение имеет единственную неподвижную точку.

Напомним, что отображение $F : X \to R$ называют *сжимающим*, если существует такое число $r \in (0,1)$, что для любых двух точек $x_1, x_2 \in X$

$$d\big(F(x_1), F(x_2)\big) \leqslant r\, d(x_1, x_2).$$

Эквивалентным является следующее определение. Пусть (X, d) — полное метрическое пространство. Отображение $S : X \to X$ удовлетворяет *условию Липшица*, если для всех $x, y \in X$ существует такое число $c \geqslant 0$, что

$$d\big(S(x), S(y)\big) \leqslant c d(x, y).$$

Постоянной Липшица $Lip\, S$ для S называют минимальное c, для которого выполняется условие Липшица. Отображение S называют сжатием, если $Lip\, S < 1$.

Неподвижная точка a отображения F — это решение уравнения $F(x) = x$. Неподвижная точка притягивает последовательность точек, полученных итерациями F, т. е. начиная с любой начальной точки, последовательность

$$x,\ F(x),\ F^{\circ 2}(x) = F\big(F(x)\big),\ F^{\circ 3}(x) = F\Big(F\big(F(x)\big)\Big),\ \ldots$$

сходится к a. Используя неравенство треугольника (4), можно оценить скорость сходимости таких итераций к неподвижной точке.

Пусть $d(x, F(x)) \leqslant \varepsilon$ и r — коэффициент сжатия для F. Тогда

$$\begin{aligned} & d(x, F^{\circ k}(x)) \leqslant \\ & \leqslant d\big(x, F(x)\big) + d\big(F(x), F^{\circ 2}(x)\big) + \ldots + d\big(F^{\circ k-1}(x), F^{\circ k}(x)\big) \leqslant \\ & \leqslant d\big(x, F(x)\big)(1 + r + \ldots + r^{k-1}) \leqslant \varepsilon(1 - r^{k-1})/(1 - r). \end{aligned}$$

При $k \to \infty$ из последнего неравенства получаем оценку:

$$d(x,\ a) \leqslant \frac{\varepsilon}{1 - r}.$$

В стандартных курсах анализа теорема Банаха получается из утверждения, что график любой непрерывной функции, определенной в замкнутом интервале $[a, b]$ для всех значений $x \in [a, b]$, по меньшей мере один раз пересекает диагональ $x = g(x)$.

Действительно, график такой функции не может начинаться в a и заканчиваться в b, в противном случае они уже будут являться его неподвижными точками, и тогда уже нечего доказывать. Поэтому, полагая, например, что $g(a) \geqslant a$, $g(b) \leqslant b$, мы получим, что непрерывная функция $g(x) - x$ удовлетворяет на границах интервала неравенствам: $g(a) - a \geqslant 0$, $g(b) - b \leqslant 0$. Но тогда, по теореме о промежуточном значении, всегда найдется точка $c \in [a, b]$, такая что $g(c) - c = 0$. Если g — линейная функция с $|g'| < 1$, т. е. является сжатием, ее графиком является прямая, которая пересекает диагональ только в одной точке! Заметим, что в нашем примере метрическое пространство является полным — это замкнутый интервал. Рассмотрим теперь другие метрические пространства и Принцип сжимающих отображений в них.

Примечания

1. Объем n-мерного шара равен $V_n(r) = C_n r^n$. В этой формуле $C_n = \pi^{n/2}/\Gamma(n/2+1)$ и Γ — гамма-функция. Если взять например $n = 2k$, то получим $V_{2k}(r) = (\pi r^2)^k (k!)^{-1}$. Легко видеть, что при $k > \pi r^2$ увеличение k (т. е. размерности n) приводит к уменьшению объема! Доля объема в шаровом ε-слое будет равна

$$\frac{\textit{объем слоя}}{\textit{объем шара}} = \frac{C_n r^n - C_m (r-\varepsilon)^n}{C_n r^n} = 1 - \left(1 - \frac{\varepsilon}{r}\right)^n \to 1,$$

   т. е. стремится к 1 при возрастании n. Поэтому практически весь объем шара большой размерности сосредоточен вблизи его поверхности. Так что относительно любой точки внутри шара почти все пары точек выборки образуют остроугольные треугольники.

2. *Банах Стефан*, он же *Степан Степанович Гречек* (30.03.1892–31.08.1945) Родился в Кракове. Член Польской АН и член-корреспондент АН УССР, профессор Львовского университета. Один из создателей современного функционального анализа. Во время оккупации находился в нацистском институте, где разрабатывалась противотифозная сыворотка, а ученого использовали для кормления вшей. *Банах* один из авторов так называемой *Шкотской книги* — реликвии математического мира. Рукопись, написанная в пивной, содержала 193 важных математических проблемы, придуманных *Банахом*, *Мазуром*, *Штейнхаусом* и *Уламом* на встречах в *Шкотской кавьярне*

(г. Львов). Многие решения утеряны: авторы записывали их сперва химическим карандашом на мраморных столиках. Призами, в зависимости от сложности задачи, были: пять малых кружек пива, вино, ужин и живой гусь. Книга была переведена на английский язык и издана *Уламом* в 1957 году.

Путеводитель по литературе. Приведенные определения можно найти в любом курсе функционального анализа. Популярное введение в теорему Банаха и ее варианты содержит книга [6]. Пример парадокса с многомерным объемом взят из замечательной книги *Хемминга* [7]. Использование ультраметрики в различных задачах описано в статье *Муртага* [5]; другие работы на эту тему есть на его страничке.

Пространство компактов и IFS

Ничто не строится на камне, все на песке, но долг человеческий строить, как если бы камнем был песок.

Х. Л. Борхес
«Фрагменты апокрифического Евангелия»

Ограниченные и замкнутые множества называются *компактными* или просто *компактами*. Ограниченность означает, что такое множество можно поместить внутрь шара конечного радиуса, а замкнутое множество содержит в себе все свой предельные (граничные) точки. Компактами, например, являются замкнутые интервалы вида $[a, b]$ на вещественной оси, но не полуоткрытые и открытые интервалы, такие как $[a, b)$, $(a, b]$, (a, b).

Рассмотрим пространство H, точками которого являются компакты из R^n. Если использовать такие множества из R^2, то точками в H будут сегменты, квадраты, круги, треугольники и т. д. Расстояние в H измеряется с помощью *метрики Хаусдорфа* (1).

Рассмотрим два множества A и B. Обозначим через $d(x, B)$ расстояние самой удаленной от B точки $x \in A$, где удаленность понимается в смысле обычного евклидова расстояния:

$$d(x, B) = \sup_{x \in A} \inf_{y \in B} d(x, y).$$

Это выражение описывает две операции. Первая состоит в том, что для произвольной *фиксированной* точки $x \in A$ мы находим расстояния до всех точек $y \in B$ и выбираем среди них кратчайшее. Символически это записывается как $\inf\limits_{y \in B} d(x, y)$. Затем переходим в найденную, самую близкую точку множества $y \in B$ и ищем наиболее удаленную от нее точку $x \in A$, т. е. берем верхнюю грань $\sup\limits_{x \in A} d(x, y)$ расстояний для всех точек, в предположении, что точка y фиксирована. Аналогично, обозначим через $d(y, A)$ расстояние самой удаленной от A точки $y \in B$ (рис. 1). Оно получается из выражения: $d(y, A) = \sup\limits_{y \in B} \inf\limits_{x \in A} d(x, y)$. В результате мы получим два, вообще говоря, разных числа. Расстояние Хаусдорфа можно определить как максимальное из них:

$$d_H(A, B) = \max\Big\{d(x, B),\ d(y, A)\big|\ x \in A,\ y \in B\Big\}$$

или как сумму этих двух чисел:

$$d_H(A, B) = \sup_{x \in A} \inf_{y \in B} d(x, y) + \sup_{y \in B} \inf_{x \in A} d(x, y).$$

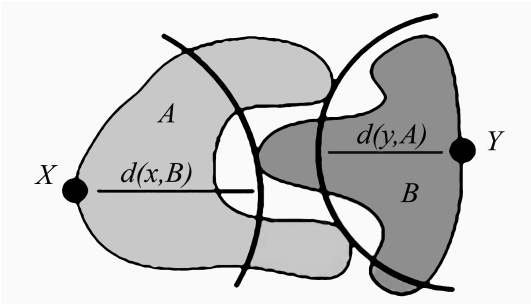

Рис. 1. Определение метрики Хаусдорфа

Существует и геометрический эквивалент этого определения. Покроем каждую точку множества A диском с радиусом ε. Возьмем объединение всех этих дисков. Те, что декорируют внутренние точки A совпадут с этим множеством. Диски с центрами на граничных точках A увеличат исходное множество на величину ε. Для того, чтобы уяснить это, представьте себе диск, центр которого движется по границе A и окрашивает

своей поверхностью полоску шириной ε, за счет которой и увеличивается площадь множества. Формально, то что получилось, записывают как множество точек удаленных от A на расстояние, не превышающее ε, т. е. $A_\varepsilon = \{x \mid d(x, A) \leqslant \varepsilon\}$. Множество A_ε называют *параллельным телом* для A или его *дилатацией*.

Будем увеличивать радиус диска до тех пор, пока дилатация A_ε не поглотит множество B, т. е. $B \subseteq A_\varepsilon$. Эта ситуация возникнет при некотором радиусе ε_1. Теперь вернемся к исходным множествам и, оставив A в покое, дилатируем B так, чтобы $A \subseteq B_\varepsilon$. Это произойдет при некотором другом радиусе ε_2. Тогда *метрикой Хаусдорфа* называют минимальный из полученных радиусов:

$$d_H(A, B) = \min\Big\{\varepsilon : A \subseteq B_\varepsilon \ \text{ и } \ B \subseteq A_\varepsilon\Big\}.$$

Все приведенные определения удовлетворяют свойствам метрики (1–4). Метрика $d_H(A, B)$ имеет два важных свойства. Пусть каждое из множеств A и B составлено из объединения конечного числа компонентов. Тогда расстояние между двумя такими кластерами не превышает максимального расстояния между их компонентами:

$$d_H\Big(\bigcup_{i=1}^{N} A_i, \quad \bigcup_{i=1}^{N} B_i\Big) \leqslant \max_{1 \leqslant i \leqslant N} d_H\big(A_i, B_i\big),$$

и в этом состоит первое свойство.

Прежде чем пояснить второе свойство, определим действие сжимающего отображения F с коэффициентом r на компакт. Запись $F(A)$, где A — компактное множество, мы будем понимать в смысле действия F на каждую точку компакта, т. е. $F(A) \equiv \Big\{F(x) \;\Big|\; \forall x \in A\Big\}$. Так вот, второе свойство метрики Хаусдорфа заключается в том, что d_H сохраняет свойство сжатия в пространстве H:

$$d_H\Big(F\big(A\big), F\big(B\big)\Big) \leqslant r\, d_H\big(A, B\big).$$

Пара (H, d_H) является полным метрическим пространством. Следовательно, любое сжимающее отображение будет иметь в этом пространстве единственную неподвижную точку. Все это выглядит слишком тривиальным.

Для того чтобы прийти к более интересным ситуациям, вместо одного сжатия рассмотрим их конечный набор. Пусть, например, таким набором будут $(W_1, W_2, \ldots, W_N)$, где каждое из W_i имеет свой собственный коэффициент сжатия r_i. Такой набор называют гиперболической[6] *Системой Итеративных Функций (Iterated Function System — IFS и* `IFS`*)*.

Определим их коллективное действие на компакт A *оператором Хатчинсона*

$$\boldsymbol{W}(A) = \bigcup_{i=1}^{N} W_i(A).$$

Эта запись означает, что на A независимо действует каждый из W_i, а затем берется объединение их образов. Последние являются просто компактными множествами и, следовательно, используя два описанных выше свойства метрики Хаусдорфа, легко убедиться, что $\boldsymbol{W}$ является сжатием в H с коэффициентом $r = \max\{r_i\}$. Действительно, для любых $A_1, A_2 \in H$

$$d_H\big(\boldsymbol{W}(A_1), \boldsymbol{W}(A_2)\big) = d_H\Big(\bigcup_{i=1}^{N} W_i(A_1),\ \bigcup_{i=1}^{N} W_i(A_2)\Big) \leqslant$$
$$\leqslant \max_{1\leqslant i\leqslant N} d_H\big(W_i(A_1), W_i(A_2)\big) \leqslant r\, d_H(A_1, A_2).$$

Используя *Принцип сжимающих отображений*, мы тотчас получаем, что существует единственная неподвижная точка в H, т. е. непустое компактное множество, которое называется *аттрактором и* `IFS`, такое что

$$A = \boldsymbol{W}(A) = \bigcup_{i=1}^{N} W_i(A) \quad \text{и} \quad \forall B \in H \lim_{k\to\infty} \boldsymbol{W}^{\circ k}(B) = A.$$

Первое выражение примечательно тем, что выражает неподвижную точку — компактное множество A — как объединение своих собственных уменьшенных копий. Иначе говоря, A является *коллажем*. Например, для $N = 2$ мы получим

$$A = W_1(A) \cup W_2(A) = W_1(W_1(A) \cup W_2(A)) \cup W_2(W_1(A) \cup W_2(A)) =$$
$$= W_{11}(A) \cup W_{12}(A) \cup W_{21}(A) \cup W_{22}(A),$$

[6]Здесь «гиперболическая» — синоним «сжимающей».

где $W_{ij} = W_i \circ W_j$. Повторяя это разложение k раз, получим

$$A = \bigcup_{1 \leqslant \sigma_1 \ldots \sigma_k \leqslant 2} W_{\sigma_1 \ldots \sigma_k}(A),$$

где $W_{\sigma_1 \ldots \sigma_k} = W_{\sigma_1} \circ \ldots \circ W_{\sigma_k}$.

Рассмотрим простой пример. Пусть $\boldsymbol{W} = W_1 \cup W_2$, где $W_1 = x/3$, $W_2 = x/3+2/3$. Выберем в качестве B единичный интервал $B = I = [0,1]$. Первая итерация дает $\boldsymbol{W}(I) = \big\{[0,1/3] \cup [2/3,1]\big\}$. Вычислим вторую итерацию $\boldsymbol{W}^{\circ 2}(I)$. Применение сжатия W_1 к двум полученным интервалам дает $W_1\big([0,1/3]\big) = [0,1/9]$ и $W_1\big([2/3,1]\big) = [2/9,1/3]$. Независимо применяя к этим же интервалам W_2 получаем $W_2\big([0,1/3]\big) = [2/3,7/9]$ и $W_1([2/3,1]) = [8/9,1]$. Таким образом, $\boldsymbol{W}^{\circ 2}(I) = \big\{[0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]\big\}$. Нетрудно догадаться, что продолжая эту процедуру *ad infinitum*, мы получим классический фрактал — множество Кантора. Оно и является неподвижной точкой выбранного оператора!

Большую роль в теории и `IFS` играет *Теорема о коллаже*, которая следует из неравенства треугольника для метрики Хаусдорфа. Пусть $B \in H$ и $\{W_i\}$, $i = \overline{1,N}$ — `IFS` с максимальной константой Липшица $c = \max\{c_i\}$, которая имеет аттрактор $\mathcal{A} = \mathbf{W}(\mathcal{A})$. Тогда

$$d_H(\mathcal{A}, B) \leqslant \frac{d_H\big(B, \mathbf{W}(B)\big)}{1-c}.$$

Иными словами, чем меньше расстояние (его называют *коллаж-расстоянием*) между произвольным начальным множеством B и его образом $\mathbf{W}(B)$, тем ближе B к аттрактору $\mathcal{A}$.

Примечания

1. *Феликс Хаусдорф* (*Felix Hausdorff*, 8.11.1868–26.01.1942) — немецкий математик, один из основоположников современной топологии. Вывел и исследовал такие понятия, как хаусдорфовы пространства и хаусдорфова размерность. Внес большой вклад в теорию множеств, функциональный анализ, теорию топологических групп и теорию чисел. Выступал как писатель под псевдонимом *Поль Монгре*. Профессор университетов в Лейпциге, Грейфсвальде и Бонне. В 1942 году, когда отправка его и его семьи в концлагерь стала неизбежной, покончил жизнь самоубийством вместе с женой и ее сестрой.

Путеводитель по литературе. Лучшее ведение в метрику Хаусдорфа можно найти в статьях *Джона Хатчинсона*, доступных на его страничке [8], в замечательной книге *Михаила Барнсли* [9], а также в монографиях *Фальконера* [10, 11], сканированные копии которых можно разыскать в библиотеках электронных книг. Популярное изложение содержится в моих лекциях [3, 12].

Пространство кодов

> Я тоже играю символами, но я играю, не забывая, что речь-то идет лишь об игре.
>
> *Иоганн Кеплер*

Рассмотрим слова, образованные из символов или букв некоторого алфавита $\mathcal{A}$, емкость которого равна L. Когда алфавит цифровой — $\mathcal{A} = \{0, 1, 2, \ldots, L-1\}$, словами являются конечные или бесконечные числовые последовательности. Например, если $\mathcal{A} = \{0, 1\}$, то слова имеют вид $s = 01000110\ldots$. Если алфавит составлен из букв $\mathcal{A} = \{\sigma_1, \sigma_2, \ldots, \sigma_{L-1}\}$, то словами являются бесконтекстные (т. е. не имеющие семантического смысла) буквенные последовательности $s = \sigma_{i_1}, \sigma_{i_2}, \ldots$, где $i_k \in \{1, 2, \ldots, L\}$.

Мы часто будем рассматривать слова одинаковой длины K. В этом случае, число всех возможных слов составляет величину L^K. Обозначим через Σ пространство, точками которого являются описанные выше слова. Такие пространства часто используют на практике. В генетике, например, ДНК представляет собой текст, состоящий из нуклеотидов, каждый из которых содержит одну из четырех букв — азотистых оснований: A (аденин), G (гуанин), T (тимин), C (цитозин).

Известно несколько способов получения символической последовательности из временных рядов. Можно, например, использовать *поворотные точки* ряда, т. е. максимумы и минимумы его графика. Пометив минимумы символом 0, а максимумы символом 1, мы получим бинарную последовательность.

Другой способ берет начало из символической динамики [13]. Проще всего преобразовать временной ряд $\{x_i\}$ в бинарный «текст», состоящий из слов $S = s_1, s_2, \ldots, s_k$, где $s_i \in \{0, 1\}$, хотя обобщение на $L > 2$ тривиально. Выберем для этого некоторое пороговое значение ординаты

графика временного ряда $x_i = h$ и трансформируем все отсчеты ряда в символы по правилу: $s_i = 0$, если $x_i < h$, и $s_i = 1$, если $x_i \geqslant h$. Чтобы получить отдельные слова, прочитаем полученный текст с помощью шаблона, содержащего K окон, сдвигая его вдоль текста на один символ. Например, для текста $011010010110\ldots$ и $K = 3$ мы получим слова 011, 110, 101 и т. д.

Можно изучить статистику полученных слов, построив гистограмму их распределения. Разделим единичный отрезок $[0, 1)$ на 2^K интервалов, и каждому слову S из 2^K возможных слов длины K поставим в соответствие точку (адрес слова) из $[0, 1)$ по правилу

$$x(S) = \frac{s_1}{2} + \frac{s_2}{2^2} + \ldots + \frac{s_K}{2^K}.$$

Например, слово 101 является *адресом* точки $x = 1/2 + 0/2^2 + 1/2^3 = 5/8$.

Распределение адресов дает нам гистограмму встречаемости слов. Понятно, что в случае алфавита, содержащего 3 символа, следует использовать троичное разложение и т. п. Если полученная гистограмма *стационарна*, т. е. не меняет своей формы для различных фрагментов текста, ее можно использовать как оценку *эмпирической меры* случайной динамической системы, которая получается следующим образом.

Рассмотрим `IFS` с аттрактором $I = [0, 1]$: $W_1(x) = x/2$, $W_2(x) = x/2 + 1/2$. Возьмем произвольную точку $x_0 \in [0, 1]$. Напомним, что это адрес некоторого слова. Будем выбирать случайно каждое из двух отображений W_i, $i = 1, 2$ с вероятностями p_1 и p_2, соответственно. Заметим, что адрес, начинающийся с нуля, соответствует точке из левой половины I: $0s_2 \ldots s_K \in [0, 1/2]$, тогда как слово, начинающееся с единицы, из правой половины: $1s_2 \ldots s_K \in [1/2, 1]$. Следовательно, применение `IFS` индуцирует появление суффикса $s_K = 1$ с вероятностью p_1, и суффикса $s_K = 0$ с вероятностью p_2.

Для следующего двоичного разряда, который уточняет адрес в каждом из 2^2 интервалов, мы получим вероятности $\{p_1p_1;\ p_1p_2;\ p_2p_1;\ p_2p_2\}$. Продолжая этот процесс, мы придем, как мы убедимся позже, к единственной гистограмме, которую называют *биноминальной мерой*. Она может рассматриваться как *теоретическая* относительно эмпирической оценки, полученной выше, если конечно вероятности выбраны правильно.

Рассмотрим теперь некоторые способы определения расстояний в пространстве кодов.

Метрика Хемминга (1) между словами равной длины определяется как

число позиций, для которых соответствующие символы различны. Например, расстояние Хемминга между словами 1011101 и 1001001 равно двум, а расстояние между словами 2143896 и 2233796 равно трем.

Метрика Левенштейна (2), которую называют еще *дистанцией редактирования*, определяется как минимальное число элементарных преобразований (удаление, вставка и замена) символов, необходимых для преобразования одного слова в другое. Так, например, для того чтобы трансформировать слово МАША в САРАЙ необходимо заменить М на С, Ш на Р и вставить Й. Следовательно, расстояние Левенштейна в данном случае равно трем.

Метрика d_Σ определяется только для пар разных $(s_1 \neq s_2)$ бесконечных слов выражением $d_\Sigma(s_1, s_2) = 1/L^k$. Здесь k — индекс первого символа, по которому различаются два слова. Например, слова $s_1 = 110010100100001\ldots$ и $s_2 = 110011011011101\ldots$ впервые различаются для $k = 6$, так что $d_\Sigma(s_1, s_2) = 1/2^6$. Можно показать, что (Σ, d_Σ) является компактным метрическим пространством.

Наша метрика обладает еще одним приятным свойством. Определим преобразование сдвига $s_m : \Sigma \to \Sigma$ на словах следующим образом. Для каждого $m = \{0, 1, 2, 3, \ldots, L-1\}$ положим $s_m(\sigma_1, \sigma_2, \ldots) = m, \sigma_1, \sigma_2, \ldots$. Иными словами, сдвиг — это добавление к слову в качестве префикса одной из букв алфавита.

Рассмотрим в качестве примера два бинарных слова $s_1 = 01001\ldots$ и $01100\ldots$, так что $d_\Sigma(s_1, s_2) = 1/4$. Добавим к каждому слову префикс, скажем 0. Расстояние между полученными новыми словами очевидно уменьшится $d_\Sigma(s_1, s_2) = 1/8$. Так что наш сдвиг в пространстве (Σ, d_Σ) является сжатием! К этому результату мы еще вернемся.

Метрика d_k, учитывающая память в словах, определяется для конечных L–слов выражением

$$d_k(s_1, s_2) = \sum_{i=1}^{L} k^{L-i+1} \delta(\sigma_i(s_1), \sigma_i(s_2)),$$

где величина $k \leqslant 1/2$; $\delta(i, j) = 0$ при $i = j$ и $\delta(i, j) = 1$ при $i \neq j$. Например, для $s_1 = 1100$, $s_2 = 1110$ и $k = 1/2$

$$d_k(s_1, s_2) = 1/4.$$

С другой стороны, расстояние между $s_1 = 1100$ и $s_2 = 1101$, которые

различаются последним символом, равно

$$d_k(s_1, s_2) = 1/2.$$

Расстояние между двумя тождественными бинарными L–словами в этой метрике будет, очевидно, равно нулю.

Примечания

1. *Ричард Весли Хэмминг* (11.02.1915–07.01.1998) — почетный член IEEE и член Национальной Академии инженерных наук, лауреат премий Тьюринга, Пиорее, Пендера и Рейма. Родился в Чикаго. В 1945 году участвовал в знаменитом Манхэттенском проекте. Затем, в течение 30 лет конструировал компьютеры в лаборатории Белла, где работал и *Клод Шеннон.* Хэмминга называют гением одной идеи. Она была изложена в 1950 году, в его единственной научной статье, посвященной кодам, и содержала конструкцию блочного кода, корректирующего одиночные ошибки, которые возникают при передаче сообщений. Эта работа Хэмминга была отмечена многими наградами. Существует медаль IEEE в честь *Р. Хемминга.*

2. *Владимир Иосифович Левенштейн* (род. в 1935 г.) — российский ученый, доктор физико-математических наук, работает в Институте прикладной математики им. М. В. Келдыша. Метрика, названная его именем, введена в 1965 году. Она с успехом применяется, например, в сейсмологии для идентификации коллективных движений литосферных блоков по синхронным записям различных сейсмических станций.

Путеводитель по литературе. Метрика Левенштейна была введена в работе [14], а ее применение в сейсмологии описано в статье [15]. Метрику d_Σ использовал *М. Барнсли* [9]. Метрика с памятью используется в серии работ *Петера Тино* [16, 17]; на его страничке можно найти также применение сжимающих отображений к проблемам математической лингвистики и анализу ДНК [18].

Меры

Бесконечность больше, чем самое большое из всего, что есть, и еще чуть-чуть.

Дуглас А.
«Автостопом по Галактике»

Абстрактные меры. Мера в определенном смысле обобщает понятие числа: это функция, которая ставит в соответствие множеству его «размер». Представим себе бесконечную массу, равномерно распределенную в евклидовом пространстве R^n. Количество массы внутри единичного куба соответствует его (*лебеговой*) мере $0 \leqslant \mu \leqslant \infty$. *Меры Радона*, к которым относится и лебегова мера, приписывают каждому ограниченному множеству *конечную* меру. Наиболее общими являются *борелевы регулярные меры*; к последним относятся k–мерные меры *Хаусдорфа*, которые не являются радоновыми мерами при $k < n$.

Мы начнем с наиболее популярной меры Лебега. Для прямоугольного множества $R \subset R^n$: $R = [a_1, b_1] \times \ldots \times [a_n, b_n]$ определим меру Лебега как его евклидов объем: $m(R) = [a_1 - b_1] \times \ldots \times [a_n - b_n]$. Тогда, для любого множества $A \subset R^n$ мера определяется как нижняя граница мер объединения прямоугольных множеств, содержащих A:

$$\mu(A) = \inf_{A \in \bigcup R_i} \sum m(R_i). \tag{1}$$

Мера Лебега имеет следующие полезные свойства:

1) $\mu(\varnothing) = 0$ для пустого множества;
2) если A является объединением счетного числа множеств $A \subset \bigcup\limits_{i=1}^{\infty} A_i$, то $\mu(A) \leqslant \sum\limits_{i=1}^{\infty} \mu(A_i)$, где равенство достигается если множества A_i попарно не пересекаются.

В качестве примера рассмотрим множество нулевой лебеговой меры. Традиционный прием построения фрактального множества Кантора заключается в рекуррентном удалении средней трети из единичного интервала $I = [0, 1]$, затем двух фрагментов по $1/9$ из двух получившихся фрагментов

и так далее, *ad infinitum*:

$$\begin{aligned}
&K_0 = [0,1];\\
&K_1 = [0,1/3] \cup [2/3,1];\\
&K_2 = [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1]; \ldots
\end{aligned}$$

Предельное множество $\mathcal{K} = \bigcap\limits_{n=0}^{\infty} K_n$ состоит из общей части всех точек K_n и называется *множеством Кантора*. На i-ом шаге каждое из множеств содержится внутри 2^i прямоугольников размером 3^{-i}. Поэтому, согласно (1), $\mu(\mathcal{K}) \leqslant 2^i \times 3^{-i}$ для всех i, так что при $i \to \infty$ $\mu(\mathcal{K}) \to 0$.

В общем случае, если в пространстве выделен класс подмножеств $\mathcal{F}$, замкнутый относительно операций конечных пересечений и объединений, то функция $\mu : \mathcal{F} \to [0,\infty]$ называется *конечно-аддитивной мерой*, при условиях

1) $\mu(\varnothing) = 0$;
2) если $\{E_i\}_{i=1}^N \subset \mathcal{F}$ — конечное семейство попарно непересекающихся множеств, т. е. $E_i \cap E_j = \varnothing$, $i \neq j$, то $\mu\Big(\bigcup\limits_{i=1}^{N} E_i\Big) = \sum\limits_{i=1}^{N} \mu(E_i)$.

Если условия (1)–(2) выполняются для *счетного* семейства $\mathcal{F}$, меру называют *счетно-аддитивной*.

Множество A называют *μ–измеримым в X*, если меру любого множества B можно представить в виде

$$\mu(B) = \mu(B \cap A) + \mu(B \cap A^C),$$

где $A^C = X \setminus A$ — дополнение множества A в X. Иными словами, измеримое множество разбивается на части, так что его можно представить в виде пересечения с любым другим множеством и его дополнением. Класс μ-измеримых множеств образует так называемую *σ-алгебру*, т. е. класс, замкнутый относительно операций взятия дополнения, счетных пересечений и объединений.

Итак, основная идея введения меры в пространстве X состоит в выделении некоторого класса $\mathcal{A}$ элементарных подмножеств, например, интервалов для меры Лебега. Пустое множество, само пространство X, а также объединение и пересечение конечного (счетного) числа множеств принадлежит этому же классу, представители которого объявляются *измеримыми множествами*. Пару $(X, \mathcal{A})$ называют *измеримым* пространством. Тогда

мера $\mu : \mathcal{A} \to [0, \infty]$ — это функция, которая ставит в соответствие каждому множеству $\mathcal{A} \subset X$ неотрицательное число из интервала $[0, \infty]$. Мера должна удовлетворять, по меньшей мере, условию конечной (счетной) аддитивности [(1)]. Тройку $(X, \mathcal{A}, \mu)$ называют *пространством с мерой*.

Динамические меры. До сих пор мы рассматривали меры как «длины». Однако, в общем случае, мера вовсе не обязана иметь геометрические атрибуты. *Борелеву меру* можно ввести, используя динамику некоторого непрерывного отображения единичного отрезка $f : [0, 1] \to [0, 1]$. *Орбитой* отображения f длины $n + 1$ называют последовательность точек $x_0, x_1, x_2, \ldots, x_n$, где каждая последующая точка — результат итерации предыдущей:

$$x_n = f^{\circ n}(x_0) = f\big(f^{\circ(n-1)}(x_0)\big).$$

Разобьем единичный интервал $I = [0, 1]$ на N сегментов

$$I_k = [(k-1)/N,\, k/N), \quad k = \overline{1, N}.$$

В результате итераций $f^{\circ n}(x_0)$ нашей функции некоторое число точек из орбиты длиной $n + 1$ попадает в I_k. Обозначим это число как

$$n_k = I_k \cap x_0,\, x_1,\, x_2, \ldots,\, x_n.$$

Его можно интерпретировать как время пребывания орбиты в I_k, если каждой итерации поставить в соответствие единицу времени. Можно надеяться, что при большом n и хороших[(2)] начальных значениях x_0 числа $p_k(n, x_0) = n_k/(n + 1)$ не зависят от n и выбора x_0. В этом случае можно доказать[(3)], что *почти для всех* орбит существует *инвариантная вероятностная мера* $\mu_k(I_k)$:

$$\mu_k = \lim_{n\to\infty} \frac{|I_k \cap \{x_0,\, x_1,\, x_2, \ldots,\, x_n\}|}{n}, \tag{2}$$

такая что $p_k(n, x_0) \approx \mu_k$ и $\sum\limits_k \mu_k = 1$. Выражение *почти для всех* понимается в том смысле, что исключения составляют множество нулевой лебеговой меры.

Заметим, что точка $y = f(x)$ для нашей функции $f : [0, 1] \to [0, 1]$ принадлежит некоторому объединению интервалов $U = I_m \cup I_p \cup \ldots \cup I_s$ тогда и только тогда, если $x \in f^{-1}(U)$. Это означает, что число точек последовательности $x_1, \ldots, x_{n+1}$, которые попали в U, совпадает с числом

тех точек орбиты $x_0, x_1, \ldots, x_n$, которые принадлежали $f^{-1}(U)$. Числа μ_k не должны зависеть от сдвига последовательности, так что выражение $\mu(U) = \mu\big(f^{-1}(U)\big)$ утверждает инвариантность меры.

Пусть $m(x_k)$ — мера, определенная в точке x_k. Тогда интеграл $\int g\,\mu$ от произвольной функции $g(x)$ по мере μ понимается как сумма [4]:

$$\int g\,\mu = \sum_{x_k} g(x_k)\ m(x_k), \tag{3}$$

где ряд сходится в абсолютном смысле.

Действие функций на меру. Пусть y_l — значения функции $y = f(x_k)$; таких точек может быть несколько[5] для каждой точки x_k. Тогда меру точек y_l естественно определить как меру их прообразов x_k — другого пути просто нет:

$$m(y_l) = \sum_{x_k = f^{-1}(y_l)} m(x_k). \tag{4}$$

Разумеется, это выражение имеет смысл, если число точек $f^{-1}(y_l)$ не бесконечно велико. Приведенные рассуждения помогают понять как функция действует на меру. Символически это действие называется [6] *push forward map* и записывается как $f \circ \mu = \mu(f^{-1})$, где $(\circ)$ — знак композиции. Иначе говоря, действие функции f на меру — это просто мера тех точек x, которые отобразились с помощью f в точки $y = f(x)$.

Для приложений полезно рассмотреть коллективное действие набора функций на меру. Рассмотрим, например, и `IFS`, снабженную вероятностями:

$$\{W_i; p_i\}, \quad i = \overline{1, N}, \quad p_1 + \ldots + p_N = 1.$$

Их коллективное действие на меру μ описывается *марковским оператором*

$$\mathbf{M}(\mu) = \sum_{i=1}^{N} p_i\,\mu \circ W_i^{-1}. \tag{5}$$

Для любой непрерывной функции $f : X \to R$ справедливо очень важ-

ное для нас выражение:

$$\int_X f\,d(M(\mu)) = \sum_{i=1}^{N} p_i \int_X f\,d(\mu \circ W_i^{-1}) =$$
$$= \sum_{i=1}^{N} p_i \int_{W_i(X)} f(x)\,d(\mu \circ W_i^{-1}) = \sum_{i=1}^{N} p_i \int_X f \circ W_i(x)\,d\mu \tag{6}$$

Поясним, как оно получается. Заметим для этого, что $\mu \circ W_i^{-1}$ — мера тех точек $x \in X$, которые генерируют образы $x \to W_i(x)$. Это обстоятельство можно учесть, используя дельта-функцию и запись $\delta(x - W_i(x))$. Тогда и получается замена $f(x) \to f \circ W_i(x)$ в последнем члене (6).

Мера называется *самоподобной*, если ее можно представить как линейную взвешенную комбинацию ее самой:

$$\mu = p_1\,\mu \circ W_1^{-1} + \ldots + p_N\,\mu \circ W_N^{-1}. \tag{7}$$

Поясним, что это означает. Представим себе один грамм массы $\mu = 1$, равномерно распределенной на интервале $I = [0, 1]$. Пусть и `IFS`

$$W_1 = x/2, \quad W_2 = x/2 + 1/2$$

выбираются с вероятностями p_1, p_2. После действия W_1 мы получим $W_1(I) = [0, 1/2]$, на котором соберется вся масса, т. е. $\mu[0, 1/2] = \mu \circ W_1^{-1}$. Независимое действие второго сжатия соберет всю массу на второй половине интервала: $W_2(I) = [1/2, 1] \to \mu[1/2, 1] = \mu \circ W_2^{-1}$. Поскольку масса должна сохраняться, мы распределяем ее в соответствии с выбранными весами, т. е. вероятностями по двум интервалам:

$$\mu = p_1 \mu \circ W_1^{-1} + p_2 \mu \circ W_2^{-1}.$$

Вторая итерация выбранной пары сжатий приведет к «перевзвешиванию» меры на каждом из четырех интервалов: $\mu[0, 1/4] = p_1\,p_1$; $\mu[1/4, 1/2] = p_1\,p_2$; $\mu[1/2, 3/4] = p_2\,p_1$; $\mu[0, 1/4] = p_2\,p_2$. Продолжая этот процесс, мы будем получать все более тонкую структуру самоподобной меры.

В общем случае обозначим для краткости $\mu \circ W_1^{-1} = S_1$ и $\mu \circ W_2^{-1} = S_2$. Тогда

$$\begin{aligned}\mu = p_1 S_1(\mu) + p_2 S_2(\mu) = \\ = p_1 S_1(p_1 S_1(\mu) + p_2 S_2(\mu)) + p_2 S_2(p_1 S_1(\mu) + p_2 S_2(\mu)) = \\ = p_{11} S_{11}(\mu) + p_{12} S_{12}(\mu) + p_{21} S_{21}(\mu) + p_{22} S_{22}(\mu),\end{aligned}$$

или, продолжая этот процесс, получим

$$\mu = \sum_{1 \leqslant \sigma_1, \ldots, \sigma_k \leqslant 2} p_{\sigma_1 \ldots \sigma_k} S_{\sigma_1 \ldots \sigma_k}(\mu),$$

где $p_{\sigma_1 \ldots \sigma_k} = p_{\sigma_1} \cdot \ldots \cdot p_{\sigma_k}$ и $S_{\sigma_1 \ldots \sigma_k} = S_{\sigma_1} \circ \ldots \circ p_{\sigma_k}$.

Очевидно, что самоподобная мера (7) возникает как неподвижная точка марковского оператора (5), т. е. $\mathrm{M}(\mu) = \mu$. Как мы уже знаем, это следствие сжимающих свойств оператора в соответствующем пространстве мер, которым мы и займемся в следующем разделе.

Примечания

1. Обычно на меру накладывают дополнительно условия инвариантности относительно группы движений. Так привычные лебеговы меры из евклидовой геометрии (длина, площадь, объем) инвариантны относительно группы твердотельных движений в R^3 — поворотов и сдвигов. В этом случае группа *просто транзитивна*: существует лишь одно преобразование $g \in G : y = g(x)$, которое преобразует x в y. Такая группа гарантирует существование единственной (с точностью до мультипликативного множителя) инвариантной меры. Если группа *сильно транзитивна*, то существуют, например, $g_1, g_2 \in G$, такие что $y = g_1(x) = g_2(x)$. В этом случае инвариантной меры может не существовать вообще. Примером может служить «резиновая» плоскость: преобразование $x \to y$ можно выполнить на ней как сдвигом, так и растяжением.

2. Речь идет о типичных начальных значениях. Так, точка $x_0 = 0$ для $f(x_n) = x_n(1 + x_n)$ — «плохая», поскольку она не генерирует орбиту. Таких точек может быть бесконечно много, но их множество должно иметь нулевую лебегову меру.

3. Представим себе путника, который блуждает по храму, состоящему из некоторого числа залов разного размера. Выбор зала случаен, но попадая в него, возможно неоднократно, путник посещает все места помещения. Тогда относительная доля времени, которую путник проводит в каждом помещении, почти для всех вариантов блуждания пропорциональна его размеру. Это утверждение связано с эргодической теоремой, которую в 1987 году доказал *Джон Элтон*. Пусть $\{w_i\}_{i=1}^N$ — набор сжимающих отображений на компактном метрическом пространстве X, снабженных вероятностями $\{p_i\}_{i=1}^N$. Тогда для всех непрерывных функций $f : X \to R$, почти всех начальных точек $x_0 \in X$ и почти всех орбит справедливо выражение:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\leqslant n} f(x_n) = \int_X f(x)\, d\mu.$$

4. $m(x_k)$ можно рассматривать как количество «массы», содержащейся в точке, т. е. как меру Радона. Традиционная запись $\int g\, d\mu$ не имеет особого смысла для дискретного случая, когда не определен дифференциал меры. Интеграл понимается поэтому как сумма произведений $g(x_k)\,\mu(x_k)$ значений функции на значения распределения «массы», т. е. гистограммы, которая задает меру.

5. Иными словами, функция f — *сюръективна*, т. е. для каждого x существует *по меньшей мере* один y. Представьте себе, что наше отображение — это соответствие между точками на нити электрической лампочки в фонаре и точками светового пятна на стене. Очевидно, что для каждой точки пятна можно найти один или даже более прообраз на нити.

6. В общем случае, для гладкого отображения $F : M \to N$ между *любыми* дифференцируемыми многообразиями *push forward* в точке p является лучшей линейной аппроксимацией F вблизи p. Это может быть, например, линейное отображение (дифференциал dF) между касательными пространствами $F_* : T_pM \to T_{F(p)}N$. Для пары измеримых пространств $(X, \mathcal{A})$ и $(X, \mathcal{B})$ функция $f : X \to Y$ называется *измеримой*, если прообразом любого измеримого множества в Y является измеримое множество в X, т. е. $B \in \mathcal{B} \Longrightarrow f^{-1}(B) \in \mathcal{A}$. Если μ — мера в $(X, \mathcal{A})$, то *push forward* с помощью измеримой функции

$f : X \to Y$ — это мера $f_*\mu$ в $(X, \mathcal{B})$, определенная выражением: $f_*\mu(B) = \mu(f^{-1}(B))$.

Путеводитель по литературе. Изложение теории меры можно найти практически в каждом учебнике по функциональному анализу, смотри, например, [19, 20]. Краткое и очень понятное введение в этот предмет содержится в небольшой книжке *Брудно* [21] и лекциях *Джона Хатчинсона* [22]. Отличное введение в меры на фракталах содержится в главе 9 книги *Барнсли* [9], которая, к сожалению, малодоступна. Кратко необходимый аппарат изложен в монографии *Фальконера* [11], которую можно попытаться скачать с известного сайта www.lib.org.by. Для лучшего понимания действия сжатия на меру следует посмотреть замечательные работы *Эдварда Врская* [23,24], доступные на его web-страничке. О самоподобных мерах можно прочесть в статье [8] и в других публикациях *Хатчинсона*, которые можно найти на его сайте.

Пространство мер: метрика Монжа-Канторовича

Расстояние между задачей и целью
должно меряться не прямизной
соединяющей линии, а кривизной
нейтрализации замысловатости.

П. Таранов
«Маневры общения»

В замечательной книге *Дагласа Хофштадтера*[7] рассматривается следующая задача, которая называется «Собака и кость». На глазах собаки хозяин бросает кость через забор в соседний двор. Кость видна в щели забора, в котором, на довольно значительном расстоянии, есть открытая калитка. Некоторые собаки подбегают к забору и начинают лаять, глядя на кость. Другие, более умные, бегут к калитке. Им становится ясно, что кажущееся увеличение *физического расстояния* между начальным и желанным положением — путь через калитку — на самом деле уменьшает *расстояние проблемы*! Очень многие задачи, в некотором смысле, являются вариантами задачи «Собака и кость», потому что проблемы обычно формулируют не

[7] *Хофштадтер Д.* Гёдель, Эшер, Бах: Эта бесконечная гирлянда. – Самара: Барах–М, 2001.

в физическом, а в концептуальном пространстве. Поэтому для преодоления абстрактных «заборов» очень важно найти внутреннюю интерпретацию задачи с подходящей метрикой между проблемой и целью. Одной из таких метрик и является *расстояние Монжа–Канторовича*.

Эту метрику открывали по меньшей мере трижды. Первым был французский геометр и общественный деятель *Гаспар Монж*. Он занимался, в частности, устройством крепостей и задачами оптимального перемещения грунта из одного места в другой. Идеи решения этой задачи были основаны на мере, близкой к той, которую впоследствии использовал для транспортных задач *Леонид Витальевич Канторович*[(2)], называя ее *метрикой Монжа*. Наконец, ее независимо ввел *Джон Хатчинсон* (1981 г.) в своей основной работе «Фракталы и самоподобие», после чего в ряде зарубежных работ ее стали называть *метрикой Хатчинсона*.

Начнем с формального определения. Рассмотрим пространство борелевых мер $\mathcal{M} = \mathcal{M}(\mathcal{B}(X))$. Если этот термин вызывает неприятные ощущения, представьте себе пространство, точками которого, являются различные распределения масс на компактных носителях, т. е. гистограммы.

Пусть $g : X \to R$ — функции с $Lip\ g \leqslant 1$. Метрика *Монжа–Канторовича* в $\mathcal{M}$ определяет расстояние между двумя мерами μ и ν как

$$d_M(\mu, \nu) = \sup\left\{ \left| \int g\, d\mu - \int g\, d\nu \right| \right\}, \tag{8}$$

где $|g(x) - g(y)| \leqslant |x - y|$, $\forall x, y \in X$.

Согласно определению, для того чтобы вычислить расстояние между двумя мерами, необходимо располагать множеством и `IFS`. Подставляя каждую из них в выражение (8), следует найти верхнюю границу разности двух интегралов. Она и будет искомым расстоянием.

Позже мы рассмотрим способ практического вычисления этой метрики, а пока читателю придется поверить, что пара $(\mathcal{M}, d_M)$ образует полное метрическое пространство. Более того, для любых $\mu, \nu \in \mathcal{M}$ марковский оператор $\mathbf{M}$ (5) определяет сжатие в $(\mathcal{M}, d_M)$, т. е.

$$d_M(\mathbf{M}(\nu_1), \mathbf{M}(\nu_2)) \leqslant c\, d_M(\nu_1, \nu_2). \tag{9}$$

Но в этом случае из *Принципа сжимающих отображений* следует, что для семейства и `IFS` с вероятностями $\{W_i; p_i\}$, $\sum_{i=1} p_i = 1$, существует единственная борелева вероятностная мера μ, такая что для любого под-

множества $A \in \mathcal{B}(X)$

$$\mu(A) = \mathbf{M}\,\mu(A) \equiv \sum_{i=1}^{N} p_i\,\mu \circ W_i^{-1}(A), \tag{10}$$

и для любой меры $\nu_0 \in \mathcal{M}$ последовательность

$$\nu_0,\ \mathbf{M}\nu_0,\ \mathbf{M}^{\circ 2}\nu_0, \ldots, \mathbf{M}^{\circ n}\nu_0$$

сходится к μ в метрике d_M, когда $n \to \infty$. Носителем инвариантной меры является аттрактор оператора Хатчинсона $\mathcal{A} = W(\mathcal{A}) = \bigcup_{i=1} W_i(\mathcal{A})$. Из выражения (10) следует, что мера μ инвариантна относительно действия $\mathbf{M}$, т. е. является его неподвижной точкой.

Earth-Mover's Distance. Численный вариант метрики Монжа–Канторовича известен как EMD–расстояние (**E**arth-**M**over's **D**istance)[8]. Идея его вычисления заключается в следующем. Аппроксимируем две меры μ, ν зеркальными гистограммами, имеющими общее основание. Пусть верхняя из них представляет собой кучки грунта μ, насыпанного в каждый бин гистограммы. Бины второй, «опрокинутой» гистограммы, моделируют емкость ям, т. е. меру ν. Пусть площади гистограмм совпадают, т. е. ямы ν способны вместить весь грунт из кучек μ. Задача заключается в перемещении грунта из верхней гистограммы в ямы нижней с *минимальными затратами*. Последние зависят от выбранного «плана» перемещения и его «стоимости», которая определяется расстоянием[9] d_{ij} между i-м бином, где находится грунт, и j-м бином ямы. Таким образом, вычисление EMD сводится к следующей стандартной транспортной задаче линейного программирования. Необходимо минимизировать транспортные расходы:

$$\sum_{i=1}^{m}\sum_{j=1}^{n} f_{ij} \cdot d_{ij} \to \min;$$

при следующих ограничениях:

$$\sum_{j=1}^{n} x_{ij} = a_i; \quad \sum_{i=1}^{m} f_{ij} = b_j; \quad f_{ij} \geqslant 0, \quad i = \overline{1,m}, \quad j = \overline{1,n}\,.$$

[8]Earth mover — это экскаватор, так что EMD — «экскаваторное расстояние».
[9]d_{ij} называют *ground distance*.

Здесь d_{ij} — стоимость перевозки единицы продукции из пункта i в j; f_{ij} — план перевозок $i \to j$; b_j — потребности в продукте в пункте j; a_i — запасы в пункте i. Для модели *закрытого* типа

$$\sum_{j=1}^{m} b_j = \sum_{i=1}^{n} a_i.$$

Основной трудностью решения задачи является большое число переменных и ограничений. Поэтому используются специальные алгоритмы, адаптированные, в частности, для вычисления EMD–расстояния между изображениями.

Примечания

1. *Монж Гаспар* (10.05.1746–28.07.1818) — академик, творец начертательной геометрии и один из организаторов Политехнической школы в Париже. Занимался математическим анализом, химией, метеорологией, практической механикой. В период Французской революции был членом комиссии по системе мер и весов, а также морским министром. Будучи одним из организаторов национальной обороны Республики, он описал способы извлечения селитры, необходимой для выделки пороха, из земли в хлевах, погребах и кладбищах. Устроил множество литейных пушечных заводов. Во время Директории *Монж* — участник похода *Наполеона* в Египет. Во времена Империи он сенатор, граф Пелузский и кавалер ордена Почетного легиона. *Монж* сделал ряд важных открытий в дифференциальной геометрии, создал свою теорию интегрирования уравнений с частными производными и нашел свое решение задачи о колебании струны.

2. *Леонид Витальевич Канторович* (19.01.1912–07.04.1986) — Нобелевский лауреат, экономист, академик, лауреат Сталинской (1949 г.) и Ленинской (1965 г.) премий. Родился в Санкт-Петербурге. В 14 лет поступил в Ленинградский университет. В 1938 году профессор *Л. В. Канторович* был назначен консультантом в лабораторию фанерной фабрики. Для решения проблем, связанных с производством фанеры, он разработал метод максимизации линейной функции при большом числе ограничителей, известный теперь как линейное программирование. В работе «Математические методы организации и планирования производства» (1939 г.) *Л. В. Канторович* показал, что такой

подход универсален для экономических проблем. Нобелевскую премию получил в 1975 г. совместно с *Тьяллингом Чарлзом Купмансом* за «вклад в теорию оптимального распределения ресурсов».

Путеводитель по литературе. Метрика Монжа–Канторовича–Хатчинсона подробно описана в монографиях [9, 11]. EMD–расстоянию, его свойствам и применению к анализу изображений посвящены работы [25, 26] и глава в диссертации [27]. В статье [28] описан улучшенный алгоритм вычисления EMD между гистограммами. О транспортных задачах и алгоритмах для их решения можно прочесть в великолепных лекциях *Б. Лемешко* [29]. Метод вычисления расстояния Монжа–Канторовича с помощью искусственных нейронных сетей предложен в статье *Ярослава Штарка* [30].

Обратная задача и IFS

... иногда чувствуешь, что Идея,
которая казалась Вдумчивой, при
ближайшем рассмотрении, когда
выходит наружу, и дает на себя
посмотреть со стороны,
оказывается совсем Другим Делом.

А. Милн
«Винни-Пух»

Мы описали выше три ипостаси$^{(1)}$ *Теоремы о сжимающем отображении*. В евклидовом метрическом пространстве (X, d) сжатием является отображение Липшица с $Lip < 1$, которое имеет единственную неподвижную точку. В пространстве компактов (H, d_H) сжатием в метрике Хаусдорфа d_H является оператор Хатчинсона с неподвижной точкой — аттрактором IFS. Наконец, в пространстве борелевых мер $(\mathcal{M}, d_M)$ сжатие в метрике Монжа–Канторовича задает оператор Маркова для IFS с вероятностями. Неподвижной точкой является здесь единственная инвариантная мера, носитель которой — аттрактор IFS.

Предположим, что мы знаем носитель, т. е. аттрактор $\mathcal{A}$ и его «раскраску», т. е. меру μ на нем. Мы знаем, на основе вышеизложенной теории, что аттрактор и мера на нем *однозначно* определяются IFS с вероятностями.

Обратная задача теории Систем Итеративных Функций как раз и состоит в нахождении подходящей IFS $\{W_i\}$, $W_i = c_i x + a_i$, $i = \overline{1, N}$, и

соответствующих вероятностей $\{p_i\}$ или переходных вероятностей P_{ij} в марковском случае.

Рассмотрим сперва случай, когда выбор отображений в динамике случайных IFS производится независимо, с помощью *испытаний Бернулли* (2), и опишем схему решения обратной задачи традиционным методом моментов, предложенным в пионерской работе *М. Барнсли* и др. [31].

Исходным является выражение (6) для инвариантной меры, которое мы перепишем в виде

$$\int g(x)\,d\mu(x) = \sum_{i=1}^{N} p_i \int g\circ W_i(x)\;d\mu(x). \tag{11}$$

Для случайной переменной $x \in [0,1]$ определим набор статистических моментов:

$$\mu_k = \int x^k\,d\mu, \quad k = 0,1,2,\dots, \tag{12}$$

где интеграл понимается в смысле формулы (4). Для вероятностной меры, очевидно, $\mu_0 = 1$. В уравнении (11) выберем $g(x) = x^k$. Тогда,

$$\big(g\circ w_i\big)(x) = (c_i x + a_i)^k = \sum_{j=o}^{k} \left(\begin{array}{c} k \\ j \end{array}\right) c_i^{k-j}\,x^{k-j}\,a_i^k. \tag{13}$$

После этого уравнение (11) принимает вид

$$\mu_k = \int g(x)\,d\mu(x) = \sum_{i=1}^{N} p_i \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array}\right) c_i^{k-j}\,a_i^k\,\mu_{k-j}. \tag{14}$$

Последнее выражение приводит к рекурсивной формуле для моментов [10]:

$$\Big(1 - \sum_{i=1}^{N} p_i\,c_i^k\Big)\mu_k = \sum_{j=0}^{k} \left(\begin{array}{c} k \\ j \end{array}\right) \mu_{k-j}\Big(\sum_{i=1}^{N} p_i\,c_i^{k-j}\,a_i^j\Big). \tag{15}$$

Эта формула содержит и коэффициенты IFS и вероятности. Однако, чтобы использовать уравнения (15) нужны значения теоретических моментов.

[10] Для марковской схемы соответствующие формулы можно найти в работе [32].

Поэтому функционал обратной задачи основан на сравнении этих неизвестных моментов с эмпирическими моментами $\hat{\mu}_k$, вычисленными по эмпирической оценке мере $\hat{\mu}$ для $k = 1, 2, \ldots, M$. О том, как можно получить такие оценки из временного ряда методами символической динамики, рассказывалось в разделе «Пространство кодов».

Для численного решения (3) используют стандартные оптимизационные пакеты или генетический алгоритм [33]. Критерием оптимального решения служит либо разность моментов (эмпирических и модельных), либо разности гистограмм эмпирической и модельной меры.

Для большинства практических задач проблема существенно упрощается. Так, например, если речь идет о предсказании пороговых значений некоторой величины, естественно использовать бинарный алфавит $\{0, 1\}$. Следовательно, в уравнениях (11)–(15) достаточно ограничиться $N = 2$. Эмпирическая мера, полученная представлением слов в двоичном коде, имеет в качестве носителя единичный интервал. Поэтому можно выбрать `IFS` с фиксированными коэффициентами в виде: $W_1(x) = x/2$, $W_1(x) = x/2 + 1/2$ и аттрактором $[0, 1]$. Обратная задача сводится тогда либо к нахождению вероятностей p_1, p_2, либо, в марковском варианте, вероятностей p_{11} и p_{22} переходов символов $0 \to 0$ и $1 \to 1$, соответственно. Сопряженная пара вероятностей вычисляется из соотношений $p_{12} = 1 - p_{11}$ и $p_{21} = 1 - p_{22}$.

Примечания

1. Кроме того, существует еще сжатие в пространстве кодов (Σ, d_Σ), о котором мы уже упоминали. Пусть $\mathcal{A}$ — аттрактор `IFS` с вероятностями. Тогда существует непрерывное отображение $F : \Sigma \to \mathcal{A}$ из пространства кодов Σ, определенное для всех последовательностей $\sigma_1, \sigma_2, \ldots \in \Sigma$ как [34]

$$F(\sigma_1, \sigma_2, \ldots) = \lim_{k \to \infty} w_{\sigma_1} \circ w_{\sigma_2} \circ \ldots \circ w_{\sigma_k}(x). \tag{16}$$

Сдвиги в пространстве (Σ, d_Σ), определенные для каждого m, принадлежащего алфавиту $\{1, 2, \ldots, M\}$, как

$$s_m(\sigma_1 \sigma_2 \sigma_3 \ldots) = m \sigma_1 \sigma_2 \sigma_3 \ldots$$

являются сжатиями и для `IFS`:

$$S = \{\Sigma, s_1, \ldots, s_M;\ p_1, \ldots, s_M\}.$$

Ее аттрактором является Σ. Единственная вероятностная мера на аттракторе π определяется как вероятность того, что в слове $\omega_1\omega_2\omega_3\ldots$ символы принимают значения $\omega_1 = \sigma_1, \ldots \omega_k = \sigma_k$. Мера π с мерой μ на аттракторе связаны отображением (16): $\mu = F(\pi)$.

2. Возьмите урну содержащую N_1 черных и N_2 белых шара. Выберите случайно один шар. Предположим, что он оказался черным. Если вы вернете его назад в урну, то вероятности извлечения шаров при следующим испытании останутся неизменными. Это и есть *испытание Бернулли*. При марковском процессе извлеченный шар не возвращается и вероятность достать вторично черный шар становится равной $(N_1 - 1)/(N_1 - 1 + N_2)$.

3. Вопросы существования и сходимости решения подробно обсуждаются в статьях [35]– [37]. В случае линейной оптимизационной задачи для функционала можно использовать L_1–метрику:

$$\sum_{i=1}^{M} |\mu_i - \hat{\mu}_i| \to \min,$$

L_2–метрику, коллаж-расстояние или *EMD*. Задачу можно свести даже к квадратичной, используя «расстояние в шаре» [38]:

$$\sum_{k=1}^{M} (\mu_k - \hat{\mu}_k)^2/k \to \min\,.$$

Предсказание магнитных бурь

По сие время не можно было человеку столько проникнуть в уставы естества, чтобы с вероятностью можно было сказать, от каких причин погода переменяется.

С. Я. Румовский
«Рассуждения о предсказании погод»

Для иллюстрации изложенной техники полезно привести практический пример. Прогноз магнитных бурь является достойной проблемой для марковского прогноза: слишком много случайных факторов *Космической погоды* должны сложится вместе, чтобы итогом стала наша головная боль.

Геомагнитные возмущения отслеживаются различными локальными и планетарными геомагнитными индексами. Причинами наиболее сильных из них — магнитных бурь — являются геоэффективные межпланетные возмущения солнечного происхождения: солнечные вспышки, выбросы корональных масс и межпланетные направленные взрывы, связанные с инверсиями межпланетного магнитного поля. Главная фаза геомагнитной бури представляет собой уменьшение H–компоненты поля от 50 до 400 nT (*нанотеслов*) и может продолжаться от нескольких часов до суток и даже более того [(1)].

Таким образом, магнитная буря — это сложный динамические процесс. Однако, ради простоты, при построении предиктора под «бурей» понимают лишь экстремальное событие — выброс индекса за фиксированный уровень, который принимают за «штатный» режим.

Одним из наиболее популярных индексов при исследовании бурь является среднесуточный D_{st}–индекс, который является мерой интенсивности кольцевого тока (рис. 2). Он вычисляется как усредненная величина возмущений, отсчитываемых от спокойного уровня по данным четырех магнитных обсерваторий, расположенных приблизительно вдоль магнитного экватора [11]. В магнитоспокойные дни величина D_{st} лежит в пределах ± 20 nT. Обычно, магнитной бурей считают значение $D_{st} \geqslant |-30|$ nT; для сильных бурь $D_{st} \geqslant |-50|$ nT и очень сильных $D_{st} \geqslant |-100|$ nT. Например, во время сильной бури 11 февраля 1958 года индекс D_{st} достиг величины -409 nT!

На рис. 3 приведена оценка эмпирической меры, полученная по 16801 значениям временного ряда среднесуточных значений D_{st}-индекса для бинарного алфавита при пороге -30 nT) и длине слова $K = 12$. Буре соответствовал символ 1, а ее отсутствию символ 0. Полученная гистограмма оказалась практически стационарной — она не меняет своей формы независимо от того, по какой половине ряда мы ее построим. Следовательно, в предположении эргодичности, ее можно считать оценкой инвариантной меры.

[11] D_{st}–индекс доступен на сайте: http://swdcdb.kugi.kyoto-u.ac.jp/dstdirV

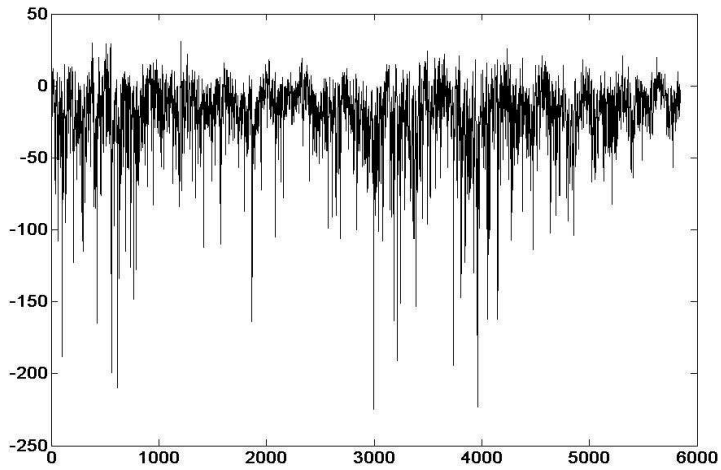

Рис. 2. График среднесуточных значений D_{st}–индекса (1957–2002 гг.). По вертикали — напряженность поля в nT

Для проверки самоподобия меры в пакете `FracLab`[12] были получены мультифрактальные спектры больших отклонений (2) этой меры для двух различных интервалов времени 1981–1991 гг. и 1992–2002 гг. Они фактически совпадают (рис. 4). При построении модели использовался фрагмент ряда с 1981 г. по 1996 г. (длина ряда — 5844 значения).

Для бинарного алфавита достаточно выбрать всего два сжимающих отображения, аттрактор которых — единичный отрезок. Так что проблема сводилась лишь к нахождению переходных вероятностей. Обратная задача для `IFS` решалась в среде `MATLAB 7.0` методом моментов (3) с использованием не менее 15 моментов с помощью `Pattern Search Tool`. Этот пакет содержит 5 методов поиска минимума оптимизационной задачи, включая генетический алгоритм. Все методы дали оценки переходных вероятностей, совпадающие с точностью до 5 знаков после запятой. Для моделирования меры использовались 2×10^5 рекуррентных точек с полученными вероятностями.

[12]Этот пакет доступен на сайте http://www.irccyn.ec-nantes.fr/hebergement/FracLab/

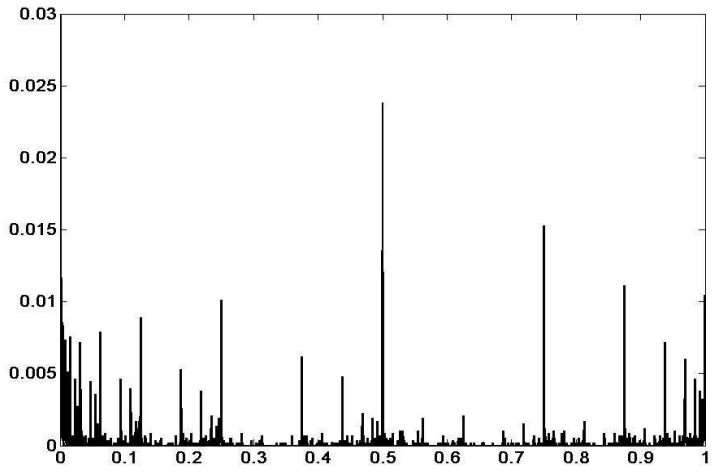

Рис. 3. Гистограмма распределения бинарных слов $K = 12$ для ряда D_{st}–индекса (1957–2002 гг.) с порогом -30 nT

В качестве альтернативы для решения обратной задачи использовалась *Теорема о Коллаже* в следующей модифицированной форме. Оценка эмпирической меры $\hat{\mu}$ рассматривалась как неподвижная точка марковского оператора. Переходные вероятности находились минимизацией функционала $d(\hat{\mu}, \mathbf{M}^{\circ n}\hat{\mu}) \to \min$, где расстоянием между двумя гистограммами $H = \{h_i\}$ и $K = \{k_i\}$ служила L_1–метрика:

$$d(H, K) = \sum_i |h_i - k_i|.$$

Сравнение модели с эмпирической мерой приведено на рис. 5, где показаны графики соответствующих кумулятивных гистограмм: $F = \sum_i |h - \overline{h}|$, построенных по аналогии с моделью случайного блуждания.

Для тестирования модели был получен *эпигноз* $^{(3)}$ 1095 значений нулей и единиц, на временном интервале 1997–1999 гг.

С практической точки зрения, ошибки в предсказании бури и ее отсутствия не равнозначны. Поэтому, для оценки качества прогноза использовался коэффициент $r = n_1/n_2$, предложенный в работе [32]. Он равен

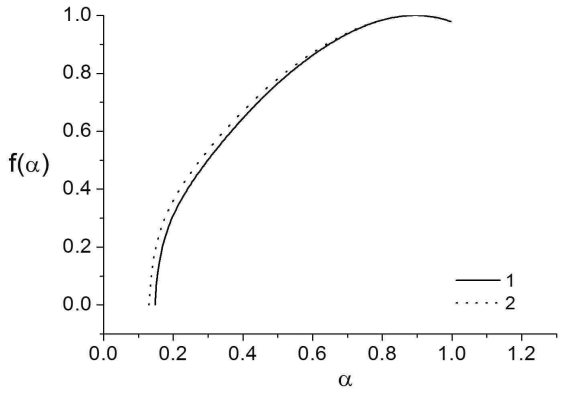

Рис. 4. Мультифрактальные спектры больших отклонений для эмпирической меры D_{st}–индекса 1981–1991 гг. (кривая 1) и 1992–2002 гг. (кривая 2).

отношению n_1 — числа *правильно* предсказанных суффиксов, при условии что соответствующие реальные значения содержали единицу, к общему количеству прогнозов слов n_2, реально содержащих единицу. Слова, содержащие одни нули, в статистике не участвовали. Метод моментов[13] дал следующие оценки эффективности предсказания (в скобках, для сравнения, указаны результаты работы [32]): на 1 день — $70.00(67.24)\%$, на 2 дня — $34.55(34.89)\%$ и на 3 дня — $19.63(19.47)\%$. Использование *Теоремы о Коллаже* дает эффективность: на 1 день — 70.00%, на 2 дня — 35.08% и на 3 дня — 19.63%.

Приведенные оценки показывают, что результаты вероятностного предсказания воспроизводимы и не слишком зависят от метода решения обратной задачи. В случае однодневного предсказания результаты вполне пригодны для практических задач, и даже для предсказания на 2 дня они лучше, чем случайный прогноз, который составляет лишь 25%.

[13] Максимальное число моментов было равно 30.

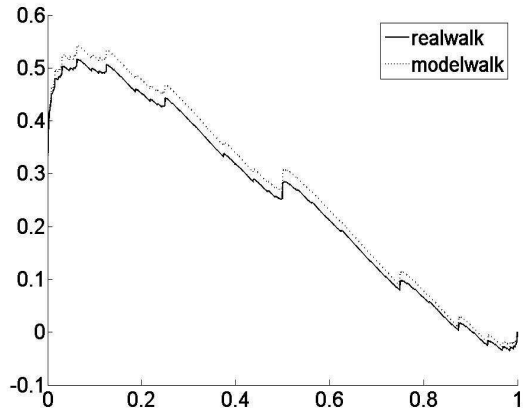

Рис. 5. Кумулятивные гистограммы для эмпирической меры и и IFS–модели (кружки — эмпирическая мера, сплошная линия — модель)

Примечания

1. Депрессия вызвана пересоединением силовых линий межпланетного и геомагнитного полей и флуктуациями размеров магнитосферы под действием потоков солнечного ветра. Эти процессы приводят к ускорению имеющейся в ней плазмы до энергий порядка нескольких тысяч электронвольт, формируя, таким образом, *кольцевой ток* на расстоянии от 3–5 радиусов Земли. Он течет в западном направлении (в северном полушарии) в окрестности Земли и порождает магнитное поле, направленное противоположно геомагнитному, и, следовательно, ослабляет его [39, 40].

2. Мультифрактальные свойства этого индекса с помощью вычислений *размерностей Реньи* были обнаружены *Ванлиссом, Аном и др.* [41].

3. Впервые метод марковского предсказания, основанный на IFS, для магнитных бурь применили *Ан, Ю и Ванлисс* [32]. Они использовали метод моментов. Мы воспроизвели их результаты и сравнили их с прогнозом, полученным с помощью *Теоремы о Коллаже*. Совсем недавно теми же авторами был предложен вариант оценки эмпириче-

ской меры, основанной на *Chaos Game* [42]. Он позволяет исключить зависимость модели от длины слова.

4. В футурологии существует следующая необщепринятая терминология. Термин *прогноз* используется в количественном контексте, скажем, когда требуется указать момент времени и значение некоторой величины. Его синонимом является обычно *предсказание*, хотя последний иногда используют в более общей ситуации. Например, прогноз землетрясений требует одновременного ответа на три вопроса: *где, когда, как сильно*? Но когда пытаются ответить на один или два вопроса, говорят о предсказании. *Эпигнозом* называют предсказание уже известных значений в предположении, что вы их не знаете. *Палегнозом* (*Postdiction*) называют прогноз неизвестного прошлого для временного ряда.

Путеводитель по литературе. Подробнее о физике магнитных бурь и геомагнитных индексах можно прочесть в монографиях [39, 40]. Разные подходы к предсказанию магнитных бурь описаны в статьях [42]– [45].

Заключение

Многое мы знали бы лучше, если бы не стремились узнать столь точно.

Иоганн Вольфганг Гёте
«Годы странствий Вильгельма Мейстера, или Отрекающиеся»

Резюмируем схему марковского предсказания. Предположим, что мы встретились с ситуацией, когда нет оснований полагать, что динамическая система, которая продуцирует временной ряд, допускает гладкое детерминированное объяснение. Более того, есть основания полагать, что мы имеем дело со стохастическим сценарием, когда значения временного ряда являются результатом суперпозиции множества независимых факторов.

Пусть статистика отсчетов демонстрирует долговременные зависимости, или, что то же самое, подчиняется распределению с «тяжелыми» хвостами.

Чтобы формализовать ситуацию, мы вынуждены снизить требования к точности идентификации фазовых точек, т. е. перейти к «крупнозернистому» пространству состояний. Это можно сделать разными способами, и один из самых удачных заключается в использовании символической динамики.

Предположим, что реальная динамика по меньшей мере *рекуррентна*, т. е. ее случайные траектории достаточно часто возвращаются и пересекают некоторую плоскость — сечение Пуанкаре, транверсальное к фазовому потоку.

Маркируем ячейки крупнозернистого разбиения этой плоскости буквами или символами некоторого алфавита Σ с объемом $|\Sigma|$; в простейшем случае символов может быть всего два: 0 и 1. В наблюдаемом временном ряде эта процедура соответствует выбору нескольких или всего одного уровня. Рекуррентные траектории «напечатают» для нас некоторый текст.

Любой осмысленный текст должен содержать слова. Мы получим их, выбрав шаблон фиксированной длины K, посредством которого прочтем множество $|\Sigma|^K$ слов, последовательно сдвигая шаблон вдоль текста на один символ. В бинарном варианте число возможных слов составит величину 2^K.

Используя $|\Sigma|$–ичное представление числа, выбранное в соответствие с объемом алфавита, поставим в соответствие каждому слову точку на единичном отрезке. В итоге получится гистограмма встречаемости слов. Предположим, что такие гистограммы, построенные для всего ряда и для его двух половин, практически совпадают. Иными словами, гистограмма стационарна. Нормируем ее и получим частоту встречаемости (вероятность) для каждого слова из нашего текста.

Почему бы не использовать ее для предсказания K–го символа (суффикса) в последнем известном $(K-1)$–буквенном слове? Казалось бы проблема выбора возможной альтернативы для последней буквы легко решается с помощью полученной гистограммы! Однако, представьте себе реальную лингвистическую ситуацию. Мы имеем в своем распоряжении фрагмент какого-либо текста из сборника, содержащего, например, произведения *Александра Сергеевича Пушкина* и *Велимира Хлебникова*. Составим частотный словарь встречаемости слов, состоящих, скажем из двух букв, используя пушкинский текст. С помощью этой гистограммы мы вряд ли сможем предсказать удивительные дифтонги из текста *Велимира Хлебникова*!

С аналогичной ситуацией мы встречаемся для временных рядов. Фраг-

мент ряда, который мы использовали для построения гистограммы, может содержать чрезмерно малую долю некоторых слов лишь потому, что их было мало в том динамическом сценарии, который мы использовали. Однако в будущем такие слова вполне могут появиться! Вот поэтому и нужна модель, которая способна генерировать статистически содержательную выборку слов, сохранив масштабные свойства эмпирической гистограммы.

Честно говоря, я не могу привести корректные доводы, почему модель, основанная на IFS с вероятностями, является лучшим из возможных вариантов. Однако эвристические резоны в ее пользу существуют. Предположим, что мы проверили скейлинговые свойства гистограммы, оценив лежандровский или какой-либо другой мультифрактальный спектр, и нашли, к своему удовольствию, что эмпирическая гистограмма статистически самоподобна. Таким образом, мы можем считать, что располагаем оценкой стационарной меры, а имея в виду эргодичность, и инвариантной меры с носителем на аттракторе — единичном интервале.

С другой стороны, мы знаем, что такой аттрактор можно получить с помощью пары сжимающих отображений, которые являются *линейным* аналогом диссипативной системы; кстати, недавно *Брумхед* и др. [46] указали на интригующие связи между IFS и нелинейными дифференциальными уравнениями. Мы знаем также, что единственную инвариантную меру на единичном интервале можно получить, используя случайную динамику и IFS с вероятностями.

Вот пожалуй и все, потому что обратная задача — это уже дело техники. Итак, выбираем число отображений N, входящих в IFS, равным объему алфавита: $N = |\Sigma|$. При фиксированных коэффициентах IFS в методе моментов свободным параметром является номер M максимального момента. Увеличение M уменьшает ошибку функционала. Однако слишком большие M могут приводить к неустойчивости численного алгоритма.

Результатом решения является матрица переходных вероятностей. Генерируем модельную меру с помощью рекуррентной последовательности IFS с полученными значениями переходных вероятностей, пока не наберем объем, превышающий объем эмпирической меры.

Модельная мера, на которой основано предсказание, должна хорошо представлять все возможные слова. В случае бинарного алфавита, успешное предсказание получается тогда, когда гистограмма имеет форму «корыта», приводящего к сравнимым вероятностям переходов $0 \to 0$ и $1 \to 1$. Для прогноза на один символ альтернатива 0 или 1 выбирается на основе модельной меры.

Схема выглядит очень красивой и математически безупречной: прогноз делается на твердой основе — инвариантной вероятностной мере. Ведь *Будущее* генетически связано с *Вероятностью*. Как говорил профессор *Бенедикт Коуска* [14]: «*То, что случается, если на самом деле случается, — то и случается. Вероятность появляется лишь там, где нечто еще не успело случиться. Так утверждает наука*».

Литература

1. *Alefeld G., Koshelev M., Mayer G.* Why it is computationally harder to reconstruct the Past than to predict the Future // *Intern. J. Theor. Phys.*. – 1997. – v. 36. – pp. 1683–1689.
2. *Макаренко Н. Г.* Эмбедология и нейропрогноз // Сб. научн. тр. Всероссийской научно-техн. конференции *«Нейроинформатика–2003»*. Часть 1. – М.: Изд-во МИФИ, 2003. – с. 86–148.
3. *Макаренко Н. Г.* Фракталы, аттракторы, нейронные сети и все такое // Сб. научн. тр. Всероссийской научно-техн. конференции *«Нейроинформатика–2002»*. Часть 2. – М.: Изд-во МИФИ, 2002. – с. 121–169.
4. *Stark J.* Iterated Function Systems as Neural Networks // *Neural Networks*. – 1991. – v. 4. – pp. 679–690.
5. *Murtagh F.* Identifying the ultrametricity of time series // *European Physical Journal B*. – 2005. – v. 43 — pp. 573–579.
URL: http://www.cs.rhul.ac.uk/home/fionn/papers/
6. *Шашкин Ю. А.* Неподвижные точки. – Серия «Популярные лекции по математике», Вып. 60. – М.: Наука, 1989. – 77 с.
7. *Хемминг Р. В.* Теория кодирования и теория информации. Пер. с англ. – М.: Радио и связь, 1983. – 176 с.
8. *Hutchinson J.* Fractals: a mathematical framework. Deterministic and random fractals.
URL: http://www.maths.anu.edu.au/∼john/publications.html
9. *Barnsley M.* Fractals everywhere. – N.Y.: Academic Press, 1988. – 531 p.
10. *Falconer K.* Fractal geometry: Mathematical foundations and applications. – Wiley, 2003. – 337 p.
11. *Falconer K.* Techniques in fractal geometry. – John Wiley & Sons, 1997. – 256 p.

[14] *Станислав Лем.* «О невозможности жизни, О невозможности прогнозирования».

12. *Макаренко Н. Г.* Фракталы, мультифрактальные меры и аттракторы // *Нелинейные волны'2002*. – Нижний Новгород, 2003. – с. 381.

13. *Daw C. S., Finney C. E. A., Tracy E. R.* A review of symbolic analysis of experimental data // *Rev. of Scientific Instruments*. – 2003. – v. 74 — pp. 916–930.

14. *Левенштейн В. И.* Двоичные коды с исправлением выпадений, вставок и замещений символов // ДАН СССР, 1965. Вып. 163. Т. 4. – с. 845–848.

15. *Гвишиани А. Д., Жижин М. Н., Иваненко Т. И.* Синтаксический анализ записей сильных движений // *Компьютерный анализ геофизических полей (Вычислительная сейсмология)*. 1990, Вып. 23. – с. 235–253.

16. *Tino P.* Spatial Representation of Symbolic Sequences through Iterative Function Systems // *IEEE Transactions on Systems, Man, and Cybernetics*. Part A: Systems and Humans, 1999. – **29**(4) – p. 386–392.

17. *Tino P., Dorffner G.* Predicting the future of discrete sequences from fractal representations of the past // *Machine Learning*, 2001. – **45** (2) – pp. 187–218.

18. *Tino P.* Multifractal properties of Hao's geometric representations of DNA sequences // *Physica A: Statistical Mechanics and its Applications*, 2002. – **304** (3–4) – pp. 480–494.
URL: http://www.cs.bham.ac.uk/∼pxt/my.publ.html

19. *Колмогоров А. Н., Фомин С. В.* Элементы теории функций и функционального анализа. – М: Наука, 1989. – 496 с.

20. *Титчмарш Е.* Теория функций. – М: Наука, 1980. – 454 с.

21. *Брудно А. Л.* Теория функций действительного переменного. – М: Наука, 1971. – 120 с.

22. *Hutchinson J.* Measure theory. – 1995.
URL: http://www.maths.anu.edu.au/∼john/lecture_notes.html

23. *Vrscay E. R.* From fractal image compression to fractal-based methods in mathematics // *Fractals in Multimedia*, ed. by *M. F. Barnsley*, *D. Saupe* and *E. R. Vrscay*. – Springer-Verlag, N. Y. – 2002.
URL: http://links.uwaterloo.ca/person.ed.html

24. *Vrscay E. R.* A Hitchhiker's guide to "fractal-based" function approximation and image compression.
URL: http://links.uwaterloo.ca/person.ed.html

25. *Rubner Y., Tomasi C., Guibas L. J.* The Earth mover's distance as a metric for image retrieval. – Technical Rep. STAN-CS-TN-98-86.

26. *Kaijser T.* Computing the Kantorovich distance for images // *J. Mathematical Imaging and Vision.* – 1998. – v. 9. – pp. 173–191.

27. *Wadstromer N.* Coding of fractal binary images with contractive set mappings composed of affine transformations. – PhD Thesis. – Linkopings Univer, 2001.

28. *Ling H.. Okada K.* EMD-L1: An efficient and robust algorithm for comparing histogram-based descriptors // *European Conference on Computer Vision 2006*. URL: http://www.cs.umd.edu/∼hbling/main.htm

29. *Лемешко Б. Ю.* Методы оптимизации. Конспект лекций. URL: http://www.ami.nstu.ru/∼headrd/

30. *Stark J.* A neural network to compute the Hutchinson metric in fractal image processing // *IEEE Trans. Neural Networks.* – 1991. – v. 2. – pp. 56–158.

31. *Barnsley M. F., Ervin V., Hardin D., Lancaster J.* Solution of an inverse problem for fractals and other sets // *Proc. Natl. Acad. Sci. USA*. – 1986. – v. 83. – pp. 1975–1977.

32. *Ahn V. V., Yu Z. G., Wanliss J. A., Watson S. M.* Prediction of magnetic storm events using the D_{st} index // *Nonlinear Processes in Geophysics.* – 2005. – v. 12. – pp. 799-806.

33. *Lutton E., Levy-Vehel J., Cretin G., Glevarec Ph., Roll C.* Mixed IFS: Resolution of the inverse problem using genetic programming // *Complex Systems.* – 1995. – v. 9. – pp. 375–398.

34. *Barnsley M., Hutchinson J., Stenflo O.* A fractal valued random iteration algorithm and fractal hierarchy. URL: http://wwwmaths.anu.edu.au/∼barnsley/pages/publication_list.htm

35. *Forte B., Vrscay E. R.* Solving the inverse problem for function/image approximations using iterated function systems, I. Theoretical basis; II. Algorithm and computations // *Fractals.* – 1994. – v. 2, 3. – pp. 325–346. URL: http://citeseer.ist.psu.edu/forte94solving.html

36. *Handy C. R., Mantica G.* Inverse problems in fractal construction: moment method solution // *Phys. D.* – 1990. – v. 43. – pp. 17–36.

37. *Abendat S., Demko S., Turchetti G.* Local moments and inverse problem for fractal measures // *Inverse Problems.* – 1992. – v. 8. – pp. 739–750.

38. *Iacus St. M., Torre D. L.* Approximating distribution functions by iterated function systems // *Departemental Working Papers 2002–03*, Department of Economics University of Milan Italy. URL: http://ideas.repec.org/e/pla155.html

39. *Яновский Б. М.* Земной магнетизм. – Ленинград: ЛГУ, 1978. – 592 с.

40. *Пудовкин М. И., Распопов О. М., Клейменова Н. Т.* Возмущения электромагнитного поля Земли. – Ленинград: ЛГУ, 1976. – 247 с.

41. *Wanliss J. A., Ahn V. V., Yu Z. G., Watson S.* Multifractal modeling of magnetic storms via symbolic dynamics analysis // *J. Geopys. Res.* – 2005. – v. 110. – p. 814.

42. *Yu Z. G., Ahn V.V., Wanliss J.A., Watson S.M.* Chaos game representation of the Dst index and prediction of geomagnetic storm events // *Chaos, Solitons and Fractals.* – 2006. – v. 31. – pp. 736–746.

43. *Watanabe Sh., Sagawa E., Ohtaka K., Shimazu H.* Prediction of the D_st index from solar wind parameters by a neural network method // *Earth Planets Space.* – 2002. – v. 54. – pp. 1263–1275.

44. *Stepanova M., Antonova E., Troshichev O.* Prediction of Dst variations from Polar Cap indices using time-delay neural network // *J. Atmosph. and Solar-Terrestrial Phys.* – 2005. – v. 67. – pp. 1658–1664.

45. *Strivastava N.* A logistic regression model for predicting the occurrence of intense geomagnetic storms // *Ann. Geophys.* – 2005. – v. 23. – pp. 2969–2974.

46. *Broomhead D. S., Huke J. P., Muldoon M. R., Stark J.* Iterated function system models of digital channels. // Proc. Roy. Soc. London, Ser. A. –2004. – v. 460. – pp. 3123-3142.

Николай Григорьевич МАКАРЕНКО, доктор технических наук, главный научный сотрудник Лаборатории компьютерного моделирования (Институт математики, Алма–Ата, Казахстан); доктор физико-математических наук и ведущий научный сотрудник Главной астрономической обсерватории РАН (Пулково, Санкт-Петербург, Россия). Области научных интересов: фрактальная геометрия, вычислительная топология, алгоритмическое моделирование, детерминированный хаос, нейронные сети, физика Солнца. Имеет более 90 научных публикаций.