ОРДЕНА ЛЕНИНА
ИНСТИТУТ ПРИКЛАДНОЙ МАТЕМАТИКИ ИМ. М.В.КЕЛДЫША РАН

А.В. Ворожцов

Прогнозирование и теория сложности

Москва
2001

Прогнозирование и теория сложности

Аннотация Данная работа посвящена не столько прогнозированию или разработке конкретного метода прогноза, сколько поиску математически точной и корректной формулировки задачи о прогнозе. Это тот самый случай, когда условие задачи содержит в себе половину решения. Оказывается для корректной постановки задачи необходимо на множестве марковских отображений (цепей Маркова) построить функцию сложности. Эта сложность соответствует "вычислительной сложности" динамической системы, которую мы моделируем для прогнозирования. Предложенная ниже модель включает в себя целый ряд частных случаев методов прогноза — линейное, гармоническое, квантовое прогнозирование и др.

Predicting and complexity theory

Abstract Here we turn onto the predicting task. The purpose of this paper is much of the setting the task correctly. It's just the case when the correctly formulated task is half of it's solution. It's showed that the measure of complexity on the set of dynamical systems is needed. And firstly Affine maps (Markov chains) must be endowed with this measure. This complexity corresponds to the complexity of modeling or the complexity of calculating dynamic system.

Suggested approach includes a lot of special cases — linear, harmonic and quantum approximations.

Содержание

1. Постановка задачи 3
1.1. Главная идея прогнозирования. 3
1.2. Специфика данного подхода. 3
1.3. Первые приближения 4
2. Описание модели 5
2.1. Абстрактная конструкция 5
2.2. Примеры. 8
3. Некоторые выводы 15
4. Сложность на множестве марковских отображений 15

1. Постановка задачи

1.1. **Главная идея прогнозирования.** Прогноз связан с моделированием реальных динамических систем на вычислительных машинах. Большинство динамических систем, встречающихся в жизни, очень сложны (например, рынок) и практически не поддаются моделированию. И проблемы возникают не только оттого, что невозможно абсолютно точно описать состояние системы (то есть описать все параметры) и правила эволюции, но также из-за нехватки вычислительных мощностей. Погрешность моделирования определяется грубостью и неточностью математической модели и несовершенностью её реализации на компьютере. Как известно реальные системы очень чувствительны к разного рода погрешностям, — большинство из них *хаотично*. А потому возникает такая идея – давайте не будем даже пытаться смоделировать саму систему, а постаемся извлечь полезную информацию об интересующем нас параметре из данных о его предыдущих значениях. Мы верим, что чем больше у нас имеется данных о прошлом, тем точнее будет наш прогноз. Как осуществить его? Каков критерий оптимального прогноза?

1.2. **Специфика данного подхода.** Прежде, чем двигаться дальше, следует отметить необычность предлагаемого подхода. А именно, мы будем "жить" в следующих предположениях:

- Имеются сколь угодно большие вычислительные мощности (быстродействие, память).
- Полностью отсутствуют какие-либо знания о природе динамической системы, которая породила исследуемый ряд.

Эти предположения позволяют перейти на необходимый уровень абстракции и не думать раньше времени о практической реализации. Отсутствие ограничений на выч. мощности позволяет нам создавать модели, сколь угодно точно (и абсолютно точно) приближающие динамику наблюдаемых нами параметров. Какую из них выбрать? Если бы у нас было хоть какое-нибудь

представление о законах реальной динамической системы, мы могли бы, исходя из каких-либо физических соображений, остановить выбор на одной из моделей. Но предположение 2 отсекает эту возможность. Зато взамен мы получаем очень интересную проблему, исследованием которой мы и займёмся: формулировка задачи о прогнозировании без физики.

1.3. **Первые приближения.** Дадим несколько пробных вариантов формулировки задачи о прогнозе. В самом простом случае, нам дано несколько чисел и нужно дописать несколько следующих:

Задача (упрощённая постановка). Пусть нам известна некоторая часть бесконечной последовательности действительных чисел, а именно, первые L членов x_1, x_2, ..., x_L. Найти наиболее вероятное продолжение этого ряда и определить надёжность прогноза.

Её усложнённая и максимально приближённая к практике переформулировка выглядит так:

Задача (приближённая к практике). Дана динамическая система. Z — множество её состояний, а α ($\alpha : Z \to R$) — какой-то её действительный параметр. Нам известно значения этого параметра $x_1, x_2, \ldots, x_L$ в последовательные моменты времени t_1, t_2, $\ldots, t_L$ с погрешностями $r_1, r_2, \ldots, r_L$. Найти наиболее вероятное будущее параметра α и определить надёжность прогноза.

Наличие погрешностей измерения здесь означает, что в моменты времени t_1, t_2, ..., t_L параметр α является нормально распределённой случайной величиной с соответствующими значениями среднего и дисперсии. А потому более точная и общая формулировка имеет следующий вид:

Задача (общая). Дана динамическая система. Z — множество её состояний, а α ($\alpha : Z \to R$) — её действительный параметр. Нам известны распределения значения этого параметра ρ_1, ρ_2, ..., ρ_L в последовательные моменты времени t_1, t_2, ..., t_L. Найти наиболее вероятное распределение $\rho(t)$ параметра α в остальные моменты времени.

Конечно, ни в одной из этих формулировок, задача о прогнозировании не может быть решена, пока не определён смысл слов "наиболее вероятное будущее", то есть пока не задано какое-то *фундаментальное распределение на множестве динамических систем.* То, что такое распределение должно существовать, интуитивно ясно. Действительно, если, например значения параметра некоторой динамической системы в первые сто моментов времени лежит в ε-окрестности a, то скорее всего — то есть более вероятно, что в сто первый момент времени параметр снова не выйдет за пределы этой окрестности. Необходимо научиться получать численную оценку этой вероятности.

Следует отметить, что перед нами стоит, в некотором смысле, задача, являющаяся обобщением задачи об аппроксимации функции. Обобщение заключается в том, что начальная информация о функции имеет вероятностный характер, а в качестве результата аппроксимации нужно получить не функцию, а распределение на множестве функций. Нам, собственно, нужно определиться с методами введения фундаментального распределения на множестве функций, — некоторое условное распределение и будет решением задачи о прогнозировании.

2. Описание модели

2.1. **Абстрактная конструкция.** Первое, что делается при построении теории прогноза временных рядов, — это *предполагается существование устройства фиксированного вида*, которое якобы генерирует рассматриваемый ряд. Это может быть цепь Маркова, линейная динамическая система, нейросеть, машина Тьюринга или др. Мы будем предполагать наличие конфигурации у этого устройства, или, если хотите, будем считать его программируемым, так что после задания конфигурации (ввода программы), оно интерпретируется как динамическая система (с дискретным или непрерывным временем) и один из её параметров соответствует величине, которую мы прогнозируем. Меняя конфигурацию, можно получать различные прогнозы. В случае линейной системы конфигурацией является набор коэффициентов в линейной рекурентной формуле, для цепей Маркова – стохастическая матрица, для нейросети — набор всех весов и параметров взвешивающих функций.

На основе этого предположения и информации о некотором участке ряда строится оптимальный прогноз и делаются оценки надёжности прогнозирования.

Остановимся на модели, использующей в качестве такого устройства цепь Маркова. Эта модель интересна тем, что она математически чиста и самая общая.

Определение 1. *Марковское отображение из A в B есть аффинное отображение аффинного пространства $P(A)$, натянутого на элементы множества A как на вершины, в аналогичное пространство $P(B)$ для B. Эти пространства — множества всех распределений на A и B соответственно. Марковское отображение из A в B однозначно задаётся матрицей p_{ij}, где $i = 1, 2, ..|A|$, $j = 1, 2, \ldots, |B|$, состоящей из положительных чисел от нуля до единицы, которые равны вероятности выхода $b_j \in B$, при условии, что на вход подан элемент $a_i \in A$. Такие матрицы называются стохастическими.*

Цепь Маркова (или стохастическая динамическая система с дискретным временем) — достаточно простая и естественная вещь. Состояние системы в

каждый момент времени t_n есть случайная величина с определённым распределением p_n. Вероятностное распределение в следующий момент времени получается из предыдущего с помощью фиксированного для данной цепи марковского отображения f:

$$p_{n+1} = f(p_n)$$

Марковскую цепь можно получить, замыкая на себя стохастический автомат с общим расширенным состоянием[1]

Идея того, что реальное состояние замкнутого автомата шире, чем наблюдаемый нами параметр out, может быть представлена в более общем виде. А именно: пусть Z общее расширенное состояние стохастического автомата. В результате замыкания его на себя получается случайная последовательность состояний z_1, z_2, ..., но наблюдаем мы не сами состояния z_i, а значения x_i некоторого параметра α, который несёт лишь частичную информацию о состоянии z и задаётся проекцией $\alpha : Z \to A$. В случае замкнутых автоматов нет смысла делить их на автоматы с памятью и без. Правильнее говорить об открытой и закрытой части информации о состоянии системы. Такое представление наглядно демонстрируют идею памяти: память — это та часть расширенного состояния стохастической динамической системы, которую мы не наблюдаем.

---

[1]Справка об автоматах и марковских отображениях.

Автомат представляет собой устройство, способное воспринимать и перерабатывать поток данных. Оно имеет внутреннее состояние $s \in \{s_1, \ldots, s_N\} = S$, и на каждый вход in выдаёт выход out и изменяет своё внутреннее состояние на $\tilde{s}$. Значения $\tilde{s}$ и out есть функции от входа in и текущего внутреннего состояния s. Можно ввести расширенное состояние входа $z = (s, in) \in Z$ и расширенное состояние выхода $\tilde{z}$=($\tilde{s}$, out), тогда действия автомата сведутся к отображению $f : Z_{in} \to Z_{out}$, а именно$f : (s, in) \mapsto (\tilde{s}, out)$. Причём в общем случае отображение f случайно, то есть $\tilde{z}$ — измерение случайной величины зависящей от z. Такие отображения называются **марковскими**. Если отображение, соответствующее автомату, марковское то он называется **стохастическим**. Если у автомата множество внутренних состояний S пусто, то его называют **автоматом без памяти**. Таким образом, стохастический автомат без памяти — это устройство с входом и выходом, которое на каждое значение входа выдаёт одно значение выхода, являющееся измерением случайной величины, зависящей только от значения входа. Если положить, что вход случаен, то получим две марковски связанные случайные величины. Обычно, вход и выход — это элементы R, R^n или некоторого конечного множества.

Рассмотрим автоматы, у которых вход in и выход out принадлежат одному и тому же множеству A, — их мы будем называть автоматами с общим расширенным состоянием. Такие автоматы можно замкнуть, то есть каждый выход использовать как следующий вход, получая в итоге последовательность $out_1, out_2, \ldots$, где марковское отображение $f : Z \to Z$ ($Z = S \times A$) задаёт вероятности переходов $(s_i, out_i) \to (s_{i+1}, out_{i+1})$ Поскольку $out \in A$ есть только часть расширенного состояния автомата, то чёткой зависимости элемента out_i от предыдущих элементов последовательности не прослеживается. Но неопределённость следующего элемента связана не только с нашим незнанием скрытой части состояния, но и с тем принципиальным моментом, что f – случайное (марковское) отображение.

Итак, рассмотрим такую модель:

Модель. Пусть интересующая нас последовательность состоит из элементов множества A. Предположим, что эта последовательность порождена некоторой стохастической динамической системой, и пусть эта система последовательно проходит состояния $z_1,\ z_2,\ \ldots, z_L$, где $z_i \in Z$,

Z — некоторое множество. Для простоты можно считать, что Z конечно. Мы будем предполагать, что наблюдаемый нами ряд $x_1,\ x_2,\ \ldots, x_L$ — это значения параметра α, то есть

$$x_i = \alpha(z_i)$$

В общем случае α — марковское отображение, то есть $a : P(Z) \to P(A)$.

Задавая марковскую цепь с элементами из Z, то есть задавая отображение $f : P(Z) \to P(Z)$, мы получаем вероятностную меру μ на множестве последовательностей элементов из Z. Если мы зафиксируем начальное распределение p_1 на Z, то есть точку в $P(Z)$, то последовательно применяя f, получаем целый ряд распределений $p_2 = f(p_1), p_3 = f(p_2) \ldots$. В итоге вероятностная мера μ на конечных последовательностях элементов из Z задаётся произведением отдельных вероятностей:

$$\mu\left(\{z_1,\ z_2, \ldots, z_L\}\right) = p_1(z_1)p_2(z_2)\ldots p_L(z_L) \tag{1}$$

Итак, f и p_1 порождают вероятностную меру μ на множестве конечных рядов элементов из Z, а наблюдаемый нами ряд состоит из значений $x_i \in A$. Понятно, что с помощью μ и отображения α можно получить вероятностную меру на последовательностях $x_1, x_2, \ldots, x_L$, но прежде следует ещё решить какие следует брать начальные распределения p_1.

Если марковский процесс, задаваемый f, эргодичен, то какое бы p мы не взяли, существует предельное распределение $p_f = \lim_{n\to\infty} \frac{p+f(p)+\cdots+f^n(p)}{n}$ и оно инвариантно: $f(p_f) = p_f$. Множество инвариантных распределений p_f обозначим P_f. Естественно p_1 брать из P_f. Это соответствует предположению, что процесс, который мы собираемся прогнозировать, идёт уже достаточно давно. Тогда все $p_i = p_1 = p_f$. Обычно марковское отображение f порождает один смежный класс на Z, и тогда P_f состоит из одного элемента. Это сильно упрощает ситуацию — нам не нужно будет подбирать подходящее f, достаточно раз и навсегда зафиксировать $p_1 = p_i$. Но в результате мы теряем возможность эффективно прогнозировать динамику — состояние стохастической системы не меняется. Поэтому в дальнейшем под P_f мы будем понимать некоторый широкий класс распределений, зависящий от f, в частности класс всех возможных распределенийй.

Задача о поиске оптимального прогноза посредством цепей Маркова при фиксированном множестве состояний Z сводится к следующему:

Задача 1 (Главная)**.** *Пусть нам дана последовательность* $\mathbf{x} = \{x_1, x_2, \ldots, x_L\}$ *элементов из* A, $\mathbf{x} \in A^L$. *Найдите оптимальные марковские отображения* $f : P(Z) \to P(Z)$, $\alpha : P(Z) \to P(A)$ *и* $p_1 \in P_f$. *Оптимальные в том смысле, что вероятность наблюдения той последовательность значений параметра* α, *которое мы наблюдали на практике, максимальна. То есть*

$$\mathcal{P}(\mathbf{x}) = \underset{f,\alpha}{MAX} \prod_{i=1}^{L} \alpha(p)(x_i) = \underset{f,\,\alpha}{MAX}(\mu_A(\mathbf{x})) \tag{2}$$

где μ_A *есть вероятностная мера на* A^L, *полученная из меры* μ, *заданной на* Z^L *(формула 1), с помощью проекции* α, *то есть* $\mu_A = \alpha^L \mu$

Если p — распределение на Z, то $\alpha(p)$ — распределение на A, а $\alpha(p)(x)$ — значение этого распределения в точке $x \in A$. Величина $\mathcal{P}$ есть вероятность того, что наилучшая при данном Z стохастическая модель продемонстрирует динамику, которую мы наблюдаем у реальной системы.

Собственно говоря, эта главная задача и не задача вовсе, а определение функции $\mathcal{P}(\mathbf{x})$. Это определение превращается в задачу, когда дело доходит до практики. Вычисление функции $\mathcal{P}(\mathbf{x})$ и нахождение оптимальных f, α — задача практически нерешаемая при надлежащих размерах Z.

2.2. **Примеры.** Рассмотрим примеры:

Пример 1. (вырожденный) Пусть прогнозируемая последовательность состоит из элементов множества A. Самый простой вырожденный случай прогнозирующей модели — это когда Z состоит из одного элемента. Тогда прогнозирование полностью лежит на случайном отображении $\alpha : P(Z) \to P(A)$, которое при таком Z есть просто элемент $P(A)$, то есть случайная, ни от чего не зависящая величина. Решение задачи по нахождению максимума 2 представляет собой α, которая с вероятностью 1 равна одному и тому же элементу a, а именно тому, который имеет максимальную частоту встречаемости в $\mathbf{x}$. Если это элемент e, то $\alpha(z)(e) = 1$ при любом z.

Под $\alpha(z)$ здесь понимается значение α на распределении p_z таком, что вероятность z равна 1.

Пример 2. (простой) Если положить, что Z состоит из двух элементов, то задача о нахождении максимального прогноза сведётся к следующему : каждому элементу x_n данной нам последовательности мы должны приписать число z_n, равное либо 0, либо 1. Мы должна подобрать такие α и f, на которых достигается максимум 2.

Пример 3. ("булева" память) Пусть элементы прогнозируемой последовательности есть бинарные слова длины 4. То есть множество A представляет собой множество состояний четырёх бит. В качестве памяти можно взять ещё

несколько бит, например 2, и тогда, аффинное отображение f определяется стохастической матрицей размера $2^6 \times 2^6$. Проектирующим отображением в таком случае обычно берут

$$\alpha : (b_1, b_2, b_3, b_4, b_5, b_6) \mapsto (b_1, b_2, b_3, b_4).$$

Пример 4. ("без памяти") Пусть прогнозируемая последовательность состоит из элементов множества A. В качестве множества расширенных состояний можно взять то же самое множество (память отсутствует). Тогда аффинное отображение f определяется стохастической матрицей вероятностей перехода из A в A, и как нетрудно получить, оптимальные значения этих вероятностей, то есть решение рассмотренной выше задачи, соответствуют наблюдаемым частотам переходов.

Но всё не так просто как может показаться с первого взгляда. То, что в качестве расширенного состояния взято само множество параметров, не означает, что наилучшей проекцией является идентичное отображение. Обычно в случаях, когда Z содержит A ($Z \supseteq A$) предполагают, что ограничение отображения $\alpha : Z \to A$ на A как на подмножество Z равно идентичному, но это лишь частный, и довольно специальный случай. Правильнее считать, что элементы Z и A разной природы. И их мощностные характеристики говорят о степени расширения или сужения множества значений параметров A до множества состояний Марковской цепи Z. В принципе отображение α может быть таким, что разным элементам соответствуют прообразы разной мощности, некоторым — ровно одно значение, другим — несколько значений или бесконечное множество, а третьим — пустое множество. В частности, в предыдущем случае α таково, что каждому значению из A соответствует одинаковое количество элементов из Z, а именно 2^2. А в принципе, размеры прообразов могли быть различными. Можно даже сказать точно: каково множество состояний, в объединении всех русел[2] имеющих параметр α

[2]Русла и Джокеры. Идея представления динамической системы в виде русел и джокеров, находящихся на концах русел, наглядно демонстрирует какое должно быть множество состояний Z.

Русло–это участок в фазовом пространстве стохастической динамической системы движение внутри которого происходит детерминированным или почти детерминированным образом.

Для дискретного случая в упрощённом виде — это просто конечный упорядоченный набор состояний, такой, что попав в его первое состояние, система почти с единичной вероятностью пройдёт последовательно все остальные состояния русла. Для непрерывных систем под руслом уже удобнее понимать ориентированную трубку, попав внутрь которой, мы будем двигаться вдоль неё пока не выйдем из второго конца. Конечно можно дать более общее определение русла, вводить меру его надёжности, эффективную длину и ширину. Но нам довольно и такого упрощённого представления. В процессе нахождения максимума вероятности $\mathcal{P}$ в приведённой выше задачи мы , в принципе, должно автоматически выявить и использовать существующие русла. Другое дело, что нам может не

равным x, таков и прообраз элемента x. Следующий пример демонстрирует наипростейший случай "неоднородности" памяти.

Пример 5. (неоднородная память). Пусть элементы прогнозируемой последовательности есть элементы из конечного множества $A = \{0, 1, 2\}$. В качестве множества расширенных состояний можно взять $Z = \{0_0, 0_1, 1, 2\}$, а проекцию $a : Z \to A$ такую, что 0_0 и 0_1 она отображает в 0, 1 в 1, а 2 в 2. Тогда, аффинное отображение f определяется стохастической матрицей размера 4×4. Здесь скрытая часть состояния (память) существует только тогда, когда значение параметра равно 0.

Пример 6. (стандартное расширение) Пусть элементы прогнозируемой последовательности есть элементы некоторого множества A. В качестве расширенного состояния можно взять $Z = A^{m+1}$ и проекцию $a : Z \to A$, определить так $\alpha(\{a_1,\ a_2,\ \ldots,\ a_m\}) = a_1$.

Именно такое расширенное состояние обычно и вводится в различных методах прогнозирования. При этом память, представляющая собой A^m, используется для хранения предыдущих m значений параметра. Рассматриваются условные вероятности следующего значения параметра, при условии что известны предыдущие m значений. Постепенно увеличивая число m мы останавливаемся на том значении, при котором величина $-\frac{log(p(m))}{m}$ имеет максимальное значение, где $p(m)$ есть $\mathcal{P}(\mathbf{x})$ для $Z = A^{m+1}$ и α, определённой указанным выше способом. Собственно говоря, для каждого m этот пример сводится к примеру 1, где вместо A следует взять A^{m+1}. То, что мы таким стандартным образом использовали память (хранили в ней предыдущие m значений), не означает, что это наиболее эффективный способ использования памяти. Мы зафиксировали отображение α и не варьировали по нему.

Введение ограничений на α и f — довольно распространённый приём, позволяющий упростить процедуру поиска максимума. Одним из показательных примеров — это прогнозирование, базирующееся на фурье-преобразовании. Ограничивая себя суммой гармоник, мы существенно упрощаем себе задачу

хватить для этого размера выбранного Z. Когда мы выберем Z таким, что оно будет допускать однозначное вложение объединения всех русел, значение вероятности $\mathcal{P}$ резко возрастёт, и именно на таком Z мы и должны остановиться. Вопрос состоит в том, как автоматически выделять эти русла, то есть как оптимизировать процесс нахождения максимума $\mathcal{P}$, и как научиться сразу же определять размер множества Z — которое равно объединению всех русел.

Итак, задачу о нахождении наилучшего расширенного состояния — не слишком большого, но достаточного для прогнозирования — можно интерпретировать как нахождение множества всех "русел". И то сколько памяти следует выделять отдельному значению, зависит от того насколько часто это значение фигурирует в "руслах".

поиска наилучшего приближения, — нужно лишь по известной формуле подсчитать коэффициенты в фурье-разложении. Модель, обобщающая фурье-методы прогнозирования, опирается на предположение, что данный параметр есть фазовая координата некоторой линейной системы, чьи собственные числа равны по модулю единице. Этот метод автор назвал "гармоническим прогнозированием". Квантование соответствующей линейной системы приводит к идее "квантового прогнозирования". Но мы начнём рассмотрение с "линейного прогнозирования", где не накладывается никаких ограничений на собственные числа линейной системы.

Пример 7. (линейное прогнозирование) Пусть прогнозируемая последовательность $\mathbf{x} = \{x_1,\ x_2, \ldots, x_L\}$ состоит из комплексных чисел, то есть $A = \mathbb{C}$. В качестве множества расширенных состояний возьмём линейное комплексное пространство $Z = \mathcal{L}(\mathbb{C}, m)$ размерности m. Положим такие ограничения: α принадлежит множеству линейных форм на Z, то есть $\alpha \in \mathcal{L}^*(\mathbb{C}, m)$, а f является обычной, не марковской, функцией и, более того, линейна, то есть $f \in Aut(\mathcal{L}(\mathbb{C}, m))$.

К линейному прогнозированию относятся например методы, подбирающие линейную рекуррентную формулу для значений параметра.

В описанных условиях для некоторых значений m необходимых α и f может вообще не найтись. То есть при маленьких m, меньших некоторого натурального числа — $m < K < L$, $K \in \mathbb{N}$, имеем $\mathcal{P}(\mathbf{x}) = 0$. А при $m \geqslant K$ получим $\mathcal{P}(\mathbf{x}) = 1$. Значение K, безусловно, имеет самостоятельный интерес, но очевидно, что правильнее вводить метрику на множестве последовательностей и о близости модели к реальной системе судить по величине удаления их траекторий. Расстояние между двумя последовательностями $\mathbf{x}$ и $\mathbf{x}' \in A^L$ определяет степень приближения одной последовательности к другой. И тогда задача о поиске оптимальной модели сведётся к поиску минимума среднего удаления последовательности, порождаемой моделью, от реальной.

Итак, предположим, что на множестве последовательностей элементов из A длины L, которое мы обозначили как A^L, введено расстояние ρ. Тогда после фиксирования ограничений на f, α и Z, мы можем приступать к поиску минимума величины погрешности модели

$$\varepsilon = \underset{f,\, p1 \in P_f,\, \alpha}{MIN} \left(\int\limits_{A^L} \rho(x, x') d\mu_A(x') \right), \tag{3}$$

где $x \in A^L$ — данный нам кусок изучаемой последовательности из L элементов, $x' \in A^L$ — переменная по которой ведётся интегрирование, μ_A– вероятностная мера на A^L, зависящая от f, α и p_1.

Во всех предыдущих примерах, со второго по шестой, мы имели дело с вырожденной метрикой: $\rho(\mathbf{x},\mathbf{x}')=\delta(\mathbf{x},\mathbf{x}')$, зато мера μ была не тривиальной. В примере 7 наоборот, вероятностная мера μ вырождена — она сосредоточена в одной точке, а метрика нетривиальна. Один из случаев невырожденных μ и ρ рассмотрен в примере 9 (квантово-механическая модель), а сейчас приведём ещё один богатый частными случаями пример, близкий по сути как к линейному, так и квантовому прогнозированию.

Пример 8. (гармоническое прогнозирование) Пусть прогнозируемая последовательность состоит из комплексных чисел — $A=\mathbb{C}$, её длина равна L и это значения некоторого параметра в моменты времени $t_1,\ t_2,\ \ldots,\ t_L$. В качестве множества расширенных состояний возьмём линейное комплексное пространство $Z=\mathcal{L}(\mathbb{C},m)$ размерности m. Положим такие ограничения: α принадлежит множеству линейных форм на Z, то есть $\alpha\in\mathcal{L}^*(\mathbb{C},m)$, а $f\in U(m)$, и имеет вид $e^{H\cdot t}$, где H антиэрмитова матрица. Более точно: $z(t_{i+1})=e^{H\cdot(t_{i+1}-t_i)}\circ z(t_i)=e^{H\cdot(t_{i+1}-t_0)}\circ z(t_0)$.

Приятно, что здесь предусмотрена возможность неравномерности временных промежутков. В качестве метрики естественно взять

$$\rho^2(\mathbf{x},\mathbf{x}')=\sum_{i=1}^{L}|x_i-x_i'|^2.$$

Тогда при поиске минимума 3 вариация будет осуществляется по α, $z_1=z(t_1)$ и антиэрмитовой матрице H, то есть по $2L+2L+L^2$ действительным параметрам. Минимизироваться будет величина (частный случай формулы 3 с вырожденной вероятностной мерой)

$$\varepsilon=\sum_{i=1}^{L}\rho^2(x_i,\alpha(z_i))=\sum_{i=1}^{L}\rho^2\left(x_i,\ \alpha\left(e^{H\cdot t_i}z_1\right)\right)$$

Но, очевидно, их число переменных варьирования можно уменьшить. Например, фиксация значения $z(t_1)$ не нарушает общности. При эволюции гармонической системы сохраняется норма вектора состояния — такие системы в принципе не могут иметь монотонно возрастающего или убывающего тренда, то есть никакой параметр не может расти или убывать на всей временной оси.

Замечание к примерам 7 и 8. Если данная нам последовательность есть значения нелинейного параметра α некоторой линейной системы (α — нелинейная функция на $Z=\mathcal{L}(\mathbb{C},m)$), то предложенные методы оказываются нерабочими. Вариация по всем возможным нелинейным α слишком непомерный и бессмысленный труд.

? **Вопрос 1**: Как просто научиться прогнозировать нелинейные параметры линейной системы, то есть как вычислить эту нелинейность по опытным данным?

Пример 9. (квантовое прогнозирование) Пусть прогнозируемая последовательность состоит из элементов конечного множества A, $|A| = m$, и её длина равна L. В качестве множества расширенных состояний возьмём единичную сферу в линейном комплексном пространстве размерности m (вектора с единичной длиной) $Z = \{z | z \in \mathcal{L}(\mathbb{C}, m),\ |z| = 1\}$ (строго говоря $Z = \mathbb{C}P^m$ — комплексное проективное пространство). Проекция α будет у нас недетерминированным отображением, то есть результатом его применения будет не конкретный элемент из множества A, а распределение на A. Определим его так: зафиксируем некоторый ортонормированный базис в $\mathcal{L}(\mathbb{C}, m)$ и сопоставим взаимооднозначным образом элементы этого базиса и элементы A; квадраты модулей коэффициентов разложения произвольного элемента из $\mathcal{L}(\mathbb{C}, m)$, модуль которого равен 1, и будут соответствовать вероятностям элементов A. Если обозначить множество квантовых состояний получнной системы как $Quant(A) = Z = \mathbb{C}P^{|A|}$, то $\alpha : Quant(A) \to P(A)$.

На f накладываются те же ограничения, что и в предыдущем примере. На этот раз мы имеем вырожденную метрику (δ-метрику) и задача по нахождению минимума 3 вырождается в задачу о максимуме 2, который в данном случае выглядит так

$$\mathcal{P} = \prod_{i=1}^{L} \alpha(z_i)(x_i) = \prod_{i=1}^{L} |\langle x_i | z_i \rangle|^2 = \prod_{i=1}^{L} \left| \left\langle x_i | e^{H \cdot t_i} z_1 \right\rangle \right|^2 \tag{4}$$

Варьировать придётся только по антиэрмитовой матрице H и начальному состоянию z_1. В данном случае мы прогнозировали непосредственно независимые состояния квантовой системы. Если мы хотим прогнозировать некоторую физическую величину G квантовой системы, то картина будет несколько сложнее. Пусть $\mathbf{G}$ – эрмитов оператор, соответствующий этой физической величине, а $|g\rangle$ собственный вектор оператора $\mathbf{G}$, соответствующие собственному значению g. В качестве α следует взять

- $\alpha(z)$ — случайная величина равная g с вероятностью $|\langle g | z_i \rangle|^2$, либо
- $\alpha(z) = \langle z,\ G \circ z \rangle \in \mathbb{R}$ — действительная функция, равная среднему значению физической величины G.

Тогда величина отклонения значения $\alpha(z_i)$ в момент t_i от реальной величины x_i будет в первом случае равна

$$r_i^2 = \sum_g |g - x_i|^2 \ |\langle g | z_i \rangle|^2 = \sum_g |g - x_i|^2 \ |\langle g | e^{H \cdot (t_i - t_1)} z_1 \rangle|^2.$$

А во втором случае

$$r_i^2 = |x_i - \alpha(z_i)|^2 = |x_i - \alpha(e^{H\cdot(t_i - t_1)} z_1)|^2.$$

Тогда формула 3 примет вид

$$\varepsilon(\mathbf{x}) = \underset{G,H,z_1}{MIN} \sum_{i=1}^{L} r_i^2 \tag{5}$$

Добавиться m^2 действительных переменных варьирования определяющих оператор $\mathbf{G}$, но их число можно сократить до m, сразу же перейдя в базис, состоящий из собственных векторов $\mathbf{G}$.

Итак, после фиксирования множества состояний Z и множества значений параметра A можно поставить задачу о максимизации вероятности при варьировании проецирующего отображения α и стохастического оператора f. Но какие именно ограничения следует накладывать на Z, f и α?

У нас есть ещё один важный и принципиальный вопрос о том, насколько сильно следует расширять A до Z. Следующий пример показывает, что можно взять настолько большое Z, что при максимизации вероятности достижима вероятность 1. То есть можно подобрать такие α и f, что данная нам последовательность для обучения прогнозирующего устройства будет порождаться нашей марковской цепью с вероятностью 1. При этом, конечно f выродится до обычного отображения из Z в Z, а марковская цепь будет обычной детерминированной динамической системой с дискретным временем.

Пример 10. (слишком большая память) Пусть прогнозируемая последовательность состоит из элементов множества A и её длина равна L. В качестве множества расширенных состояний можно взять $Z = A^L \times C$, где $C = \{1, 2, \ldots, L\}$. Тогда, если в качестве начального состояния взять $(x_1, x_2, \ldots, x_L, 1)$, и положить f и α обычными, не марковскими отображениями определёнными как

$$f : (x_1, x_2, \ldots, x_L,\ i) \mapsto (x_1, x_2, \ldots, x_L,\ i + 1\ mod\ L)$$

$$\alpha : (x_1, x_2, \ldots,\ x_L,\ i) \mapsto x_i,$$

тогда вместо марковской цепи получаем детерминированную систему, значения параметра α которой в первые L моментов времени пробегают последовательность $x_1, x_2, \ldots, x_L$. Таким образом данная система максимально точно приближает исследуемую систему в первые L моментов времени; далее она просто циклически повторяет эти L состояний. Такое прогнозирующее устройство довольно нелепо. Это говорит о том, что бессмысленно выбирать всё более и более мощные Z, — где-то нужно остановиться.

3. Некоторые выводы

Подведём итоги рассмотрения данных примеров:

◇ **Нужно определиться с ограничениями на Z, f, и α.**

Как выбрать оптимальное множество расширенных состояний? Искать среди конечных Z или рассматривать континуальные множества? Следует ли при этом переходить на линейные пространства и ограничиваться линейными операторами f? Трудно ответить на эти вопросы, но понятно, что чем больше мы ограничений наложим на f и α, тем проще будет решаться задача о поиске минимума 3, но тем меньше мы охватим возможных динамических систем при данном Z и, скорее всего, найденный условный минимум будет не абсолютным.

◇ **Нужно найти критерий выбора размера Z.**

Как было замечено в примере 10 нет смысла брать слишком большие Z. Максимальное значение вероятности фиксированном Z есть $\mathcal{P}(Z)$. Число $S = -\log \mathcal{P}(Z)/L$, можно назвать *хаотичностью*, которую мы вынуждены вводить в модель, будучи ограниченными множеством состояний Z, но $\mathcal{P}$ может быть мала в силу хаотичности самой реальной системы, и нет смысла моделировать эту хаотичность. С увеличением мощности Z величина хаотичности уменьшается ($\mathcal{P}(Z)$ растёт) и в какой-то момент достигает нуля (например $Z = A^L \cdot \{1,\ 2, \ldots,\ L\}$). Но нам нужно остановиться на каком-то промежуточном значении, поскольку реальная система скорее всего имеет ненулевую составляющую шума и пытаться найти для неё детерминированную формулу — значить тратить зря вычислительную мощность.

4. Сложность на множестве марковских отображений

Сконцентрируемся на втром пункте отмеченных выше выводов. И как и раньше будем размышлять максимально общё. Вычислительные мощности растут не только с увеличением Z, но также зависят и от сложности моделирующего эволюцию отображения f. А потому естественно говорить не о размере Z, а о сложности модели. Эта сложность — функция на множестве моделей, то есть на множестве пар $\{Z,\ f : P(Z) \to P(Z)\}$

Итак, пусть мы имеем такие две алгебраические конструкции:

- отношения сложности на множестве прогнозирующих моделей,
- расстояния между последовательностями.

Тогда мы можем ввести следующую важную функцию — зависимость сложности модели от точности приближения.

В нашем случае множество моделей это — множество марковских отображений.

Уточним сказанное. Предположим, что на множестве последовательностей элементов из A длины L, которое мы обозначили как A^L введено расстояние ρ, а на множестве отображений f — функция сложности $C(f)$ принимающая положительные действительные значения. Тогда можно поставить задачу о нахождении отображения f со сложностью меньшей C_0 и начального состояния $p_1 \in P_f$ таких, что среднее удаление порождаемой в итоге случайной последовательности длины L от реально измеренной было минимально.

То есть ищется

$$(6) \qquad \underset{\substack{f,\ p_1 \in P_f,\ \alpha \\ C(f) < C_0}}{MIN} \left(\int\limits_{A^L} \rho(\mathbf{x}, \mathbf{x}') d\mu_A(\mathbf{x}') \right),$$

где $\mathbf{x} \in A^L$ — данный нам кусок изучаемой последовательности из L элементов, $\mathbf{x}' \in A^L$ — переменная по которой ведётся интегрирование, μ_A — вероятностная мера на A^L, зависящая от f, α и p_1. Можно поставить и сопряжённую ей задачу о нахождении максимально простого f, при условии, что погрешность приближения не должна превышать величины ε.

Нам осталось только ввести упомянутые алебраические структуры.

Выбор метрики на множестве конечных последовательностей зависит от конкретного случая и принципиальных проблем с её введением нет. В данной работе мы сассматривали метрики $\rho(\mathbf{x}, \mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}')$, для случая, когда на A не задано никакой топологии, и стандартную евклидову метрику, когда динамика происходит в линейном пространстве.

Выбор метрики полностью лежит на нашей ответственности. Она отражает то, что мы понимаем под “приближением”.

А вот со сложностью дело обстоит не так просто. Универсального метода определения сложности до сих пор не существует.

Первые идеи введения сложности на различных объектах были описаны Колмогоровым А.Н. в его работе “Комбинаторные основания теории информации”, §4 “Определение сложности”. Определение Колмогорова построено на понятии *языка описания*. Особое внимание он уделял языкам описания алгоритмов и выделению сложностных классов на множестве алгоритмических задач. Этот метод здесь был бы вполне уместен, если бы не одно “но”. Реализация функции на компьютере есть не что иное как ее описание на ассемблере. Увеличивая допустимую сложность описания (наращивая выч. мощности), можно получать все более точное приближение модели к реальности. Неприменимость такого подхода вызвана тем, что нет выделенного языка описания. Если мы выберем какой-нибудь из языков описания, то возникает

класс эквивалентных ему языков и другие классы. И класс самых эффективных языков оказывается зависимым от нашего начального выбора. Конечно, этот выбор можно зафиксировать, взяв какой-нибудь “естественный” язык описания функции. Можно убедиться, что большинство “естественных” языков порождают одну и ту же иерархию классов языков описания по их описательной способности. Но это не даёт решение проблемы — класс самых эффективных языков описания оказывается настолько широк, что значения минимума 6, грубо говоря, может быть любым.

Нужна абсолютная мера сложности, а теория алгоритмов рассматривает исключительно ассимптотическое сложности, выделяя, например, классы полиномиально вычислимых функций и функций с экспоненциальной сложностью.

Введению меры сложности на множестве аффинных отображений посвящена работа автора “Алгебраические методы определения сложности”