Построение предиктора разладки в нестационарных временных рядах

Колотова Арина Александровна

Научный руководитель:
д.ф.-м.н., проф. КВМ МФТИ, зав. отд. вычислительной физики и кинетических уравнений ИПМ им. М.В. Келдыша РАН
Орлов Юрий Николаевич

- **Колотова А.А., Орлов Ю.Н.** Моделирование предикторов разладок в нестационарном временном ряде.
- Препринты ИПМ им. М.В.Келдыша. 2025. № 22. 23 с. EDN: LXLFJY https://library.keldysh.ru/pre 22

ИПМ им.М.В.Келдыша РАН • Электронная библиотека

Препринты ИПМ • Препринт № 22 за 2025 г.

ISSN 2071-2898 (Print)
ISSN 2071-2901 (Online)

А.А. Колотова, Ю.Н. Орлов

Моделирование предикторов разладок в нестационарном временном ряде

Рекомендуемая форма библиографической ссылки: Колотова А.А., Орлов Ю.Н. Моделирование предикторов разладок в нестационарном временном ряде // Препринты ИПМ им. М.В.Келдыша. 2025. № 22. 23 с. EDN: LXLFJY https://library.keldysh.ru/preprint.asp?id=2025-22

Определение

Временной ряд - последовательность наблюдений некоторого признака (случайной величины) x в последовательные моменты времени.

Рис. 1. Индекс RTS, июнь 2012, шаг цены закрытия 5 мин (пунктиром показан коридор среднеквадратичного отклонения одношаговых приростов цен)

Определение

Распределение называется **стационарным** в широком смысле, если среднее значение и дисперсия случайной величины не меняются во времени.

```
(б) Сплошная последовательность текстов разных авторов, требуется распознать
авторов фрагментов и указать начало и конец фрагмента.
```

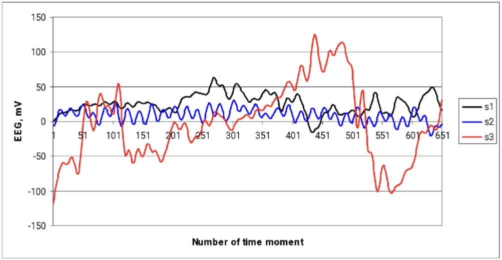


Рис. 2. Ряды электроэнцефалограмм по трем отведениям для одного пациента

```
(в) Ряды биометрических данных, образующих поток с высокой интенсивностью
(порядка 100–1000 в секунду — Рис. 2).
```

О разладках

- **Нестационарность** порождается тем, что в случайные моменты времени происходит переключение между состояниями (**разладка**), причем распределение моментов разладок не является стационарным.
- **Эмпирической** или **выборочной функцией распределения (ВФР)** называется ступенчатая неубывающая функция $F_T(x)$ такая, что

$$F_T(x) \equiv F_T(k) = \sum_{i=1}^{k} v_i = \sum_{i=1}^{k} \frac{n_i(T)}{T},$$

где T - объем выборки. При условии, $x_k \leq x < x_{k+1}, k \leq n-1$. Если $x < x_1$, то $F_T(x) = 0$.

Что мы имеем?

- Много методов анализа стационарных временных рядов (в частности, регрессионные).
- Интересно же работать с реальными системами, которые лишены стационарности.

Предположения насчет ряда

1. ВР конечной размерности состоит из данных, которые производятся некоторой реальной системой, т.е. *физическим объектом.*
2. Идеальное функционирование системы отвечает ее пребыванию в определенном состоянии, которое описывается стационарной ФР.
3. Число возможных состояний заранее неизвестно, но считается, что каждому состоянию отвечает определенная ФР.
4. Переход от одного состояния к другому в реальной физической системе происходит за некоторое конечное ненулевое время.

Формулировка общей задачи

1. построить модель идентификации локального состояния временного ряда
2. разработать **предиктор** - численный алгоритм, позволяющий моделировать системы, для которых характерен набор дискретных состояний.

- Классические подходы (например, ARIMA) фиксируют момент разладки — то есть, **реагируют постфактум**, когда система уже изменила своё состояние.
- Однако во многих приложениях важна **заблаговременная индикация**.
- Это и есть задача построения предиктора.

Для чего это нам? (отчет 29 ноября 2024)

1. на основе наблюдения за одной реальной траекторией можно моделировать целый набор нестационарных траекторий, характерных для некоторой систем.
2. на этом наборе будем тестировать управляющий или распознающий функционал.

Статья (апрель 2025):

1. Формулируется методика моделирования ансамбля траекторий временного ряда, в котором происходят случайные переключения между заданным набором состояний.
2. Предложена **модель предиктора** перехода как минимального расстояния от текущего выборочного распределения до эталонов промежуточного базиса.
3. Описана **структура программного комплекса**, позволяющего проводить моделирование и анализ нестационарных временных рядов

Модель

Пространство состояний. Пусть существует фиксированное множество стационарных состояний

$$\mathcal{M} = \{1, 2, \ldots, n\},$$

где каждому состоянию $j \in \mathcal{M}$ соответствует фиксированная функция распределения $\Phi_j(x)$ с плотностью $\phi_j(x)$. Эти функции задают поведение системы в стационарном режиме.

Модель

Пространство состояний. Пусть существует фиксированное множество стационарных состояний

$$\mathcal{M} = \{1, 2, \ldots, n\},$$

где каждому состоянию $j \in \mathcal{M}$ соответствует фиксированная функция распределения $\Phi_j(x)$ с плотностью $\phi_j(x)$. Эти функции задают поведение системы в стационарном режиме.
Модель состоит из чередующихся фаз:

- стационарной фазы в состоянии j, и
- переходной фазы между состояниями $j \to k$.

Модель

Пусть:

- p_j — априорная вероятность начала траектории в состоянии $j \in \mathcal{M}$,
- $\chi_j(T)$ — функция распределения длительности пребывания в состоянии jj,
- p_{jk} — вероятность перехода из состояния j в состояние k,
- $\Lambda_{jk}(\tau)$ — функция распределения длительности перехода от j к k.

Модель

Пусть:

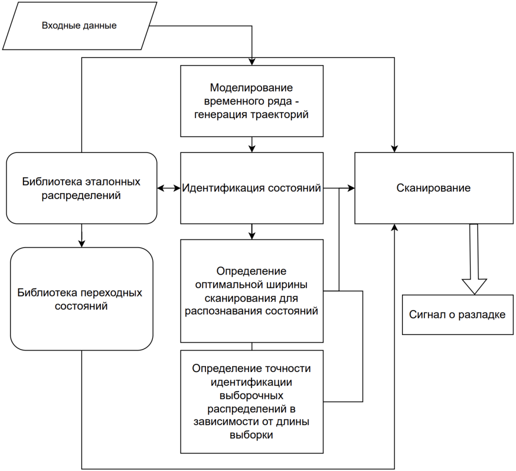
- p_j — априорная вероятность начала траектории в состоянии $j \in \mathcal{M}$,
- $\chi_j(T)$ — функция распределения длительности пребывания в состоянии jj,
- p_{jk} — вероятность перехода из состояния j в состояние k,
- $\Lambda_{jk}(\tau)$ — функция распределения длительности перехода от j к k.

Переход осуществляется через промежуточное состояние, описываемое функцией:

$$\Psi_{jk}(x) = \frac{1}{2}\left(\Phi_j(x) + \Phi_k(x)\right)$$

которая аппроксимирует распределение выборки в середине перехода.

Общая схема подхода



Базис эмпирических распределений

Для моделирования переходного распределения используется рекуррентная формула:

$$f(i,k+1) = f(i,k) + \frac{f(i,k)}{T}\frac{F_0(i-1) - F_1(i-1)}{\left(1-\frac{k}{T}\right) f_0(i) + \frac{k}{T} f_1(i)}$$

$$-\frac{f(i+1,k)}{T}\frac{F_0(i) - F_1(i)}{\left(1-\frac{k}{T}\right) f_0(i+1) + \frac{k}{T} f_1(i+1)}$$

Это решение динамически корректно, так как сохраняет полугруппу.

Базис эмпирических распределений

Таким образом,

- В каждый момент перехода мы точно знаем, из какого распределения нужно генерировать значения.
- Это позволяет моделировать переходную фазу так же осмысленно, как и стационарные участки.

Структура программного комплекса

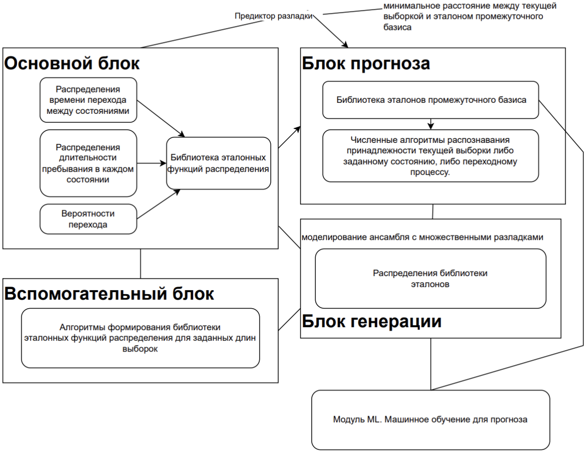
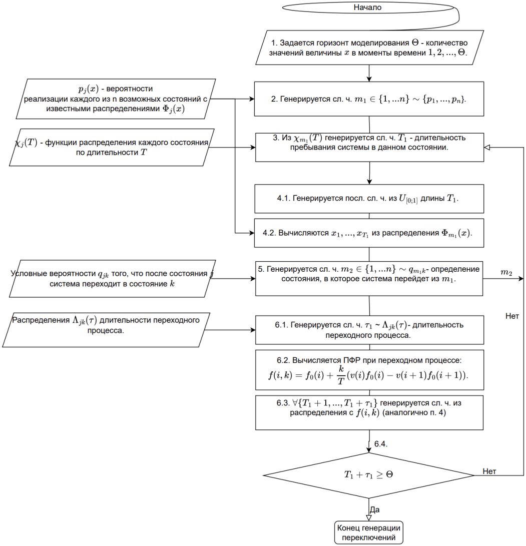


Рис.: Схема программного комплекса

Алгоритм генерации траектории



Генерация траектории

1. Сначала случайным образом выбирается начальное состояние $j \sim p_j$.
2. Из закона $\chi_j(T)$ генерируется длительность T — сколько система пробудет в этом состоянии.
3. На этом интервале T по заданному распределению $\Phi_j(x)$ **генерируются случайные значения** — стационарный участок.
4. Затем выбирается следующее состояние $k \sim p_{jk}$ и длительность перехода $\tau \sim \Lambda_{jk}(\tau)$.
5. В течение перехода используется формула вычисляется ПФР на переходном процессе. Из него пошагово генерируются значения.
6. Если суммарная длина траектории меньше заданной, цикл продолжается.

Результат генерации

1. **Итог**
   - Через конечное число шагов построена траектория случайного процесса со случайными переключениями.
   - Для построения ансамбля таких траекторий описанная процедура повторяется N раз.

Результат генерации

1. **Итог**
   - Через конечное число шагов построена траектория случайного процесса со случайными переключениями.
   - Для построения ансамбля таких траекторий описанная процедура повторяется N раз.
2. **Далее...**
   - построение **предиктора разладк**
   - оптимизация управляющих функционалов вдоль траекторий.

Предиктор разладки

Задача

Как можно раньше обнаружить момент времени t, при котором на заданном уровне значимости будет спрогнозирован переход.

Решение: построить модель предиктора перехода из одного состояния в другое на основе анализа изменения выборочной функции распределения в скользящем окне.

1. Сканируем ряд окном длины N, классифицируем фрагменты по состояниям, сопоставляя с уже имеющимися эталонами методом ближайшего соседа.
   - К этому моменту уже есть библиотека эталонов $f_k^{et}(x;N)$, где N - длина выборки, по которой собирались эталоны, k - номер состояния, x - случайная величина.

$$r_k(n;N) = \left\| f(x;N;n-N+1) - f_k^{et}(x;N) \right\| .$$

Переходный эталон

- функция, получающейся при сдвиге на промежуток времени, равный половине длины выборки:

$$\frac{1}{2}\left\| f_i^{et}(x;N) - f_j^{et}(x;N) \right\| > \varepsilon_0(N)$$

Итеративно в скользящем окне

- Вычисляем минимальное по всем состояниям расстояние, и по методу ближайшего соседа определяем, к какому состоянию относится текущий фрагмент.
- Если это расстояние больше, чем критерий - считаем, что фрагмент находится в переходном состоянии
  $\Rightarrow$ сравниваем с переходным базисом.

Это и есть момент регистрации смены состояния.

Новизна

До настоящего момента не моделировались нестационарные переходы в виде

- динамически корректного уравнения Лиувилля
- и возможности тестирования на этих траекториях некоторого функционала управления.

Выводы

Программный продукт позволит...

- выделять из временного ряда базисные паттерны;
- строить промежуточный базис для создания предиктора перехода;
- тестировать различные алгоритмы идентификации разгадок;
- генерировать ансамбль траекторий с нестационарным поведением

For further research

Перспективы

- реализация модели в ПО
- тестирование на специально заданных примерах
- применение к анализу реального ряда (например, последовательности промежутков времени между землетрясениями)

На данный момент происходит обсуждение и подбор наиболее подходящего временного ряда для дальнейшей реализации модели.

Список литературы

- Андрианова Е.Г., Головин С.А., Зыков С.В., Лесько С.А., Чукалина Е.Р. Обзор современных моделей и методов анализа временных рядов динамики процессов в социальных, экономических и социотехнических системах. Российский технологический журнал. 2020;8(4):7-45.
- Воронина М.Ю., Кислицын А.А., Орлов Ю.Н. Алгоритм коррекции метода биграмм в задаче идентификации автора текста // Математическое моделирование. 2022. Т. 34. № 9. С. 3-20.
- Гнеденко Б.В. Курс теории вероятностей. – М.: Физматлит, 1961. – 406 с
- Кислицын А.А., Кислицына М.Ю. Распознавание выборочных распределений среди системы эталонов: метод ближайшего соседа // Препринты ИПМ им. М.В. Келдыша. 2023. № 29. С. 1-21.
- Кислицын А.А., Козлова А.Б., Корсакова М.Б., Машеров Е.Л., Орлов Ю.Н. Стационарная точка уровня значимости для нестационарных функций распределения // Препринты ИПМ им. М.В. Келдыша. 2018. № 113. 20 с
- Кислицына М.Ю., Орлов Ю.Н. Статистический анализ полного корпуса художественной литературы на русском языке и распознавание автора // Препринты ИПМ им. М.В. Келдыша. 2024. № 17. С. 1-24.
- Кизбикенов, К. О. Прогнозирование и временные ряды [Электронный ресурс] : учебное пособие / К. О. Кизбикенов. – Барнаул : АлтГПУ, 2017.
- Орлов Ю.Н. Кинетические методы исследования нестационарных временных рядов. – М.: МФТИ, 2014. – 276 с.
- De Gooijer, J. G., Hyndman, R. J. (2006). 25 years of time series forecasting. International Journal of Forecasting, 22(3), 443-473.
- Januschowski, T., Gasthaus, J., Wang, Y., Salinas, D., Flunkert, V., Bohlke-Schneider, M., Callot, L. (2019). Criteria for classifying forecasting methods. International Journal of Forecasting.
- Hyndman, R. J., Athanasopoulos, G. (2021). Forecasting: Principles and Practice (3rd ed.).

Спасибо за внимание!
kolotova.aa@phystech.edu

Приложение. О генерации

- Имеется 2 набора эталонных частот $f(y)$ и $g(y)$, также дана выборка x_i из равномерного распределения и 2 функции эмпирического распределения - $F(y_i), G(y_k)$.
- Есть задачи: 1) понимать, правильно ли прошло распознавание вообще, и 2) научиться сравнивать эталонные частоты между собой, чтобы узнавать, какой эталон лучше описывает ситуацию.
- Мы умеем генерировать выборки из прообразов $\{y_1, \dots, y_N\}$, принадлежащие непрерывному распределению, по имеющейся выборке $\{x_1, \dots, x_N\}$ из равномерного распределения (будем считать, что она имеет некоторые выборочные частоты $p_N(j), j = 1, \dots, n.$).

Приложение. О генерации

- Как мы генерируем? Путем обращения F и G соответственно (так как $y_k = F^{-1}(x_k)$). Обратив F, мы получим игреки, которые можно "загрузить"в G - получим новую выборку $\{z_1, \ldots, z_N\}$ с частотами, обозначаемыми через $q_N(j), j = 1, \ldots, n$. Далее идет сравнение с теоретическими частотами $1/n$ в норме L_1:
$S_1 = \sum_{j=1}^{n} \left| p_N(j) - 1/n \right|, S_2 = \sum_{j=1}^{n} \left| q_N(j) - 1/n \right|$
- Если $S_1 > S_2$, то выборка, которую мы сгенерировали по эталону иксов с частотами f, ближе к функции G, чем к функции F, и мы считаем, что распознавание прошло неверно: то есть, не должно быть ситуации, что G ближе к равномерному распределению, чем F. Если $S_1 < S_2$, то функция $F(y)$ непрерывного распределения адекватно подгоняет исходную выборку.