

Анализ временных рядов на основе алгебраических методов распознавания образов (концепции активного сенсора и многомасштабного анализа)

Алексей Кузьмин, Михаил Кумсков
Механико-математический факультет

Московский Государственный Университет имени М.В.Ломоносова, Москва, Россия
Aleksej.Kyzmin@gmail.com, kumskov@mail.ru

Аннотация

Задача прогнозирования временных рядов – одна из наиболее востребованных в современной математике. Её решения применяются в различных областях: начиная от задач обработки сигналов и заканчивая финансовыми рынками.

В статье предлагается применить к решению данной задачи алгоритмы и методы из теории распознавания образов. В ней будут введены основные определения, показан общий алгоритм действий, а так же сформулированы проблемы и трудности, которые могут возникнуть в ходе её решения.

Ключевые слова: Временные ряды, прогнозирование, распознавание образов.

1. ПОСТАНОВКА ЗАДАЧИ

Опр. Временной ряд - собранный в разные моменты времени статистический материал о значении каких-либо параметров (в простейшем случае одного) исследуемого процесса. Каждая единица статистического материала называется измерением или отсчётом, также допустимо называть его уровнем на указанный с ним момент времени.

Опр. Измерение t_0 временного ряда $x(t)$ называется **локальным минимумом (локальным максимумом)**, если $x(t_0 - 1) > x(t_0)$ и $x(t_0 + 1) > x(t_0)$ ($x(t_0 - 1) < x(t_0)$ и $x(t_0 + 1) < x(t_0)$).

Опр. Временной ряд называется **растущим** на отрезке $[a, b]$, если каждый следующий локальный минимум на этом отрезке больше, чем предыдущий.

Опр. Временной ряд называется **убывающим** на отрезке $[a, b]$, если каждый следующий локальный максимум на этом отрезке меньше, чем предыдущий.

Обозначим как $\hat{z}_t(l)$ – прогноз, сделанный в момент времени t , относительно значения временного ряда z_{t+l} в момент времени $t + l$ (предполагается, что $l > 0, l \in \mathbb{N}$).

Функция $\hat{z}_t(l)$, дающая в момент времени t , прогнозы для всех будущих времен l , будет называться прогностической.

Целью работы является разработка алгоритма построения прогностической функции временного ряда z_t , дающей оптимальные прогнозы в ближайших к текущему моменту N точках. В простейшем случае требуется ответ на вопрос – будет ли расти или убывать временной ряд на следующем измерении.

2. ПОСТРОЕНИЕ ПРОГНОСТИЧЕСКОЙ ФУНКЦИИ

Построение прогностической функции мы разобьём на несколько этапов:

2.1 Выбор временного разрешения ряда

Пусть нам дан временной ряд $x(t)$, определенный на отрезке $t \in [a, b]$.

Построим на основе него другой временной ряд по следующему правилу:

Зафиксируем временную единицу M (M , к примеру, 1 минута, 2 часа, день, неделя и т.д.)

Разобьём интервал $[a, b]$ на отрезки длины M . Получим систему отрезков

$[a + (k-1)*M, a + k*M]$, $k \in \mathbb{N}, k > 1, k < \frac{b-a}{M}$ и отрезок $[a + \text{floor}(\frac{b-a}{M}) * M, b]$, где $\text{floor}(\frac{b-a}{M})$ – максимальное целое, не превосходящее $\frac{b-a}{M}$.

Из каждого полученного отрезка выберем измерение $t_k = \max_t \{t \in [a + (k-1) * M, a + k * M]\}$. На последнем отрезке $t_{\text{floor}(\frac{b-a}{M})+1} = \max_t \{t \in [a + \text{floor}(\frac{b-a}{M}) * M, b]\}$.

Построим новый временной ряд, включив в него следующие пары {измерение, значение}:

$\{a + M, x(t_1)\}; \{a + 2M, x(t_2)\}; \{a + 3M, x(t_3)\}; \dots \{b, x(t_{\text{floor}(\frac{b-a}{M})+1})\}$;

Полученный временной ряд мы будем называть M -тым срезом временного ряда X и обозначать $X^M(t)$

Примеры: одноминутный срез временного ряда X : $X^{1\text{мин}}(t)$

Введение такого определения обусловлено тем, что один и тот же временной ряд может демонстрировать разное поведение на разных временных срезах. Например, на одноминутном срезе ряд может быть убывающим, а на часовом срезе – возрастающим.

2.2 Выбор временного окна для проведения анализа ряда

Опр. Временным окном размера k измерения t_0 временного ряда X называется подряд ряда X , состоящий из измерений $t_0 - k + 1; t_0 - k + 2; t_0 - k + 3; \dots t_0$.

Обозначение: $U_k(x; t_0)$

Рассматривание всего временного ряда, особенно если количество измерений велико, может серьезно ухудшить вычислимость. Более того, если временной ряд содержит очень старые элементы, то они могут негативно сказываться на качестве прогноза (с течением времени факторы, определяющие значения ряда, могут изменяться. В таком случае старые измерения будут основаны на факторах, которые уже не играют роли в формировании текущих

значений, и в тоже время они могут не учитывать факторы, которые теперь актуальны.).

Поэтому выбор размера временного окна для прогноза – важная задача, которую требуется решить.

Выбор временного окна, по сути, является результатом применения концепции активного сенсора. Когда для построения прогноза мы смотрим на часть данных, отбрасывая остальные за ненадобностью. В действительности, обладая достаточно производительной машиной, мы можем строить анализ на основе динамически изменяемого временного окна. Однако алгоритм этого динамического изменения пока ещё не разработан.

2.3 Построение индикаторов

Опр. Индикатор – временной ряд $g(t)$ построенный на основе преобразования (возможно, не взаимно-однозначного) исходного временной ряда $x(t)$.

Индикаторы – характеристики исходного временного ряда. Классическим примером индикатора можно считать, например, скользящее среднее.

Опр. Скользящее среднее с периодом N – индикатор, значение которого в каждом измерении t_k определяется по

$$g(t_k) = \frac{\sum_{i=t_k-N}^{t_k} x(t_i)}{N}.$$

Основная задача этого этапа – составить набор индикаторов $\{g_1; \dots; g_n\}$, на основании которых мы будем прогнозировать наш временной ряд.

2.4 Построение разметки каждого индикатора

Для каждого индикатора будет построена разметка $R_g(U_k(g; t))$, превращающая временное окно $U_k(g; t)$ индикатора g в строку вида $\{a_1; \dots; a_l\}$.

В рамках данного преобразования требуется разработать алфавит, которым можно задавать изменение индикатора. В качестве примера такого алфавита можно рассмотреть, например, следующий:

I – значения временного ряда за промежутки значимо не изменялись

G – значения временного ряда выросли, без дальнейшего отката назад

F – значения временного ряда уменьшились, без дальнейшего отката к исходным значениям

A – значения временного ряда возросли, а затем вернулись к исходному уровню

V – значения временного ряда уменьшились, а затем вернулась к исходному значению

Примеры:



Figure 1. Пример построения алфавита.



Figure 2. Пример построения алфавита.

В задаче построения алфавита встает вопрос о выборе минимального значимого изменения временного ряда (т.е. когда I должна превращаться в другие буквы), а так же выбор максимальной и минимальной длины участка. В простейшем случае, их можно считать константными. Например, 5 букв на все временное окно.

Опр. Полученная таким образом строка $\{a_1; \dots; a_l\}$ будет называться характеристической строкой индикатора g .

Далее, на основе функций разметки каждого из индикаторов будет построена общая функция разметки по следующему правилу:

$$R(U_k(g_1; t_0); \dots; U_k(g_n; t_0);) = \{R_{g_1}; R_{g_2}; \dots; R_{g_n}\}$$

Опр. Итоговая строка, состоящая из характеристических строк каждого из временных окон индикаторов $g_1; \dots; g_n$, будет называться характеристической строкой временного окна $U_k(x; t_0)$ исходного временного ряда x .

2.5 Построение функции расстояния

На множестве полученных выше характеристических строк временного окна $U_k(x; t_0)$ будет построена функция расстояния. А точнее, некая мера.

Для её построения, можно, например, ввести матрицу расстояний, значения которой можно вычислять эмпирически.

2.6 Проведение кластерного анализа

На основе функции расстояния будет проведен кластерный анализ, где множество характеристических строк φ_i будет разбито на кластеры $K_1; \dots; K_n$. Для решения этой задачи можно использовать алгоритмы кластеризации, используемые в биологии (1).

2.7 Исследование зависимости полученных кластеров от разрешения ряда, временного окна, набора индикаторов, функции расстояния и разметки

Будет произведена аналитическая оценка зависимости полученных кластеров от данных параметров.

2.8 Выбор оптимальных параметров

Пусть в измерении t_0 временного ряда X нас интересует прогноз на измерение t .

Тогда на основе исторических данных мы можем максимизировать следующие вероятности:

$$\forall i \exists x_t(K_i) \forall \varphi \in K_i | P(x\varphi(t) = x_t(K_i)) \rightarrow \max$$

Где $x\varphi(t)$ – временной ряд, характеристическая строка которого для измерения t_0 равна φ .

Т.е. принадлежность характеристической строки кластеру должно с максимальной точностью определять прогнозируемое значение временного ряда на момент t .

Максимизация будет проводиться по разрешению временного ряда, временному окну, набору индикаторов, функции расстояния и разметки.

3. ЗАКЛЮЧЕНИЕ

В статье был сформулирован общий алгоритм применения теории распознавания образов (алгебраического подхода) к решению задачи прогнозирования временных рядов. Были сформулированы основные определения, а так же указаны некоторые трудности, которые могут возникнуть в ходе применения данного алгоритма на практике.

4. ССЫЛКИ

[1] Hand D., Mannila H., Smyth P. *Principles of Data Mining* The MIT Press - 2001.

[2] Журавлев Ю.И. *Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть I*, Кибернетика. — 1977. — № 4. — С. 5-17.

[3] Журавлев Ю.И. *Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть II*, Кибернетика. — 1977. — № 6. — С. 21-27.

[4] Журавлев Ю.И. *Корректные алгебры над множествами некорректных (эвристических) алгоритмов. Часть III*, Кибернетика. — 1978. — № 2. — С. 35-43.

[5] Журавлев Ю.И. *Об алгебраическом подходе к решению задач распознавания или классификации*, Проблемы кибернетики. — 1978. — Т. 33. — С. 5-68.

[6] Журавлев Ю.И. *Об алгебраических методах в задачах распознавания и классификации*, Распознавание, классификация, прогноз. — 1988. — Т. 1. — С. 9-16.

Об авторах

Михаил Кумсков – профессор Механико-математического факультета МГУ имени М.В. Ломоносова. kumskov@mail.ru

Алексей Кузьмин – аспирант Механико-математического факультета МГУ имени М.В. Ломоносова.

Aleksej.Kyzmin@gmail.com