

Авторегрессионные алгоритмы прогнозирования*

И. В. Фадеев, Н. П. Ивкин, Н. А. Савинов, А. И. Корниенко, Д. С. Кононенко,
Р. Б. Джамтырова

fadeev.mipt@mail.ru, ivkinnikita@gmail.com, nikolay.savinov@phystech.edu,
aleksej.kornienko@phystech.edu, daniilkononenko@gmail.com, RaisaD90@yandex.ru

Московский физико-технический институт, ФУПМ, каф. «Интеллектуальные системы»

Работа посвящена исследованию свойств авторегрессионных алгоритмов прогнозирования и включает ряд кратких описаний и вычислительных экспериментов. В работу включены следующие алгоритмы: сингулярный структурный анализ (алгоритм «Гусеница»), авторегрессионное интегрированное скользящее среднее, метод гибких наименьших квадратов, локальное прогнозирование временных рядов, авторегрессионное прогнозирование событий, и многомерная авторегрессия.

Ключевые слова: авторегрессия, прогнозирование, временной ряд.

Метод SSA для прогнозирования временных рядов

Описание алгоритма.

Наблюдается система функций дискретного аргумента $(f_i^{(k)})_{i=1}^N$, где $k = 1, \dots, s$, s — число временных рядов, k — номер ряда, N — длина временного ряда, i — номер отсчета. Требуется разложить ряд в сумму компонент (используя метод главных компонент, см. описание алгоритма), интерпретировать каждую компоненту, и построить продолжение ряда $(f_i^{(k)})_{i=1}^{N+M}$ по выбранным компонентам.

Рассмотрим сначала одномерный временной ряд $(f_i)_{i=1}^N$. Выберем n такое, что $0 < n \leq N - 1$ — время жизни многомерной гусеницы. Пусть $\sigma = N - n + 1$ — длина гусеницы. Построим последовательность из n векторов в R^σ следующего вида:

$$Y^{(l)} \in R^\sigma,$$

$$Y^{(l)} = (f_{i+l-1})_{i=1}^\sigma$$

Обозначим

$$Z = (Y^{(1)}, \dots, Y^{(n)}):$$

Будем называть Z нецентрированной матрицей наблюдений, порождённой гусеницей со временем жизни n . В случае многомерного временного ряда матрицей наблюдения называется столбец из матриц наблюдений, соответствующих каждой из компонент. Проводимый в дальнейшем анализ главных компонент может проводиться как по центрированной, так и по нецентрированной выборкам. Для упрощения выкладок рассмотрим простейший нецентрированный вариант.

Рассмотрим ковариационную матрицу полученной выборки:

$$C = \frac{1}{n} Z Z^T.$$

Выполним её svd-разложение:

$$C = V \Lambda V^T,$$

Научный руководитель В. В. Стрижов

где $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_r)$ — диагональная матрица собственных чисел, $V = (v^{(1)}, \dots, v^{(r)})$, $(v^{(i)})^T v^{(j)} = \delta_{ij}$ — ортогональная матрица собственных векторов. Далее рассмотрим систему главных компонент:

$$U = V^T Z, U = (U^{(1)}, \dots, U^{(r)})^T.$$

После проведения анализа главных компонент обычно предполагается проведение операции восстановления исходной матрицы наблюдений по некоторому поднабору главных компонент, т. е. для $V' = (v^{(i_1)}, \dots, v^{(i_r)})$ и $U' = V'^T Z$ вычисляется матрица $Z' = V'U'$. Далее восстанавливаются исходные последовательности. В одномерном случае i -ая компонента восстановленного ряда есть среднее значение по i -ой диагонали восстановленной матрицы наблюдений Z' .

В многомерном случае усреднение проводится с учётом того, что матрица наблюдений состоит из подматриц, соответствующих каждой компоненте ряда:

$$f'_m{}^{(k)} = \left\{ \begin{array}{ll} \frac{1}{m} \sum_{i=1}^m x_i^{(m-i+1,k)} & 1 \leq m \leq \sigma, \\ \frac{1}{\sigma} \sum_{i=1}^{\sigma} x_i^{(m-i+1,k)} & \sigma \leq m \leq n, \\ \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} x_{i+m-n}^{(n-i+1,k)} & n \leq m \leq N. \end{array} \right\}$$

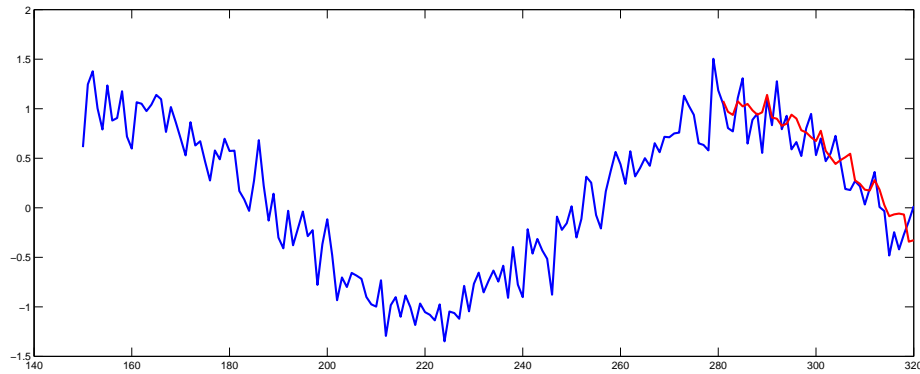


Рис. 1. Прогноз алгоритмом SSA

Числовой ряд $(f_i)_{i=1}^{N+1}$ называется продолжением ряда $(f_i)_{i=1}^N$, если порождаемая им при гусеничной обработке выборка лежит в той же гиперплоскости, что и у исходного ряда. Пусть у нас есть некоторый набор выбранных главных компонент i_1, i_2, \dots, i_r . Определим

$$w = \begin{pmatrix} v_{\sigma}^{(i_1)} & v_{\sigma}^{(i_2)} & \dots & v_{\sigma}^{(i_r)} \\ v_{2\sigma}^{(i_1)} & v_{2\sigma}^{(i_2)} & \dots & v_{2\sigma}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau}^{(i_1)} & v_{\tau}^{(i_2)} & \dots & v_{\tau}^{(i_r)} \end{pmatrix}$$

и

$$V_* = \begin{pmatrix} v_1^{(i_1)} & v_1^{(i_2)} & \dots & v_1^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\sigma-1}^{(i_1)} & v_{\sigma-1}^{(i_2)} & \dots & v_{\sigma-1}^{(i_r)} \\ v_{\sigma+1}^{(i_1)} & v_{\sigma+1}^{(i_2)} & \dots & v_{\sigma+1}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{2\sigma-1}^{(i_1)} & v_{2\sigma-1}^{(i_2)} & \dots & v_{2\sigma-1}^{(i_r)} \\ \vdots & \vdots & \ddots & \vdots \\ v_{\tau-1}^{(i_1)} & v_{\tau-1}^{(i_2)} & \dots & v_{\tau-1}^{(i_r)} \end{pmatrix}$$

Также положим

$$Q = \left(f_{N-\sigma+2}^{(1)}, \dots, f_N^{(1)}, f_{N-\sigma+2}^{(2)}, \dots, f_N^{(2)}, \dots, f_{N-\sigma+2}^{(s)}, \dots, f_N^{(s)} \right)^T$$

Тогда прогнозируемые значения системы в точке $N + 1$ вычисляются по формуле:

$$f_{N+1} = w(V_*^T V_*)^{-1} V_*^T Q.$$

Тестовый прогноз. Протестируем алгоритм на тестовом временном ряде, синтезируемом как сумма линейной компоненты (тренда), двух сезонных компонент (синусоид разной частоты) и нормального шума. На графике представлен исходный временной ряд (синий) и прогнозируемый с помощью алгоритма SSA (красный):

Авторегрессионное интегрированное скользящее среднее

Постановка задачи. Решается задача прогнозирования временных рядов. Обучающей выборкой является временной ряд, называемый историей. Требуется спрогнозировать дальнейшее поведение некоторого показателя данного временного ряда.

Для нахождения решения предлагается использовать модель ARIMA[1]. Модель ARIMA является обобщением модели авторегрессионного скользящего среднего (ARMA). Модель ARMA является синтезом авторегрессионной модели и модели скользящего среднего.

Пусть задан временной ряд X_t , где t — целое число и X_t — вещественные числа. Авторегрессионная модель порядка p задается следующим образом:

$$X_t = c + \sum_{i=1}^p \varphi_i X_{t-i} + \varepsilon_t, \tag{1}$$

где φ_i — параметры модели, c — константа, ε_t — белый шум.

Модель скользящего среднего порядка q задается следующим образом:

$$X_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t, \tag{2}$$

где θ_i — параметры модели и $\varepsilon_t \propto N(0, \sigma^2)$.

Модель ARMA(p, q) задается следующим образом:

$$X_t = \sum_{i=1}^p \varphi_i X_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + c + \varepsilon_t. \tag{3}$$

Модель ARMA работает в предположение, что ряд стационарен. Приводить ряд к стационарному предлагается методом взятия последовательной разности. Данный метод не обязательно приводит исходный ряд к стационарному, но неплохо решает проблему наличия тренда. Метод заключается в построение нового ряда Z_t такого, что:

$$Z_t = \Delta X_t = X_t - X_{t-1}. \tag{4}$$

Модель ARIMA(p,q,d) заключается в приведение ряда к стационарному виду взятием последовательной разности d раз и в применение к новому временному ряду модели ARMA(p,q,d).

Определение параметров p, q, d . Для определения параметра d будем брать последовательную разность до тех пор, пока тест стационарности Дики-Фуллера [1] не опровергнет нулевую гипотезу для получившегося временного ряда. Параметры p и q предлагается находить минимизирующими функционал качества SSE. На практике значения этих параметров редко бывают больше трех.

Результаты работы алгоритма ARIMA, одномерный случай.

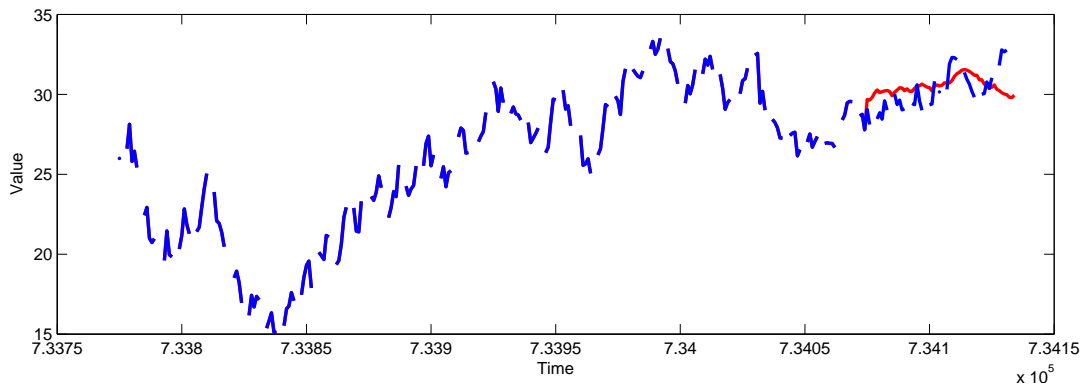


Рис. 2. Одномерный случай

Метод гибких наименьших квадратов

Общая постановка задачи и алгоритм решения. Решается задача восстановления нестационарной регрессии. Пусть задана линейная нормальная модель сигнала со скрытой компонентой:

$$\mathbf{x}_t = \mathbf{V}_t \mathbf{x}_{t-1} + \xi_t, \quad t = 2 \dots T, \tag{5}$$

$$y_t = \mathbf{c}_t^T \mathbf{x}_t + \eta_t, \quad t = 1 \dots T, \tag{6}$$

где $\mathbf{x}_t \in \mathbb{R}^n$ — скрытая компонента (вектор коэффициентов регрессии), $y_t \in \mathbb{R}$ — наблюдаемая компонента (прогнозируемая переменная); $\xi_t \in \mathbb{R}^n$ — внутренний шум, $\eta_t \in \mathbb{R}$ — внешний шум, $cov(\xi_t) = \frac{1}{\lambda} \mathbf{U}_t^{-1}$, λ — отношение интенсивностей внешнего и внутреннего шумов (гиперпараметр модели). Параметры $\mathbf{V}_t \in \mathbb{R}^{n \times n}$, $\mathbf{U}_t \in \mathbb{R}^{n \times n}$ задаются априори, а $\mathbf{c}_t \in \mathbb{R}^n$ — значение многомерного временного ряда, по которому выполняется прогнозирование.

Алгоритм получает на вход выборку $\{y_t, \mathbf{c}_t\}, t = 1 \dots T$, на выходе выдает набор значений коэффициентов регрессии $\{\mathbf{x}_t\}, t = 1 \dots T$. Для оценки коэффициентов используется

квадратичная функция потерь:

$$J(\mathbf{x}_1, \dots, \mathbf{x}_T) = \sum_{t=1}^n \psi_t(\mathbf{x}_t|y_t) + \lambda \sum_{t=2}^n \gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t), \text{ где} \quad (7)$$

$$\psi_t(\mathbf{x}_t|y_t) = (y_t - \mathbf{c}_t^T \mathbf{x}_t)^2 = (\mathbf{x}_t - \mathbf{x}_t^0)^T \mathbf{Q}_t^0 (\mathbf{x}_t - \mathbf{x}_t^0) \text{ — отклонение от модели сигнала } y_t, \quad (8)$$

$$\mathbf{x}_t^0 = \frac{y_t}{\mathbf{c}_t^T \mathbf{c}_t} \mathbf{c}_t, \quad \mathbf{Q}_t^0 = \mathbf{c}_t \mathbf{c}_t^T \text{ — точка минимума и матрица квадратичной формы } \psi_t(\mathbf{x}_t|y_t), \quad (9)$$

$$\gamma_t(\mathbf{x}_{t-1}, \mathbf{x}_t) = (\mathbf{x}_t - \mathbf{V}_t \mathbf{x}_{t-1})^T \mathbf{U}_t (\mathbf{x}_t - \mathbf{V}_t \mathbf{x}_{t-1}) \text{ — отклонение от модели скрытых переменных.} \quad (10)$$

Данная функция потерь объединяет 2 противоречивых требования к коэффициентам регрессии \mathbf{x}_t : точность восстановления сигнала y_t (первое слагаемое) и гладкость изменения самих коэффициентов регрессии (второе слагаемое). Компромисс между этими требованиями достигается выбором гиперпараметра λ .

Процедура оценивания сводится к применению фильтра-интерполятора Калмана-Бьюси и выполняется в 2 этапа.

1. Фильтрация:

- (а) $\mathbf{x}_{1|1} = \mathbf{x}_1^0, \mathbf{Q}_{1|1} = \mathbf{Q}_1^0, \quad t = 1,$
- (б) $\mathbf{Q}_{t|t} = \mathbf{Q}_t^0 + (\mathbf{V}_t \mathbf{Q}_{t-1|t-1} \mathbf{V}_t^{-1} + \frac{1}{\lambda} \mathbf{U}_t^{-1}), \quad t = 2 \dots T,$
- (в) $\mathbf{k}_t = \mathbf{Q}_{t|t}^{-1} \mathbf{c}_t, \quad t = 2 \dots T,$
- (г) $\mathbf{x}_{t|t} = \mathbf{V}_t \mathbf{x}_{t-1|t-1} + \mathbf{k}_t (y_t - \mathbf{c}_t^T \mathbf{V}_t \mathbf{x}_{t-1|t-1}), \quad t = 2 \dots T.$

2. Интерполяция:

- (а) $\mathbf{H}_t = (\mathbf{I} + \frac{1}{\lambda} \mathbf{U}_{t+1}^{-1} (\mathbf{V}_{t+1}^{-1})^T \mathbf{Q}_{t|t})^{-1}, \quad t = T - 1 \dots 1,$
- (б) $\hat{\mathbf{x}}_t = \mathbf{x}_{t|t} + \mathbf{H}_t (\hat{\mathbf{x}}_{t+1} - \mathbf{x}_{t|t}), \quad t = T - 1 \dots 1.$

Для оценки гиперпараметра λ используется метод скользящего контроля LOO. В статье [10] получены следующие соотношения для значений параметров $\hat{\mathbf{x}}_t$ на основании данных с пропущенным значением y_t :

$$\hat{\mathbf{x}}_t(\mathbf{Y}^{(t)}, \lambda) = (\mathbf{Q}_{t|T}^{(t)})^{-1} (\mathbf{Q}_{t|T} \mathbf{x}_{t|T} - \mathbf{Q}_t^0 \mathbf{x}_t^0), \quad (11)$$

$$\mathbf{Q}_{t|T}^{(t)} = \mathbf{Q}_{t|T} - \mathbf{Q}_t^0, \text{ где } (t) \text{ означает выброшенное измерение } y_t; \quad (12)$$

$$\mathbf{Q}_{t|T} = (\mathbf{H}_t \mathbf{V}_{t+1}^{-1} \mathbf{Q}_{t+1|T}^{-1} (\mathbf{V}_{t+1}^{-1})^T \mathbf{H}_t^T + (\mathbf{Q}_{t|t} + \lambda \mathbf{V}_{t+1}^T \mathbf{U}_{t+1} \mathbf{V}_{t+1})^{-1})^{-1}. \quad (13)$$

Тогда внешний критерий LOO вычисляется по определению:

$$LOO(\lambda) = J(\hat{\mathbf{x}}_1(\mathbf{Y}^{(1)}, \lambda), \dots, \hat{\mathbf{x}}_T(\mathbf{Y}^{(T)}, \lambda)).$$

При решении задачи значение λ выбирается из условия минимизации внешнего критерия:

$$\hat{\lambda} = \arg \min_{\lambda} LOO(\lambda).$$

Процедура обладает линейной сложностью по длине T временного ряда.

Для применения алгоритма следует задать $\mathbf{U}_t, t = 1 \dots T$ (на практике можно использовать диагональные матрицы), а также $\mathbf{V}_t, t = 1 \dots T$, которые определяют модель изменения скрытой компоненты. В исходной работе [13] используются единичные матрицы.

Применение для прогнозирования. Сначала производится оценка λ по известной истории временных рядов $hist$. Далее выполняется прогнозирование в заданном временном интервале frc .

1. Прогнозируется \mathbf{c}_t , $t \in frc$ (с помощью экспоненциального сглаживания).
2. Прогнозируется предварительное значение $y_{t|t-1}$ (с помощью экспоненциального сглаживания).
3. FLS оценивает сглаженный коэффициент регрессии $\mathbf{x}_{t|t}$.
4. Вычисляется сглаженное значение $y_{t|t} = \mathbf{c}_t^T \mathbf{x}_{t|t}$.
5. $t := t + 1$.

См. исходный код: [9]. Тестовый прогноз представлен на рис. 3.

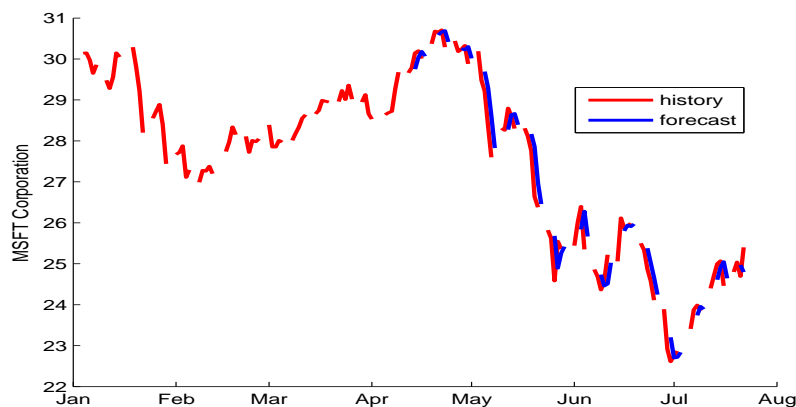


Рис. 3. Цена акций компании Microsoft

Локальные методы прогнозирования временных рядов

В работе исследуется метод прогнозирования временных рядов. Все методы построения прогноза временных рядов можно разделить на локальные и глобальные. Глобальные методы используют всю известную историю для построения прогноза. Локальные используют только часть истории. Мы исследуем алгоритм, основанный на методе k ближайших соседей. Мы решаем задачу настройки параметров алгоритма и параметров метрики.

К параметрам алгоритма относятся:

1. Число ближайших соседей.
2. Метрика.
3. Длина рассматриваемого вектора предыстории.

Рассмотренные методы были ранее предложены в работах [14, 15].

Постановка задачи прогнозирования. Заданы несколько временных рядов, один из которых требуется спрогнозировать на h шагов вперед. Предполагается, что прогнозируемый отрезок лежит сразу же за окончанием известных нам значений рядов, т.е. за концом истории. Будем считать, что в истории нет пропусков. Тогда в качестве объекта будем рассматривать последовательности точек длины l .

Требуется найти k наиболее похожих объектов на окончание истории в смысле выбранной метрики. Прогноз строится как линейная комбинация прогнозов соседей.

Описание алгоритма. Основную сложность составляет определение метрики, в смысле которой определяется близость объектов. Схема алгоритма:

1. Выделяем объект длины l .
2. Нормализуем выделенный объект. Вычитаем минимум и делим на разницу максимального и минимального значений по всем данным рядам.
3. Далее, находим расстояние между выделенным и эталонным объектами:

$$\text{dist}(x, y) = \min_{A, b} \rho(x, Ay + b).$$

В данной работе рассматриваются две метрики:

$$\rho(x, y, \omega) = \sum_{k=1}^l \omega^{2k} (x(k) - y(k))^2$$

и

$$\rho(x, y, \omega) = \sum_{k=1}^l \omega^{2k} |x(k) - y(k)|.$$

В случае, когда ряд — многомерный, вычисляется метрика для каждого ряда со своим параметром ω , и затем значения складываются. Параметр ω настраивается в ходе обучения модели.

4. Если объект оказался среди k ближайших на данный момент, запоминаем полученные параметры и расстояние. Выбрасываем из списка самый далекий от эталонного объект.
5. По результатам работы выделяем k лучших объектов и выделяем за ними ряд длиной достаточной, чтобы покрыть историю.
6. Из полученных рядов строим выпуклую линейную комбинацию с равными весами.
7. Выбираем те элементы спрогнозированного ряда, по которым необходим прогноз.

Вычислительный эксперимент. Для примера использовался модельный ряд

$$\mathbf{x} = \sin(8 \sin(\mathbf{a})) + \xi,$$

где \mathbf{a} — вектор арифметической прогрессии из 1000 точек, у которой $a_1 = 1$, $a_{1000} = 500$; $\xi \in \mathcal{N}(0, 0.1)$ — нормально распределенная случайная величина.

Обучение — первые 600 элементов ряда. Прогноз строится для 50 следующих временных отсчетов, причем для всех точек сразу.

На рисунке: синий (Time series) — исходный ряд, красный (Forecast) — спрогнозированные значения.

Прогнозирование событий

Рассматривается метод прогнозирования событий на основании информации о наличии событий во временных рядах, данной экспертом. Предлагается способ порождения и отбора признаков, описывающих временной ряд. Алгоритм основан на разметке интервалов роста и падения временного ряда и применении логистической регрессии для настройки параметров линейной модели и оценки ее качества.

Дана выборка временных рядов, размеченных экспертом:

$$\{\mathbf{x}^i, y^i\}, \mathbf{x}^i = (x_1^i, \dots, x_T^i) \in R^T, i = 1 \dots l.$$

Здесь целевая переменная:

$$y^i = \begin{cases} 1, & \text{в данном ряде есть событие;} \\ 0, & \text{в данном ряде нет события.} \end{cases}$$

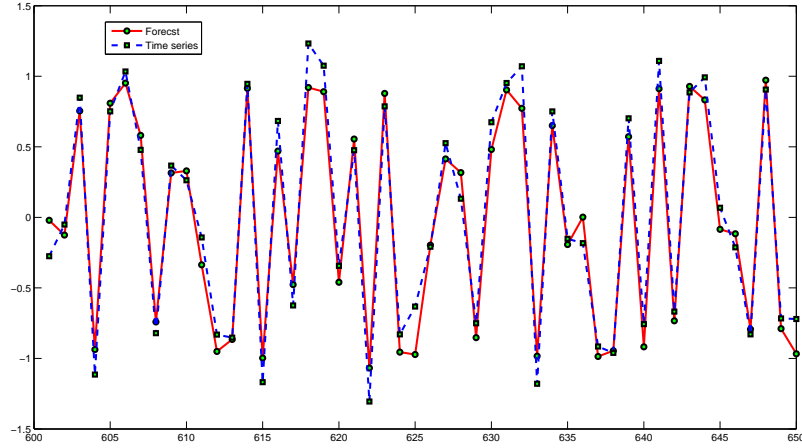


Рис. 4. Результаты прогнозирования

Необходимо предложить признаковое описание временного ряда:

$$g(x^i) = (g_1(x^i), \dots, g_n(x^i)).$$

На основании этого описания требуется решить задачу классификации — построить модель $f : R^w \times R^T \rightarrow \{0, 1\}$, где R^w — пространство параметров модели.

Задача разбивается на три этапа.

1. Порождение множества числовых признаков $(g_1(x), \dots, g_n(x))$, описывающих временной ряд.
2. Предложение критерия качества модели.
3. Выбор наилучшей модели.

Порождение множества признаков. Зафиксируем множество меток $\mathbf{M} = \{'U', 'D'\}$: 'U' — повышение значения ряда, 'D' — понижение (см. (14)). Размечаем временной ряд (m_1, \dots, m_T) :

$$m_k = \begin{cases} 'U', & \text{если } x_k - x_{k-1} > 0, \\ 'D', & \text{иначе.} \end{cases} \quad (14)$$

Таким образом, для каждого временного ряда \mathbf{x}^i получаем строчку из $(T - 1)$ символа 'U' или 'D': $\mathbf{m}^i = m_1^i \dots m_{T-1}^i$ — слово длины $(T - 1)$ в алфавите \mathbf{M} .

Рассмотрим теперь все слова длины не более w в алфавите \mathbf{M} . Таких слов $N = (2^w - 1)$ (исключаем пустое слово). Пронумеруем их: $R = (r_1, \dots, r_N)$. Слову $r_k = r_k^1 \dots r_k^{n_k}$ поставим в соответствие 2 числовых признака.

1. Бинарный признак

$$g_{2k-1}(x_i) = \begin{cases} 1, & \text{если в строке } \mathbf{m}^i \text{ есть подстрока } r_k; \\ 0, & \text{иначе.} \end{cases}$$

2. Действительный признак

$$g_{2k}(x_i) = \begin{cases} x_{j+n_k}^i - x_j^i, & \text{если в строке } \mathbf{m}^i \text{ есть подстрока } r_k : (m_{j+1}^i, \dots, m_{j+n_k}^i) = (r_k^1, \dots, r_k^{n_k}); \\ 0, & \text{если в строке } \mathbf{m}^i \text{ нет подстроки } r_k. \end{cases}$$

Если комбинация встречается в ряде несколько раз, то мы рассматриваем только ее последнее по времени вхождение (и записываем в действительный признак суммарное изменение цены на нем). Таким образом, мы получаем $2N = 2(2^w - 1)$ числовых признаков и матрицу объект-признак размера $l \times 2N$:

$$G = \begin{pmatrix} g_1(x^1) & \dots & g_{2N}(x^1) \\ \vdots & \ddots & \vdots \\ g_1(x^l) & \dots & g_{2N}(x^l) \end{pmatrix}.$$

Отбор наилучшей модели. Строить классификатор и оценивать качество модели будем с помощью логистической регрессии [17]. За оценку качества модели примем площадь под ROC-кривой данного классификатора.

Отбор наилучшей модели будем производить с помощью генетического алгоритма [16].

Фиксируем четное число — длину модели n (количество признаков в ней). Ранее мы построили множество слов $R = (r_1, \dots, r_N)$ длины не более w в алфавите \mathbf{M} и каждому слову r_k поставили в соответствие признаки g_{2k-1} и g_{2k} . Активное множество признаков \mathbf{A} — признаки, на основании которых строится текущая модель, $|\mathbf{A}| = n$. Поставим ему в соответствие активное множество индексов $\mathbf{I} = \{i_1, \dots, i_{\frac{n}{2}}\}$, $i_k \in \{1, \dots, N\}$, $|\mathbf{I}| = \frac{n}{2}$, следующим образом:

$$\mathbf{I} = \{i_1, \dots, i_{\frac{n}{2}}\} \Leftrightarrow \mathbf{A} = \{g_{2i_1-1}, g_{2i_1}, \dots, g_{2i_{\frac{n}{2}}-1}, g_{2i_{\frac{n}{2}}}\}.$$

Таким образом, признаки g_{2k-1} и g_{2k} всегда используются или не используются в модели вместе. Модель полностью описывается активным множеством индексов, поэтому далее генетический алгоритм опишем в терминах этих множеств.

Генетический алгоритм содержит следующие параметры для отбора моделей: F — число лучших моделей в популяции, F_1 — число моделей для скрещивания, P_2 — вероятность выбора модели для мутации.

Начальное множество моделей (популяцию) выбираем случайным образом. Делим все множество индексов $\{1, \dots, N\}$ случайным образом на подмножества размера $\frac{n}{2}$ (для простоты считаем, что $2N$ делится на n):

$$\{1, \dots, N\} = \{(k_1, \dots, k_{\frac{n}{2}}), (k_{\frac{n}{2}+1}, \dots, k_n), \dots\} = (\mathbf{I}_1, \dots, \mathbf{I}_P),$$

где $P = \frac{2N}{n}$ — размер популяции.

Итеративно выполняются следующие операции.

1. Отбор.

Модели сортируются по убывания качества (площадь под ROC-кривой), и в следующее поколение попадают F лучших.

2. Случайным образом выбираются F_1 моделей для скрещивания и мутации.

3. Скрещивание.

Выбранные модели случайным образом делятся на пары. В каждой паре активные множества индексов $\mathbf{I}_p = (k_1^p, \dots, k_{\frac{n}{2}}^p)$ и $\mathbf{I}_q = (k_1^q, \dots, k_{\frac{n}{2}}^q)$ разбиваются точкой кроссинговера m , выбираемой случайно из множества $\{1, \dots, \frac{n}{2}\}$, на 2 части. Происходит обмен элементов множеств \mathbf{I}_p и \mathbf{I}_q так, чтобы в новых множествах не было одинаковых индексов:

$$\begin{cases} (k_1^p, \dots, k_m^p, k_{m+1}^p, \dots, k_{\frac{n}{2}}^p) \\ (k_1^q, \dots, k_m^q, k_{m+1}^q, \dots, k_{\frac{n}{2}}^q) \end{cases} \mapsto \begin{cases} (k_1^p, \dots, k_m^p, k_{m+1}^q, \dots, k_{\frac{n}{2}}^q) \\ (k_1^q, \dots, k_m^q, k_{m+1}^p, \dots, k_{\frac{n}{2}}^p) \end{cases}.$$

4. Мутация.

Каждая модель $\mathbf{I} = (k_1, \dots, k_{\frac{n}{2}})$ из полученного множества с вероятностью P_2 подвергается мутации: случайным образом из множества $\{1, \dots, \frac{n}{2}\}$ выбирается число j , и значение k_j меняется на случайно выбранный индекс из множества $\{1, \dots, N\}$. При этом контролируется, чтобы в новом множестве индексов не было одинаковых индексов.

5. Оценка полученных моделей.

Новые модели настраиваются и оцениваются (см. пункт Оценивание качества модели).

После заданного числа шагов генетического алгоритма выбирается лучшая модель, таким образом получаем набор признаков (g_1, \dots, g_n) , на основании которых строится итоговый классификатор.

Многомерная авторегрессия

Многомерная авторегрессия является эконометрической моделью, которую используют для прогнозирования временными рядами. Многомерная авторегрессия является обобщением модели авторегрессии (AR). Даны временной ряд $\mathbf{s}_1 = [x_1, \dots, x_{T-1}]^T$, $x_i \in \mathbb{R}^1$ и матрица признаков, столбцами которой являются временные ряды $\mathbf{s}_2, \mathbf{s}_3 \dots \mathbf{s}_m$. Необходимо спрогнозировать следующую величину x_T ряда \mathbf{s}_1 .

Предполагается, что

- отсчеты сделаны через равные промежутки времени,
- ряд имеет периодическую составляющую,
- ряд не имеет пропущенных значений,
- длина ряда кратна периоду k .

Составляется $(m \times k)$ -матрица значений временного ряда:

$$S = \begin{pmatrix} x_T & x_{T-1} & \dots & x_{T-k+1} \\ x_{(m-1)k} & x_{(m-1)k-1} & \dots & x_{(m-2)k+1} \\ \dots & \dots & \dots & \dots \\ x_{nk} & x_{nk-1} & \dots & x_{n(k-1)+1} \\ \dots & \dots & \dots & \dots \\ x_k & x_{k-1} & \dots & x_1 \end{pmatrix}.$$

Введем обозначения

$$\left(\begin{array}{c|c} x_T & \mathbf{x}^T \\ \hline \mathbf{y} & X \end{array} \right).$$

В случае, когда учитываются временные ряды $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_m$, для каждого j -го временного ряда строится авторегрессионная матрица S_j и присоединяется справа. Полученная матрица

$$A = [S_1 \ S_2 \ \dots \ S_m].$$

Пусть задан набор функций $G = \{g_1, \dots, g_r\}$, например, $g_1 = 1$, $g_2 = \sqrt{x}$, $g_3 = x$, $g_4 = x\sqrt{x}$. Матрица порождённых признаков имеет вид

$$A = \left(\begin{array}{c|cccccc} x_T & g_1 \circ x_{T-1} & \dots & g_r \circ x_{T-1} & \dots & g_1 \circ x_{T-k+1} & \dots & g_r \circ x_{T-k+1} \\ x_{(m-1)k} & g_1 \circ x_{(m-1)k-1} & \dots & g_r \circ x_{(m-1)k-1} & \dots & g_1 \circ x_{(m-2)k+1} & \dots & g_r \circ x_{(m-2)k+1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{nk} & g_1 \circ x_{nk-1} & \dots & g_r \circ x_{nk-1} & \dots & g_1 \circ x_{n(k-1)+1} & \dots & g_r \circ x_{n(k-1)+1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_k & g_1 \circ x_{k-1} & \dots & g_r \circ x_{k-1} & \dots & g_1 \circ x_1 & \dots & g_r \circ x_1 \end{array} \right).$$

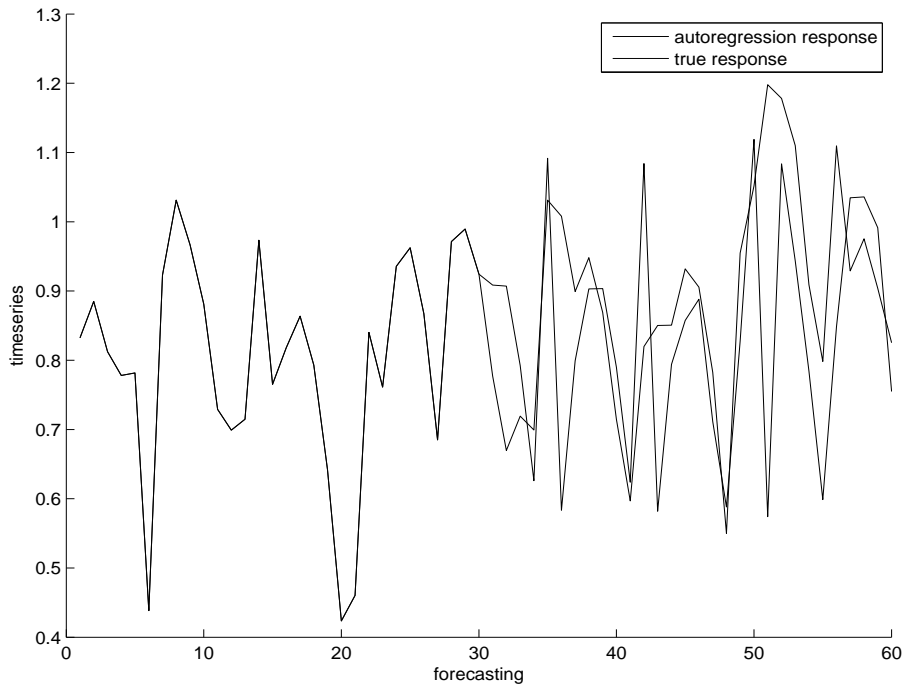


Рис. 5. Прогноз с помощью авторегрессионной матрицы

Пути решения задачи. Требуется решить задачу линейной регрессии $\|X\mathbf{w} - \mathbf{y}\|^2 \rightarrow \min$. Искомый вектор параметров имеет вид:

$$\mathbf{w} = (X^T X)^{-1} (X^T \mathbf{y}).$$

В терминах линейной регрессии

$$\mathbf{y} = X\mathbf{w}$$

$$y_T = \langle \mathbf{x}^T, \mathbf{w} \rangle$$

Возможны два подхода к нахождению значения \mathbf{x}_T . Первый заключается в последовательном построении прогнозируемых значений. В качестве вектора \mathbf{x}^T выбираются $K - 1$ предыдущих значений временного ряда, и для каждого следующего значения $x_{T+1}, x_{T+2}, \dots, x_{T+l}$ необходимо заново строить авторегрессионную матрицу X . Второй подход заключается в прогнозировании периода ряда по истории. В качестве вектора \mathbf{x}^T выбираются $K - 1$ значений ряда предшествующего периода. При данном подходе перестраивать авторегрессионную матрицу X не нужно. Результат работы алгоритма представлен на рис. 5.

Литература

- [1] Тихонов, А. Н. *Решение некорректно поставленных задач и метод регуляризации*, ДАН, 1963, Vol.151, pp.501-504.
- [2] Tibshirani, R. *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society, 1996, Vol.58, pp.267-288.
- [3] Hastie, T. and Taylor, J. and Tibshirani, R. and Walther, G. *Forward stagewise regression and the monotone lasso*, Electronic Journal of Statistics, 2007, Vol.1, pp.1-29.
- [4] Jerome H. Friedman *Greedy Function Approximation: A Gradient Boosting Machine*, Annals of Statistics, 2000, Vol.29, pp.1189-1232.

- [5] Jerome Friedman and Trevor Hastie and Robert Tibshirani *Additive Logistic Regression: a Statistical View of Boosting*, Annals of Statistics, 1998, Vol.28, p.2000.
- [6] Efron, B. and Hastie, T. and Johnstone, I. and Tibshirani, R. *Least angle regression*, Annals of Statistics, 2004, Vol.32, pp.407-499.
- [7] Draper, N. and Smith, H. *Applied Regression Analysis*, Wiley, 1996.
- [8] Магнус Я.Р., Эконометрика. Начальный курс: Учеб. - 6 изд. — М.: Дело, 2004. — 576 с.
- [9] Nikolay Savinov *Flexible Least Squares*, <https://mlalgorithms.svn.sourceforge.net/svnroot/mlalgorithms/TSForecasting/FlexibleLeastSquares/>, 2011.
- [10] Vadim Mottl and Olga Krasotkina and Michael Markov and Ilya Muchnik *Time-varying regression model with unknown time-volatility for nonstationary signal analysis*, 2006.
- [11] Vadim Mottl and Olga Krasotkina and Michael Markov and Ilya Muchnik *Dynamic analysis of hedge funds*, 2006.
- [12] Vadim Mottl *Lectures on statistical methods of signal and data array analysis*.
- [13] Kalaba, R. and Tesfatsion, L. *Time-varying linear regression via flexible least squares*, 1989.
- [14] Федорова В. П. *Локальные методы прогнозирования временных рядов*, МГУ им. М.В.Ломоносова, 2009.
- [15] J. McNames *Innovations in local modeling for time series prediction*, Stanford University, 1999.
- [16] В. В. Стрижов *Методы выбора регрессионных моделей*, ВЦ РАН, 2010.
- [17] К. В. Воронцов *Машинное обучение (курс лекций)*, 2009.
Стрижов В. В. *Методы выбора регрессионных моделей*, 2010.
- [18] V. Strijov. *Model Generation and its Applications in Financial Sector*, 2009.