Прогнозирование временных рядов

К. В. Воронцов vokov@forecsys.ru
А.А. Романенко alexromsput@gmail.com

29 Октября 2016

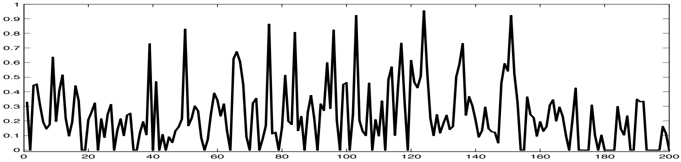
Содержание

Временной ряд

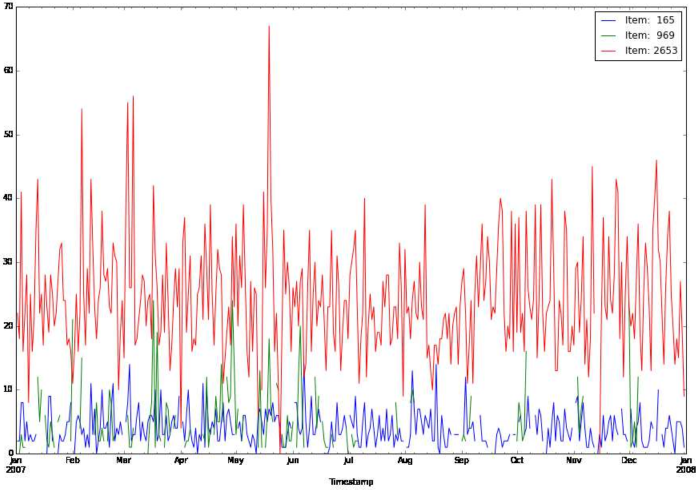
$y_0, y_1, \ldots, y_t, \ldots$ — временной ряд, $y_i \in \mathbb{R}$
$\hat{y}_{t+d}(w) = f_t(y_1, \ldots, y_t; w)$ — модель временного ряда,
где $d = 1, \ldots, D$, D — горизонт (отсрочка, delay),
w — вектор параметров модели

Особенности задачи:

- количество временных рядов $10^6 - 10^8$;
- пропуски в данных;
- нестационарность (непостоянство модели);
- несимметричная, кусочно гладкая функция потерь.



Примеры реальных временных рядов продаж



Эконометрика — основной источник задач прогнозирования

Примеры эконометрических временных рядов:

- объёмы продаж в торговых сетях
- объёмы грузовых и пассажирских перевозок
- рыночные цены
- дорожный трафик (прогнозирование пробок)
- объёмы потребления и цены электроэнергии

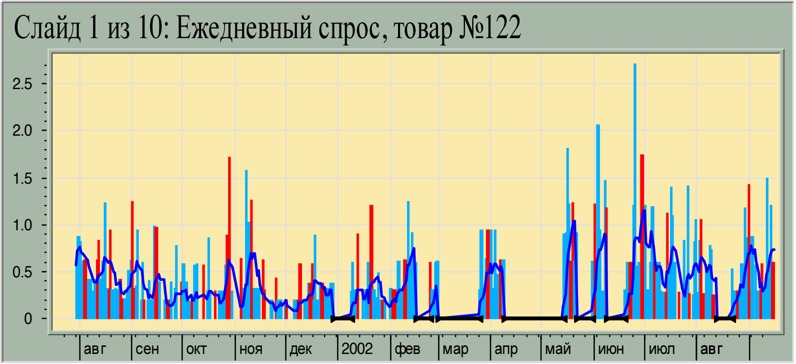
Основные явления в эконометрических временных рядах:

- тренды
- сезонности
- разладки (смены модели ряда)

Марно Вербик. Путеводитель по современной эконометрике, 2008.

Пример. Задача прогнозирования объёмов продаж

Ежедневные объёмы продаж товара



Особенности задачи: продажи зависят от типа товара, тренды, сезонность, пропуски, праздники, промо–акции, скачки, плохо работают сложные модели

Простое ЭСС

Простое ЭСС:

$$\hat{y}_{t+1} := \alpha y_t + (1-\alpha)\hat{y}_t = \hat{y}_t + \alpha \cdot e_t$$

$e_t = y_t - \hat{y}_t$ — ошибка прогноза отсчёта вр t
Чем больше α, тем больше вес последних отсчётов,
при $\alpha \to 1$ тривиальный прогноз $\hat{y}_{t+1} = y_t$.

Чем меньше α, тем сильнее сглаживание,
при $\alpha \to 0$ — $\hat{y}_{t+1} = \bar{y}$ (или скользящее среднее).

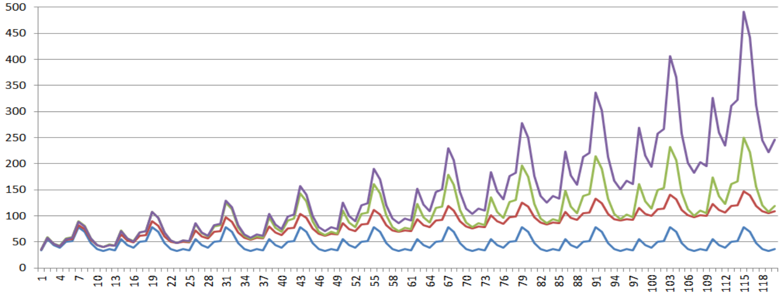
Оптимальное α^* находим по скользящему контролю:

$$Q(\alpha) = \sum_{t=T_0}^{T_1} \left(\hat{y}_t(\alpha) - y_t\right)^2 \to \min_{\alpha}$$

Эмпирические правила:
если $\alpha^* \in (0, 0.3)$, то ряд стационарен, ЭС работает;
если $\alpha^* \in (0.3, 1)$, то ряд нестационарен, нужна модель тренда.

Примеры трендов и сезонностей

Пример: сочетание тренда и сезонности (модельные данные)



Ряд 1 — сезонность без тренда

Ряд 2 — линейный тренд, аддитивная сезонность

Ряд 3 — линейный тренд, мультипликативная сезонность

Ряд 4 — экспоненциальный тренд, мультипликативная сезонность

Модель Хольта = линейный тренд

Линейный тренд без сезонных эффектов:

$$\hat{y}_{t+d} = a_t + b_t d,$$

где a_t, b_t — адаптивные коэффициенты линейного тренда

Рекуррентная формула:

$$a_t := \alpha_1 y_t + (1-\alpha_1)(a_{t-1} + b_{t-1}) = \hat{y}_t + \alpha_1 e_t;$$
$$b_t := \alpha_2 (a_t - a_{t-1}) + (1-\alpha_2) b_{t-1} = b_{t-1} + \alpha_1 \alpha_2 e_t.$$

Частный случай — модель линейного роста Брауна:

$$\alpha_1 = 1 - \beta, \quad \alpha_2 = 1 - \beta \ \text{(или } 1 + \beta\text{)}.$$

Модель Уинтерса = мультипликативная сезонность

Мультипликативная сезонность периода s:

$$\hat{y}_{t+d} = a_t \cdot \theta_{t-s+(d \bmod s)},$$

$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.
Рекуррентная формула:

$$a_t := \alpha_1(y_t/\theta_{t-s}) + (1-\alpha_1)a_{t-1} = a_{t-1} + \alpha_1 e_t/\theta_{t-s};$$
$$\theta_t := \alpha_2(y_t/a_t) + (1-\alpha_2)\theta_{t-s} = \theta_{t-s} + \alpha_2(1-\alpha_1)e_t/a_t.$$

Доказательство последнего равенства:

$$\theta_t := \theta_{t-s} + \alpha_2\left(y_t/a_t - \theta_{t-s}\right) = \theta_{t-s} + \alpha_2\left(y_t - \theta_{t-s}a_t\right)/a_t =$$
$$\theta_{t-s} + \alpha_2\Big(y_t - \theta_{t-s}(a_{t-1} + \alpha_1 e_t/\theta_{t-s})\Big)/a_t = \theta_{t-s} +$$
$$+\alpha_2\left(\underbrace{y_t - \theta_{t-s}a_{t-1}}_{e_t} - \alpha_1 e_t\right)/a_t$$

Модель Тейла–Вейджа

Линейный тренд с аддитивной сезонностью периода s:

$$\hat{y}_{t+d} = (a_t + b_t d) + \theta_{t+(d \bmod s)-s}.$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1(y_t - \theta_{t-s}) + (1-\alpha_1)(a_{t-1} + b_{t-1}) = a_{t-1} + b_{t-1} + \alpha_1 e_t;$$
$$b_t := \alpha_2(a_t - a_{t-1}) + (1-\alpha_2)b_{t-1} = b_{t-1} + \alpha_1\alpha_2 e_t;$$
$$\theta_t := \alpha_3(y_t - a_t) + (1-\alpha_3)\theta_{t-s} = \theta_{t-s} + \alpha_3(1-\alpha_1)e_t.$$

Модель Уинтерса с линейным трендом

Мультипликативная сезонность периода s с линейным трендом:

$$\hat{y}_{t+d} = (a_t + b_t d) \cdot \theta_{t+(d \bmod s)-s},$$

$a_t + b_t d$ — тренд, очищенный от сезонных колебаний,
$\theta_0, \ldots, \theta_{s-1}$ — сезонный профиль периода s.

Рекуррентная формула:

$$a_t := \alpha_1 (y_t / \theta_{t-s}) + (1 - \alpha_1)(a_{t-1} + b_{t-1}) = a_{t-1} + b_{t-1} + \alpha e_t / \theta_{t-s};$$
$$b_t := \alpha_2 (a_t - a_{t-1}) + (1 - \alpha_2) b_{t-1} = b_{t-1} + \alpha_1 \alpha_2 e_t / \theta_{t-s};$$
$$\theta_t := \alpha_3 (y_t / a_t) + (1 - \alpha_3) \theta_{t-s} = \theta_{t-s} + \alpha_3 (1 - \alpha_1) e_t / a_t.$$

Следящий контрольный сигнал

$e_t = y_t - \hat{y}_t$ — ошибка прогноза $\hat{y}_t$, сделанного на шаге $t-1$
Следящий контрольный сигнал (tracking signal [Trigg, 1964])

$$K_t = \frac{\hat{e}_t}{\tilde{e}_t} \qquad \begin{aligned} \hat{e}_{t+1} &:= \gamma e_t + (1-\gamma)\hat{e}_t; \\ \tilde{e}_{t+1} &:= \gamma|e_t| + (1-\gamma)\tilde{e}_t. \end{aligned}$$

Рекомендация: $\gamma = 0.05 \ldots 0.1$

Статистический тест адекватности (при $\gamma \geqslant 0.1,\ \ t \to \infty$):
гипотеза H_0: $\mathsf{E}\varepsilon_t = 0,\ \ \mathsf{E}\varepsilon_t\varepsilon_{t+d} = 0$
принимается на уровне значимости δ, если

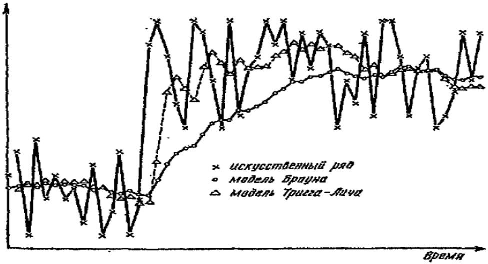
$$|K_t| \leqslant 1.2\Phi_{1-\delta/2}\sqrt{1-\gamma/(1+\gamma)},$$

$\Phi_{1-\delta/2}$ — квантиль нормального распределения,
$\Phi_{1-\delta/2} = \Phi_{0.975} = 1.96$ при $\delta = 0.05$

Модель Тригга–Лича [Trigg, Leach, 1967]

Проблема: адаптивные модели плохо приспосабливаются к резким структурным изменениям

Решение: $\alpha = |K_t|$



Недостатки:

1) плохо реагирует на одиночные выбросы; $(\alpha_t = |K_{t-1}|)$
2) требует подбора γ, при рекомендации $\gamma = 0.05 \ldots 0.1$.

Модель ARMA

ARMA(p,q): $y_1, \dots, y_t$

- $y_t = c + \underbrace{\sum_{i=1}^{p} \alpha_i y_{t-i}}_{AR} + \underbrace{\sum_{j=1}^{q} \beta_j \varepsilon_{t-j}}_{MA} + \varepsilon_t;$
- $L : Ly_t = y_{t-1}$ — лаговый оператор;
  $L^i : L^i y_t = L^{i-1}(Ly_t) = y_{t-i};$
- $\underbrace{(1 - \sum_{i=1}^{p} \alpha_i L^i)}_{H(L)} y_t = c + \underbrace{(1 + \sum_{j=1}^{q} \beta_j L^j)}_{F(L)} \varepsilon_t;$

$$y_t = \mu + \frac{F(L)}{H(L)} \varepsilon_t \text{ — каноническая запись ARMA}$$

ε_t — случайная компонента, $E\varepsilon_t = 0, E\varepsilon_i \varepsilon_j = 0$

Ряд y_t является стационарным, если корни $H(z) = 0$ лежат вне единичного круга комплексной плоскости.

Модель ARIMA

y_t — НЕстационарный, т.е. $H(z)$ — имеет d единичных корней;

- $H(L) = (1 - \sum_{i=1}^{p} \alpha_i L^i) = (1 - \sum_{i=1}^{\tilde{p}} \alpha_i L^i)(1 - L)^d$
- $(1 - \sum_{i=1}^{\tilde{p}} \alpha_i L^i)(1 - L)^d y_t = c + (1 + \sum_{j=1}^{q} \beta_j L^j)\varepsilon_t;$

$z_t = (1-L)^d y_t = \mu + \frac{F(L)}{H(L)}\varepsilon_t$ — каноническая запись ARIMA(p,q,d)

Можно также выписывать аналог для сезонных временных рядов ARIMA$(p, q, d) \times (P, Q, D)_s$:

$$(1 - L)^d (1 - L^s)^D y_t = \mu + \frac{F(L)}{H(L)} \frac{(1 + \sum_{j=1}^{Q} \gamma_j L^{s \cdot j})}{(1 - \sum_{i=1}^{P} \delta_i L^{s \cdot i})} \varepsilon_t$$

Эквивалентность некоторым моделям ЭСС

ARIMA эквивалентна следующим моделям из семейства ЭС

- простое ЭСС: ARIMA(p=0,q=1,d=1)

$$(1-L)y_t = (1-\beta L)\varepsilon_t$$

$$\beta = 1-\alpha$$

Доказательство:

$$y_t - y_{t-1} = \varepsilon_t - \beta\varepsilon_{t-1} = y_t - \hat{y}_t - (1-\alpha)\cdot(y_{t-1} - \hat{y}_{t-1})$$

$$\hat{y}_t = y_{t-1} - y_{t-1} + \alpha y_{t-1} + (1-\alpha)\cdot\hat{y}_{t-1} = \hat{y}_{t-1} + \alpha\cdot e_t + \varepsilon_t$$

- модель линейного тренда (Хольта): ARIMA(p=0,q=2,d=2)

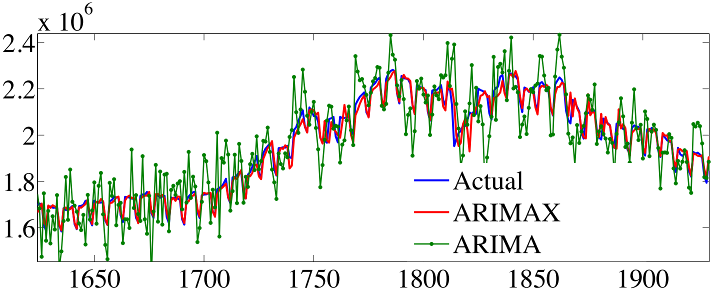
$$(1-L)^2 Y_t = (1-\Theta_1 L - \Theta_2 L^2)\varepsilon_t$$

$$\Theta_1 = 2-\alpha-\alpha\gamma,\ \Theta_2 = \alpha - 1$$

Модель ARIMAX

y_t — НЕстационарный, X_t — вектор регрессоров из $\mathbb{R}^N$, известный до начала момента прогнозирования;
ARIMAX(p,q,d):

$$z_t = \mu + \sum_{n=1}^{N} \frac{v_n(L)}{u_n(L)} X_{n,t} + \frac{F(L)}{H(L)} \varepsilon_t$$



Настройка ARIMA. Идея ММП

Настройка параметров для ARIMA сводится к ARMA модели

$$H(L)(y_t - \mu_t) = F(L)\varepsilon_t,$$

$\mu = H(L)\mu_t$ — преобразованная константа
$\varepsilon_t \in N(0, \sigma^2)$ — независимые для разных t

Правдоподобие:

$$-Log(P) = \frac{1}{2\sigma^2}\mathbf{x}'\mathbf{\Omega}\mathbf{x} - \frac{1}{2}\ln(|\mathbf{\Omega}|) - \frac{t}{2}\ln\sigma^2 \to \min_{\sigma^2, \mathbf{\Omega}}$$

$\mathbf{x} = y_t - \mu_t$ $\mathbf{\Omega}$ — ковариационная матрица (функция от $H(L)$ и $F(L)$)

Настройка ARIMA. Идея ММП (продолжение)

Оценка ММП σ^2:

$$\hat{\sigma}^2 = \frac{1}{t}\mathbf{x}'\boldsymbol{\Omega}\mathbf{x}$$

Далее оценивается $\boldsymbol{\Omega}$ с учётом $\hat{\sigma}^2$:

$$-\frac{t}{2}\ln(\mathbf{x}'\boldsymbol{\Omega}\mathbf{x}) - \frac{1}{2}\ln(|\boldsymbol{\Omega}|) \to \min_{\boldsymbol{\Omega}}.$$

Финальное решение ищется в виде нижнетреугольной матрицы $\mathbf{H}$ такой, что

$$\mathbf{H}\mathbf{H}' = \boldsymbol{\Omega}, \quad h_{ii} > 0,$$

$$\mathbf{e} = \mathbf{H}^{-1}\mathbf{x},$$

Тогда оценка Ω ищется в виде решения задачи

$$|\mathbf{H}|^{1/t}\mathbf{e}'\mathbf{e}|\mathbf{H}|^{1/t} \to \min_{\mathbf{H}}.$$

Pro&Cons ARIMA

Заключение по ARIMA–моделям

- обобщают модели а-ля ЭСС (ЭСС $=$ ARIMA(0,1,1) при $\mu = 0$)
- позволяют учитывать внешние факторы (акции, скачки цен, температуру и т.д.)
- не работает при наличии пропусков в данных;
- тяжело обучить (на практике используют перебор $p, q, d = \overline{0, 3}$;
- методы обучения опираются на нормальность ε_t
  - не для всех временных рядов удаётся найти соответствующую модель;
  - плохо работает на разреженных и коротких временных рядах;

Box, G. E. P. – Jenkins, G. M. – Reinsel, G. C.: Time Series Analysis: Forecasting and Control. John Wiley & Sons Inc., New York, 2008

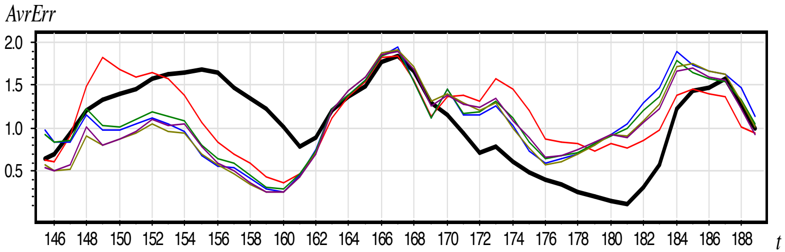
Заключение по базовым методам

- Методы прогнозирования типа ЭСС
- ARMA, ARIMA, ARIMAX, ARFIMA
- ARCH, GARCH, EGARCH ...
- VAR (векторная авторегрессия)
- Gaussian State Space Models (UCM)
- GAS, GASX (обобщённая авторегрессия)
- Гусеница [Голяндина, 2003]

Лукашин Ю. П. Адаптивные методы краткосрочного прогнозирования временных рядов. Финансы и статистика, 2003
Jonathan D. Cryer, Kung-Sik Chan Time Series Analysis With Applications in R. Second Edition. Springer, 2008
Магнус Я.Р. и др. Эконометрика. Начальный Курс М.: Дело, 2007

Пример

Динамика средних ошибок прогнозов для 6 моделей



Адаптивная селекция, [Лукашин, 2003; Timmermann, 2006]
Адаптивная композиция,
ЛАВР , [Воронцов К.В., 2006]
AFTER [Yang Y., 2004]

Адаптивная селективная модель

Пусть имеется N моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$e_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{e}_{jt} := \gamma|e_{jt}| + (1-\gamma)\tilde{e}_{j(t-1)}$ — сглаженная ошибка.

Неотличимо лучшие модели в момент времени t для порога качества $e^* \geqslant 0$:

$$\Omega_t^* = \left\{ j \mid e_{jt} - \arg\min_{j=1,\dots,N} \tilde{e}_{jt} < e^* \right\}.$$

Адаптивная селективная модель:

$$\hat{y}_{t+d} := \frac{1}{|\Omega_t^*|} \sum_{j \in \Omega_t^*} \hat{y}_{jt,t+d}$$

Требуется подбор γ, рекомендация: $\gamma = 0.01 \dots 0.1$.

Адаптивная композиция моделей

Пусть имеется N моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$e_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{e}_{jt} := \gamma|e_{jt}| + (1-\gamma)\tilde{e}_{jt}$ — экспоненциально сглаженная ошибка.

Линейная (выпуклая) комбинация моделей:

$$\hat{y}_{t+d} = \sum_{j=1}^{N} w_{jt}\hat{y}_{j,t+d}, \qquad \sum_{j=1}^{N} w_{jt} = 1, \ \ \forall t.$$

Адаптивный подбор весов [Лукашин, 2003]:

$$w_{jt} = \frac{(\tilde{e}_{jt})^{-1}}{\sum_{s=1}^{N}(\tilde{e}_{st})^{-1}}.$$

Требуется подбор γ, рекомендация: $\gamma = 0.01\ldots0.1$.

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

$$\begin{cases} \sum\limits_{i=0}^{t} \beta^{t-i} \Bigl(\sum\limits_{j=1}^{k} w_j \hat{y}_{j,i} - y_i \Bigr)^2 + \lambda \sum\limits_{j=1}^{k} \bigl(w_j - w_{j,t-1} \bigr)^2 \to \min\limits_{w_j, j=1,\dots,k} \\ \sum_{j=1}^{k} w_j = 1. \end{cases}$$

β — коэффициент «забывания» предыстории,
λ — коэффициент регуляризации.

Дополнительные варианты:

- $\beta \to 0$ — локальная адаптация весов с регуляризацией (оставляем в функционале только одно слагаемое, $i = t$)
- $w_j \geqslant 0$ — монотонный корректор

Воронцов К. В., Егорова Е. В. Динамически адаптируемые композиции алгоритмов прогнозирования // Искусственный Интеллект, Донецк, 2006. №2. С. 277–280.

Сравнение моделей

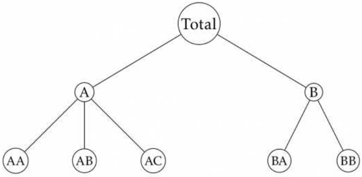
Средняя ошибка прогнозов на скользящем контроле ($T = 620$)

базовый-1	**0.7142**
базовый-2	0.7294
базовый-3	0.7534
базовый-4	0.7624
базовый-5	0.7624
базовый-6	0.7664
базовый-7	0.7793
базовый-8	0.7793

ЛАВР+Монот	0.5899
селекция+сглаживание, γ_{opt}	0.5956
МНК+Монот, β=0.7	0.6314
ЛАВР без Монот	0.6591
МНК без Монот, β=0.7	0.6834
МНК по всем данным	0.7142
среднее	0.7294
селекция без сглаживания	0.9107

- Базовые модели, их усреднение, неадаптивный МНК по всем данным — работают плохо
- Адаптивная селекция работает хорошо, если подобрать γ
- $\gamma_{\text{opt}} = 0.2 \dots 0.3$ — усреднение по $3 \dots 5$ дням

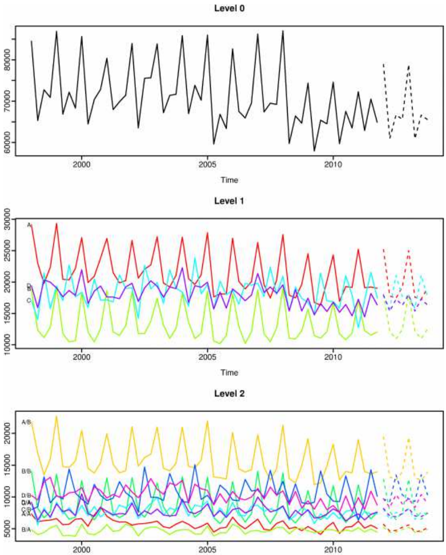
Иерархия прогноза



Основная идея: согласование независимо построенных прогнозов между разными уровнями иерархии

$$\begin{Vmatrix} \hat{y}_{total} \\ \hat{y}_A \\ \hat{y}_B \\ \hat{y}_{AA} \\ \hat{y}_{AB} \\ \hat{y}_{AC} \\ \hat{y}_{BA} \\ \hat{y}_{BB} \end{Vmatrix} = \begin{Vmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{Vmatrix} \cdot \begin{Vmatrix} \hat{y}_{AA} \\ \hat{y}_{AB} \\ \hat{y}_{AC} \\ \hat{y}_{BA} \\ \hat{y}_{BB} \end{Vmatrix} \quad \rightarrow \quad \hat{\mathbf{y}}_{\mathbf{t}} = \mathbf{S} \cdot \hat{\mathbf{y}}_{\mathbf{t},\mathbf{H}}$$

Пример согласования прогнозов



Согласование прогнозов bottom-up

$h = 0, 1, \ldots, H$ — номера уровни иерархии (сверху вниз)
$1_h, \ldots, n_h$ — временные ряды на уровне h
$i_h \in j_{h-1}$ — временной ряд на уровне h относится согласно иерархии к ряду j_{h-1}
$\hat{y}_{t|}, \tilde{y}_{t|}$ — исходный и скорректированный прогнозы;

Bottom-up (основной уровень иерархии — листовой):

$$\tilde{y}_{t|j_{h-1}} = \sum_{i_h \in j_{h-1}} \hat{y}_{t|i_h}$$

Недостатки:

- согласованный прогноз обычно низкой точности;
- на нижнем уровне плохо наблюдаются закономерности;

Замечание: используется для проверки адекватности прогнозов

Согласования прогнозов top-down

Top-down (основной уровень иерархии — корень дерева):

- Proportional tow-down

$$\tilde{y}_{t|total} = \hat{y}_{t|total} = \hat{y}_t$$

$$\tilde{y}_{t|j_h} = k_{j_h} \cdot \hat{y}_t$$

1. средняя пропорция на горизонте (в истории)
   $k_{j_h} = \frac{1}{d} \sum_{t=T+1}^{T+d} y_{t|j_h} / y_t$
2. пропорции по средним на горизонте (в истории)
   $k_{j_h} = \sum_{t=T+1}^{T+d} \hat{y}_{t|j_h} / \sum_{t=T+1}^{T+d} \hat{y}_t$

- Optimal top-down $\tilde{\mathbf{y}}_{t|H} = \mathbf{S}(\mathbf{S}'\mathbf{S})^{-1}\mathbf{S}' \cdot \hat{\mathbf{y}}_t$
  S — матрица иерархии

Rob Jhyndman etc. Optimal combination forecasts for hierarchical time series, Computational Statistics & Data Analysis, 2011

Резюме по иерархическому прогнозированию

Что нужно, чтобы согласовать прогнозы?

- выбрать иерархию прогнозирования — экспертно или сегментированием временных рядов;
- назначить основной уровень иерархии;
- независимо построить прогноз по каждому временному ряду;
- выбрать метод согласования;

Бонусы от согласования прогнозов

- возможность учитывать факторы, влияющие на прогноз, на разных уровнях иерархии;
- учёт кросс-зависимостей прогнозов (эффекты каннибализации и гало);
- дополнительные проверка адекватности прогнозов;