Прикладной статистический анализ данных. 12. Анализ временных рядов, часть третья.

Рябенко Евгений
riabenko.e@gmail.com

28 ноября 2014 г.

Причинность по Грейнджеру

Между рядами $x_1, \dots, x_T$ и $y_1, \dots, y_T$ существует **причинная связь Грейнджера** $x_t \to y_t$, если дисперсия ошибки оптимального прогноза $\hat{y}_{t+1}$ по $y_1, \dots, y_t, x_1, \dots, x_t$ меньше, чем только по $y_1, \dots, y_t$.

Причинность по Грейнджеру является необходимым, но не достаточным условием причинно-следственной связи.

$x_1, \dots, x_T$ и $y_1, \dots, y_T$ **взаимосвязаны**, если $x_t \to y_t$ и $y_t \to x_t$.

Критерий Грейнджера

$$y_t = \alpha + \sum_{i=1}^{k_1} \phi_{1i} y_{t-i} + \sum_{i=1}^{k_2} \phi_{2i} x_{t-i} + \varepsilon_t.$$

k_1 и k_2 выбирается по информационному критерию.

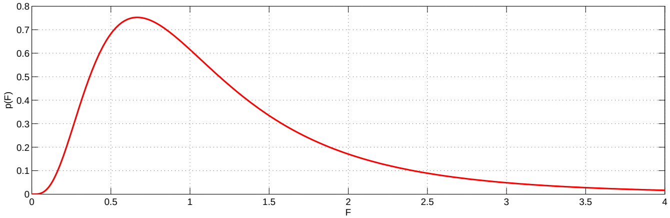
$$x_t \to y_t \quad \Rightarrow \quad \exists \phi_{2i} \neq 0.$$

нулевая гипотеза: $H_0\colon \phi_{21} = \dots = \phi_{2k_2} = 0;$

альтернатива: $H_1\colon H_0$ неверна;

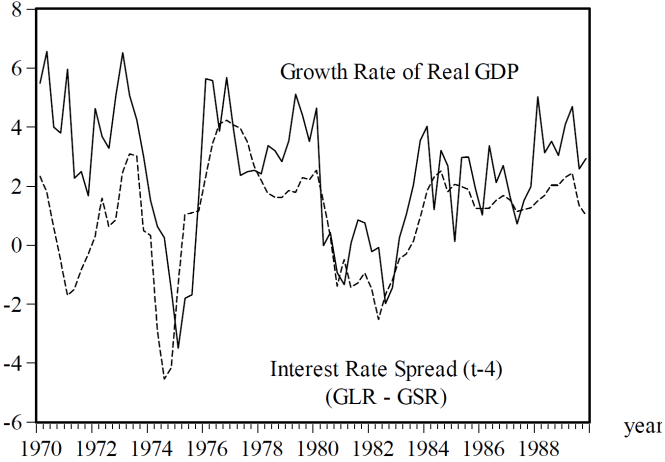
статистика: $F = \frac{(RSS_r - RSS_{ur})/k_2}{RSS_{ur}/(T-k_1-k_2-1)};$

$F \sim F(k_1, T-k_1-k_2-1)$ при H_0.



Критерий Грейнджера

Маржа сверх процентной ставки и рост ВВП, ФРГ:



Критерий Грейнджера

y	x	k_1	k_2	$F(y \leftarrow x)$	$F(y \rightarrow x)$
$\triangle_4 \ln GDP_r$	$\triangle_4 \ln M1_r$	4	4	6.087^{***}	1.918
		8	8	3.561^{**}	1.443
$\triangle_4 \ln GDP_r$	$GLR - GSR$	4	4	3.160^{*}	3.835^{**}
		8	8	$1.927^{(*)}$	2.077^{*}
$\triangle_4 \ln M1_r$	$GLR - GSR$	4	4	5.615^{***}	1.489
		8	8	2.521^{*}	1.178

$\triangle_4 \ln GDP_r$ — годовой прирост ВВП в процентах,
$\triangle_4 \ln M1_r$ — годовой прирост фактического количества денег в процентах,
GLR — рост государственных облигаций,
GSR — трёхмесячная ставка денежного рынка во Франкфурте.
(*), *, **, *** — значимость на уровне 0.1, 0.05, 0.01 и 0.001.

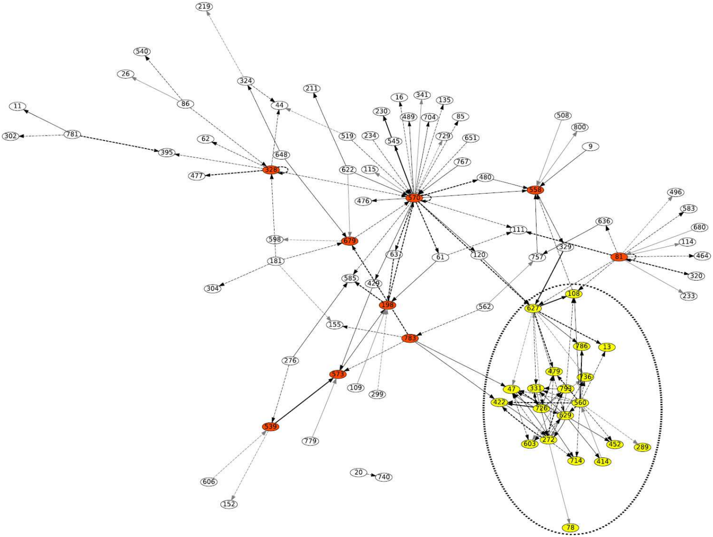
Многомерный критерий Грейнджера

Зависимость между признаками x и y может оцениваться с учётом возможной зависимости от всех остальных признаков:

$$y_t = \alpha + \sum_{i=1}^{k_1} \phi_{1i} y_{t-i} + \sum_{i=1}^{k_2} \phi_{2i} x_{t-i} + \sum_{j=1}^{m} \sum_{i=1}^{k_{j+2}} \phi_{(j+2)i} z^j_{t-i} + \varepsilon_t.$$

Для задач с большим количеством признаков могут использоваться регуляризаторы (лассо, ридж).

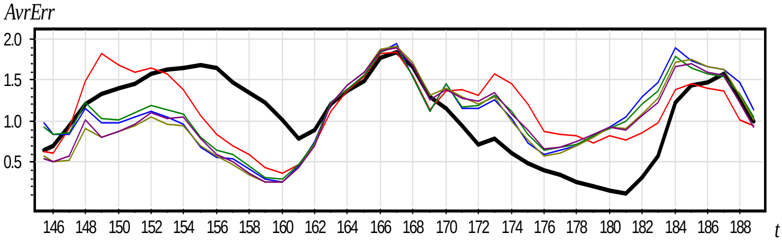
Граф причинности по Грейнджеру



Критерий Грейнджера + поправка на множественную проверку гипотез

Адаптивная селекция и композиция

Динамика средних ошибок прогнозов для 6 моделей (по реальным данным объёмов продаж в супермаркете):



Идея: включить наиболее удачные модели и отключить менее удачные.

Адаптивная селективная модель

Пусть имеется k моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$\varepsilon_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{\varepsilon}_{jt} = \sum\limits_{t=R}^{T-d} \gamma^{T-d-t} |\varepsilon_{jt}|$ — экспоненциально сглаженная ошибка.

Лучшая модель в момент времени t:

$$j_t^* = \operatorname*{argmin}_{j=1,\dots,k} \tilde{\varepsilon}_{jt}.$$

Адаптивный селективный прогноз:

$$\hat{y}_{j,t+d} := \hat{y}_{j_t^*,t+d}.$$

Адаптивная композиция моделей

Пусть имеется k моделей прогнозирования,
$\hat{y}_{j,t+d}$ — прогноз j-й модели на момент $t+d$,
$\varepsilon_{jt} = y_t - \hat{y}_{jt}$ — ошибка прогноза в момент t,
$\tilde{\varepsilon}_{jt} = \sum\limits_{t=R}^{T-d} \gamma^{T-d-t} |\varepsilon_{jt}|$ — экспоненциально сглаженная ошибка.

Линейная комбинация моделей:

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

Адаптивный подбор весов:

$$w_{jt} = \frac{\left(\tilde{\varepsilon}_{jt}\right)^{-1}}{\sum\limits_{s=1}^{k} \left(\tilde{\varepsilon}_{st}\right)^{-1}}.$$

Адаптация весов с регуляризацией

На каждом шаге t веса определяются по МНК и сглаживаются:

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

β — коэффициент забывания предыстории,
λ — коэффициент регуляризации.

Дополнительные варианты:

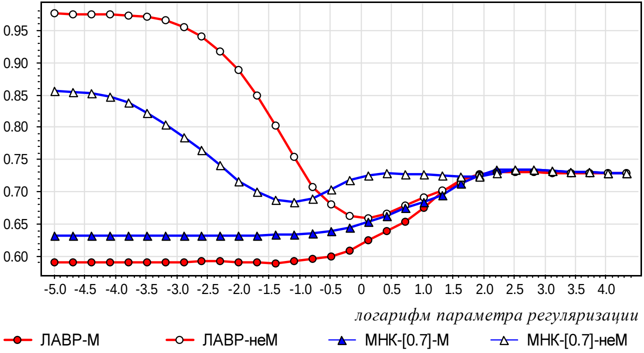
- $\beta \to 0$ — локальная адаптация весов с регуляризацией (оставляем в функционале только одно слагаемое, $i = t$);
- $w_j \geq 0$ — монотонный корректор;
- если прогнозов слишком много, индивидуальные прогнозы можно кластеризовать, усреднить по кластерам и подбирать веса средних.

Другие способы комбинирования

- регрессия ряда на несколько первых главных компонент прогнозов в скользящем окне;
- МНК с байесовским регуляризатором на веса (априорное распределение — равномерное);
- экспоненциально сглаживаются не модули, а квадраты ошибок;
- вместо экспоненциального сглаживания используется усреднение в окне;
- ...

Сравнение способов комбинирования

(Воронцов, Егорова, 2006): лучший результаты даёт адаптация весов с монотонным корректором и без регуляризации; хорошо работает адаптивная селекция с подбором параметра сглаживания.
Средняя ошибка прогнозов на скользящем контроле ($T = 620$)



(Genre et al., 2013): относительный выигрыш сложных методов по сравнению с простым усреднением прогнозов невелик и в большинстве случаев статистически незначим (“forecast combination puzzle”).

Аггрегирующий алгоритм Вовка

Пусть $y \in [0,1]$; рассмотрим адаптивную композицию со следующими весами:

$$
\begin{aligned}
w_{j1} &= \frac{1}{k},\\
w_{jt}^* &= e^{-\eta\varepsilon_{jt}^2} w_{j(t-1)},\\
w_{jt} &= \frac{w_{jt}^*}{\sum\limits_{j=1}^{k} w_{jt}^*},
\end{aligned}
$$

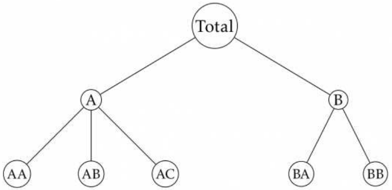
$j = 1, \dots, k, \quad \eta \in (0, \infty)$ — скорость обучения.

Пусть $\varepsilon_{AGG,t}$ — остатки такого алгоритма. При $0 < \eta \leq 2$

$$
\sum_{t=1}^{T} \varepsilon_{AGG,t}^2 \leq \sum_{t=1}^{T} \varepsilon_{jt}^2 + \frac{\ln k}{\eta} \qquad \forall j = 1, \dots, k.
$$

Иерархия рядов

Часто необходимо прогнозировать совокупности временных рядов иерархической структуры. Например, продажи могут группироваться по товарным группам, складам, поставщикам и т. д.



$$\begin{pmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{pmatrix}, \qquad \mathbf{y}_t = S\mathbf{y}_{K,t}.$$

Подходы к прогнозированию

Снизу вверх: прогнозы рядов более высоких уровней иерархии получаются суммированием прогнозов нижнего уровня.

- информация не теряется из-за агрегирования, но
- прогнозировать ряды нижнего уровня часто сложнее.

Сверху вниз: прогноз суммарного ряда y_t распределяется согласно средним долям:

$$p_j = \frac{1}{T}\sum_{t=1}^{T}\frac{y_{j,t}}{y_t}$$

или долям средних:

$$p_j = \sum_{t=1}^{T}\frac{y_{j,t}}{T} \Big/ \sum_{t=1}^{T}\frac{y_t}{T}.$$

- прогнозировать суммированный ряд легко, но
- из-за агрегирования теряется информация (например, если компоненты имеют разную сезонность).

Подходы к прогнозированию

Оптимальная комбинация: ряд каждого уровня прогнозируется отдельно, затем прогнозы корректируются в сторону большей согласованности с помощью регрессии

$$\hat{\mathbf{y}}_h = S\beta_h + \varepsilon_h, \;\; \mathbb{E}\varepsilon_h = 0, \;\; \operatorname{cov}\varepsilon = \Sigma_h;$$

$$\varepsilon_h \approx S\varepsilon_{K,h} \Rightarrow \tilde{\mathbf{y}}_h = S\left(S^T S\right)^{-1}\hat{\mathbf{y}}_h.$$

Метод реализован в пакете hts.

Метод с теоретическими гарантиями: (Стенина, Стрижов, 2015) если суммарные потери при прогнозировании всех рядов иерархии измеряются с помощью функции из класса дивергенций Брегмана, проецирование вектора прогнозов на множество векторов, удовлетворяющих структуре иерархии, не увеличивает суммарные потери.

TBATS

Метод Хольта-Уинтерса (ETS(A,A,A)):

$$\begin{aligned}
\hat{y}_{t+d|t} &= l_t + db_t + s_{t-m+(d \mod m)}, \\
l_t &= \alpha \left(y_t - s_{t-m}\right) + (1-\alpha)\left(l_{t-1} + b_{t-1}\right), \\
b_t &= \beta \left(l_t - l_{t-1}\right) + (1-\beta)\, b_{t-1}, \\
s_t &= \gamma \left(y_t - l_{t-1} - b_{t-1}\right) + (1-\gamma)\, s_{t-m}.
\end{aligned}$$

Если m велико, число сезонных коэффициентов можно сократить с помощью разложения Фурье:

$$\begin{aligned}
s_t &= \sum_{j=1}^{k} s_{j,t}, \\
s_{j,t} &= s_{j,t-1} \cos \lambda_j + s^*_{j,t-1} \sin \lambda_j + \gamma_1 d_t, \\
s^*_{j,t} &= -s_{j,t-1} \sin \lambda_j + s^*_{j,t-1} \cos \lambda_j + \gamma_2 d_t, \\
\lambda_j &= \frac{2\pi j}{m},
\end{aligned}$$

d_t — ARMA(p,q)-процесс, γ_1, γ_2 — параметры сглаживания.

TBATS

- реализован с автоматическим подбором параметров в пакете forecast, функция tbast;
- позволяет моделировать ряды с нецелым периодом сезонности;
- может учитывать сезонность с несколькими разными периодами (для этого ряд должен быть объектом типа msts, а не ts);
- наследует недостаток методов экспоненциального сглаживания — не может включать регрессоры.

Пример — прогнозирование недельных данных:

```
y   <- ts(x, frequency=52.18)
fit <- tbats(y)
fc  <- forecast(fit)
```

regARIMA

regARIMA позволяет моделировать сложные сезонности за счёт включения фурье-гармоник по длинным сезонным периодам в регрессоры. Кроме того, можно добавить индикаторы плавающих праздников.

Пример — прогнозирование дневных данных с годовой сезонностью:

```
y   <- ts(x, frequency=7)
z   <- fourier(ts(x, frequency=365.25), K=5)
zf  <- fourierf(ts(x, frequency=365.25), K=5, h=100)
fit <- auto.arima(y, xreg=cbind(z,holiday))
fc  <- forecast(fit, xreg=cbind(zf,holidayf), h=100)
```

K можно выбрать, минимизируя AIC.

Литература

- причинность по Грейнджеру (Granger causality) — Kirchgassner, глава 3;
- пример построения графа причинности — Opgen-Rhein;
- адаптивная селекция и композиция прогнозирующих алгоритмов — Воронцов, Genre;
- аггрегирующий алгоритм Вовка — Калнишкан;
- иерархические ряды — Hyndman, 9.4;
- TBATS — De Livera.

Литература

Воронцов К.В., Егорова Е.В. (2006). *Динамически адаптируемые композиции алгоритмов прогнозирования*. Искусственный Интеллект, Донецк, 2, 277–280.

Калнишкан Ю.А. *Введение в методы конкурентного предсказания*. `http://www.clrc.rhul.ac.uk/people/yura/concurrent_rus.pdf`

De Livera A.M., Hyndman R.J., Snyder R.D. (2011). *Forecasting Time Series With Complex Seasonal Patterns Using Exponential Smoothing*. Journal of the American Statistical Association, 106(496), 1513–1527.

Genre V., Kenny G., Meyler A., Timmermann A. (2013). *Combining expert forecasts: Can anything beat the simple average?* International Journal of Forecasting, 29(1), 108–121.

Hyndman R.J., Athanasopoulos G. *Forecasting: principles and practice*. — OTexts, 2014. `https://www.otexts.org/book/fpp`

Kirchgassner G., Wolters J., Hassler U. *Introduction to modern time series analysis*. — Heidelberg: Springer, 2013.

Opgen-Rhein R., Strimmer K. (2007). *Learning causal networks from systems biology time course data: an effective model selection procedure for the vector autoregressive process*. BMC Bioinformatics, 8 Suppl 2, S3.