Динамически адаптируемые композиции алгоритмов прогнозирования

К. В. Воронцов, Е. В. Егорова
(Москва)

При прогнозировании зашумленных нестационарных временных рядов возникает проблема выбора адекватной модели временного ряда. Модели нестационарных процессов, такие как GARCH, несомненно, расширяют область применимости классических статистических моделей. Однако они опираются на априорные предположения о природе нестационарности (например, гипотезу о непостоянстве дисперсии) и потому являются в той же степени эвристическими, что и классические стационарные модели.

Более универсальной представляется идея совместного применения нескольких эвристических алгоритмов прогнозирования [1]. В этом случае принимается следующая гипотеза: ряд может переходить из одного состояния в другое, причем в каждом состоянии его поведение неплохо описывается одной из стандартных моделей. На Рис. 1 показана динамика средних ошибок прогнозирования одного и того же временного ряда шестью моделями типа ARIMA (по реальным данным объёмов продаж в супермаркете). Видны три интервала, когда точность прогнозов одной из моделей соответственно хуже, сравнима, и лучше остальных. Длительность этих интервалов представляется вполне достаточной, чтобы успеть включить наиболее удачные модели и отключить менее удачные.

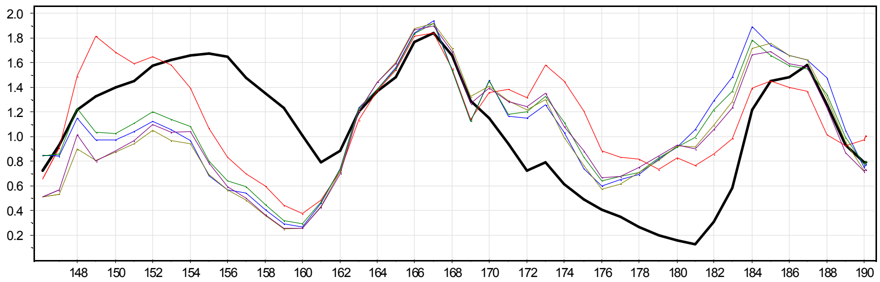

Рис. 1. Экспоненциально сглаженные средние ошибки 6 базовых алгоритмов.

Пусть для прогнозирования временного ряда $y(t)$, $t = 1,...,T$, используются p базовых алгоритмов $b_1,...,b_p$. Прогноз $b_i(t)$ в момент времени t вычисляется алгоритмом b_i с использованием информации, доступной на интервале $[1,...,t-1]$. Постоим алгоритм $a(t)$ как линейную комбинацию базовых алгоритмов:

$$a(t) = \sum_{i=1}^{p} w_{it} b_i(t), \quad \sum_{i=1}^{p} w_{it} = 1, \quad t = 1,...,T. \tag{1}$$

Веса алгоритмов w_{it}, используемые для прогнозирования в момент t, вычисляются по данным предыстории $[1,...,t-1]$. Будем называть алгоритмическую композицию (1) *динамически адаптируемой*, если веса обновляются в каждый момент времени непосредственно перед вычислением прогноза.

Рассмотрим следующие методы динамической адаптации весов.

1. $\mathrm{MS}(\theta)$ — выбор наилучшей модели (model selection) по последнему наиболее актуальному отрезку ряда. Критерием выбора служит функционал экспоненциально сглаженной средней ошибки прогнозов:

$$w_{it} = \begin{cases} 1, & i = I(t) \\ 0, & i \neq I(t) \end{cases}, \text{ где } I(t) = \arg\min_{i=1,...,p} \sum\nolimits_{\tau=1}^{t-1} \theta^{t-\tau-1} \left(b_i(\tau) - y(\tau)\right)^2 ,$$

т. е. $I(t)$ — номер модели, выбираемой в качестве наилучшей в момент времени t. Параметр $\theta \in [0,1]$ задаёт «скорость забывания» предыстории. При $\theta = 1$ все точки ряда учитываются с одинаковым весом, при $\theta = 0$ учитывается только последняя точка, при $0 < \theta < 1$ веса более старых точек убывают по геометрической прогрессии.

2. $\mathrm{LS}(\theta, \lambda)$ — метод наименьших квадратов с регуляризацией:

$$w_t = \arg\min_{w} \sum\nolimits_{\tau=1}^{t-1} \theta^{t-\tau-1} \left(\sum\nolimits_{i=1}^{p} w_i b_i(\tau) - y(\tau) \right)^2 + \lambda \sum\nolimits_{i=1}^{p} \left(w_i - w_{i,t-1} \right)^2 ,$$

где минимум берётся по вектору весов $w = (w_1,...,w_p)$, удовлетворяющему условию нормировки из (1). Параметр $\theta \in [0,1]$ задаёт «скорость забывания» предыстории, аналогично методу MS. Второе слагаемое представляет собой штраф за отклонение вектора весов w_t от вектора весов w_{t-1} в предыдущий момент времени. Параметр регуляризации $\lambda \geq 0$ позволяет найти компромисс между точностью прогнозов на обучающих данных и устойчивостью весов во времени.

3. $\mathrm{NNLS}(\theta, \lambda)$ — метод наименьших квадратов с регуляризацией и неотрицательными весами алгоритмов (non-negative least squares). Отличается от предыдущего метода введением дополнительного условия $w_{it} \geq 0$. В этом случае корректирующая операция (1) является монотонным отображением. Требование монотонности является вполне естественным для корректирующих операций и означает, что если мы не доверяем базовому алгоритму, то лучше вовсе отключить его, чем пытаться инвертировать его прогнозы [1].

В качестве эталонов для сравнения будем рассматривать ещё два, в некотором смысле тривиальных, метода.

4. AVR — усреднение базовых прогнозов:

$$w_{it} = 1/p \text{ для всех } i = 1,\dots,p,\ t = 1,\dots,T .$$

5. LS-ALL — линейная комбинация, построенная по всему ряду:

$$w_t = \arg\min_w \sum_{\tau=1}^{T}\left(\sum_{i=1}^{p} w_i b_i(\tau) - y(\tau)\right)^2 .$$

Точность всех перечисленных методов сравнивалась на реальных данных объёмов продаж товаров в супермаркете. В качестве базовых алгоритмов использовались 6 стандартных моделей типа ARIMA. Точность прогнозирования оценивалась по функционалу скользящего контроля на интервале времени $[T_0, T]$ длиной 620 точек:

$$Q = \sum_{t=T_0}^{T}\left(a(t) - y(t)\right)^2 .$$

Таблица 1. Средняя точность прогнозов базовых алгоритмов и методов динамической адаптации при оптимальных значениях параметров θ и λ.

базовые алгоритмы
0.714
0.729
0.753
0.762
0.766
0.779

методы динамической адаптации весов	
NNLS(0, λ)	**0.590**
MS(θ)	0.596
LS(0, λ)	0.659
LS-ALL	0.714
AVR	0.729
MS(1)	0.911

Результаты и выводы

1. Использование динамически адаптируемой линейной композиции способно увеличить точность базовых прогнозов на 20–30% в данной прикладной задаче.

2. Наилучшие результаты дают методы NNLS(0, λ) и MS(θ) при оптимальном значении $\theta \approx 0.7$, что приблизительно соответствует выбору наилучшей модели по 3–5 последним точкам. Это как раз тот интервал времени, на котором представляется возможной идентификация момента переключения с одной модели на другую, см. Рис. 1.

3. При ограничении неотрицательности весов точность прогнозирования оказалась выше. Интересные результаты показал анализ зависимости точности прогнозов Q от параметра регуляризации λ, Рис. 2. Без ограничения неотрицательности весов график $Q(\lambda)$ имеет классическую форму с чётко выраженным минимумом. При наличии ограничения функция $Q(\lambda)$ становится практически монотонно возрастающей. Чем меньше λ, тем лучше точность прогнозов. Она не ухудшается даже в пределе при $\lambda \to +0$, когда регуляризация фактически отключается. Таким образом,

требование монотонности само по себе является хорошим регуляризатором. Оно не только обеспечивает отбор значимых предикторов и лучшую интерпретируемость корректора, но и повышает устойчивость весов. Наилучшее качество прогнозирования достигается именно монотонным корректором.

Работа выполнена при поддержке Российского Фонда Фундаментальных Исследований, проекты 04-01-08063-офи, 05-01-00877.

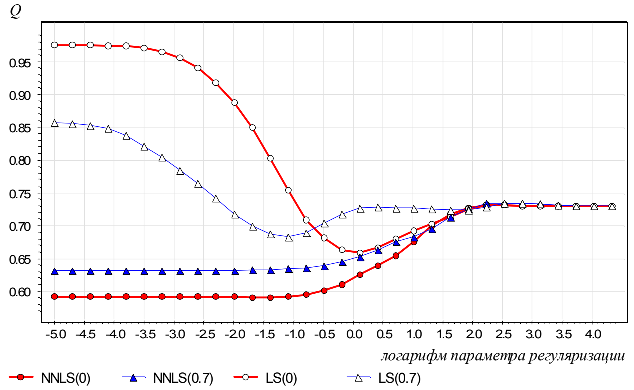

Рис. 2. Зависимость точности прогнозов от параметра регуляризации.

Литература

1. Воронцов К. В. Оптимизационные методы линейной и монотонной коррекции в алгебраическом подходе к проблеме распознавания // ЖВМ и МФ. — 2000. — Т. 40, № 1. — С. 166–176.