**УДК 519.216.3**
***К. В. Воронцов, Е. В. Егорова***
Вычислительный центр РАН,
г. Москва, Россия
voron@ccas.ru

# ДИНАМИЧЕСКИ АДАПТИРУЕМЫЕ КОМПОЗИЦИИ АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ

## АННОТАЦИЯ

В статье рассматриваются три различных метода динамической перенастройки весов в линейных композициях алгоритмов прогнозирования. Экспериментально показано, что требование неотрицательности весов не только улучшает качество прогнозов, но и выступает в роли регуляризатора, повышая устойчивость композиции.

## ВВЕДЕНИЕ

При прогнозировании зашумленных нестационарных временных рядов возникает проблема выбора адекватной модели временного ряда. Модели нестационарных процессов, такие как GARCH [1] 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 типа ARIMA [1]. Более универсальной представляется идея совместного применения нескольких эвристических алгоритмов прогнозирования [2]. 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 ARIMA. 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.

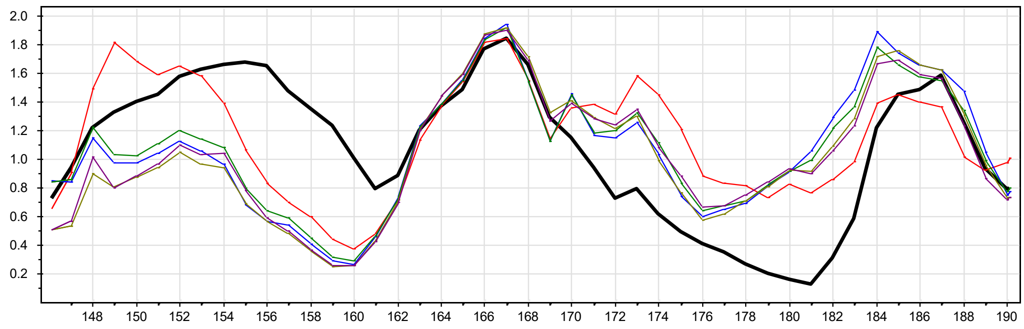


Рис. 1. Экспоненциально сглаженные средние ошибки прогнозирования для 6 базовых алгоритмов типа ARIMA (по реальному временному ряду объёмов продаж).

## ДИНАМИЧЕСКИ АДАПТИРУЕМЫЕ ЛИНЕЙНЫЕ КОМПОЗИЦИИ

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

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

Веса алгоритмов $w_{it}$, используемые для прогнозирования в момент *t*, вычисляются по данным предыстории $[1,\ldots,t-1]$. 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. MS($\theta$) — выбор наилучшей модели (model selection) по последнему наиболее актуальному отрезку ряда. Критерием выбора служит функционал экспоненциально сглаженной средней ошибки прогнозов:

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

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:

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

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 композицию (1) можно рассматривать как линейную монотонную корректирующую операцию над эвристическими алгоритмами прогнозирования [2]. Требование монотонности является вполне естественным для корректирующих операций и означает, что если базовый алгоритм имеет низкое качество, то лучше вовсе отключить его, чем пытаться инвертировать его прогнозы.

В качестве эталонов для сравнения будем рассматривать ещё два метода, не обладающих свойством динамической адаптации весов.

4. AVR — среднее базовых прогнозов: $w_{it} = 1/p$ для всех $i = 1, ..., p$, $t = 1, ..., T$.

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

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

**РЕЗУЛЬТАТЫ ЭКСПЕРИМЕНТОВ**

Точность перечисленных методов сравнивалась на реальных данных ежедневных объёмов продаж товаров в супермаркете. В качестве базовых алгоритмов использовались 6 стандартных моделей типа ARIMA. Точность прогнозирования оценивалась по функционалу скользящего контроля $Q = \sum_{t=T_0}^{T} \left( a(t) - y(t) \right)^2$ на интервале времени $[T_0, T]$ длиной 620 точек.

В следующей таблице приведена средняя ошибка прогнозов базовых алгоритмов и методов динамической адаптации при оптимальных значениях параметров $\theta$ и $\lambda$.

| Базовые алгоритмы | Методы динамической адаптации весов | |
|---|---|---|
| 0.714 | **NNLS(0, λ)** | **0.590** |
| 0.729 | MS($\theta$) | 0.596 |
| 0.753 | LS(0, $\lambda$) | 0.659 |
| 0.762 | LS-ALL | 0.714 |
| 0.766 | AVR | 0.729 |
| 0.779 | MS(1) | 0.911 |

Наилучшие результаты дают методы $\mathrm{NNLS}(0,\lambda)$ и MS($\theta$) 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.

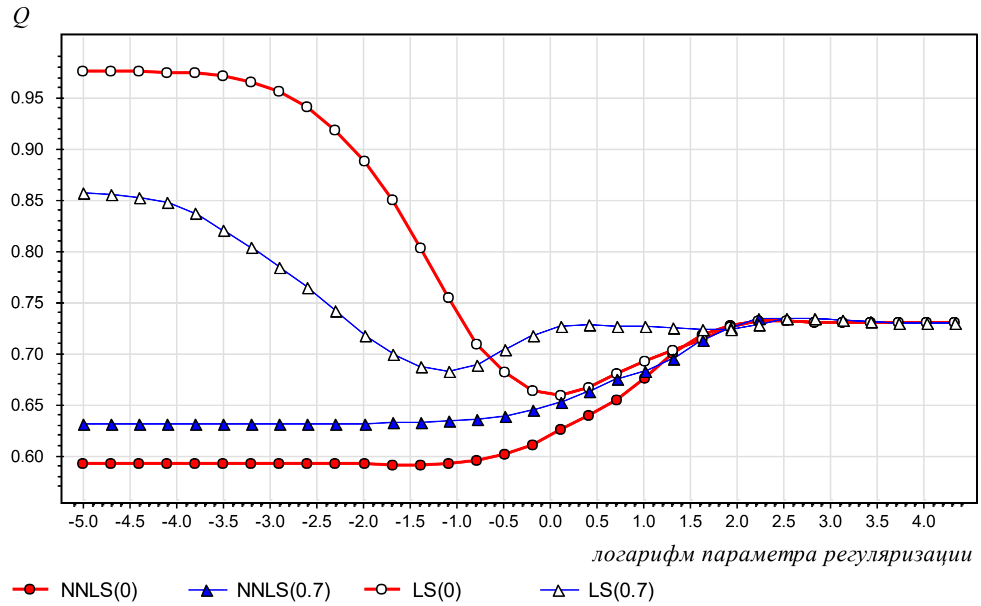


Рис. 2. Зависимость точности прогнозов от параметра регуляризации.

Интересные результаты показал анализ зависимости точности прогнозов $Q$ от параметра регуляризации $\lambda$, Рис. 2. Без ограничения неотрицательности весов график $Q(\lambda)$ имеет классическую форму с чётко выраженным минимумом, соответствующим оптимальному значению параметра регуляризации. При ограничении неотрицательности весов функция $Q(\lambda)$ становится практически монотонно возрастающей. Чем меньше $\lambda$, тем лучше точность прогнозов. Она не ухудшается даже в пределе при $\lambda \to +0$, т. е. когда регуляризация фактически отключается. Таким образом, требование монотонности само по себе является хорошим регуляризатором. Оно не только обеспечивает отбор значимых предикторов, но и повышает устойчивость весов. Наилучшее качество прогнозирования достигается именно монотонным корректором.

## ЗАКЛЮЧЕНИЕ

Использование динамически адаптируемой линейной композиции способно увеличить точность базовых прогнозов на 20–30% в данной прикладной задаче. Точность прогнозирования оказывается наилучшей при наложении ограничения неотрицательности на веса алгоритмов в композиции. Показано, что данное ограничение играет роль регуляризатора, повышая устойчивость весов. Рассмотренные методы построения композиций применяются в системе Forecsys Goods4Cast, предназначенной для прогнозирования объемов продаж в ритейлинговых сетях.

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

## ЛИТЕРАТУРА

1. Магнус Я. Р., Катышев П. К., Пересецкий А. А. Эконометрика. Начальный курс. М.: Дело. 2004.

2. 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.

***К. В. Воронцов, Е. В. Егорова***

**Динамічно адаптируемые композиції алгоритмів прогнозування**

У статті розглядаються три різних методи динамічного перенастроювання ваг у лінійних композиціях алгоритмів прогнозування. Експериментально показане, що вимога незаперечності ваг не тільки поліпшує якість прогнозів, але і вне-ступає в ролі регуляризатора, підвищуючи стійкість композиції.

***K. V. Vorontsov, E. V. Egorova***

**Dynamically updating compositions of forecasting algorithms**

The linear ensembles of forecasting algorithms are considered with three different methods of dynamic weights updating. It is experimentally shown that a requirement to weights be nonnegative favors better generalization and plays a regularization role increasing the weights stability.