



О НЕКОТОРЫХ МОДЕЛЯХ КРАТКОСРОЧНОГО ПРОГНОЗИРОВАНИЯ

Е.Г. ЖИЛЯКОВ
В.В. СКУБИЛИН

*Белгородский
государственный
национальный
исследовательский
университет*

*e-mail:
Zhilyakov@bsu.edu.ru
skuvvv@gmail.com*

В статье рассмотрен анализ алгоритмов краткосрочного прогнозирования временных рядов. Приведены результаты вычислительных экспериментов сравниваемых алгоритмов.

Ключевые слова: прогнозирование, временные ряды, доли энергий сигнала, сравнительный анализ.

Прогнозирование все чаще используется во всех сферах деятельности человека. Наверное, наиболее, широкое практическое применение прогнозирование получило в экономике. Страховые компании изучают информацию своих клиентов для прогнозирования будущих выплат, а следовательно и оптимизации самих страховок для конкретных клиентов. Владея такой информацией, страховые компании более эффективно управляют своими рисками и расходами, тем самым улучшают прибыль компании. Таким образом, становится ясна актуальность разработки новых перспективных алгоритмов прогнозирования.

В статье рассмотрен сравнительный анализ некоторых популярных алгоритмов прогнозирования, а так же разрабатываемая автором модель:

1. Наивная модель [4]
2. Модель авторегрессии [4]
3. Модель линейного предсказания [5]
4. Модель Гусеница [1]
5. Модель субполосного предсказания

Модель субполосного предсказания

Данная модель прогнозирования предполагает наличие инвариантных составляющих во временном ряду, которые используются для составления прогноза. Одним из основных моментов является вычисление распределения долей энергии по частотным интервалам

$$P_r = \frac{\bar{x}^T A_r \bar{x}}{\|\bar{x}\|^2} \quad (1)$$

Здесь A – квадратная, симметричная субполосная матрица с элементами a_{ij} [2],

$\|\bar{x}\|^2$ - норма сигнала.

$$\|\bar{x}\|^2 = \sum_{k=1}^N x_k^2 \quad (2)$$

$$A = \{a_{ij}\}, i, j = 1, \dots, N \quad (3)$$

$$\begin{cases} a_{ij} = \frac{\{\sin[f_2(i-j)] - \sin[f_1(i-j)]\}}{[\pi(i-j)]}, i \neq j \\ a_{ij} = \frac{(f_2 - f_1)}{\pi}, i = j \end{cases} \quad (4)$$

После вычисления долей энергии (1), они сортируются в убывающем порядке, при этом сохраняются индексы соответствующие частотным интервалам.

Зная доли энергии в заданных частотных интервалах, мы можем выделить наиболее значимые из них

$$A_{\Sigma} = \sum_{r=1}^R \begin{cases} A_{\Sigma} + A_r; P_{\Sigma} < \text{порог}; P_{\Sigma} = P_{\Sigma} + P_r \\ A_{\Sigma} + 0; P_{\Sigma} \geq \text{порог} \end{cases} \quad (5)$$



где R – количество частотных интервалов. P_{Σ} - аккумулярованное значение долей энергии, порог – пороговое значение суммы выбранных долей энергий, выбирается экспериментальным методом.

Предположим, что в значимых частотных интервалах доли энергии сохраняются, тогда прогнозируемое значение можно вычислить по следующей формуле

$$\tilde{x}_{N+1} = x_N + \sum_{k=1}^{N-1} (x_{k+1} - x_k) a_{N,k}^{\Sigma} \tag{6}$$

Методы расчета ошибки прогнозирования

Для вычисления ошибки прогнозирования будем использовать среднеквадратическую ошибку (MSE)

$$MSE = \frac{1}{M} \sum_{i=1}^M (x_i - \tilde{x}_i)^2 \tag{7}$$

где x_i - истинное значение, \tilde{x}_i - предсказанное значение, M – количество шагов предсказания. Также введем относительную ошибку (MAPE), вычисляемую по формуле

$$MAPE = \frac{1}{M} \sum_{i=1}^M \left| \frac{x_i - \tilde{x}_i}{\sigma} \right| * 100 \tag{8}$$

$$\sigma = \sqrt{\frac{1}{M} \sum_{i=1}^M (x_i - \tilde{x}_i)^2} \tag{9}$$

Исходные данные

Для проведения эксперимента было подготовлена 4 временных ряда, предоставленных на рис. 1.

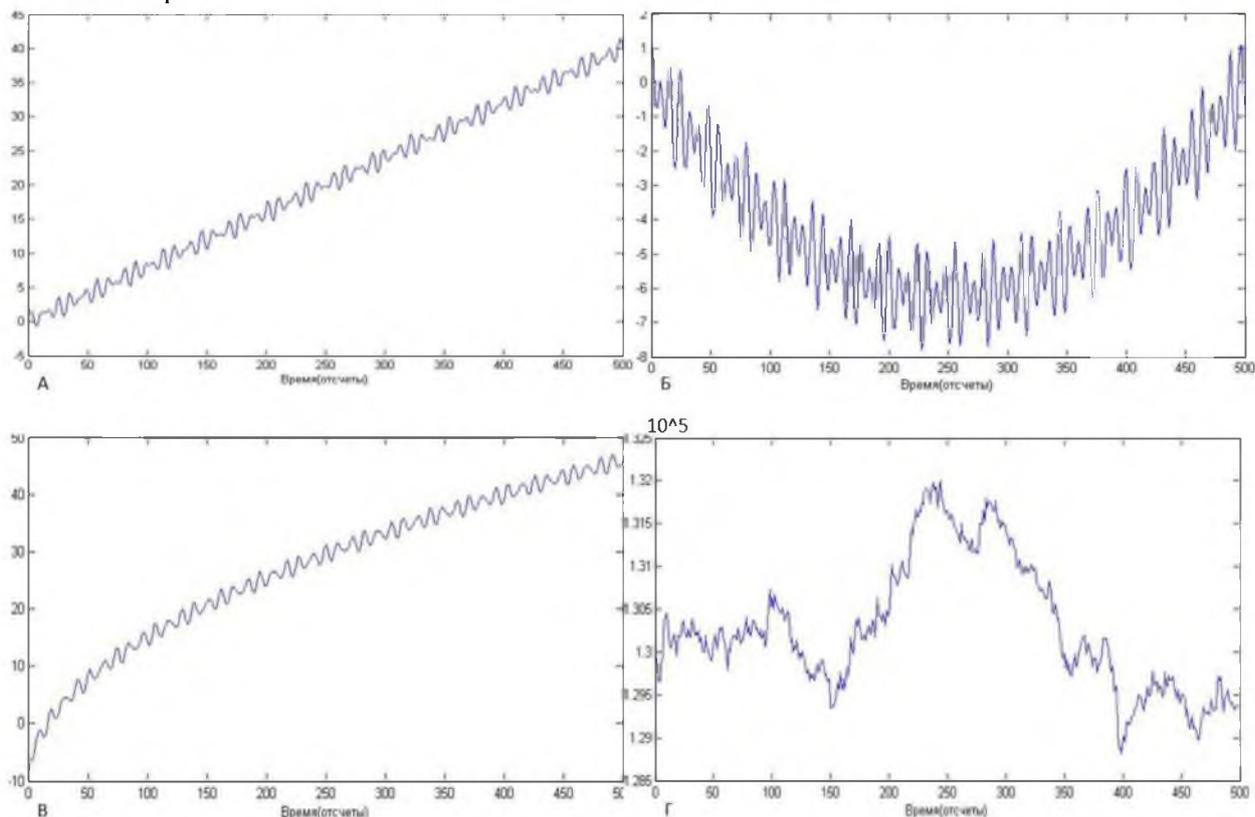


Рис. 1. Исходные данные: А – $\bar{x} = 0,8t + 0,9\sin(\frac{2\pi}{11}) + 0,8\sin(\frac{2\pi(t + 0,09)}{8})$;

Б – $\bar{x} = 0,0001t^2 - 0,05t + 0,6\sin(\frac{2\pi(t - 0,14)}{11}) + \cos(\frac{2\pi}{8})$;

В – $\bar{x} = -10 + 2,5\sqrt{t} - 1,2\sin(\frac{2\pi(t + 0,24)}{11}) + 0,5\sin(\frac{2\pi(t - 0,35)}{8})$;

Г – реальные рыночные котировки.



Временной ряд, изображенный на рисунке 1 (Г) был взят из реальных рыночных котировок фьючерса на индекс РТС (Биржа РТС) с целью проверки поведения алгоритмов прогнозирования на нестационарных данных.

Вычислительный эксперимент.

Каждый временной ряд состоит из 500 отсчетов, первые 250 из них используются для расчетов алгоритмов и являются апостериорными данными для прогноза. Остальные 250 значений временного ряда прогнозируются. На каждом шаге прогнозирования вычисляются ошибки предсказания (7,8), после чего истинное значение временного ряда на текущем шаге предсказания добавляется в ряд исходных данных и используется для следующих шагов прогнозирования.

Параметры моделей прогнозирования были подобраны экспериментальным путем. Результаты экспериментов приведены в табл. 1.

Таблица 1

Результаты экспериментов прогнозирования временных рядов

| Модель предсказания | Ряд А | | Ряд Б | | Ряд В | | Ряд Г | |
|---|------------|-------------|------------|---------|--------|--------|--------|---------|
| | MSE | MAPE | MSE | MAPE | MSE | MAPE | MSE | MAPE |
| Модель 1 (Наивная) | 0,325 | 4,040% | 0,359 | 25,130% | 0,306 | 3,546% | 9500 | 9,618% |
| Модель 2 (Авторегрессия) | 0,134 | 2,657% | 0,167 | 17,454% | 0,106 | 2,039% | 9695 | 9,772% |
| Модель 3 (Линейное предсказание, порядок модели = 30) | 0,146 | 2,743% | 0,006 | 3,269% | 0,198 | 2,781% | 309930 | 72,392% |
| Модель 4 (Гусеница, L=50) | 1,719e-024 | 7,494e-012% | 4,682e-005 | 0,333% | 0,0006 | 0,108% | 30275 | 18,046% |
| Модель 5 (Субполосное предсказание, R=30, сумма долей энергий = 0,95) | 0,332 | 4,080% | 0,235 | 20,261% | 0,316 | 3,613% | 9729 | 9,768% |

Влияние неопределенностей временного ряда на ошибку прогнозирования.

В качестве исходного временного ряда будем использовать временной ряд представленный на рисунке 1 (А) к которому добавим шумоподобную составляющую

$$Noise = A * Rand(N) \quad (10)$$

где A – амплитуда шума, Rand(N) – равномерно распределенные псевдослучайные числа.

Результаты эксперимента приведены в табл. 2.

Таблица 2

Результаты эксперимента по влиянию шумов на ошибку прогнозирования

| Название модели предсказания | MAPE, A=0 (без шумов) | MAPE, A=0.5 | MAPE, A=1 | MAPE, A=2 |
|---|-----------------------|-------------|-----------|-----------|
| Модель 1 (Наивная) | 4,0406% | 4,3124% | 4,6840% | 7,0029% |
| Модель 2 (Авторегрессия) | 2,6575% | 3,2175% | 4,2982% | 6,9472% |
| Модель 3 (Линейное предсказание, порядок модели = 30) | 2,7437% | 3,0153% | 3,8407% | 6,4929% |
| Модель 4 (Гусеница, L=50) | 7,4942e-012% | 1,1025% | 2,3294% | 4,5888% |
| Модель 5 (Субполосное предсказание, R=30, сумма долей энергий = 0,95) | 4,0803% | 4,3694% | 4,7336% | 7,1245% |



В ходе проведения вычислительных экспериментов было выявлено, что предлагаемый метод прогнозирования временных рядов нуждается в дальнейшем исследовании и улучшении показателя ошибки. На временных рядах А, Б, В метод Гусеница показал наименьшую ошибку прогнозирования, но временной ряд Г этот метод прогнозирует значительно хуже, что говорит о невозможности применения метода для всех временных рядов.

Эксперимент по влиянию неопределенностей временного ряда показал, что шумоподобная составляющая слабо влияет на предлагаемую модель вплоть до $A=1$, что означает хорошую устойчивость модели. Модель Гусеница имеет наименьшее увеличение ошибки предсказания при росте уровня шума.

Список литературы

1. Голяндина Н.Э. Метод «Гусеница»-SSA: прогноз временных рядов [Текст] / Голяндина Н.Э. – СПб., 2004. – 52 с.
2. Жилияков Е.Г., Белов С.П. Обнаружение звуков речи на фоне шумов [Статья] / «Научные ведомости Белгородского государственного университета» № 7(126) выпуск 22/1 Белгород, 2012. С. 182-189.
3. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов [Текст] / Ю.П. Лукашин. – М.: Финансы и статистика, 2003. – 416 с.
4. Ханк Д.Э. Бизнес-Прогнозирование [Текст] / Ханк Д.Э, Уичерн Д.У, Райтс А.Дж. – М.: Вильямс, 2003. – 656 с.
5. Alan O.C. Linear Prediction – The Technique, Its Solution and Application to Speech – Ireland: Dublin Institute of Technology, 2008. – 19 p.

ABOUT SOME SHORT TERM PREDICTION MODELS

E.G. ZHYLYAKOV
V.V. SKUBILIN

*Belgorod National
Research University*

*e-mail:
Zhilyakov@bsu.edu.ru
skuvvv@gmail.com*

The paper considers analysis of algorithms short time series prediction. The results of numerical experiments compared algorithms.

Keywords: prediction, time series, comparative analysis, part of signal energy.