

Оценивание эффективности статистических методов прогнозирования в условиях динамического хаоса

Р. М. Юсупов

Санкт-Петербургский институт информатики и
автоматизации Российской академии наук (СПИИРАН)
yusupov@iias.spb.su

А. А. Мусаев

Санкт-Петербургский институт информатики и
автоматизации Российской академии наук (СПИИРАН)
amusaev@technolog.edu.ru

Д. А. Григорьев

Центр эконометрики и бизнес-аналитики СПбГУ (ЦЭБА)
d.a.grigoriev@spbu.ru

Аннотация. Статья посвящена анализу возможности использования традиционных технологий статистического синтеза прогностических решений в условиях динамического хаоса, характерного для задач управления в нестабильных средах погружения. Показано принципиальное отличие свойств рядов наблюдений за состоянием нестабильных систем от вероятностных описаний традиционных моделей, основанных на статистической парадигме. Рассмотрена аддитивная модель с хаотической системной составляющей и нестационарными шумами, наиболее адекватно описывающая указанные ряды наблюдений. Предложена методика прагматического оценивания функциональной эффективности технологий прогнозирования в условиях хаотической неопределенности.

Ключевые слова: статистические методы прогнозирования; теория эффективности; прагматический подход; динамический хаос; нестабильные системы

I. ВВЕДЕНИЕ

Характерной чертой многих открытых нелинейных систем, связанных с нестабильными средами погружения, является наличие множества трудно прогнозируемых гетерогенных факторов влияния, не фиксируемых в явном виде системами мониторинга. При этом в каждый момент времени тот или иной фактор, часто имеющий латентный характер, может оказаться доминирующим, что приводит к возникновению множества явно выраженных локальных трендов. Наличие таких трендов, а также больших участков рядов наблюдений, динамические свойства которых представляют собой колебательные неперiodические функции, является характерным для процессов, описываемых моделью динамического хаоса [1–5]. Достаточно понятно, что свойства таких процессов не соответствуют совокупности традиционных ограничений, в рамках которых статистические оценки отвечают требованиям их состоятельности, эффективности

и несмещенности. В частности, имеются в виду такие свойства наблюдаемых процессов, как стационарность шумовой составляющей, независимость линейных приращений, нормальности (или гауссовости) вариаций значений и др.

В то же время традиционный анализ качества прогнозирования [6–11], основанный на квадратичной и других мерах подобия реальной и прогнозируемой траектории эволюции вектора состояния, для многих практических задач сам по себе является неэффективным. Для реальных приложений важна не степень подобия траекторий эволюции вектора состояния, и даже, до определенной степени, не величина дисперсии терминальной оценки. Вполне достаточным, с точки зрения результативности проактивного (прогностического) управления [12], является корректный прогноз факта достижения вектором состояния заданной области фазового пространства. При этом переходный процесс может носить достаточно произвольный характер, при условии нахождения фазовой траектории в области допустимых значений. Такой подход снимает с задачи прогнозирования требования к соответствию традиционным ограничениям подобия модели, и позволяет более рационально и прагматично оценивать эффективность прогностических вычислительных схем в сложных условиях динамического хаоса.

II. ОСОБЕННОСТИ ФОРМИРОВАНИЯ ПРОАКТИВНЫХ УПРАВЛЯЮЩИХ РЕШЕНИЙ В УСЛОВИЯХ ДИНАМИЧЕСКОГО ХАОСА

Наличие возмущающих воздействий, оказываемых средой взаимодействия на любую открытую систему, приводит к тому, что имеющиеся наблюдения $y \in Y$ всегда содержат случайную составляющую $v \in V$. При этом связь временных рядов наблюдений $Y_t, t = 1, \dots, n$ с вектором параметров состояния системы $x \in X$ может быть задана с помощью некоторого, в общем случае нелинейного оператора, имеющего для одномерной задачи вид

Работа выполнена при финансовой поддержке РФФИ, проекты №№19–08–00989, 20–08–01046 и Центра ЦЭБА СПбГУ

отображения $\Phi: y = \Phi(x, v)$. Формирование управляющего решения $d \in D$ осуществляется на основе наблюдений $Y_t, t = 1, \dots, n$ и совокупности априорной информации $I_a(X, v)$, имеющейся в распоряжении системы управления (СУ). Причем качество реализации проактивного решения в значительной степени зависит от наличия и достоверности априорной информации $I_a(X, v)$. Недостаток объема достоверной априорной информации о текущей и прогнозируемой ситуациях задает уровень неопределенности, при котором принимается управляющее решение.

Естественно полагать, что чем больше достоверной априорной информации о текущей ситуации используется процедурой анализа данных, тем выше качество найденного решения. Однако это справедливо лишь при наличии достоверной априорной информации о распределении вектора состояния и статистических характеристиках среды взаимодействия. Даже при незначительных отклонениях реальных характеристик от ошибочно принятой исходной модели их эффективность может существенно снизиться, т.е. указанные оценки могут оказаться статистически неэффективными и неустойчивыми, что неизбежно скажется на достоверности результатов анализа торговой ситуации и, в конечном счете, существенно снизит эффективность формируемых управляющих решений.

III. МОДЕЛЬ НАБЛЮДЕНИЙ В ЗАДАЧЕ УПРАВЛЕНИЯ В НЕСТАБИЛЬНЫХ СРЕДАХ

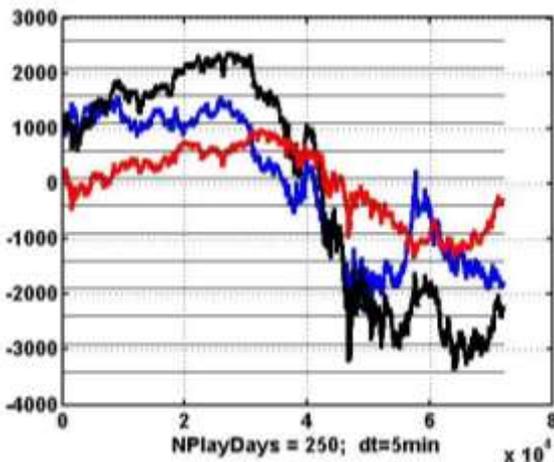


Рис. 1. Динамика изменения параметров нестабильной системы в течение 250 дней наблюдения

На рис. 1 приведены примеры изменения наблюдений за параметрами нестабильной системы (в данном случае плотностей турбулентных газодинамических потоков) в течение 250 дней. Визуальное изучение рядов наблюдений приводит к выводу, что их природа определяется хаотической динамикой и описывается аддитивной моделью вида [13]

$$y_k = x_k + v_k, k = 1, \dots, n \quad (1)$$

в которой системная составляющая наблюдаемого процесса $x_k, k = 1, \dots, n$, используемая в процессе

формирования управляющего решения, представляет собой реализацию динамического хаоса и обычно имеет вид колебательного непериодического процесса с множеством ложных локальных трендов неопределенной длительности. Шумовая составляющая наблюдаемого процесса $v_k, k = 1, \dots, n$, как показали численные исследования, основанные на проверке статистических гипотез, представляет собой нестационарный гетероскедастический процесс с невырожденной функцией автокорреляции, зависящей от времени. Задача состоит в том, чтобы оценить эффективность применения решений, формируемых на основе традиционных методов статистического прогнозирования, к задаче проактивного управления в условиях, описываемых моделью наблюдений (1). С этой целью в рамках проведенных исследований разработана методика, основанная на подходе к оцениванию эффективности информационных технологий (ИТ), изложенная в [14, 15].

IV. МЕТОДИКА АНАЛИЗА ЭФФЕКТИВНОСТИ СТАТИСТИЧЕСКИХ РЕШЕНИЙ В ЗАДАЧЕ УПРАВЛЕНИЯ НЕСТАБИЛЬНЫМИ СИСТЕМАМИ

В соответствии с указанной общей теорией эффективности ИТ, результативность алгоритмов анализа данных, должна оцениваться через результативность метасистемы, в интересах которой она функционирует. Применение технологий проактивного управления базируется на прогнозировании развития наблюдаемого процесса. Аналитическая оценка эффективности статистических методов в условиях хаотической и нестационарной динамики наблюдаемых процессов оказывается не реализуемой. В связи с этим в настоящей работе проведены численные исследования, основанные на терминальной оценке эффективности алгоритмов прогнозирования и формируемых на их основе проактивных управляющих решений.

Статистическая идентификация модели формируется на участке наблюдения, отвечающего интервалу времени изменения значения прогнозируемого параметра на заданную величину dL . Корректный статистический прогноз отвечает способности исследуемого процесса сохранить ранее обнаруженный тренд от уровня, на котором он был фиксирован, до момента его пересечения следующего (по направлению установленного тренда) уровня. В случаях, когда процесс разворачивается и достигает противоположного уровня, соответствующий фрагмент эксперимента будет рассматриваться как ошибочное решение. При использовании такой методики, в качестве показателя эффективности статистических решений будет использоваться оценка вероятности (точнее, частота) события, заключающегося в том, что процесс пересечет уровень подтверждения корректности прогноза раньше, чем уровня принятия решений об его ошибочности.

В. ЭКСПЕРИМЕНТЫ

А. Описание вычислительного эксперимента

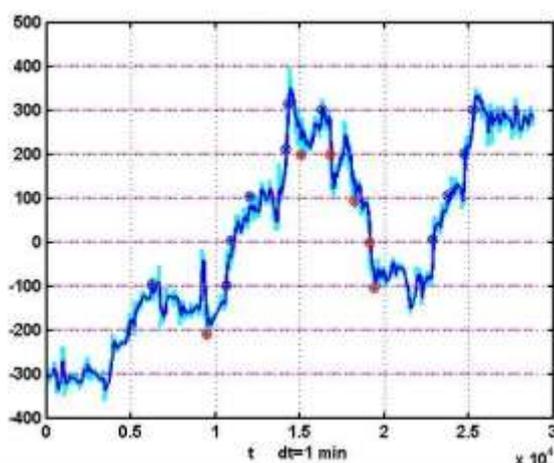


Рис. 2. Пример результатов прогнозирования хаотического процесса на 10-дневном интервале

Разделим область изменения исследуемого процесса на равномерные сектора размера dL . Значения процесса измеряются в унифицированных величинах, называемых пунктами (п.). Предположим, что процесс перешел с уровня dL до вышерасположенного уровня L_{k+1} . Данный факт можно грубо интерпретировать как наличие положительного тренда. Обозначим такой переход $L_k \rightarrow L_{k+1}$. И наоборот, переход $L_k \rightarrow L_{k-1}$ можно рассматривать, как переход на динамику с отрицательным трендом. Вопрос состоит в подтверждении гипотезы инерционности процесса Y_t , определяемой достижением последующего уровня в направлении, соответствующем установленному тренду. Будем оценивать вероятность положительных исходов, то есть перехода процесса $L_k \rightarrow L_{k+1}$ после того, как он осуществил переход $L_{k-1} \rightarrow L_k$. Отрицательным исходом будет обратный переход на нижестоящий уровень $L_{k+1} \rightarrow L_k$ сразу после ранее осуществленного возрастающего перехода $L_{k-1} \rightarrow L_k$. В силу симметрии аналогичные оценки исходов имеют место и для убывающего перехода. Таким образом, полная группа событий состоит из двух положительных исходов $(L_k \rightarrow L_{k+1} | L_{k-1} \rightarrow L_k)$ и двух отрицательных $(L_k \rightarrow L_{k-1} | L_{k+1} \rightarrow L_k)$ и $(L_k \rightarrow L_{k+1} | L_{k+1} \rightarrow L_k)$.

В качестве иллюстрации на рис. 2 приведен пример последовательности измерений на 10-дневном интервале наблюдения с границами зон сегментации и отметками о пересечении соответствующих границ. Заметим, что на графике представлен как сам процесс Y_t , $t = 1, \dots, n$, так и его сглаженная версия \tilde{Y}_t , $t = 1, \dots, n$. Сглаживание осуществлялось простым экспоненциальным фильтром

$$\tilde{Y}_t = \alpha Y_t + \beta \tilde{Y}_{t-1}, t = 1, \dots, n, \alpha \in (0, 1), \beta = 1 - \alpha \quad (2)$$

Предположим, что проведено N экспериментов, в каждом из которых зафиксирован тренд, как направление

перехода с одного уровня на другой. В случае, если направление движения сохранится вплоть до пересечения со следующим по направлению тренда уровнем, то такой исход будем воспринимать как событие, подтверждающее правильность прогнозирования. Наоборот, если процесс развернется и достигнет предшествующего уровня, то такой исход следует рассматривать как ошибочный прогноз. Предположим, что в результате N последовательно проведенных экспериментов, m исходов подтверждают гипотезу о наличии тренда, а $N - m$ исходов ее отрицают. Тогда предположение о наличии значимого тренда может трактоваться как альтернатива $H_1: p \neq 0.5$ к нулевой гипотезе $H_0: p = 0.5$ об его отсутствии.

В. Вычислительные эксперименты

Эксперименты по оценке эффективности статистического прогноза хаотического процесса проводились большим (до 100 дней) участкам наблюдений. Рассматривались ряды наблюдений, отвечающие физическим и эконометрическим хаотическим процессам. Полиномиальная идентификация модели осуществлялась на основе МНК многочленами 1–3 степени. Прогноз осуществлялся на основе прямой статистической экстраполяции. Прагматическая оценка эффективности осуществлялась в соответствии с описанной выше методикой. Оценка вероятности осуществляется через частоту положительных исходов. Значимость отличия найденной частоты $h = m/N$ (где m – число положительных исходов в N экспериментах) от величины p_0 осуществлялось на основе статистики $u = (h - p_0) \sqrt{N} / \sqrt{p_0(1 - p_0)}$ для уровня доверия $\alpha = 0.99$. Для двухсторонней критической области u_{kp} определяется по таблице значений функции Лапласа. Использовались области сегментации $dL = 100, 75, 50n$.

В качестве примера в табл. 1 приведены данные, демонстрирующие частоту положительных исходов для эконометрических данных, полученных в результате мониторинга котировок для трех наиболее распространенных валютных пар евро-доллар США, евро-йена и доллар США-йена. Прогностические решения формировались на пяти 100-дневных интервалах с частотой отсчетов 1 мин.

ТАБЛИЦА 1 ЧАСТОТА ПОЛОЖИТЕЛЬНЫХ ИСХОДОВ ПРОГНОЗА

Временной интервал, сутки	Валютные инструменты		
	EURUSD	EURJPY	USDJPY
1-100	0.552	0.484	0.444
101-200	0.507	0.536	0.465
201-300	0.533	0.552	0.560
301-400	0.494	0.452	0.465
401-500	0.446	0.545	0.444

Важной особенностью валютного рынка является его высокая волатильность и крайне низкая инерционность протекающих процессов. В течение нескольких минут нормированное значение котировок может измениться в несколько раз. В этих условиях традиционные методы статистического прогноза, как видно из приведенных в табл. 1 данных, оказываются совершенно неэффективны,

вероятность выигрышных управляющих решений крайне незначительно колеблется вокруг наилучшего исхода, т. е. величины $p = 0.5$.

Для физического процесса картина оказывается несколько лучше. В табл. 2 приведены нормированные и центрированные результаты мониторинга плотности турбулентного потока жидкости для трех различных сечений русла S_1, S_2, S_3 и трех интервалах наблюдения длительностью по три часа каждый и частотой отсчетов 3 мин.

ТАБЛИЦА 2 Частота положительных исходов прогноза

Временной интервал, часы	Сечения		
	S_1	S_2	S_3
0-3	0.68	0.72	0.80
3-6	0.71	0.77	0.79
6-9	0.65	0.78	0.75

Дополнительная серия экспериментов проводилась для сглаженных процессов с использованием экспоненциальных фильтров (2).

Из рассмотренных вычислительных экспериментов можно сделать выводы, что для безынерционных хаотических сред (например, для котировок цен на рынках капитала) статистические методы для выбранного уровня доверия имеют вероятность верных прогнозов около 0.5. В то же время для материальных хаотических процессов, обладающих определенной инерционностью, частота правильных прогнозов колеблется в интервале 0.65–0.8. Степень пригодности такого уровня доверия определяется содержательным аспектом решаемой задачи.

VI. ЗАКЛЮЧЕНИЕ

Вопрос применимости статистических технологий прогнозирования является ключевым с точки зрения решения прикладных задач проактивного управления в условиях хаотических процессов с нестационарной случайной составляющей. Построить аналитические оценки вариантов управления для нестабильных сред не удастся в силу параметрической неустойчивости наблюдаемых процессов. В связи с этим подобные исследования осуществляются численно, на больших интервалах наблюдений, способных охватить многообразие возможных вариаций стохастического хаоса.

Предложенная в статье технология оценивания эффективности статистических решений позволяет прагматически подойти к вопросу о целесообразности применения этих методов, отказавшись от нерациональных критериев пригодности, основанных на мерах подобия реальной и прогнозируемой траекторий. Обобщение данного подхода ведет к предположению о целесообразности перехода от критериев подобия математических моделей, к критериям их полезности для значительного числа прикладных задач управления.

СПИСОК ЛИТЕРАТУРЫ

- [1] Smith L. Chaos: A Very Short Introduction. Oxford University Press: Oxford, UK, 2007; 180 p.
- [2] Broer Henk, and Floris Takens. Dynamical systems and chaos. Springer Science & Business Media, Vol. 172, 2010.
- [3] Manneville P. Instabilities, Chaos and Turbulence: An Introduction to Nonlinear Dynamics And Complex Systems. World Scientific Publishing Company, 2004.
- [4] Kautz R. Chaos: The Science of Predictable Random Motion. Oxford University Press, 2011.
- [5] Klimontovich Yu.L. Turbulent motion and the structure of chaos. A new approach to the statistical theory of open systems. URSS Publ., 2nd ed., 2010.
- [6] A. Stuart, M. G. Kendall, et al. The advanced theory of statistics. Griffin, 1963.
- [7] Balagurin P., Makarikhin P., Grigorieva A. Assessment of the Integrity of the Object Based on the Correlation of Super-Signals // Innovations Induced by Research in Technical Systems. IIRTS 2019. Lecture Notes in Mechanical Engineering. Springer, Cham., 2020.
- [8] Bolch B.W., Huang C.J. Multivariate statistical methods for business and economics. Prentice Hall, 1973.
- [9] Maddala G.S., Lahiri K. Introduction to econometrics // Macmillan, vol. 2., New York, 1992.
- [10] Asteriou D., Hall S.G. Applied econometrics. Macmillan International Higher Education, 2015.
- [11] Montgomery D.C., Peck E.A., Vining G.G. Introduction to linear regression analysis. John Wiley & Sons, 2021.
- [12] Мусаев А.А., Фенин М.М. Проактивное стабилизационное управление для нестационарных сред на основе гибридных алгоритмов анализа данных // Известия Санкт-Петербургского государственного технологического института (технического университета). 2020. № 52(79). С. 100–106.
- [13] Мусаев А.А. Quod est veritas. Трансформация взглядов на системную составляющую наблюдаемого процесса // Информатика и автоматизация. 2010. №. 15. С. 53-74.
- [14] Юсупов Р.М., Мусаев А.А. Особенности оценивания эффективности информационных систем и технологий // Труды СПИРАН. 2017. Вып. 2 (51). С. 5-34.
- [15] Юсупов Р.М., Мусаев А.А. К оцениванию эффективности информационных систем. Методологические аспекты // Информационные технологии. 2017. Том 23. №5. С. 323-332.