

ОТЗЫВ ОФИЦИАЛЬНОГО ОППОНЕНТА

доктора технических наук, профессора Авдеенко Татьяны Владимировны на диссертационную работу Тимошенко Юлии Сергеевны на тему «Разработка методики интеграции формальных методов прогнозирования временных рядов и метода ассимиляции данных», представленной на соискание ученой степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика

Актуальность темы исследования

Прогнозирование нелинейных динамических систем является одной из важных задач, прогнозирование применяется в различных областях науки и техники, данная задача сводится к прогнозированию временных рядов. На сегодняшний день существует большое число подходов и методов прогнозирования временных рядов. Это методы, которые используют только знания о временном ряде, методы на основе нейронных сетей, методы, позволяющие учитывать математическое описание системы, порядившей временной ряд. Анализ научной литературы по данной теме позволяет сделать вывод, что прогнозирование временных рядов, порождённых сложными системами, не всегда возможно при использовании только методов, которые используют знания только об отсчетах временного ряда. Есть также системы, данные о состоянии которых можно получать в реальном времени и использовать для коррекции прогноза. На практике часто возникают проблемы, которые в теории и практике прогнозирования освещены в недостаточной степени. Одной из таких проблем является вопрос возможности применения методов прогнозирования временных рядов, использующий математические модели к прогнозу временных рядов, для которых таких моделей не существует путем комбинирования с методами, использующими феноменологические модели временных рядов. В этой связи разработка новых комбинированных методов прогнозирования временных рядов является актуальной задачей, а диссертационное исследование выполнено на актуальную тему.

Научная новизна полученных результатов

В диссертационной работе получены следующие новые научные результаты:

1. Обоснована возможность интеграции формальных методов прогнозирования временных рядов, порожденных динамическими системами, описываемыми феноменологическими моделями, и метода ассимиляции данных.
2. Разработана методика интеграции формальных методов прогнозирования временных рядов в метод ассимиляции данных и ее программной реализации.
3. Обоснован выбор метрик и набора показателей, обеспечивающих количественную оценку качества прогнозирования временных рядов.

Данные результаты подтверждаются автором в публикациях и в тексте диссертации.

Обоснованность и достоверность научных положений, сформулированных в диссертации

Обоснованность и достоверность научных результатов исследования достигнута благодаря корректной постановке цели и задач исследования, использованием известных математических методов, непротиворечивостью результатов прогнозирования известных временных рядов, полученных с использованием разработанного программного комплекса, а также их согласованностью с результатами, полученными другими авторами.

Результаты исследования опубликованы в 14 научных работах, 9 из которых представлены в ведущих рецензируемых журналах, рекомендованных ВАК РФ и Аттестационным советом УрФУ. Автором получено три свидетельства о регистрации программы для ЭВМ.

Практическая значимость результатов диссертации

Практическая значимость заключается в разработке программного комплекса «Прогнозирование временных рядов», обеспечивающего использование разработанной методики интеграции формальных методов прогнозирования и метода ассимиляции данных, в котором реализованы:

- формальные методы прогнозирования ВР: ARIMA, SSA, GMDH, LSTM-net;
- классический метод Data Assimilation;
- методика интеграции формальных методов прогнозирования временных рядов и метода Data Assimilation, разработанная автором
- модельные временные ряды и реальные временные ряды, приведенные в комплексе для возможности проведения эксперимента.

Полученные результаты внедрены в ФГАОУ ВО «Уральский федеральный университет имени первого Президента России Б.Н. Ельцина», (Екатеринбург, Россия); в ООО «Эйрбэйс», (Екатеринбург, Россия).

Оценка содержания диссертации и её оформления

Диссертационная работа содержит 146 страниц основного текста (всего 162 с.), 77 рисунков (без приложений) и 6 таблиц (без приложений). Работа состоит из введения, четырех глав, заключения, списка сокращений и условных обозначений, списка литературы из 105 наименований, 2 приложения.

В **первой главе** представлен анализ современного состояния предметной области, основанный на результатах исследований отечественных и зарубежных учёных. Автор

вводит основные определения, описывает что такое прогноз, какие методы прогноза бывают. Описываются формальные методы прогноза такие как авторегрессионные методы, авторегрессии проинтегрированного скользящего среднего, метод сингулярного спектрального анализа («Гусеница»), метод группового учета аргументов и метод на основе использования нейронной сети LSTM. Также автором представлен классический метод Data Assimilation, представлено описание алгоритма и модификации на основе фильтра Калмана. Проведен анализ преимуществ и недостатков методов. Здесь же на основе анализа состояния предметной области автор формулирует цель и задачи исследования.

Вторая глава посвящена обоснованию необходимости разработки методики интеграции формальных методов прогнозирования временных рядов и метода Data Assimilation. Представлен алгоритм работы методики и результаты сравнения точности прогнозирования системы Лоренца с использованием классического метода Data Assimilation и с помощью представленной методики.

В третьей главе диссертационной работы автор описывает разработанный программный комплекс «Прогнозирование временных рядов». Подробно описывается выбор инструмента разработки, функции, реализованные в программных библиотеках. Представлен графический пользовательский интерфейс.

Четвертая глава посвящена апробации разработанной методики с интеграции формальных методов прогнозирования временных рядов, представленных в главе 1, в метод Data Assimilation. Апробация выполнена на примере реальных временных рядов: «Air Passengers», набора данных солнечной активности и курса валют доллара и евро к рублю. Проведен анализ точности прогноза при применении представленных формальных методов. Представленные автором результаты показывают работоспособность разработанной методики.

Замечания и вопросы по работе

Вместе с этим следует отметить некоторые замечания и вопросы по содержанию диссертации:

1. Предлагаемый в диссертации метод прогнозирования временных рядов является по своей сути эвристической комбинацией существующих методов прогнозирования для моделей, которые автор называет «феноменологическими», с методом ассимиляции данных, являющимся частным случаем линейного фильтра Калмана. При этом автор не приводит какого-либо аналитического доказательства или объяснения оптимальности метода. Поэтому тот факт, что предложенный подход показал меньшие ошибки прогноза для всех рассмотренных в 4 главе временных рядов, отнюдь не гарантирует оптимальности

прогноза во всех потенциальных случаях (в отличие от фильтра Калмана, гарантирующего минимальную среднеквадратичную ошибку прогноза).

2. Представленные преимущества предложенного в диссертации метода перед методом ARIMA-прогнозирования, очевидно, являются результатом некорректного применения последнего:

2.1. Уравнения состояния (1.6), (1.7), (1.8) для моделей AR, MA и ARIMA соответственно, некорректно используются для вычисления прогнозов. Непонятно, как производятся эти вычисления, если в правой части содержатся неизвестные аддитивные случайные компоненты. В действительности, в теории ARIMA-моделирования предлагаются математически строго выведенные формулы для вычисления прогнозов с минимальной среднеквадратичной ошибкой)^{1, 2}, в том числе, формулы для обновления прогнозов, и они не совпадают с уравнениями (1.6), (1.7), (1.8).

2.2. Оптимальный результат прогнозирования с использованием методологии ARIMA предполагает первоочередную роль технологии построения адекватной модели, начинающейся с этапа идентификации, когда определяется тип модели и порядок полиномов AR и MA, последующего оценивания параметров первичной модели методом максимального правдоподобия, и, наконец, диагностики модели, включающей такую важную процедуру, как анализ остатков. Модель считается адекватной, т.е. соответствующей наблюдаемому процессу, если остатки подтверждают исходное предположение о белом шуме в уравнении модели. В этой связи классический временной ряд авиаперевозок, рассмотренный в главе 4, был идентифицирован как соответствующий модели AR(12), являющейся, по всей видимости, не вполне адекватной (анализ остатков модели в диссертации не приводится). В литературе^{1,2} убедительно показано, что наиболее подходящей моделью для ряда авиаперевозок является мультипликативная модель ARIMA (0,1,1) x (0,1,1)₁₂, учитывающая и нестационарность временного ряда, и сезонную составляющую (кроме того, эта модель содержит всего 2 оцениваемых параметра, а не 12, как AR(12), то есть соответствует лучшему значению критериев качества моделей AIC и BIC). Очевидно, таким образом, что преимущество предлагаемого в диссертации метода для временного ряда авиаперевозок объясняется в большой степени неадекватностью используемой модели ARIMA и некорректно применяемым аппаратом прогнозирования. В

¹. Wei W.S. Time Series Analysis (Univariate and multivariate methods). – Addison-Wesley Publishing Company, 1990. - 554 p.;

² Авдеев Т. В. Компьютерные методы анализа временных рядов и прогнозирование: учеб. пособие, Изд-во НГТУ, 2008. - 272 с.

неадекватностью используемой модели ARIMA и некорректно применяемым аппаратом прогнозирования. В случае адекватной модели и метода прогнозирования ARIMA заведомо должен получиться оптимальный прогноз с минимальной среднеквадратичной ошибкой.

3. Основной целью диссертации является достижение наиболее точного прогноза. Однако в погоне за точностью возникает риск переобучения, когда сложность модели (например, число оцениваемых параметров) не соответствует информационному содержанию данных. Вопрос переобучения в диссертации упоминается лишь в связи с моделью нейронной сети, хотя данный вопрос актуален для всех типов моделей (например, регрессионных), и заслуживает отдельного рассмотрения.

4. В диссертации представлено описание разработанного программного комплекса, однако не представлены скоростные характеристики. Позволяет ли предлагаемый комплекс решать задачи в реальном времени? Являются ли скоростные характеристики дополнительным преимуществом предложенного метода прогнозирования?

Данные замечания имеют дискуссионный характер и не снижают научной ценности работы.

Заключение по работе

Исследовательская работа изложена грамотным научно-техническим языком, в полной мере отвечает требованиям по актуальности, научной новизне, практической значимости, личному вкладу автора, отражению результатов в публикациях, а также полностью соответствует п. 9 Положения о присуждении ученых степеней в УрФУ и специальности 2.3.1. Системный анализ, управление и обработка информации, статистика. Автор диссертации Тимошенко Юлия Сергеевна заслуживает присуждения учёной степени кандидата технических наук по специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Официальный оппонент:

Доктор технических наук, профессор
Профессор кафедры теоретической и прикладной информатики
Федеральное государственное бюджетное образовательное учреждение высшего образования «Новосибирский государственный технический университет»
Тел.: +7 (383) 346-06-00
e-mail: avdeenko@corp.nstu.ru
Адрес: 630073, г. Новосибирск, проспект Карла Маркса, 20, 1 учебный корпус


(подпись)

17.11.2012
(дата)

ПОДПИСЬ ЗАВЕРЯЮ
Начальник отдела кадров
ФГБОУ ВО НГТУ



О. К. Пустовалова

Авдеенко Татьяна Владимировна