

# Отзыв официального оппонента

на диссертацию Чирихина Константина Сергеевича

**«Использование методов теории информации и искусственного интеллекта для разработки и исследования высокоточных методов прогнозирования временных рядов»**, представленной на соискание ученой степени кандидата технических наук по специальности 05.13.18 – «Математическое моделирование, численные методы и комплексы программ»

**Актуальность.** Временные ряды повсеместно встречаются на практике. Несмотря на то, что методы прогнозирования временных рядов начали разрабатываться достаточно давно (по меньшей мере с первой половины 20 века), а также на то, что в настоящее время для решения данной задачи активно используются методы машинного обучения, задача все ещё далека от своего разрешения. Зачастую в широко используемых методах предполагается стационарность временного ряда, что часто нарушается в практических случаях. Кроме того, даже небольшое повышение точности прогнозов может иметь важное прикладное значение, которое может выражаться, например, в оптимизации производственных процессов и повышении прибыли предприятия. В представленной диссертационной работе разрабатываются и экспериментально исследуются методы прогнозирования, основывающиеся на комбинировании методов теории информации, теории формальных языков и искусственного интеллекта и учитывающие некоторые закономерности, которые не могут быть учтены моделями для стационарных временных рядов. Значительная часть диссертации уделена разработке и развитию теоретико-информационного метода объединения различных методов сжатия и прогнозирования. К настоящему времени в ряде работ показано, что при комбинировании различных методов удаётся получать более точные прогнозы, чем с помощью отдельных методов. Поэтому актуальность исследования не вызывает сомнений.

**Цели и задачи работы.** Целями диссертации Константина Сергеевича Чирихина являются разработка, программная реализация и экспериментальное исследование методов прогнозирования временных рядов, способных обнаруживать «сложные» скрытые закономерности во временных рядах для повышения точности прогнозов. В работе решаются следующие задачи:

1. Построение методов прогнозирования временных рядов, использующих алгоритмы искусственного интеллекта для обнаружения скрытых закономерностей в данных с целью повышения точности прогнозов.
2. Разработка метода сведения алгоритмов прогнозирования к методам сжатия данных с целью использования алгоритмов искусственного интеллекта, применяемых в методах сжатия, а также для объединения различных методов прогнозирования в один комплекс.
3. Решение задачи простого выбора лучшего метода прогнозирования из имеющихся.

**Структура и основное содержание диссертации.** Полный объем диссертации Константина Чирихина — 107 страниц. В работе подробно излагаются полученные результаты, а также содержится довольно подробный обзор существующих методов прогнозирования временных рядов.

В *главе 1* содержатся основные сведения о теоретико-информационном подходе к прогнозированию временных рядов. Предлагается метод прогнозирования, построенный на базе объединения нескольких методов сжатия данных в один метод прогнозирования. Автор отмечает, что заранее неизвестно, какой из доступных методов сжатия окажется наиболее точным при прогнозировании того или иного временного ряда, и данный подход позволяет автоматизировать выбор такого метода. Все последующие результаты в той или иной форме опираются на этот подход. Например, метод на основе конечных автоматов из главы 2 хорошо приспособлен для прогнозирования только определённого класса закономерностей, с целью получения универсального метода его предлагается объединять с методами сжатия. Результаты главы 3 также имеют прямое отношение к этому подходу — повышают его универсальность и сокращают объем необходимых вычислений. Далее, отмечается, что напрямую теоретико-информационные методы можно использовать только для прогнозирования целочисленных временных рядов. Вещественные временные ряды необходимо преобразовывать к целочисленным с помощью квантования. В работе предлагается достаточно интересный подход к автоматическому выбору числа уровней при квантовании: по исходному вещественному ряду строятся  $k$  целочисленных рядов с помощью квантования с использованием  $2^i, i = 1, 2, \dots, k$ , уровней, каждый из рядов сжимается с его возможными продолжениями заданной длины и затем полученные длины кодовых слов используются при формировании итогового распределения вероятностей для будущих значений временного ряда. В конце главы приводится обзор методов сжатия на основе грамматических моделей, поясняется, каким образом контекстно-свободные грамматики могут быть использованы при прогнозировании.

В *главе 2* предлагается метод прогнозирования временных рядов на основе конечных автоматов. За основу берётся алгоритм Смита прогнозирования полилинейных слов для 10-головочного автомата. Предлагается модификация этого алгоритма, позволяющая с помощью него получать длины кодовых слов для входных последовательностей, что в свою очередь делает возможным включение этого метода в комбинацию с методами сжатия.

В *главе 3* далее развивается предлагаемый подход к объединению методов прогнозирования. С целью уменьшения дополнительных затрат времени, возникающих при использовании набора методов сжатия вместо одного метода, на базе универсальных по времени кодов строится адаптивный метод прогнозирования. Основной его идеей является использование тестового фрагмента данных, на котором пробуются все доступные методы и выбирается тот,

который обеспечивает наилучшее сжатие. Также уделяется внимание проблеме включения в комбинацию с методами сжатия методов прогнозирования, не основанных на сжатии. Для этого разрабатывается метод оценки размера сжатого представления рассматриваемого временного ряда с помощью любого метода прогнозирования, способного построить прогноз на один шаг для последовательности вещественных чисел.

В главе 4 содержатся результаты вычислений на реальных данных, выполненных с использованием предложенных в диссертации методов. В начале соискатель приводит описание методологии вычислений, используемой в данной главе. Затем рассматриваются: прогнозирование временных рядов, участвовавших в конкурсе M3 Competition (3003 временных ряда преимущественно из областей экономики и бизнеса), прогнозирование временных рядов значений Т-индекса, планетарного К-индекса, временного ряда среднемесячных чисел солнечных пятен, а также прогнозирование набора показателей Новосибирской области. Наиболее интересные результаты получились при прогнозировании среднемесячных чисел солнечных пятен и Т-индекса (эти временные ряды связаны): для них доступны прогнозы, построенные метеорологическим бюро, и точность прогнозов на один шаг, построенных в диссертации, оказалась выше, чем точность прогнозов бюро, что подтверждает эффективность предложенных методов.

Глава 5 посвящена описанию программной реализации разработанных в диссертации методов. Программа состоит из библиотеки, в которой реализованы основные вычислительные процедуры, и пакета для Python, основанного на этой библиотеке. К диссертации приложено свидетельство о государственной регистрации этой программы.

**Научная новизна.** Научная новизна работы К. С. Чирихина состоит в разработке метода прогнозирования временных рядов, способного обнаруживать закономерности, которые не могут быть обнаружены другими методами, в реализации программного комплекса, применении его к исследованию реальных временных рядов. В диссертации разработан метод интеграции различных методов сжатия и прогнозирования, предложен способ повышения его вычислительной эффективности.

**Научная и практическая значимость.** Полученные в диссертации К.С. Чирихина результаты могут быть интересны широкому кругу специалистов, сталкивающихся в своей работе с задачей прогнозирования временных рядов. В частности, рассматриваемые в диссертации методы позволяют выявлять скрытые закономерности, не учитываемые моделями для стационарных временных рядов, в экономических и астрономических данных.

**Достоверность результатов.** Достоверность полученных автором результатов обеспечивается адекватным применением методов математической статистики, теории информации и дискретной математики.

**Апробация результатов исследования.** Основные результаты по теме диссертации изложены в 9 печатных изданиях, в т.ч. 1 статья в журнале, индексируемом в Scopus, 2 статьи в журналах, входящих в перечень ВАК, 6 публикаций, включённых в сборники трудов конференций. Получено свидетельство о государственной регистрации программы для ЭВМ. Результаты диссертации хорошо отражены в рецензируемых научных изданиях и докладывались на конференциях.

Недостатки работы:

1. В работе представлены многочисленные примеры успешного применения предложенных подходов для прогнозирования различных временных рядов. Вместе с тем, в работе отсутствует описание класса тех рядов, для которых применение предложенного подхода может быть эффективным в смысле точности или сложности.
2. На стр. 46 предлагается вычислять время сжатия блока из  $t$  букв как  $T=tv$ , где  $v$  - максимальное время сжатия одной буквы. Такая оценка может оказаться весьма неточной при обработке больших блоков данных на процессорах общего назначения, т.к. в этом случае на время вычислений могут оказывать влияние эффекты, связанные с кэш-промахами. Эти эффекты возникают в том случае, если все обрабатываемые данные не помещаются в кэш процессора, и могут привести к значительному увеличению времени вычислений.
3. На стр. 27 указано, что формулы (1.6) и (1.7) могут быть использованы совместно. Неясно, как именно это может быть сделано.
4. В работе отсутствует сравнение предлагаемого подхода с некоторыми существующими методами предсказания, основанными на конечных автоматах:
  - Boné, R., Daguin, C., Georgevail, A., Maurel, D. (1997). Time series forecasting by finite-state automata. In: Raymond, D., Wood, D., Yu, S. (eds) Automata Implementation. WIA 1996. Lecture Notes in Computer Science, vol 1260. Springer, Berlin, Heidelberg
  - Chen, Y., Wu, Z., Li, Z., Zhang, Y. (2010). Research on Time Series Forecasting Model Based on Moore Automata. In: Cao, L., Feng, Y., Zhong, J. (eds) Advanced Data Mining and Applications. ADMA 2010. Lecture Notes in Computer Science, vol 6440. Springer, Berlin, Heidelberg.
5. В формулах (1.2), (1.4), (1.6) используются длины кодовых слов в битах. Учитывается ли в программной реализации, что в последнем байте после сжатия могут быть использованы не все восемь бит?
6. В описании программной реализации указано, что функции автоматического построения прогнозов на тренировочной выборке и вычисления доверительных интервалов реализованы в пакете `itr` на Python. Не было бы целесообразнее реализовать эти функции в библиотеке `itr_core`? В текущем варианте при реализации доступа к `itr_core` из каждой новой среды программирования эти функции необходимо переписывать заново.
7. В параграфе 4.3 иногда используется термин «временной ряд солнечных пятен» (стр. 59, 62, 63 и др.). Он видится некорректным. Вероятно, в местах его использования автор имеет в виду термин «временной ряд среднемесячных чисел солнечных пятен», который также широко используется.

Отмеченные недостатки не снижают общего положительного впечатления от работы.

**Заключение.** Диссертационная работа Константина Сергеевича Чирихина «Использование методов теории информации и искусственного интеллекта для разработки и исследования высокоточных методов прогнозирования временных рядов» является законченной научно-квалификационной работой, выполненной на актуальную тему, в которой получены новые результаты, имеющие важное научное и прикладное значение. Работа выполнена на высоком научном уровне,

приведённые выводы и рекомендации являются обоснованными. Автореферат хорошо отражает содержание диссертации и в полной мере передаёт полученные результаты. Основные результаты диссертации неоднократно обсуждались на международных и российских конференциях, достаточно полно отражены в рецензируемых научных изданиях. Считаю, что диссертация «Использование методов теории информации и искусственного интеллекта для разработки и исследования высокоточных методов прогнозирования временных рядов» удовлетворяет критериям пунктов 9-14 Положения о присуждении учёных степеней (утверждено постановлением Правительства Российской Федерации от 24 сентября 2013 г., N 842), а её автор – Чирихин Константин Сергеевич заслуживает присуждения ему учёной степени кандидата технических наук по специальности 05.13.18 – «Математическое моделирование, численные методы и комплексы программ».

Официальный оппонент,  
профессор факультета безопасности  
информационных технологий  
Университета ИТМО, д.т.н., доцент

  
Трифонов Петр Владимирович


31.08.2022

Трифонов Петр Владимирович доктор технических наук по специальности 05.13.17 – Теоретические основы информатики, профессор факультета безопасности информационных технологий Федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский университет ИТМО».

Адрес: 197101, г. Санкт-Петербург, Кронверкский пр., д. 49, лит. А.

Рабочий тел.: +7 (812) 480-00-00.

Электронный адрес: pvtrifonov@itmo.ru

Подпись   
удостоверяю  
Менеджер ОПС  
Шипик В.А.

