

УДК 517

АНАЛИЗ ХАРАКТЕРИСТИК ВРЕМЕННЫХ РЯДОВ СОЦИОДИНАМИЧЕСКИХ ПРОЦЕССОВ В СЕТЯХ КОММЕНТАРИЕВ ПОЛЬЗОВАТЕЛЕЙ МАСС-МЕДИА

К. К. Отраднов, старший преподаватель кафедры информационного противоборства института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет», Москва, Россия;

orcid.org/0000-0002-0055-9651 e-mail: const.otradnov@yandex.ru

Ю. П. Перова, старший преподаватель кафедры телекоммуникаций института радиоэлектроники и информатики ФГБОУ ВО «МИРЭА – Российский технологический университет», Москва, Россия;

orcid.org/0000-0003-4028-2842 e-mail: jul-np@yandex.ru

В. Р. Григорьев, к.т.н., доцент, заведующий кафедрой информационного противоборства института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет», Москва, Россия;

orcid.org/0000-0003-0279-6391 e-mail: grigorjev_vr@mail.ru

Д. О. Жуков, д.т.н., профессор кафедры информационного противоборства института кибербезопасности и цифровых технологий ФГБОУ ВО «МИРЭА – Российский технологический университет», Москва, Россия;

orcid.org/0000-0002-1211-5214 e-mail: zhukovdm@yandex.ru

Целью работы является анализ наблюдаемой динамики и характеристик временных рядов, описывающих процессы в сложных системах, характеризующихся наличием человеческого фактора, приводящего не только к проявлению неопределенности и стохастичности, но и к возможности реализации в системах памяти и самоорганизации. В качестве объекта исследований были выбраны динамические процессы, наблюдаемые в сетях комментариев пользователей масс-медиа. Обработка наблюдаемых данных показала, что в распределении амплитуд изменения активности пользователей при комментировании новостей наблюдается небольшая величина асимметрии и распределение амплитуд является почти симметричным, но наблюдается так называемый «тяжелый хвост» – графики плотности распределений лежат выше графика нормального распределения. Проведенные исследования позволяют сделать ряд важных выводов о том, что изученные временные ряды являются нестационарными, а описываемые ими процессы и системы обладают краткосрочной памятью (показатель Хёрста значительно меньше 0,5). Существенное отличие определенных для них показателей Хёрста от 0,5 говорит о том, что описываемые ими процессы или системы не только обладают памятью, но и их структура имеет фрактальный характер. Фрактальность может быть связана с тем, что наблюдаемые процессы, характеризуются дробными переменными измерения, а значит при выводе аппроксимирующих функций плотности распределения их параметров с целью построения прогностических моделей целесообразно использовать дробно – дифференциальные уравнения.

Ключевые слова: социодинамические процессы, активность пользователей сетевых масс-медиа, временные ряды, показатель Хёрста, память, самоподобие, фрактальность временного ряда.

DOI: 10.21667/1995-4565-2022-82-147-161

Введение

Новости и блоги, под которыми пользователи сети интернет оставляют свои комментарии, являются одним из важнейших сетевых объектов, поскольку могут выполнять в онлайн-режиме роль индикаторов общественного мнения. Если в новостях или блогах затрагивается общественно значимая тема, то это, как правило, притягивает, как её сторонников, так и противников, которые вступают в дискуссии и оставляют взаимные комментарии. И чем более резонансной является новость или блог, тем выше активность

пользователей и больше число комментаторов (появляется многоуровневая структура комментариев к комментариям). Анализ активности пользователей новостных сообщений и блогов является одной из практически значимых и актуальных научных задач, решение которой позволяет обеспечивать стабильность развития общества.

Изменения активности пользователей в социальных сетях и их комментарии по общественно значимым темам образуют временные ряды, анализ динамики которых и их характеристик играет важную роль с точки зрения прогнозирования направленности социодинамических процессов.

Сложные системы (а социальные сети и сетевые масс-медиа относятся именно к таким системам) можно определить как структуры, в которых в качестве хотя бы одного из элементов выступает человек. С одной стороны, действие множества случайных факторов приводит к стохастичности свойств таких систем. Кроме того, помимо стохастичности, ещё возникает и неопределенность, связанная с иррациональностью поведения людей в некоторых случаях. А с другой стороны, наличие человеческого фактора создает предпосылки для самоорганизации системы и может определять существование памяти о её предыдущих состояниях. Всё это приводит к появлению организованной сложности (эмерджентности), возникновение которой нельзя рассматривать как итог простого суммирования характеристик элементов, так как она является результатом возникновения системных связей и адаптивного перераспределения функций между элементами. Следует отметить, что большинство процессов, наблюдаемых в сложных социальных и экономических системах, может быть отнесено к классу нестационарных временных рядов. Из-за нелинейности и изменчивости характеристик нестационарных временных рядов процессов, протекающих в сложных социальных и экономических системах, традиционные методы анализа и моделирования, например, такие как интегрированная модель авторегрессии – скользящего среднего (ARIMA, модель Бокса – Дженкинса) и многие другие часто приводят к неточным или ошибочным результатам.

Поэтому, на наш взгляд, анализ особенностей поведения характеристик временных рядов социодинамических процессов в сетях комментариев пользователей масс-медиа является очень важным для понимания их природы и может быть предпосылкой для создания прогностических моделей их эволюции.

Обзор современных исследований динамики процессов в социальных сетях

В процессе исследования были изучены статьи, на основании которых можно сделать вывод, что методы прогнозирования временных рядов являются важнейшим аспектом понимания сложных сетей и управления ими. Можно рассмотреть целый ряд близких к тематике наших исследований работ.

Авторы статьи [1] подробно изучают две концепции временного анализа настроений и причинно-следственной связи настроений. Анализ временных настроений полезен для обобщения событий на основе настроений и времени. Причинно-следственная связь полезна для определения причинно-следственных связей событий, а также полезна для прогнозирования событий. Эти две концепции в сочетании приводят к модели прогнозирования событий, которая может прогнозировать период времени между событиями и настроение предстоящих событий. Эта работа вводит обобщенную модель прогнозирования, основанную на анализе временных настроений, чтобы определить причинно-следственную связь между событиями, которая может быть использована для прогнозирования настроения события и продолжительности между событиями. Предлагаемый метод оценивается с использованием показателей производительности, точности и отзыва.

Авторы статьи [2] рассматривают различные проблемы прогнозирования, связанные с сетью, обобщают и классифицируют соответствующие методы прогнозирования, анализируют их преимущества и недостатки и указывают на передовые, а также критические проблемы в этой области.

В статье [3] представлена модель прогнозирования новостных событий в новостных лентах, разработанная на основе стохастической динамики изменений структуры новостных кластеров с памятью предыдущих состояний и самоорганизацией их структуры. Прогнозирование событий выполняется на основе их текстового описания, векторизации текста и вычисления косинусов углов между вектором данного текста и центроидами различных семантических кластеров в информационной системе (ИС). При создании модели строятся вероятностные схемы переходов из состояния в состояние на ИС, учитывающие набор предыдущих шагов. На основе предложенного подхода получено нелинейное дифференциальное уравнение второго порядка, сформулирована и решена краевая задача прогнозирования новостных событий. Это позволяет получить теоретическую зависимость от времени функции распределения плотности вероятности параметров нестационарного временного ряда, описывающих эволюцию ИБ. Экспериментальная проверка разработанной модели показала ее надежность и адекватность.

В статье [4] описана модель прогнозирования событий в новостных лентах, которая основана на использовании стохастической динамики изменений структуры нестационарных временных рядов в новостных кластерах (состояниях информационного пространства) на основе использования диффузионной аппроксимации. Прогнозирование событий в ленте новостей основано на их текстовом описании, векторизации и нахождении значения косинуса угла между заданным вектором и центроидами различных семантических кластеров информационного пространства.

Изменения со временем значения косинуса таких углов между вектором и центроидами можно представить в виде точки, блуждающей по отрезку $[0, 1]$. Этот сегмент содержит ловушку в пороговой точке возникновения события, в которую в конечном итоге может попасть блуждающая точка. При создании модели авторы учитывали вероятностные закономерности переходов между состояниями в информационном пространстве.

На основе этого подхода было выведено нелинейное дифференциальное уравнение второго порядка, сформулирована и решена краевая задача прогнозирования новостных событий, что позволило получить теоретическую временную зависимость функции плотности вероятности распределения параметров нестационарных временных рядов, описывающих эволюцию информационного пространства. Результаты исследования, которые были описаны в статье, позволили сделать ряд очень важных выводов. Разработанная модель прогнозирования событий в ленте новостей является адекватной и последовательной [все новостные события, использованные для тестирования модели, произошли с высокой вероятностью (около 80 %) или, если они являются вымышленными новостями, они могут произойти только через неприемлемо долгое время]. Анализ модели прогнозирования событий в ленте новостей на основе простой диффузионной модели подтверждает возможность прогнозирования событий ленты новостей на основе их текстового описания, векторизации и нахождения значения косинуса угла между этим вектором и центроиды различных информационных кластеров. Изменение этого косинуса с течением времени можно рассматривать как блуждание точки на отрезке $[0, 1]$, который включает в себя ловушку, в которую блуждающая точка может в конечном итоге попасть. Результат моделирования временной зависимости вероятности наступления события с экспериментально определенными наборами значений параметров разработанной модели не является противоречивым с точки зрения вероятностного поведения (помимо прочего, на больших временах вероятности асимптотически стремятся к единице). Оценки точности и надежности прогнозирования новостей позволяют предположить, что разработанная модель может быть использована для прогнозирования, а относительная точность прогнозирования может быть выше 70 %.

В работе [5] авторы показали на примере финансового временного ряда возможность получения важных повсеместных физических характеристик при учете агрегации многих систем. Обсудили возможность получения не только аномального процесса диффузии, но и нелинейного уравнения диффузии, которое приводит к распределению вероятности при ис-

пользовании набора немарковских процессов. Это распределение вероятности демонстрирует поведение по закону мощности в структуре своих хвостов. Оно также отражает аномальные транспортные характеристики ансамбля частиц. Это повсеместное поведение, с законом мощности в диффузионном переносе и структуре распределения вероятности, связано с быстрым флуктуационным явлением, представленным в параметре шума.

Выбор источников данных и сбор статистики активности пользователей новостных ресурсов

Для наших исследований были выбраны несколько интернет-ресурсов (в частности, был взят портал российской радиостанции «Эхо Москвы» – советом директоров ЗАО «Эхо Москвы» от 3 марта 2022 года было принято решение о ликвидации радиоканала и одноимённого электронного периодического издания «Эхо Москвы»; информационный ресурс «Риа Новости») и одно из сетевых сообществ социальной сети «ВКонтакте», посвященное обсуждению новостей (<https://vk.com/ria>).

«РИА-новости» был выбран исходя из его узнаваемости и популярности в российском обществе, занимает 1 место среди медиаресурсов (за март 2022) по версии «br-analytics». (<https://br-analytics.ru/mediatrends/media/?period=202203>), входит в ТОП-3 самых цитируемых информационных агентств в СМИ и социальных медиа (за март 2022) и занимают 1-е место. (<https://www.mlg.ru/ratings/media/federal/11110/#internet>).

Выбор новостного портала российской радиостанции «Эхо Москвы» определяется следующими причинами.

1. Новостной портал (радиостанция) входил в топ – 10 новостных сайтов России и на июль 2021 года и занимал 9-е место по посещаемости и 7-е по активности пользователей, также на конец июля 2021 входил в топ – 8 цитируемых радиостанций и занимал 1-е место (по гиперссылкам в социальных медиа) на конец августа 2021 г. и 4-е место (по индексу цитирования в СМИ).

2. Портал имел разнообразную тематику (присутствуют новости политической, спортивной, экономической, научной, культурной направленности и т.д.).

3. Новостной портал существовал с 1990 года и публиковал блоги известных медийных личностей.

4. Практически не проводилась премодерация комментариев (премодерация применялась только к новым пользователям или пользователям, которые ранее нарушали правила новостного портала), но существовала постмодерация дискуссий (с требованиями к комментариям и запретами к их размещению). Пользователи могут высказывать различные мнения (не совпадающие с официальной позицией), и их комментарии удаляются только за нарушение правил.

Сначала мы с помощью специального программного приложения (парсера) скачивали интересующий нас диапазон новостей начиная с 1 января 2019 года по апрель 2022 года с ресурса «ВКонтакте», используя разработанный нами парсер и API сети (<https://dev.vk.com/guide>) и Внутри социальной сети у каждого поста есть свой уникальный адрес (https://vk.com/ria?w=wall{owner_id}_{post_id}), где {owner_id} – уникальный идентификатор сообщества, в случае с «РИА-новости» это "-15755094", а {post_id} – уникальный идентификатор поста (новости). У каждого поста (новости) присутствует ряд основных параметров таких как: уникальный идентификатор поста в социальной сети; текст поста; дата и время публикации; количество просмотров и комментариев пользователей. А комментарии имеют следующие параметры: уникальный идентификатор в сообществе социальной сети; уникальный идентификатор пользователя; текст комментария; дата и время появления; уровень иерархии комментария; связь по уровню комментирования с родительским комментарием (кто из пользователей комментировал кого из других пользователей при обсуждении новости).

Портал российской радиостанции «Эхо Москвы» распределяет новости по дням, и с каждым отдельным днем можно ознакомиться по ссылке (<https://echo.msk.ru/news/{year}/>

{month}/{day}/", где day – день, month – месяц, year – год). У каждой новости существует ряд параметров, таких как: текст новости, уникальный номер новости на портале, заголовок, адрес страницы (URL), и также метаданные (дата и время публикации, количество просмотров и комментариев), тексты комментариев пользователей (а также уникальный номер комментария, уникальный номер пользователя, дата и время появления комментария, уровень иерархии комментария, связь по уровню комментирования с родительским комментарием) и доступные сведения об их авторах (уникальный номер пользователя на сайте, город, род занятий, место работы, имя или “никнэйм”, дата регистрации, количество рекомендаций и просмотров профиля пользователя, общее число комментариев за наблюдаемый период и т.д.). В среднем количество новостей варьировалось от 160 до 190 в день. В процессе сбора данных также были загружены сведения о том, кто из пользователей комментировал кого из других пользователей при обсуждении новости. Нами были загружены данные с портала российской радиостанции «Эхо Москвы» за последние 10 лет.

Поскольку комментарии могли оставлять чат-боты, спамеры и недобросовестные пользователи, которые пишут комментарии на профессиональной основе, необходимо было ввести правила отчистки данных. К недобросовестным были отнесены те, кто написали за год более 7365 комментариев (в среднем более 20 за сутки) или писали с частотой более 1-го комментария в 5 минут.

На основании полученных данных была создана база данных архива новостей, которую мы использовали для анализа и разработки моделей.

На рисунке 1 в качестве примера представлен исходный (ненормированный) временной ряд суточной активности пользователей портала «РИА Новости» по комментированию новостей с января 2013 по декабрь 2020 года. Активность определялась как отношение общего числа уникальных комментариев, оставленных пользователями ко всем новостям в течение суток к общему количеству уникальных просмотров пользователями в течение суток всех опубликованных новостей.

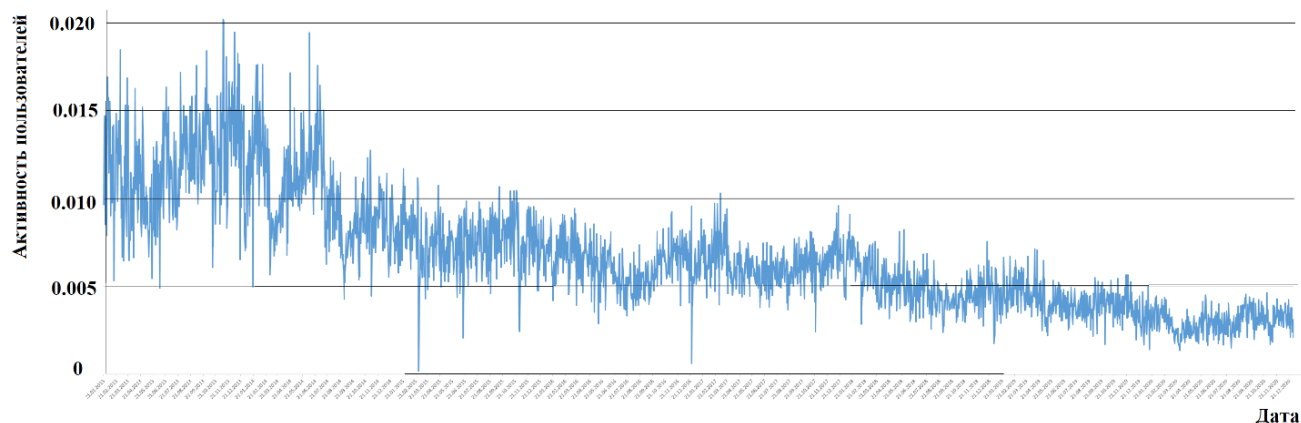


Рисунок 1 – Ненормированный временной ряд активности пользователей портала «РИА Новости» по комментированию новостей с января 2013 по декабрь 2020 года
Figure 1 – Non-normalized time series of RIA Novosti portal users' activity in commenting on news from January 2013 to December 2020

Также можно использовать и нормированные временные ряды. При нормировке может быть использован следующий подход. Вычисляется разность между двумя ближайшими активностями (последующей и предыдущей) и делится на величину предыдущей активности.

Обработка наблюдаемых данных и анализ статистики активности пользователей новостных ресурсов

Для исследования динамики временных рядов и определения их характеристик можно использовать метод нормированного размаха Хёрста [6], позволяющий определить их фрактальную размерность и классифицировать их тип поведения. Можно пронормировать размах величин уровней ряда R стандартным отклонением значений величин уровней ряда S и представить их отношение (нормированный размах) в виде уравнения: $R / N = C\tau^H$, где C – некоторая константа, τ – число наблюдений (уровней ряда), составляющих рассматриваемый временной ряд (ВР), а показатель степени H – это так называемый коэффициент или показатель Хёрста. Наличие изломов в зависимости $R/S(\tau)$ свидетельствует о наличии характерных временных масштабов и/или периодичностей. Величина H позволяет дать классификацию временных рядов по характеру их поведения.

При $H = 0.5$ (белый шум) уровни ряда случайны и некоррелированы. Размах накопленных отклонений должен увеличиваться пропорционально квадратному корню из времени.

При $0 < H < 0.5$ (черный шум) ряд является антиперсистентным (эргодическим). В этом случае рост в прошлом означает уменьшение в будущем, а тенденция к уменьшению в прошлом делает вероятным увеличение в будущем. И чем меньше H , тем больше эта вероятность. В таких процессах после возрастания переменной обычно происходит её уменьшение, а после уменьшения – возрастание. Неустойчивость такого антиперсистентного ряда зависит от близости H к нулю: приближение влечёт усиление изменчивости ряда. Это можно рассматривать, как краткосрочную память.

При $0.5 < H < 1$ (розовый шум) временной ряд является персистентным (трендоустойчивым). Если ряд возрастает (убывает) в предыдущий период, то вероятно, что он будет сохранять эту тенденцию еще какое-то время в будущем. Это можно рассматривать как долгосрочную память.

При наличии выраженной персистентности или антиперсистентности можно говорить о том, что уровни ряда не являются независимыми и каждое наблюдение несёт память обо всех предшествующих событиях, а недавние события оказывают более сильное влияние, по сравнению с отдалёнными событиями.

Алгоритм метода Хёрста для анализа временных рядов заключается в следующем.

1. Пусть $\xi(t)$ – последовательность измерений значений наблюдаемого временного ряда (например, число значений равно N). Разделим временной ряд на непересекающиеся последовательности некоторой длины τ (величину τ можно интерпретировать как интервал времени или как число последовательных отсчетов значений ряда или величины выборки значений или уровней ряда). Общее число таких последовательностей будет равно N/τ . Величину τ выбирают таким образом, чтобы, с одной стороны, оно было статистически значимым, а с другой – общее число последовательностей тоже было бы статистически значимым. Вычислим среднее значение величины $\langle \xi \rangle_j$ за выбранный интервал τ для каждой из последовательностей $j (1 \leq j \leq \text{integer}[N/\tau])$: $\langle \xi \rangle_j = \frac{1}{\tau} \sum_i^{\tau} \xi_i(j, t)$.

2. Далее для каждой из последовательностей j найдём $X(j, \tau)$ – накопленные отклонения от среднего значения: $X(j, \tau) = \xi_i(j, t) - \langle \xi \rangle_j$.

3. Затем для каждой из последовательностей j вычисляем размах $R(j, \tau)$, как разность между максимальным и минимальным из найденных накопившихся отклонений $X(j, \tau)$: $R(j, \tau) = \max X(j, \tau) - \min X(j, \tau)$.

4. Далее для каждой из последовательностей j найдём $S(j, \tau)$ стандартное отклонение

(квадратный корень из дисперсии): $S(j, \tau) = \sqrt{\frac{\sum_{i=1}^{\tau} \{\xi_i(j, t) - \xi_j\}^2}{\tau}}$. И вычисляем значение $R(j, \tau) / S(j, \tau)$.

5. Проводим усреднение величины $R(j, \tau) / S(j, \tau)$ по всем последовательностям j для данного интервала τ : $\langle R / S \rangle_{\tau} = \frac{\tau}{N} \sum_{j=1}^{N/\tau} \{R(j, \tau) / S(j, \tau)\}$.

6. Изменяем величину интервала разделения временного ряда τ и повторяем выполнение шагов 1 – 6. Затем строим зависимость логарифма величины $\langle R / S \rangle$ от значений величины τ и находим методом наименьших квадратов угловой коэффициент прямой линейной регрессии, который даёт значение показателя Хёрста.

Для зависимости величины логарифма R/S от логарифма величины выборки уровней временного ряда активности пользователей портала «РИА Новости» по комментированию новостей (см. рисунок 2, прямая 1) может быть получено следующее линейное уравнение: $y = 0,25x - 1,12$ со значением коэффициента корреляции $R^2 = 0,99$.

В данном случае $H = 0,25$, и, следовательно, наблюдаемый временной ряд является антиперсистентным (эргодическим). Поскольку величина коэффициента Хёрста существенно отличается от 0,5, то из этого следует, что структура данного ряда обладает фрактальностью, а описываемые им процессы могут иметь краткосрочную память.

Для зависимости величины логарифма R/S от логарифма величины выборки уровней временного ряда активности пользователей портала «Эхо Москвы» по комментированию новостей (см. рисунок 2, прямая 2) может быть получено следующее линейное уравнение: $y = 0,18x + 0,27$ со значением коэффициента корреляции $R^2 = 0,89$. В данном случае $H = 0,18$ и, следовательно, наблюдаемый временной ряд также является антиперсистентным (эргодическим).

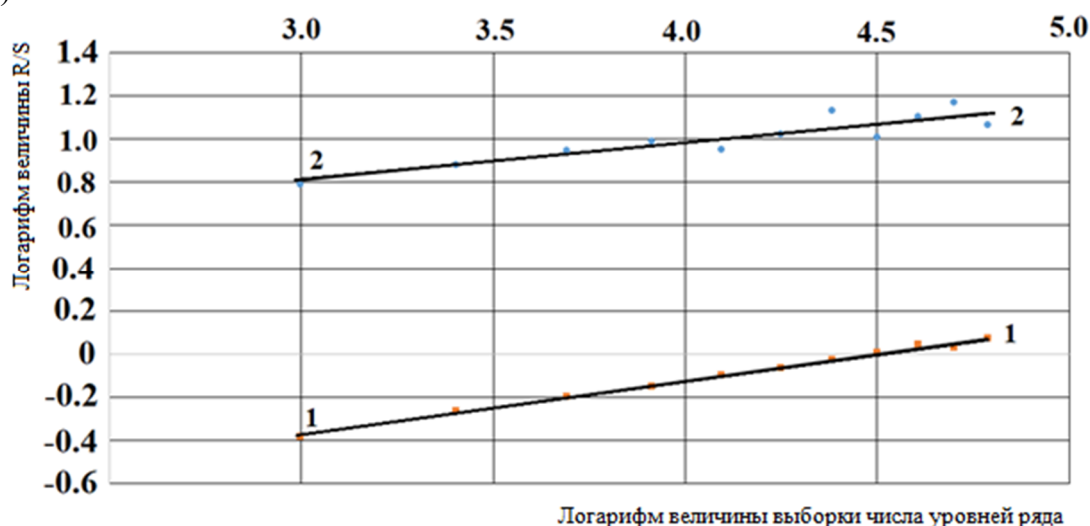


Рисунок 2 – Определение показателя Хёрста ненормированных временных рядов активности пользователей портала «РИА Новости» – прямая 1 и «Эхо Москвы» – прямая 2 по комментированию новостей за наблюдаемые периоды времени
Figure 2 – Determination of the Hurst indicator of non-normalized time series of activity of users of the portal «RIA Novosti» – direct 1 and «Echo of Moscow» – direct 2 for commenting on news for the observed time periods

Существенное отличие показателя Хёрста от величины равной 0,5 говорит о том, что наблюдаемые характеристики исследуемых временных рядов активности пользователей должны существенно отличаться от нормального закона распределения. Однако это требует дополнительной проверки.

Для дальнейшей обработки наблюдаемых данных и определения свойств функций плотности вероятности величины активности пользователей сетевых новостных ресурсов (это необходимо для анализа характеристик и динамики наблюдаемых процессов) можно использовать следующий алгоритм.

1. Сначала можно задать размер «скользящего окна» (интервал времени между наблюдаемыми значениями активности пользователей сетевых ресурсов), например, одни сутки, двое суток, трое и т.д. и выбрать из временного ряда данные по активности для заданного временного диапазона (размера «скользящего окна»).

2. Далее по выбранным данным можно рассчитать амплитуды изменения величины активности пользователей сетевых новостных ресурсов для различных выбранных временных интервалов (размеров «скользящего окна»). Эти данные представлены на рисунке 3.

3. Затем, наборы значений, полученные для каждого из измеряемых интервалов, необходимо отсортировать по возрастанию (от отрицательного к положительному), и для каждого размера «скользящего окна» построить гистограммы плотности распределения амплитуд отклонений величины активности пользователей сетевых новостных ресурсов.

4. Потом можно по полученным гистограммам рассчитать моменты распределения (среднее значение: математическое ожидание, дисперсию, асимметрию, эксцесс) для выбранных интервалов времени расчета амплитуд отклонений активности пользователей сетевых ресурсов (размеров «скользящего окна»).

5. Далее можно построить зависимости для математического ожидания (см. рисунок 4), дисперсии (см. рисунок 5), асимметрии (см. рисунок 6) и эксцесса (см. рисунок 7) амплитуд отклонений активности пользователей сетевых ресурсов от интервалов времени расчета (размеров «скользящего окна»).

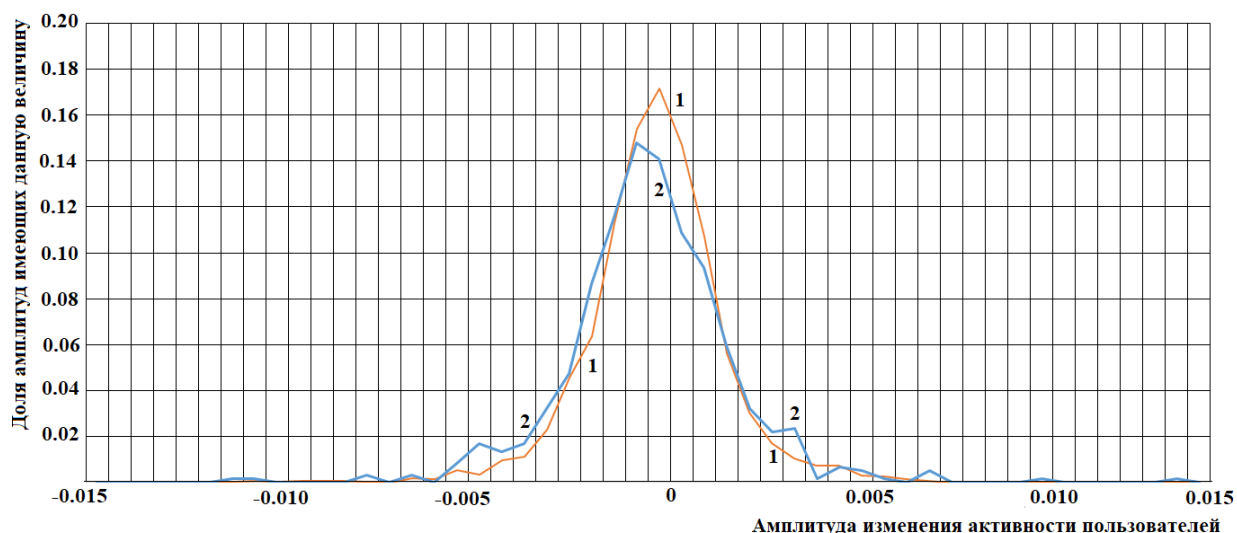


Рисунок 3 – Гистограммы распределения амплитуд изменения активности пользователей портала «РИА Новости» – линия 1 и «Эхо Москвы» – линия 2 при комментировании новостей по своим величинам для «скользящего окна» равного одному дню

Figure 3 – Histograms of the distribution of amplitudes of changes in the activity of users of the portal «RIA Novosti» – line 1 and «Echo of Moscow» – line 2 when commenting on news by their values for a «sliding window» equal to one day

Исследование гистограмм, а также зависимости математического ожидания, дисперсии, асимметрии (*третий момент распределения*) и эксцесса (*четвертый момент распределения*) амплитуд отклонений активности пользователей сетевых ресурсов от интервалов времени расчета (размеров «скользящего окна») позволяет определить, являются ли изучаемые временные ряды нестационарными и построить модели их эволюции, что необходимо для прогнозирования их динамики.

Проанализируем наблюдаемые данные. На рисунке 3 представлены гистограммы распределения по своим величинам амплитуд изменения активности пользователей при комментировании новостей, для «скользящего окна» равного одному дню. Как видно из рисунка 3, распределения амплитуд имеют острые пики, что мало характерно, если бы мы имели дело с нормальным законом распределения.

При увеличении интервала времени расчета амплитуд отклонений активности пользователей сетевых ресурсов (размеров «скользящего окна») вид гистограмм по ширине несколько изменяется, однако их положение относительно нуля остается практически постоянным (не наблюдается явление «сноса»).

Рисунок 4 показывает, что величины математического ожидания амплитуд изменения активности пользователей по комментированию новостей зависят от интервала времени расчета этих амплитуд («скользящего окна»), что указывает на то, что исследуемые временные ряды являются не стационарными, а их параметры не могут быть описаны нормальным законом распределения. Например, если обозначить величину амплитуды как x , а интервал времени её расчета, как t (величина «скользящего окна»), то при выполнении нормального закона

на распределения $\rho(x, t) = \frac{1}{2\sqrt{\pi at}} \cdot e^{-\frac{x^2}{4at}}$, величина математического ожидания $\mu(t)$ будет равна нулю:

$$\mu(t) = \int_{-\infty}^{+\infty} x \cdot \rho(x, t) dx = \frac{1}{2\sqrt{\pi at}} \int_{-\infty}^{+\infty} x \cdot e^{-\frac{x^2}{4at}} dx = 0$$

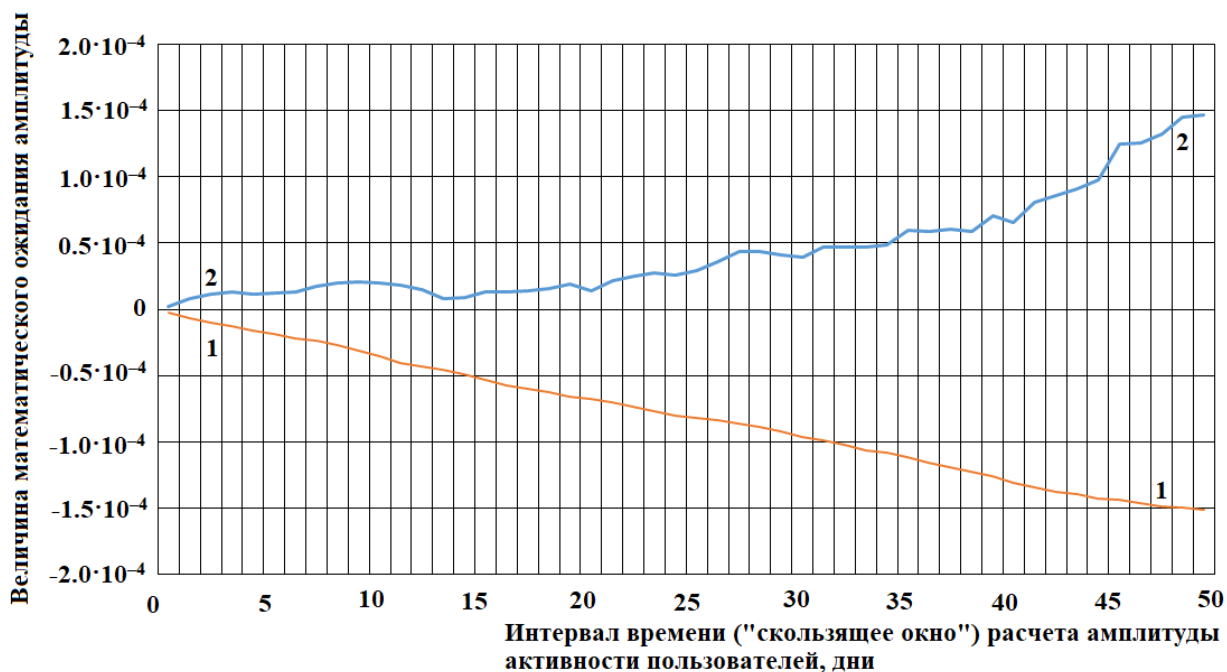


Рисунок 4 – Зависимость величины математического ожидания амплитуд изменения активности пользователей портала «РИА Новости» – линия 1 и «Эхо Москвы» – линия 2 при комментировании новостей от интервала времени расчета этих амплитуд («скользящего окна»)

Figure 4 – The dependence of mathematical expectation value of changing amplitudes in the activity of users of «RIA Novosti» portal – line 1 and «Echo of Moscow» – line 2 when commenting on the news from the time interval for calculating these amplitudes («sliding window»)

Рисунок 5 показывает, что зависимости величин дисперсии амплитуд изменения активности пользователей по комментированию новостей от интервала времени расчета этих амплитуд («скользящего окна») имеют сложный нелинейный характер. В данном случае дис-

персия $\sigma^2(t)$ была рассчитана, как среднее значение квадрата амплитуды. При выполнении нормального закона для дисперсии амплитуд должна была бы наблюдаться линейная зависимость от величины «скользящего окна»:

$$\sigma^2(t) = \int_{-\infty}^{+\infty} x^2 \rho(x, t) dx = \frac{1}{2\sqrt{\pi at}} \int_{-\infty}^{+\infty} x^2 e^{-\frac{x^2}{4at}} dx = 2at = at.$$

Отклонение от линейного закона в данном случае также говорит о том, что наблюдаемые временные ряды являются нестационарными.

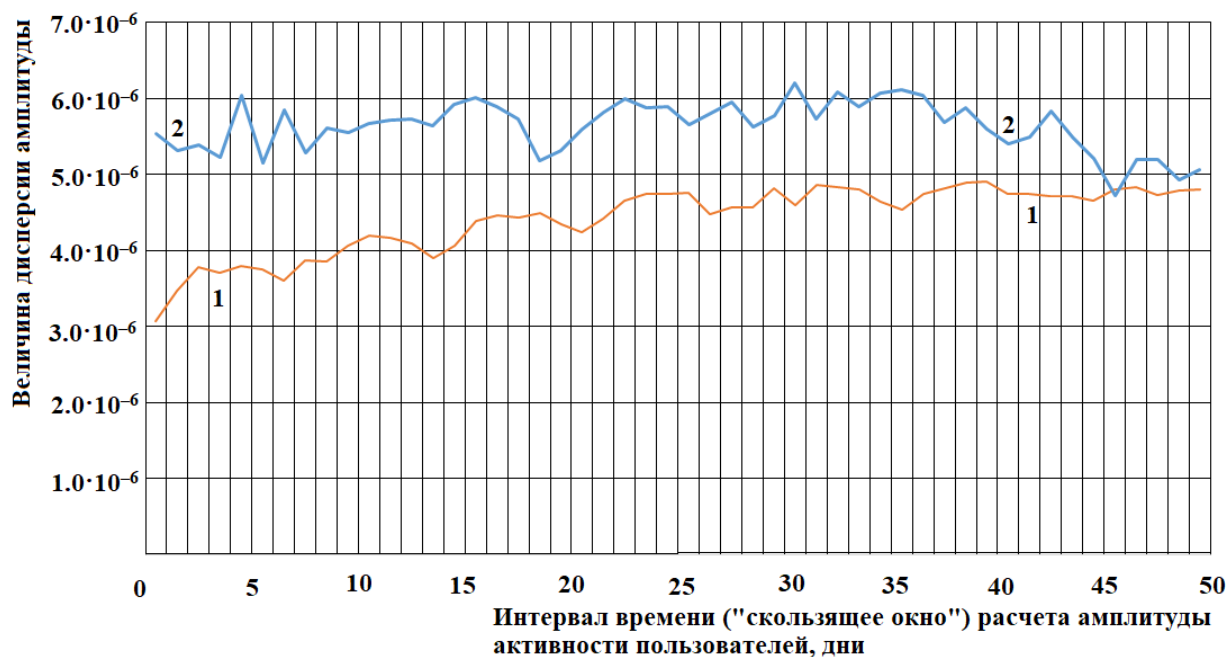


Рисунок 5 – Зависимость величины дисперсии амплитуд изменения активности пользователей портала «РИА Новости» – линия 1 и «Эхо Москвы» – линия 2 при комментировании новостей от интервала времени расчета этих амплитуд («скользящего окна»)

Figure 5 – The dependence of changing amplitudes variance magnitude in the activity of users of «RIA Novosti» – line 1 and «Echo of Moscow» – line 2 portal when commenting on the news from the time interval for calculating these amplitudes («sliding window»)

При расчете представленных на рисунке 6 зависимостей асимметрии ($As(t)$ – третий момент распределения) и представленных на рисунке 7 зависимостей эксцесса ($Ex(t)$ – четвертый момент) распределения величин амплитуд изменения активности пользователей от интервала времени расчета этих амплитуд («скользящего окна»:

$$As(t) = \frac{\sum_{j=1}^N \{x_j - (t)\}^3}{\sigma^3 \sum_{l=1}^M n_l} \text{ и } Ex(t) = \frac{\sum_{j=1}^N \{x_j - (t)\}^4}{\sigma^4 \sum_{l=1}^M n_l},$$

где $\sigma(t) = \sqrt{\frac{\sum_{j=1}^N \{x_j - \mu(t)\}^2}{\sum_{l=1}^M n_l}}$, $\mu(t) = \frac{\sum_{j=1}^N x_j}{\sum_{l=1}^M n_l}$, а $\sum_{l=1}^M n_l$ – вычисляется по числу значений амплитуд n_l .

При обсуждении данных, представленных на рисунках 6 и 7 трудно выделить какую-либо определенную зависимость асимметрии и эксцесса распределения величин амплитуд изменения активности пользователей от интервала времени расчета этих амплитуд («скользящего окна»). Поэтому можно обсудить вопрос об их средних значениях (для асимметрии «РИА Новости» среднее значение равно – 0,17; для асимметрии «Эхо Москвы» среднее зна-

чение равно $+0,046$, для эксцесса «РИА Новости» 3,14; для эксцесса «Эхо Москвы» 5,16) и независимости от интервала времени расчета.

Для этого необходимо исследовать поведение остатков, и проверить гипотезу о том, что они нормально распределены со средним значением равным нулю и имеют однородную дисперсию. Вычисление остатков можно провести на основе представленных на рисунках 6 и 7 данных, и полученных для них средних значений. Рассчитанная величина математического ожидания для распределения остатков для «РИА Новости» при исследовании асимметрии равна нулю; а дисперсия 0,01; асимметрия практически равна нулю; эксцесс 0,0002. Для эксцесса «РИА Новости» исследование остатков дает следующие значения: математическое ожидание равно нулю, дисперсия 0,15, асимметрия 0,06 и эксцесс 0,11. Рассчитанная величина математического ожидания для распределения остатков для «Эхо Москвы» при исследовании асимметрии равна нулю; а дисперсия 0,04; асимметрия практически равна нулю; эксцесс 0,005. Для эксцесса «Эхо Москвы» исследование остатков дает следующие значения: математическое ожидание равно нулю, дисперсия 0,81, асимметрия 0,017 и эксцесс 1,48. Таким образом, на основании полученных данных, можно принять гипотезу о почти независимости асимметрии и эксцесса распределения величин амплитуд изменения активности пользователей от интервала времени расчета этих амплитуд.

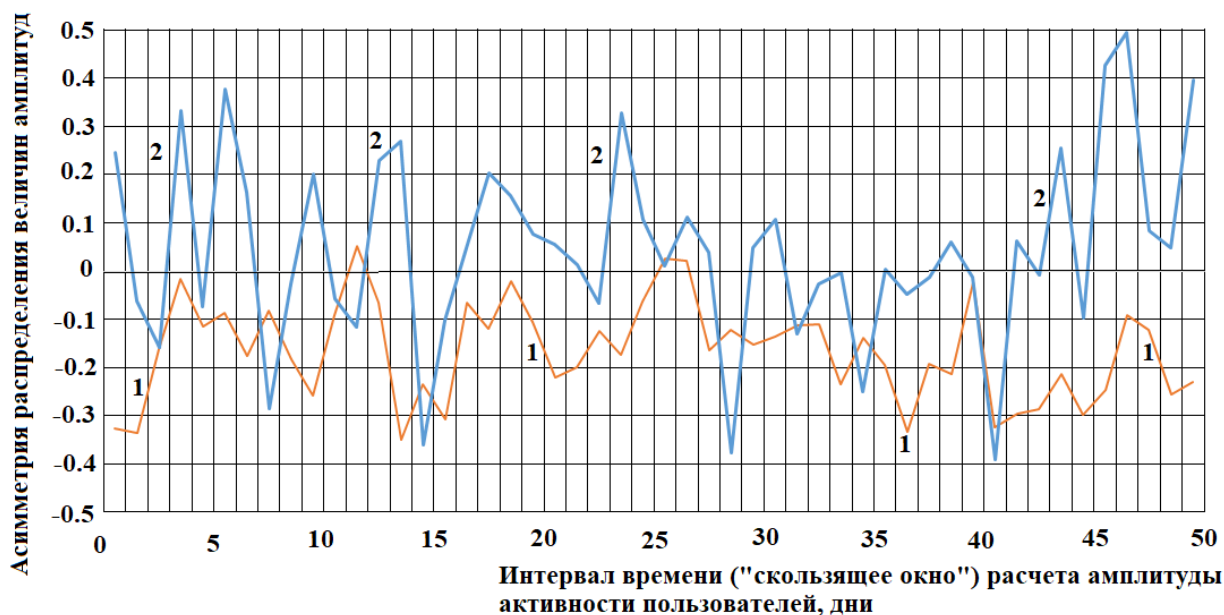


Рисунок 6 – Зависимость асимметрии распределения величин амплитуд изменения активности пользователей портала «РИА Новости» – линия 1 и «Эхо Москвы» – линия 2 при комментировании новостей от интервала времени расчета этих амплитуд («скользящего окна»)

Figure 6 – Dependence of distribution asymmetry of changing amplitudes in the activity of users of «RIA Novosti» – line 1 and «Echo of Moscow» portal – line 2 when commenting on the news from the time interval for calculating these amplitudes («sliding window»)

Асимметрия характеризует «скошенность» функции распределения, и для симметричных функций (например, распределение Пуассона) она равна 0. В данном случае (см. рисунок 6) наблюдается небольшая величина асимметрии и распределение амплитуд является почти симметричным.

Эксцесс характеризует «хвост» распределения. При больших положительных величинах эксцесса функция распределения медленнее убывает при удалении от среднего значения, чем при малых. Для нормального распределения эксцесс равен 3. При величине эксцесса больше трех график плотности распределения будет лежать выше графика нормального распределения, а меньше трех – ниже. В рассмотренных случаях наблюдается так называемый «тяже-

лый хвост» – графики плотности распределений лежат выше графика нормального распределения.



Рисунок 7 – Зависимость эксцесса распределения величин амплитуд изменения активности пользователей портала «РИА Новости» – линия 1 и «Эхо Москвы» – линия 2 при комментировании новостей от интервала времени расчета этих амплитуд («скользящего окна»)

Figure 7 – Dependence of distribution kurtosis of changing amplitudes in the activity of users of «RIA Novosti» portal – line 1 and «Echo of Moscow» – line 2 when commenting on the news from the time interval for calculating these amplitudes («sliding window»)

Также отметим, что при обработке нормированных временных рядов наблюдаются сходные с описанными зависимости.

Закключение и выводы

Можно сделать ряд общих выводов о динамике наблюдаемых процессов.

1. Временные ряды, описывающие рассмотренные процессы являются нестационарными (величины математического ожидания и дисперсии амплитуд изменения активности пользователей при комментировании новостей сложным образом зависят от интервала времени расчета этих амплитуд («скользящего окна»)).

2. Анализ наблюдаемых нестационарных временных рядов показывает, что описываемые ими процессы и системы обладают краткосрочной памятью (величина показателя Хёрста существенно меньше чем 0,5).

3. В распределении амплитуд изменения активности пользователей при комментировании новостей наблюдается небольшая величина асимметрии и распределение амплитуд является почти симметричным.

4. В распределении амплитуд изменения активности пользователей при комментировании новостей наблюдается так называемый «тяжелый хвост» – графики плотности распределений лежат выше графика нормального распределения и, следовательно, данные процессы не могут быть описаны этим законом.

5. При моделировании динамики рассматриваемых процессов и прогнозировании эволюции, описывающих их временных рядов необходим поиск иных подходов и моделей, которые бы в полной мере учитывали их описанные свойства.

Для создания модели, которая в полной мере будет учитывать обнаруженные характеристики наблюдаемых временных рядов необходимо рассматривать их свойства фрактальности. Существенное отличие определенных для них показателей Хёрста от 0,5 говорит о том, что описываемые ими процессы или системы не только обладают памятью, но и то, что их структура имеет фрактальный характер [7]. Фрактальность может быть связана с тем, что наблюдаемые процессы, характеризуются дробными переменными измерения. Соответственно можно предположить, что это должно быть учтено при выводе аппроксимирующих функций плотности распределения их параметров.

Нормальный закон распределения $\rho(x, t) = \frac{1}{2\sqrt{\pi at}} \cdot e^{-\frac{x^2}{4at}}$ может быть получен из рассмотрения изменения состояния x некоторой системы, при котором эти состояния могут случайным образом изменяться с течением времени на бесконечном множестве (например величины амплитуд отклонений в зависимости от выбранного интервала времени их расчета). Такие процессы принято относить к процессам обычного диффузионного типа и они описываются дифференциальным уравнением типа: $\frac{\partial \rho(x, t)}{\partial t} = D \frac{\partial^2 \rho(x, t)}{\partial x^2}$. В этом случае для дисперсии $\sigma^2(t)$ будет наблюдаться линейный закон её зависимости от времени ($\sigma^2(t) \sim t$). Однако в ряде случаев [8 – 12] наблюдается существенное отклонение от линейного закона. При этом может происходить как более «медленный» рост $\sigma^2(t)$ от t ($\sigma^2(t) \sim \sqrt[t]{t}$, где $0 < \alpha < 1$) и такой процесс квалифицируется как субдиффузия [8, 9], так и более «быстрый» рост $\sigma^2(t)$ от t ($\sigma^2(t) \sim t^\beta$, где $1 < \beta < 2$), и такой процесс квалифицируется как супердиффузия [10]. Функция распределения, описывающая процессы субдиффузионного типа получается из решения дробно – дифференциального уравнения вида: $\frac{\partial^\alpha \rho(x, t)}{\partial t^\alpha} = D \frac{\partial^2 \rho(x, t)}{\partial x^2}$, а супердиффузионного из уравнения: $\frac{\partial \rho(x, t)}{\partial t} = D \frac{\partial^\beta \rho(x, t)}{\partial x^\beta}$, где α и β показатели степеней дробных производных, а D – некоторый коэффициент (коэффициент диффузии). Также могут наблюдаться и процессы аномальной диффузии [11, 12] описываемые дробно – дифференциальным уравнением общего вида: $\frac{\partial^\alpha \rho(x, t)}{\partial t^\alpha} = D \frac{\partial^\beta \rho(x, t)}{\partial x^\beta}$. Следует особо отметить, что в данных случаях можно рассматривать фрактальность, как по переменной t (время), так и по переменной x , которая в диффузионных процессах рассматривается как пространственная координата, однако в общем случае, переменную x можно рассматривать, как некий параметр, описывающий состояние процесса или системы (например, в нашем случае, как величину амплитуды изменения активности пользователей при комментировании новостей). Следует отметить, что существует множество различных дробно – дифференциальных уравнений, все из которых относятся к классу стохастических уравнений, обладающих памятью, что позволяет описывать не Марковские процессы.

Представленные на рисунке 5 наблюдаемые зависимости величин дисперсии амплитуд изменения активности пользователей при комментировании новостей от интервала времени расчета этих амплитуд («скользящего окна») существенно отличаются от линейного закона. Кроме того, рассматриваемые временные ряды характеризуются сильным отличием показателя Хёрста от значения 0,5. Всё это говорит о фрактальности наблюдаемых процессов, а значит при выводе аппроксимирующих функций плотности распределения их параметров целесообразно использовать дробно – дифференциальные уравнения.

Таким образом, учитывая все вышеизложенное, можно сформулировать основные требования, которым должна отвечать модель аппроксимирующих функций плотности распреде-

ления параметров нестационарных временных рядов: она должна учитывать фрактальность наблюдаемых процессов и наличие памяти.

Благодарности

Исследование выполнено при поддержке Российского научного фонда (РНФ), грант № 22-21-00109 «Разработка моделей прогнозирования динамики социальных настроений на основе анализа временных рядов текстового контента социальных сетей с использованием уравнений Фоккера-Планка и нелинейной диффузии».

Библиографический список

1. **P. G Preethi, V. Uma, Ajit Kumar** Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction Procedia Computer Science Open Access, vol. 48, pp. 84-89, 2015 International Conference on Computer, Communication and Convergence, ICC3 2015 Bhubaneshwar, OD 27 December 2014 28 December 2014, 117138 ISSN 18770509 DOI 10.1016/j.procs.2015.04.154
2. **Zhuo-Ming Ren, An Zeng, Yi-Cheng Zhang** Structure-oriented prediction in complex networks Open Access, vol. 750, pp. 1-5, 13 August 2018 ISSN 03701573 DOI 10.1016/j.physrep.2018.05.002
3. **Zhukov, Dmitry Khvatova, Tatiana Aleshkin, Anton Schiavone, Francesco** Forecasting news events based on the model accounting for self-organisation and memory
4. **Dmitry Zhukov, Elena Andrianova and Olga Trifonova**, Stochastic Diffusion Model for Analysis of Dynamics and Forecasting Events in News Feeds
5. **Fuentes, Miguel A.** Non-Linear Diffusion and Power Law Properties of Heterogeneous Systems: Application to Financial Time Series Entropy Open Access, vol. 20, no. 930 August 2018 № 649 ISSN 10994300 DOI 10.3390/e20090649
6. **H. E. Hurst.** Long - term storage capacity of reservoirs. Transactions of American Society of Civil Engineers. 1951, vol. 116. 770 p.
7. **Mandelbrot B. B.** The Fractal Geometry of Nature. W. H. Freeman, San Francisco, 1982.
8. **Mainardi F.** Waves and Stability in Continuous Media. Ed. by Rionero S., Ruggeri T. Singapore: World Scientific, 1994.
9. **Wyss W. J.** The fractional diffusion equation. J. Math. Phys. 27, 2782 (1986); <https://doi.org/10.1063/1.527251Math>.
10. **Schneider W.R., Wyss W.J.** Fractional diffusion and wave equations J. Math. Phys. 30, 134 (1989); <https://doi.org/10.1063/1.528578Math>.
11. **Ilic, M., Liu, F., Turner, I., Anh, V.** Numerical approximation of a fractional-in-space diffusion equation, I. Fractional Calculus and Applied Analysis 8(3), pp. 323-341 (2005)
12. **Feldman R., Sanger J.** The Text Mining Handbook. Cambridge: Cambridge University Press, 2007. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008, 544 p.

UDC 517

ANALYSIS OF SOCIODYNAMIC PROCESSES TIME SERIES CHARACTERISTICS IN COMMENT NETWORKS OF MASS MEDIA USERS

K. K. Otradnov, lecturer, the Department of Information Warfare of the Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University, Moscow, Russia;
orcid.org/0000-0002-0055-9651 e-mail: const.otradnov@yandex.ru

J. P. Perova, lecturer, the Chair of Telecommunications, Institute of Radio Electronics and Computer Science, MIREA – Russian Technological University, Moscow, Russia;
orcid.org/0000-0003-4028-2842 e-mail: jul-np@yandex.ru

V. R. Grigoriev, Ph.D. (Tech.), Associate Professor, Head of the Department of Information Warfare, Institute of Cybersecurity and Digital Technologies MIREA – Russian Technological University, Moscow, Russia;
orcid.org/0000-0003-0279-6391 e-mail: grigorjev_vr@mail.ru

D. O. Zhukov, Dr. Sc. (Tech.), full professor, Professor of the Department of Information Warfare of the Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University Moscow, Russia;
 orcid.org/0000-0002-1211-5214 e-mail: zhukovdm@yandex.ru

The article analyzes the dynamics and characteristics of sociodynamic processes time series in comment networks of mass media users. There is a small amount of asymmetry in the distribution of amplitudes of changes in user activity when commenting on news and the distribution of amplitudes is almost symmetrical, but there is a so-called «heavy tail» – distribution density graphs lie above the normal distribution graph. The studies conducted allow us to draw a number of important conclusions that the time series studied are non-stationary, and the processes and systems described by them have short-term memory (the Hurst index is significantly less than 0.5). A significant difference between the Hurst indicators defined for them from 0.5 indicates that the processes or systems described by them not only have memory, but also that their structure has a fractal character. Fractality may be related to the fact that the observed processes are characterized by fractional measurement variables, which means that it is advisable to use fractional differential equations in the derivation of approximating density functions of their parameters distribution in order to build predictive models.

Key words: sociodynamic processes, activity of users of network mass media, time series, Hurst index, memory, self-similarity, fractality of time series.

DOI: 10.21667/1995-4565-2022-82-147-161

References

1. **P. G Preethi, V. Uma, Ajit Kumar** Temporal Sentiment Analysis and Causal Rules Extraction from Tweets for Event Prediction *Procedia Computer Science Open Access*, vol. 48, pp. 84-89, 2015 *International Conference on Computer, Communication and Convergence, ICC3 2015 Bhubaneshwar, OD 27 December 2014 28 December 2014*, 117138 ISSN 18770509 DOI 10.1016/j.procs.2015.04.154
2. **Zhuo-Ming Ren, An Zeng, Yi-Cheng Zhang** Structure-oriented prediction in complex networks *Open Access*, vol. 750, pp. 1-5, 13 August 2018 ISSN 03701573 DOI 10.1016/j.physrep.2018.05.002
3. **Zhukov, Dmitry Khvatova, Tatiana Aleshkin, Anton Schiavone, Francesco** Forecasting news events based on the model accounting for self-organisation and memory
4. **Dmitry Zhukov, Elena Andrianova and Olga Trifonova**, Stochastic Diffusion Model for Analysis of Dynamics and Forecasting Events in News Feeds
5. **Fuentes, Miguel A.** Non-Linear Diffusion and Power Law Properties of Heterogeneous Systems: Application to Financial Time Series *Entropy Open Access*, vol. 20, no. 930 August 2018 № 649 ISSN 10994300 DOI 10.3390/e20090649
6. **H. E. Hurst.** Long - term storage capacity of reservoirs. *Transactions of American Society of Civil Engineers*. 1951, vol. 116. 770 p.
7. **Mandelbrot B. B.** *The Fractal Geometry of Nature*. W. H. Freeman, San Francisco, 1982.
8. **Mainardi F.** *Waves and Stability in Continuous Media*. Ed. by Rionero S., Ruggeri T. Singapore: World Scientific, 1994.
9. **Wyss W.J.** The fractional diffusion equation. *J. Math. Phys.* 27, 2782 (1986); <https://doi.org/10.1063/1.527251Math>.
10. **Schneider W.R., Wyss W.J.** Fractional diffusion and wave equations *J. Math. Phys.* 30, 134 (1989); <https://doi.org/10.1063/1.528578Math>.
11. **Ilic, M., Liu, F., Turner, I., Anh, V.** Numerical approximation of a fractional-in-space diffusion equation, *I. Fractional Calculus and Applied Analysis* 8(3), pp. 323-341 (2005)
12. **Feldman R., Sanger J.** *The Text Mining Handbook*. Cambridge: Cambridge University Press, 2007. Christopher D. Manning, Prabhakar Raghavan, Hinrich Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008, 544 p.