

Сентимент-анализ комментариев пользователей сетевых масс-медиа и прогнозирование динамики их настроений на основе уравнения Фоккера – Планка

Ю.П. Перова, С.А. Лесько, В.Н. Калинин

Аннотация— Предложена вероятностная социодинамическая модель прогнозирования динамики настроений пользователей новостных интернет-ресурсов, основанная на решении нестационарного уравнения Фоккера – Планка. Динамика изменения состояния графа комментариев описывается как стохастический процесс с двумя компонентами: целенаправленной (коэффициент сноса) и случайной (коэффициент диффузии). Сформулирована и аналитически решена краевая задача с граничными условиями поглощения; получены выражения для плотности вероятности и вероятности достижения порогового состояния (порога перколяции). Рассмотрены два случая расположения начального состояния относительно порога перколяции. Показано, что при превышении порога для выхода из негативной области необходимо обеспечить преобладание целенаправленного воздействия над стохастической составляющей. Разработана методика определения параметров модели по наблюдаемым данным с использованием метода оптимизации Нелдера – Мида. Предложен алгоритм прогнозирования и управления достижением заданных пороговых состояний сети, включающий мониторинг, параметрическую идентификацию и оценку времени достижения критических состояний. В качестве источника данных для верификации использована социальная сеть «ВКонтакте»: выгружены публикации и комментарии за период 90 дней с привязкой ко времени, проведён сентимент-анализ текстов и построены временные ряды изменения долей узлов с различной сентимент-окраской и средних значений коэффициента посредничества. Точность прогнозирования долей сентимент-состояний составила 5–19% (средняя – от 5% до 13%), а средних значений коэффициента посредничества — 29–35%, что определяет границы применимости модели для различных характеристик графа.

Ключевые слова— сентимент-анализ, уравнение Фоккера – Планка, граф комментариев, порог перколяции, прогнозирование, социальные сети, коэффициент посредничества.

I. ВВЕДЕНИЕ

Анализ пользовательских комментариев в новостных

интернет-ресурсах представляет значительный интерес для понимания динамики общественных настроений. Комментарии пользователей образуют сложные сетевые структуры, количественные характеристики которых изменяются во времени. Для описания этих процессов может быть применён аппарат стохастической динамики.

Структура и динамика сложных сетей исследуются в работах [1, 2], где показано, что социальные сети обладают свойствами масштабной инвариантности. Ритмы взаимодействий пользователей в крупных онлайн-платформах изучены в работах [3, 4], а подходы к обнаружению сообществ исследованы в [5, 6]. Задачи классификации контента и мониторинга мнений в масс-медиа рассматриваются в работах [7–9]. Моделирование распространения информации и мнений в сетях на основе дифференциальных уравнений в частных производных и стохастических подходов исследовано в [10–13]. Уравнение Фоккера – Планка как инструмент анализа временных рядов и стохастических процессов применяется в [14–16]. Применение перколяционного подхода к социальным сетям обосновано в работах [17–19]. Стохастическое моделирование переходов между состояниями в социотехнических системах на основе уравнения Фоккера – Планка развито в [20, 21].

Ранее авторами было показано, что из стационарного решения уравнения Фоккера – Планка следует наблюдаемый степенной закон распределения числа комментариев [22], а также предложены подходы к анализу динамики комментариев на основе теории графов [23–25]. В настоящей работе развиваются эти результаты: предлагается модель на основе нестационарного уравнения Фоккера – Планка для прогнозирования динамики сентимент-состояний, разрабатывается алгоритм прогнозирования и оценки времени достижения критических состояний сети, а также методика определения параметров модели по наблюдаемым данным.

Ю.П. Перова, Институт радиоэлектроники и информатики, Российский технологический университет, Москва (e-mail: jul-pr@yandex.ru).

С.А. Лесько, Институт кибербезопасности цифровых технологий, МИРЭА – Российский технологический университет, Москва (e-mail: lesko@mirea.ru).

В.Н. Калинин, Институт радиоэлектроники и информатики, РТУ МИРЭА, Москва, Россия (e-mail: vkalininz@mail.ru)

II. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ЭВОЛЮЦИИ СОСТОЯНИЙ ГРАФА КОММЕНТАРИЕВ

A. Постановка краевой задачи

Для описания изменения состояния x сети комментариев с течением времени рассмотрим решение нестационарного уравнения Фоккера – Планка [14, 15]. Обозначим через $\rho(x, t)$ плотность вероятности обнаружения состояния x в момент времени t , а через L_{\min} и L_{\max} — границы области допустимых состояний.

Сформулируем граничные условия. Состояние $x = L_{\min}$ определяет границу, переход через которую необходимо предотвратить. Граничные условия поглощения на обеих границах формализуются как обращение в ноль плотности вероятности:

$$\rho(x, t)|_{x=L_{\min}} = 0, \tag{2.1}$$

$$\rho(x, t)|_{x=L_{\max}} = 0. \tag{2.2}$$

Начальное условие задаётся в виде δ -функции Дирака, отражающей начальное состояние системы x_0 :

$$\rho(x, t = 0) = \delta(x - x_0). \tag{2.3}$$

B. Решение краевой задачи

Решение сформулированной задачи [26, 20] записывается в виде двух выражений. При $L_{\min} \leq x \leq x_0$: где

$$\rho_1(x, t) = -\frac{2x_0^\alpha x^{-(1+\alpha)} e^{-\frac{D_0 \alpha^2}{2} t}}{\ln\left(\frac{L_{\max}}{L_{\min}}\right)} \sum_{n=1}^M \frac{1}{\cos(n\pi)} \times \tag{2.4}$$

$$\times \sin\left\{\pi n \frac{\ln(L_{\max}/x_0)}{\ln(L_{\max}/L_{\min})}\right\} \sin\left\{\pi n \frac{\ln(x/L_{\min})}{\ln(L_{\max}/L_{\min})}\right\} e^{-\frac{\pi^2 n^2 D_0 t}{2[\ln(L_{\max}/L_{\min})]^2}},$$

при $x_0 \leq x \leq L_{\max}$:

$$\rho_2(x, t) = \frac{2x_0^\alpha x^{-(1+\alpha)} e^{-\frac{D_0 \alpha^2}{2} t}}{\ln\left(\frac{L_{\max}}{L_{\min}}\right)} \sum_{n=1}^M \frac{1}{\cos(n\pi)} \times \tag{2.5}$$

$$\times \sin\left\{\pi n \frac{\ln(x_0/L_{\min})}{\ln(L_{\max}/L_{\min})}\right\} \sin\left\{\pi n \frac{\ln(x/L_{\max})}{\ln(L_{\max}/L_{\min})}\right\} e^{-\frac{\pi^2 n^2 D_0 t}{2[\ln(L_{\max}/L_{\min})]^2}},$$

$$\alpha = \frac{1}{2} - \frac{\mu_0}{D_0}, \tag{2.6}$$

μ_0 — коэффициент сноса, характеризующий направленное изменение состояния; D_0 — коэффициент диффузии, определяющий стохастическую составляющую; M — число членов ряда, обеспечивающее требуемую точность.

C. Вероятность достижения порогового состояния

Вероятность того, что к моменту t состояние системы не выйдет за пределы отрезка $[L_{\min}, L_{\max}]$:

$$P(\theta, t) = \int_{L_{\min}}^{x_0} \rho_2(x, t) dx + \int_{x_0}^{L_{\max}} \rho_1(x, t) dx. \tag{2.7}$$

Вероятность достижения или превышения порогового состояния θ (например, порога перколяции) к моменту t :

$$Q(\theta, t) = 1 - P(\theta, t). \tag{2.8}$$

В качестве L_{\max} используется порог перколяции сети комментариев [17, 18], определяемый по формуле

$$\ln P_c(x) = 4,02x - 2,2, \tag{2.9}$$

где x — обратная величина средней плотности сети.

D. Имитационное моделирование

Для верификации непротиворечивости модели

проведено имитационное моделирование при различных наборах параметров. Таблица 1 содержит значения μ_0 и D_0 , полученные по наблюдаемым данным для нескольких новостей [23].

ТАБЛИЦА 1. ЗНАЧЕНИЯ ПАРАМЕТРОВ МОДЕЛИ ДЛЯ ВЫБРАННЫХ НОВОСТЕЙ

ID новости	μ_0	D_0	Ошибка
2613301	0,20	0,35	8,45%
2660775	0,24	0,43	9,88%
2626290	0,22 / 0,41	0,36 / 0,60	9,29%
2703843	0,73 / 0,41	0,85 / 0,35	9,90%

На рис. 1 представлены зависимости вероятности достижения порога перколяции $\theta = 0,30$ от времени при различных начальных долях негативных узлов x_0 (кривые 1 – 3 соответствуют $x_0 = 0,05; 0,10; 0,15$) для двух наборов параметров.

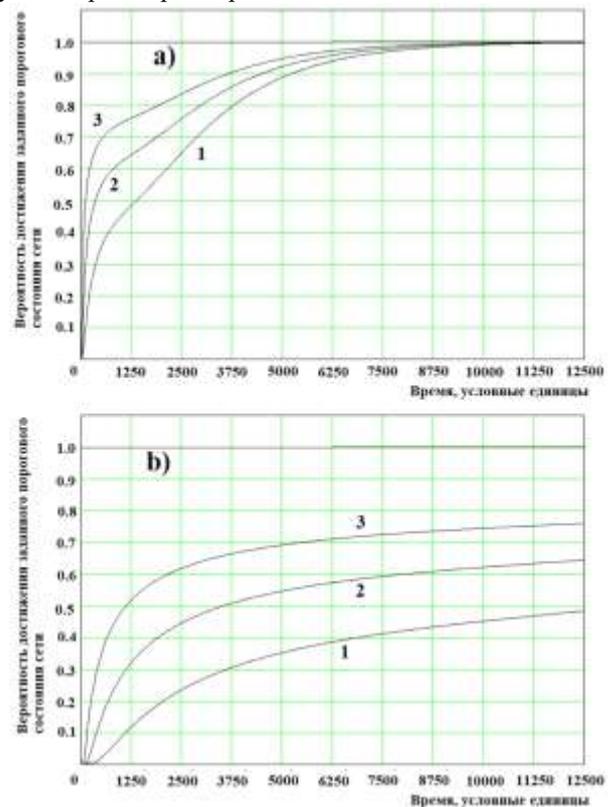


Рис. 1. Зависимость вероятности достижения порогового состояния от времени при $x_0 < \theta$: а) $D_0 = 0,015, \mu_0 = 0,005$; б) $D_0 = 0,0015, \mu_0 = 0,0005$

Расчёты показывают, что вероятность достижения порогового состояния монотонно возрастает от 0 до 1 при любых наборах параметров, причём рост происходит тем быстрее, чем больше начальная доля x_0 негативных узлов и чем больше значения D_0 и μ_0 .

При $x_0 > \theta$ и $\mu_0 < D_0$ вероятность достижения порога перколяции равна нулю. Для выхода из негативной области необходимо обеспечить условие $\mu_0 > D_0$ [21], то есть целенаправленное воздействие должно преобладать над стохастической составляющей. Рассмотрим этот случай подробнее. Когда начальная доля негативных узлов x_0 превышает порог перколяции θ , вероятность того, что к моменту времени t состояние системы будет находиться на отрезке от θ до 1 (т.е. пороговое

состояние не будет достигнуто снизу), вычисляется по формуле:

$$P(\theta, t) = \int_0^{x_0} \rho_2(x, t) dx + \int_{x_0}^1 \rho_1(x, t) dx, \quad (2.10)$$

где $\rho_2(x, t)$ и $\rho_1(x, t)$ определяются выражениями (2.4) – (2.5) при замене L_{\min} на θ и L_{\max} на 1. Условие $\mu_0 > D_0$ обеспечивает положительный снос в направлении уменьшения доли негативных узлов, что физически означает целенаправленное управляющее воздействие на информационное пространство.

Для иллюстрации рассмотрен случай $\mu_0 = 0,01$, $D_0 = 0,0075$ ($\mu_0 > D_0$); порог перколяции принят равным 0,30. На рис. 2 представлены результаты имитационного моделирования: зависимость вероятности достижения порогового состояния сети от времени для трёх начальных долей негативных узлов (кривая 1 — $x_0 = 0,40$, кривая 2 — $x_0 = 0,50$, кривая 3 — $x_0 = 0,80$).

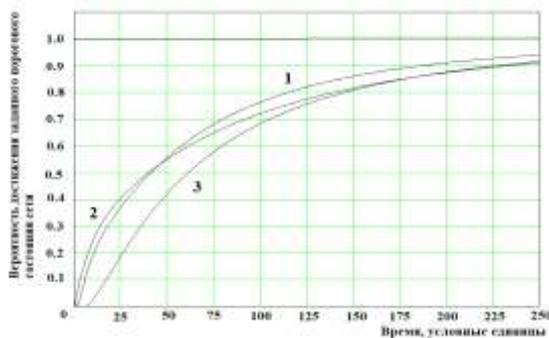


Рис. 2. Зависимость вероятности достижения порогового состояния от времени для случая $x_0 > \theta$ при $\mu_0 = 0,01$, $D_0 = 0,0075$

Расчёты показывают, что чем дальше находится начальная доля x_0 от порога перколяции, тем медленнее растёт вероятность его достижения с течением времени, что согласуется с физической интерпретацией модели. Данные результаты подтверждают непротиворечивость разработанной модели для обоих рассмотренных случаев расположения начального состояния относительно порога перколяции.

III. МЕТОДИКА ОПРЕДЕЛЕНИЯ ПАРАМЕТРОВ МОДЕЛИ

A. Определение μ_0 и D_0

Параметры модели определяются по наблюдаемым данным следующим образом. Из социальной сети собираются публикации и комментарии с привязкой ко времени. Для каждого поста фиксируются: идентификатор сообщества, идентификатор поста, временная метка и метаданные. Для комментариев — идентификатор, временная метка, текст и связь с родительским объектом.

Из социальной сети «ВКонтакте» были выгружены публикации сообществ и комментарии к ним. Для каждого комментария фиксировались: идентификатор поста ($post_id$), идентификатор комментария ($comment_id$), временная метка ($timestamp$), текст, число ответов ($thread_count$) и связь с родительским объектом:

$parent_id = 0$ для комментариев верхнего уровня, иначе — идентификатор родительского комментария.

На рис. 3 представлен пример динамики дневного количества комментариев за июнь 2025 г., демонстрирующей характерную волнообразную структуру с периодом, близким к недельному циклу активности пользователей.

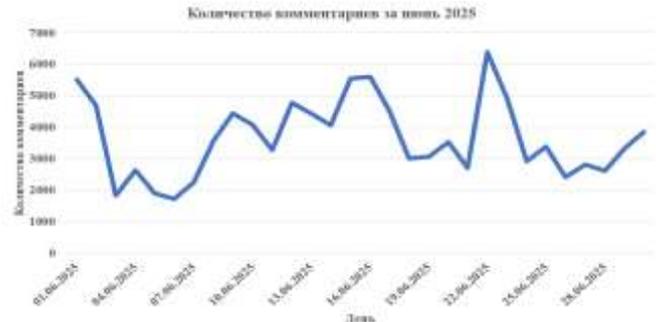


Рис. 3. Дневное количество комментариев за июнь 2025 г.

На основе собранных данных строится наблюдаемая динамика изменения числа комментариев за выбранный период. Затем методом Нелдера – Мида [23, 22] осуществляется подбор параметров μ_0 и D_0 , минимизирующих расхождение между теоретической кривой (2.4) – (2.8) и наблюдаемыми данными. На рис. 4 показан результат аппроксимации для двух временных интервалов (1 и 2 суток); относительная ошибка составляла 10 – 20%.

Диапазон полученных значений параметров μ_0 и D_0 составляет от сотых до десятых долей, при этом характер аппроксимации существенно зависит от выбранного временного интервала. Для интервала в одни сутки (рис. 4а) наблюдается более выраженная S-образная кривая роста числа комментариев, тогда как для двухсуточного интервала (рис. 4б) динамика приобретает более плавный характер, что может быть связано с усреднением суточных колебаний активности пользователей.

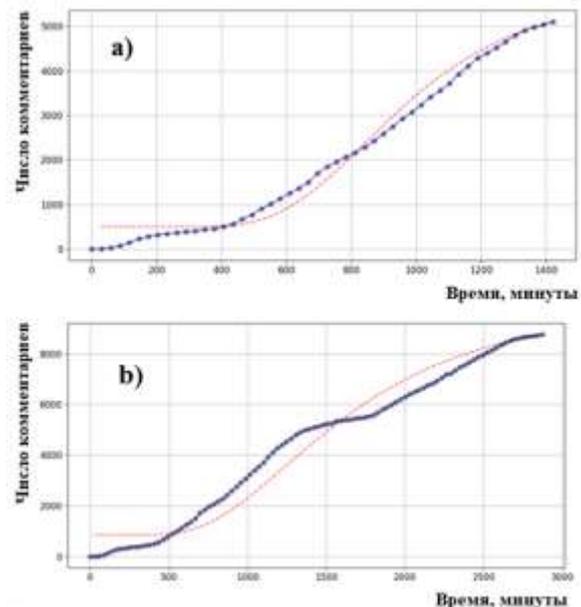


Рис. 4. Аппроксимация наблюдаемых данных по динамике числа комментариев теоретической моделью: а) интервал 1 день; б) интервал 2 дня. Точки — наблюдения, штриховые линии — расчёт

В. Определение порога перколяции

Все комментарии объединяются (через сделавших их пользователей) в единый граф, для которого рассчитывается средняя плотность. По формуле (2.9) определяется порог перколяции θ [19], который интерпретируется как предельное значение коэффициента посредничества: при его достижении информация от узлов с данной сентимент-окраской может беспрепятственно распространяться по всей сети.

Порог перколяции определяет критическую долю узлов определённого типа (например, негативно окрашенных), при достижении которой эти узлы образуют связный кластер, охватывающий всю сеть[17]. Применительно к сети комментариев это означает, что информация от пользователей с данной сентимент-окраской может беспрепятственно распространяться по всей структуре без существенных потерь. Величину порога перколяции можно использовать как предельное значение коэффициента посредничества (его значение лежит в диапазоне от 0 до 1). Если средняя величина коэффициента посредничества для узлов сети с определённой сентимент-окраской достигает значения порога перколяции, то информация от этих узлов может распространяться по сети практически без препятствий.

В качестве x_0 может использоваться как начальная доля узлов, находящихся в негативном состоянии, так и отношение средних значений коэффициентов посредничества или коэффициентов кластеризации (попарное отношение негативных к позитивным) в данный момент времени.

IV. АЛГОРИТМ ПРОГНОЗИРОВАНИЯ

На основе разработанной модели предложен алгоритм прогнозирования динамики настроений и оценки времени достижения критических состояний сети.

Алгоритм 1 Прогнозирование достижения заданного состояния графа комментариев

Вход: ряд комментариев $\{c_i\}_{i=1}^T$, надёжность Q^*

Выход: время достижения порога t^*

- 1: Собрать комментарии и метаданные с привязкой к дате и времени
- 2: Провести сентимент-анализ; построить граф $G = (V, E)$
- 3: Рассчитать плотность d , посредничество BC , долю негативных узлов x_0
- 4: Определить порог перколяции θ : $\ln P_c = 4,02/d - 2,2$
- 5: Построить динамику числа комментариев за период Δt
- 6: Методом Нелдера – Мида определить μ_0 и D_0
- 7: если $x_0 \geq \theta$ то
- 8: Управляющее воздействие: обеспечить $\mu_0 > D_0$
- 9: **конец если**
- 10: По (4) – (8) при $L_{\max} = \theta$ рассчитать $Q(\theta, t)$
- 11: Найти t^* : $Q(\theta, t^*) = Q^*$
- 12: вернуть t^* , $Q(\theta, t)$ для горизонта прогноза

В качестве управляющих воздействий при достижении или превышении порога принятия решений могут применяться: модерация трафика, создание искусственной задержки негативных комментариев, а также привлечение дополнительных пользователей с позитивным контентом.

Время достижения заданных пороговых состояний от текущего состояния x_0 можно оценить, приняв в формулах (2.4) – (2.10) в качестве L_{\max} значение порога перколяции или величину заданного порога принятия решения, а в качестве значимой вероятности

достижения — числовое значение, равное, например, 0,90 или 0,95.

Среди возможных управляющих инструментов можно выделить два основных. Первый — это искусственное замедление трафика, исходящего от пользователей, дающих негативные комментарии, что приводит к уменьшению величины μ_0 для негативной составляющей (например, негативный комментарий появляется в сети с некоторой задержкой по сравнению с положительным). Второй — появление дополнительных пользователей, дающих позитивные комментарии и таким образом средствами информационного воздействия влияющих на соотношение сентимент-состояний в сети. Выбор конкретного инструмента определяется соотношением текущих значений μ_0 и D_0 , а также расстоянием от текущего состояния x_0 до порогового.

V. ВЕРИФИКАЦИЯ МОДЕЛИ И ОЦЕНКА ТОЧНОСТИ ПРОГНОЗИРОВАНИЯ

А. Исходные данные

Из социальной сети «ВКонтакте» были выгружены публикации и комментарии за период 90 дней (июнь – август 2025 г.). Для каждого дня были построены: временные ряды доли пользователей с различной сентимент-окраской (рис. 5а) и временные ряды средних значений коэффициента посредничества (рис. 5б). На обоих графиках кривые 1, 2, 3 соответствуют узлам с позитивной, негативной и нейтральной окраской.

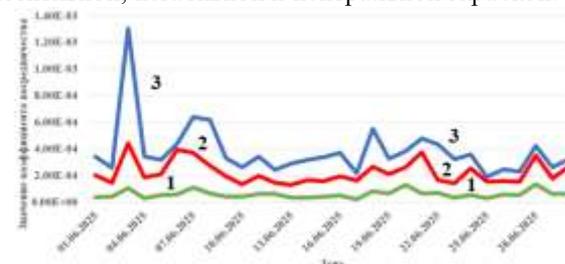


Рис. 5а) Доли сентимент-окраски

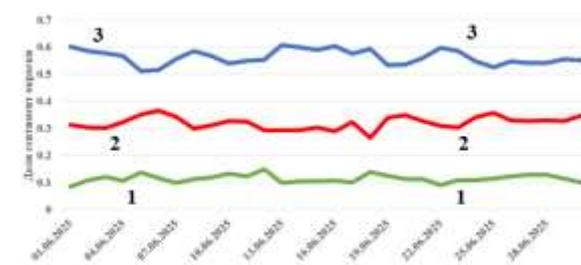


Рис. 5б) Коэффициент посредничества

Рис. 5. Временные ряды характеристик комментариев к новостям за июнь 2025 г. Кривые: 1 — позитивная, 2 — негативная, 3 — нейтральная окраска

В. Методика оценки точности

Для оценки точности прогнозирования применён метод скользящего окна. На каждом шаге из временного ряда длины T выделяется интервал размера W , по которому методом Нелдера – Мида определяются μ_0 и D_0 . Затем для глубины прогноза τ вычисляется прогнозируемое значение и сравнивается с наблюдаемым. Относительная ошибка усредняется по всем положениям скользящего окна. Процедура

повторяется при различных W и τ .

Подробнее, процедура оценки точности состоит из следующих шагов.

1. Из временного ряда длины T выделяется скользящее окно размера Δt_n ($n = 1, 2, 3, \dots$). По данным этого окна строится наблюдаемая динамика числа комментариев и методом Нелдера – Мида определяются параметры μ_0 и D_0 .
2. После определения μ_0 и D_0 выполняется прогноз на глубину τ_i ($i = 1, 2, 3, \dots$). Вероятность $Q(\theta, t)$ того, что заданное sentiment-состояние L будет достигнуто к моменту t , вычисляется по формуле (2.8). В качестве x_0 используется последнее значение анализируемой характеристики на окне Δt_n . Принимая надёжность $Q(t) = 0,95$, рассчитывается прогнозируемое значение X для каждой глубины прогноза τ_i .
3. Сравнивая прогнозируемые значения X с наблюдаемыми уровнями временного ряда в момент t_i , находим относительную ошибку (по модулю) в процентах: разность прогнозируемого и наблюдаемого значений делится на величину наблюдаемого.
4. Скользящее окно Δt_n сдвигается по временному ряду на один шаг вправо, и шаги 1 – 3 повторяются до последнего допустимого положения ($T - \Delta t_n$).
5. Относительная ошибка усредняется для каждой глубины прогноза τ_i по всем положениям скользящего окна Δt_n .
6. Размер скользящего окна увеличивается на единицу ($n \rightarrow n + 1$), и шаги 1 – 5 повторяются. Таким образом, для каждой пары (W, τ) получается усреднённая оценка относительной ошибки прогноза, формирующая трёхмерную поверхность ошибок.

С. Результаты

На рис. 6 представлена трёхмерная поверхность зависимости средней относительной ошибки прогноза от размера скользящего окна и глубины прогноза для позитивной sentiment-окраски. Аналогичные поверхности для негативной и нейтральной окрасок обладают качественно сходным характером.

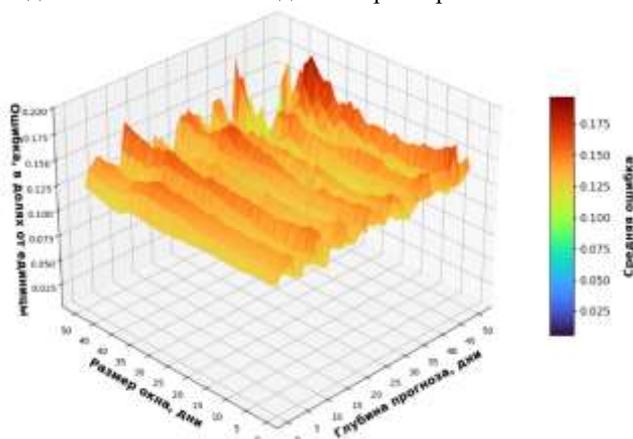


Рис. 6. Зависимость относительной ошибки прогноза от размера скользящего окна и глубины прогноза для позитивной sentiment-окраски

На рис. 7 представлены сечения указанной поверхности: а) при фиксированной глубине прогноза τ ; б) при фиксированном размере скользящего окна W .

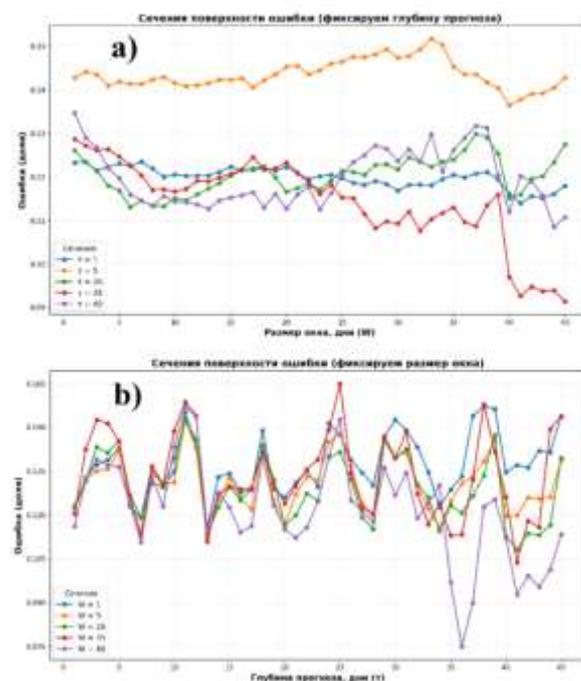


Рис. 7. Сечения поверхности ошибки для позитивной sentiment-окраски: а) зависимость от W при различных τ ; б) зависимость от τ при различных W

Анализ результатов позволяет сделать следующие выводы.

Зависимость ошибки от глубины прогноза при различных размерах скользящего окна имеет осциллирующий характер и слабо зависит от W . С увеличением τ ошибка в среднем возрастает, что является ожидаемым поведением.

Итоговые оценки точности прогнозирования долей sentiment-состояний приведены в табл. 2.

Таблица 2. Точность прогнозирования долей sentiment-состояний

Сентимент-окраска	Диапазон ошибки	Средняя ошибка
Позитивная	5 – 19%	13%
Негативная	3 – 16%	8%
Нейтральная	4 – 11%	5%

Аналогичный анализ проведён для временных рядов средних значений коэффициента посредничества узлов с различной sentiment-окраской. Качественный характер зависимости ошибки от параметров скользящего окна и глубины прогноза аналогичен случаю долей sentiment-состояний, однако абсолютные значения ошибок существенно выше — от 29% до 35%, что объясняется большей нестационарностью данных метрик.

Итоговые оценки точности прогнозирования средних значений коэффициента посредничества приведены в табл. 3.

Таблица 3. Точность прогнозирования средних значений коэффициента посредничества

Сентимент-окраска	Диапазон ошибки	Средняя ошибка
Позитивная	25 – 40%	32%
Негативная	22 – 38%	29%
Нейтральная	28 – 45%	35%

Сравнение табл. 2 и 3 показывает, что точность прогнозирования долей sentiment-состояний существенно выше, чем для коэффициентов посредничества. Это объясняется тем, что доли sentiment-состояний являются интегральными характеристиками графа и обладают большей устойчивостью к локальным флуктуациям, тогда как коэффициент посредничества чувствителен к топологическим изменениям в структуре сети, которые могут происходить при добавлении или удалении даже небольшого числа узлов.

Следует также отметить, что поведение временных рядов для средних значений коэффициентов посредничества узлов с различной sentiment-окраской (см. рис. 5б) имеет характер, схожий с поведением временных рядов для изменения долей узлов с различной sentiment-окраской (см. рис. 5а), что свидетельствует о наличии корреляции между этими показателями. Нейтральная окраска доминирует по доле узлов, однако её коэффициент посредничества наиболее волатилен, что указывает на менее устойчивую структурную роль нейтрально окрашенных узлов в сети.

В целом, разработанная модель демонстрирует достаточно высокую точность прогнозирования для практических задач мониторинга общественных настроений: средняя ошибка прогноза долей sentiment-состояний не превышает 13%, что сопоставимо с погрешностью традиционных социологических опросов. Для коэффициентов посредничества точность ниже (средняя ошибка 29 – 35%), что определяет границы применимости модели для структурных характеристик графа.

VI. ЗАКЛЮЧЕНИЕ

В работе предложена вероятностная модель прогнозирования динамики настроений пользователей сетевых масс-медиа, основанная на решении нестационарного уравнения Фоккера – Планка. Основные результаты:

1. Сформулирована и решена краевая задача для уравнения Фоккера – Планка, описывающая эволюцию состояний графа комментариев; получены аналитические выражения для плотности вероятности и вероятности достижения порогового состояния.
2. Разработана методика определения параметров модели (μ_0 , D_0) по наблюдаемым данным с использованием метода Нелдера – Мида, а также способ определения порога перколяции по плотности сети.
3. Предложен алгоритм прогнозирования и управления достижением заданного состояния графа сети комментариев, включающий мониторинг, параметрическую идентификацию и оценку времени достижения критических состояний.
4. Проведена верификация модели на данных социальной сети «ВКонтакте» за 90 дней. Точность прогнозирования долей sentiment-состояний составила 5 – 19% (средние — от 5% до 13%), а

средних значений коэффициента посредничества — 29 – 35%.

Полученные результаты свидетельствуют о непротиворечивости и применимости разработанной модели для задач мониторинга и прогнозирования динамики общественных настроений в сетевых масс-медиа.

БИБЛИОГРАФИЯ

- [1] Dorogovtsev S. N., Mendes J. F. F. Evolution of networks // *Advances in Physics*. — 2002. — Vol. 51. — P. 1079–1187.
- [2] Newman M. E. J. The structure and function of complex networks // *SIAM Review*. — 2003. — Vol. 45. — P. 167–256.
- [3] Golder S. A., Wilkinson D. M., Huberman B. A. Rhythms of social interaction: messaging within a massive online network // *Communities and Technologies*. — 2007. — P. 41–66.
- [4] Sekara V., Stopczynski A., Lehmann S. Fundamental structures of dynamic social networks // *Proceedings of the National Academy of Sciences*. — 2016. — DOI: 10.1073/pnas.1602803113.
- [5] Airoldi E. M., Blei D. M., Fienberg S. E., Xing E. P. Mixed membership stochastic blockmodels // *Journal of Machine Learning Research*. — 2008. — Vol. 9. — P. 1981–2014.
- [6] Tran C., Shin W.-Y., Spitz A. Community detection in partially observable social networks // *ACM Transactions on Knowledge Discovery from Data*, 16(2), Article 22 (2022).. — DOI: 10.1145/3461339.
- [7] Mukhamediev R. I., Yakunin K., Mussabayev R. et al. Classification of negative information on socially significant topics in mass media // *Symmetry*. — 2020. — Vol. 12. — P. 1945. — DOI: 10.3390/sym12121945.
- [8] Willaert T., Van Eecke P., Beuls K., Steels L. Building social media observatories for monitoring online opinion dynamics // *Social Media + Society*. — 2020. — Vol. 6.
- [9] Bandari R., Asur S., Huberman B. A. The pulse of news in social media: forecasting popularity // *arXiv preprint arXiv:1202.0332*. — 2012.
- [10] Du B., Lian X., Cheng X. Partial differential equation modeling with Dirichlet boundary conditions on social networks // *Boundary Value Problems*. — 2018. — Vol. 2018, No. 50. — DOI: 10.1186/s13661-018-0964-4.
- [11] Liu X., He D., Liu C. Modeling information dissemination and evolution in time-varying online social network based on thermal diffusion motion // *Physica A*. — 2018. — Vol. 510. — P. 456–476.
- [12] Bomba A., Kunanets N., Pasichnyk V., Turbal Y. Mathematical and computer models of message distribution in social networks based on the space modification of Fermi–Pasta–Ulam approach // *Advances in Intelligent Systems and Computing*. — 2019. — Vol. 836. — P. 257–266.
- [13] Sahafzadeh E., Tork Ladani B. The impact of group propagation on rumor spreading in mobile social networks // *Physica A*. — 2018. — Vol. 506. — P. 412–423.
- [14] Lux T. Inference for systems of stochastic differential equations from discretely sampled data: a numerical maximum likelihood approach // *Annals of Finance*. — 2013. — Vol. 9, No. 2. — P. 217–248. — DOI: 10.1007/s10436-012-0219-9.
- [15] Hurn A., Jeisman J., Lindsay K. Teaching an old dog new tricks: improved estimation of the parameters of stochastic differential equations by numerical solution of the Fokker–Planck equation // *Financial Econometrics Handbook* / eds. Gregoriou G., Paschalau R. — London: Palgrave, 2010.
- [16] Орлов Ю. Н., Федоров С. Л. Генерация нестационарных траекторий временного ряда на основе уравнения Фоккера – Планка // *Труды МФТИ*. — 2016. — Т. 8, 2. — С. 126–133.
- [17] Grimmett G. *Percolation*. — Berlin: Springer-Verlag, 1999. — 2nd ed.
- [18] Stauffer D., Aharony A. *Introduction to Percolation Theory*. — London: Taylor & Francis, 1992.
- [19] Zhukov D., Khvatova T., Lesko S., Zaltsman A. Managing social networks: applying the percolation theory methodology to understand individuals' attitudes and moods // *Technological Forecasting and Social Change*, 129 (2018), 297–307 — DOI: 10.1016/j.techfore.2017.09.039.
- [20] Sigov A. S., Zhukov D. O., Khvatova T. Yu., Andrianova E. G. A model of forecasting of information events on the basis of the solution of a boundary value problem for systems with memory and self-organization // *Journal of Communications Technology and*

- Electronics. — 63, 1478–1485 (2018).— DOI: 10.1134/S1064226918120227.
- [21] Zhukov D., Khvatova T., Millar C., Zaltzman A. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organization and memory // *Technological Forecasting and Social Change*. — 2020. — Vol. 158. — P. 120134. — DOI: 10.1016/j.techfore.2020.120134.
- [22] Zhukov D., Perova J., Kalinin V. Description of the distribution law and non-linear dynamics of growth of comments number in news and blogs based on the Fokker–Planck equation // *Mathematics*. — 2022. — Vol. 10. — P. 989. — DOI: 10.3390/math10060989.
- [23] Перова Ю. П., Лесько С. А., Жуков Д. О., Чечурин А. В. Анализ и моделирование процессов в сложных социальных сетевых структурах на основе уравнения Фоккера–Планка // *Вестник Томского государственного университета. Управление, вычислительная техника и информатика*. — 2022. — 60. — С. 32–41. — DOI: 10.17223/19988605/60/4.
- [24] Перова Ю. П., Григорьев В. Р., Жуков Д. О. Модели и методы анализа сложных сетей и социальных сетевых структур // *Russian Technological Journal*. — 2023. — Т. 11, 2. — С. 33–49. — DOI: 10.32362/2500-316X-2023-11-2-33-49.
- [25] Отрадных К. К., Перова Ю. П., Григорьев В. Р., Жуков Д. О. Анализ характеристик временных рядов социодинамических процессов в сетях комментариев пользователей масс-медиа // *Вестник РГРТУ*. — 2022. — 82. — С. 147–161. — DOI: 10.21667/1995-4565-2022-82-147-161.
- [26] Perova J. P., Kalinin V. N., Lesko S. A. Modeling the dynamics of changes in the number of comments of mass media users based on the Fokker–Planck equation and parameters of networks of their connections // *International Journal of Open Information Technologies*. — 2023. — Vol. 11, No. 4.

Об авторах:

Перова Юлия Петровна, старший преподаватель, институт радиоэлектроники и информатики, МИРЭА - Российский технологический университет

Лесько Сергей Александрович, доктор технических наук, доцент, институт кибербезопасности цифровых технологий, МИРЭА - Российский технологический университет

Калинин Владимир Николаевич, лаборант, институт радиоэлектроники и информатики, МИРЭА - Российский технологический университет

Sentiment Analysis of User Comments in Online Mass Media and Forecasting the Dynamics of Their Sentiment Based on the Fokker–Planck Equation

J.P. Perova, S.A. Lesko, V.N. Kalinin,

Abstract. A probabilistic sociodynamic model for forecasting user sentiment dynamics in online news media is proposed, based on the solution of the non-stationary Fokker–Planck equation. The dynamics of the comment graph state is described as a stochastic process comprising two components: a directed one (drift coefficient) and a random one (diffusion coefficient). A boundary value problem with absorbing boundary conditions describing the evolution of the comment graph state is formulated and solved analytically; expressions for the probability density and the probability of reaching a threshold state (percolation threshold) are obtained. Two cases of the initial state position relative to the percolation threshold are considered. It is shown that when the threshold is exceeded, a purposeful control action is required to ensure the predominance of directed influence over the stochastic component. A methodology for estimating model parameters from observed data using the Nelder–Mead optimization method is developed. An algorithm for forecasting and controlling the attainment of given threshold states of the network is presented, including monitoring, parametric identification, and estimation of the time to reach critical states. The model is verified using data from the VKontakte social network collected over 90 days. Forecasting accuracy for sentiment proportions is 5–19% (average 5% to 13%), and for average betweenness centrality values — 29–35%, which defines the applicability limits of the model for different graph characteristics.

Keywords — sentiment analysis, Fokker–Planck equation, comment network, percolation threshold, forecasting, social networks, betweenness centrality.

БИБЛИОГРАФИЯ

- [1] Dorogovtsev S. N., Mendes J. F. F. Evolution of networks // *Advances in Physics*. — 2002. — Vol. 51. — P. 1079–1187.
- [2] Newman M. E. J. The structure and function of complex networks // *SIAM Review*. — 2003. — Vol. 45. — P. 167–256.
- [3] Golder S. A., Wilkinson D. M., Huberman B. A. Rhythms of social interaction: messaging within a massive online network // *Communities and Technologies*. — 2007. — P. 41–66.
- [4] Sekara V., Stopczynski A., Lehmann S. Fundamental structures of dynamic social networks // *Proceedings of the National Academy of Sciences*. — 2016. — DOI: 10.1073/pnas.1602803113.
- [5] Airoldi E. M., Blei D. M., Fienberg S. E., Xing E. P. Mixed membership stochastic blockmodels // *Journal of Machine Learning Research*. — 2008. — Vol. 9. — P. 1981–2014.
- [6] Tran C., Shin W.-Y., Spitz A. Community detection in partially observable social networks // *ACM Transactions on Knowledge Discovery from Data*, 16(2), Article 22 (2022).. — DOI: 10.1145/3461339.
- [7] Mukhamediev R. I., Yakunin K., Mussabayev R. et al. Classification of negative information on socially significant topics in mass media // *Symmetry*. — 2020. — Vol. 12. — P. 1945. — DOI: 10.3390/sym12121945.
- [8] Willaert T., Van Eecke P., Beuls K., Steels L. Building social media observatories for monitoring online opinion dynamics // *Social Media + Society*. — 2020. — Vol. 6.
- [9] Bandari R., Asur S., Huberman B. A. The pulse of news in social media: forecasting popularity // *arXiv preprint arXiv:1202.0332*. — 2012.
- [10] Du B., Lian X., Cheng X. Partial differential equation modeling with Dirichlet boundary conditions on social networks // *Boundary Value Problems*. — 2018. — Vol. 2018, No. 50. — DOI: 10.1186/s13661-018-0964-4.
- [11] Liu X., He D., Liu C. Modeling information dissemination and evolution in time-varying online social network based on thermal diffusion motion // *Physica A*. — 2018. — Vol. 510. — P. 456–476.
- [12] Bomba A., Kunanets N., Pasichnyk V., Turbal Y. Mathematical and computer models of message distribution in social networks based on the space modification of Fermi–Pasta–Ulam approach // *Advances in Intelligent Systems and Computing*. — 2019. — Vol. 836. — P. 257–266.
- [13] Sahafizadeh E., Tork Ladani B. The impact of group propagation on rumor spreading in mobile social networks // *Physica A*. — 2018. — Vol. 506. — P. 412–423.
- [14] Lux T. Inference for systems of stochastic differential equations from discretely sampled data: a numerical maximum likelihood approach // *Annals of Finance*. — 2013. — Vol. 9, No. 2. — P. 217–248. — DOI: 10.1007/s10436-012-0219-9.
- [15] Hurn A., Jeisman J., Lindsay K. Teaching an old dog new tricks: improved estimation of the parameters of stochastic differential equations by numerical solution of the Fokker–Planck equation // *Financial Econometrics Handbook* / eds. Gregoriou G., Pascalau R. — London: Palgrave, 2010.
- [16] Orlov Yu. N., Fedorov S. L. Generation of nonstationary trajectories of a time series based on the Fokker–Planck equation. *Proceedings of MIPT*, 2016, vol. 8, no. 2, pp. 126–133.
- [17] Grimmett G. *Percolation*. — Berlin: Springer-Verlag, 1999. — 2nd ed.
- [18] Stauffer D., Aharony A. *Introduction to Percolation Theory*. — London: Taylor & Francis, 1992.
- [19] Zhukov D., Khvatova T., Lesko S., Zaltsman A. Managing social networks: applying the percolation theory methodology to understand individuals' attitudes and moods // *Technological Forecasting and Social Change*, 129 (2018), 297–307 — DOI: 10.1016/j.techfore.2017.09.039.
- [20] Sigov A. S., Zhukov D. O., Khvatova T. Yu., Andrianova E. G. A model of forecasting of information events on the basis of the solution of a boundary value problem for systems with memory and self-organization // *Journal of Communications Technology and Electronics*. — 63, 1478–1485 (2018). — DOI: 10.1134/S1064226918120227.
- [21] Zhukov D., Khvatova T., Millar C., Zaltsman A. Modelling the stochastic dynamics of transitions between states in social systems incorporating self-organization and memory // *Technological Forecasting and Social Change*. — 2020. — Vol. 158. — P. 120134. — DOI: 10.1016/j.techfore.2020.120134.
- [22] Zhukov D., Perova J., Kalinin V. Description of the distribution law and non-linear dynamics of growth of comments number in news and blogs based on the Fokker–Planck equation // *Mathematics*. — 2022. — Vol. 10. — P. 989. — DOI: 10.3390/math10060989.
- [23] Perova Yu. P., Lesko S. A., Zhukov D. O., Chechurin A. V. Analysis and modeling of processes in complex social network structures based

on the Fokker–Planck equation. Bulletin of Tomsk State University. Management, Computer Engineering and Informatics, 2022, no. 60, pp. 32–41. DOI: 10.17223/19988605/60/4.

- [24] Perova Yu. P., Grigoriev V. R., Zhukov D. O. Models and methods for the analysis of complex networks and social network structures. Russian Technological Journal, 2023, vol. 11, no. 2, pp. 33–49. DOI: 10.32362/2500-316X-2023-11-2-33-49.
- [25] Otradnov K. K., Perova Yu. P., Grigoriev V. R., Zhukov D. O. Analysis of time-series characteristics of sociodynamic processes in mass media user comment networks. Bulletin of RSREU, 2022, no. 82, pp. 147–161. DOI: 10.21667/1995-4565-2022-82-147-161.
- [26] Perova J. P., Kalinin V. N., Lesko S. A. Modeling the dynamics of changes in the number of comments of mass media users based on the

Fokker–Planck equation and parameters of networks of their connections // International Journal of Open Information Technologies. — 2023. — Vol. 11, No. 4.

About the authors:

Perova Yulia Petrovna, Senior Lecturer, Institute of Radio Electronics and Informatics, MIREA - Russian Technological University

Lesko Sergey Alexandrovich, Doctor of Technical Sciences, Associate Professor, Institute of Cybersecurity and Digital Technologies, MIREA – Russian Technological University.

Kalinin Vladimir Nikolaevich, laboratory assistant, Institute of Radioelectronics and Informatics, MIREA - Russian Technological University