

Научная статья

УДК 519.765, 004.912

DOI 10.25205/1818-7935-2024-22-1-50-64

**Опыт анализа социального самочувствия горожан:
соотнесение характеристик городской инфраструктуры здоровья
и тональности текстов из социальных сетей
(на примере Санкт-Петербурга)**

**Анна Владимировна Чижик¹
Александр Петрович Садохин²**

Национальный исследовательский университет ИТМО
Санкт-Петербург, Россия

¹chizhik@itmo.ru, <https://orcid.org/0000-0002-4523-5167>

²sadalpetr@yandex.ru, <https://orcid.org/0000-0002-6420-6601>

Аннотация

В статье описаны результаты исследования социального настроения горожан на основе текстовых данных из социальной сети ВКонтакте. Объект исследования – город Санкт-Петербург. Предлагаемый метод предполагает сопоставление тональности дискуссий пользователей социальных медиа с количественными данными, описывающими инфраструктуру города. Такое сопоставление возможно благодаря тому, что текстовые данные могут быть привязаны к району. Целью исследования было проанализировать возможности предлагаемого подхода, поэтому в качестве количественных данных взяты датасеты, описывающие инфраструктуру города по маркеру «поддержание здоровья»: количество поликлиник, больниц, площадок для занятия спортом. В статье описываются подходы к анализу тональности специфических текстовых данных (приводятся метрики качества используемых моделей), обосновывается актуальность связывания подобного рода данных с количественными показателями (статистиками, участвующими традиционно в методиках оценки благополучия городской среды). В работе продемонстрированы итоговые прикладные результаты: данные исследования нанесены на карту города.

Ключевые слова

обработка естественного языка, компьютерная лингвистика, машинное обучение, векторные представления текста, анализ тональности

Финансирование

Исследование выполнено при поддержке Российского научного фонда и Санкт-Петербургского научного фонда, грант № 23-28-10069 «Прогнозирование социального самочувствия с целью оптимизации функционирования экосистемы городских цифровых сервисов Санкт-Петербурга» (<https://rscf.ru/project/23-28-10069/>).

Для цитирования

Чижик А. В., Садохин А. П. Опыт анализа социального самочувствия горожан: соотнесение характеристик городской инфраструктуры здоровья и тональности текстов из социальных сетей (на примере Санкт-Петербурга) // Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2024. Т. 22, № 1. С. 50–64. DOI 10.25205/1818-7935-2024-22-1-50-64

© Чижик А. В., Садохин А. П., 2024

ISSN 1818-7935

Вестник НГУ. Серия: Лингвистика и межкультурная коммуникация. 2024. Т. 22, № 1

Vestnik NSU. Series: Linguistics and Intercultural Communication, 2024, vol. 22, no. 1

Analysis of the Citizen Social Well-Being: Correlation between Urban Public Health Infrastructure and Tonality of Texts from Social Networks (based on the Example of St. Petersburg)

Anna V. Chizhik¹
Aleksandr P. Sadokhin²

ITMO University
St.-Petersburg, Russian Federation

¹chizhik@itmo.ru, <https://orcid.org/0000-0002-4523-5167>

²sadalpetr@yandex.ru, <https://orcid.org/0000-0002-6420-6601>

Abstract

This paper presents the results of the social mood study of St. Petersburg citizens based on text data from the social network VKontakte. The proposed method involves juxtaposition of the social media users' tone of discussions with quantitative data describing the city infrastructure. The comparison is possible due to the fact that text data can be tied to a region. The article substantiates the method by which the concepts "social mood" and "social well-being" are combined. Approaches to sentiment analysis of specific textual data are also described (quality metrics of the models used). The paper demonstrates the final applicable results: the data are plotted on the map of the city.

Keywords

natural language processing, natural language understanding, word embeddings, text vectorization, sentiment analysis

Funding

This research is supported by Russian Science Foundation and Saint Petersburg Science Foundation, grant No. 23-28-10069 «Forecasting social well-being in order to optimize the functioning of the urban digital services ecosystem in St. Petersburg» (<https://rscf.ru/project/23-28-10069/>).

For citation

Chizhik A. V., Sadokhin A. P. Analysis of the citizen social well-being: correlation between urban public health infrastructure and tonality of texts from social networks (based on the example of St. Petersburg). *Vestnik NSU. Series: Linguistics and Intercultural Communication*, 2024, vol. 22, no. 1, pp. 50–64. (in Russ.) DOI 10.25205/1818-7935-2024-22-1-50-64

Введение

Городская среда является социально конструируемым феноменом: оказываясь в нем, люди образуют уникальные социальные сети; на его почве развиваются разнообразные культурные практики и создаются специфические пространства и символы, выражающие и отражающие менталитет и идентичность горожан. Таким образом, город не может описываться только географическими или административными категориями, и, скорее, представляет собой социокультурную среду, которая в своем движении входит в синхронизацию или, наоборот, рассинхронизацию с системой инфраструктуры региона (включающей политический, социальный и экономический компоненты). Географически город является, в первую очередь, особым многомерным пространством, которое можно описать рядом количественных данных. Стоит отметить, что субъективное восприятие этого пространства не менее важно. Ряд ученых [Duarte, 2017; Fang, 2017; Gibson, 1992; Almusaed, 2019; Riaz, 2014; Montgomery, 1998], размышляя над феноменом города, обращают внимание на то, городская среда темпоральна по своей сути, поэтому индивиды воспринимают город на уровне абстракций и воображения. С. Милгрэм отмечал [Milgram, 1970; Blass, 2005] в своих исследованиях, что город не является исключительно пространственной реальностью, а скорее – ее отражением в умах горожан. Г. Зиммель [Simmel, 1903] в своих исследованиях обращал внимание на то, что городам характерно особое

ментальное пространство. Основные тезисы его исследовательской работы на тему урбанизации [Borden, 1997] ведут к следующему выводу: жизнь города погружает индивида в достаточно нервное состояние за счет большого количества неопределенностей и необходимости быстрого реагирования на динамичную среду мегаполиса; это приводит к стремлению людей к объединению в социальные группы для обретения конструкции из общих настроений и мнений, уравнивающих непрерывно движущийся хаос. Эту же мысль продолжают исследования чикагской социологической школы, так, например, Р. Парк утверждает [Парк, 2002], что городское сообщество представляет собой территориально организованный элемент, индивидуальные единицы которого взаимодействуют между собой, руководствуясь эмоциональной близостью идей. При этом, как подчеркивает Э. Берджесс [Берджесс, 2002], между социальными группами существует конкуренция, которая рождает каскады изменения социального настроения. Итак, город – это не только некая среда обитания, но и сфера постоянной жизнедеятельности человека, результат которой сводится к двум феноменам: 1) создание индивидами образов города; 2) объединение индивидов в социальные группы, между членами которых, как правило, образы обладают схожими чертами. Следовательно, город есть территориальная общественная система, ядром которой являются социальные группы, складывающиеся исходя из специфики повседневных практик индивидов: работа, район проживания, увлечения и т. п. Таким образом, анализ ощущения комфортности городской среды жителями может стать критерием качественной оценки социального самочувствия горожан. Представляется целесообразным определять доминирующие эмоции социальных групп, функционирующих внутри городского пространства, а затем попробовать найти корреляцию между социальным настроением и критериями социальной благоустроенности городской среды.

1. Гипотеза

В рамках данного исследования мы попытались оценить настроения жителей города Санкт-Петербурга. В качестве базиса оценки мы выбрали категорию социального настроения. При этом нам хотелось, выяснив стабильные эмоции социальных групп, сопоставить их с маркерами, входящими в метрики классического измерения социального самочувствия.

Современный статистический ландшафт позволяет наблюдать поведение аудитории через ее реакцию на события городской жизни внутри виртуальной реальности, которая в формате коммуникативного поля взаимодействий доступна в социальных сетях. Концепция открытых данных, в свою очередь, предполагает наличие большого объема данных, описывающих разные стороны социального благополучия региона. Таким образом, опираясь на вычислительные модели эмоций и анализируя взаимосвязь реальности (статистик) и эмоционального фона индивидов (анализ тональности текстов), становится возможным приблизиться к оценке социального самочувствия в обход стандартным социологическим методам (формат опроса) и при этом более репрезентативно учесть социальное настроение как основной фактор формирования самоощущения индивида. Итак, если зафиксировать географическую привязку комментаторов в социальных сетях, то появляется возможность связать результаты анализа текстовых данных [Wu, 2020; Tan, 2019; Ramadhani, 2017], полученных из социальных сетей, и статистику, описывающую конкретные локации города.

Цель данного исследования – проверка гипотезы о возможности связать статистики о районах и анализ текстовых данных, поэтому на этом этапе было принято решение выбрать одну наиболее важную группу индикаторов социального самочувствия жителей и провести проверку идеи, связав только несколько маркеров с тональностями дискуссий в социальных сетях.

Одним из глобальных индикаторов социального самочувствия [Keynes, 1971] жителей города является группа показателей, описывающих доступность возможностей поддержания здоровья (медицинская помощь, объекты для занятий спортом и т. п.).

Далее будут описаны результаты анализа контента дискуссий в социальных сетях (определение настроений людей с привязкой к районам) и рефлексии над успешностью их связывания к данным, характеризующим доступность медицинской помощи (больницы и поликлиники) и возможности для занятий спортом (общественные спортивные площадки). Район был взят за единицу городского ландшафта. Статистика инфраструктуры районов была взята из открытых данных на сайтах администраций города.

2. Методы: обзор удачных практик

Есть несколько успешных примеров использования автоматического анализа данных, полученных из социальных сетей. Например, Graph Neural Network (GCN) использовалась для исследования влияния виртуального социального общения на предпочтения пользователей [Song, 2006; Tang, 2012; Peng, 2016]. Говоря об обработке естественного языка, следует также отметить, что большое значение имеют работы, посвященные тематическому моделированию и анализу настроений. Например, сверточная нейронная сеть (CNN) использовалась для создания языковой модели для анализа текстовых данных Twitter с целью последующего прогнозирования общественного резонанса вокруг новостной повестки (модель обучалась для английского языка, однако ее можно масштабировать для любого языка, включив в обучающие данные необходимый набор текстов) [Severyn, 2015]. Стоит отметить, что в исследовании приводятся метрики качества модели, а также ее сравнение с еще несколькими подобными разработками. Анализ настроений с использованием нейронных сетей стал предметом нескольких крупных исследований в Twitter, Facebook и Weibo [Volkova, 2015; Hao, 2015; Li, 2020]. Эмпирический материал этих трех исследований разный, но объединяются они за счет единой логики анализа текстов: за основу взяты шесть базовых эмоций [Ekman, 2007] (гнев, радость, удивление, страх, отвращение и печаль), которые с использованием моделей оценки настроений детектируются в текстах пользователей социальных сетей. Выявленный эмоциональный фон используется для предсказания поведения социальной группы.

В последние годы появился еще один подход к анализу эмоционального состояния индивидов с целью последующего прогнозирования и/или интерпретации социального настроения. Он заключается в том, что базовые эмоции необходимо связывать с набором ключевых слов, отражающих контекст темы. Таким образом, используется объединение тематического моделирования и анализа тональности текстов. Удачными примерами применения этой практики являются следующие исследования – [Pilato, D’Avanzo, 2018; Turón, 2023].

Также для улучшения метода анализа тональности можно использовать словари оценочной лексики, большая их часть разработана для английского языка, но имеет переводные версии (в том числе на русский). Примером успешного проекта может являться тональный словарь PolSentiLex [Koltsova, Alexeeva, Kolcov, 2016], исследователи отобрали 7546 слов на основе списка высокочастотных прилагательных, словаря ProductSentiRus, толкового словаря и перевода англоязычного словаря оценочной лексики SentiStrength. Затем не менее трех аннотаторов каждому слову присваивали оценки от -2 до $+2$.

Стоит отметить, что помимо оценки настроений, также можно использовать теорию графов, пытаясь найти взаимосвязь пользователей между собой, для того чтобы затем изучить поведение онлайн-пользователей и дальнейшее поведение социальной группы [Rykov, Koltsova, Sinyavskaya, 2020].

Что касается использования карт как научного метода, то в рамках междисциплинарных исследований он появился достаточно давно. Еще в середине XIX века британский врач Джон Сноу прибегнул к методу социальной картографии, выяснив с использованием карты причины распространения холеры в Лондоне; позже социолог Чарльз Бут создал социальную карту Лондона, в которой отразил благополучность условий жизни и труда (ученый разделил положение

горожан на 7 классов) [Volkova, 2015; Hao, 2015; Li, 2020]. Затем этот метод активно использовался социологами чикагской школы Робертом Парком и Эрнстом Берджесом (20–30-е годы XX века).

В большинстве исследовательских работ по этой теме выделяется идея о том, что социальные группы, функционирующие на территории города и объединяющиеся в городское общество, в своей организации представляют собой совокупность направлений и событий, которые поддаются фиксации на карте, а это, в свою очередь, помогает выделить основную тенденцию исследуемого явления (обоснованную некоторыми географическими закономерностями). Это помогает объяснить последовательность, казалось бы, не связанных между собой событий.

В этой связи стоит отметить ряд отечественных исследований. Например, интересным представляется исследование [Tikhonova, Yakimuk, Mityagin, 2021], в котором актуализируется ценностный подход к концепции умного города, а также предлагается методика анализа потребностей горожан через изучение их цифрового следа в социальных сетях (комментарии и посты). В результате такого анализа появляется возможность связать эмоции и контексты горожан с географическими координатами. Проблемами разработки методологической базы социологических исследований общественных пространств с применением картирования и анализа эмоций также успешно занимается А. Е. Ненько [Ненько, 2020]. Одним из ярких примеров полезности анализа текстов с целью исследования городской среды являются работы [Ненько, Недосека, 2022; Ненько, Недосека, Галактионова, 2021], в которых объектом исследования выбраны соседские онлайн-сообщества Санкт-Петербурга, а в качестве результата продемонстрировано развитие системы отношений «человек – среда» через процесс коллективного осмысления городской пространства как контекста жизни. Ученые показывают закономерности формирующегося вокруг городских и районных тем дискурса, а также актуализируют методы анализа текстовых данных. Также уже классическим исследованием является разметка карты Санкт-Петербурга с использованием теплового градиента [Ненько, Курилова, Подкорытова, 2020]. Тепловые карты скопления эмоций можно посмотреть как для конкретной эмоции, так и с использованием всех слоев одновременно. Стоит отметить, что в данном случае вместо анализа данных из социальных сетей исследователи прибегли к техникам классического социологического опроса. На наш взгляд, это также актуализирует интересное направление, в рамках которого можно попробовать сопоставить данные, получаемые из социальных сетей, с данными из опросов.

В настоящее время социальная картография представляет собой группу методов, включающих в группу методов как пространственно-графического, так и графического характера, объединенных одним объектом изучения – социальной реальностью [Вавилина, 2015]. Помимо фиксации особенностей территориальной организации различных объектов, социальное картирование позволяет упорядочить и визуализировать сложные и многочисленные данные социальной статистики применительно к пространству [Lynch, 1984; Горелова, 2020].

Четкого разграничения между методами нет, однако можно выделить социально-статистический, социально-антропологический, проектный и социально-сетевой подходы к картированию. В первую очередь обращает на себя внимание социально-антропологический подход, предусматривающий анализ территории с акцентом на антропоцентрическое содержание и субъектную составляющую пространственного измерения [Барбаш, 1994]. Социально-сетевой подход прекрасно дополняет эту основу, поскольку позволяет дополнить статистические данные о городском жизнеобеспечении данными из социальных сетей (в обоих случаях с привязкой к территории) [Вавилина, 2015]. Это дает возможность исследовать корреляции между эмоциями горожан и объективными событиями, окружающими их в момент их повседневной социальной траектории.

3. Описание набора данных

Для того чтобы узнать эмоции горожан, мы обратились к постам и комментариям в региональных онлайн-сообществах (социальная сеть ВКонтакте), в выборку вошли 18 публичных

групп. Количество групп продиктовано разделением Санкт-Петербурга на 18 административно-территориальных единиц (закон Санкт-Петербурга от 25 июля 2005 года «О территориальном устройстве Санкт-Петербурга»), таким образом, запрос к API VK подразумевал поиск по ключевым биграммам (типа «название – район»), также учитывались параметры количества подписчиков, активности постинга и динамики ответов комментаторов. В итоге каждый район был представлен одной группой: численность подписчиков в среднем варьировалась от 50 тыс. до 80 тыс.; посты в выбранных группах появляются с частотой 1–2 раза в сутки, динамика комментариев к ним варьируется от 3 до 20 (в среднем).

Первичный анализ текстовых данных (с использованием модели LDA) выявил ключевые темы, которые обсуждались в группах. Был проведен ряд экспериментов с параметром количества тем, для качества модели LDA мы использовали относительную перплексию, которая принимает значения из отрезка $[0, 1]$ (чем меньше, тем лучше). Отметим, что достаточно часто для оценки успешности обученной модели используют обычную перплексию $P = \exp(-L/N)$. Однако она зависит от мощности словаря и распределения частот слов в коллекции. В силу специфики наших данных было необходимо найти критерий, нечувствительный к изменению состава словаря. Итак, в ходе проведенных экспериментов для обучаемой нами модели LDA оптимальным оказался параметр $\text{num_topics}=5$. Изучение топа ключевых слов в каждой теме и проверка содержания отдельных текстов показали, что пятью самыми популярными темами оказались (рис. 1): проблемы района (2), вопросы здоровья (3), спорт (1), семья (6) и мероприятия/встречи (4). Это подтвердило гипотезу значимости темы поддержания здоровья в сознании горожан.

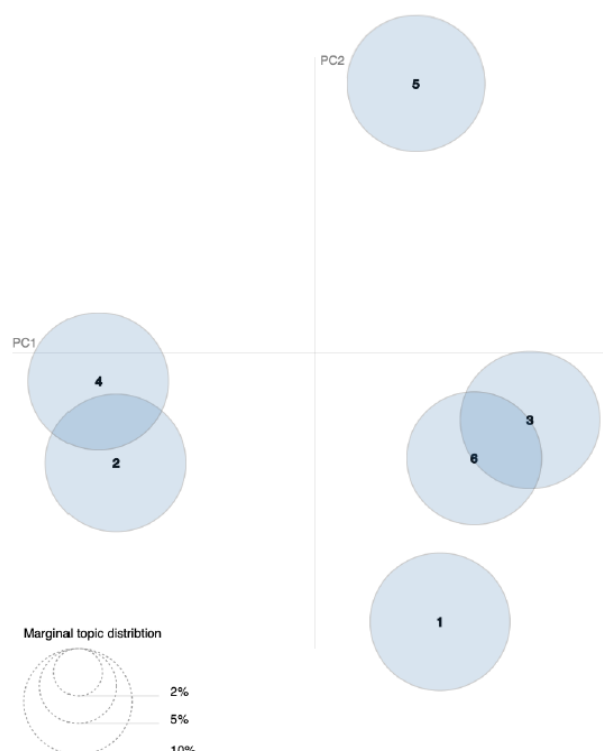


Рис. 1. Соотношение тематических кластеров
Fig. 1. Location and volume of thematic topics in relation to each other

Стоит отметить, что выделился еще один тематический кластер, в котором содержались разрозненные тексты о ежедневных событиях (потерянные ключи и кошельки, графики ра-

боты администрации и т. д.). Простейшее измерение тональности сообщений (библиотека Dostoevsky, обученная на наборе данных RuSentiment) дало представление о том, что в большинстве случаев сообщения и комментарии к ним в этом тематическом кластере имеют нейтральный тон ($\{ \text{'neutral': 0.65, 'positive': 0.35} \}$, это среднее значение по всем данным в теме). Поэтому было принято решение об удалении этого кластера из анализируемых данных.

Также было принято решение оценить посты (начало обсуждения темы) с точки зрения количества комментариев к ним. Если пост содержал менее 5 комментариев длиной не менее 30 символов, то он удалялся из набора данных. Это правило позволило очистить текстовые данные от неприносящих особого вклада восторженных коротких реплик типа «класс!» или «потрясающе!» (и антонимичных междометий) и тем самым сбалансировать данные с точки зрения содержания.

Отметим, что мы собирали данные за 2018–20-е годы. Это было сделано потому, что было важно проверить идею о возможности считывания эмоций, описывающих городское пространство, без шума в данных. Мировая и федеральная новостная повестка после пандемии COVID-19 могла существенно повлиять на темы и тон комментариев. Поэтому ближайшие годы (2021–2023 гг.) были исключены из анализа.

Средняя длина комментариев в итоговой выборке – 525 слов (на пост оставляют 10–15 комментариев); средняя длина постов – 41 слово. Так как целью была оценка общего тона реакции на пост, то комментарии к нему суммировались и составляли одну единицу текстовых данных.

Собранные текстовые данные были предварительно обработаны в соответствии со следующими этапами: 1) токенизация; 2) удаление спецсимволов, эмодзи, ссылок и знаков препинания; 3) удаление стоп-слов; 4) нормализация токенов. Окончательный набор обучающих данных содержал 104 107 строк, общее количество содержащихся в датасете токенов – 53 924 562.

Нами была проведена серия экспериментов по векторизации текстовых данных, с целью выяснить, какой метод лучше подходит к специфическим текстам из социальной сети.

На первом этапе эксперимента было принято решение посмотреть способность анализируемых моделей эмбедингов (TF-IDF, Word2Vec и Laser) к разделению на кластеры (использовался алгоритм К-средних, $k = 2$). Результаты разбиения на два кластера представлены на рис. 2. Для визуализации результатов использовался алгоритм понижения размерности PCA.

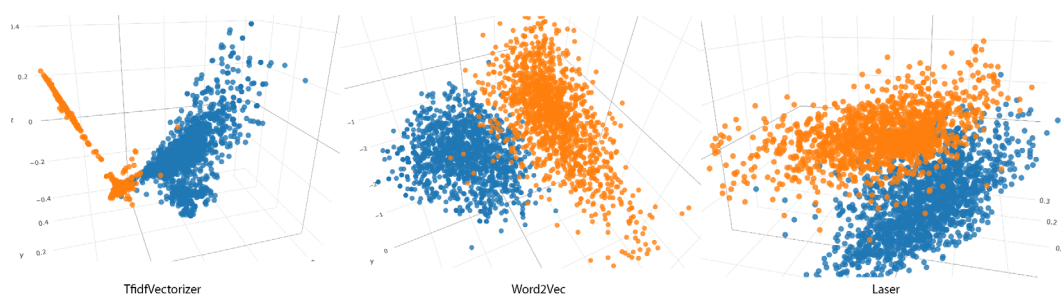


Рис. 2. Анализ разделимости данных на кластеры с использованием 3D-сжатия векторного пространства с использованием алгоритма PCA

Fig. 2. Analysis of data separability into clusters using 3D vector space compression using the PCA algorithm

Как видно из графика, TF-IDF неплохо учитывает разницу в данных. Далее мы разметили треть имеющихся данных вручную на два класса. В разметке участвовали 10 респондентов (студенты-филологи), что позволило избежать эффекта усталости от текстов. Разметчики получили инструкцию, которая объясняла, на какие параметры ориентироваться при отнесении текста к одному из двух классов. Мы предложили нейтральные тексты относить к классу «по-

зитивных»; для детекции негативных коннотаций был создан шаблон правил, учитывающий специфику построения восклицательно-негативных фраз, контексты сарказма и т. п. Стоит отметить, что процесс разметки предполагал также, что сами респонденты имеют необходимые навыки определения тональности, поэтому мнение разметчика априорно считалось выше вспомогательных тезисов инструкции. В итоге размеченная часть датасета позволила посмотреть содержание «негативного» и «позитивного» контента. Так, было выяснено, что «негативные» тексты гораздо важнее правильно детектировать нежели «позитивные», так как они явно более содержательны с точки зрения возможностей дальнейшего анализа контекстов (рис 3).

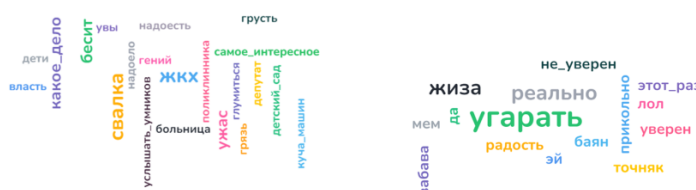


Рис. 3. Облако слов, вносящих наибольший вклад в «негативный» класс (слева); облако слов, вносящих наибольший вклад в «позитивный» класс (справа)
Fig. 3. The word cloud that contributes the most to the “negative” class (left); cloud of words contributing the most to the “positive” class (right)

Далее мы провели эксперимент по бинарной классификации текстов. В качестве алгоритма классификации была выбрана логистическая регрессия. На рис. 4 представлены результаты работы логистической регрессии при анализе датасета районных сообществ.

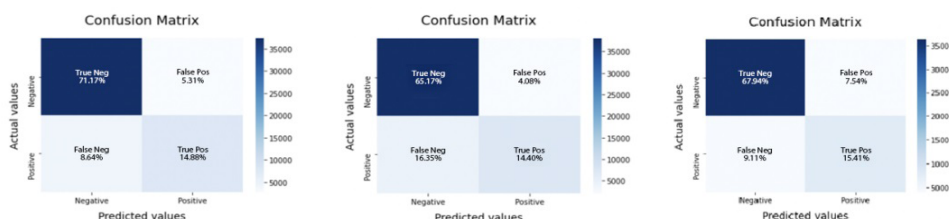


Рис. 4. Результаты работы модели логистической регрессии при отправленных в нее векторных представлениях, полученных тремя способами (слева направо: tf-idf, w2v, Laser)
Fig. 4. The results of the logistic regression model with vector representations sent to it, obtained in three ways (from left to right: tf-idf, w2v, Laser)

Результаты экспериментов показали, что все три метода работают с текстовыми данными этого типа достаточно хорошо, однако по приведенным матрицам видно, что лучше всего «негативный» класс детектируется логистической регрессией, работающей на векторном представлении tf-idf. Таким образом, для нашего дальнейшего исследования мы остановились именно на нем.

4. Результаты анализа тональности

Мы решили использовать модель ruGPT-3 для анализа тональности окончательного набора данных. Это авторегрессионная генеративная языковая модель, обученная на наборе данных русского языка. Модель основана на GPT-3 и может повысить точность анализа настроений за счет использования возможностей генерации естественного языка. Мы решили использовать ruGPT-3 Large (760 млн параметров) без дополнительного обучения. В будущем мы планируем придумать шкалы, учитывающие плавные градиенты эмоций, а в рамках текущих экспериментов использовали бинарную классификацию на положительные и отрицательные маркеры

настроений. Перплексия использовалась как мера определения тона настроения. В теории информации перплексия – это мера того, насколько хорошо распределение вероятностей предсказывает выборку, таким образом, идея была основана на обратной вероятности тестового набора, нормализованной по количеству слов. В качестве входных данных модели подавалась пара в формате <уровень счастья> + <текст> ('счастливый:' + текст и 'печальный:' + текст). Далее для каждого варианта рассчитывалась перплексия. Затем значения сравнивались друг с другом, чтобы найти наименьшее значение. Так присваивался класс, к которому тяготел текст.

Для оценки качества алгоритма использовался набор данных, размеченный вручную. На проверочной выборке были получены следующие результаты: Accuracy = 0,74 и F1-мера = 0,78.

После подтверждения того, что общий подход с использованием перплексии оказался полезным, был введен третий класс настроений, «умеренно счастливый» (так мы определили класс, тяготеющий к нейтральной тональности).

Таким образом, мы получили следующую шкалу настроения горожан: счастливый (2) > умеренно счастливый (0) > печальный (1). В результате карта Санкт-Петербурга была размечена (рис. 5).

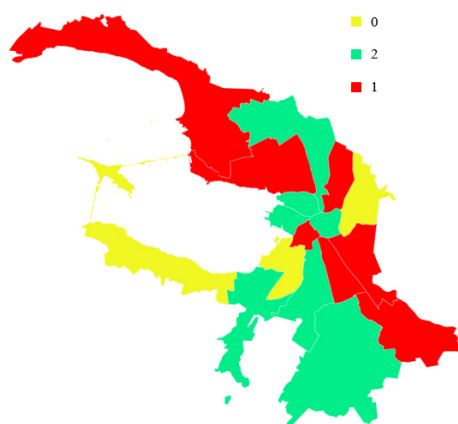


Рис. 5. Размеченная карта Санкт-Петербурга, три класса эмоций:
счастливый (2) > умеренно счастливый (0) > грустный (1)

Fig. 5. Marked map of St. Petersburg, three classes of emotions: happy (2) > moderately happy (0) > sad (1)

Следующим шагом стала маркировка количественных данных. К инфраструктурам, напрямую связанным с возможностями поддержания здоровья, мы отнесли три компонента: 1) наличие поликлиник (количество городских поликлиник в районе); 2) количество больниц в районе; 3) наличие возможностей для занятия спортом (открытые спортивные площадки, футбольные и хоккейные поля, наличие фитнеса и тренажерных залов). Количественные данные были взяты из общедоступных наборов данных и дополнены информацией, предоставленной районными администрациями по запросу.

Далее нужно было составить правило приближения статистики к трем классам (чтобы эти данные можно было сопоставить с маркировкой эмоций жителей районов, полученной на предыдущем шаге). Были выделены три класса: (1) нет возможности поддерживать здоровье в районе; (0) средние возможности; (2) в районе созданы все условия для профилактики и сохранения здоровья. Далее была создана формула расчета класса (количество_поликлиник + 0,5 * количество_больниц + 0,2 * возможности_спорта). Максимальное значение представленной формулы = 14. Шаг классов мы сделали в 4 единицы. Учитывалась важность необходимого количества поликлиник по отношению к количеству жителей района. Также

мы предположили, что наличие возможностей для занятия спортом не так сильно должно влиять на настроение жителей района (так как достаточно часто люди находят локации для этих целей за пределами места жительства и не считают это обременительным). Таким образом, был введен коэффициент этого фактора.

На рис. 6 представлены результаты.

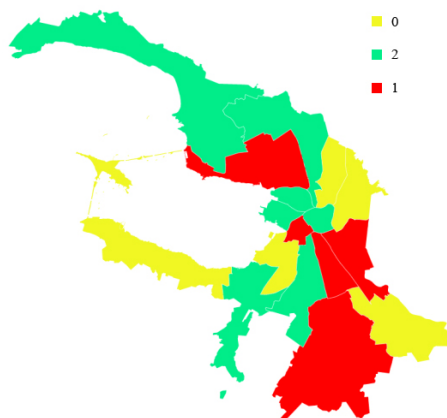


Рис. 6. Карта Санкт-Петербурга с нанесенными данными о возможностях поддержания здоровья (поликлиники, больницы, места для занятий спортом)

Fig. 6. Map of St. Petersburg marked with data on health opportunities (polyclinics, hospitals, places for sports)

Результаты показывают, что существует связь между доступом к возможностям для здоровья и притяжением эмоций жителей к отрицательным или положительным полюсам. Однако в двух (из 18) районах корреляции не выявлены. Количественные данные показывают, что в районах все хорошо с доступом к медицине и спорту, но дискурсы в онлайн-сообществах имеют негативную коннотацию. С одной стороны, это допустимая ошибка. С другой стороны, это говорит о необходимости поиска дополнительного статистического параметра, объясняющего, с чем связана негативность онлайн-общения. Также это показывает, что трех выделенных классов недостаточно в случае, когда нет точного совпадения количественных данных и результатов анализа тональности.

Заключение

Полученные результаты свидетельствуют о применимости метода для поиска факторов, влияющих на социальное настроение жителей районов. Однако необходимы некоторые дополнительные эксперименты. Мы планируем разработать определение настроений с использованием нескольких эмоциональных шкал. Если на данный момент используется самая распространенная палитра (от негативных коннотаций к позитивным), то в будущем представляется целесообразным проверить куб Лёвхейма [Колмагорова, 2023]. Мы также планируем пересмотреть используемую формулу корреляции количественных данных с тремя выбранными классами. Также целесообразным представляется дополнить количественные данные еще несколькими статистиками: количеством аптек, спортивных школ, статистикой преступлений по районам и зафиксированных проблем с ЖКХ. Необходимо проанализировать нагрузку на спортивный объект с точки зрения количества людей, его использующих (такая статистика есть в районных администрациях).

Список литературы

- Duarte F., Firmino R. J.** Unplugging the city: The urban phenomenon and its sociotechnical controversies. 2017.
- Ekman P., Revealed E.** Recognizing faces and feelings to improve communication and emotional life // *Emotions revealed*. 2007.
- Fang C., Yu D.** Urban agglomeration: An evolving concept of an emerging phenomenon // *Landscape and urban planning*. 2017. Vol. 162. P. 126–136.
- Gibson D. V., Kozmetsky G., Smilor R. W. (ed.)**. The technopolis phenomenon: Smart cities, fast systems, global networks. 1992.
- Almusaed A., Almssad A.** City phenomenon between urban structure and composition // *Sustainability in Urban Planning and Design*. 2019. Vol. 3.
- Koltsova O. Yu., Alexeeva S. V., Kolcov S. N.** An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media // *Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference "Dialogue-2016"*. 2016. Vol. 15(22). P. 277–287.
- Pilato G., D'Avanzo E.** Data-driven social mood analysis through the conceptualization of emotional fingerprints // *Procedia computer science*. 2018. Vol. 123. P. 360–365.
- Riazi S. A. H.** City as an Interdisciplinary Phenomenon // *Interdisciplinary Studies in Humanities*. 2014. Vol. 6, №. 1. C. 101–115.
- Rykov Y., Koltsova O., Sinyavskaya Y.** Effects of user behaviors on accumulation of social capital in an online social network // *Plos one*. 2020. Vol. 15. №. 4. P. e0231837. URL: <https://doi.org/10.1371/journal.pone.0231837>
- Turón A.** Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis // *Public health*. 2023. Vol. 215. P. 83–90.
- Tikhonova O., Yakimuk I., Mityagin S. A.** Detection the Relevance of Urban Functions for Value-Based Smart City Management // *International Conference on Digital Transformation and Global Society*. Cham: Springer International Publishing, 2021. P. 193–206.
- Montgomery J.** Making a city: Urbanity, vitality and urban design // *Journal of urban design*. 1998. Vol. 3, № 1. C. 93–116.
- Milgram S.** The experience of living in cities: A psychological analysis // *Annual Meeting of the American Psychological Association*, Sep, 1969, Washington, DC, US; This paper is based on an Invited Address presented to the Division of General Psychology at the aforementioned meeting. American Psychological Association. 1970.
- Blass T.** The urban psychology of Stanley Milgram // *Journal of Social Distress and the Homeless*. 2005. Vol. 14, № 1–2. P. 12–22.
- Simmel G.** The metropolis and mental life // *The urban sociology reader*. 1903. P. 23–31.
- Borden I.** Space beyond: Spatiality and the city in the writings of Georg Simmel // *The Journal of Architecture*. 1997. Vol. 2, № 4. P. 313–335.
- Парк Р.** Город как социальная лаборатория // *Социологическое обозрение*. 2002. Vol. 2, № 3. P. 3–12.
- Берджесс Э.** Рост города: введение в исследовательский проект // *Личность. Культура. Общество*. 2002. Т. 4. №. 1–2. С. 168–181.
- Ядов В. А.** Социальные и социально-психологические механизмы формирования социальной идентичности личности // *Мир России. Социология. Этнология*. 1995. Т. 4, № 3–4. С. 158–181.
- Поршнев Б. Ф.** Социальная психология и история. М., 1966.
- Бергер П., Лукман Т.** Социальное конструирование реальности. М., 1995.
- Бурдьё П.** Экономическая антропология. Курс лекций в Коллеж де Франс (1992–1993). М., 2019.

- Бурдые П. Социология политики. М., 1993.
- Гидденс Э. Устроение общества. Очерк теории структуризации. М., 2005.
- Keynes J. M. et al. The Collected Writings of John Maynard Keynes. London: Macmillan. 1971. Vol. 30. P. 1971–1989.
- Wu L., Zhang Q., Chen C.-H., Guo K., Wang D. Deep learning techniques for community detection in social networks. IEEE Access. 2020. Vol. 8. P. 96016–96026. DOI: 10.1109/ACCESS.2020.2996001
- Tan Q., Liu N., Hu X. Deep representation learning for social network analysis. Front. Big Data. 2019. DOI: 10.3389/fdata.2019.00002
- Ramadhani A. M., Goo H. S. Twitter sentiment analysis using deep learning methods; Proceedings of the 2017 7th International Annual Engineering Seminar (InAES); Yogyakarta, Indonesia, 2017. P. 1–4.
- Song X., Tseng B. L., Lin C.-Y., Sun M.-T. Personalized recommendation driven by in-formation flow; Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; Seattle, WA, USA. 2006. P. 509–516.
- Tang X., Yang C.C. Ranking user influence in healthcare social media. ACM Trans. Intell. Syst. Technol. (TIST). 2012. Vol. 3. P. 1–21. DOI: 10.1145/2337542.2337558
- Peng S., Wang G., Xie D. Social influence analysis in social networking big data: Opportunities and challenges. IEEE Netw. 2016. Vol. 31. P. 11–17. DOI: 10.1109/MNET.2016.1500104NM
- Severyn A., Moschitti A. Twitter sentiment analysis with deep convolutional neural networks; Proceedings of the 38th international ACM SIGIR Conference on Research and Development in Information Retrieval; Santiago, Chile. 2015. P. 959–962.
- Volkova S., Bachrach Y., Armstrong M., Sharma V. Inferring latent user properties from texts published in social media // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (Austin, TX: Association for the Advancement of Artificial Intelligence). 2015. P. 4296–4297.
- Hao B., Li L., Gao R., Li A., Zhu T. Sensing subjective well-being from social media // Active Media Technology, eds D. Ślęzak, G. Schaefer, S. T. Vuong, and Y. Kim (Cham: Springer). 2014. P. 324–335. DOI: 10.1007/978-3-319-09912-5_27
- Li S., Wang Y., Xue J., Zhao N., Zhu T. The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users // Int. J. Environ. Res. Public Health. 2020. Vol. 17. Art. 2032. DOI: 10.3390/ijerph17062032
- Ненько А. Е. Социологические методы изучения общественных пространств. СПб.: Университет ИТМО 2020.
- Ненько А. Е., Недосека Е. В. Ценности городской среды в дискурсе соседских онлайн-сообществ // Журнал социологии и социальной антропологии. 2022. Т. 25, № 1. С. 217–251.
- Ненько А. Е., Недосека Е. В., Галактионова А. А. Возможности семантического анализа ключевых биграмм для исследования дискурса соседского онлайн-сообщества // International Journal of Open Information Technologies. 2021. Т. 9, № 12. С. 111–118.
- Ненько А. Е., Курилова М. А., Подкорытова М. И. Анализ эмоционального восприятия территорий и развитие «Умного города» // International Journal of Open Information Technologies. 2020. Т. 8, № 11. С. 128–136.
- Вавилина Н. Д., Скалабан И. А. Социальное картирование: метод исследования и инструмент развития территории. М., 2015.
- Lynch K. Reconsidering the image of the city. 1984. P. 151–161.
- Горелова Ю. Образ города в восприятии горожан. М., 2020.
- Барбаш Н. Б., Крючков Ю. А. Социальное картографирование как способ пространственной организации данных о социально-территориальной дифференциации состава и качества жизни населения // Прогнозное социальное проектирование: теоретико-методологические и методические проблемы. М.: Наука. 1994. С. 115–129.

Колмогорова, А. В., Калинин А. А. Визуализация данных в эмоциональном анализе русскоязычных интернет-текстов на основе модели «Куб Лёвхейма» // Язык и искусственный интеллект: Сб. ст. по итогам конференции «Лингвистический форум 2020.–2023». С. 167–181. DOI 10.37892/978-5-907498-47-1-8

References

- Duarte F., Firmino R. J.** Unplugging the city: The urban phenomenon and its sociotechnical controversies. 2017.
- Ekman P., Revealed E.** Recognizing faces and feelings to improve communication and emotional life // Emotions revealed. 2007.
- Fang C., Yu D.** Urban agglomeration: An evolving concept of an emerging phenomenon // Landscape and urban planning. 2017. Vol. 162. P. 126–136.
- Gibson D. V., Kozmetsky G., Smilor R. W. (ed.).** The technopolis phenomenon: Smart cities, fast systems, global networks. 1992.
- Almusaed A., Almssad A.** City phenomenon between urban structure and composition // Sustainability in Urban Planning and Design. 2019. Vol. 3.
- Koltsova O. Yu., Alexeeva S. V., Kolcov S. N.** An Opinion Word Lexicon and a Training Dataset for Russian Sentiment Analysis of Social Media // Computational Linguistics and Intellectual Technologies: Papers from the Annual International Conference “Dialogue-2016”. 2016. Vol. 15(22). P. 277–287.
- Pilato G., D’Avanzo E.** Data-driven social mood analysis through the conceptualization of emotional fingerprints // Procedia computer science. 2018. Vol. 123. P. 360–365.
- Riazi S. A. H.** City as an Interdisciplinary Phenomenon // Interdisciplinary Studies in Humanities. 2014. Vol. 6, № 1. С. 101–115.
- Rykov Y., Koltsova O., Sinyavskaya Y.** Effects of user behaviors on accumulation of social capital in an online social network // Plos one. 2020. Vol. 15. № 4. P. e0231837. URL: <https://doi.org/10.1371/journal.pone.0231837>
- Turón A.** Evolution of social mood in Spain throughout the COVID-19 vaccination process: a machine learning approach to tweets analysis // Public health. 2023. Vol. 215. P. 83–90.
- Tikhonova O., Yakimuk I., Mityagin S. A.** Detection the Relevance of Urban Functions for Value-Based Smart City Management // International Conference on Digital Transformation and Global Society. Cham: Springer International Publishing, 2021. P. 193–206.
- Montgomery J.** Making a city: Urbanity, vitality and urban design // Journal of urban design. 1998. Vol. 3, № 1. С. 93–116.
- Milgram S.** The experience of living in cities: A psychological analysis // Annual Meeting of the American Psychological Association., Sep, 1969, Washington, DC, US; This paper is based on an Invited Address presented to the Division of General Psychology at the aforementioned meeting. American Psychological Association. 1970.
- Blass T.** The urban psychology of Stanley Milgram // Journal of Social Distress and the Homeless. 2005. Vol. 14, № 1–2. P. 12–22.
- Simmel G.** The metropolis and mental life // The urban sociology reader. 1903. P. 23–31.
- Borden I.** Space beyond: Spatiality and the city in the writings of Georg Simmel // The Journal of Architecture. 1997. Vol. 2, № 4. P. 313–335.
- Park R.** The city as a social laboratory. Sociological Review. 2002. T. 2. No. 3. P. 3–12.
- Burgess E.** Urban growth: an introduction to the research project // Personality. Culture. Society. 2002. T. 4. No. 1–2. P. 168–181.
- Yadov V. A.** Social and socio-psychological mechanisms for the formation of social identity of the individual // World of Russia. Sociology. Ethnology. 1995. T. 4. No. 3–4. P. 158–181.
- Porshnev B.F.** Social psychology and history. 1966.
- Berger P., Lukman T.** Social construction of reality. 1995.

- Bourdieu P.** Economic anthropology. Course of lectures at the College de France (1992–1993). 2019.
- Bourdieu P.** Sociology of politics. 1993.
- Giddens E.** The Structure of Society. Essay on the theory of structuration. 2005.
- Keynes J. M. et al.** The Collected Writings of John Maynard Keynes. London: Macmillan. 1971. Vol. 30. P. 1971–1989.
- Wu L., Zhang Q., Chen C.-H., Guo K., Wang D.** Deep learning techniques for community detection in social networks. *IEEE Access*. 2020. Vol. 8. P. 96016–96026. DOI: 10.1109/ACCESS.2020.2996001
- Tan Q., Liu N., Hu X.** Deep representation learning for social network analysis. *Front. Big Data*. 2019. DOI: 10.3389/fdata.2019.00002
- Ramadhani A. M., Goo H. S.** Twitter sentiment analysis using deep learning methods; Proceedings of the 2017 7th International Annual Engineering Seminar (InAES); Yogyakarta, Indonesia, 2017. P. 1–4.
- Song X., Tseng B. L., Lin C.-Y., Sun M.-T.** Personalized recommendation driven by in-formation flow; Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval; Seattle, WA, USA. 2006. P. 509–516.
- Tang X., Yang C. C.** Ranking user influence in healthcare social media. *ACM Trans. Intell. Syst. Technol. (TIST)*. 2012. Vol. 3. P. 1–21. DOI: 10.1145/2337542.2337558
- Peng S., Wang G., Xie D.** Social influence analysis in social networking big data: Opportunities and challenges. *IEEE Netw*. 2016. Vol. 31. P. 11–17. DOI: 10.1109/MNET.2016.1500104NM
- Severyn A., Moschitti A.** Twitter sentiment analysis with deep convolutional neural networks; Proceedings of the 38th international ACM SIGIR Conference on Research and Development in Information Retrieval; Santiago, Chile. 2015. P. 959–962.
- Volkova S., Bachrach Y., Armstrong M., Sharma V.** Inferring latent user properties from texts published in social media // Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (Austin, TX: Association for the Advancement of Artificial Intelligence). 2015. P. 4296–4297.
- Hao B., Li L., Gao R., Li A., Zhu T.** Sensing subjective well-being from social media // *Active Media Technology*, eds D. Ślęzak, G. Schaefer, S. T. Vuong, and Y. Kim (Cham: Springer). 2014. P. 324–335. DOI: 10.1007/978-3-319-09912-5_27
- Li S., Wang Y., Xue J., Zhao N., Zhu T.** The impact of COVID-19 epidemic declaration on psychological consequences: a study on active Weibo users // *Int. J. Environ. Res. Public Health*. 2020. Vol. 17. Art. 2032. DOI: 10.3390/ijerph17062032
- Nenko A.E.** Sociological methods for studying public spaces. St. Petersburg, ITMO University, 2020.
- Nenko A. E., Nedoseka E. V.** Values of the urban environment in the discourse of neighborhood online communities // *Journal of Sociology and Social Anthropology*. 2022. T. 25, № 1. P. 217–251.
- Nenko A. E., Nedoseka E. V., Galaktionova A. A.** Possibilities of semantic analysis of key bigrams for studying the discourse of a neighborhood online community // *International Journal of Open Information Technologies*. 2021. Vol. 9. No. 12. P. 111–118.
- Nenko A. E., Kurilova M. A., Podkorytova M. I.** Analysis of emotional perception of territories and the development of a “Smart City” // *International Journal of Open Information Technologies*. 2020. Vol. 8. No. 11. P. 128–136.
- Vavilina N. D., Skalaban I. A.** Social mapping: a research method and a tool for territory development. 2015.
- Lynch K.** Reconsidering the image of the city. 1984. P. 151–161.
- Gorelova Yu.** The image of the city in the perception of citizens. 2020.
- Barbash N. B., Kryuchkov Yu. A.** Social mapping as a way of spatial organization of data on the socio-territorial differentiation of the composition and quality of life of the population //

Predictive social design: theoretical, methodological and methodological problems. Moscow, Nauka publ., 1994. P. 115–129.

Kolmogorova, A. V., Kalinin A. A. Data visualization in the emotional analysis of Russian-language Internet texts based on the Lövhheim Cube model // Language and artificial intelligence: Collection. Art. following the results of the conference “Linguistic Forum 2020–2023”. pp. 167–181. DOI 10.37892/978-5-907498-47-1-8

Информация об авторах

Чижик Анна Владимировна, кандидат культурологии, старший научный сотрудник, центр технологий электронного правительства Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики

Садохин Александр Петрович, доктор культурологии, профессор, эксперт Центра технологий электронного правительства Санкт-Петербургского национального исследовательского университета информационных технологий, механики и оптики

Information about the Authors

Anna V. Chizhik, Senior Researcher, eGovernment Center, ITMO University, St. Petersburg, Russian Federation

Aleksandr P. Sadokhin, Doctor of Cultural Studies, Professor, eGovernment Center expert, ITMO University, St. Petersburg, Russian Federation

*Статья поступила в редакцию 24.07.2023;
одобрена после рецензирования 12.10.2023; принята к публикации 17.01.2024*

*The article was submitted 24.07.2023;
approved after reviewing 12.10.2023; accepted for publication 17.01.2024*