

ИДЕАЛЬНЫЙ ПОЛИТИК ДЛЯ СОЦИАЛЬНОЙ СЕТИ: ПОДХОД К АНАЛИЗУ ИДЕОЛОГИЧЕСКИХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ¹

Л.Г. Бызов, Д.А. Губанов, И.В. Козицин, А.Г. Чхартишвили

Аннотация. Рассмотрена задача определения идейно-политических предпочтений пользователей онлайн-социальных сетей. Предложена модель, позволяющая оценивать идейно-политические предпочтения пользователей онлайн-социальной сети ВКонтакте на основе информации из их аккаунтов и учитывающая основные идеологические направления в современной России (ДЛС-модель). В основе метода оценки лежит обучение с учителем, при помощи которого выявление предпочтений сводится к задаче многоклассовой классификации и расчету апостериорных вероятностей принадлежности классам. Для реализации данного метода создана база размеченных анонимизированных данных, проведен ее анализ, разработаны соответствующие алгоритм и комплекс программ. Далее сформулирована и решена задача выбора политическим деятелем стратегии идеологического позиционирования в модели стохастических предпочтений (частным случаем которой является ДЛС-модель), которая потенциально позволит получить наибольшую поддержку рассматриваемого множества пользователей. На основании решения этой задачи и с помощью анализа данных получены стратегии позиционирования для некоторых популярных сетевых сообществ.

Ключевые слова: онлайн-социальная сеть, идейно-политические взгляды, ДЛС-модель, большие данные, задача поиска оптимальной идеологической позиции.

ВВЕДЕНИЕ

Решение задачи определения идейно-политических предпочтений пользователей онлайн-социальных сетей имеет огромную важность в теориях информационного управления и информационной войны [1–4]. Знание, какую позицию занимает тот или иной пользователь, какие идейно-политические ценности имеют для него большое значение, а какие несущественны, позволяет политикам, анализирующим подобную информацию, корректировать свои действия в зависимости от конкретных целей.

К настоящему моменту предложено множество методов для решения этой задачи [5–7]. В рамках данной работы мы предлагаем подход, позволяющий не только оценивать идейные предпочте-

ния пользователей онлайн-социальной сети (на примере сети ВКонтакте), но и использовать полученную оценку для выбора политическим деятелем той стратегии идеологического позиционирования, которая потенциально позволит получить наибольшую поддержку в рамках рассматриваемого множества пользователей. Ключевую роль здесь играет базис (набор основных идеологических направлений), в котором рассматриваются идейно-политические взгляды.

Отметим, что в данной работе мы не рассматриваем вопрос предпочтений пользователей в аспекте выбора между различными политическими деятелями (этот выбор зависит от ряда факторов, включая интенсивность политической рекламы, ораторское мастерство политика и др.). Предмет нашего рассмотрения — идейные (идейно-политические) предпочтения индивидов [8].

Далее в § 1 дан краткий обзор известных классификаций идейно-политических направлений и введена новая классификация, используемая в дан-

¹ Работа выполнена при частичной поддержке РФФИ (проект 18-29-22042).

ной работе. В § 2 описывается подход к определению идеологических предпочтений пользователей онлайн-социальной сети ВКонтакте (этот подход является развитием методологии, описанной в [9, 10]). В § 3 поставлена и решена задача поиска оптимальной идеологической позиции, минимизирующей сумму расстояний до позиций пользователей, а также приведены примеры расчета «идеальной» позиции. В Заключении приведены основные выводы и намечены направления дальнейших исследований.

1. ОСНОВНЫЕ ИДЕОЛОГИЧЕСКИЕ НАПРАВЛЕНИЯ В СОВРЕМЕННОЙ РОССИИ

Поиск основания для классификации респондентов по ценностям, идеологическим направлениям, идейно-политическим взглядам давно представляет собой актуальную задачу социологических и политологических исследований. Одна из первых классификаций предложена М. Рокичем в книге «Природа человеческих ценностей» [11]. Рокич выдвигает идею о том, что содержание главных четырех идеологических течений XX в. — социализма, коммунизма, фашизма и капитализма — можно представить в виде двумерной шкалы, образованной системами координат двух ценностей: свободы и равенства. Эта модель содержит четыре ячейки, образованные высоким и низким положением каждой из этих ценностей. Социализм соответствует ячейке с высокой степенью свободы и равенства, фашизм — положению с низкой степенью свободы и равенства, коммунизм занимает кластер с высокой степенью равенства и малой свободой, а при капитализме равенства нет, зато свобода велика. Для проверки этой модели Рокич в 1973 г. провел контент-анализ ключевых теоретических работ каждой из четырех идеологий, что позволило оценить позитивные и негативные упоминания свободы и равенства.

Теория Рокича вызвала ряд нареканий, заключающихся в том, что «...значимость таких ценностей, как свобода и равенство, может изменяться от одной среды к другой» [12, с. 266]. Отмечалось, что свобода и равенство воспринимаются неодинаково людьми из разных общественных групп и культур, что делает эти категории неоднозначными эталонами для систематизации всего спектра человеческих ценностей. «Для капиталиста свобода — это отсутствие принуждения, особенно от органов власти. Для социалиста свобода означает способность достичь своих целей, а это предполагает усилия органов власти по устранению проблем, например, бедности и расизма. Следовательно, сторонники этих идеологий могут оценить свободу только в их собственном понимании» [12, с. 266].

Схожим образом строит модель ценностей Р. Инглхарт [13, 14]. Он исходит из принципа, близкого концепции «пирамиды А. Маслоу», когда провозглашается иерархия ценностей, идущая от биологических к социальным потребностям высшего порядка. В основе ее лежит разделение ценностей на четыре составляющих: поддержание порядка в государстве, наделение людей большим влиянием на решение правительства, борьба с растущими ценами, защита свободы слова. Первая и третья составляющие в теории Инглхарта отнесены к материалистическим потребностям, вторая и четвертая — к постматериалистическим.

В статье [8] была предложена классификация, состоящая из шести политических направлений: социал-государственники, социал-демократы, радикальные националисты, державники, радикальные рыночники, западники.

В статье [15] рассмотрены шесть типов идеологических сторонников: националисты, сторонники стабильности, интернационалисты, неуверенные, рыночники и гуманисты. В той же статье были предложены четыре типа политических ценностей: рыночные, демократические, социалистические и державнические.

Далее идейно-политические взгляды россиян предлагается рассматривать как смесь трех ярко выраженных базисных установок: Державник-Либерал-Социалист [16]. Компонента «Державник» описывает, насколько индивид привержен идее о едином, сильном и независимом государстве. Компонента «Либерал» выражает важность для индивида уважения со стороны государства его (индивида) личных прав и свобод (в первую очередь политических и экономических). Компонента «Социалист» отвечает за стремление индивида к социально-экономической справедливости.

В 2018 г. Институт социологии РАН провел опрос [17], в котором респондентам предлагалось определить ценности, на которых должно строиться будущее России. Наибольшей поддержкой пользуются «Социальная справедливость» (59 %), «Права человека демократия, свобода самовыражения личности» (37 %), «Россия должна стать великой державой» (32 %). Нетрудно видеть, что эти ценности хорошо соответствуют трем базисным установкам, упомянутым выше: Социалист, Либерал, Державник соответственно.

Взгляды отдельного индивида (российского гражданина) в рамках предлагаемой в данной работе модели, которую будем далее называть *ДЛС-моделью*, представим в виде *ДЛС-вектора* (вектора-строки) из трех неотрицательных компонент, в сумме равных единице:

$$p = (p_1, p_2, p_3), \quad p_1, p_2, p_3 \geq 0, \quad p_1 + p_2 + p_3 = 1.$$

Например, индивид, описываемый ДЛС-вектором $(1, 0, 0)$, является чистым державником, которому не близки ценности либералов и социалистов. Напротив, индивид, характеризуемый ДЛС-вектором $(0, 1/2, 1/2)$, отвергает идеологию державников, разделяя при этом воззрения либералов и социалистов примерно в равной пропорции. Случай $(1/3, 1/3, 1/3)$ соответствует индивиду, не отдающему предпочтения ни одному из трех идеологических направлений.

В заключение § 1 отметим, что конкретный набор базисных взглядов, актуальный для нынешнего российского идейно-политического ландшафта, со временем может потребовать корректировки. Однако, как будет видно далее, предлагаемые в данной работе методы достаточно универсальные в том смысле, что их можно применить и по от-

ношению к другому базисному набору (в том числе к набору с числом базисных взглядов, отличным от трех).

2. ВЫЯВЛЕНИЕ ИДЕОЛОГИЧЕСКИХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ СЕТИ ВКОНТАКТЕ

2.1. Маркерные аккаунты

Отправной точкой для выявления предпочтений пользователей онлайн-социальной сети ВКонтакте служит экспертное определение маркерных аккаунтов (учетных записей) — тех аккаунтов сети ВКонтакте, контент которых может быть однозначно отнесен ровно к одному из трех идеологических направлений. Далее предполагается, что эти аккаунты характеризуются единич-

Таблица 1

Маркерные аккаунты

Наименование аккаунта	Идеологическое направление	URL
Проханов Александр Андреевич — творчество Александр Дугин Сергей Марков Наталия Нарочницкая Никита Михалков Наталья Поклонская Стрелков Игорь Иванович Новороссия / Игорь Стрелков Андрей Николаевич Савельев	Державники	https://vk.com/alexandr_prohanov https://vk.com/public2789767 https://vk.com/id382620481 https://vk.com/narochnitskaia https://vk.com/public26662151 https://vk.com/public333999 https://vk.com/public129997795 https://vk.com/public50414414 https://vk.com/id21503935
Николай Подосокорский Григорий Явлинский Явлинский. Группа сторонников Команда Явлинского Борис Вишневский Борис Вишневский Интернет-приемная Андрей Кураев Профессор богословия, протоиерей Андрей Кураев Станислав Белковский Владимир Кара-Мурза Борис Надеждин Илья Яшин	Либералы	https://vk.com/id2008214 https://vk.com/id146037120 https://vk.com/public30071970 https://vk.com/yavlinsky_yabloko https://vk.com/id2480895 https://vk.com/public40181447 https://vk.com/public22797031 https://vk.com/diak_kuraev https://vk.com/public30873220 https://vk.com/id263391319 https://vk.com/id40143548 https://vk.com/id3686745
Максим Шевченко Сергей Шаргунов СЕРГЕЙ ШАРГУНОВ Сергей Удальцов Сергей Удальцов Сергей Глазьев Сергей Глазьев Сергей Глазьев Анатолий Вассерман (фан-страница) Михаил Хазин Михаил Хазин ФЭИ Михаил Делягин Михаил Делягин (официальная страница)	Социалисты	https://vk.com/public28029604 https://vk.com/id5355252 https://vk.com/shargunoff https://vk.com/id43050243 https://vk.com/public39238475 https://vk.com/glazevs https://vk.com/public75299666 https://vk.com/club258208 https://vk.com/awasyk https://vk.com/id45286626 https://vk.com/m_khazin https://vk.com/id95640814 https://vk.com/public52638381

ными базисными ДЛС-векторами: $e_1 = (1, 0, 0)$ (чистый державник), $e_2 = (0, 1, 0)$ (чистый либерал) и $e_3 = (0, 0, 1)$ (чистый социалист).

Список маркерных аккаунтов с экспертной оценкой их идеологических направлений приведен в табл. 1.

2.2. Формирование перечня пользователей

После определения перечня маркерных аккаунтов был зафиксирован перечень пользователей, являющихся их друзьями (в случае, если маркерный аккаунт относится к типу «пользователь») или подписчиками. При этом отбирались только пользователи, удовлетворяющие следующим критериям:

- пользователь не удален и не заблокирован;
- пользователь посещал социальную сеть ВКонтакте хотя бы раз за последний месяц;
- аккаунт пользователя не был скрыт настройками приватности.

Для каждого из этих пользователей были выгружены такие данные его аккаунта:

- пол;
- возраст;
- политические предпочтения (далее — ПП), выбранные пользователем из списка, предоставляемого функционалом сети ВКонтакте (коммунистические, социалистические, умеренные, либеральные, консервативные, монархические, ультраконсервативные, индифферентные, либертарианские);
- паблики (тип аккаунта «page»), группы (тип аккаунта «group»), события (тип аккаунта «event») и блогеры (обладатели более 1000 подписчиков, тип аккаунта «profile»), на которые подписан пользователь;
- аккаунты друзей пользователя.

Отдельно рассматривались пользователи, связанные одновременно с несколькими маркерными аккаунтами, характеризующимися различными ДЛС-векторами (кросс-идеологические пользователи). В результате была сформирована выборка I , состо-

Таблица 2

Кросс-идеологические пересечения

Множество идеологических направлений маркерных аккаунтов, на которые подписан пользователь	Число пользователей
{Д}	64 139
{Л}	59 951
{С}	53 614
{Д, Л}	3179
{Д, С}	7671
{Л, С}	4205
{Д, Л, С}	1915
Всего пользователей	194 674

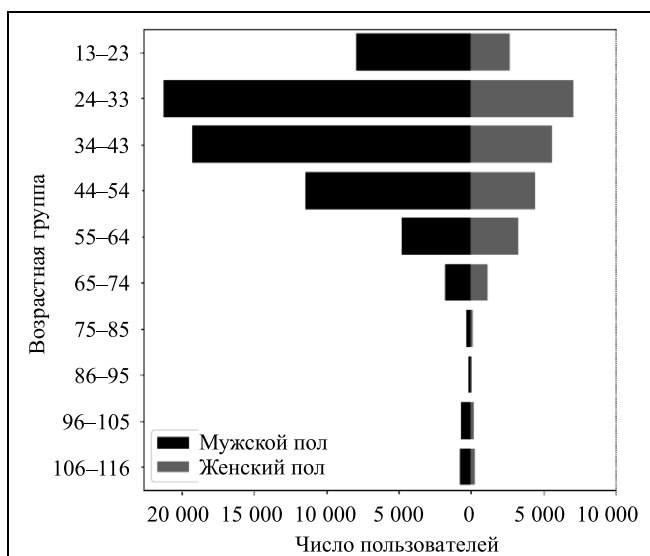


Рис. 1. Половозрастная структура пользователей (тех, кто указал свой пол и возраст): число пользователей, у которых не указан пол, 115, возраст — 119 235; минимальный указанный возраст 13 лет, максимальный — 116

ящая из 194 674 пользователей. Информация по кросс-идеологическим пересечениям в рассматриваемой выборке приведена в табл. 2. Отметим, что в выборку I попали некоторые маркерные аккаунты, поскольку они, будучи страницами пользователей (тип аккаунта «profile»), были друзьями других маркерных аккаунтов или были подписаны на них.

2.3. Разведочный анализ

Половозрастная структура пользователей выборки I приведена на рис. 1. Как видно из рисунка, большую часть выборки формируют аккаунты пользователей, указавших при регистрации мужской пол. Кроме того, наблюдается смещение по возрасту в сторону более молодых пользователей. Распределение пользователей относительно ПП представлено на рис. 2.

Описательная статистика числа подписок пользователей на паблики, группы, аккаунты событий и блогеров приведена в табл. 3, на рис. 3 представлены соответствующие гистограммы.

2.4. Анализ связей дружбы

Для анализа связей дружбы между пользователями выборки были выполнены следующие операции. Для каждого пользователя $i \in I$ был определен список тех его друзей V_i , которые также принадлежат множеству I . Данная информация была представлена в виде квадратной матрицы смежности $A = (a_{ij})$ (ее размерность совпадает с числом пользователей), в которой $a_{ij} \in \{0, 1\}$, при этом ра-

венство $a_{ij} = 1$ выполнено тогда и только тогда, когда $j \in V_i$.

Далее матрица **A** была симметризована: для каждой пары (i, j) , $i \neq j$, если хотя бы один из симметричных относительно главной диагонали элементов a_{ij} и a_{ji} равен 1, второй также полагался равным 1. Данный шаг позволил заполнить некоторые пропуски в данных, связанные с настройками приватности (когда из двух пользователей, соединенных связью дружбы, только у одного эта связь наблюдаема). Полученная в результате матрица (для удобства также будем обозначать ее **A**) была симметричной: $\mathbf{A} = \mathbf{A}^T$. Таким образом, пользователи $i \in I$, для которых соответствующая строка матрицы **A** оказалась нулевой (что соответствует отсутствию у пользователя i друзей среди пользователей J), далее не рассматривались.

В результате были сформированы множество $J \subset I$, $|J| = 156\,514$ (через $|X|$ здесь и далее обозначено число элементов в множестве X) и квадратная матрица $\mathbf{B} \in \{0, 1\}^{|J| \times |J|}$, являющаяся симметричной подматрицей матрицы **A** со строго положительными строчными (и, следовательно, столбцовыми) суммами. С помощью матрицы **B**, рассматривая ее как матрицу смежности, был построен граф $G[\mathbf{B}]$, фиксирующий связь дружбы между пользователями из множества J . Распределение степеней вершин этого графа представлено на рис. 4.

Для графа $G[\mathbf{B}]$ тремя различными способами был рассчитан коэффициент ассортативности [18]. (Напомним, что коэффициент ассортативности лежит в отрезке $[-1, 1]$ и показывает, насколько значения некоторого признака вершин графа коррелируют для соседних вершин (в данном случае число значений признака равно 3). Если этот коэффициент близок к нулю, то корреляции нет; если положителен, то соседние вершины чаще имеют одинаковые метки; если отрицателен, то соседние вершины чаще имеют разные метки).

Таблица 3

Дескриптивная статистика подписок пользователей

	Число подписок на:			
	паблики	группы	события	блогеров
Среднее	167,77	34,45	5,37	40,6
Стандартное отклонение	322,05	91,82	32,55	140,12
Минимум, %	0	0	0	0
25	11	0	0	0
50	66	1	0	3
75	180	15	0	13
Максимум	4894	955	962	1000

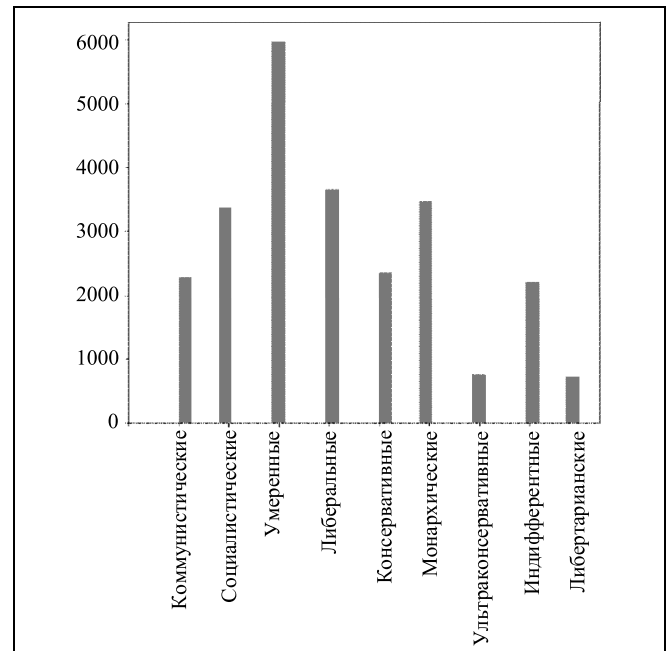


Рис. 2. Распределение пользователей относительно ПП для тех, у кого они указаны: доля пользователей, у которых ПП не указаны, составляет 0,87

В первом случае метки вершин были сгенерированы в соответствии с табл. 2: всего было семь типов меток, каждая соответствовала одной из семи возможных комбинаций подписок пользователей на маркерные аккаунты. Полученный коэффициент ассортативности оказался равным 0,14. Во втором случае рассматривались только четыре типа меток: все кросс-идеологические пользователи ассоциировались с одним и тем же типом меток. Коэффициент ассортативности при таком способе раздачи меток равен 0,15. В третьем случае кросс-идеологические пользователи не рассматривались вообще. Коэффициент ассортативности в этом случае равен 0,35.

Полученные результаты свидетельствуют о присутствии эффекта гомофилии в рассматриваемом графе: подписчики маркерных аккаунтов предпочитают иметь в друзьях подписчиков маркерных аккаунтов того же идеологического направления. При этом для кросс-идеологических пользователей данный эффект менее выражен. На рис. 5 (см. третью страницу обложки) изображена часть графа $G[\mathbf{B}]$, соответствующая третьему способу раздачи меток и наглядно демонстрирующая указанную тенденцию.

2.5. Оценивание ДЛС-векторов

Задача оценивания ДЛС-вектора российского пользователя сети ВКонтакте может быть сведена

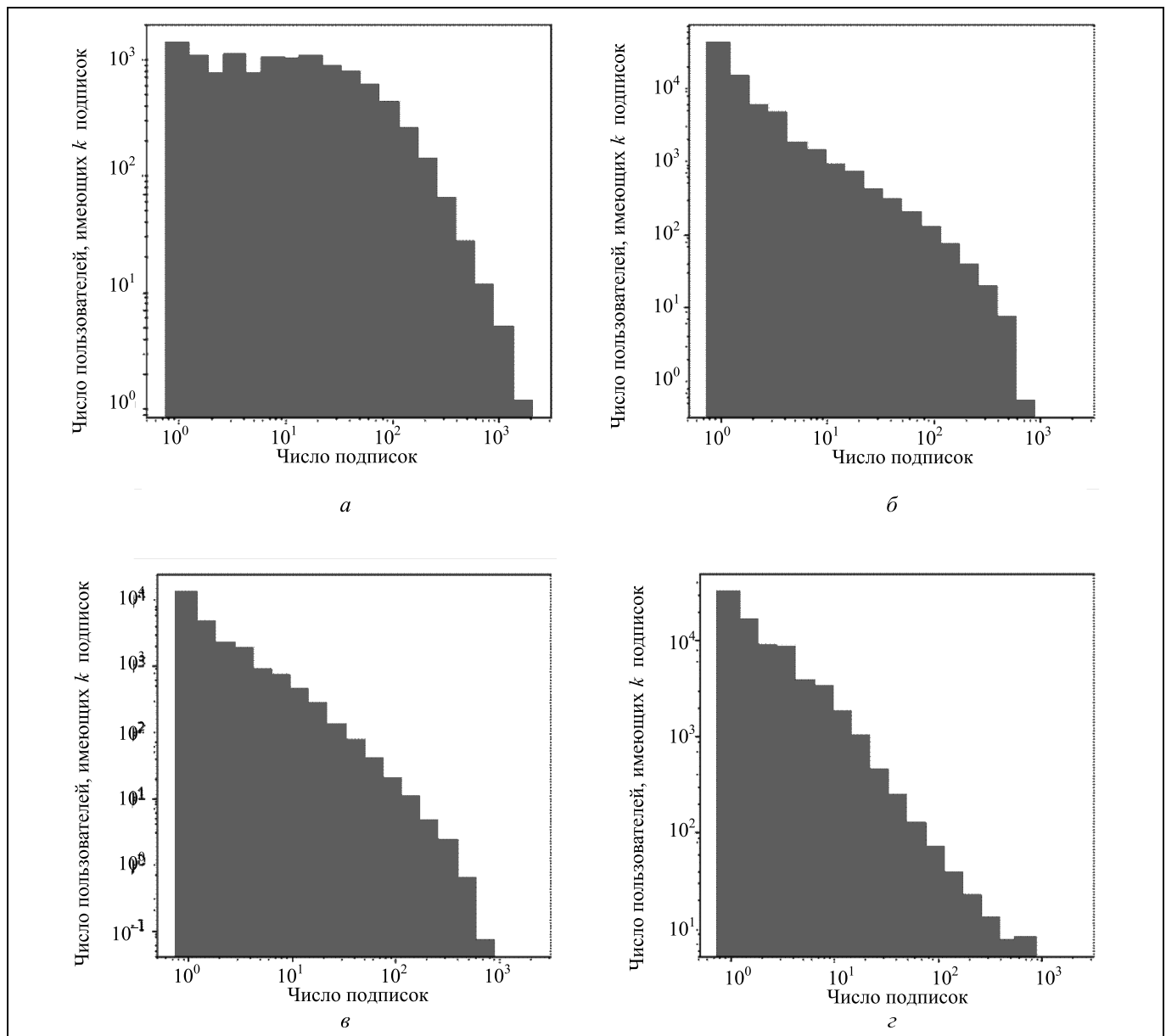


Рис. 3. Гистограммы распределений числа подписок пользователей на паблики (а), группы (б), блогеров (в) и события (г)

к задаче трехклассовой классификации. Рассматриваются классы: державники (класс Д), либералы (класс Л) и социалисты (класс С). Множество пользователей, не входящих в кросс-идеологические пересечения, обозначим через $I_* \subset I$. Число элементов множества I_* обозначим через N , $N = |I_*| = 177\,704$. Каждый пользователь из множества I_* может быть однозначно отнесен к одному из трех классов. Далее под словосочетанием «размеченные данные» будем понимать данные аккаунтов (см. п. 2.2), а также метки классов пользователей множества I_* . В соответствии с табл. 2 численности классов составляют 64 139, 59 951 и

53 614 соответственно. В качестве признаков (независимых переменных) принимались ПП пользователей, а также их подписки на паблики, группы, страницы событий и блогеров. Признаки были формализованы с помощью one-hot кодирования (унитарный код) [19]. В результате каждому пользователю ставился в соответствие вектор из нулей и единиц, размерность которого совпадала с числом значений всех признаков, а единица на соответствующей позиции означала, что данный признак для данного пользователя принимает именно это значение. Формально это можно описать таким образом: была сформирована матрица объек-

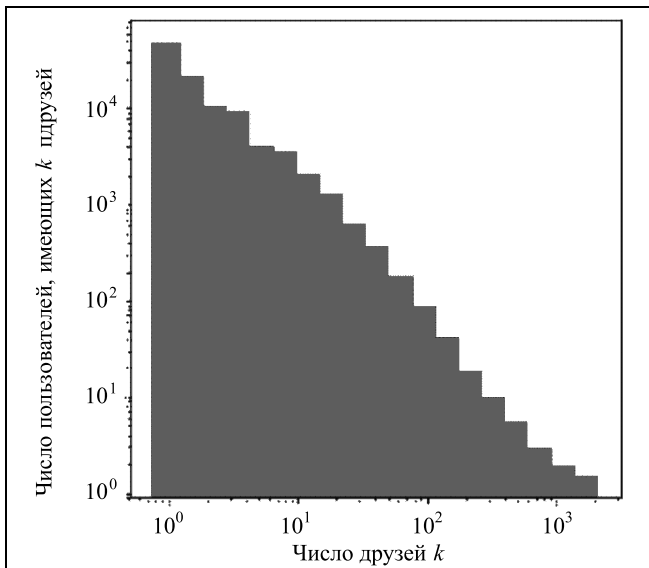


Рис. 4. Распределение числа степеней вершин графа $G[B]$: оси абсцисс и ординат представлены в логарифмическом масштабе

ты-признаки $\mathbf{X} = (\mathbf{X}_{PP} \mathbf{X}_{sub}) \in \{0, 1\}^{N \times 4\,917\,363}$. Соответствующий ей вектор классов обозначим через $\mathbf{y} \in \{1, 2, 3\}^N$. Матрица $\mathbf{X}_{PP} \in \{0, 1\}^{N \times 11}$ формализует ПП пользователей, число 11 — это число значений признака ПП (включая случай, когда пользователь не указал ПП). В свою очередь, матрица $\mathbf{X}_{sub} \in \{0, 1\}^{N \times 4\,917\,352}$ фиксирует подписки пользователей, число столбцов этой матрицы соответствует числу уникальных пабликов, групп, страниц событий и аккаунтов блогеров, на которых подписаны пользователи I_* . В силу построения каждый столбец матрицы $\mathbf{X}_{sub}$ имеет хотя бы один ненулевой элемент. Каждая строка матрицы $\mathbf{X}_{PP}$ также содержит хотя бы одну единицу.

В ходе экспериментов было выявлено, что аккаунты пользователей, у которых менее 11 и более 300 подписок, значительно снижают качество классификации (данные числа были определены опытным путем как максимизирующие точность классификации). Поэтому было принято решение не включать их ни в обучающую, ни в тестовую выборки.

Далее каждый аккаунт был «клонирован», т. е. был добавлен новый аккаунт (будем называть новые аккаунты *синтетическими*), отличающийся от исходного тем, что были обнулены компоненты в one-hot представлении, соответствующие подпискам на маркерные аккаунты. Данная мера позволяла провести более беспристрастную оценку качества классификации, поскольку в противном случае классификатор переобучился бы на маркерных аккаунтах и демонстрировал бы близкую к

единице верность классификации. Метки классов синтетических аккаунтов оставались неизменными. «Клонирование» изменило матрицу объекты-признаки $\mathbf{X} = (\mathbf{X}_{PP} \mathbf{X}_{sub})$ и вектор классов $\mathbf{y}$ (формально, это уже не те матрицы, что были введены ранее, поскольку, как уже было сказано, в результате отсеивания аномалий часть строк этих матриц была удалена, однако для удобства мы оставим те же обозначения) таким образом: матрица, описывающая реальных пользователей, была дополнена по строкам матрицей $\mathbf{X}_{trunc} = (\mathbf{X}_{PP} \mathbf{X}_{sub_trunc})$, а вектором $\mathbf{y}$ был дополнен таким же (т. е. продублирован). Матрицы $\mathbf{X}_{sub}$ и $\mathbf{X}_{sub_trunc}$ имеют одинаковую размерность и единственное различие между ними состоит в том, что у последней в столбцах, соответствующих маркерным аккаунтам, расположены только нули.

С помощью одного и того же случайного зерна $\text{random_state} = 42$ (которое дается на вход генератору псевдослучайных чисел) каждая из матриц $(\mathbf{X}_{PP} \mathbf{X}_{trunc} | \mathbf{y})$ и $(\mathbf{X}_{PP} \mathbf{X}_{sub_trunc} | \mathbf{y})$ была разбита на две подматрицы псевдослучайным образом в пропорции 2:1. Две трети строк каждой из матриц $(\mathbf{X}_{PP} \mathbf{X}_{sub} | \mathbf{y})$ и $(\mathbf{X}_{PP} \mathbf{X}_{sub_trunc} | \mathbf{y})$ были выбраны случайным образом — сгруппированы в матрицы $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub}^1 | \mathbf{y}^1)$ и $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub_trunc}^1 | \mathbf{y}^1)$, а оставшиеся строки каждой из матриц были сгруппированы в матрицы $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub}^2 | \mathbf{y}^2)$ и $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub_trunc}^2 | \mathbf{y}^2)$:

$$(\mathbf{X}_{PP} \mathbf{X}_{sub} | \mathbf{y}) \rightarrow (\mathbf{X}_{PP}^1 \mathbf{X}_{sub}^1 | \mathbf{y}^1), (\mathbf{X}_{PP}^2 \mathbf{X}_{sub}^2 | \mathbf{y}^2);$$

$$(\mathbf{X}_{PP} \mathbf{X}_{sub_trunc} | \mathbf{y}) \rightarrow (\mathbf{X}_{PP}^1 \mathbf{X}_{sub_trunc}^1 | \mathbf{y}^1), (\mathbf{X}_{PP}^2 \mathbf{X}_{sub_trunc}^2 | \mathbf{y}^2).$$

Применение совпадающих зерен для генерации разбиений гарантировало, что матрицы $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub}^1 | \mathbf{y}^1)$ и $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub_trunc}^1 | \mathbf{y}^1)$, а также матрицы $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub}^2 | \mathbf{y}^2)$ и $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub_trunc}^2 | \mathbf{y}^2)$ «соответствовали» друг другу в том смысле, что в каждой из этих двух пар i -я строка каждой из матриц соответствовала одному и тому же пользователю: для матриц $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub}^1 | \mathbf{y}^1)$ и $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub}^2 | \mathbf{y}^2)$ — «натуральному», а для матриц $(\mathbf{X}_{PP}^1 \mathbf{X}_{sub_trunc}^1 | \mathbf{y}^1)$ и $(\mathbf{X}_{PP}^2 \mathbf{X}_{sub_trunc}^2 | \mathbf{y}^2)$ — синтетическому.

В качестве обучающей выборки принимались данные

$$(\mathbf{X}_{train} | \mathbf{y}_{train}) = \begin{pmatrix} \mathbf{X}_{PP}^1 \mathbf{X}_{sub}^1 | \mathbf{y}^1 \\ \mathbf{X}_{PP}^1 \mathbf{X}_{sub_trunc}^1 | \mathbf{y}^1 \end{pmatrix},$$

Рисунок к статье [Л.Г. Бызова], Д.А. Губанова, И.В. Козицина, А.Г. Чхартишвили

**«ИДЕАЛЬНЫЙ ПОЛИТИК ДЛЯ СОЦИАЛЬНОЙ СЕТИ:
ПОДХОД К АНАЛИЗУ ИДЕОЛОГИЧЕСКИХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ»
(с. 15–26)**

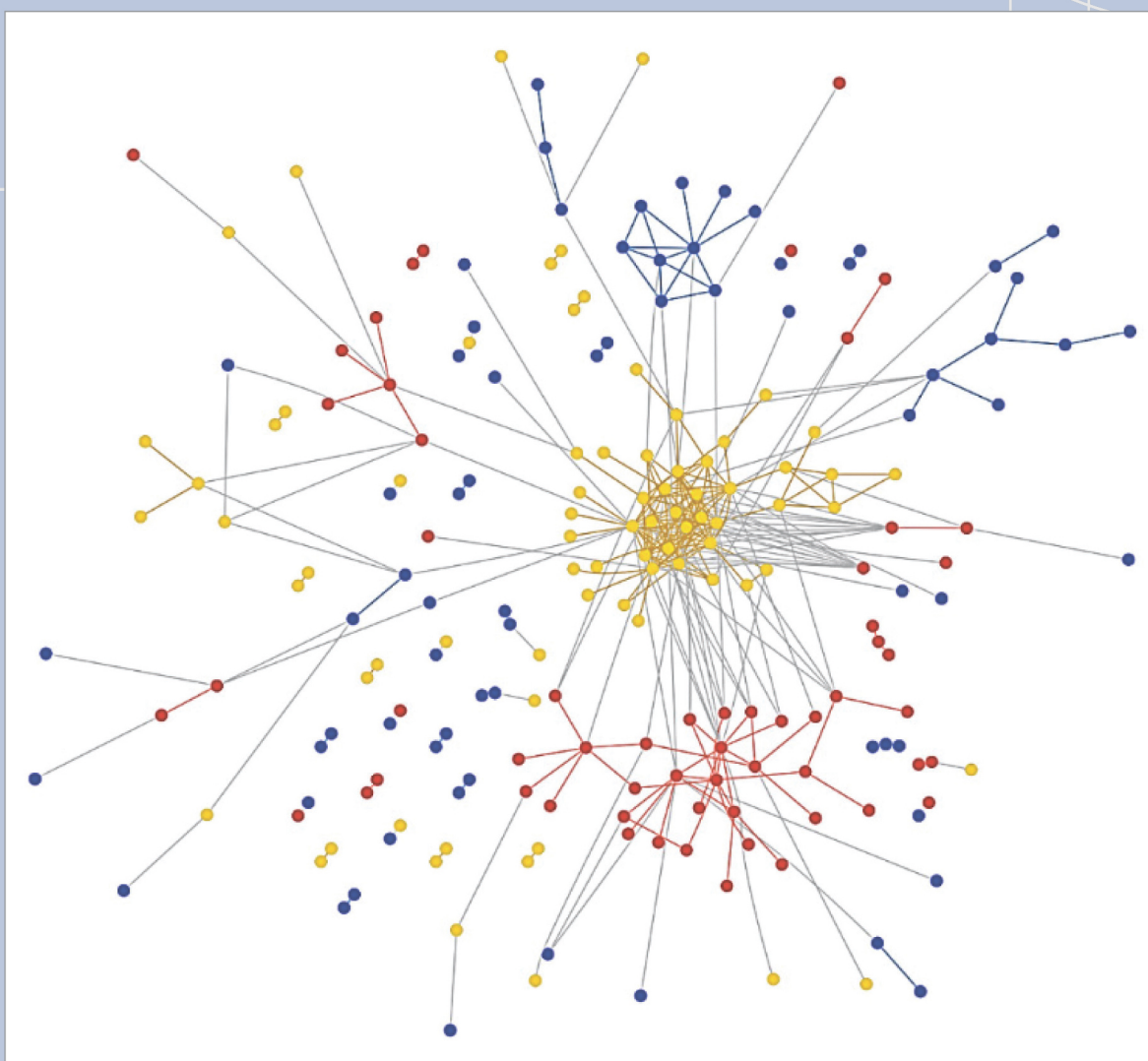


Рис. 5. Подграф графа $G[V]$, полученный отбрасыванием кросс-идеологических пересечений и случайным выбором вершин последнего. Число пользователей в подграфе 206. Желтый цвет соответствует пользователям, подписанным только на маркерные аккаунты державников, синий – либералов, красный – социалистов. Коэффициент ассортативности для подграфа 0,37

а в качестве тестовой — данные

$$(\mathbf{X}_{\text{test}} | \mathbf{y}_{\text{test}}) = (\mathbf{X}_{\text{PP}}^2 \mathbf{X}_{\text{sub_trunc}}^2 | \mathbf{y}^2).$$

В результате, качество классификации проверялось на синтетических пользователях, которые не были представлены в обучающей выборке ни в каком виде.

Матрица объекты-признаки обучающей выборки $\mathbf{X}_{\text{train}}$ применялась для обучения декомпозитора пространства (принадлежащего к классу TruncatedSVD, в качестве метода решения применялся алгоритм ARPACK), снижающий размерность последнего до 350 компонент (это число также было определено опытным путем как максимизирующее точность классификации). Далее обученный декомпозитор трансформировал матрицы $\mathbf{X}_{\text{train}}$ и $\mathbf{X}_{\text{test}}$, после чего эти матрицы вместе с векторами классов $\mathbf{y}_{\text{train}}$ и $\mathbf{y}_{\text{test}}$ использовались для обучения и тестирования классификатора. Наилучшее качество показал классификатор, основанный на статистической модели логистической регрессии с коэффициентом $C = 0,5$ L2-регуляризации, который применялся к задаче трехклассовой классификации по схеме one-versus-rest. Для решения проблемы несбалансированности классов каждому из классов был поставлен в соответствие вес (веса подбирались эмпирически: $D \rightarrow 1,5$, $L \rightarrow 1$, $S \rightarrow 1,1$), что в итоге снизило риск переобучения классификатора на мажоритарном классе.

Для оценивания качества классификации мы пользуемся показателем «средняя полнота». (Напомним, что полнота класса — это доля верно классифицированных представителей этого класса, средняя полнота — это значение показателя полноты класса, усредненное по всем классам.) Класс оцененного ДЛС-вектора (p_1, p_2, p_3) соот-

ветствует тому значению $i \in \{1, 2, 3\}$, для которого соответствующая компонента p_i принимает наибольшее значение из трех (если два наибольших значения одинаковы либо все три значения одинаковы, выбираем одно из них с равной вероятностью).

Матрица ошибок, демонстрирующая качество работы классификатора на тестовой выборке, представлена на рис. 6. Средняя полнота классификации составляет 0,674, что более, чем в два раза превышает точность при случайной классификации (с учетом несбалансированности классов).

3. ПОИСК ОПТИМАЛЬНОЙ ИДЕОЛОГИЧЕСКОЙ ПОЗИЦИИ

3.1. Постановка и решение оптимизационной задачи

В предыдущих разделах было показано, как можно оценить идейно-политические предпочтения пользователей сети ВКонтакте. Понятно, что эту информацию можно использовать для различных целей. В данной работе мы остановимся на одной из возможных: рассмотрим ситуацию, когда политик (или политическая партия) хочет занять идеологическую позицию, максимально близкую к позициям некоторого подмножества пользователей. Для случая двух политических направлений эта задача была исследована в классической работе Даунса [20]. В настоящей статье мы в основном рассматриваем случай трех направлений, однако предложенный в данном пункте метод может быть применен и в случае произвольного числа направлений k .

Для формализации задачи нахождения «идеальной» идеологической позиции требуется для начала уточнить, что понимается под расстоянием между двумя идеологическими позициями, задаваемыми стохастическими векторами $p_1 = (p_{11}, \dots, p_{1k})$ и $p_2 = (p_{21}, \dots, p_{2k})$. Мы будем пользоваться следующим расстоянием (см. [16, 21]):

$$d(p_1, p_2) = 1 - \sum_{j=1}^k \min(p_{1j}, p_{2j}) = \frac{1}{2} \sum_{j=1}^k |p_{1j} - p_{2j}|. \quad (1)$$

Отметим, что справедливость второго равенства в выражении (1) для стохастических векторов легко показать:

$$\begin{aligned} \frac{1}{2} \sum_{j=1}^k |p_{1j} - p_{2j}| &= \frac{1}{2} \sum_{j=1}^k (\max(p_{1j}, p_{2j}) - \min(p_{1j}, p_{2j})) \\ &= \frac{1}{2} \left(2 - \sum_{j=1}^k \min(p_{1j}, p_{2j}) - \sum_{j=1}^k \min(p_{1j}, p_{2j}) \right) = \\ &= 1 - \sum_{j=1}^k \min(p_{1j}, p_{2j}). \end{aligned}$$

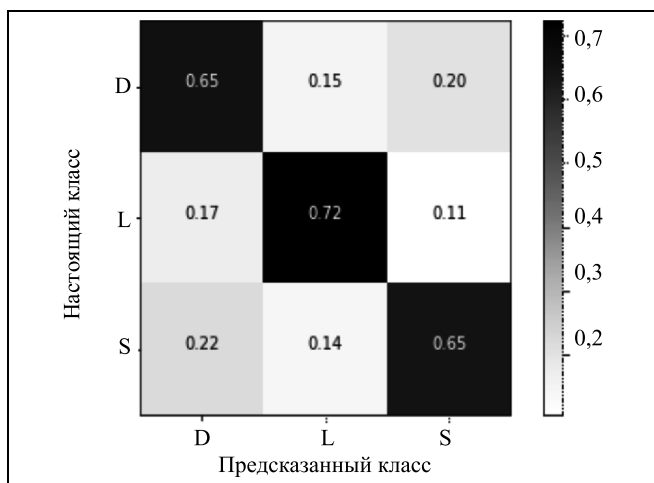


Рис. 6. Матрица ошибок, визуализирующая качество классификации обученной предсказательной модели: средняя полнота 0,674

Метрика (1) является частным случаем хорошо известного в теории вероятностей расстояния общей вариации. Рассмотрим на примере вопрос, в чем ее преимущество в предлагаемой модели по сравнению с более традиционной для расчетов в конечномерных пространствах евклидовой метрикой

$$d_E(p_1, p_2) = \sqrt{\sum_{j=1}^k (p_{1j} - p_{2j})^2}.$$

Предположим, что имеются четыре вектора взглядов (принадлежащих четырем индивидам — первому, второму, третьему и четвертому соответственно): $p_1 = (1, 0, 0, 0)$, $p_2 = (0, 1, 0, 0)$, $p_3 = (0, \frac{1}{2}, \frac{1}{2}, 0)$, $p_4 = (0, \frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Первый индивид полностью отдает приоритет ценностям, отраженным первой компонентой вектора (равной 1); ценности, отраженные в прочих компонентах, для него неважны. Поэтому второй, третий и четвертый индивиды одинаково далеки от него, поскольку для них, наоборот, неважны ценности первой компоненты. Это соответствует соотношению $d(p_1, p_2) = d(p_1, p_3) = d(p_1, p_4)$, в справедливости которого легко убедиться. Поэтому метрика (1) представляется более адекватной, чем евклидова метрика, в которой расстояния между этими же векторами попарно различны.

Сформулируем задачу поиска оптимальной идеологической позиции $a = (a_1, \dots, a_k)$ для заданного множества k -мерных векторов $p_1, \dots, p_n$:

$$L(a_1, \dots, a_k) = \sum_{i=1}^n d(p_i, a) \xrightarrow{a} \min, \quad (2)$$

$$a_j \geq 0, \quad j \in K, \quad \sum_{j=1}^k a_j = 1.$$

Отметим, что задача (2) — это, по сути, задача нахождения многомерной медианы множества точек (точнее говоря, одной из многомерных медиан — см. статью [22]). В случае евклидовой метрики эта задача довольно сложная и поддается лишь численному решению [23]. Однако для метрики (1), как будет показано далее, существует алгоритм получения точного решения.

Для решения задачи (2) сначала переупорядочим по неубыванию числа p_{ij} , $i \in N = \{1, \dots, n\}$, для каждого $j \in K = \{1, \dots, k\}$. Переобозначим эти числа через x_{ij} :

$$\forall j \in K \quad x_{1j} \leq \dots \leq x_{nj}. \quad (3)$$

Задачу (2) можно переписать в эквивалентном виде:

$$L(a_1, \dots, a_k) = \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^k |x_{ij} - a_j| \xrightarrow{a} \min, \quad (4)$$

$$a_j \geq 0, \quad j \in K, \quad \sum_{j=1}^k a_j = 1. \quad (5)$$

Решение задачи (4) (и, значит, задачи (2)) существует в силу непрерывности функции $L(a_1, \dots, a_k)$ и компактности множества (5). Его можно получить, опираясь на следующие два утверждения.

Утверждение 1. Пусть для чисел a_q и a_r выполнены условия

$$x_{iq} \leq a_q < x_{i+1,q}, \quad (6)$$

$$x_{mr} < a_r \leq x_{m+1,r}, \quad (7)$$

для некоторых $i, m \in N$, $i < m$. Тогда содержащий компоненты a_q и a_r вектор a не может быть решением задачи (4), (5).

Доказательство. Пусть оптимальным вектором — решением задачи (4), (5) является вектор a , удовлетворяющий условиям (5) и содержащий компоненты a_q и a_r , удовлетворяющие условиям (6), (7). Обозначим через $L_1 = L(a)$ значение целевой функции задачи (4) на этом векторе.

Увеличим a_q на величину $\Delta > 0$, а a_r уменьшим на ту же величину Δ , притом выберем Δ достаточно малым для того, чтобы числа $a_q + \Delta$ и $a_r - \Delta$ не вышли за границы интервалов (6) и (7) соответственно (для определенности можно взять $\Delta = 0,5 \min\{x_{i+1,q} - a_q, a_r - x_{mr}\}$). В результате получим новый вектор (удовлетворяющий ограничениям (5)) и новое значение целевой функции, которое обозначим через L_2 .

Посчитаем (для удобства удвоенную) разность $L_2 - L_1$ с учетом неравенств

$$x_{1q} \leq \dots \leq x_{iq} \leq a_q < a_q + \Delta < x_{i+1,q} \leq \dots \leq x_{nq},$$

$$x_{1r} \leq \dots \leq x_{mr} < a_r - \Delta < a_r \leq x_{m+1,r} \leq \dots \leq x_{nr},$$

а также того факта, что в двойной сумме в выражении (4) изменились только слагаемые, содержащие компоненты a_q и a_r :

$$\begin{aligned} 2(L_2 - L_1) &= \sum_{j=1}^n (|x_{jq} - (a_q + \Delta)| + |x_{jr} - (a_r - \Delta)|) - \\ &- \sum_{j=1}^n (|x_{jq} - a_q| + |x_{jr} - a_r|) = \sum_{j=1}^i ((-x_{jq} + a_q + \Delta) - \\ &- (-x_{jq} - a_q)) + \sum_{j=i+1}^n ((x_{jq} - a_q - \Delta) - (x_{jq} - a_q)) + \\ &+ \sum_{j=1}^m ((-x_{jr} + a_r - \Delta) - (-x_{jr} + a_r)) + \\ &+ \sum_{j=m+1}^n ((x_{jr} - a_r + \Delta) - (x_{jr} - a_r)) = \\ &= i\Delta - (n-i)\Delta - m\Delta + (n-m)\Delta = 2(i-m)\Delta < 0. \end{aligned}$$

Таким образом, в результате изменения аргумента целевая функция задачи (4) уменьшилась, что противоречит оптимальности исходного вектора a . ♦

Утверждение 2. Пусть для векторов $a = (a_1, \dots, a_k)$ и $b = (b_1, \dots, b_k)$ выполнены ограничения (5), а также условия для некоторого $i \in N$:

$$\begin{aligned} x_{ij} &\leq a_j \leq x_{i+1,j}, \\ x_{ij} &\leq b_j \leq x_{i+1,j} \end{aligned} \quad (8)$$

для всех $j \in K$. Тогда значения целевой функции L задачи (4) на этих векторах совпадают: $L(a) = L(b)$.

Доказательство. Покажем, что разность $L(a) - L(b)$ равна нулю:

$$\begin{aligned} L(a) - L(b) &= \frac{1}{2} \sum_{j=1}^k \left(\sum_{m=1}^i (a_j - x_{mj}) + \sum_{m=i+1}^n (x_{mj} - a_j) \right) - \\ &\quad - \frac{1}{2} \sum_{j=1}^k \left(\sum_{m=1}^i (b_j - x_{mj}) + \sum_{m=i+1}^n (x_{mj} - b_j) \right) = \\ &= \frac{1}{2} \sum_{j=1}^k (a_j - b_j) (i - (n - i)) = \\ &= \left(i - \frac{n}{2} \right) \left(\sum_{j=1}^k a_j - \sum_{j=1}^k b_j \right) = 0. \quad \spadesuit \end{aligned}$$

Заметим теперь, что все вектора $a = (a_1, \dots, a_k)$, удовлетворяющие условиям (5), разбиваются на два непересекающихся подмножества: для них условия (6) и (7) либо выполняются либо не выполняются. Векторы из первого подмножества, в соответствии с утверждением 1, не могут быть оптимальными. Следовательно, оптимальный вектор (решение задачи (4), (5)) следует искать во втором подмножестве, для векторов из которого выполняется условие (8) для некоторого $i \in N$. Для нахождения решения следует найти (например, путем последовательного перебора по i) значение $i \in N$, для которого выполняется условие

$$\sum_{j=1}^k x_{ij} \leq 1 \leq \sum_{j=1}^k x_{i+1,j} \quad (9)$$

В силу условия (3) сумма в левой части неравенства (9) не убывает по i , поэтому такое значение существует и единственное — за исключением случаев, когда имеет место двойное равенство

$$\sum_{j=1}^k x_{ij} = 1 = \sum_{j=1}^k x_{i+1,j} \quad (10)$$

Если в условии (9) оба неравенства для некоторого i обращаются в равенства (т. е. выполняется (10)), то единственным решением задачи (3) является вектор

$$a_j = x_{ij}, \quad j \in K. \quad (11)$$

В противном случае имеется множество векторов, в которых, согласно утверждению 2, целевая функция принимает одно и то же значение; для определенности предлагается считать решением вектор

$$a_j = (1 - \gamma)x_{ij} + \gamma x_{i+1,j}, \quad j \in K. \quad (12)$$

где γ — решение уравнения $\sum_{j=1}^k a_j = 1$:

$$\gamma = \frac{1 - \sum_{j=1}^k x_{ij}}{\sum_{j=1}^k x_{i+1,j} - \sum_{j=1}^k x_{ij}}. \quad (13)$$

Итак, решение задачи (2) имеет вид (12), (13) или (11), где i определяется соотношением (9). ♦

3.2. Примеры

Приведем примеры расчета «идеальных» политических позиций для четырех пабликов онлайн-социальной сети ВКонтакте:

- «Православная держава» (<https://vk.com/club63342269>);
- «СССР Ленин Сталин Марксизм Коммунизм Коммунисты» (<https://vk.com/public53550079>);
- «Новости RT на русском» (https://vk.com/rt_russian);
- «Медуза» (<https://vk.com/meduzaproject>).

Для каждого из пабликов был зафиксирован перечень пользователей — их подписчиков, далее из перечней были исключены пользователи, имеющие менее 11 подписок. Рассмотрим отдельно для каждого паблика распределение идейно-политических предпочтений пользователей.

Пример 1. Паблик «Православная держава». Рассмотрены предпочтения 23 749-ти пользователей. На рис. 7 (см. первую страницу вклейки) представлены тепловые карты (heatmap) на симплексе идейно-политических предпочтений пользователей. Каждая точка на симплексе характеризует соотношение между компонентами вектора идейно-политической позиции. Координаты точки можно определить при помощи линий, параллельных сторонам симплекса. Интенсивностью цвета показаны частоты встречаемости пользователей, попадающих в шестиугольники симплекса. «Идеальной» идеологической позицией паблика (или оптимальным ДЛС-вектором) является вектор $(0,5151, 0,3497, 0,1352)$, который обозначен на рис. 7 красным плюсом. Среднее расстояние позиций пользователей до идеальной позиции составляет 0,2684. Исходя из рис. 7, можно отметить, что пользователи паблика придерживаются в большей степени державных позиций, в меньшей — либеральных.

Пример 2. Паблик «СССР Ленин Сталин Марксизм Коммунизм Коммунисты». Рассмотрены предпочтения 6179-ти пользователей. На рис. 8 (см. первую страницу

**«ИДЕАЛЬНЫЙ ПОЛИТИК ДЛЯ СОЦИАЛЬНОЙ СЕТИ:
ПОДХОД К АНАЛИЗУ ИДЕОЛОГИЧЕСКИХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ»
(с. 15–26)**

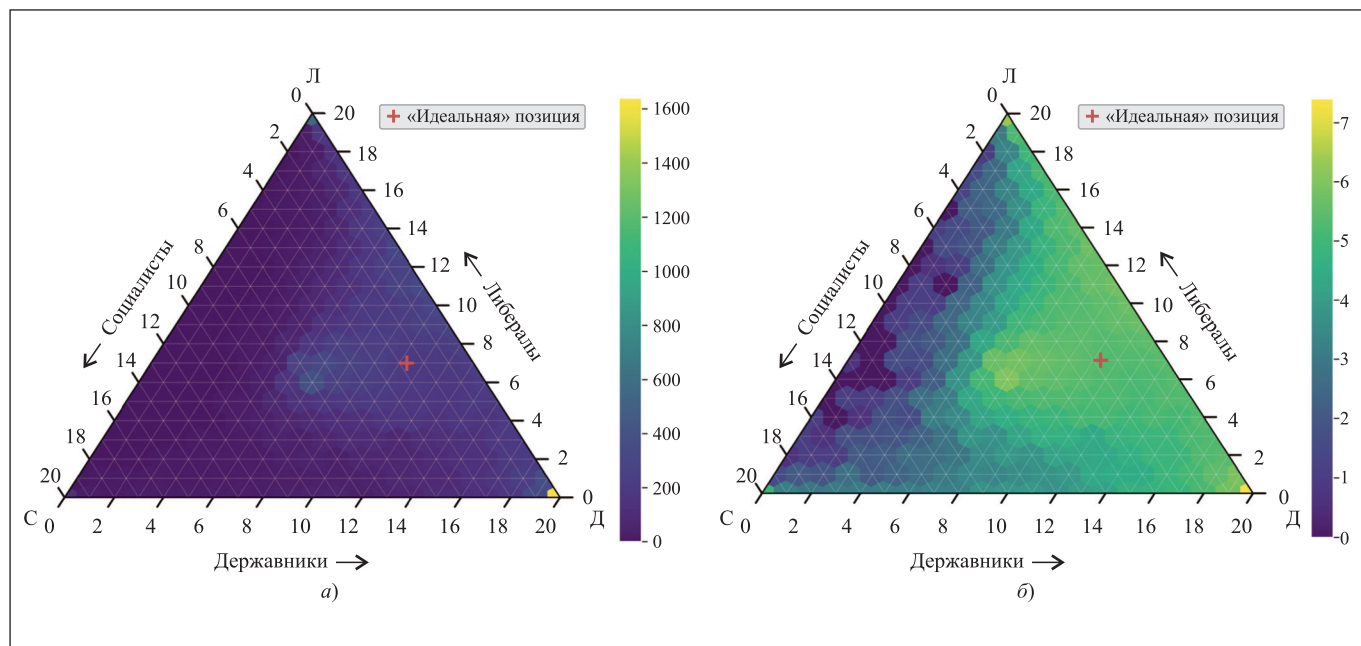


Рис. 7. Тепловая карта предпочтений подписчиков «Православная держава»:
а — частоты; *б* — логарифм частот

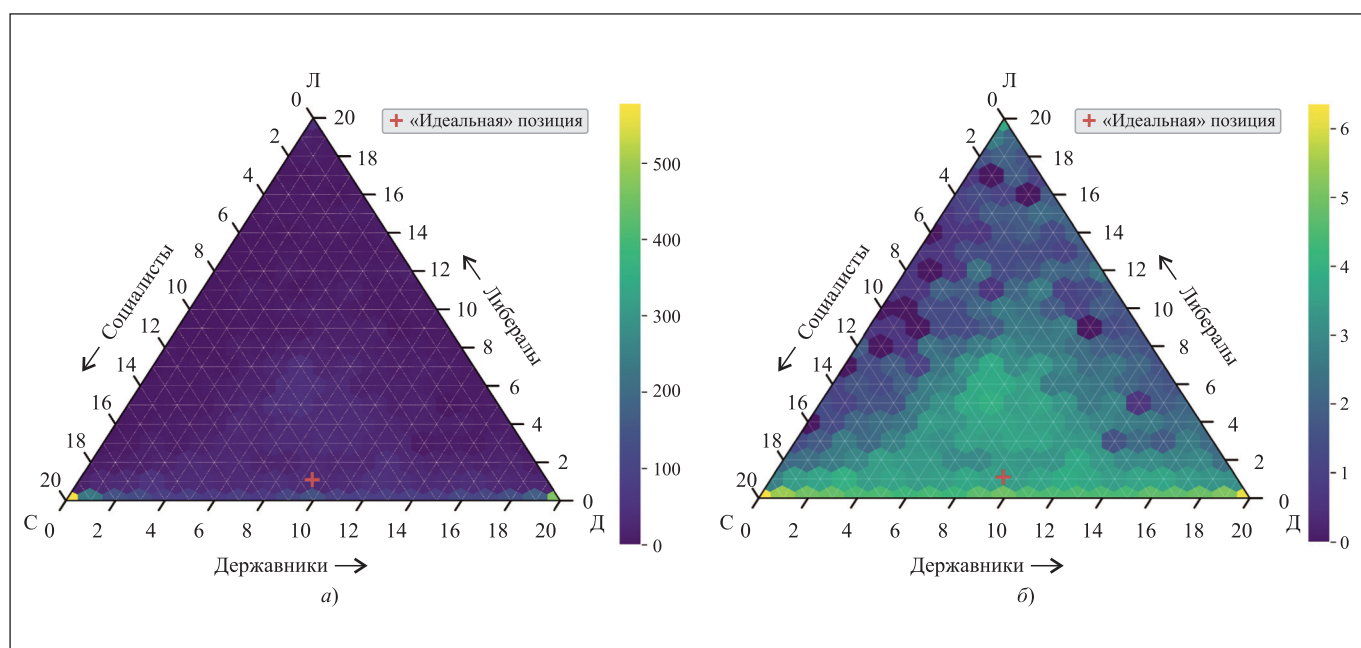


Рис. 8. Тепловая карта предпочтений подписчиков «СССР»:
а — частоты; *б* — логарифм частот

вклейки) представлены тепловые карты предпочтений пользователей. Идеальной идеологической позицией паблика является вектор **(0,4708, 0,0549, 0,4743)**. Среднее расстояние позиций пользователей до идеальной позиции составляет 0,3341. Максимальная концентрация пользователей достигается в полюсах, соответствующих социалистам и державникам. В целом можно сказать, что пользователи паблика придерживаются державных и социалистических позиций.

Пример 3. Паблик «Новости RT на русском». Рассмотрены предпочтения 129 141-го пользователя. На рис. 9 (см. вторую страницу вклейки) представлены тепловые карты предпочтений. Идеальной идеологической позицией паблика служит вектор **(0,4596, 0,1537, 0,3866)**. Среднее расстояние позиций пользователей до идеальной позиции составляет 0,2341. Пользователи паблика придерживаются в большей степени державной и в меньшей степени социалистической позиции.

Пример 4. Паблик «Медуза». Для паблика «Медуза» рассмотрены предпочтения 134 705-ти пользователей. На рис. 10 (см. вторую страницу вклейки) представлены тепловые карты предпочтений пользователей. Идеальной идеологической позицией паблика служит вектор **(0,2462, 0,5135, 0,2403)**. Среднее расстояние позиций пользователей до идеальной позиции составляет 0,2321. В целом можно сказать, что пользователи паблика придерживаются в большей степени либеральной позиции.

ЗАКЛЮЧЕНИЕ

Решение задачи определения идейно-политических представлений пользователей онлайн-социальной сети может найти различные применения. В настоящей работе внимание акцентировано на одном из них: нахождение оптимальной (в смысле минимизации суммарного расстояния до пользователей социальной сети) позиции политика или политической партии. Для решения этой задачи были предприняты следующие шаги: определение идейно-политического базиса (ДЛС-модель), подбор характеризующих этот базис направлений популярных сетевых сообществ (маркерных аккаунтов), решение задачи трехклассовой классификации для нахождения ДЛС-векторов пользователей, решение задачи минимизации суммарного расстояния.

Довольно высокое качество классификации на тестовой выборке косвенно свидетельствует в пользу адекватности выбранного нами идеологического базиса. При этом предлагаемый подход универсален и может быть адаптирован к базисам иного состава и иной размерности (см., например, [9]).

В заключение отметим, что в дальнейших исследованиях мы планируем воспользоваться построенной ДЛС-моделью для анализа динамики мнений пользователей сети ВКонтакте. В частности, интерес представляет вопрос, насколько хорошо формальные модели динамики мнений согла-

суются с динамикой ДЛС-векторов пользователей [24–26]. Кроме того, ДЛС-модель позволит проверить гипотезы о наличии в сети ВКонтакте таких явлений, как поляризация [27] и эхо-камеры [28].

ЛИТЕРАТУРА

1. *Gayo-Avello, D.* A meta-analysis of state-of-the-art electoral prediction from Twitter data // *Social Science Computer Review*. — 2013. — Vol. 31, no. 6. — P. 649–679.
2. *Petrov, A., Proncheva, O.* Modeling Propaganda Battle: Decision-Making, Homophily, and Echo Chambers / *Conf. on Artificial Intelligence and Natural Language*. — Springer, Cham, 2018. — P. 197–209.
3. *Chkhartishvili, A.G., Gubanov, D.A., Novikov, D.A.* Social Networks: Models of information influence, control and confrontation. — Springer, 2018. — 158 p.
4. *Petrov, A.P., Proncheva, O.G.* Modeling Position Selection by Individuals during Informational Warfare with a Two-Component Agenda // *Mathematical Models and Computer Simulations*. — 2020. — Vol. 12, no. 2. — P. 154–163.
5. *Barberà, P.* Birds of the Same Feather Tweet Together: Bayesian Ideal Point Estimation Using Twitter Data // *Political Analysis*. — 2015. — Vol. 23, no. 1. — P. 76–91.
6. *Makazhanov, A., Rafiei, D., Waqar, M.* Predicting Political Preference of Twitter Users // *Social Network Analysis and Mining*. — 2014. — Vol. 4, no. 1. — P. 193.
7. *Pennacchiotti, M., Popescu, A.M.* Democrats, republicans and starbucks aficionados: user classification in twitter // *Proc. of the 17th ACM SIGKDD Intern. Conf. on Knowledge Discovery and Data Mining*. — 2011. — P. 430–438.
8. *Бызов Л.Г.* Динамика идейно-политических предпочтений за 25 лет. Три этапа трансформации общественного сознания // *Россия XXI*. — 2019. — № 2. — С. 6–21. [*Byzov, L.G.* Dinamika ideino-politicheskikh predpochtenii za 25 let. Tri etapa transformacii obshchestvennogo soznaniia // *Rossia XXI*. — 2019. — No. 2. — S. 6–21. (In Russian)]
9. *Kozitsin, I.V., Chkhartishvili, A.G., Marchenko, A.M., et al.* Modeling Political Preferences of Russian Users Exemplified by the Social Network Vkontakte // *Mathematical Models and Computer Simulations*. — 2020. — Vol. 12, no. 2. — P. 185–194.
10. *Kozitsin, I.V., Marchenko, A.M., Goiko, V.L., et al.* Symmetric Convex Mechanism of Opinion Formation Predicts Directions of Users' Opinions Trajectories // *Twelfth Intern. Conf. «Management of Large-scale System Development» (MLSD)*. — IEEE, 2019.
11. *Rokeach, M.* The Nature of Human Values. — NY: Free Press, 1973.
12. *Фелдман С.* Ценности, идеология и структура политических установок / *Политическая психология. Хрестоматия по ред. Е.Б. Шестопал*. — М.: Аспект Пресс, 2007. — С. 256–260. [*Feldman, S.* Tsennosti, ideologiya i struktura politicheskikh ustanovok / *Politicheskaya psikhologiya. Khrestomatia po red. E.B. Shestopal*. — M.: Aspekt Press, 2007. — S. 256–260. (In Russian)]
13. *Inglehart, R.* The silent revolution. — Princeton: Princeton University Press, 1977.
14. *Inglehart, R.* Cultural shift in advanced industrial society. — Princeton: Princeton University Press, 1990.
15. *Блинов В.В.* Типы идеологических сторонников в современной России // *Мониторинг общественного мнения: экономические и социальные перемены*. — 2010. — № 6 (100). [*Blinov, V.V.* Tipy ideologicheskikh storonnikov v sovremennoi Rossii // *Monitoring obshchestvennogo mneniya: ekonomicheskie i sotsialnye peremeny*. — 2010. — No. 6 (100). (In Russian)]

**«ИДЕАЛЬНЫЙ ПОЛИТИК ДЛЯ СОЦИАЛЬНОЙ СЕТИ:
ПОДХОД К АНАЛИЗУ ИДЕОЛОГИЧЕСКИХ ПРЕДПОЧТЕНИЙ ПОЛЬЗОВАТЕЛЕЙ»
(с. 15–26)**

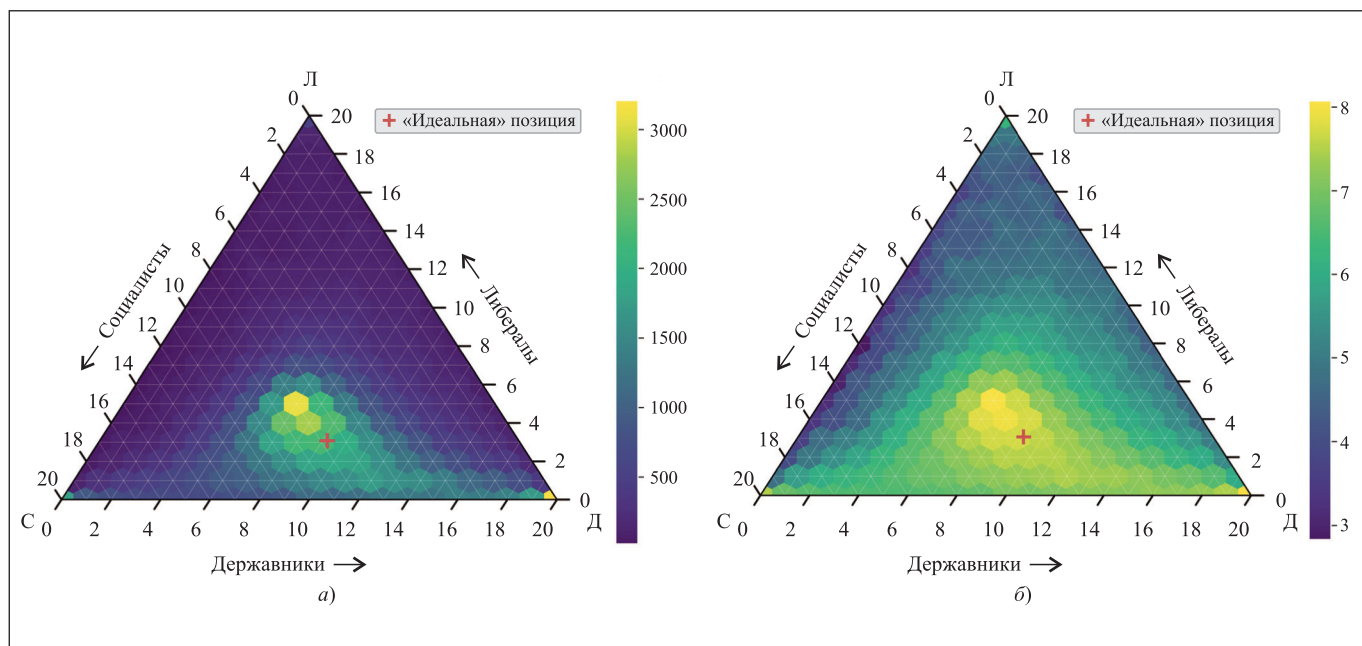


Рис. 9. Тепловая карта предпочтений подписчиков «RT»:

а — частоты; *б* — логарифм частот

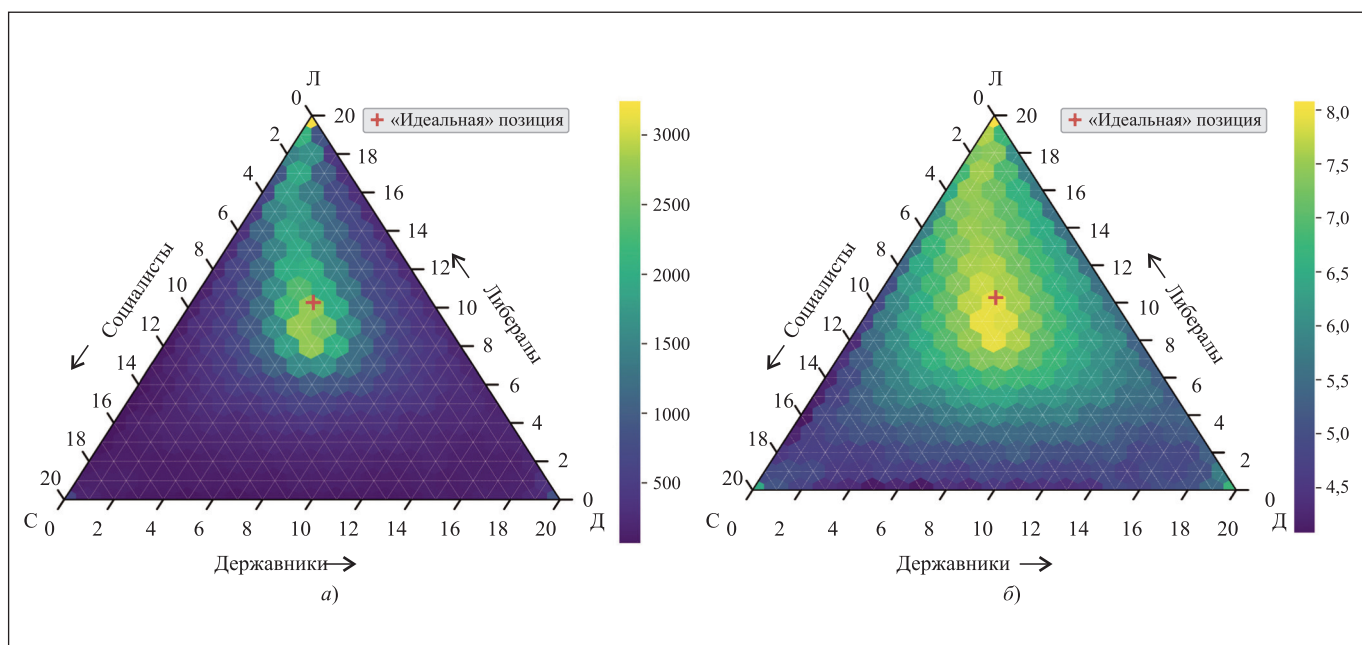


Рис. 10. Тепловая карта предпочтений подписчиков «Медуза»:

а — частоты; *б* — логарифм частот

16. Козицин И.В., Чхартишвили А.Г. Об одной модели идейно-политических предпочтений пользователей онлайн-социальной сети // Теория активных систем — 50 лет / Материалы междунар. науч.-практ. конф., 18–19 ноября 2019 г., Москва, ИПУ РАН. Под общ. ред. В.Н. Буркова. — С. 520–527. [Kozitsin, I.V., Chkhartishvili, A.G. Ob odnoi modeli ideino-politicheskikh predpochtenii polzovatelei onlain-ovoi sotcialnoi seti // Teoriia aktivnykh sistem — 50 let / Materialy mezhhdunar. nauch.-prakt. konf., 18–19 noiabria 2019 g., Moskva, IPU RAN. Pod obshch. red. V.N. Burkova. — S. 520–527. (In Russian)]
 17. URL: <https://www.kommersant.ru/doc/3792003>
 18. Newman, M. Networks. — Oxford: Oxford University Press, 2018.
 19. Flach, P. Machine Learning: The Art and Science of Algorithms that Make Sense of Data. — Cambridge: Cambridge University Press, 2012.
 20. Downs, A. An Economic Theory of Political Action in a Democracy // Journal of Political Economy. — 1957. — Vol. 65, no. 2. — P. 135–150.
 21. Gubanov, D., Petrov, I. Multidimensional Model of Opinion Polarization in Social Networks // Twelfth Intern. Conf. «Management of Large-scale System Development» (MLSD). — IEEE, 2019.
 22. Small, C.G. A Survey of Multidimensional Medians // International Statistical Review. — 1990. — No. 58. — 1990. — P. 263–277.
 23. Croux, C., Filzmoser, P., Fritz, H. A Comparison of Algorithms for the Multivariate L1-median // Computational Statistics. — 2010. — Vol. 27, no 3. — P. 1–18.
 24. Kozitsin, I.V., Belolipetskii, A.A. Opinion Convergence in the Krasnoshchekov Model // The Journal of Mathematical Sociology. — 2019. — Vol. 43, no. 2. — P. 104–121.
 25. Proskurnikov, A.V., Tempo, R. A Tutorial on Modeling and Analysis of Dynamic Social Networks. Part I // Annual Reviews in Control. — 2017. — Vol. 43. — P. 65–79.
 26. Fedyanin, D., Giliyazova, A. Influence of Deactivated Agents in Social Networks: Switching Between French-De Groot Models and Friedkin-Johnsen Model // Twelfth Intern. Conf. «Management of Large-scale System Development» (MLSD). — IEEE, 2019.
 27. Chhartishvili, A.G., Kozitsin, I.V., Goiko, V.L., Saifulin, E.R. On an Approach to Measure the Level of Polarization of Individuals' Opinions // Ibid.
 28. Chkhartishvili, A., Kozitsin, I. Binary Separation Index for Echo Chamber Effect Measuring // Ibid.
- Статья представлена к публикации членом редколлегии чл.-корр. РАН Д.А. Новиковым.*
- Поступила в редакцию 2.03.2020, после доработки 26.03.2020.
Принята к публикации 28.03.2020.*

Бызов Леонтий Георгиевич

— канд. экон. наук,
Федеральный научно-исследовательский социологический центр РАН, г. Москва,

Губанов Дмитрий Алексеевич — канд. техн. наук,
Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва, ✉ dmitry.a.g@gmail.com,

Козицин Иван Владимирович — мл. науч. сотрудник,
Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва, ✉ kozitsin.ivan@mail.ru,

Чхартишвили Александр Гедванович — д-р физ.-мат. наук,
Институт проблем управления им. В.А. Трапезникова РАН,
г. Москва, ✉ sandro_ch@mail.ru.

PERFECT POLITICIAN FOR SOCIAL NETWORK: AN APPROACH TO ANALYSIS OF IDEOLOGICAL PREFERENCES OF USERS

[L.G. Byzov]¹, D.A. Gubanov², I.V. Kozitsin³, A.G. Chkhartishvili⁴

¹ Federal Center of Theoretical and Applied Sociology of Russian Academy of Sciences

^{2,3,4} V.A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia

² ✉ dmitry.a.g@gmail.com, ³ ✉ kozitsin.ivan@mail.ru, ⁴ ✉ sandro_ch@mail.ru

Abstract. In this paper, we consider the problem of determining politico-ideological preferences of users of online social networks. We propose a model (DLS model) that allows assessing politico-ideological preferences of VKontakte users by using information taken from their accounts (digital footprints). The model incorporates the main ideological directions in modern-day Russia. Our approach is based on the supervised learning methodology whereby one solves a classification problem by calculating posterior probabilities of class membership. We make an anonymized labeled dataset and develop corresponding software. Then we formulate and solve the problem of the politician's choice of an ideological positioning strategy, that will potentially receive the most support from the considered set of users. We exemplify our methodology by finding «ideal» political positioning for some popular online communities in VKontakte.

Keywords: online social networks, ideological views, DLS-model, big data, finding the optimal ideological position.

Funding. The work was performed with financial support of Russian Foundation of Basic Research (project no. 18-29-22042).