ВЛИЯНИЕ АЛГОРИТМОВ РАНЖИРОВАНИЯ, БОТОВ И МОДЕРАЦИИ КОНТЕНТА НА ФОРМИРОВАНИЕ МНЕНИЙ В СОЦИАЛЬНОЙ СЕТИ[1]

Губанов Д. А.[2], Чхартишвили А. Г.[3]
(*ФГБУН Институт проблем управления им. В.А. Трапезникова РАН, Москва*)

Рассматривается модель формирования информационных каскадов в онлайновых социальных сетях, учитывающая влияние алгоритмов ранжирования контента, действий ботов и модерации контента. Особое внимание уделено динамике мнений, которая критически важна для прогнозирования и управления социальными процессами. В отличие от традиционных моделей, здесь мнения агентов (пользователей) не наблюдаемы напрямую: их действия, такие как публикация комментариев, служат косвенными индикаторами взглядов. Эти действия влияют на мнения других пользователей, приводя к формированию информационного каскада в сети. Модель дополнена такими факторами, как алгоритмы показа комментариев, поведение ботов и модерация контента администратором информационного ресурса. Вычислительные эксперименты показывают, что алгоритмы ранжирования существенно влияют на динамику мнений и действий, особенно при ограниченной глубине просмотра пользователей. Кроме того, введение ботов и модерации может существенно изменить ход обсуждений. В работе исследуется взаимодействие стратегических игроков, включая модератора и ботов с противоположными позициями, и прогнозируется результат их взаимодействия на основе равновесий Нэша. Наконец, формализована и решена для частного случая задача управляющего органа, который, стремясь продвинуть нужную ему точку зрения в сети, осуществляет влияние на количество ботов.

Ключевые слова: онлайновые социальные сети, алгоритмы социальной сети, формирование мнений пользователей, боты, модерация контента, информационное противоборство, имитационное моделирование.

---

[1] *Работа выполнена при финансовой поддержке Российского научного фонда в рамках проекта №23-21-00408.*
[2] *Дмитрий Алексеевич Губанов, д.т.н., в.н.с.* (*dmitry.a.g@gmail.com*).
[3] *Александр Гедеванович Чхартишвили, д.ф.-м.н., г.н.с.* (*sandro_ch@mail.ru*).

1. Введение

В последние десятилетия большой интерес теоретиков и практиков привлекают социальные сети. Выявление и прогнозирование динамики предпочтений пользователей онлайновых социальных сетей имеет огромную важность при моделировании информационного управления и информационного противоборства [10]. Эти предпочтения могут иметь социально-политическую, экономическую, психологическую или какую-либо другую природу.

Одним из методов исследования динамики мнений в социальных сетях является имитационное моделирование. Оно основано на задании на микроуровне правила изменения мнения пользователей (*агентов*) в зависимости от наблюдаемых ими действий тех соседних узлов сети – агентов, с которыми имеется та или иная связь (см., например, [16]). Отметим, что альтернативным подходом является моделирование динамики на макроуровне, например при помощи системы дифференциальных уравнений [17].

При моделировании динамики мнений в социальных сетях традиционно считается, что мнение и действие агента (индивида, являющегося узлом сети) отождествляются – см., например, [11, 12]. Применительно к онлайновым социальным сетям это означает, что агент (в данном случае – пользователь сети) без искажения транслирует свое внутреннее состояние и другие агенты имеют возможность это состояние наблюдать. В последнее десятилетие ситуация меняется в сторону разработки более сложных и реалистичных моделей [1, 4, 6, 14, 15]. Кроме того, предлагаются модели, в которых рассматривается влияние средств коммуникации на информационные процессы в сетях [13, 18].

В данной работе развивается ранее предложенная авторами модель [2, 3, 5, 7, 8] именно такого класса, где мнения (или предпочтения) агентов не наблюдаемы, а наблюдаемые действия не полностью отражают их мнения. Модель дополнена факторами алгоритма показа комментариев пользователям социальной сети, действий виртуальных агентов (ботов)

и модерации контента со стороны администратора страницы информационного источника.

В первом разделе статьи описана модель формирования информационного каскада и стратегические игроки, влияющие на него. Во втором разделе исследовано влияние ранжирования комментариев со стороны сети. В третьем разделе исследовано влияние ботов и модерации контента со стороны администратора. Наконец, в четвертом разделе рассмотрена возможность влияния управляющего органа, который может влиять на количество ботов, являющееся важным параметром ситуации информационного противоборства.

2. Формирование информационного каскада и стратегические игроки

В данном разделе мы опишем, следуя [9], модель формирования последовательности комментариев к сообщению (посту) в социальной сети.

Сначала приведем краткое вербальное описание. Будем считать, что имеется фиксированное множество пользователей социальной сети, являющихся подписчиками информационного источника. В информационном источнике публикуется пост, который рано или поздно увидят все пользователи-подписчики. У каждого пользователя есть мнение, которое он корректирует, прочитав часть ранее оставленных комментариев (отметим, что сам пост мы рассматриваем лишь как точку для сбора комментариев, т.е. он может как выражать какое-либо мнение, так и носить чисто информационный характер). После этого пользователь выбирает действие (в соответствии со своим сформированным мнением), ставит лайк соответствующим комментариям (из числа просмотренных), затем с некоторой вероятностью сам пишет комментарий (мы считаем, что каждый пользователь оставляет не более одного комментария).

Приведем теперь формальное описание модели. В начальный момент имеется множество $N = \{1, \ldots, n\}$ пользователей, которые подписаны в онлайновой сети на данный информационный источник. Считаем, что у каждого пользователя $i \in N$ в начальный момент времени имеются

следующие параметры: мнение $x_i \in [0;1]$, вероятность написать комментарий p_i, а также n_i – максимальное количество комментариев, которые пользователь просмотрит перед выбором своего действия (число комментариев / глубина просмотра – это параметр, отражающий характеристики как самого агента, так и характеристики поста). Также заданы неотрицательные числа $b_{ij}, i,j \in N$, характеризующие степень доверия пользователя i пользователю j.

Формирование последовательности комментариев после появления в информационном источнике сообщения (поста) осуществляется посредством выполнения n однотипных шагов.

На каждом шаге i с равной вероятностью выбирается любой из еще не видевших сообщение пользователей, не ограничивая общности будем считать его i-м. Он просматривает сообщение и либо первые n_i комментариев, либо, если количество всех имеющихся к данному шагу комментариев меньше n_i, все комментарии. Обозначим множество авторов просмотренных i-м пользователем комментариев через N_i. Каждый комментарий j-го пользователя является отражением его действия $y_j \in \{0,1\}$ – выбора позиции «за» (действие $y_j = 1$) или «против» (действие $y_j = 0$).

Будем считать, что под влиянием просмотренных комментариев i-й пользователь корректирует свое мнение x_i следующим образом:

(1) $x_i' = \frac{b_{ii}x_i + \sum_{j \in N_i} b_{ij} y_j}{b_{ii} + \sum_{j \in N_i} b_{ij}}$.

После этого i-й пользователь выбирает свое действие $y_i \in \{0,1\}$ в соответствии с параметром x_i': действие 1 («за») с вероятностью x_i' и действие 0 («против») с вероятностью $(1 - x_i')$. Далее i-й пользователь ставит лайк тем из просмотренных комментариев, которые соответствуют выбранному им действию (т.е. ставит лайк комментарию j-го пользователя при условии $y_i = y_j$). Наконец, в завершение шага i пользователь с вероятностью p_i сам пишет комментарий «за» или «против» в соответствии с выбранным действием (соответственно, с вероятностью $(1 - p_i)$ i-й пользователь не оставляет комментарий под сообщением). Введем параметр

$z_i \in \{0,1\}$, который равен 1, если i-й пользователь оставил комментарий, и равен 0 в противоположном случае.

В результате n шагов алгоритма формируется последовательность комментариев. Обозначим через N_z множество оставивших комментарий пользователей. Будем считать, что наиболее важной характеристикой последовательности является доля комментариев «за», т.е. $\delta = \sum_{i \in N_z} y_i / N_z$.

Пользователи социальной сети не являются стратегическими игроками, однако на них нацелено воздействие стратегических игроков. Будем рассматривать стратегических игроков трех типов, целью каждого их которых является либо максимизация, либо минимизация величины δ.

Первый тип – боты. Будем считать, что боты составляют множества $M^0 = \{n+1, \dots, n+m^0\}$ и $M^1 = \{n+m^0+1, \dots, n+ +m^0+m^1\}$, как бы дополняющие множество пользователей N. Бот отличаются от обычного пользователя тем, что всегда (с вероятностью 1) пишет комментарий, и его действие предопределено заранее: для j-го бота, $j \in M^0$, это действие $y_j = 0$ (и проставление лайков комментариям, соответствующим этому действию); для k-го бота, $k \in M^1$, это действие $y_k = 1$ (и, аналогично, проставление лайков комментариям, соответствующим этому действию). Таким образом, боты из множества M^1 (далее для краткости будем называть их 1-ботами) стремятся увеличить долю комментариев «за», а боты из множества M^0 (их будем называть 0-ботами) – уменьшить.

Второй тип стратегического игрока – администратор страницы информационного источника в сети (далее – администратор). Будем считать, что администратор источника может удалять нежелательные для него (по какой-либо причине) комментарии и лайки пользователей и ботов.

Наконец, третий тип стратегического игрока – сама онлайновая социальная сеть (далее – онлайновая сеть). Будем считать, что онлайновая сеть может управлять параметрами алгоритма ранжирования комментариев к постам.

Опишем теперь стратегии игроков, которые мы будем рассматривать.

Для ботов (т.е., по сути, для команд однотипных ботов) будем рассматривать два варианта:

(б1) боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи);

(б2) боты являются первыми комментаторами.

Администратор может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев (нежелательными являются либо комментарии «за», либо комментарии «против»). Будем считать, что администратор стирает каждый нежелательный комментарий, как только тот появляется с фиксированной вероятностью $q \in [0; 1]$, особо будем рассматривать два случая:

(а1) q=0,2;

(а2) q=0,8.

Онлайновая сеть может применять один из трех вариантов показа комментариев:

(с1) в обратном хронологическом порядке – сначала новые, потом старые;

(с2) в порядке убывания количества лайков (при одинаковом количестве лайков – в обратном хронологическом порядке, как в п. (с1));

(с3) сначала комментарии «за», затем комментарии «против» (внутри обоих множеств – в обратном хронологическом порядке, как в п. (с1)).

3. Влияние ранжирования комментариев

Для введенной выше модели будем оценивать характеристики информационных каскадов при помощи имитационного моделирования, позволяющего рассчитать усредненную долю комментариев «за» в зависимости от варианта показа (ранжирования) комментариев (действия ботов и администрации информационного источника рассмотрены в следующем разделе). Будем считать, что в сети $n = 100$ участников (например, подписчиков данной информационного ресурса) и она представляет собой полный граф, в котором каждый участник одинаково доверяет всем агентам (в том числе

самому себе). Мнения агентов в начальный момент времени равномерно распределены на отрезке [0; 1], кроме того, для всех агентов одинакова как вероятность написать комментарий $p_i = 0{,}5$, так и «глубина» просмотра $n_i = 7$.

Сначала в качестве иллюстрации приведем результаты одиночных экспериментов (см. рис. 1, вертикальными линиями обозначены моменты высказывания комментариев, начальные мнения агентов в последовательности обозначены звездочками).

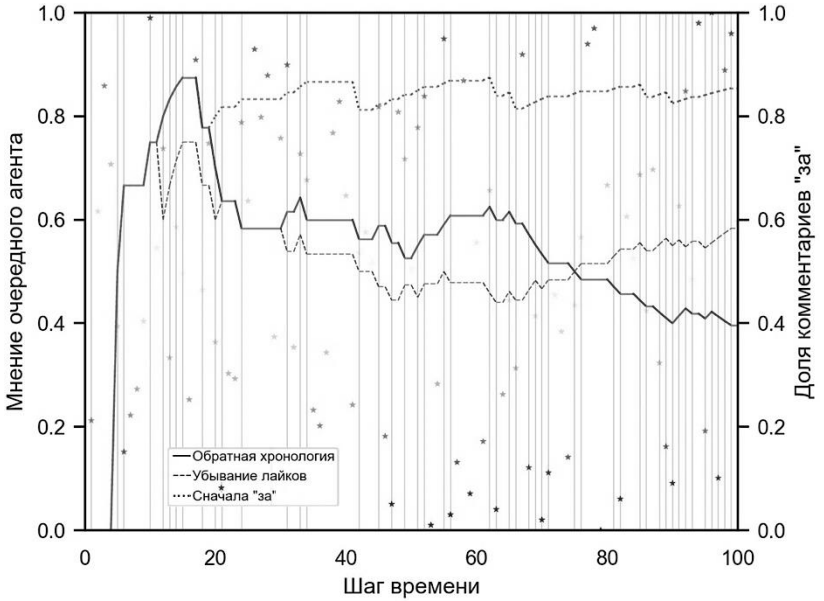


Рис. 1. Динамика доли комментариев «за»

Для случая показа в обратном хронологическом порядке (см. п.1) доля комментариев «за» резко растет, а затем спадает до уровня 0,40. Для случая показа в порядке убывания лайков (см. п.2) после резкого роста происходит падение, а затем рост восстанавливается до уровня 0,58. Показ сначала комментариев «за» приводит к удержанию уровня 0,85.

Для каждого варианта показа выполним 1000 запусков, а затем усредним результаты. Для варианта показа в обратном хронологическом порядке доля комментариев «за» составила 0,5,

для варианта показа в порядке убывания лайков – 0,5, а для показа сначала комментариев «за» – 0,9.

Введем теперь еще одну характеристику информационного каскада – долю комментариев «за» для первых 10 комментариев, которую дискретизируем следующим образом:

$$(2)\quad d = \begin{cases} 0, & \text{доля} \le 1/3, \\ 1, & 1/3 < \text{доля} \le 2/3, \\ 2 & \text{иначе.} \end{cases}$$

Увеличим количество агентов до $n = 500$ и проведем анализ доли «за» в зависимости от алгоритма показа, вероятности написать комментарий (p_i), глубины просмотра (n_i) и доли «за» (d), см. рис. 2.

	n_i	1			5			10			50			Вся история		
	d	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
Вариант показа	***p***															
Обратная хронология	0,1	0,45	0,50	0,57	0,32	0,46	0,66	0,21	0,53	0,77	0,13	0,50	0,86	0,13	0,50	0,86
	0,5	0,49	0,50	0,51	0,47	0,50	0,53	0,39	0,49	0,59	0,17	0,52	0,85	0,15	0,51	0,87
	0,9	0,49	0,50	0,50	0,48	0,51	0,52	0,44	0,49	0,55	0,18	0,50	0,82	0,14	0,49	0,86
Сначала "за"	0,1	0,63	0,71	0,76	0,60	0,83	0,91	0,43	0,82	0,93	0,13	0,50	0,86	0,13	0,50	0,86
	0,5	0,73	0,74	0,75	0,85	0,90	0,92	0,83	0,93	0,95	0,48	0,84	0,96	0,15	0,51	0,87
	0,9	0,74	0,74	0,75	0,88	0,91	0,92	0,89	0,94	0,95	0,66	0,90	0,97	0,14	0,49	0,86
Убывание лайков	0,1	0,24	0,48	0,76	0,15	0,51	0,85	0,13	0,51	0,86	0,13	0,50	0,86	0,13	0,50	0,86
	0,5	0,25	0,50	0,75	0,14	0,49	0,87	0,13	0,50	0,89	0,14	0,51	0,88	0,15	0,51	0,87
	0,9	0,25	0,52	0,75	0,12	0,48	0,87	0,10	0,47	0,88	0,13	0,49	0,86	0,14	0,49	0,86

Рис. 2. Доли комментариев «за» в зависимости от параметров

Сильно влияет на динамику вариант показа «сначала за», причем увеличение глубины просмотра до определенного момента позволяет усилить воздействие алгоритма (поскольку влияние окружения усиливается), но затем – по мере приближения n_i к n – воздействие ослабляется (поскольку агент начинает видеть все разнообразие позиций в сети). Кроме того, сложившееся в начале каскада «усредненное общественное мнение» (характеристика d) оказывает существенное влияние на итоговую долю комментариев «за» для всех алгоритмов показа. Особенно это верно для варианта показа «убывание лайков», что объясняется подкреплением влияния начального состояния лайками и возникновением положительной обратной связи.

Содержательно, ранний «захват» обсуждений, например, ботами приведет к достижению цели их владельца.

Рассмотрим теперь случай, когда распределение начальных мнений подчиняется бета-распределению (с параметрами $\alpha = 1$, $\beta = 5$). Проведем анализ доли «за» в зависимости от алгоритма показа, вероятности написать комментарий (p_i), глубины просмотра (n_i) и доли «за» (d), см. рис. 3.

	n_i	1			5			10			50			Вся история		
	d	0	1	2	0	1	2	0	1	2	0	1	2	0	1	2
Вариант показа	***p***															
Обратная хронология	0,1	0,16	0,25	0,28	0,11	0,35	0,52	0,08	0,45	0,69	0,06	0,46	0,80	0,06	0,46	0,80
	0,5	0,17	0,18	0,18	0,16	0,19	0,23	0,14	0,26	0,34	0,06	0,46	0,73	0,05	0,48	0,78
	0,9	0,17	0,17	0,18	0,16	0,17	0,19	0,15	0,23	0,22	0,08	0,42	0,70	0,06	0,46	0,79
Сначала "за"	0,1	0,42	0,57	0,63	0,35	0,79	0,86	0,21	0,78	0,90	0,06	0,46	0,80	0,06	0,46	0,80
	0,5	0,55	0,58	0,59	0,72	0,85	0,86	0,65	0,90	0,92	0,21	0,83	0,93	0,05	0,48	0,78
	0,9	0,57	0,58	0,59	0,78	0,86	0,86	0,77	0,91	0,92	0,40	0,89	0,96	0,06	0,46	0,79
Убывание лайков	0,1	0,09	0,56	0,63	0,07	0,44	0,79	0,06	0,48	0,82	0,06	0,46	0,80	0,06	0,46	0,80
	0,5	0,08	0,54	0,59	0,05	0,45	0,78	0,04	0,47	0,80	0,05	0,48	0,78	0,05	0,48	0,78
	0,9	0,09	0,54	0,59	0,05	0,45	0,80	0,04	0,46	0,81	0,05	0,45	0,80	0,06	0,46	0,79

Рис. 3. Доли комментариев «за» в зависимости от параметров

Предыдущие выводы сохраняются, однако мы видим общее для всех случаев естественное уменьшение доли «за».

Дадим содержательную интерпретацию полученным результатам.

Вариант показа по убыванию лайков. Нужно отметить, что агенты могут не публиковать комментарии, но ставят лайки. На начальном этапе появляется больше агентов с мнением «против», поэтому в показе будут доминировать комментарии с соответствующей позицией. Доминирование усугубляется в случае увеличения вероятности написать комментарий (p_i), поскольку агенты с мнением «против» не будут молчать и тем самым будут влиять на развитие информационного каскада (особенно при увеличении глубины просмотра агентов).

Вариант с обратной хронологией ведет к не столь быстрому доминированию доли комментариев «против». Агенты в последовательности поддаются влиянию уже высказавшихся участников сети. И здесь увеличение глубины наблюдения комментариев

«усугубляет» ситуацию, поскольку «консенсус» сложился, агенты поддаются влиянию большинства.

Вариант показа «сначала за» ведет в итоге к доминированию доли комментариев «за», но в случае агентов с небольшой глубиной наблюдений; в противном случае алгоритм может и не найти достаточное число комментариев «за» для перелома ситуации.

4. Влияние воздействия ботов и модерации комментариев

В данном разделе учтем действия ботов и действия администратора, который может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев.

Рассмотрим следующие параметры имитационного моделирования:

- доля m_0 от n: {0; 0,1; 0,3};
- доля m_1 от n: {0; 0,1; 0,3};
- «глубина» просмотра всех агентов n_i: 5 или $(n + m_0 + m_1)$;
- вероятность написать комментарий $p_i = 1$;
- позиция администратора: {«за», «против»};
- алгоритм сети: $\{c1, c2, c3\}$.

Таким образом, имеется следующий набор изменяемых в ходе имитационного моделирования параметров: количество 0-ботов, количество 1-ботов, глубина просмотра, алгоритм сети c, позиция администратора ({«за», «против»}). В этой ситуации имеется три стратегических игрока: администратор (его стратегии (а1), (а2)), 0-боты (стратегии (б1), (б2)) и 1-боты (стратегии (б1), (б2)). Выигрыш каждого игрока определяется итоговой долей δ комментариев «за» пользователей (не ботов!) – со знаком плюс для администратора с позицией «за» и 1-ботов, со знаком минус для администратора с позицией «против» и 0-ботов. Будем считать, что имеется игра в нормальной форме – игроки принимают решение одновременно и независимо, стремясь максимизировать свой выигрыш.

Решением игры будем считать, как обычно, равновесие Нэша – ситуацию (набор стратегий игроков), в которой ни один игрок не может увеличить свой выигрыш, изменив свою стратегию при фиксированных стратегиях остальных игроков.

Зафиксируем каждый набор значений параметров и оценим выигрыши игроков при тех или иных стратегиях (проведя 100 запусков каждой конфигурации). Построив матрицу выигрышей, найдем решение игры для каждой ситуации – равновесия Нэша в чистых стратегиях (напомним, что равновесием Нэша называется ситуация, в которой ни один игрок не может увеличить свой выигрыш, изменив стратегию при неизменных стратегиях других игроков).

Нетрудно видеть, что всего возможно $2^2 3^3 = 108$ ситуаций. Оказывается, что в 105 из них существует равновесие Нэша в чистых стратегиях. Упорядочим ситуации с равновесиями по убыванию итоговой доли «за». Максимум достигается при максимальной глубине просмотра, максимальном числе ботов «за», отсутствии ботов «против» и позиции администратора «за». Минимум достигается при максимальной глубине просмотра, максимальном числе ботов «против», отсутствии ботов «за» и позиции администратора «против». Глубина просмотра уменьшает влияние алгоритма ранжирования, поскольку пользователи видят всю историю сообщений.

Во всех ситуациях доминирующая стратегия администратора – удалять как можно чаще «нежелательные» сообщения. Для ботов, как правило, лучше стратегия (б2), при которой боты стремятся быть первыми комментаторами, задавая начальный тон обсуждения. Начальное преобладание мнений может существенно повлиять на последующих агентов, склоняя их к тому же мнению. Однако есть случаи, когда равновесной является стратегия (б1), при которой боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи). Во всех этих случаях глубина просмотра пользователей маленькая, при этом: а) либо сетью используется алгоритм «сначала новые»; б) либо алгоритм «по убыванию лайков» и как ботов с противоположной позицией больше, так и модератор имеет противоположную позицию; в) либо алгоритм «сначала за» и ботов с противоположной позицией («за») максимальное число (здесь равновесие (б1, б2)).

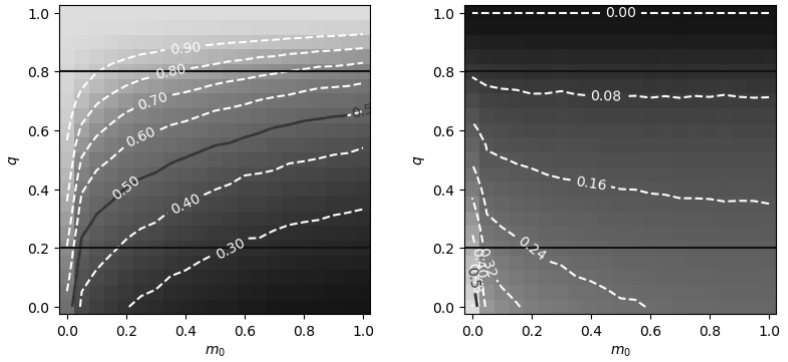
Теперь рассмотрим ситуацию, когда администратор по каким-то причинам фиксирует стратегию (а1) – это означает,

напомним, что он удаляет 20% нежелательных для него комментариев и это известно ботам. И снова лучшей стратегией является (б2).

Равновесия со стратегиями ботов «как обычные пользователи» (б1) возникают в случае, когда глубина просмотра пользователей маленькая (n_i = 5) и применяется алгоритм «сначала новые» (с1). Вариант равновесия «как обычные пользователи» (стратегия 0-ботов), «первые комментаторы» (стратегия 1-ботов) возникает в том случае, когда:

- позиция администратора «за»;
- доля 0-ботов – 10% или 30 %, доля 1-ботов – 10 %;
- глубина просмотра пользователей n_i = 5;
- алгоритм «сначала за» (с3).

Отметим случаи, когда изменение стратегии администратора приводит (см. рис. 4а и 4б, соответствующие стратегии линии q = 0,2 и 0,8):



(а) (б)

Рис. 4. Тепловые карты δ: а) позиции администратора ($y = 1$) и ботов противоположны; б) позиции администратора ($y = 0$) и ботов совпадают

а) к значительному изменению δ (>0,5) – это случай противоборства администратора с ботами с противоположной позицией, когда поддерживающие администратора боты отсутствуют, а глубина просмотра пользователей максимальна;

б) к незначительному изменению δ (<0,007) – это случай совпадения интересов администратора и ботов, когда боты с противоположной позицией отсутствуют, а глубина просмотра пользователей максимальна.

На рис. 4а видно, что ситуация условного баланса выигрышей противоборствующих сил ($\delta = 0{,}5$) описывается вогнутой кривой: при малом количестве ботов его увеличение «требует» от администратора значительного повышения усилий по удалению комментариев.

5. Воздействие управляющего органа

Выше были рассмотрены возможности стратегических игроков в зависимости от значений параметров ситуации. Эти игроки действуют оптимально в своих интересах, и равновесным (в теоретико-игровом смысле) результатом информационного процесса является средняя доля δ голосов «за», которая при каждом наборе параметров может быть оценена на основе многократных экспериментов. Таким образом, величина δ является функцией параметров ситуации. При этом важнейшими параметрами являются количества 0-ботов и 1-ботов – m_0 и m_1.

Предположим теперь, что в ситуации участвует некий управляющий орган (далее будем называть его *центром*), который может влиять на параметры ситуации. Целевой функцией центра является максимизация доли голосов «за». Центр не управляет ботами (т.е. не определяет их стратегию), но может влиять на их количество, неся при этом определенные затраты.

Формализуем задачу центра. Пусть, понеся затраты $c_0 \geq 0$, он может добиться количества 0-ботов $m_0(c_0)$, а понеся затраты $c_1 \geq 0$ – количества 1-ботов $m_1(c_1)$; если соответствующего результата добиться невозможно, то затраты на его достижение условно можно считать бесконечными. Суммарные затраты ограничены величиной c, т.е. бюджетное ограничение имеет вид $c_0 + c_1 \leq c$. При этом выгода центра от доли «за» δ составляет $f(\delta)$. При этих условиях задача максимизации целевой функции центра F имеет следующий вид:

(1) $F(P, c_0, c_1) = f(\delta(P, m_0(c_0), m_1(c_1))) - c_0 - c_1 \underset{c_0, c_1}{\longrightarrow} \max,$

при ограничениях $c_0 \geq 0, c_1 \geq 0, c_0 + c_1 \leq c$, где через P обозначены остальные параметры ситуации. Она может быть решена при помощи имитационного моделирования.

В конкретных частных ситуациях принятие решений выбор центра может формулироваться более просто на основе той же задачи (1). Пусть, например, при данном наборе значений параметров P имеется 30% от n 0-ботов и отсутствуют 1-боты, а у центра имеются только три возможности, удовлетворяющие бюджетному ограничению:

- не влиять на ситуацию, тогда затраты отсутствуют и $F = f(\delta(P; 0{,}3n; 0))$;
- уменьшить количество 0-ботов до нуля, тогда затраты c_0' и $F = f(\delta(P; 0; 0)) - c_0'$;
- уменьшить количество 0-ботов до 10% от n (затраты c_0''), при этом одновременно увеличив количество 1-ботов до тех же 10% от n (затраты c_1''), тогда $F = f(\delta(P; 0{,}1n; 0{,}1n)) - c_0'' - c_1''$.

Тогда оптимальное действие центра соответствует максимальному значению целевой функции.

Пусть теперь $c_0 = c_0' = \frac{3}{2}c_0''$, $c = 1$ (т.е. $c_0 + c_1 \leq 1$), $f\big(\delta(\cdot)\big) = k\delta(\cdot)$, где $k \in \{0{,}5; 1; 2; 20\}$.

Тогда целевая функция:

- в случае первой стратегии (А1): $k\delta_1$;
- в случае второй стратегии (А2): $k\delta_2 - c_0$;
- в случае третьей стратегии (А3): $k\delta_3 - \frac{2}{3}c_0 - c_1$.

Область доминирования стратегии А1:

$k\delta_1 > k\delta_2 - c_0$, $k\delta_1 > k\delta_3 - \frac{2}{3}c_0 - c_1$, т.е.

$c_0 > k(\delta_2 - \delta_1)$, $c_1 > k(\delta_3 - \delta_1) - \frac{2}{3}c_0$.

Область доминирования стратегии А2:

$k\delta_2 - c_0 > k\delta_1$, $k\delta_2 - c_0 > k\delta_3 - \frac{2}{3}c_0 - c_1$, т.е.

$c_0 < k(\delta_2 - \delta_1)$, $c_1 > k(\delta_3 - \delta_2) + \frac{1}{3}c_0$.

Область доминирования стратегии А3:

$k\delta_3 - \frac{2}{3}c_0 - c_1 > k\delta_1$, $k\delta_3 - \frac{2}{3}c_0 - c_1 > k\delta_2 - c_0$, т.е.

$c_1 < k(\delta_3 - \delta_1) - \frac{2}{3}c_0$, $c_1 < k(\delta_3 - \delta_2) + \frac{1}{3}c_0$.

Численные расчеты показывают, что $\delta_1 = 0{,}115$, $\delta_2 = 0{,}509$, $\delta_3 = 0{,}498$.

Рассмотрим области доминирования для различных k.

Пусть k = 0,5 (см. рис. 5). Тогда область доминирования A1 определяется

- $c_0 > 0{,}197$, $c_1 > 0{,}192 - 0{,}666c_0$.

Область доминирования A2:

- $c_0 < 0{,}197$, $c_1 > -0{,}006 + 0{,}333c_0$.

Область доминирования A3:

- $c_1 < 0{,}192 - 0{,}666c_0$, $c_1 < -0{,}006 + 0{,}333c_0$.

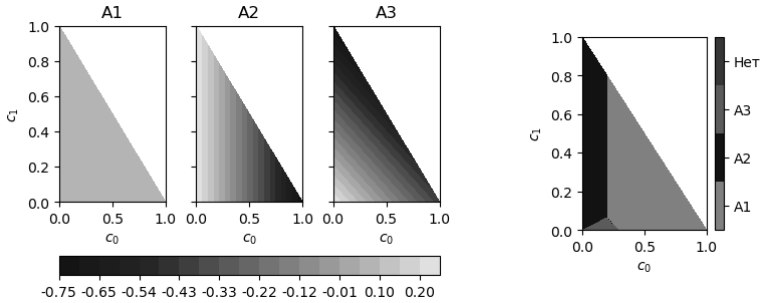


Рис. 5. k = 0,5. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть k = 1,0 (см. рис. 6). Тогда области доминирования определены справа на рис. 6.

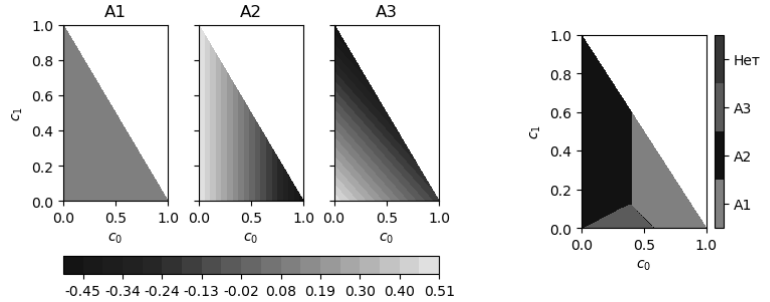


Рис. 6. k = 1,0. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть $k = 2{,}0$ (см. рис. 7). Тогда области доминирования определены справа на рис. 7.

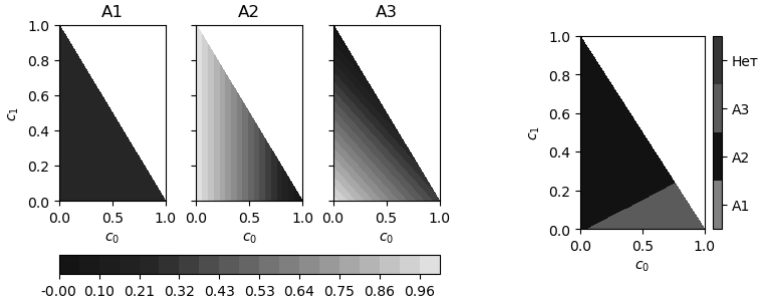


Рис. 7. k = 2. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

Пусть $k = 2{,}0$ (см. рис. 8). Тогда области доминирования определены справа на рис. 8.

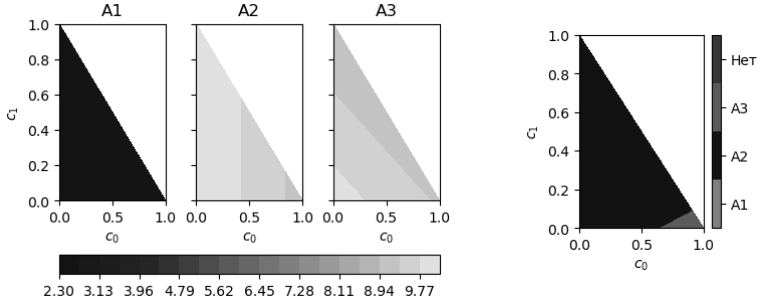


Рис. 8. k = 20. Значения целевой функции при разных действиях центра: A1, A2, A3; области доминирования представлены справа

По мере роста коэффициента k (определяющего выигрыш от числа голосов «за») увеличивается область допустимых затрат, при которых выгоднее действовать (т.е. выполнять A2 или A3). Выбор A2 или A3 определяется соотношением между c_0 и c_1 (см. рис. 5–8).

6. Заключение

В работе рассмотрена модель формирования информационных каскадов, в которой мнения (относительно некоторого вопроса) агентов не наблюдаемы, а наблюдаемые действия не полностью отражают их мнения. Совершаемые агентами действия (написание комментариев) влияют на мнения действующих впоследствии агентов, тем самым формируя информационный каскад мнений и действий. Как показали вычислительные эксперименты, существенное влияние на такой каскад оказывает алгоритм показа предшествующих действий агенту сети: в обратном хронологическом порядке, по убыванию лайков, или сначала комментарии с заданной позицией. Особенно это верно в том случае, когда агенты просматривают небольшое число комментариев (возможно в силу когнитивных ограничений). Следовательно, относительно простые изменения в алгоритмах онлайновой социальной сети могут оказать косвенное, но решающее воздействие на мнения и предпочтения пользователей в сети, и в итоге – на их действия.

Существенное влияние на мнения агентов также оказывают действия ботов и действия администратора, который может осуществлять модерацию, стирая определенное количество нежелательных для него комментариев. В этих случаях при различных параметрах модели рассматривается противоборство, для которого рассчитываются равновесия Нэша в чистых стратегиях. Во всех ситуациях доминирующая стратегия администратора – удалять как можно чаще «нежелательные» сообщения. Для ботов, как правило, лучше стратегия, при которой боты стремятся быть первыми комментаторами, задавая начальный тон обсуждения.

Начальное преобладание мнений может существенно повлиять на последующих агентов, склоняя их к тому же мнению, – это подтверждает важность первоначальной реакции пользователей для итогового результата обсуждения. Однако есть случаи, когда равновесной является стратегия «распределения», при которой боты с равной вероятностью оказываются на любом месте в последовательности комментаторов (как обычные пользователи). Это случаи характеризуются малой глубиной просмотра пользо-

вателей, когда администратор занимает противоположную позицию и (или) когда преобладают боты с противоположной позицией, – тогда ботам эффективнее предпринимать действия в случайные моменты времени, а не пытаться переломить ситуацию в самом начале (условно говоря, это «стратегия слабых»).

Формализована задача управляющего органа (центра), который может влиять на параметры ситуации, при этом неся определенные затраты и стремясь к максимизации доли голосов «за». Показаны области значений доступных центру затрат, где ему выгоднее отказаться от влияния либо предпринять то или иное действие.

Литература

1. ГУБАНОВ Д.А., НОВИКОВ Д.А. *Модели совместной динамики мнений и действий в онлайновых социальных сетях. Ч. 2. Линейные модели* // Проблемы управления. – 2023. – №3. – С. 40–64.
2. ГУБАНОВ Д.А., ПЕТРОВ И.В. *О модели поляризации мнений в социальных сетях* // Материалы 12-й Междунар. конф. «Управление развитием крупномасштабных систем» (MLSD'2019), Москва, ИПУ РАН. – М., 2019. – С. 1200–1202.
3. ГУБАНОВ Д.А., ПЕТРОВ И.В., ЧХАРТИШВИЛИ А.Г. *Многомерная модель динамики мнений в социальных сетях: индексы поляризации* // Проблемы управления. – 2020. – №3. – С. 26–33.
4. НОВИКОВ Д.А. *Модели динамики психических и поведенческих компонент деятельности в коллективном принятии решений* // Управление большими системами. – 2020. – Вып. 85. – С. 206–237.
5. ЧХАРТИШВИЛИ А.Г. *Задача нахождения медианного предпочтения индивидов в стохастической модели* // Автоматика и телемеханика. – 2021. – №5. – С. 139–150.
6. BANISCH S., OLBRICH E. *An argument communication model of polarization and ideological alignment* // arXiv:1809.06134. – 2018.
7. CHKHARTISHVILI A.G., GUBANOV D.A. *A Study on the Control of the Dynamics of Multidimensional Opinions in Social Networks* // Proc. of the 14th Int. Conf. "Management of Large-

Scale System Development" (MLSD-2021). – Moscow: IEEE. – 2021. – DOI: 10.1109/MLSD52249.2021.9600250.

8. CHKHARTISHVILI A.G., GUBANOV D.A. *Forming Opinions in Social Networks: The Confrontation of Several Information Sources* // Proc. of the 15th Int. Conf. Management of Large-Scale System Development (MLSD-2022). – Moscow: IEEE. – 2022. – DOI: 10.1109/MLSD55143.2022.9934221.
9. CHKHARTISHVILI A.G., GUBANOV D.A. *The Impact of Online Social Network Algorithms on User Opinion Formation* // Proc. of the 16th Int. Conf. Management of Large-Scale System Development (MLSD-2023). – Moscow: IEEE. – 2023. – DOI: 10.1109/MLSD58227.2023.10303932.
10. CHKHARTISHVILI A.G., GUBANOV D.A., NOVIKOV D.A. *Social Networks: Models of information influence, control and confrontation*. – Cham, Switzerland: Springer International Publishing, 2019. – 158 p.
11. DEGROOT M.H. *Reaching a Consensus* // Journal of American Statistical Association. – 1974. – No. 69. – P. 118–121.
12. FLACHE A., MÄS M. et al. *Models of Social Influence: Towards the Next Frontiers* // The Journal of Artificial Societies and Social Simulation. – 2017. – Vol. 20, No. 4. – URL: https://jasss.soc.surrey.ac.uk/20/4/2.html.
13. HUSZÁR F. et al. *Algorithmic amplification of politics on Twitter* // Proc. of the National Academy of Sciences. – 2022. – Vol. 119, No. 1. – e2025334119.
14. KOZITSIN I.V. *A general framework to link theory and empirics in opinion formation models* // Scientific Reports. – 2022. – Vol. 12, No. 1. – URL: https://www.nature.com/articles/s41598-022-09468-3.
15. MÄS M., FLACHE A. *Differentiation without distancing. Explaining bi-polarization of opinions without negative influence* // PloS one. – 2013. – Vol. 8, No. 11. – e74516.
16. PERRA N., ROCHA L.E. C. *Modelling opinion dynamics in the age of algorithmic personalization* // Scientific reports. – 2019. – Vol. 9, No. 1. – P. 1–11.
17. PETROV A.P., LEBEDEV S.A. *Online Political Flashmob: The Case of 632305222316434* // Computational Mathematics and Information Technologies. – 2019. – No. 1. – P. 17–28.

18. ROSSI W.S., POLDERMAN J.W., FRASCA P. *The Closed Loop Between Opinion Formation and Personalized Recommendations* // IEEE Trans. on Control of Network Systems. – 2021. – Vol. 9, No. 3. – P. 1092–1103.

THE INFLUENCE OF RANKING ALGORITHMS, BOTS AND CONTENT MODERATION ON OPINION FORMATION IN SOCIAL NETWORKS

Dmitry Gubanov, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor.Sc. (dmitry.a.g @gmail.com).
Alexander Chkhartishvili, V.A. Trapeznikov Institute of Control Sciences of RAS, Moscow, Doctor.Sc. (sandro_ch @mail.ru).

Abstract: This paper considers a model of information cascade formation in online social networks, accounting for the influence of content ranking algorithms, bot actions, and content moderation. Special attention is given to opinion dynamics, which is critically important for predicting and managing social processes. Unlike traditional models, in this case, the opinions of agents (users) are not directly observable: their actions, such as posting comments, act as indirect indicators of their opinions. These actions influence the opinions of other users, leading to the formation of an information cascade within the network. The model incorporates additional factors such as comment ranking algorithms, bot behavior, and content moderation. Computational experiments demonstrate that ranking algorithms significantly affect the dynamics of opinions and actions, particularly when users have a limited view depth. Moreover, the introduction of bots and moderation can substantially alter the course of discussions. The study explores the interaction of strategic players, including the moderator and bots with opposing views, and predicts the outcome of their interactions based on Nash equilibria. Finally, the problem of a control subject (principal) is formalized and solved for a specific case, where it seeks to advance a specific viewpoint in the network by influencing bots within the network.

Keywords: online social networks, social network algorithms, user opinion generation, bots, content moderation, information warfare, simulation modeling.

УДК 519.8+51-77
ББК 22.18

Статья представлена к публикации членом редакционной коллегии Ф.Т. Алескеровым.

Поступила в редакцию 28.09.2024.
Опубликована 30.11.2024.