ПРОГРАММИРОВАНИЕ, 2021, № 4, с. 45–55

ИНФОРМАЦИОННЫЙ ПОИСК

УДК 004.9

МОДЕЛЬ И МЕТОД ОБНАРУЖЕНИЯ ИНФОРМАЦИОННЫХ КАМПАНИЙ

© 2021 г. Д. Ю. Турдаков[a,b,*], С. В. Гарбук[c,], П. В. Хенкин[d,***], И. С. Козлов[a,****], А. В. Лагута[a,*****], М. И. Варламов[a,******]**

[a] *Институт системного программирования им. В.П. Иванникова РАН*
109004 Москва, ул. А. Солженицына, д. 25, Россия

[b] *Московский государственный университет имени М.В. Ломоносова*
119991 Москва, Ленинские горы, д. 1, Россия

[c] *Национальный исследовательский университет "Высшая школа экономики"*
101000 Москва, ул. Мясницкая, д. 20, Россия

[d] *Фонд перспективных исследований*
121059 Москва, Бережковская наб., д. 22, стр. 3, Россия

**E-mail: turdakov@ispras.ru*

***E-mail: sgarbuk@hse.ru*

****E-mail: pkhenkin@yandex.r*

*****E-mail: kozlov-ilya@ispras.ru*

******E-mail: laguta@ispras.ru*

*******E-mail: varlamov@ispras.ru*

Поступила в редакцию 10.03.2021 г.
После доработки 15.03.2021 г.
Принята к публикации 19.03.2021 г.

Статья посвящена исследованию возможности автоматического выявления информационных кампаний в условиях отсутствия априорных знаний о факте проведения, целях, затрагиваемых объектах и целевой аудитории. В статье предлагается общая модель информационной кампании, а также выделяются признаки проведения скрытых информационных кампаний. Модель подходит для описания информационных кампаний как в социальных медиа, так и в традиционных СМИ, в том числе за пределами сети Интернет. На основе описанных признаков предложен метод обнаружения информационных кампаний, позволяющий решать задачу в автоматическом режиме.

Для подтверждения работоспособности метода было проведено экспериментальное исследование на данных, собранных из социальных медиа. Мы привлекли экспертов в смежных областях для разметки сообщений и создания тестового корпуса. С целью анализа сложности задачи мы оценили степень их согласия. Результаты анализа подтвердили первоначальную гипотезу, что даже для профессионалов, задача обнаружения скрытых информационных кампаний является нетривиальной. Тем не менее, используя метод голосования, мы построили тестовую коллекцию на которой провели исследование отдельных признаков, а также сравнения предложенного метода с отдельными ответами экспертов. Результат экспериментов подтвердил перспективность предложенного подхода к решению задачи обнаружения информационных кампаний.

DOI: 10.31857/S0132347421040063

1. ВВЕДЕНИЕ

Информационная кампания – это совокупность информационных сообщений, целенаправленно публикуемых в некотором промежутке времени, направленных на определенную целевую аудиторию, с целью побуждения этой аудитории к конкретным действиям.

Наиболее распространенными информационными кампаниями являются рекламные кампании, целью которых является повышение продаж товаров. Реклама может осуществляться *явно* в выделенных для этого блоках (отведенное время в сетке телевизионного вещания, рекламные блоки на веб-страницах) или *скрыто* с помощью более сложных техник, таких как продакт-плейсмент [1] и публикация заказных статей в СМИ. В отличие от рекламных, политические информационные кампании чаще используют скрытые методы

доведения информации до целевой аудитории. В работе [2] отмечается, что сообщения “в поддержку политического решения должны предоставлять информацию, помогающую целевой аудитории сделать свои собственные выводы”. В частности, информационные кампании являются эффективным инструментом для лоббизма политических решений [3].

В области исследования информационных войн, для обозначения схожего процесса используется термин “информационная операция” [4]. Информационные операции могут состоять из одной или нескольких повторяющихся информационных кампаний, между которыми делается пауза для сбора и анализа реакции целевого объекта [5]. Специфика информационных операций заключается в необходимости быть скрытыми, так как сам факт обнаружения информационной операции может помешать достижению ее цели. Актуальность проблемы выявления информационных операций отмечается в Стратегии научно-технологического развития Российской Федерации, утверждённой Указом Президента Российской Федерации от 1 декабря 2016 г. № 642, и Доктрине информационной безопасности Российской Федерации, утвержденной Указом Президента Российской Федерации от 5 декабря 2016 г. № 646.

Обнаружение скрытых информационных кампаний, а также анализ их целей, является важным шагом для понимания текущего состояния и динамики развития общества, без чего, в свою очередь, невозможно принятие эффективных стратегических решений, как на уровне бизнеса, так и уровне государственного управления.

В статье исследуются методы обнаружения скрытых информационных кампаний. Мы предлагаем формальную модель информационной кампании. Модель определяет стадии жизненного цикла информационных кампаний, для каждой стадии описываются косвенные признаки проведения информационных кампаний. Мы выбрали наиболее распространенные признаки и реализовали метод обнаружения скрытых информационных кампаний на их основе. Для подтверждения работоспособности представленного метода была разработана методика проведения экспериментального исследования. Основной проблемой при проведении экспериментального исследования оказалась сложность решения задачи обнаружения информационных кампаний для людей, даже для тех кто является экспертами в близких областях.

В следующей секции приведен обзор релевантных работ. Затем предлагается модель информационной кампании (секция 3) и признаки выявления скрытых информационных кампаний (секция 4). Далее описывается разработанный метод (секция 5). В секциях 6 и 7 описывается методика построения проверочного корпуса, анализируются ответы экспертов и приводятся результаты экспериментального исследования разработанного метода и отдельных признаков.

2. ОБЗОР РЕЛЕВАНТНЫХ РАБОТ

Наиболее релевантной работой среди отечественных авторов, является работа А.В. Потемкина [6]. Автор анализирует информационные операции в Интернет СМИ. Основная используемая гипотеза состоит в том, что информационные операции отличаются от естественного распространения тем, что делается информационный вброс на малоизвестных новостных ресурсах (“активная фаза”), а потом, после короткого затишья, новость появляется в более известных новостных ресурсах (“пассивная фаза”). Для естественного распространения – наоборот. Автор строит граф распространения новостей, где ребрами соединяются новости с похожим текстом. Похожесть считается методом шинглов. Направление задается временем опубликования. Новости группируются по сюжетам, объединяющим сообщения по ключевым словам в заданном промежутке времени. Далее ищутся шаблоны, удовлетворяющие основной гипотезе.

Наибольшее число статей в зарубежной литературе посвящено исследованию методов обнаружения информационных кампаний в социальных медиа. Под термином “социальные медиа” принято понимать социальные сети, форумы, блоги и другие сервисы, предоставляющие своим пользователям возможность взаимодействия друг с другом путём обмена сообщениями, комментирования, выставления оценок и др. Возрастающую роль социальных медиа, как средства воздействия на массовое сознание людей, отмечают многие исследователи в различных областях. Особенностями социальных медиа, в отличие от традиционных СМИ, являются простота публикации сообщений и моментальное доведение информации до целевой аудитории, независимо от границ государств. При этом информация, полученная от других людей в виде комментариев, сообщениях на форумах и в социальных сетях, постах в блогах, вызывает наибольшее доверие. В условиях отсутствия у пользователей инструментов проверки актуальности и достоверности полученной информации, социальные медиа открывают беспрецедентные возможности манипулирования людьми.

Наиболее релевантной нашему исследованию работой является статья [7] коллектива авторов из Техасского университета A&M и Университета штата Огайо США. Авторы подробно описывают метод поиска информационных кампаний в Twitter, основанный на группировке похожих сообщений и авторов. Изучаются методы анализа графов для выделения скоординированных кампа-

ний (coordinated campaigns) и методы вычисления схожести коротких текстовых сообщений.

Авторы работы [8] группируют сообщения в сюжеты с помощью алгоритма потоковой кластеризации, а затем предоставляют их на проверку эксперту. Авторы предполагают, что имея статистическую информацию о сюжете, представленном в виде дашборда, эксперт сможет сказать, является ли этот сюжет информационной кампанией или нет. Наше исследование показывает, что задача выявления информационных кампаний является сложной даже для экспертов и требует более продвинутых методов автоматизации.

Также стоит отметить работы, изучающие отдельные признаки информационных кампаний. Так в работе [9] авторы изучают поведение пользователей сети Twitter для выявления групп вредоносных ретвитеров. В работе [10] предлагается метод выявления информационных кампаний за счет выявления ботов влияния.

В этой работе мы ставим задачу разработки метода обнаружения информационных кампаний, который будет работать как для традиционных СМИ, так и для социальных медиа. Целенаправленное сокрытие осуществления информационных кампаний и огромный объем информационных потоков в социальных медиа делает разработку средств их обнаружения сложной теоретической и практической задачей.

3. МОДЕЛЬ ИНФОРМАЦИОННОЙ КАМПАНИИ

В наиболее широком понимании, **информационная кампания** – это любая *деятельность A* в *информационном пространстве F*, направленная на достижение *цели информационной кампании C*.

В общем случае *целью информационной кампании* является побуждение аудитории к определенному действию или бездействию [11]. Это может быть совершение определенной покупки, выход на площадь для протестов, или, наоборот, желание остаться дома для ограничения эпидемии. При этом особенность современной коммуникации такова, что призыв не будет выполнен аудиторией, если она к этому эмоционально не готова. Поэтому информационные кампании направлены именно на осуществления такой эмоциональной подготовки.

Таким образом, **целью информационной кампании** C является создание у целевой аудитории $aud \subset U$ определенного эмоционального отношения к одному или нескольким целевым объектам $obj \in O$. Формально цель информационной кампании можно определить как множество

$$C = \{\{obj, s\}, aud\},$$

где $\{obj, s\} \subset \{O, S\}$ – множество пар объект–эмоциональная окраска. U – множество пользователей. O – множество целевых объектов. S – множество типов эмоциональной окраски. Заметим, что в частном случае, целью информационной кампании может быть увеличение числа упоминаний объекта в информационном пространстве вне зависимости от эмоциональной окраски.

Деятельность A по осуществлению информационной кампании заключается в создании информационных сообщений $M_{orig} = \{m_{orig}\} \subset M$, помогающих достичь цель, и поддержке распространения информации, для ее *доведения* до целевой аудитории и обеспечении ее *принятия*.

По аналогии с моделями информационных войн [11] возможно выделить следующие стадии жизненного цикла информационных кампаний:

1. Подготовка кампании. На этом шаге формируется план информационной кампании: анализируется отношение аудитории к целевым объектам, задается целевое отношение.

2. Подготовка инфраструктуры. Создается инфраструктура для распространения информации: создаются ресурсы для публикации материалов, на них привлекается целевая аудитория, производится “накрутка” популярности, в том числе создаются искусственные аккаунты (боты влияния). Так как создание новой инфраструктуры является сложной и дорогостоящей задачей, часто одна и та же инфраструктура используется для нескольких информационных кампаний.

3. Публикации первоначальной информации, создание информационного повода. Чаще всего информационный повод не создается “с нуля”, а ожидается подходящий информационный повод (реальное событие), информация о нем представляется в виде, способствующем достижению цели.

4. Поддержка доведения информации до целевой аудитории, с целью максимального распространения среди целевой аудитории. Сюда же отвлечение внимания аудитории на “более важные” события.

5. Анализ результатов информационной кампании и, при необходимости, повторение цикла.

Заметим, что для отдельно взятой информационной кампании обязательными являются только шаги 4 и 5. При этом, последний шаг может не оставлять наблюдаемых артефактов.

4. ПРИЗНАКИ ИНФОРМАЦИОННЫХ КАМПАНИЙ

Так как специфика изучаемых информационных кампаний предполагает скрытность, их выявление возможно только по косвенным признакам, нами разработан перечень признаков, проявляющихся при реализации информационных

Таблица 1. Признаки скрытых информационных кампаний на каждой стадии

Подготовка кампании	Проведение социологических опросов; Тестирование реакции небольшой фокус-группы на публикацию информации
Подготовка инфраструктуры	Признаки существующей инфраструктуры: • использование ресурсов и аккаунтов в других, уже известных, информационных кампаниях; • выявление контроля над ресурсами или аккаунтами организатора информационной кампании; Признаки создания новой инфраструктуры: • создание ресурсов (веб-сайтов, групп в социальных сетях и т.п.) для поддержки информационной кампании; • создание искусственных аккаунтов, в т.ч. со специфическими характеристиками (пол, возраст, фотографии и т.п.); • создание и искусственное увеличение аудитории и популярности ресурса (накрутка подписчиков, лайков и репостов); • резкая смена тематики ресурса (при сохранении аудитории);
Публикация первоначальной информации и поддержка доведения информации (шаги 3, 4)	Дубликаты сообщений, направленных на изменения отношения к целевым объектам, в различных ресурсах от разных авторов; Использование искусственных аккаунтов (ботов); Всплеск числа негативных комментариев о целевых объектах; Использование скомпрометированных ресурсов для публикации информации; Ссылки на скомпрометированные ресурсы; Использование характерных манипулятивных техник в тексте сообщений; Публикация сообщений не соответствующих основной тематике ресурса; Накрутка лайков для сообщений информационной кампании; Массовые репосты и установка ссылок на сообщения информационной кампании.
Анализ результатов	Этап анализа результатов информационной кампании в общем случае не оставляет наблюдаемых артефактов, по которым можно было понять наличие и длительность этого этапа. Либо деятельность аналогична первому этапу

кампаний различных типов и инвариантных к их конкретному содержанию (таблица 1). Приведенный список признаков не претендует на полноту. Однако заметим, что каждый элемент этого списка может быть обнаружен автоматизированными средствами.

Для полноты описания необходимо определить, что относится к использованию манипулятивных техник. Наиболее полный список перечислен в работе [12] и включает следующие техники: использование лжи, клеветы и дезинформации; провокация; аналитика и “квазианалитика”: статьи, оценивающие, интерпретирующие происходящие события, в т.ч. с учетом исторического контекста, направленные на изменение отношения к объекту; апеллирование к авторитету; подмена терминов; упрощенная “двуполярность” в интерпретации ситуации: «мы» и «враг»; использование специфичных дискурсивных конструкций: лозунги, агитация, пропаганда; акцентирование проблем; манипулирование ценностями; создание образа врага; поиск виновных; возложение вины на конкретную группу; формирование и развитие осознания идентичности; наставничество; двойные стандарты; создание коннотаций. Для определения некоторых из них предложены автоматические методы [13, 14]. Однако нам не известны работы по обнаружению большей части манипулятивных техник. Тем не менее, использование большинства из них может быть обнаружено с помощью методов машинного обучения.

Все признаки способны в той или иной мере определять информационные кампании на ранних стадиях. Однако признак на основе анализа всплеска числа негативных сообщений, в большинстве случаев, работает уже с реакцией на информационную кампанию, когда пользователи начинают выражать свой негатив по отношению к “информационному вбросу”. Таким образом способность признака определять информационную кампанию на ранней стадии ограничена. Тем не менее, в случае, если начальные сообщения информационной кампании выражают негатив по отношению к объекту мониторинга, они могут быть обнаружены с помощью данного подхода. Кроме того, признак на основе анализа всплеска числа негативных сообщений может быть полезен для увеличения полноты определения инфор-

мационных кампаний, если они не будут обнаружены на основе других признаков.

5. МЕТОД ОБНАРУЖЕНИЯ ИНФОРМАЦИОННЫХ КАМПАНИЙ

Мы ограничились обнаружением скрытых информационных кампаний на стадиях публикации первоначальной информации и поддержки доведения информации (стадии 3, 4). Нас интересовали признаки, инвариантные к конкретному содержанию информационной кампании (в частности, не использующие ключевые слова). Кроме того, в нашем исследовании мы ограничились только политической тематикой. Для этого был натренирован бинарный классификатор сообщений (логистическая регрессия), который позволил уменьшить объем данных для последующего анализа и увеличить их содержательность.

Предложенный метод обнаружения информационных кампаний состоит из следующих шагов:

1. Объединение отдельных сообщений в сюжеты;

2. Выявление признаков информационной кампании в сюжете.

Для объединения сообщений в сюжеты мы использовали метод, представленный в статье [15]. Метод состоит из двух шагов. На первом сообщения объединяются в кластеры на основе использования метода шинглов и “наивной кластеризации” по наличию общих специфичных именованны сущностей. На втором шаге содержимое кластеров уточняется, используя бинарный классификатор на парах сообщений.

Среди всех признаков мы выбрали пять наиболее распространенных (по нашим наблюдениям):

• Дубликаты сообщений, касающихся репутации целевых объектов, в различных ресурсах от разных авторов;

• Использование искусственных аккаунтов (ботов);

• Всплеск числа негативных комментариев о целевых объектах;

• Использование скомпрометированных ресурсов для публикации информации;

• Ссылки на скомпрометированные ресурсы.

Мы считали сообщения дубликатами, если их схожесть по мере Жаккара была больше 0.5. Для обнаружения таких сообщений мы воспользовались методом шинглов, по аналогии с алгоритмом выделения сюжетов. Для работы этого признака требуется только текст сообщений,однако полнота собираемых данных должна быть максимально высокой, так как сообщения-дубликаты публикуются в комментариях на различных ресурсах от имени разных людей. Обычный человек не может без специализированных инструментальных средств обнаружить такие публикации.

Для выявления искусственных аккаунтов использовался метод, описанный в работе [16]. Метод использует векторное представление вершин графа, для предсказания времени, через которое аккаунт может быть заблокирован. Мы считали, что признак сработал, если в дискуссии приняло участие не менее 3 ботов и доля сообщений от них среди всех сообщений была не менее 30%. Использование этого признака требует связи сообщений с аккаунтами авторов, а для работы используемого метода выявления ботов требуется граф социальных связей между этими аккаунтами.

В основе метода определения эмоционального отношения автора сообщения к упомянутому в этом сообщении объекту лежит модификация метода аспектно-ориентированного анализа эмоциональной окраски, предложенного на конференции SemEval-2016 [17]. Модификация заключается в удалении признаков, не релевантных для объектно-ориентированного анализа эмоциональной окраски (признаки, непосредственно связанные с аспектами).

Определение всплеска числа негативных комментариев о целевых объектах осуществлялось по следующему алгоритму:

1. Для каждого политического сообщения определяются объекты, упомянутые в сообщении и эмоциональное отношение к выявленным объектам. Для выявления объектов использовалась система Текстерра [18];

2. Сообщения группируются по времени написания (по часу);

3. Для каждого часа вычисляется число сообщений, написанных в этот час;

4. Для каждого объекта за каждый час вычисляется число сообщений, содержащих упоминания объекта;

5. Производится нормировка числа упоминаний: число упоминаний объекта делится на число сообщений за данный час

$$\frac{objectNum(hour)}{messageNum(hour)};$$

6. Для каждого объекта вычисляется среднее нормированное число упоминаний m и стандартное отклонение нормированного числа упоминаний σ;

7. Число негативных сообщений по отношению к заданному объекту считается аномально большим если нормированное число упоминаний объекта превысило среднее более чем на 3 стандартных отклонения

$$num > m + 3 * \sigma.$$

Для получения списка скомпрометированных ресурсов брались ресурсы, в которых предложенным методом были не менее трех раз обнаружены сюжеты, содержащие другие признаки информа-

Таблица 2. Ответы экспертов при разметке корпуса для определения точности

Номер эксперта	Номера сюжетов, отмеченные как информационные кампании
1	1, 3, 5, 7, 8, 9, 10, 11, 13, 14, 15, 17, 19, 21, 22, 24, 25, 26, 27, 28, 33, 34, 35, 36, 37, 43, 44, 45, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 79, 81, 83, 84, 85, 87
2	1, 4, 5, 6, 8, 11, 13, 15, 39, 40, 42, 47, 70, 75, 77
3	1, 2, 3, 4, 5, 8, 10, 11, 13, 14, 15, 17, 19, 20, 21, 22, 24, 25, 26, 27, 29, 31, 33, 34, 36, 37, 38, 39, 40, 41, 45, 46, 48, 49, 52, 53, 54, 55, 59, 60, 62, 63, 64, 65, 67, 68, 69, 70, 71, 72, 73, 75, 76, 77, 79, 83, 84, 87
4	1, 2, 3, 4, 5, 8, 10, 11, 14, 17, 21, 22, 24, 26, 27, 29, 34, 35, 36, 37, 40, 42, 44, 46, 49, 51, 53, 54, 56, 57, 60, 62, 63, 64, 65, 67, 68, 69, 70, 73, 75, 81, 86
5	2, 3, 4, 5, 10, 11, 13, 17, 21, 24, 25, 27, 33, далее разметка не проводилась
6	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 31, 33, 34, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 67, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 81, 82, 83, 84, 86, 87
7	2, 8, 11, 13, 16, 23, 28, 33, 35, 59, 62, 64, 65, 67, 68, 75, 79, 83, 87

ционных кампаний. Публикация сообщений в этих ресурсах и ссылки из сообщений на эти ресурсы, в свою очередь, считались признаками информационных кампаний.

6. КОРПУС ТЕКСТОВ ДЛЯ ЭКСПЕРИМЕНТАЛЬНОГО ИССЛЕДОВАНИЯ

При проведении экспериментального исследования мы ограничились сообщениями социальной сети "Вконтакте[1]" и "Живого Журнала (ЖЖ)[2]". Были собраны все сообщения и комментарии с одного миллиона наиболее активных групп социальной сети Вконтакте и журналы 110 тысяч активных пользователей и сообществ ЖЖ за январь и февраль 2017 года. Суммарный объем собранных данных составил 168 Гб. Мы не собирали информацию с сайтов СМИ, так как многие из них имеют представительства в сети "Вконтакте", которые являются зеркалами основных сайтов. При этом пользователи могут оставлять комментарии к новостям. Таким образом, мы проанализировали существенный срез информационных потоков в сети Интернет.

Для измерения точности и полноты системы необходим корпус текстов, объединенных в сюжеты, в котором эксперты разметили эти сюжеты, на предмет, являются ли они информационными кампаниями. При этом, для измерения точности методов, в таком корпусе должно быть не менее нескольких десятков информационных кампаний. Однако, в реальности доля информационных кампаний среди всех сюжетов крайне мала, и необходимо разметить выборку из нескольких тысяч сюжетов, чтобы в итоговом корпусе набралось достаточное количество примеров. При этом, предполагается, что выявление информационной кампании является сложной для эксперта задачей, поэтому стандартный подход с разметкой большого корпуса в данном случае неприменим.

Для того, чтобы уменьшить объем работы экспертов было решено разметить два независимых корпуса, отдельно для *точности* и *полноты*. Для тестирования *полноты* экспертам был предложен список всех сюжетов по темам внешнеполитической деятельности РФ за 20–28 февраля 2017 года. Список содержал 70 сюжетов для разметки. Экспертам была поставлена задача выбрать, какие из этих сюжетов, по их мнению, имеют признаки информационной кампании. Для тестирования *точности* результатов была предложена случайная выборка сюжетов за февраль 2017 года, найденных автоматически различными методами. Выборка состояла из 87 сюжетов. Данные за январь использовались для составления базы скомпрометированных ресурсов.

Разметка корпусов производилась с 15 по 22 марта 2017 года. В разметке участвовало девять экспертов, им назначены номера 1–9 в описании экспериментов далее. С корпусом для тестирования *точности* работали эксперты 1–7, корпус для тестирования *полноты* размечали эксперты 2–9. Результаты разметки корпусов экспертами представлены в таблицах 2 и 3.

Эталонная выборка строится в зависимости от степени согласия экспертов: при высокой степени согласия в эталонную выборку попадают все сюжеты, которые эксперты посчитали информационной кампанией; при низкой степени согласия экспертов в выборку попадают сюжеты, которые заданная доля экспертов посчитала информационной кампанией. Так как определение информационной кампании – сложная для экс-

[1] https://vk.com
[2] https://www.livejournal.com

перта задача, мы ожидали, что степень согласия экспертов окажется низкой. В этом случае, необходимо исследовать зависимость точности и полноты результатов метода от числа согласившихся экспертов.

Возможны следующие случаи:

• Сюжет считается информационной кампанией, если хотя бы один эксперт посчитал его таковым (порог равен 0);

• Сюжет считается информационной кампанией, если все эксперты посчитали его таковым (порог равен 1);

• Сюжет считается информационной кампанией, если более одного эксперта посчитали его таковым (порог между 0 и 1).

Точность результатов работы метода высчитывается на эталонном корпусе для точности как

$$P = \frac{|S \cap E|}{|S|} = \frac{|E|}{|S|},$$

где S – множество сюжетов–кандидатов, найденных методом и предложенных к разметке, E – множество сюжетов–кандидатов, попавших в эталонную выборку. При таком определении точность может только уменьшаться при увеличении согласия экспертов, так как знаменатель не меняется, а в число сюжетов, на которых эксперты согласны, что это информационная кампания, может только уменьшиться при появлении новых экспертов.

Полнота результатов работы метода высчитывается на эталонном корпусе для полноты по формуле

$$R = \frac{|S \cap E|}{|E|}.$$

При такой постановке, полнота может как расти, так и уменьшаться в зависимости от порога согласия экспертов.

Для определения степени согласия экспертов был измерен коэффициент каппа Флейса (Fleiss' kappa), который позволяет определить согласие для любого фиксированного числа экспертов и любого числа оцениваемых объектов.

$$\kappa_F = \frac{\overline{P} - \overline{P_e}}{1 - \overline{P_e}},$$

где $1 - \overline{P_e}$ – степень согласия, достижимая случайно, $\overline{P} - \overline{P_e}$ – прирост в достигнутой степени согласия относительно случайного уровня.

На разметке данных для определения *точности* результатов метода каппа Флейса была равна $\kappa_F = 0.05$. Согласие при разметке корпуса для определения *полноты* – $\kappa_F = 0.22$.

Таблица 3. Ответы экспертов при разметке корпуса для определения полноты

Номер эксперта	Номера сюжетов, отмеченные как информационные кампании
4	3, 5, 13, 14, 16, 17, 20, 31, 39, 43, 46, 55, 69
9	1
5	2, 3, 5, 6, 17, 18, 23, 31, 35, далее разметка не проводилась
2	31, 34
8	1
7	3, 4, 13, 31, 34, 39, 43, 46, 66, 69, 70
6	2, 3, 4, 5, 6, 13, 15, 16, 23, 35, 38, 43, 46, 59, 66, 69, 70
3	2, 3, 5, 6, 13, 14, 16, 23, 31, 43, 46, 66, 69, 70

Таблица 4. Референсные значения коэффициента каппа Флейса

κ_F	Интерпретация
<0	Отсутствие согласия
0.01–0.20	Крайне низкое согласие
0.21–0.40	Низкое согласие
0.41–0.60	Умеренное согласие
0.61–0.80	Существенное согласие
0.81–1.00	Почти полное согласие

Сравнение с референсными значениями каппы Флейса (таблица 4) показывает, что согласие экспертов при разметке было *крайне низким* при разметке корпуса для определения точности и *низким* при разметке корпуса для определения полноты, что говорит о сложности задачи определения информационных кампаний для человека.

Для более детального понимания проблемы было изучено попарное согласие экспертов. Результаты представлены на рисунках 1–4. Для измерения согласия использовались следующие меры:

• Коэффициент каппа Коэна $\kappa_C(A, B) = \frac{p_0 - p_e}{1 - p_e}$,

где p_0 – относительное наблюдаемое согласие между двумя экспертами, p_e – вероятность случайного согласия экспертов.

• Мера Жаккара $Jaccard(A, B) = \frac{|A \cap B|}{|A \cup B|}$

• Точность $Precision(A, B) = \frac{|A \cap B|}{|A|}$

• Полнота $Recall(A, B) = \frac{|A \cap B|}{|B|}$, где A и B – сюжеты, отмеченные первым и вторым эксперта-

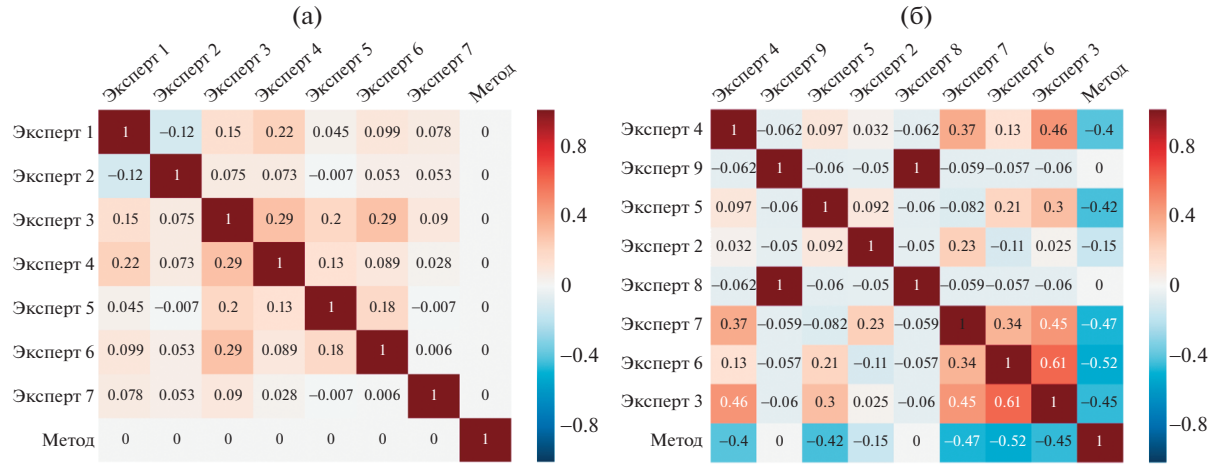

Рис. 1. Попарное согласие экспертов (и метода) по коэффициенту **каппа Коэна** при разметке данных для определения **точности** (а) и **полноты** (б) результатов метода.

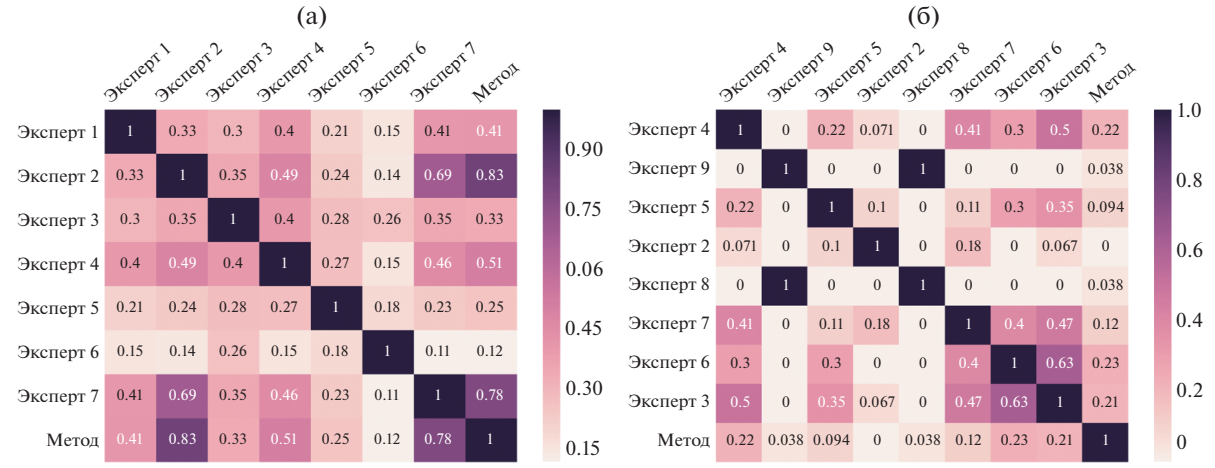

Рис. 2. Попарное согласие экспертов (и метода) по мере **Жаккара** при разметке данных для определения **точности** (а) и **полноты** (б) результатов метода.

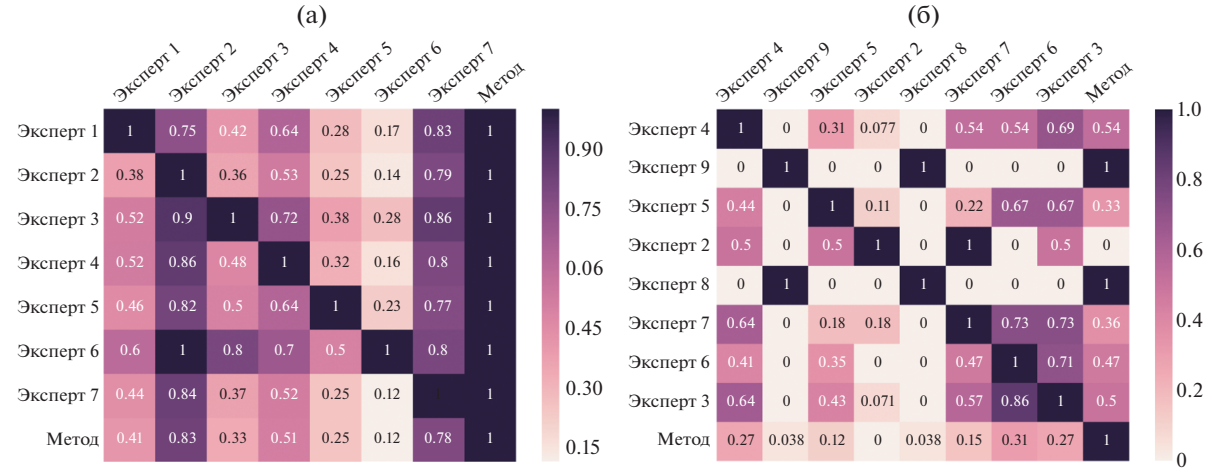

Рис. 3. Попарное согласие экспертов (и метода) по мере **Точности** при разметке данных для определения **точности** (а) и **полноты** (б) результатов метода.

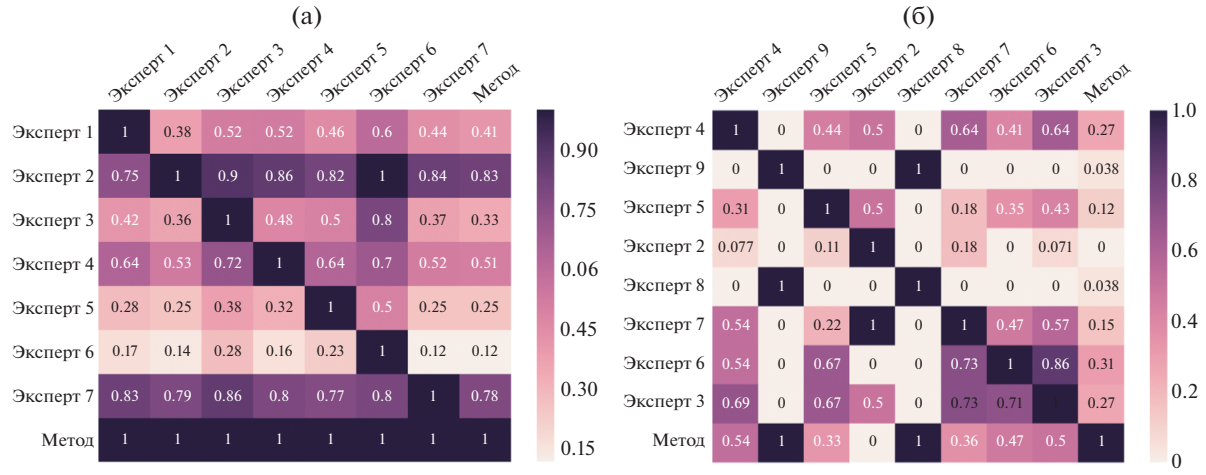

Рис. 4. Попарное согласие экспертов (и метода) по мере **Полноты** при разметке данных для определения **точности** (а) и **полноты** (б) результатов метода.

ми соответственно, как информационные кампании.

На рисунке 1 представлено попарное согласие экспертов (и предлагаемого метода) по коэффициенту каппа Коэна при разметке данных для определения точности (а) и полноты (б).

Каппа Коэна для предложенного метода и любого эксперта на эталонной коллекции для измерения точности всегда равна нулю. Так получается из-за того, что эта коллекция была сформирована на основе ответов, выдаваемых разработанным методом, поэтому наблюдаемое согласие с ним не отличается от случайного.

Для интерпретации остальных значений можно воспользоваться таблицей 4. В большинстве случаев попарное согласие экспертов является низким, что характеризует задачу выявления информационных кампаний, как крайне сложную для экспертов.

Стоит также отметить полное согласие двух экспертов (8 и 9) при разметке данных для определения **полноты**. Однако если обратиться к таблице 3, содержащей ответы экспертов, то можно увидеть, что оба эксперта отметили только первый сюжет, как информационную кампанию. При этом другие эксперты не посчитали этот сюжет содержащим признаки информационных кампаний. Более того в разметке точности упомянутые эксперты не участвовали. Исходя из этого можно сделать предположение, что эксперты только ознакомились с системой разметки, однако саму разметку не производили. В связи с этим, было решено убрать их ответы из эталонного корпуса. Это увеличило согласованность экспертов при разметке данных для определения полноты: $\kappa_F = 0.35$. Тем не менее, согласие экспертов осталось *низким*.

7. ЭКСПЕРИМЕНТАЛЬНОЕ ИССЛЕДОВАНИЕ МЕТОДА

На рисунке 5 представлены значения точности (а) и полноты (б) ответов экспертов и метода на соответствующих эталонных корпусах, в зависимости от значения порога согласия экспертов: в эталонный корпус попадали ответы, если с ним было согласно не менее порогового числа экспертов. Точность в таблице 5а вычислялась как

$$P = \frac{|S \cap E|}{|S|},$$

где S – множество ответов эксперта (или системы), E – множество информационных кампаний в эталонной выборке.

В разметке данных для определения *точности* участвовало 7 экспертов. Поэтому рисунок 5а содержит 7 столбцов. В разметке данных для определения *полноты* участвовало 6 экспертов (те же за исключением 1), однако не оказалось ни одного сюжета, где бы все 6 экспертов согласились, что он является информационной кампанией. Поэтому на рисунке 5б представлено только 5 столбцов.

Точность и полнота ответов системы на эталонных корпусах представлена в последней строке рисунков 5а и 5б соответственно.

На рисунке 6 показаны значения точности (а) и полноты (б) отдельных признаков на соответствующих эталонных корпусах, полученных при заданном значении порога согласия экспертов.

- *dublicates* – дубликаты сообщений;
- *badSources* – публикация в скомпрометированных источниках;
- *botComments* – участие ботов;
- *negative* – всплеск числа негативных сообщений;

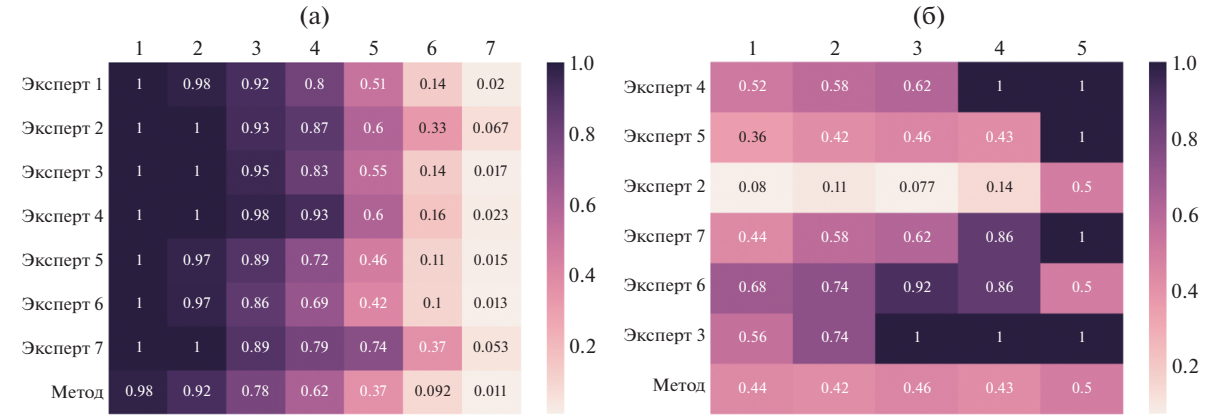

Рис. 5. Зависимость точности (а) и полноты (б) ответов экспертов и метода на соответствующих эталонных корпусах, полученных при заданном значении порога согласия экспертов.

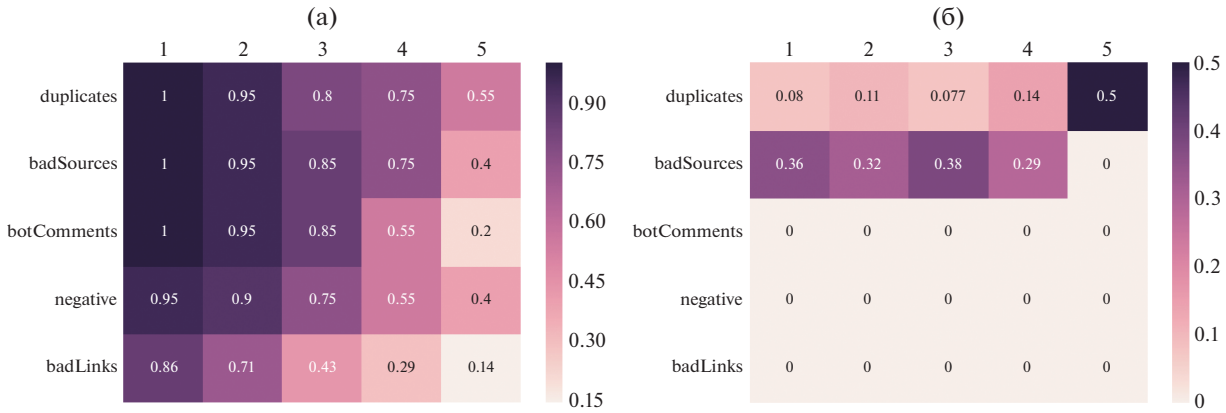

Рис. 6. Зависимость точности (а) и полноты (б) методов на эталонных корпусах, полученных при заданном значении порога согласия экспертов.

• *badLinks* – ссылки на скомпрометированные источники.

Лучшую точность показывают методы на основе поиска дубликатов и на основе недостоверных источников. В эталонный корпус для проверки *полноты* не попали информационные кампании, обнаруживаемые тремя последними методами, поэтому значения полноты нулевые. Исходя из этого можно сделать вывод, что улучшение методов выявления и анализ инфраструктуры распространения информации является перспективным направлением для исследований.

8. ЗАКЛЮЧЕНИЕ

В работе мы предложили формализацию понятия “информационная кампания”. Было выделено пять этапов проведения информационной кампании и для каждого из этапов определены признаки, которые позволяют обнаруживать скрытые информационные кампании. На основе наиболее распространенных из этих признаков был реализован метод выявления информационных кампаний, позволяющий получать результат в автоматическом режиме независимо от содержания кампании.

Особое внимание было уделено проведению экспериментального исследования предложенного метода. Задача выявления информационных кампаний оказалась крайне сложной для людей, о чем свидетельствует низкая степень согласованности ответов экспертов при разметке эталонных корпусов. В связи со сложностью получения эталонных данных, мы разделили задачу на две части, и отдельно измерили точность и полноту. Кроме того мы измерили точность и полноту метода и отдельных признаков в зависимости от порога согласия экспертов. Результаты измерений показали, что разработанный метод сравним по качеству с экспертной оценкой, однако сами признаки не обладают достаточной полнотой для их использования отдельно от других. Таким об-

разом, расширение предложенного метода путем добавления алгоритмов автоматического выявления остальных описанных в работе признаков является перспективным направлением работы.

СПИСОК ЛИТЕРАТУРЫ

1. *Березкина О.П.* Product Placement: технология скрытой рекламы. Издательский дом “Питер”, 2008.
2. *Годдард Б.* Кампании поддержки политических решений. Справочник по политическому консультированию / под ред. Д.Д. Перлматтера, 2002.
3. *Павроз А.В.* Информационные кампании в современном лоббизме. Вестник Пермского университета. Серия: Политология, 2014. № 2. С. 66–74.
4. *Расторгуев С.П.* Планирование и моделирование информационной операции. Информационные войны, 2014. № 1. С. 2–10.
5. *Манойло А.В.* “Дело Скрипалей” как операция информационной войны // Вестник Московского государственного областного университета, 2019. № 1.
6. *Потемкин А.* Распознавание информационных операций средств массовой информации сети интернет. Интернет-журнал Науковедение. 2015. Т. 3. № 28. С. 14.
7. *Lee K., Caverlee J., Cheng Z., Sui D.Z.* Campaign extraction from social media // ACM Trans. Intell. Syst. Technol. 2014. V. 5. № 1. P. 9:1–9:28. https://doi.org/10.1145/2542182.2542191
8. *Assenmacher D., Clever L., Pohl J.S., Trautmann H., Grimme C.* A two-phase framework for detecting manipulation campaigns in social media // International Conference on Human-Computer Interaction, Springer, 2020. P. 201–214.
9. *Vo N., Lee K., Cao C., Tran T., Choi H.* Revealing and detecting malicious retweeter groups. В 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), IEEE, 2017. P. 363–368.
10. *Abu-El-Rub N., Mueen A.* Botcamp: bot-driven interactions in social campaigns // The World Wide Web Conference, 2019. P. 2529–2535.
11. *Неждановв И.Ю.* Технологии информационных войн в интернете. [PDF] http://bash.rosmu.ru/activity/attach/events/1283/01.pdf, 2001.
12. *Кара-Мурза С.* Манипуляция сознанием. Век XXI. 2017.
13. *Zhou X., Zafarani R.* Fake news: a survey of research, detection methods, and opportunities. arXiv preprint arXiv:1812.00315, 2, 2018.
14. *Snefjella B., Lana N., Kuperman V.* How emotion is learned: semantic learning of novel words in emotional contexts // Journal of Memory and Language. 2020. V. 115. P. 104171.
15. *Скорняков К.А., Ласкина А.С., Турдаков Д.Ю.* Двухшаговый метод объединения новостей в сюжеты // Труды Института системного программирования РАН. 2020. Т. 32. № 4.
16. *Skorniakov K., Turdakov D., Zhabotinsky A.* Make social networks clean again: graph embedding and stacking classifiers for bot detection // Proceedings of the 27th ACM International Conference on Information and Knowledge Management 2018.
17. *Mayorov V., Andrianov I.* Mayand at semeval-2016 task 5: syntactic and word2vec-based approach to aspect-based polarity detection in russian. В Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016), pp. 325–329, San Diego, California. Association for Computational Linguistics, 2016.
18. *Турдаков Д., Астраханцев Н., Недумов Я., Сысоев А., Андрианов И., Майоров В., Федоренко Д., Коршунов А., Кузнецов С.* Texterra: инфраструктура для анализа текстов. Труды Института системного программирования РАН, 2014. Т. 26. № 1.