

ЦИФРОВЫЕ АЛГОРИТМЫ ВЫЯВЛЕНИЯ МОТИВОВ ПРИ ПУБЛИКАЦИИ ИНФОРМАЦИИ В СМИ

Лопатинский А.Г., студент группы СГНЗ-13Б

Московский государственный технический университет им. Н.Э. Баумана

*Научный руководитель: Карась Р.А., кандидат политических наук,
доцент кафедры «Информационная аналитика и политические технологии»,
karas@bmstu.ru*

Аннотация: Современные средства массовой информации (СМИ) играют ключевую роль в формировании общественного мнения и распространении информации. С развитием цифровых технологий, включая обработку больших данных, машинное обучение и методы анализа естественного языка, появилась возможность выявлять скрытые мотивы публикаций, анализируя их содержание и контекст. В данной статье рассматриваются цифровые алгоритмы, которые позволяют выявить основные мотивы публикации информации в СМИ, а также исследуются их преимущества и недостатки.

Ключевые слова: цифровые алгоритмы, обработка естественного языка, мотивы публикации, анализ СМИ, дезинформация, медиаграмотность.

В наше время, количество информации, циркулирующей в цифровом пространстве, стремительно возрастает. Тем самым, задачи по её фильтрации и оценке её достоверности, усложняются. В условиях распространения фейков, политической пропаганды в медиа и СМИ разработка эффективных инструментов для выявления скрытых мотивов публикаций становится критически важна. Цифровые алгоритмы, использующие методы машинного обучения, обработки природного языка и анализа социальных сетей способствуют не только автоматизации процесса, но и обеспечению более высокой степени объективности в оценке информации. В условиях информационных войн и роста дезинформации в медиапространстве такие технологии играют ключевую роль в защите информационного суверенитета государства.

Прежде чем перейти к цифровым методам, нужно понять, что такое мотив в контексте публикаций в СМИ. В этой статье мотив рассматривается как явная или скрытая цель публикации. Эта цель может быть политической, социальной, экономической или культурной. СМИ традиционно выполняют несколько функций, включая информационную, воспитательную, развлекательную и манипулятивную, и каждая из них однозначно связана с определенными мотивами [1].

Тем самым мотивы можно разделить на несколько типов:

1. Информационные мотивы. Публикации, целью которых является осведомление аудитории о тех или иных событиях, фактах и новостях. При этом сообщения могут представлять как объективную действительность, так и субъективную.

2. Пропагандистские мотивы. Использование СМИ для продвижения определенных идей, политических взглядов. Могут включать в себя манипулирование фактами, преувеличение, искажение информации с целью воздействия на общественное мнение.

3. Экономические мотивы. В СМИ нередко встречаются публикации, главной целью которых является продвижение экономических интересов, что выражается в наличии явной рекламы и создании информационных поводов для привлечения аудитории.

4. Социальные мотивы. Публикации, направленные на освещение социальных проблем, улучшение имиджа некоторых социальных групп, или на продвижение и нормализацию изменений в обществе.

5. Культурные мотивы. В СМИ могут быть опубликованы материалы, главной целью которых является популяризация определенной культурной продукции, искусства, музыки, кино.

Выявление мотивов публикаций в СМИ позволяет не только анализировать существующие публикации, но и предсказывать возможные тенденции, манипуляции или попытки влияния на общественное сознание со стороны элит. С развитием технологий обработки данных одним из эффективных инструментов для выявления мотивов публикаций стали цифровые алгоритмы. Рассмотрим основные виды алгоритмов, применимые для анализа контента в медиaprостранстве.

В первую очередь стоит отметить применимость машинного обучения для анализа текстов публикаций. Машинное обучение (ML) это область искусственного интеллекта, которая занимается разработкой алгоритмов, позволяющих компьютерам «обучаться» на основе данных, без явного программирования. То есть, алгоритмы могут автоматически улучшать свои результаты, находя закономерности в больших объемах данных. Подразделяется на два основных типа в зависимости от того, какие данные доступны для обучения модели. Машинное обучение с учителем (supervised learning) подразумевает обучение модели на размеченных данных, где каждому входному примеру уже соответствует правильный ответ (метка), и задача модели – научиться предсказывать эти метки для новых данных. В отличие от этого, машинное обучение без учителя (unsupervised learning) работает с неразмеченными данными, где модель должна самостоятельно выявить структуру или закономерности в данных, например, путем кластеризации или выделения скрытых факторов. ML представляет собой важный инструмент для автоматического анализа текстов и публикаций [3]. Используя алгоритмы машинного обучения, можно «обучить» систему распознавать определенные закономерности в тексте, что позволяет выявлять не только очевидный подтекст, но и скрытые смыслы.

1. Классификация текста. С помощью алгоритмов классификации, таких как наивный байесовский классификатор или SVM (машина

опорных векторов), можно автоматически классифицировать публикации по различным категориям мотивов, например, «информационные», «пропагандистские» или «экономические». Это позволяет заранее определить, какие из публикаций могут носить манипулятивный характер, и какие из них больше ориентированы на объективное освещение событий.

2. Нейронные сети. Более сложные нейронные сети, такие как сверточные нейронные сети (CNN) и рекуррентные нейронные сети (RNN), могут использоваться для более глубокого анализа текста и выявления сложных взаимосвязей в контексте целых медиа-экосистем. Например, нейронные сети могут помочь распознавать скрытые закономерности или тонкие изменения в риторике, которые могут быть характерны для определенной социальной, экономической или политической идеологии.

Обработка природного языка (NLP) также является ключевым инструментом для анализа текста в цифровом виде. С помощью NLP можно извлекать информацию из текстов, анализировать их содержание и выявлять важные элементы, которые могут указывать на мотивы публикации [5]. Стоит выделить несколько алгоритмов NLP:

1. Анализ тональности (sentiment analysis). Один из самых распространенных методов анализа в NLP – это анализ тональности, который позволяет определять эмоциональный окрас текста. Тексты с положительным, нейтральным или отрицательным окрасом могут указывать на различные мотивы. Например, публикация с преобладанием негативной тональности вполне может быть направлена на дискредитацию какого-либо объекта или лица, что может являться свидетельством наличия в публикации пропаганды или манипуляции.

2. Извлечение сущностей (Named Entity Recognition, NER). Эта технология позволяет идентифицировать важные сущности в тексте (например, имена людей, компании, географические объекты и т. д.). Такой подход позволяет выявить, о чем или о ком именно идет речь в публикации, а также установить связи между различными участниками, что важно для анализа скрытых мотивов и идеологических направлений.

3. Тематика текста и тематическое моделирование. Алгоритмы тематического моделирования позволяют выявлять скрытые мотивы в большом объеме текста. Примером такого алгоритма является Latent Dirichlet Allocation (LDA). Основная идея данного алгоритма заключается в том, что документы представляются в виде случайных смесей по скрытым темам, где каждая тема характеризуется распределением по словам [6]. Использование таких алгоритмов помогает понять, о каких проблемах или событиях идет речь, и как СМИ могут манипулировать темами для продвижения определенных мнений.

4. Анализ семантики текста (semantic analysis). Включает использование алгоритмов для анализа содержательной структуры элементов смысла, связанной с количеством смысловых ядер, тем в тексте. С его помощью можно выявить скрытые сообщения или предвзятость, присутствующие в тексте [2].

Социальные сети как социальная структура для социальных коммуникаций между социальными объектами постепенно стали важнейшей средой для распространения информации, и в настоящее время все СМИ имеют присутствие в социальных сетях. Их анализ позволяет выявить, как мотивы публикаций в СМИ адаптируются в цифровом пространстве. Современные алгоритмы позволяют отслеживать, как пользователи реагируют на публикации: какие комментарии они оставляют, кому они эту публикацию пересылают и т. п. [7]

1. Анализ социальных сетей. С помощью алгоритмов анализа социальных сетей можно определить, какие именно публикации становятся вирусными, какие темы обсуждаются с наибольшим количеством комментариев и лайков. Это помогает выявить основные мотивационные направления, в которых развивается общественное обсуждение.

2. Анализ эмоций и мимики. Видеоматериалы и изображения также являются важным компонентом публикаций в СМИ, и анализ этих материалов с помощью цифровых технологий позволяет выявить, какую реакцию эмоциональный визуальный контент вызывает у аудитории.

В рамках данного исследования была проанализирована статья американского новостного телеканала CNN с использованием алгоритмов NLP. В представленной статье освещается штурм здания Капитолия США сторонниками экс-президента Дональда Трампа. Сам Трамп в данной статье фигурирует под ярлыком «угроза демократии».

Результаты анализа экстракта статьи с использованием алгоритмов NLP:

1. Sentiment Analysis:

Общий тон:

Положительные аспекты: Призывы к защите демократии, восстановление справедливости, и признание героизма защитников Капитолия.

Негативные аспекты: Осуждение насилия, критика Трампа и его сторонников.

Общая тональность: сбалансированная с уклоном в негативное (из-за описания угроз демократии и насилия).

2. Latent Dirichlet Allocation (LDA):

Выделены следующие темы:

Защита демократии: «rule of law», «soul of America», «voting rights.»

Осуждение насилия: «violent mob,» «storm the Capitol.»

Политическая ответственность: «former President,» «election lies.»

Законодательные реформы: «Voting Rights Advancement Act,» «Freedom to Vote Act.»

3. Semantic Analysis:

Ключевые концепты:

Демократия как идеал и система: «preserve the flame of democracy.»

Угрозы: «lies», «violence», «subvert our elections.»

Исторический контекст: отсылки к прошлым кризисам, символизирующим стойкость демократии.

Байден представлен как защитник национальных ценностей. Трамп и его сторонники ассоциируются с ложью и угрозой демократическим институтам.

4. Named Entity Recognition (NER):

Лица: Joe Biden, Donald Trump, Kamala Harris, Stephanie Grisham, John Lewis.

Организации: US Capitol Police, Congress, Senate.

Места: US Capitol, Washington, DC.

События: January 6 insurrection, 2020 election.

Законы: John Lewis Voting Rights Advancement Act, Freedom to Vote Act.

Анализ текста выявил признаки кризиса демократии в США, вызванного событиями 6 января 2021 года, с акцентами на защите демократических принципов и осуждении политических манипуляций. Он подчеркивает историческую значимость момента и необходимость законодательных реформ.

Данный пример показывает, что цифровые алгоритмы весьма точны при анализе текста публикации. При этом важно отметить, что даже при наличии такого разнообразия алгоритмов, пригодных для выявления скрытых мотивов публикаций, алгоритмы не безупречны. Тем самым можно выделить как преимущества алгоритмов, так и их недостатки.

Алгоритмы имеют следующие преимущества:

1. Автоматизация анализа. Цифровые алгоритмы позволяют ускорить процесс анализа и обработки данных.

2. Объективность. В отличие от человеческого анализа, алгоритмы могут выявлять мотивы и закономерности, не поддаваясь личным предвзятостям и эмоциям. Здесь важно отметить, что алгоритмы анализируют текст относительно нейтрально, которая задаётся человеком, и тем самым полной объективности при анализе достичь практически невозможно.

Недостатки алгоритмов:

1. Ограниченность контекста. Алгоритмы не всегда учитывают культурные, исторические или социальные контексты, в которых происходит публикация информации. Это может привести к ошибочным выводам о мотивах. Тем самым, если анализировать русскоязычную публикацию, переводя её на английский язык, вполне возможно, что часть смысловой нагрузки будет утеряна.

2. Проблемы с интерпретацией. Даже самые современные алгоритмы не всегда могут точно интерпретировать скрытые или многозначные мотивы, которые могут быть явными для человеческого анализатора. Чаще всего выбирает преобладающий мотив в публикации как мотив всей публикации. К примеру, при анализе новостной публикации, посвященной торгово-экономическому противостоянию между Китаем и США, алгоритм учтёт тот факт, что в данной публикации есть как политический мотив, так и экономический, но выделит один [8].

Цифровые алгоритмы являются мощным инструментом для анализа публикаций в СМИ, позволяя выявлять их скрытые мотивы. На основе проведённого исследования можно прийти к следующим умозаключениям: Автоматизация анализа публикаций повышает эффективность обработки данных. Современные алгоритмы способны выявлять манипуляции и эмоциональные акценты, что делает их важным инструментом в борьбе с дезинформацией. Тем не менее, для достижения более точных результатов необходимо учитывать контекст и специфику анализируемых данных. В будущем внедрение технологий глубокого обучения и мультимодального анализа может значительно повысить качество таких исследований.

Литература и источники:

1. Пенков В.И. Масс-медиа и общественное мнение.
2. Ремарчук В.Н. Информационная аналитика: теория, методология, технологии: учебник для вузов. «Лань» 2022.
3. Blei D.M., Jordan M.I. Latent Dirichlet Allocation. Journal of Machine Learning Research 2003.
4. Goodfellow I., Bengio Y., Courville A. Deep Learning. MIT Press. 2016. Chapter 1
5. Шевченко, А.В., Соколова Н.В. Анализ социальных сетей: методы и инструменты // Вестник ТГУ. 2019.
6. Соловьев В.М. Обработка естественного языка в СМИ: алгоритмы и приложения. М.: Издательство МГУ (2020). С. 45–65.
7. Иванов В.И. Методы и технологии медианализов в эпоху цифровых трансформаций // Технологии и общество. 2022. С. 45–50
8. Косов А.А., Щербаков В.В. Модели и методы анализа текста в информационных системах. М.: Научный мир, 2019. С. 45–72.

9. Бочарников И.В., Овсянникова О.А. Риски и вызовы информационной работы при проведении специальных военных и полицейских операций // Вестник Академии военных наук. 2024. № 1. С. 22–29.

10. Информационная аналитика в современном социально-политическом процессе: теория и практика Ремарчук В.Н., Бочарников И.В., Артемьев А.А., Галаганова С.Г., Гришнова Е.Е., Егоров В.Г., Карась Р.А., Катков О.Н., Ламинина О.Г., Смутьский С.В., Шевчун В.Н., Урсул В.И. Москва, 2024.

11. Украинский кризис в условиях трансформации современного миропорядка: тенденции развития, угрозы и вызовы для России / Баранов В.П., Бартош А.А., Бочарников И.В., Дульнев П.А., Караваев И.Н., Кардаш И.Л., Карпович О.Г., Корабельников А.А., Кулаков А.А., Манойло А.В., Овсянникова О.А., Петренко А.И., Ремарчук В.Н., Стригунов К.С., Сурма И.В., Суханов П.В. (2-е издание, исправленное) Москва, 2022.