DOI 10.37882/2223-2966.2024.10.31

СРАВНИТЕЛЬНЫЙ АНАЛИЗ АЛГОРИТМОВ ФИЛЬТРАЦИИ КОНТЕНТА В СОЦИАЛЬНЫХ СЕТЯХ

COMPARATIVE ANALYSIS OF CONTENT FILTERING ALGORITHMS IN SOCIAL NETWORKS

N. Nekrasov

Summary. In this article, the author conducted a review of content filtering algorithms in social networks, such as TF-IDF and Naive Bayes. Each algorithm is examined in the context of its advantages, disadvantages, and potential areas for improvement. The presented comparative analysis demonstrates which algorithm is better suited for content filtering in social networks, by comparing them based on a specific example of classifying comments on a post in the social network VKontakte.

Keywords: TF-IDF, Naive Bayes, content filtering, social networks.

Некрасов Никита Михайлович
аспирант, Финансовый университет при Правительстве РФ, г. Москва
nekrasovnm@ya.ru

Аннотация. В данной статье автором был проведён обзор алгоритмов фильтрации контента в социальных сетях, таких как TF-IDF и Naive Bayes. Каждый алгоритм рассматривается в контексте его преимуществ, недостатков и потенциальных областей улучшения. Представленный сравнительный анализ показывает, какой алгоритм больше подходит для фильтрации контента в социальных сетях, проведя сравнение на конкретном примере классификации комментариев к посту в социальной сети Вконтакте.

Ключевые слова: TF-IDF, Naive Bayes, фильтрация контента, социальные сети.

В современном обществе социальные сети играют важную роль в передаче информации и взаимодействии между пользователями. Однако с ростом популярности социальных платформ увеличивается количество нежелательного контента, такого как: спам, фейковые новости, оскорбительные сообщения и другие виды. Для борьбы с данным типом широко применяются алгоритмы фильтрации, которые позволяют автоматически выявлять и удалять нежелательные материалы, обеспечивая безопасность и комфорт пользователей.

Существует несколько алгоритмов фильтрации контента в социальных сетях, наиболее популярными из которых являются методы на основе *TF-IDF* для анализа текста и Naive Bayes. Каждый из этих методов обладает своими особенностями и преимуществами, что открывает возможности для их использования в различных ситуациях.

Term Frequency-Inverse Document Frequency (*TF-IDF*) вычисляет важность каждого слова в документе относительно количества его употреблений в данном документе и во всей коллекции текстов.

TF (*Term Frequency*) — относительная частота слова в документе. Она измеряет, насколько часто слово появляется в документе. Чем чаще слово встречается в документе, тем выше его *TF*.

$$TF(t,d) = \frac{\text{число раз, когда слово } t \text{ встречается в документе } d}{\text{общее число слов в документе } d} \quad (1)$$

IDF (*Inverse Document Frequency*) — обратная частота документов, содержащих слово. Она измеряет, насколько уникально слово в контексте всего корпуса документов. Чем реже слово встречается в других документах, тем выше его *IDF*.

$$IDF(t,D) = \log \frac{\text{общее число документов в коллекции } D}{\text{число документов в коллекции, содержащих слово } t} \quad (2)$$

После расчета *TF* и *IDF* для каждого слова, *TF-IDF* для слова *t* в документе *d* вычисляется как произведение *TF* и *IDF*:

$$TF - IDF(t,d) = TF(t,d) * IDF(t,D) \quad (3)$$

В целом, метод *TF-IDF* является важным инструментом для оценки важности слов в контексте документа, однако он имеет свои преимущества и недостатки.

Существует несколько способов улучшить метод *TF-IDF*, один из них — вместо использования простого подсчета числа вхождений слова в документе для вычисления *TF*, можно применить деление на общее количество слов в документе. Это позволит сделать *TF* независимым от длины документа и более точно оценивать важность слова.

Применение *TF-IDF* предоставляет множество возможностей для анализа и понимания текстовых данных. В сочетании с современными методами анализа он способствует формированию информативных моделей и раскрытию семантических взаимосвязей в текстах.

Таблица 1.
Преимущества и недостатки метода TF-IDF

Преимущества	Недостатки
Учитывает важность слова в контексте документа: *TF-IDF* выделяет слова, которые часто встречаются в документе, но редко встречаются в других документах, что делает их более значимыми для содержания данного документа;	Чувствительность к редким словам: редкие слова, которые встречаются в небольшом количестве документов, могут получить завышенные значения *TF-IDF*, что может привести к искажению оценки их важности.
Позволяет учитывать длину документа: *TF-IDF* корректирует *Term Frequency* в зависимости от длины документа, что позволяет более точно оценивать важность слова;	Не учитывает семантическую связь слов: *TF-IDF* рассматривает каждое слово независимо от контекста, что может привести к недооценке или переоценке важности слов;
Прост в вычислении: расчет *TF-IDF* относительно прост и может быть эффективно реализован.	Не подходит для обработки коротких текстов: в случае коротких текстов или документов, *TF-IDF* может оказаться менее эффективным из-за недостаточного объема данных.

Наивный байесовский классификатор (Naive Bayes) представляет собой один из наиболее используемых инструментов в области фильтрации контента в социальных сетях. Данный алгоритм машинного обучения может применяться для автоматической идентификации и классификации разнообразных типов контента, таких как сообщения, изображения и видео, с целью защиты пользователей от вредоносного, неприемлемого или нежелательного контента.

Формула Байеса для машинного обучения выглядит следующим образом:

$$P(Ck \mid X) = \frac{P(Ck)P(X \mid Ck)}{P(X)} \quad (4)$$

где: $P(Ck|X)$ — апостериорная вероятность принадлежности образца к классу Ck с учётом его признаков X;

$P(X|Ck)$ — правдоподобие, то есть вероятность признаков X при заданном классе Ck;

$P(Ck)$ — априорная вероятность принадлежности случайно выбранного наблюдения к классу Ck;

$P(X)$ — априорная вероятность признаков X.

При использовании не одного, а нескольких признаков для описания объекта, формула будет:

$$P(Ck \mid X1, X2, ..., Xn) = \frac{P(Ck)\prod_{i=1}^{n} P(Xi \mid Ck)}{P(X1, X2, ..., Xn)} \quad (5)$$

Существует несколько разновидностей наивного байесовского классификатора, которые отличаются основой вероятностной модели и типом распределения вероятностей.

1. Мультиномиальный наивный Байес (Multinomial Naive Bayes): этот вариант наивного Байеса часто используется для классификации текстовых данных, где признаки представлены частотой появления слов в документах.
2. Бернуллиев наивный Байес (Bernoulli Naive Bayes): в отличие от мультиномиального наивного Байеса, этот вариант используется для бинарных данных, где признаки могут принимать только значения 0 или 1. Он также часто используется в задачах анализа текста, но моделирует каждый класс как распределение Бернулли.
3. Гауссов наивный Байес (Gaussian Naive Bayes): этот вариант применяется, когда признаки представлены непрерывными значениями и считается, что они имеют гауссово распределение в каждом классе. Гауссов наивный Байес может быть эффективен для задач классификации с числовыми признаками.
4. Комплементарный наивный Байес (Complement Naive Bayes): этот вариант разработан специально для сбалансированных датасетов с несбалансированными классами, что делает его полезным для задач с неодинаковым распределением классов.

Таблица 2.
Преимущества и недостатки метода Naive Bayes

Преимущества	Недостатки
Простота и высокая скорость обучения: Naive Bayes обладает простой структурой и не требует сложных вычислений для обучения, что позволяет эффективно обрабатывать большие объемы данных и достигать высокой скорости работы.	Неэффективность при неправильном выборе модели: выбор неподходящей вероятностной модели может существенно снизить производительность наивного байесовского классификатора.
Эффективность при небольшом объеме данных: даже при небольшом количестве обучающих данных наивный байесовский классификатор может продемонстрировать хорошую производительность	Предположение о независимости признаков: одним из основных недостатков является его предположение о независимости между признаками.

Один из способов улучшения производительности наивного байесовского классификатора — это использование методов сглаживания, для избегания нулевых вероятностей при отсутствии какого-либо признака в обучающих данных. Еще одним вариантом повышения качества Naive Bayes является проведение предварительной обработки данных, т.к. нормализация признаков, удаление шума, отбор признаков и другие методы обработки данных могут улучшить качество классификации и сделать классификатор менее чувствительным к шуму.

Наивный байесовский классификатор — это простой и эффективный метод машинного обучения, обладаю-

щий хорошей обобщающей способностью и быстрой скоростью работы. Он остается важным инструментом в машинном обучении, особенно в задачах классификации текстовых данных в социальных сетях.

Рассмотрев теоретические аспекты двух алгоритмов, проведем обработку и анализ набора комментариев к посту в социальной сети вконакте на тему того, что из-за искусственного интеллекта исчезают профессии, используя оба подхода, что позволит оценить их эффективность в контексте фильтрации спам-контента.

Проводя анализ *TF-IDF*, первым делом проводим предварительную обработку данных, которая включает в себя очистку текста от специальных символов, приведение всех слов к нижнему регистру, токенизацию и удаление стоп-слов, таких как предлоги, союзы. Следующим шагом преобразовываем комментарии в *TF-IDF* векторы, используя TfidfVectorizer из библиотеки sklearn. Слова обрабатываются с учётом частоты их появления в каждом комментарии и во всём наборе данных. В результате была получена итоговая таблица, где демонстрируется значимость слов, основываясь на их весах *TF-IDF*. Слова с наибольшими значениями *TF-IDF* являются наиболее важными в обсуждении и характеризуют тему диалога.

Таблица 3.

Итоговые значения *TF-IDF*

Слово	TF-IDF
Интеллект	1.456
Искусственный	1.2341
Люди	1.1015
Человек	0.9874
Заменит	0.8762
Рабства	0.7654
Задачи	0.6543
Хозяйства	0.5432
Компании	0.4321
Данные	0.3210

Можно заметить, что анализ *TF-IDF* выделяет наиболее важные слова в комментариях, но не может классифицировать, является комментарий спамом или нет.

Для анализа методом Naive Bayes необходимо создание обучающего набора данных, в котором комментариям присваиваются метки «спам» и «не спам». Следующим шагом преобразовываем текст в числовые признаки при помощи CountVectorizer и обучаем модель на размеченных данных. В результате получаем итоговую таблицу с определением спам и не спам комментариев.

Таблица 4.

Значения *TF-IDF*

Текс комментария	Метка
Нейросеть уже сама коды пишет и намного лучше, чем люди-айтишники	Не спам
Люди, воспитанные в духе рабства, будут бояться потерять работу	Не спам
Может ли ИИ заменить инфоцыган?	Не спам
Присоединяйтесь к нашему бесплатному вебинару и не дайте себя заменить	Спам
Калькулятор тоже считает быстрее и точнее человека, но это не значит, что он умнее	Не спам
Мне ничего не грозит, я печник	Не спам
Всё по плану	Спам

Naive Bayes показал высокую точность в классификации комментариев по заданным меткам. Спам-комментарии корректно были определены, а остальные отнесены к категории «не спам».

Подводя итог сравнительного анализа по нахождение спам комментариев под постом в социальной сети вконтактее, можно сделать вывод, что *TF-IDF* эффективен для анализа содержания текста и выявления ключевых слов, превосходит Naive Bayes в задачах анализа тематики и выделения важных понятий в тексте. Однако он не пригоден для фильтрации спама, в то время как Naive Bayes отлично справляется с классификацией комментариев и превосходит *TF-IDF*, когда требуется именно классификация на «не спам» или «спам».

В рамках данного исследования были проведены сравнительные анализы двух популярных методов обработки текста — *TF-IDF* и Naive Bayes. Метод *TF-IDF* оказался эффективным для выявления ключевых слов в текстах, что позволяет оценить значимость каждого слова в контексте всей коллекции документов.

Метод Naive Bayes продемонстрировал свою силу в классификации текстов на категории «спам» и «не спам». С помощью этого метода была создана модель, обученная на наборе комментариев, что позволило эффективно различать спамовые сообщения от нормальных. Результаты классификации подтвердили высокую точность модели, что делает Naive Bayes подходящим инструментом для задач фильтрации контента в реальных приложениях.

Оба метода имеют свои особенности и преимущества. Комбинирование этих подходов может предложить еще более глубокое понимание и более точную обработку текстовых данных.

ЛИТЕРАТУРА

1. Акбархужаев С.А. Сравнительный анализ методов Наивного Байеса и SVM алгоритмов при классификации текстовых документов — «Молодой ученый» № 29 (267). — с. 8–10.
2. Батура Т.В. Методы автоматической классификации текстов // Программные продукты и системы. — 2017. — Т. 30, № 1. — С. 85–99.
3. Корюкин А.В. Исследование влияния настроек TF-IDF векторизации текста на результаты бинарной классификации тональности — «Математические методы в технологиях и технике» №5, — 2021 — с. 126–130.
4. Мельниченко С.С. Анализ основных методов веб-фильтрации контента на примере detox browzer и алгоритмов машинного обучения — «Экономика и качество системы связи» №3, — 2022 — с. 60–66.
5. Мутаиро Ш.И., Бушмелева К.И. Алгоритмы обработки и вычисления сходства текстовых данных пользователей социальных сетей — «Успехи кибернетики» №4, — 2023 — с.33–38.
6. Сабуров В.С. Байесовский классификатор в машинном обучении // Шаг в науку. — 2024. — № 1. — С. 78–81.
7. Сидорова Е.А., Кононенко И.С., Загорулько Ю.А. Подход к фильтрации запрещенного контента в веб-пространстве — Труды XIX Международной конференции «Аналитика и управление данными в областях с интенсивным использованием данных», Москва, Россия, 10–13 октября 2017 г.
8. Wang Bin, Si Yang Tao, Fu Jun Tao. News classification based on improved TF-IDF and Bayesian algorithm [J]. Science and technology wind, 2020 (31): 9–10.
9. Sharma N., Singh M. Modifying Naive Bayes Classifier for Multinomial Text Classification. 2016 International Conference on Recent Advances and Innovations in Engineering (ICRAIE). IEEE., 2016 p. 1–7.

Журнал «Современная наука: актуальные проблемы теории и практики»