ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНЫХ ТЕХНОЛОГИЙ ДЛЯ ПОВЫШЕНИЯ ЭФФЕКТИВНОСТИ КОНТЕНТНОЙ ФИЛЬТРАЦИИ

Бахрачева Ю.С.

Волгоградский государственный университет,
Россия, г. Волгоград, Университетский пр., д. 100
bakhracheva@volsu.ru

Аннотация: Целью работы является повышение эффективности контентной фильтрации. Предложен программный комплекс контентной фильтрации с применением интеллектуальных технологий Проведён анализ результатов экспериментов, в результате которого был сделан вывод, что предлагаемый подход, по сравнению с классическим, обладает большей точностью классификации и следовательно эффективностью контентной фильтрации.

Ключевые слова: информационная безопасность, контентная фильтрация, интеллектуальные технологии.

Введение

Для эффективной борьбы с распространением опасной информации, помимо законодательных инициатив, необходима реализация средств фильтрации, которые позволят отсеивать запрещенную информацию из общего потока пользовательских данных. Информация в глобальной сети представлена в виде контента, который потребляется ее пользователями, и именно применение контентной фильтрации является основным способом решения данной проблемы.

Целью работы является повышение эффективности контентной фильтрации.

Фильтрацию осуществляет контент-фильтр — устройство или программное обеспечение, которое реализует алгоритмы контентной фильтрации.

Контент-фильтры для своей работы могут использовать различные алгоритмы, отличающиеся своей сложностью, эффективностью, временем анализа и другими характеристиками. Основными являются следующие алгоритмы контентной фильтрации:

1) использование черных и белых списков;
2) блокировка поиска по ключевым словам/фразам;
3) использование предопределённых баз категорий ресурсов;
4) использование машинного обучения;
5) поиск и анализ регулярных выражений;
6) статистический анализ

Под интеллектуальными технологиями понимают такие информационные технологии, в которых предусмотрены следующие возможности:

- Наличие баз знаний, отражающих опыт конкретных людей, групп, обществ, человечества в целом, при решении творческих задач в определенных сферах деятельности, традиционно считавшихся прерогативой интеллекта человека (например, такие плохо формализуемые задачи, как принятие решений, проектирование, извлечение смысла, объяснение, обучение и прочие).
- Наличие моделей мышления на основе баз знаний: правил и логических выводов, аргументации и рассуждения, распознавания и классификации ситуаций, обобщения и понимания.
- Способность формировать достаточно четкие решения на основе нечетких, нестрогих, неполных данных.
- Способность объяснять выводы и решения, то есть наличие механизма объяснений.
- Способность к обучению, переобучению и, следовательно, к развитию.[1]

Исходя из приведенного перечисления можно сделать вывод, что машинное обучение обладает большинством из этих возможностей, и это позволяет отнести его к группе интеллектуальных технологий.

Уже сейчас машинное обучение широко используется в интеллектуальном анализе данных (datamining), при классификации и кластеризации различных объектов, к которым так же относятся тексты, изображения, звуки и прочее, что в совокупности представляет собой контент, предлагаемый пользователям сети Интернет. Именно применение интеллектуальных технологий позволит автоматически классифицировать веб-ресурсы, как нежелательные, или безопасные методами машинного обучения и осуществлять фильтрацию пользовательского контента.

1 Разработка математической модели контентной фильтрации с применением интеллектуальных технологий

Задачей разрабатываемой модели является реализация, оценка и сравнение эффективности описанных ранее подходов к контентной фильтрации.

В общем виде процесс контентной фильтрации можно представить в виде функциональной модели (рисунок 1).

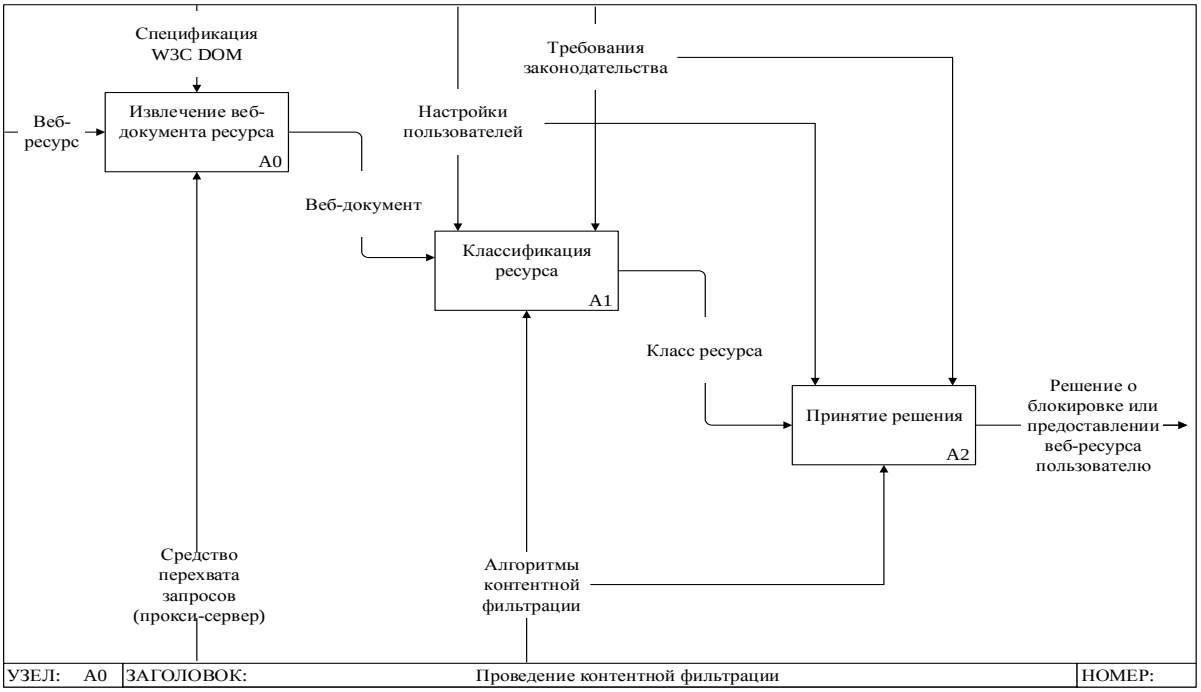

Рис. 1. Функциональная модель контентной фильтрации данных

Реализуемое программное средство должно осуществлять фильтрацию путём классификации веб-ресурсов, получаемых пользователем из сети [2, 3]. То есть получать значение:

$$K = \{k1, k2\} \quad (1)$$

где: K – класс, к которому относится анализируемый документ; k_1 – класс допустимых документов; k_2 – класс запрещённых документов.

Любой веб-документ, который был получен из сети представляет собой объект

$$D = (A, T) \quad (2)$$

Где: A –адрес источника документа; T – текст, содержащийся в документе.Для определения показателей эффективности рассматриваемых подходов к фильтрации контента, классификация веб-ресурсов должна осуществляться по отдельности для каждого из них.

Математическая модель классической контентной фильтрации:

В классической модели контентной фильтрации данных сперва осуществляется поиск веб-адреса источника документа в списках разрешённых и запрещённых адресов, осуществляемый функцией:

$$X(A) = x \quad (3)$$

где: x – флаг, обозначающий наличие искомого адреса в базе.

В результате происходит классификация веб-документа:

$$f = \begin{cases} k_1, \text{true} \\ k_2, \text{false} \\ \text{null}, \text{undefined} \end{cases} \quad (4)$$

В случае, если x = undefined, продолжаем анализ документа.

1) Выполняется поиск ключевых слов в тексте документа при помощи функции:

$$S(T, Q) = s \quad (5)$$

где: Q – множество ключевых слов, определённых в базе настроек программы; s – флаг, обозначающий наличие в тексте анализируемого документа ключевых слов.

Далее выполняется классификация веб-документа:

$$s = \begin{cases} k_1, \text{false} \\ k_2, \text{true} \end{cases} \quad (6)$$

В зависимости от значения K принимается решение о блокировании контента, или предоставлении его пользователю.

Предлагаемая математическая модель контентной фильтрации данных:

1) В предлагаемой модели контентной фильтрации классификация документа начинается с функции

$$F(A) = f \quad (7)$$

где:F(A) - функция поиска адреса источника анализируемого документа в базе категорий веб-ресурсов и базе настроек программы; f – флаг, обозначающий наличие искомого адреса в базе.

В результате, как и при классическом подходе, происходит попытка классификации веб-документа:

$$f = \begin{cases} k_1, \text{true} \\ k_2, \text{false} \\ \text{null}, \text{undefined} \end{cases} \quad (8)$$

Если f = undefined, продолжаем анализ документа.

2) Применяем семантический анализ – процесс выявления смыслового содержания слов и словосочетаний в предложении. Он обеспечивает нормализацию синтаксической структуры предложений, распознавание терминов, классификацию терминов по семантическим признакам, с учетом синонимических и гипонемических (отношение «общее – частное») классов [5]. Семантический анализ содержимого веб-ресурса в данном случае реализуется при помощи инструмента Word2Vec, принцип работы которого основан на использовании одного из методов машинного обучения – искусственной нейронной сети и позволяет осуществлять автоматическую классификацию сайтов при проведении контентной фильтрации [5]. Она состоит из трех слоев (рисунок 2):

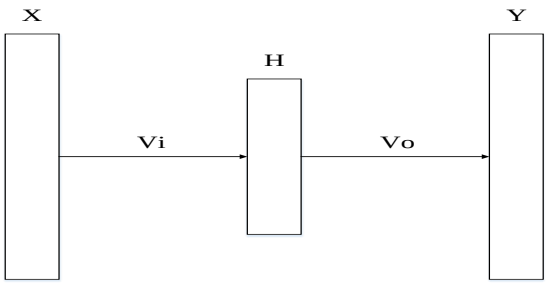

Рис. 2. Схема нейронной сети Word2Vec

- размер первого слоя X равен длине поступающего на вход вектора;
- скрытый слой $H = X * V_i$ имеет линейную функцию активации;
- выходной слой $Y = \text{softmax}(H * V_o)$ имеет функцию активации softmax, его размер равен размеру входного слоя X.

Последовательно выполняются кластеризация и векторизация T средствами Word2Vec, путем подачи его на вход нейронной сети.

Функция P(T) разбивает текст документа по кластерам:

$$P(T)=(c_1, c_2, c_3 \dots c_i) \quad (9)$$

где: i – количество кластеров; c_i– множество термов, принадлежащих i-му кластеру.

Функция$V(c_i)$ векторизует полученные кластеры:

$$V(c_i) = \overline{c_i} \quad (10)$$

где: $\overline{c_i}$–вектор, описывающий расстояние от соответствующего кластера до центра класса в векторном пространстве.

Семантический вектор всего текста получается путем сложения всех векторов [5].

$$\bar{t} = \sum_{1}^{i} \bar{c_i} \quad (11)$$

3) В результате вычисляется косинусное сходство между $\bar{t}$ и эталонными векторами $\bar{e_1}$ и $\bar{e_2}$ классов k_1и k_2соответственно. К классу с большим сходством относится анализируемый документ.

$$K = \begin{cases} k_1, \dfrac{\bar{t} * \bar{e_1}}{\|t\|\|e_1\|} \\ k_2, \dfrac{\bar{t} * \bar{e_2}}{\|t\|\|e_2\|} \end{cases} \quad (12)$$

В зависимости от значения K принимается решение о блокировании контента, или предоставлении его пользователю.

Полученные результаты работы программного средства, включая время анализа веб-ресурсов, используются для вычисления показателей эффективности каждого из рассматриваемых подходов.

2 Эксперимент

Задачей экспериментального исследования является получение значений эффективности рассматриваемых подходов к контентной фильтрации при помощи программы разработанной и описанной ранее.

Для всех подходов к фильтрации контента планируется одинаковый перечень экспериментов с последующим расчетом и сравнением их эффективности.

Планируемые эксперименты:

- Для классического подхода:

1) Анализ веб-документа с отмеченным в черном списке адресом;

2) Анализ 50 веб-документов с отсутствующими в списках адресами, и содержащих разрешенный контент;

3) Анализ 50 веб-документов с отсутствующими в списках адресами, и содержащих запрещенный контент;

4) Расчёт эффективности классического подхода.

- Для предлагаемого подхода:

5) Анализ веб-документа с отмеченным в черном списке адресом;

6) Анализ 50 веб-документов с отсутствующими в списках адресами и содержащих разрешенный контент;

7) Анализ 50 веб-документов с отсутствующими в списках адресами и содержащих запрещенный контент;

8) Расчёт эффективности контентной фильтрации с применением интеллектуальных технологий.

3 Обсуждение результата экспериментов

Было проведено 8 экспериментов, разделённых на две группы, в результате которых получены следующие результаты:

Для классического подхода к контентной фильтрации:

- среднее значение времени, необходимое контент-фильтру для анализа веб-ресурса в классическом режиме равно 1,92 секунды;
- количество верно распознанных веб-ресурсов равно 67;
- общее количество ошибок фильтрации равно 34;
- эффективность классического подхода к контентной фильтрации данных 1,02.
- Для предлагаемого подхода к контентной фильтрации:
- среднее значение времени, необходимое контент-фильтру для анализа веб-ресурса при предлагаемом подходе равно 4,17 секунд;
- количество верно распознанных веб-ресурсов равно 91;
- общее количество ошибок фильтрации равно 10;
- эффективность предлагаемого подхода к контентной фильтрации данных равна 2,18.

Проанализировав полученные результаты, можно прийти к выводу, что предлагаемый подход, по сравнению с классическим, при увеличении времени анализа веб-ресурсов, обладает большей точностью классификации и, следовательно, эффективностью контентной фильтрации.

Заключение

Проанализировано применение интеллектуальных технологий в области контентной фильтрации. Установлено, что применение интеллектуальных технологий позволит автоматически классифицировать веб-ресурсы, как нежелательные, или безопасные методами машинного обучения и осуществлять фильтрацию пользовательского контента.

Определено понятие эффективности контентной фильтрации. Под эффективностью контентной фильтрации понимается соотношение между количеством верных срабатываний фильтра, общим количеством анализируемых веб-документов и затрачиваемым на анализ временем.

Разработана математическая модель контентной фильтрации. Ее задачей является реализация, оценка и сравнение эффективности описанных ранее подходов к контентной фильтрации.

Поставлена задача на проведение экспериментальных исследований:

Задачей экспериментального исследования является получение значений эффективности рассматриваемых подходов к контентной фильтрации при помощи программы разработанной и описанной ранее.

Проведён анализ результатов экспериментов, в результате которого был сделан вывод, что предлагаемый подход, по сравнению с классическим обладает большей точностью классификации.

Литература

1. *Андреев Л. Ю.* Законодательное и нормативно-правовое обеспечение функционирования закона «О защите детей от информации, причиняющей вред их здоровью и развитию» в сети Интернет// Молодой ученый. — 2016. — №6.1. — С. 4-7. — URL: https://moluch.ru/archive/110/27047/ (дата обращения: 20.02.2021)
2. *Стрекалов И.Э.* Система формирования безопасного контента// Вестник Тамбовского университета. Серия: Естественные и технические науки.— 2015.— №2— С.462-464.
2. *Аносов А.Е.* Методы фильтрации «Стихийного» трафика в динамических интернет-ресурсах // Вестник РГГУ. Серия «Документоведение и архивоведение. Информатика. Защита информации и информационная безопасность». —2013. —№14 (115).—URL: https://cyberleninka.ru/article/n/metody-filtratsii-stihiynogo-trafika-v-dinamicheskih-internet-resursah-1 (дата обращения: 07.03.2021).
3. *Чемодуров А.С., Карпутина А.Ю.*Обзор средств фильтрации трафика в корпоративной сети// Концепт. — 2015. —№2.— URL: https://cyberleninka.ru/article/n/obzor-sredstv-filtratsii-trafika-v-korporativnoy-seti (дата обращения: 18.02.2021).
4. *Соловьев Н.А., Чернопрудова Е.Н.* Формирование устойчивых словосочетаний в задаче контентной фильтрации электронных сообщений // Вестник ОГУ. — 2013. — №11 — С.84-90.