Слайд 1. Тема ВКР

Слайд 2. Цель и задачи

Количество пользователей в интернете растёт с каждым днём, а вместе с этим увеличивается и количество выкладываемых изображений на веб-ресурсах. В большинстве случаев такой контент не проходит должную проверку и в сети появляются материалы деструктивного содержания, которые находятся в открытом доступе. Существующие средства фильтрации криминального материала имеют ряд недостатков:

- либо требуют больше аналитиков, которые могут ошибаться в следствии человеческого фактора
- либо имеют ограниченный функционал (проверка URL или ключевых слов)
- или недоступны простым пользователям сети (гос. Средства).

Поэтому, в рамках этой работы, была поставлена цель: разработать расширение для браузера, анализирующего посещаемый веб-ресурс на криминальный контент с применением средств машинного обучения и туманных вычислений.

Для реализации такой системы были выделены следующие задачи:

1. Обзор моделей машинного обучения

2. Обучение выбранной модели под задачи данной работы

3. Проектирование архитектуры системы

4. Исследование разработанной системы

Слайд 3. Обзор моделей классификации изображений

Использование методов машинного обучения в анализе контента позволяет быстрее и надежнее определять деструктивное содержание на изображениях, чем ручной мониторинг ресурсов.

Существует множество различных видов нейронных сетей, которые ис-

пользуются для анализа иллюстраций, но наиболее распространёнными считаются сверточные нейронные сети - это тип моделей для классификации

изображений. Они специально разработаны для обработки многомерных данных, таких как изображения.

Среди сетей для классификации изображений выделяют 6 самых популярных видов: EfficientNet, VGG16, InseptionV3, ResNet50, MobileNetV2.

Опираясь на результаты сравнения этих моделей между собой, можно выделить нейронную сеть EfficientNet, которая обладает самой высокой точностью классификации. Именно она была использована в разработке, поскольку в работе показатель точности анализа является главным приоритетом.

Слайд 4. Обучение модели EfficientNet

Чтобы адаптировать выбранную модель под задачи данной работы, было принято решение о трансферном обучении.

Трансферное обучение - это метод обучения глубоких нейронных сетей, который позволяет использовать знания, полученные при обучении на одной задаче, для решения другой задачи. Вместо обучения сети с нуля на новом наборе данных, трансферное обучение позволяет использовать заранее обученную модель как базовую, затем модифицировать ее для решения новой задачи.

В данном случае была использована модель EfficientNet с весами, предварительно обученными на датасете ImageNet. (набор более 14 млн изображений разбитых на более чем 21 тыс. классов).

Затем, для адаптации под задачи поиска негативного материала, собран набор данных, включающий изображения деструктивного содержания в категориях: алкоголь, наркотики, порнография, оружие, а также иллюстрации не криминального вида.

После, на этом наборе, EfficientNet была натренирована классифицировать изображения в пяти упомянутых категориях и достигла точности в 92% на проверочных данных.
Обучение нейронной сети происходило на базе фреймворка тензорфлоу с помощью языка программирования Python. и чтобы иметь возможность использовать переобученную модель в браузере, библиотека предоставляет методы импорта данных из исходного формата HDF5 в JSON.

(Hierarchical Data Format, HDF (Иерархический формат данных) — название формата файлов, разработанного для хранения большого объема цифровой информации. Первоначально был разработан Национальным центром суперкомпьютерных приложений, сейчас поддерживается некоммерческой организацией HDF Group.

JSON (англ. JavaScript Object Notation, обычно произносится как /ˈdʒeɪsən/ JAY-sən[3]) — текстовый формат обмена данными, основанный на JavaScript. Как и многие другие текстовые форматы, JSON легко читается людьми. Формат JSON был разработан Дугласом Крокфордом[4].)

Слайд 5. Архитектура системы

Программная реализация расширения подразумевает клиент-серверную архитектуру. Всего имеется 3 модуля:

1. Расширение: оно реализовано как часть браузера и имеет доступ к загружаемым в браузер страницам. Также модуль ответственен за сбор информации о загруженной странице и изображении, с которым произошло взаимодействие пользователя. Далее эта информация передаётся сервис-воркеру для последующей обработки. Расширение имеет интерфейс пользователя, который будет отображать результаты анализа иллюстрации на наличие криминального контента.

2. Сервис-воркер: он предоставляет API расширению для связи с сервером и управляется событиями расширения.

3. Сервер: он хранит модели машинного обучения и использует их при анализе страниц. Также на сервере происходит обработка изображения, что бы его можно было передать на вход нейронной сети.

Слайд 6. Интеграция с облачными технологиями

Поскольку выбранная модель машинного обучения EfficientNet обладает большой задержкой в обработке данных, было проведено тестирование разработанного продукта при интеграции с облачными технологиями.

Исследование проводилось на личном компьютере Acer aspire с Озу в 8гб, двух облачных средах Яндекс с озу 12 и 20 гб соответственно, а также в среде гугл коллаб с графическим процессором. Синим цветом представлено время, затраченное на предобработку изображения перед анализом, а красным указано время, за которое программа классифицирует изображение.

Мы видим, что применение облачных сред с большими вычислительными мощностями помогает снизить задержки предсказания в 10 раз, по сравнению с размещением разработанного ПО на персональном компьютере.

Слайд 7. Заключение

Все поставленные задачи были достигнуты:

Были изучены существующие технологии машинного обучения для анализа и обработки изображений.

В рамках сравнения нейронных сетей была выбрана модель, обладающая наибольшей точностью предсказаний.

В процессе работы была разработана архитектура программной системы, включающая расширение для браузера, сервис-воркер и сервер.

Тестирование показало, что разработанное расширение обладает высокой точностью и эффективностью в обнаружении негативного контента на

изображениях,а также обладает возможностью делать предсказания менее чем за 2 секунды.

Слайд 8. Дальнейшее развитие

Существует несколько способов дальнейшего развития разработанного

расширения для браузера. Некоторые из них включают:

1. Расширение функциональности: можно добавить дополнительные

возможности и функции в расширение. Будет полезным внедрить возможности анализа текстового контента и мониторинг нескольких изображений за единицу времени.

2. Улучшение точности и производительности: работа над оптимизацией моделей машинного обучения и алгоритмов анализа может помочь улучшить точность обнаружения негативного контента и снизить ложные срабатывания. Также можно работать над оптимизацией производительности для ускорения процесса анализа изображений и улучшения отзывчивости расширения.

3. Расширение может быть адаптировано и расширено для поддержки разных браузеров и платформ.

4. Оптимизация и улучшение пользовательского интерфейса расширения поможет сделать его более интуитивно понятным и удобным в использовании.

5. Регулярное обновление моделей машинного обучения и базы данных негативного контента поможет улучшить работу расширения и его способность обнаруживать новые типы негативного содержания.

Слайд 9. Апробация

Исходный код разработанного расширения доступен на гитхаб по ссылке и Qr-коду на слайде. В описании репозитория имеется подробная иснтрукция по установке и запуску программы.