

ОСНОВНЫЕ ВОПРОСЫ РАЗРАБОТКИ МОДУЛЯ ОПРЕДЕЛЕНИЯ СПЕЦИАЛЬНОЙ СИМВОЛИКИ НА ИЗОБРАЖЕНИЯХ СОЦИАЛЬНОЙ СЕТИ

Колесников И.Е.¹, Смирнова О.С.² Email: Kolesnikov17106@scientifictext.ru

¹Колесников Илья Евгеньевич – магистрант,
Московский технологический университет;

²Смирнова Ольга Сергеевна – старший специалист,
Центр разработки специального программного обеспечения ЗАО «НПК «ВТ и СС»,
ассистент,
кафедра инструментального прикладного программного обеспечения,
Московский технологический университет,
г. Москва

Аннотация: в данной статье рассматриваются основные вопросы разработки программного модуля для распознавания образов на изображениях из социальной сети «ВКонтакте» с целью выявления причастности пользовательских профилей к квестовой игре суицидального характера «Синий кит». Рассматриваются процессы проектирования и реализации программного модуля. В статье представлены интерфейс и показатели точности работы модуля.

Исследование выполнено федеральным государственным бюджетным образовательным учреждением высшего образования «Московский технологический университет» (МИРЭА) в рамках гранта Российского фонда фундаментальных исследований (проект № 16-37-00492).

Ключевые слова: анализ изображений; распознавание образов; искусственные нейронные сети; сверточные нейронные сети; анализ социальных сетей.

THE MAIN ISSUES OF DEVELOPING A MODULE FOR DETERMINING SPECIAL SYMBOLS ON SOCIAL NETWORK PICTURES

Kolesnikov I.E.¹, Smirnova O.S.²

¹Kolesnikov Ilya Evgenyevich – Master's Student,
MOSCOW TECHNOLOGICAL UNIVERSITY;

²Smirnova Olga Sergeevna – Senior Specialist,
SPECIAL SOFTWARE DEVELOPMENT CENTER ZAO NPK "VT AND SS",
ASSISTANT,
DEPARTMENT OF INSTRUMENTAL APPLIED SOFTWARE,
MOSCOW TECHNOLOGICAL UNIVERSITY,
MOSCOW

Abstract: this article discusses the main issues of developing a software module for image recognition on images from the social network "VKontakte" in order to identify the involvement of user profiles in the suicidal quest game "Blue Whale". The processes of design and implementation of the software module are considered. The article presents the interface and indicators of the accuracy of the module.

The study was carried out by the federal state budget educational institution of higher education "Moscow Technological University" (MIREA) in the framework of the grant of the Russian Foundation for Basic Research (project No. 16-37-00492).

Keywords: image analysis; pattern recognition; artificial neural networks; convolutional neural networks; analysis of social networks.

УДК 004.4
ВАК 05.13.11

Введение

Одна из актуальных на сегодняшний день тем для анализа данных является анализ открытых данных социальных сетей [1], в частности изображений [2]. Изображения в социальной сети часто содержат в себе информационные образы или символику, относящиеся к тем или иным явлениям.

Примером задачи для анализа изображений социальных сетей является задача распознавания символики квестовой игры суицидального характера «Синий кит», имевшей популярность в конце 2016 – начале 2017 гг. в социальной сети «ВКонтакте» [3]. С целью выявления пользователей и сообществ, распространяющих такие изображения, спроектирован и разработан программный модуль для анализа изображений профилей социальных сетей. Основные задачи данного модуля:

- получение всех доступных изображений профиля;
- определение причастности профиля к квестовой игре «Синий кит» посредством распознавания соответствующей символики на изображениях [4].

Проектирование

Разработка представляет собой веб-сервис и является модулем общей системы анализа данных социальной сети «ВКонтакте» [5 – 12]. Таким образом, система должна обеспечивать возможность взаимодействия с ней при помощи HTTP запросов. Также, для удобства работы с системой, необходимо реализовать пользовательский интерфейс.

Разрабатываемая система состоит из клиентской части, которая включает в себя пользовательский интерфейс в виде веб-страницы, и серверной части, которая включает в себя сервер и компонент нейронной сети, отвечающий за анализ изображений на предмет наличия символики игры «Синий кит». В свою очередь сервер взаимодействует с клиенткой частью посредством интерфейса HTTP запросов, а также с сервером «ВКонтакте» посредством интерфейса HTTP запросов к открытому API, а также запросов для загрузки файлов. Схема развертки приложения представлена на рисунке 1.

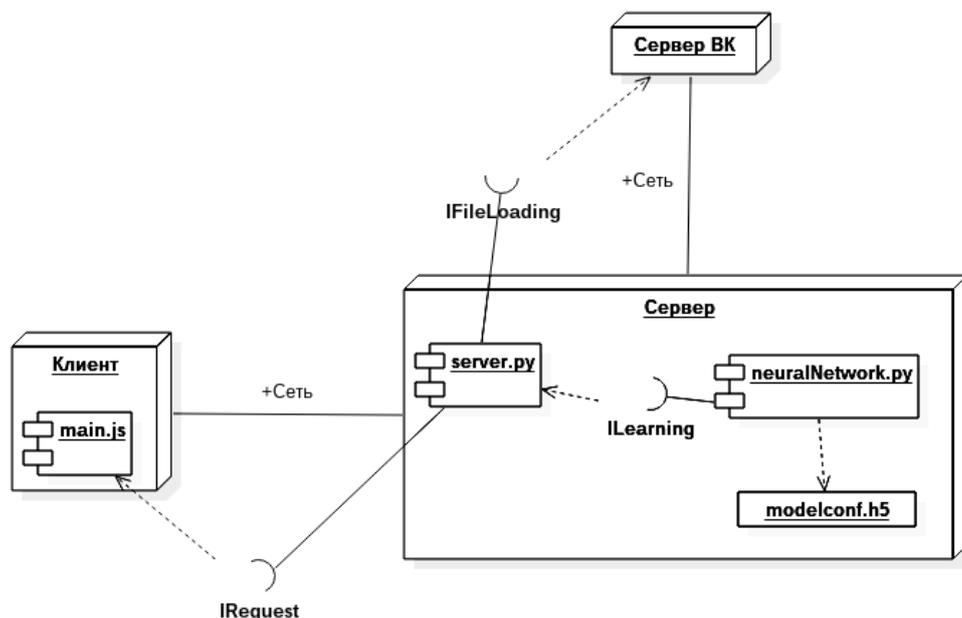


Рис. 1. Диаграмма развертывания

На данном рисунке файлы main.js, server.py, и neuralNetwork.py представляют собой компоненты клиента, сервера, и нейронной сети соответственно. Файл modelconf.h5 представляет собой образ модели нейросети, который можно сохранять после дообучения сети. Интерфейсы IRequest, IFileLoading и ILearning представляют собой интерфейсы HTTP-запросов, загрузки файлов и обучения нейросети соответственно.

Разработка

В качестве платформы для реализации пользовательского интерфейса была выбрана платформа веб-технологий (HTML5, CSS3, JavaScript). Для эффективной реализации работы с HTTP запросами с сервером используются средства, являющиеся современным стандартом возможностей JavaScript и HTML5 API – Fetch API и Promises.

Модуль сервера взаимодействует со всеми остальными компонентами системы, т.е. является «точкой входа» системы, и поэтому этот модуль выполняет большое количество функций в процессе работы, которые необходимо рассмотреть.

Сервер взаимодействует с клиентом посредством приема и передачи HTTP запросов, тем самым этот сервер должен выполнять функции HTTP-сервера. Для реализации этой задачи используется микрофреймворк Flask, который позволяет быстро и эффективно создать HTTP-сервер с поддержкой REST API.

Для реализации модуля нейронной сети, которая позволяет распознавать изображения, требуется создать сверточную нейронную сеть. Сверточная нейронная сеть является нейросетью с наиболее эффективной топологией для распознавания изображений [13], что необходимо для работы разрабатываемой системы.

В качестве средства для реализации сверточной нейронной сети была выбрана библиотека Theano. Данная библиотека позволяет создавать модели нейросетей, показатели быстродействия которых являются одними из самых высоких. Для более эффективной и простой работы с ней выбрана библиотека Keras, которая является высокоуровневым интерфейсом для основной библиотеки нейронной сети.

В качестве конфигурации нейронной сети выбраны следующие параметры:

- количество входных нейронов – 10;
- размер ядра свертки – 3x3;
- количество слоев – 8;
- количество эпох – 50.

Результаты разработки

На рисунке 2 представлен разработанный пользовательский интерфейс.

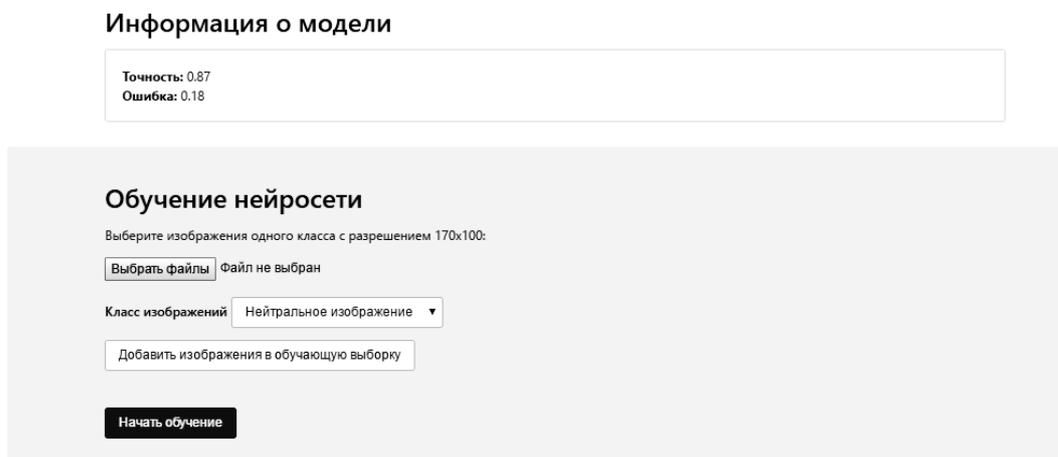


Рис. 2. Пользовательский интерфейс

При открытии веб-страницы пользователь в первой секции страницы видит данные о точности и ошибке модели нейросети.

Для обучения нейросети пользователь должен выбрать файлы во второй секции, выбрать класс изображения в выпадающем списке, и нажать на кнопку «Добавить фотографии». После добавления нужных фотографий следует нажать на кнопку «Начать обучение» для обучения модели.

Для обработки профиля в социальной сети пользователь должен ввести адрес в текстовое поле в третьей секции (показана на рисунке 3), выбрать тип профиля – группа или пользователь, и нажать на кнопку «Обработать».

В итоге, снизу будет выведен список изображений с результатом оценки наличия на изображении искомой символики, а также данные об аналогичной оценке группы или пользовательского профиля в целом. Пример такого отображения результата представлен на рисунке 3.

Обработка группы



Рис. 3. Пример результата обработки группы

На текущий момент, для обучающей выборки собрано суммарно 182 изображения. Выборка состоит из изображений 2-х классов: нейтральный образ и образ синего кита.

С такой выборкой получены следующие результаты модели:

- точность – 0.737;
- ошибка – 0.497.

Относительно невысокая точность и высокий показатель ошибки связаны с тем, что собрана слишком малая обучающая выборка, так как большинство изображений с символикой, присущей игре «Синий кит», на данный момент удалено из социальной сети «ВКонтакте». Однако, модель нейросети способна дообучаться, поэтому, при дополнении обучающей выборки, точность будет расти.

Заключение

Дальнейшее обучение разработанного модуля определения специальной символики на изображениях социальной сети будет способствовать уменьшению ошибки и увеличению точности распознавания. Использование разработанного модуля как отдельно, так и в составе системы анализа данных социальной сети, позволит выявлять на основе распознанных изображений определенные группы и пользователей, принадлежащих к различным тематическим направлениям.

Список литературы / References

- 1 *Smirnova Olga, Petrov Alexey, Babiychuk Georgy*. Common techniques for social network analysis study // Selected Papers of the XI International Scientific-Practical Conference Modern Information Technologies and IT-Education (SITITO 2016). Moscow, Russia, November 25-26, 2016. P. 262–269.
- 2 *Ишин И.А., Смирнова О.С.* Технологии анализа изображений в рамках исследования социальных сетей // Российская научная конференция «Интеллектуальные системы в информационном противоборстве». М.: МИРЭА, 22 – 24 ноября 2016 г.
- 3 Синий кит (игра). Википедия [Электронный ресурс] // Режим доступа: [https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BD%D0%B8%D0%B9_%D0%BA%D0%B8%D1%82_\(%D0%B8%D0%B3%D1%80%D0%B0\)](https://ru.wikipedia.org/wiki/%D0%A1%D0%B8%D0%BD%D0%B8%D0%B9_%D0%BA%D0%B8%D1%82_(%D0%B8%D0%B3%D1%80%D0%B0))/ (дата обращения: 11.06.2017).
- 4 *Смирнова О.С., к.т.н. Чумак Б.Б., Колесников И.Е.* Распознавание информационных образов за счет анализа изображений социальной сети // НТК ИИТ МГУ МИРЭА, 16 мая 2017 г.
- 5 *Смирнова О.С., Шишков В.В.* Выбор топологии нейронных сетей и их применение для классификации коротких текстов // International Journal of Open Information Technologies. Том 4. № 8, 2016. С. 50–54.
- 6 *Смирнова О.С., Алымов А.С., Баранюк В.В.* Детектирование бот-программ, имитирующих поведение людей в социальной сети «ВКонтакте» // International Journal of Open Information Technologies. Том 4. № 8, 2016. С. 55–60.
- 7 *Смирнова О.С., Баранюк В.В., Десяткова А.Д.* Подходы к определению психоэмоциональных особенностей информационного образа пользователя социальных сетей // International Journal of Open Information Technologies. Том 4. № 8, 2016. С. 61–65.
- 8 *Петров А.И., Смирнова О.С.* Анализ аудио- и видеообъектов пользовательского пользователя социальной сети // Российская научная конференция «Интеллектуальные системы в информационном противоборстве». М.: МИРЭА. 22 – 24 ноября 2016 г.
- 9 *Бабийчук Г.А., Смирнова О.С.* Анализ социальных графов и определение тональности текстовых сообщений пользователей // Российская научная конференция «Интеллектуальные системы в информационном противоборстве». М.: МИРЭА. 22 – 24 ноября 2016 г.
- 10 *Петров А.И., Смирнова О.С., Чумак Б.Б.* Анализ контента социальной сети на примере квестовой игры суицидального характера, направленной на детей и подростков // International Journal of Open Information Technologies. Том 5. № 6, 2017. С. 16–19.
- 11 *Смирнова О.С., Шишков В.В.* Графовый подход при составлении характеристики социального объекта // International Journal of Open Information Technologies. Том 5. № 6, 2017. С. 12–15.
- 12 *Алымов А.С., Смирнова О.С., к.т.н., с.н.с. Баранюк В.В.* Определения группы риска аккаунтов социальной сети «ВКонтакте», попадающих под влияние квестовой игры суицидального характера // НТК ИИТ МГУ МИРЭА, 16 мая 2017 г.
- 13 *LeCun Y. and Bengio Y.* Convolutional Networks for Images, Speech, and Time-Series, in Arbib, M.A. (Eds), The Handbook of Brain Theory and Neural Networks. MIT Press, 1995.