

# Определение личностных черт у пользователей Вконтакте на основе анализа изображений\*

Н. А. Игнатъев<sup>I</sup>, М.А. Станкевич<sup>II</sup>, Н. В. Кисельникова<sup>III</sup>, О. Г. Григорьев<sup>II</sup>

<sup>I</sup> Российский университет дружбы народов, г. Москва, Россия

<sup>II</sup> Федеральное государственное учреждение "Федеральный исследовательский центр "Информатика и управление" Российской академии наук, г. Москва, Россия

<sup>III</sup> Психологический институт РАО, г. Москва, Россия

**Аннотация.** Психометрический анализ информации из интернета является одним из быстрорастущих трендов современных исследований. На основе интеллектуального анализа данных из социальных сетей выявляются психологические особенности пользователей социальных сетей и наличие у них признаков психологического неблагополучия. В настоящей работе решается задача выявления личностных черт – предикторов депрессии, у пользователей социальной сети Вконтакте на основе интеллектуального анализа публикуемых ими изображений. Описаны предлагаемые методы решения задачи и результаты экспериментальной проверки методов на данных из сети Вконтакте. Наши исследования показали, что учет наличия объектов на изображениях позволяет решать задачу с наилучшим уровнем качества.

**Ключевые слова:** личностные черты, социальные сети, анализ изображений, машинное обучение.

DOI 10.14357/20718594190404

## Введение

Психометрический анализ информации из социальных сетей является одним из трендов современных научных исследований. Такой анализ позволяет выявлять, например, психологические особенности пользователей социальных сетей, наличие у них признаков психологического неблагополучия и психических заболеваний. Решение указанных задач востребовано в таких отраслях, как здоровьесбережение и здравоохранение, которые остро нуждаются в технологиях, позволяющих удаленно собирать информацию о состоянии здоровья пациентов для своевременного реагирования на отклонения в здоровье.

Известно, что гораздо эффективнее предупредить болезнь, чем лечить. Этим обусловлено быстрое развитие здоровьесбережения, т.е. раннего обнаружения рисков заболеваний или отклонений в здоровье и профилактика этих рисков. Проблема оценки рисков и профилактики заболеваний является сегодня вызовом как для научного сообщества, так и для медицинской индустрии в целом. Бурный рост массивов данных о человеке с одной стороны и развитие средств их аналитической обработки с другой позволяет создавать системы оценки состояния здоровья человека и выдачи рекомендаций по уменьшению рисков заболеваний без посещения поликлиники и проведения дорогостоящих обследований.

\* Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации. Соглашение № 05.604.21.0226 (Уникальный идентификатор проекта RFMEFI60419X0226)

✉ Станкевич Максим Алексеевич. E-mail: stankevich@isa.ru

Рассмотрим, например, такое психическое заболевание как депрессия. Это - ведущее по распространенности ментальное заболевание в Европе. Ежегодные расходы на лечение депрессии и трудовые издержки, связанные с ней, оцениваются ВОЗ в 136,3 миллиардов евро. Сейчас сделаны существенные усилия, чтобы сделать доступным лечение депрессии, однако меры по предотвращению депрессии практически не реализуются. В то же время многие исследования показывают, что профилактика депрессии играет более важную роль, чем лечение, с экономической, социальной и психологической точек зрения. Поэтому проблема раннего обнаружения признаков психологических заболеваний, в том числе на основе анализа информации из социальных сетей, приобрела сегодня большую актуальность.

В ряде исследований показано, что личностные черты (в частности, черты, выделяемые в модели «Большая пятерка») могут выступать в качестве предикторов различных психических расстройств [1] и использоваться для их диагностики [2]. Кроме того, за рубежом широко известны работы, направленные на выявление личностных черт пользователей социальных сетей на основе информации об активности пользователя в социальных сетях, его текстовых сообщений и изображений, публикуемых им в социальных сетях.

В работе [3] решается задача автоматического определения личностных черт у пользователей социальной сети ВКонтакте на основе текстовых и нетекстовых данных. В настоящей работе решается задача выявления личностных черт у пользователей социальной сети ВКонтакте на основе интеллектуального анализа публикуемых ими изображений. В первой части статьи представлен обзор исследований по теме, во второй части описаны предлагаемые методы решения задачи, в третьей части представлены результаты экспериментальной проверки методов на данных из социальной сети ВКонтакте.

## 1. Обзор работ

Социальные сети рассматриваются исследователями как уникальный источник информации об индивидах и их социальных отношениях, а современные способы анализа данных позволяют строить точные предсказательные

модели поведения человека. Многочисленные исследования показывают, что анализ персональных страниц пользователей может являться источником информации не только о социально-демографических характеристиках пользователя, но и о его личностных чертах и характеристиках, психологических предпочтениях и текущем психологическом состоянии. Например, для решения задачи выявления пятифакторной модели личностных черт человека проводился анализ крупного текстового корпуса сообщений пользователей Facebook. [4]. В другой работе оценивалась возможность выявления личностных черт и социально-демографических характеристик на основе информации о понравившемся пользователям контенте [5]. Данные посвященного музыке онлайн ресурса last.fm, в свою очередь, были использованы для выявления достоверных корреляций между личностными чертами и музыкальными предпочтениями пользователей [6]. Стоит отметить, что наиболее популярным подходом в подобных исследованиях является создание набора данных, состоящего из данных пользователей социальной сети и результатов специализированного опросника.

Наибольший интерес у исследователей вызывает задача оценки ментального здоровья на основе информации из социальных сетей и онлайн ресурсов [7,8]. Например, в рамках проекта CLEF/eRisk 2018 участникам предоставлялись коллекции текстовых сообщений пользователей Reddit для задачи выявления депрессии и анорексии [9]. Задача представлялась в виде бинарной классификации, где участникам на основе тренировочного набора данных необходимо было подготовить предсказывающую модель. По результатам проекта, лучшая F1-мера для задачи выявления депрессии составила 64%, для задачи определения анорексии F1-мера составила 85%. В другом исследовании был предложен подход к динамической оценке выраженности 9 основных симптомов депрессии на основе метода машинного обучения с частичным привлечением учителя [10]. Подход был оценен на наборе данных, состоящем из 23 миллионов сообщений Twitter, и продемонстрировал 68% усредненной точности предсказания наличия симптомов депрессии. С наибольшей точностью определялись следующие симптомы: утеря интереса, подавленное настроение и расстройства пищевого поведения.

Несмотря на тот факт, что оценка психологических характеристик человека в подобных

задачах наиболее часто происходит на основе анализа текстовой информации, существует ряд исследований, которые рассматривают эту проблему на основе анализа изображений [11]. Изображения из популярной социальной сети Instagram использовались для задачи выявления депрессии у ее пользователей [12]. В рамках этой работы при помощи краудсорсинговой платформы Amazon's Mechanical Turk был собран набор данных, состоящий из 43950 фотографий от 166 добровольцев. Из собранных данных извлекались следующие признаки: показатели активности в социальной сети, цветовые параметры фото, наличие цветочных фильтров и количество лиц на изображениях. Точность предложенной модели выявления депрессии составила около 65% F1-меры.

Согласно исследованиям, эмоции и ментальные характеристики человека имеют определенные связи с его цветовыми предпочтениями [13,14]. В свою очередь, связи между личностными чертами пользователей Flickr и цветовыми характеристиками выкладываемых ими изображений были обнаружены на основе анализа набора данных из 32056 изображений и результатов стандартного опросника личностных черт [15]. Также можно выделить исследование, где авторы искали корреляции между личностными чертами человека и частотой встречаемости определенных групп объектов, полученных при помощи инструмента Google Vision API [16].

Данные ресурса Flickr использовались для создания регрессионных моделей, способных предсказывать выраженность личностных черт пользователей [17]. Авторы работы проанализировали 60000 «избранных» изображений (по 200 изображений от каждого из 300 пользователей) и выделили множество цветовых, композиционных и текстурных характеристик, которые использовались в качестве признаков для обучения моделей. Всего было проведено 2 отдельных эксперимента: предсказание уровня выраженности личностных черт с использованием результата опросника в качестве целевого параметра, и предсказание уровня личностных черт с помощью других пользователей, которые давали оценку личностных черт вручную, изучив изображения, выложенные волонтерами. Если второй эксперимент оказался удачным, то первый дал коэффициент детерминации  $R^2$  меньше 0.1 среди всех личностных

черт. Данное исследование было продолжено в другой работе, где авторы поставили задачу бинарной классификации между высоким и низким уровнем выраженности личностных черт на этих же данных, с помощью дообученной сверточной нейронной сети для выделения признаков [18]. Точность классификации в таком эксперименте составила от 61% до 69% по разным личностным чертам.

Проанализировав данную проблему, можно судить, что большинство работ рассматривают задачу выявления психологических характеристик и личностных черт человека отдельно по каждому типу данных, который используется для анализа: общая информация профилей, текстовые сообщения и изображения. Также можно заметить, что подход к решению этих задач на основе анализа изображений не уступает по своему качеству подходам, основанным на анализе текста. Для получения более точных моделей предсказания необходимо использовать комплексный подход, где для анализа будут использоваться все возможные типы данных социальной сети. Применительно к русскоязычным социальным сетям, для подобного исследования подходит онлайн ресурс Вконтакте. Но исследований по выявлению психологических характеристик пользователей на основе анализа изображений с использованием данных Вконтакте еще не проводилось, а результаты существующих исследований, проведенных на других ресурсах, причем не в рамках русскоязычной аудитории, нельзя напрямую применить к формату Вконтакте, хотя они и представляют интерес с точки зрения предложенных подходов и методов.

## 2. Методы определения личностных черт на основе анализа изображений

Для решения задачи выявления личностных черт на основе анализа изображений был создан набор данных пользователей Вконтакте. Для сбора данных был разработан онлайн ресурс, где пользователям предоставлялась возможность авторизоваться через социальную сеть Вконтакте и пройти адаптированную для русского языка версию опросника личностных черт [19]. В данной модели представления личностных черт выделяется 5 отдельных составляющих: нейротизм, экстраверсия, готовность к согласию, открытость опыту и сознательность. Собранные результаты опросника позволяют

получить значение каждой из личностных черт в виде целого числа на шкале от 0 до 48 баллов. По аналогии с другой схожей работой [18], для проведения экспериментов пользователи разбивались на 2 класса по уровню выраженности личностных черт: низкий и высокий. Классу низких значений соответствовали пользователи со значениями личностной черты меньше нижней квантили, классу высоких значений - выше верхней квантили. Для каждой из личностных черт проводились независимые эксперименты по бинарной классификации.

Данные пользователей, которые дали доступ к своим фотографиям в ВК и прошли опросник, были использованы для составления набора данных, в который вошли 310071 изображений от 576 пользователей Вконтакте. Все фотографии пользователей Вконтакте разделены на отдельные альбомы. Мы сгруппировали их в 4 отдельные выборки:

- **Wall.** Изображения с персональной стены пользователя;
- **Ava.** Изображения, которые когда-либо были фотографией профиля пользователя («Аватарки»);
- **Other.** Изображения в остальных альбомах, к которым пользователь открыл доступ;
- **All.** Все изображения вместе.

В каждую из выборок вошло разное количество пользователей, так как у некоторых из них, например, есть изображения на стене, но нет изображений в остальных альбомах. Статистика по выборкам представлена в Табл. 1.

После этапа группировки альбомов, для каждого пользователя формировался вектор объектов (люди, животные, предметы быта, автомобили и т.д.), присутствующих на изображениях пользователей. Вычисления векторов происходило по трём различным стратегиям, независимо от выбранной модели-детектора. **(I)**

При первой стратегии объект считался присутствующим на изображении, если модель детектор выдавала вероятность наличия этого объекта больше установленного порогового значения. **(II)** При второй стратегии также вычислялась частота присутствия объекта, но суммировались сами значения вероятности присутствия объекта, где это значение было выше порогового. **(III)** Третий подход заключался в оценке вероятности присутствия объекта на фотографиях пользователя.

Были использованы 2 модели-детектора, которые были обучены на наборе данных COCO [20]: Faster R-CNN [21] и Mask R-CNN [22]. Данные модели позволяют детектировать 80 различных классов объектов на изображениях. Некоторые из этих объектов, например тостер или пожарный гидрант, встречаются очень редко на фотографиях пользователей, поэтому из признаков были убраны данные об объектах, средняя частота которых была меньше порогового значения (от 0.0001 до 0.005, в зависимости от эксперимента). Также из выборок были убраны пользователи, у которых было менее пяти фотографий.

В качестве других признаков изображений были использованы усреднённые цветовые характеристики фотографий. С помощью библиотеки OpenCV [23] были извлечены средние значения и стандартное отклонение компонент цветовых моделей RGB и HSV (Hue, Saturation, Value – тон, насыщенность, яркость) [24]. Для каждого пользователя полученные вектора усреднялись по всем фотографиям.

### 3. Результаты экспериментов

Целью экспериментов было выявления наиболее точной модели предсказания личностных черт вне зависимости от используемой выборки. Для этого было проведено множество

Табл. 1. Статистика по выборкам

mean – среднее количество на пользователя; std – среднеквадратическое отклонение

Выборка	N пользователей	N изображений	mean (std)	Медиана
All	576	310,071	538.3 (1021.2)	113
Other	343	247,684	722.1 (1140.2)	230
Wall	446	52,753	118.3 (218.2)	43
Ava	565	9,634	17.1 (24.3)	8

экспериментов по бинарной классификации, где для каждой личностной черты и каждой из 4 выборок данные сначала делились на 2 класса (низкие и высокие значения), затем делились на обучающую (80%) и тестовую (20%) выборки и, в конечном итоге, представлялись в виде векторного пространства признаков, который подавался на вход классификационной модели. Для обучения моделей использовались следующие классификационные алгоритмы: логистическая регрессия (**LR**), метод опорных векторов (**SVM**), наивный байесовский классификатор (**NB**), алгоритм случайного леса (**RF**), метод k-ближайших соседей (**KNN**), линейный дискриминантный анализ (**LDA**) и многослойный перцептрон (**MLP**). Все эти методы доступны в библиотеке `scikit-learn` [25]. В качестве метрики для оценки качества классификации использовалась макро усреднение F1-меры по двум классам.

Сначала проводилась классификация только с использованием признаков объектов, которые были получены при помощи моделей-детекторов **Faster R-CNN** и **Mask R-CNN**. Мы использовали различные стратегии по расчету векторного пространства (**I, II или III**), а также варьировали значение параметра  $th$  от 0.1 до 0.7. Параметр  $th$  является пороговым значением, где тот или иной объект считался присут-

ствующем на изображении, если модель-детектор давала оценку вероятности наличия этого объекта больше значения  $th$ . В Табл. 2 представлены наилучшие результаты классификации личностных черт с использованием признакового пространства объектов.

Исходя из результатов в Табл. 2, признаки объектов, полученные при помощи модели-детектора **Faster R-CNN** работали лучше, чем полученные при помощи **Mask R-CNN**. Признаки **Mask R-CNN** показали незначительно большую точность для экстраверсии и готовности к согласию, однако признаки **Faster R-CNN** показали заметно большую точность для сознательности и открытости опыту. Также можно заметить, что лучшие результаты не были получены ни разу на выборке **All**. В большинстве случаев, наибольшую точность показывают модели, которые обучались на выборке **Other**, самой маленькой выборке с точки зрения количества пользователей. В Таблице 3 представлены результаты схожих экспериментов, которые проводились на цветовых признаках и на комбинации признаков объектов и цветовых параметров изображений.

Сравнивая результаты в Табл. 2 и Табл. 3, можно заметить, что добавление цветовых признаков к признакам объектов дает прирост в

Табл. 2. Лучшие результаты классификации на признаковом пространстве объектов

Личностная черта	Faster R-CNN			Mask R-CNN		
	F1	Алгоритм	Параметры	F1	Алгоритм	Параметры
Нейротизм	.71	LR	$th=0.5$ (I) Other	.70	RF	$th=0.7$ (I) Other
Экстраверсия	.71	RF	$th=0.5$ (III) Ava	.72	SVM	$th=0.7$ (I) Other
Сознательность	.77	RF	$th=0.7$ (I) Other	.73	SVM	$th=0.6$ (I) Other
Готовность к согласию	.71	NB	$th=0.5$ (III) Wall	.72	NB	$th=0.6$ (II) Wall
Открытость к опыту	.67	LR	$th=0.5$ (III) Other	.63	LR	$th=0.6$ (I) Other

Табл. 3. Лучшие результаты классификации на цветовых признаках

Личностная черта	Цветовые признаки			Цветовые признаки + объекты		
	F1	Модель	Параметры	F1	Модель	Параметры
Нейротизм	.71	MLP	Wall	.74	LR	F. R-CNN $th=0.5$ (I) Other
Экстраверсия	.67	LDA	Wall	.70	RF	M. R-CNN $th=0.7$ (I) Other
Сознательность	.68	SVM	Wall	.72	NB	M. R-CNN $th=0.6$ (I) Other
Готовность к согласию	.70	NB	Other	.68	NB	F. R-CNN $th=0.5$ (II) Wall
Открытость к опыту	.58	SVM	Other	.68	MLP	F. R-CNN $th=0.6$ (III) Other

Табл. 4. Результаты лучших моделей

Личностная черта	F1 объекты	F1 цвета	F1 вместе	Лучший эксперимент	
				Модель	Параметры признаков
Нейротизм	.71	.71	<b>.74</b>	<i>LR</i>	<i>Цвета + F. R-CNN th=0.5 (I) Other</i>
Экстраверсия	<b>.72</b>	.67	.70	<i>SVM</i>	<i>M. R-CNN th=0.7 (I) Other</i>
Сознательность	<b>.77</b>	.68	.72	<i>RF</i>	<i>F. R-CNN th=0.7 (I) Other</i>
Готовность к согласию	<b>.72</b>	.70	.68	<i>NB</i>	<i>M. R-CNN th=0.6 (II) Wall</i>
Открытость опыту	.67	.58	<b>.68</b>	<i>MLP</i>	<i>Цвета + F. R-CNN th=0.6 (III) Other</i>

точности только для нейротизма и открытости опыту. В целом цветовые признаки работают хуже, чем признаки объектов, однако также показывают сравнительно неплохой результат по некоторым личностным чертам. В Табл. 4 представлена сводка по всем проведенным экспериментам с указанием наилучшей модели и наилучшего подбора параметров признаков для каждой из личностных черт.

С точки зрения данных, которые использовались для составления признаков, лучшие результаты получались при использовании выборки **Other** и **Wall**, то есть изображения из альбомов пользователей и с персональной стены Вконтакте. Согласно результатам наших экспериментов, с наибольшей точностью классифицируется уровень сознательности и нейротизма пользователей Вконтакте, а хуже всего определяется открытость опыту. Стоит заметить, что такие же результаты были получены в похожей работе [18], в одном из экспериментов по бинарной классификации личностных черт пользователей Flickr, где уровень личностных черт устанавливался путем оценки от 12 волонтеров. С другой стороны, точность классификации пользователей в работе [18], с использованием результатов опросника личностных черт, составила от 54% до 55%, что заметно ниже результатов, представленных в нашем исследовании.

## Заключение

В работе рассмотрена задача выявления личностных черт пользователей Вконтакте на основе анализа опубликованных ими изображений. На основе набора данных изображений от 576 пользователей Вконтакте, были оценены призна-

ки, с помощью которых объекты на изображениях разбиваются на 80 классов, а также опробованы несколько стратегий по их расчету. Были рассмотрены различные источники изображений, которые можно получить из данных Вконтакте: фотографии профиля, изображения с персональной стены профиля и другие альбомы. Для выявления личностных черт лучшим образом подошли изображения с персональной стены и альбомов пользователей. Использование цветовых признаков изображений позволяет улучшить качество выявления нейротизма и открытости опыту. С наибольшим качеством определяются сознательность и нейротизм.

В дальнейшей работе планируется рассмотреть задачу выявления личностных черт пользователей Вконтакте на основе анализа изображений, как задачу регрессионного анализа. Для этого потребуется расширить существующую выборку. Планируется опробовать на данных более сложные цветовые параметры и характеристики изображений. В итоге планируется создать комплексный подход к анализу личностных черт и психологических особенностей пользователей социальных сетей, где будут рассматриваться изображения, текстовые сообщения, а также активность пользователей одновременно, причем с соответствующими коэффициентами, в зависимости от анализируемой черты.

## Литература

1. Widiger T. A., Costa Jr P. T. Personality and personality disorders //Journal of abnormal psychology. – 1994. – Т. 103. – №. 1. – С. 78.
2. Widiger T. A., Mullins-Sweatt S. N. Clinical utility of a dimensional model of personality disorder //Professional Psychology: Research and Practice. – 2010. – Т. 41. – № 6. – С. 488.

3. Stankevich M., Latyshev A., Kiselnikova N., Smirnov I. Predicting Personality Traits from Social Network Profiles // In: Kuznetsov S., Panov A. (eds) Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science. – Springer, Cham – 2019. – vol. 1093. – pp. 177-188.
4. Schwartz H. A. et al. Personality, gender, and age in the language of social media: The open-vocabulary approach //PloS one. – 2013. – T. 8. – №. 9. – C. E73791.
5. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior //Proceedings of the National Academy of Sciences. – 2013. – T. 110. – №. 15. – C. 5802-5805.
6. Ferwerda B., Tkalcic M., Schedl M. Personality Traits and Music Genres: What Do People Prefer to Listen To? //Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. – ACM, 2017. – C. 285-288.
7. Shatte A. B. R., Hutchinson D. M., Teague S. J. Machine learning in mental health: a scoping review of methods and applications //Psychological medicine. – 2019. – T. 49. – №. 9. – C. 1426-1448.
8. Kursuncu U. et al. Predictive Analysis on Twitter: Techniques and Applications //Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. – Springer, Cham, 2019. – C. 67-104.
9. Losada D. E., Crestani F., Parapar J. Overview of eRisk: early risk prediction on the internet //International Conference of the Cross-Language Evaluation Forum for European Languages. – Springer, Cham, 2018. – C. 343-361.
10. Yazdavar A. H. et al. Semi-supervised approach to monitoring clinical depressive symptoms in social media //Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. – ACM, 2017. – C. 1191-1198.
11. Wongkoblap A., Vadillo M. A., Curcin V. Researching mental health disorders in the era of social media: systematic review //Journal of medical Internet research. – 2017. – T. 19. – №. 6. – C. e228.
12. Reece A. G., Danforth C. M. Instagram photos reveal predictive markers of depression //EPJ Data Science. – 2017. – T. 6. – №. 1. – C. 15.
13. Nolan R. F., Dai Y., Stanley P. D. An investigation of the relationship between color choice and depression measured by the Beck Depression Inventory //Perceptual and motor skills. – 1995. – T. 81. – №. 3\_suppl. – C. 1195-1200.
14. Valdez P., Mehrabian A. Effects of color on emotions //Journal of experimental psychology: General. – 1994. – T. 123. – №. 4. – C. 394.
15. Wieloch M. et al. Profiling user colour preferences with BFI-44 personality traits //International Conference on Business Information Systems. – Springer, Cham, 2018. – C. 63-76.
16. Ferwerda B., Tkalcic M. You are what you post: What the content of instagram pictures tells about users' personality //The 23rd International on Intelligent User Interfaces, March 7-11, Tokyo, Japan. – 2018.
17. Segalin C. et al. The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits //IEEE Transactions on Affective Computing. – 2016. – T. 8. – №. 2. – C. 268-285.
18. Segalin C., Cheng D. S., Cristani M. Social profiling through image understanding: Personality inference using convolutional neural networks //Computer Vision and Image Understanding. – 2017. – T. 156. – C. 34-50.
19. Costa P. T., McCrae R. R. Normal personality assessment in clinical practice: The NEO Personality Inventory //Psychological assessment. – 1992. – T. 4. – №. 1. – C. 5.
20. Lin T. Y. et al. Microsoft coco: Common objects in context //European conference on computer vision. – Springer, Cham, 2014. – C. 740-755.
21. Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – C. 91-99.
22. He K. et al. Mask r-cnn //Proceedings of the IEEE international conference on computer vision. – 2017. – C. 2961-2969.
23. Bradski G., Kaehler A. Learning OpenCV: Computer vision with the OpenCV library. – " O'Reilly Media, Inc.", 2008.
24. Tkalcic M., Tasic J. F. Colour spaces: perceptual, historical and applicational background. – IEEE, 2003. – T. 1. – C. 304-308.
25. Pedregosa F. et al. Scikit-learn: Machine learning in Python //Journal of machine learning research. – 2011. – T. 12. – №. Oct. – C. 2825-2830.

## Predicting Personal Traits from Vkontakte Images

N. A. Ignatiev<sup>I</sup>, M. A. Stankevich<sup>II</sup>, N. V. Kiselnikova<sup>III</sup>, O. G. Grigoriev<sup>II</sup>

<sup>I</sup>RUDN University, Moscow, Russia

<sup>II</sup>Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia

<sup>III</sup>Psychological Institute of Russian Academy of Education, Moscow, Russia

**Abstract.** The psychometric analysis of information from the Internet is one of the fastest growing trends in modern research. Using social networks, one can identify psychological characteristics and mental disorders. In this paper, we solve the problem of identifying personal qualities - predictors of depression among users of the social network VKontakte, analyzing the images published by them. We describe our methods and approaches for solving this problem and present the results of experimental verification on Vkontakte data. Our study shows that you can use object detection methods to create effective features for predicting personality traits.

**Keywords:** personality traits, social media, image recognition, machine learning.

**DOI** 10.14357/20718594190404

## References

1. Widiger T. A., Costa Jr P. T. Personality and personality disorders //Journal of abnormal psychology. – 1994. – Т. 103. – №. 1. – С. 78.
2. Widiger T. A., Mullins-Sweatt S. N. Clinical utility of a dimensional model of personality disorder //Professional Psychology: Research and Practice. – 2010. – Т. 41. – №. 6. – С. 488.
3. Stankevich M., Latyshev A., Kiselnikova N., Smirnov I. Predicting Personality Traits from Social Network Profiles // In: Kuznetsov S., Panov A. (eds) Artificial Intelligence. RCAI 2019. Communications in Computer and Information Science. – Springer, Cham – 2019. – vol. 1093. – pp. 177-188.
4. Schwartz H. A. et al. Personality, gender, and age in the language of social media: The open-vocabulary approach //PloS one. – 2013. – Т. 8. – №. 9. – С. E73791.
5. Kosinski M., Stillwell D., Graepel T. Private traits and attributes are predictable from digital records of human behavior //Proceedings of the National Academy of Sciences. – 2013. – Т. 110. – №. 15. – С. 5802-5805.
6. Ferwerda B., Tkalcic M., Schedl M. Personality Traits and Music Genres: What Do People Prefer to Listen To? //Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization. – ACM, 2017. – С. 285-288.
7. Shatte A. B. R., Hutchinson D. M., Teague S. J. Machine learning in mental health: a scoping review of methods and applications //Psychological medicine. – 2019. – Т. 49. – №. 9. – С. 1426-1448.
8. Kursuncu U. et al. Predictive Analysis on Twitter: Techniques and Applications //Emerging Research Challenges and Opportunities in Computational Social Network Analysis and Mining. – Springer, Cham, 2019. – С. 67-104.
9. Losada D. E., Crestani F., Parapar J. Overview of eRisk: early risk prediction on the internet //International Conference of the Cross-Language Evaluation Forum for European Languages. – Springer, Cham, 2018. – С. 343-361.
10. Yazdavar A. H. et al. Semi-supervised approach to monitoring clinical depressive symptoms in social media //Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017. – ACM, 2017. – С. 1191-1198.
11. Wongkoblap A., Vadillo M. A., Curcin V. Researching mental health disorders in the era of social media: systematic review //Journal of medical Internet research. – 2017. – Т. 19. – №. 6. – С. e228.
12. Reece A. G., Danforth C. M. Instagram photos reveal predictive markers of depression //EPJ Data Science. – 2017. – Т. 6. – №. 1. – С. 15.
13. Nolan R. F., Dai Y., Stanley P. D. An investigation of the relationship between color choice and depression measured by the Beck Depression Inventory //Perceptual and motor skills. – 1995. – Т. 81. – №. 3\_suppl. – С. 1195-1200.
14. Valdez P., Mehrabian A. Effects of color on emotions //Journal of experimental psychology: General. – 1994. – Т. 123. – №. 4. – С. 394.
15. Wieloch M. et al. Profiling user colour preferences with BFI-44 personality traits //International Conference on Business Information Systems. – Springer, Cham, 2018. – С. 63-76.
16. Ferwerda B., Tkalcic M. You are what you post: What the content of instagram pictures tells about users' personality //The 23rd International on Intelligent User Interfaces, March 7-11, Tokyo, Japan. – 2018.
17. Segalin C. et al. The pictures we like are our image: continuous mapping of favorite pictures into self-assessed and attributed personality traits //IEEE Transactions on Affective Computing. – 2016. – Т. 8. – №. 2. – С. 268-285.
18. Segalin C., Cheng D. S., Cristani M. Social profiling through image understanding: Personality inference using convolutional neural networks //Computer Vision and Image Understanding. – 2017. – Т. 156. – С. 34-50.
19. Costa P. T., McCrae R. R. Normal personality assessment in clinical practice: The NEO Personality Inventory //Psychological assessment. – 1992. – Т. 4. – №. 1. – С. 5.
20. Lin T. Y. et al. Microsoft coco: Common objects in context //European conference on computer vision. – Springer, Cham, 2014. – С. 740-755.
21. Ren S. et al. Faster r-cnn: Towards real-time object detection with region proposal networks //Advances in neural information processing systems. – 2015. – С. 91-99.
22. He K. et al. Mask r-cnn //Proceedings of the IEEE international conference on computer vision. – 2017. – С. 2961-2969.
23. Bradski G., Kaehler A. Learning OpenCV: Computer vision with the OpenCV library. – " O'Reilly Media, Inc.", 2008.
24. Tkalcic M., Tasic J. F. Colour spaces: perceptual, historical and applicational background. – IEEE, 2003. – Т. 1. – С. 304-308.
25. Pedregosa F. et al. Scikit-learn: Machine learning in Python //Journal of machine learning research. – 2011. – Т. 12. – №. Oct. – С. 2825-2830.