

## Использование нейросетей для прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям

А. А. Браницкий<sup>а</sup>, канд. техн. наук, [orcid.org/0000-0003-3104-0622](https://orcid.org/0000-0003-3104-0622)

Е. В. Дойникова<sup>а</sup>, канд. техн. наук, [orcid.org/0000-0001-6707-9153](https://orcid.org/0000-0001-6707-9153)

И. В. Котенко<sup>а</sup>, доктор техн. наук, профессор, [orcid.org/0000-0001-6859-7120](https://orcid.org/0000-0001-6859-7120), [ivkote@comsec.spb.ru](mailto:ivkote@comsec.spb.ru)

<sup>а</sup>Санкт-Петербургский институт информатики и автоматизации РАН, 14-я линия В. О., 39, Санкт-Петербург, 199178, РФ

**Постановка проблемы:** в рамках социальных сетей пользователи имеют возможность удаленно общаться, само-выражаться, искать людей со схожими интересами. В то же время социальные сети как источник информации могут оказывать негативное влияние на поведение и мышление их пользователей. **Цель исследования:** разработка методики прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям на основе использования искусственных нейронных сетей. **Результаты:** разработана методика и выполнена экспериментальная оценка прогнозирования результатов теста Аммона по профилю пользователя в социальной сети при помощи искусственных нейронных сетей. За основу методики взяты результаты теста Аммона для студентов медицинского вуза. Для обучения нейронной сети сформирован набор признаков на базе информации, предоставляемой пользователями в социальных сетях. По результатам экспериментов подтверждено наличие зависимости между информацией, предоставляемой пользователями в социальных сетях, и психологическими характеристиками личности. Разработан механизм, предназначенный для своевременного выявления деструктивных воздействий и профилей пользователей в социальных сетях, указывающих на подверженность таким воздействиям, упрощающий работу специалистов-психологов. Результаты экспериментов показали, что наибольшей точностью среди исследованных четырех типов нейронных сетей обладает многослойная нейронная сеть. В дальнейшем планируется расширить набор признаков для повышения точности нейронной сети. **Практическая значимость:** полученные результаты могут быть использованы для построения систем, предназначенных для выявления в интернет-пространстве воздействий, неблагоприятных для психического здоровья молодого поколения и нации в целом.

**Ключевые слова** – нейронная сеть, тест Аммона, профиль пользователя, социальная сеть, психологические шкалы, деструктивное воздействие, признаки.

**Для цитирования:** Браницкий А. А., Дойникова Е. В., Котенко И. В. Использование нейросетей для прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям. *Информационно-управляющие системы*, 2020, № 1, с. 24–33. doi:10.31799/1684-8853-2020-1-24-33

**For citation:** Branitskiy A. A., Doynikova E. V., Kotenko I. V. Use of neural networks for forecasting of the exposure of social network users to destructive impacts. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 1, pp. 24–33 (In Russian). doi:10.31799/1684-8853-2020-1-24-33

### Введение

В настоящее время социальные сети являются одним из наиболее распространенных интернет-ресурсов среди представителей молодого поколения. Возможность общения, нахождения людей с похожими интересами, а также удаленного обмена информацией способствует росту популярности таких сетевых сервисов. Информация, размещаемая в профилях пользователей социальной сети, может включать переписку с друзьями, сведения о личных интересах, а также фотографии. Эти данные могут использоваться в качестве источника для построения психологического портрета пользователя. В частности, то, что пользователь считает нужным сообщить о себе в социальной сети, может многое о нем сказать [1].

Стоит отметить, что в социальных сетях может присутствовать вредоносная информация:

пропаганда насильственных действий, популяризация нездорового образа жизни и вредных привычек, призывы к нарушению религиозных и морально-идеологических устоев в обществе и т. д. Подобная информация оказывает негативное влияние, уровень которого для разных пользователей может существенно различаться. Тем самым рассматриваемая задача, а именно построение систем, определяющих степень подверженности пользователей социальных сетей деструктивным воздействиям, является актуальной. Своевременное выявление таких воздействий в интернет-пространстве за счет мониторинга информации в социальных сетях, а также целевых профилей, подверженных подобным влияниям, позволит ослабить или даже предотвратить их негативную роль. В настоящее время для определения подверженности индивидуумов негативным воздействиям используются специа-

лизированные тесты, которые необходимо проводить регулярно, и дорогостоящее оборудование, что затрудняет этот процесс.

Для решения поставленной задачи может использоваться аппарат искусственных нейронных сетей (НС), которые широко распространены в целях классификации объектов и успешно рекомендовали себя как инструмент поиска аномалий.

## Релевантные работы

Анализ работ в рассматриваемой области позволил выделить следующие схожие группы задач: 1) классификация клинических данных; 2) выявление психологических нарушений; 3) определение эмоционального настроения; 4) выделение социальных групп; 5) защита от вредоносной информации в сети Интернет.

Для многоклассовой категоризации клинических диагнозов на основе различных медицинских показаний авторы работы [2] применяют рекуррентную НС в целях обработки многомерных временных рядов медицинских наблюдений. Результаты экспериментов подтвердили высокую эффективность применения модели, содержащей 128 ячеек долгой краткосрочной памяти.

Мониторинг психологического состояния пользователей сети Интернет привлекает внимание многих ИТ-специалистов. В частности, исследовалось [3] применение сверточных НС для выявления психологических нарушений. Для обучения и тестирования нейросетевой модели в качестве исходных выступали данные, полученные из сервисов социальных микроблогов. Выделены два типа атрибутов, которые формируют анализируемый вектор признаков и обрабатываются НС: 1) низкоуровневые атрибуты, извлеченные из единичного сообщения (твита), включая текст или изображение; 2) статистические атрибуты, касающиеся активности аккаунта (количество твитов, количество комментариев к сообщениям) в течение заданного периода времени. Результаты экспериментов показали целесообразность использования четырехслойной НС, в которой в качестве функции активации применялся особый вид логистической функции.

Большой объем полезной информации, пригодной для составления психологического профиля человека, может быть получен из учетной записи, размещенной на его персональной странице в социальной сети. В частности, авторы статьи [4] рассматривают вопрос построения системы распознавания психологического профиля и особенностей характера человека по его фотографиям из социальной сети Facebook. Авторы отмечают, что для общительных и открытых людей

характерны фотографии, в которых присутствует множество человеческих лиц и преобладают светлые тона, в то время как для людей с депрессивным мышлением чаще встречаются картинки с замкнутыми пространствами.

Исследования, представленные в работах [5, 6], посвящены анализу настроений интернет-пользователей. На основе данных, извлекаемых из коротких сообщений, авторы предлагают строить глубокую НС, способную предсказывать бинарное настроение (положительное или отрицательное) адресанта. Обрабатываемый вектор признаков строится в многоуровневом стиле: для анализа используется информация, извлеченная как из отдельных символьных последовательностей, так и из более высокоуровневых синтаксических конструкций, а именно предложений [5].

В работе [7] определяется уровень эмоционального воздействия, производимого на человека в результате просмотра цифровых изображений. В качестве признаков, характеризующих изображения, подлежащие обработке, использовались параметры яркости цветовых схем, относительное преобладание темных цветов, параметры текстур и композиции. Для автоматической классификации изображений по уровню эмоционального воздействия применялся наивный байесовский классификатор.

Основой методики обнаружения групп в социальных сетях, представленной в работе [8], является генетический алгоритм, оптимизируемая фитнес-функция которого направлена на поиск плотно сгруппированных узлов внутри графа связей, причем разные образовавшиеся таким образом кластеры должны разреженно соединяться друг с другом или вовсе не иметь общих связей. Одним из прикладных назначений данной методики можно назвать выделение общих интересов у исследуемой группы людей и построение усредненного психологического профиля участника обнаруженной группы.

Предложено [9] рассматривать генетический алгоритм как эвристический инструмент для решения NP-полной задачи — кластеризации фрагмента социальной сети. В работе [10] для определения количества кластеров внутри социальной сети предлагается применять байесовский подход.

В работе [11] решается задача прогнозирования повторной публикации сообщений в социальной сети Twitter. Для вычисления показателей точности и полноты использованы линейная машина опорных векторов и логистическая регрессия.

Вопросы защиты пользователей от вредоносной и нежелательной информации рассматривались в статьях [12–15]. Одним из целевых назначений разработанной авторами методики в рамках рассматриваемой предметной области является

сокращение (или вовсе предотвращение) случаев обращения доступа представителей молодого поколения к информации, носящей нелегитимный характер. Данный фактор имеет также важное значение при построении систем родительского контроля, что позволит уберечь несовершеннолетних детей от информации, способной отрицательно воздействовать на психологическое состояние ребенка. Архитектура разработанной на базе этой методики программной системы организована в иерархическом стиле. На первом уровне размещаются бинарные классификаторы, определяющие принадлежность анализируемого контента к определенной категории. Второй уровень включает классификаторы, ориентированные на использование отдельных структурных элементов html-контента и предназначенные для определения темы. На третьем уровне располагается агрегирующий классификатор.

Рассмотренные работы подтверждают актуальность затронутой проблемы и необходимость совершенствования существующих методов. В отличие от представленных работ данное исследование направлено на разработку комплексной методики, предназначенной для выявления подверженности пользователей социальных сетей деструктивным воздействиям и сочетающей в себе технологии искусственных НС при анализе разнородных данных (скалярных величин, текста, изображений).

### Методика прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям

При разработке методики прогнозирования подверженности пользователей социальных сетей деструктивным воздействиям были выбраны многослойные НС. Такой выбор обусловлен способностью искусственных НС к аппроксимации разделяющих гиперплоскостей между сложными линейно неразделимыми множествами [16–18]. Предлагаемая методика включает три шага: 1) сбор данных из профилей пользователей социальной сети; 2) формирование вектора признаков; 3) прогнозирование результатов теста Аммона по собранному данным из профиля пользователя социальной сети при помощи обученных НС.

#### Сбор данных из социальных профилей

Полученные в результате проведенного онлайн-тестирования Аммона среди студентов-медиков записи были помечены 18 шкалами, представляющими собой комбинации шести Я-функций (агрессии, тревоги, внешнего Я-отграничения и пр.) и трех компонентов (конструктивного, де-

структивного, дефицитарного). Для каждой из таких пар вычислялось значение Т-балла (Т), интерпретируемое как низкое для  $0 \leq T \leq 39$ , среднее для  $40 \leq T \leq 60$  или высокое для  $61 \leq T \leq 110$ . Исследуемый набор данных содержал 460 записей, описывающих уникальные профили пользователей социальной сети.

#### Формирование вектора признаков

В роли признаков, описывающих профили пользователей в социальной сети, использовались три группы исходных данных: 1) численные параметры (скаляры), включающие месяц рождения, количество подписчиков, друзей и фотографий и пр.; 2) последовательность слов (например, параметры, направленные на формирование набора наиболее употребительных слов в рамках каждой психологической шкалы); 3) параметры, вычисляемые на уровне бинарных потоков данных (например, результат классификации изображения при помощи НС imagenet [19]). В табл. 1 приведены сведения о вычисляемых параметрах.

Примеры наиболее популярных слов, встречающихся в постах, и наиболее частых категорий фотографий, распознанных НС imagenet, приведены в табл. 2.

#### Прогнозирование результатов теста Аммона

При проведении экспериментов учитывался только компонент деструктивности. В роли нейросетевых классификаторов рассматривались многослойная нейронная сеть (МНС) с тремя скрытыми слоями нейронов, машина опорных векторов (МОВ) с радиально-базисным ядром, линейная регрессия (ЛР), а также сверточная нейронная сеть (СНС). При обучении первых трех классификаторов использовалось подмножество признаков, представленных в табл. 1. Для обучения последнего классификатора использовались посты, размещенные на стенах профилей в социальной сети. Функционирование этого классификатора (рис. 1) включает следующие шаги.

1. Преобразование текста на распределительном слое. Текст разбивается на отдельные предложения. Каждому слову приписывается уникальный численный идентификатор: чем больше частота встречаемости слова, тем меньше это значение. Полученные значения размещаются внутри прямоугольной решетки таким образом, чтобы ячейка под номером  $(i, j)$  соответствовала  $j$ -му слову  $i$ -го предложения. Выравнивание обеспечивается путем приписывания нулей справа до получения строки максимальной длины.

2. Применение свертки. Каждая подматрица, составленная из элементов построенной на шаге 1 решетки, поэлементно умножается на ядро свертки. Полученные произведения суммируются.

■ **Таблица 1.** Параметры, формирующие вектор признаков

■ **Table 1.** Parameters forming the feature vector

№	Параметр	Тип исходных данных
1	Количество подписчиков	Скаляр
2	Количество друзей	То же
3	Количество групп	— " —
4	Количество фотографий	— " —
5	Количество подписок	— " —
6	Количество видео	— " —
7	Пол (м/ж)	— " —
8	Количество постов	— " —
9	Количество репостов	— " —
10	Месяц рождения	— " —
11–55	Результат применения word2vec [20] между пятью наиболее употребительными словами, характерными для постов каждого класса, и тремя наиболее употребительными словами в постах анализируемого профиля	Последовательность слов
56–100	Результат применения word2vec между пятью наиболее частыми категориями фотографий в рамках каждого класса и тремя наиболее частыми категориями фотографий в анализируемом профиле	Байтовая последовательность (изображение)
101	Результат применения сверточной НС над последовательностью слов в постах	Последовательность слов

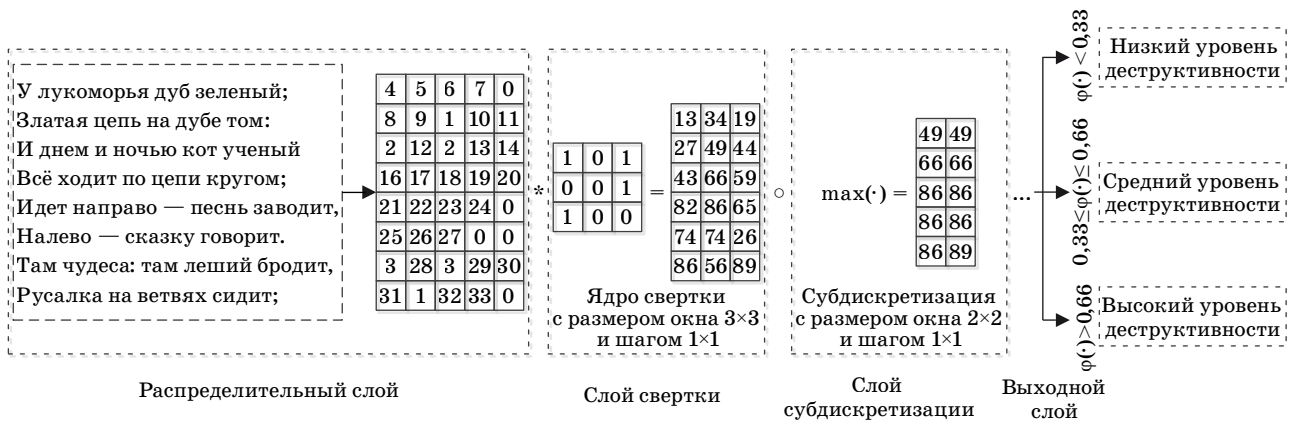
■ **Таблица 2.** Примеры наиболее популярных слов и категорий фотографий, характерных для различных уровней деструктивного нарциссизма

■ **Table 2.** Examples of the most popular words and categories of photos characteristic for various levels of destructive narcissism

Слово	Частота	Слово	Частота	Слово	Частота
Уровень деструктивного нарциссизма					
низкий		средний		высокий	
<b>В постах</b>					
Очень	59	Club	106	Подарки	58
Спасибо	56	Vkfeed	106	Очень	46
Всегда	34	Iphone	106	The	42
Мат*	34	Очень	71	Можно	41
Даже	32	Спасибо	64	Просто	39
<b>На фотографиях</b>					
Dust	174	Suit	453	Site	6222
Site	159	Tie	433	Web	1975
Wig	136	Site	417	Website	1960
Shirt	128	Miniskirt	413	Internet	1874
Boo	112	Pajama	409	Book	1151

3. Применение субдискретизации. Из матрицы, полученной на шаге 2, извлекаются наиболее весомые элементы внутри окна заданного размера. С этой целью применяется операция максимума.

4. Интерпретация выходных результатов. Сигналы, полученные после шага 3, оформляются в виде вектора, который скалярно умножается на веса выходного слоя; результат — величина  $y$ . Низкий уровень деструктивности соответствует



■ **Рис. 1.** Функционирование СНС, обученной на постах профилей в социальной сети (применение нормализации на распределительном слое и функции активации на слое свертки опущено)

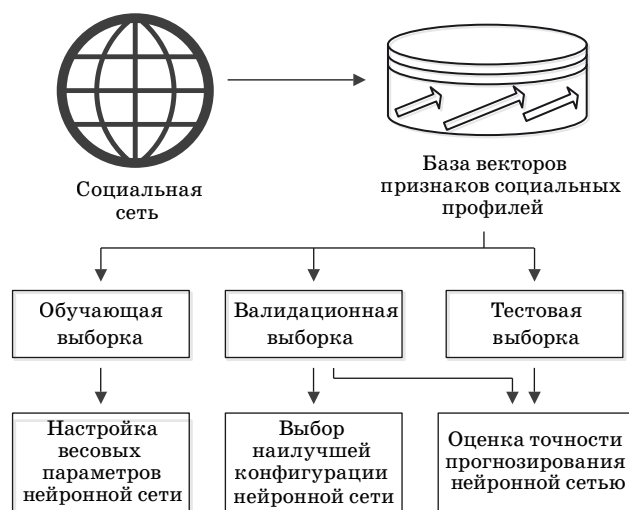
■ **Fig. 1.** The functioning of the convolutional neural network trained at the posts of profiles in a social network (the use of normalization on the distribution layer and activation functions on the convolution layer is omitted)

случаю  $\varphi(y) < 0,33$ , средний уровень — случаю  $0,33 \leq \varphi(y) \leq 0,66$ , высокий уровень — случаю  $\varphi(y) > 0,66$ , где  $\varphi$  — сигмоидная функция активации.

**Эксперименты по оценке методики**

Эксперименты проводились по схеме, представленной на рис. 2.

В экспериментах использовалась 10-блочная перекрестная проверка. Исходный набор признаков данных разбивался на 10 частей таким образом, чтобы в каждой из этих частей содержались примерно равноразмерные подвыборки, состоящие из элементов одного и того же класса. Девять частей использовалось для обучения.



■ **Рис. 2.** Схема эксперимента

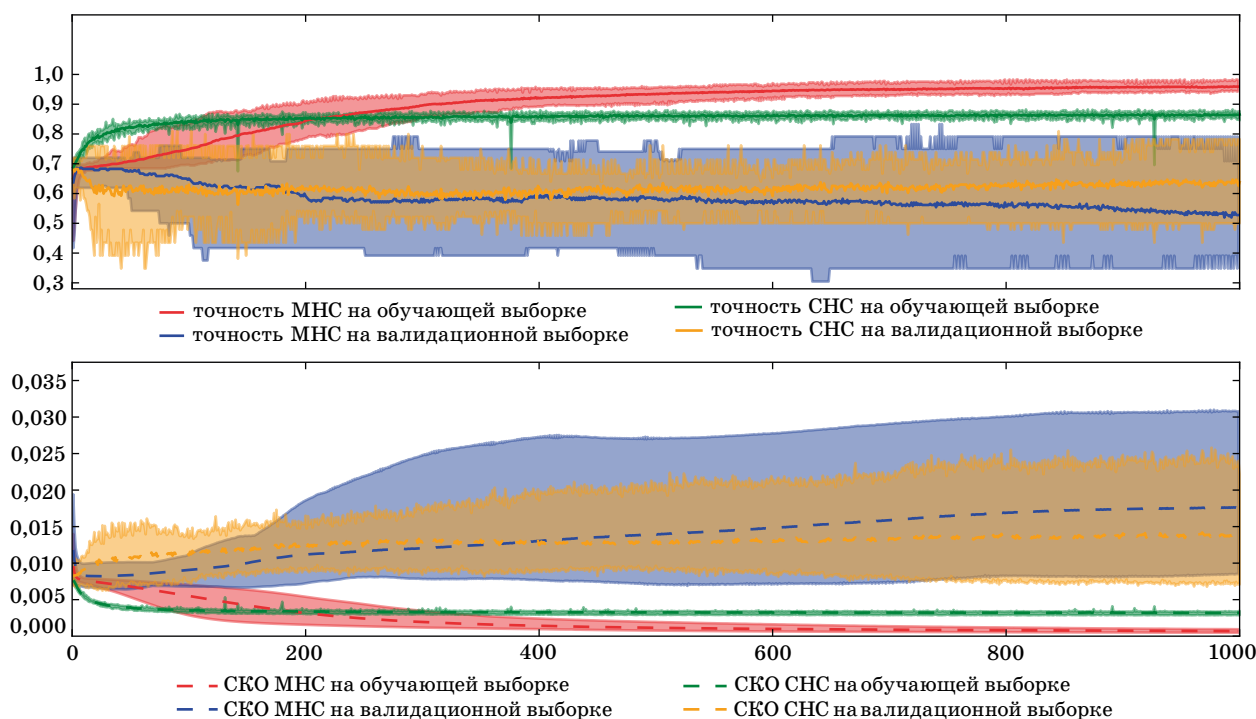
■ **Fig. 2.** The scheme of the experiment

Оставшаяся часть использовалась в качестве тестовой, половина элементов которой формировала валидационное множество.

Итеративная настройка весовых коэффициентов НС осуществлялась при помощи элементов обучающей выборки. Валидационная выборка использовалась для вычисления среднеквадратичной ошибки (СКО) после каждой эпохи обучения и последующего выбора конфигурации НС с наименьшей СКО. Тестовая выборка использовалась для вычисления точности результатов прогнозирования. На рис. 3 представлена зависимость точности и СКО МНС и СНС от номера эпохи обучения на обучающей и валидационной выборках для психологической шкалы деструктивной агрессии.

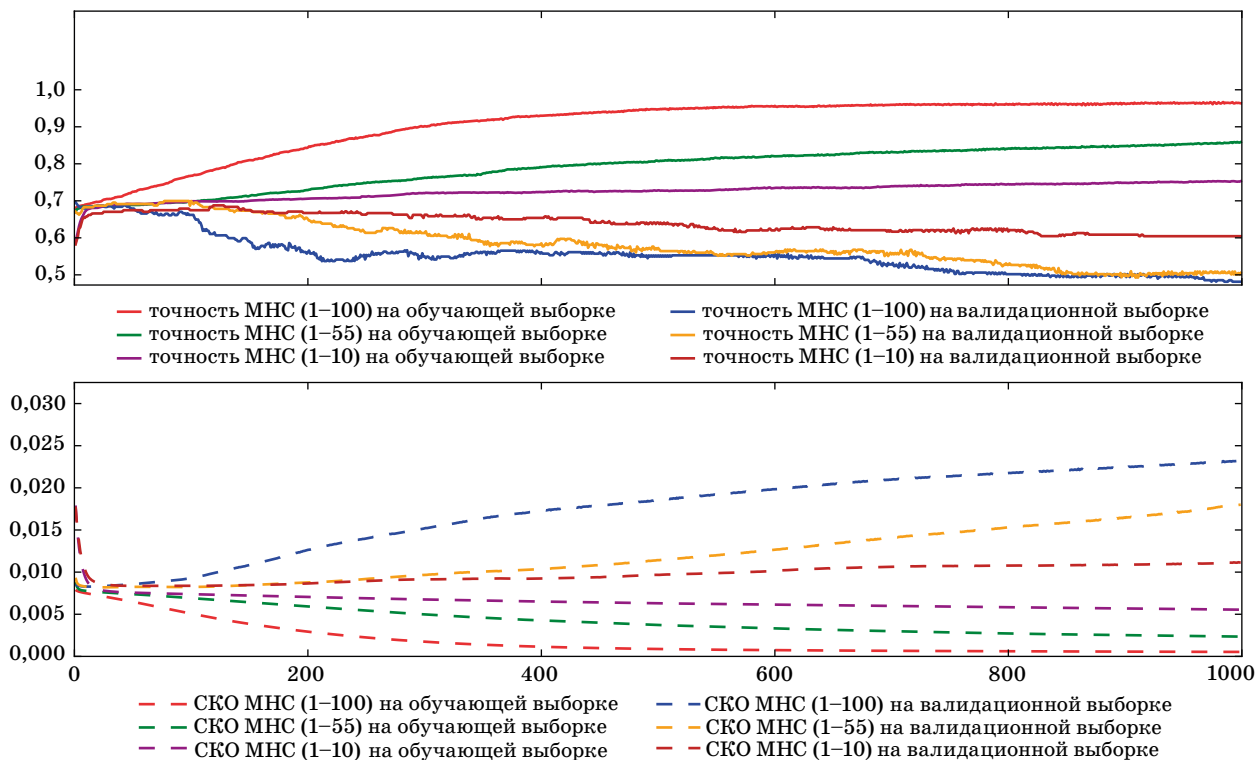
Многослойная НС характеризуется более быстрой сходимостью алгоритма обучения по сравнению с СНС. В то же время с увеличением числа эпох обучения для первой наблюдается заметное снижение точности прогнозирования уровня психологической шкалы на валидационной выборке, что обуславливается детальным представлением признаков и появлением эффекта переобучения. На рис. 4 отражено влияние размерности вектора признаков, обрабатываемого при помощи МНС, на точность и СКО прогнозирования элементов обучающей и валидационной выборок. Увеличение числа признаков позволяет существенно снизить СКО, вычисленную на обучающей выборке, но по мере увеличения числа эпох обучения, как и в случае рис. 3, также наблюдается уменьшение обобщающей способности МНС на валидационной выборке.

Корреляция признаков с меткой прогнозируемых уровней деструктивного нарциссизма приведена на рис. 5. Наибольшее абсолютное значение корреляции доставляет последний признак,



■ **Рис. 3.** Зависимость точности и СКО МНС и СНС от номера эпохи обучения для психологической шкалы деструктивной агрессии

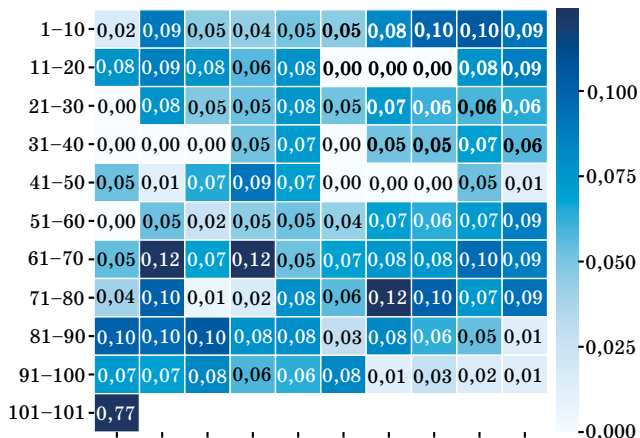
■ **Fig. 3.** The dependence of accuracy and mean-square error of the multilayer neural network and the convolutional neural network on the number of the epoch of training for the psychological scale destructive aggression



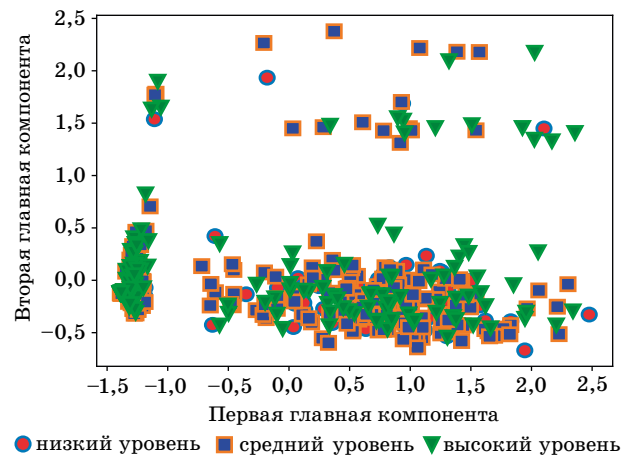
■ **Рис. 4.** Зависимость точности и СКО МНС от номера эпохи обучения для психологической шкалы деструктивной агрессии

■ **Fig. 4.** The dependence of accuracy and mean-square error of the multilayer neural network on the number of the epoch of training for the psychological scale destructive aggression





■ **Рис. 5.** Корреляция признаков с меткой прогнозируемых уровней деструктивного нарциссизма  
 ■ **Fig. 5.** Correlation of features with a label of predicted levels of destructive narcissism



■ **Рис. 6.** Первые две главные компоненты признаков для психологической шкалы деструктивного нарциссизма  
 ■ **Fig. 6.** The first two main components of features for the psychological scale of destructive narcissism

■ **Таблица 3.** Значения точности прогнозирования результатов теста Аммона для четырех типов классификаторов на тестовой выборке  
 ■ **Table 3.** The accuracy of predicting the results of the Ammon's test for four types of classifiers in the test sample

Классификатор	Деструктивный компонент Я-функции, %						Среднее значение
	первой	второй	третьей	четвертой	пятой	шестой	
МНС (1-10)	68,74	57,87	48,96	66,23	50,43	67,39	59,94
МНС (1-55)	69,62	62,18	49,89	66,47	48,51	67,62	60,71
МНС (1-100)	70,10	59,36	52,32	66,65	51,39	67,80	61,27
МНС (1-101)	70,15	59,76	52,82	66,12	51,59	67,60	61,34
МОВ (1-10)	68,49	55,4	48,35	65,07	48,57	67,79	58,95
ЛР (1-10)	30,89	41,84	39,21	30,41	41,12	47,74	38,54
СНС	64,48	46,94	44,51	58,99	43,85	59,06	52,97

представляющий собой выходное значение СНС, предварительно обученной на постах.

Первые две главные компоненты признаков представлены на рис. 6. Области, содержащие объекты из разных уровней, перекрывают друг друга, поэтому сужение размерности анализируемых векторов признаков для обучения НС может приводить к увеличению ошибок прогнозирования результатов теста Аммона.

Представленные в табл. 3 значения точности, вычисленные на тестовой выборке для четырех классификаторов: МНС, МОВ, ЛР, СНС, — усреднялись по 10 блокам перекрестной проверки. Для обучения МОВ и ЛР использовались только 10 скалярных параметров, представленных в табл. 1. Для МНС дополнительно выполнялось обучение также с привлечением параметров второй и третьей групп исходных данных.

Наибольшее значение точности, полученное посредством усреднения по шести психологическим шкалам деструктивности, принадлежит МНС, которая обучалась с применением 101-мерного вектора признаков. Наименьшее значение точности было получено с использованием ЛР, что подтверждает необходимость введения слоя нелинейности для корректного обособления объектов из разных классов.

**Заключение**

В статье рассмотрено приложение НС к задаче прогнозирования результатов теста Аммона по профилю пользователей социальной сети. Было выполнено экспериментальное сравнение четырех различных типов НС, среди которых наилуч-

шими показателями точности обладает четырех-слойный перцептрон.

Полученные результаты экспериментов по шкалам деструктивности теста Аммона для тестовой группы молодежи указывают на то, что зависимость между информацией, предоставляемой пользователями в социальных сетях, и психологическими характеристиками личности существует. В то же время необходимо учитывать, что эксперимент проводился только для ограниченной группы молодых людей и для ограниченного набора характеристик личности. Для более точных результатов, и тем более формирования выводов о конкретной личности, необходим более тщательный анализ личности экспертами, а также анализ условий, в которых данная личность существует.

К достоинствам предлагаемой методики можно отнести то, что она позволяет автоматически обрабатывать большие объемы данных в социальных сетях и сосредоточиться эксперту на ограниченной группе тех личностей, которые

могут быть подвержены деструктивным воздействиям.

К недостаткам относится тот факт, что психологические характеристики личности являются очень индивидуальными, и разработка единого автоматического механизма их анализа является сложной задачей, требующей комплексного учета большого количества факторов, далеко не все из которых находят отражение в социальной сети. Поэтому в дальнейшей работе предполагается сосредоточиться на выделении новых признаков подверженности деструктивности, определяемых на основе информации, предоставляемой в социальных сетях, а также на расширении методики на другие шкалы теста Аммона.

### Финансовая поддержка

Работа выполнена при частичной финансовой поддержке проекта РФФИ 18-29-22034 мк и бюджетной темы 0073-2019-0002.

### Литература

1. Казнова Н. Н., Овчинникова И. Г. Специфика коммуникации в социальных сетях по сравнению с блогосферой. *Вопросы психолингвистики*, 2014, № 21, с. 86–97.
2. Lipton Z. C., et al. *Learning to diagnose with LSTM recurrent neural networks*. arXiv preprint arXiv: 1511.03677, 2015 (дата обращения: 02.12.2019).
3. Lin H., et al. User-level psychological stress detection from social media using deep neural network. *Proceedings of the 22nd ACM International Conference on Multimedia*, ACM, 2014, pp. 507–516.
4. Segalin C., et al. What your Facebook profile picture reveals about your personality. *Proceedings of the 2017 ACM on Multimedia Conference*, ACM, 2017, pp. 460–468.
5. dos Santos C., Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.
6. Socher R., et al. Semi-supervised recursive autoencoders for predicting sentiment distributions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2011, pp. 151–161.
7. Machajdik J., Hanbury A. Affective image classification using features inspired by psychology and art theory. *Proceedings of the 18th ACM International Conference on Multimedia*, ACM, 2010, pp. 83–92.
8. Pizzuti C. Ga-net: A genetic algorithm for community detection in social networks. *International Conference on Parallel Problem Solving from Nature*, Springer, Berlin, Heidelberg, 2008, pp. 1081–1090.
9. Firat A., Chatterjee S., Yilmaz M. Genetic clustering of social networks using random walks. *Computational Statistics & Data Analysis*, 2007, vol. 51, no. 12, pp. 6285–6294.
10. Handcock M. S., Raftery A. E., Tantrum J. M. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2007, vol. 170, no. 2, pp. 301–354.
11. Yang Z., et al. Understanding retweeting behaviors in social networks. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ACM, 2010, pp. 1633–1636.
12. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet. *International Journal of Internet Protocol Technology (IJIPT)*, 2017, vol. 10, no. 1, pp. 61–71.
13. Котенко И. В., Чечулин А. А., Комашинский Д. В. Автоматизированное категорирование веб-сайтов для блокировки веб-страниц с неприемлемым содержанием. *Проблемы информационной безопасности. Компьютерные системы*, 2015, № 2, с. 62–68.
14. Котенко И. В., Саенко И. Б., Чечулин А. А. *Защита от нежелательной и вредоносной информации в глобальных информационных сетях. Информационно-психологическая и когнитивная безопасность: коллективная монография/ под ред. И. Ф. Кефели, Р. М. Юсупова*. СПб., Аврора, 2017. 345 с.
15. Новожилов Д. А., Чечулин А. А., Котенко И. В. Улучшение категорирования веб-сайтов для бло-



кировки неприемлемого содержимого на основе анализа статистики HTML-тэгов. *Информационно-управляющие системы*, 2016, № 6, с. 65–73. doi:10.15217/issn1684-8853.2016.6.65

16. Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCCS)*, 1989, vol. 2, no. 4, pp. 303–314.
17. Funahashi K. I. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 1989, vol. 2, no. 3, pp. 183–192.

18. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 1989, vol. 2, no. 5, pp. 359–366.
19. Krizhevsky A., Sutskever I., Hinton G. E. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, pp. 1097–1105.
20. Mikolov T., et al. *Efficient estimation of word representations in vector space*. arXiv preprint arXiv:1301.3781, 2013 (дата обращения: 02.12.2019).

UDC 004.056

doi:10.31799/1684-8853-2020-1-24-33

### Use of neural networks for forecasting of the exposure of social network users to destructive impacts

A. A. Branitskiy<sup>a</sup>, PhD, Tech., orcid.org/0000-0003-3104-0622, branitskiy@comsec.spb.ru

E. V. Doynikova<sup>a</sup>, PhD, Tech., orcid.org/0000-0001-6707-9153, doynikova@comsec.spb.ru

I. V. Kotenko<sup>a</sup>, Dr. Sc., Tech., Professor, orcid.org/0000-0001-6859-7120, ivkote@comsec.spb.ru

<sup>a</sup>Saint-Petersburg Institute for Informatics and Automation of the RAS, 39, 14 Line, V. O., 199178, Saint-Petersburg, Russian Federation

**Introduction:** In social networks, the users can remotely communicate, express themselves, and search for people with similar interests. At the same time, social networks as a source of information can have a negative impact on the behavior and thinking of their users. **Purpose:** Developing a technique of forecasting the exposure of social network users to destructive influences, based on the use of artificial neural networks. **Results:** A technique has been developed and experimentally evaluated for forecasting Ammon's test results by a social network user's profile using artificial neural networks. The technique is based on the results of Ammon's test for medical students. For training the neural network, a set of features was generated based on the information provided by social network users. The results of the experiments have confirmed the dependence between the data provided by social network users and their psychological characteristics. A mechanism has been developed aimed at prompt detection of destructive impacts or social network users' profiles indicating the susceptibility to such impacts, in order to facilitate the work of psychologists. The experiments have shown that out of the four investigated types of neural networks, the highest accuracy is provided by a multilayer neural network. In the future, it is planned to expand the set of features in order to achieve a better accuracy. **Practical relevance:** The obtained results can be used to develop systems for monitoring the Internet environment, detecting the impacts potentially dangerous for mental health of the young generation and the nation as a whole.

**Keywords** — neural network, Ammon's test, user profile, social network, psychological scales, destructive impact, features.

**For citation:** Branitskiy A. A., Doynikova E. V., Kotenko I. V. Use of neural networks for forecasting of the exposure of social network users to destructive impacts. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2020, no. 1, pp. 24–33 (In Russian). doi:10.31799/1684-8853-2020-1-24-33

### References

1. Kaznova N. N., Ovchinnikova I. G. Specific features of computer mediated communication: social networks vs. weblogs. *Journal of Psycholinguistics*, 2014, no. 21, pp. 86–97 (In Russian).
2. Lipton Z. C., et al. *Learning to diagnose with LSTM recurrent neural networks*. Available at: arXiv preprint arXiv:1511.03677, 2015 (accessed 2 December 2019).
3. Lin H., et al. User-level psychological stress detection from social media using deep neural network. *Proceedings of the 22nd ACM International Conference on Multimedia*, ACM, 2014, pp. 507–516.
4. Segalin C., et al. What your Facebook profile picture reveals about your personality. *Proceedings of the 2017 ACM on Multimedia Conference*, ACM, 2017, pp. 460–468.
5. dos Santos C., Gatti M. Deep convolutional neural networks for sentiment analysis of short texts. *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, 2014, pp. 69–78.
6. Socher R., et al. Semi-supervised recursive autoencoders for predicting sentiment distributions. *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 2011, pp. 151–161.
7. Machajdik J., Hanbury A. Affective image classification using features inspired by psychology and art theory. *Proceedings of the 18th ACM International Conference on Multimedia*, ACM, 2010, pp. 83–92.
8. Pizzuti C. Ga-net: A genetic algorithm for community detection in social networks. *International Conference on Parallel Problem Solving from Nature*, Springer, Berlin, Heidelberg, 2008, pp. 1081–1090.
9. Firat A., Chatterjee S., Yilmaz M. Genetic clustering of social networks using random walks. *Computational Statistics & Data Analysis*, 2007, vol. 51, no. 12, pp. 6285–6294.
10. Handcock M. S., Raftery A. E., Tantrum J. M. Model-based clustering for social networks. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 2007, vol. 170, no. 2, pp. 301–354.
11. Yang Z., et al. Understanding retweeting behaviors in social networks. *Proceedings of the 19th ACM International Conference on Information and Knowledge Management*, ACM, 2010, pp. 1633–1636.
12. Kotenko I., Chechulin A., Komashinsky D. Categorisation of web pages for protection against inappropriate content in the internet. *International Journal of Internet Protocol Technology (IJIPT)*, 2017, vol. 10, no. 1, pp. 61–71.

13. Kotenko I., Chechulin A., Komashinsky D. Automated categorization of web-sites for inappropriate content blocking. *Information Security Problems. Computer Systems*, 2015, no. 2, pp. 62–68 (In Russian).
14. Kotenko I., Saenko I., Chechulin A. *Zashchita ot nezhelatel'noj i vredonosnoj informacii v global'nyh informacionnyh setyah. Informacionno-psihologicheskaya i kognitivnaya bezopasnost'* [Protection from unwanted and harmful information in global information networks. Information-psychological and cognitive security]. Ed. I. F. Kefeli, R. M. Yusupova. Saint-Petersburg, Avrora Publ., 2017. 345 p. (In Russian).
15. Novozhilov D. A., Chechulin A. A., Kotenko I. V. Improving website categorization based on HTML tag statistics for blocking unwanted content. *Informatsionno-upravliaiushchie sistemy* [Information and Control Systems], 2016, no. 6, pp. 65–73 (In Russian). doi:10.15217/issn1684-8853.2016. 6.65
16. Cybenko G. Approximation by superpositions of a sigmoidal function. *Mathematics of Control, Signals, and Systems (MCSS)*, 1989, vol. 2, no. 4, pp. 303–314.
17. Funahashi K. I. On the approximate realization of continuous mappings by neural networks. *Neural Networks*, 1989, vol. 2, no. 3, pp. 183–192.
18. Hornik K., Stinchcombe M., White H. Multilayer feedforward networks are universal approximators. *Neural Networks*, 1989, vol. 2, no. 5, pp. 359–366.
19. Krizhevsky A., Sutskever I., Hinton G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 2012, pp. 1097–1105.
20. Mikolov T., et al. *Efficient estimation of word representations in vector space*. Available at: arXiv preprint arXiv:1301.3781, 2013 (accessed 2 December 2019).

---

### УВАЖАЕМЫЕ АВТОРЫ!

Научная электронная библиотека (НЭБ) продолжает работу по реализации проекта SCIENCE INDEX. После того как Вы зарегистрируетесь на сайте НЭБ (<http://elibrary.ru/defaultx.asp>), будет создана Ваша личная страничка, содержание которой составят не только Ваши персональные данные, но и перечень всех Ваших печатных трудов, имеющих в базе данных НЭБ, включая диссертации, патенты и тезисы к конференциям, а также сравнительные индексы цитирования: РИНЦ (Российский индекс научного цитирования), h (индекс Хирша) от Web of Science и h от Scopus. После создания базового варианта Вашей персональной страницы Вы получите код доступа, который позволит Вам редактировать информацию, помогая создавать максимально объективную картину Вашей научной активности и цитирования Ваших трудов.

---