

УДК 004-032-26

<https://www.doi.org/10.47813/rosnio.4.2025.2003>

EDN [QBANAP](#)

Применение генеративно-сопоставительной сети для выявления суицидального поведения

О.А. Митина*, И.В. Черенков

МИРЭА – Российский технологический университет, проспект Вернадского, 78,
Москва, Россия

*E-mail: alogmi@yandex.ru

Аннотация. Изучены виды нейронных сетей для выявления суицидального поведения путём решения задачи классификации текстовых данных. Проведено обучение генеративно-сопоставительной сети для генерации примеров суицидального поведения. Проведена аугментация текстовых данных и оценка качества классификаторов до и после внедрения в обучающее множество искусственных данных. Россия занимает одно из первых мест в мире по частоте самоубийств среди детей и подростков. Это вторая по распространённости причина смерти после несчастного случая. При этом ряд врачей-суицидологов считает, что многие несчастные случаи на деле являются суицидами. Если это так, то риск самоубийства является основным риском смерти среди молодёжи, поэтому профилактика суицидального риска становится необходимой миссией в воспитании нового поколения. В наши дни самоубийства стали распространены как причина смерти, что заставляет искать способы, как можно определить заранее, что человек планирует покончить собой. Сообщения в социальных сетях, находящиеся в открытом доступе, могут служить источником информации о человеке и его намерениях. Чтобы автоматически анализировать большой поток таких сообщений, внедряются интеллектуальные системы на основе нейронных сетей. Но такие системы имеют существенный недостаток – их тяжело обучить ввиду недостатка данных. Решением проблемы является привлечение для создания обучающих образцов суицидального контента генеративных нейросетевых моделей, в том числе генеративно-сопоставительных сетей.

Ключевые слова: суицидальное поведение, нейронные сети, генеративно-сопоставительные сети, синтетические данные.

Application of generative adversarial network to detect suicidal behavior

O.A. Mitina*, I.V. Cherenkov

MIREA – Russian Technological University, Vernadsky Avenue, 78, Moscow, Russia

*E-mail: alogmi@yandex.ru

Abstract The types of neural networks for detecting suicidal behavior by solving the problem of classifying text data were studied. A generative adversarial network was trained to generate examples of suicidal behavior. Text data was augmented and the quality of classifiers was assessed before and after introducing artificial data into the training set. Russia ranks among the world's leaders in the frequency of suicides among children and adolescents. This is the second most common cause of death after accidents. At the same time, a number of suicidologists believe that many accidents are actually suicides. If this is so, then the risk of suicide is the main risk of death among young people, so the prevention of suicide risk becomes a necessary mission in raising a new generation. Nowadays, suicide has become a common cause of death, which makes it necessary to look for ways to determine in advance that a person is planning to commit suicide. Messages on social networks that are publicly available can serve as a source of information about a person and his intentions. To automatically analyze a large flow of such messages, intelligent systems based on neural networks are being implemented. But such systems have a significant drawback - they are difficult to train due to a lack of data. The solution to the problem is to use generative neural network models, including generative adversarial networks, to create training samples of suicidal content.

Keywords: suicidal behavior, neural networks, generative adversarial networks, synthetic data.

1. Введение

Самоубийства – растущая современная проблема всего мира. Для снижения смертности создаются системы надзора на основе нейронных сетей, которые автоматически выявляют склонности к суициду по текстовому контенту в соцсетях. Нейросетевые модели могут классифицировать такие сообщения, но собрать достаточное количество примеров реального суицидального поведения, чтобы обучить эти модели, сложно.

Решением проблемы является аугментация данных с использованием генеративных нейронных сетей. Примером такой сети может являться генеративно-состязательная сеть.

Результатом исследования является обученная генеративно-состязательная сеть, а также сравнение метрик классификации до и после аугментации данных. Это исследование может помочь модераторам социальных сетей, специалистам по обнаружению деструктивного контента, более эффективно обучать модели выявления суицидального поведения, а также способствовать популяризации применения генеративно-состязательных сетей для генерации текста.

2. Постановка задачи (Цель исследования)

Цель исследования – изучить существующие модели выявления суицидального поведения, применимость генеративно-состязательных сетей для генерации примеров такого поведения и их использование в качестве синтетических данных в задаче классификации.

Исследование может помочь модераторам социальных сетей, специалистам по обнаружению деструктивного контента, более эффективно обучать модели выявления суицидального поведения, а также способствовать популяризации применения генеративно-состязательных сетей для генерации текста.

3. Методы и материалы исследования

Суицидальное поведение проявляется как мысли, фантазии, представления или действия, направленные на самоповреждение или самоуничтожение и, по крайней мере, в минимальной степени мотивируемое явными или неявными интенциями к смерти. [1]

Самоубийства являются глобальной проблемой во всем мире и от того актуальной становится тема поиска способов распознать суицидальное поведение в человеке до того, как от мыслей он перейдет к действию. С переходом общения в виртуальное пространство стал возможен анализ общедоступного текстового контента для психологического профилирования автора. Современные исследования показывают, что сверточные нейронные

сети и сети LSTM эффективно справляются с психометрической классификацией и анализом настроения и эмоций. [2] С принципами работы этих нейронных в сетей в задачах анализа и классификации текста можно ознакомиться в работе [3].

Глубокое обучение требует большого объёма данных, сбор которых затруднен из-за чувствительности темы самоубийства [4]. В этой ситуации применяют генеративные модели для создания синтетических данных. Примером таких моделей являются генеративно-сопоставительные сети (GAN).

Обучение GAN – игра между генератором, создающим данные, и дискриминатором, различающим реальные и искусственные данные. Мы будем использовать GAN Вассерштейна с градиентным штрафом (WGAN-GP), который минимизирует расстояние Вассерштейна между распределениями сгенерированных и реальных данных. Обоснование работы такой сети впервые было описано в [5]. В качестве генератора и дискриминатора мы используем LSTM сети.

Задача генерации текста заключается в обучении нейронной сети предсказывать вероятность каждого токена как следующего в последовательности. Токен с наибольшей вероятностью добавляется, и процесс повторяется. Распределение вероятностей получается на выходе с помощью функции активации Софтмакс. Препятствием для обучения такой сети становится недифференцируемость операции выбора следующего токена. Однако, описанная в [6] аппроксимация процесса выборки Софтмаксом Гумбеля позволяет обойти это ограничение. Для установки значения гиперпараметра этой функции мы воспользуемся описанной в [7] стратегией перехода от генерации близкого к равномерного к почти one-hot вектору по ходу обучения.

Исходный набор данных сформируем на основе данных с платформы Kaggle [8]. Мы сократим число образцов целевого класса, чтобы внести ожидаемый дисбаланс. В результате, из 25547 записей в обучающем множестве, суицидальное поведение будет в 5310. В тестовом множестве это отношение составит 6144 на 6638.

В ходе предобработки, выполнены следующие шаги: удаление HTML тэгов, ссылок, адресов электронной почты, отделение буквенных и небуквенных символов друг от друга пробелом, замена всех «ё» на «е», замена пробелом всех символов, не являющихся буквой русского или английского алфавита, знаком пунктуации или эмодзи, удаление лишних пробелов между словами.

Для классификации также удалены стоп-слова и проведён стемминг. Для CNN проведено усечение или дополнение текстов, чтобы добиться постоянного размера входа. В

тексте оставлены 20000 самых распространённых токенов, остальные заменены специальным токеном.

WGAN-GP обучена генерировать последовательности длиной в пять токенов. Для сравнения с реальными данными, исходные последовательности были разбиты на подпоследовательности соответствующей длины. Показанная на рисунке 1 динамика расстояния Вассерштейна по ходу обучения позволяет оценить качество генерации.

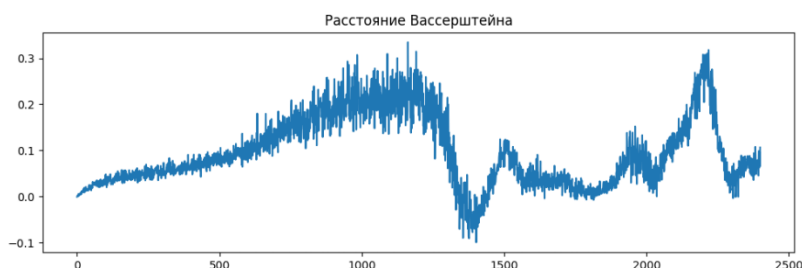


Рисунок 1. Обучение WGAN-GP.

Обученная модель использовалась для аугментации данных. Тестовое множество осталось неизменным, а в обучающем появилось ещё 12000 записей.

4. Полученные результаты

В Таблице 1 представлена информация о результатах классификации до и после добавления синтетических данных.

Таблица 1. Результаты классификации.

Метрика	До аугментации		После аугментации	
	Свёрточная нейронная сеть	Сеть долгой краткосрочной памяти	Свёрточная нейронная сеть	Сеть долгой краткосрочной памяти
Accuracy	0.9204	0.9360	0.9173	0.9312
Precision	0.8038	0.8661	0.8405	0.7961
Recall	0.8358	0.8328	0.7620	0.9171
F1	0.8195	0.8491	0.7994	0.8523

Для CNN наблюдается заметное уменьшение recall, но рост precision. Обратная картина в результатах обучения LSTM-сети. Применение синтетических данных, созданных WGAN-GP, оправдано, если используется CNN и стоит задача уменьшить количество «ложных тревог», или если используется LSTM-сеть и стоит задача уменьшить вероятность, что будет пропущен суицидальный контент.

5. Выводы

Результаты, полученные в рамках данной работы, могут быть использованы модераторами социальных сетей, специалистами по выявлению деструктивного поведения и многими другими.

В целом, результаты данной работы показывают, как можно упростить процесс обучения нейронных сетей в качестве инструментов выявления суицидального поведения. А также способствуют популяризации использования генеративно-состязательных сетей в задачах, требующих генерации текстовых данных.

Список литературы

1. Мищенко В.А. Психология аутодеструктивного поведения и его профилактика: курс лекций: учебное пособие / В.А. Мищенко. – Ставрополь: СКФУ, 2022. – 232 с.
2. Malhotra Anshu. Deep learning techniques for suicide and depression detection from online social media: A scoping review / Anshu Malhotra, Rajni Jindal // Applied Soft Computing. – 2022. – Volume 130: 109713.
3. Бузин Д.С. Алгоритмы машинного обучения для анализа тональности высказываний / Д.С. Бузин, М.Т. Азизов, Л.В. Лабунец // Вестник Российского нового университета. Серия Сложные системы модели, анализ и управление. – 2022. – № 2. – С. 129-139.
4. Ghanadian H. Socially Aware Synthetic Data Generation for Suicidal Ideation Detection Using Large Language Models / H. Ghanadian, I. Nejadgholi and H. A. Osman // IEEE Access. – 2024. – vol. 12. – pp. 14350-14363.
5. Gulrajani Ishaan. Improved training of Wasserstein GANs / Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron C. Courville // Advances in Neural Information Processing Systems. – 2017. – pp. 5767-5777.
6. Kusner M.J. Gans for sequences of discrete elements with the gumbel-softmax distribution / M.J. Kusner, J.M.H. Lobato. – URL: <http://arxiv.org/abs/1611.04051> (дата обращения: 21.04.2025). – 2016.
7. Tao Zhou. Password Guessing Based on GAN with Gumbel-Softmax / Zhou Tao, Wu Hao-tian, Lu Hui, Xu Peiming, Cheung Yiu-ming // Security and Communication Networks. – 2022. – Volume 2022. – pp. 1-12.
8. Kaggle. Depression-Suicide posts RU. – URL: <https://www.kaggle.com/datasets/eldamar1717/depression-suicide-twits-ru> (дата обращения: 21.04.2025).