

ТЕХНИЧЕСКИЕ НАУКИ



УДК 004.8

Применение методов искусственного интеллекта в задаче автоматизации процесса модерации объявлений

А.М. Израелян, Д.В. Фатхи

Донской государственный технический университет (г. Ростов-на-Дону, Российская Федерация)

Аннотация. Проанализировано применение методов и алгоритмов на основе искусственного интеллекта для решения задачи автоматизированной модерации объявлений. Рассмотрены методы кластеризации, анализа тональности и классификации текста. Использована библиотека TensorFlow для решения задач классификации. Разработана простая прогнозирующая модель на основе натренированного алгоритма машинного обучения и нейронной сети. Полученная точность модели составляет 76,8 %, что показывает эффективность использования искусственного интеллекта для решения поставленной задачи.

Ключевые слова: модерация текста, искусственный интеллект, нейронные сети, классификация текста, TensorFlow, точность.

Application of Artificial Intelligence Methods in the Problem of Automation of the Announcement Moderation Process

Ararat M Israelyan, Denis V Fatkhi

Don State Technical University (Rostov-on-Don, Russian Federation)

Abstract. The article analyzes the application of methods and algorithms based on artificial intelligence to solve the problem of automated ad moderation. Methods of clustering, tonality analysis and text classification are considered. The TensorFlow library is used to solve classification problems. A simple predictive model based on a trained machine learning algorithm and a neural network has been developed. The obtained accuracy of the model is 76.8%, which shows the effectiveness of using artificial intelligence to solve the task.

Keywords: text moderation, artificial intelligence, neural networks, text classification, Tensorflow, accuracy.

Введение. Современный интернет предоставляет пользователям огромное количество информации и контента. Однако рядом с полезным и качественным контентом существует также и множество нежелательного и незаконного, который может нанести вред как пользователям, так и платформам, на которых он распространяется, в частности это касается компаний, которые занимаются электронной коммерцией. В связи с этим задача автоматизации процесса модерации объявлений становится все более актуальной. Традиционно модерацией занимаются люди, однако с увеличением объема контента этот процесс становится все более сложным и требует большого количества времени и ресурсов.

В этой связи возникает необходимость в автоматизации процесса модерации контента. Современные методы искусственного интеллекта, такие как машинное обучение и нейронные сети, могут быть использованы для создания автоматизированных систем модерации контента. Эти системы способны автоматически определять и фильтровать нежелательный контент, классифицировать текстовые объявления, выявлять аномалии, что позволяет сократить затраты на модерацию и повысить ее эффективность.

Целью данной работы является анализ применения методов и алгоритмов на основе искусственного интеллекта для автоматизации процесса модерации объявлений. Для достижения этой цели необходимо рассмотреть существующие методы искусственного интеллекта, возможность их применения в задаче модерации объявлений по категориям и разработать прогнозирующую модель.

Основная часть. В задаче модерации текста объявлений можно использовать следующие методы [1]:

1. Кластеризация. Метод кластеризации в задаче модерации объявлений используется для автоматического разбиения текстовых объявлений на группы (кластеры) с похожим содержанием. Это позволяет упростить

процесс автоматической модерации, так как алгоритму-модератору не нужно проверять каждое объявление по отдельности, а только группу объявлений внутри кластера.

Для проведения кластеризации необходимо выбрать подходящий алгоритм, который сможет разбить объявления на группы. Можно использовать иерархическую кластеризацию, когда объявления постепенно объединяются на основе их сходства, или *k*-средние, когда изначально задается количество кластеров, а затем объявления распределяются между ними. После того как объявления были разбиты на кластеры, необходимо провести модерацию только для представителей каждого кластера. Это позволяет существенно ускорить процесс модерации и сократить количество времени и ресурсов, необходимых для проведения модерации.

Метод кластеризации может быть полезным, например, для обнаружения объявлений, содержащих информацию о нелегальных товарах или услугах, которые могут быть распространены под различными названиями.

2. Анализ тональности. Метод анализа тональности в задаче модерации объявлений используется для определения эмоциональной окраски текста объявлений, чтобы выявлять нежелательные объявления с негативной эмоциональной окраской или негативным содержанием.

Основной подход к анализу тональности — это машинное обучение на основе размеченных данных. Для этого необходимо подготовить набор данных, содержащий объявления с указанием их тональности (положительная, отрицательная, нейтральная). Предварительная обработка текста включает в себя удаление стоп-слов и пунктуации, лемматизацию и токенизацию. Затем на этом наборе данных обучается модель, которая сможет самостоятельно классифицировать тексты по тональности. Если текст имеет негативную тональность, то объявление может быть помечено как нежелательное и отклонено алгоритмом модерации или модератором.

3. Классификация текста. Метод классификации в задаче модерации объявлений используется для автоматического определения категории или типа объявления. Он является одним из основных методов машинного обучения, который позволяет классифицировать данные на основе известных признаков.

Для применения метода классификации необходимо подготовить набор данных, содержащий текст объявлений и соответствующие им категории. Затем происходит предварительная обработка данных, включающая токенизацию (разделение текста на слова или токены), лемматизацию (приведение слов к базовой форме), удаление стоп-слов (часто используемых слов, которые не несут значимой информации), а также приведение всех слов к нижнему регистру.

Далее происходит обучение модели, в котором на вход модели подается предобработанный текст объявления, а на выходе получается класс или категория, к которой относится объявление. Обучение модели происходит на основе алгоритма машинного обучения, такого как наивный Байес, решающее дерево, метод опорных векторов (SVM), нейронные сети, которые способны обрабатывать тексты и выявлять в них признаки, соответствующие заданным категориям, например «недвижимость», «транспорт», «электроника» и т. д.

Применение метода классификации в задаче модерации объявлений позволяет автоматически классифицировать тексты объявлений на основе заранее известных категорий, что упрощает процесс модерации и повышает эффективность работы алгоритмов-модераторов.

В зависимости от конкретных требований и характеристик задачи модерации объявлений можно применять различные методы и их комбинации для достижения наилучшего результата.

В данной работе для разработки прогнозирующей модели для модерации объявлений по категориям был выбран метод классификации, а для реализации была выбрана библиотека машинного обучения с открытым исходным кодом TensorFlow [2].

Прогнозирующая модель — алгоритм, который на основе предыдущих данных строит модель для прогнозирования будущих значений. Она используется в различных областях, таких как финансы, экономика, медицина, техника и многие другие. В данной работе она необходима для классификации текста по категориям.

В контексте машинного обучения прогнозирующая модель является нейронной сетью, которая на основе обучающих данных настраивает свои веса и соединения между нейронами для прогнозирования будущих значений. Она состоит из следующих этапов:

1. Подготовка данных — сбор данных и их предобработка для использования в модели.
2. Разработка модели — выбор типа нейронной сети и ее параметров для построения модели.
3. Обучение модели — использование обучающих данных для настройки весов и соединений между нейронами в модели.
4. Оценка модели — использование тестовых данных для оценки точности и эффективности модели.
5. Использование модели — использование модели для прогнозирования будущих значений на основе новых данных.

Для реализации модели была определена архитектура. Схема процесса создания прогнозирующей модели представлена на рис. 1. Она состоит из алгоритма машинного обучения, «натренированного» на тестовых данных, взятых из открытых источников. В тренировочном наборе данных все тексты помечены метками, указывающими, к какой категории он принадлежит.

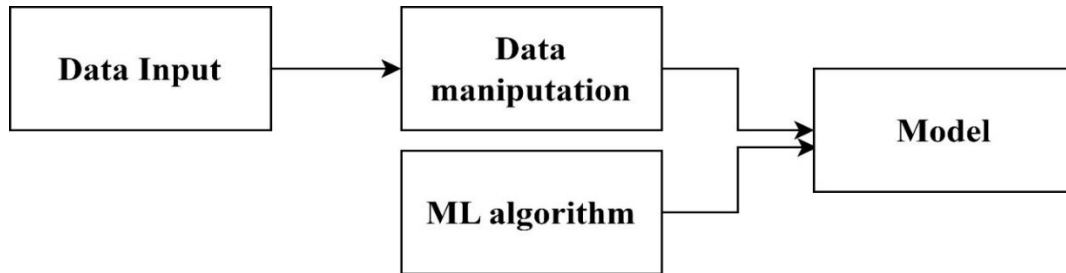


Рис. 1. Схема процесса создания прогнозирующей модели

Архитектура нейронной сети представлена в виде двух скрытых слоев и слоя вывода. Для получения результатов слоя вывода используется унитарное кодирование. Унитарное кодирование — это метод кодирования категориальных признаков, при котором каждая категория закодирована вектором, в котором только один элемент равен 1, а остальные равны 0. Категории, которые мы кодируем: «электроника», «одежда» и «автомобили». Каждый скрытый слой содержит несколько нейронов, которые выполняют линейные и нелинейные преобразования входных данных, а слой вывода производит классификацию данных на основе выходных значений скрытых слоев и унитарного кодирования категорий. Функция Softmax преобразует вывод в вероятностное распределение по нескольким категориям [3]. Она принимает на вход вектор значений и вычисляет вероятность каждого значения в этом векторе. Softmax функция определяется следующим образом:

$$\sigma(z)_i = \frac{e^{z_i}}{\sum_{k=1}^k e^{z_k}}$$

В результате функция Softmax преобразует вектор \square в новый вектор размерности \square , где каждый элемент является вероятностью относительно других элементов вектора. Кроме того, эти вероятности всегда суммируются до 1.0, что делает Softmax функцию полезной для задач классификации с несколькими категориями.

Таким образом, полученный в результате расчётов вывод отражает вероятность принадлежности текста к каждой категории. Результат работы функции представлен на рис. 2.

$$\begin{array}{|l} 1.2 \\ 0.9 \\ 0.4 \end{array} \rightarrow [\text{softmax}] \rightarrow \begin{array}{|l} 0.50 \\ 0.32 \\ 0.18 \end{array}$$

Рис. 2. Принадлежность текста к каждой категории

После выполнения обучения и получения вероятностных предсказаний натренированная модель оценивается на основе двух параметров — точность и ошибки. Параметр accuracy (точность) определяет долю правильных предсказаний в классе относительно всех предсказаний в этом классе. Она измеряет, как часто модель правильно предсказывает метки категорий.

Параметр loss (ошибка) отражает степень отклонения прогнозируемых значений модели от истинных значений. Он является мерой того, насколько хорошо модель работает во время обучения. Чем меньше значение потерь, тем более точные предсказания делает модель. Однако максимизация точности не всегда соответствует минимизации потерь, и выбор метрики зависит от конкретной задачи и требований к модели.

Результаты оценки производительности данной модели, полученные на основе тестовых данных, сведены в графики, которые отображены на рис. 3. и 4.

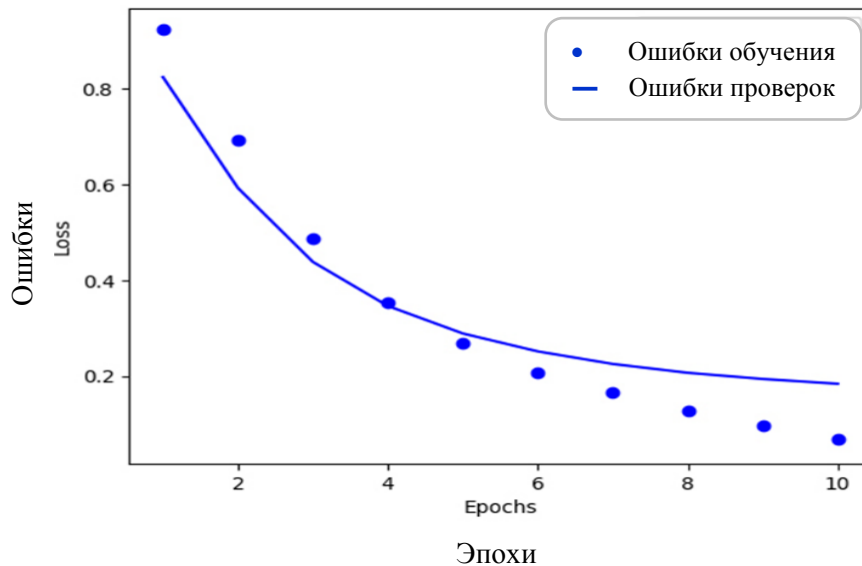


Рис. 3. График потерь обучения

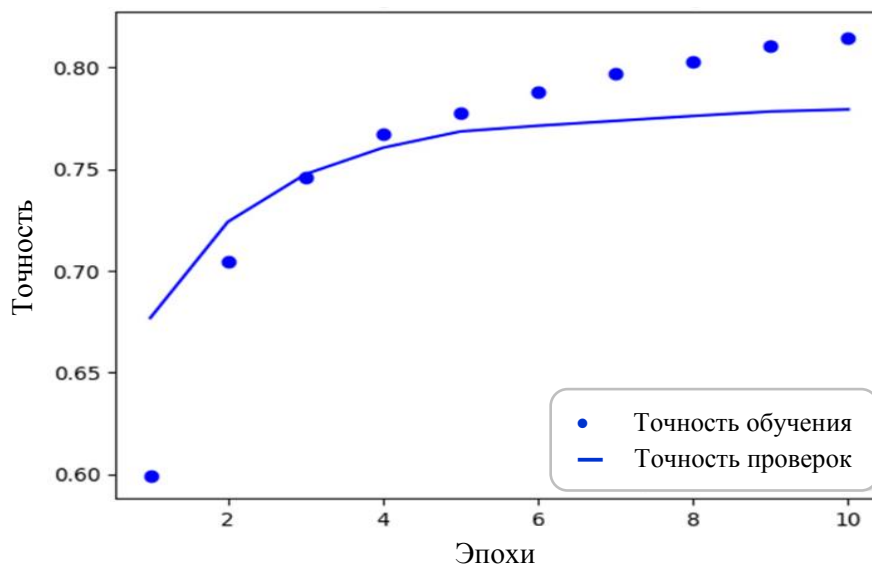


Рис. 4. График точности обучения

Таким образом, создана модель с использованием нейронной сети для классификации текстов объявлений по заданным категориям. Полученная точность составляет 0.768 или 76,08 %, что является хорошим показателем производительности модели.

Выводы. В заключение следует отметить, что цель работы была достигнута — создана прогнозирующая модель с использованием нейронной сети для классификации текстов объявлений по категориям с хорошей точностью. Были рассмотрены различные методы искусственного интеллекта, которые можно применять в задаче модерации контента, а также использованы инструменты и алгоритмы из библиотеки Tensorflow для решения задач классификации. Данная работа может быть полезной для применения на интернет-платформах, в которых актуальна задача модерации объявлений.

Библиографический список

1. Srivastava A.N., Sahami M. *Text Mining: Classification, Clustering and Applications*. New York: CRC Press; 2021. 328 p. <https://doi.org/10.1201/9781420059458>
2. Pramod S., Avinash M. *Learn TensorFlow 2.0: Implement Machine Learning and Deep Learning Models with Python*. Karnataka, India: Apress; 2019. 164 p.
3. Классификация данных при помощи нейронных сетей. *Аналитическая платформа Loginom*. URL: <https://loginom.ru/blog/neural-classification> (дата обращения: 26.03.2023).

4. *Understand the Softmax Function in Minutes*. Medium. URL: <https://medium.com/data-science-bootcamp/understand-the-softmax-function-in-minutes-f3a59641e86d> (accessed 26.03.2023).

Об авторах:

Фатхи Денис Владимирович, доцент кафедры «Информационные технологии» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), кандидат технических наук, zmey2257@mail.ru

Израелян Арарат Михайлович, магистрант кафедры «Информационные технологии» Донского государственного технического университета (344003, РФ, г. Ростов-на-Дону, пл. Гагарина, 1), ararat.israelyan@mail.ru

About the Authors:

Denis V Fatkhi, associate professor of the Information Technologies Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, RF), Cand. Sci. (Eng.), zmey2257@mail.ru

Ararat M Israelyan, Master's degree student of the Information Technologies Department, Don State Technical University (1, Gagarin Sq., Rostov-on-Don, 344003, RF), ararat.israelyan@mail.ru