Применение методов искусственного интеллекта для анализа и фильтрации текстового контента

Н.А. Семыкина[1,2], И.А. Шаповалова[1]

[1]Тверской государственный университет
[2]АО НИИ ЦентрПрограммСистем, г. Тверь

Аннотация: Одно из основных условий обеспечения информационной безопасности состоит в предупреждении распространения ложных и намеренно искаженных сведений. Решением данной задачи может служить фильтрация контента информационных ресурсов Интернета. Последнее время для анализа и классификации распространяемых данных все чаще рассматривается подход, использующий методы и математические модели искусственного интеллекта. Использование нейросетей позволяет автоматизировать процесс обработки большого массива информации и подключать человека только на этапе принятия решения. В работе основное внимание уделено процессу обучения нейронной сети. Рассмотрены различные алгоритмы обучения: стохастический градиентный спуск, Adagrad, RMSProp, Adam, AdaMax и Nadam. Приведены результаты реализации распознавания тематики текста с помощью рекуррентной нейронной сети модели LSTM. Представлены результаты вычислительных экспериментов, проведен анализ и сделаны выводы.

Ключевые слова: информационная безопасность, анализ текста, метод искусственного интеллекта, искусственная нейронная сеть, рекуррентная сеть LSTM.

Введение

По данным различных международных организаций одной из главных опасностей для общества и личности признана угроза распространения заведомо ложной информации, которая может оказывать деструктивное воздействие на информационно-психологическую безопасность человека и страны в целом. Если в 2021 году было зафиксировано 1,5 млн. ссылок с фейками, то в 2023 году количество ссылок выросло до 11 млн. [1-3].

Решение задачи обеспечения защиты от недостоверной информации реализуется в различных направлениях: юридическом, техническом, психологическом и воспитательном. В частности, России в марте 2022 года внесены изменения в Уголовный и Административный кодексы с целью уголовно-правовой борьбы с публичным распространением ложной информации [4]. Один из технических методов предотвращения доступа к антисоциальной информации в сети Интернет является автоматическое

блокирование сайтов. Для этого требуется анализировать и классифицировать содержание ресурсов сети и документов информационного контента. Последнее время для фильтрации текста активно используются методы и математические модели искусственного интеллекта [5-9]. Этот подход обладает такими преимуществами, как самообучаемость, адаптируемость и автономность.

Исследуя различные источники, можно сделать вывод, что большинство авторов для задач обработки естественного языка использовали рекуррентные нейронные сети (RNN) [7-9]. Этот тип сетей обладает следующими достоинствами: учет последовательности при обработки входных данных, возможность ввода данных переменной длины, распознавание закономерностей и взаимосвязи в документах [9,10]. Однако у RNN есть и недостатки. Один из которых – проблема с исчезающим градиентом, когда значение градиента экспоненциально уменьшается во времени [9]. Для решения этой проблемы используют усложненный тип архитектуры – рекуррентная нейронная сеть с долгой краткосрочной памятью (LSTM), которую будем далее использовать в экспериментах [10].

В статье уделим внимание процессу обучения, цель которого состоит в нахождении оптимальных весовых коэффициентов, которые доставляют минимум функции потерь. Будем рассматривать обучение искусственной нейронной сети (ИНС) с учителем. Существуют различные алгоритмы обучения (оптимизаторы). Рассмотрим их и сравним результаты использования.

Алгоритмы обучения ИНС

Для исследования рассмотрим различные методы обучения ИНС: стохастический градиентный спуск, Adagrad, RMSProp, Adam, AdaMax и Nadam [11,12].

В основе алгоритма стохастического градиентного спуска (SGD) лежит классический градиентный метод, с учетом обновления параметров на каждом шаге. В этом методе большое значение играет выбор коэффициента скорости обучения (шаг градиентного спуска).

Алгоритм Adagrad позволяет настроить коэффициент скорости обучения по следующему принципу: если в обучающей выборке встречаются типичные признаки, то обновление происходит реже, в обратном случае – чаще. Это позволяет исключить подбор шага градиентного спуска, как в предыдущем оптимизаторе. Однако оптимизатор Adagrad устремляет коэффициент скорости обучения бесконечно малым значением, что в дальнейшем не дает использовать алгоритм [12].

Модификацией Adagrad является метод RMSProp, который, за счет использования экспоненциально скользящей средней, дает возможность затормозить уменьшение коэффициента скорости обучения [12].

Оптимизатор Adam объединяет в себе метод RMSProp и метод Adagrad – применяют экспоненциально скользящую среднюю для значения градиентов и правило обновления параметров сети [12].

Метод AdaMax является модификацией алгоритма Adam. Для этого в формуле оценивания изменения градиента используем инерционный момент распределения градиентов произвольной степени [12].

Алгоритм Nadam использует идею применения ускоренного градиента Нестерова, которая заключается в вычислении градиента вперед по вектору обновления, для оптимизатора Adam [12].

Сравнительный анализ оптимизаторов для обучения ИНС в задаче фильтрации текста

Задача вычислительного эксперимента будет состоять в определении оптимизатора, который даст наилучший результат при обучении ИНС для анализа и фильтрации текстового документа.

Требуется выяснить, содержит ли текст информацию о конкретной предметной области. В нашем исследовании определим тему – «Искусственные нейронные сети».

По сути постановки решается задача бинарной классификации. Поэтому рассмотрим в качестве функции потерь – бинарную перекрестную энтропию.

Для достижения поставленной цели требуется решить частные задачи: формирование набора данных для обучения нейронной сети; построение выбранной модели нейронной сети; проведение вычислительных экспериментов; оценка качества обучения для разных видов оптимизатора.

Программная реализация проведена с помощью языка Python, библиотеки глубокого обучения Keras и библиотеки NumPy. Для обработки текста, создания и сохранения частотного словаря использованы библиотеки Pymorphy2 и JSON [11].

Сеть долгой краткосрочной памяти (LSTM) в слое активации использует сигмоидальную функцию. Для предотвращения переобучения в методе Dropout доля отключения определена 20%. Параметрами сети являются размерность словаря – 5594 слова, длина высказывания – 80 слов.

Так как на качество обучения ИНС влияет число эпох обучения, то этот параметр также будет исследоваться в численном эксперименте. Рассматриваются количество эпох обучения на тренировочной выборке от 5 до 10.

Ниже приведены результаты численных экспериментов для перечисленных оптимизаторов обучения ИНС и различного числа эпох. В таблице 1 использованы следующие обозначения: E_o – точность обучения ИНС с единицами измерения %; $E_т$ – точность тестовых данных с единицами измерения %.

Как можно видеть из таблицы 1, применение алгоритма SGD для различного числа эпох обучения точность обучения и точность тестовых данных не изменились, и при этом данный метод показал наихудший результат.

Таблица № 1

Результаты обучения ИНС

E_o (%) / $E_т$ (%)		Типы оптимизаторов					
		SGD	Adagrad	RMSProp	Adam	AdaMax	Nadam
Число эпох обучения	5	60.67 52.00	99,33 98,00	99,00 98,00	98,33 96,00	71,00 52,00	99,67 98,00
	6	60.67 52.00	99,67 99,00	99,67 99,00	99,33 99,00	90,33 66,00	99,00 94,00
	7	60.67 52.00	100,00 99,00	99,67 99,00	99,33 99,00	98,67 96,00	99,67 98,00
	8	60.67 52.00	100,00 99,00	99,67 99,00	99,33 99,00	98,67 98,00	99,67 98,00
	9	60.67 52.00	99,67 98,00	100,00 99,00	99,33 99,00	99,00 99,00	99,00 96,00
	10	60.67 52.00	100,00 99,00	99,67 99,00	99,67 98,00	99,00 99,00	98,67 98,00

Заметим, что число эпох также влияет на время обучения ИНС. И может возникнуть ситуация переобучения, когда нейронная сеть с достаточно высокой точностью работает на тренировочной выборке и с низкой точностью – на тестовых данных.

Поэтому при выборе лучшего оптимизатора ориентируемся на точность тестовых данных. В этом случае методы Adagrad и RMSProp показали одинаковый лучший результат. А именно, на 6 эпохах обучения ИНС результат теста достиг 99 %.

Заключение

В работе были представлены результаты тестирования различных типов оптимизаторов для применения в задаче бинарной классификации текстовых документов. По совокупности характеристик (точность и время обучения, точность тестовых данных) наиболее удачными алгоритмами являются Adagrad и RMSProp.

Изучение типов оптимизаторов, их анализ и сравнение позволит в дальнейшем оптимизировать процесс обучения рекуррентной нейронной сети, что даст возможность уменьшить вычислительные ресурсы системы.

Литература

1. Количество фейковых сообщений за три года выросло в 10 раз // Ведомости. URL: vedomosti.ru/society/news/2023/11/02/1003940-kolichestvo-feikovih-soobschenii-viroslo. (Дата обращения 24.06.2024)

2. Количество фейков в сети в 2022 году выросло в шесть раз // ТАСС. URL: tass.ru/obschestvo/16642301. (Дата обращения 24.06.2024)

3. Ложная информация - это не проблема. Это опасность для всего человечества // SecurityLab.ru by Positive Technologies. URL: securitylab.ru/news/545390.php. (дата обращения 24.06.2024)

4. Абдулатипов А.М. Уголовно-правовая характеристика публичного распространения под видом достоверных сообщений заведомо ложной информации // Юридический вестник Дагестанского государственного университета. 2022. №2. URL: cyberleninka.ru/article/n/ugolovno-pravovaya-harakteristika-publichnogo-rasprostraneniya-pod-vidom-dostovernyh-soobscheniy-zavedomo-lozhnoy-informatsii (дата обращения 27.06.2024).

5. Ларионова А. В., Хорев П. Б. Метод фильтрации спама на основе искусственной нейронной сети // Вестник евразийской науки. 2016. №3 (34). URL: cyberleninka.ru/article/n/metod-filtratsii-spama-na-osnove-iskusstvennoy-neyronnoy-seti (дата обращения 27.06.2024).

6. Красников И.А., Никуличев Н.Н. Гибридный алгоритм классификации текстовых документов на основе анализа внутренней связности текста // Инженерный вестник Дона, 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1773 (дата обращения 24.06.24).

7. Akhter M., Jiangbin, Zh., Naqvi, S.I., Abdelmajeed M., etc. (2020). Document-level text classification using single-layer multisize filters convolutional neural network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2976744. URL: researchgate.net/publication/339547689 (дата обращения 27.06.2024).

8. RNN for Text Classifications in NLP. URL: geeksforgeeks.org/rnn-for-text-classifications-in-nlp/.

9. Text Classification with RNN. URL: scaler.com/topics/tensorflow/rnn-for-text-classification/.

10. Федоров В.Х., Васюков Д.Ю., Лаута О.С, Баленко Е.Г., Иванов Д.А. Подход к работе системы защиты сети передачи данных от компьютерных атак на основе гибридной нейронной сети // Инженерный вестник Дона, 2023, №1. URL: ivdon.ru/ru/magazine/archive/n1y2023/8163 (дата обращения 24.06.24).

11. Keras Documentations Optimizers. URL: keras.io/optimizers/ (дата обращения 15.06.2024)

12. Садовников П. Методы оптимизации нейронных сетей. URL: habr.com/ru/post/318970/ (дата обращения 25.06.2024).

References

1. Kolichestvo fejkovyh soobshchenij za tri goda vyroslo v 10 raz [The number of fake messages has increased 10 times in three years]. Vedomosti. URL: vedomosti.ru/society/news/2023/11/02/1003940-kolichestvo-feikovih-soobschenii-viroslo. (date accessed 24.06.2024)

2. Kolichestvo fejkov v seti v 2022 godu vyroslo v shest' raz [The number of fakes on the web has increased six fold in 2022]. TASS. URL: tass.ru/obschestvo/16642301. (date accessed 24.06.2024)

3. Lozhnaya informaciya - eto ne problema. Eto opasnost' dlya vsego chelovechestva [False information is not a problem. This is a danger to all mankind]. SecurityLab.ru by Positive Technologies. URL: securitylab.ru/news/545390.php. (date accessed 24.06.2024)

4. Abdulatipov A. M. Yuridicheskij vestnik Dagestanskogo gosudarstvennogo universiteta, 2022. №2. URL: cyberleninka.ru/article/n/ugolovno-pravovaya-harakteristika-publichnogo-rasprostraneniya-pod-vidom-dostovernyh-soobscheniy-zavedomo-lozhnoy-informatsii. (date accessed 27.06.2024)

5. Larionova A. V., Horev P. B. Vestnik evrazijskoj nauki, 2016. №3 (34). URL: cyberleninka.ru/article/n/metod-filtratsii-spama-na-osnove-iskusstvennoy-neyronnoy-seti. (date accessed 27.06.2024)

6. Krasnikov I.A., Nikulichev N.N. Inzhenernyj vestnik Dona, 2013, №3. URL: ivdon.ru/ru/magazine/archive/n3y2013/1773. (date accessed 24.06.24)

7. Akhter M., Jiangbin, Zh., Naqvi, S.I., Abdelmajeed M., etc. (2020). Document-level text classification using single-layer multisize filters convolutional neural network. IEEE Access. PP. 1-1. 10.1109/ACCESS.2020.2976744. URL: researchgate.net/publication/339547689. (date accessed 27.06.2024)

8. RNN for Text Classifications in NLP. URL: geeksforgeeks.org/rnn-for-text-classifications-in-nlp/.

9. Text Classification with RNN. URL: scaler.com/topics/tensorflow/rnn-for-text-classification/.

10. Fedorov V.X., Vasyukov D.Yu., Lauta O.S, Balenko E.G., Ivanov D.A. Inzhenernyj vestnik Dona, 2023, №1. URL: ivdon.ru/ru/magazine/archive/n1y2023/8163. (date accessed 24.06.2024)

11. Keras Documentations Optimizers. URL: keras.io/optimizers/. (date accessed 15.06.2024)

12. Sadovnikov P. Metody optimizacii nejronnyh setej [Methods for optimizing neural networks]. URL: habr.com/ru/post/318970/. (date accessed 25.06.2024)

Дата поступления: 3.06.2024

Дата публикации: 11.07.2024