

УДК 681.3

DOI: [10.26102/2310-6018/2021.35.4.012](https://doi.org/10.26102/2310-6018/2021.35.4.012)

Сравнительный анализ результатов, полученных при решении задачи анализа тональности текста с помощью сверточной и рекуррентной нейронных сетей

Д.В. Меняйлов, А.П. Преображенский
Воронежский институт высоких технологий,
Воронеж, Российская Федерация

Резюме. В современном мире существуют различные средства коммуникации: электронные устройства, web- и мобильные приложения (интернет-форумы, чаты, блоги, социальные сети). В результате возникает огромное количество информации о самих пользователях, их отношении к другим людям, событиям, происходящим в мире. Данная информация может быть полезна при моделировании процессов, происходящих в обществе, прогнозировании поведения людей. В связи с тем, что информация, находящаяся в интернете, во многих случаях представлена в виде текста на естественном языке, необходимо использовать методы компьютерной лингвистики. Например, по тексту требуется определить, какую эмоцию он несет. Для этого достаточно будет классифицировать эмоцию на позитивную и негативную. В работе дан анализ подходов, которые могут при этом использоваться: сверточные нейронные сети (Convolutional Neural Networks) и рекуррентные нейронные сети (Recurrent Neural Network). Авторами составлен алгоритм обработки текста на естественном языке. Созданный в работе программным образом алгоритм реализуется с помощью Word2Vec, SQLite, Python: Gensim, Keras, функция активации ReLU, функция оптимизации Adam. Даны рекомендации по предварительной обработке текста на естественном языке (приведение к нижнему регистру, замена «ё» на «е», замена ссылок на токен «URL», замена упоминания пользователя на токен «USER», удаление знаков пунктуации), чтобы осуществлять полноценный анализ тональности текста. Показаны основные этапы алгоритма анализа тональности текста. В статье продемонстрирован пример фрагментов программного кода, описывающих работу ключевых этапов алгоритма. Проведено сравнение результатов применения сверточной нейронной сети и рекуррентной нейронной сети. Они продемонстрировали, что, с точки зрения определения позитивной, негативной, нейтральной тональности, лучшими характеристиками по точности обладает сверточная нейронная сеть.

Ключевые слова: анализ тональности, обработка записей и комментариев, информация, текст, сверточные нейронные сети, рекуррентная нейронная сеть.

Для цитирования: Меняйлов Д.В., Преображенский А.П. Сравнительный анализ результатов, полученных при решении задачи анализа тональности текста с помощью сверточной и рекуррентной нейронных сетей. *Моделирование, оптимизация и информационные технологии*. 2021;9(4). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1039> DOI: 10.26102/2310-6018/2021.35.4.012

Comparative analysis of the results obtained when solving the problem of analyzing the text sentiment using convolutional and recurrent neural networks

D.V. Menyaylov, A.P. Preobrazhenskiy

Voronezh Institute of High Technologies, Voronezh, Russian Federation

Abstract: In the modern world, there are various means of communication: electronic devices, web and mobile applications (Internet forums, chats, blogs, social networks). As a result, there is a huge amount of

information about the users themselves, about their attitude to other people, to events taking place in the world. This information can be useful in modeling the processes occurring in society, predicting people's behavior. Due to the fact that the information on the Internet is in many cases presented in the form of a text in a natural language, it is necessary to use the methods of computational linguistics. For example, according to the text, it is required to determine what emotion it carries. To do this, it will be enough to classify the emotion into positive and negative. The paper analyzes the approaches that can be utilized in this case: Convolutional Neural Networks and Recurrent Neural Networks. The authors have compiled an algorithm for processing text in natural language. The algorithm, programmatically created in the work, is implemented using Word2Vec, SQLite, Python: Gensim, Keras, ReLU activation function, Adam optimization function. Recommendations are given for the pre-processing of natural language text (converting to lower case, replacing "ё" with "e", replacing links to the "URL" token, replacing the username with the "USER" token, removing punctuation marks) in order to carry out a complete tonality analysis of the text. The main stages of the text sentiment analysis algorithm are shown. The paper demonstrates an example of code fragments describing the operation of the algorithm key stages. Comparison of the results of a convolutional neural network and a recurrent neural network application is carried out. They demonstrated that in terms of determining positive, negative, neutral tonality, the convolutional neural network has the best characteristics in terms of accuracy.

Keywords: sentiment analysis, processing of notes and comments, information, text, convolutional neural networks, recurrent neural network

For citation: Menyaylov D.V., Preobrazhenskiy A.P. Comparative analysis of the results obtained when solving the problem of analyzing the sentiment of a text using convolutional and recurrent neural networks. *Моделирование, оптимизация и информационные технологии*. 2021;9(4). Доступно по: <https://moitvvt.ru/ru/journal/pdf?id=1039> DOI: 10.26102/2310-6018/2021.35.4.012 (In Russ).

ВВЕДЕНИЕ

Тональность рассматривается в виде эмоционального отношения авторов высказываний к некоторым объектам в реальном мире, событиям, процессам или их характеристикам, которые выражаются в текстах. В ходе поиска эмоций в тексте, например, грусти, гнева или радости, можно разделить итог поиска на негативную, позитивную и нейтральную тональность. Сверточные нейронные сети хорошо справляются с подобными задачами.

Данной проблеме посвящено множество работ. В работе [1] представлен первый этап разработки классификатора корпуса текста по тональности эмоциональной окраски на основе частей речи. Авторы анализируют термин «тон» с точки зрения эмоциональной окраски сообщения. То есть представлена демонстрация положительного отношения к чему-либо, отрицательного отношения или нейтрального. На первом шаге при разработке классификатора происходит процесс выявления паттернов применения определенных частей речи. Это зависит от того, какой тоновый цвет сообщения. Далее по шаблонам строится классификатор. В начале работы собран обучающий корпус текстов. Корпус автоматически разделен на положительные, отрицательные и нейтральные сообщения. Чтобы выделить закономерности и зависимости эмоционально окрашенных текстов, образцы отмечены частями речи. Затем ведется анализ и происходит выявление основных закономерностей применения соответствующих форм частей речи в позитивно, негативно и нейтрально окрашенных сообщениях, далее алгоритм классификации строится на основе шаблонов.

В [2] представлен обзор работ, связанных с анализом тональности. Это одно из перспективных направлений в автоматической обработке естественного языка. В настоящее время можно говорить об анализе тональности в виде отдельной области компьютерной лингвистики. Автором обозначены основные термины, связанные с этой теорией. Представлено краткое описание происхождения анализа настроений, описаны уровни анализа настроений. Показаны проблемы, связанные с анализом сантиментов,

включая выявление неявной оценки, сарказма и иронии. Также указаны проблемы устранения неоднозначности, монотематических систем, кореферентности и референции. Представлены компьютерные подходы к улучшению результатов программ анализа тональности, но больше внимания уделяется лингвистическим подходам. Рассмотрены вопросы формирования специальных лингвистических ресурсов. Они необходимы для сентимент-анализа. Отмечаются, например, корпуса и лексиконы. Кроме того, показаны теории, которые связаны с синтаксическими отношениями, риторическими структурами текста и дополнением блока правил.

В работе [3] обсуждается проблема классификации изображений и дается краткое описание структуры сверточных нейронных сетей. Проведен обзор сверточных нейронных сетей для задачи классификации изображений и проведено сравнение их точности на примере аннотированной базы данных изображений ImageNet.

В [4] рассматривается инструмент, позволяющий проводить сбор и анализ корпуса коротких текстов. В нем первый модуль представлен в виде текстового сборщика (парсера). Второй модуль связан с обработкой корпуса. В нем есть две компоненты. Первая осуществляет морфологическую разметку корпуса. Во второй формируется общий тематический словарь эмоциональной лексики. Тогда обозначаются списки по наиболее часто применяемым словам и фразам. Это дает возможности выделять значимые слова по положительным и отрицательным текстам. При этом ведется расчет набора статистических характеристик по каждому слову и фразе. Они являются характерными для положительных или отрицательных текстовых окрасок. Проведено построение словаря эмоциональной лексики с привлечением корпуса текстов. Осуществлен расчет статистического веса значимости слов по каждому из терминов или фраз в словарях при учете класса положительного или отрицательного текста. Для того, чтобы строить и обучать классификатор тонов, могут быть полезными полученные результаты.

В [5] можно увидеть результаты, связанные с аналитическим исследованием рекуррентных нейронных сетей (РНС). Рассмотрена их обобщающая классификация. Она осуществлена с точки зрения динамических систем. При этом учитывается новый класс РНС – резервуарные РНС. За счет обобщения результатов были выделены ключевые динамические режимы работы РНС. Кроме того, указаны достаточно перспективные направления по развитию способов обучения РНС при учете выявленных достоинств и недостатков в действующих способах.

В [6] дано рассмотрение общих представлений, связанных с анализом необработанных многоязычных текстов. Представлена созданная нейронная сеть, базирующаяся на долгой краткосрочной памяти (LSTM). Она позволяет размечать и дополнительным образом создавать символьные последовательности. Проведено обучение нейронной сети выделению лемм, формированию частеречной разметки и определению морфологических признаков. Использовалась программа UDPipe 1.2 для того, чтобы разбивать тексты по предложениям, вести токенизацию и синтаксический анализ. Из результатов работы видна актуальность использования рассматриваемой архитектуры в настоящее время.

Цель. Оценка целесообразности применения сверточной или рекуррентной нейронных сетей в задаче анализа тональности текста при проведении анализа обращений по сервису на предмет качества (выявлении мнений пользователей о сервисе).

Для достижения цели в работе были решены **задачи**:

1. Выявление факта частого использования анализа тональности текста в большинстве систем анализа текстовой информации и на практике.
2. Рассмотрены алгоритмы анализа тональности текста с помощью сверточной и рекуррентной нейронных сетей.

3. Проведена экспериментальная оценка эффективности для каждой нейронной сети на примере естественного текста на русском языке.

4. Сравнение результатов при применении сверточной нейронной сети с результатами, полученными при использовании рекуррентной нейронной сети.

Проблемы:

1. Сарказм рассматривается в виде одного из самых сложных чувств, с точки зрения того, чтобы его автоматическим образом отслеживать и правильным образом интерпретировать.

2. Нейтральные настроения следует делить по конкретным темам. Это позволит вести раскрытие каждого из мнений более подробным образом.

3. Для смешанных или многомерных настроений внутри одного предложения можно наблюдать как позитивы, так и негативы.

4. Настроение является условным, поскольку есть его связь с последующими действиями.

5. Не во всех случаях основная тема касается положительных чувств.

6. Негативные слова могут быть неоднозначными. Следует провести разбор их контекста, а также сделать соответствующие пометки.

Метод может быть полезен на практике:

1. Фирмы стремятся к тому, чтобы определить мнение потребителей относительно их продукта или услуг. Полученные данные применяют для обеспечения роста в качестве их работы, определения характеристик по целевым аудиториям и при распознавании ключевых недостатков и достоинств существующих конкурентов.

2. Специалисты интересуются мнением общества, чтобы проголосовать за кандидата на политических выборах, и, учитывая мнения других, принимают свое решение.

3. Для обработки потоков информации, выявление мнений клиентов, присутствующих в тексте.

4. Применение в рекомендательных системах, рекомендации для пользователей и участников.

МАТЕРИАЛЫ И МЕТОДЫ

Анализ тональности текста с помощью сверточной нейронной сети

Рассмотрим, предложенный алгоритм анализа тональности текста с помощью сверточной нейронной сети. Рисунок 1.

Предварительно обрабатываем «Выгрузку обращений и отзывов» за определенный период по сервису, для дальнейшего проведения анализа на предмет тональности.

Считаем данные с выгрузки и сформируем набор данных:

Считывание данных:

- положительные данные;
- отрицательные данные.

Формирование сбалансированного датасета (набора данных):

- размер образца (начальная форма положительных и отрицательных данных);
- первоначальные данные (объединение значений размеров выборок положительных и отрицательных данных, с начальной осью).

Проведем следующие действия: нижний регистр, замена «ё» на «е», удаление знаков пунктуации, ссылок на токен «URL», токен «USER» вместо участника.

СТРУКТУРА АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТОВ С ПОМОЩЬЮ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ

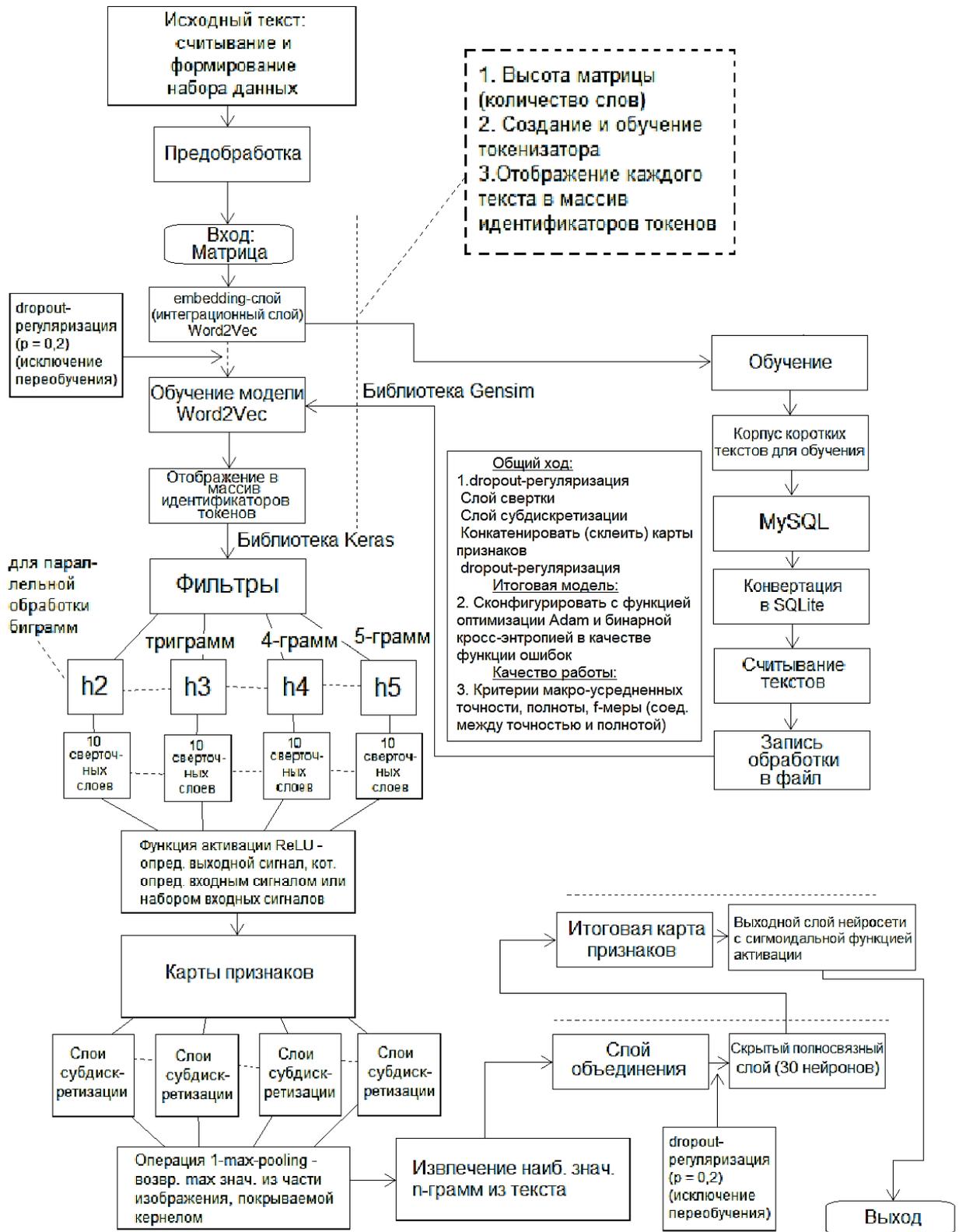


Рисунок 1 – Структура анализа тональности текстов с помощью сверточных нейронных сетей

Figure 1 – The structure of sentiment analysis of texts using convolutional neural networks

Процедура предварительной обработки:

- приведение к нижнему регистру;
- замена «ё» на «е»;
- замена ссылок на токен «URL»;
- замена упоминания пользователя на токен «USER»;
- удаление знаков пунктуации.

Разделение по выборкам. Данные мы также разделили на обучающую (по которой производится настройка, оптимизация параметров модели зависимости) и тестовую выборку (по которой оценивается качество построенной модели) в соотношении 4:1:

Разбивка набора данных на обучающую и тестовую выборку в соотношении 4:1, параметры разбивки:

- данные;
- метки;
- значение размера теста = 0,2;
- значение случайного состояния = 1.

На входе – матрица, имеющая высоту, строка – векторное отображение слова в признаковое пространство размерности k . Для интеграционного слоя нейросети также использовалась утилита Word2Vec (отображает значение слов в векторное пространство и находит связи между словами при условии, если в подобных контекстах есть семантически близкие слова). В данном случае определить границы предложений было сложно, поэтому допустим, что один комментарий – это одно предложение.

База существует в формате SQL, сконвертируем в SQLite. Дамп MySQL DB. И преобразуем дампы в SQLite DB.

Работа с базой данных:

- дампы базы данных MySQL;
- преобразование дампа в базу данных SQLite.

Откроем SQLite базу данных, считаем тексты комментариев, и запишем обработку в файл.

Запись обработки в файл:

- открытие SQLite базы данных;
- считывание текстов твитов;
- запись преобработанных твитов в файл.

Используя библиотеку Gensim, обучим модель Word2Vec (Таблица 1).

Таблица 1 – Параметры обучения модели Word2Vec
Table 1 – Parameters of training the Word2Vec model

Размерность признакового пространства	Количество слов из контекста, которое анализирует алгоритм	Минимальное значение встречаемости слова
200	5	3

Считаем файл с обработкой и обучим модель.

Считывание файла и обучение модели:

- считывание файла с предобработанными твитами;
- обучение модели: *Word2Vec* (размерность признакового пространства = 200, количество слов из контекста, которое анализирует алгоритм = 5, минимальное значение встречаемости слова = 3).

Далее текст отображается в массив идентификаторов токенов. При размерности вектора текста = 26 покрывается 99 % всех комментариев. Если количество слов будет превышать высоту матрицы, то остаток не будет учитываться при классификации, размер матрицы предложения 26*200.

Высота матрицы (макс. кол-во слов), создание и обучение токенизатора, отображение каждого текста в массив идентификаторов токенов.

Описание матрицы:

- высота матрицы (максимальное количество слов в твите) = 26;
- размер словаря = 100000;
- получение последовательности;
- создаем и обучаем токенизатор;
- отображаем каждый текст в массив идентификаторов токенов.

Для построения нейросети использовалась библиотека Keras, интеграционный слой был инициализирован весами, полученными при обучении Word2Vec. Чтобы свести к минимуму изменения в интеграционном слое, нужно его остановить на 1-ом этапе обучения.

В ходе работы мы применяли фильтры, имеющие высоту 2, 3, 4, 5. Они требуются для осуществления параллельной обработки по биграммам, триграммам, 4-граммам и 5-граммам. К нейросети осуществим добавление по 10 сверточных слоев по каждой из высот фильтров. При этом применяем функцию ReLU. Когда обработка слоями свертки завершена, происходит движение карт признаков к слоям субдискретизации. Относительно них использовалась операция 1-max-pooling. При этом происходил процесс извлечения наиболее значимых n-грамм из текстов. После этого векторы будут объединены в общий вектор. Это соответствует слою объединения, подаваемому к скрытому полносвязному слою, имеющему 30 нейронов. После этого осуществлялся процесс подачи итоговой карты признаков к выходному слою нейросети, соответствующему сигмоидальной функции активации. Поскольку нейросети могут быть переобучены, проведено добавление dropout-регуляризации, имеющей вероятность выброса вершины $p=0.2$ после интеграционного слоя и до скрытого полносвязного слоя.

Осуществляем добавление dropout-регуляризации, слоя свертки, слоя субдискретизации. Проводим конкатенацию карт признаков и осуществляем добавление dropout-регуляризации.

Работа по dropout-регуляризации:

- добавляем dropout-регуляризацию: (*Dropout* (0,2), кодировщик твитов);
- для размера фильтры подсчитываются в [(2, 10), (3, 10), (4, 10), (5, 10)]:
 - добавляем слой свертки;
 - добавляем слой субдискретизации;
- конкатенируем карты признаков;
- добавляем dropout-регуляризацию: (исключение (0,2); плотность (30, активация = 'relu'); плотность (1)).

Проведено конфигурирование итоговой модели с функцией оптимизации Adam, а также бинарной кросс-энтропией в качестве функции ошибок. В классификаторе характеристики качества работы следующие: критерии точности, полноты и F-меры.

Характеристики качества:

- конфигурирование итоговой модели;
- качество работы классификатора оценивается в критериях точности, полноты и f-меры.

На 1-ом этапе обучения остановлен интеграционный слой. Проведено обучение всех остальных слоев в течение 10 эпох. По размеру групп примеров при обучении имеем: 32, по размеру валидационной выборки имеем: 25 %:

Характеристики обучения:

- "контрольно-пропускной пункт": контрольная точка модели (модель. hdf5, монитор, истина при только лучшем сохранении, максимальный режим, 1-й период);
- история: соответствие модели (обучение, размер пакета = 32, эпохи = 10, разделение проверки = 0,25, обратные вызовы = контрольная точка).

Далее рассматриваем модель с лучшими показателями F-меры на валидационном наборе данных, т. е. полученную на 8-ой эпохе обучения. У модели включим интеграционный слой, затем запустим еще 5 эпох обучения.

Загружаем веса модели, делаем интеграционный слой способным к обучению и уменьшаем скорость обучения.

Работа с параметрами модели:

- загружаем веса модели: (. hdf5);
- делаем embedding слой способным к обучению: (истина);
- уменьшаем скорость обучения;
- "контрольно-пропускной пункт": контрольная точка модели (модель. hdf5, монитор, истина при только лучшем сохранении, максимальный режим, 1-й период);
- история обучения: соответствие модели (обучение, размер пакета = 32, эпохи = 5, разделение проверки = 0,25, обратные вызовы = контрольная точка).

Таким образом, лучший результат был достигнут на 3-й эпохе обучения. Результаты приведены ниже в Таблице 2.

Анализ тональности текста с помощью рекуррентной нейронной сети

Рассмотрим, предложенный алгоритм анализа тональности текста с помощью рекуррентной нейронной сетей (Рисунок 2).

Аналогично, как и ранее, предварительно обработаем «Выгрузку обращений и отзывов» за определенный период по сервису, для дальнейшего проведения анализа на предмет тональности.

СТРУКТУРА АНАЛИЗА ТОНАЛЬНОСТИ ТЕКСТА С ПОМОЩЬЮ РЕКУРРЕНТНОЙ НЕЙРОННОЙ СЕТИ

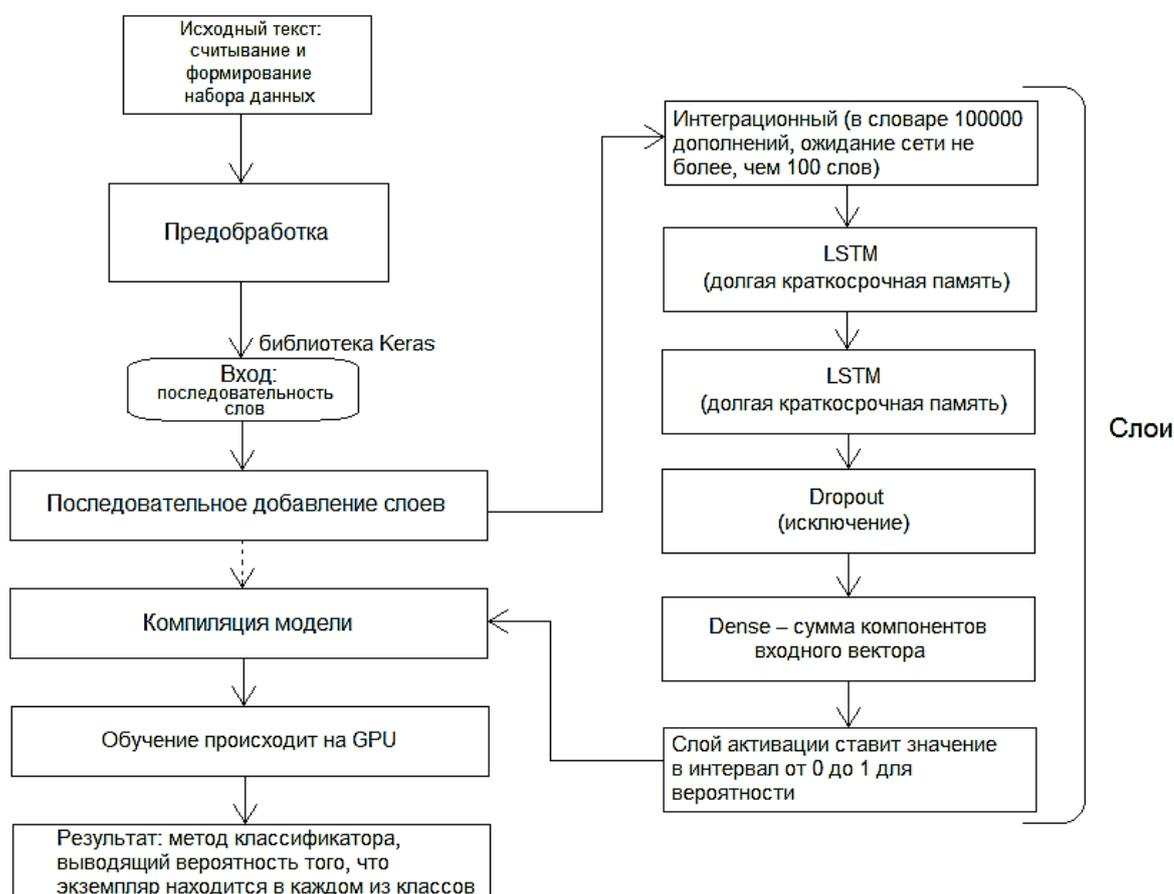


Рисунок 2 – Структура анализа тональности текста с помощью рекуррентной нейронной сети
Figure 2 – Structure of text sentiment analysis using a recurrent neural network

Для построения нейросети использовалась библиотека Keras. Так как на входе принимается последовательность слов, поэтому тип модели: Sequential. Далее, последовательно добавляются слои, компилируется модель, проводится обучение и получается результат.

Работа модели:

- максимальное количество функций = 100000;
- макс. длина = 100;
- размер партии = 32;
- последовательная модель:
 - добавление модели (встраивание (максимальное количество функций, 128, длина ввода = максимальная длина));
 - добавление модели (Долгая краткосрочная память (64, возвращаемые последовательности = истина));
 - добавление модели (Долгая краткосрочная память (64));
 - добавление модели (Исключение (0,5));
 - добавление модели (Плотность (1));

- добавление модели (Активация (Сигмоидальная функция активации));
- компиляция модели (потеря = бинарная кросс-энтропия, оптимизатор = Adam, режим класса = "двоичный");
- соответствие модели (обучение, размер партии = размер партии, эпоха = 1, скрыть точность = ложь;
- результат = модель прогнозирует вероятность.

Сеть имеет 6 слоев: интеграционный (в словаре 100000 дополнений, ожидание сети не более, чем 100 слов); 2 слоя LSTM (на выход batch_size / length of a sequence / units in LSTM и batch_size / units in LSTM, флаг return_sequences=True); Dropout; Dense – сумма компонентов входного вектора; слой активации ставит значение в интервал от 0 до 1 для вероятности.

Обучение происходит на GPU. Обучение осуществлялось всего за 1 эпоху, поскольку далее было зафиксировано переобучение. Результаты также приведены ниже в таблице (Таблица 2).

РЕЗУЛЬТАТЫ

Таблица 2 – Результаты, полученные при решении задачи анализа тональности текста с помощью сверточной и рекуррентной нейронных сетей

Table 2 – Results obtained in solving the problem of analyzing the sentiment of a text using convolutional and recurrent neural networks

№ п/п	Нейронная сеть	Эпоха обучения	Качество классификации: точность (accuracy)		
			Негативная тональность	Позитивная тональность	Нейтральная тональность
1	Сверточная нейронная сеть (CNN)	3 (лучший результат)	0.82195	0.82090	0.82143
2	Сверточная нейронная сеть (CNN)	1 (при сравнении с рекуррентной нейронной сетью)	0.80178	0.80073	0.80126
3	Рекуррентная нейронная сеть (RNN)	1 (лучший результат, так как далее фиксируется переобучение)	0.79180	0.79075	0.79128

Для сверточной нейронной сети по 1-му этапу обучения была остановка для интеграционного слоя. Обучение остальных слоев велось в течение 10 эпох. После этого, у модели, имеющей лучшие показатели F-меры при валидационном наборе данных, созданной для 8-ой эпохи обучения, был включен интеграционный слой, и были запущены еще 5 эпох обучения. Загрузили веса модели, сделали интеграционный слой способным к обучению и уменьшили скорость обучения. Таким образом, лучший результат был достигнут на 3-й эпохе обучения.

Для рекуррентной нейронной сети процесс обучения осуществлялся всего за 1 эпоху, поскольку далее было зафиксировано переобучение.

Сравнив результаты в ходе применения сверточной нейронной сети с результатами, полученными в ходе применения рекуррентной нейронной сети, можно сделать вывод, сверточная нейронная сеть при решении задачи анализа тональности текста при проведении анализа обращений по сервису на предмет качества (выявлении мнений пользователей о сервисе) более целесообразна с точки зрения применения. Точность негативной тональности составила 0.82195, позитивной тональности 0.82090, нейтральной тональности 0.82143; при использовании рекуррентных нейронных сетей точность негативной тональности составила 0.79180, позитивной тональности 0.79075, нейтральной тональности 0.79128.

ЗАКЛЮЧЕНИЕ

Проведен анализ и выявлен факт частого использования анализа тональности текста в большинстве систем анализа текстовой информации и на практике. Рассмотрены алгоритмы анализа тональности текста с помощью сверточной и рекуррентной нейронных сетей. Проведена экспериментальная оценка эффективности для каждой нейронной сети на примере естественного текста на русском языке. Произведен сравнительный анализ результатов в ходе применения сверточной нейронной сети с результатами, которые были получены с привлечением рекуррентной нейронной сети. Было продемонстрировано, что на практике представляет интерес использование сверточной нейронной сети, поскольку она обладает лучшими характеристиками по точности.

СПИСОК ИСТОЧНИКОВ

1. Рубцова Ю. Автоматическое построение и анализ корпуса коротких текстов (постов микроблогов) для задачи разработки и тренировки тонового классификатора. *Инженерия знаний и технологии семантического веба*. 2012;1(2):109-116.
2. Семина Т.А. Анализ тональности текста: современные подходы и существующие проблемы. *Прикладное языкознание, перевод, переводоведение*. 2020;2(2):47-64.
3. Сикорский О.С. Обзор свёрточных нейронных сетей для задачи классификации изображений. *Новые информационные технологии в автоматизированных системах*. 2017;17(4):17-26.
4. Рубцова Ю.В. Построение корпуса текстов для настройки тонового классификатора. *Программные продукты и системы*. 2015;1(109):72-78.
5. Бендерская Е.Н., Никитин К.В. Рекуррентная нейронная сеть как динамическая система и подходы к ее обучению. *Информатика, телекоммуникации и управление*. 2013;3(3):29-40.
6. Немальцев А.С. Использование рекуррентных нейронных сетей для анализа необработанного многоязычного текста. *Международный журнал гуманитарных и естественных наук*. 2020;3(2):55-59.
7. Cliche M. Twitter Sentiment Analysis with CNNs and LSTMs. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017;11(24):573-580.
8. Mikolov T. Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems*. 2013;5(11):3111-3119.
9. Rosenthal S., Farra N., Nakov P. Sentiment Analysis in Twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017;11(24):502-518.
10. Zhang Y., Wallace B. Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *arXiv preprint arXiv: 1510.03820*. 2015.

REFERENCES

1. Rubtsova Yu. Automatic construction and analysis of the corpus of short texts (microblogging posts) for the task of developing and training a tone classifier. *Knowledge Engineering and Semantic Web Technologies*. 2012;1(2):109-116.
2. Semina T.A. Sentiment analysis of the text: modern approaches and existing problems. *Applied linguistics, translation, translation studies*. 2020;2(2):47-64.
3. Sikorsky O.S. An overview of convolutional neural networks for the image classification problem. *New information technologies in automated systems*. 2017;17(4):17-26.
4. Rubtsova Yu.V. Building a text corpus to customize a tone classifier. *Software products and systems*. 2015;1(109):72-78.
5. Benderskaya E.N., Nikitin K.V. Recurrent neural network as a dynamic system and approaches to its training. *Computer science, telecommunications and management*. 2013;3(3):29-40.
6. Nemaltsev A.S. Using recurrent neural networks to analyze raw multilingual text. *International Journal of Humanities and Natural Sciences*. 2020;3(2):55-59.
7. Cliche M. Twitter Sentiment Analysis with CNNs and LSTMs. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017;11(24):573-580.
8. Mikolov T. Distributed Representations of Words and Phrases and Their Compositionality. *Advances in Neural Information Processing Systems*. 2013;5(11):3111-3119.
9. Rosenthal S., Farra N., Nakov P. Sentiment Analysis on Twitter. *Proceedings of the 11th International Workshop on Semantic Evaluation*. 2017;11(24):502-518.
10. Zhang Y., Wallace B. Sensitivity Analysis of (and Practitioners' Guide to) Convolutional Neural Networks for Sentence Classification. *arXiv preprint arXiv: 1510.03820*. 2015.

ИНФОРМАЦИЯ ОБ АВТОРАХ / INFORMATION ABOUT THE AUTHORS

Меняйлов Дмитрий Владимирович, аспирант, Воронежский институт высоких технологий, Воронеж, Российская Федерация.
e-mail: dmitriy.menyaylov111@yandex.ru

Dmitriy V. Menyaylov, postgraduate student, Voronezh Institute of High Technologies, Voronezh, Russian Federation.

Преображенский Андрей Петрович, доктор технических наук, профессор, Воронежский институт высоких технологий, Воронеж, Российская Федерация.
e-mail: app@vvt.ru

Andrey P. Preobrazhenskiy, Doctor Of Technical Science, Professor, Voronezh Institute of High Technologies, Voronezh, Russian Federation.

Статья поступила в редакцию 21.08.2021; одобрена после рецензирования 10.10.2021; принята к публикации 28.10.2021.

The article was submitted 21.08.2021; approved after reviewing 10.10.2021; accepted for publication 28.10.2021.