МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет компьютерных технологий и прикладной математики

Кафедра вычислительных технологий

**ОТЧЕТ О ПРОХОЖДЕНИИ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ (НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА)**

по направлению подготовки

02.03.02 Фундаментальная информатика и информационные технологии

период с 15.04.2022 г. по 28.04.2022 г

Выполнил студент \_\_46\_\_ гр. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_Хван К.Л.\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись) (Ф.И.О. студента)*

Руководитель практики (**производственная практика (научно-исследовательская работа)**)

д.т.н, профессор Вишняков Ю. М.

*(ученое звание, должность) (подпись) (Ф.И.О)*

Оценка по итогам защиты практики: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

«\_\_\_28\_\_» \_\_\_\_апреля\_\_\_\_\_\_ 2022 г.

Краснодар

2022 г.

ФГБОУ ВО «КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

Факультет компьютерных технологий и прикладной математики

Кафедра вычислительных технологий

**ИНДИВИДУАЛЬНОЕ ЗАДАНИЕ, ВЫПОЛНЯЕМОЕ В ПЕРИОД**

**ПРОВЕДЕНИЯ   
ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ**

**(НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА)**

Студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Хван Константин Леонидович\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(фамилия, имя, отчество полностью)*

Направление подготовки 02.03.02 Фундаментальная информатика и информационные технологии

Место прохождения практики \_\_\_ ФГБОУ ВО «Кубанский государственный университет»\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Срок прохождения практики с 15.04.2022 г. по 28.04.2022 г

Цель практики – формирование навыков самостоятельного осуществления научно-исследовательской работы, направленной на решение профессиональных задач; развитие профессиональных знаний в области прикладной математики и информатики, закрепление полученных теоретических знаний по дисциплинам направления и специальным дисциплинам программы бакалавриата Фундаментальная информатика информационные технологии, овладение необходимыми профессиональными компетенциями по избранному направлению подготовки.

Формирование компетенций, регламентируемых ФГОС ВО:

| **Код компетенции** | **Содержание компетенции (или её части)** | **Планируемые результаты при прохождении практики** |
| --- | --- | --- |
| ПК-1 | способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям |  |
| ПК-2 | способностью понимать, совершенствовать и применять современный математический аппарат, фундаментальные концепции и системные методологии, международные и профессиональные стандарты в области информационных технологий |  |
| ПК-3 | способностью использовать современные инструментальные и вычислительные средства |  |
| ПК-4 | способностью решать задачи профессиональной деятельности в составе научно-исследовательского и производственного коллектива |  |
| ПК-5 | способностью критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности |  |

Перечень вопросов (заданий, поручений) для прохождения практики

1. Изучение документации библиотеки глубокого обучения Keras и инструментов обработки текстовых данных.
2. Изучение способов сбора, предобработки и преобразования исходных данных для дальнейшего их использования в обучении.
3. Сбор, предобработка и преобразование данных.
4. Изучение разнообразных архитектур нейронных сетей и их обучение.
5. Проектирование архитектур нейронных сетей.

**План-график выполнения работ**:

|  |  |  |  |
| --- | --- | --- | --- |
| № | Этапы работы (виды деятельности) при прохождении практики | Сроки | Отметка руководителя практики о выполнении (подпись) |
| 1 | Инструктаж по технике безопасности, охраны труда, пожарной безопасности, а также правилами внутреннего распорядка обучающихся.  Выбор и обоснование темы исследования | 15.04.2022 |  |
| 2 | Анализ актуальности темы исследований в рамках работы над ВКР и анализ предметной области исследования | 16.04.2022-18.04.2022 |  |
| 3 | Обзор литературы и других источников | 19.04.2022-20.04.2022 |  |
| 4 | Анализ полноты требований к разрабатываемому в рамках ВКР программного приложения | 21.04.2022-22.04.2022 |  |
| 5 | Тестирование программного приложения, выполнение численных экспериментов и анализ результатов. | 23.04.2022-27.04.2022 |  |
| 6 | Защита отчета | 28.04.2022 |  |

Ознакомлен \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*(подпись студента) ( расшифровка подписи)*

«15» апреля 2022 г.

Руководитель производственной практики (научно-исследовательская работа)

заведующий кафедры вычислительных технологий

факультета компьютерных технологий

и прикладной математики, д.т.н. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Вишняков Ю.М.

*(подпись) (Ф.И.О. руководителя)*

\*Инструктаж по ознакомлению с требованиями охраны труда, техники безопасности, а также правилами внутреннего трудового распорядка.

**ОЦЕНОЧНЫЙ ЛИСТ**

**результатов прохождения   
ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ**

**(НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА)**

**по направлению подготовки**

**02.03.02 Фундаментальная информатика и информационные технологии**

Фамилия И.О студента \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Курс \_\_\_\_\_\_\_\_\_\_

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | ОБЩАЯ ОЦЕНКА  (отмечается руководителем практики) | Оценка | | | |
| 5 | 4 | 3 | 2 |
|  | Уровень подготовленности студента к прохождению практики |  |  |  |  |
|  | Умение правильно определять и эффективно решать основные задачи |  |  |  |  |
|  | Степень самостоятельности при выполнении задания по практике |  |  |  |  |
|  | Оценка трудовой дисциплины |  |  |  |  |
|  | Соответствие программе практики работ, выполняемых студентом в ходе прохождении практики |  |  |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | СФОРМИРОВАННЫЕ В РЕЗУЛЬТАТЕ УЧЕБНОЙ ПРАКТИКИ КОМПЕТЕНЦИИ  (отмечается руководителем практики от университета) | Оценка | | | |
| 5 | 4 | 3 | 2 |
|  | ПК–1: способностью собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по соответствующим научным исследованиям |  |  |  |  |
|  | ПК–2: способностью понимать, совершенствовать и применять современный математический аппарат, фундаментальные концепции и системные методологии, международные и профессиональные стандарты в области информационных технологий |  |  |  |  |
|  | ПК–3: способностью использовать современные инструментальные и вычислительные средства |  |  |  |  |
|  | ПК–4: способностью решать задачи профессиональной деятельности в составе научно-исследовательского и производственного коллектива |  |  |  |  |
|  | ПК–5: способностью критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности |  |  |  |  |

Руководитель практики \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_Вишняков Ю.М.\_\_\_\_\_\_\_\_\_\_

*(подпись) (расшифровка подписи)*

Сведения о прохождении инструктажа по ознакомлению с требованиями охраны труда, технике безопасности, пожарной безопасности, а также правилами внутреннего трудового распорядка

Предприятие Федеральное государственное бюджетное образовательное учреждение высшего образования «Кубанский государственный университет»

Факультет компьютерных технологий и прикладной математики

Кафедра вычислительных технологий

|  |  |
| --- | --- |
| Студент | Хван К.Л., 21 год |
|  | (ФИО, возраст) |

Дата 15 апреля 2022 г.

**1.     Инструктаж по требованиям охраны труда**

|  |  |
| --- | --- |
| Провел | профессор Вишняков Ю.М. |
|  | (должность, ФИО сотрудника, проводившего инструктаж, подпись) |

|  |  |
| --- | --- |
| Прослушал | Хван К.Л. |
|  | (ФИО, подпись студента) |

**2.     Инструктаж по технике безопасности**

|  |  |
| --- | --- |
| Провел | профессор Вишняков Ю.М |
|  | (должность, ФИО сотрудника, проводившего инструктаж, подпись) |

|  |  |
| --- | --- |
| Прослушал | Хван К.Л. |
|  | (ФИО, подпись студента) |

**3.     Инструктаж по пожарной безопасности**

|  |  |
| --- | --- |
| Провел | профессор Вишняков Ю.М |
|  | (должность, ФИО сотрудника, проводившего инструктаж, подпись) |

|  |  |
| --- | --- |
| Прослушал | Хван К.Л. |
|  | (ФИО, подпись студента) |

**4. Инструктаж по правилам внутреннего трудового распорядка**

|  |  |
| --- | --- |
| Провел | профессор Вишняков Ю.М |
|  | (должность, ФИО сотрудника, проводившего инструктаж, подпись) |

|  |  |
| --- | --- |
| Прослушал | Хван К.Л. |
|  | (ФИО, подпись студента) |

**Отзыв руководителя**

**О ПРОХОЖДЕНИИ ПРОИЗВОДСТВЕННОЙ ПРАКТИКИ**

**(НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА)**

|  |  |
| --- | --- |
| Студент |  |
|  | *(фамилия, имя, отчество полностью)* |

Направление подготовки 02.03.02. Фундаментальная информатика и информационные технологии

За время прохождения производственной практики (НАУЧНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА) студент \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_............ Студентом были решены следующие задачи:

1

2

3

Все поставленные задачи выполнены, степень проработанности материала считаю достаточной.

Результаты практики оцениваю на оценку «\_\_\_\_\_\_\_\_\_».

|  |  |  |
| --- | --- | --- |
| Руководитель студента |  | Вишняков Ю.М. |
|  | *(подпись)* | *(расшифровка подписи)* |

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 9](#_Toc102398090)

[1 Способы предобработки текстовых данных 10](#_Toc102398091)

[2 Способы преобразования текстовых данных 12](#_Toc102398092)

[2.1 Прямое кодирование слов 12](#_Toc102398093)

[2.2 Векторное представление слов 12](#_Toc102398094)

[2.3 Word2Vec 14](#_Toc102398095)

[3 Архитектуры нейронных сетей 17](#_Toc102398096)

[3.1 Полносвязанная сеть 17](#_Toc102398097)

[3.2 Сверточные сети 20](#_Toc102398098)

[3.3 LSTM 22](#_Toc102398099)

[3.4 GRU 23](#_Toc102398100)

[4 Применение 24](#_Toc102398101)

[ЗАКЛЮЧЕНИЕ 25](#_Toc102398102)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 26](#_Toc102398103)

# ВВЕДЕНИЕ

Цель практики: формирование навыков самостоятельного осуществления научно-исследовательской работы, направленной на решение профессиональных задач; развитие профессиональных знаний в области информатики, закрепление полученных теоретических знаний по дисциплинам направления и специальным дисциплинам программы бакалавриата «Фундаментальная информатика информационные технологии», овладение необходимыми профессиональными компетенциями по избранному направлению подготовки.

Место прохождения практики: Россия, Краснодарский край, г. Краснодар, ФГБОУ ВО «Кубанский государственный университет».

Дата начала: 15 апреля 2021.

Продолжительность практики: 2 недели.

Перечень основных работ: исследование способов работы с текстовыми данными и построения различных архитектур нейронных сетей для задачи классификации.

# 1 Способы предобработки текстовых данных

Текст – это один их самых распространенных видов информации. Его можно интерпретировать как последовательности символов, либо же как последовательности слов. В нашей случае текст будет использовать на уровне слов.

Модели глубокого обучения не понимают текст, как это делает человек. Они лишь показывают статистическую структуру письменного языка, что вполне достаточно для многих простых задач обработки текста.

Поэтому перед началом обучения модели необходимо выполнить начальную предобработку и преобразование текстовых данных в так называемые тензоры (численные векторы представления слов).

Под начальной предобработкой текста понимают выполнение следующих шагов:

1. лексеммизация: весь текст разбивается на список отдельных элементов (слов (токенов), n-грамм);
2. преобразование в нижний регистр: слово, начинающееся с заглавной буквы (например, «Собака»), и то же слово, находящееся в середине (например, «собака»), не имеют между собой смысловых различий. Таким образом, преобразовывая все слова в нижний регистр, игнорируется выделение на заглавные и строчные буквы;
3. удаление стоп-слов: стоп-слова – это слова, которые очень часто встречаются в тексте, но не несут большого смыслового значения, например «и», «а», «которые», «нет», «не может», «не» и т. д. В некоторых случаях, например в задачах классификации по тональности, рекомендуется делать исключения и не удалять стоп-слова, делающие высказывание отрицательным;
4. удаление знаков препинания и цифр: практически всегда знаки препинания и цифры не вносят точности в модели естественного языка, поэтому их и удаляют;
5. стемминг: стемминг – это усечение слов до их основы. Он позволяет объединить слова с одинаковыми значениями в одну лексемму, что способствует выявлению большего количества примеров контекстов для этих слов. Обычно стемминг используется только для небольших корпусов, для больших его использовать не имеет смысла;
6. обработка n-грамм: n-граммы представляют собой n подряд идущих слов. В естественном языке часто встречаются n-граммы, например биграммы (). Поэтому их лучше учитывать как единое целое, а не каждый по отдельности.
7. лемматизация: использование стемминга иногда может приводить к потери изначального смысла некоторых слов, вследствие чего может нарушиться целостная картина восприятия этого предложения в корпусе. Например, при стемминге слово «генеральный» и «генеральский» преобразуются сразу в слово «генерал», хотя значения двух этих слов разнове. Поэтому лучшим способом избежать этого является использование лемматизации, т. е. приведение к начальной форме слова. В русском языке нормальными формами считаются следующие морфологические формы:
   1. для существительных – именительный падеж, единственное число;
   2. для прилагательных – именительный падеж, единственное число, мужской род;
   3. для глаголов, причастий, деепричастий – глагол в инфинитиве (неопределённой форме) несовершенного вида.

# 2 Способы преобразования текстовых данных

## 2.1 Прямое кодирование слов

Прямое кодирование слов (one-hot encoding) – это один из самых простых методов преобразования слов (токенов) в численные векторы (тензоры). Этот метод заключается в том, что каждому слову присваивается уникальный целочисленный идентификатор , который преобразуется в бинарный вектор размера (размер словаря всех слов). Все элементы этого вектора являются нулями, кроме  элемента, который является единицей.

Пример кодирования предложения «What is the answer to everything?»:

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 1 – Представление слов в виде векторов (one-hot encoding)

Очевидный минусом такого представления является большое количество памяти необходимое для такого кодирования большого объема данных и разреженность полученных векторов.

## 2.2 Векторное представление слов

Другим популярным и мощным способом связывания вектора со словом является использование плотных векторов слов, или векторного представления слов (word embeddings). В отличие от векторов, полученных прямым кодированием, — бинарных, разреженных (почти полностью состоящих из нулей) и с большой размерностью (их размерность совпадает с количеством слов в словаре) — векторные представления слов являются малоразмерными векторами вещественных чисел (то есть плотными (dense) векторами, в противоположность разреженным), как показано на рисунке 2.

В отличие от векторов, полученных прямым кодированием, векторные представления слов конструируются из данных. При работе с огромными словарями размерность векторов слов нередко может достигать 256, 512 или 1024. С другой стороны, прямое кодирование слов обычно влечет за собой создание векторов с числом измерений 20 000 или больше (при использовании словаря с 20 000 токенов). Иначе говоря, векторное представление слов позволяет уместить больший объем информации в меньшее число измерений.

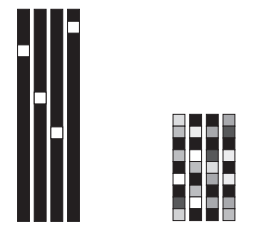


Рисунок 2 – векторы (one-hot encoding) и векторные представления

Получить векторные представления можно двумя способами:

1. создавать собственные векторные представления на основе собственных данных при решении основной задачи. В этом случае изначально создаются случайные векторы слов, которые потом обучаются по аналогии весов в нейронных сетях;
2. загрузить уже векторные представления, полученные при решении других задач, отличных от решаемой. Такие векторные представления называют предварительно обученными векторными представлениями.

В своей задаче же планируется создавать свое векторное представление слов на основе собственного корпуса с использованием алгоритма Word2Vec, GloVe и сравнить с уже предобученным.

# 2.3 Word2Vec

Word2Vec – это алгоритм обучения без учителя, не использующих никаких меток, которые могут или не могут быть в корпусе. Таким образом, на вход этому алгоритму можно подавать любой набор текстовых данных на естественном языке. Это является огромный плюсом, потому что для разных предметных областей мы имеем возможность создавать собственные векторные представлеия.

Алгоритм имеет две разные базовые архитектуры обучения – *skip-gram* (SG) или *continiuos bag of words* (CBOW).

Суть этого алгоритма можно объяснить на примере предложения «What is the answer to everything?»

Рассмотрим слово «answer» как *целевое,* а два слова слева и справа как *контекстные* (количество рассматриваемых слов, как контекстных, является гиперпараметром, который необходимо учитывать при запуске алгоритма). В архитектуре CBOW целевое слово предсказывается на основе контекстных слов. В архитектуре SG, наоборот, контекстные слова предсказываются на основе целевого. На рисунке 3 изображены эти отличия.

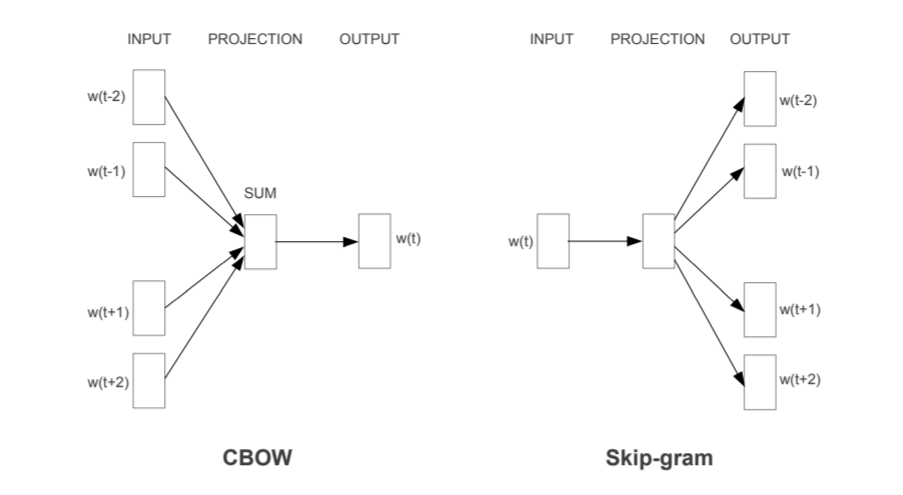


Рисунок 3 – обобщенная схема работы моделей с архитектурами CBOW и SG

Как было сказано раннее в архитектуре CBOW целевое слово предсказывается как среднее из окружающих его контекстов, причем не имеет значения в каком месте стоят контекстные слова относительно друг друга и целевого слова, т.е. здесь используется механизм BOW («bag of words» или «мешок слов»):

1. берутся все контекстные слова в окне, справа и слева от целевого;
2. добавляются эти слова в «мешок» (словарь);
3. вычисляется среднее значение для всех контекстных слов в мешке и на основании среднего оценивается, какое целевое слово должно быть.

Буква C, или «continuos» означает, что скользящее окно, включающее целевое и контекстные слова, непрерывно скользит по тексту, слово за словом, от первого слова в корпусе до последнего. В каждой позиции на этом пути целевое слово оценивается с учетом контекстных слов. С помощью стохастического градиентного спуска расположение слов в векторном пространстве может сдвигаться и соответственно, оценки целевых слов могут постепенно улучшаться.

Многие специалисты рекомендуют использовать модели с архитектурой SG для небольших корпусов. Но в этой работе будут использоваться обе модели и сравниваться между собой.

# 3 Архитектуры нейронных сетей

Выполнив предобработку данных и преобразовав их в тот вид, который поймет нейросеть, необходимо будет выбрать архитектуру нейронной сети для классификации жанров произведений. Всего планируется сравнить 4 архитектуры: обычную полносвязанную сеть, сверточную сеть, LSTM и GRU -сети.

# 3.1 Полносвязанная сеть

Выбор архитектуры сети является важным пунктом в решении задачи классификации. От того какая архитектура была выбрана, будет зависеть скорость, точность и качество обучения. Поэтому необходимо просмотреть несколько их них и выбрать наиболее подходящую.

Точкой же отсчета будет являться полносвязанная нейронная сеть, структура которой будет изображена на рисунке 4:

Изображение выглядит как стол

Автоматически созданное описание

Рисунок 4 – Структура полносвязанной нейронной сети

Это последовательная модель полносвязанной нейронной сети, которая состоит из семи слоев:

1. embedding – этот слой позволяет создавать векторы слов из корпуса документов, в нашем случае из подготовленных отрывков книг, образующих обучающую выборку. Так как этот слой является первым, то для него также встраивается входной слой, который принимает вектор слов определенной длины. Создать векторы слов можно и с помощью алгоритма word2vec, что в дальнейшем будет сделано;
2. flatten – слой, который преобразует 2D вектор в 1D;
3. Далее следуют полносвязанные слои dense и dense\_1 с функцией активации ReLu, к которым добавили слои прореживания dropout и dropout\_1, чтобы уменьшить шанс переобучения. Каждый из слоев состоит из 64 нейронов;
4. И наконец выходной слой dense\_2 с функцией активацией softmax и количеством нейронов равным 3.

Далее следует скомпилировать сеть. Параметры компиляции изображены на рисунке 5.



Рисунок 5 – Комплирование полносвязанной нейронной сети

В качестве функции стоимости берется categorical\_crossentropy, в качестве оптимизатора – ADAM, а в качестве метрики – точность.

# 3.2 Сверточные сети

Сверточные сети особенно хорошо подходят для распознавания пространственных шаблонов. Поэтому рассмотрим подробнее пример архитектуры сверточной сети на рисунке 6.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 6 – Архитектура сверточной нейронной сети

Более подробное описание модели необходимо ниже:

1. embedding – слой тот же, что и в предыдущей модели, отличие лишь в том, что здесь добавлено прореживание за счет слоя spatial\_droupout1d и увеличен объем входных данных до 400 элементов.
2. conv1d – одномерный сверточный слой с функцией активации ReLu. Также этот слой имеет 256 уникальных фильтров, каждый из которых может специализироваться на активациях при прохождении определенной последовательности из трех лексем. Карта активаций для каждого из 256 фильтров имеет длину 398 для выхода с формой 256 × 398.
3. Следом слой global\_max\_pooling1d, который уменьшает размерность карты активации с 256 × 398 до 256 × 1. В этом случае сохраняется порядок наибольшей активации для данного сверточного фильтра.
4. После предыдущего слоя на выходе получится одномерный вектор, который можно будет передать в полносвязанный слой с функцией активации ReLu и слоем dropout.
5. Выходной слой же состоит из 3 нейронов, как и в предыдущей модели.

Компиляция модели точно такая же, как и в модели с полносвязанной архитектурой.

# 3.3 LSTM

Архитектуру сети с долгой краткосрочной памятью проиллюстрирована на рисунке 7.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 7 – Архитектура LSTM-сети

В этой модели после слоя embedding вместо сверточного или полносвязанного слоя идет слой lstm, который и является рекуррентным слоем с долгой краткосрочной памятью. Практически всегда после рекуррентного слоя, в отличие от сверточного, не добавляется полносвязанный слой из-за того, что это не дает почти никаких преимуществ.

# 3.4 GRU

Архитектура сети с управляемыми блоками памяти изображена на рисунке 8.

Изображение выглядит как текст

Автоматически созданное описание

Рисунок 8 – Архитектура GRU-сети

Эта модель ничем не отличается от предыдущей, кроме того, что вместо слоя lstm присутствует слой gru.

# 4 Применение

Знания, полученные в данном исследовании, применяются в выпускной квалификационной работе для реализации автоматической жанровой классификации произведений.

Проект представляет собой сравнительный анализ разных методик классификации текстовых произведений с использованием различных архитектур и модель с лучшими показателями качества предсказаний.

Построенную модель жанровой классификации можно применять в:

1. Коммерческих организациях, занимающихся продажей книжных изделий.
2. В интернет-маркетинге.
3. В библиотеках для осуществления автоматической сортировки на жанры.

# ЗАКЛЮЧЕНИЕ

За время практики удалось научиться собирать, обрабатывать и интерпретировать данные современных научных исследований, необходимые для формирования выводов по исследованию классификации жанров произведений (ПК-1). Также удалось научиться понимать, совершенствовать и применять современный математический аппарат, фундаментальные концепции и системные методологии, международные и профессиональные стандарты в области информационных технологий (ПК-2). Для исследования библиотеки Keras были разработаны модели нейронной сети на современном языке программирования Python (ПК-3). Удалось решать задачи профессиональной деятельности в составе научно-исследовательского и производственного коллектива (ПК-4). А также критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности (ПК-5).

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Крон Джон, Бейлевельд Грант, Аглаэ Бассенс: Глубокое обучение в картинках. Визуальный гид по искусственному интеллекту. — СПб.: Питер, 2020. — 400 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-1574-7
2. Шолле Франсуа: Глубокое обучение на Python. — СПб.: Питер, 2018. — 400 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-4461-0770-4
3. Бенгфорт Бенджамин, Билбро Ребекка, Охеда Тони: Прикладной анализ текстовых данных на Python. Машинное обучение и создание приложений обработки естественного языка. — СПб.: Питер, 2019. — 368 с.: ил. — (Серия «Бестселлеры O’Reilly»). ISBN 978-5-4461-1153-4
4. Документация пакета Keras [Электронный ресурс] URL: https://keras.io/ (дата ображения: 15.04.2022)
5. Mohammad S Sorower: A Literature Survey on Algorithms for Multi-label Learning
6. Tomas Mikolov, Kai Chen, Greg Corrado, Jeffrey Dean: Efficient Estimation of Word Representations in Vector Space.
7. Word2Vec [Электронный ресурс] URL: https://www.tensorflow.org/tutorials/text/word2vec (дата обращения 25.04.2022)