МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

**«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г. ЧЕРНЫШЕВСКОГО»**

Кафедра теоретических основ компьютерной безопасности и криптографии

**Нейронные сети для обработки текста на естественном языке**

РЕФЕРАТ

студента 5 курса 531 группы

специальности 10.05.01 Компьютерная безопасность

факультета компьютерных наук и информационных технологий

Латанова Кирилла Вячеславовича

|  |  |  |  |
| --- | --- | --- | --- |
| Проверил:  доцент | \_\_\_\_\_\_\_\_\_\_ | И. И. Слеповичев | |
|  | подпись, дата | |  |
|  |  | |  |

Саратов 2022

СОДЕРЖАНИЕ

[ВВЕДЕНИЕ 3](#_Toc123428367)

[1 Постановка задачи 4](#_Toc123428368)

[2 Возможность реализации 5](#_Toc123428369)

[3 История развития нейронных сетей 6](#_Toc123428370)

[4 Архитектуры нейронных сетей для решения задач NLP 7](#_Toc123428371)

[4.1 Многослойный перцептрон 7](#_Toc123428372)

[4.2 Сверточная нейронная сеть 7](#_Toc123428373)

[4.3 Рекурсивная нейронная сеть 8](#_Toc123428374)

[4.4 Рекуррентная нейронная сеть 8](#_Toc123428375)

[4.5 Сеть долгой краткосрочной памяти 8](#_Toc123428376)

[4.6 Sequence-to-sequence 9](#_Toc123428377)

[4.7 Неглубокие нейронные сети 9](#_Toc123428378)

[ЗАКЛЮЧЕНИЕ 11](#_Toc123428379)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 12](#_Toc123428380)

# ВВЕДЕНИЕ

По мере роста объемов информации, с которым приходится взаимодействовать каждый день, возникает потребность в её быстрой обработке за разумное время, что приводит к большому спросу на специалистов в области больших данных, машинного обучения и нейронных сетей. Данные направления имеют широкий диапазон применения - от подсчета различных статистик и их анализа, до создания различных программ, позволяющих прогнозировать определенные события или совершать инновационные открытия.

В этой работе речь пойдет о том для чего может быть необходима обработка текста на естественном языке.

# Постановка задачи

Обработка естественного языка (англ. Natural Language Processing, NLP) – одно из направлений искусственного интеллекта, задачей которого являются компьютерный анализ и синтез естественных языков. Под этим понимается возможность для компьютера понимать и создавать осмысленные тексты на любых языках, с помощью которых общаются люди, например, китайский, английский или русский.

Примеры реализации данной задачи можно часто встретить в нашей повседневной жизни – лингвистический анализ, выделение важной информации из текста, поиск синонимичных запросов в Интернете, генерация субтитров, распознавание речи и голосовой ввод, автоматизированный перевод, генерация текста для различных целей (в том числе полноценных осмысленных и законченных произведений), генерация уникального текста на основе уже имеющегося, словообразование и многое другое [1, 2, 3].

Рассмотрим проблемы, которые могут возникнуть при реализации подобной задачи на примере генерации текста в русском языке. Как известно, современные компьютеры позволяют вычислять сложные выражения с большими числами за довольно короткие промежутки времени, однако в данном случае мы имеем дело не с цифрами. Значит, для начала необходимо придумать способ конвертации слов, букв, правил и всего, что связано с этим языком в биты. Более того, нам придется научить компьютер правилам данного языка, что далеко не всегда по силам некоторым людям, особенно, если этот язык для них не родной.

В случае русского языка подобные трудности связаны с обилием слов и его различных форм, поскольку, например, существительные можно склонять, а глаголы – спрягать. Кроме того каждое существительное /глагол имеет склонение/спряжение, которое далеко не всегда можно определить по правилам, а, значит, часть слов придется отдельно запоминать как исключения. При изменении формы у слов начинают меняться окончания в зависимости от части речи, склонения, падежа и числа, которые в свою очередь также необходимо уметь правильно определять или запоминать. При конструировании предложений порядок слов имеет значение, а некоторые слова и даже выражения могут иметь различный смысл в зависимости от контекста, что автоматически обязывает нас однозначно решать эту задачу.

Таким образом, мы имеем широкий спектр подзадач, которые нужно правильно классифицировать, обработать и только потом начать их вычислять, что, в свою очередь, займет довольно долгое время, поскольку каждая подзадача может потребовать громоздких расчетов.

# Возможность реализации

Реализация данной задачи была выполнена по тому же принципу, по которому это делает человек – с помощью нейронных сетей. Разница состоит лишь в том, что у людей это происходит при помощи органов биологической системы, а у компьютера за счет технических и аппаратных средств.

Главным в этом процессе у человека является мозг, поскольку именно он выполняет функции центра управления. Мозг состоит из нейронов, которые в свою очередь через дендриты по аксону получают сигналы от других нейронов. Нейрон даже в случае неактивности создает электрический импульс, который в случае активности других нейронов увеличивает частоту. Связь в синапсе и дендрите может быть как положительной, так и отрицательной. Все это в совокупности позволяет моделировать работу нейрона пуассоновским процессом, интенсивность которого будет зависеть от входных параметров, а, значит, дает возможность создавать подобные модели искусственно, в том числе при помощи компьютера.

Мозг способен обучаться, что означает наличие некоторого алгоритма, по которому этот процесс происходит, а, значит, его также можно смоделировать, поняв суть. Для этого нейробиологи продолжительное время занимались исследованиями мозга, выявляя определенные закономерности, и классифицируя их, а также определяя некоторые физические характеристики мозга (рис. 1).

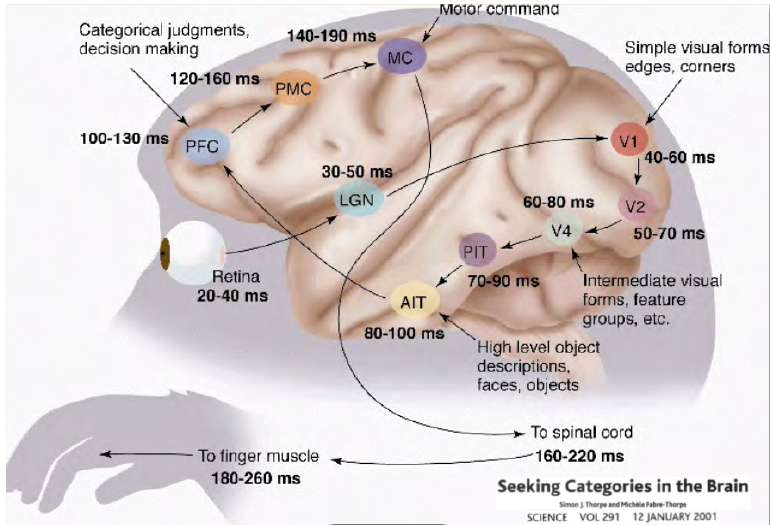


Рисунок 1 – Строение мозга человека

# История развития нейронных сетей

Опираясь на модель работы мозга человека, была разработана искусственная нейронная сеть. Сама идея такой возможности была впервые предложена Аланом Тьюрингом в 1948 году, после чего спустя несколько лет появился первый алгоритм обучения методом градиентного спуска. После этого стали появляться ламповые реализации перцептрона для распознавания букв, но на ранних стадиях серьезных результатов это не имело. В 1960-х годах ученые стали активно изучать перцептрон, что в 1969 году привело Мински и Паперта к осознанию невозможности модуляции операции битового сложения по модулю два, а это в свою очередь на некоторое время замедлило развитие нейронных сетей. Однако спустя несколько лет появилась резкая динамика в данной области:

* Во второй половине 1970-х годов была разработана современная
* В первой половине 1980-х начали создавать глубокие модели обучения
* В 1988 году была создана первая нейронная сеть для распознавания речи
* В 1990 году появились первые рекуррентные нейронные сети

После этого был достигнут вычислительный предел и обучение занимало несколько недель, но по мере развития технологий эта проблема перестала быть критичной, что позволило выйти на новый уровень.

# Архитектуры нейронных сетей для решения задач NLP

## Многослойный перцептрон

Многослойный перцептрон состоит из 3 и более слоев и использует нелинейную функцию активации (часто тангенциальную или логистическую), позволяющую классифицировать линейно неразделимые данные. Каждый отдельный узел в слое соединен с каждым узлом в последующем слое, что делает сеть полностью связанной. Такая архитектура применяется в задачах распознавания речи и машинном переводе (рис. 2).

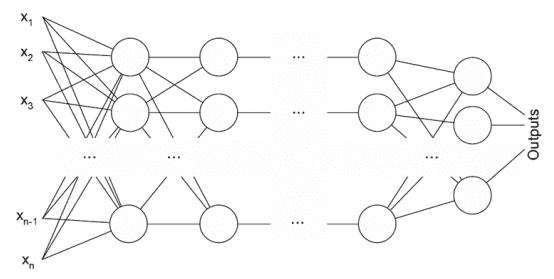


Рисунок 2 – Архитектура многослойного перцептрона

## Сверточная нейронная сеть

Сверточная нейронная сеть содержит несколько объединенных сверточных слоев и использует вариацию многослойного перцептрона. Сами слои используют операцию свертки на входных данных, а затем передают их в следующий слой, что позволяет сети становиться глубже с меньшим количеством параметров. Данная модель эффективно работает на задачах классификации текста [4].

## 4.3 Рекурсивная нейронная сеть

Данная модель сформирована при применении одних и тех же наборов весов рекурсивно над структурой с целью сделать структурированное предсказание над входной структурой переменного размера через активацию структуры в топологическом порядке.

## 4.4 Рекуррентная нейронная сеть

Эта нейронная сеть в основе связей между нейронами имеет направленные циклы, что означает зависимость выходной информации не только от входной, но и от состояния нейрона на предыдущем шаге. Это позволяет эффективно справляться с задачами распознавания рукописного текста или речи (рис. 3).

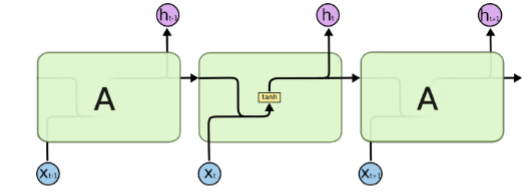


Рисунок 3 – Архитектура рекуррентной нейронной сети.

## 4.5 Сеть долгой краткосрочной памяти

Эта модель является разновидностью архитектуры рекуррентной нейросети, но при этом не использует функцию активации в рекуррентных компонентах, а сохраненные значения не модифицируются. Для данного подхода характерно использование в блоках по несколько элементов, состоящих из затворов, контролирующих построение потока информации на основе логистической функции (рис. 4).

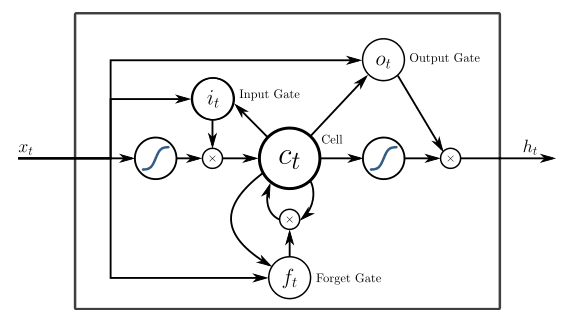


Рисунок 4 – Архитектура LSTM.

## 4.6 Sequence-to-sequence

Архитектура подразумевает наличие двух рекуррентных сетей – кодировщика, обрабатывающего входные данные, и декодера, осуществляющего вывод. Такой подход используется в чат-ботах и машинном переводе.

## 4.7 Неглубокие нейронные сети

Word2vec – один из самых известных подходов в области NLP, разработанный Google, суть которого состоит в преобразовании слов в вектор, который отражает семантические свойства слова. Таким образом, слова близкие по значению находятся рядом в рамках N-мерного пространства результирующего вектора. Пары слов со схожим отношением будут иметь схожее смещение в конечном векторе. В то же время эти векторы обладают свойством получения схожих векторов при трансформациях в рамках языка. Однако существенным недостатком данной модели является неэффективность при изменяющемся размере словаря, хранящего слова (рис. 5) [3].

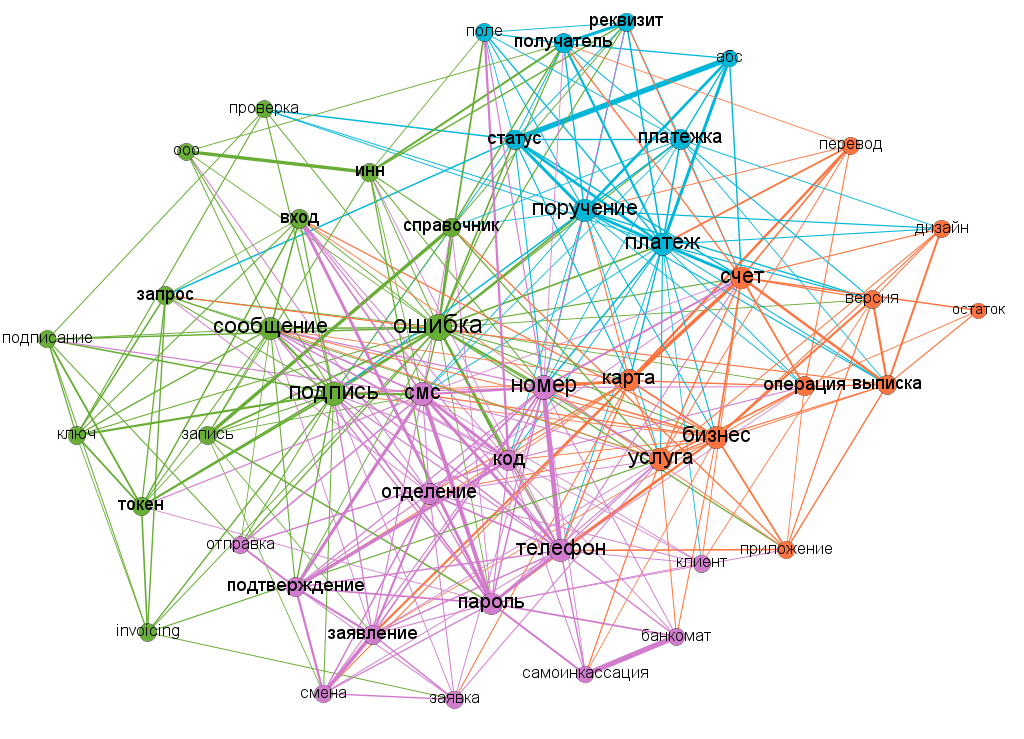


Рисунок 5 – Пример семантически близких слов в конечном поле

# ЗАКЛЮЧЕНИЕ

В данной работе были подробно рассмотрены основные архитектуры нейронных сетей, а также краткая история их развития. На основании этого и беря во внимание растущие объемы информации (в том числе и текстовой), которые необходимо ежедневно перерабатывать, очевидно, что последующее десятилетие будет активно совершенствоваться и прогрессировать разработка нейронных сетей в различных сферах деятельности.

# СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. 7 архитектур нейронных сетей для решения задач NLP. [Электронный ресурс] / URL: https://neurohive.io/ru/osnovy-data-science/7-arhitektur-nejronnyh-setej-nlp/ https://www.elibrary.ru/item.asp?id=41384033 (дата обращения: 24. 12. 22). – Яз. рус.
2. Применение сверточных нейросетей для задач NLP. [Электронный ресурс] / URL: https://habr.com/ru/company/ods/blog/353060/ (дата обращения: 24. 12. 22). – Яз. рус.
3. Богомолов, Ю.А Обзор моделей нейронных сетей для обработки естественного языка. [Текст] / Ю.А. Богомолов // Научно-образовательный журнал – Санкт-Петербург – 2020 – №4 – URL: https://cyberleninka.ru/article/n/obzor-modeley-neyronnyh-setey-dlya-obrabotki-estestvennogo-yazyka/viewer (дата обращения: 24. 12. 22). – Яз. рус.
4. Convolutional Neural Networks for Sentence Classification. [Электронный ресурс] / Yoon Kim // статья – 2014 – URL: https://aclanthology.org/D14-1181/ https://towardsdatascience.com/understanding-rmsprop-faster-neural- (дата обращения: 24. 12. 22). – Яз. англ.