МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра информационных технологий**

**ДОКЛАД**

**RNN В МАШИННОМ ПЕРЕВОДЕ: АРХИТЕКТУРЫ И МЕТОДЫ**

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_В.В.Несветайлов

Направление подготовки 02.03.03 «Математическое обеспечение и администрирование информационных систем» курс 3

Направленность Системное программирование и компьютерные технологии

Научный руководитель

канд. физ-мат. наук, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.К Курамагомедова

Краснодар

2025

СОДЕРЖАНИЕ

[СОДЕРЖАНИЕ 2](#_Toc196171149)

[ВВЕДЕНИЕ 3](#_Toc196171150)

[1 Основы RBF-сетей 4](#_Toc196171151)

[1.1 Архитектура и принцип работы RBF-сетей 4](#_Toc196171152)

[1.2 Обучение RBF-сетей 4](#_Toc196171153)

[2 Применение RBF-сетей в классификации 6](#_Toc196171154)

[2.1 Классификация многоклассовых данных 6](#_Toc196171155)

[2.2 Сравнение RBF с другими методами 7](#_Toc196171156)

[ЗАКЛЮЧЕНИЕ 8](#_Toc196171157)

[СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ 9](#_Toc196171158)

# ВВЕДЕНИЕ

Машинный перевод – одна из ключевых задач обработки естественного языка (NLP), которая заключается в автоматическом переводе текста с одного языка на другой. Традиционные методы, такие как статистический машинный перевод (SMT), уступили место нейронным подходам благодаря их способности учитывать контекст и семантику предложений.

Рекуррентные нейронные сети (RNN) сыграли важную роль в развитии нейронного машинного перевода (NMT). Они позволяют обрабатывать последовательности переменной длины, что делает их идеальными для работы с текстом. Однако у классических RNN есть ограничения, такие как проблема исчезающего градиента, что привело к разработке более сложных архитектур, таких как LSTM и GRU.

В данном докладе рассматриваются:

* Основные архитектуры RNN, применяемые в машинном переводе.
* Модель Seq2Seq и ее улучшение с помощью механизма внимания.
* Современные методы и тенденции в NMT.

1. ****Основы машинного перевода с использованием RNN****

## **Рекуррентные нейронные сети (RNN) и их архитектуры**

RNN – это класс нейронных сетей, предназначенных для обработки последовательных данных. Их ключевая особенность – наличие скрытого состояния (hidden state), которое сохраняет информацию о предыдущих элементах последовательности.

Основные свойства RNN:

Рекуррентные связи – информация передается от одного шага к другому.

Обработка последовательностей – подходят для текста, речи, временных рядов.

Проблемы:

* Короткая память (vanishing gradient problem).
* Сложность обучения на длинных последовательностях.
* Улучшенные архитектуры:

LSTM (Long Short-Term Memory) – содержит механизм "ворот" (forget, input, output gates), что позволяет лучше сохранять долгосрочные зависимости.

GRU (Gated Recurrent Unit) – упрощенная версия LSTM с двумя воротами, быстрее обучается.

## **Модель Seq2Seq**

Seq2Seq (Sequence-to-Sequence) — это архитектура нейронной сети, предназначенная для преобразования одной последовательности в другую. Она широко применяется в задачах обработки естественного языка (NLP), таких как машинный перевод, генерация текста, summarization, чат-боты и другие.

Основные компоненты Seq2Seq

1. Encoder (Кодировщик)

* Принимает входную последовательность (например, предложение на исходном языке).
* Преобразует её в вектор фиксированной длины — контекстный вектор (context vector), который кодирует смысл входных данных.

Обычно использует RNN (LSTM, GRU) или Transformer.

1. Decoder (Декодировщик)

* Получает контекстный вектор от encoder.
* Генерирует выходную последовательность (например, перевод на целевой язык) пошагово, учитывая скрытое состояние и предыдущие предсказания.

Может использовать механизм внимания (attention), чтобы фокусироваться на разных частях входной последовательности при генерации каждого элемента выхода.

Как работает Seq2Seq?

1. Входные данные:

Входная последовательность (например, слова или символы) подаётся в encoder.

Каждый элемент последовательности обрабатывается RNN/Transformer, и скрытое состояние обновляется.

1. Контекстный вектор:

Final hidden state encoder становится контекстным вектором — сжатым представлением всей входной последовательности.

1. Генерация выхода:

Decoder начинает генерацию, используя контекстный вектор как начальное скрытое состояние.

На каждом шаге decoder предсказывает следующий элемент (например, слово) и передаёт своё скрытое состояние на следующий шаг.

1. Механизм внимания (опционально):

Позволяет decoder "смотреть" на разные части входной последовательности при генерации каждого элемента выхода, улучшая качество модели для длинных последовательностей.

1. ****Улучшение моделей машинного перевода****

## **Механизм внимания (Attention Mechanism)**

Механизм внимания (Attention Mechanism) — это ключевая технология в современных нейронных сетях, особенно в задачах обработки последовательностей (NLP, компьютерное зрение и др.). Он позволяет модели динамически фокусироваться на наиболее важных частях входных данных при генерации каждого элемента выходной последовательности.

Зачем нужен Attention?

1. Проблема "бутылочного горлышка" в Seq2Seq:

В классической Seq2Seq (без внимания) encoder сжимает всю входную последовательность в один контекстный вектор, что приводит к потере информации, особенно для длинных текстов.

Attention решает это, давая decoder доступ ко всем скрытым состояниям encoder (а не только к последнему).

1. Улучшение интерпретируемости:

Attention-веса показывают, какие слова входной последовательности влияют на текущее предсказание (например, выравнивание слов при переводе).

## **Современные архитектуры на основе RNN (LSTM, GRU, Bidirectional RNN)**

Современные архитектуры на основе рекуррентных нейронных сетей (RNN) включают в себя несколько популярных вариантов, таких как LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit) и Bidirectional RNN. Давайте рассмотрим каждую из этих архитектур подробнее.

1. LSTM (Long Short-Term Memory)

LSTM была разработана для решения проблемы исчезающего градиента, характерной для стандартных RNN. Она включает в себя специальные механизмы управления потоком информации, которые помогают сохранять долгосрочные зависимости. Основные компоненты LSTM:

• Ячейка памяти: хранит информацию на протяжении длительного времени.

• Входные ворота: определяют, какая информация будет добавлена в ячейку памяти.

• Ворота забывания: решают, какая информация будет удалена из ячейки памяти.

• Выходные ворота: контролируют, какая информация будет передана на выход.

2. GRU (Gated Recurrent Unit)

GRU является более простой альтернативой LSTM и также решает проблемы, связанные с исчезающим градиентом. Она объединяет входные и забывающие ворота в одно и использует меньше параметров, что делает ее более быстрой и менее ресурсоемкой. Основные компоненты GRU:

• Обновляющие ворота: определяют, какую информацию сохранять и обновлять в состоянии скрытого слоя.

• Сбросные ворота: контролируют, насколько предыдущая информация важна для текущего состояния.

3. Bidirectional RNN

Bidirectional RNN состоит из двух слоев RNN: один обрабатывает последовательность данных в прямом направлении, а другой — в обратном. Это позволяет модели учитывать контекст как до, так и после текущего элемента последовательности. Это особенно полезно в задачах, где контекст имеет значение (например, в обработке естественного языка).

# ****ЗАКЛЮЧЕНИЕ****

RNN сыграли ключевую роль в развитии нейронного машинного перевода, особенно с появлением моделей Seq2Seq и механизма внимания. Несмотря на то, что сегодня трансформеры (например, BERT, GPT) вытесняют RNN в многих задачах, понимание принципов работы рекуррентных сетей остается важным для специалистов в области NLP.

Перспективы развития:

Использование RNN в lightweight-моделях для устройств с ограниченными ресурсами.

Гибридные подходы, сочетающие RNN с другими архитектурами.

Таким образом, RNN продолжают оставаться важным инструментом в машинном переводе, хотя их применение постепенно сужается в пользу более современных архитектур.

# ****СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ****

1. Этические вопросы и проблемы нейросетей [Электронный источник] // URL - https://sky.pro/wiki/python/eticheskie-voprosy-i-problemy-nejrosetej/
2. Генеративные нейросети: восстание машин или новая экономика [Электронный источник] // URL - https://roscongress.org/materials/generativnye-neyroseti-vosstanie-mashin-ili-novaya-ekonomika/
3. Искусственный интеллект: друг или враг инвестора? Что ждет финансовые рынки по мере внедрения нейросетевых технологий [Электронный источник] // URL - https://habr.com/ru/companies/finam\_broker/articles/874466/
4. Направления влияния нейросетей на экономику, бизнес и образование [Электронный источник] // URL - https://cyberleninka.ru/article/n/napravleniya-vliyaniya-neyrosetey-na-ekonomiku-biznes-i-obrazovanie
5. Кто владеет результатами работы нейросетей? [Электронный источник] // URL - https://www.uplab.ru/blog/avtorskoe-pravo-i-neyroseti-kto-vladeet-rezultatami-ii-tvorchestva/