Столыпинский вестник

Научная статья

Original article

УДК 004.58

АНАЛИЗ ПАРАМЕТРОВ НЕЙРОННЫХ СЕТЕЙ В РАСПОЗНАВАНИИ РЕЧИ

PARAMETER ANALYSIS OF NEURAL NETWORKS IN SPEECH RECOGNITION

Горячкин Борис Сергеевич, кандидат технических наук, доцент; Московский государственный технический университет им. Н.Э. Баумана (105005, Москва, 2-я Бауманская ул., д. 5, с. 1), тел. 8(499) 263-63-91, bsgor@bmstu.ru

Назаров Максим Михайлович, магистрант; Московский государственный технический университет им. Н.Э. Баумана (105005, Москва, 2-я Бауманская ул., д. 5, с. 1), тел. 8(499) 263-63-91, nmm18u400@student.bmstu.ru

Boris S. Goryachkin, candidate of technical sciences, associate professor; Moscow State Technical University named after. N.E. Bauman (105005, Moscow, 2nd Baumanskaya st., 5, bldg. 1), tel. 8(499) 263-63-91, bsgor@bmstu.ru

Maksim M. Nazarov, master's student; Moscow State Technical University named after. N.E. Bauman (105005, Moscow, 2nd Baumanskaya st., 5, bldg. 1), tel. 8(499) 263-63-91, nmm18u400@student.bmstu.ru

Аннотация. В последние десятилетия нейронные сети приобрели статус ключевого инструмента в области обработки речи, стимулируя исследователей обратить внимание на вопросы оптимизации и оценки их параметров. Главная цель данной работы - провести тщательный анализ теоретических основ и разработать методы оценки параметров нейронных сетей, с особым вниманием

к их воздействию на процессы распознавания речи. Результаты: Основными аспектами исследования являются методы измерения воздействия параметров нейронных сетей на процессы распознавания речи и разработка критериев, определяющих оптимальные настройки для достижения максимальной точности и эффективности. Особое внимание уделяется экспериментальному подтверждению теоретических выводов на конкретных сценариях. Результаты исследования визуализированы в виде инфографической модели, систематизирующей теоретическую базу и иллюстрирующей влияние различных параметров на точность и скорость распознавания речи. Практическая значимость: Инфографическая модель представляет наглядное руководство для выбора оптимальных параметров и настроек нейронных сетей в задачах распознавания речи, способствуя повышению эффективности и точности систем обработки речи.

Abstract. In recent decades, neural networks have acquired the status of a key tool in the field of speech processing, stimulating researchers to pay attention to the issues of optimization and estimation of their parameters. The main goal of this work is to conduct a thorough analysis of the theoretical foundations and develop methods for estimating neural network parameters, with special attention to their impact on speech recognition processes. Results: The main aspects of the study are methods for measuring the impact of neural network parameters on speech recognition processes and the development of criteria that define optimal settings to maximize accuracy and efficiency. Emphasis is placed on experimental validation of theoretical findings on specific scenarios. The results of the study are visualized in the form of an infographic model that systematizes the theoretical basis and illustrates the influence of various parameters on the accuracy and speed of speech recognition. Practical significance: The infographic model provides a visual guide for selecting optimal parameters and settings of neural networks in speech recognition tasks, contributing to improving the efficiency and accuracy of speech processing systems.

Ключевые слова: *Нейронные сети, распознавание речи, оптимизация параметров, оценка параметров, точность распознавания, скорость распознавания.*

Keywords: *Neural networks, speech recognition, parameter optimization, parameter estimation, recognition accuracy, recognition rate.*

Введение

В последние десятилетия нейронные сети приобрели статус ключевого инструмента в области обработки речи, побуждая исследователей уделять пристальное внимание вопросам оптимизации и оценки их параметров. Основная цель данного исследования – провести тщательный анализ теоретических основ, разработать методы оценки параметров нейронных сетей и изучить их влияние на процессы распознавания речи.

Ключевыми аспектами работы являются методы измерения воздействия различных параметров нейронных сетей на производительность систем распознавания речи, а также разработка критериев для определения оптимальных настроек, обеспечивающих максимальную точность и эффективность. Особое внимание уделяется экспериментальной валидации теоретических выводов на практических примерах и сценариях, демонстрирующих влияние конкретных параметров и настроек на точность и скорость распознавания речи.

Для систематизации теоретической базы и визуализации полученных результатов разработана инфографическая модель, обобщающая взаимосвязи между различными параметрами нейронных сетей и их воздействие на процессы распознавания речи. Такая наглядная модель представляет практическую ценность, выступая в качестве руководства для специалистов при выборе оптимальной конфигурации нейронных сетей.

Параметры нейронных сетей

Распознавание речи — это задача компьютерного зрения, которая заключается в автоматическом преобразовании звуков речи в текст [1]. Эта задача

имеет большое значение для различных приложений, таких как голосовой ввод, голосовой поиск и автоматический перевод.

В последние годы распознавание речи добилось значительных успехов благодаря развитию нейронных сетей. Нейронные сети — это класс алгоритмов машинного обучения, которые способны обучаться на больших наборах данных и обнаруживать сложные закономерности.

В этой работе будут рассмотрены основные параметры, которые влияют на работу нейронных сетей в задачах распознавания речи. Эти параметры включают в себя:

Архитектура нейронной сети: под архитектурой нейронной сети понимается способ, как нейроны (блоки элементов нейросети) организованы и взаимодействуют друг с другом [2]. Каждая архитектура имеет свои особенности и различные блоки из которых она состоит, что влияет на сам принцип работы нейронной сети.

Глубина сети: Глубина нейронной сети отвечает за количество слоев в данной нейронной сети. Чем больше слоев, тем больше нейронов и связей между ними задействовано в ходе работы сети, что позволяет улучшить скорость обучения и точность нейронной сети.

Размерность входных данных: Размерность входных данных — это количество и структура параметров (функций или признаков), предоставляемых в качестве входных данных модели [2]. В контексте задачи распознавания речи входные данные представляют собой аудиофайл, который будет преобразован в спектрограмму, чтобы нейронная сеть могла обучать на фрагментах спектрограммы.

Использование функций активаций: Функции активации играют ключевую роль в нейронных сетях, определяя выход нейрона в зависимости от его входных данных. Это позволяет по-разному задействовать целые слои нейронное сети изменяя процесс ее работы, что позволяет улучшить точность работы сети.

Оптимизаторы: Оптимизаторы в контексте нейронных сетей — это алгоритмы, которые используются для минимизации функции потерь путем

обновления параметров модели. Оптимизаторы позволяют минимизировать разницу между прогнозами модели и фактическими значениями, что позволяет увеличить скорость достижения наилучшей точности модели.

Скорость обучения: Скорость обучения в контексте нейронных сетей — это параметр, который определяет, насколько сильно обновляются параметры (нейроны) модели на каждом шаге обучения. Это влияет на то, насколько быстро или медленно модель адаптируется к тренировочным данным и находит оптимальные значения параметров. Скорость обучения модели позволяет быстрее достигнуть минимума градиентного спуска, что позволит достичь желанного результата с наименьшим затраченным временем.

Регуляризация: Регуляризация — это метод контроля за сложностью модели с целью предотвращения переобучения. Переобучение происходит, когда модель слишком хорошо подстраивается под тренировочные данные и теряет обобщающую способность на новых, ранее не виденных данных.

Использование предобученных моделей: В задаче распознавания речи можно воспользоваться предварительно обученными моделями, которые были обучены на больших наборах данных. Такие модели могут быть предварительно обучены на задачах, связанных с распознаванием речи. Затем эти предобученные модели могут быть дообучены на более узкой задаче распознавания речи в текст, чтобы улучшить производительность и сэкономить вычислительные ресурсы.

Методы и средства оценки параметров

Для оценки параметров нейронных сетей распознающих речь используются различные методы и средства. Выбор конкретного метода зависит от типа параметра, который необходимо оценить.

Для анализа архитектуры нейронной сети можно использовать следующие методы:

1. Word Error Rate (WER) - WER измеряет процент ошибок в распознанном тексте по сравнению с эталонным текстом на уровне слов. WER рассчитывается как отношение суммы вставок (вставленные слова),

удалений (пропущенные слова) и замен (замененные слова) к общему числу слов в эталонном тексте [3].

Формула:

$$WER = \frac{I+D+S}{N} (1)$$, где:

I - количество вставок,

D - количество удалений,

S - количество замен,

N - общее количество слов в эталонном тексте.

2. Character Error Rate (CER) - CER измеряет процент ошибок в распознанном тексте по сравнению с эталонным текстом на уровне символов. CER рассчитывается как отношение суммы вставок (вставленные символы), удалений (пропущенные символы) и замен (замененные символы) к общему числу символов в эталонном тексте [3].

Формула:

$$CER = \frac{I+D+S}{N} (2)$$, где:

I - количество вставок,

D - количество удалений,

S - количество замен,

N - общее количество символов в эталонном тексте.

Для анализа глубины сети будем использовать кривую обучения. Кривая обучения – это график, отображающий изменение производительности модели (точность или ошибка) на тренировочном и тестовом наборах данных в зависимости от количества прошедших эпох обучения. Анализ кривой обучения может помочь в оценке эффективности обучения модели, выявлении таких проблем как переобучения или недообучения и оптимизации процесса обучения [4].

Для анализа размерности входных данных используется точность при разных размерах входных данных. Этот процесс обычно включает в себя

изменение размера входных данных и измерение точности модели на каждом размере [4].

Для анализа использования функций активаций используется точность модели. Точность модели используется для оценки того, как различные функции активации влияют на производительность нейронной сети. Функции активации определяют, как нейроны реагируют на входные данные и какой выход они производят [5].

Анализ оптимизаторов с использованием скорости достижения наивысшей точности может быть эффективным способом оценки того, как различные оптимизаторы влияют на обучение вашей нейронной сети. Оптимизаторы отвечают за обновление весов сети в процессе обучения с целью минимизации функции потерь.

Проанализировать скорость обучения модели нейронной сети можно с помощью WER и точности. Эти метрики могут предоставить информацию о том, насколько быстро модель обучается и как хорошо она обобщает на новых данных [6].

Для анализа регуляризации можно использовать уменьшение переобучения и увеличение точности. Регуляризация предотвращает переобучение модели, улучшая её обобщающую способность.

При использовании предобученных моделей в ансамбле (“склейке”) можно ожидать улучшения в общей производительности, поскольку разные модели могут обладать различными сильными сторонами и способностью к обнаружению разных паттернов в данных. Комбинирование этих моделей может привести к лучшему обобщению и снижению риска переобучения.

Таблица 1 является сводной и содержит информацию о наиболее важных параметрах нейронных сетей, которые могут быть оценены. Для каждого параметра указаны методы и средства оценки, а также критерии оценивания, рассмотренные ранее.

Таблица 1. Параметры нейронной сети, методы и средства оценки, критерии оценивания.

Параметры нейронной сети	**Методы и средства оценки**	**Критерии оценивания**
архитектура	WER, CER	минимальные значения WER, CER
глубина сети	кривая обучения	количество слоев и итоговая точность
размерность входных данных	точность при разных размерах входных данных	максимальная точность с определенным размером входных данных
использование функций активаций	точность	максимальная точность
оптимизаторы	скорость достижения наибольшей точности	максимальная точность при минимальном времени
скорость обучения	WER, точность	скорость достижения минимального значения WER
регуляризация	уменьшение переобучения, увеличение точности	при заданной регуляризации значение переобучения уменьшается и достижение наибольшей точности
использование предобученных моделей	разница между базовой моделью и «склейки» моделей	сравнение результатов точности обычной модели и модели в «склейке»

Данные о выбранных нейронных сетях

Распознавание речи слабослышащих отличается от стандартного распознавания речи из-за особенностей произношения и речи. Для анализа нейронных сетей, предназначенных для распознавания речи слабослышащих, необходимо уделить особое внимание фонетическим признакам. Это поможет улучшить эффективность работы нейронных сетей. При выборе нейронных сетей для сравнения в данной задаче основное внимание уделяется статье «A Comparison of Transformer, Convolutional, and Recurrent Neural Networks on Phoneme Recognition» [7]. В данной статье подчеркивается, что способность извлекать фонологически значимые признаки является ключевой для задачи распознавание речи. Подходы к извлечению признаков в нейронных сетях, таких как сверточные (CNN), рекуррентные (RNN), трансформерные (transformer) и conformer, представлены в статье. В статье отмечается, что распознавание фонем рассматривается как задача, зависящая от очень короткого временного интервала,

по сравнению с лингвистической обработкой. Авторы отмечают, что задача распознавания фонем является наиболее подходящей для оценки способности извлечения фонетических признаков, так как она требует от модели меньше информации, чем другие задачи обработки речи, и может быть легко измерена по точности.

RNN (Рекуррентная нейронная сеть) [8]: использует рекуррентные связи для обработки последовательных данных. Обладает внутренним состоянием, которое позволяет учитывать контекст предыдущих моментов времени. Была выбрана классическая архитектура RNN без дополнительных модификационных блоков.

CNN (Сверточная нейронная сеть) [9]: применяет операции свертки для выделения локальных паттернов в данных. Алгоритм работы сети представляет собой поэтапное сворачивание входных данных до минимального значения, заданного при построении, для выделения ключевых паттернов в данных. После операции свертывания нейронная сеть начинает развертывание данных до момента совпадения размера с поступившими входными данными без потери, так как присутствуют специальные блоки, отвечающие за связь операций свертывания и развертывания.

Transformer [10]: использует механизм внимания для обработки входных данных параллельно, без рекуррентных связей. Используется три основных блока в своей архитектуре: входной (энкодер) блок, механизм внимания (Self-Attention блок), выходной (декодер) блок. Во входном блоке звуковой сигнал анализируется на предмет важных фонем или звуковых фрагментов. Механизм внимания позволяет входному блоку обращать больше внимания на определенные части входных данных, в зависимости от их важности для контекста. Выходной блок генерирует последовательность слов, представляющую распознанный текст.

Conformer [11]: комбинирует идеи из RNN, CNN и Transformer. Использует слои внимания, свертку во времени и контекстуальные блоки для обработки последовательных данных. Входной уровень преобразует аудиосигнал в

спектрограмму [12], чтобы представить частотные характеристики звука. Затем используются сверточные блоки для обработки временных и частотных характеристик входных данных. Это помогает извлекать локальные признаки и улавливать шаблоны в аудиосигнале. Self-Attention блок, в отличие от классических Transformer, включает слой внимания, который помогает модели улавливать долгосрочные зависимости в данных. Этот слой внимания дополнительно взаимодействует со сверточными слоями. Далее идут трансформерные блоки. Эти блоки обрабатывают последовательности признаков, учитывая как локальные, так и глобальные зависимости. Выходы последнего трансформерного блока подаются на линейный слой для генерации финального распознанного текста.

Экспериментальные исследования и анализ параметров нейронных сетей для распознавания речи

1. Информация об используемых данных

Для сравнения нейронных сетей был выбран набор данных LibriSpeech [13] с дополнительным разделом речи слабослышащих [14]. Этот обширный и бесплатно доступный набор содержит аудиозаписи и соответствующие текстовые транскрипции для задач распознавания и обработки речи. Он включает более 1000 часов записей, выполненных профессионально с использованием различных микрофонов и дикторов. Набор данных состоит из пяти поднаборов, каждый из которых содержит аудиозаписи книг из библиотеки LibriVox, разбитых на главы.

2. Результаты эксперимента

Параметр 1 (Архитектура нейронной сети). В ходе 1 эксперимента были обучены 4 модели нейронных сетей. В таблице 2 занесены результаты эксперимента, в данном эксперименте учитывались только методы оценки WER и CER.

Таблица 2. Результаты сравнения архитектур нейронных сетей

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**	
		WER	CER
Архитектура	RNN	10,4	9,8
	CNN	8,8	7,2
	Transformer	7,5	6,9
	Conformer	5,4	3,7

Параметр 2 (Глубина сети). В рамках эксперимента 2, нацеленного на изучения глубины сети, были получены два графика. Первый график, изображенный на рисунке 1, является кривой обучения, которая содержит информацию о точности (accuracy) обученных моделей нейронных: RNN, CNN, Transformer, Conformer на обучающей и тестовой выборках данных. На данном графике отображается прогресс обучения нейронной в зависимости от количества нейронов внутри каждой нейронной сети.

Формула точности:

$$accuracy = \frac{\text{Количество правильных предсказаний}}{\text{Общее количество примеров}} \quad (3)$$

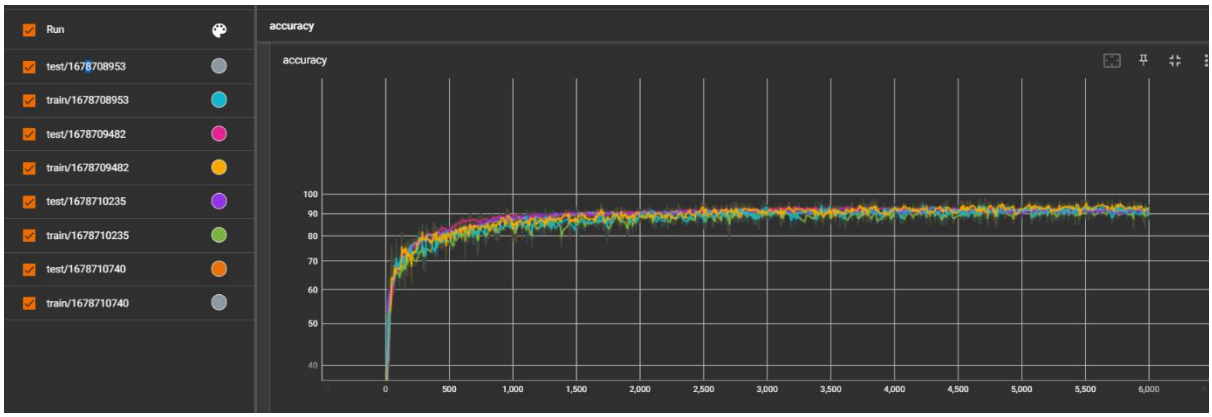

Рис. 1. Кривая обучения - по точности.

Параметр 3 (Размерность входных данных). В рамках 3 эксперимента исследовалось влияние размерности входных данных на точность моделей нейронной сети. Исследуемые модели нейронных сетей: RNN, CNN, Transformer, Conformer. Основным показателем является точность из формулы 3. Результаты представлены в таблице 3.

Таблица 3. Результаты сравнения точности нейронных сетей по размерности входных данных

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**		
		Точность для размер вх. д. 25	Точность для размер вх. д. 50	Точность для размер вх. д. 75
Размерность входных данных	RNN	0,81	0,82	0,76
	CNN	0,82	0,82	0,71
	Transformer	0,85	0,85	0,90
	Conformer	0,87	0,89	0,93

Параметр 4 (Использование функций активаций). В 4 эксперименте проводилось исследование связанной с использованием функций активаций. Была измерена и занесена в таблицу 4 точность в зависимости от активационных функций [15].

Таблица 4. Результаты сравнения точности нейронных сетей с разными функциями активации

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**	
		ReLU [15]	Tanh [15]
Использование функций активаций	RNN	0,82	0,84
	CNN	0,78	0,80
	Transformer	0,90	0,89
	Conformer	0,90	0,91

Параметр 5 (Оптимизаторы). В 5 эксперименте исследовалось влияние оптимизаторов на скорость обучения нейронных сетей. В ходе данного эксперимента были получены результаты влияния оптимизатора на время обучения нейронной сети (в часах) и эти результаты были занесены в таблицу 5. В рамках данного эксперимента было использовано ограничивающее значение точности нейронной сети, после которого она останавливала процесс обучения. Значение было выставлено на точности 0.70.

Таблица 5. Результаты сравнения времени обучения нейронных сетей с оптимизаторами и без оптимизаторов

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**		
		Без оптимизатора	Adam	SGD
Оптимизаторы	RNN	2	2	1
	CNN	3	2	2
	Transformer	5	3	6
	Conformer	5	3	6

Параметр 6 (Скорость обучения). В рамках данного эксперимента исследовалось, скорость обучения влияет на показатель WER. Результаты данного эксперимента занесены в таблицу 6.

Таблица 6. Результаты сравнения обучения нейронных сетей по показателю WER от времени обучения

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**		
		WER при вр. об 1 ч.	WER при вр. об 3 ч.	WER при вр. об 5 ч.
Скорость обучения	RNN	0,71	0,81	0,61
	CNN	0,69	0,82	0,82
	Transformer	0,47	0,81	0,91
	Conformer	0,45	0,8	0,92

Параметр 7 (Регуляризация). В 7 эксперименте исследовалось влияние использования регуляризации на производительность нейронных сетей. Регуляризация в контексте нейронных сетей обычно применяется для предотвращения переобучения модели на тренировочных данных, что приводит к увеличению точности нейронной сети. Данные эксперимента представлены в таблице 7.

Таблица 7. Результаты сравнения точности нейронных сетей с разными функциями активации

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**	
		Без регуляризации	С регуляризацией

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**	
Регуляризация	RNN	0,78	0,84
	CNN	0,80	0,87
	Transformer	0,45	0,92
	Conformer	0,85	0,93

Параметр 8 (Использование предобученных моделей). В 8 эксперименте исследовалось влияние использование предобученных моделей. В третьем столбце идет значения нейронной сети без предобученных моделей. В 5 столбце идут значение точности после «склейки» двух разных моделей нейронных сетей [17]. Данные эксперимента представлены в таблице 8.

Таблица 8. Результаты сравнения точности нейронных сетей с использование предобученных и без использования предобученных моделей

Параметры нейронной сети	**Нейронная сеть**	**Методы и средства оценки**	**Нейронная сеть**	**Методы и средства оценки**
		Точность [16]		Точность [16]
Использование предобученных моделей	RNN	0,81	RNN+LSTM [17]	0,85
	CNN	0,82	CNN+Jasper [18]	0,88
	Transformer	0,81	Transformer+Whisper[19]	0,94
	Conformer	0,8	Conformer+SpeechStew [20]	0,95

Инфографическая модель анализа параметров нейронных сетей для распознавания речи

Для наглядного представления результатов проведенного исследования и обобщения полученных данных была разработана инфографическая модель анализа параметров нейронных сетей для распознавания речи. Инфографика позволяет визуализировать сложные взаимосвязи между различными компонентами процесса и способствует лучшему пониманию влияния параметров на эффективность нейронных сетей.

В основе инфографической модели, представленной на рисунке 2, лежит детальная таблица 1, содержащая параметры нейронных сетей, используемых

для распознавания речи, методы и средства оценки, а также критерии оценивания, полученные в ходе последующих экспериментов. Инфографическая модель объединяет и систематизирует результаты предыдущих разделов статьи, иллюстрируя взаимосвязи между различными аспектами исследования и обеспечивая более полное понимание изменений, происходящих в процессе работы.

На инфографической модели наглядно представлены различные параметры нейронных сетей, включая количество слоев, число нейронов в каждом слое, используемые методы активации и другие ключевые характеристики, применяемые при распознавании речи. Основные параметры из таблицы 1 выделены оранжевым цветом, что позволяет легко оценить их влияние на конечные результаты эксперимента. Такая визуализация представляет практическую ценность для специалистов, работающих с нейронными сетями в задачах распознавания речи, помогая им выбирать оптимальные параметры и настройки.

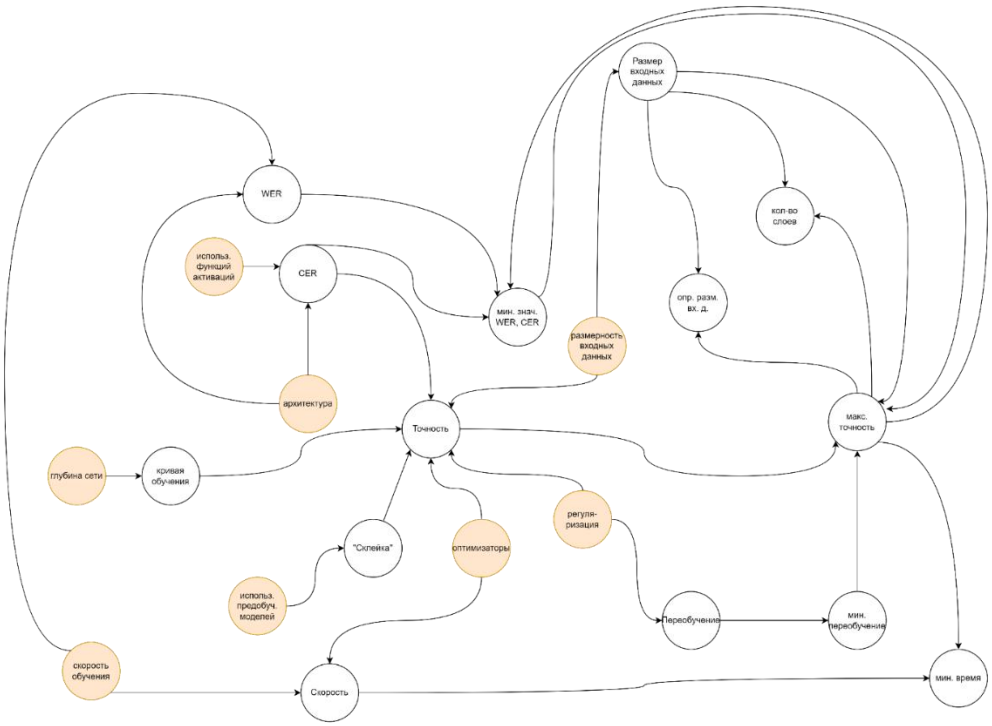

Рис. 2. Инфографическая модель анализа параметров нейронных сетей для распознавания речи

Таким образом, инфографическая модель выступает в качестве наглядного руководства и обобщения результатов исследования, подчеркивая важность

визуализации и моделирования сложных процессов для более глубокого понимания и практического применения.

Заключение

Данное исследование посвящено всестороннему изучению влияния параметров и настроек нейронных сетей на их эффективность в задачах распознавания речи. Рассмотрение различных аспектов, таких как архитектура сети, функции активации, скорость обучения и регуляризация, подчеркивает их ключевую роль в обеспечении высокой производительности нейронных сетей в этой области.

Экспериментальные исследования и анализ данных выявили основные факторы, влияющие на точность и скорость распознавания речи с помощью нейронных сетей. Одним из важнейших выводов стала необходимость тщательного подбора параметров для максимально эффективной работы нейросетевых моделей.

В ходе исследования были разработаны методы оценки параметров нейронных сетей и критерии для определения их оптимальных значений, обеспечивающих наилучшие результаты в задачах распознавания речи.

Для наглядной систематизации теоретической базы и экспериментальных данных была создана инфографическая модель, визуализирующая взаимосвязи между различными параметрами нейронных сетей и их влияние на процессы распознавания речи. Такая модель представляет практическую ценность, выступая в качестве руководства для специалистов при выборе оптимальной конфигурации нейросетевых моделей и способствуя глубокому пониманию принципов их работы.

Conclusion

This study is devoted to a comprehensive investigation of the influence of neural network parameters and settings on their performance in speech recognition tasks. Consideration of various aspects such as network architecture, activation functions, learning rate and regularization highlights their key role in ensuring high performance of neural networks in this domain.

Experimental studies and data analysis have revealed the main factors affecting the accuracy and speed of speech recognition using neural networks. One of the most important findings was the need for careful selection of parameters to maximize the performance of neural network models.

During the study, methods for estimating the parameters of neural networks and criteria for determining their optimal values that provide the best results in speech recognition tasks were developed.

For visual systematization of the theoretical basis and experimental data, an infographic model was created to visualize the relationships between different parameters of neural networks and their influence on speech recognition processes. Such a model is of practical value, acting as a guide for specialists in choosing the optimal configuration of neural network models and contributing to a deep understanding of the principles of their operation.

Литература

1. Ле Н.В., Панченко Д.П. Распознавание речи на основе искусственных нейронных сетей // Технические науки в России и за рубежом: материалы I Междунар. науч. конф. (г. Москва, май 2011 г.). – Москва: Ваш полиграфический партнер, 2011. – С. 8–11.
2. Беседин И.Ю. Анализ проблем автоматического распознавания речи // Наука. Инновации. Технологии. – 2010. – №70.
3. Evaluate OCR Output Quality with Character Error Rate (CER) and Word Error Rate (WER) / Kenneth Leung. URL: https://towardsdatascience.com/evaluating-ocr-output-quality-with-character-error-rate-cer-and-word-error-rate-wer-853175297510 (дата обращения: 20.12.2023).
4. Как работают системы распознавания речи / URL: https://habr.com/ru/companies/amvera/articles/691288/ (дата обращения: 20.12.2023).
5. Метрики оценки для моделей распознавания устной речи / URL: https://learn.microsoft.com/ru-ru/azure/ai-services/language-

service/conversational-language-understanding/concepts/evaluation-metrics (дата обращения: 20.12.2023).

6. Measure and improve speech accuracy / URL: https://cloud.google.com/speech-to-text/docs/speech-accuracy (дата обращения: 21.12.2023)
7. A Comparison of Transformer, Convolutional, and Recurrent Neural Networks on Phoneme Recognition / Kyuhong Shim, Wonyong Sung. URL: https://arxiv.org/pdf/2210.00367.pdf (дата обращения: 20.12.2023).
8. A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement / Jean-Marc Valin. URL: https://arxiv.org/pdf/1709.08243.pdf (дата обращения: 20.12.2023).
9. Чучупал В.Я., Коренчиков А.А. Моделирование вариативности произношения для уменьшения уровня ошибок при распознавании речи // Моделирование вариативности произношения. – 2014. – С. 1168–1179.
10. Zhang Q., Lu H., Sak H., Tripathi A., McDermott E., Koo S., Kumar S. Transformer transducer: A streamable speech recognition model with transformer encoders and rnn-t loss // ICASSP 2020, 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2020. – P. 7829–7833.
11. Gulati A., Qin J., Chiu C.-C., Parmar N., Zhang Y., Yu J., Han W., Wang S., Zhang Z., Wu Y., Pang R. Conformer: Convolution augmented Transformer for Speech Recognition // Interspeech 2020, 2020. – 5 p.
12. Suppression of acoustic noise in speech using spectral subtraction / S. Boll. URL: https://ieeexplore.ieee.org/document/1163209/metrics#metrics (дата обращения: 22.12.2023).
13. Panayotov V., Chen G., Povey D., Khudanpur S. LibriSpeech: an ASR corpus based on public domain audio books // ICASSP 2015, 2015. – 5 p.
14. Пахомкин К.С., Торжков М.С., Алехин С.С., Канев А.И., Рогозин Д.Р. Анализ библиотек автоматического распознавания речи // ИИАСУ 2022, 2022.

15. Чучупал В.Я., Коренчиков А.А. Моделирование вариативности произношения для уменьшения уровня ошибок при распознавании речи // Моделирование вариативности произношения. – 2014. – С. 1168–1179.
16. Self-training and Pre-training are Complementary for Speech Recognition / Qiantong Xu, Alexei Baevski, Tatiana Likhomanenko/ URL: https://arxiv.org/pdf/2010.11430v1.pdf (дата обращения: 22.12.2023).
17. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. Attention is all you need. – 2017. – 15 p.
18. Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition / Yu Zhang, James Qin, Daniel S. Park/ URL: https://arxiv.org/pdf/2010.10504v2.pdf (дата обращения: 22.12.2023).
19. OpenAI открыла код системы распознавания речи Whisper / URL: https://habr.com/ru/news/690104/ (дата обращения: 22.12.2023).
20. SpeechStew: Simply Mix All Available Speech Recognition Data to Train One Large Neural Network / William Chan, Daniel S. Park, Chris A. Lee. URL: https://arxiv.org/pdf/2104.02133v3.pdf (дата обращения: 22.12.2023).

References

1. Le N.V., Panchenko D.P. (2011) Raspoznavanie rechi na osnove iskusstvennykh neironnykh setei [Speech Recognition Based on Artificial Neural Networks]. Tekhnicheskie nauki v Rossii i za rubezhom: materialy I Mezhdunar. nauch. konf. [Technical Sciences in Russia and Abroad: Materials of the 1st International Scientific Conference] (Moskva, mai 2011 g.), Moskva:Vash poligraficheskii partner, pp. 8–11.
2. Besedin I.Yu. (2010) Analiz problem avtomaticheskogo raspoznavaniya rechi [Analysis of Problems in Automatic Speech Recognition]. Nauka. Innovatsii. Tekhnologii [Science. Innovation. Technologies], no 70.
3. Leung K. Evaluate OCR Output Quality with Character Error Rate (CER) and Word Error Rate (WER). Available at: https://towardsdatascience.com/evaluating-ocr-output-quality-with-character-error-rate-cer-and-word-error-rate-wer-853175297510.

4. Kak rabotayut sistemy raspoznavaniya rechi [How Speech Recognition Systems Work]. Available at: https://habr.com/ru/companies/amvera/articles/691288/.
5. Metriki otsenki dlya modelei raspoznavaniya ustnoi rechi [Evaluation Metrics for Speech Recognition Models]. Available at: https://learn.microsoft.com/ru-ru/azure/ai-services/language-service/conversational-language-understanding/concepts/evaluation-metrics (accessed 20 December 2023).
6. Measure and improve speech accuracy. Available at: https://cloud.google.com/speech-to-text/docs/speech-accuracy.
7. Shim K., Sung W. (2023) A Comparison of Transformer, Convolutional, and Recurrent Neural Networks on Phoneme Recognition. Available at: https://arxiv.org/pdf/2210.00367.pdf (accessed 20 December 2023).
8. Valin J.-M. (2017) A Hybrid DSP/Deep Learning Approach to Real-Time Full-Band Speech Enhancement. Available at: https://arxiv.org/pdf/1709.08243.pdf.
9. Chuchupal V.Ya., Korenchikov A.A. (2014) Modelirovanie variativnosti proiznosheniya dlya umen'sheniya urovnya oshibok pri raspoznavanii rechi [Pronunciation Variation Modeling for Reducing Speech Recognition Errors]. Modelirovanie variativnosti proiznosheniya [Pronunciation Variation Modeling], pp. 1168–1179.
10. Zhang Q., Lu H., Sak H., Tripathi A., McDermott E., Koo S., Kumar S. (2020) Transformer transducer: A streamable speech recognition model with transformer encoders and rnn-t loss. ICASSP 2020, 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, pp. 7829–7833.
11. Gulati A., Qin J., Chiu C.-C., Parmar N., Zhang Y., Yu J., Han W., Wang S., Zhang Z., Wu Y., Pang R. (2020) Conformer: Convolution augmented Transformer for Speech Recognition. Interspeech 2020, 5 p.
12. Boll S. Suppression of acoustic noise in speech using spectral subtraction. Available at: https://ieeexplore.ieee.org/document/1163209/metrics#metrics.
13. Panayotov V., Chen G., Povey D., Khudanpur S. (2015) LibriSpeech: an ASR corpus based on public domain audio books. ICASSP 2015, 5 p.

14. Pakhomkin K.S., Torzhkov M.S., Alekhin S.S., Kanev A.I., Rogozin D.R. (2022) Analiz bibliotek avtomaticheskogo raspoznavaniya rechi [Analysis of Automatic Speech Recognition Libraries]. IIASU 2022.
15. Chuchupal V.Ya., Korenchikov A.A. (2014) Modelirovanie variativnosti proiznosheniya dlya umen'sheniya urovnya oshibok pri raspoznavanii rechi [Pronunciation Variation Modeling for Reducing Speech Recognition Errors]. Modelirovanie variativnosti proiznosheniya [Pronunciation Variation Modeling], pp. 1168–1179.
16. Xu Q., Baevski A., Likhomanenko T. (2020) Self-training and Pre-training are Complementary for Speech Recognition. Available at: https://arxiv.org/pdf/2010.11430v1.pdf.
17. Vaswani A., Shazeer N., Parmar N., Uszkoreit J., Jones L., Gomez A.N., Kaiser L., Polosukhin I. (2017) Attention is all you need, 15 p.
18. Zhang Y., Qin J., Park D.S. (2020) Pushing the Limits of Semi-Supervised Learning for Automatic Speech Recognition. Available at: https://arxiv.org/pdf/2010.10504v2.pdf.
19. OpenAI otkryla kod sistemy raspoznavaniya rechi Whisper [OpenAI Open-Sourced Whisper Speech Recognition System]. Available at: https://habr.com/ru/news/690104.
20. Chan W., Park D.S., Lee C.A. (2021) SpeechStew: Simply Mix All Available Speech Recognition Data to Train One Large Neural Network. Available at: https://arxiv.org/pdf/2104.02133v3.pdf.

Для цитирования: Горячкин Б.С., Назаров М.М. Анализ параметров нейронных сетей в распознавании речи// Научный сетевой журнал «Столыпинский вестник» №5/2024.