

СРАВНЕНИЕ АРХИТЕКТУР НЕЙРОСЕТЕЙ ДЛЯ АНАЛИЗА ТЕКСТОВЫХ ДАННЫХ

COMPARISON OF NEURAL NETWORK ARCHITECTURES FOR TEXT DATA ANALYSIS

Роднин А.В., аспирант, ФГБОУ ВО «Вологодский государственный университет», Вологда, Россия
Rodnin A.V., Post graduate student, Vologda State University, Vologda, Russia

Горбунов В.А., доктор физико-математических наук, профессор, профессор, ФГБОУ ВО «Вологодский государственный университет», Вологда, Россия
Gorbunov V.F., DrSc in Physics and Mathematics, professor, professor, Vologda State University, Vologda, Russia

Аннотация. Данная статья представляет сравнительный анализ архитектур нейросетей для обработки текстовых данных. Рассматриваются особенности работы различных моделей (RNN, CNN, трансформеры) и их алгоритмы обучения, а также эффективность в извлечении семантической информации из больших текстовых корпусов. Основной акцент делается на выявлении преимуществ и недостатков каждой архитектуры для оптимизации задач анализа текста.

Ключевые слова: нейросети, архитектуры нейросетей, анализ текстовых данных.

Abstract. The article presents a comparative analysis of neural network architectures for processing text data. The features of various models (RNN, CNN, transformers) and their learning algorithms are considered, as well as their effectiveness in extracting semantic information from large text corpora. The main focus is on identifying the advantages and disadvantages of each architecture to optimize text analysis tasks.

Keywords: neural networks, neural network architectures, text data analysis.

В образовательной сфере объем и разнообразие информационных источников постоянно растут, учебные пособия, научные статьи, книги и новости требуют оперативного анализа и обработки. Создание тестовых материалов, лабораторных работ и мини пособий зачастую сопряжено с большими трудозатратами, что затрудняет обеспечение качественного и своевременного образовательного контента. Автоматизация этих процессов становится особенно актуальной, поскольку традиционные методы ручной подготовки не справляются с возросшей нагрузкой [2].

Нейросетевые модели, способные выполнять задачи классификации, извлечения информации, анализа тональности и генерации текстов, представляют собой мощный инструмент для решения подобных задач. Их применение позволяет ускорить разработку тестовых заданий, лабораторных работ и кратких обучающих материалов, обеспечивая при этом высокую точность и адаптивность создаваемых образовательных ресурсов.

Цель данной работы заключается в проведении сравнительного анализа различных архитектур нейросетей, используемых для обработки текстовых данных, с акцентом на выявление их сильных и слабых сторон. Результаты исследования позволят определить оптимальные подходы для реализации систем анализа текстовой информации в условиях постоянно растущего информационного потока.

Современные нейросети стали мощным инструментом для работы с текстовой информацией, обеспечивая более глубокий анализ данных по сравнению с традиционными методами. В отличие от статистических алгоритмов, которые обрабатывают слова изолированно, нейросетевые модели способны распознавать контекст, выявлять взаимосвязь понятий и учитывать смысловую структуру предложений [1]. Это делает их особенно эффективными при работе с большими объемами текстов, таких как учебные материалы, статьи и книги.

Благодаря этим возможностям нейросети успешно применяются в задачах классификации текстов, извлечения ключевой информации и даже генерации осмысленных ответов на вопросы. В образовательной сфере такие технологии могут сыграть важную роль в автоматизации создания тестов, резюмировании длинных документов или адаптировании материалов под уровень подготовки учащихся. Однако разные нейросетевые архитектуры решают эти задачи по-разному, обладая своими преимуществами и ограничениями.

Для эффективного анализа и сравнения различных архитектур нейросетей, применяемых к обработке текстовых данных, необходимо учитывать несколько ключевых критериев. Во-первых, важна точность и способность модели к обобщению – насколько она хорошо справляется с классификацией, извлечением информации и пониманием сложных зависимостей в тексте. Это особенно критично в образовательных задачах, где необходимо корректное выявление тем, ключевых идей и контекстных связей в учебных материалах.

Во-вторых, значимыми характеристиками являются вычислительные затраты и время обучения [1]. Разные архитектуры требуют различного количества ресурсов: одни модели могут быть быстрее и легче в обучении, но менее точными, тогда как другие обеспечивают высокую точность за счет значительных вычислительных мощностей. Это особенно важно при обработке больших объемов текстовых данных, таких как книги или статьи.

Кроме того, модели различаются по способности работать с длинными последовательностями. В образовательных приложениях важно, чтобы нейросеть могла учитывать глобальный контекст документа, а не ограничивалась анализом отдельных фрагментов. Современные трансформерные модели решают эту проблему с помощью специальных механизмов, позволяющих эффективно обрабатывать длинные тексты.

Также важным критерием является гибкость и возможность дообучения, поскольку модели должны адаптироваться под конкретные образовательные задачи, такие как анализ тональности текстов, генерация тестов или ответы на вопросы. И, наконец, немаловажным остается уровень интерпретируемости модели – возможность объяснить, почему она пришла к тем или иным выводам. Это играет ключевую роль в образовательных и научных приложениях, где важна прозрачность результатов.

Эти критерии позволяют объективно сравнить различные архитектуры нейросетей и определить их эффективность в задачах анализа текстовых данных. Рассмотрим, как они проявляются на практике при использовании популярных типов нейросетей.

Простые нейронные сети, или многослойные персептроны (MLP), представляют собой один из самых базовых типов нейросетевых архитектур. В таких моделях данные проходят от входного слоя к выходному через один или несколько скрытых слоев, где происходит обработка информации [1]. MLP используют функции активации для преобразования данных, что позволяет моделям выявлять нелинейные зависимости и улучшать точность предсказаний при решении различных задач.

Преимущества таких сетей включают их простоту в реализации и сравнительно низкие вычислительные затраты, что делает их удобными для небольших задач с ограниченным объемом данных. Они могут эффективно работать при решении задач классификации, например, для анализа текстов с ограниченным количеством категорий или при использовании с небольшими объемами данных, такими как классификация коротких предложений или идентификация ключевых слов.

Однако ограничения MLP становятся очевидными при анализе текстов, где важен порядок слов и контекст. Модели MLP не учитывают последовательность входных данных, что критично для понимания сложных текстов или извлечения информации из длинных документов. Например, для анализа текста, где важно понять, как одно слово связано с другими в пределах предложения или даже абзаца, многослойный персептрон не способен эффективно обрабатывать такие зависимости. Поэтому для более сложных задач анализа текстов, таких как извлечение смысла из длинных фрагментов текста, такие сети оказываются недостаточными.

Рекуррентные нейронные сети (RNN, LSTM, GRU) были разработаны специально для работы с последовательными данными, такими как текст. В отличие от многослойных персептронов, они обладают циклическими связями, которые позволяют учитывать предыдущую информацию при обработке текущего элемента последовательности. Это делает их особенно полезными для задач анализа текстов, где важен порядок слов и контекст [2].

Основное преимущество RNN заключается в способности запоминать предыдущие элементы последовательности, что делает их подходящими для обработки предложений и абзацев. Однако стандартные RNN страдают от проблемы исчезающего градиента, из-за чего они слабо справляются с запоминанием длинных зависимостей в тексте.

Для решения этой проблемы были разработаны усовершенствованные версии RNN — LSTM (Long Short-Term Memory) и GRU (Gated Recurrent Unit). Они используют специальные механизмы (гейты), которые помогают эффективно хранить и передавать информацию на более дальние расстояния в тексте [3]. Это делает их полезными для задач, где требуется анализировать взаимосвязь слов в длинных документах, например, при обработке диалогов или при автоматическом резюмировании текста.

Несмотря на улучшения, ограничения таких моделей остаются. Они требуют значительных вычислительных ресурсов, особенно при обработке очень длинных текстов, и обучаются медленнее по сравнению с более современными архитектурами. Кроме того, даже LSTM и GRU имеют ограничения по длине контекста, который они могут эффективно

учитывать. В связи с этим на смену рекуррентным сетям в задачах анализа больших объемов текстовых данных пришли трансформеры, которые обеспечивают лучшее понимание контекста на глобальном уровне.

Сверточные нейронные сети (CNN) для текста применяются для выявления локальных зависимостей в текстовых данных. Первоначально разработанные для обработки изображений, они оказались эффективными и в анализе текстов благодаря способности извлекать ключевые признаки (n-граммы, фразы, шаблоны словосочетаний) с помощью сверточных фильтров [4].

Основной принцип работы CNN в обработке текста заключается в том, что специальные сверточные фильтры сканируют последовательности слов, выявляя характерные паттерны, например, фразы определенной длины, сочетания слов или важные лексические конструкции. Затем с помощью операций пулинга (pooling) сеть выбирает наиболее значимые признаки, что снижает размерность данных и увеличивает вычислительную эффективность.

Ключевое преимущество сверточных нейросетей – их высокая производительность при анализе локальных структур текста. Они работают быстрее, чем рекуррентные сети, и могут быть полезны в задачах, где важно выявление ключевых фраз, классификация текстов или анализ тональности.

Однако ограничения CNN связаны с их неспособностью учитывать глобальный контекст. Они хорошо определяют локальные зависимости, но могут упускать смысловые связи между удаленными словами и предложениями. Это делает их менее подходящими для обработки длинных текстов, где критически важно понимать взаимоотношения слов в разных частях документа. Именно поэтому сверточные сети часто используются в комбинации с другими архитектурами, например, с рекуррентными сетями или трансформерами, чтобы учесть как локальные, так и глобальные зависимости в тексте.

И, наконец, трансформеры и их производные (BERT, GPT, T5 и др.). Они стали революционным решением в обработке текстовых данных благодаря механизму self-attention (самовнимание), который позволяет учитывать контекст слов в тексте вне зависимости от их позиции [5]. В отличие от рекуррентных сетей, трансформеры анализируют весь текст одновременно, выявляя глубокие связи между словами и предложениями.

Одним из ключевых преимуществ трансформеров является их способность работать с длинными последовательностями текста, сохраняя понимание глобального контекста. Они успешно решают широкий спектр задач: от классификации текстов и выделения ключевой информации до генерации осмысленных ответов на вопросы.

Среди популярных моделей на основе трансформеров можно выделить:

- BERT (Bidirectional Encoder Representations from Transformers) – модель, ориентированная на анализ текста и извлечение информации. Она хорошо справляется с задачами классификации, поиска релевантных фрагментов и анализа смысловых связей.

- GPT (Generative Pre-trained Transformer) – генеративная модель, которая используется для создания связного текста, ответов на вопросы и диалоговых систем.

- T5 (Text-to-Text Transfer Transformer) – универсальная модель, которая преобразует любую задачу обработки естественного языка в формат «текст в текст» и успешно применяется для перевода, суммаризации и генерации текстов.

Несмотря на высокую точность и масштабируемость, трансформеры имеют существенные ограничения. Их обучение требует огромных вычислительных ресурсов, особенно для крупных моделей с миллиардами параметров. Это делает их труднодоступными для индивидуального обучения с нуля, однако использование предобученных моделей и их дообучение на специфических задачах позволяет эффективно применять трансформеры даже в образовательных проектах.

Благодаря своей универсальности и точности, трансформеры стали стандартом в современных задачах анализа текстов и широко применяются для создания интеллектуальных систем обработки информации, включая автоматическую генерацию учебных материалов, анализ студенческих работ и построение диалоговых помощников для обучения.

На основе рассмотренных архитектур нейронных сетей, трансформеры и их производные представляются наиболее эффективными для обработки текстовых данных. Это связано с следующими признаками:

1. Учет глобального контекста: Трансформеры используют механизм самовнимания (self-attention), позволяющий анализировать весь текст целиком и выявлять глубокие смысловые связи между словами и предложениями. Это особенно важно для понимания сложных языковых конструкций и длинных документов.

2. Высокая точность в задачах НЛП: Модели на основе трансформеров демонстрируют превосходные результаты в различных задачах обработки естественного языка, включая классификацию текстов, извлечение информации и генерацию связного текста.

3. Гибкость и масштабируемость: Трансформеры легко адаптируются к различным задачам и могут быть дообучены на специфических данных, что повышает их универсальность и применимость в разных областях.

В данной статье проведен сравнительный анализ различных архитектур нейросетей, применяемых для обработки текстовых данных. Было рассмотрено, как простые многослойные перцептроны, рекуррентные сети, сверточные модели и трансформеры справляются с задачами классификации, извлечения информации и генерации текстов.

Анализ показал, что трансформеры обладают наибольшей эффективностью благодаря механизму самовнимания, позволяющему учитывать глобальный контекст. Они демонстрируют высокую точность, хорошо масштабируются и подходят для большинства задач анализа текстов, включая обработку длинных документов. Однако их применение связано с высокими вычислительными затратами, что делает такие модели менее доступными для небольших проектов без специализированного оборудования.

При выборе архитектуры важно учитывать компромисс между вычислительной эффективностью, точностью и способностью обрабатывать длинные последовательности. Например, CNN можно использовать для быстрого анализа локальных признаков текста, RNN – для работы с последовательными данными, а BERT и GPT – для сложных задач понимания и генерации текста.

Полученные результаты и рекомендации помогут исследователям и разработчикам в области НЛП принять обоснованное решение при выборе нейросетевой архитектуры для конкретных задач, таких как анализ научных публикаций, создание образовательного контента или генерация ответов на вопросы.

СПИСОК ЛИТЕРАТУРЫ

1. Бережнов, Н.И. Совершенствование механизмов внимания для архитектуры трансформер в задачах повышения качества изображений / Н.И. Бережнов, А.А. Сирота // Компьютерная оптика. – 2024. – №5. – С. 726-734.
2. Алчинов, В.И. А55 Основы нейросетевого искусственного интеллекта. Краткий вводный курс : учеб.-метод. пособие / В.И. Алчинов, А.И. Иванов. – Пенза : Изд-во ПГУ, 2024. – 68 с.
3. Козлов, С.В. Анализ LSTM и GRU моделей для построения прогнозов временных рядов / С.В. Козлов, С.А. Седенков // International Journal of Open Information Technologies. – 2024. – Т. 12, №7. – С. 43-50.
4. Гайнетдинов, А.Ф. Исследование влияния трансформеров на улучшение генерации изображений / А.Ф. Гайнетдинов // Universum: технические науки. – 2024. – №4 (121). – С. 44-49.
5. Бобков, И.А. Моделирование неопределенности при помощи нейронных сетей / И.А. Бобков, А.А. Бурдина, А.А. Нехрест-Бобкова // Ars Administrandi. – 2023. – Т. 15, №1. – С. 45-59.

REFERENCES

1. Berezhnov, N.I. Improving attention mechanisms for the transformer architecture in image quality enhancement problems / N.I. Berezhnov, A.A. Sirota // Computer Optics. – 2024. – No. 5. – P. 726-734.
2. Alchinov, V.I. A55 Fundamentals of neural network artificial intelligence. A brief introductory course: textbook-method. manual / V.I. Alchinov, A.I. Ivanov. – Penza: Publishing house of PSU, 2024. – 68 p.
3. Kozlov, S.V. Analysis of LSTM and GRU models for constructing time series forecasts / S.V. Kozlov, S.A. Sedenkov // International Journal of Open Information Technologies. – 2024. – Vol. 12, No. 7. – P. 43-50.
4. Gainetdinov, A.F. Study of the influence of transformers on improving image generation / A.F. Gainetdinov // Universum: technical sciences. – 2024. – No. 4 (121). – P. 44-49.
5. Bobkov, I.A. Modeling uncertainty using neural networks / I.A. Bobkov, A.A. Burdina, A.A. Nekhrest-Bobkova // Ars Administrandi. – 2023. – Vol. 15, No. 1. – P. 45-59.