

Научная статья

УДК 004

DOI 10.25205/1818-7900-2024-22-4-49-61

Эффективность нейросетевых алгоритмов в автоматическом реферировании и суммаризации текста

Кирилл Вячеславович Ребенок

Московский финансово-юридический университет МФЮА,
Москва, Россия

rebenokkv@gmail.com, <https://orcid.org/0009-0003-2015-033X>

Аннотация

Статья посвящена анализу роли и эффективности нейросетевых алгоритмов в задачах автоматического реферирования и суммаризации текстов, которые являются ключевыми в области обработки естественного языка (NLP). Основная цель автоматического реферирования — извлечение и генерация важнейшей информации из текстов для обеспечения быстрого доступа к основному содержанию без необходимости читать весь документ. В статье рассматриваются основные проблемы, с которыми сталкиваются разработчики при реализации алгоритмов реферирования, включая понимание контекста, иронии, сохранение связности текста, адаптацию к разным языкам и стилям. Особое внимание уделяется нейросетевым моделям, таким как Transformer, BERT и GPT, которые благодаря своей способности обучаться на больших объемах данных показали выдающуюся эффективность в автоматическом реферировании текстов. Статья также освещает вклад ведущих ученых в области глубокого обучения и анализирует методы, лежащие в основе современных алгоритмов NLP, подчеркивая значимость непрерывного технологического прогресса в улучшении качества реферирования и доступности информации. Статья будет интересна широкому кругу читателей, включая исследователей в области искусственного интеллекта и NLP, разработчиков программного обеспечения, занимающихся автоматизацией обработки текстов, а также специалистов в областях, где требуется быстрая обработка и анализ больших объемов текстовой информации, таких как юридическая практика, медицинская диагностика и научные исследования. Кроме того, материал статьи будет полезен преподавателям и студентам, изучающим технологии обработки данных и искусственного интеллекта, предоставляя им актуальные примеры применения теоретических знаний в практических проектах.

Ключевые слова

естественный язык, NLP, нейросетевые алгоритмы, метрики, автоматическое реферирование, суммаризация текста

Для цитирования

Ребенок К. В. Эффективность нейросетевых алгоритмов в автоматическом реферировании и суммаризации текста // Вестник НГУ. Серия: Информационные технологии. 2024. Т. 22, № 4. С. 49–61. DOI 10.25205/1818-7900-2024-22-4-49-61

© Ребенок К. В., 2024

Efficiency of Neural Network Algorithms in Automatic Abstracting and Summarization Text

Kirill V. Rebenok

Moscow University of Finance and Law MFUA,
Moscow, Russian Federation

rebenokkv@gmail.com, <https://orcid.org/0009-0003-2015-033X>

Abstract

The article is devoted to the analysis of the role and efficiency of neural network algorithms in the tasks of automatic abstracting and summarization of texts, which are key in the field of natural language processing (NLP). The main goal of automatic abstracting is to extract and generate essential information from texts to provide quick access to the main content without having to read the whole document. The paper discusses the main challenges faced by developers in implementing abstracting algorithms, including understanding context, irony, maintaining text cohesion, and adapting to different languages and styles. Special attention is given to neural network models such as Transformer, BERT, and GPT, which have shown outstanding performance in automatic text abstracting due to their ability to learn on large amounts of data. The article also highlights the contributions of leading researchers in the field of deep learning and analyzes the methods underlying state-of-the-art NLP algorithms, highlighting the importance of continuous technological progress in improving abstracting quality and information accessibility. The article will be of interest to a wide range of readers, including researchers in the field of artificial intelligence and NLP, software developers engaged in automation of text processing, as well as specialists in areas where fast processing and analysis of large amounts of textual information is required, such as legal practice, medical diagnostics and scientific research. In addition, the material of the article will be useful for teachers and students studying data processing and artificial intelligence technologies, providing them with actual examples of applying theoretical knowledge in practical projects.

Keywords

natural language, NLP, neural network algorithms, metrics, automatic abstracting, text summarization

For citation

Rebenok K. V. Efficiency of neural network algorithms in automatic abstracting and summarization text. *Vestnik NSU. Series: Information Technologies*, 2024, vol. 22, no. 4, pp. 49–61 (in Russ.) DOI 10.25205/1818-7900-2024-22-4-49-61

Введение

Автоматическое реферирование и суммирование текстов – важнейшая из задач в области обработки естественного языка (NLP), призванная упростить доступ к информации и улучшить ее восприятие. В условиях развития Интернета и цифровых технологий объем доступной текстовой информации стремительно растет, и ее полноценное изучение и анализ в ручном режиме становится невозможным. В этой связи автоматическое реферирование текста – это процесс создания краткого и сжатого резюме длинного документа с сохранением его основного содержания и ключевых идей.

Цель автоматического реферирования заключается в том, чтобы извлекать или генерировать наиболее важную информацию из текста таким образом, чтобы конечный пользователь мог быстро получить представление о его содержании без необходимости читать весь документ. В особенности это актуально для новостных статей, научных публикаций, юридических документов и любых других областей, где требуется быстрый доступ к сжатой форме информации.

Многие разработчики и пользователи технологий NLP нередко сталкиваются с проблемами при автоматическом реферировании и резюмировании текстов, включая понимание контекста и иронии, сохранение связности и логичности изложения, адаптацию к различным языкам и стилям письма.

Нейросетевые модели играют одну из ключевых ролей в решении этих задач благодаря своей способности обучаться на больших объемах данных и выявлять сложные зависимости

в тексте. Модели глубокого обучения, такие как Transformer, BERT, GPT и их производные, продемонстрировали потрясающую производительность в задачах автоматического реферирования, научившись генерировать краткие и осмысленные резюме текстов на основе обучающих примеров. Технологический прогресс непрерывно стремится повысить качество реферирования и сделать автоматическое реферирование максимально доступным и эффективным инструментом для обработки информации в различных областях.

Проблема автоматического реферирования и резюмирования текстов активно изучается множеством исследователей, так как она имеет ключевое значение для развития области обработки естественного языка (NLP). В их числе следует выделить таких ученых, как Дж. Хинтон, Я. ЛеКун и Й. Бенджио, внесших значительный вклад в развитие глубокого обучения и нейронных сетей. Исследования И. Суцкевера, А. Крижевски и Д. Сильвера позволяют лучше понять методы нейросетевого обучения, лежащие в основе современных алгоритмов NLP. Важное место в области машинного обучения занимают работы таких авторов, как С. Рудер и Т. Вольф, развивающих и адаптирующих такие алгоритмы, как Transformer и BERT, с целью эффективного извлечения и создания кратких изложений текста. Примечателен также вклад Л. Куна и его работа над моделями GPT, показавшая отличные результаты в генеративных задачах NLP, включая резюмирование.

В отечественной научной среде проблема автоматического реферирования и резюмирования текстов не получила широкого распространения, несмотря на мировой интерес к этой области. Недостаток ресурсов и доступа к большим данным, а также трудности, связанные с особенностями русского языка, тормозят прогресс в разработке соответствующих алгоритмов. В то же время потенциал дальнейшего развития все же существует благодаря наличию квалифицированных специалистов и растущему интересу к технологиям искусственного интеллекта. Упор на развитие специализированных образовательных программ и укрепление связей между академическими, исследовательскими и коммерческими организациями может стать ключом к преодолению существующих барьеров и стимулированию развития этой области.

Материалы и методы

В рамках данного исследования были использованы актуальные научные публикации, освещающие вопросы нейросетевых алгоритмов, больших данных, машинного обучения, реферирования и суммаризации текстов, методов интеллектуального анализа данных. В этой статье активно применялись различные методологические подходы: монографический анализ для детального исследования темы, оценочный подход для анализа и интерпретации данных, а также метод рефлексии, позволяющий глубоко осмыслить и критически оценить полученные результаты.

Результаты исследования

Анализ моделей обработки естественного языка (NLP) имеет решающее значение для оценки их эффективности, удобства использования и надежности в реальных приложениях. Метрики, включая BLEU, ROUGE, METEOR и BERTScore, играют ключевую роль в этом процессе, предоставляя количественные показатели эффективности модели. Метрики помогают оценить, насколько эффективно модель справляется с задачами перевода, обобщения, генерации или понимания текста в сравнении с человеческими оценками или эталонными данными. Такая оценка жизненно важна не только для точной настройки и улучшения моделей, но и для обеспечения их соответствия необходимым стандартам, необходимым для развертывания в чувствительных приложениях, таких как медицинская диагностика, юридический анализ или автоматизация обслуживания клиентов.

Обсуждение результатов

За последние годы текстовые генеративные модели ИИ добились значительных успехов в решении задач обработки естественного языка, таких как перевод, обобщение текста и создание диалогов. Они способны генерировать текст, зачастую неотличимого от человеческого, что делает их все более популярными в различных отраслях, включая обслуживание клиентов, создание контента и анализ данных. И хотя эти модели могут быть невероятно мощными и полезными, они также могут выдавать неожиданные или даже пагубные результаты, в связи с чем за ними необходимо внимательно следить.

При автоматическом реферировании и суммировании текстов используются различные нейросетевые алгоритмы, каждый из которых обладает уникальными особенностями и принципами работы. Традиционно эти алгоритмы можно разделить на генеративные и извлекающие методы резюмирования.

Извлекающий метод резюмирования работает путем выделения и копирования ключевых фраз или предложений из исходного текста для формирования сжатого содержания. При таком подходе исходный текст не изменяется, а фильтруется и изымается наиболее значимая его часть. При этом методы извлечения основываются на анализе важности слов и предложений в документе с помощью таких метрик, как частота слов, положение предложения в тексте и связность предложений [1].

Генеративные методы обобщения, напротив, создают новые предложения, не обязательно присутствующие в исходном тексте, чтобы отразить суть содержания. При использовании этих методов глубокое обучение позволяет генерировать связные и последовательные сводки, которые могут включать перефразирование или обобщение информации. Модели генеративного типа требуют более сложных алгоритмов и значительных вычислительных ресурсов, но способны создавать гораздо качественные и естественные обобщения [2].

Определение эффективности нейросетевых алгоритмов в задаче автоматического реферирования и резюмирования текстов включает в себя оценку различных аспектов, таких как точность, качество генерируемых обобщений, их релевантность и связность, а также способность алгоритмов адаптироваться к различным текстовым данным.

Рассмотрим более подробно каждую из метрик, которые применяются для автоматического реферирования и суммирования текста.

Это набор показателей, используемых для оценки автоматического обобщения и машинного перевода. Он сравнивает автоматически созданное резюме или перевод с набором справочных резюме (обычно написанных человеком). ROUGE [3] измеряет качество резюме путем подсчета количества перекрывающихся единиц, таких как n -граммы¹, последовательности слов и пары слов между текстом, созданным моделью, и справочными текстами.

Наиболее распространенные варианты ROUGE:

- ROUGE-N, фокусирующийся на n -граммах (фразах из N слов);
- ROUGE-1 и ROUGE-2 наиболее распространенные, которые фокусируются на униграммах² и биграммах³ соответственно;

¹ N -граммы – это последовательности из n слов, извлеченных из текста. Например, в предложении «Я иду домой» биграммы (2-граммы) будут «Я иду» и «иду домой». N -граммы используются для оценки степени сходства между двумя текстами на уровне словесных последовательностей.

² Униграммы – это одиночные слова, извлеченные из текста. В контексте обработки текста и анализа данных униграммы представляют собой самые простые элементы, используемые для анализа и сравнения текстов. Они могут быть использованы для статистического анализа частотности слов, оценки сходства текстов и других задач, связанных с языковыми моделями.

³ Биграммы – это последовательности из двух слов, следующих друг за другом в тексте. Они являются основным инструментом в текстовом анализе и помогают уловить связи между словами, что важно для понимания структуры предложений и для создания статистических моделей языка.

- ROUGE-L, основанный на самой длинной общей подпоследовательности (LCS), учитывающий сходство структуры на уровне предложений и автоматически определяющий самые длинные последовательные n-граммы⁴.

Рассмотрим пример применения метрики ROUGE (табл. 1).

Таблица 1

ROUGE: практическая реализация (составлено автором)

Table 1

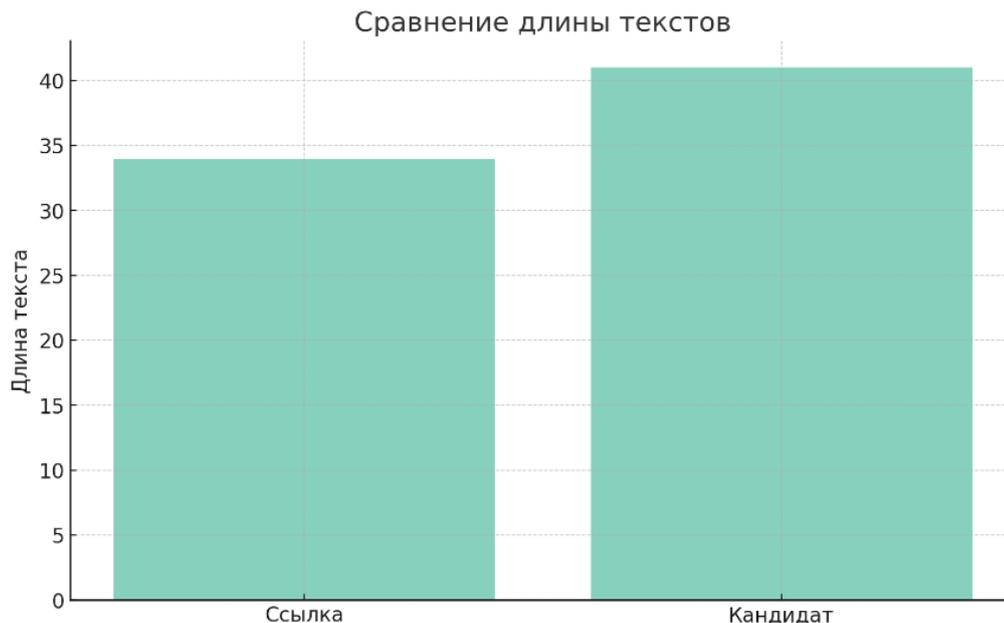
ROUGE: practical implementation (compiled by the author)

Тип текста	Текст	Ключевые аспекты	Сравнение с эталонным суммарием
Исходный текст	«Глобальное потепление вызывает серьёзные изменения в климате Земли, приводя к увеличению частоты и интенсивности экстремальных погодных явлений. Эти изменения угрожают сельскому хозяйству, водным ресурсам и здоровью человека, требуя незамедлительных действий по сокращению выбросов парниковых газов и адаптации к уже неизбежным последствиям»	Глобальное потепление, климатические изменения, угрозы экосистемам, сельскому хозяйству, водным ресурсам, здоровью, требование действий	Охватывает весь спектр проблем, связанных с глобальным потеплением
Эталонное резюме	«Глобальное потепление угрожает экосистемам, увеличивая экстремальные погодные явления и требует срочных мер по сокращению выбросов и адаптации»	Угроза экосистемам, увеличение экстремальных явлений, необходимость срочных мер по сокращению выбросов и адаптации	Суммирует основные точки исходного текста, упрощая детали
Сгенерированное суммарие	«Увеличение частоты экстремальных погодных явлений из-за глобального потепления требует действий для снижения выбросов парниковых газов»	Увеличение экстремальных погодных явлений, необходимость снижения выбросов парниковых газов	Сфокусировано на погодных явлениях и снижении выбросов

Реализация данного примера на языке Python будет выглядеть, как изображено на рисунке.

BLEU – метрика, разработанная исследователями IBM для оценки текста, которая позволяет оценить точность перевода путем измерения совпадения n-грамм между машинногенерированным текстом и набором высококачественных эталонных переводов. Основное внимание уделяется точности. BLEU славится своей простотой и эффективностью, что делает его эталоном в области машинного перевода. Вместе с тем он оценивает лексическое сходство на поверхностном уровне, часто упуская из виду более глубокие семантические и контекстуальные нюансы языка [4–6].

⁴ ROUGE and BLEU scores for NLP model evaluation. URL: <https://clementbm.github.io/theory/2021/12/23/rouge-bleu-scores.html> (дата обращения: 12.03.2024).



Сравнение длины текстов для примера использования метрики ROUGE
Comparison of text lengths for an example using the ROUGE metric

Среди основных недостатков использования таких метрик, как BLEU или ROUGE, можно назвать тот факт, что эффективность работы моделей генерации текстов зависит от точных совпадений. Возможно, точные совпадения важны для таких сценариев использования, как машинный перевод, но для генеративных моделей ИИ, которые пытаются генерировать осмысленные и похожие тексты на основе имеющегося массива данных, точные совпадения могут быть не очень верными.

Метрика METEOR, разработанная для более глубокой оценки машинного перевода, призвана устранить такие недостатки BLEU, как недооценка семантической связности текста. В отличие от BLEU, METEOR учитывает не только точные совпадения слов, но также включает основы и синонимы для оценки переводов, что позволяет охватить более широкий диапазон лингвистических сходств [7]. METEOR позволяет точно и эффективно оценивать качество переводов текстов. Он учитывает не только точность перевода, но и то, насколько легко запомнить переведённый текст, добавляя штрафы за изменения в порядке слов. METEOR отличается высокой согласованностью с оценками, которые дают люди, особенно при анализе отдельных предложений. Это делает его подходящим для тщательной оценки качества переводов. Однако стоит отметить, что METEOR более сложен и требует больше ресурсов для расчётов по сравнению с более простыми методами, такими как BLEU [8; 9].

Для реализации метрики METEOR в Python можно использовать библиотеку NLTK, которая, помимо прочего, предоставляет инструменты для работы с метрикой BLEU. В отличие от BLEU, стандартная поддержка METEOR в NLTK отсутствует, что требует более сложной реализации с использованием внешних инструментов или создания собственной функции расчёта.

METEOR, как и BLEU, предназначена для оценки качества машинного перевода, сравнивая сгенерированный текст с одним или несколькими эталонными переводами. Она учитывает точность и полноту, а также синонимичность и порядок слов. Тем не менее можно адаптировать подходы к расчёту METEOR для оценки суммаризации текстов. Прямая реализация METEOR на Python может быть достаточно сложной из-за необходимости учитывать синони-

мы, морфологический анализ и порядок слов. В качестве альтернативы можно использовать готовые реализации или обращаться к инструментам, таким как Meteor Universal Tool, который предоставляется в виде Java-приложения.

BERTScore – это новая метрика для оценки качества созданных текстов, которая опирается на передовые технологии в области искусственного интеллекта. Она использует модель глубокого обучения под названием BERT для анализа текстов. Модель способна учитывать контекст каждого слова в тексте, что помогает оценить, насколько хорошо сгенерированный текст соответствует оригинальному или эталонному тексту. BERTScore не просто измеряет поверхностное совпадение слов, но и анализирует глубокое семантическое сходство, используя метод подсчёта косинусного сходства, что позволяет более точно оценить качество текста, учитывая его смысловое содержание [10]. Такой подход позволяет ИТ-специалистам оценивать качество создания текста с акцентом на семантическое содержание и контекст, что делает его более чувствительным к смыслу, передаваемому в тексте. Несмотря на то что BERTScore предлагает более детальную оценку, чем традиционные метрики, основанные на перекрытии, ее использование требует больших вычислительных затрат и ресурсов, поскольку она опирается на большие, предварительно обученные языковые модели.

Выбор наиболее эффективной метрики зависит от специфических целей оценки суммаризации. Если важно оценить точность воспроизведения конкретных фактов и данных, ROUGE может быть наиболее подходящей. Для более общего анализа качества перефразирования и семантической близости текстов лучше подойдет METEOR и BERTScore.

Преимущества и недостатки рассмотренных метрик представлены в табл. 2. Чтобы контролировать работу генеративных моделей, важно применять комплексный подход. Например, когда есть эталонный текст для сравнения, можно использовать BLEU для оценки точности перевода или ROUGE для измерения полноты.

Meteor, с другой стороны, учитывает и точность, и полноту, показывает хорошую корреляцию с результатами человеческих оценок как на уровне предложений, так и на уровне сегментов. BERTScore полезен для оценки семантической близости между генерируемым и эталонным текстом с помощью контекстуализированных вкраплений слов⁵.

Недавнее исследование метрик BLEU, METEOR и BERTScore показало, что они не всегда эффективно различают критические и некритические ошибки перевода, особенно когда ошибка перевода изменяет сентимент⁶ сообщения⁷ [11; 12]. Это показывает, что при оценке качества перевода важно использовать комплексный подход и сочетать различные метрики для получения наиболее полной картины. С учетом этих факторов, важно подходить к выбору метрик для оценки качества суммаризации и машинного перевода с учетом специфики задачи и ограничений каждой метрики. При обработке естественного языка (NLP) ключевую роль играют методы, основанные на вычислении векторных представлений (embeddings). С их помощью слова, предложения или документы преобразуются в векторы чисел, что делает их пригодными для компьютерной обработки.

Нейросетевые алгоритмы суммаризации текстов находят широкое применение в различных отраслях, значительно увеличивая эффективность работы специалистов и делая информацию более доступной для общественности. Так, в юридической практике суммаризация текстов облегчает анализ и обработку большого количества юридических документов, таких

⁵ Evaluating NLP Models: A Comprehensive Guide to ROUGE, BLEU, METEOR, and BERTScore Metrics. URL: <https://plainenglish.io/community/evaluating-nlp-models-a-comprehensive-guide-to-rouge-bleu-meteor-and-bertscore-metrics-d0f1b1> (дата обращения: 12.03.2024); Tekgul H. Monitoring Text-Based Generative AI Models Using Metrics Like Bleu Score. URL: <https://arize.com/blog-course/generative-ai-metrics-bleu-score/> (дата обращения: 12.03.2024).

⁶ Сентимент сообщения, или анализ тональности текста, – это процесс определения эмоциональной окраски текста.

⁷ BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text. URL: <https://ar5iv.labs.arxiv.org/html/2109.14250> (дата обращения: 12.03.2024).

Таблица 2

Преимущества и недостатки метрик для суммаризации и реферирования текста
(составлено автором)

Table 2

Advantages and disadvantages of metrics for text summarization and abstracting
(compiled by author)

Метрика	Преимущества	Недостатки
BLEU	Простая и широко используемая. Хорошо подходит для оценки точности на уровне слов и фраз	Не учитывает семантическую связность и может приводить к искаженной оценке качества при слишком буквальном переводе
ROUGE	Учитывает как точность, так и полноту, предоставляя более сбалансированную оценку	Может быть менее эффективной для языков с сложной структурой из-за фокуса на совпадении слов и фраз
METEOR	Более сбалансированная метрика, включающая как точность, так и полноту, с учетом синонимов и парадигм	Сложнее в реализации и вычислении, чем BLEU или ROUGE
BERTScore	Использует контекстные вложения для оценки семантической близости, захватывая более тонкие нюансы языка	Требует больших вычислительных ресурсов и может быть зависимой от качества предварительно обученной модели BERT
Self-BLEU	Помогает оценить разнообразие в сгенерированном тексте, предотвращая избыточное повторение	Не учитывает качество содержания; фокусируется только на разнообразии
WMD	Эффективно оценивает семантическую близость на основе расстояния между словами в векторном пространстве	Может быть вычислительно затратной для длинных текстов и менее точной для очень коротких фраз

как законы, судебные решения и договоры, что позволяет юристам экономить время на подготовку к делам, улучшая понимание существующих прецедентов и законодательных требований. А в академическом мире алгоритмы суммаризации текстов значительно упрощают литературный обзор и анализ научных публикаций. Исследователи могут быстрее ознакомиться с последними достижениями в своей области, выявляя ключевые идеи и результаты из больших объемов научной литературы.

Техники резюмирования текстов делятся на две категории: экстрактивную и абстрактивную. Каждая из этих техник используется для создания краткого изложения длинного текста, но они делают это разными способами.

Экстрактивное резюмирование — это процесс выбора ключевых предложений или фраз непосредственно из исходного текста и их компиляции для создания резюме. Эта техника не вносит изменений в текст: она просто извлекает наиболее значимые части. Основные характеристики:

- система определяет и извлекает наиболее информативные предложения или фразы из текста на основе различных метрик, таких как частотность слов, важность темы и так далее;
- извлеченные предложения остаются неизменными, сохраняя оригинальный стиль и структуру автора;
- так как не требуется генерация нового текста, экстрактивное резюмирование может быть более простым и быстрым в реализации.

Таблица 3

Table 3

Оценка эффективности различных языковых моделей в задачах резюмирования и суммаризации текста*

Evaluating the effectiveness of different language models in text summarization and summarization tasks

Название модели	Разработчик	Год выпуска	Количество параметров	Архитектура	Преимущества	Недостатки
BERT	Google AI	2018	110 миллионов	Transformer	Хорошо справляется с контекстом	Требует тонкой настройки
T5	Google AI	2019	11 миллиардов	Transformer	Гибкость в задачах	Ограничен в длине текста
RoBERTa	Facebook AI	2019	355 миллионов	Transformer	Улучшенная обработка текста	Требует больших вычислительных ресурсов
XLNet	Google/CMU	2019	340 миллионов	Transformer	Отличные результаты на различных задачах	Сложность в использовании
GPT-2	OpenAI	2019	1.5 миллиарда	Transformer	Хорошая генерация текста	Ограничения по контексту
GPT-3	OpenAI	2020	175 миллиардов	Transformer	Мощная генерация текста	Высокая стоимость
GPT-4	OpenAI	2023	более 100 миллиардов	Transformer	Улучшенное понимание и генерация текста	Требует значительных вычислительных ресурсов

* Составлена автором на основе [13–15].

Абстрактное резюмирование переформулирует исходный текст, создавая новые предложения, которые могут не встречаться напрямую в исходном материале. Этот метод часто считается более сложным и продвинутым, так как требует глубокого понимания текста и способности к его творческой переработке. Основные характеристики:

- модель создает новые предложения, которые резюмируют оригинальный контент, используя передовые NLP-модели;
- лучше справляются с передачей основных идей текста в сжатой форме, поскольку они не ограничены только тем, что написано в исходнике;
- необходимы развитые алгоритмы понимания языка и генерации текста, такие как трансформеры и модели на основе искусственного интеллекта.

В табл. 3 представлены основные характеристики и возможности нескольких передовых языковых моделей, используемых для резюмирования текстов.

Изучение сравнительной таблицы языковых моделей для резюмирования и суммаризации текстов выявляет ключевые различия и потенциальные сферы применения каждой модели:

- BERT идеально подходит для задач, где требуется глубокое понимание контекста и точность в экстрактивном резюмировании;
- GPT-4 выделяется в абстрактном резюмировании и суммаризации, предлагая высококачественную генерацию текста, хотя и с высокими требованиями к ресурсам;
- T5 обеспечивает выдающуюся гибкость и адаптируемость, что делает её подходящей для широкого спектра задач резюмирования/суммаризации и других задач NLP;
- BART эффективно справляется с абстрактным резюмированием/суммаризацией благодаря своей способности к восстановлению и переформулировке текста.

Выбор подходящей модели зависит от специфических требований проекта, включая язык, на котором представлен контент, требуемую скорость обработки, доступные вычислительные ресурсы и предпочтения в стилях резюмирования/суммаризации.

Заключение

Использование нейросетевых алгоритмов в автоматическом реферировании и резюмировании текстов является перспективным направлением развития систем обработки информации. Нейросети открывают новые возможности для работы с большими объемами данных, повышают эффективность поиска и анализа информации, а также имеют широкий спектр практического применения. Существенно продвинуться в этой области позволяет использование таких технологий, как векторные представления слов (Word2Vec, GloVe), трансформаторы (BERT, GPT) и различные нейросетевые архитектуры. Рассмотренные модели способны обрабатывать и обобщать информацию, сохраняя при этом семантическую целостность и релевантность содержания, что особенно важно в областях, требующих быстрой обработки больших объемов данных, таких как новостные публикации или анализ научных текстов.

Из-за сложности языка, в том числе особенностей контекста, идиоматических выражений и культурных аллюзий, возникают проблемы с точной интерпретацией моделей и определением качественных показателей оценки. Традиционные метрики фокусируются на поверхностных характеристиках текста, таких как совпадение слов, которые могут не полностью отражать способность модели понимать или генерировать семантически и синтаксически корректный язык. Использование эталонных наборов данных для оценки может привести к предвзятости или ограничить область оценки, поскольку эти наборы данных могут не отражать всего многообразия употребления языка в реальном мире.

Перспективы нейросетевых алгоритмов в области автоматического реферирования весьма многообещающи. Непрерывный рост вычислительных мощностей, создание новых алгоритмов обучения и оптимизации моделей, а также улучшение предварительной обработки

и последующего анализа текстовых данных могут значительно повысить точность и адаптивность систем реферирования. Кроме того, применение технологий искусственного интеллекта для анализа эмоциональной окраски, контекста и стиля текста может открыть новые горизонты для разработки более совершенных и ориентированных на человека систем обработки естественного языка.

Список литературы

1. **Divakar Y., Jalpa D., Arun K. Y.** Automatic Text Summarization Methods: A Comprehensive Review. 2020. <https://doi.org/10.48550/arXiv.2204.01849>
2. **Salchner M. F., Adam A.** A Survey of Automatic Text Summarization Using Graph Neural Networks // In Proceedings of the 29th International Conference on Computational Linguistics. Gyeongju, Republic of Korea. International Committee on Computational Linguistics. 2022. P. 6139–6150.
3. **Vamvas J., Domhan T., Trenous S., Sennrich R., Hasler E.** Trained MT Metrics Learn to Cope with Machine-translated References // Conference: Proceedings of the Eighth Conference on Machine Translation. 2023. <https://doi.org/10.18653/v1/2023.wmt-1.95>.
4. **Mathur N., Baldwin T., Cohn T.** Tangled up in BLEU: Reevaluating the Evaluation of Automatic Machine Translation Evaluation Metrics // Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 4984–4997.
5. **Reiter E.** A Structured Review of the Validity of BLEU // Computational Linguistics. 2018. № 44 (3). P. 393–401.
6. **Tianyi Z., Kishore V., Wu F., Weinberger K. Q., Artzi Y.** BERTScore: Evaluating Text Generation with BERT. ArXiv abs/1904.09675. 2019.
7. **Guo Y., Hu J.** Meteor++ 2.0: Adopt syntactic level paraphrase knowledge into machine translation evaluation // In Proceedings of the Fourth Conference on Machine Translation. 2019. Vol. 2. P. 501–506. <https://doi.org/10.18653/v1/W19-5357>
8. **Ayub S.A., Gaol F.L., Matsuo T.** A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization. IEEE Access. 2021. № 9. P. 13248–13265. <https://doi.org/10.1109/ACCESS.2021.3052783>
9. **Al E. W., Awajan A. A.** SemG-TS: Abstractive Arabic Text Summarization Using Semantic Graph Embedding // Mathematics. 2022. № 10 (18). P. 3225. <https://doi.org/10.3390/math10183225>
10. **Tianyi Zhang T., Kishore V., Wu F., Weinberger K.Q., Artzi Y.** BERTSCORE: Evaluating Text Generation with BERT. Department of Computer Science and Cornell Tech, Cornell University. 2019.
11. **Saadany H., Orasan C.** BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text // Conference: TRITON (TRanslation and Interpreting Technology Online). 2021. https://doi.org/10.26615/978-954-452-071-7_006
12. **Sudoh K., Takahashi K., Nakamura S.** Is this translation error critical?: Classification-based human and automatic machine translation evaluation focusing on critical errors // Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval). 2021. P. 46–55.
13. **Siddhant A., Johnson M., Tsai T., Ari N.** Evaluating the Cross-Lingual Effectiveness of Massively Multilingual Neural Machine Translation // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. № 34(05). P. 8854–8861. <https://doi.org/10.1609/aaai.v34i05.6414>.
14. **Lin J., Nogueira R., Yates A.** Pretrained Transformers for Text Ranking: BERT and Beyond // Synthesis Lectures on Human Language Technologies. 2021. № 14 (4). P. 1–325. <https://doi.org/10.2200/S01123ED1V01Y202108HLT053>.
15. **Chistyakova K., Kazakova T.** Grammar in Language Models: BERT Study // National research university higher school of economics. 2023. № 115.

References

1. **Divakar Y., Jalpa D., Arun K. Y.** Automatic Text Summarization Methods: A Comprehensive Review. 2020. <https://doi.org/10.48550/arXiv.2204.01849>
2. **Salchner M.F., Adam A.** A Survey of Automatic Text Summarization Using Graph Neural Networks. In: *Proceedings of the 29th International Conference on Computational Linguistics*. Gyeongju, Republic of Korea. International Committee on Computational Linguistics, 2022, pp. 6139–6150.
3. **Vamvas J., Domhan T., Trenous S., Sennrich R., Hasler E.** Trained MT Metrics Learn to Cope with Machine-translated References. *Conference: Proceedings of the Eighth Conference on Machine Translation*. 2023. <https://doi.org/10.18653/v1/2023.wmt-1.95>.
4. **Mathur N., Baldwin T., Cohn T.** Tangled up in BLEU: Reevaluating the Evaluation of Automatic Machine Translation Evaluation Metrics. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 2020, pp. 4984–4997.
5. **Reiter E.** A Structured Review of the Validity of BLEU. *Computational Linguistics*, 2018, № 44 (3), pp. 393–401.
6. **Tianyi Z., Kishore V., Wu F., Weinberger K. Q., Artzi Y.** BERTScore: Evaluating Text Generation with BERT. ArXiv abs/1904.09675. 2019.
7. **Guo Y., Hu J.** Meteor++ 2.0: Adopt syntactic level paraphrase knowledge into machine translation evaluation. In: *Proceedings of the Fourth Conference on Machine Translation*, 2019, vol. 2, pp. 501–506. <https://doi.org/10.18653/v1/W19-5357>
8. **Ayub S. A., Gaol F. L., Matsuo T.** A Survey of the State-of-the-Art Models in Neural Abstractive Text Summarization. *IEEE Access*, 2021, № 9, pp. 13248–13265. <https://doi.org/10.1109/ACCESS.2021.3052783>
9. **AIE. W., Awajan A. A.** SemG-TS: Abstractive Arabic Text Summarization Using Semantic Graph Embedding. *Mathematics*, 2022, № 10 (18), p. 3225. <https://doi.org/10.3390/math10183225>
10. **Tianyi Zhang T., Kishore V., Wu F., Weinberger K. Q., Artzi Y.** BERTSCORE: Evaluating Text Generation with BERT. Department of Computer Science and Cornell Tech, Cornell University, 2019.
11. **Saadany H., Orasan C.** BLEU, METEOR, BERTScore: Evaluation of Metrics Performance in Assessing Critical Translation Errors in Sentiment-oriented Text. *Conference: TRITON (Translation and Interpreting Technology Online)*, 2021. https://doi.org/10.26615/978-954-452-071-7_006
12. **Sudoh K., Takahashi K., Nakamura S.** Is this translation error critical?: Classification-based human and automatic machine translation evaluation focusing on critical errors. In: *Proceedings of the Workshop on Human Evaluation of NLP Systems (HumEval)*, 2021, pp. 46–55.
13. **Siddhant A., Johnson M., Tsai T., Ari N.** Evaluating the Cross-Lingual Effectiveness of Massively Multilingual Neural Machine Translation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, № 34 (5), pp. 8854–8861. <https://doi.org/10.1609/aaai.v34i05.6414>.
14. **Lin J., Nogueira R., Yates A.** Pretrained Transformers for Text Ranking: BERT and Beyond. *Synthesis Lectures on Human Language Technologies*, 2021, № 14 (4), pp. 1–325. <https://doi.org/10.2200/S01123ED1V01Y202108HLT053>.
15. **Chistyakova K., Kazakova T.** Grammar in Language Models: BERT Study. *National research university higher school of economics*, 2023, № 115.

Сведения об авторах

Ребенок Кирилл Вячеславович, аспирант Московского финансово-юридического университета МФЮА

Information about the Author

Kirill V. Rebenok, Postgraduate Student of the Moscow University of Finance and Law MFUA,
Moscow, Russian Federation

*Статья поступила в редакцию 10.04.2024;
одобрена после рецензирования 07.11.2024; принята к публикации 07.11.2024*

*The article was submitted 10.04.2024;
approved after reviewing 07.11.2024; accepted for publication 07.11.2024*