ВЕСТНИК ТОМСКОГО ГОСУДАРСТВЕННОГО УНИВЕРСИТЕТА

2022 Управление, вычислительная техника и информатика № 59

Tomsk State University Journal of Control and Computer Science

Научная статья
УДК 81’322.2
doi: 10.17223/19988605/59/11

Подход к преобразованию обучающей выборки для повышения качества генерации заголовков научных текстов

Анна Валерьевна Глазкова

Тюменский государственный университет, Тюмень, Россия, a.v.glazkova@utmn.ru

Аннотация. Предлагается подход к улучшению качества генерации заголовков, основанный на ранжировании примеров обучающей выборки в соответствии со значениями метрики ROUGE-1, вычисленных для текстов и заголовков, фильтрации данных и генерации искусственных обучающих примеров. Предложенный подход, протестированный на примере нейросетевой модели BART, показал улучшение качества генерации заголовков на материале двух англоязычных корпусов.

Ключевые слова: обработка естественного языка; автоматическое реферирование; анализ научных текстов, генерация заголовков; BART

Благодарности: Работа выполнена в рамках программы «Михаил Ломоносов» Германской службы академических обменов (DAAD) и Минобрнауки (№ 121040100251-1).

Для цитирования: Глазкова А.В. Подход к преобразованию обучающей выборки для повышения качества генерации заголовков научных текстов // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2022. № 59. С. 99–107. doi: 10.17223/19988605/59/11

Original article
doi: 10.17223/19988605/59/11

Approach to transforming training data for improving the title generation performance for scientific texts

Anna V. Glazkova

University of Tyumen, Tyumen, Russian Federation, a.v.glazkova@utmn.ru

Abstract. Due to the significant increase in the availability of scientific resources and their expansion, the analysis and systematization of scientific documents become an important task of natural language processing. Scientific articles contain much significant diverse information. Besides, their amount is constantly increasing, and tracking actual scientific publications takes a lot of time. Reducing the number of viewed documents and their generalization is possible using special tools for automatic text processing, including text classification, information extraction, and text summarization.

As regards the summarization of scientific documents, one of the particular problems is the generation of the title for the scientific paper. Taking into account the large volumes of scientific resources, the title is especially significant. The title accuracy affects the visibility of the paper by the scientific community and therefore the number of prospective readers. Moreover, some recent studies showed that the quality of the paper title influences the number of citations. Despite this, the authors often spend not enough time creating a good title, which makes it non-informative and non-reflecting the content of the article. To overcome this weakness, the methods of automatic title generation for scientific texts can be developed and used.

In this work, we propose an approach to improving the quality of title generation for scientific texts. The proposed approach uses training data filtering and generates new training examples. We consider the following steps: 1) determining recall-oriented *ROUGE*-1 scores between titles and source texts from the training set. These scores show how many words from the title came from the text. Thus, we can conclude the content correspondence between the title and the source text; 2) ranking examples of the training sample by the recall-oriented scores; 3) filtering examples having scores less than the threshold value k ($k \in [0; 1)$); 4) training model for title generation on the filtered training sample; 5) enriching the filtered training sample to the original size with the pseudo examples generated from the trained model. These examples are generated only for examples removed in the previous step.

The approach was tested on two English corpora of scientific texts (SciTLDR and arXiv). We used scientific abstracts as a source for text summarization. We evaluated the values of k in the range from 0,3 to 0,9 in increments of 0,1. In most cases, the results showed that the use of a training sample consisting of filtered and pseudo examples increases the performance of the title generation in comparison with the generation using the original training sample. In our experiments, the most preferred values of the threshold k were 0,7 and 0,8. Experiments were conducted using the BART-base model.

Keywords: natural language processing; automatic text summarization; analysis of scientific texts; title generation; BART

Acknowledgments: The research was carried out with the financial support of the Ministry of Science and Higher Education of the Russian Federation (no. 121040100251-1).

***For citation*:** Glazkova, A.V. (2022) Approach to transforming training data for improving the title generation performance for scientific texts. *Vestnik Tomskogo gosudarstvennogo universiteta. Upravlenie, vychislitelnaja tehnika i informatika – Tomsk State University Journal of Control and Computer Science*. 59. pp. 99–107. doi: 10.17223/19988605/59/11

Целью автоматического реферирования является создание (генерация) более краткой версии исходного текста [1]. Методы автоматического реферирования в целом подразделяются на извлекающие (extractive) и генерирующие (abstractive). Первые строят реферат на основании алгоритмов отбора наиболее важных частей исходного текста [2–3], в то время как вторые обобщают содержание оригинального текста, создавая совершенно новый документ [4–5]. Ввиду сложности создания качественного текста на естественном языке при помощи компьютерных средств, извлекающие методы в настоящее время имеют более широкое практическое применение. Однако в последние годы наблюдается рост интереса к генерирующим методам, потенциально имеющим более широкий круг возможностей и вариантов использования [6].

Оценка качества алгоритмов автоматического реферирования выполняется с помощью специально разработанных метрик схожести текстов. Одной из распространенных метрик является ROUGE-*N* (Recall-Oriented Understudy for Gisting Evaluation), которая оценивает схожесть как долю пересекающихся *n*-грамм двух текстов [7]. Соответственно, ROUGE-1 измеряет соответствие униграмм (отдельных слов или токенов), ROUGE-2 – биграмм (пар слов или токенов) и т.д. Другими распространенными метриками качества автоматического реферирования являются BERTScore [8], выражающая сходство текстов на основе их контекстуализированных представлений, BLEU [9], определяющая меру качества реферирования одновременно для нескольких размеров *n*-грамм и др. Лучшие результаты в области генерирующего автоматического реферирования в последние годы демонстрируют модели, основанные на применении нейросетевой архитектуры Transformer. В частности, к ним относятся BART [10], Pegasus [11], T5 [12], BertSumAbs [13].

Генерация заголовков является важной задачей автоматического реферирования. Автоматическая генерация заголовков позволяет сократить временные затраты на написание текстов и обеспечить создание заголовков, объективно отражающих содержание текста [14]. К настоящему моменту предложен ряд подходов к автоматическому созданию заголовков, эти работы выполнены в основном на материале новостных текстов [15–17]. Тем не менее исследование методологии генерации заголовков для других жанров также представляется актуальной задачей. В частности, инструменты генерации заголовков научных текстов, широко представленных в многочисленных электронных библиотеках, способны значительно ускорить систематизацию материалов научных электронных ресурсов. Кроме того, исследования подтвердили, что качество заголовка научного текста влияет на видимость работы научному сообществу и количество цитирований [18–19].

Заголовок не всегда точно отражает содержание текста. Так, в работе [20] отмечено, что новостные заголовки зачастую на соответствуют исходным текстам с точки зрения пересечения *n*-грамм. Это затрудняет применение алгоритмов машинного обучения, которым для создания качественной модели необходима репрезентативная обучающая выборка. Авторы предложили использовать нейросетевые модели, обученные для автоматического определения логической связи между текстами (Natural Language Inference; NLI), чтобы очистить обучающую выборку от новостных заголовков, не являющихся логическими следствиями исходного текста, и тем самым повысить качество моделей автоматического реферирования. В отличие от заголовков текстов новостных порталов, заголовки научных статей являются более абстрактными и реже представляют собой пересказ содержания текста в рамках одного предложения [21–22]. Авторы работ [23–24] подчеркивают, что многие заголовки научных статей информируют о смысле научного текста, уведенном в подтекст, и понимаются ретроспективно, после прочтения текстов. Таким образом, применение моделей NLI для аналогичной операции над научными текстами видится менее перспективным.

В данной работе предлагается подход к повышению качества генерации заголовков для научных текстов, основанный на преобразованиях обучающих данных. Автор оценивает качество заголовков из обучающей выборки с помощью показателей полноты (*recall*) метрики ROUGE-1, рассчитанной для текстов заголовков и аннотаций научных статей. Полученные результаты показывают, какая доля слов заголовка получена из текста аннотации, т.е. насколько заголовок соответствует ее содержанию. Заголовки, имеющие низкий уровень соответствия, могут быть отсеяны (отфильтрованы) при обучении модели генерации заголовка. Так, «шумные» данные исключаются из процесса обучения. Также автор экспериментирует с генерацией искусственных обучающих примеров для отсеянных текстов. Результаты сравниваются с результатами модели, обученной на полном корпусе текстов (без фильтрации), и модели, обученной на заголовках, отобранных с помощью модели NLI. Эксперименты проведены на материале англоязычных корпусов с помощью модели BART для генерирующего реферирования текстов.

1. Текстовые корпуса

Исследование проводилось на материале двух текстовых корпусов, содержащих заголовки научных статей и их аннотации. Корпус SciTLDR [25] включает в себя фрагменты англоязычных научных статей компьютерной и информационной тематики, собранные на открытой платформе OpenReview.org. Корпус arXiv представляет собой фрагмент датасета arXiv[1], размещенного на платформе Kaggle. Он включает в себя фрагменты англоязычных препринтов, опубликованных в электронной библиотеке arXiv.org, относящиеся к следующим тематикам: статистика, биология, физика, экономика, компьютерные науки, математика. В качестве источника для генерации заголовков использованы тексты аннотаций научных статей. Основные статистические характеристики корпусов представлены в табл. 1.

Таблица 1

Характеристики текстовых корпусов

Характеристика	SciTLDR	arXiv
Общий объем (количество пар заголовок–аннотация)	3 229	20 000
Объем обучающей выборки	1 992	16 000
Объем валидационной выборки	619	2 000
Объем тестовой выборки	618	2 000
Средняя длина аннотации (в словах)	172,38	130,69
Средняя длина заголовка (в словах)	8,74	10,14
Количество уникальных слов (в аннотациях)	6 997	15 200
Количество уникальных слов (в заголовках)	1 597	4 529
Доля новых слов (*)	20,34	22,01

[1] https://www.kaggle.com/Cornell-University/arxiv

Для расчета доли новых слов (*) использовалась следующая формула:

$$New\_words = \frac{|W_t \setminus W_a|}{|W_t|}, \quad (1)$$

где W_t и W_a – множества уникальных слов для заголовков и аннотаций соответственно. В (1) $|\cdot|$ обозначает мощность множества, знак \ обозначает разность множеств.

2. Подход к преобразованию обучающей выборки

Предлагаемый подход к фильтрации и генерации обучающих примеров состоит в применении следующих шагов.

1. Определить значения показателей полноты метрики ROUGE-1, вычисленной для пар заголовков и аннотаций из обучающей выборки *D*.

Значение метрики ROUGE-1 относительно задачи автоматического реферирования характеризует степень сходства униграмм сгенерированного текста (в данном случае заголовка) и исходного текста (аннотации) и вычисляется по принципу нахождения *F*-меры:

$$\text{ROUGE-1} = \frac{2 \times \text{ROUGE-1}(recall) \times \text{ROUGE-1}(precision)}{\text{ROUGE-1}(recall) + \text{ROUGE-1}(precision)}. \quad (2)$$

При этом значение показателей полноты (*recall*) метрики показывает, какая доля слов, присутствующих в заголовке, присутствует в тексте аннотации.

$$\text{ROUGE-1}(recall) = \frac{\text{количество совпадающих униграмм для заголовка и текста}}{\text{количество униграмм заголовка}}. \quad (3)$$

Значение показателей точности (precision) метрики характеризует долю слов текста, присутствующих в заголовке.

$$\text{ROUGE-1}(precision) = \frac{\text{количество совпадающих униграмм для заголовка и текста}}{\text{количество униграмм текста}}. \quad (4)$$

2. Ранжировать примеры обучающей выборки в соответствии с попарными значениями ROUGE-1 (*recall*).

3. Отсеять примеры, имеющие значения ROUGE-1 (*recall*), меньшие порогового коэффициента *k*, т.е. такие пары аннотаций и заголовков, в которых заголовок характеризуется наименьшим содержанием слов, присутствующих в аннотации. Значение *k* находится в диапазоне [0; 1).

4. Обучить модель на фильтрованной обучающей выборке $D_{filtered}$.

С помощью обученной модели сгенерировать новые заголовки для примеров, которые были отфильтрованы, получив выборку $D_{generated}$ размера, соответствующего размеру выборки *D*.

3. Эксперименты и результаты

3.1. Модель

В качестве модели генерации заголовков использовалась BART-base, вариация модели BART для автоматического реферирования, комбинирующей в себе энкодер BERT (Bidirectional Encoder Representations from Transformers) [26] и декодер GPT-2 [27]. Модель имеет следующие характеристики: количество слоев – 12 (6 слоев энкодера и 6 слоев декодера), размер скрытого слоя – 768, количество параметров – 139 млн, стратегия декодирования – лучевой поиск (количество шагов – 5), кросс-энтропийная функция потерь.

В рамках экспериментов каждая модель была обучена (fine-tuned) на обучающей выборке в течение 3 эпох с размером батча, равным 4, и максимальной длиной входной последовательности, равной 256 токенам. После каждой итерации обучения проводилась проверка модели на валидационной

выборке. Лучшая модель оценивалась на тестовой выборке. Для реализации использовались библиотеки PyTorch [28] и Transformers [29].

3.2. Преобразования обучающей выборки

На начальном этапе проведения экспериментов для текстов обучающих выборок обоих корпусов были вычислены значения ROUGE-1 (*recall*) (п. 2, шаг 1). Как видно из данных, проиллюстрированных рис. 1, большая часть примеров в обучающих выборках обоих корпусов имеет значение метрики, большее либо равное 0,8 (более 64% примеров из корпуса SciTLDR и более 70% из arXiv). Превалирующей категорией при этом являются пары аннотаций и заголовков, для которых значение метрики равно или превышает 0,9 (40,4% и 47,7% соответственно). Это значит, что в большинстве случаев заголовок научного текста преимущественно состоит из слов, входящих в состав аннотации. Однако в случае обоих корпусов существует значительное количество примеров, значение ROUGE-1 (*recall*) для которых не превышает 0,5 (4,7% для SciTLDR и 5,2% для arXiv). Примеры с низким значением метрики можно рассматривать как «шумные» и попытаться улучшить качество модели автоматического реферирования, исключив их из процесса обучения.

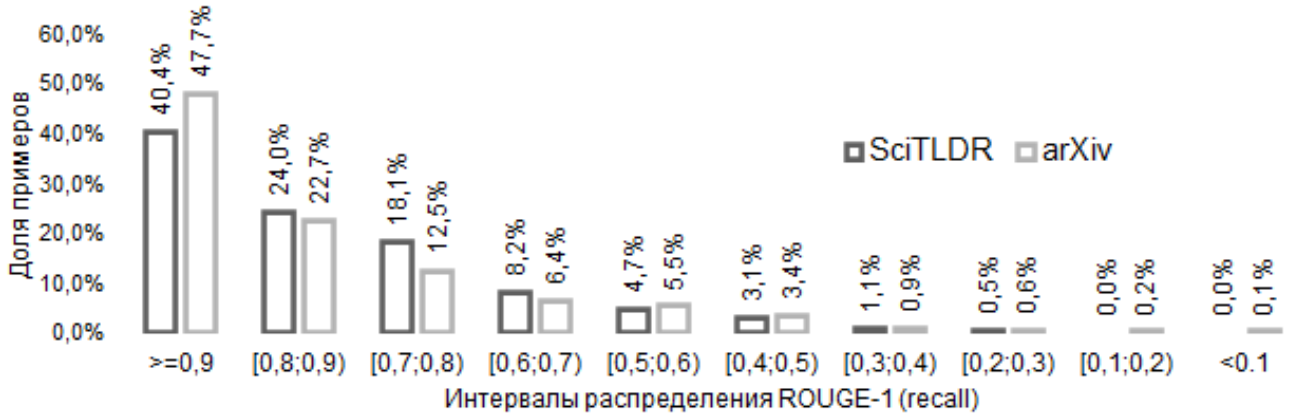

Рис. 1. Распределение значений показателей полноты ROUGE-1 для примеров обучающей выборки.
Fig. 1. Distribution of recall-oriented ROUGE-1 values for the training set

После ранжирования примеров обучающей выборки в соответствии со значениями ROUGE-1 (*recall*) (п. 2, шаг 2) для отсеивания обучающих примеров были выбраны значения $k \in [0,5;1)$ с шагом 0,1 (п. 2, шаг 3). Меньшие значения k не оценивались, поскольку для используемых корпусов количество примеров, отсеиваемых при $k \leq 0,4$ достаточно мало (менее 3% от размера D). Размеры фильтрованной выборки $D_{filtered}$ для двух корпусов и различных значений k представлены в табл. 2.

Таблица 2

Размеры фильтрованной выборки $D_{filtered}$ (в скобках указана доля от размера исходной выборки)

k	SciTLDR	arXiv
0,5	1 901 (95,4%)	15 164 (94,8%)
0,6	1 808 (90,8%)	14 292 (89,3%)
0,7	1 644 (82,5%)	13 267 (82,9%)
0,8	1 284 (64,5%)	11 261 (70,4%)
0,9	805 (40,4%)	7 636 (47,7%)

Фильтрованная выборка $D_{filtered}$ использовалась для обучения модели генерации заголовков (п. 2, шаг 4). Далее с помощью обученной модели были сгенерированы искусственные заголовки для отсеянных примеров (п. 2, шаг 5). Была сформирована обучающая выборка $D_{generated}$, состоящая из примеров, входящих в выборку $D_{filtered}$, и искусственных примеров.

3.4. Результаты

В табл. 3 сравниваются результаты моделей, обученных на разных типах обучающих выборок:

1) обучение на исходной выборке (D), *базовая модель*;

2) обучение на фильтрованной обучающей выборке ($D_{filtered}$) при различных значениях k;

3) обучение на выборке, состоящей из отфильтрованных и искусственных примеров ($D_{generated}$) при различных значениях k.

Полученные результаты сравнены с результатами моделей, обученных на выборках, фильтрованных с помощью модели NLI. По аналогии с [20] в данной работе использовалась модель RoBERTa-large-mnli [30] для определения логической связи между двумя текстами, обученная на корпусе MultiNLI [31]. Для формирования фильтрованной выборки $D_{NLI\text{-}filtered}$ были отсеяны примеры, в которых тексты заголовков не являются логически связанными с текстами аннотаций. Размер выборки $D_{NLI\text{-}filtered}$ составил 1 047 текстов для SciTLDR и 8 705 текстов для arXiv. Далее по аналогии с $D_{generated}$ была сформирована выборка $D_{NLI\text{-}generated}$, включающая в себя фильтрованные и искусственные примеры. Сравнение качества генерации заголовков проводится с помощью четырех метрик: ROUGE-1, ROUGE-2, ROUGE-*L* (ROUGE-*N*, рассчитанная для наибольшей общей последовательности), BERTScore. Лучшие значения с точки зрения каждой метрики выделены полужирным шрифтом.

Таблица 3

Результаты, %

Корпус	Обучающая выборка	ROUGE-1	ROUGE-2	ROUGE-*L*	BERTScore
SciTLDR	D	45,3	26,81	41,81	88,94
	$D_{filtered}$ (k = 0,5)	45,33	26,7	41,8	88,93
	$D_{filtered}$ (k = 0,6)	45,18	26,56	41,66	88,87
	$D_{filtered}$ (k = 0,7)	44,6	26,23	41,05	88,85
	$D_{filtered}$ (k = 0,8)	44,53	26,27	41,15	88,95
	$D_{filtered}$ (k = 0,9)	43,7	24,41	39,49	88,72
	$D_{NLI\text{-}filtered}$	43,27	24,56	39,51	88,75
	$D_{generated}$ (k = 0,5)	45,25	26,8	41,78	88,92
	$D_{generated}$ (k = 0,6)	45,14	26,8	41,08	88,8
	$D_{generated}$ (k = 0,7)	**45,36**	26,79	**42,2**	**88,97**
	$D_{generated}$ (k = 0,8)	45,32	**26,87**	42,11	88,96
	$D_{generated}$ (k = 0,9)	44,01	25,17	39,56	88,83
	$D_{NLI\text{-}generated}$	44,05	25,44	40,87	88,77
arXiv	D	42,35	23,4	38,3	87,83
	$D_{filtered}$ (k = 0,5)	42,3	23,42	38,3	87,82
	$D_{filtered}$ (k = 0,6)	42,85	24,13	38,41	87,88
	$D_{filtered}$ (k = 0,7)	43,1	24,25	39,19	87,91
	$D_{filtered}$ (k = 0,8)	43,14	24,22	39,24	87,94
	$D_{filtered}$ (k = 0,9)	42,14	22,88	37,99	87,78
	$D_{NLI\text{-}filtered}$	42,25	23,08	38,03	87,81
	$D_{generated}$ (k = 0,5)	42,44	23,7	38,56	87,84
	$D_{generated}$ (k = 0,6)	42,57	23,88	39	87,84
	$D_{generated}$ (k = 0,7)	43,4	**24,32**	39,22	87,88
	$D_{generated}$ (k = 0,8)	**43,44**	24,31	**39,25**	**87,95**
	$D_{generated}$ (k = 0,9)	42,11	22,93	38,01	87,77
	$D_{NLI\text{-}generated}$	42,3	22,9	38,03	87,79

Обучение на выборке $D_{filtered}$ для корпуса arXiv показало улучшение результатов в сравнении с базовой моделью при значениях k от 0,6 до 0,8. В остальных случаях качество сопоставимо с качеством базовой модели. Для SciTLDR качество постепенно снижается с уменьшением размера $D_{filtered}$.

Обучение на выборке $D_{generated}$ в большинстве случаев улучшило результаты на корпусе arXiv (лучшие результаты получены при k = 0,7 и k = 0,8). Для SciTLDR улучшение заметно также при k = 0,7 и k = 0,8. При этом при k = 0,9 качество всех моделей резко ухудшается. Модели, использующие фильтрацию данных с помощью NLI, не демонстрируют улучшения качества в сравнении с базовой моделью. Это подтверждает гипотезу о неэффективности такого подхода для научных текстов. Выявленные различия в результатах на двух корпусах обусловлены, вероятно, меньшим размером корпуса SciTLDR.

Заключение

В работе предложен подход к повышению качества генерации заголовков для научных текстов, использующий фильтрацию обучающей выборки на основе оценки показателей полноты метрики ROUGE-1 и генерации искусственных примеров. Тестирование подхода на материале двух текстовых корпусов показало его результативность и позволило выявить наиболее предпочтительные значения порогового коэффициента. Полученные результаты могут быть применены в системах автоматического реферирования и электронных научных библиотеках. Путями дальнейшего развития данного исследования являются, с одной стороны, автоматизация определения порогового коэффициента и, с другой стороны, тестирование предложенного подхода с помощью других моделей автоматического реферирования и на других корпусах (в том числе русскоязычных).

Список источников

1. El-Kassas W. S. et al. Automatic text summarization: a comprehensive survey // Expert Systems with Applications. 2021. V. 165. Art. 113679.
2. Nallapati R., Zhai F., Zhou B. Summarunner: A recurrent neural network based sequence model for extractive summarization of documents // Thirty-First AAAI Conference on Artificial Intelligence. 2017. P. 2101–2110.
3. Chen J., Zhuge H. Extractive summarization of documents with images based on multi-modal RNN // Future Generation Computer Systems. 2019. V. 99. P. 186–196.
4. Song S., Huang H., Ruan T. Abstractive text summarization using LSTM-CNN based deep learning // Multimedia Tools and Applications. 2019. V. 78 (1). P. 857–875.
5. Hanunggul P.M., Suyanto S. The impact of local attention in LSTM for abstractive text summarization // International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). 2019. P. 54–57.
6. Allahyari M. et al. Text Summarization Techniques: a Brief Survey // International Journal of Advanced Computer Science and Applications (IJACSA). 2017. V. 8 (10). P. 397–405.
7. Lin C.Y. Rouge: A package for automatic evaluation of summaries // Text summarization branches out. Barcelona, 2004. P. 74–81.
8. Zhang T. et al. BERTScore: Evaluating Text Generation with BERT // International Conference on Learning Representations. 2020. URL: https://arxiv.org/pdf/1904.09675v1.pdf
9. Papineni K. et al. BLEU: a method for automatic evaluation of machine translation // Proc. of the 40th annual meeting of the Association for Computational Linguistics. 2002. P. 311–318.
10. Lewis M. et al. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 7871–7880.
11. Zhang J. et al. Pegasus: Pre-training with extracted gap-sentences for abstractive summarization // International Conference on Machine Learning. 2020. P. 11328–11339.
12. Raffel C. et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer // Journal of Machine Learning Research. 2020. V. 21. P. 1–67.
13. Liu Y., Lapata M. Text Summarization with Pretrained Encoders // Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. P. 3730–3740.
14. Shen S.Q. et al. Recent advances on neural headline generation // Journal of Computer Science and Technology. 2017. V. 32 (4). P. 768–784.
15. Zhang R. et al. Question headline generation for news articles // Proc. of the 27th ACM international conference on information and knowledge management. 2018. P. 617–626.
16. Gavrilov D., Kalaidin P., Malykh V. Self-attentive model for headline generation // European Conference on Information Retrieval. 2019. P. 87–93.
17. Bukhtiyarov A., Gusev I. Advances of Transformer-Based Models for News Headline Generation // Conference on Artificial Intelligence and Natural Language. 2020. P. 54–61.
18. Putra J.W.G., Khodra M.L. Automatic title generation in scientific articles for authorship assistance: a summarization approach // Journal of ICT Research and Applications. 2017. № 11 (3). P. 253–267.

19. Fox C.W., Burns C.S. The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal // Ecology and Evolution. 2015. № 5 (10). P. 1970–1980.
20. Matsumaru K., Takase S., Okazaki N. Improving Truthfulness of Headline Generation // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics. 2020. P. 1335–1346.
21. Harmon J.E., Gross A.G. The structure of scientific titles // Journal of Technical Writing and Communication. 2009. V. 39 (4). P. 455–465.
22. Soler V. Writing titles in science: An exploratory study // English for specific purposes. 2007. V. 26 (1). P. 90–102.
23. Суворова С.А. Лексическая детерминированность заголовков научных статей // Ученые записки Крымского федерального университета им. В.И. Вернадского. Филологические науки. 2011. Т. 24, № 1-1. С. 163–166.
24. Филоненко Т.А. Аттрактивные заголовки в научной речи // Известия Самарского научного центра Российской академии наук. Социальные, гуманитарные, медико-биологические науки. 2008. Т. 10, № 6-2. С. 290–296.
25. Cachola I. et al. TLDR: Extreme Summarization of Scientific Documents // Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings. 2020. P. 4766–4777.
26. Devlin J. et al. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding // Proc. of NAACL-HLT. 2019. P. 4171–4186.
27. Radford A. et al. Language models are unsupervised multitask learners // OpenAI blog. 2019. V. 1 (8). P. 9.
28. Paszke A. et al. Pytorch: An imperative style, high-performance deep learning library // Advances in Neural Information Processing Systems. 2019. V. 32. P. 8026–8037.
29. Wolf T. et al. Transformers: State-of-the-art natural language processing // Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. 2020. P. 38–45.
30. Liu Y. et al. Roberta: a robustly optimized bert pretraining approach // arXiv preprint arXiv:1907.11692. 2019.
31. Williams A., Nangia N., Bowman S. A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference // Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2018. V. 1: Long Papers. P. 1112–1122.

References

1. El-Kassas, W.S. et al. (2021) Automatic text summarization: A comprehensive survey. *Expert Systems with Applications*. 165. pp. 113679.
2. Nallapati, R., Zhai, F. & Zhou, B. (2017) Summarunner: A recurrent neural network based sequence model for extractive summarization of documents. *Thirty-First AAAI Conference on Artificial Intelligence*. pp. 2101–2110.
3. Chen, J. & Zhuge, H. (2019) Extractive summarization of documents with images based on multi-modal RNN. *Future Generation Computer Systems*. 99. pp. 186–196. DOI: 10.1016/j.future.2019.04.045
4. Song, S., Huang, H. & Ruan, T. (2019) Abstractive text summarization using LSTM-CNN based deep learning. *Multimedia Tools and Applications*. 78(1). pp. 857–875. DOI: 10.1007/s11042-018-5749-3
5. Hanunggul, P.M. & Suyanto, S. (2019) The impact of local attention in LSTM for abstractive text summarization. *International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*. pp. 54–57.
6. Allahyari, M. et al. (2017) Text Summarization Techniques: A Brief Survey. *International Journal of Advanced Computer Science and Applications (IJACSA)*. 8(10). pp. 397–405.
7. Lin, C.Y. (2004) *Rouge: A package for automatic evaluation of summaries*. Association for Computational Linguistics. pp. 74–81.
8. Zhang, T. et al. (2020) BERTScore: Evaluating Text Generation with BERT. *International Conference on Learning Representations.*
9. Papineni, K. et al. (2002) BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics.* July. pp. 311–318.
10. Lewis, M. et al. (2020) BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 7871–7880.
11. Zhang, J. et al. (2020) Pegasus: Pre-training with extracted gap-sentences for abstractive summarization. *International Conference on Machine Learning*. pp. 11328–11339.
12. Raffel, C. et al. (2020) Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. *Journal of Machine Learning Research*. 21. pp. 1–67.
13. Liu, Y. & Lapata, M. (2019) Text Summarization with Pretrained Encoders. *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP).* pp. 3730–3740.
14. Shen, S. Q. et al. (2017) Recent advances on neural headline generation. *Journal of Computer Science and Technology*. 32(4). pp. 768–784.
15. Zhang, R. et al. (2018) Question headline generation for news articles. *Proceedings of the 27th ACM International Conference on Information and Knowledge Management*. pp. 617–626.
16. Gavrilov, D., Kalaidin, P. & Malykh, V. (2017) Self-attentive model for headline generation. *European Conference on Information Retrieval*. pp. 87–93.
17. Bukhtiyarov, A. & Gusev, I. (2020) Advances of Transformer-Based Models for News Headline Generation. *Conference on Artificial Intelligence and Natural Language*. pp. 54–61.

18. Putra, J.W.G. & Khodra, M.L. (2017) Automatic title generation in scientific articles for authorship assistance: a summarization approach. *Journal of ICT Research and Applications*. 11(3). pp. 253–267. DOI: 10.5614/itbj.ict.res.appl.2017.11.3.3
19. Fox, C.W., & Burns, C.S. (2015) The relationship between manuscript title structure and success: editorial decisions and citation performance for an ecological journal. *Ecology and Evolution*. 5(10). pp. 1970–1980. DOI: 10.5614/itbj.ict.res.appl.2017.11.3.3
20. Matsumaru, K., Takase, S. & Okazaki, N. (2020) Improving Truthfulness of Headline Generation. *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. pp. 1335–1346.
21. Harmon, J.E. & Gross, A.G. (2009) The structure of scientific titles. *Journal of Technical Writing and Communication.* 39(4). pp. 455–465.
22. Soler, V. (2007) Writing titles in science: An exploratory study. *English for Specific Purposes*. 26(1). pp. 90–102. DOI: 10.1016/j.esp.2006.08.001
23. Suvorova, S.A. (2011) Leksicheskaya determinirovannost' zagolovkov nauchnykh statey [Lexical Determination of Scientific Article Headlines]. *Uchenye zapiski Krymskogo federal'nogo universiteta im. V.I. Vernadskogo. Filologicheskie nauki*. 24(1-1). pp. 163–166.
24. Filonenko, T.A. (2008) Attractive Scientific Discourse Headings. *Izvestiya Samarskogo nauchnogo tsentra Rossiyskoy akademii nauk. Sotsial'nye, gumanitarnye, mediko-biologicheskie nauki – Izvestiya of Samara Science Centre of the Russian Academy of Sciences. Social, Humanitarian, Medicobiological sciences*. 10(6-2). pp. 290–296.
25. Cachola, I. et al. (2020) TLDR: Extreme Summarization of Scientific Documents. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: Findings*. pp. 4766–4777.
26. Devlin, J. et al. (2019) BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT*. pp. 4171–4186.
27. Radford, A. et al. (2019) Language models are unsupervised multitask learners. *OpenAI blog*. 1(8). p. 9.
28. Paszke, A. et al. (2019) Pytorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*. 32. pp. 8026–8037. DOI: 10.48550/arXiv.1912.01703
29. Wolf, T. et al. (2020) Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. pp. 38–45.
30. Liu, Y. et al. (2019) *Roberta: A robustly optimized bert pretraining approach*. arXiv preprint arXiv:1907.11692.
31. Williams, A., Nangia, N. & Bowman, S. (2018) A Broad-Coverage Challenge Corpus for Sentence Understanding through Inference. *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* Vol. 1. pp. 1112–1122.

***Информация об авторе*:**
Глазкова Анна Валерьевна – кандидат технических наук, доцент кафедры программного обеспечения Института математики и компьютерных наук Тюменского государственного университета (Тюмень, Россия). E-mail: a.v.glazkova@utmn.ru

Автор заявляет об отсутствии конфликта интересов.

***Information about the author*:**
Glazkova Anna V. (Candidate of Technical Sciences, Associate Professor, University of Tyumen, Tyumen, Russian Federation). E-mail: a.v.glazkova@utmn.ru

The author declares no conflicts of interests.

Поступила в редакцию 20.11.2021; принята к публикации 30.05.2022

Received 20.11.2021; accepted for publication 30.05.2022