
Суммаризация научных подборок в задаче автоматической генерации обзоров

A Preprint

Дегтев Василий Денисович
МГУ им. М.В. Ломоносова
Москва, Россия
DegtevVD@my.msu.ru

Ищенко Роман Валерьевич
МГУ им. М.В. Ломоносова
Москва, Россия
ischenkorv@my.msu.ru

Abstract

Автоматическая генерация обзора литературы является важной задачей в области автоматической обработки научных публикаций и направлена на создание релевантных и контекстуально уместных цитат для поддержки аргументации в академических текстах. Появление сверхбольших языковых моделей значительно продвинуло исследования в этой области: на сегодняшний день большинство существующих подходов так или иначе основаны на использовании больших языковых моделей (БЯМ). В данной работе исследуется возможность улучшения качества генерации обзоров литературы с помощью БЯМ, благодаря дообучению небольшой модели на генерацию целей цитирования методом обучения с подкреплением с наградой, выдаваемой БЯМ (RLAIF).

1 Введение

Качественный обзор литературы является важной частью любой научной работы, обеспечивая контекст и обосновывая актуальность исследования. Однако его подготовка — трудоемкий процесс. В связи с этим автоматическая генерация обзора литературы, направленная на создание релевантных и контекстуально уместных цитат для поддержки аргументации, стала одной из важных задач в области обработки научных публикаций.

На сегодняшний день большинство передовых подходов в генерации текста цитирования опираются на использование Больших Языковых Моделей (БЯМ), число параметров которых может насчитывать десятки или сотни миллиардов.

Цели цитирования (citation intents) представляют из себя кратко сформулированные причины, о которой автор академической работы включает ссылку на другую научную публикацию. В области обработки естественного языка цель цитирования часто классифицируется по определенным категориям (например, методология, результат). [Wu et al., 2021] первые применили цели цитирования в задаче генерации текста цитирования и показали их эффективность в решении данной задачи. Позднее, Şahinuç et al. [2024] показали, что цели цитирования, сформулированные в свободной форме, а не как категориальные признаки, позволяют добиться лучшего качества генерации текста цитирования.

Şahinuç et al. [2024] использовали модель T5 с zero-shot генерацией целей цитирования. В данной работе мы усовершенствуем данный подход, посредством дообучения модели Qwen2.5-0.5B-Instruct [Qwen: An Yang, 2024] методами обучения с подкреплением на выходах модели, генерирующей текст цитирования.

Существует значительная разница в качестве генерации моделей разного размера. Обучение моделей на десятки или сотни миллиардов параметров является очень вычислительно-затратным процессом. Мы полагаем, что наш подход позволит добиться заметного улучшения качества генерации

цитирующего текста БЯМ без необходимости из дообучения на данную задачу.

2 Обзор литературы

2.1 Подходы к решению

Ранние работы, такие как [Hu and Wan, 2014], использовали тематическое моделирование и классические методы машинного обучения для экстрактивного выделения текста цитирования. Современные же решения основаны на БЯМ с трансформерной архитектурой, которые впервые были применены в [Xing et al., 2020]. Значительный прогресс в качестве генерации цитирующего текста связан с внедрением намерений цитирования (citation intents), предложенные [Wu et al., 2021] с целью повышения качества генерируемых цитат. Этот подход развит в [Jung and Lin, 2022] и [Gu and Hahnloser, 2023a] с использованием обучения с подкреплением (PPO).

Контекстное обогащение стало важным направлением: в [Arita et al., 2022] на вход модели подаются релевантные фрагменты из цитируемых статей, а в [Li et al., 2024] вводится концепция cited text spans — автоматически извлечённых информативных отрывков. Генерация обзоров для нескольких работ одновременно решена в [Anand et al., 2023] с помощью моделей с увеличенным контекстом, способных формировать связные абзацы.

С появлением БЯМ на миллиарды параметров активно применяются техники промптинга: Chain-of-Thought в [Ji et al., 2024], seq2seq-модели с графами знаний в [Anand et al., 2024]. Графовые энкодеры и контрастивное обучение использованы в [Chen et al., 2022] и [Chen et al., 2021] для моделирования связей между документами. Интегрированные пайплайны, такие как SciLit [Gu and Hahnloser, 2023b], сочетают поиск, суммаризацию и генерацию цитирований. Систематическое исследование промптинга и контроля намерений проведено в [Şahinuç et al., 2024].

2.2 Данные

Качество генерации во многом зависит от специализированных датасетов. SciReviewGen [Kasanishi et al., 2023] предоставляет масштабный корпус для генерации обзоров литературы. CORWA [Li et al., 2022] содержит аннотации связанных работ с указанием семантических ролей цитирований. В [Anand et al., 2024] представлен новый датасет с мультидокументными примерами и улучшенными аннотациями. Многие работы используют абстракты или полные тексты из открытых научных репозиторий (arXiv, ACL Anthology), а также предварительно обработанные фрагменты [Li et al., 2024]. В работе [Martin Funkquist, 2023] представлен бенчмарк, основанный на корпусе CORWA. Датасет Multi-XScience [Yao Lu, 2020] представляет из себя набор данных для задачи мультидокументной суммаризации подборок аннотаций статей с целью генерации обзора литературы.

2.3 Критерия оценки качества

Оценка качества генерации остаётся сложной задачей. Стандартной метрикой является ROUGE [Lin, 2004], измеряющая пересечение n-грамм, однако она недостаточна для оценки семантической точности. Метрика BERTScore [Tianyi Zhang, 2019], основанная на энкодере BERT, позволяет более точно оценить семантическую близость между эталонным текстом и выходами модели. [Şahinuç et al., 2024] использовали вместо обычного BERT модель SciBERT [Iz Beltagy, 2019] для подсчёта BERTScore. Работы [Anand et al., 2024], [Gu and Hahnloser, 2023a] предложили собственные метрики, учитывающие точность цитирований и релевантность контента.

3 Постановка задачи

3.1 Дано

Набор данных $\{(A_j, a_j, c_j, G_j), j_{=1}^N\}$ состоит из A_j - аннотации цитирующей статьи, a_j, \dots, a_{kj} - аннотации цитируемой статьи, c_j (citation intent) - цель цитирования, G_j - эталонных текстов цитирований.

3.2 Найти

Обозначим модель за $M = M(\Theta)$, $\Theta = (\theta_1, \dots, \theta_s)$ - вектор параметров модели, размер словаря $|V|$ и длину выходной последовательности T . Отображение модели M из пространства входов X в пространство предсказанных распределений вероятностей токенов текста цитирования Y :

$$M : X \longrightarrow Y$$

где X – пространство входных данных, а Y – последовательность векторов вероятностей, предсказанных моделью:

$$Y = (y_1, y_2, \dots, y_T)$$

Каждый элемент y_t в этой последовательности является вектором (распределением) вероятностей по всему словарю V :

$$y_t \in \mathbb{R}^{|V|}$$

3.3 Внешние критерии оценки качества

3.3.1 ROUGE-L

Одними из основных метрик для оценки качества суммаризации являются ROUGE-based метрики, оценивающие пересечение n -грамм в референсном и сгенерированном тексте. Мы используем ROUGE-L для оценки качества.

Пусть:

- G : Эталонный текст длины m .
- Y : Сгенерированный текст длины n .
- $LCS(G, Y)$: Длина наибольшей общей подпоследовательности между G и Y .

$$R_{lcs} = \frac{LCS(G, Y)}{m} \tag{1}$$

$$P_{lcs} = \frac{LCS(G, Y)}{n} \tag{2}$$

$$F_{lcs} = \frac{(1 + \beta^2) R_{lcs} P_{lcs}}{R_{lcs} + \beta^2 P_{lcs}} \tag{3}$$

3.3.2 SciBERTScore

Поскольку ROUGE-L недостаточно, чтобы корректно оценить семантическую близость между сгенерированным текстом и эталонным мы используем BERTScore. В качестве эмбеддера мы берем вместо стандартного BERT его дообученную на научной литературе версию SciBERT.

Пусть:

- $G = (g_1, \dots, g_t)$ – эталонная последовательность.
- $Y = (y_1, \dots, y_t)$ – сгенерированная последовательность.
- $\hat{y}_j, \hat{g}_j \in \mathbb{R}^{|V|}$ – векторы эмбеддингов для токенов y_i и g_j .
- $w(t)$ – IDF-вес для токена t .
- $s(\hat{a}, \hat{b}) = \frac{\hat{a} \cdot \hat{b}}{\|\hat{a}\| \|\hat{b}\|}$.

$$R_{BERT} = \frac{\sum_{i=1}^m w(y_i) \max_{j=1}^n s(\hat{y}_i, \hat{g}_j)}{\sum_{i=1}^m w(y_i)} \tag{4}$$

$$P_{BERT} = \frac{\sum_{j=1}^n w(g_j) \max_{i=1}^m s(\hat{y}_i, \hat{g}_j)}{\sum_{j=1}^n w(g_j)} \quad (5)$$

$$F_{BERT} = 2 \frac{P_{BERT} \cdot R_{BERT}}{P_{BERT} + R_{BERT}} \quad (6)$$

4 Предлагаемое решение

В связи с отсутствием размеченных данных и возможности собрать их из открытых источников, мы решили обучать модель-генератора цели цитирования на логитах модели-генератора текста цитирования. Поскольку наша конечная цель использовать сгенерированные намерения с моделями типа "черный ящик" которые принимают входной текст и генерируют выходной, мы отказались от возможности моделирования эмбедингов токенов цели цитирования в пользу генерации самих токенов. В силу недифференцируемости операции $\arg\max$ мы не можем обучить модель простым supervised-fine-tuning. Поэтому мы используем подход обучения с подкреплением методом GRPO [DeepSeek-AI, 2024] для решения поставленной задачи.

Пусть:

- x обозначает контекст цитирования (например, аннотация цитирующей статьи, заголовок цитируемой статьи).
- y^* обозначает истинный текст цитирования.
- π_θ — обучаемая малая модель (Политика / Policy).
- \mathcal{M}_ϕ — замороженная большая языковая модель (Модель Награды / Reward Model).

Малая модель генерирует цель цитирования (добавку к промпту) z на основе контекста:

$$z \sim \pi_\theta(\cdot|x) \quad (7)$$

Замороженная большая модель получает на вход объединенные данные $[x; z]$ (контекст + цель) и предсказывает текст цитирования y .

4.1 Функция Награды

Награда $r(z)$ определяется как отрицательная перекрестная энтропия (negative Cross-Entropy loss) большой модели на истинном тексте y^* при условии сгенерированного намерения z :

$$r(z) = \log P_{\mathcal{M}_\phi}(y^*|x, z) = \sum_{t=1}^{|y^*|} \log P_{\mathcal{M}_\phi}(y_t^*|y_{<t}^*, x, z) \quad (8)$$

Более низкое значение функции потерь (более высокое правдоподобие генерации истинного текста) дает более высокую награду.

4.2 Преимущество (Advantage)

Для каждого входного примера x мы сэмплируем группу из G намерений $\{z_1, z_2, \dots, z_G\}$ из текущей политики $\pi_{\theta_{old}}$.

Мы вычисляем награды для каждого варианта: $R = \{r(z_1), \dots, r(z_G)\}$.

Преимущество (Advantage) A_i для каждого сгенерированного намерения z_i нормализуется относительно группы:

$$A_i = \frac{r(z_i) - \text{mean}(R)}{\text{std}(R) + \epsilon} \quad (9)$$

4.3 Целевая функция

Целевая функция максимизирует преимущество политики, одновременно штрафуя за отклонение от референсной модели π_{ref} через KL-дивергенцию:

$$\mathcal{L}_{GRPO}(\theta) = \mathbb{E}_{x \sim \mathcal{D}, z \sim \pi_{\theta_{old}}} [L_i^{CLIP}(z_i) - \beta D_{KL}(\pi_{\theta}(z_i|x) || \pi_{ref}(z_i|x))] \quad (10)$$

Где L^{CLIP} — стандартная функция потерь PPO [John Schulman, 2017]:

$$L_i^{CLIP}(z) = \min(\rho_i A_i, \text{clip}(\rho_i, 1 - \epsilon, 1 + \epsilon) A_i), \quad \rho_i = \frac{\pi_{\theta}(z_i|x)}{\pi_{\theta_{old}}(z_i|x)} \quad (11)$$

5 Эксперименты

5.1 Данные

Для обучения и валидации моделей использовался датасет собранный [Gu and Hahnloser, 2023a] из аннотаций статей, контекстов и категориальных целей цитирований, размеченных классификатором.

5.2 Гипотезы

В результате экспериментов мы стремимся проверить две гипотез:

1. Получится ли добиться значительного прироста качества генерации текста цитирования при дообучении модели, генерирующей цели цитирования.
2. Большинство современных моделей выпускаются в вариациях с различным числом параметров. Поскольку модели меньшего размера из одной серии обучаются дистилляцией знаний из более крупных моделей, мы хотим проверить возможно ли улучшить качество модели с большим числом параметров, при дообучении на выходах малой модели, если они принадлежат одной серии.

5.3 Обучение

Мы обучили модель Qwen2.5-0.5B-Instruct на выборке из 9600 примеров 5 эпох на видеокарте Nvidia Tesla P100. Была выбрана именно Instruct версия: поскольку обучение проходило без предварительно SFT, модель должна была уже иметь следовать инструкциям, чтобы генерировать осмысленные цели. На этапе обучения в качестве модели, генерирующей цитаты, мы использовали Qwen2.5-1.5B-Instruct. В качестве оптимизатора использовался AdamW с гиперпараметрами $lr = 1e - 5$, $\beta = (0.9, 0.95)$; warmup-ratio = 0.05. Коэффициент β для KL-дивергенции был выбран равным 0.3, поскольку при обучении на выходах другой модели высока вероятность сильно сбить выравнивание исходной модели.

5.4 Результаты

Результаты сравнения подходов на тестовой выборке из 1080 примеров для моделей Qwen2.5-1.5B-Instruct и Qwen2.5-72B-Instruct приведены в Таб. 1,2.

Таблица 1: Оценка генерации текста цитирования моделью Qwen2.5-1.5B-Instruct.

Intent generator	ROUGE-L			SciBERTScore		
	Precision	Recall	F-measure	Precision	Recall	F-measure
No intents	0.048	0.165	0.069	0.458	0.512	0.482
Categorical intents	0.062	0.205	0.089	0.481	0.504	0.496
Zero-shot Qwen2.5-0.5B-Instruct	0.058	0.188	0.082	0.472	0.531	0.498
GRPO-tuned Qwen2.5-0.5B-Instruct (our)	0.069	0.231	0.099	0.495	0.563	0.526

Наш подход позволяет добиться значительного прироста в SciBERTScore на модели Qwen2.5-1.5B-Instruct и немного меньшего на модели Qwen2.5-72B-Instruct. Значения же Rouge-L не столь однозначны,

Таблица 2: Оценка генерации текста цитирования моделью Qwen2.5-72B-Instruct на тестовой выборке.

Intent generator	ROUGE-L			SciBERTScore		
	Precision	Recall	F-measure	Precision	Recall	F-measure
No intents	0.072	0.254	0.105	0.511	0.592	0.545
Categorical intents	0.093	0.276	0.116	0.519	0.603	0.558
Zero-shot Qwen2.5-0.5B-Instruct	0.085	0.283	0.118	0.526	0.611	0.564
GRPO-tuned Qwen2.5-0.5B-Instruct (our)	0.083	0.304	0.123	0.537	0.627	0.577

однако улучшение в качестве так ж заметно. При этом по Rouge-L подход с категориальными целями превосходит zero-shot генерацию намерений, однако по SciBERTScore сгенерированные цели оказываются лучше, чем просто категориальные.

6 Заключение

В работе представлен метод улучшения качества генерации текста цитирования путем дообучения Qwen2.5-0.5B-Instruct методом RLAIIF с GRPO на задачу генерации целей цитирования. По результатам исследования подтвердились обе гипотезы выдвинутые в 5.2. Обе модели (Qwen2.5-1.5B-Instruct и Qwen2.5-72B-Instruct) показали прирост в качестве на тестовой выборке по сравнению с категориальными целями и zero-shot генерацией целей. Данный подход уместен при невозможности дообучения большой модели, он позволяет добиться улучшения в качестве, затрачивая минимум ресурсов. Однако при отсутствии ограничения по ресурсам для обучения SFT на размеченной выборке будет более разумным выбором.

Список литературы

- Jia-Yan Wu, Alexander Te-Wei Shieh, Shih-Ju Hsu, and Yun-Nung Chen. Towards generating citation sentences for multiple references with intent control. 2021.
- Furkan Şahinuç, Iliia Kuznetsov, Yufang Hou, and Iryna Gurevych. Systematic task exploration with llms: A study in citation text generation. 2024.
- Beichen Zhang Binyuan Hui Bo Zheng Bowen Yu Chengyuan Li Qwen: An Yang, Baosong Yang. Qwen2.5 technical report. 2024.
- Yue Hu and Xiaojun Wan. Automatic generation of related work sections in scientific papers: An optimization approach. 2014.
- Xinyu Xing, Xiaosheng Fan, and Xiaojun Wan. Automatic generation of citation texts in scholarly papers: A pilot study. 2020.
- Shing-Yun Jung and Ting-Han Lin. Intent-controllable citation text generation. 2022.
- Nianlong Gu and Richard H. R. Hahnloser. Controllable citation sentence generation with language models. 2023a.
- Akito Arita, Hiroaki Sugiyama, Kohji Dohsaka, Rikuto Tanaka, and Hirotoishi Taira. Citation sentence generation leveraging the content of cited papers. 2022.
- Xiangci Li, Yi-Hui Lee, and Jessica Ouyang. Cited text spans for scientific citation text generation. 2024.
- Avinash Anand, Kritarth Prasad, Ujjwal Goel, Mohit Gupta, Naman Lal, Astha Verma, and Rajiv Ratn Shah. Context-enhanced language models for generating multi-paper citations. 2023.
- Bin Ji, Huijun Liu, Mingzhe Du, and See-Kiong Ng. Chain-of-thought improves text generation with citations in large language models. 2024.
- Avinash Anand, Ashwin R. Nair, Kritarth Prasad, Vrinda Narayan, Naman Lal, Debanjan Mahata, Yaman K. Singla, and Rajiv Ratn Shah. Advances in citation text generation: Leveraging multi-source seq2seq models and large language models. 2024.
- Xiuying Chen, Hind Alamro, Mingzhe Li, Shen Gao, Rui Yan, Xin Gao, and Xiangliang Zhang. Target-aware abstractive related work generation with contrastive learning. 2022.

- Xiuying Chen, Hind Alamro, Mingzhe Li, Shen Gao, Xiangliang Zhang, Dongyan Zhao, and Rui Yan. Capturing relations between scientific papers: An abstractive model for related work section generation. 2021.
- Nianlong Gu and Richard H. R. Hahnloser. Scilit: A platform for joint scientific literature discovery, summarization and citation generation. 2023b.
- Tetsu Kasanishi, Masaru Isonuma, Junichiro Mori, and Ichiro Sakata. Scireviewgen: A large-scale dataset for automatic literature review generation. 2023.
- Xiangci Li, Biswadip Mandal, and Jessica Ouyang. Corwa: A citation-oriented related work annotation dataset. 2022.
- Yufang Hou Iryna Gurevych Martin Funkquist, Ilia Kuznetsov. Citebench: A benchmark for scientific citation text generation. 2023.
- Laurent Charlin Yao Lu, Yue Dong. Multi-xscience: A large-scale dataset for extreme multi-document summarization of scientific articles. 2020.
- Chin-Yew Lin. Rouge: A package for automatic evaluation of summaries. 2004.
- Felix Wu Kilian Q. Weinberger Yoav Artzi Tianyi Zhang, Varsha Kishore. Bertscore: Evaluating text generation with bert. 2019.
- Arman Cohan Iz Beltagy, Kyle Lo. Scibert: A pretrained language model for scientific text. 2019.
- Bei Feng Bin Wang Bingxuan Wang Bo Liu Chenggang Zhao Chengqi Dengr DeepSeek-AI, Aixin Liu. Deepseek-v2: A strong, economical, and efficient mixture-of-experts language model. 2024.
- Prafulla Dhariwal Alec Radford Oleg Klimov John Schulman, Filip Wolski. Proximal policy optimization algorithms. 2017.