Применение Self-Refinement для повышения точности и объективности генерации ответов в системах вопросно-ответного типа (QA)

Трофимова Маргарита Сергеевна

Московский физико-технический институт
Физтех-школа прикладной математики и информатики

2024

Self-Refine: Итеративное самоусовершенствование

Обзор метода: В статье *"Self-Refine: Iterative Refinement with Self-Feedback"* (Madaan et al., 2023) предложен новый подход для больших языковых моделей (LLMs), позволяющий им итеративно улучшать выходные данные с использованием самообратной связи.
Ключевые моменты:

- Применен на 7 задачах: генерация текста, обработка кода, математические задачи.
- Использованы современные LLMs: GPT-3.5, ChatGPT, GPT-4.
- Повышение точности от 5% до 40% в различных задачах.

Принцип работы: Модели анализируют и корректируют свои ответы без вмешательства человека, выявляя ошибки и генерируя обратную связь.
Обновления в 2024 году: Предложены модернизации метода и выявлены пробелы. Рассматриваются ключевые исследования и примеры применения в системах вопросно-ответного типа (QA).

Мотивация: Почему важно решать проблему?

- **Растущая роль LLM:** Основы современных технологий, от виртуальных ассистентов до автоматизации обработки текстов. Без самоусовершенствования модели склонны к ошибкам и предвзятости.
- **Снижение зависимости от ручной разметки:** Автоматизация обучения через самооценку ускоряет развитие моделей и снижает затраты на разметку данных.
- **Экономия на разметке данных:** Модели генерируют качественные ответы без участия человека, что делает разработку ИИ более доступной.
- **Повышение точности малых моделей:** Методы self-refinement улучшают способности малых моделей, приближая их к большим по качеству.
- **Улучшение качества взаимодействия:** Методы self-refinement помогают моделям предоставлять более точные и согласованные ответы, что особенно важно в критических областях.
- **Проблема предвзятости:** Устранение предвзятости через самосовершенствование способствует более справедливому использованию моделей.
- **Необходимость адаптации:** Self-refinement позволяет моделям адаптироваться к специализированным областям знаний, используя доступные данные и механизмы оптимизации.

Постановка задачи

Общая цель: Найти уточнённую модель M^*, которая максимизирует метрики релевантности, точности и полноты для генерации ответов на запросы пользователей.

Формулировка:

$$\arg\max_{M^*} \frac{1}{n} \sum_{i=1}^{n} \left[R(a_i^*, q_i) + P(a_i^*, K) + C(a_i^*, K) \right],$$

где:

- $R(a_i^*, q_i)$ — релевантность ответа, оценка соответствия ответа запросу,
- $P(a_i^*, K)$ — точность ответа, доля корректных фактов в ответе,
- $C(a_i^*, K)$ — полнота ответа, доля фактов из ответа, найденных в источнике знаний.

Этапы метода:

1. Генерация начального ответа.
2. Самоанализ с использованием метрик.
3. Генерация обратной связи для выявленных ошибок.
4. Уточнение ответа на основе обратной связи.
5. Итеративное уточнение до достижения оптимума.

SELF-REFINE: Iterative Refinement with Self-Feedback

Self-Refine улучшает выходные данные больших языковых моделей (LLMs) через итеративное самоусовершенствование, основанное на обратной связи. **Процесс:**

1. Генерация начального ответа.
2. Оценка ответа по метрикам (релевантность, точность, полнота).
3. Генерация обратной связи для выявленных ошибок.
4. Уточнение ответа на основе обратной связи.
5. Повторение процесса до достижения удовлетворительного результата.

Преимущества:

- Повышение точности и релевантности ответов.
- Адаптация к сложным задачам с множеством целей.
- Эффективность в задачах, таких как оптимизация кода и анализ настроений.

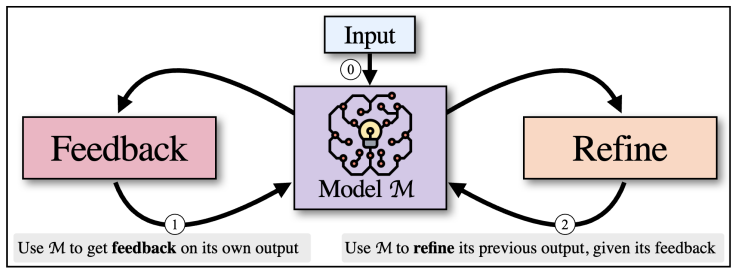


Рис.: При наличии входных данных (0) САМООЧИЩЕНИЕ начинается с генерации выходных данных и передачи их обратно в ту же модель M для получения обратной связи (1). Обратная связь передается обратно в M, которая уточняет ранее сгенерированные выходные данные (2). Шаги (1) и (2) повторяются до тех пор, пока не будет выполнено условие остановки. САМООПРЕДЕЛЕНИЕ выполняется с помощью языковой модели, такой как GPT-3.5, и не требует участия человека.

Pride and Prejudice: Self-Bias in Self-Refinement

- Исследует проблему "самопредвзятости"LLM, когда модели склонны предпочитать собственные генерации, что может снижать качество ответов.
- Предлагаются решения, такие как увеличение размера модели и привлечение внешней обратной связи для уменьшения этой предвзятости.

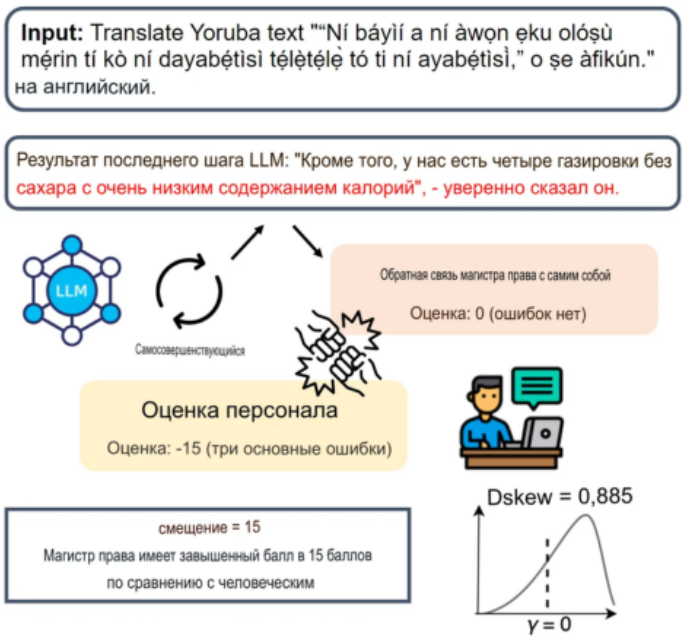


Рисунок 1: Как самооценка LLM завышает баллы по сравнению с оценкой человека. Смещение — это средняя разница между баллами LLM и человека, а асимметрия (Dskew) измеряет асимметрию их распределения вокруг нуля. При unbiased-оценке Dskew=0.

Оценка смещения

$$\text{Bias}(\hat{\theta}) = \frac{1}{n}\sum_{i=1}^{n}(\mathbb{E}[\hat{\theta}_i] - \theta_i),$$

где $\mathbb{E}[\theta\hat{}i]$ — ожидаемый прогноз качества LLM для выборки i, а θi — истинное качество выборки i. В идеале θi должно быть получено на основе аннотаций человека. Чем больше значение Bias(θ^), тем сильнее LLM предвзято относится к собственным образцам.

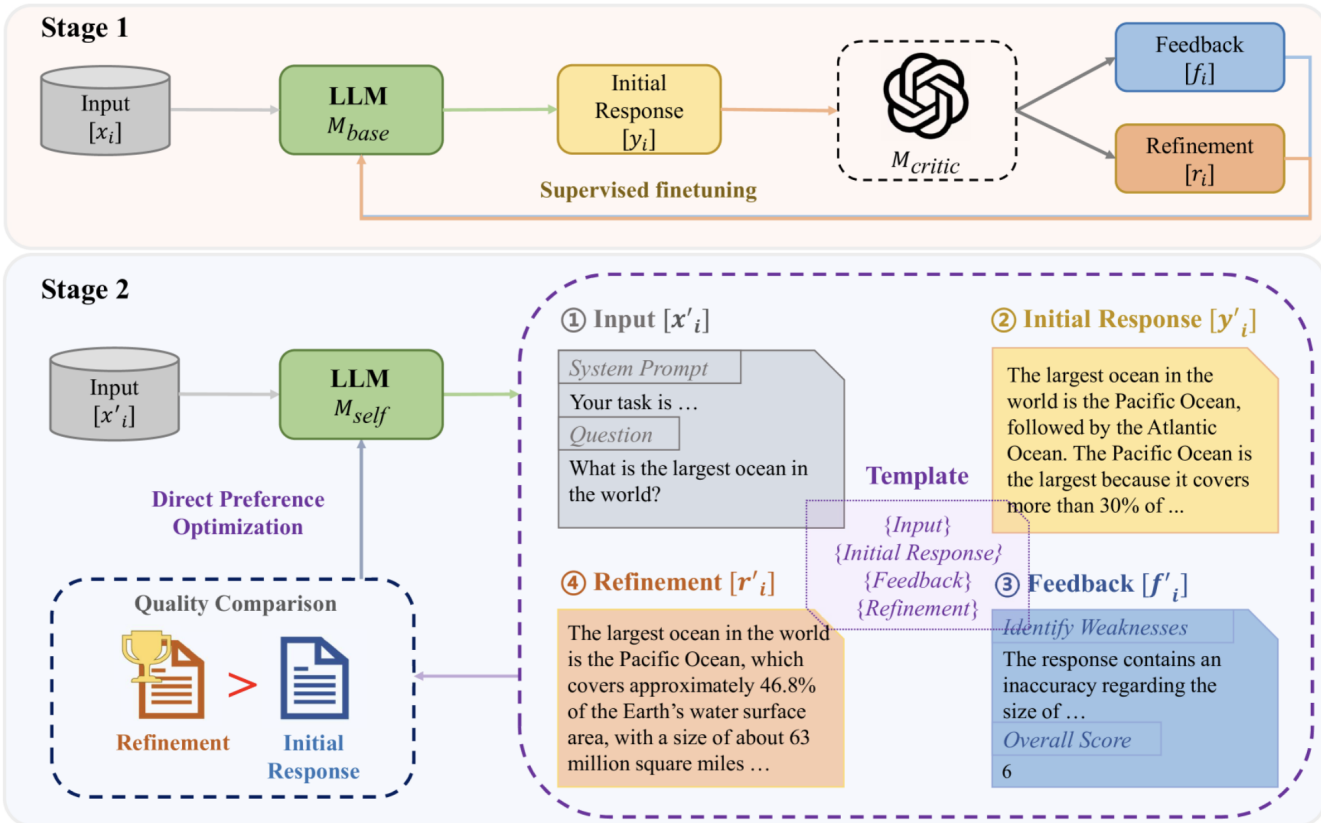
Оценка Асимметрии расстояния

$$d\text{Skew}_n(X) = 1 - \frac{\sum_{i,j}\|x_i - x_j\|}{\sum_{i,j}\|x_i + x_j - 2\gamma\|},$$

где xi и xj - два независимых идентичных случайных примера, взятых из $\mathbb{E}[\theta\hat{}]-\theta$. dSkewn(X) измеряет асимметрию X по отношению к γ. Асимметрия расстояния варьируется от 0 и 1. dSkewn(X) равно 0 тогда и только тогда, когда X распределено по диагонали относительно γ. dSkewn(X) равно 1 тогда и только тогда, когда X распределено с постоянной величиной на одной стороне γ. Большая асимметрия расстояния указывает на более асимметричное распределение $\mathbb{E}[\theta\hat{}]-\theta$.

Teaching Language Models to Self-Improve by Learning from Language Feedback

Представляет метод Self-Refinement Tuning (SRT), где базовая модель генерирует ответы, которые затем критикуются и уточняются более продвинутой моделью. Это создает цикл обратной связи, способствуя непрерывному обучению и улучшению модели.



Self-Refine Instruction-Tuning for Aligning Reasoning in Language Models

- Фокусируется на передаче навыков пошагового рассуждения от больших моделей к меньшим, используя Direct Preference Optimization (DPO) для самооптимизации.
- Это улучшает производительность и согласованность ответов малых моделей.

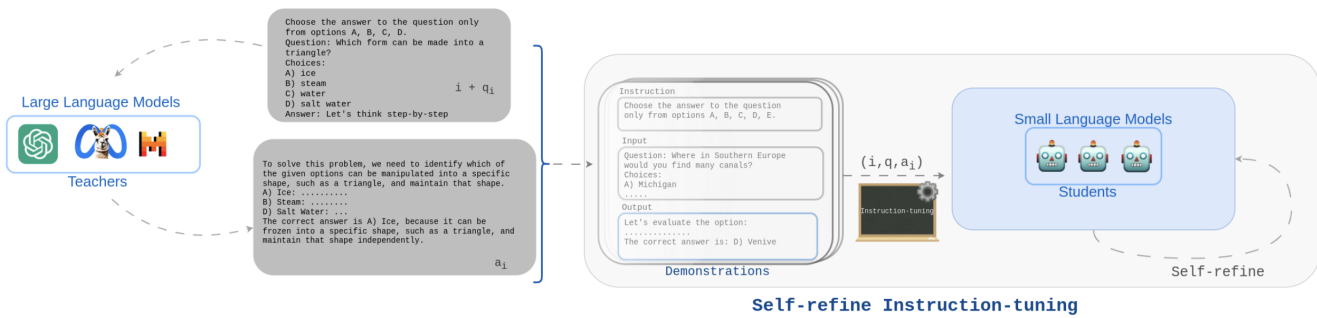


Рис.: При самостоятельной настройке инструкций демонстрации, проводимые моделями учителей, используются для согласования способностей к объяснению в обстановке учитель-ученик. После передачи пошаговых знаний посредством настройки структуры учащиеся самостоятельно совершенствуют свои способности при поддержке методов прямой оптимизации предпочтений.

Примеры применения

- **SimRAG: Self-Improving Retrieval-Augmented Generation for Adapting Large Language Models to Specialized Domains:** Предлагает адаптацию LLM к специализированным областям с использованием немаркированных корпусов. Адаптация больших языковых моделей к специализированным областям через самообучение. Метод включает настройку модели на retrieval-oriented данных и генерацию синтетических QA-пар на основе специализированных данных, что приводит к значительному приросту точности в задачах QA. Применяет самообучение для улучшения результатов, как и Self-Refinement Tuning.
- **Boosting Conversational QA with Fine-Grained Retrieval and Self-Check:** Ориентирован на улучшение диалоговых систем путем интеграции точного поиска информации и механизма самопроверки, что повышает качество и релевантность ответов в разговорных системах. Улучшение разговорного вопросно-ответного взаимодействия с помощью самопроверки. Использует механизмы самопроверки, что схоже с подходами в Self-Refinement и SimRAG.

Заключение

- Методы self-refinement демонстрируют высокую эффективность для улучшения качества языковых моделей (LLM), снижая их зависимость от ручной разметки и повышая точность генерации.
- Общие черты подходов:
  - Все методы направлены на улучшение производительности LLM через механизмы самооценки и самокоррекции.
  - Использование обратной связи, будь то от самой модели или от более продвинутых систем, для повышения качества ответов.
  - Стремление уменьшить зависимость от человеческого вмешательства, автоматизируя процесс совершенствования.
- Потенциал для применения:
  - Медицинские системы: точная диагностика и рекомендации.
  - Образовательные платформы: автоматическая проверка и помощь в обучении.
  - Бизнес-аналитика: улучшение взаимодействия в системах поддержки клиентов и анализа данных.
- Заключение:
  - Несмотря на общую цель повышения эффективности LLM, каждый из подходов предлагает уникальные методы и решения, подчеркивая богатство возможностей для дальнейшего развития области.

Сравнительный анализ методов Self-Refinement

Метод	Описание	Сходства	Метрика для оценки	Результаты из статей
Self-Refinement Tuning	Итеративное улучшение выходных данных модели через самооценку и обратную связь.	Все методы направлены на улучшение качества выходных данных. Используют итеративные подходы к обучению.	Точность (Accuracy)	Увеличение точности на 5-10% по сравнению с базовой моделью.
SimRAG	Адаптация больших языковых моделей к специализированным областям через самообучение.	Применяет самообучение для улучшения результатов, как и Self-Refinement Tuning.	F1-мера (F1 Score)	Увеличение F1-меры на 8% в специализированных задачах.
ConvRAG	Улучшение разговорного вопросно-ответного взаимодействия с помощью самопроверки.	Использует механизмы самопроверки, что схоже с подходами в Self-Refinement и SimRAG.	BLEU или ROUGE	Повышение ROUGE на 12% по сравнению с предыдущими версиями.
Self-Refine Instruction-Tuning	Обучение на инструкциях с последующим самоулучшением через оптимизацию предпочтений.	Все методы используют итеративные процессы и обратную связь для улучшения результатов.	NDCG (Normalized Discounted Cumulative Gain) | Увеличение NDCG на 15% при использовании инструкций для обучения.	
Self-Refinement	Общий подход к улучшению выходных данных через внутренние механизмы проверки и доработки.	Общая цель — улучшение качества ответов, использование обратной связи и итеративность.	Общая точность и качество ответов	Повышение общей точности на 7% в различных задачах.

Список литературы

- madaan2024self - Self-refine: Iterative refinement with self-feedback.
- xu-etal-2024-pride - Pride and Prejudice: LLM Amplifies Self-Bias in Self-Refinement.
- hu2024selfimprove - Teaching Language Models to Self-Improve by Learning from Language Feedback.
- ranaldi2024self - Self-Refine Instruction-Tuning for Aligning Reasoning in Language Models.
- xu2024simragselfimprovingretrievalaugmentedgeneration - SimRAG: Self-Improving Retrieval-Augmented Generation for Adapting Large Language Models to Specialized Domains.
- selfrefine2024 - SELF-REFINE: Iterative Refinement with Language Models Without Additional Supervision.