УДК 004.8

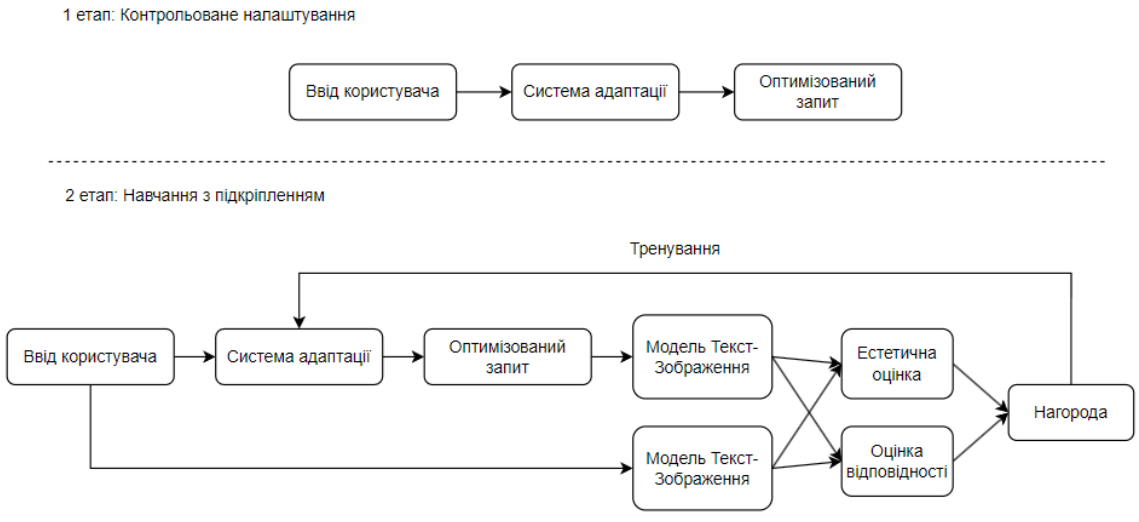
МЕТОДИ ОПТИМІЗАЦІЇ ГЕНЕРАЦІЇ ЗОБРАЖЕНЬ ЗА ДОПОМОГОЮ НЕЙРОННИХ МЕРЕЖ

Фесенко А.В.
Науковий керівник – к.т.н., доц. Рожнова Т. Г.
Харківський національний університет радіоелектроніки
(61166, Харків, просп. Науки,14, каф. АПОТ, тел. (057) 702-13-06)
e-mail: anton.fesenko1@nure.ua

The work analyzes an automatic system for adapting initial user input to model-preferred prompts using a pre-trained language model with supervised fine-tuning and reinforcement learning.

Розробка запитів має велике значення для генеративних моделей щоб слідувати інструкціям користувача та створювати високоякісний контент. Однак невелика ємність текстових кодерів у моделях «Текст-Зображення» часто вимагає редагування запиту. Попередні підходи реалізували ручну розробку запитів, але страждають від непрактичності та несумісності між версіями генеративної моделі. Таким чином, виникає потреба в систематичному способі автоматичної адаптації користувацьких намірів до бажаних запитів моделей. *Мета дослідження* – автоматична оптимізація запитів для моделей «Текст-Зображення», що дозволить узгодити введений користувачем текст з переважними запитами моделі та підвищить її продуктивність. *Джерела* – науково-технічна література на тему нейронних мереж та поточні моделі генерації зображень [1-5].

Модель оптимізації побудована на основі попередньо навченої моделі. Спочатку збирається набір розроблених користувачами прикладів для контрольованого налаштування. Згодом навчання з підкріпленням використовується для максимізації цільової винагороди, підвищуючи як відповідність, так і якість створених зображень.



Процес тренування системи

Розроблені людиною запити можна зібрати з результатів існуючих генеративних нейромереж, після чого з них можна виділити 2 компоненти: основний контент, що описує намір користувача, та модифікатори, які налаштовують стиль мистецтва. Для створення паралельних даних будуть використані три методи побудови початкових входів. По-перше, витягується основний вміст шляхом видалення модифікаторів, розглядаючи його як вхідні данні користувача. По-друге, деякі модифікатори випадково видаляються або перемішуються, залишивши решту тексту як початкові входи. По-третє, використовуючи API «Davinci», отриманий вміст та розроблені людиною запити перефразуються.

Оцінка ефективності оптимізованих запитів проходить за двома критеріями: відповідність та естетичність. Спочатку оцінюється актуальність створених зображень до початкового запиту після адаптації, тобто обирається результат моделі «Текст-Зображення», яка працювала з оптимізованим запитом. Згодом буде обчислена оцінка подібності CLIP, щоб оцінити чи актуальне створене зображення та оригінальний запит. Далі використовуватиметься естетичний предиктор для кількісної оцінки естетичних переваг. Він конструює лінійний оцінювач поверх фіксованої моделі CLIP, попередньо навченої з використанням людських оцінок з набору даних Естетичного Візуального Аналізу. Процес завершується визначенням загальної винагороди шляхом поєднання розрахованих оцінок з додатковим штрафом, який необхідний для зниження проблеми надмірної оптимізації.

Розроблена система автоматичної оптимізації запитів для генерації зображень, що використовує навчання з підкріпленням та контрольоване налаштування на попередньо навченій моделі. Гнучкість методу забезпечує безшовне узгодження між людськими намірами і мовою, яку сприяє генеративна модель.

Список використаних джерел:

1. Paul F Christiano, Jan Leike, Tom Brown, Miljan Martic, Shane Legg, & Dario Amodei. (2017). Deep reinforcement learning from human preferences. https://arxiv.org/abs/1706.03741.

2. John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, & Oleg Klimov. (2017). Proximal policy optimization algorithms. https://arxiv.org/abs/1707.06347.

3. Ashwin K. Vijayakumar, Michael Cogswell, Ramprasaath R. Selvaraju, Qing Sun, Stefan Lee, David J. Crandall, & Dhruv Batra. (2016). Diverse beam search: Decoding diverse solutions from neural sequence models. https://arxiv.org/abs/1610.02424.

4. Vivian Liu & Lydia B. Chilton. (2021). Design guidelines for prompt engineering text-to-image generative models. https://arxiv.org/abs/2109.06977.

5. Yongchao Zhou, Andrei Ioan Muresanu, Ziwen Han, Keiran Paster, Silviu Pitis, Harris Chan, & Jimmy Ba. (2022). Large language models are human-level prompt engineers. https://arxiv.org/abs/2211.01910.