УДК 004.8:514.1

Идрисов А.
учащийся 11 класса
Кокшетауский лицей «Білім-инновация»
Казахстан, г. Кокшетау
Турсын Г.
ментор
образовательная программа ADMIT
Казахстан

КООРДИНАТЫ ВМЕСТО ИДЕИ: КАК НЕБОЛЬШИЕ ЯЗЫКОВЫЕ МОДЕЛИ РЕШАЮТ ОЛИМПИАДНЫЕ ЗАДАЧИ ПО ГЕОМЕТРИИ

Аннотация: Небольшие открытые языковые модели всё чаще используются школьниками для проверки решений олимпиадных задач. По сохранённым ответам трёх моделей семейства Qwen2.5 (1,5, 3 и 7 млрд параметров) на 75 геометрических задач олимпиадного уровня из набора Omni-MATH рассмотрено, какими методами модели решают геометрию. Вычислительные методы – координаты, тригонометрию или векторы – использовали 17,3% решений самой маленькой модели, 24,0% модели на 3 млрд параметров и 46,7% модели на 7 млрд параметров. На задачах средней сложности модель на 7 млрд параметров перешла к вычислениям в 19 задачах из 25 и решила одну. Ни в одном из 75 решений этой модели не встретились инверсия, проективные преобразования, гомотетия или подобные треугольники. Решения с вычислениями были длиннее и чаще обрывались, не дойдя до ответа. Результаты согласуются с наблюдением автора-олимпиадника: модель реже замечает идею, которая сокращает решение, и вместо этого считает.

Ключевые слова: языковые модели, искусственный интеллект, олимпиадная математика, геометрия, метод координат, Omni-MATH

Idrissov A.
11th-grade student
Kokshetau Bilim-Innovation Lyceum
Kazakhstan, Kokshetau
Tursyn G.
mentor
ADMIT educational program
Kazakhstan

COORDINATES INSTEAD OF AN IDEA: HOW SMALL LANGUAGE MODELS SOLVE OLYMPIAD GEOMETRY PROBLEMS

Abstract: Small open language models are increasingly used by students to check solutions to olympiad problems. Using the saved answers of three Qwen2.5 models (1.5, 3 and 7 billion parameters) to 75 olympiad-level geometry problems from the Omni-MATH benchmark, this article examines which methods the models use in geometry. Computational methods (coordinates, trigonometry or vectors) appeared in 17.3% of the solutions of the smallest model, 24.0% of the 3-billion model and 46.7% of the 7-billion model. On middle-difficulty problems the 7-billion model turned to computation in 19 of 25 problems and solved one. None of this model's 75 solutions used inversion, projective transformations, homothety or similar triangles. Solutions with computation were longer and more often cut off before reaching an answer. The results agree with the observation of the author, an olympiad student: the model rarely notices the idea that shortens a solution and computes instead.

Keywords: language models, artificial intelligence, olympiad mathematics, geometry, coordinate method, Omni-MATH

Введение

Автор, решая олимпиадную геометрию, сначала старается нарисовать максимально точный чертёж, потому что по нему уже видна вся конструкция. Потом он часто идёт от обратного: думает, что должно быть верно, чтобы получить ответ, например какие углы или треугольники должны оказаться равными, и постепенно идёт назад, пока не заметит знакомую конфигурацию. Если ничего не получается, иногда переходит к координатам. По наблюдению автора, языковые модели в своих решениях, наоборот, часто сразу берутся за вычисления, координаты и простые методы и реже замечают места, где можно применить более сильную идею, например инверсию или проективное преобразование.

В предыдущей работе автора три небольшие открытые модели семейства Qwen2.5 были проверены на 287 олимпиадных задачах из набора Omni-MATH [1, 2]. Оказалось, что модели гораздо сильнее в алгебре, чем в геометрии: модель на 7 млрд параметров решила 50,7% задач по алгебре и лишь 17,3% по геометрии, причём разрыв не объясняется задачами, для которых нужен чертёж [3]. Та работа проверяла только итоговые ответы. Настоящая статья проверяет наблюдение автора по самим решениям: какими методами модели решают геометрию и как это связано с результатом.

Методы

Использованы сохранённые полные ответы трёх моделей Qwen2.5-Instruct (1,5, 3 и 7 млрд параметров, 4-битные версии, запуск на ноутбуке, жадная генерация, не более 1 024 новых токенов) на 75 геометрических задач из Omni-MATH: по 25 лёгких, средних и сложных [1–3]. Правильность ответов взята из проверки, выполненной в работе [3].

Методы в каждом решении определялись по ключевым словам. Вычислительными считались координаты (размещение точек в системе координат, упоминание осей и координат точек), тригонометрия (синусы, косинусы, теоремы синусов и косинусов) и векторы или комплексные числа. Отдельно искались синтетические приёмы олимпиадной геометрии: подобные и равные треугольники, вписанные четырёхугольники, степень точки, гомотетия, инверсия, проективные преобразования и другие. Одно решение может содержать несколько методов. Для каждого решения записаны его длина в токенах и то, оборвалось ли оно на ограничении длины.

Результаты

Чем больше модель, тем чаще она решает геометрию вычислениями. Хотя бы один вычислительный метод встретился в 13 решениях из 75 (17,3%) у модели на 1,5 млрд параметров, в 18 (24,0%) у модели на 3 млрд и в 35 (46,7%) у модели на 7 млрд (табл. 1). Чаще всего это координаты: у самой большой модели они есть в 27 решениях (36,0%).

Таблица 1. Методы в решениях 75 геометрических задач, число решений (доля, %)

Метод	**Qwen2.5 1,5 млрд**	**Qwen2.5 3 млрд**	**Qwen2.5 7 млрд**
Координаты	6 (8,0)	15 (20,0)	27 (36,0)
Тригонометрия	6 (8,0)	6 (8,0)	12 (16,0)
Векторы или комплексные числа	2 (2,7)	0	0
Хотя бы один вычислительный метод	13 (17,3)	18 (24,0)	35 (46,7)
Решено верно, всего	9 (12,0)	7 (9,3)	13 (17,3)

Синтетические приёмы, о которых пишет автор, модель на 7 млрд параметров почти не использует. Ни в одном из её 75 решений не встретились инверсия, проективные преобразования, гомотетия или

подобные треугольники; степень точки упоминается в трёх решениях, вписанный четырёхугольник – в одном.

Модель переходит к вычислениям прежде всего на задачах средней сложности: так она решала 19 задач из 25, и верный ответ получила в одной (табл. 2). На лёгких задачах вычисления встречались только в 4 решениях из 25. Внутри одной группы сложности решения с вычислениями и без них различаются мало, а общий разрыв в доле верных ответов (3 из 35 против 10 из 40) в основном объясняется тем, что вычисления модель выбирает на более трудных задачах.

Таблица 2. Модель на 7 млрд параметров: решения с вычислительными методами и без них по сложности задач

Сложность	**С вычислениями: решений / верных**	**Без вычислений: решений / верных**
Лёгкие (25 задач)	4 / 2	21 / 9
Средние (25 задач)	19 / 1	6 / 0
Сложные (25 задач)	12 / 0	13 / 1
Все (75 задач)	35 / 3	40 / 10

Решения с вычислениями длиннее: медианная длина 827 токенов против 628 у остальных решений модели на 7 млрд параметров. Из 15 её геометрических решений, оборвавшихся на ограничении длины без ответа, 10 приходятся на решения с вычислениями.

Иногда переход к координатам меняет саму задачу. В одной из задач средней сложности (HMMT) пятиугольник вписан в произвольный треугольник NOW; модель поместила точки N, O и W в (0, 0), (1, 0) и (0, 1), то есть сама сделала треугольник прямоугольным и равнобедренным, хотя в условии этого нет, и получила неверный ответ.

Обсуждение

Из результатов следуют три вывода. Во-первых, наблюдение автора подтверждается: небольшие модели решают олимпиадную геометрию в

основном вычислениями и почти не используют приёмы, которые олимпиадники применяют, чтобы сократить решение. Во-вторых, с ростом модели доля вычислений не падает, а растёт: модель на 7 млрд параметров считает в почти половине задач. В-третьих, вычисления не спасают модель на средних и сложных задачах: решения становятся длиннее, чаще обрываются и иногда опираются на допущения, которых нет в условии. Для школьников практический вывод тот же, что и в работе [3]: небольшой модели нельзя доверять проверку олимпиадной геометрии без самостоятельной проверки решения.

Ограничения работы. Методы определялись по ключевым словам, а не по полному прочтению каждого решения, поэтому отдельные решения могли быть отнесены к методу неточно. В каждой группе сложности 25 задач, и различия в один-два ответа случайны. Проверены только модели одного семейства и только с ограничением в 1 024 токена; с более длинным ответом часть вычислительных решений могла бы дойти до ответа. Вывод о связи метода с результатом наблюдательный: модель могла выбирать вычисления именно потому, что не находила другой идеи.

Заключение

Небольшие открытые языковые модели решают олимпиадную геометрию прежде всего координатами и тригонометрией, и чем больше модель, тем чаще она так делает. Синтетические приёмы, сокращающие решение, – инверсия, проективные преобразования, гомотетия, подобные треугольники – в решениях самой большой из проверенных моделей не встречаются вовсе. Это согласуется с наблюдением автора и объясняет, почему геометрия остаётся для таких моделей одним из самых трудных разделов олимпиадной математики.

Вклад авторов и использование ИИ. А. Идрисов написал первый абзац введения о том, как он сам решает геометрию и как, по его наблюдению, это делают модели; из этого наблюдения и выросла статья.

Он самостоятельно сверил долю решённых геометрических задач с опубликованной работой [3] и вручную пересчитал доли решений с вычислительными методами. Г. Турсын (ментор образовательной программы ADMIT) сформулировал вопрос статьи, выполнил анализ решений, написал программный код и текст с помощью ИИ-ассистента (Claude). Оба автора прочитали статью и согласны с ней.

Использованные источники:

1. Gao B., Song F., Yang Z., Cai Z., et al. Omni-MATH: A universal Olympiad level mathematic benchmark for large language models // arXiv. – 2024. – arXiv:2410.07985. URL: https://arxiv.org/abs/2410.07985 (дата обращения: 28.10.2026).

2. Qwen Team (Yang A., Yang B., et al.). Qwen2.5 technical report // arXiv. – 2024. – arXiv:2412.15115. URL: https://arxiv.org/abs/2412.15115 (дата обращения: 28.10.2026).

3. Idrissov A. Where do small open language models break on olympiad mathematics? Domain, difficulty and model size on Omni-MATH // Central Asian Scientific Journal. – 2026. – № 4(32), Т. 1. – С. 403–407. URL: https://cajournal.kz/article.php?id=48 (дата обращения: 28.10.2026).