

Бандиты в Query Selection

Шестаков Владимир Павлович

Московский физико-технический институт

Курс: Автоматизация научных исследований

Эксперт: Ю. В. Дорн

Консультант: И. М. Латыпов

2025



Задача и цель исследования

Задача

Воспроизвести результаты алгоритма Бао, модифицировать его различными способами и проверить

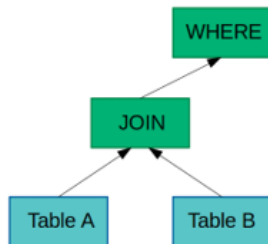
Цель

Исследовать влияние изменений алгоритма на качество

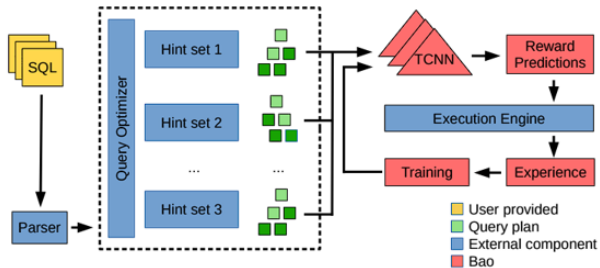
-  Marcus, Ryan и др. (июнь 2021). “Бао: Making learned query optimization practical”. В: *Proceedings of the 2021 International Conference on Management of Data*.
-  Slivkins, Aleksandrs (апр. 2019). *Introduction to Multi-Armed Bandits*. arXiv: 1904.07272 [cs.LG].

Начальная проблема: Query Optimization

1. Рост данных приводит к замедлению исполнения SELECT запросов
2. Решение: оптимизаторы запросов. Два вида оптимизаторов: статические и динамически изменяемые

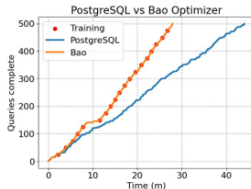
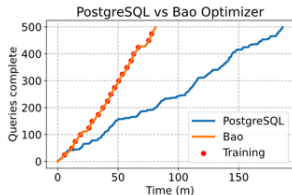


Алгоритм Bao



Первые результаты

Используется датасет, содержащий часть данных IMDb и 40 различных SELECT запросов. Алгоритм запускается с выбором 500 запросов из заданных с повторениями.



Дальнейшие эксперименты

1. Подбор гиперпараметров (grid search)
2. Замена последнего слоя на байесовский (распределения — логарифм от равномерного и нормальное)

Дальнейшие эксперименты

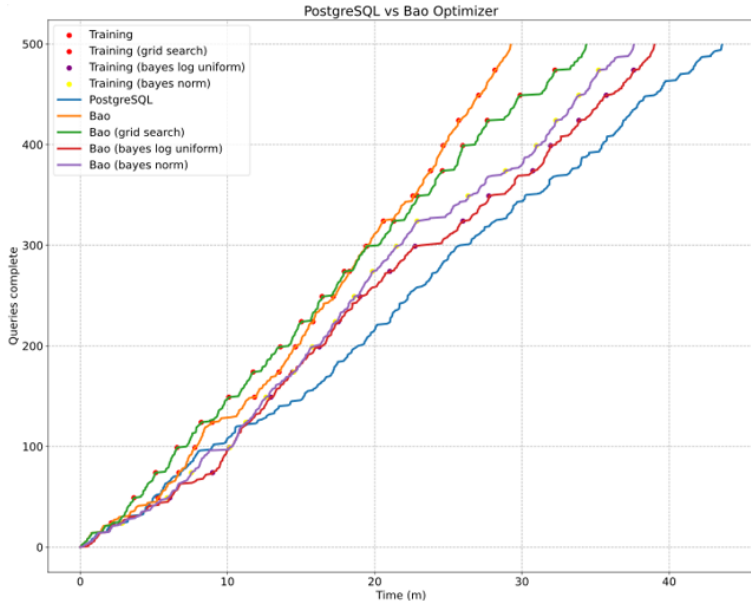


График CDF

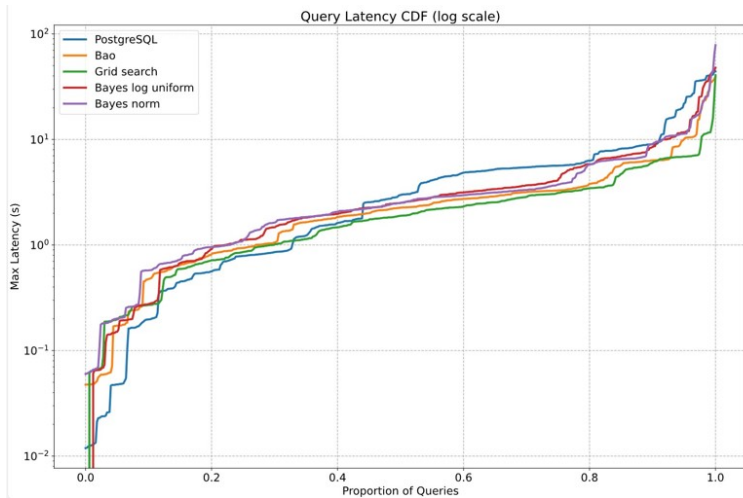


Таблица сравнения результатов экспериментов по некоторым запросам

	PG	Bao worst	Bao best	Grid search worst	Grid search best	Bayes log uniform worst	Bayes log uniform best	Bayes norm worst	Bayes norm best
Q									
q1	6.06	5.81	2.25	2.57	2.41	6.83	1.56	6.27	2.15
q2	38.62	4.01	3.23	3.54	1.96	44.90	2.99	42.53	3.17
q3	1.60	3.25	2.77	3.51	1.50	3.68	2.02	62.80	1.83
q4	1.04	2.28	0.88	1.16	1.03	2.26	1.01	2.15	0.95
q5	8.57	1.99	1.77	2.39	0.63	10.47	1.93	2.03	1.80
q6	8.20	2.71	2.52	13.41	2.69	10.56	2.68	12.53	2.49
q7	17.42	19.51	3.11	15.93	1.35	18.39	1.66	17.96	1.60
q8	6.31	2.49	2.13	2.38	2.17	6.92	0.70	6.22	2.13
q9	5.71	2.82	2.72	3.12	1.42	5.85	2.97	5.52	2.79
q10	3.57	3.67	3.33	3.73	1.41	4.05	1.54	3.69	3.35
q11	7.31	2.02	2.02	2.09	2.09	2.22	2.22	2.04	2.04
q12	4.93	1.90	1.64	1.99	0.30	5.94	1.78	5.05	1.72

- ▶ Получены результаты работы модификаций алгоритмов
- ▶ Некоторые из модификаций лучше изначального алгоритма по определённым качествам
- ▶ Есть потенциал дальнейшего улучшения