Work Progress

kNN Search with Parallel Incremental Query Answering

Jaouhara Chanchaf

Tuesday Nov. 17th, 2022

1. Summary

Done:

AI 1	Kashif PQA: use an array and a job counter instead of a job queue.
AI 2	Write a pseudo code for Kashif PQA.

In progress:

AI 3	Plot NN distance and recall.
AI 4	Measure query time when #threads changes.

Issues:

- Increasing recall plot is only true when measuring recall based on brute force results for identical NN search (recall != precision).
- Parallel QA, coordinator blocks on barrier but worker thread has finished NN search.

2. Performance

Experiment over 1GB of data: 100k tables, 494k columns, 5M vectors. 10 Queries of size [50 - 100].

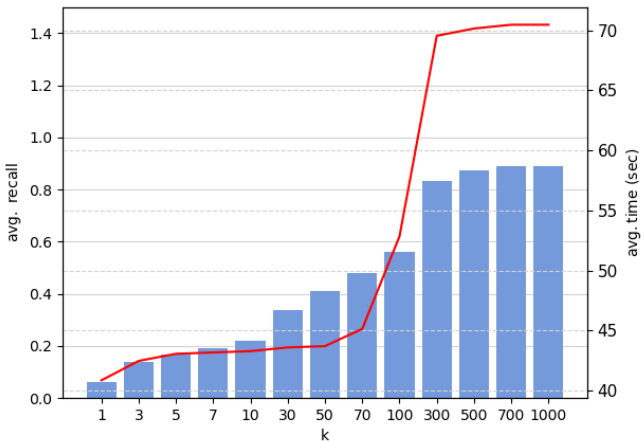

Figure: Kashif Recall based on identical NN

2. Performance

Experiment over 1GB of data: 100k tables, 494k columns, 5M vectors. 10 Queries of size [50 - 100].

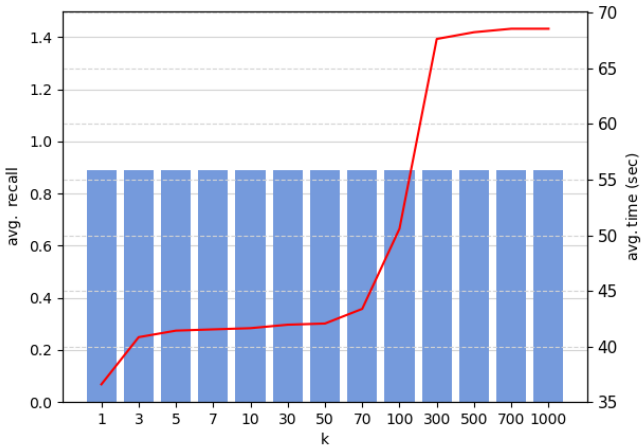

Figure: Kashif Recall

3. Discussion

1. Recall for Non identical NN is always close to 1, even for small k values.
2. Measuring recall based on the NN distance (how many NNs of the same distance to the query kashif has missed).
3. Review internship report structure (it is not complete).