Work Progress

Demo: KNN Search with Incremental Query Answering

Jaouhara Chanchaf

October 11th, 2022

Summary

In progress:

AI 2	Kashif: Only store a recall matrix at the coordinator thread and update NN recall incrementally.
AI 3	Search in the literature for techniques to estimate recall.

Not started:

AI 1	Kashif: Implement Parallel Incremental Query Answering with multiple worker-threads.

Discussion | Recall guarantees

1. Recall guarantee in PUFFINN (M. Aumüller et al. LIPIcs, 2019):

   - **What:** How to perform parameter-less, probabilistic KNN search?
   - **Why:** Exact search is a hard problem, probabilistic approaches suffer from low recall. One must find the optimal parameter values to achieve good performance.
   - **How:** An LSH-based Index that only requires recall guarantee and index memory size as parameters to perform KNN search.

   PUFFINN performs $k - \delta$ KNN search. each NN is guaranteed to be exact with probability 1 - δ, the expected recall is $(1 - \delta)k$.
   To satisfy the recall guarantee, the algorithm doesn't stop until j' trees are process at level i', j' is expressed in terms of δ.

Discussion | Kashif Parallel IQA - Code optimization

2. Reducing the knn array size in Parallel IQA:
   - Approximate search requires the full KNN array.
   - Some approximate results appear in the final KNN array. Hence, reducing the size of the KNN array in approximate search might decrease result accuracy.