CS754 Project Proposal

FLINNG: Fast Locality-Sensitive Hashing for Approximate Near Neighbor Search via Group Testing

Saksham Rathi (22B1003), Ekansh Ravi Shankar (22B1032), Kshitij Vaidya (22B1829)

Description

This project aims to implement the FLINNG (Filters to Identify Near-Neighbor Groups) algorithm for approximate near neighbor search as described in the NeurIPS 2021 paper. FLINNG combines group testing with locality-sensitive hashing (LSH) to provide a fast, memory-efficient solution for high-dimensional nearest neighbor search. The algorithm transforms the near neighbor search problem into a group testing problem by using distance-sensitive Bloom filters to identify groups containing near neighbors. Unlike traditional LSH approaches, FLINNG avoids expensive distance computations during queries and can be constructed in a single pass through the data. The implementation will focus on building the core components of the FLINNG algorithm, including the group testing framework, distance-sensitive Bloom filters, and the threshold relaxation algorithm.

Objectives

- Implement the FLINNG algorithm for approximate near neighbor search as described in the paper
- Optimize the implementation for practical performance using the techniques outlined in Section 6
- Evaluate the algorithm's performance on real-world datasets and compare with baseline methods
- Analyze the trade-offs between query time, memory usage, and precision/recall

Datasets

- RefSeqG
- RefSeqP
- PromethION
- URL
- Webspam
- YFCC100M