Fast Approximated Nearest Neighbor Joins For Relational Database Systems

Vector Data in Database Systems

- Word2Vec / Glove: Word Embeddings
- SIFT / GIST: Image Descriptors
- Deepwalk / Node2Vec: Graph Embeddings

Enables
- Similarity Queries
- Inductive Reasoning (Analogies)

Requires
- Integration of Fast Vector Search Operations

Fast Inverted Search Algorithm

1. Preprocessing
   - Precompute distance values
   - Adaptive to target set size

2. Query Construction
   - Determine \( \omega \) (Number of Part.)
   - Estimate number of retrieved entries (Probabilistic Approach)

3. Data Retrieval
   - Retrieve kNN candidates

4. Distance Calculation
   - User can adjust trade-off between precision and response time

Integration of kNN-Join Operation

\[
\text{kNN}(R \bowtie T) = \{(r, t) | t \in \text{kNN}(r, T), r \in R\}
\]

Challenges
- Batch-wise kNN Search for large query sets
  - Reduce interface and retrieval times
- High Dimensional Data
  - Previous Work mainly focus on low dimensional data (e.g. geographical data)
- Adaptive kNN-Join Algorithm
  - Different cardinalities of join operands
  - Only one index for all vectors
- Different Demands on Precision and Response Time

Inverted Indexing

- Token Vector Inception
- Coarse Quantization
- Database Retrieval
- kNN Results

Evaluation of Algorithm with Different Distance Calculation Methods

Product Quantization

Fast Approximated Search Method

Exact Distance Calculation

Fast Approximated Search Method

Post Verification

Post verify candidate solutions obtained by product quantization

Setup

- 5,000 query vectors
- 50,000 target vectors

Parameter

- Population Size \( N \): Number of all index vectors
- Number of success states in population \( K \): Number of targets
- Number of draws \( d \): Number of vectors in selected partitions
- Number of Successes \( m \): Number of targets in selected partitions

Web Demo

Word Embedding operations that use the kNN-Join Index can be explored in our web demo.

Contact: Michael.Guenther@tu-dresden.de  https://www.db.inf.tu-dresden.de/research-projects/freddy/