A Fast Image Retrieval System with Adjustable Objectives

TU Berlin IDA, TU Berlin/DFKI DIMA, Fraunhofer HHI
IDA’s Role in BBDC

AP1: Material science
AP2: Video mining
AP3: Language processing
AP4: Image analysis in medicine

AP5: Statistical DA & ML
- Summarize requirements from application partners.

AP6: Scalable ML
- Develop new methods to be added to Technology X.

Data
- Huge
- Complex-structured
- Multi-modal
- Distributed
- Streaming
- Partially-observed

Non-trivial, further scalability, justification

Trivially parallelizable (pre-processing, linear regression, etc.)

IDM (AP7/8)

DIMA (AP7/8)

Flink

© BBDC
Requirements & Solutions

AP1: Material science
AP2: Video mining
AP3: Language processing
AP4: Image analysis in medicine

Data
- Huge
- Complex-structured
- Multi-modal
- Distributed
- Streaming
- Partially-observed

Requirements
- Efficient
- Distributed
- Handy
- Deep
- Online
- Transductive

+ Classification
  - Regression
  - Clustering
  - Inference
  - Retrieval

Solutions
- Data indexing
- Deep learning
- Decomposable optimization
- Multi-modal analysis
Projects done

Data indexing:
- Multi-purpose Locality Sensitive Hashing (mpLSH)

Deep learning:
- Deep Tensor Neural Networks
- Explaining Non-linear Machine Learning

Decomposable Optimization:
- Multi-class SVM for Extreme Classification
- Parallel Matrix Factorization
- Polynomial-time Message Passing for High-order Potentials
- Performance Guarantee of Approximate Bayesian Learning

Multi-modal Analysis:
- Multi-modal Source Power Co-modulation (mSPoC)
- Transductive Conditional Random Field Regression (TCRFR)

Please see our poster!
Multiple-purpose
Locality Sensitive Hashing
Nearest neighbor search (NNS)

Naïve implementation (linear scan) requires $O(N)$ time.

$$\hat{x} = \arg\min_n \| q - x_n \|^2.$$
Locality sensitive hashing (LSH)

LSH enables approximate NNS in $O(N^\rho \log N)$ time for $0 < \rho < 1$

$h : \mathbb{R}^L \to \mathbb{N}$ s.t. $P(h(x) = h(x')) > P(h(x) = h(x''))$ if $||x - x'||^2 < ||x - x''||^2$.

Random projections provide good LSH functions.
Locality sensitive hashing (LSH)

LSH enables approximate NNS in $O(N^\rho \log N)$ time for $0 < \rho < 1$

$h : \mathbb{R}^L \rightarrow \mathbb{N}$ s.t. $\mathbb{P}(h(x) = h(x')) > \mathbb{P}(h(x) = h(x''))$ if $\|x - x'\|^2 < \|x - x''\|^2$.

Only samples in the same bucket should be evaluated.
Motivation

Different similarity requires different LSH codes.

- \( L_{L2}(q, x) = ||q - x||_2 \) for L2 similarity [Datar et al. 2004]
  \[
  h_{a}^{\text{sign}}(x) = \text{sign}(a^\top x)
  \]

- \( L_{\text{cos}}(q, x) = 1 - \frac{q^\top x}{||q||_2 ||x||_2} \) for cosine similarity [Coemans & Williamson 1995]
  \[
  h_{a,b}^{L2}(x) = \left[ R^{-1}(a^\top x + b) \right]
  \]

- \( L_{\text{ip}}(q, x) = -q^\top x \) for inner product (IP) similarity [Neyshabur & Srebro 2015]
  \[
  h_{a}^{\text{emp-a}}(q) = \text{sign}(a^\top (q; 0))
  h_{a}^{\text{emp-x}}(x) = \text{sign}(a^\top (x; \sqrt{1 - ||x||_2^2}))
  \]
Motivation

Different similarity requires different LSH codes.

Multi-purpose similarity:

$$L_{mp}(\{q^{(w)}\}, x) = \sum_{w=1}^{W} \sum_{g=1}^{G} \left( \gamma_g^{(w)} |q_g^{(w)} - x_g|^2 - 2\eta_g^{(w)} \frac{q_g^{(w)\top} r_g}{|q_g^{(w)}||x_g||} - 2\lambda_g^{(w)} q_g^{(w)\top} x_g \right)$$

General LSH coding for multi-purposes?
Multi-purpose LSH (mpLSH)

Different similarity requires different LSH codes.

Multi-purpose similarity:

$$\mathcal{L}_{mp}({q}^{(w)}, x) = \sum_{w=1}^{W} \sum_{g=1}^{G} \left( \gamma_g^{(w)} ||q_g^{(w)} - x_g||_2^2 - 2\delta_g^{(w)} \frac{q_g^{(w)^\top} x_g}{||q_g^{(w)}||_2 ||x_g||_2} - 2\lambda_g^{(w)} q_g^{(w)^\top} x_g \right)$$

Feature augmentation for metric transform

$${\tilde{q}}$$  

$${\tilde{x}}$$

LSH coding by random projection

$$\tilde{h}(\tilde{q})$$  

$$\tilde{h}(\tilde{x})$$

Common part

Similarity dependent part

Multi-metric search by cover tree

Multi-metric in code space:

$$\mathcal{D}_{\text{GA}}({q}^{(w)}, x) = \sum_{g=1}^{G} \left( \alpha_g \sum_{l=1}^{L} |h^{(w)}_{g,l}(\tilde{q}) - h^{(w)}_{g,l}(\tilde{x})| + \sum_{m=1}^{2} \beta_{g,m} |j^{(w)}_{g,m}(\tilde{q}) - j^{(w)}_{g,m}(\tilde{x})| \right)$$
Multi-purpose LSH (mpLSH)

Theoretical and empirical performance validation:

- **(Conditional) LSH property** theoretically guaranteed.
- **Approximate search performance** validated on several read data.
- **Computational and memory efficiency** proven on large data (100M samples).

Theorem 1 For \( q^{(w)} = \lambda^{(w)} = 0 \), \( \forall w \), i.e., \( L_{\text{mp}}(\{q^{(w)}\}, x) \) is the cosine similarity, it holds that \( P(\mathcal{D}_{\text{CA}}(\{q^{(w)}\}, x) = 0) = F_{\text{mp}}(1 + \frac{1}{2}L_{\text{mp}}(\{q^{(w)}\}, x))^T \).

Theorem 2 For \( q^{(w)} = \eta^{(w)} = 0 \), \( \forall w \), i.e., \( L_{\text{mp}}(\{q^{(w)}\}, x) \) is the IP similarity, the expectation of the mp-LSH-CA code similarity is bounded as

\[
\frac{2p}{\sigma}(2 + L_{\text{mp}}(\{q^{(w)}\}, x)) \leq \mathbb{E}(\mathcal{D}_{\text{CA}}(\{q^{(w)}\}, x)) \leq \sqrt{\frac{T}{T}}(2 + L_{\text{mp}}(\{q^{(w)}\}, x)) + 1.
\]
Multi-purpose LSH (mpLSH)

Characteristics
- Random projection based (data-independent).
- Query-time weight adjustment supported by cover tree.
- Similar memory requirement to sign-LSH.

Applications:
- General purpose indexing at data collection phase (without fixed analysis plan).
- Material search with optimized properties (e.g., stability, utility, etc.)
- Image/video retrieval with adjustable query (e.g., preference + closeness).
Demo: Image retrieval with adjustable objective

http://bbdcdemo.bbdc.tu-berlin.de

Retrieve similar images to user-provided query (L2-query) with user-preference query (IP-query) taken into account. Mixing weight is adjusted in real-time.

Closest to L2-query

Best match IP-query

Most relevant to mixed query

Query search in ~100msec!

Please try the demo!