Neighbor-Sensitive Hashing

Yongjoo Park   Michael Cafarella   Barzan Mozafari

University of Michigan, Ann Arbor

Nov 05, 2015

k-Nearest Neighbors Search (kNN)

k-Nearest Neighbors Search (kNN)

$q = \{120, 65, \ldots, 212\}$

Database

k-Nearest Neighbors Search (kNN)

k-Nearest Neighbors Search (kNN)

Database

$q = \{120, 65, \ldots, 212\}$
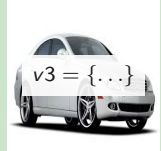
$v1 = \{\ldots\}$ $v2 = \{\ldots\}$ $v3 = \{\ldots\}$

$v4 = \{\ldots\}$ $v5 = \{\ldots\}$ $v6 = \{\ldots\}$

q v_1 v_2 v_3 v_4 v_5 v_6

0

distance from q

k-Nearest Neighbors Search (kNN)

Database

$q = \{120, 65, \ldots, 212\}$
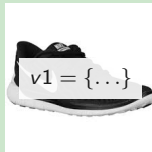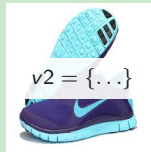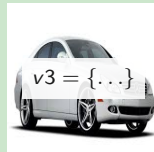
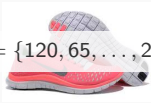$v1 = \{\ldots\}$ $v2 = \{\ldots\}$ $v3 = \{\ldots\}$
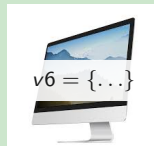
$v4 = \{\ldots\}$ $v5 = \{\ldots\}$ $v6 = \{\ldots\}$
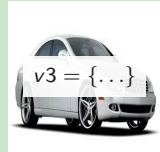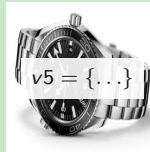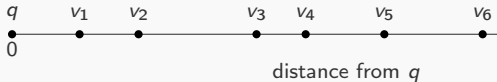
Exact kNN for high-dimensional vectors is *Slow*.

► (Kashyap *KDD*'11) took more than 1 min for 10M objects.

Hashing-based kNN

$\{201, 101, 43, 188, \ldots\}$

Regular Search

$\{$answer set$\}$
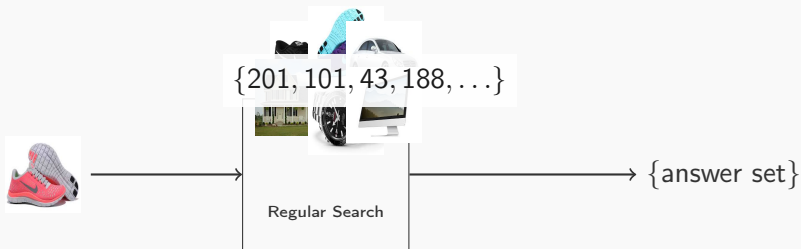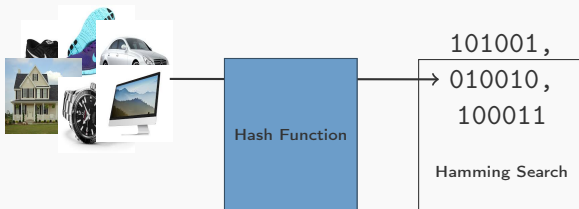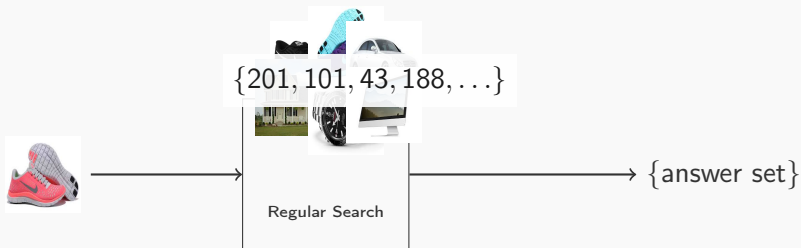
Hashing-based kNN
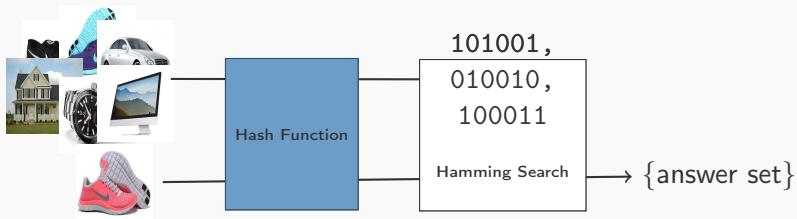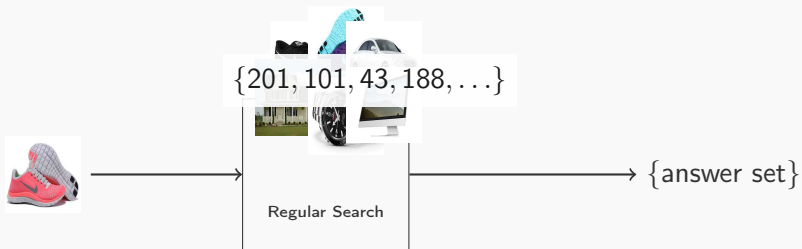
Hashing-based *k*NN

Hashing-based kNN
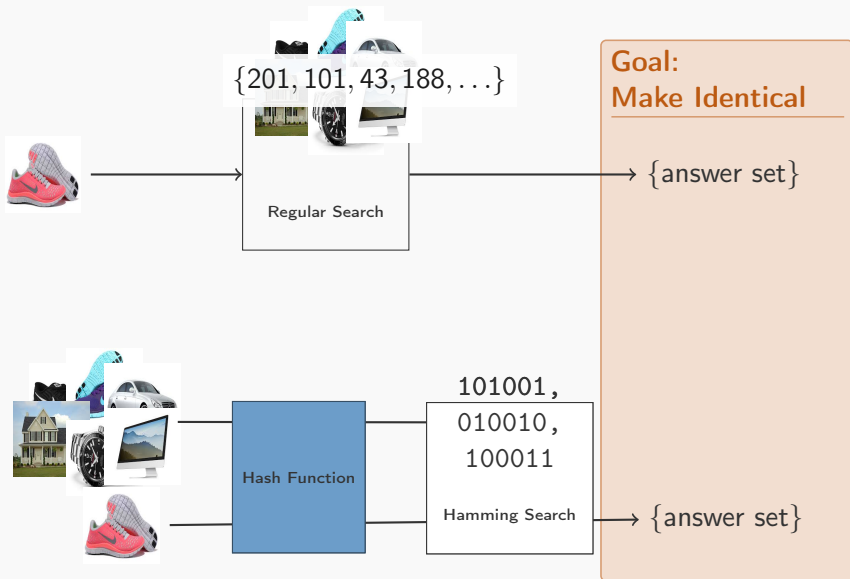
Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.

- 2012, Spherical Hashing (SpH), Heo at al.
    Kernelized Supervised Hashing (KSH), Liu et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.

- 2012, Spherical Hashing (SpH), Heo at al.
  Kernelized Supervised Hashing (KSH), Liu et al.

- 2013, Compressed Hashing (CH), Lin et al.
  Complementary Projection Hashing (CPH), Jin et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.

- 2012, Spherical Hashing (SpH), Heo at al.
  Kernelized Supervised Hashing (KSH), Liu et al.

- 2013, Compressed Hashing (CH), Lin et al.
  Complementary Projection Hashing (CPH), Jin et al.

- 2014, Data Sensitive Hashing (DSH), Jagadish et al.

Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.

- 2012, Spherical Hashing (SpH), Heo at al.
  Kernelized Supervised Hashing (KSH), Liu et al.

- 2013, Compressed Hashing (CH), Lin et al.
  Complementary Projection Hashing (CPH), Jin et al.

- 2014, Data Sensitive Hashing (DSH), Jagadish et al.

LSH

q v_1 v_2 v_3 v_4 v_5 v_6 v_7 v_8
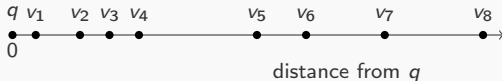
0

distance from q
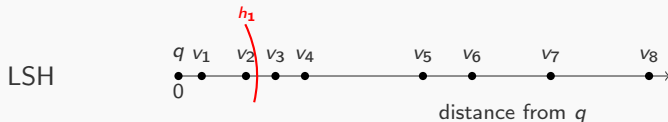
Previous Approaches

- 2004, Locality Sensitive Hashing (LSH), Datar et al.

- 2009, Spectral Hashing (SH), Weiss et al.

- 2011, Anchor Graph Hashing (AGH), Liu et al.
- 2012, Spherical Hashing (SpH), Heo at al.
      Kernelized Supervised Hashing (KSH), Liu et al.

- 2013, Compressed Hashing (CH), Lin et al.
      Complementary Projection Hashing (CPH), Jin et al.

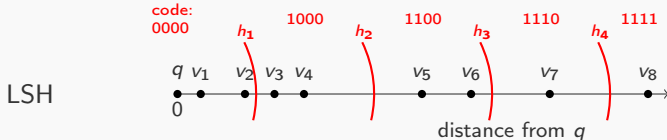- 2014, Data Sensitive Hashing (DSH), Jagadish et al.

LSH

Previous Approaches

2004, Locality Sensitive Hashing (LSH), Datar et al.

2009, Spectral Hashing (SH), Weiss et al.

2011, Anchor Graph Hashing (AGH), Liu et al.

2012, Spherical Hashing (SpH), Heo at al.
        Kernelized Supervised Hashing (KSH), Liu et al.

2013, Compressed Hashing (CH), Lin et al.
        Complementary Projection Hashing (CPH), Jin et al.

2014, Data Sensitive Hashing (DSH), Jagadish et al.

LSH

Previous Approaches

2004, Locality Sensitive Hashing (LSH), Datar et al.

> **Good idea for:**
>
> ► Sorting all data items

2012, Spherical Hashing (SpH), Heo at al.
    Kernelized Supervised Hashing (KSH), Liu et al.

2013, Compressed Hashing (CH), Lin et al.
    Complementary Projection Hashing (CPH), Jin et al.
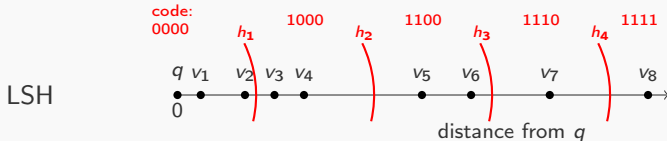
2014, Data Sensitive Hashing (DSH), Jagadish et al.

LSH

Previous Approaches

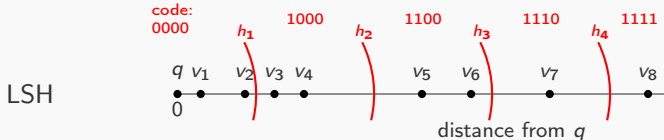2004, Locality Sensitive Hashing (LSH), Datar et al.

Good idea for:

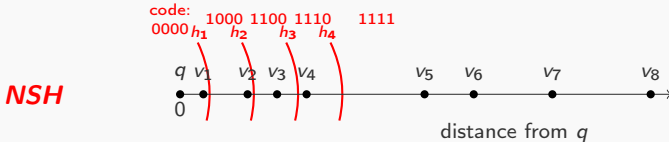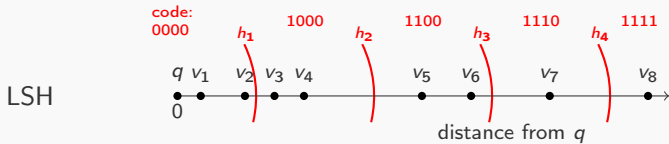▸ Sorting all data items

2012, Spherical Hashing (SpH), Heo at al.

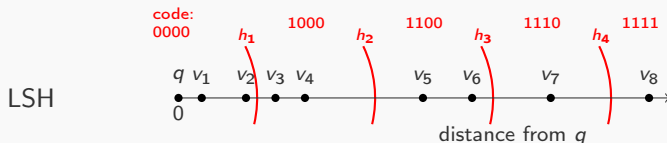*Remember:*

▸ We are interested in only *k* items.

2014, Data Sensitive Hashing (DSH), Jagadish et al.
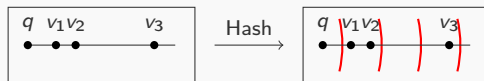
LSH

Our Approach

LSH
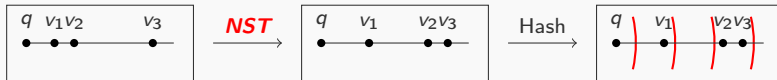
NSH

Our Approach

LSH

NSH
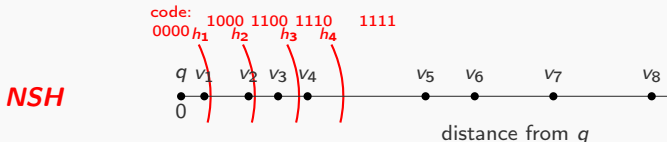
More Separators between Neighbors

- Neighbors close to the query are better distinguished
- Low Resolution for distant items $\implies$ *No Problem!*
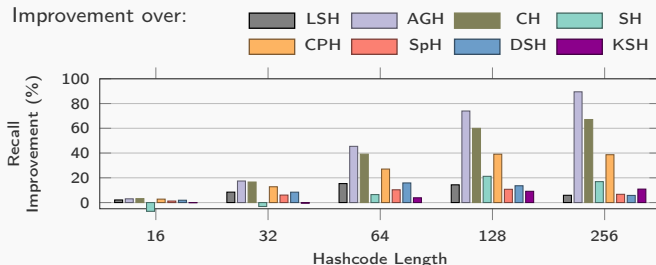
Algorithmic Details

(a) Regular Hashing

(b) Hashing with NST

A special Non-Linear Transformation

▶ Expand the distance between Neighbors

▶ Effectively, more hash functions for Neighbors
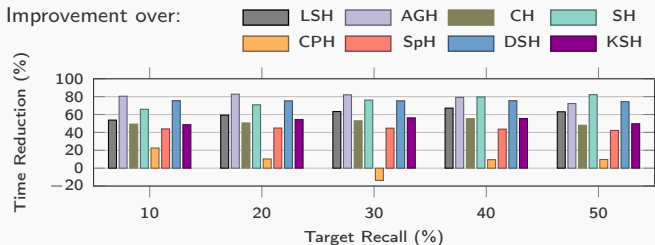
Recall Improvement over State-of-the-art

Hashcode size vs. Recall Improvement

- ▸ Up to 10% recall improvement over SpH [2]
- ▸ Up to 15% recall improvement over LSH [1]

(The SIFT dataset)

Speed Improvement over State-of-the-art

Improvement over: LSH, AGH, CH, SH, CPH, SpH, DSH, KSH

Target Recall vs. Time Reduction

- ▶ Up to 22% time reduction over CPH [3]
- ▶ Up to **67%** time reduction over LSH [1]

(The SIFT dataset)

Questions

[1] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni.
Locality-sensitive hashing scheme based on p-stable distributions.
In *SoCG*, 2004.

[2] J.-P. Heo, Y. Lee, J. He, S.-F. Chang, and S.-E. Yoon.
Spherical hashing.
In *CVPR*, 2012.

[3] Z. Jin, Y. Hu, Y. Lin, D. Zhang, S. Lin, D. Cai, and X. Li.
Complementary projection hashing.
In *ICCV*, 2013.