Multilinear Hyperplane Hashing

Xianglong Liu[†] Xinjie Fan[†] Cheng Deng[‡] Zhujin Li[†] Hao Su[§] Dacheng Tao[♮]

[†]State Key Lab of Software Development Environment, Beihang University, Beijing, China

[‡]School of Electronic Engineering, Xidian University, Xi'an, China

[§]Department of Computer Science, Stanford University, Stanford, CA, USA

[♮]QCIS and FEIT, University of Technology, Sydney, Australia

xlliu@nlsde.buaa.edu.cn chdeng.xd@gmail.com dacheng.tao@uts.edu.au

1. The Proof

Theorem 1 *Given a database $\mathcal{D}$ with n points and a hyperplane query $\mathcal{P}_{\boldsymbol{w}}$, if there exists a database point $\boldsymbol{x}^*$ such that $d(\boldsymbol{x}^*, \mathcal{P}_{\boldsymbol{w}}) \leq r$, then with $\rho = \frac{\ln p_1}{\ln p_2}$ (1) using n^ρ hash tables with $\log_{1/p_2} n$ hash bits, the random multilinear hyperplane hash of an even order is able to return a database point $\hat{\boldsymbol{x}}$ such that $d(\hat{\boldsymbol{x}}, \mathcal{P}_{\boldsymbol{w}}) \leq r(1+\epsilon)$ with probability at least $1 - \frac{1}{c} - \frac{1}{e}$, $c \geq 2$; (2) the query time is sublinear to the entire data number n, with $n^\rho \log_{1/p_2} n$ bit generations and cn^ρ pairwise distances computation.*

This can be completed easily following prior research [1,2].

Proof 1 *Denote the number of hash tables to be L. For the l-th hash table, the proposed MH-Hash algorithm randomly samples k hash functions $h^m_{l,1}, \cdots, h^m_{l,k}$ with replacement from M_m, which will generate a k-bit hash key for each input data vector $\boldsymbol{x}$ We denote $\boldsymbol{x}$'s hash code by $\mathrm{H}^m_l(\boldsymbol{x}) = [h^m_{l,1}(\boldsymbol{x}), \cdots, h^m_{l,k}(\boldsymbol{x})]$. The main observation is that using $\mathrm{L} = n^\rho$ independent hash tables, a $(1+\epsilon)$-appropriate nearest neighbor is achieved with a nontrivial constant probability. Moreover, the query (search) time complexity is proved to be sub-linear with respect to the entire data number n.*

To complete the proof, we define the following two events $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$. It suffices to prove the theorem by showing that both $\boldsymbol{F}_1$ and $\boldsymbol{F}_2$ hold with probability larger than 0.5. The two event are defined as follows:

$\boldsymbol{F}_1$: If there exists a database point $\boldsymbol{x}^*$ such that $d(\boldsymbol{x}^*, \mathcal{P}_{\boldsymbol{w}}) \leq r$, then $\mathrm{H}^m_l(\boldsymbol{x}^*) = \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})$ for some $1 \leq l \leq \mathrm{L}$.

$\boldsymbol{F}_2$: Provided with a false alarm set

$$\mathscr{S} = \{\check{\boldsymbol{x}} \mid \check{\boldsymbol{x}} \in \mathscr{X} \text{ such that } \mathrm{D}(\check{\boldsymbol{x}}, \mathcal{P}_{\boldsymbol{w}}) > r(1+\epsilon) \text{ and } \exists l \in [1:\mathrm{L}], \mathrm{H}^m_l(\check{\boldsymbol{x}}) = \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})\},$$

where $\epsilon > 0$ is the given small constant. Then the set cardinality $|\mathscr{S}| < c\mathrm{L}$.

*First, we prove that $\boldsymbol{F}_1$ **holds with probability at least** $1 - \frac{1}{e}$.*

Let us consider the converse case that $\mathrm{H}^m_l(\boldsymbol{x}^*) \neq \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})$ for $\forall l \in [1:\mathrm{L}]$ whose probability is

$$\begin{aligned}
&\mathbf{P}[\mathrm{H}^m_l(\boldsymbol{x}^*) \neq \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}}), \forall l \in [1:\mathrm{L}]] \\
&= (\mathbf{P}[\mathrm{H}^m_l(\boldsymbol{x}^*) \neq \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})])^{\mathrm{L}} \\
&= (1 - \mathbf{P}[\mathrm{H}^m_l(\boldsymbol{x}^*) = \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})])^{\mathrm{L}} \\
&\leq (1 - {p_1}^k)^{\mathrm{L}} = (1 - {p_1}^{log_{\frac{1}{p_2}} n})^{n^\rho} = (1 - n{-}\rho)^{n^\rho} \\
&= ((1 - n^{-\rho})^{-n^\rho})^{-1} \leq \frac{1}{e},
\end{aligned}$$

where inequality (1) follows the inequality $(1 - n^{-\rho})^{-n^\rho} \geq e$. Herewith we derive

$$\begin{aligned}
&\mathbf{P}[\mathrm{H}^m_l(\boldsymbol{x}^*) = \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}}), \exists l \in [1:\mathrm{L}]] \\
&= 1 - \mathbf{P}[\mathrm{H}^m_l(\boldsymbol{x}^*) \neq \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}}), \forall l \in [1:\mathrm{H}]] \\
&\geq 1 - \frac{1}{e}.
\end{aligned}$$

*Second, we prove that $\boldsymbol{F}_2$ **holds with probability at least** $1 - \frac{1}{c}$.*

For every false alarm point $\check{\boldsymbol{x}}$ conforming to $\mathrm{D}(\check{\boldsymbol{x}}, \mathcal{P}_{\boldsymbol{w}}) > r(1+\epsilon)$, in any hash table $l \in [1:\mathrm{L}]$ we have

$$\mathbf{P}[\mathrm{H}^m_l(\check{\boldsymbol{x}}) = \mathrm{H}^m_l(\mathcal{P}_{\boldsymbol{w}})] < {p_2}^k = (p_2)^{log_{\frac{1}{p_2}} n} = \frac{1}{n}.$$

Therefore the expected number of false alarm points, which fall into the same hash bucket with the query $\mathcal{P}_{\boldsymbol{w}}$ in hash table l, is smaller than $n \times 1/n = 1$. Immediately, we *conclude* $\mathbf{E}[| \mathscr{S} |] < \mathrm{L}$. *Subsequently, we further apply Markov's inequality to derive the following result:*

$$\mathbf{P}[| \mathscr{S} | \geq c\mathrm{L}] \leq \frac{\mathbf{E}[| \mathscr{S} |]}{c\mathrm{L}} < \frac{\mathrm{L}}{c\mathrm{L}} = \frac{1}{c},$$

which leads to

$$\mathbf{P}[| \mathscr{S} | < c\mathrm{L}] = 1 - \mathbf{P}[| \mathscr{S} | \geq c\mathrm{L}] > 1 - \frac{1}{c}.$$

Third, we prove that $\boldsymbol{F}_1$ *and* $\boldsymbol{F}_2$ ***simultaneously hold with probability at least*** $1 - \frac{1}{c} - \frac{1}{e}$.

References

[1] P. Jain, S. Vijayanarasimhan, and K. Grauman. Hashing Hyperplane Queries to Near Points with Applications to Large-Scale Active Learning. In *NIPS*, pages 928–936. 2010. 1

[2] W. Liu, J. Wang, Y. Mu, S. Kumar, and S.-F. Chang. Compact hyperplane hashing with bilinear functions. In *ICML*, pages 1–8, 2012. 1