CSE 254 Handout

Nearest Neighbor Preserving Embeddings

(paper by Piotr Indyk and Assaf Naor)

Konstantin Pervyshev
pervyshev@cs.ucsd.edu

1 Nearest Neighbor Search

Problem 1 (The nearest neighbor problem)**.** *Given a set $X \subset \mathbb{R}^n$, build a data structure which given any query $x \in \mathbb{R}^n$, quickly reports the point x' in X that is (approximately) closest to x.*

There are efficient data structures for this problem. In particular, there exist $(1+\epsilon)$-approxmate nearest neighbor data structures with space complexity $O(|X|/\epsilon^n)$ and query time $O(n \log(|X|/\epsilon))$ [HP01, AM02]. However, in case the dimension n is big, the complexity of these data structures is not satisfactory.

Indeed, some special methods are known for the case the "intrinsic" dimension of X is smaller than the dimension of its containing space $\mathbb{R}^n$. Their complexity depends on the "intrinsic" dimension k rather than explicitly given dimension n.

In contrast, Indyk and Naor do not provide any new data structure or algorithm for the nearest neighbor problem. What they do is they construct a function f that embeds set $X \subset \mathbb{R}^n$ into space $\mathbb{R}^k$, where k corresponds to the "intrinsic" dimension of X. This embedding somewhat preserves the property of being the closest neighbor. Consequently, in order to find the nearest neighbor of x, we may find the nearest neighbor $f(x')$ of $f(x)$ in $f(X)$. The property of embedding guarantees that x' is a $(1+\epsilon)$-nearest neighbor of x.

2 Nearest Neighbor Preserving Embedding

Definition 1. *Let $\mathbb{R}^n$ and $\mathbb{R}^k$ be spaces with Euclidean metric. Let $X \subset \mathbb{R}^n$. We say that a distribution over mappings $f : R^n \to R^k$ is a nearest neighbor preserving embedding (or NN-preserving), if for any $x_q \in R^n$, with probability at least $1/2$, if $f(x')$ is the nearest neighbor of $f(x)$ in $\mathbb{R}^k$, then x' is $(1+\epsilon)$-nearest neighbor of x in $\mathbb{R}^n$.*

Obviously, such an embedding suffices for the purpose of searching for $(1+\epsilon)$-nearest neighbor.

Recall that the general notion of embedding requires that all pairwise distances between points in X are somewhat preserved. One may see that, therefore, the introduced notion of NN-preserving embedding is significantly weaker than the general one. The main question here is: Can we construct weak embeddings that beat general embeddings in some aspect?

In particular, it would have been nice to obtain an embedding that maps $X \subset R^n$ into R^k where k if roughly the "intrinsic" dimension of n.

Definition 2 (Doubling Dimension)**.** *Let (X, d_x) be a metric space. In what follows, $B_X(x; r)$ denotes the ball in X of radius r centered at $x \in X$, i.e. $B_X(x; r) = \{y \in X : d_x(x, y) < r\}$. The doubling constant of*

X, denoted λ_X, is the least integer $\lambda \geq 1$ such that for every $x \in X$ and $r > 0$ there is $S \subseteq X$ with $|S| \leq \lambda$ such that

$$B_X(x; 2r) \subseteq \cup_{s \in S} B_X(s; r). \tag{1}$$

The doubling dimension is defined as $\log \lambda_X$.

Example 1. *Imagine a line in R^n. Apparently, its doubling constant is 2. Now, curve it a little bit. Does its doubling constant grow? What if we curve it in different places and in different directions? What if these places are far from each other?*

The weak definition of embedding allows Indyk and Naor to come up with a embedding that maps $\mathbb{R}^n$ to $\mathbb{R}^k$ where $k = O(\frac{\log(1/\epsilon)}{\epsilon^2} \cdot \log \lambda_X)$.

3 Main Result

Not surprisingly, Indyk and Naor use the embedding f from $\mathbb{R}^n$ to $\mathbb{R}^k$ that has been employed for so many times already:

$$f(x) = Gx \tag{2}$$

$$G = \frac{1}{\sqrt{k}}(g_{i,j}) \quad \text{where} \quad 1 \leq i \leq k \text{ and } 1 \leq j \leq n, \tag{3}$$

where $g_{i,j}$ are independent standard gaussian variables.

Actually, a little bit weaker result is shown:

Theorem 3. *For $X \in \mathbb{R}^n$ and $\epsilon \in (0,1)$ there exists $k = O(\frac{\log(1/\epsilon)}{\epsilon^2} \cdot \log \lambda_X)$ such that for every $x \in X$ with probability at least $1/2$ (over the distribution of mappings),*

1. *$d(Gx_0, G(X \setminus \{x_0\})) \leq (1+\epsilon) d(x_0, X \setminus \{x_0\})$*

2. *Every $x \in X$ with $\|x_0 - x\| > (1+2\epsilon) d(x_0, X \setminus \{x_0\})$ satisfies $\|Gx_0 - Gx\| > (1+\epsilon) d(x_0, X \setminus \{x_0\})$.*

This result guarantees that if $f(x)$ is the nearest neighbor of $f(x_0)$, then x is $(1+\epsilon)$-nearest neighbor of x. A stronger modification of this theorem that allows $f(x)$ to be only $(1+\epsilon/2)$-nearest neighbor of $f(x_0$ can also be proven.

4 "Neighborhoods do not grow much"

In the case of small doubling dimension, we have a nice result. It essentially says that bounded sets do not like to expand under the action of G:

Lemma 4. *Assume a set $X \subseteq B(0,1)$ in $\mathbb{R}^n$ having doubling dimension λ_X. If $k \geq C \log \lambda_X + 1$ and $D > 6$, then*

$$\mathbf{Pr}\{\exists x \in X : \|Gx\| \geq D\} \leq e^{-ckD^2}. \tag{4}$$

First, let use recall two very useful concentration results for the distribution of $\|Ga\|$, where $\|a\| = 1$. These results roughly say that vectors do not like to change their lengths under the action of G:

$$\mathbf{Pr}\{|\,\|Ga\| - 1| \geq D\} \leq e^{-kD^2/8} \tag{5}$$

$$\mathbf{Pr}\{|\,\|Ga\| \leq 1/D\} \leq (3/D)^k. \tag{6}$$

One can scale these bounds for the case of arbitrary $\|a\|$.

Before giving the actual proof of the lemma, let us try a simple approach. Actually, one can try to derive the lemma applying union bound to (5):

$$\mathbf{Pr}\{\exists x \in X : \|Gx\| \geq D\} \leq |X| \cdot e^{-ckD^2} \overset{(?)}{\leq} e^{-c'kD^2}. \tag{7}$$

Indeed, the last inequality holds when $k \geq \log |X|$. To obtain the result for smaller k we should take into account that, for points x and y being close to each other, if $\|Gx\|$ is not very big, then $\|Gy\|$ is likely not to be very big too.

The proof of this lemma follows below. Interestingly, the technique from this proof will be extended later to prove the main result.

PROOF: The main idea is as follows. If $\|Gx'\|$ is not big and $\|x - x'\|$ is small, then $\|Gx\|$ is not big.

Net construction. First, we construct a number of nets $\{N_i\}_{i\geq 1}$ over X so that each N_i is a net of radius 2^{-i}. Indeed, let $N_0 = \{B(0,1)\}$. Then consider a cover of $X \cap B(0,1)$ that consists of λ_X balls of radius $1/2$. These balls constitute our next net N_1. Note that the existence of the above cover is guaranteed by the definition of doubling dimension. Analogously, we obtain all the other nets N_i of size no more $(\lambda_X)^i$.

One may perceive that every ball from N_i is refined into λ_X balls from N_{i+1}.

Chaining. For every $x \in X$, there exists a chain $x_0, x_1, x_2, \ldots$ with $x_0 = 0$ and $x_i \in N_i$, such that

$$\|x_{i+1} - x_i\| \leq 2^{-i}.$$

This chain can be easily obtained from our nets.

Good pairs of vectors. We say that a pair (x_{i-1}, x_i), where $x_{i-1} \in N_{i-1}$ and $x_i \in N_i$ is good if

$$\|G(x_{i-1} - x_i)\| \leq 2^{-(i-1)} \cdot \delta_i \quad , \quad \text{where} \quad \delta_i = \frac{D}{6} \cdot (3/2)^i. \tag{8}$$

Our choice of δ_i will become clear soon. Since $\|x_i - x_{i-1}\| \leq 2^{-(i-1)}$, we have

$$\mathbf{Pr}\{(x_{i-1}, x_i) \text{ is bad}\} \leq e^{-k \cdot D^2 (3/4)^{2i}}. \tag{9}$$

Further,

$$\begin{aligned}
\mathbf{Pr}\{\exists i : \text{ some pair } (x_{i-1}, x_i) \text{ is bad}\} &\leq \sum_{i\geq 1} \mathbf{Pr}\{\text{some } (x_{i-1}, x_i) \text{ is bad}\} && (10)\\
&\leq \sum_{i\geq 1} (\lambda_X)^{2i-1} \cdot e^{-kD^2(3/4)^{2i}} && (11)\\
&\leq \sum_{i\geq 1} e^{(2i-1)\log \lambda_X - kD^2(3/4)^{2i}} && (12)\\
&\leq \sum_{i\geq 1} e^{-kD^2(3/4)^{2i}} && (13)\\
&\leq e^{-kD^2}. && (14)
\end{aligned}$$

Big union bound. Recall that our aim is to show that, with high probability, $\|Gx\|$ is not big for all x simultaneously. By chaining,

$$\|Gx\| \le \left\| \sum_{i\ge 1} (Gx_i - Gx_{i-1}) \right\| \le \sum_{i\ge 1} \|G(x_i - x_{i-1})\| \tag{15}$$

In particular, if all pairs are good, then

$$\|Gx\| \le \sum_{i\ge 1} 2^{-(i-1)} \cdot \delta_i = \sum_{i\ge 1} \frac{D}{3}(3/4)^i \le D. \tag{16}$$

□

5 Main Theorem

(For simplicity, in the proof below we omit constants in all exponents, therefore e^{-100k} becomes simply e^{-k} and so on.)

PROOF: Without loss of generality, we may assume for this proof that $x_q = 0$ and $d(x_q, X \setminus \{x_q\}) = 1$. Then, in particular, for some $x_a \in X$, we have $\|x_a - x_q\| = 1$, therefore

$$\mathbf{Pr}\{\|Gx_a - Gx_q\| \ge 1 + \epsilon\} \le e^{-k\epsilon^2}. \tag{17}$$

The first part of the theorem follows as $k > 1/\epsilon^2$.

To certain extent, the second part of the theorem resembles the proof of the lemma above. In that lemma we proved that X contained in a ball doesn't expand too much. Now, we have to prove that X contained in the complement of a ball doesn't contract too much.

Pivoting. Split the complement of a ball $B(0; 1 + 2\epsilon)$ into rings

$$B(0; r_i) \setminus B(0; r_{i-1}) \quad , \quad \text{where} \quad r_i = 1 + 2\epsilon + i \cdot \frac{\epsilon}{4}. \tag{18}$$

Then let $X_i = X \cap B(0; r_i) \setminus B(0; r_{i-1})$. Note that with growth of i, vectors from X_i are less likely to contract downto $1 + \epsilon$. But, on the other hand, with growth of i the number of vectors in X_i may also significanlty increase.

By the definition of doubling dimension, for every X_i, there exists an $\epsilon/4$-net S_i such that

$$|S_i| \le \lambda_X^{\log(r_i/\epsilon)} \approx e^{\log \lambda_X \cdot (1/\epsilon + i)}. \tag{19}$$

Let centers of the net be our pivots in X_i. Certainly, the number of pivots in X_i increases when i grows.

Good vectors. For every vector $x \in X_i$ consider a closest pivot $p(x)$. We call vector x good if

$$\|Gx - Gp(x)\| \le \epsilon\sqrt{i}/4. \tag{20}$$

Since it holds that $\|x - x'\| < \epsilon/4$, the above lemma implies that

$$\mathbf{Pr}\{\text{some } x \in X_i \text{ is bad}\} \le |S_i| \cdot e^{-k(\sqrt{i})^2} \le e^{\log \lambda_X \cdot (1/\epsilon + i)} \cdot e^{-ki} \le e^{-ki}. \tag{21}$$

The last inequality holds since $k > \frac{\log 1/\epsilon}{\epsilon^2} \log \lambda_x$.

Good pivots. We call a pivot y good if

$$\|Gy\| \geq 1 + \epsilon + \frac{\epsilon\sqrt{i}}{4}.$$

Recall that y is a center of a ball from S_i, therefore $\|y\| \geq r_{i-1} - \epsilon/4$. Hence

$$\frac{\|Gy\|}{\|y\|} \leq \frac{1 + \epsilon + \epsilon\sqrt{i}/4}{1 + 2\epsilon + \epsilon(i-2)/4}. \tag{22}$$

Depending on value of i, this expression is governed either by i itself or by ϵ:

$$\frac{\|Gy\|}{\|y\|} \leq \begin{cases} \frac{\epsilon\sqrt{i}/4}{\epsilon i/4} \leq 1/\sqrt{i} & , i \geq 1/\epsilon^2 \\ \frac{1+\epsilon}{1+2\epsilon} \leq 1 - \epsilon & , i < 1/\epsilon^2. \end{cases} \tag{23}$$

Applying the concentration results,

$$\mathbf{Pr}\{\text{pivot } y \text{ is bad}\} \leq \begin{cases} e^{-k \log i} & , i \geq 1/\epsilon^2 \\ e^{-k\epsilon^2} & , i < 1/\epsilon^2. \end{cases} \tag{24}$$

The probability that some pivot in S_i is bad is approximately the same for $|S_i|$ is just $e^{\log \lambda_X \cdot (1/\epsilon + i)}$.

Union bound. Assume y is the pivot for x. Then it holds that $\|Gx\| \geq \|Gx - Gy\| - \|Gy\|$. For x and y being a good vector and a good pivot correspondingly, we have $\|Gx\| \leq 1 + 2\epsilon$. This allows us to derive the following bound:

$$\mathbf{Pr}\{\exists x \in X : \|Gx\| < 1 + \epsilon\} \leq \sum_{i \geq 1} \mathbf{Pr}\{\exists x \in X_i : \|Gx\| < 1 + \epsilon\} \tag{25}$$

$$\leq \sum_{i \geq 1} \mathbf{Pr}\{\text{some } x \in X_i \text{ is bad}\} + \sum_{i \geq 1} \mathbf{Pr}\{\text{some pivot } y \in S_i \text{ is bad}\} \tag{26}$$

$$\leq \sum_{i \geq 1} e^{-ki} + \sum_{i \leq 1/\epsilon^2} e^{-k\epsilon^2} + \sum_{i \geq 1/\epsilon^2} e^{-k \log i} \tag{27}$$

$$\leq e^{-k} + 1/\epsilon^2 e^{-k\epsilon^2} + e^{-k \log 1/\epsilon}. \tag{28}$$

This gonna be much smaller than 1 provided $k \geq \frac{\log 1/\epsilon}{\epsilon^2} \log \lambda_x$. □

6 Open Problems

Indyk and Naor give a NN-preserving embedding for l_2-metric. But there are other metric spaces that which support fast approximate nearest neighbor search (e.g. low dimensional l_∞ [Ind98]). Are there NN-preserving embeddings for those metrics?

Indyk and Naor present a NN-preserving embedding from $\mathbb{R}^n$ to $\mathbb{R}^{O(\log \lambda_x)}$. Is there an embedding $\mathbb{R}^n$ to $\mathbb{R}^{O(\log \lambda_x)}$ that approximately preserves all distances between vectors? (For more discussion, see the original paper [IN07])

References

[AM02] S. Arya and T. Malamatos. Linear-size approximate voronoi diagrams. In *SODA*, 2002.

[HP01] S. Har-Peled. A replacement for voronoi diagrams of near linear size. In *FOCS*, 2001.

[IN07] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. In *ACM Transactions on Algorithms (to appear)*, 2007.

[Ind98] P. Indyk. On approximate nearest neighbors in non-euclidean spaces. In *FOCS*, 1998.