Embeddings into l_∞

Piotr Indyk

Stanford University

Universality of l_∞

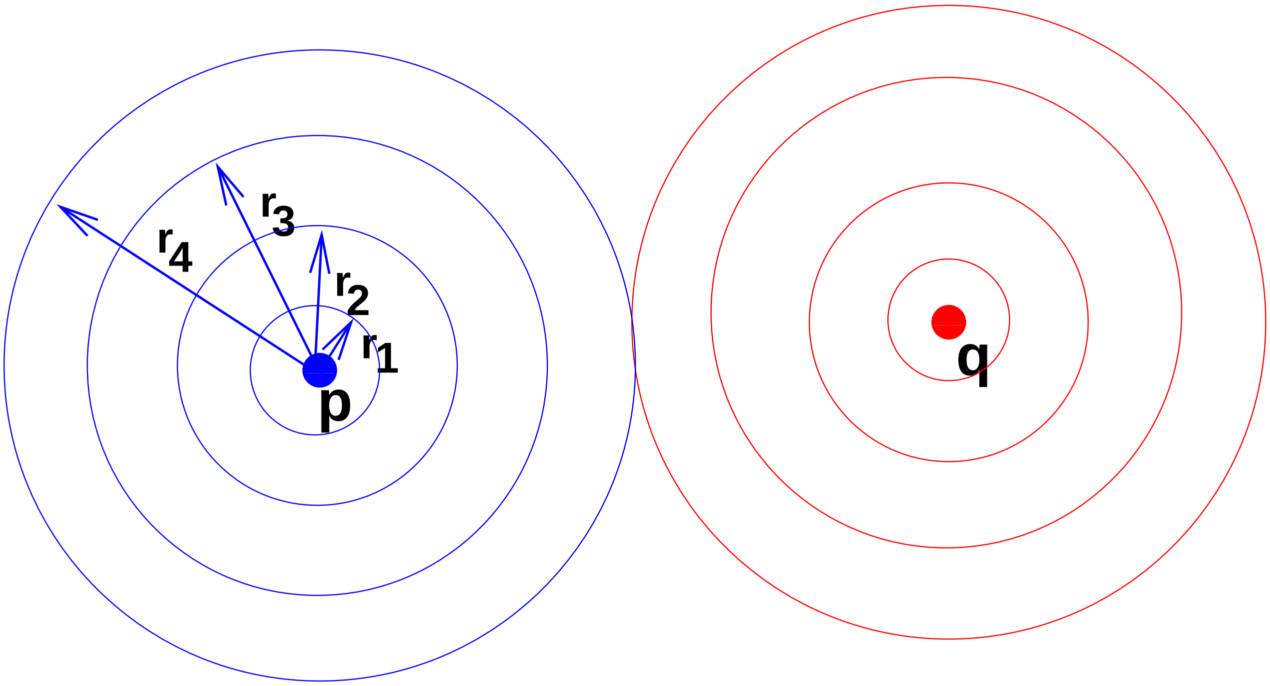
Theorem[Matousek]: For any metric (X, d) and $t \geq 1$ with $|X| = n$ there is a contraction $f : X \rightarrow l_\infty^k$ where

$$k = O(tn^{\frac{1}{\lfloor (t+1)/2 \rfloor}} \log n)$$

such that for any $p, q \in X$

$$|f(p) - f(q)|_\infty \geq d(p, q)/t$$

Main idea



Claim: Let $r_i = i \cdot d(p,q)/t$. Then there exists r_i such that

$$|B(p, r_{i+1})| \leq k|B(q, r_i)|$$

(or vice versa).

Sampling

Algorithm:

- guess $B = |B(p, r_{i+1})|$
- sample $l = k \log n$ sets $A_1, \ldots A_l$ including each metric point with probability $1/2B$
- map p to $(d(p, A_1), \ldots, d(p, A_l))$, where $d(p, A) = \min_{a \in A} d(p, a)$

Claim: For each A_i

$$\Pr[B(p, r_i) \cap A_i \neq \emptyset \wedge B(p, r_{i+1}) \cap A_i = \emptyset] \geq \Omega(1/k)$$

Then

$$|f(p) - f(q)|_\infty \geq |d(p, A_i) - d(q, A_i)| \geq r_{i+1} - r_i = d(p, q)/t$$

It is a contraction

Claim: For any j

$$|d(p, A_j) - d(q, A_j)| \leq d(p, q)$$

Proof: Follows from $d(p, A_j) \leq d(p, q) + d(q, A_j)$.

Corollary: $d(p, q)/t \leq |f(p) - f(q)|_\infty \leq d(p, q)$.

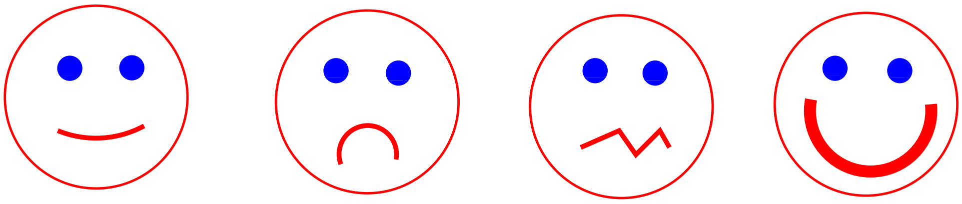
Algorithmic implications

- ultimate nearest neighbor algorithm ? not really ...
- can be used for product metrics

$$D((p_1 \ldots p_k), (q_1 \ldots q_k)) = \max[D_1(p_1, q_1), \ldots D_k(p_k, q_k)]$$

- approximate shortest path queries

Shape metrics



Hausdorff metric

For sets A,B:

$$\vec{d}_H(A, B) = \max_{a \in A} \min_{b \in B} d(a, b)$$

$$d_H(A, B) = \max(\vec{d}_H(A, B), \vec{d}_H(B, A))$$

Embeddings[FI'99]

Define $H_s(X)$ to be a Hausdorff metric over $\leq s$-subsets of X.

Theorem: There exists an isometry $f : H_s(X) \to l_\infty^{|X|}$, for any s.

Theorem: For any $R > 0$ there exists a distribution over contractions $F : H_s(l_p^d) \to l_\infty^D$ with

$$D = 1/\epsilon^{O(d)} \cdot s \cdot \log 1/\delta$$

such that for any sets A, B s.t. $d_H(A, B) \geq R$

$$\Pr[d(F(A), F(B)) \geq (1-\epsilon)R] \geq 1-\delta$$

Isometry

Let $X = \{p_1 \dots p_n\}$.
We map

$$A \to (d(p_1, A), d(p_2, A) \dots d(p_n, A))$$

Lower bound: let $d_H(A, B) = d(a, B)$. Then

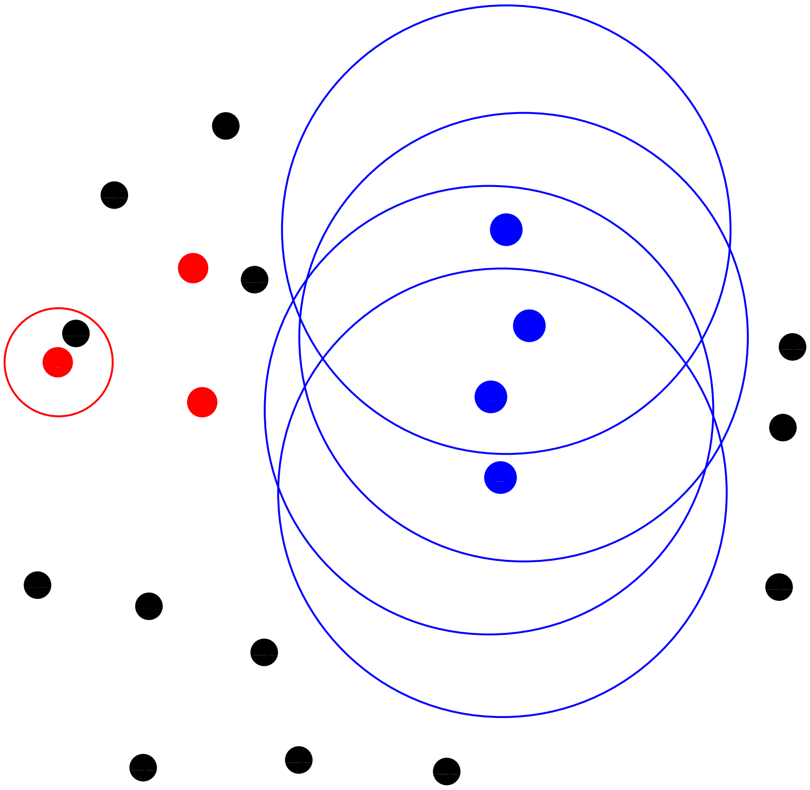
$$|f(A) - f(B)|_\infty \geq |d(a, A) - d(a, B)| = d(a, B) = d_H(A, B)$$

Upper bound (i.e. $|d(p, A) - d(p, B)| \leq d_H(A, B)$): follows from

$$d(p, A) \leq d(p, B) + d_H(A, B)$$

Approximate embedding

$$A \to (d(A, S_1) \ldots d(A, S_D))$$



Want S_i to hit red circle and avoid blue circles (as before).

Complexity

Blue volume: $s \cdot CR_2^d$

Red volume: CR_1^d.

$\rightarrow$ Each S_i is good with probability $\frac{1}{s(R_2/R_1)^d}$.