max planck institut
informatik

Lecture 6:Testing Sparse images

Themis Gouleakis

May 18, 2021

Models of testing graph properties

Dense graph model

- **Adjacency predicate**: $g : \binom{V}{2} \to \{0,1\}$
  - $g(\{u,v\}) = [\{u,v\} \in E]$ (truth value)
- **Alternatively**: $g : V \times V \to \{0,1\}$
  - $g(u,v) = [\{u,v\} \in E]$ (truth value)
- **Distance**: $\delta(G,G') = \delta(g,g') = \frac{|\{u,v\} : g(u,v) \neq g'(u,v)|}{|V|^2}$
  - $\delta_\Pi(G) = \min_{G' \in \Pi, \pi} \left\{ \frac{|\{u,v\} : g(u,v) \neq g'(\pi(u),\pi(v))|}{|V|^2} \right\}$

Models of testing graph properties

Bounded degree graph model

- **Incidence function**: $g : g : V \times [d] \to V \cup \{\perp\}$
  - $g(\{u, i\}) = v$, where v is the i-th neighbor of u.
  - $g(\{u, i\}) = \perp$ if $d(u) < i$.
- **Distance**: G is ϵ- far from a graph property Π if for any permutation π over V, the following holds:

$$\sum_{u \in V} |\{v : \exists i \ g(u,i) = v\} \triangle \{v : \exists i \ g'(\pi(u), i) = \pi(v)\}| \ > \ \epsilon d N \,,$$

where $N = |V|$.

Models of testing graph properties

General graph model

- Both representations and types of queries used.
- **Distance**: $\delta(G, G') = \frac{|E \Delta E'|}{\max\{|E|,|E'|\}}$
- More general that the other models.
- Not as easy to use.

Which model to use?

- Depends on the graphs contained in Π and the type of queries used.

Definitions

Images:

- An image will be represented by a $0/1$-valued $n \times n$ matrix M.
  - Dense if it contains $\Omega(n^2)$ 1-entries/pixels.

Access models:

- Dense image model: (analog to dense graph model)
  - Query access to entries
- Sparse image model: (analog to sparse graph model)
  - Query access to entries
  - Sample access to 1-entries

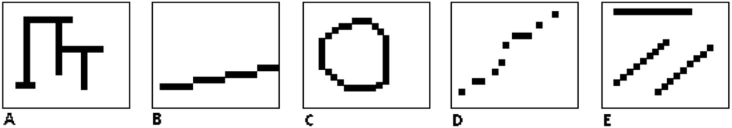
Distance:

- Dense image model: $\delta(M, M') = \frac{d_H(M,M')}{n^2}$
- Sparse image model: $\delta(M, M') = \frac{d_H(M,M')}{w(M)}$

where $w(M)$ is the number of 1-pixels in M

Example properties

- Connectivity: Graph of M is connected
- Line imprint: $\exists$ a line segment such that $M(i,j) = 1$ iff the line intersects the pixel.
- Convexity: Similar for a convex shape
- Monotonicity: $\forall$ (i_1, j_1) and (i_2, j_2) 1-pixels it holds: $i_1 < i_2 \Rightarrow j_1 \leq j_2$.



VC dimension and ϵ-nets

Definition (ϵ-nets)

Let X be a set, μ a probability measure on X, $\mathcal{F}$ be a system of μ-measurable subsets of X and $\epsilon > 0$. A subset $N \subseteq X$ is called an ϵ-net for $(X, \mathcal{F})$ with respect to μ if $N \cap S \neq \emptyset$ for all $S \in \mathcal{F}$ with $\mu(S) \geq \epsilon$.

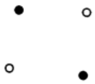
In our case (ϵ-net on the differences of images):

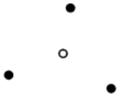
- $X = \mathbb{R}^d$
- $\{M_1, \ldots, M_k\}$ is a set of images.
  - $\mathcal{F} = \{M_i \Delta M_j\}_{i \neq j \leq k}$
- If $\Pr_{(i,j)\sim\mu}[M_1(i,j) \neq M_2(i,j)] > \epsilon$, then the ϵ-net N contains at least one pixel on which M_1, M_2 differ.

Definition (*VC*-**dimension**)

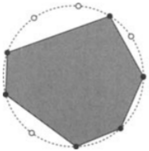
Let $\mathcal{F}$ be a set system on X We say that $A \subseteq X$ is shattered by $\mathcal{F}$ if each of the subsets of A can be obtained as the intersection of some $S \in \mathcal{F}$ with A. We call VC-dimension of a set system $\mathcal{F}$(denoted by $dim(\mathcal{F})$) the supremum of the sizes of all shattered subsets of $\mathcal{F}$.

Example 1: Halfspaces in $\mathbb{R}^2$: VC dimension = 3





Example 2: Convex sets in $\mathbb{R}^2$: VC dimension = ∞



Theorem (ϵ-net theorem)

Let $\mathcal{C}$ be a class of images with VC dimension d. There exists a constant c_1 such that for any distribution D the following holds: If N consists of $c_1 \frac{d \log(1/\epsilon)}{\epsilon}$ pixels drawn according to D, then it is an ϵ-net for $\mathcal{C}$ with high constant probability.

Corollary

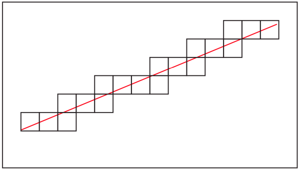
Let $w = w(M)$ denote the number of 1-pixels of an image M. Then, there exists a property tester for membership in $\mathcal{C}$, that given $w(M)$ has sample complexity $O(d \log(1/\epsilon)/\epsilon)$.

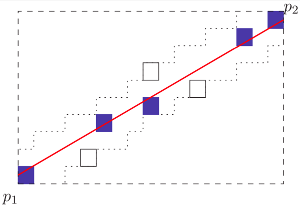
Proof:

Testing Line imprints

Definition

The imprint of a line segment in $\mathbb{R}^2$ is the set of pixels it intersects.





Definition

The sleeve defined by 2 pixels p_1, p_2 is the union of all line imprints of line segments that start in p_1 and end in p_2.

Property tester

1. Take a sample S_1 of size $m_1 = \Theta(1/\epsilon)$ uniformly from the 1-pixels of M and find 2 pixels p_1, p_2 with maximum distance between them.
2. Take another sample S_2 of size $m_2 = \Theta(1/\epsilon)$ 1-pixels of M
   - we distinguish this from S_1 for analysis purposes.
3. Let P be the sleeve of p_1, p_2. Query M on a set Q of size $m_3 = \Theta(\frac{\log(1/\epsilon)}{\epsilon})$ pixels of P. If there exists a line imprint that is consistent with the 1-pixels is $S_1 \cup S_2$ and the queries in Q, then ACCEPT. Otherwise REJECT.

Theorem

The above property tester is an one-sided error tester for the property of being a line imprint. with sample and query complexity of $O(\frac{\log(1/\epsilon)}{\epsilon})$ *and running time of* $O(1/\epsilon)$.

Proof: