Wasserstein Propagation for Semi-Supervised Learning

Justin Solomon, Raif M. Rustamov, Leonidas Guibas, Adrian Butscher. ICML-2014

Zoltán Szabó

Gatsby Unit, Tea Talk

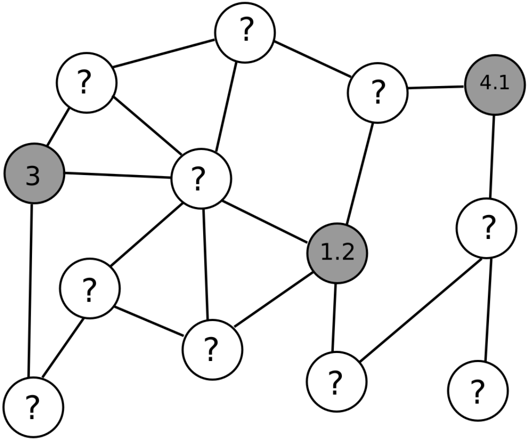
March 21, 2014

Outline

- Motivation.
- Problem formulation.
- Numerical illustration.

Goal: label propagation

Task: $\{(x_i, y_i)\}_{i=1}^{l}, \{x_i\}_{i=l+1}^{l+u} \Rightarrow \{y_i\}_{i=l+1}^{l+u}$.

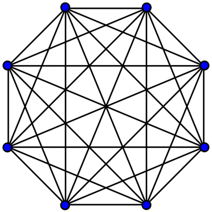


Graph-based semi-supervised learning

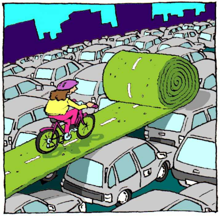
- Difficulty: obtaining labels can be
  - time consuming,
  - expensive.
- Protein shape classification: 1 month/label for an expert.

Exploit the underlying graph structure

- Idea:
  - combine unlabelled + labelled data.
  - weights: similarities of instances.
  - labelling, which is smooth w.r.t. the similarity graph.



Labels: real number $\rightarrow$ distribution



- Example:
  - node = location,
  - label = traffic density over a day.
- Problem: limited number of sensors.
- Goal: propagate densities to the entire map.

Domain of distributions

- Real line, integers ($\mathbb{Z}$ or $\{1, \ldots, m\}$).
- $\mathbb{S}^1$: unit circle
  - Example: wind direction prediction

- Metric space: $(\mathbb{D}, r)$.

Formulation: label = real number

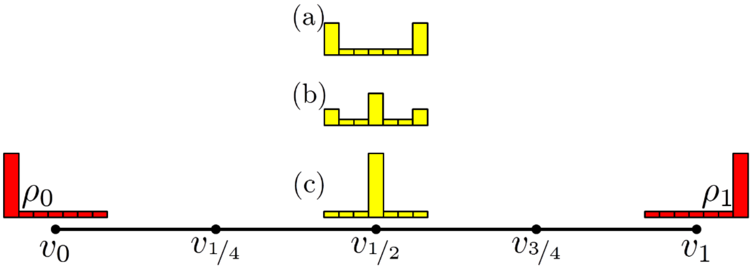
- Given: graph $G = (V, E)$ with edge weights w_e.
- Label function: $f : V \to \mathbb{R}$, known on $V_0 \subseteq V$.
- Goal: extend f to $V \setminus V_0$.
- Objective function: Dirichlet energy

$$J(f) := \sum_{(v,w) \in E} w_e (f_v - f_w)^2 \to \min_f, \quad (1)$$

with prescribed values on V_0. $\Leftrightarrow \Delta f = 0$ on $V \setminus V_0$.

Formulation: requirements

- Spread/uncertainty preservation.
- Peakness preservation around the propagated means.
- (a): bins-, (b): samples separately, (c): ideal.



Optimal transportation, Wasserstein distance

- In case of $f_v \in \mathbb{R}$ ($v \in V$): $(f_v - f_w)^2$.
- Now: $f_v \in \mathcal{M}_1^+(\mathbb{D})$, distance between distributions
  - $(\mathbb{D}, r) = (\mathbb{R}, |\cdot|)$:

$$W_2^2(a,b) = \inf_{\pi \in P(a,b)} \int_{\mathbb{R}^2} |x-y|^2 \mathrm{d}\pi(x,y). \tag{2}$$

Optimal transportation, Wasserstein distance

- In case of $f_v \in \mathbb{R}$ $(v \in V)$: $(f_v - f_w)^2$.
- Now: $f_v \in \mathcal{M}_1^+(\mathbb{D})$, distance between distributions
  - $(\mathbb{D}, r) = (\mathbb{R}, |\cdot|)$:

$$W_2^2(a,b) = \inf_{\pi \in P(a,b)} \int_{\mathbb{R}^2} |x-y|^2 \mathrm{d}\pi(x,y). \tag{2}$$

  - $(\mathbb{D}, r) = (\{1,\ldots,m\}, r)$:

$$W_2^2(a,b) = \inf_{\pi \in P(a,b)} \int_{\mathbb{R}^2} r^2(x,y) \mathrm{d}\pi(x,y) \tag{3}$$

$$= \inf_{\pi: \sum_j \pi_{ij} = a_i, \sum_i \pi_{ij} = b_j, \pi_{ij} \geq 0 (\forall i,j)} \sum_{i,j=1}^{m} r_{ij}^2 \pi_{ij}. \tag{4}$$

Formulation: label = distribution

- Given: graph $G = (V, E)$.
- Label function (distribution-valued map):
  - $f : V \to \mathcal{M}_1^+(\mathbb{D})$, known on $V_0 \subseteq V$.
- Goal: extend f to $V \setminus V_0$.
- Wasserstein propagation:

$$J(f) := \sum_{(v,w)\in E} W_2^2(f_v, f_w) \to \min_f, \tag{5}$$

with prescribed distributions on V_0.

Solution: $\mathbb{D} = \mathbb{R}$

- Result:
  - F_v: cdf of f_v ($v \in V_0$).
  - For each $s \in [0,1]$, let $g_s : V \to \mathbb{R}$ be the solution of

$$\Delta g_s = 0 \quad (\forall v \in V \setminus V_0), \tag{6}$$

$$g_s(v) = F_v^{-1}(s) \quad (\forall v \in V_0). \tag{7}$$

  - For each v: $s \mapsto g_s(v)$ = icdf of a distribution f_v; $v \mapsto f_v$ is the Wasserstein propagation.
- Trick:

$$W_2^2(f_c, f_d) = \left\| F_c^{-1} - F_d^{-1} \right\|_{L^2[0,1]}^2 = \int_0^1 \left[F_c^{-1}(s) - F_d^{-1}(s) \right]^2 \mathrm{d}s.$$

Solution: $(\mathbb{D}, r) = (\{1, \ldots, m\}, r) \Rightarrow$ LP

$$W_2^2(a,b) = \inf_{\pi: \sum_j \pi_{ij} = a_i, \sum_i \pi_{ij} = b_j, \pi_{ij} \geq 0 (\forall i,j)} \sum_{i,j=1}^{m} r_{ij}^2 \pi_{ij}. \tag{8}$$

$$J(f) = \sum_{(v,w)\in E} W_2^2(f_v, f_w) \Rightarrow J(f,\pi) = \sum_{e\in E} \sum_{i,j=1}^{m} r_{ij}^2 \pi_{ij}^{(e)}, \tag{9}$$

$$\text{s.t.} \quad \sum_j \pi_{ij}^{(e)} = f_{vi} \quad \forall e = (v,w) \in E, i \in \mathbb{D}, \tag{10}$$

$$\sum_i \pi_{ij}^{(e)} = f_{wj} \quad \forall e = (v,w) \in E, j \in \mathbb{D}, \tag{11}$$

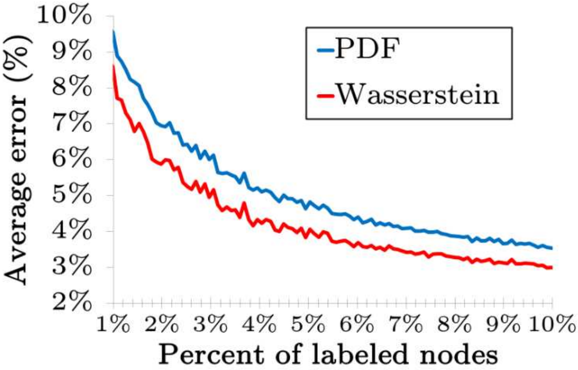
$$\sum_i f_{vi} = 1 \quad (\forall v \in V), \quad f_{vi} \text{ fixed } (\forall v \in V_0), \tag{12}$$

$$f_{vi} \geq 0 \quad (\forall v \in V, i \in \mathbb{D}), \quad \pi_{ij}^{(e)} \geq 0 (\forall i,j \in \mathbb{D}, e \in E). \tag{13}$$

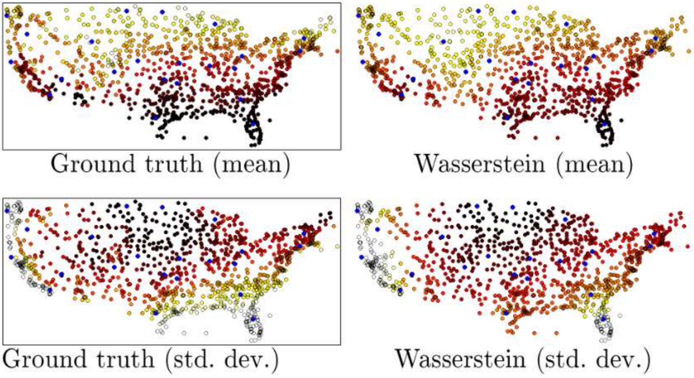
Demo: temperature distribution propagation

- $|V| = 1113$ weather stations.
- f_v: daily temperature histograms ($m = 100$ bins).
- Baseline: bin-by-bin propagation.
- Performance measure: one-Wasserstein error averaged over $V \setminus V_0$.

Temperature distribution propagation



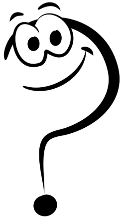
Temperature distribution propagation



Summary

- Semi-supervised learning with distribution labels.
- Smoothness: Wasserstein distance.
- Domain:
  - $\mathbb{R}$: boils down to the classical Dirichlet problem (F^{-1}).
  - $\{1, \ldots, m\}$: LP.
- Application: temperature/wind prediction.

Thank you for the attention!



Continuous Dirichlet problem: given V domain

- Dirichlet energy minimization:

$$J(f) = \frac{1}{2}\int_V \|\nabla f(\mathbf{u})\|^2 \to \min_f, \tag{14}$$

with boundary conditions on f.

- Laplacian equation (st. the boundary conditions):

$$0 = (\Delta f)(\mathbf{u}) = \nabla^2 f(\mathbf{u}) = \sum_i \frac{\partial^2 f}{\partial^2 u_i}(\mathbf{u}), \quad (\mathbf{u} \in V). \tag{15}$$

- Equivalent problems; solutions =: harmonic functions.

Discrete Dirichlet problem

- $\mathbf{W} = [w_{ij}] \in \mathbb{R}^{|V|\times|V|}$, $\mathbf{D} = diag(d_i)$, $d_i = \sum_j w_{ij}$.
- $\mathbf{L} = \mathbf{D} - \mathbf{W} \in \mathbb{R}^{|V|\times|V|}$: combinatorial Laplacian.
- $\mathbf{f} = [\mathbf{f}_L; \mathbf{f}_U] \in \mathbb{R}^{|V_0|+|V\setminus V_0|}$.
- Objective:

$$J(\mathbf{f}) = \frac{1}{2}\mathbf{f}^T\mathbf{L}\mathbf{f} \to \min_{\mathbf{f}_U}. \tag{16}$$

- $\mathbf{L} = \begin{bmatrix} \mathbf{L}_{L,L} & \mathbf{L}_{L,U} \\ \mathbf{L}_{U,L} & \mathbf{L}_{U,U} \end{bmatrix}$. Solution ($\mathbf{f}_U$):

$$-\mathbf{L}_{U,U}\mathbf{f}_U = \mathbf{L}_{U,L}\mathbf{f}_L. \tag{17}$$