

Robust Ordinal Embedding from Contaminated Relative Comparisons

Ke Ma^{1,2}, Qianqian Xu³, Xiaochun Cao¹

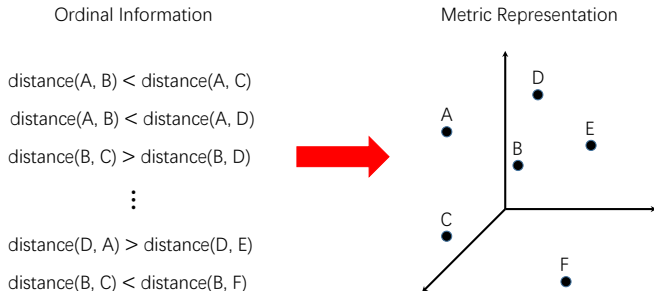
¹ SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences

² School of Cyber Security, University of Chinese Academy of Sciences

³ VIPL, Institute of Computing Technology, Chinese Academy of Sciences

November 14, 2018

Ordinal Embedding

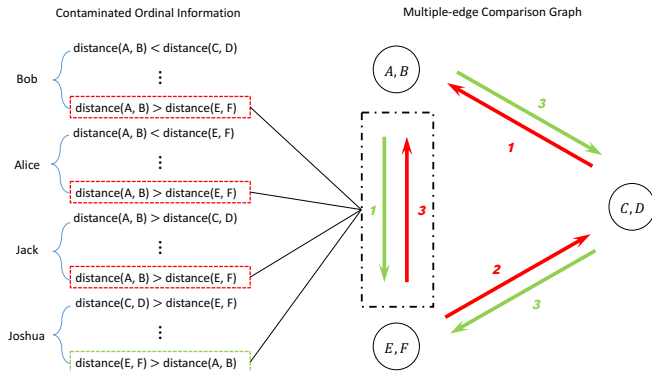


Suppose the relative comparisons, *e.g.* (i, j, l, k) , are consistent with a low-dimensional embedding, our goal is to associate each item with a point $\mathbf{x} \in \mathbb{R}^p$ such that

$$\|\mathbf{x}_i - \mathbf{x}_j\|_2 < \|\mathbf{x}_l - \mathbf{x}_k\|_2.$$

Wait a minute

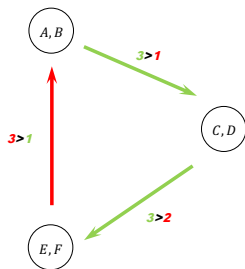
Actually, we only collect the contaminated relative comparisons from different annotators



Multi-stage Method

- ▶ The multiple-edges are aggregated, e.g. majority voting.

Condorcet's paradox !

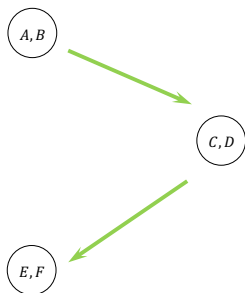


Cons:

- ▶ pruning the right comparison direction
- ▶ leading to the well-known 'Condorcet's paradox'

Multi-stage Method

- ▶ Maximum acyclic subgraph approximation.



Cons:

- ▶ NP-complete problem
- ▶ the available data would be rare.

Preliminaries

- ▶ object set $\mathcal{O} = \{\mathbf{o}_1, \dots, \mathbf{o}_n\}$
- ▶ similarity function $\zeta : \mathcal{O}^2 \rightarrow \mathbb{R}^+$
- ▶ relative comparison set

$$\mathcal{C}_{\mathcal{U}} = \left\{ (i, j, l, k)_u \mid \begin{array}{l} i, j, l, k \in [n], u \in \mathcal{U} \\ i \neq j, l \neq k, (i, j) \neq (l, k) \end{array} \right\}$$

and the corresponded comparison graph $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ where $\mathcal{V} = \{v_{ij} \mid i, j \in [n]\}$ and $\mathcal{E} = \{\mathbf{e}_c^{\mathcal{U}}, \mathbf{e}_{\bar{c}}^{\mathcal{U}}\}$.

- ▶ multiple edge $\mathbf{e}_c^{\mathcal{U}} = \{e_c^u, c_u \in \mathcal{C}_{\mathcal{U}}\}$, and the indicator y_c^u on e_c^u is $y_c^u = 1$ if e_c^u existed. Furthermore, the weight of $\mathbf{e}_c^{\mathcal{U}}$ is

$$w_c = \sum_{u \in \mathcal{U}} [y_c^u = 1], \quad c_u \in \mathcal{C}_{\mathcal{U}}$$

The Proposed Unified Framework

- ▶ Detecting the outliers in the edge set \mathcal{E} . A set of unknown variables $\gamma = \{\gamma_c\} \in \mathbb{R}^{|\mathcal{E}|}$ indicate whether the edge $\mathbf{e}_c^{\mathcal{U}}$ is an outlier or not. The outlier detection task in $\mathcal{C}_{\mathcal{U}}$ thus becomes the problem of estimating γ with \mathcal{G} .
- ▶ Obtaining an embedding $\mathbf{X} \in \mathbb{R}^{p \times n}$. Given $\mathbf{e}_c^{\mathcal{U}} \in \mathcal{E}$ which represents the correct relative similarity measurement, we hope the embedding \mathbf{X} satisfies

$$\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 < \|\mathbf{x}_l - \mathbf{x}_k\|_2^2, \quad c_u = (i, j, l, k)_u \in \mathcal{C}_{\mathcal{U}}.$$

A Linear Model

- ▶ Given an edge $\mathbf{e}_c^{\mathcal{U}} \in \mathcal{E}$, its corresponding direction indicator y_c is modeled as

$$\begin{aligned} y_c &= \|\mathbf{x}_i - \mathbf{x}_j\|_2^2 - \|\mathbf{x}_l - \mathbf{x}_k\|_2^2 + \gamma_c + \varepsilon_c \\ &= d_{ij}^2 - d_{lk}^2 + \gamma_c + \varepsilon_c. \end{aligned} \quad (1)$$

- ▶ It is known that there is a map between the distance matrix \mathbf{D} and the Gram matrix $\mathbf{G} = \{g_{ij}\} = \mathbf{X}^\top \mathbf{X}$ as

$$d_{ij} = g_{ii} - 2g_{ij} + g_{jj}$$

and (1) can be written as

$$\mathbf{y} = \mathbf{Z} \odot \mathbf{G} + \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \quad (2)$$

where

$$\mathbf{Z} \odot \mathbf{G} = \mathbf{Z}\mathbf{g} = \mathbf{Z} \cdot \text{vec}(\mathbf{G}).$$

Robust Ordinal Embedding

$$\begin{aligned} & \underset{\mathbf{G}, \gamma}{\text{minimize}} \quad \mathcal{L}_{\mathbf{w}}(\mathbf{G}, \gamma) + \lambda \|\gamma\|_{1, \mathbf{w}} \\ & \text{subject to} \quad \text{rank}(\mathbf{G}) = p, \quad \mathbf{G} \succeq 0 \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathcal{L}_{\mathbf{w}}(\mathbf{G}, \gamma) &= \frac{1}{2} \|\mathbf{y} - \mathbf{Z} \odot \mathbf{G} - \gamma\|_{2, \mathbf{w}}^2 \\ &= \frac{1}{2} \sum_{\mathbf{e}_c^{\mathcal{U}} \in \mathcal{E}} w_c^2 (y_c - \gamma_c - d_{ij} + d_{lk})^2 \\ &= \frac{1}{2} \|\mathbf{W}\mathbf{y} - (\mathbf{W}\mathbf{Z}) \odot \mathbf{G} - \mathbf{W}\gamma\|_2^2 \end{aligned} \tag{4}$$

and

$$\|\gamma\|_{1, \mathbf{w}} = \sum_{\mathbf{e}_c^{\mathcal{U}} \in \mathcal{E}} w_c |\gamma_c| = \|\mathbf{W}\gamma\|_1 \tag{5}$$

Optimization

$$\begin{aligned}\mathcal{L}_{\mathbf{w}}(\mathbf{G}, \gamma) &= \frac{1}{2} \|\mathbf{y}_{\mathbf{w}} - \mathbf{WZ} \cdot \mathbf{g} - \mathbf{W}\gamma\|_2^2 \\ &= \frac{1}{2} \left\| \mathbf{y}_{\mathbf{w}} - \begin{bmatrix} \mathbf{WZ} & \mathbf{W} \end{bmatrix} \begin{pmatrix} \mathbf{g} \\ \gamma \end{pmatrix} \right\|_2^2 \\ &= \frac{1}{2} \|\mathbf{y}_{\mathbf{w}} - \mathbf{A} \cdot \beta\|_2^2 := f(\beta)\end{aligned}\tag{6}$$

$$\lambda \|\gamma\|_{1, \mathbf{w}} = \lambda \left\| \begin{bmatrix} \mathbf{0} & \mathbf{W} \end{bmatrix} \begin{pmatrix} \mathbf{g} \\ \gamma \end{pmatrix} \right\|_1 = \lambda \|\mathbf{B} \cdot \beta\|_1 := g(\beta).\tag{7}$$

With (6), (7) and ignoring the constraints on \mathbf{G} , (3) is equivalent to a *LASSO* formulation

$$\arg \min_{\beta} F(\beta) := f(\beta) + g(\beta)\tag{8}$$

Let \mathcal{F} be the feasible set of (8) and suppose \mathbf{G}^* is a optimal solution of (3), it holds that $\mathbf{G}^* \in \mathcal{F}$. As a consequence, we come to a semi-definite programming with rank equality constraint

$$\begin{aligned} & \text{find } \mathbf{G}, \gamma \\ & \text{subject to } \mathbf{G}, \gamma \in \mathcal{F}, \mathbf{G} \succeq 0, \text{rank}(\mathbf{G}) = p. \end{aligned} \tag{9}$$

- ▶ Solving a SDP with rank equality constraints like (9) is notoriously difficult.
- ▶ Here we adopt the ran-reduction for semi-definite programming.

Solving the SDP

First, we solve the following optimization problem

$$\begin{aligned} & \text{find } \mathbf{G}, \gamma \\ & \text{subject to } \mathbf{G}, \gamma \in \mathcal{F}, \mathbf{G} \succeq 0. \end{aligned} \tag{10}$$

For any $L > 0$, consider the following quadratic approximation of $F(\beta) := f(\beta) + g(\beta)$ at a given point β_0 :

$$Q_L(\beta, \beta_0) = f(\beta_0) + \langle \beta - \beta_0, \nabla f(\beta_0) \rangle + \frac{L}{2} \|\beta - \beta_0\|^2 + g(\beta) \tag{11}$$

which admits a unique minimizer

$$P_L(\beta) = \arg \min_{\beta} \left\{ g(\beta) + \frac{L}{2} \left\| \beta - \left(\beta_0 - \frac{1}{L} \nabla f(\beta_0) \right) \right\|^2 \right\}.$$

$$\mathbf{G}_{t+1} = \Pi_{\mathcal{S}_+^n} \left(\mathbf{G}_t - \frac{1}{L} \nabla_{\mathbf{G}} f(\beta_t) \right) \tag{12a}$$

$$\gamma_{t+1} = \mathcal{T}_{\mu} \left(\gamma_t - \frac{1}{L} \nabla_{\gamma} f(\beta_t) \right) \tag{12b}$$

Rank Reduction

Here we reformulate (10) as a standard SDP

$$\begin{aligned} & \text{minimize} \quad \|\gamma\|_{1,\mathbf{w}} \\ & \text{subject to} \quad \langle \mathbf{G}, \mathbf{A}_c \rangle + \gamma_c = y_c, \quad \mathbf{e}_c^{\mathcal{U}} \in \mathcal{E}, \\ & \quad \quad \quad \mathbf{G} \succeq 0, \end{aligned} \tag{13}$$

Readable Mathematics

Let X_1, X_2, \dots, X_n be a sequence of independent and identically distributed random variables with $E[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$, and let

$$S_n = \frac{X_1 + X_2 + \dots + X_n}{n} = \frac{1}{n} \sum_i^n X_i$$

denote their mean. Then as n approaches infinity, the random variables $\sqrt{n}(S_n - \mu)$ converge in distribution to a normal $\mathcal{N}(0, \sigma^2)$.