Landmark Ordinal Embedding

Nikhil Ghosh† **Yuxin Chen**‡ **Yisong Yue**⋆

†University of California, Berkeley ‡University of Chicago ⋆California Institute of Technology

Motivation

- Want representations respecting object similarity
- Difficult to obtain quantitative measurements
- Instead use triplet preference feedback from humans

Is  closer to  than  ?

Problem Statement

- Given n objects with unknown embeddings $\mathbf{x}_1^*, \dots, \mathbf{x}_n^* \in \mathbb{R}^d$
- $\mathbf{D}^*$ is the Euclidean Distance Matrix (EDM) i.e. $\mathbf{D}_{ij}^* = \left\|\mathbf{x}_i^* - \mathbf{x}_j^*\right\|_2^2$.
- Receive noisy answers to the triplet query "is object j closer to i than k?"
- Given $\langle i, j, k \rangle$ receive "yes" or "no" label where $\mathbb{P}[\text{"yes"}] = f(\mathbf{D}_{ij}^* - \mathbf{D}_{ik}^*)$
- We consider the Bradley-Terry-Luce model: $f(\theta) = \frac{1}{1+\exp(-\theta)}$
- **Goal**: using m queries, estimate $\hat{\mathbf{x}}_1, \dots, \hat{\mathbf{x}}_n$ minimizing Fröbenius norm error

$$\left\|\mathbf{X}^* - \hat{\mathbf{X}}\right\|_F$$

Primary Related Work

Let $\mathcal{T}$ be the set of triplet queries $\langle i, j, k \rangle$ that received label "yes".

- Stochastic Triplet Embedding (STE) [VDMW12]

$$\max_X \sum_{\forall \langle i,j,k \rangle \in \mathcal{T}} f(D_{ij} - D_{ik})$$

- Generalized Non-metric Multidimensional Scaling (GNMDS) [Aga+07]

$$\min_{X, \xi_{ijk} \geqslant 0} \sum_{\forall \langle i,j,k \rangle \in \mathcal{T}} \xi_{ijk} \text{ subject to } D_{ik} - D_{ij} \geqslant 1 - \xi_{ijk}$$

Issues with Existing Approaches

- Existing algorithms require solving expensive optimization problems
- (Projected) Gradient Descent algorithms require operations that are take $\Omega(n^2)$ time
- Becomes prohibitively slow when n is large ($\geqslant 10^4$)
- We would like to compute embeddings for large n

Landmark Multidimensional Scaling [DST04]

Let $\mathcal{L} \subset [n]$ be a set of *landmark* points. Given distances $\mathbf{D}_{ij}^*$, $(i,j) \in \mathcal{L} \times [n]$, Landmark Multidimensional Scaling (LMDS) allows us to recover the full embedding $\mathbf{X}$.

Our Contributions

- A fast embedding algorithm, *Landmark Ordinal Embedding* (LOE), inspired from *Landmark Multidimensional Scaling* (LMDS).
- A thorough analysis in both sample complexity and computational complexity.
- LOE allows us to *warm-start* existing state-of-the-art embedding approaches that are statistically more efficient but computationally more expensive.
- By warm-starting with LOE we can find accurate embeddings on massive datasets much more quickly than existing methods.

Landmark Ordinal Embedding

Algorithm Sketch

Input: # of items n; # of landmarks ℓ; # of samples m; dimension d;

1. Randomly select ℓ landmarks from $[n]$
2. Compute rankings $R_1, \dots, R_\ell$ of landmark cols.
3. Estimate $\ell \times \ell$ landmark submatrix of $\mathbf{D}^*$
4. Estimate landmark column shifts mean$(\mathbf{D}_i^*)$
5. Recover $\mathbf{D}_i^*$, output embedding using LMDS

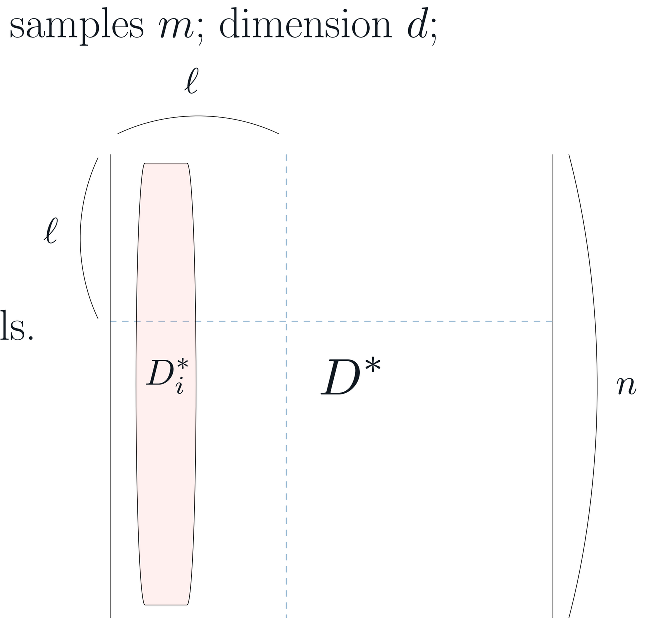


Figure 1: LOE

Key Ideas

1. Queries $\langle i, j, k \rangle$ correspond to pairwise comparisons $\langle j, k \rangle$ based on distance from i
2. View recovering column $\mathbf{D}_i^*$ as a ranking problem
3. Ranking model is not identifiable, but we can recover $R_i = \mathbf{D}_i^* - \text{mean}(\mathbf{D}_i^*) \cdot \mathbf{1}$
4. Submatrix of landmarks is a distance matrix, allows identification of mean$(\mathbf{D}_i^*)$
5. LMDS can compute an embedding using only $\ell = O(d)$ columns of $\mathbf{D}^*$

Theoretical Analysis

Theorem 1 (Consistency) *$\Omega(\text{poly}(d) n \log n)$ triplets suffice for LOE to recover the embedding in Fröbenius norm with high probability.*

- *Proof sketch*: first bound the propagated error of the landmark columns estimate; combined with a perturbation bound for LMDS to obtain the above result.

Theorem 2 (Computational Complexity) *Recovering the embedding up to a fixed error using LOE requires time $O(m + nd^2 + d^3)$ which is linear in n.*

Experimental Results

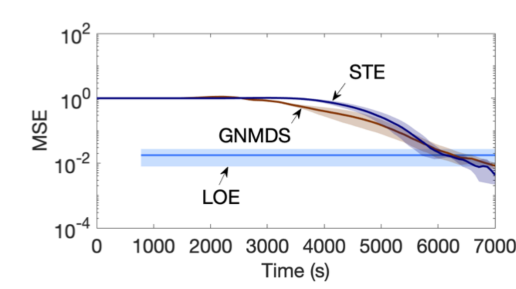
- STE, GNMDS: baselines
- LOE: our algorithm
- LOE-STE: using LOE to warm-start STE

Food Embedding using LOE-STE

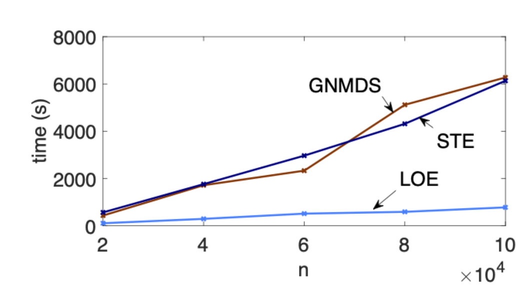
Datasets

- Synthetic ($\mathbf{x}_i^* \sim$ normal dist.)
- MNIST (handwritten digits)
- FOOD (images of food)

We sample a total number of $cn \log n$ triplets per experiment

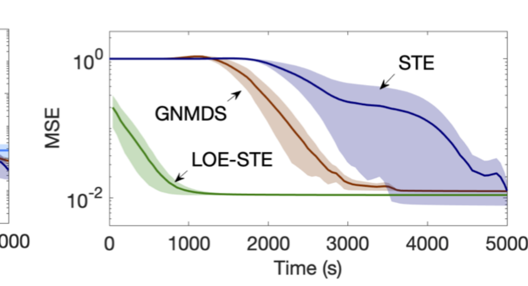


(a) $(n, d, c) = (10^5, 2, 200)$

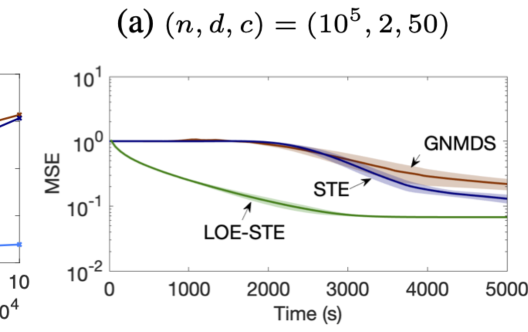


(b) Time to LOE error vs n

Figure 2: Scalability

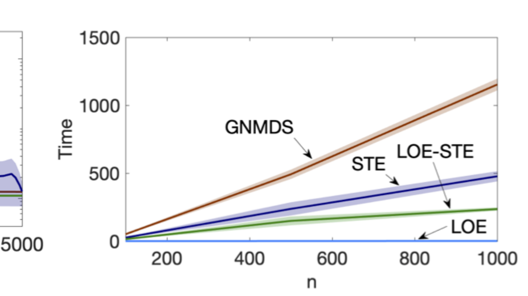


(a) $(n, d, c) = (10^5, 2, 50)$

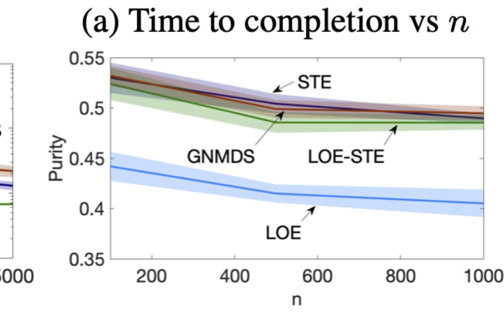


(b) $(n, d, c) = (2 \times 10^4, 10, 200)$

Figure 3: Warm-start



(a) Time to completion vs n



(b) Purity vs n

Figure 4: MNIST

References

[VDMW12] Laurens Van Der Maaten and Kilian Weinberger. "Stochastic triplet embedding". In: *Machine Learning for Signal Processing (MLSP)*. 2012, pp. 1–6.

[Aga+07] Sameer Agarwal, Josh Wills, Lawrence Cayton, Gert Lanckriet, David Kriegman, and Serge Belongie. "Generalized non-metric multidimensional scaling". In: *AISTATS*. 2007, pp. 11–18.

[DST04] Vin De Silva and Joshua B Tenenbaum. *Sparse multidimensional scaling using landmark points*. Tech. rep. 2004.