Recent methods for dimensionality reduction: A brief comparative analysis

D.H. Peluffo-Ordóñez[1], John A. Lee [1,2], Michel Verleysen [1]

[1] **Université catholique de Louvain**, Machine Learning Group - ICTEAM,
[2] **Université catholique de Louvain**, Molecular Imaging Radiotherapy and Oncology - IREC.

Dimensionality reduction (DR)

- DR is a key stage for both the design of a pattern recognition system or data visualization.
- Formally, the goal of DR is to embed a high dimensional (HD) data matrix into a low-dimensional (LD), latent data matrix.
- Recently, there has been an increasing interest in those methods aimed at preserving the data topology.
- Among them, Laplacian eigenmaps (LE) and stochastic neighbour embedding (SNE) are the most representative.

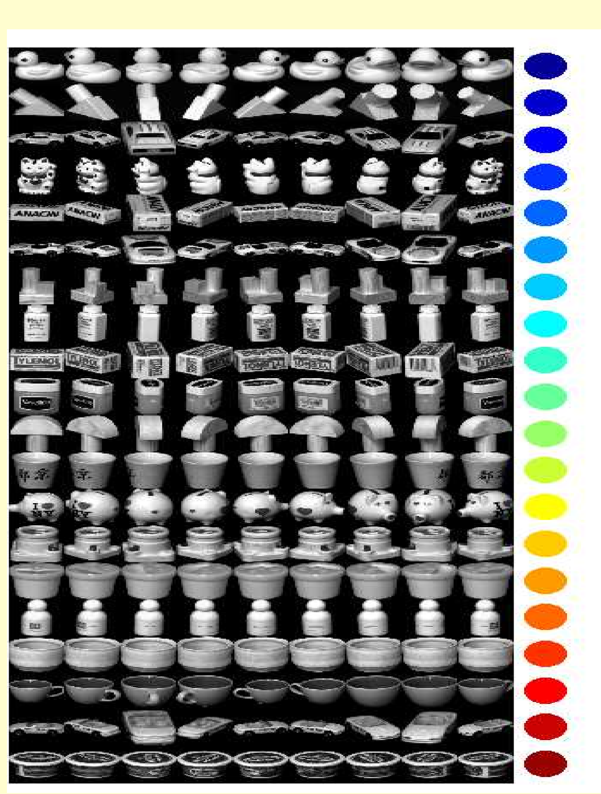
Scope of this work

- In this work, we present a brief comparative among very recent methods being alternatives to LE and SNE.
- Comparisons are done mainly on two aspects: algorithm implementation, and complexity.
- Also, relations between methods are depicted.
- The goal of this work is to provide researches on this field with some discussion and criteria decision to choose a method according to the user's needs.

Review of DR methods

- Notation:
  - HD data matrix: $\boldsymbol{Y} = [\boldsymbol{y}_i]_{1\leq i \leq N}$, such that $\boldsymbol{y}_i \in \mathbb{R}^D$
  - LD data matrix: $\boldsymbol{X} = [\boldsymbol{x}_i]_{1\leq i \leq N}$, being $\boldsymbol{x}_i \in \mathbb{R}^d$, where $d < D$
  - Similarity: $\boldsymbol{W} = [w_{nm}]_{1\leq n \leq N}$, Laplacian: $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$, $\boldsymbol{D} = \mathrm{Diag}(\boldsymbol{W}\boldsymbol{1}_N)$.
  - $\boldsymbol{P}_n = [p_{nm}]_{1\leq m \leq N}$ and $\boldsymbol{Q}_n = [q_{nm}]_{1\leq m \leq N}$.
- Spectral methods:
  - LE [Belkin and Niyogi, 2003]: $E_{\mathrm{LE}} = \mathrm{tr}(\boldsymbol{X}\boldsymbol{L}\boldsymbol{X}^\top)$s. t. $\boldsymbol{X}\boldsymbol{D}\boldsymbol{X}^\top = \boldsymbol{I}_d$.
  - A fast LE version can be reached by using only a subset of L locally linear landmark (LLL) $\widetilde{\boldsymbol{Y}} \in \mathbb{R}^{L\times N}$ such that: $\boldsymbol{Y} \approx \widetilde{\boldsymbol{Y}}\boldsymbol{Z}$, being $\boldsymbol{Z}$ the projection matrix minimizing $||\boldsymbol{Y} - \widetilde{\boldsymbol{Y}}\boldsymbol{Z}||^2$ [Vladymyrov and Carreira-Perpinán, 2013].
- Divergence-based methods:
  - SNE [Hinton and Roweis, 2002]: $E = \sum_{n=1}^{N} \mathrm{D}_{\mathrm{KL}}(\boldsymbol{P}_n||\boldsymbol{Q}_n) = \sum_{n,m=1}^{N} p_{nm} \log \frac{p_{nm}}{q_{nm}}$, where D_{KL} denotes the Kullback-Leibler directed divergence.
  - Symmetric SNE (SSNE): By symmetrizing $\boldsymbol{P}$ and $\boldsymbol{Q}$.
  - t-SNE [Van der Maaten and Hinton, 2008]: By forcing that $\boldsymbol{Q}$ follows a t-Student distribution.
  - Jensen-Shannon embedding (JSE) [Lee et al., 2013]: $E_{\mathrm{JSE}} = \sum_{n=1}^{N} \mathrm{D}^{\beta}_{\mathrm{KL}}(\boldsymbol{Q}_n||\boldsymbol{S}_n)$ where $\mathrm{D}^{\beta}_{\mathrm{KL}} = (1-\beta)\,\mathrm{D}_{\mathrm{KL}}(\boldsymbol{P}_n||\boldsymbol{S}_n) + \beta\,\mathrm{D}_{\mathrm{KL}}(\boldsymbol{Q}_n||\boldsymbol{S}_n)$, and $\boldsymbol{S}_n = (1-\beta)\boldsymbol{P}_n + \beta\boldsymbol{Q}_n$.

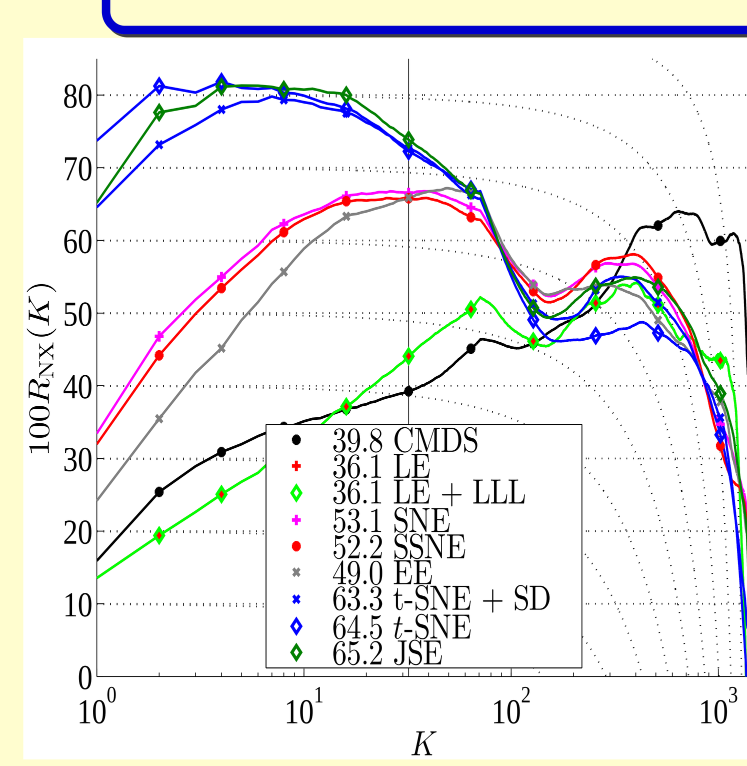
Experimental setup

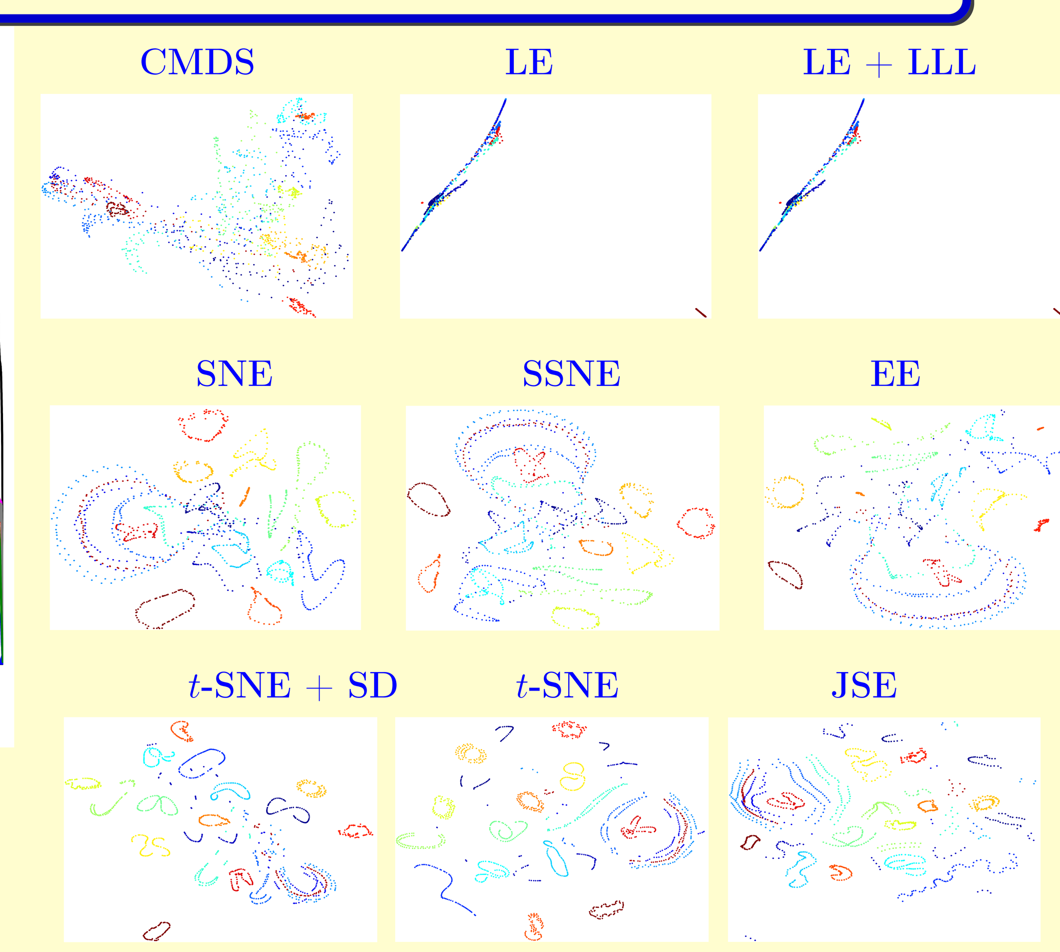


Data set: COIL-20

- DR are tested over the same data set.
- To quantify the performance, the scaled average agreement rate $R_{NX}(K)$ [Lee et al., 2013] is used. This measure means the relative position between between a perfect embedding and a random one.
- A numerical indicator of the overall performance is given by the area under the curve (AUC) of $R_{NX}(K)$.

Experimental results





Conclusion

This work gathers some key aspects to compare dimensionality reduction methods. Namely, relations between them, algorithm implementation, and complexity/processing time. Very recent methods were studied such as elastic embedding, locally linear landmarks for Laplacian eigenmaps and Jensen-Shanon embedding. Discussion and hints provided here may facilitate users to chose a method according the trade-off between performance and complexity.

Comparative and discussion

- Spectral methods, in general, attempt to preserve the global structure.
- LLL is a good alternative to initialize LE meaning a decreasing of the processing time when $\mathcal{O}(N^2 d) + \mathcal{O}(\frac{1}{3}N) + \mathcal{O}(L^3) < \mathcal{O}(N^3)$.
- SNE-like methods perform a better embedding preserving smaller neighbours (local structure). We can notice that SNE, SSNE and EE have a similar performance.
- Spectral direction (SD) makes that SNE and EE behave as a symmetrized version due to strong assumption on the gradient calculation.
- t-SNE + SD yields a better embedding since t-distributed neighborhoods may improve the separation of underline clusters. JSE outperforms the remaining considered methods thanks to its symmetric divergence.

Main references

[Belkin and Niyogi, 2003] Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

[Hinton and Roweis, 2002] Hinton, G. E. and Roweis, S. T. (2002). Stochastic neighbor embedding. In *Advances in neural information processing systems*, pages 833–840.

[Lee et al., 2013] Lee, J. A., Renard, E., Bernard, G., Dupont, P., and Verleysen, M. (2013). Type 1 and 2 mixtures of kullback-leibler divergences as cost functions in dimensionality reduction based on similarity preservation. *Neurocomputing*.

[Van der Maaten and Hinton, 2008] Van der Maaten, L. and Hinton, G. (2008). Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(2579-2605):85.

[Vladymyrov and Carreira-Perpinán, 2013] Vladymyrov, M. and Carreira-Perpinán, M. Á. (2013). Locally linear landmarks for large-scale manifold learning. In *Machine Learning and Knowledge Discovery in Databases*, pages 256–271. Springer.