Automatic Model Selection in Subspace Clustering via Triplet Relationships

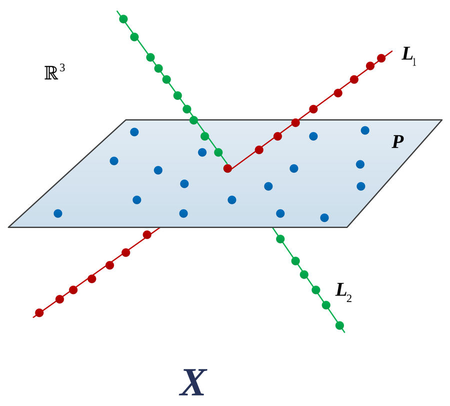
Jufeng Yang, Jie Liang, Kai Wang, Yong-Liang Yang, Ming-Ming Cheng
杨巨峰，梁杰(liang27jie@163.com)，王恺，杨永亮，程明明

南開大學 Nankai University

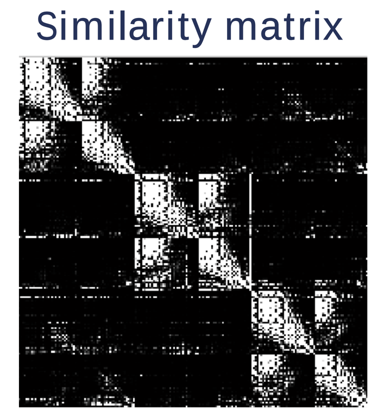
1. Overview of subspace clustering

Subspace clustering models high-dimensional data samples into a union of low-dimensional linear subspaces. Traditional subspace clustering methods first optimize the following self-representation problem followed by employing a spectral clustering to calculate the final assignments.

Dataset

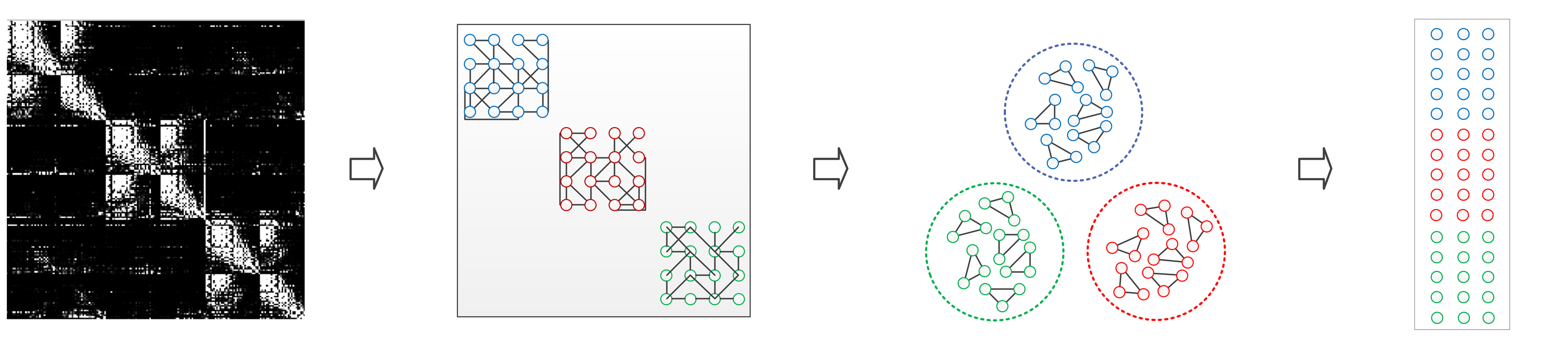


Similarity matrix



$$\underbrace{\min_{C} L(XC, X)}_{\text{Reconstruction Loss}} + \underbrace{\lambda \|C\|_{\xi}}_{\text{Regularization}} \quad \underbrace{s.t.\ c_{ii} = 0}_{\text{Constraint}}$$

2. Triplet Relationship



Similarity Matrix ⇨ Neighboring Relevance ⇨ Triplet Relationships ⇨ Triplet Matrix

m Nearest Neighbor: $N_m(x_j) = \arg\max_{\{x_{i_l}\}} \sum_{l=1}^{m} |c_{i_l,j}|$

Triplet Relationship: $\underbrace{\mathbf{1}_{x_i \in N_m(x_j)} \times \mathbf{1}_{x_j \in N_m(x_k)} \times \mathbf{1}_{x_k \in N_m(x_i)}}_{\text{Indicator Functions}} = 1 \Rightarrow \underbrace{\{x_i, x_j, x_k\} \in \tau}_{\text{Triplet}}$

Proposition: **Given arbitrary three samples in same triplet, we have that these three samples belong to same subspace.**

Advantages against pairwise correlation:

1. Triplet relationship is more robust when partitioning the inter-cluster samples near the intersection of two subspaces due to the complementarity of multiple constrains.
2. It evokes mutual restrictions of neighbored samples thus depicts a local geometrical structure, by which we can calculate the segmentation greedily.

3. Clustering Procedure

Algorithm 1 : Automatic Subspace Clustering (autoSC)

Input: $\boldsymbol{X} = [\boldsymbol{x}_1, \cdots, \boldsymbol{x}_N] \in \mathbb{R}^{D \times N}$.

Calculate the similarity matrix $\boldsymbol{C}$ with self-representation optimization;
for $i = 1 : N$ **do**
  Calculate the m Nearest Neighbors $\boldsymbol{N}_m(\boldsymbol{x}_i)$;
end for
Discovering the Triplet Relationships and generate the triplet matrix $\boldsymbol{T} \in \mathbb{R}^{n \times 3}$;
Reshape $\boldsymbol{T}$ to $\boldsymbol{X}_{out} \in \mathbb{R}^{3n}$, $\boldsymbol{X}_{in} = \varnothing$;
$\widetilde{K} = 1$;
Find the initialized triplet $\boldsymbol{\tau}_{ini}^{\widetilde{K}}$ according to local density;
while $\rho(\boldsymbol{\tau}_{ini}^{\widetilde{K}}, \boldsymbol{X}_{out}) > \rho(\boldsymbol{\tau}_{ini}^{\widetilde{K}}, \boldsymbol{X}_{in})$ **do**
  $\mathcal{C}_{\widetilde{K}} = \boldsymbol{\tau}_{ini}^{\widetilde{K}}$;
  $\mathcal{C}_{\widetilde{K}} = \mathcal{C}_{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\}$ where $\boldsymbol{\tau}^*$ has high-relevance with $\mathcal{C}$;
  $\boldsymbol{X}_{in} = \boldsymbol{X}_{in} \cup \boldsymbol{\tau}_{ini}^{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\}$;
  $\boldsymbol{X}_{out} = \boldsymbol{X}_{out} / (\boldsymbol{\tau}_{ini}^{\widetilde{K}} \cup \{\boldsymbol{\tau}^*\})$;
  $\widetilde{K} = \widetilde{K} + 1$;
  Calculate $\boldsymbol{\tau}_{ini}^{\widetilde{K}}$;
end while
Merge $\mathcal{C}_i$ and $\mathcal{C}_j$ if we have $s_{ij} > \min(|\mathcal{C}_i|, |\mathcal{C}_j|)$; Get $\widehat{K}$ clusters;
for $j = 1 : |\boldsymbol{X}_{out}|$ **do**
  Calculate $\mathcal{C}^*$ for $\boldsymbol{x}_j$ using the fusion reward: $\mathcal{C}^* = \arg\max_{\mathcal{C}_i} R_f(\mathcal{C}_i|\boldsymbol{x}_j), \quad i \in \{1, 2, \cdots, \widehat{K}\}$;
end for

Output: The cluster assignment $\{\mathcal{C}_i\}_{i=1}^{\widehat{K}}$.

Model Selection Reward: $R_m(\mathcal{C}) = \sum_i f(\mathcal{C}_i|\boldsymbol{X}_{out}^I) - \lambda_m \sum_i f(\mathcal{C}_i|\boldsymbol{X}_{in}^I)$

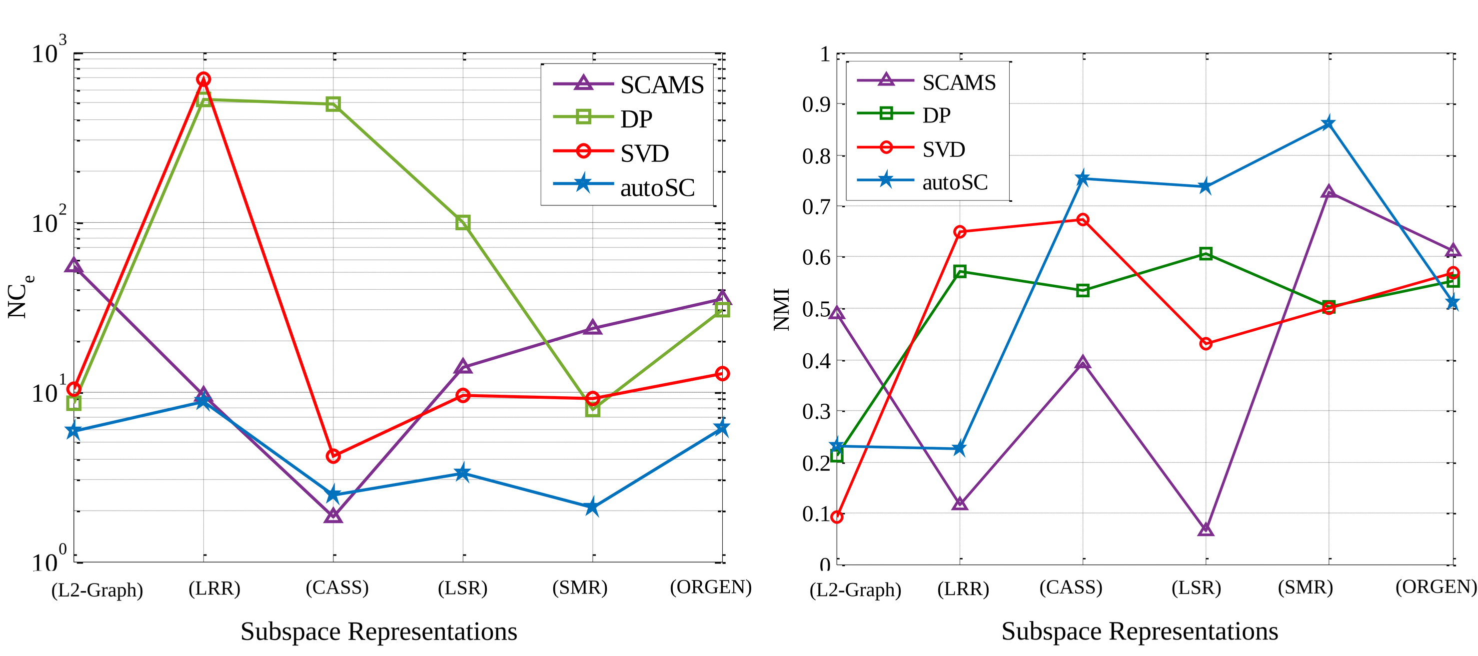
Fusion Reward: $R_f^i(\mathcal{C}_i|\boldsymbol{x}_j \in \boldsymbol{X}_{out}) = f(\boldsymbol{x}_j|\mathcal{C}_i) + \lambda_f f(\boldsymbol{N}_m(\boldsymbol{x}_j)|\boldsymbol{N}_m(\mathcal{C}_i))$

Local Density of Triplet: $\rho(\boldsymbol{\tau}, \boldsymbol{X}_{out}) = \sum_{j=1}^{|\boldsymbol{n}|} f(\boldsymbol{x}_{\boldsymbol{n}_j}|\boldsymbol{X}_{out})$

4. Performance

Datasets		LRR	CASS	LSR	SMR	ORGEN
eYaleB	8	0.0155	0.0158	0.0144	0.0135	0.0166
	15	0.0147	0.0148	0.0148	0.0149	0.0145
	30	0.0176	0.0157	0.0181	0.0181	0.0177
COIL20	5	0.0185	0.0195	0.0188	0.0175	0.0196
	10	0.0252	0.0198	0.0188	0.0182	0.0210
	15	0.0224	0.0203	0.0212	0.0196	0.0215

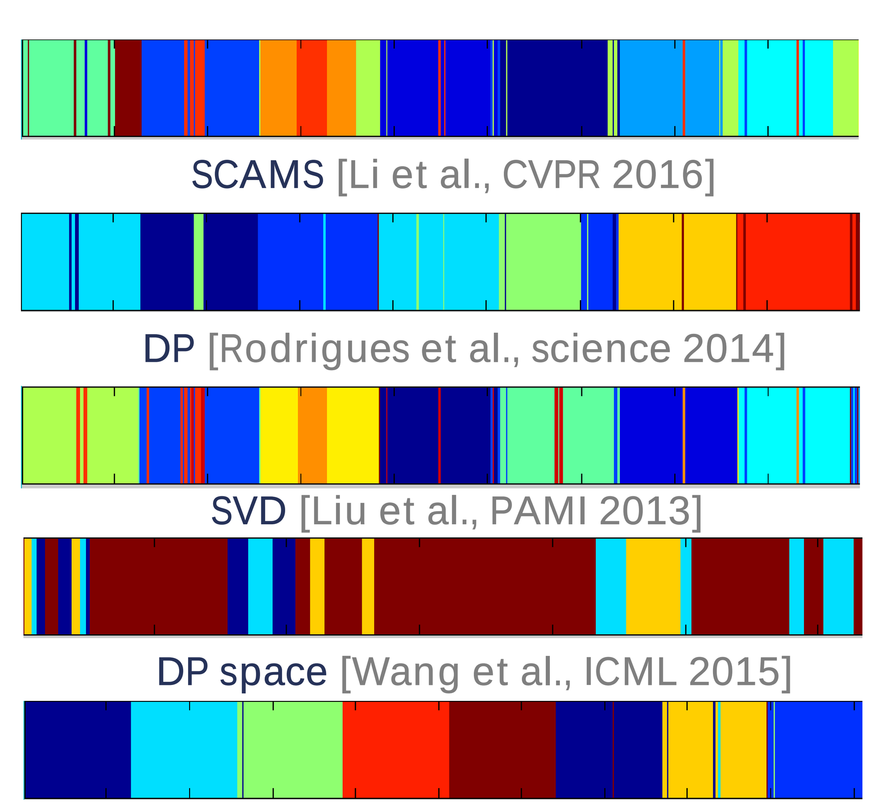
Error rate of the triplet relationships on top of different self-representation schemes. The robustness of triplet relationship guarantees the performance of model selection and subspace clustering in the unified model.



Evaluation of the extensions to different self-representations on Extended Yale B dataset with 8 subjects using NC_e (Left) and NMI (right). As shown, the proposed method achieves favorable performance against other contrastive methods.

Methods	Metrics	extended Yale B 8	extended Yale B 15	extended Yale B 30	COIL-20 5	COIL-20 10	COIL-20 15
SCAMS	NC_e	9.26	23.60	76.22	8.48	19.72	32.40
	NMI	0.7183	0.7272	0.7266	0.5885	0.6527	0.6668
DP	NC_e	3.06	7.84	24.76	2.22	5.30	9.72
	NMI	0.6196	0.5026	0.2166	0.6864	0.4467	0.3643
SVD	NC_e	2.40	9.06	24.00	0.48	2.58	8.36
	NMI	0.7078	0.4993	0.2808	0.7024	0.7127	0.7224
DP-space	NC_e	2.08	8.96	23.92	0.78	4.78	9.38
	NMI	0.0343	0.0226	0.0406	0.0904	0.0829	0.0718
autoSC	NC_e	**0.76**	**2.08**	**4.98**	**0.38**	**1.18**	**0.80**
	NMI	**0.9062**	**0.8589**	**0.8287**	**0.8315**	**0.7701**	**0.7266**

Overall comparison between autoSC and other contrastive methods on subsets of extended Yale B and COIL-20 datasets.



Visualization of the clustering labels for all contrastive methods conducted on extended Yale B dataset with 8 subjects.

5. Conclusion

1. The triplet relationship induces a high relevance and locality which is validated to be favorable against the traditional pairwise correlation.

2. Our unified framework for joint model selection and subspace clustering explores the intrinsic geometrical structures using triplet relationships, and outperforms the existing methods.

Any comments or suggestions are welcome.
Email: liang27jie@163.com
Homepage: http://cv.nankai.edu.cn/