Adaptive graph convolutional subspace clustering

自适应图卷积子空间聚类

魏莱 (Wei Lai)

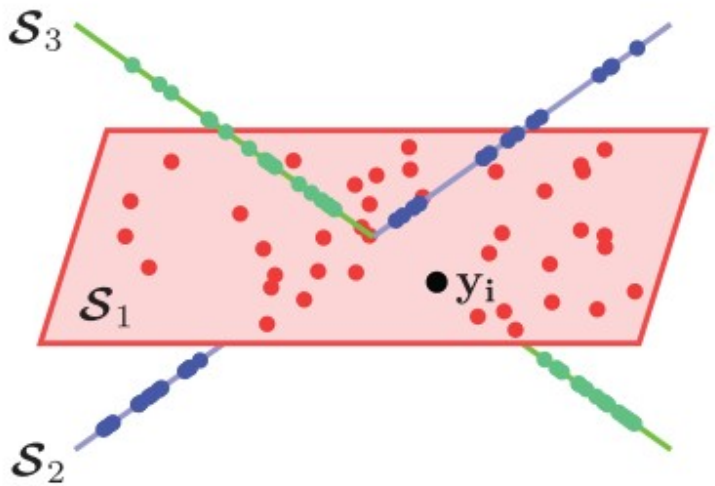
上海海事大学
Shanghai Maritime University
weilai@shmtu.edu.cn

2023 年 5 月 29 日

上海海事大学
SHANGHAI MARITIME UNIVERSITY

Conception of subspace clustering

- Subspace clustering is an attractive topic in machine learning and computer vision fields;
- Subspace clustering aims to arrange the high-dimensional data samples into a union of linear subspaces where they are generated from;



Formulation of subspace clustering problem

- Suppose a data matrix $\mathbf{X} = [\mathbf{x}_1; \mathbf{x}_2; \cdots; \mathbf{x}_n] \in \mathscr{R}^{n \times d}$ contains n data samples drawn from k subspaces, and d is the number of features. The general formulation of a spectral-type subspace clustering algorithm could be expressed as follows:

$$\begin{aligned} \min_{\mathbf{C}} \quad & \Omega\big(\Phi(\mathbf{X}) - \mathbf{C}\Phi(\mathbf{X})\big) + \lambda \Psi(\mathbf{C}), \\ s.t. \quad & \Theta(\mathbf{C}) = \mathbf{0} \end{aligned} \tag{1}$$

- $\mathbf{C} \in \mathscr{R}^{n \times n}$ is the reconstruction coefficient matrix and $\Psi(\mathbf{C})$ is usually some kind of constraint;
- $\Phi(\cdot)$ is a function used to find the meaningful latent features for original data samples;
- $\Omega(\cdot)$ is a function to measure the reconstruction residual;
- $\Theta(\mathbf{C})$ is some additional constraints.

The existing subspace clustering models

- Classical spectral-type subspace clustering algorithms mainly focus on designing $\Psi(\mathbf{C})$

Table 1: The different functions applied in some representative subspace clustering algorithms.

Algorithms	$\Phi(\cdot)$	$\Omega(\cdot)$	$\Psi(\cdot)$	$\Theta(\cdot)$
SSC	$\mathbf{X}$	$\|\cdot\|_1$	$\|\mathbf{C}\|_1$	$\delta(\mathbf{C})=\mathbf{0}_n$
LRR	$\mathbf{X}$	$\|\cdot\|_{2,1}$	$\|\mathbf{C}\|_*$	$-$
LSR	$\mathbf{X}$	$\|\cdot\|_F^2$	$\|\mathbf{C}\|_F^2$	$\delta(\mathbf{C})=\mathbf{0}_n$
BDR	$\mathbf{X}$	$\|\cdot\|_F^2$	$\|\mathbf{C}\|_{\mathscr{K}}$	$\delta(\mathbf{C})=\mathbf{0}_n$, $\mathbf{C}=\mathbf{C}^\top$, $\mathbf{C}\geq\mathbf{0}_{n\times n}$
CASS	$\mathbf{X}$	$\|\cdot\|_F^2$	$\sum_{i=1}^n \|\Delta(\mathbf{c}_i)\mathbf{X}\|_*$	$-$
LS3C	$\mathbf{XP}$	$\|\cdot\|_F^2$	$\|\mathbf{C}\|_1$	$\mathbf{P}^\top\mathbf{P}=\mathbf{I}_{n\times n}$, $\mathbf{XPP}^\top=\mathbf{X}$
RKLRR	$\phi(\mathbf{X})$	$\|\cdot\|_F^2$	$\|\mathbf{C}\|_*$	$-$
DSCN	$\mathbb{E}(\mathbf{X})$	$\|\cdot\|_F^2$	$\|\mathbf{C}\|_1$ or $\|\mathbf{C}\|_F^2$	$\mathbb{D}(\mathbb{E}(\mathbf{X}))=\mathbf{X}$

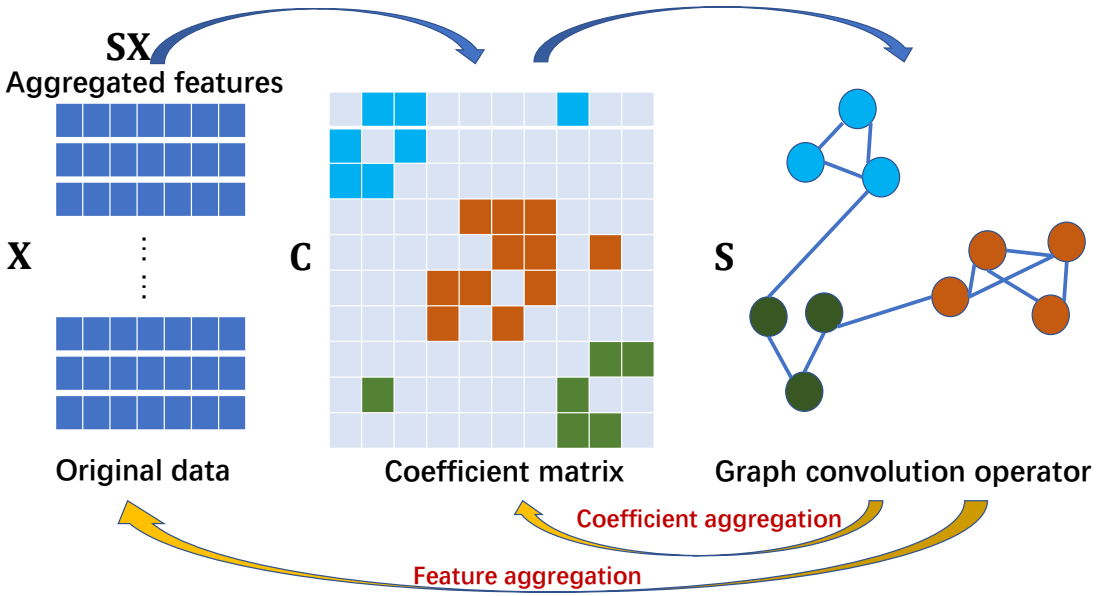
- Usually, deep subspace clustering models achieve the best results.

The weakness of current DSC models

- Deep neural networks can produce highly-expressive mapping functions, the self-expression term in the latent space becomes a regularization function;
- The embedded data geometry is trivial, so the success of deep subspace clustering algorithms may be attributed to an ad-hoc post-processing strategy.

Idea

- To use GCN technique to design a feature extraction approach and a regularize for reconstruction coefficient matrix simultaneously.
  - The obtained reconstruction coefficient matrix is used to construct a graph convolution operator $\mathbf{S}$;
  - The graph convolution operator is used to design 1: a feature extraction method and 2: a constraint of $\mathbf{C}$;
  - $\mathbf{S}$ and $\mathbf{C}$ affect each other.



Backgrouds: Graph convolutional networks (GCNs)

- Classical GCN:
  - At the l-th layer of a GCN model, the features $\mathbf{H}^{l-1}$ $(\mathbf{H}^0 = \mathbf{X})$ of each node are averaged with the feature vectors in its local neighborhood first. Then the aggregated features are transformed linearly.
  - A nonlinear activation (e.g. ReLU()) is applied to output new feature representations $\mathbf{H}^l$.
  - Objection :

$$\mathbf{H}_l \leftarrow \sigma(\mathbf{S}\mathbf{H}_{l-1}\mathbf{W}_{l-1}), \tag{2}$$

where $\mathbf{S} = \mathbf{D}'^{-\frac{1}{2}}\mathbf{A}'\mathbf{D}'^{-\frac{1}{2}}$, $\mathbf{A}' = \mathbf{A} + \mathbf{I}$, $\mathbf{D}'$ is the degree matrix of $\mathbf{A}'$.

Backgrouds: Simple graph convolution

- SGC:
  - It claims that “the nonlinearity between GCN layers is not critical - but that the majority of the benefit arises from the local averaging";
  - Objection:

$$\mathbf{H}_l = \mathbf{S}\mathbf{H}_{l-1}\mathbf{W}_{l-1}. \tag{3}$$

By integrating several graph convolutional layers, the final feature representation is

$$\mathbf{F} = \overbrace{\mathbf{S}\cdots\mathbf{S}}^{M}\mathbf{X}\mathbf{W}_1\cdots\mathbf{W}_M = \mathbf{S}^M\mathbf{X}\mathbf{W} \tag{4}$$

where M is the number of graph convolutional layers, the $\mathbf{W} = \mathbf{W}_1\cdots\mathbf{W}_M$ is also a linear transformation matrix.

Backgrouds: GCN-related subspace clustering algorithms

- GCSC (Graph Convolutional Subspace Clustering)[1]:

$$\begin{aligned} \min_{\mathbf{C}} \quad & \|\mathbf{X}-\mathbf{CSX}\|_F^2+\lambda\|\mathbf{C}\|_F^2, \\ s.t. \quad & diag(\mathbf{C})=\mathbf{0} \end{aligned} \tag{5}$$

- Graph Filter LSR[2]:

$$\min_{\mathbf{C}} \quad \|\mathbf{FX}-\mathbf{CFX}\|_F^2+\lambda\|\mathbf{C}\|_F^2, \tag{6}$$

where $\mathbf{F}=(\mathbf{I}-\frac{\mathbf{L}}{2})^k$, $\mathbf{L}=\mathbf{I}-\mathbf{S}$, k is the order of the graph filter, $\mathbf{S}$ is obtained by using $\mathbf{C}$. FLSR is an iterative method.

[1] Y. Cai, Z. Zhang, Z. Cai, X. Liu, X. Jiang, and Q. Yan, "Graph Convolutional Subspace Clustering: A Robust Subspace Clustering Framework for Hyperspectral Image," IEEE Transactions on Geoscience and Remote Sensing, pp. 1-12, 2020.
[2]Z. Ma, Z. Kang, G. Luo, L. Tian, and W. Chen, "Towards Clustering-friendly Representations: Subspace Clustering via Graph Filtering," in International Conference on Multimedia, Seattle, WA, USA, 2020: ACM, pp. 3081–3089.

The proposed method: AGCSC

- 1. The feature extraction method in AGCSC.
  - Suppose $\mathbf{C}$ is obtained, an affinity matrix $\mathbf{A}$ could be constructed: $\mathbf{A} = (|\mathbf{C}| + |\mathbf{C}^\top|)/2$;
  - With some additional constraints, $\mathbf{C} = \mathbf{C}^\top, \mathbf{C} \geqslant \mathbf{0}, \mathbf{C1} = \mathbf{1}$ and $diag(\mathbf{C}) = \mathbf{0}$, we could obtain $\mathbf{A} = \mathbf{C}$, $\mathbf{S} = \mathbf{D'}^{-\frac{1}{2}}\mathbf{A'}\mathbf{D'}^{-\frac{1}{2}} = (\mathbf{C}+\mathbf{I})/2$;
  - Aggregated feature matrix: $\mathbf{F} = \mathbf{SX} = (\mathbf{C}+\mathbf{I})/2$.
- 2. The regularizer of the coefficient matrix in AGCSC.
  - The new representation of $\mathbf{C}$: $\mathbf{SC} = \frac{1}{2}(\mathbf{C}+\mathbf{I})\mathbf{C} = \frac{1}{2}(\mathbf{C}^2+\mathbf{C})$
  - $\mathbf{C}$ and $\mathbf{SC}$ should have a similar characteristic, then $\Psi(\mathbf{C}) = \|\mathbf{C} - \mathbf{SC}\|_F^2 = \|\mathbf{C} - \frac{1}{2}(\mathbf{C}^2+\mathbf{C})\|_F^2 = \|\frac{1}{2}\mathbf{C} - \frac{1}{2}\mathbf{C}^2\|_F^2 = \frac{1}{4}\|\mathbf{C} - \mathbf{C}^2\|_F^2$.

The objective function of AGCSC

- By combing the two terms, we have:

$$\begin{aligned} \min_{\mathbf{F},\mathbf{C}} \quad & \|2\mathbf{F}-(\mathbf{C}+\mathbf{I})\mathbf{X}\|_F^2+\alpha\|\mathbf{X}-\mathbf{C}\mathbf{F}\|_F^2+\beta\|\mathbf{C}-\mathbf{C}^2\|_F^2, \\ s.t. \quad & \mathbf{C}=\mathbf{C}^\top, \mathbf{C}\mathbf{1}=\mathbf{1}, \mathbf{C}\geqslant\mathbf{0}, diag(\mathbf{C})=\mathbf{0}, \end{aligned} \tag{7}$$

The optimization of AGCSC

- The above problem could be solved by using ADMM (alternating direction method of multipliers method)

$$
\begin{cases}
\mathbf{C}_{t+1} &= \big(2\mathbf{X}\mathbf{X}^\top + 2\alpha\mathbf{F}_t\mathbf{X}^\top + 2\beta(\mathbf{I}-\mathbf{Z}_t)(\mathbf{I}-\mathbf{Z}_t)^\top \\
&\quad +\mu_t(\mathbf{I}+\mathbf{1}\mathbf{1}^\top)\big)\big(4\mathbf{F}_t\mathbf{X}^\top - 2\mathbf{X}\mathbf{X}^\top + 2\alpha\mathbf{X}\mathbf{F}_t^\top \\
&\quad +\mu_l(\mathbf{Z}_t+\mathbf{1}\mathbf{1}^\top) - \mathbf{M}_t - \mathbf{N}_t\mathbf{1}^\top\big)^{-1}, \\
\mathbf{F}_{t+1} &= \big(\alpha\mathbf{C}_{t+1}^\top\mathbf{C}_{t+1} + 2\mathbf{I}\big)^{-1}\big((\mathbf{C}_{t+1}+\mathbf{I} \\
&\quad +\alpha\mathbf{C}_{t+1}^\top)\mathbf{X}\big), \\
\mathbf{Z}_{t+1} &= \big(2\beta\mathbf{C}_{t+1}^\top\mathbf{C}_{t+1} + \mu_t\mathbf{I}\big)^{-1}\big(2\beta\mathbf{C}_{t+1}^\top\mathbf{C}_{t+1} \\
&\quad +\mathbf{M}_t + \mu_t\mathbf{C}_{t+1}\big), \\
\mathbf{M}_{t+1} &= \mathbf{M}_t + \mu_t(\mathbf{C}_{t+1} - \mathbf{Z}_{t+1}), \\
\mathbf{N}_{t+1} &= \mathbf{N}_t + \mu_t(\mathbf{C}_{t+1}\mathbf{1} - \mathbf{1}), \\
\mu_{t+1} &= \min(\mu_{max}, \rho\mu_t),
\end{cases}
\tag{8}
$$

Block diagonal property

- Problem (8) could be regarded as a relaxed problem of the following problem:

$$\begin{aligned} \min_{\mathbf{C}} \quad & \|\mathbf{C}-\mathbf{C}^2\|_F^2 \\ s.t. \quad & \mathbf{X}=\mathbf{C}\mathbf{X}, \\ & \mathbf{C}=\mathbf{C}^\top, \mathbf{C}\mathbf{1}=\mathbf{1}, \mathbf{C}\geqslant\mathbf{0}, diag(\mathbf{C})=\mathbf{0}. \end{aligned} \tag{9}$$

- The solution to Problem (9) will be block diagonal.

Doubly stochastic property

- For the i-th block of $\mathbf{C}$, $|[\mathbf{C}_i]_{p,q} - [\mathbf{C}_i]_{s,t}| \leq 1$, where $[\mathbf{C}_i]_{p,q}$ and $[\mathbf{C}_i]_{s,t}$ are the (p,q)-th and $(s,t)-$th elements in $\mathbf{C}_i$ and $p,q,s,t \in 1,2,\cdots,n_i$;
- The differences in coefficients located in the same diagonal block will be small.

Post-processing strategy

- The frequently used post-processing strategy is to keep the m-largest values for each coefficient vector and discard the relatively small ones.
- AGCSC with thresholding post-processing skill is called TAGCSC.

Clustering results: Compared with shallow models

Dataset	Metric	Method										
		SSC	LRR	LRSC	LSR	BDR	TRR	FLSR	FTRR	GCSC	AGCSC	TAGCSC
ORL	ACC	72.50	72.75	77.00	75.50	78.25	***85.75***	73.50	79.75	73.75	80.50	**86.25**
	NMI	84.52	83.26	85.02	84.97	88.46	***91.49***	83.84	87.60	83.98	88.51	**92.84**
YALEB	ACC	55.15	73.48	75.64	74.05	76.56	91.65	72.94	***91.90***	62.50	84.79	**92.31**
	NMI	55.71	77.11	78.32	78.13	80.34	93.03	76.57	***93.18***	68.04	87.37	**94.04**
Umist	ACC	52.92	64.79	63.33	64.17	64.92	74.38	60.62	69.37	79.58	***81.04***	**90.83**
	NMI	75.38	73.41	72.02	73.17	75.13	80.63	70.72	78.49	86.44	***87.46***	**94.99**
COIL20	ACC	68.61	70.14	71.81	69.17	71.71	85.97	69.93	86.53	79.79	***88.75***	**98.96**
	NMI	66.85	76.43	77.27	74.17	80.51	90.23	77.19	91.17	85.67	***93.38***	**99.11**
COIL40	ACC	63.13	60.42	58.23	56.88	57.25	65.00	62.88	71.39	73.72	***78.12***	**92.60**
	NMI	82.28	76.29	74.48	75.87	76.73	79.83	76.26	82.46	84.32	***89.21***	**97.32**
MNIST	ACC	63.70	64.60	64.30	62.80	61.30	67.70	65.10	66.40	67.70	***71.40***	**72.80**
	NMI	59.75	60.67	58.91	57.18	54.76	64.43	61.10	63.21	61.99	***65.84***	**67.54**

Figure 1: Clustering results (in %) of various methods on the used benchmark data sets. The best results are emphasized in bold and the second best results are denoted in bold and italic.

Clustering results: Compared with deep models

Dataset	Metric	Method							
		AE+SSC	DSC-L1	DSC-L2	DEC	DKM	DCCM	AGCSC	TAGCSC
ORL	ACC	75.63	85.50	86.00	51.75	46.82	62.50	80.50	**86.25**
	NMI	85.55	90.23	90.34	74.49	73.32	79.06	88.51	**92.84**
YALEB	ACC	74.80	96.80	**97.33**	86.84	-	-	84.79	92.32
	NMI	78.33	96.87	**97.03**	92.40	-	-	87.34	94.04
Umist	ACC	70.42	72.42	73.12	55.21	51.06	54.48	81.04	**90.83**
	NMI	75.15	75.56	76.62	71.25	72.49	74.40	87.49	**94.99**
COIL20	ACC	87.11	93.14	93.68	72.15	66.51	80.21	88.75	**98.96**
	NMI	89.90	93.53	94.08	80.07	79.71	86.39	93.38	**99.11**
COIL40	ACC	73.91	80.03	80.75	48.72	58.12	76.91	78.12	**92.60**
	NMI	83.18	88.52	89.41	74.17	78.40	88.90	89.21	**97.23**
MNIST	ACC	48.40	72.80	**75.00**	61.20	53.32	40.20	71.40	72.80
	NMI	53.37	72.17	**73.19**	57.43	50.02	34.68	65.84	67.54

Figure 2: Clustering results (in %) of AGCSC and TAGCSC compared with several deep clustering methods. The best results are emphasized in bold. Some results are ignored because the results are not found in corresponding literatures.

Experiments: Feature aggregation

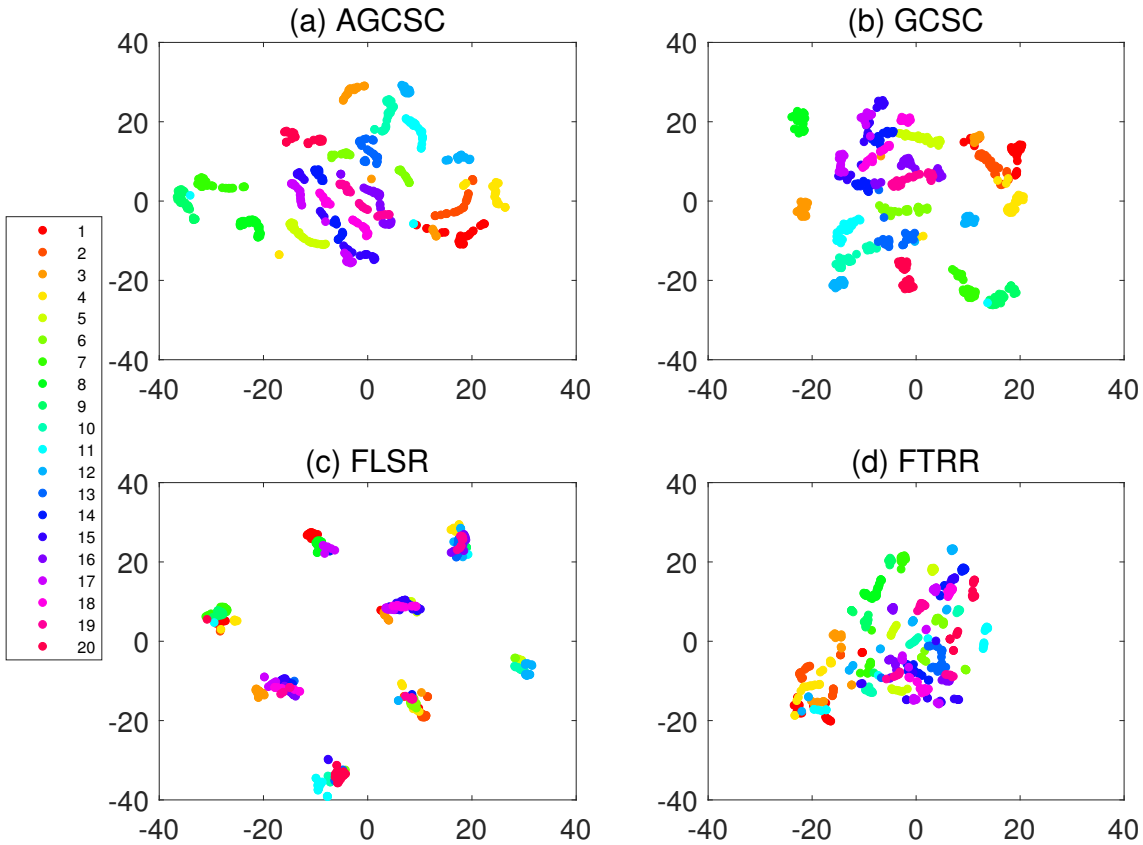


Figure 3: The visualisationa of aggregated feature representations obtained by (a) AGCSC, (b) GCSC, (c) FLSR and (d) FTRR on Umist data set.

Experiments: Block diagonal property of coefficient matrices 1

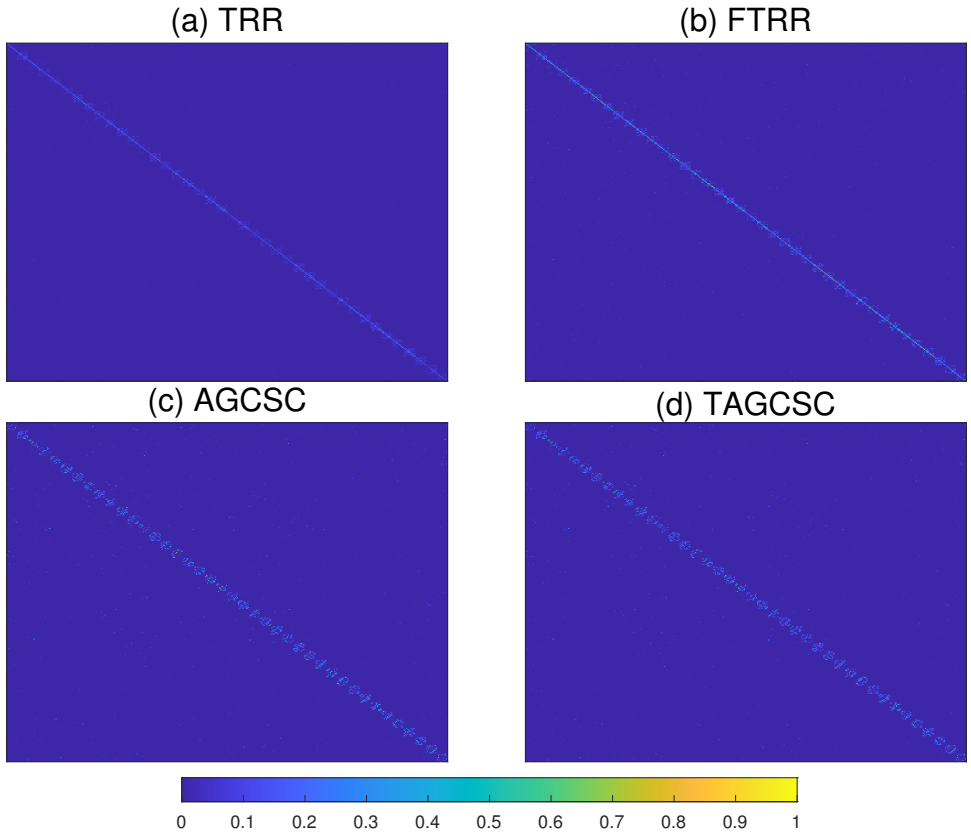


Figure 4: The obtained coefficient matrices obtained by (a) TRR, (b) FTRR, (c) AGCSC and (d) TAGCSC

Experiments: Block diagonal property of coefficient matrices 2

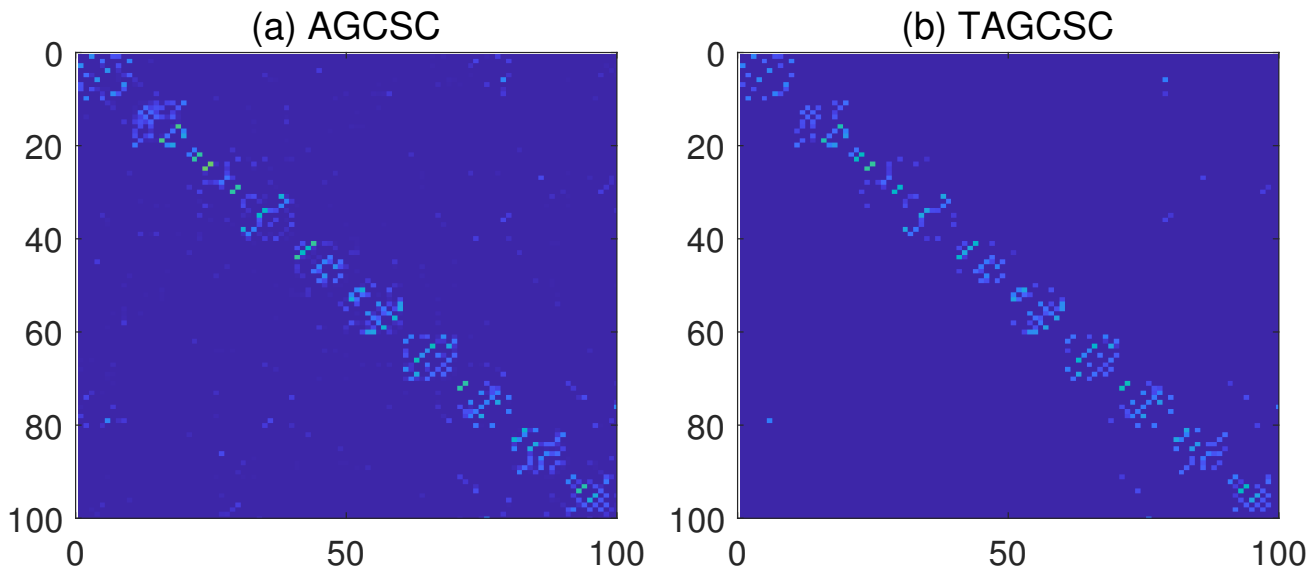


Figure 5: The partial coefficient matrices of (a) AGCSC and (b) TAGCSC corresponding to the samples from the first 10 classes.

Experiments: Parameter analyses (α and β)

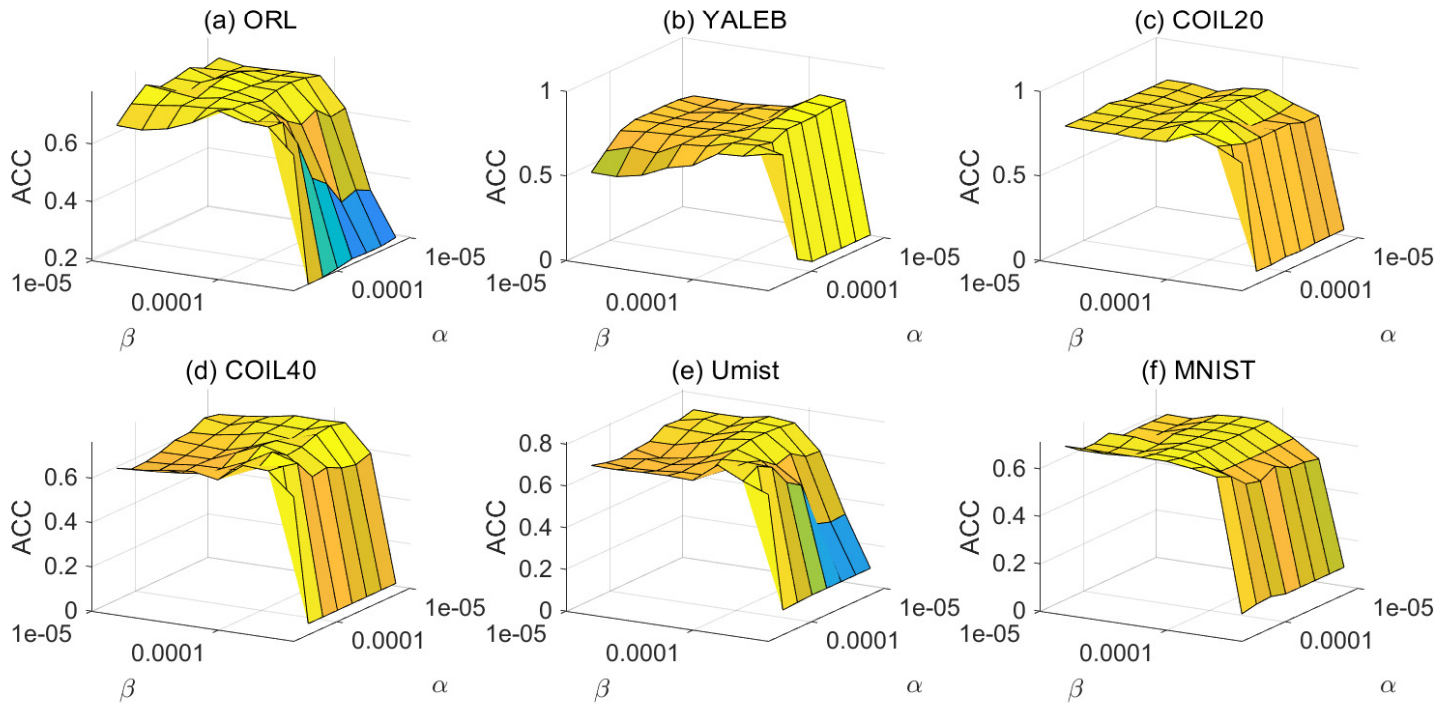


Figure 6: The influence of parameters α and β on clustering accuracy of AGCSC.

Experiments: Parameter analyses (m)

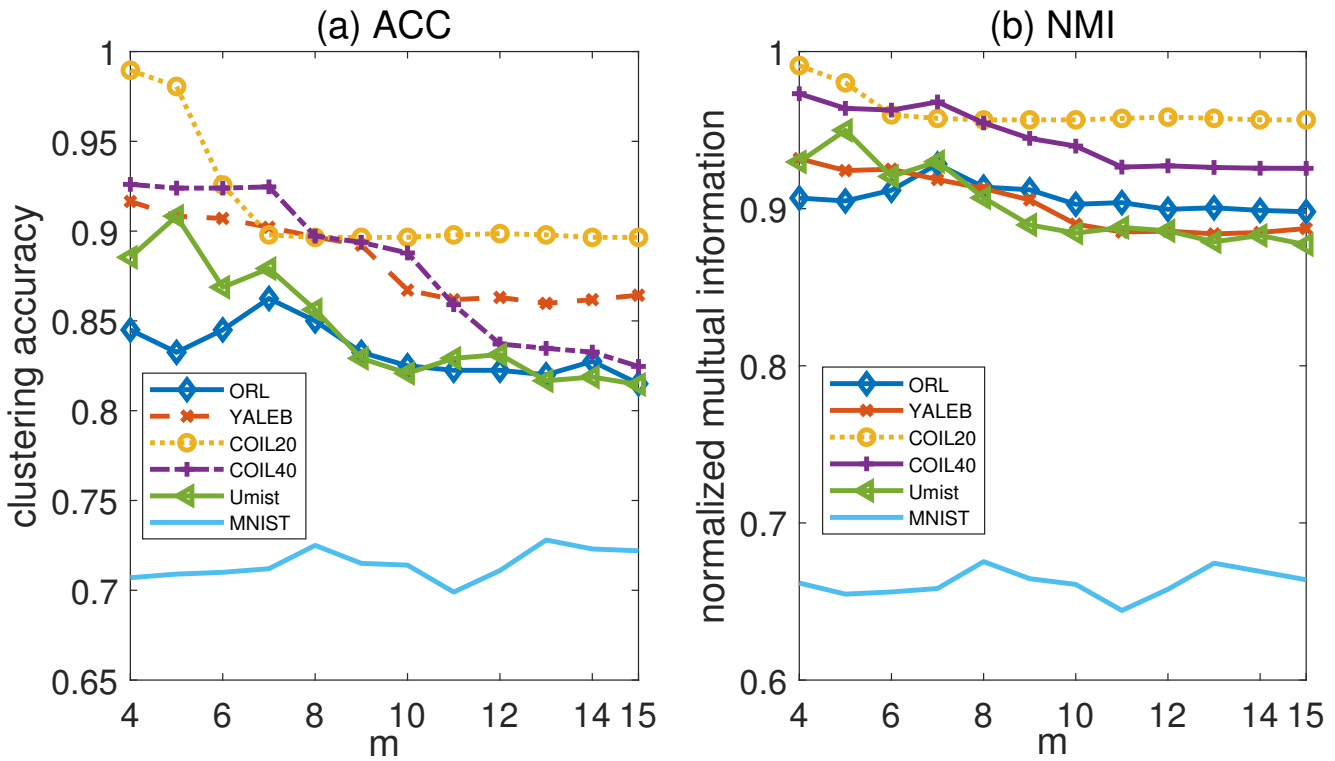


Figure 7: The influence of thresholding value m on the clustering performance of TAGCSC.

Experiments: Convergence analyses

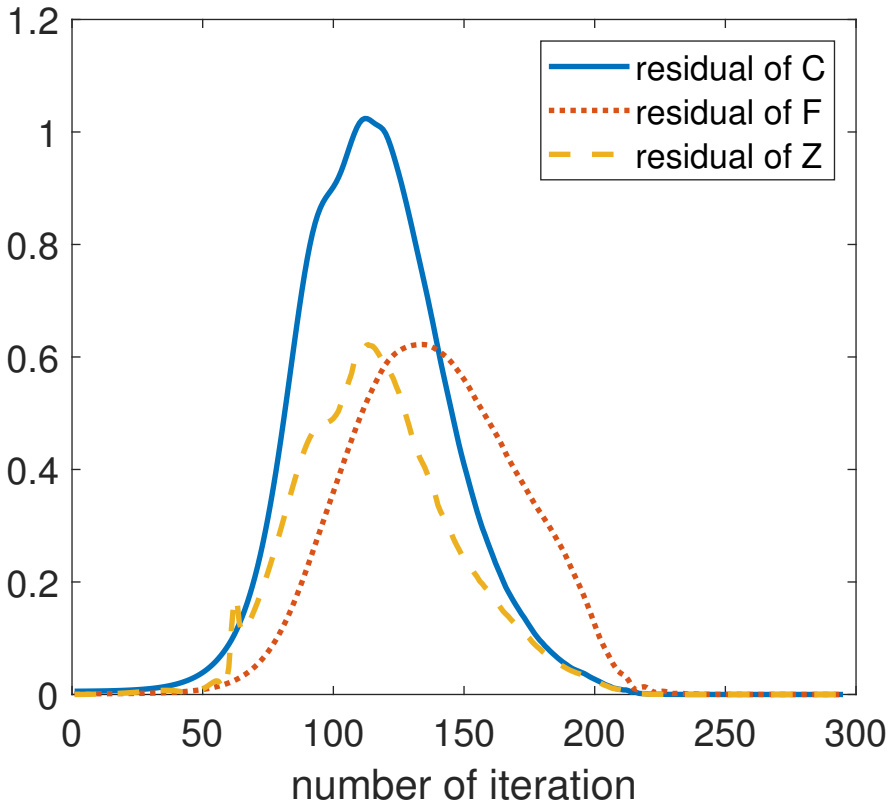


Figure 8: The residuals of variables $\mathbf{F}, \mathbf{C}, \mathbf{Z}$ versus the iterations on ORL database.

Experiments: Computation burden analyses

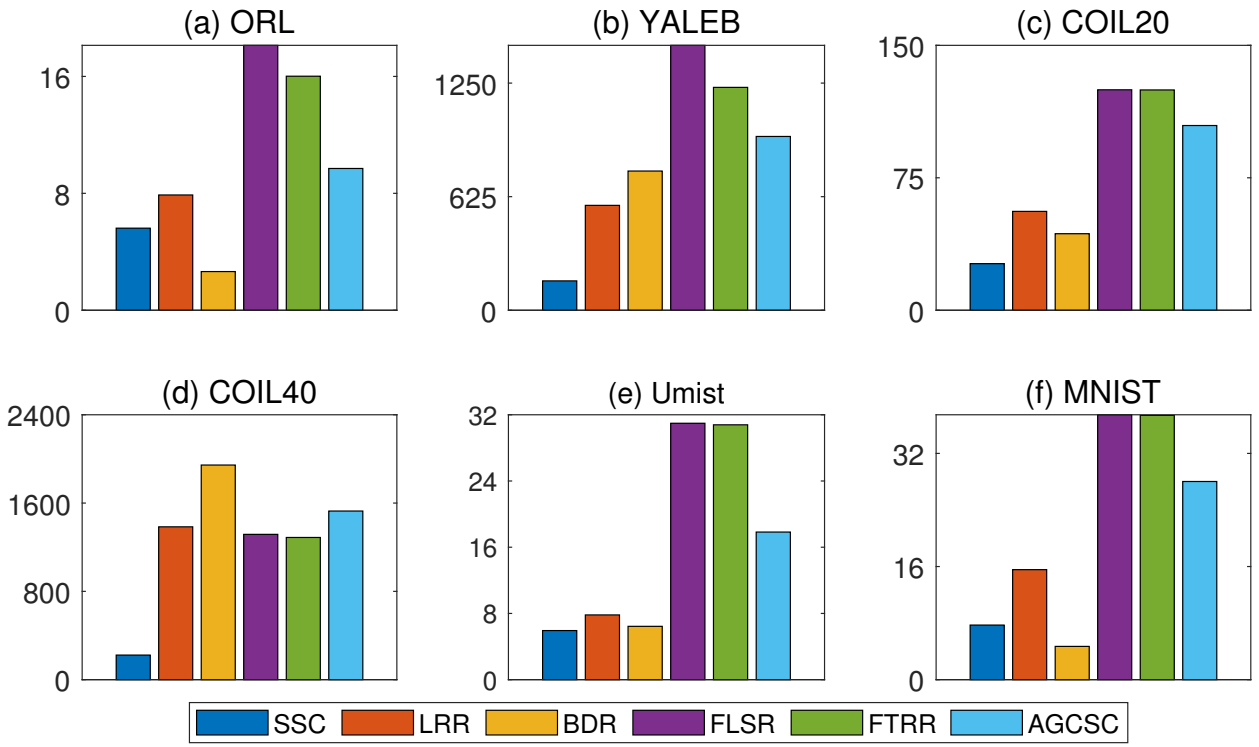


Figure 9: Computation time (seconds) for different algorithms to run on different datasets. The y-axis coordinates are in seconds.

Thanks!