Subspace Clustering via Joint Unsupervised Feature Selection

Wenhua Dong[1], Xiao-Jun Wu*[2], Hui Li, Zheng-Hua Feng, Josef Kittler
[1]wenhua_dong_jnu@jiangnan.edu.cn, [2]wu_xiaojun@jiangnan.edu.cn

No. 1104

Contribution:

We propose a novel subspace clustering method which integrates the unsupervised feature selection into subspace clustering. Different from most existing clustering methods, we uses the reconstructed feature matrix as the dictionary rather than the original data matrix, which strengthens the ability of our method. Related optimization problem is effectively solved using the half-quadratic and augmented Lagrange multiplier method. Experimental results on real datasets demonstrate the effectiveness of our method.

Model:

Given a collection $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ of n data samples drawn from a union of multiple subspaces. The left and right representation matrices w.r.t. $\boldsymbol{X}$ denote by $\boldsymbol{L} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{R} \in \mathbb{R}^{n\times n}$ called the feature and sample representation matrices, respectively. In order to remove the redundant features lying in the data, previous work in [1] design a feature selection matrix. Different from the existing approach, we focus on alleviating the influence of irrelevant features on clustering performance using the reconstructed feature matrix. The specific idea is to learn a more discriminative feature matrix $\boldsymbol{LX}$ as the dictionary, while at the same time learning a sample representation matrix $\boldsymbol{R}$ w.r.t. $\boldsymbol{LX}$ for subspace clustering. We consider the following minimization problem:

$$\begin{aligned} &\min_{\boldsymbol{R},\boldsymbol{L},\boldsymbol{E}} \ \|\boldsymbol{R}\|_F^2 + \lambda tr(\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^T) + \gamma\psi(\boldsymbol{E}) \\ &\text{s.t.} \ \ \boldsymbol{X} = \boldsymbol{L}\boldsymbol{X}\boldsymbol{R} + \boldsymbol{E}, \end{aligned} \tag{1}$$

where $\gamma > 0$ is the trade-off parameter, $\boldsymbol{E} \in \mathbb{R}^{m\times n}$ is the representation residual, and $\psi(\boldsymbol{E}) = \sum_{i,j}(1 - k_\sigma(e_{ij}))$ is the Correntropy induced regularization term with the kernel function $k_\sigma(\cdot)$. In this paper, we choose the Gaussian function $k_\sigma(x) = exp(-x^2/2\sigma^2)$ due to its simplicity and wide application, where σ is the kernel size. We note that the problem in (1) is reduced to UFS when $\boldsymbol{R}$ is set to $\boldsymbol{I}$. On the other hand, if we set $\boldsymbol{L} = \boldsymbol{I}$, (1) is the variant of LSR. We call the optimization problem in (1) Subspace Clustering via Joint Unsupervised Feature Selection (SC-UFS).

Results:

1) **We demonstrate the effectiveness of the proposed method on four real datasets: Yale B, AR, COIL-20 and USPS. Tables I and II presents the experimental results. We can see that the clustering performance of SC-UFS consistently outperforms the compared algorithms. This demonstrates the importance of reconstructing the feature matrix using representative features.**

TABLE I
CLUSTERING QUALITY (%) ON THE YALE B DATASET

Method		LSR	SSC	SSC-OMP	ℓ_0-SSC	LRR	LR-SC	Lat-LRR	FSC-NN	SC-UFS	SC-UFS(L)
2 subjects	CE	5.92	1.86	5.21	7.92	2.13	3.41	2.54	0.81	1.09	**0.70**
	Median	6.25	0.00	0.78	0.78	0.78	1.56	0.78	0.78	0.00	**0.00**
3 subjects	CE	9.31	3.28	5.38	10.97	3.50	5.80	4.21	1.20	1.47	**1.08**
	Median	9.38	0.52	2.08	4.17	2.08	4.69	2.60	1.04	0.52	**0.52**
5 subjects	CE	17.92	4.31	7.40	13.81	5.91	13.19	6.90	1.94	1.85	**1.25**
	Median	18.44	2.66	3.44	7.19	5.00	13.75	5.63	1.25	1.41	**0.94**
8 subjects	CE	29.09	5.85	9.82	14.53	11.05	29.69	14.34	2.56	2.08	**1.36**
	Median	29.49	4.49	5.86	8.59	7.42	32.13	10.06	1.95	1.76	**1.07**
10 subjects	CE	32.60	10.94	11.25	14.74	16.93	31.88	22.92	2.19	2.34	**1.51**
	Median	35.62	5.63	13.91	10.00	18.91	31.87	23.59	2.50	1.56	**1.09**

TABLE II
CLUSTERING QUALITY (%) ON THE AR, COIL AND USPS DATASETS

Method		LSR	SSC	SSC-OMP	ℓ_0-SSC	LRR	LR-SC	Lat-LRR	FSC-NN	SC-UFS	SC-UFS(L)
AR	CE	30.53	30.94	47.06	51.37	46.00	47.46	43.54	38.46	21.93	**20.79**
COIL-20	CE	37.99	12.15	49.50	14.72	41.04	39.44	40.76	24.65	15.83	**9.72**
USPS	CE	26.92	27.70	18.50	25.60	25.50	23.30	24.70	25.70	19.40	**10.70**

2) **This experiment is to test the sensitivity of the parameters on the Yale B dataset. We can observe that SC-UFS is pretty stable when λ or γ are chosen in an appropriate range.**

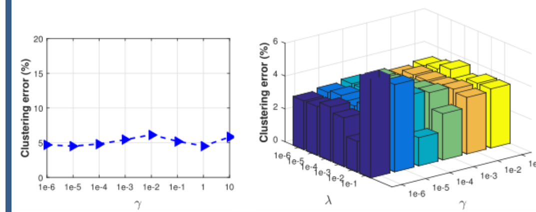

Fig. 1. Clustering error (%) versus different parameters γ and λ-γ on the Yale B dataset: SC-UFS (left) and SC-UFS(L) (right).

3) **The goal of this experiment is to demonstrate the convergence of SC-UFS on the Yale B dataset. We can see that the objective function of SC-UFS is decreased step by step and converged rapidly within 20 iterations.**

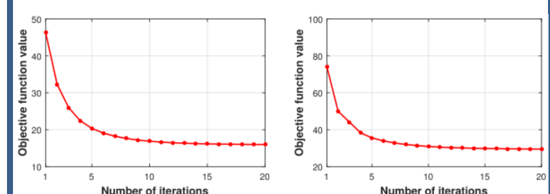

Fig. 2. Objective function values versus the number of iterations on the Yale B dataset: SC-UFS (left) and SC-UFS(L) (right).

References

[1] C. Peng, Z. Kang, M. Yang, and Q. Cheng, "Feature selection embedded subspace clustering," IEEE Signal Processing Letters, vol. 23, no. 7, pp. 1018–1022, 2016.