



# A Trace Ratio Maximization Method for Parameter Free Multiple Kernel Clustering

Yan Chen<sup>1</sup>, Lei Wang<sup>2</sup>, Liang Du<sup>2,3(✉)</sup>, and Lei Duan<sup>1(✉)</sup>

<sup>1</sup> School of Computer Science, Sichuan University, Chengdu, China  
chenyan557@stu.scu.edu.cn, leidian@scu.edu.cn

<sup>2</sup> School of Computer and Information Technology, Shanxi University,  
Taiyuan, China  
duliang@sxu.edu.cn

<sup>3</sup> Institute of Big Data Science and Industry, Shanxi University, Taiyuan, China

**Abstract.** Multiple Kernel Clustering (MKC) is helpful to leverage complementary information from various contexts and alleviate the difficulty of kernel determination. However, the key weighting strategies for optimal kernel learning around individual kernels are not derived from their optimization problems but embedded in a plug-and-play manner and lead to sub-optimal objective function value. More seriously, the hyper-parameters, introduced by the additive balance of these two coupled sub-tasks, are hard to determine in unsupervised learning scenarios and lead to inconsistent and less satisfying results. To avoid the problems mentioned above, we present a novel parameter-free MKC method with the trace ratio criterion (TRMKC in short), which minimizes the approximation errors between consensus and base kernels using the corr-entropy induced metric and maximizes the mean similarities based on the consensus kernel. The trade-off between these two coupled sub-procedures can be automatically balanced, and the performance could be mutually reinforced. To solve the trace ratio criterion and the corr-entropy induced non-quadratic function optimization problem, we present an alternative strategy with monotonic convergence proof, which reformulates it into a series of sub-problems with trace difference and quadratic programming by utilizing the half-quadratic optimization technique. Extensive MKC experimental results well demonstrate the effectiveness of TRMKC.

**Keywords:** Multiple kernel clustering · Corr-entropy induced kernel weighting · Trace ratio criterion

---

This work was supported in part by the National Key Research and Development Program of China (2018YFB0704301-1), the National Natural Science Foundation of China (61972268, 61976129, 62176001), the Sichuan Science and Technology Program (2020YFG0034).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022  
A. Bhattacharya et al. (Eds.): DASFAA 2022, LNCS 13246, pp. 681–688, 2022.  
[https://doi.org/10.1007/978-3-031-00126-0\\_52](https://doi.org/10.1007/978-3-031-00126-0_52)

## 1 Introduction

As a fundamental research field in machine learning and data mining, clustering has been widely used in various applications. The Multiple Kernel Clustering (MKC) methods are widely investigated to incorporate complementary information across kernels and avoid the kernel selection or design problem. The paradigm of MKC has received considerable attentions, and several methods have been proposed recently. In general, existing MKC algorithms can be roughly grouped into three categories. The first category takes the early fusion strategy, where one sub-task is to learn the consensus kernel from multiple candidate kernels, and another is to perform clustering on the single consensus kernel [5, 7, 8, 10]. The second category algorithms also take the early fusion strategy via the paradigm of multiple graph clustering. One sub-task is to extract multiple affinity graphs from multiple kernels, and another one is to integrate these affinity graphs to get the final consensus graph. These two sub-tasks can be concatenated separately as in [6, 15] or optimized jointly as in [14, 16]. The third category utilizes the late fusion strategy, where multiple base partitions are first generated from individual kernels and then integrated to build the final consensus partition [9, 11, 17].

Based on the inherent connection between consensus kernel learning and clustering on consensus kernel, most existing MKC algorithms jointly optimize these procedures within a unified learning framework, where these targets are often manually balanced by additional hyper-parameters. These algorithms often perform well with suitable parameters detected by the grid search strategy, which uses the ground truth to determine the proper parameters. However, the limitation of such a strategy is also clear. On the one hand, the performance of MKC algorithms is less stable and largely dependent on the choice of hyper-parameters. On the other hand, it is less applicable and even infeasible to search for the best hyper-parameters via the so-called grid search strategy for the unsupervised MKC scenario, where no label information is available. Therefore, it is much preferable to develop MKC methods without parameter tuning or even parameter-free.

To address the problems mentioned above on candidate kernel weighting and hyper-parameter sensitivity, we propose to learn the optimal neighborhood kernel called TRMKC by minimizing the approximation errors between consensus and base kernels using the corr-entropy induced metric, where the large errors caused by poor quality kernels could be largely suppressed. In contrast, good kernels corresponding to small errors could be carefully emphasized. It should be noted that the kernel weight can be estimated directly from the optimization problem, not like the plug-and-play strategy as [13, 19]. TRMKC is proposed to integrate consensus kernel learning and clustering by maximizing the trace ratio criterion without introducing additional hyper-parameters, where the balance between these two coupled sub-procedures can be automatically balanced, and the performance could be mutually reinforced. To solve the optimization problem with the trace ratio criterion and the corr-entropy induced non-quadratic function, we present an alternative optimization strategy with monotonic convergence proof, which reformulates it into a series of sub-problems with trace difference and quadratic programming by utilizing the half-quadratic optimization techniques. Extensive

MKC experimental results on 11 benchmark datasets with ten recent MKC algorithms well demonstrate the effectiveness of TRMKC.

## 2 Design of TRMKC

Given multiple kernel grammatrices  $\{\mathbf{K}^i\}_{i=1}^m$  on  $n$  samples, where  $\mathbf{K}^i \in \mathcal{R}^{n \times n}$  is the  $i$ -th kernel matrix and it is PSD as required. The MKC algorithm aims to identify the  $c$  consensus cluster structure across these candidate kernels.

### 2.1 Consensus Kernel Learning via CIM

Instead of using a traditional quadratic function like [13, 19], the approximation error between consensus kernel  $\mathbf{K}$  and single individual kernel  $\mathbf{K}^i$  is measured by the corr-entropy induced non-quadratic loss function, which leads to the following optimization problem for consensus kernel learning,

$$\min_{\mathbf{K}} \sum_{i=1}^m \ell(\mathbf{K}, \mathbf{K}^i) = \sum_{i=1}^m \left( 1 - \exp\left(-\frac{\|\mathbf{K} - \mathbf{K}^i\|^2}{\delta^2}\right) \right), \quad \text{s.t. } \mathbf{K} \succeq 0, \quad (1)$$

where  $\delta$  is the bandwidth of the Gaussian function, which is estimated automatically in this paper. From the perspective of the loss function, the benefits of the above non-quadratic loss function against the traditional quadratic one are two folds. Firstly, the CIM loss is upper bounded and change slowly for large errors. Therefore, the effect of a low-quality kernel with large error could be large suppressed. Secondly, the CIM loss changes quickly on small errors. As a result, similar high-quality kernels, which correspond to small errors, could be carefully distinguished. The different behavior of CIM function on large and small errors makes it is appropriate to characterize the expectation of consensus kernel across these candidate kernels.

From the perspective of kernel weighting, the implicit kernel weights can also be derived from the CIM loss function according to the half-quadratic optimization theory [1]. It assigns larger weights to good kernels with small errors and small weights to bad kernels with large errors. Moreover, we further introduce the explicit weight for each kernel by solving the following problem,

$$\min_{\mathbf{K}, \mathbf{s}} \sum_{i=1}^m \frac{1}{s_i} \left( 1 - \exp\left(-\frac{\|\mathbf{K} - \mathbf{K}^i\|^2}{\delta^2}\right) \right), \quad \text{s.t. } \mathbf{K} \succeq 0, \sum_{i=1}^m s_i = 1, s_i \geq 0, \forall i, \quad (2)$$

where  $\mathbf{s} \in \mathcal{R}^{m \times 1}$  and  $\frac{1}{s_i}$  can be seen as the explicit weight for the  $i$ -th kernel.

### 2.2 Clustering on Consensus Kernel

Spectral clustering and  $k$ -means are two major traditional clustering methods. Ratio cut aims to maximize the mean similarities between data points in the

same cluster while  $k$ -means aim to minimize the mean distances between points in the same cluster. Recently, it has been shown that both  $k$ -means and ratio cut can be unified into the following framework [12],

$$\max_{\mathbf{H}} \operatorname{tr}(\mathbf{H}^T \mathbf{A} \mathbf{H}), \quad \text{s.t. } \mathbf{H}^T \mathbf{H} = \mathbf{I}, \tag{3}$$

where  $\mathbf{H} \in \mathcal{R}^{n \times c}$  is the partition matrix and  $\mathbf{A} \in \mathcal{R}^{n \times n}$  is the similarity matrix. The mere difference between ratio-cut and  $k$ -means is that the former usually uses the Gaussian kernel to capture the pairwise similarity, while  $k$ -means adopts the linear kernel. Based on such a unified view of these two methods, the final clustering on consensus kernel can also be achieved by solving the optimization problem in Eq. (3), where we have  $\mathbf{A} = \mathbf{K}$ .

### 2.3 The Proposed TRMKC

We have two sub-tasks, i.e., minimizing Eq. (2) for consensus kernel learning and maximizing Eq. (3) for clustering on consensus kernel. Instead of introducing additional hyper-parameter to balance these targets, we take the trace ratio criterion to incorporate these two coupled procedures in a unified learning framework, which can be presented as follows,

$$\begin{aligned} \max_{\mathbf{s}, \mathbf{H}, \mathbf{K}} \quad & \frac{\operatorname{tr}(\mathbf{H}^T \mathbf{K} \mathbf{H})}{\sum_{i=1}^m \frac{1}{s_i} (1 - \exp(-\frac{\|\mathbf{K} - \mathbf{K}^i\|_F^2}{\delta^2}))} \\ \text{s.t.} \quad & \mathbf{H}^T \mathbf{H} = \mathbf{I}, \mathbf{K} \succeq 0, \sum_{i=1}^m s_i = 1, s_i \geq 0, \forall i. \end{aligned} \tag{4}$$

Eq. (4) learns the consensus kernel around individual kernels by minimizing the weighted corr-entropy induced non-quadratic loss in the denominator and performs clustering on the consensus kernel by maximizing the quadratic term in the nominator. The consensus kernel  $\mathbf{K}$  is involved both in the nominator and denominator. These two targets can be automatically balanced without the involvement of hyper-parameters. Based on these discussions, the optimization problem in Eq.(4) is suitable for MKC.

### 2.4 Optimization

We present an alternative learning algorithm to maximize Eq. (4).

**Update  $\mathbf{H}$  with fixed  $\mathbf{K}$**  Given  $\mathbf{s}$  and  $\mathbf{K}$ , the sub-problem with respect to  $\mathbf{H}$  can be simplified as

$$\max_{\mathbf{H}} \operatorname{tr}(\mathbf{H}^T \mathbf{K} \mathbf{H}), \quad \text{s.t. } \mathbf{H}^T \mathbf{H} = \mathbf{I}. \tag{5}$$

Since the consensus kernel  $\mathbf{K}$  is PSD, the optimal solution of  $\mathbf{H}$  can be obtained by the  $c$  eigenvectors corresponding to the  $c$  largest eigenvalues.

**Update  $\mathbf{s}$  with fixed  $\mathbf{K}$  and  $\mathbf{H}$**  Denote  $h_i = 1 - \exp(-\frac{\|\mathbf{K} - \mathbf{K}^i\|_F^2}{\delta^2})$ , the rest problem w.r.t.  $\mathbf{H}$  can be rewritten as

$$\min_{\mathbf{s}} \sum_{i=1}^m \frac{h_i}{s_i}, \quad \text{s.t.} \sum_{i=1}^m s_i = 1, s_i \geq 0, \forall i. \quad (6)$$

The optimal solution can be computed by  $s_i = \frac{\sqrt{h_i}}{\sum_{i'=1}^m \sqrt{h_{i'}}$ .

**Update  $\mathbf{K}$  with fixed  $\mathbf{H}$  and  $\mathbf{s}$**  Given  $\mathbf{H}$  and  $\mathbf{s}$ , the problem w.r.t.  $\mathbf{K}$  can be written as,

$$\max_{\mathbf{K}} \frac{\text{tr}(\mathbf{H}^T \mathbf{K} \mathbf{H})}{\sum_{i=1}^m \frac{1}{s_i} (1 - \exp(-\frac{\|\mathbf{K} - \mathbf{K}^i\|_F^2}{\delta^2}))}, \quad \text{s.t.} \quad \mathbf{K} \succeq 0. \quad (7)$$

The trace ratio problem in Eq. (7) can be solved by a series of trace difference problems [3]. Given  $\mathbf{H}$ ,  $\mathbf{s}$  and  $\mathbf{K}^{t-1}$  at the  $t-1$ -th iteration, we first introduce the following auxiliary variable at the  $t$ -th iteration,

$$\lambda^t = \frac{\text{tr}(\mathbf{H}^T \mathbf{K}^{t-1} \mathbf{H})}{\sum_{i=1}^m \frac{1}{s_i} (1 - \exp(-\frac{\|\mathbf{K}^{t-1} - \mathbf{K}^i\|_F^2}{\delta^2}))}, \quad (8)$$

where the index of  $(t-1)$  on  $\mathbf{H}$ ,  $\mathbf{s}$  is omitted. Equation (7) can be reformulated into the following trace difference problem,

$$\max_{\mathbf{K}} \text{tr}(\mathbf{H}^T \mathbf{K} \mathbf{H}) - \lambda^t \sum_{i=1}^m \frac{1}{s_i} (1 - \exp(-\frac{\|\mathbf{K} - \mathbf{K}^i\|_F^2}{\delta^2})), \quad \text{s.t.} \quad \mathbf{K} \succeq 0. \quad (9)$$

According to the HQ theory, we introduce the auxiliary variable  $\mathbf{g} \in \mathcal{R}^{m \times 1}$  with  $g_i = \exp(-\frac{\|\mathbf{K}^{t-1} - \mathbf{K}^i\|_F^2}{\delta^2})$ . Then Eq. (9) can be rewritten as,

$$\min_{\mathbf{K}} \lambda^t \sum_{i=1}^m \frac{g_i}{s_i} \|\mathbf{K} - \mathbf{K}^i\|^2 - \text{tr}(\mathbf{K}^T (\mathbf{H} \mathbf{H}^T)), \quad \text{s.t.} \quad \mathbf{K} \succeq 0. \quad (10)$$

By introducing the auxiliary variable  $\hat{\mathbf{K}} = \frac{1}{\sum_{i'=1}^m \frac{g_{i'}}{s_{i'}}} \sum_{i=1}^m \frac{g_i}{s_i} \mathbf{K}^i + \frac{1}{2\lambda^t \sum_{i'=1}^m \frac{g_{i'}}{s_{i'}}} \mathbf{H} \mathbf{H}^T$ , the problem in Eq. (10) can be simplified as,

$$\min_{\mathbf{K}} \|\mathbf{K} - \hat{\mathbf{K}}\|^2, \quad \text{s.t.} \quad \mathbf{K} \succeq 0. \quad (11)$$

The above problem is the quadratic projection of  $\hat{\mathbf{K}}$  within the positive semi-definite space. The optimal solution can be obtained by setting

$$\mathbf{K} = \mathbf{U}_{\hat{\mathbf{K}}} \mathbf{S}_{\hat{\mathbf{K}}}^+ \mathbf{V}_{\hat{\mathbf{K}}}^T, \quad (12)$$

where  $\hat{\mathbf{K}} = \mathbf{U} \mathbf{S} \mathbf{V}^T$  is the singular value decomposition (SVD) of  $\hat{\mathbf{K}}$  and  $\mathbf{S}^+$  is the non-negative projection of  $\mathbf{S}$ , i.e.,  $\mathbf{S}^+ = \max(\mathbf{S}, \mathbf{0})$ .

### 3 Experiments

We compare TRMKC<sup>1</sup> with 10 MKC algorithms to show the effectiveness and provide the convergence analysis of TRMKC on benchmark datasets.

#### 3.1 Experiment Setup

We use 11 widely used data sets for multiple kernel clustering comparison [2, 4]. We omit the details of these data sets due to the space limitation. We compare TRMKC with 10 recent developed MKC methods, including RMSC [18], RMKKM [2], MKCMR [8], LKAM [5], ONKC [10], JMKSC [19], MKCSS [20], ONALK [7], SPMKC [13], OPLF [9]. The codes are mostly provided by the authors in public repository or emails. We adopt clustering accuracy (ACC) to evaluate the clustering result.

**Table 1.** ACC(%) of 11 algorithms on 11 benchmark datasets

	RMSC	RMKKM	MKCMR	LKAM	ONKC	JMKSC	MKCSS	ONALK	SPMKC	OPLF	TRMKC
JAFFE	67.14	79.34	95.77	88.26	97.18	82.16	38.03	93.90	96.24	96.71	<b>98.59</b>
CSTR	42.74	70.11	64.84	65.26	65.26	65.26	62.53	45.89	65.89	65.05	<b>80.42</b>
TR45	37.83	61.45	69.86	67.10	72.03	68.99	65.07	25.94	74.20	68.70	<b>76.09</b>
TR41	53.08	55.47	64.92	49.43	61.50	60.82	53.19	45.44	57.18	59.34	<b>68.00</b>
MPeg7	46.29	49.00	53.21	47.86	53.93	49.36	52.50	48.21	55.71	52.29	<b>56.64</b>
BA	33.69	43.45	45.80	41.24	44.09	29.06	45.44	46.23	34.54	37.11	<b>52.28</b>
COIL	58.75	61.88	64.51	60.21	65.90	58.96	71.67	67.64	64.10	56.04	<b>78.96</b>
Wap	36.22	44.62	44.10	41.35	46.35	36.47	35.26	32.95	47.88	43.91	<b>59.81</b>
Digit	59.43	78.91	83.14	76.74	83.08	65.33	87.76	87.65	78.02	80.30	<b>88.48</b>
Palm	56.90	70.55	73.50	59.05	78.80	76.70	74.80	71.05	80.80	78.05	<b>88.15</b>
K1B	66.84	75.73	64.10	58.08	67.48	78.29	84.57	63.33	63.03	72.44	<b>84.79</b>
Avg	50.81	62.77	65.80	59.51	66.87	61.04	60.98	57.11	65.24	64.54	<b>75.65</b>

We follow the same strategy in [2, 19] to build 12 base kernels to evaluate the clustering performance of the MKL methods. There are some parameters should be set in advance for fair. So, it is non-trivial to choose the proper multiple hyper-parameters corresponding to the best clustering performance for LKAM, ONKC, JMKSC, MKCSS, ONALK, and SPMKC across different datasets. As a result, some hyper-parameters of these algorithms are determined according to the suggestions from their papers, while others are determined by observing the corresponding sensitivity study results and choosing the stable value.

The concrete setting of hyper-parameters for all these algorithms are presented as follows. For all these MKC algorithms, the number of clusters is set to the true number of classes for all the data sets. For RMSC, the regularization parameter is set to  $\lambda = 0.005$  as in [18]. For RMKKM, the parameter  $\gamma$  to control the kernel weight distribution is set to  $\gamma = 0.3$  as in [2]. For MKCMR [8],

<sup>1</sup> The code of TRMKC has been released at <https://github.com/YanChenSCU/TRMKC-DASFAA-2022>.

the regularization parameter is fixed to  $\lambda = 2^{-2}$ . For LKAM [5], the regularization parameter and the neighborhood size are set to  $\lambda = 2^{-1}$  and  $\tau = 0.05n$ . For ONKC [10], the importance of the neighborhood kernel learning and the kernel diversity balancing term are set to  $\rho = 2^{-4}$  and  $2^{-7}$ . For JMKSC, the hyper-parameters are set to  $\alpha = 10^{-1}, \beta = 20, \gamma = 5$  as suggested in [19]. For MKCSS [20], the constrained rank value, the importance of the Frobenius term, the kernel diversity balancing term and the number of neighbors are fixed as  $0.1n, 10^{-4} \|\mathbf{K}_{\text{avg}}\|_F, 2^{-2}$  and  $0.01n$ . For ONALK [7], the kernel diversity balancing term, the importance of the neighborhood kernel learning and the threshold of neighborhood similarity are set to  $\lambda = 2, \rho = 2^{-1}, \tau = 0.1$ . For SPMKC [13], the hyper-parameters are set to  $\lambda_1 = 4, \lambda_2 = 1, \lambda_3 = 100, \lambda_4 = 1000$ . For OPLF, it is free of hyper-parameters [9]. For TRMKC, we use the neighborhood kernel as [20] and the local size is fixed to  $5 \log(n)$  on all the datasets without tuning. For all these methods except for OPLF, we run k-means 10 times and obtain the final result corresponding to the minimal value of k-means objective. The variance for all the compared algorithms is 0 in our experiments.

### 3.2 Clustering Results Analysis

Based on the clustering result in Table 1, we can see that

- TRMKC outperforms all these 10 MKC algorithms on 11 benchmark data sets in most cases. We can see that our method achieves 13.59% improvement against the second-best result on averages in terms of ACC.
- The results for MKCMR, LKAM, ONKC, JMKSC, MKCSS, ONALK and SPMKC degenerate in general and exhibit poor performance in many cases. These algorithms would achieve better results by tuning the parameters according to the grid-search strategy. However, it is infeasible in practice. Since we run these algorithms with fixed hyper-parameters, the ubiquitous degeneration indeed indicates that these algorithms are less stable without careful tuning.
- Considering the difficulty of hyper-parameter determination in the MKC scenario and the degeneration of existing MKC algorithms on fixed parameters, these experimental results motivate us to develop a parameter-free method for the MKC task. Compared with baseline methods, the results of TRMKC demonstrate not only the superiority in performance but also the stableness across all data sets.

## 4 Conclusion

In this paper, we propose a novel parameter-free MKC method with the trace ratio criterion (TRMKC), which minimizes the approximation errors for optimal kernel learning and maximizes the mean similarities on the consensus one. An alternative algorithm has been developed to solve the optimization problem with trace ratio criterion and the non-quadratic CIM loss function. Comprehensive

experimental results on 11 widely-used datasets show that exiting MKC methods with hyper-parameters are indeed less stable and exhibit poor results without proper parameter determination. Moreover, TRMKC consistently outperforms the state-of-the-art competitors in terms of clustering performance.

## References

1. Du, L., Li, X., Shen, Y.D.: Robust nonnegative matrix factorization via half-quadratic minimization. In: ICDM, pp. 201–210 (2012)
2. Du, L., et al.: Robust multiple kernel k-means using l21-norm. In: IJCAI, pp. 3476–3482 (2015)
3. Jia, Y., Nie, F., Zhang, C.: Trace ratio problem revisited. *IEEE Trans. Neural Netw.* **20**(4), 729–735 (2009)
4. Kang, Z., Peng, C., Cheng, Q., Xu, Z.: Unified spectral clustering with optimal graph. In: AAAI, pp. 3366–3373 (2018)
5. Li, M., Liu, X., Wang, L., Dou, Y., Yin, J., Zhu, E.: Multiple kernel clustering with local kernel alignment maximization. In: IJCAI, pp. 1704–1710 (2016)
6. Li, X., Ren, Z., Lei, H., Huang, Y., Sun, Q.: Multiple kernel clustering with pure graph learning scheme. *Neurocomputing* **424**, 215–225 (2021)
7. Liu, J., et al.: Optimal neighborhood multiple kernel clustering with adaptive local kernels. In: TKDE, p. 1 (2021)
8. Liu, X., Dou, Y., Yin, J., Wang, L., Zhu, E.: Multiple kernel k-means clustering with matrix-induced regularization. In: AAAI, vol. 30 (2016)
9. Liu, X., et al.: One pass late fusion multi-view clustering. In: ICML, pp. 6850–6859 (2021)
10. Liu, X., et al.: Optimal neighborhood kernel clustering with multiple kernels. In: AAAI, pp. 2266–2272 (2017)
11. Liu, X., et al.: Late fusion incomplete multi-view clustering. *TPAMI* **41**(10), 2410–2423 (2019)
12. Pei, S., Nie, F., Wang, R., Li, X.: Efficient clustering based on a unified view of k-means and ratio-cut. In: NeurIPS (2020)
13. Ren, Z., Sun, Q.: Simultaneous global and local graph structure preserving for multiple kernel clustering. *TNNLS* **32**(5), 1839–1851 (2021)
14. Ren, Z., Sun, Q., Wei, D.: Multiple kernel clustering with kernel k-means coupled graph tensor learning. In: AAAI, pp. 9411–9418 (2021)
15. Ren, Z., Yang, S.X., Sun, Q., Wang, T.: Consensus affinity graph learning for multiple kernel clustering. *IEEE Trans. Cybern.* **51**(6), 3273–3284 (2020)
16. Sun, M., et al.: Projective multiple kernel subspace clustering. *IEEE Trans. Multim.* **1** (2021)
17. Wang, S., et al.: Multi-view clustering via late fusion alignment maximization. In: IJCAI, pp. 3778–3784 (2019)
18. Xia, R., Pan, Y., Du, L., Yin, J.: Robust multi-view spectral clustering via low-rank and sparse decomposition. In: AAAI, vol. 28 (2014)
19. Yang, C., Ren, Z., Sun, Q., Wu, M., Yin, M., Sun, Y.: Joint correntropy metric weighting and block diagonal regularizer for robust multiple kernel subspace clustering. *Inf. Sci.* **500**, 48–66 (2019)
20. Zhou, S., et al.: Multiple kernel clustering with neighbor-kernel subspace segmentation. *TNNLS* **31**(4), 1351–1362 (2020)