Sample Adaptive Localized Simple Multiple Kernel K-means

Xinwang Liu, *Senior Member, IEEE*, Li Liu, *Senior Member, IEEE*, Limin Peng, Yi Zhang, and Dewen Hu

Abstract—By optimally exploiting a set of pre-calculated base kernel matrices, multiple kernel clustering (MKC) aims to improve performance of clustering tasks. Different from existing MKC algorithms, simple multiple kernel k-means (SimpleMKKM) proposes a novel minimization-maximization learning paradigm, and makes remarkable achievements in some applications. However, we observe that SimpleMKKM does not considerably make use of the variation among samples, resulting in unsatisfying clustering performance. To address this issue, we propose a local kernel alignment criterion with the aim to better capture the variation among samples. It makes the clustering algorithms focus on reliable pairwise samples that shall stay together and cut off unreliable farther pairwise ones. Based on this criterion, we first propose a unified weighted localized SimpleMKKM (UWL-SMKKM), and theoretically uncover that SimpleMKKM is a special case of UWL-SMKKM. According to this connection, one can readily implement the proposed UWL-SMKKM with existing SimpleMKKM packages. We further improve UWL-SMKKM by proposing sample adaptive localized SimpleMKKM (SAL-SMKKM) where the weight of the local alignment for each sample can be adaptively adjusted, leading to an intractable tri-level minimization-minimization-maximization. To solve it, we reformulate it as a minimization of an optimal function which is a minimization-maximization optimization, prove its differentiability, and design a reduced gradient descent optimization to decrease it. We empirically evaluate the clustering performance of the proposed UWL-SMKKM and SAL-SMKKM on several widely used benchmark datasets. The experimental results have clearly indicated that our algorithms consistently outperform state-of-the-art ones. Finally, we apply UWL-SMKKM to the multi-modal parcellation of human cerebral cortex, which is essential and helpful to understand brain organization and function. As seen, UWL-SMKKM achieves accurate parcellation in an automatic and objective manner without any manual intervention, which once again demonstrates its validity and effectiveness in practical applications. The codes of UWL-SMKKM and SAL-SMKKM are publicly accessed at: https://github.com/xinwangliu/LocalizedSMKKM.

Index Terms—k-means, multiple kernel clustering, multi-view clustering

1 Introduction

As an elegant learning framework, multiple kernel clustering (MKC) maximally utilizes multiple sources information to partition samples into different clusters [1], [2], [3]. Specifically, MKC learns an optimal kernel matrix from a set of pre-calculated base kernel matrix to best serve for clustering tasks [4], [5], [6], [7], [8], [9], [10], [11], [12], [13]. It has attracted extensive study [14], [14], [15], [16] and been widely used in various practical applications such as image segmentation [17], [18], anomaly detection [19], cancer biology [20], to name just a few. Many pioneering works on MKC have been recently proposed in the literature. In [21], a regularization term is induced and incorporated into MKKM to enhance the diversity and decrease the redundancy of the selected kernel matrices. The work in [14] develops a local kernel alignment criterion with the aim to sufficiently consider the variation among sample, which is considered to be helpful in improving the clustering performance. Different from existing assumption that an optimal kernel is composed of the combination of base kernels, the optimal neighborhood MKC algorithm is designed in [15], where the optimal kernel resides in the neighbor of the combined kernels. By this way, the representability of the learned optimal kernel is enhanced, leading to improved clustering performance.

Different from the aforementioned kernel based MKC, late fusion based MKC has been more recently proposed [16]. It seeks to achieve a consensus partition matrix from a group of base partition ones, which are generated from base kernel matrices. To fulfil this goal, the work in [16] maximally aligns multiple base partitions with a consensus one, which could significantly reduce the computational complexity and improve the clustering performance. Along this direction, [22] proposes an effective and efficient late fusion MKC algorithm to address incomplete multi-view clustering.

A representative MKC algorithm, termed simple multiple kernel k-means (SimpleMKKM), is recently put forward [23]. Different from existing optimization strategy where the kernel coefficients and clustering partition matrix are jointly minimized, SimpleMKKM proposes a novel min-max optimization paradigm for MKC. Specifically, its objective is to minimize the kernel coefficients and maximize the clustering partition matrix. This results in a min-max optimization problem, which is much more intractable than existing optimization. To solve it, the work in [23] equivalently rewrites the min-max optimization as a minimization of an optimal function, proves its differentiablity, and design a reduced gradient descent algorithm to solve the resultant minimization. Further, an ablation study is design to show

- *X. Liu and Y. Zhang are with College of Computer, National University of Defense Technology, Changsha, 410073, China (E-mail: xinwangliu@nudt.edu.cn).*
- *L. Liu is with College of System Engineering, National University of Defense Technology, Changsha, 410073, China (E-mail: li.liu@oulu.fi).*
- *L. Peng and D. Hu are with College of Mechatronics and Automation, National University of Defense Technology, Changsha, China.*

Manuscript received April 10, 2022.

that both the novel min-max formulation and new optimization technique lead to the improved clustering performance.

Despite the aforementioned advantages, SimleMKKM maximally aligns the combined kernel matrix with a similarity matrix generated by the clustering partition matrix in a "global" manner. As seen, this alignment is not able to sufficiently take the variation among samples into consideration and ignores local structures among samples, result in unsatisfying clustering performance. In this paper, we design a local kernel alignment criterion to overcome this issue. To fulfil this goal, we firstly calculate the neighborhoods for each sample, which defines a local patch in the combined kernel and ideal similarity matrix, respectively. The proposed local kernel alignment criterion only requires these two patches being maximally aligned. As seen, this local criterion helps clustering algorithms to put more emphasis on closer pairwise samples that shall stay together, and takes little consideration of farther pairwise samples in kernel alignment. Consequently, our criterion is able to better utilize the variation among samples, which could be helpful to bring forth improved clustering performance.

Following the minimization-maximization optimization of SimpleMKKM, we firstly derive the objective of UWL-SMKKM. We then theoretically uncover that SimpleMKKM is a special case of UWL-SMKKM. Base on this connection, we point out that UWL-SMKKM can be readily implemented with SimpleMKKM packages after simply normalizing each base kernel. In addition, we further improve the localized SimpleMKKM by allowing the weight of local kernel alignment for each sample being adaptively optimized, leading to a more intractable tri-level optimization. We firstly reformulate the resultant problem as a minimization of an optimal value function, provide an theorem to guarantee its differentiability, and design a reduced gradient descent algorithm to optimize it. We conduct comprehensive experiments to evaluate the clustering performance of the proposed algorithms on several benchmark datasets, and the results well validate their effectiveness. Finally, in order to test the performance of our localized SimpleMKKM in practical applications, we apply it into the multi-modal parcellation of human cerebral cortex, which is essential for understanding brain organization and function. The experimental results show that our algorithm is able to achieve accurate parcellation in an automatic and objective manner without manual intervention, well indicating its superiority.

This paper bears the following main contributions.

- We, for the first time, point out that the newly developed SimpleMKKM does not well consider the variation among samples, and design a local kernel alignment criterion to overcome this issue.
- We build the theoretical connection between SimpleMKKM and the proposed UWL-SMKKM, and reveal that the former is a special case of ours.
- We further improve the localized SimpleMKKM by proposing a new variant, where the weights of local kernel alignment for each sample are allowed to be adaptively adjusted. Further, we develop a new algorithm with proved convergence to solve the resultant optimization.
- We conduct comprehensive experimental study on several benchmark datasets and the multi-modal parcellation of human cerebral cortex. As indicated, the experimental results have demonstrated that our algorithms consistently outperforms the state-of-the-art competitors, verifying its effectiveness.

Finally, we discuss this work with our previous conference paper [24]. This work substantially extends our original version from the following aspects: (1) Instead of combining the local kernel alignments for each sample with equal weights, we design a new algorithm, termed Sample Adaptive Localized SimpleMKKM, by allowing the weight of each local kernel alignment to be adaptive adjusted. (2) We then develop a new optimization with convergence to decrease the resultant optimization problem. Moreover, the newly proposed variant significantly outperforms Localized SimpleMKKM in the previous paper [24]. (3) More comprehensive experimental research has been conducted from perspectives of overall clustering performance, the evolution of learned clustering partition matrix, the learned kernel weights, the learned weights of each local kernel alignment, and the convergence of the developed solving algorithms. These results have well validated the effectiveness of the proposed algorithms. In addition, we have applied the proposed algorithm into multi-modal parcellation of human cerebral cortex, and the results demonstrate its superiority.

2 SimpleMKKM: Simple Multiple Kernel k-means

In this section, we briefly introduce SimpleMKKM [23], which is closely related to our work.

In literature, multiple kernel k-means (MKKM) [25] has been widely used to solve MKC. It assumes that the optimal kernel $\mathbf{K}_{\boldsymbol{\gamma}}$ is composed of a group of pre-specified base kernel matrices, and jointly optimizes $\boldsymbol{\gamma}$ and the clustering partition matrix $\mathbf{H}$, leading to the following objective function,

$$\min_{\boldsymbol{\gamma}\in\Delta}\min_{\mathbf{H}\in\Gamma}\ \mathrm{Tr}(\mathbf{K}_{\boldsymbol{\gamma}}(\mathbf{I}-\mathbf{H}\mathbf{H}^\top)) \tag{1}$$

where $\mathbf{K}_{\boldsymbol{\gamma}}=\sum_{p=1}^m\gamma_p^2\mathbf{K}_p$, $\Delta=\{\boldsymbol{\gamma}\in\mathbb{R}^m\mid\sum_{p=1}^m\gamma_p=1,\ \gamma_p\geq 0,\ \forall p\}$, $\Gamma=\{\mathbf{H}|\mathbf{H}\in\mathbb{R}^{n\times k},\ \mathbf{H}^\top\mathbf{H}=\mathbf{I}_k\}$, and $\mathbf{I}_k$ is an identity matrix with size k.

A two-step alternate optimization strategy is developed to solve $\boldsymbol{\gamma}$ and $\mathbf{H}$ in Eq. (1). That is, one variable is optimized with the other fixed. With fixed $\boldsymbol{\gamma}$, the optimization in Eq. (1) w.r.t. $\mathbf{H}$ reduces to traditional kernel k-means, which can be readily solved by existing off-the-shelf optimization packages. With fixed $\mathbf{H}$, the optimization in Eq. (1) w.r.t. $\boldsymbol{\gamma}$ is equivalent to

$$\min_{\boldsymbol{\gamma}\in\Delta}\ \sum_{p=1}^m\gamma_p^2\mathrm{Tr}\left(\mathbf{K}_p(\mathbf{I}_n-\mathbf{H}\mathbf{H}^\top)\right), \tag{2}$$

which can be analytically obtained. These two-step in optimizing $\mathbf{H}$ and $\boldsymbol{\gamma}$ are alternately performed until convergence.

MKKM has been intensively studied and widely adopted due to its conceptual simplification and easy-to-implementation. However, it is empirically noticed that MKKM could not be able to gain superior performance in some applications, sometimes or even worse than the

unified weighted kernel k-means. Novel clustering models are needed to alleviate this situation.

Different from $\min_{\boldsymbol{\gamma}} \min_{\mathbf{H}}$ used in existing MKKM, the newly proposed SimpleMKKM [23] adopts the following $\min_{\boldsymbol{\gamma}} \max_{\mathbf{H}}$ optimization paradigm,

$$\min_{\boldsymbol{\gamma}\in\Delta} \max_{\mathbf{H}\in\Gamma} \operatorname{Tr}\left(\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{H}\mathbf{H}^\top\right), \tag{3}$$

which optimizes $\boldsymbol{\gamma}$ and $\mathbf{H}$ to make the objective minimized and maximized, respectively. Correspondingly, the aforementioned strategy optimizing one variable with the other fixed cannot guarantee that the objective is monotonically changed. As a result, the widely used alternate optimization cannot be applied to solve the new minimization-maximization formulation in Eq. (3).

In [23], SimpleMKKM develops a reduced gradient descent algorithm to solve Eq. (3). Specifically, it firstly transforms the $\min_{\boldsymbol{\gamma}} \max_{\mathbf{H}}$ into a minimization w.r.t $\boldsymbol{\gamma}$, and proves the differentiability of the resultant optimal value function. Specifically, we equivalently rewrite Eq. (3) as,

$$\min_{\boldsymbol{\gamma}\in\Delta} \mathcal{J}(\boldsymbol{\gamma}), \tag{4}$$

with

$$\mathcal{J}(\boldsymbol{\gamma}) = \left\{\max_{\mathbf{H}\in\Gamma} \operatorname{Tr}\left(\mathbf{H}^\top\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{H}\right)\right\}. \tag{5}$$

In this manner, the formulation in Eq. (3) is reduced a minimization one, whose objective $\mathcal{J}(\boldsymbol{\gamma})$ is a kernel k-means optimal value function.

To solve Eq. (4) with gradient descent, SimpleMKKM firstly prove the differentiability of $\mathcal{J}(\boldsymbol{\gamma})$ in Eq. (4) w.r.t. $\boldsymbol{\gamma}$, and computes its gradient as $\frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_p} = 2\gamma_p \operatorname{Tr}\left(\mathbf{H}^{*\top}\mathbf{K}_p\mathbf{H}^*\right)$, where $\mathbf{H}^* = \left\{\arg\max_{\mathbf{H}\in\Gamma} \operatorname{Tr}\left(\mathbf{H}^\top\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{H}\right)\right\}$. In Eq. 4, there are equality and positivity constraints on $\boldsymbol{\gamma}$. To keep the obtained solution at each iteration satisfying such constraints, SimpleMKKM searches a descent direction which guarantees the equality and non-negativity constraints on $\boldsymbol{\gamma}$. To do so, SimpleMKKM calculates the following reduced gradient to make the equality constraint satisfied. Let γ_u be a non-zero component of $\boldsymbol{\gamma}$ and $\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})$ denote the reduced gradient of $\mathcal{J}(\boldsymbol{\gamma})$. The p-th $(1 \le p \le m)$ component of $\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})$ is

$$\left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_p = \frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_p} - \frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_u} \quad \forall\, p \neq u, \tag{6}$$

and

$$\left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_u = \sum\nolimits_{p=1,p\neq u}^{m} \left(\frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_u} - \frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_p}\right), \tag{7}$$

where u is set as the index corresponding to the largest component of vector $\boldsymbol{\gamma}$ which is considered to provide better numerical stability [23], [23], [26].

The non-negativity constraint on $\boldsymbol{\gamma}$ is then considered. If there is an index p such that $\gamma_p = 0$ and $\left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_p > 0$, updating $\boldsymbol{\gamma}$ along this direction would make the non-negativity constraint violated. In this case, one should set the descent direction for that component as 0. Together the aforementioned consideration of equality and non-negativity constraints, SimpleMKKM gives the following descent direction to update $\boldsymbol{\gamma}$ as

$$d_p = \begin{cases} 0 & \text{if } \gamma_p = 0 \text{ and } \left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_p > 0 \\ -\left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_p & \text{if } \gamma_p > 0 \text{ and } p \neq u \\ -\left[\bigtriangledown\mathcal{J}(\boldsymbol{\gamma})\right]_u & \text{if } p = u. \end{cases} \tag{8}$$

In sum, $\boldsymbol{\gamma}$ can be updated via $\boldsymbol{\gamma} \leftarrow \boldsymbol{\gamma} + \alpha\mathbf{d}$, where $\mathbf{d} = [d_1, \cdots, d_m]^\top$ is a descent direction computed by Eq. (8), and α is the optimal step size. The whole procedure in solving Eq. (3) is presented in Algorithm 1.

Algorithm 1 SimpleMKKM

Input: $\{\mathbf{K}_p\}_{p=1}^m$, k, $t = 1$.
Initialize $\boldsymbol{\gamma}^{(1)} = \mathbf{1}/m$, flag $= 1$.
while flag **do**
  calculate $\mathbf{H}^{(t)}$ via a kernel k-means with $\mathbf{K}_{\boldsymbol{\gamma}^{(t)}}$.
  calculate $\frac{\partial\mathcal{J}(\boldsymbol{\gamma})}{\partial\gamma_p}$ $(\forall p)$ and the descent direction $\mathbf{d}^{(t)}$ in Eq. (8).
  $\boldsymbol{\gamma}^{(t+1)} \leftarrow \boldsymbol{\gamma}^{(t)} + \alpha\mathbf{d}^{(t)}$.
  if $\max|\boldsymbol{\gamma}^{(t+1)} - \boldsymbol{\gamma}^{(t)}| \le 1e-4$ **then**
    flag=0.
  end if
  $t \leftarrow t + 1$.
end while

In addition, the work in [23] theoretically shows that $\mathcal{J}(\boldsymbol{\gamma})$ in Eq. (4) is a convex function w.r.t $\boldsymbol{\gamma}$, which guarantees the global optimum of Algorithm 1.

3 Sample Adaptive Localized SimpleMKKM

3.1 UWL-SMKKM: Unified Weighted Localized SimpleMKKM

As seen from Eq. (3), SimpleMKKM adopts a minimization-maximization strategy to align $\mathbf{K}_{\boldsymbol{\gamma}}$ and $\mathbf{H}\mathbf{H}^\top$. This criterion requires each K_{ij} to be indiscriminately aligned with an "ideal" value $\mathbf{h}_i^\top\mathbf{h}_j$, where K_{ij} and $\mathbf{h}_i$ denote the (i,j)-th component of $\mathbf{K}_{\boldsymbol{\gamma}}$ and i-th row of $\mathbf{H}$. It would cause K_{ij}'s with large variation to be aligned with a same clustering label. To address this issue, we develop a more reasonable alignment method, which shall cast off the less reliable farther pairwise similarity and in the mean time concentrate more on consolidating these high confidence clustering predictions. To fulfill this goal, we propose to align $\mathbf{K}_{\boldsymbol{\gamma}}$ with $\mathbf{H}\mathbf{H}^\top$ in a local way.

To take the variation of samples into consideration, we firstly define the nearest neighbors for the i-th sample, denoted as $\mathbf{S}^{(i)} \in \{0,1\}^{n\times\operatorname{round}(\tau\times n)}$, where $\operatorname{round}(\cdot)$ is a rounding function and $0 < \tau \le 1$ controls the size of neighborhood. We propose to calculate the local alignment for the i-th sample in Eq. (9),

$$\left\langle \mathbf{S}^{(i)^\top}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{S}^{(i)}, \mathbf{S}^{(i)^\top}\mathbf{H}\mathbf{H}^\top\mathbf{S}^{(i)} \right\rangle_{\mathrm{F}}, \tag{9}$$

where $\mathbf{S}^{(i)^\top}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{S}^{(i)}$ and $\mathbf{S}^{(i)^\top}\mathbf{H}\mathbf{H}^\top\mathbf{S}^{(i)}$ denote extracting the most reliable elements from $\mathbf{K}_{\boldsymbol{\gamma}}$ and $\mathbf{H}\mathbf{H}^\top$ according to the nearest neighbors of the i-th sample, respectively.

As seen from Eq. (9), only the most reliable samples that shall stay together are involved in calculating the local

alignment, which avoids samples with large variations being aligned with a same clustering label. By summing over the local alignment of each sample in a *unified weighted* manner, we obtain the objective function of unified weighted localized SimpleMKKM (UWL-SMKKM) as follow:

$$\min_{\boldsymbol{\gamma}\in\Delta}\ \max_{\mathbf{H}\in\Gamma}\ \mathrm{Tr}\left(\mathbf{H}^\top\sum\nolimits_{i=1}^{n}\left(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{A}^{(i)}\right)\mathbf{H}\right), \tag{10}$$

where $\Delta=\{\boldsymbol{\gamma}\in\mathbb{R}^m|\sum_{p=1}^m\gamma_p=1,\,\gamma_p\geq 0,\,\forall p\}$, $\Gamma=\{\mathbf{H}\in\mathbb{R}^{n\times k}|\mathbf{H}^\top\mathbf{H}=\mathbf{I}_k\}$, $\mathbf{K}_{\boldsymbol{\gamma}}=\sum_{p=1}^m\gamma_p^2\mathbf{K}_p$ and $\mathbf{A}^{(i)}=\mathbf{S}^{(i)}\mathbf{S}^{(i)^\top}$.

The following Theorem 1 uncovers the theoretical connection between SimpleMKKM and the proposed UWL-SMKKM.

Theorem 1. *The objection of SimpleMKKM is a special case of Eq. (10).*

Proof. The objective function in Eq. (10) can be written as

$$\begin{aligned}
&\sum\nolimits_{i=1}^{n}\mathrm{Tr}\left(\mathbf{H}^\top(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{A}^{(i)})\mathbf{H}\right)\\
&=\sum\nolimits_{i=1}^{n}\left\langle\mathbf{A}^{(i)}\otimes\mathbf{K}_{\boldsymbol{\gamma}},\mathbf{A}^{(i)}\otimes(\mathbf{H}\mathbf{H}^\top)\right\rangle_{\mathrm{F}}\\
&=\sum\nolimits_{i=1}^{n}\left\langle\mathbf{A}^{(i)}\otimes\mathbf{K}_{\boldsymbol{\gamma}},\mathbf{H}\mathbf{H}^\top\right\rangle_{\mathrm{F}}\\
&=\left\langle\left(\sum\nolimits_{i=1}^{n}\mathbf{A}^{(i)}\right)\otimes\mathbf{K}_{\boldsymbol{\gamma}},\mathbf{H}\mathbf{H}^\top\right\rangle_{\mathrm{F}}\\
&=\sum\nolimits_{p=1}^{m}\gamma_p^2\left\langle\left(\sum\nolimits_{i=1}^{n}\mathbf{A}^{(i)}\right)\otimes\mathbf{K}_p,\mathbf{H}\mathbf{H}^\top\right\rangle_{\mathrm{F}}\\
&=\sum\nolimits_{p=1}^{m}\gamma_p^2\left\langle\tilde{\mathbf{K}}_p,\mathbf{H}\mathbf{H}^\top\right\rangle_{\mathrm{F}}\\
&=\mathrm{Tr}\left(\mathbf{H}^\top\tilde{\mathbf{K}}_{\boldsymbol{\gamma}}\mathbf{H}\right),
\end{aligned}\tag{11}$$

where $\otimes$ represents element-wise multiplication between two matrices, $\tilde{\mathbf{K}}_p=\left(\sum_{i=1}^n\mathbf{A}^{(i)}\right)\otimes\mathbf{K}_p$ is a normalization of $\mathbf{K}_p$, and $\tilde{\mathbf{K}}_{\boldsymbol{\gamma}}=\sum_{p=1}^m\gamma_p^2\tilde{\mathbf{K}}_p$. This indicates that the objective of SimpleMKKM is a special case of Eq. (10). This completes the proof. □

According to Theorem 1, it is not difficult to verify that our formulation in Eq. (10) is reduced to SimpleMKKM when we set all elements of $\mathbf{A}^{(i)}$ as one. In that case, each sample takes all the rest ones as its neighbors. This implies that SimpleMKKM is a special case of UWL-SMKKM. The following Theorem 2 shows that each $\tilde{\mathbf{K}}_p$ still keeps positive semidefinite (PSD) under the aforementioned normalization.

Theorem 2. *Each $\tilde{\mathbf{K}}_p\,(1\leq p\leq m)$ is PSD.*

Proof. We provide the detailed proof in the appendix due to page limit. □

We then discuss how to efficiently implement UWL-SMKKM via SimpleMKKM packages. Based on Theorem 1, we firstly normalize each $\mathbf{K}_p$ with $\sum_{i=1}^n\mathbf{A}^{(i)}$ to generate $\tilde{\mathbf{K}}_p$, which is kept PSD according to Theorem 2. After that, we take $\{\tilde{\mathbf{K}}_p\}_{p=1}^m$ as the input of SimpleMKKM 1, producing the solution of the proposed UWL-SMKKM.

3.2 Sample Adaptive Localized SimpleMKKM

3.2.1 The Formulation

As seen from Eq. (10), the local kernel alignment for each sample is taken over in an unified manner. That is, the weight for each sample in calculating local kernel alignment is equal. A more reasonable approach to integrating these local alignments is that their weights shall be adaptively tuned. By this way, one can expect to learn better weights, which could further improve clustering performance. To fulfil this goal, we propose the objective of sample adaptive localized SimpleMKKM (SAL-SMKKM) as follows,

$$\min_{\boldsymbol{\beta}\in\Phi}\ \min_{\boldsymbol{\gamma}\in\Delta}\ \max_{\mathbf{H}\in\Gamma}\ \mathrm{Tr}\left(\mathbf{H}^\top\sum\nolimits_{i=1}^{n}\beta_i(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{A}^{(i)})\mathbf{H}\right), \tag{12}$$

where $\beta_i\,(1\leq i\leq n)$, denoting the weight of local alignment for the i-th sample, is also jointly optimized, and $\Phi=\{\boldsymbol{\beta}\in\mathbb{R}^n|\sum_{i=1}^n\beta_i=1,\,\beta_i\geq 0,\,\forall i\}$.

3.2.2 The Optimization

Compared with UWL-SMKKM in Eq. (10), Eq. (12) provides a more flexible and effective way to locally align $\mathbf{K}_\gamma$ with $\mathbf{H}\mathbf{H}^\top$. However, the optimization in Eq. (12) is a tri-level optimization, which is much more intractable than the one in Eq. (10). To solve Eq. (12), we firstly rewrite it as a minimization of an optimal value function as follows,

$$\min_{\boldsymbol{\beta}\in\Phi}\ \mathcal{T}(\boldsymbol{\beta}) \tag{13}$$

with

$$\mathcal{T}(\boldsymbol{\beta})=\left\{\min_{\boldsymbol{\gamma}\in\Delta}\ \max_{\mathbf{H}\in\Gamma}\ \mathrm{Tr}\left(\mathbf{H}^\top\left(\sum\nolimits_{i=1}^{n}\beta_i\mathbf{A}^{(i)}\right)\otimes\mathbf{K}_{\boldsymbol{\gamma}}\right)\mathbf{H}\right)\right\}, \tag{14}$$

where $\mathrm{Tr}\Big(\mathbf{H}^\top\left(\sum_{i=1}^n\beta_i\mathbf{A}^{(i)}\right)\otimes\mathbf{K}_{\boldsymbol{\gamma}}\Big)\mathbf{H}\Big)=\mathrm{Tr}\Big(\mathbf{H}^\top\sum_{i=1}^n\beta_i(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{A}^{(i)})\mathbf{H}\Big)$.

Eq. (13) is an optimal function optimization on simplex. In the following, we firstly prove the differentiability of $\mathcal{T}(\boldsymbol{\beta})$ w.r.t. $\boldsymbol{\beta}$, calculate its reduced gradient, and design a reduced gradient descent algorithm to optimize it. The following Theorem 3 guarantees that $\mathcal{T}(\boldsymbol{\beta})$ is differentiable w.r.t. $\boldsymbol{\beta}$.

Theorem 3. *$\mathcal{T}(\boldsymbol{\beta})$ in Eq. (13) is differentiable w.r.t $\boldsymbol{\beta}$. Furthermore, we have $\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_i}=\mathrm{Tr}\left(\mathbf{H}^{*\top}\left(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}^*}\mathbf{A}^{(i)}\right)\mathbf{H}^*\right)$, where $(\boldsymbol{\gamma}^*,\mathbf{H}^*)=\left\{\arg\min_{\boldsymbol{\gamma}\in\Delta}\ \max_{\mathbf{H}\in\Gamma}\ \mathrm{Tr}\left(\mathbf{H}^\top\left(\sum_{i=1}^n\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}}\mathbf{A}^{(i)}\right)\mathbf{H}\right)\right\}$.*

Proof. Given $\boldsymbol{\beta}$, the optimization in Eq. (14) is the one same to SimpleMKKM. Based on Theorem 2 in [23], we know that Eq. (14) has the global optimum. According to Theorem 4.1 in [27], $\mathcal{T}(\boldsymbol{\beta})$ in Eq. (13) is differentiable, and $\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_i}=\mathrm{Tr}\left(\mathbf{H}^{*\top}\left(\mathbf{A}^{(i)}\mathbf{K}_{\boldsymbol{\gamma}^*}\mathbf{A}^{(i)}\right)\mathbf{H}^*\right)$. □

According to Theorem 3, one can calculate the gradient of $\mathcal{T}(\boldsymbol{\beta})$ w.r.t $\boldsymbol{\beta}$. Let β_u be a non-zero component of $\boldsymbol{\beta}$, and $\bigtriangledown\mathcal{T}(\boldsymbol{\beta})$ denotes the reduced gradient of $\mathcal{T}(\boldsymbol{\beta})$. The i-th $(1\leq i\leq n)$ element of $\bigtriangledown\mathcal{T}(\boldsymbol{\beta})$ is

$$[\bigtriangledown\mathcal{T}(\boldsymbol{\beta})]_i=\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_i}-\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_u}\ \ \forall\, i\neq u, \tag{15}$$

and

$$[\bigtriangledown\mathcal{T}(\boldsymbol{\beta})]_u=\sum\nolimits_{i=1,\,i\neq u}^{n}\left(\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_u}-\frac{\partial\mathcal{T}(\boldsymbol{\beta})}{\partial\beta_i}\right). \tag{16}$$

Correspondingly, the descent direction for updating $\boldsymbol{\beta}$ as

$$v_i = \begin{cases} 0 & \text{if } \beta_i = 0 \text{ and } [\bigtriangledown \mathcal{T}(\boldsymbol{\beta})]_i > 0 \\ -[\bigtriangledown \mathcal{T}(\boldsymbol{\beta})]_i & \text{if } \beta_i > 0 \text{ and } i \neq u \\ -[\bigtriangledown \mathcal{T}(\boldsymbol{\beta})]_u & \text{if } i = u. \end{cases} \tag{17}$$

The whole algorithm in solving Eq. (12) is listed in Algorithm 2. Overall, Algorithm 2 updates $(\boldsymbol{\gamma}, \mathbf{H})$ via Algorithm 1 with given $\boldsymbol{\beta}$, then calculates the gradient of $\mathcal{T}(\boldsymbol{\beta})$ w.r.t $\boldsymbol{\beta}$, and update $\boldsymbol{\beta}$ with a reduced gradient descent algorithm along the direction $\mathbf{v} = [v_1, \cdots, v_n]^\top$ in Eq. (17). This procedure is iteratively performed until achieving the given stopping criterion. In Algorithm 2, $\mathrm{obj}^{(t)}$ denotes the objective value in Eq. (12) at the t-th iteration with $\beta^{(t)}$, $\boldsymbol{\gamma}^{(t)}$ and $\mathbf{H}^{(t)}$.

Algorithm 2 SAL-SMKKM: Sample Adaptive Localized SimpleMKKM

Input: $\{\mathbf{K}_p\}_{p=1}^m$, k, τ, $t = 1$.
Initialize $\boldsymbol{\beta}^{(1)} = \mathbf{1}/n$ and $t = 1$, flag $= 1$.
Calculate $\{\mathbf{A}^{(i)}\}_{i=1}^n$ according to the average kernel.
$\tilde{\mathbf{K}}_p = \left(\sum_{i=1}^n \beta_p^{(t)} \mathbf{A}^{(i)}\right) \otimes \mathbf{K}_p$, $\forall p$.
while flag **do**
    optimize $(\boldsymbol{\gamma}^{(t)}, \mathbf{H}^{(t)})$ by Algorithm 1 with $\beta^{(t)}$.
    compute $\frac{\partial \mathcal{T}(\boldsymbol{\beta})}{\partial \beta_i}$ $(1 \leq i \leq n)$ and the descent direction $\mathbf{v}^{(t)}$ in Eq. (17).
    update $\boldsymbol{\beta}^{(t+1)} \leftarrow \boldsymbol{\beta}^{(t)} + \alpha \mathbf{v}^{(t)}$.
    if $\max |\boldsymbol{\beta}^{(t+1)} - \boldsymbol{\beta}^{(t)}| \leq 1e-4$ **then**
        flag=0.
    end if
    $t \leftarrow t + 1$.
end while

3.3 Computational Complexity and Convergence

We firstly analyze the computational complexity of the proposed UWL-SMKKM and SAL-SMKKM 2. UWL-SMKKM needs calculating a neighborhood mask matrix firstly with computational complexity $\mathcal{O}(n^2 \log_2 n)$, and then performs SimpleMKKM 1 with normalized kernel matrices. Compared with SimpleMKKM, UWL-SMKKM takes extra $\mathcal{O}(n^2 \log_2 n)$ to pre-process kernel matrices. As seen, UWL-SMKKM maintains a close computational complexity of SimpleMKKM, which has complexity $\mathcal{O}(n^3)$ at each iteration. In addition, according to [23], UWL-SMKKM is proved to be a convex optimization and theoretically guaranteed to converge to the global optimum.

SAL-SMKKM solves a SimpleMKKM problem in 1 with given $\boldsymbol{\beta}$ at each iteration. Its computational complexity is $\mathcal{O}(\ell_0 * n^3)$, where ℓ_0 is the minimum of iterations to achieve convergence. Further, the complexity of calculating $\frac{\partial \mathcal{T}(\boldsymbol{\beta})}{\partial \beta_i}$ $(1 \leq i \leq n)$ is $\mathcal{O}(k * n^3)$. As a result, its overall computational complexity at each iteration is $\mathcal{O}((\ell_0 + k) * n^3)$, which is comparable to that of SimpleMKKM.

We then briefly discuss the convergence of SAL-SMKKM. Given $\boldsymbol{\beta}$, Eq. (12) reduces to SimpleMKKM in Algorithm 1, which is proved to has a global optimum. Under this circumstance, the gradient computation in Theorem 3 can be exact, and we perform reduced gradient descent on a continuously differentiable function $\mathcal{T}(\boldsymbol{\beta})$ defined on the simplex $\{\boldsymbol{\beta} \in \mathbb{R}^n | \sum_{i=1}^n \beta_i = 1, \beta_i \geq 0, \forall i\}$, which does converge to the minimum of $\mathcal{T}(\boldsymbol{\beta})$ [26]. The experimental results in Figure 2 have also empirically validated the convergence of SAL-SMKKM.

4 Experiments

In this part, we perform empirical experiments to verify the proposed localized SimpleMKKM. We first illustrate the experimental setup, including the used datasets, the evaluation metrics, and the baseline algorithms. We then present the clustering performance comparison, the variation of learned kernel weights, and the algorithm convergence. We conduct additional experiments to show the evolution of the learned $\mathbf{H}$ and consider an analysis of parameter sensitivity. We finally make a running time comparison of different algorithms to demonstrate the efficiency of our algorithm.

4.1 Experimental Setup

Several MKKM benchmarks are adopted to evaluate the clustering performance of our proposed algorithm, including *MSRA* [30], *Cal-7*[1], *PFold*[2], *Flo17*[3], *Flo102*[4], and *Reuters*[5]. The detailed information of these dataset are summarized in Table 2. As seen, the number of samples, kernels, and clusters of the used benchmarks are considerably different, which provides the clustering algorithms with a great platform for performance comparison.

It is worth noting that we set the pre-specified value k as the true number of clusters for all datasets. Three widely-used metrics are employed for evaluation, including clustering accuracy (ACC), normalized mutual information (NMI), and rand index (RI). To alleviate the impact of randomness caused by k-means, we run each algorithm 50 times with random initialization, and report the averaged results with standard deviation.

We comprehensively compare the proposed UML-SMMK and SAL-SMKKM with nine state-of-the-art multiple kernel clustering algorithms, including Average kernel (Avg-KKM), Multiple kernel (MKKM) [25], Localized multiple kernel (LMKKM) [28], Optimal neighborhood kernel clustering (ONKC) [15], Multiple kernel with matrix-induced regularization (MKKM-MR) [21], Multiple kernel clustering with local alignment maximization (LKAM) [14], Multi-view clustering via late fusion alignment maximization (LF-MVC) [16], MKKM-MM [29], and SimpleMKKM [23]. We directly reproduce the compared algorithms by following their available implementations in original literature and report the corresponding clustering results. Among the mentioned baseline algorithms, ONKC [15], MKKM-MiR [21], LKAM [14] and LF-MVC [16] have hyper-parameters to be optimized. Following the same experimental setup in original literature, we adopt the public source codes and conduct the hyper-parameter tuning process carefully to

1. http://www.vision.caltech.edu/ImageDatasets/Caltech101
2. mkl.ucsd.edu/dataset/protein-fold-prediction
3. www.robots.ox.ac.uk/~vgg/data/flowers/17/
4. www.robots.ox.ac.uk/~vgg/data/flowers/102/
5. http://kdd.ics.uci.edu/databases/reuters21578/

TABLE 1: Empirical comparison of the proposed UWL-SMKKM and SAL-MKKM with nine baseline algorithms on six datasets in terms of four publicly-used metrics. Boldface means the best clustering result.

DATASETS	AVG-KKM	MKKM [25]	LMKKM [28]	ONKC [15]	MKKM-MR [21]	LKAM [14]	LF-MVC [16]	MKKM-MM [29]	SMKKM [23]	UWL-SMKKM PROPOSED	SAL-SMKKM PROPOSED
ACC											
MSRA	83.3 ± 0.8	81.3 ± 3.1	81.9 ± 0.7	85.4 ± 0.4	88.1 ± 0.1	89.1 ± 0.2	87.8 ± 0.4	83.3 ± 0.8	86.5 ± 0.2	$\mathbf{91.2 \pm 1.0}$	$\mathbf{91.9 \pm 0.0}$
CAL-7	59.2 ± 4.9	52.2 ± 4.3	53.9 ± 1.0	69.4 ± 2.5	68.4 ± 0.3	70.4 ± 1.4	71.4 ± 1.4	59.2 ± 4.9	68.2 ± 1.5	$\mathbf{76.5 \pm 0.2}$	$\mathbf{79.3 \pm 0.1}$
PFD	29.0 ± 1.5	27.0 ± 1.1	22.4 ± 0.7	36.3 ± 1.5	34.7 ± 1.8	37.7 ± 1.2	33.0 ± 1.4	29.0 ± 1.5	34.7 ± 1.9	$\mathbf{35.9 \pm 1.5}$	$\mathbf{36.7 \pm 1.6}$
FLO17	50.8 ± 1.5	44.9 ± 2.4	37.5 ± 1.6	54.2 ± 2.2	58.5 ± 1.5	50.0 ± 0.8	61.0 ± 0.7	50.8 ± 1.5	59.5 ± 1.3	$\mathbf{61.3 \pm 1.3}$	$\mathbf{61.5 \pm 0.9}$
FLO102	27.1 ± 0.8	22.4 ± 0.5	-	39.5 ± 0.7	40.2 ± 0.9	41.4 ± 0.8	38.4 ± 1.2	27.1 ± 0.8	42.5 ± 0.8	$\mathbf{44.0 \pm 1.0}$	43.2 ± 1.1
REUTERS	45.5 ± 1.5	45.4 ± 1.5	-	40.9 ± 2.1	39.7 ± 1.5	40.0 ± 2.2	45.4 ± 1.7	45.5 ± 1.5	45.5 ± 0.7	$\mathbf{46.6 \pm 1.0}$	$\mathbf{53.2 \pm 0.0}$
NMI											
MSRA	74.0 ± 1.0	73.2 ± 1.7	75.0 ± 1.4	74.9 ± 0.7	77.6 ± 0.3	79.8 ± 0.2	79.4 ± 0.8	74.0 ± 1.0	75.2 ± 0.5	$\mathbf{82.6 \pm 1.5}$	$\mathbf{85.1 \pm 0.0}$
CAL-7	59.1 ± 2.9	51.6 ± 4.1	52.1 ± 1.3	63.5 ± 2.4	64.1 ± 0.2	65.3 ± 0.7	70.1 ± 3.0	59.1 ± 2.9	63.7 ± 0.3	$\mathbf{74.6 \pm 1.2}$	$\mathbf{75.7 \pm 0.1}$
PFD	40.3 ± 1.3	38.0 ± 0.6	34.7 ± 0.6	44.4 ± 0.9	43.7 ± 1.2	46.2 ± 0.6	41.7 ± 1.1	40.3 ± 1.3	44.4 ± 1.1	$\mathbf{45.2 \pm 1.3}$	$\mathbf{45.8 \pm 1.1}$
FLO17	49.7 ± 1.0	44.9 ± 1.5	38.8 ± 1.1	52.6 ± 1.2	56.4 ± 0.9	49.8 ± 0.6	$\mathbf{58.9 \pm 0.4}$	49.7 ± 1.0	57.8 ± 0.9	$\mathbf{58.9 \pm 0.5}$	$\mathbf{60.6 \pm 0.6}$
FLO102	46.0 ± 0.5	42.7 ± 0.2	-	56.1 ± 0.4	56.7 ± 0.5	56.9 ± 0.3	54.9 ± 0.4	46.0 ± 0.5	58.6 ± 0.5	$\mathbf{60.0 \pm 0.4}$	59.5 ± 0.4
REUTERS	27.4 ± 0.4	27.3 ± 0.4	-	21.0 ± 1.8	21.3 ± 1.3	21.5 ± 2.3	27.2 ± 0.2	27.4 ± 0.4	$\mathbf{27.7 \pm 0.2}$	27.0 ± 0.6	$\mathbf{31.0 \pm 0.0}$
PURITY											
MSRA	83.3 ± 0.8	81.5 ± 2.7	81.9 ± 0.7	85.4 ± 0.4	88.1 ± 0.1	89.1 ± 0.2	87.8 ± 0.4	83.3 ± 0.8	86.5 ± 0.2	$\mathbf{91.2 \pm 1.0}$	$\mathbf{91.9 \pm 0.0}$
CAL-7	68.0 ± 3.2	63.8 ± 3.7	66.4 ± 0.6	74.0 ± 2.1	72.9 ± 0.3	76.6 ± 0.4	79.6 ± 2.9	68.0 ± 3.2	72.3 ± 0.2	$\mathbf{81.7 \pm 1.3}$	$\mathbf{83.9 \pm 0.1}$
PFD	37.4 ± 1.7	33.7 ± 1.1	31.2 ± 1.0	42.7 ± 1.3	41.9 ± 1.4	43.7 ± 0.8	39.3 ± 1.5	37.4 ± 1.7	41.8 ± 1.5	$\mathbf{42.5 \pm 1.6}$	$\mathbf{43.9 \pm 1.5}$
FLO17	51.9 ± 1.5	46.2 ± 2.0	39.2 ± 1.3	55.4 ± 2.2	59.7 ± 1.6	51.4 ± 0.7	$\mathbf{62.4 \pm 0.7}$	51.9 ± 1.5	60.9 ± 1.2	62.0 ± 1.3	$\mathbf{62.8 \pm 0.6}$
FLO102	32.3 ± 0.6	27.8 ± 0.4	-	45.6 ± 0.7	46.3 ± 0.8	48.0 ± 0.6	44.6 ± 0.8	32.3 ± 0.6	48.6 ± 0.7	$\mathbf{50.3 \pm 0.7}$	49.7 ± 0.6
REUTERS	53.0 ± 0.4	52.9 ± 0.5	-	51.8 ± 1.5	50.9 ± 1.4	51.9 ± 1.0	52.9 ± 0.3	53.0 ± 0.4	$\mathbf{53.3 \pm 0.0}$	52.8 ± 0.2	$\mathbf{56.9 \pm 0.0}$
RI											
MSRA	68.1 ± 1.0	66.2 ± 3.1	68.0 ± 1.1	69.8 ± 0.7	74.5 ± 0.1	76.7 ± 0.4	74.5 ± 0.8	68.1 ± 1.0	71.2 ± 0.5	$\mathbf{80.6 \pm 1.8}$	$\mathbf{82.5 \pm 0.0}$
CAL-7	46.0 ± 6.5	38.3 ± 4.9	41.2 ± 1.1	56.8 ± 4.2	55.6 ± 0.6	59.4 ± 2.2	65.2 ± 3.4	46.0 ± 6.5	55.6 ± 0.3	$\mathbf{69.4 \pm 0.7}$	$\mathbf{78.9 \pm 0.1}$
PFD	14.4 ± 1.8	12.1 ± 0.7	7.8 ± 0.4	18.0 ± 1.1	17.2 ± 1.5	20.1 ± 1.1	16.1 ± 1.5	14.4 ± 1.8	17.6 ± 1.9	$\mathbf{19.8 \pm 1.2}$	$\mathbf{20.6 \pm 1.2}$
FLO17	32.2 ± 1.3	27.2 ± 1.8	20.6 ± 1.1	35.2 ± 1.5	39.9 ± 1.3	31.6 ± 0.8	$\mathbf{44.1 \pm 0.4}$	32.2 ± 1.3	41.5 ± 1.5	43.2 ± 0.9	$\mathbf{44.7 \pm 0.7}$
FLO102	15.5 ± 0.5	12.1 ± 0.4	-	24.9 ± 0.5	25.5 ± 0.6	27.2 ± 0.6	25.5 ± 1.0	15.5 ± 0.5	28.5 ± 0.8	$\mathbf{29.9 \pm 0.8}$	29.3 ± 0.9
REUTERS	21.8 ± 1.4	21.8 ± 1.4	-	18.8 ± 2.4	18.9 ± 2.0	16.9 ± 2.7	21.4 ± 1.1	21.8 ± 1.4	$\mathbf{22.1 \pm 0.8}$	21.5 ± 0.3	$\mathbf{30.1 \pm 0.0}$

TABLE 2: Benchmark datasets used in our experiments.

Dataset	Number of		
	Samples	Kernels	Clusters
MSRA	210	6	7
Cal-7	441	6	7
PFD	694	12	27
Flo17	1360	7	17
Flo102	8189	4	102
Reuters	18758	5	6

achieve the best possible clustering results on each benchmark dataset.

4.2 Empirical Study

4.2.1 Clustering Performance Analysis

As illustrated in Table 1, we report the clustering performance of all compared algorithms in terms of ACC, NMI, purity and RI. From these results, several observations can be concluded as below:

- MKKM-MM [29] is the first algorithm to improve MKKM via the minimization-maximization optimization. As seen, although the adversarial learning does benefit the MKKM, the clustering performance enhancement over MKKM is slight on all datasets. Moreover, the proposed UWL-SMKKM outperforms MKKM-MM by a large margin. The newly proposed SAL-SMKKM further improves UWL-SMKKM. This demonstrates the advantage of our formulation and the associated optimization strategy.
- Besides our UWL-SMKKM and SAL-SMKKM, SimpleMKKM achieves competitive performance compared to the aforementioned clustering algorithms on all datasets. These results have solidly verified the superiority of its novel formulation and innovative optimization algorithm.
- The developed UWL-SMKKM consistently exceeds SimpleMKKM by a large margin. Taking the ACC metric for example, UWL-SMKKM outperforms SimpleMKKM algorithm by 4.7%, 8.3%, 1.2%, 1.8%, 1.5% and 1.1% on six benchmark datasets. The improvements in terms of other criteria are similar. The newly proposed SAL-SMKKM further improves UWL-SMKKM by 0.7%, 2.8%, 0.8%, 0.2% and 6.6% on MRSA, CAL-7, PFD, FLO17 and Reuters, respectively, and achieve slight worse ACC than UWL-SMKKM on FLO102. These results clearly verify the effectiveness and superiority of the proposed UWL-SMKKM and SAL-SMKKM that benefit from extracting and preserving the local information of kernel matrix.

The proposed algorithm not only inherits the clustering-friendly formulation associated with a novel optimization from SimpleMKKM, but also conducts the kernel alignment in a localized manner, which enables the algorithm to well tackle the issue of kernel variation. These merits jointly contribute to its significant performance improvement over the existing competitors on all datasets. It is expected that the proposed simple and effective solution will have the potential to be a good option in real-world clustering applications. Moreover, please note that the clustering results of LMKKM [28] on some datasets are not presented because of the out-of-memory exception caused by the cubic computational complexity.

4.2.2 Kernel Weight Analysis

Figure 1 shows the analysis on kernel weights learned by different algorithms. From these results, we observed that

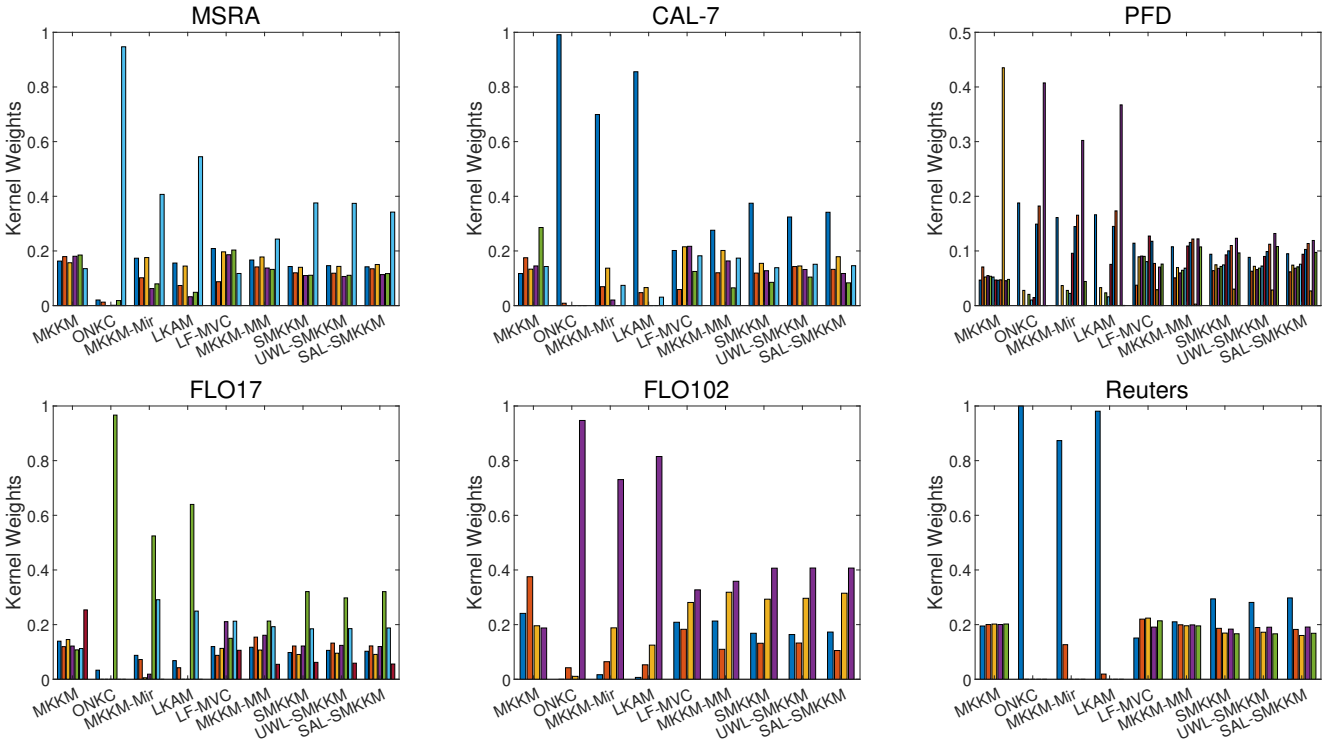

Fig. 1: The learned kernel weights of different algorithms on all datasets.

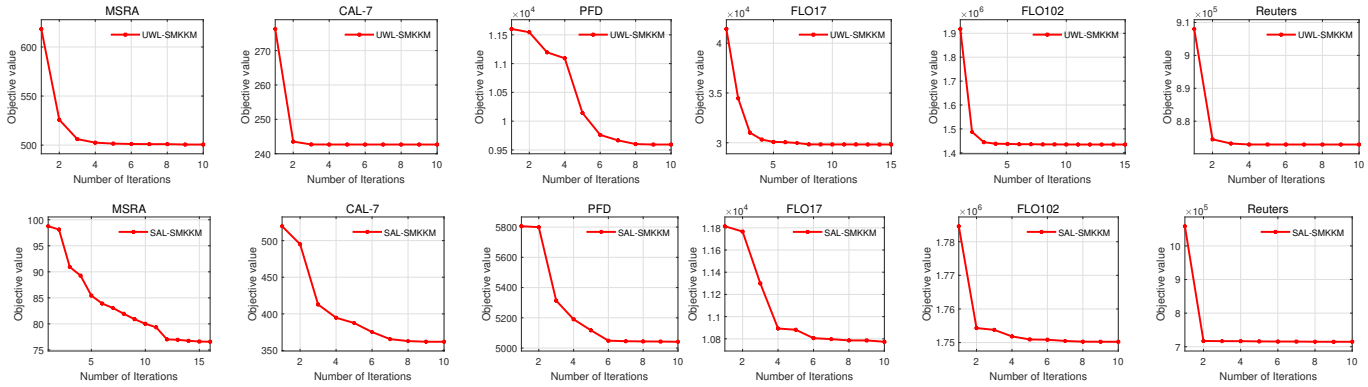

Fig. 2: The objectives of UWL-SMKKM and SAL-SMKKM vary with iterations.

the kernel weights learned by LKAM, ONKC and MKKM-MiR algorithms exhibit great variation on most of datasets, while keep highly sparse on Reuters. This sparsity would prevent the multiple kernel matrices from being fully exploited, resulting in unsatisfactory performance. Taking the results on Reuters for example, MKKM-MiR and LKAM only achieves 39.7% and 40.0% clustering accuracy. Differently, despite the ℓ_1-norm constraint on $\boldsymbol{\gamma}$, our UWL-SMKKM and SAL-SMKKM could consistently learn non-sparse kernel weights on all datasets, which leads to its promising clustering performance. This benefits from the proposed $\min_{\boldsymbol{\gamma}}$-$\max_{\mathbf{H}}$ kernel alignment objective and the resultant reduced gradient descent algorithm.

4.2.3 Convergence and Evolution of the Learned $\mathbf{H}$

In section 3.3, it has been proved that the proposed UWL-SMKKM and SAL-SMKKM are guaranteed to converge in theory. To see this point in depth, Figure 2 presents the objective curves of UWL-SMKKM and SAL-SMKKM with iterations. From these results, we can see that the objectives of UWL-SMKKM and SAL-SMKKM keep monotonically reduced until convergence in few iterations on all datasets.

In addition, to analysis the clustering performance variation of the learned $\mathbf{H}$ by UWL-SMKKM and SAL-MKKM with iterations, we calculate four metrics with iterations, and present the corresponding curves in Figure 3 and 4. As seen, the clustering performance of UWL-SMKKM and SAL-MKKM first tends to be increased with slight oscillation, and then keeps stable in a wide range of iterations. These results show the necessity and effectiveness of the optimization procedure.

4.2.4 Hyper-parameter Analysis

As observed, the proposed UWL-SMKKM and SAL-MKKM introduce a hyper-parameter τ to explore and collect more informative neighborhood information among samples. As

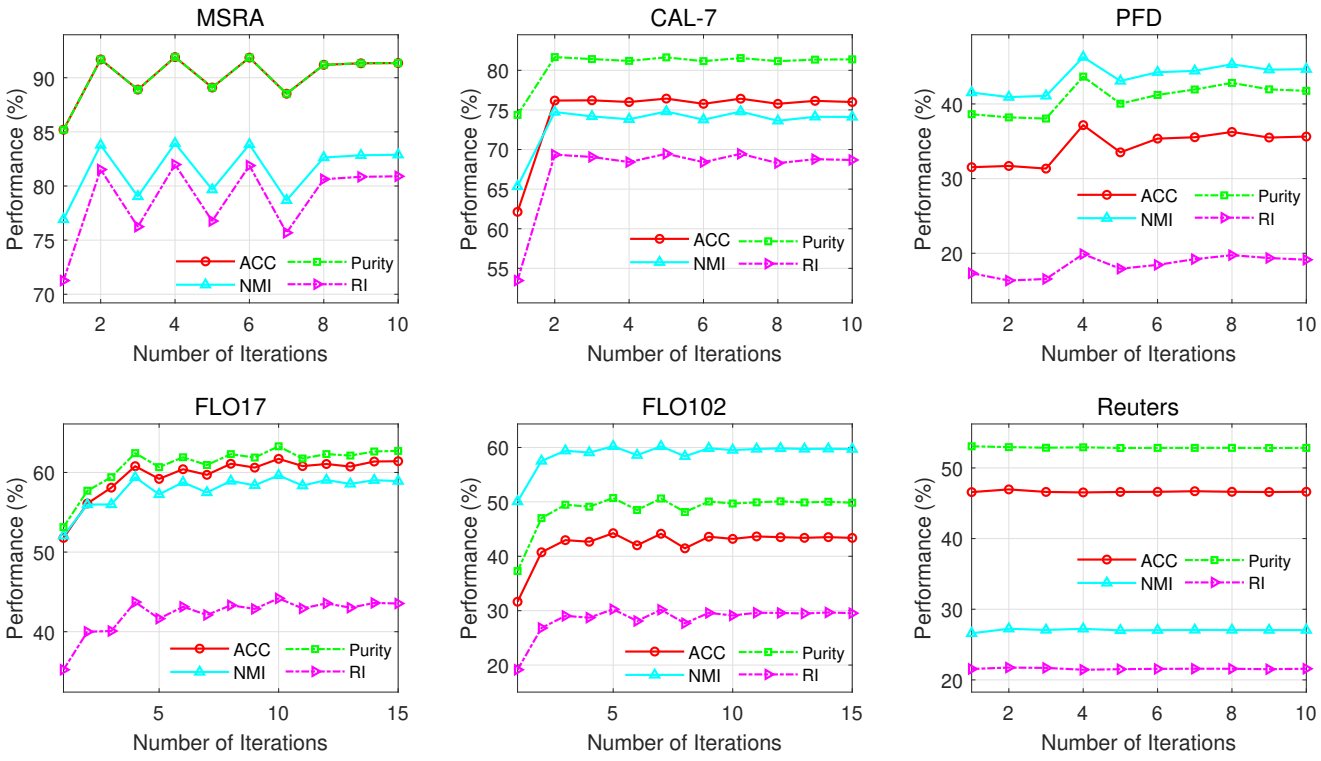

Fig. 3: The clustering performance of the learned $\mathbf{H}$ by UWL-SMKKM at each iteration on six datasets.

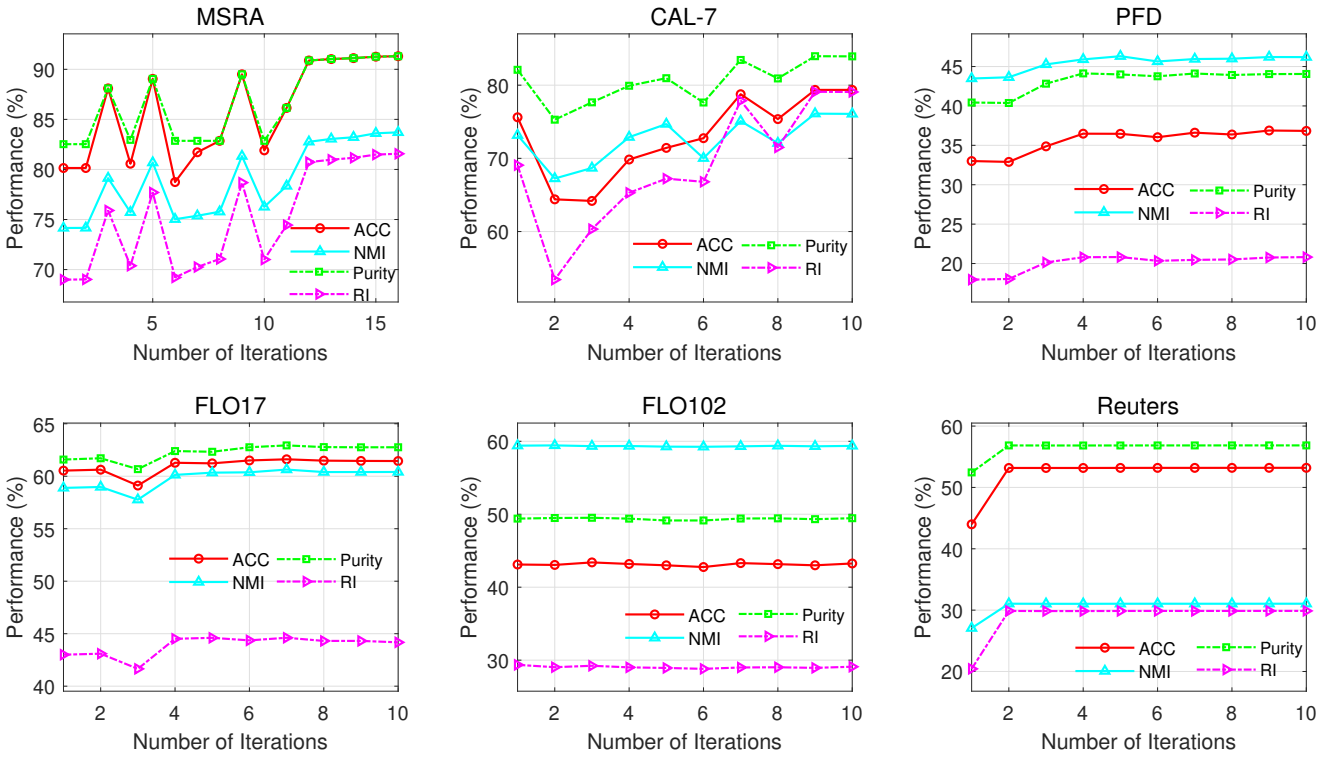

Fig. 4: The clustering performance of the learned $\mathbf{H}$ by SAL-SMKKM at each iteration on six datasets.

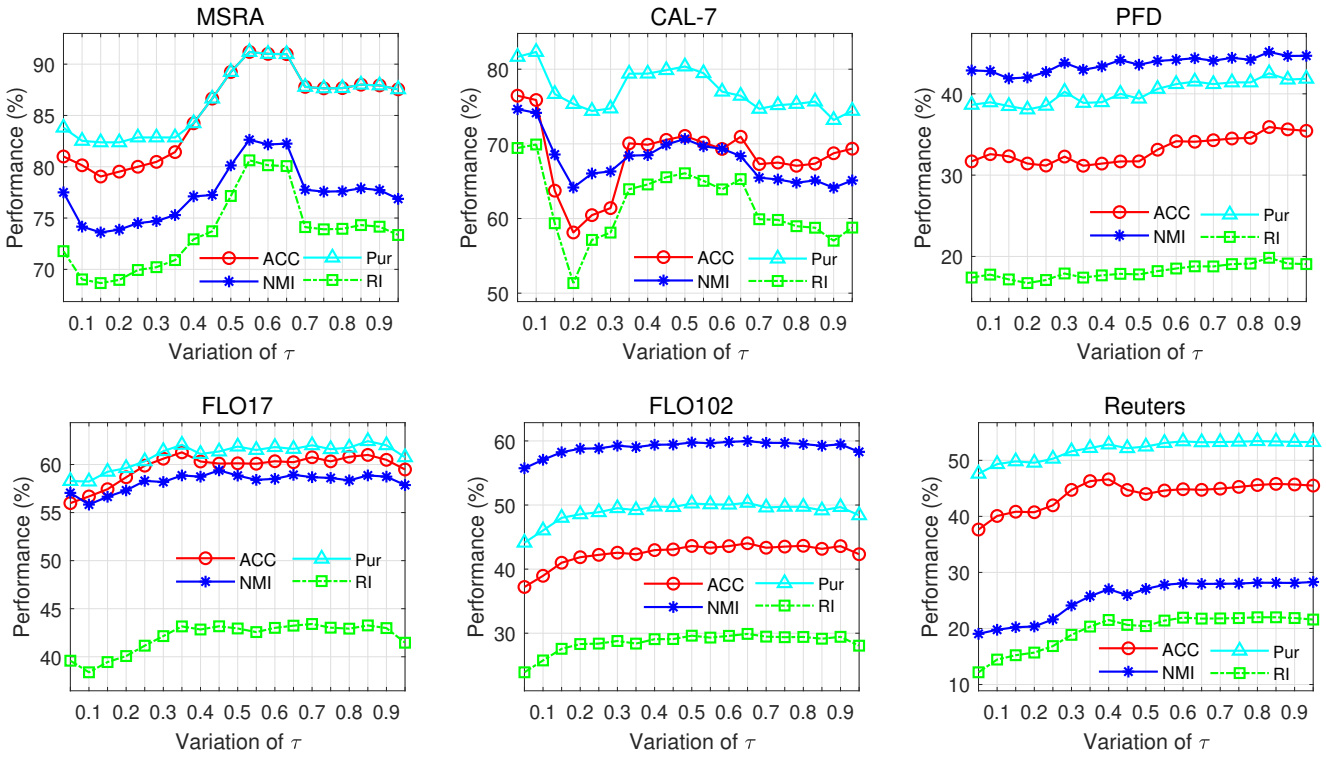

Fig. 5: The clustering performance of UWL-SMKKM with the different size of neighborhood τ on six datasets

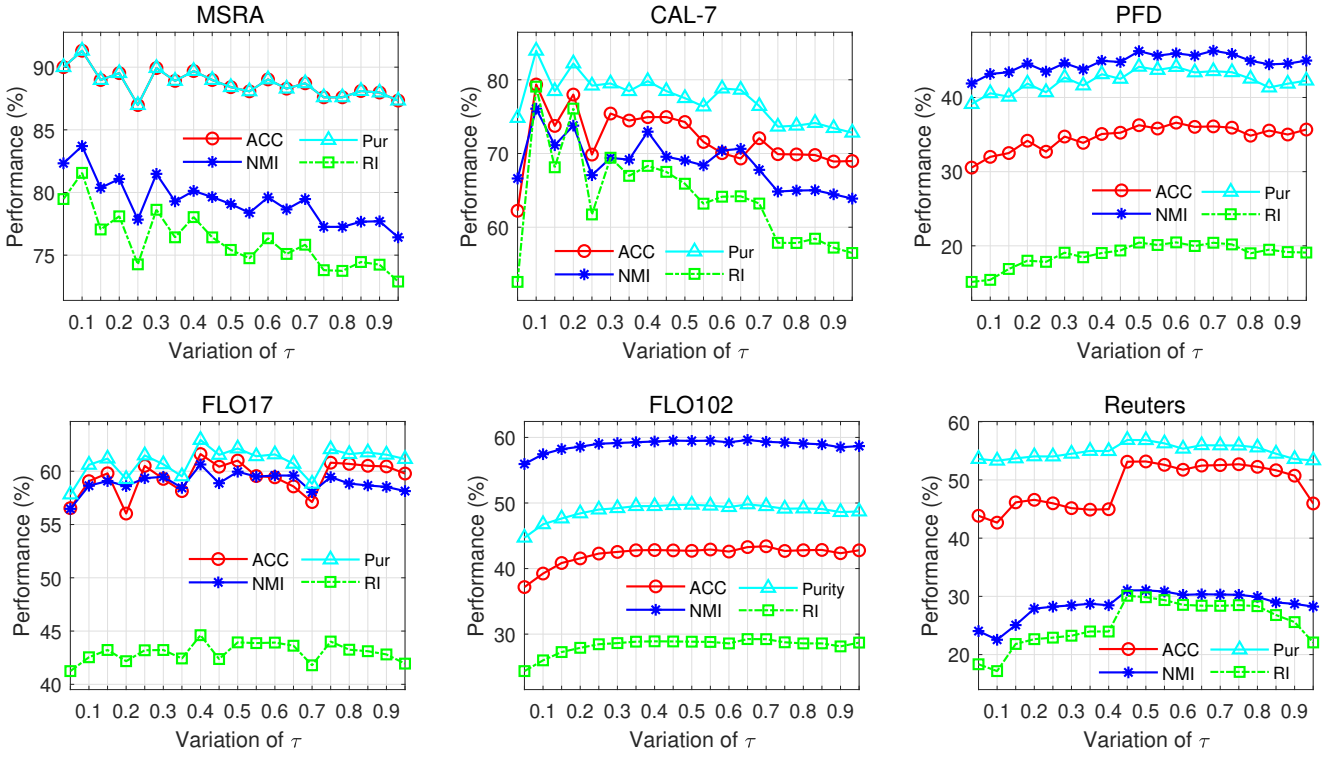

Fig. 6: The clustering performance of SAL-SMKKM with the different size of neighborhood τ on six datasets.

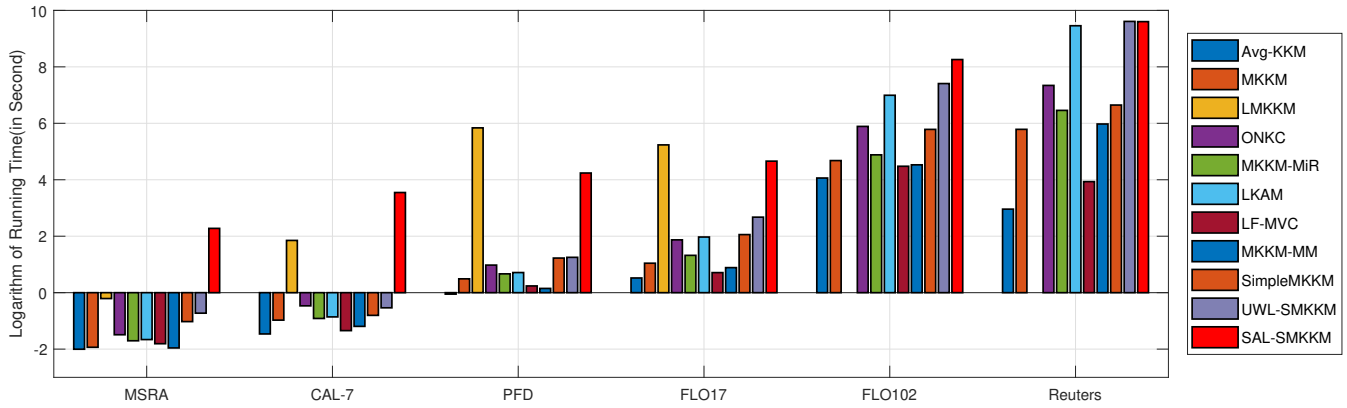

Fig. 7: Running time comparison of different algorithms (logarithm in seconds). Note that we empirically evaluate their running time on all datasets and conduct all experiments on a common PC platform in MATLAB R2020b environment.

illustrated in Figure 5 and 6, we investigate how this parameter affects the clustering performance by performing a grid search (e.g., τ varies from 0.05 to 0.95) with a step size of 0.05. From these figure, we find that the newly developed UWL-SMKKM and SAL-MKKM show stable clustering performance in a wide range of τ. This observation indicates that the proposed algorithm is robust with the variation of τ values.

4.2.5 *Running time analysis*

According to the aforementioned analysis in Subsection 3.3, the proposed UWL-SMKKM and SAL-SMKKM have the same computational complexity with SimpleMKKM in theory. For a fair comparison, we empirically evaluate their running time on all datasets and conduct all experiments on a common PC platform in MATLAB R2020b environment. The results are presented in Figure 7. As observed, the proposed UWL-SMKKM and SAL-SMKKM greatly improves the clustering performance without significantly increasing the computation burden.

5 Application in Multi-modal Parcellation of Human Cerebral Cortex

In this section, we conduct application research by employing the proposed UWL-SMKKM to the multi-modal parcellation of human cerebral cortex, with the aim to test its performance in practical applications. As known, cerebral cortex is the most complex and sophisticated system in the world. Neuroscientists has made many efforts to subdivided the cerebral cortex into functionally and anatomically distinct areas, which is essential to understand how the brain works. Accurate parcellation will not only illuminate the functional and structural properties of the brain, but also help to reduce data dimension while improving statistical sensitivity. Most previous studies parcellated the cerebral cortex by only one neurobiological feature, such as functional connectivity [31] and cortical morphology [32]. However, few study attempted to parcellate in a multi-modal manner. Since some cortical boundaries are not sensitive to a single specific modality, it is believed that integrating multiple properties will provide complementary as well as confirmatory information, which can help to generate a more correct parcellation.

Inspired by Glasser and his colleagues who have established the first multi-modal parcellation of human cerebral cortex by a semi-automatic method with manual delineation [33], we aim to achieve the more objective parcellation and accurate boundaries mapping without neuroanatomist's instruction. Multi-modal neuroimaging data are downloaded from human connectome project [6]. In practice, for each cortical region defined in Desikan–Killiany (DK) template [32], we construct five kernels as the inputs of SimpleMKKM. The first four kernels reflect the functional properties of this region, which are obtained from the 7T fMRI data of four scanning runs. In specific, for each cortical vertex within the region, the functional connectivity map to the whole brain is calculated for each run. Then, four RBF kernels are constructed based on the functional connectivity maps of four scanning runs, respectively, with the parameter gamma defined by averaged pair-similarities. More details about the calculation process can be found in our previous study [31]. The fifth kernel represents the anatomical properties, with cortical thickness, myelin content and sulcus architecture being utilized to construct the RBF kernel.

Parcellation results generated by UWL-SMKKM with five multi-modal kernels are illustrated in Figure 8, where the human cerebral cortex is subdivided into 552 cortical parcels. Most of the parcels demonstrate satisfactory symmetry, with their symmetrical parcels observed in the contralateral hemisphere. Additionally, we further compare some local parcellation results with Glasser's work. As illustrated in Figure 9, the black lines represent the empirical boundaries derived from the multi-modal gradients and anatomist's experiences [34], and the colorful parcels are the clustering results of UWL-SMKKM on paracentral gyrus, precuneus as well as parahippocampal gyrus. It is exciting that extremely high consistency is observed between the results from two approaches, while UWL-SMKKM achieves the accurate parcellation in an automatic and objective

6. https://www.humanconnectome.org/

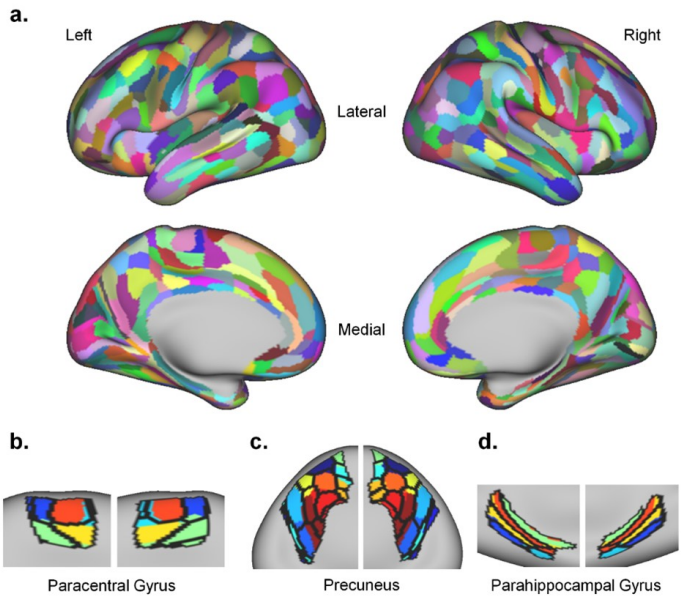

Fig. 8: The multi-modal parcellation of human cerebral cortex derived from UWL-SMKKM (a). Parcellation results of paracentral gyrus, precuneus and parahippocampal gyrus are overlaid with the empirical boundaries from neuroanatomist's delineation, respectively (b). The colorful parcels are the clustering results of UWL-SMKKM, and the black lines denote the boundaries from [34].

manner without manual intervention, which indicates the validity and effectiveness of our method.

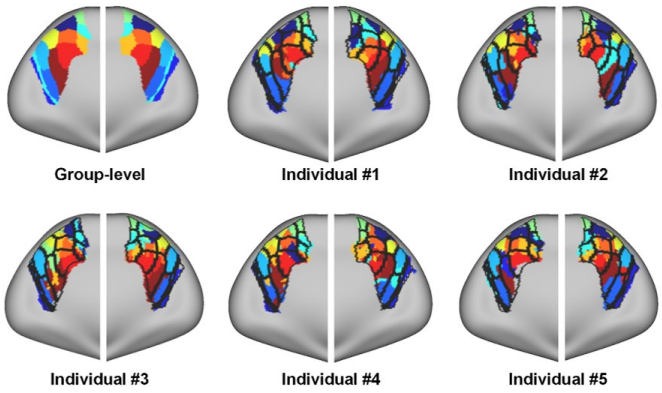

Fig. 9: The multi-modal parcellation of human precuneus on five single individuals. The black lines denote the group-level boundaries.

It should be noted that both the Glasser's and our parcellations are established on the population level, where the features of each modality are averaged across many individuals. Although the group parcellation has revealed many basic principles of cortical organization, it is limited for the clinical applications due to the lack of emphasis on meaningful individual differences [33]. Moreover, expert experiences will undoubtedly allow the exquisite delineation of cerebral cortex, but this method is impractical to extend on the individual level. As an unsupervised and automatic approach, UWL-SMKKM has shown a good performance on the group-level parcellation, and we believe that it can serve as an alternative on the individual multi-modal parcellation. In practice, we take the human precuneus as an example, which is an important medium of multiple functional networks and participates in many high-level functions. Considering the low signal-to-noise ratio of individual data, we introduce the spatial constraint by adding the sixth RBF kernel which reflects the spatial locations of cortical vertices. The calculation of each kernel is conducted on five different individuals, respectively. The clustering results are illustrated in Figure 9, where individual multi-modal parcellations of precuneus are overlaid with the group boundaries. As seen, topological differences are observed between the results of individuals, which reflects the individual-specific functional and anatomical patterns; all the individuals still partly share the similar boundaries with group, which in turn indicates the reliability of individual parcellation. In summary, the individual differences and group correspondences demonstrate the potential of UWL-SMKKM on individual multi-modal parcellation.

6 Conclusion

While the currently proposed SimpleMKKM demonstrates encouraging clustering performance, it doesn't sufficiently consider the variation information among base kernel matrices. To address this issue, we propose to calculate the kernel alignment in a local manner rather than a global one. We reveal the theoretical connection between the proposed UWL-SMKKM and SimpleMKKM. Based on this discovery, we still use the reduced gradient descent algorithm to solve the obtained optimization problem. We further improve UWL-SMKKM by allowing the weights of local kernel alignment for each sample to be adaptively adjusted, which is termed SAL-SMKKM. The proposed UWL-SMKKM and SAL-SMKKM demonstrate significantly improved clustering results via extensive experiments on both benchmark datasets and practical applications. In the future, to replace the unchanged nearest neighbors of each sample, we plan to study the automatic updating them during the optimization learning process for further improving clustering effect.

Acknowledgements

This work was supported by the National Key R&D Program of China 2020AAA0107100, the Natural Science Foundation of China (project no. 61922088).

References

[1] L. Xu, J. Neufeld, B. Larson, and D. Schuurmans, "Maximum margin clustering," in *NIPS*, 2004, pp. 1537–1544.
[2] W. Tang, Z. Lu, and I. S. Dhillon, "Clustering with multiple graphs," in *ICDM*. IEEE, 2009, pp. 1016–1021.
[3] S. Yu, L. Tranchevent, X. Liu, W. Glanzel, J. A. K. Suykens, B. De Moor, and Y. Moreau, "Optimized data fusion for kernel k-means clustering," *IEEE Transactions on Software Engineering*, vol. 34, no. 5, pp. 1031–9, 2012.
[4] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *ICML*, 2011, pp. 393–400.
[5] Z. Huang, P. Hu, J. T. Zhou, J. Lv, and X. Peng, "Partially view-aligned clustering," in *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020*.

[6] X. Peng, Z. Huang, J. Lv, H. Zhu, and J. T. Zhou, "COMIC: multi-view clustering without parameter selection," in *Proceedings of the 36th International Conference on Machine Learning, ICML 2019*, K. Chaudhuri and R. Salakhutdinov, Eds., vol. 97, pp. 5092–5101.

[7] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *Proceedings of the 30th International Conference on Machine Learning, ICML 2013*, vol. 28, pp. 352–360.

[8] S. Zhou, E. Zhu, X. Liu, T. Zheng, Q. Liu, J. Xia, and J. Yin, "Subspace segmentation-based robust multiple kernel clustering," *Information Fusion*, vol. 53, pp. 145–154, 2020.

[9] W. Liang, S. Zhou, J. Xiong, X. Liu, S. Wang, E. Zhu, Z. Cai, and X. Xu, "Multi-view spectral clustering with high-order optimal neighborhood laplacian matrix," *IEEE Transactions on Knowledge and Data Engineering*, 2020.

[10] Z. Kang, X. Zhao, C. Peng, H. Zhu, J. T. Zhou, X. Peng, W. Chen, and Z. Xu, "Partition level multiview subspace clustering," *Neural Networks*, vol. 122, pp. 279–288, 2020.

[11] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1582–1590.

[12] C. Zhang, H. Fu, J. Wang, W. Li, X. Cao, and Q. Hu, "Tensorized multi-view subspace representation learning," *International Journal of Computer Vision*, pp. 1–18, 2020.

[13] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 330–344, 2022. [Online]. Available: https://doi.org/10.1109/TPAMI.2020.3011148

[14] M. Li, X. Liu, L. Wang, Y. Dou, J. Yin, and E. Zhu, "Multiple kernel clustering with local kernel alignment maximization," in *International Joint Conference on Artificial Intelligence*, 2016, pp. 1704–1710.

[15] X. Liu, S. Zhou, Y. Wang, M. Li, Y. Dou, E. Zhu, and J. Yin, "Optimal neighborhood kernel clustering with multiple kernels," in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, S. P. Singh and S. Markovitch, Eds. AAAI Press, 2017, pp. 2266–2272. [Online]. Available: http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14761

[16] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, and J. Yin, "Multi-view clustering via late fusion alignment maximization." in *IJCAI*, 2019, pp. 3778–3784.

[17] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, and W. Gao, "Multiple kernel k-means with incomplete kernels," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 5, pp. 1191–1204, 2020. [Online]. Available: https://doi.org/10.1109/TPAMI.2019.2892416

[18] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao, "Late fusion incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 10, pp. 2410–2423, 2019. [Online]. Available: https://doi.org/10.1109/TPAMI.2018.2879108

[19] X. Liu, L. Wang, X. Zhu, M. Li, E. Zhu, T. Liu, L. Liu, Y. Dou, and J. Yin, "Absent multiple kernel learning algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1303–1316, 2020. [Online]. Available: https://doi.org/10.1109/TPAMI.2019.2895608

[20] M. Gönen and E. Alpaydin, "Localized multiple kernel learning," in *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 352–359.

[21] X. Liu, Y. Dou, J. Yin, L. Wang, and E. Zhu, "Multiple kernel k -means clustering with matrix-induced regularization," in *Thirtieth AAAI Conference on Artificial Intelligence*, 2016, pp. 1888–1894.

[22] X. Liu, M. Li, C. Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, and E. Zhu, "Efficient and effective regularized incomplete multi-view clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 8, pp. 2634–2646, 2021. [Online]. Available: https://doi.org/10.1109/TPAMI.2020.2974828

[23] X. Liu, E. Zhu, J. Liu, T. M. Hospedales, Y. Wang, and M. Wang, "Simplemkkm: Simple multiple kernel k-means," *CoRR*, vol. abs/2005.04975, 2020. [Online]. Available: https://arxiv.org/abs/2005.04975

[24] X. Liu, S. Zhou, L. Liu, C. Tang, S. Wang, J. Liu, and Y. Zhang, "Localized simple multiple kernel k-means," in *IEEE/CVF International Conference on Computer Vision, ICCV*. IEEE, 2021, pp. 9273–9281. [Online]. Available: https://doi.org/10.1109/ICCV48922.2021.00916

[25] H.-C. Huang, Y.-Y. Chuang, and C.-S. Chen, "Multiple kernel fuzzy clustering," *IEEE Transactions on Fuzzy Systems*, vol. 20, no. 1, pp. 120–134, 2012.

[26] A. Rakotomamonjy, F. R. Bach, S. Canu, and Y. Grandvalet, "Simplemkl," *JMLR*, vol. 9, pp. 2491–2521, 2008.

[27] J. F. Bonnans and A. Shapiro, "Optimization problems with perturbations: A guided tour," *SIAM Review*, vol. 40, no. 2, pp. 228–264, 1998.

[28] M. Gönen and A. A. Margolin, "Localized data fusion for kernel k-means clustering with application to cancer biology," in *Advances in Neural Information Processing Systems*, 2014, pp. 1305–1313.

[29] S. Bang and W. Wu, "Multiple kernel k-means clustering using min-max optimization with ℓ_2 regularization," *CoRR*, vol. abs/1803.02458, 2018. [Online]. Available: http://arxiv.org/abs/1803.02458

[30] J. M. Winn and N. Jojic, "LOCUS: learning object classes with unsupervised segmentation," in *10th IEEE International Conference on Computer Vision (ICCV 2005), 17-20 October 2005, Beijing, China*. IEEE Computer Society, 2005, pp. 756–763. [Online]. Available: https://doi.org/10.1109/ICCV.2005.148

[31] Z. Luo, L.-L. Zeng, J. Qin, C. Hou, H. Shen, and D. Hu, "Functional parcellation of human brain precuneus using density-based clustering," *Cerebral Cortex*, vol. 30, no. 1, pp. 269–282, 05 2019.

[32] R. S. Desikan, F. Ségonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, M. S. Albert, and R. J. Killiany, "An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest," *NeuroImage*, vol. 31, no. 3, pp. 968–980, 2006.

[33] D. Wang, R. L. Buckner, M. D. Fox, D. J. Holt, A. J. Holmes, S. Stoecklein, G. Langs, R. Pan, T. Qian, K. Li *et al.*, "Parcellating cortical functional networks in individuals," *Nature neuroscience*, vol. 18, no. 12, pp. 1853–1860, 2015.

[34] M. F. Glasser, T. S. Coalson, E. C. Robinson, C. D. Hacker, J. Harwell, E. Yacoub, K. Ugurbil, J. Andersson, C. F. Beckmann, M. Jenkinson *et al.*, "A multi-modal parcellation of human cerebral cortex," *Nature*, vol. 536, no. 7615, pp. 171–178, 2016.

Xinwang Liu received his PhD degree from National University of Defense Technology (NUDT), China. He is now Professor at School of Computer, NUDT. His current research interests include kernel learning and unsupervised feature learning. Dr. Liu has published 90+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IEEE T-KDE, IEEE T-IP, IEEE T-NNLS, IEEE T-MM, IEEE T-IFS, ICML, NeurIPS, CVPR, ICCV, AAAI, IJCAI, etc. More information can be found at https://xinwangliu.github.io/.

Li Liu received her Ph.D. degree in information and communication engineering from the National University of Defense Technology (NUDT), China, in 2012. She is currently a Professor with the College of System Engineering, NUDT. She was a cochair of seven International Workshops at CVPR, ICCV, and ECCV. She was a guest editor of special issues for IEEE TPAMI and IJCV. Her current research interests include facial behavior analysis, texture analysis, image classification, object detection and recognition. Her papers have currently over 6000 citations in Google Scholar. She currently serves as Associate Editor of the Visual Computer Journal and Pattern Recognition.

Limin Peng received the B.E. degree in control science and engineering from Hunan University, Changsha, China, in 2016, and the M.E. degree in control science and engineering from National University of Defense Technology, Changsha, China, in 2018, from where he is currently pursuing the Ph.D. degree in control science and engineering. His research interests include unsupervised learning and functional parcellation of the brain.

Yi Zhang is pursuing his P.H.D degree in National University of Defense Technology (NUDT), China. His current research interests include kernel learning and unsupervised multi-view learning. He has published several peer-reviewed papers in conferences and journals such as ICML, ICCV, ACM MM, AAAI, IEEE TIP, ACM TOMM, etc.

Dewen Hu received the B.S. and M.S. degrees from Xi'an Jiaotong University, Xi'an, China, in 1983 and 1986, respectively, and the Ph.D. degree from the National University of Defense Technology, Changsha, China, in 1999. Since 1986, he has been with the National University of Defense Technology. From 1995 to 1996, he was a Visiting Scholar with the University of Sheffield, Sheffield, U.K. He was promoted as a Professor in 1996. His research interests include image processing, system identification and control, neural networks, and cognitive science. Dr. Hu is an Action Editor of Neural Networks.