Adaptive Feature-weighted Co-clustering with Local Coordinate Coding

Yiyan Wang, Bin Yang, Zhaohu Liu, Junhua Li, *Senior Member, IEEE* and Yong Peng, *Senior Member, IEEE*

***Abstract*—Co-clustering enables the simultaneous clustering of features and samples by exploiting their associations. Based on the commonly used matrix tri-factorization objective function in co-clustering, we propose an adaptive feature-weighted co-clustering with local coordinate coding (AFC-LCC) model in this paper, by making two improvements to enhance the data clustering performance. On one hand, a local coordinate constraint is introduced to enforce the smoothness of the sample cluster indicators along the scaled feature cluster spaces; on the other hand, a quantitative measurement is incorporated to adaptively learn the feature contributions in co-clustering for model discriminative ability enhancement. By co-optimizing the involved variables in the AFC-LCC model objective function, the experimental results not only show competitive clustering performance in comparison with related clustering models, but also depict the rationality and effectiveness of the local coordinate constraint and the feature importance descriptor.**

***Index Terms*—Adaptive feature weighting, co-clustering, local coordinate coding, probabilistic feature cluster indicator.**

I. Introduction

CO-CLUSTERING achieves simultaneous clustering of features and samples which respectively correspond to the rows and columns of a data matrix, aiming to mine bidirectional associations and reveal data structures with feature semantics [1]. Co-clustering appears in many fields such as the topic clustering and keyword extraction via word-document matrix partitioning [2], [3], the optimization of recommendation systems through personalized strategies based on user-item interaction matrices [4], [5], the identification of disease-related gene modules from expression data [6], [7], image identification [8] and target detection [9], while clustering for high-dimensional data remains an active research topic [10].

In recent years, significant progress has been made in co-clustering research. As a variant of the classic bipartite spectral graph partitioning (BSGP) method [11], a weighted bilateral k-means algorithm was proposed by reformulating the bipartite graph clustering problem as a non-negative matrix factorization task to achieve co-clustering, significantly enhancing its computational efficiency [12]. This method was further extended to multi-view scenario by leveraging the fast matrix factorization techniques to adaptively assign weights to different feature views, which effectively addresses the heterogeneity issues in multi-view data [13]. In [14], a fast bipartite graph model was developed to handle sparse data through co-occurrence pattern mining. To improve the robustness and structure-preserving ability of clustering models, Lin *et al.* introduced a robust model fitting-based loss function in bipartite graph-based co-clustering [15], while collaborative structure learning strategies have also been explored in fuzzy clustering [16]. As a semi-supervised co-clustering extension, a sparse neighbor-constrained framework was proposed where prior information was integrated with regularization techniques to provide new insights for semi-supervised data clustering [17]. For graph data analysis, a joint NMF and network embedding framework was proposed for graph co-clustering. To handle large-scale data sets, Xie *et al.* proposed an anchor-guided label propagation algorithm that utilizes sparse anchor sets to reduce the computational burden in clustering high-dimensional data, which has proven effective in processing massive data sets [18].

In many cases, the co-clustering objective takes the form of matrix tri-factorization, i.e., decomposing the data matrix into the product of a feature cluster indicator matrix, a scaling matrix, and a sample cluster indicator matrix [19], [20]. Therefore, it is desirable to characterize the correlations among these three factor matrices in a unified manner. Since the product of the first two matrices and the last one can be viewed as the dictionary and coding matrices, respectively, the Local Coordinate Coding (LCC) principle [21] naturally characterizes their interactions while preserving local structures. To this end, we propose the Adaptive Feature-weighted Co-clustering with Local Coordinate Coding (AFC-LCC) model. AFC-LCC not only constrains the learned sample cluster indicators to be smoother along the dictionary space (defined by the product of feature cluster indicator and scaling matrices) by following the local invariance data property, but also incorporate a feature importance descriptor to achieve quantitative measurement of features for enhancing the model discriminative ability.

II. Methodology

To facilitate the following model formulation and derivations, we summarize the main notations in Table I.

TABLE I
The main notations used in this paper.

Not.	Description	Not.	Description
$\mathbf{X}$	data matrix	$\mathbf{P}$	feature cluster indicator matrix
$\mathbf{S}$	scaling matrix	$\mathbf{Q}$	sample cluster indicator matrix
d	feature dimension	c	number of clusters
n	number of samples	$\boldsymbol{\theta}$	feature importance descriptor

This work was supported in part by the ‘Pioneer’ and ‘Leading Goose’ R&D Program of Zhejiang (2025C04001), the National Natural Science Foundation of China (62571171), and the MoE Key Laboratory of Embedded System and Services Computing (ESSCKF2024-11) (*Corresponding author: Yong Peng*).

Y. Wang, B. Yang, Z. Liu, and Y. Peng are with School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, China (email: yongpeng@hdu.edu.cn).

J. Li is with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ Colchester, U.K.

A. Motivation and Formulation

By replacing the bipartitioning normalized cuts in BSGP [11] with multi-partitioning normalized cuts, the Bilateral k-means (BKM) [22] performs clustering on data collection matrix $\mathbf{X} \in \mathbb{R}^{d \times n}$ by optimizing the following objective, i.e.,

$$\min_{\mathbf{P} \in \text{Ind}^{d \times c}, \mathbf{Q} \in \text{Ind}^{n \times c}, \mathbf{S} \in \text{diag}^{c \times c}} \|\mathbf{X} - \mathbf{PSQ}^T\|_F^2, \tag{1}$$

where $\mathbf{P}$ and $\mathbf{Q}$ are the feature and cluster indicator matrices, respectively. $\mathbf{S}$ is a diagonal scaling matrix in-between $\mathbf{P}$ and $\mathbf{Q}$. (1) can be regarded as a mathematical approximation of the data matrix using the product of three factor matrices. If $\mathbf{PS}$ is regarded as a whole, it can be understood as a set of basis vectors, while $\mathbf{Q}$ serves as the coefficient matrix.

Motivated by the LCC principle that encourages larger coding coefficients for nearby basis vectors and preserves local geometric structures [21], we introduce the following local coordinate constraint

$$\min_{\mathbf{P},\mathbf{Q},\mathbf{S}} \sum\nolimits_{i=1}^{n} \sum\nolimits_{j=1}^{c} q_{ij} \|\mathbf{x}_i - (\mathbf{PS})_j\|_2^2. \tag{2}$$

To capture the fine-grained relationships among these factor matrices, we first relax $\mathbf{S}$ from diagonal to regular to provide additional flexibility for modeling the interactions between feature clusters and sample clusters, treat $\mathbf{PS}$ as the basis, and incorporate LCC into (1) to enforce the local invariance property of data. That is, the closer of a certain sample to a specific column of the basis, the larger the corresponding element of the coefficient vector. Additionally, we introduce a feature importance descriptor $\boldsymbol{\theta}$ (which satisfies $\boldsymbol{\theta} \geq \mathbf{0}$ and $\boldsymbol{\theta}^T \mathbf{1} = 1$) to quantify the contribution of each feature to the clustering process, where larger values indicate stronger discriminative ability [23]. The final objective function of our proposed AFC-LCC model is formulated as

$$\begin{aligned} \min_{\mathbf{P},\mathbf{Q},\mathbf{S},\boldsymbol{\Theta}} & \|\boldsymbol{\Theta}\mathbf{X} - \mathbf{PSQ}^T\|_F^2 + \alpha \sum_{i,j=1}^{n,c} q_{ij} \|\boldsymbol{\Theta}\mathbf{x}_i - (\mathbf{PS})_j\|_2^2, \\ & s.t.\ \mathcal{C}(\boldsymbol{\Theta}), \mathbf{P} \geq 0, \mathbf{P1} = \mathbf{1}, \mathbf{Q} \in \text{Ind}^{n \times c}, \end{aligned} \tag{3}$$

where $\mathcal{C}(\boldsymbol{\Theta}) \triangleq \{\boldsymbol{\Theta} = \text{diag}(\boldsymbol{\theta}), \boldsymbol{\theta}^T \mathbf{1} = 1, \boldsymbol{\theta} \geq \mathbf{0}\}$ and we relax $\mathbf{P}$ from one-hot to probabilistic encoding to improve its flexibility. The regularizer in (3) emphasizes that q_{ij} could be larger if $\mathbf{x}_i$ is closer to the j-th column of $\mathbf{PS}$, i.e., $(\mathbf{PS})_j$.

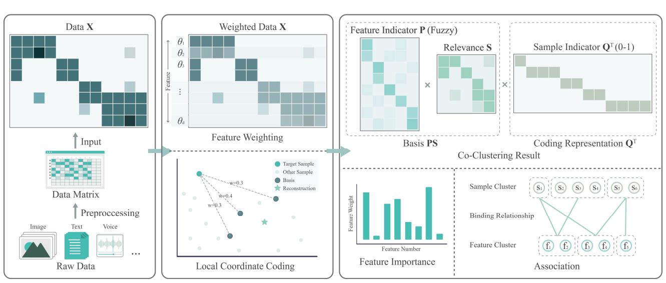

Fig. 1. The general framework of the proposed AFC-LCC model.

Fig. 1 illustrates the overall framework of AFC-LCC. The primary idea is the introduction of a feature weighting diagonal matrix $\boldsymbol{\Theta}$ to adaptively weight the data matrix $\mathbf{X}$, and decompose the weighted matrix $\boldsymbol{\Theta}\mathbf{X}$ into three factor matrices, *i.e.*, probabilistic feature cluster indicator matrix $\mathbf{P}$, scaling matrix $\mathbf{S}$, and sample cluster indicator matrix $\mathbf{Q}$, where their coupling relationship is characterized by the LCC-based regularizer. Finally, based on the learned model, feature importance weights and co-clustering results can be obtained, which support various post-analyses such as the correlation analysis between feature clusters and sample clusters.

B. Optimization

Each variable in (3) can be updated by fixing the others.

■ Update $\mathbf{Q}$. Denoting $a_{ij} \triangleq \|\boldsymbol{\Theta}\mathbf{x}_i - (\mathbf{PS})_j\|_2^2$ as the (i,j)-th element of $\mathbf{A} \in \mathbb{R}^{n \times c}$, we have

$$\min_{\mathbf{Q} \in \text{Ind}^{n \times c}} \|\boldsymbol{\Theta}\mathbf{X} - \mathbf{PSQ}^T\|_F^2 + \alpha \text{Tr}(\mathbf{AQ}^T). \tag{4}$$

By expanding the first term into its trace form, we get

$$\min_{\mathbf{Q} \in \text{Ind}^{n \times c}} \text{Tr}(\mathbf{S}^T\mathbf{P}^T\mathbf{PSQ}^T\mathbf{Q} - 2(\mathbf{S}^T\mathbf{P}^T\boldsymbol{\Theta}\mathbf{X} - \frac{\alpha}{2}\mathbf{A}^T)\mathbf{Q}). \tag{5}$$

Then, we convert it into the perfect square form as

$$\min_{\mathbf{Q} \in \text{Ind}^{n \times c}} \|\mathbf{B}^{\frac{1}{2}}(\mathbf{B}^{-1}\mathbf{C} - \mathbf{Q}^T)\|_F^2 + \text{Const}, \tag{6}$$

where $\mathbf{B} \triangleq \mathbf{S}^T\mathbf{P}^T\mathbf{PS}$ and $\mathbf{C} \triangleq \mathbf{S}^T\mathbf{P}^T\boldsymbol{\Theta}\mathbf{X} - \frac{\alpha}{2}\mathbf{A}^T$. It is observed that (6) is independent for each row in $\mathbf{Q}$; then, the optimization to $\mathbf{Q}$ can be decomposed into n independent subproblems, and the $j|_{j=1}^{n}$-th one is expressed as

$$\min_{\mathbf{Q} \in \text{Ind}^{n \times c}} \|(\mathbf{B}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{C})_j - \mathbf{B}^{\frac{1}{2}}\mathbf{q}^{jT}\|_2^2, \tag{7}$$

where $(\mathbf{B}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{C})_j$ denotes its j-th column, and $\mathbf{q}^{jT}$ denotes the transpose of the j-th row of $\mathbf{Q}$. We find that only one element of $\mathbf{q}^j$ is one and all the others are zeros, leading to the following coordinate descent-based updating rule

$$q_{ij} = \begin{cases} 1, & i = \arg\min_k \|(\mathbf{B}^{\frac{1}{2}}\mathbf{B}^{-1}\mathbf{C})_j - (\mathbf{B}^{\frac{1}{2}})_k\|_2^2, \\ 0, & \text{otherwise}. \end{cases} \tag{8}$$

■ Update $\mathbf{P}$. We rewrite the objective function (3) as

$$\min_{\mathbf{P} \geq 0, \mathbf{P1} = \mathbf{1}} \|\boldsymbol{\Theta}\mathbf{X} - \mathbf{PSQ}^T\|_F^2 + \alpha \sum_{i=1}^{n} \|\left(\boldsymbol{\Theta}\mathbf{x}_i\mathbf{1}^T - \mathbf{PS}\right)\mathbf{V}_i^{\frac{1}{2}}\|_2^2, \tag{9}$$

where $\mathbf{V}_i = \text{diag}(\mathbf{q}^i)$ is a diagonal matrix and $\mathbf{q}^i$ is the i-th row of the indicator matrix $\mathbf{Q}$. Equivalently, we obtain

$$\begin{aligned} \min_{\mathbf{P} \geq 0, \mathbf{P1} = \mathbf{1}} & \text{Tr}(\mathbf{P}^T\mathbf{PSQ}^T\mathbf{QS}^T + \alpha \sum\nolimits_{i=1}^{n} \mathbf{P}^T\mathbf{PSV}_i\mathbf{S}^T \\ & -2\mathbf{P}^T\boldsymbol{\Theta}\mathbf{XQS}^T - 2\alpha \sum\nolimits_{i=1}^{n} \mathbf{P}^T\boldsymbol{\Theta}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i\mathbf{S}^T). \end{aligned} \tag{10}$$

By denoting $\mathbf{D} \triangleq \mathbf{SQ}^T\mathbf{QS}^T$, $\mathbf{E} \triangleq \boldsymbol{\Theta}\mathbf{XQS}^T$, $\mathbf{F} \triangleq \sum_{i=1}^{n} \mathbf{SV}_i\mathbf{S}^T$, and $\mathbf{G} \triangleq \sum_{i=1}^{n} \boldsymbol{\Theta}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i\mathbf{S}^T$, we have

$$\min_{\mathbf{P} \geq 0, \mathbf{P1} = \mathbf{1}} \text{Tr}(\mathbf{P}^T\mathbf{P}\left(\mathbf{D} + \alpha\mathbf{F}\right) - 2\mathbf{P}^T\left(\mathbf{E} + \alpha\mathbf{G}\right)). \tag{11}$$

Then, by completing the square form of (11), we have

$$\min_{\mathbf{P} \geq 0, \mathbf{P1} = \mathbf{1}} \|\left(\mathbf{P} - \mathbf{H}\right)\mathbf{J}^{\frac{1}{2}}\|_F^2 + \text{Const}, \tag{12}$$

where $\mathbf{H} \triangleq \left(\mathbf{E} + \alpha\mathbf{G}\right)\left(\mathbf{D} + \alpha\mathbf{F}\right)^{-1}, \mathbf{J} \triangleq \mathbf{D} + \alpha\mathbf{F}$. Likewise, each row of $\mathbf{P}$ is also an independent subproblem; and for the i-th row of $\mathbf{P}$, i.e., $\mathbf{p}^i$, we aim to solve

$$\min_{\mathbf{p}^i \geq 0, \mathbf{p}^i\mathbf{1} = 1} \mathbf{p}^i\mathbf{J}\mathbf{p}^{iT} - 2\mathbf{p}^i\mathbf{J}\mathbf{h}^{iT}, \tag{13}$$

which is a quadratic optimization problem with simplex constraints, and can be effectively solved by the Lagrange multiplier method as described in [24].

■ Update $\mathbf{S}$. The objective function is the same as (9) but with $\mathbf{S}$ as variable. Let $\mathcal{J}(\mathbf{S})$ denote (9) and then we compute the derivative of $\mathcal{J}(\mathbf{S})$ with respect to $\mathbf{S}$ as

$$\begin{aligned}\frac{\partial \mathcal{J}(\mathbf{S})}{\partial \mathbf{S}} &= 2\mathbf{P}^T\mathbf{P}\mathbf{S}\mathbf{Q}^T\mathbf{Q} - 2\mathbf{P}^T\mathbf{\Theta}\mathbf{X}\mathbf{Q} \\ &+ \alpha \sum\nolimits_{i=1}^{n} \left(2\mathbf{P}^T\mathbf{P}\mathbf{S}\mathbf{V}_i - 2\mathbf{P}^T\mathbf{\Theta}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i^T\right).\end{aligned} \tag{14}$$

Upon setting the resulting formula (14) to zero, we get

$$s_{ij} = \frac{(\mathbf{P}^T\mathbf{\Theta}\mathbf{X}\mathbf{Q} + \alpha\sum_{i=1}^{n}\mathbf{P}^T\mathbf{\Theta}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i^T)_{ij}}{(\mathbf{P}^T\mathbf{P})_{ii}(\mathbf{Q}^T\mathbf{Q} + \alpha\sum_{i=1}^{n}\mathbf{V}_i)_{jj}}. \tag{15}$$

■ Update $\mathbf{\Theta}$. The objective function can be rewritten as

$$\min_{\mathcal{C}(\mathbf{\Theta})} \|\mathbf{\Theta}\mathbf{X} - \mathbf{P}\mathbf{S}\mathbf{Q}^T\|_F^2 + \alpha \sum\nolimits_{i=1}^{n} \|\left(\mathbf{\Theta}\mathbf{x}_i\mathbf{1}^T - \mathbf{P}\mathbf{S}\right)\mathbf{V}_i^{\frac{1}{2}}\|_2^2, \tag{16}$$

which is equivalent to

$$\begin{aligned}\min_{\mathcal{C}(\mathbf{\Theta})} &\mathrm{Tr}(\mathbf{\Theta}\mathbf{\Theta}\mathbf{X}\mathbf{X}^T + \alpha\mathbf{\Theta}\mathbf{\Theta}\sum\nolimits_{i=1}^{n}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i\mathbf{1}\mathbf{x}_i^T \\ &-2\mathbf{X}\mathbf{Q}\mathbf{S}^T\mathbf{P}^T\mathbf{\Theta} - 2\alpha\sum\nolimits_{i=1}^{n}\mathbf{P}\mathbf{S}\mathbf{V}_i\mathbf{1}\mathbf{x}_i^T\mathbf{\Theta}).\end{aligned} \tag{17}$$

Setting $\mathbf{K} \triangleq \sum_{i=1}^{n}\mathbf{x}_i\mathbf{1}^T\mathbf{V}_i\mathbf{1}\mathbf{x}_i^T$ and $\mathbf{L} \triangleq \sum_{i=1}^{n}\mathbf{P}\mathbf{S}\mathbf{V}_i\mathbf{1}\mathbf{x}_i^T$, the above equation is equivalent to

$$\min_{\boldsymbol{\theta}\geq\mathbf{0},\boldsymbol{\theta}^T\mathbf{1}=1} \boldsymbol{\theta}^T((\mathbf{X}\mathbf{X}^T + \alpha\mathbf{K}))\boldsymbol{\theta} - \boldsymbol{\theta}^T\mathbf{m}, \tag{18}$$

where $\mathbf{m} = 2\mathrm{diag}(\mathbf{X}\mathbf{Q}\mathbf{S}^T\mathbf{P}^T + \alpha\mathbf{L})$, and (18) can be solved similarly as (13).

The above updating procedure is summarized in Alg. 1.

Algorithm 1 Optimization of AFC-LCC objective (3)

Input: Data $\mathbf{X} \in \mathbb{R}^{d\times n}$, # clusters c and parameter α;
Output: Feature cluster indicator matrix $\mathbf{P}$ and sample cluster indicator matrix $\mathbf{Q}$.

Initialize $\boldsymbol{\theta} = [\frac{1}{d}, \cdots, \frac{1}{d}]^T$, generate $\mathbf{P}$ and $\mathbf{Q}$ randomly;
while not converged **do**
  Update feature cluster indicator $\mathbf{P}$ by solving (13);
  Update diagonal scaling matrix $\mathbf{S}$ by solving (15);
  Update sample cluster indicator $\mathbf{Q}$ by equation (8);
  Update feature importance descriptor θ by solving (18);
end while

III. Experiments

A. Data and Experimental Setup

The following experiments were conducted on three benchmark datasets, ORL (400 samples, 4096 dimensions, 40 clusters), COIL20 (1440 samples, 1024 dimensions, 20 clusters), and MSRA (1799 samples, 256 dimensions, 12 clusters). Since the MSRA dataset exhibits inter-class sample imbalance, two versions are adopted including the original full version to simulate real imbalanced scenarios, and a balanced version constructed by randomly selecting 64 samples from each cluster to evaluate the robustness of AFC-LCC under different class distribution scenarios.

To demonstrate the effectiveness of AFC-LCC, we compared it with several relevant methods including K-means, fuzzy c-means (FCM), BKM, and three degraded versions of AFC-LCC, namely AFC (without the LCC regularization), C_P-LCC (excluding feature-weighted learning), and C_P (without LCC or feature weighting). The involved parameters in respective models were tuned within the range $\{10^{-5}, 10^{-4}, \cdots, 10^{1}\}$. The fuzzy order index in FCM is searched from [1.1, 2] with a step size 0.1. For each data set, different randomly chosen clusters were used for performance evaluation but they were kept the same for these models.

B. Results and Analysis

Table II lists the clustering results of all methods in terms of three metrics, i.e., accuracy (Acc), normalized mutual information (NMI), and purity [25]. Our AFC-LCC achieves the best performance in most cases which is marked in bold.

The compared models exhibit a clear performance hierarchy. AFC-LCC achieves the best performance on most datasets and metrics. Traditional clustering methods (k-means, FCM, BKM) perform relatively weakly, basic fuzzy variants (C_P, AFC) yield moderate results, while models integrated with local constraints (C_P-LCC, AFC-LCC) achieve significant improvements over the basic versions, fully verifying the effectiveness of the proposed model.

The integration of local invariance constraints and adaptive feature weighting greatly boosts clustering performance. Local constraints effectively capture the local neighborhood structure of data and reduce over-reliance on global distribution, while adaptive weighting flexibly handles noisy and outlier features in data. Their combination achieves both global optimization and local structure preservation, which is the primary reason for the superior performance of AFC-LCC.

The performance ranking of all models is highly consistent across different datasets. AFC-LCC remains optimal not only on standard datasets but also on the balanced MSRA dataset, showing strong generalization ability to diverse data distributions. Moreover, it exhibits significantly smaller performance fluctuations when the cluster number c deviates from the ground truth, demonstrating stronger robustness and higher practical value in real-world applications.

C. Discussions

Fig. 2. The original images from ORL (left), and basis vectors learned by AFC (middle) and AFC -LCC (right).

1) Comparison of the Learned Basis: To explore the role of local coordinate encoding, we randomly selected one face image from four subjects in ORL. Fig. 2 shows the original face images, the basis vectors learned respectively by AFC and AFC-LCC. It is obvious that the basis vectors learned with the LCC strategy resemble more to the original face

TABLE II
CLUSTERING PERFORMANCE OF DIFFERENT METHODS ON THE BENCHMARK DATA SETS.

	metrics	Acc (%)						NMI (%)						Purity (%)					
ORL	c	8	16	24	32	40	Avg.	8	16	24	32	40	Avg.	8	16	24	32	40	Avg
	k-means	65.00	55.63	51.25	49.38	50.00	54.25	63.41	65.29	65.12	69.42	72.35	67.12	68.75	60.00	53.33	55.94	55.25	58.65
	FCM	46.25	46.25	49.58	49.06	45.50	47.33	32.83	53.45	61.41	63.71	64.00	55.08	46.25	47.50	51.67	49.38	47.75	48.51
	BKM	46.25	51.25	52.50	46.88	43.50	48.08	48.48	64.11	68.07	65.14	65.62	62.28	52.50	53.75	56.25	48.44	47.25	51.64
	C_P	58.75	54.37	52.50	50.63	46.00	52.45	59.36	65.34	68.38	70.06	68.60	66.35	62.50	56.25	54.17	54.38	49.50	55.36
	C_P-LCC	67.50	56.25	57.50	54.38	53.00	57.73	66.18	67.85	72.22	71.67	72.35	70.05	67.50	58.75	59.58	56.56	56.25	59.73
	AFC	61.25	58.13	52.50	51.56	46.75	54.04	58.92	67.40	68.60	70.55	70.72	67.24	61.25	60.62	55.42	55.31	53.00	57.12
	AFC-LCC	**72.50**	**60.00**	**58.33**	**55.00**	**54.50**	**60.07**	**71.58**	**68.26**	**72.47**	**72.17**	**72.93**	**71.48**	**72.50**	**61.88**	**62.08**	**57.19**	**57.00**	**62.13**
COIL20	c	4	8	12	16	20	Avg.	4	8	12	16	20	Avg.	4	8	12	16	20	Avg
	k-means	67.01	55.90	42.48	60.07	53.68	55.83	52.71	57.06	59.16	74.46	74.94	63.67	68.40	60.24	49.19	66.23	61.18	61.05
	FCM	66.67	51.22	45.02	52.43	42.92	51.65	46.92	47.05	48.38	57.45	54.42	50.84	66.67	52.26	45.02	52.43	42.92	51.86
	BKM	72.92	49.48	58.75	52.58	56.46	58.04	59.34	57.86	64.90	67.89	75.69	65.14	72.92	56.08	63.61	53.77	63.26	61.93
	C_P	77.78	41.32	45.14	51.65	53.40	53.86	61.74	44.45	52.14	61.34	61.33	56.20	75.83	43.23	51.62	59.46	58.19	57.67
	C_P-LCC	81.60	57.12	63.77	61.55	58.89	64.59	65.33	56.71	70.91	74.60	75.34	68.58	81.60	57.99	68.98	63.63	63.75	67.19
	AFC	80.21	54.51	51.27	55.03	54.65	59.13	64.30	54.48	63.74	70.17	72.53	65.04	80.21	54.69	55.44	58.85	61.60	62.16
	AFC-LCC	**87.85**	**58.51**	**67.25**	**69.18**	**59.44**	**68.45**	**78.24**	**58.52**	**72.64**	**78.99**	**76.04**	**72.89**	**87.85**	**61.46**	**69.21**	**71.70**	**64.10**	**70.86**
MSRA	c	3	5	7	9	12	Avg.	3	5	7	9	12	Avg.	3	5	7	9	12	Avg
	k-means	64.03	58.56	51.14	51.25	48.30	54.66	47.24	55.02	57.72	57.75	61.84	55.91	64.03	58.56	56.64	54.88	53.31	57.48
	FCM	64.25	60.05	42.05	52.31	45.53	52.84	47.27	46.84	33.62	48.07	41.26	43.41	64.25	60.06	42.05	53.31	45.80	53.09
	BKM	59.05	61.04	42.92	50.60	41.52	51.03	37.61	49.46	42.08	51.00	47.92	45.61	59.05	61.04	48.51	53.88	47.86	54.07
	C_P	64.25	58.44	55.77	53.10	46.30	55.57	48.47	54.46	55.88	58.36	55.24	54.48	64.25	58.44	58.65	56.94	51.36	57.93
	C_P-LCC	82.13	64.27	62.59	61.28	54.70	64.99	60.69	59.16	59.93	65.56	60.02	61.07	82.13	64.27	62.76	61.42	56.20	65.36
	AFC	73.53	60.42	59.62	54.73	54.86	60.63	45.43	41.37	55.12	57.19	56.30	51.08	73.53	60.42	59.62	55.16	56.09	60.96
	AFC-LCC	**86.07**	**67.87**	**64.16**	**63.20**	**57.31**	**67.72**	**67.31**	**59.72**	**62.14**	**66.81**	**66.88**	**64.57**	**85.07**	**67.87**	**64.16**	**63.20**	**60.64**	**68.19**
MSRA (Balanced)	c	3	5	7	9	12	Avg.	3	5	7	9	12	Avg.	3	5	7	9	12	Avg
	k-means	99.48	99.38	80.13	85.07	74.74	87.76	97.55	97.99	85.49	89.07	86.34	91.29	99.48	99.38	85.27	86.28	82.55	90.59
	FCM	99.48	93.44	91.52	85.24	62.76	86.49	97.55	89.32	86.82	83.65	63.46	84.16	99.48	93.44	91.52	86.11	63.15	86.74
	BKM	99.48	80.94	86.16	74.13	74.22	82.99	97.55	81.82	81.74	74.81	79.13	83.01	99.48	80.94	86.16	76.22	80.08	84.58
	C_P	99.48	81.25	80.58	81.47	88.15	86.19	97.55	81.85	84.80	86.73	91.62	88.51	99.48	81.25	85.04	85.27	90.76	88.36
	C_P-LCC	99.48	99.38	94.87	87.05	88.54	93.86	97.55	97.99	91.88	87.90	92.43	93.55	99.48	99.38	94.87	87.05	91.15	94.39
	AFC	99.48	99.38	94.20	85.59	91.80	94.09	97.55	97.99	91.17	90.38	92.23	93.86	99.48	99.38	94.20	88.37	91.80	94.65
	AFC-LCC	**99.48**	**99.69**	**96.65**	**96.35**	**91.93**	**96.82**	**97.55**	**99.00**	**94.76**	**94.49**	**92.65**	**95.69**	**99.48**	**99.69**	**94.20**	**96.35**	**91.93**	**96.33**

images, demonstrating that effectively characterizing the local invariance between data points and their associated basis vectors helps preserve more details and therefore improves the clustering performance.

2) Rationality of the Learned Feature Importance: To verify the rationality of the feature importance learned by AFC-LCC, we conduct experiments on the Yale64 face dataset, which contains 165 face images from 15 subjects, and each is represented by a 4096-dimensional feature vector. We reshape the original features into an image structure, rank the learned feature importance, and visualize the top 60, 120, and 180 important features respectively. As shown in Fig. 3, the high-weight regions are mainly concentrated around the eyes, nose, and mouth, which are the most discriminative areas for face characterization. The results verify that AFC-LCC assigns larger weights to discriminative features.

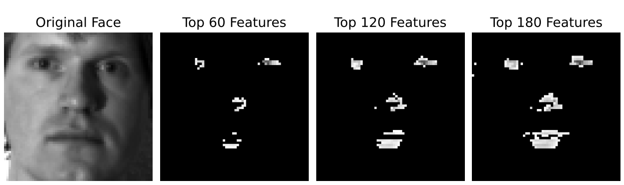

Fig. 3. Visualization of different numbers of selected features on Yale64.

3) Convergence and Parameter Sensitivity: Fig. 4 presents the objective function convergence curves, accuracy curves (left) and hyper-parameter α sensitivity (right) of the AFC-LCC method on MSRA. As the number of iterations increases, the objective function value decreases monotonically while the accuracy gradually stabilizes, demonstrating the excellent convergence of the proposed optimization algorithm. The right subfigure illustrates the clustering performance with varying hyper-parameter α. It can be observed that our model prefers small values of α and achieves the optimal performance at 10^{-2}.

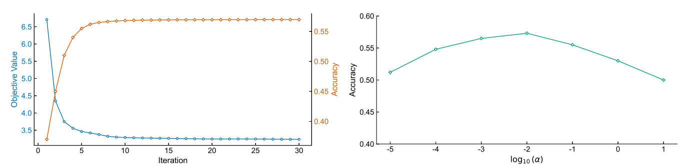

Fig. 4. Convergence of accuracy and objective value (left) and parameter sensitivity of α (right) on MSRA.

IV. CONCLUSION

This paper proposed AFC-LCC to improve co-clustering performance from two aspects. One was the introduction of the local coordinate constraint to enforce the local invariance data property by explicitly characterizing the underlying semantic connection among the three factor matrices in co-clustering; the other was the incorporation of feature importance descriptor to adaptively quantify the feature contributions in co-clustering for model discriminative ability enhancement. Experimental results demonstrated the competitive performance of AFC-LCC in data clustering; besides, the rationality and effectiveness of the local coordinate constraint and feature importance descriptor were illustrated by intuitive results. Future work will focus on extending AFC-LCC to multi-view and semi-supervised clustering scenarios.

References

[1] E. Battaglia, F. Peiretti, and R. G. Pensa, "Co-clustering: A survey of the main methods, recent trends, and open problems," *ACM Computing Surveys*, vol. 57, no. 2, pp. 1–33, 2024.

[2] Y. Chen, Z. Lei, Y. Rao, H. Xie, F. L. Wang, J. Yin, and Q. Li, "Parallel non-negative matrix tri-factorization for text data co-clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 5, pp. 5132–5146, 2023.

[3] B. Amiri and R. Karimianghadim, "A novel text clustering model based on topic modelling and social network analysis," *Chaos Solitons & Fractals*, vol. 181, p. 114633, 2024.

[4] X. He, S. Liu, J. Keung, and J. He, "Co-clustering for federated recommender system," in *Proceedings of the ACM Web Conference*, 2024, pp. 3821–3832.

[5] M. Rezghi and E. Baratnezhad, "A novel fuzzy co-clustering method for recommender systems via inverse stereographic NMF," *Expert Systems with Applications*, vol. 259, p. 125301, 2025.

[6] W. Yang, H. Wang, Y. Zhang, Z. Luo, and T. Li, "A nondominated sorting genetic model for co-clustering," *Information Sciences*, vol. 667, p. 120459, 2024.

[7] X. Chen, J. Z. Huang, Q. Wu, and M. Yang, "Subspace weighting co-clustering of gene expression data," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 16, no. 2, pp. 352–364, 2019.

[8] F. Nie, S. Pei, R. Wang, and X. Li, "Fast clustering with co-clustering via discrete non-negative matrix factorization for image identification," in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2020, pp. 2073–2077.

[9] Z. Pang, J. Guo, W. Sun, and S. Li, "Biclustering collaborative learning for cross-domain person re-identification," *IEEE Signal Processing Letters*, vol. 28, pp. 2142–2146, 2021.

[10] B. Jiang, Z. Wang, J. Yang, G.-K. Xu, W. Chen, C. Zhang, X. Liang, P. Zhou, W. Sheng, and W. Ding, "Self-enhanced density clustering for high dimension and low sample size data," in *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2026, p. 508–519.

[11] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2001, pp. 269–274.

[12] K. Song, X. Yao, F. Nie, X. Li, and M. Xu, "Weighted bilateral k-means algorithm for fast co-clustering and fast spectral clustering," *Pattern Recognition*, vol. 109, p. 107560, 2021.

[13] F. Nie, S. Shi, and X. Li, "Auto-weighted multi-view co-clustering via fast matrix factorization," *Pattern Recognition*, vol. 102, p. 107207, 2020.

[14] W. Chen, H. Wang, Z. Long, and T. Li, "Fast flexible bipartite graph model for co-clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 7, pp. 6930–6940, 2023.

[15] S. Lin, H. Luo, Y. Yan, G. Xiao, and H. Wang, "Co-clustering on bipartite graphs for robust model fitting," *IEEE Transactions on Image Processing*, vol. 31, pp. 6605–6620, 2022.

[16] B. Jiang, C. Zhang, Z. Wang, X. Liang, P. Zhou, L. Du, Q. Zhang, W. Ding, and Y. Liu, "Scalable fuzzy clustering with collaborative structure learning and preservation," *IEEE Transactions on Fuzzy Systems*, vol. 33, no. 9, pp. 3047–3060, 2025.

[17] X. Li, X. Lu, and X. Fan, "Semi-supervised sparse neighbor constrained co-clustering with dissimilarity and similarity regularization," *Engneering Applications of Artificial Intelligence*, vol. 114, p. 104989, 2022.

[18] F. Xie, F. Nie, W. Yu, and X. Li, "Fast co-clustering via anchor-guided label spreading," *Neural Networks*, vol. 185, p. 107187, 2025.

[19] Z. Lu, S. Wang, G. Liu, and F. Nie, "Robust weighted co-clustering with global and local discrimination," *Pattern Recognition*, vol. 138, p. 109405, 2023.

[20] C. Liu, C. Chang, and F. Nie, "Self-weighted multi-view fuzzy clustering with multiple graph learning," *IEEE Signal Processing Letters*, vol. 32, pp. 1585–1589, 2025.

[21] K. Yu, T. Zhang, and Y. Gong, "Nonlinear learning using local coordinate coding," in *Proceedings of Advances in Neural Information Processing Systems*, 2009, pp. 2233–2241.

[22] J. Han, K. Song, F. Nie, and X. Li, "Bilateral k-means algorithm for fast co-clustering," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 31, no. 1, 2017.

[23] Y. Wang, M. Jin, Y. Chen, Y. Peng, Z. Yang, F. Nie, A. Cichocki, and W. Kong, "Effective feature-sample co-clustering by adaptive feature-sample co-weighting," *Information Sciences*, vol. 718, p. 122427, 2025.

[24] Y. Peng, X. Zhu, F. Nie, W. Kong, and Y. Ge, "Fuzzy graph clustering," *Information Sciences*, vol. 571, pp. 38–49, 2021.

[25] Y. Peng, W. Huang, W. Kong, F. Nie, and B.-L. Lu, "JGSED: An end-to-end spectral clustering model for joint graph construction, spectral embedding and discretization," *IEEE Transactions on Emerging Topics on Computational Intelligence*, vol. 7, no. 6, pp. 1687–1701, 2023.