Simple yet Effective Incomplete Multi-view Clustering: Similarity-level Imputation and Intra-view Hybrid-group Prototype Construction

Shengju Yu · Zhibin Dong · Siwei Wang · Pei Zhang · Yi Zhang · XinwangLiu · Thomas Guan · Tiejun Li · Yiu-ming Cheung

School of Computer Science

National University of Defense Technology

Motivation

Most of IMVC methods (1) only utilize observed unpaired samples to construct bipartite similarity (2) employ a single quantity of prototypes to extract the information of all views.

- Miss out latent useful information from missing samples, resulting in the generated similarity not that accurate & Lead to unbalanced cluster distribution and deteriorate the graph structure.
- Not competent to adequately characterize all views, accordingly weakening the view information diversity.

Our solution

- Transform partial bipartition learning under prototype orthogonality into the form containing original samples by utilizing the data reconstruction concept to split out of observed similarity.
- Relax conventional non-negative constraints through a sample regularization skill to make the measure of similarity freer.
- Introduce the learnable consensus graph which is shared for all views to provide unified structure.
- Built the connection between all view-specific bipartition similarities and the consensus graph.
- Gather the information from other views at the similarity level to assist imputing the incomplete parts of similarity on each view.
- Associate a group of hybrid prototype quantities for each individual view so that it can flexibly exploit features according to the characteristics of each view.

Motivation

Most of IMVC methods (1) choose to ignore the missing samples and only utilize observed unpaired samples to construct bipartite similarity; (2) employ a single quantity of prototypes to extract the information of all views.

- Miss out latent useful information from missing samples, resulting in the generated similarity not that accurate & Lead to unbalanced cluster distribution and deteriorate the graph structure.
- Not competent to adequately characterize all views, accordingly weakening the view information diversity.

Our solution

- Transform partial bipartition learning under prototype orthogonality into the form containing original samples by utilizing the data reconstruction concept to split out of observed similarity.

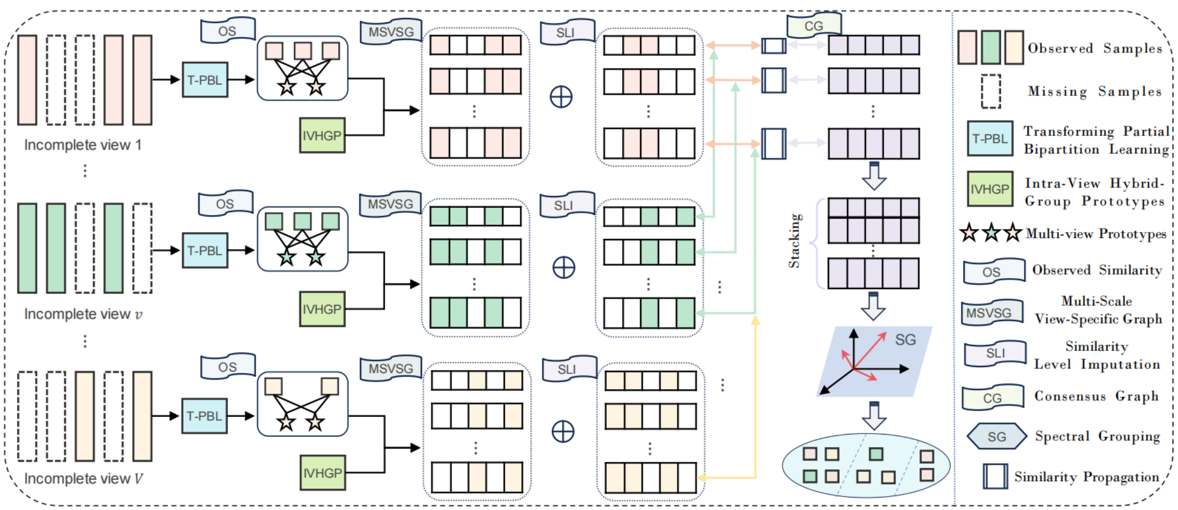


Figure 1: The overall framework of proposed SIIHPC. It firstly utilizes T-PBL to split out of observed similarity and then conducts SLI based on the connection between CG and view-specific similarities. Further, combined with IVHGP, it generates MSVSG to flexibly extract features on respective view.

Contributions

- Successfully impute incomplete parts at the similarity level. Not only does this alleviate the adverse impacts caused by the unpairing of observed samples but also can take advantages of the potential useful information of missing samples to help characterize the similarity more accurately.
- Successfully generate a group of hybrid prototype quantities for each individual view to flexibly extract data features according to the characteristics of each view itself. The resulting graphs are with diverse scales and besides balancing views, they are also able to more comprehensively describe the overall similarity.
- Carefully designs an alternate solving scheme which decomposes the entire problem into four parts and solves the sub-problem via an ingenious auxiliary function with theoretically proven monotonic-increasing properties.

Methodology

Let matrices $\{\mathbf{D}_v \in \mathbb{R}^{d_v \times n}\}_{v=1}^{V}$ and vectors $\{r_v \in \mathbb{R}^{n_v}\}_{v=1}^{V}$ denote the overall data and indexes of observed data respectively, then, the basic IMVC framework can be expressed as

$$\min_{\mathbf{X}_v} \sum_{v=1}^{V} \|\mathbf{D}_v\mathbf{W}_v - \mathbf{H}_v\mathbf{X}_v\mathbf{W}_v\|_F^2 + \lambda \|\mathbf{X}_v\|_F^2 \quad s.t. \;\; \mathbf{X}_v^\top \mathbf{1} = \mathbf{1}, \mathbf{X}_v \geq 0, \tag{1}$$

Rethinking (1), its nature is to reconstruct $\mathbf{D}_v$ using $\mathbf{H}_v\mathbf{X}_v$ under given $\mathbf{H}_v$. Unlike the fixing strategy, we firstly make prototype learnable and then introduce orthogonal constraint to strengthen its discrimination, i.e., $\mathbf{H}_v^\top\mathbf{H}_v = \mathbf{I}$. On this basis, we have $\mathbf{H}_v^\top\mathbf{D}_v\mathbf{W}_v = \mathbf{X}_v\mathbf{W}_v$. Then, the observed parts can be splited out through $\mathbf{X}_v\mathbf{W}_v\mathbf{W}_v^\top = \mathbf{H}_v^\top\mathbf{D}_v\mathbf{W}_v\mathbf{W}_v^\top$. Notice that the item $\mathbf{H}_v^\top\mathbf{D}_v$ can be regarded as the cosine similarity between $\mathbf{H}_v^\top$ and $\mathbf{D}_v$ when all columns of $\mathbf{D}_v$ are unit vectors. Hence, we choose to do normalization on $\mathbf{D}_v$, which expands the similarity range from $[0, 1]$ to $[-1, 1]$,

more freely measuring the similarity. Subsequently, we introduce a consensus graph $\mathbf{G}$ to aggregate information from different views, and impute the incomplete parts by utilizing $\mathbf{H}_v^\top\mathbf{D}_v\mathbf{W}_v\mathbf{W}_v^\top$ and $\mathbf{G}$. Further, to avoid a single prototype quantity for all views, we provide a group of hybrid prototype

$\mathbf{G}$. Further, to avoid a single prototype quantity for all views, we provide a group of hybrid prototype quantities $\{m_1, m_2, \cdots, m_s, \cdots, m_S\}$ for each view v to flexibly extract features according to the characteristics of each view itself. Consequently, we have $\mathbf{X}_{v,s}\mathbf{W}_v\mathbf{W}_v^\top = \mathbf{H}_{v,s}^\top\mathbf{D}_v\mathbf{W}_v\mathbf{W}_v^\top$. Besides, to adaptively adjust the importance between prototype quantities, we associate a learnable weight, $a_{v,s}$, for each prototype quantity on each view. Finally, our SIIHPC can be formulated as

Methodology

Let matrices $\{\mathbf{D}_v \in \mathbb{R}^{d_v \times n}\}_{v=1}^{V}$ and vectors $\{r_v \in \mathbb{R}^{n_v}\}_{v=1}^{V}$ denote the overall data and indexes of observed data respectively, then, the basic IMVC framework can be expressed as

$$\min_{\mathbf{X}_v} \sum_{v=1}^{V} \|\mathbf{D}_v\mathbf{W}_v - \mathbf{H}_v\mathbf{X}_v\mathbf{W}_v\|_F^2 + \lambda\|\mathbf{X}_v\|_F^2 \quad s.t. \;\; \mathbf{X}_v^\top\mathbf{1} = \mathbf{1}, \mathbf{X}_v \geq 0, \tag{1}$$

Rethinking (1), its nature is to reconstruct $\mathbf{D}_v$ using $\mathbf{H}_v\mathbf{X}_v$ under given $\mathbf{H}_v$. Unlike the fixing strategy, we firstly make prototype learnable and then introduce orthogonal constraint to strengthen its discrimination, i.e., $\mathbf{H}_v^\top\mathbf{H}_v = \mathbf{I}$. On this basis, we have $\mathbf{H}_v^\top\mathbf{D}_v\mathbf{W}_v = \mathbf{X}_v\mathbf{W}_v$. Then, the observed parts can be splited out through $\mathbf{X}_v\mathbf{W}_v\mathbf{W}_v^\top = \mathbf{H}_v^\top\mathbf{D}_v\mathbf{W}_v\mathbf{W}_v^\top$. Notice that the item $\mathbf{H}_v^\top\mathbf{D}_v$ can be regarded as the cosine similarity between $\mathbf{H}_v^\top$ and $\mathbf{D}_v$ when all columns of $\mathbf{D}_v$ are unit vectors. Hence, we choose to do normalization on $\mathbf{D}_v$, which expands the similarity range from $[0,1]$ to $[-1,1]$,

$$\begin{aligned} \min_{\mathbf{A},\mathbf{H}_{v,s},\mathbf{Q}_{v,s},\mathbf{G}_s} & \sum_{v=1}^{V}\sum_{s=1}^{S} a_{v,s}\left(\left\|\mathbf{H}_{v,s}^\top\mathbf{D}_v\mathbf{W}_v\mathbf{W}_v^\top + \mathbf{Q}_{v,s}\mathbf{M}_v\mathbf{M}_v^\top - \mathbf{G}_s\right\|_F^2 + \lambda\|\mathbf{G}_s\|_F^2\right) + \beta\|\mathbf{A}\|_F^2 \\ & s.t. \;\; \mathbf{H}_{v,s}^\top\mathbf{H}_{v,s} = \mathbf{I}_{m_s}, -1 \leq \mathbf{Q}_{v,s} \leq 1, -1 \leq \mathbf{G}_s \leq 1, \mathbf{A}\mathbf{1} = \mathbf{1}, 0 \leq \mathbf{A}, \end{aligned} \tag{2}$$

where $\mathbf{H}_{v,s} \in \mathbb{R}^{d_v \times m_s}$ denotes the prototype matrix with the s-th quantity on view v. $\mathbf{M}_v$ consists of $[\mathbf{M}_v]_{i,j} = 1 \;\; when \;\; [h_v]_j == i \;\; otherwise \;\; [\mathbf{M}_v]_{i,j} = 0, \;\forall j = 1, 2, \cdots, n - n_v; i = 1, 2, \cdots, n$. $h_v = \{z | z \in T_a \; and \; z \notin T_o\}$ where $T_a = \{1, 2, \cdots, n\}$ and $T_o = \{[r_v]_1, [r_v]_2, \cdots, [r_v]_{n_v}\}$. $\mathbf{Q}_{v,s} \in \mathbb{R}^{m_s \times n}$ is the imputation matrix with the s-th scale on view v. $\mathbf{A} \in \mathbb{R}^{V \times S}$ consists of $a_{v,s}$.

Methodology

Let matrices $\{\mathbf{D}_v \in \mathbb{R}^{d_v \times n}\}_{v=1}^{V}$ and vectors $\{r_v \in \mathbb{R}^{n_v}\}_{v=1}^{V}$ denote the overall data and indexes of observed data respectively, then, the basic IMVC framework can be expressed as

$$\min_{\mathbf{X}_v} \sum_{v=1}^{V} \|\mathbf{D}_v \mathbf{W}_v - \mathbf{H}_v \mathbf{X}_v \mathbf{W}_v\|_F^2 + \lambda \|\mathbf{X}_v\|_F^2 \quad s.t. \;\; \mathbf{X}_v^\top \mathbf{1} = \mathbf{1}, \mathbf{X}_v \geq 0, \tag{1}$$

Denote the function $g(\mathbf{H}_{v,s}) = \mathrm{Tr}(\mathbf{H}_{v,s}^\top \widehat{\mathbf{L}}_v \mathbf{H}_{v,s} + \mathbf{H}_{v,s}^\top \mathbf{P}_{v,s})$, its derivative as $\nabla g((\mathbf{H}_{v,s}))$, the value of $\mathbf{H}_{v,s}$ at the r-th iteration as $(\mathbf{H}_{v,s})^r$, the singular value decomposition results of $\nabla g((\mathbf{H}_{v,s})^r)$ as $(\mathbf{U}_{v,s})^r\,(\mathbf{\Sigma}_{v,s})^r\,(\mathbf{V}_{v,s}^\top)^r$. Then, we have the following two lemmas hold.

Theorem 1. *For the function g, under any $(\mathbf{H}_{v,s})^r$ and $(\mathbf{H}_{v,s})^{r+1} = (\mathbf{U}_{v,s})^r\,(\mathbf{V}_{v,s}^\top)^r$, we have $g(\mathbf{H}_{v,s})$ is monotonically increasing.*

Algorithm 1 The procedure of optimizing $\mathbf{H}_{v,s}$ in (3).

Input: The matrices $\mathbf{H}_{v,s}$, $\mathbf{Q}_{v,s}$, $\mathbf{G}_s$, $\mathbf{A}$, $\mathbf{D}_v$, $\mathbf{W}_v$, $\mathbf{M}_v$.
Construct the function $g(\mathbf{H}_{v,s})$.

while $g((\mathbf{H}_{v,s})^{r+1}) - g((\mathbf{H}_{v,s})^r)/g((\mathbf{H}_{v,s})^r) \leq 1e-3$ **do**
  Compute the derivative function $\nabla g\left((\mathbf{H}_{v,s})^r\right)$.
  Generate the singular matrices $(\mathbf{U}_{v,s})^r$ and $(\mathbf{V}_{v,s}^\top)^r$.
  Assign $(\mathbf{H}_{v,s})^{r+1}$ by $(\mathbf{H}_{v,s})^{r+1} = (\mathbf{U}_{v,s})^r\,(\mathbf{V}_{v,s}^\top)^r$.
  $r = r + 1$.
end while

Output: The prototype matrices $\{\mathbf{H}_{v,s}\}_{v=1,s=1}^{V,\;S}$.

Algorithm 2 The procedure of solving the problem (2).

Input: Data matrix $\mathbf{D}_v$, index vectors b_v, hyper-parameters λ and β, $v = 1, 2, \cdots, V$.
Construct indicator matrices $\mathbf{W}_v$ and $\mathbf{M}_v$.

while $(f_{obj}(t) - f_{obj}(t+1))/f_{obj}(t) <= 1e-4$ **do**
  Optimize the variable $\mathbf{H}_{v,s}$ by **Algorithm** 1.
  Optimize the variable $\mathbf{Q}_{v,s}$ by solving (6).
  Optimize the variable $\mathbf{G}_s$ by solving (7).
  Optimize the variable $\mathbf{A}$ by (11).
end while

Output: The unified representation matrices $\{\mathbf{G}_s\}_{s=1}^{S}$.

Experimental Performance

✓ Clustering Results

Table 2: Clustering Results on Benchmark Datasets

Method	BDGPFEA									NUSOBJECT								
	30%			50%			70%			30%			50%			70%		
	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
LSIMVC	26.56	6.14	26.56	26.81	7.18	26.81	27.40	8.35	28.07	21.70	9.03	31.23	21.81	8.43	30.99	20.57	7.43	30.52
GSRIMC	39.84	14.22	39.96	38.30	13.15	36.78	34.41	11.22	34.17	23.03	8.34	32.73	20.91	7.58	32.04	20.46	7.86	31.20
HCPIMSC	34.12	12.65	36.36	32.20	12.41	35.24	33.16	**11.81**	34.58	21.56	6.38	29.31	21.33	8.93	30.18	**22.87**	8.07	29.81
EEIMVC	35.70	14.47	36.34	33.23	12.39	36.05	31.02	10.37	33.64	21.51	6.17	13.51	21.07	8.97	13.42	20.13	8.14	13.17
LRGRIMVC	34.71	12.58	35.74	31.43	9.12	32.63	27.25	4.80	27.27	21.43	7.25	30.17	**22.62**	7.83	30.99	21.73	8.16	30.15
IMVCCBG	**40.05**	15.01	**40.17**	38.10	11.49	36.76	34.78	10.03	34.06	22.59	8.36	31.96	21.35	7.90	32.85	22.05	7.43	31.16
BGIMVSC	22.65	3.19	23.26	26.88	9.68	27.08	24.04	4.72	24.56	19.06	0.30	22.86	19.13	0.34	22.91	19.06	0.31	22.86
OSLFIMVC	30.38	9.58	36.47	31.84	9.12	35.37	31.68	8.80	35.73	21.88	7.67	32.34	20.68	6.91	33.00	18.41	4.63	29.58
NGSPCGL	29.95	6.64	31.13	29.54	6.07	29.89	27.15	5.70	27.64	23.09	7.83	30.78	19.97	4.62	28.32	17.47	2.22	25.20
PIMVC	34.03	14.62	35.99	33.04	12.92	33.54	34.41	11.60	35.41	21.29	**9.36**	31.07	21.08	8.52	31.12	19.44	7.89	31.36
PSIMVC	34.00	12.51	35.58	31.98	9.55	33.65	30.12	9.15	32.45	19.65	8.25	29.11	20.09	8.91	30.21	22.07	7.91	30.35
SAGL	23.76	1.69	23.92	23.03	1.41	23.60	28.52	4.09	29.56	20.48	7.67	27.46	20.39	6.91	26.16	18.39	6.63	26.47
HCLSCGL	29.80	7.12	31.40	24.28	3.12	25.08	28.25	4.31	28.55	21.93	7.54	30.68	21.59	7.79	31.49	20.28	7.81	31.31
Ours	38.80	**15.21**	39.97	**40.31**	**13.88**	**40.31**	**35.04**	11.54	**37.30**	**23.30**	9.14	**32.87**	22.38	**9.21**	**33.92**	21.46	**8.36**	**31.49**
	VGGFACEFIFTY									VGGFACEHUND								
LSIMVC	8.45	10.78	8.65	7.30	9.83	7.56	6.94	9.08	7.18									
GSRIMC					N/A									N/A				
HCPIMSC	10.33	12.36	12.14	10.54	11.90	10.85	10.85	9.23	10.16									
EEIMVC	6.05	14.03	5.94	5.60	14.15	5.50	5.33	13.29	5.23	3.37	7.32	4.78	3.41	6.89	5.67	3.20	6.27	5.74
LRGRIMVC	9.21	13.23	11.37	10.02	12.48	11.45	9.15	11.58	**12.56**					N/A				
IMVCCBG	12.13	14.25	13.11	11.52	13.29	12.40	10.80	12.35	11.66	8.12	14.23	8.92	7.52	13.25	8.25	6.80	12.20	7.06
BGIMVSC	6.49	9.83	6.83	7.34	9.45	7.19	6.76	9.85	7.19					N/A				
OSLFIMVC	8.50	8.79	8.96	6.98	6.70	7.58	6.01	5.09	6.60	5.54	9.59	5.97	4.62	7.54	5.05	3.60	5.81	4.08
NGSPCGL	6.50	6.47	7.19	6.24	6.54	6.74	6.08	6.24	6.75					N/A				
PIMVC	9.40	13.36	11.07	9.06	12.52	11.12	8.78	11.89	12.06	6.10	13.42	7.32	5.97	12.91	7.11	5.68	12.36	6.72
PSIMVC	10.63	12.50	11.58	9.54	11.33	10.49	9.06	10.45	9.92	6.17	11.04	6.71	5.28	10.58	5.89	5.51	9.91	6.04
SAGL	8.25	9.33	9.75	6.54	9.65	6.75	5.84	9.65	8.86	5.84	10.54	6.36	4.85	10.13	4.74	3.84	9.32	4.54
HCLSCGL	5.65	9.55	5.74	4.18	8.68	4.62	4.67	8.55	5.01	3.05	10.32	4.26	3.05	10.12	4.13	3.15	9.51	4.02
Ours	**12.52**	**14.91**	**13.44**	**12.31**	**14.48**	**13.21**	**11.18**	**13.35**	11.96	**8.26**	**14.94**	**9.13**	**7.55**	**13.85**	**8.39**	**6.82**	**12.69**	**7.55**

✓ Overhead Comparison

Table 3: Running Time and Memory Overhead Comparison

Method	BDGPFEA		NUSOBJECT		VGGFACEFIFTY		VGGFACEHUND		YOUTUBEFACE		FASHMINST	
	Time	Memo	Time	Memo	Time	Memo	Time	Memo	Time	Memo	Time	Memo
LSIMVC	0.03	0.32	0.13	2.42	0.55	15.18						
GSRIMC	2.10	2.82	28.88	28.07	N/A		N/A					
HCPIMSC	2.79	1.93	43.33	17.25	741.45	104.84			N/A		N/A	
EEIMVC	0.02	0.80	0.33	5.06	2.99	30.97	238.01	98.84				
LRGRIMVC	3.35	1.11	60.59	10.09	492.37	61.87	N/A					
IMVCCBG	0.02	0.17	0.05	0.20	0.56	1.73	2.46	**3.96**	**1.23**	6.16	**1.27**	6.41
BGIMVSC	4.85	1.27	15.76	8.75	132.67	135.70	N/A		N/A		N/A	
OSLFIMVC	0.09	0.30	0.18	1.51	3.23	10.02	20.00	41.96	13.54	126.28	12.27	123.42
NGSPCGL	0.97	1.79	11.42	13.43	111.29	89.09	N/A		N/A		N/A	
PIMVC	**0.01**	0.46	0.40	2.53	**0.36**	17.20	3.37	76.58				
PSIMVC	0.02	0.16	**0.04**	**0.15**	0.40	**1.62**	**1.30**	5.22	1.57	6.42	1.95	6.41
SAGL	0.22	0.40	0.41	2.22	7.54	16.29	45.06	26.96	38.39	82.96	30.07	99.01
HCLSCGL	0.17	1.84	4.14	13.92	309.87	91.51	4838.55	133.20	N/A		N/A	
Ours	0.06	**0.14**	0.11	0.22	1.86	2.84	6.22	10.32	3.31	**6.01**	3.57	**5.13**

Ablation Study

✓ Similarity-level Imputation Ablation

Table 4: Similarity-level Imputation Effectiveness

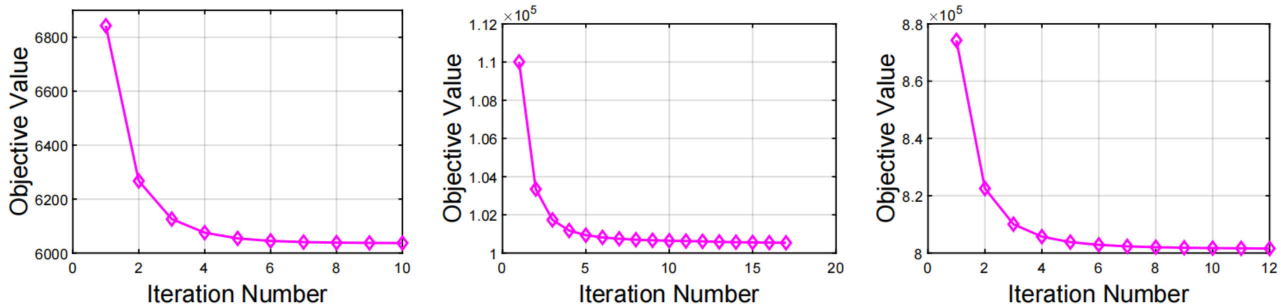
AB	MR	BDGPFEA			NUSOBJECT			VGGFACEFIFTY			VGGFACEHUND			YOUTUBEFACE			FASHMINST		
		ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
NSLI	30%	28.15	3.83	30.38	22.60	7.29	31.85	6.71	7.22	7.51	4.83	9.93	5.53	46.19	40.95	51.69	46.99	33.74	49.41
SLI		**38.80**	**15.21**	**39.97**	**23.30**	**9.14**	**32.87**	**12.52**	**14.91**	**13.44**	**8.26**	**14.94**	**9.13**	**76.29**	**82.27**	**80.81**	**61.24**	**59.52**	**62.69**
NSLI	50%	29.74	3.97	30.98	21.09	6.20	31.10	5.33	4.45	6.00	3.73	6.69	4.26	26.07	16.16	28.60	37.85	24.03	40.85
SLI		**40.31**	**13.88**	**40.31**	**22.38**	**9.21**	**33.92**	**12.31**	**14.48**	**13.21**	**7.55**	**13.85**	**8.39**	**72.60**	**79.60**	**77.65**	**62.51**	**60.22**	**64.64**
NSLI	70%	26.39	1.68	26.80	18.05	3.25	27.86	5.01	3.76	5.64	3.12	4.98	3.56	15.82	15.40	17.46	25.15	9.26	27.31
SLI		**35.04**	**11.54**	**37.30**	**21.46**	**8.36**	**31.49**	**11.18**	**13.35**	**11.96**	**6.82**	**12.69**	**7.55**	**71.05**	**79.19**	**76.75**	**60.59**	**58.77**	**63.18**

✓ Hybrid-group Prototype Quantity Ablation

Table 5: Hybrid-group Prototype Quantity Effectiveness

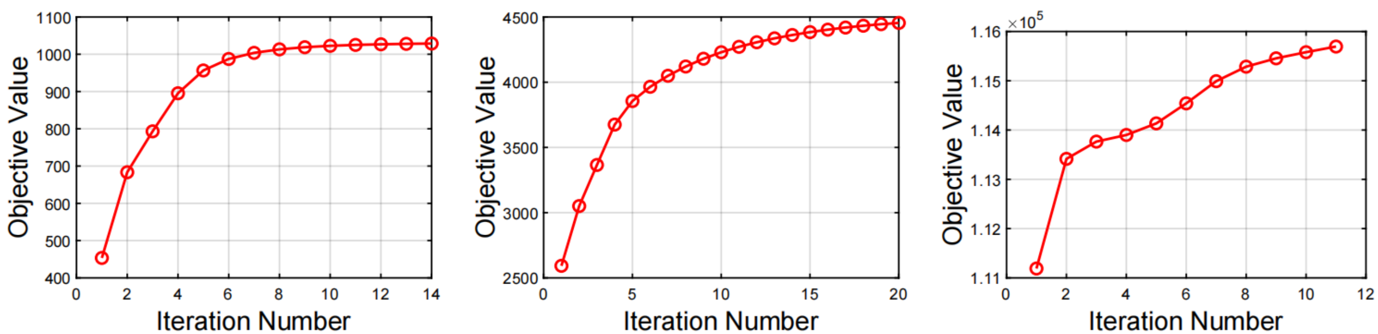
AB	MR	PQ	BDGPFEA			NUSOBJECT			VGGFACEFIFTY			VGGFACEHUND			YOUTUBEFACE			FASHMINST		
			ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR	ACC	NMI	PUR
SPQ	30%	m=1k	29.77	6.07	30.45	12.75	1.30	23.03	9.48	11.21	10.45	5.15	10.63	6.03	65.28	72.52	70.51	53.16	56.70	58.30
		m=2k	34.46	9.79	35.89	13.93	2.08	23.22	10.42	12.57	11.23	5.67	11.54	6.59	69.74	78.24	75.29	55.01	58.94	59.26
		m=3k	28.71	6.70	30.60	15.15	3.64	24.36	11.29	13.35	12.06	6.41	12.27	7.27	70.68	80.63	76.57	53.50	58.61	57.34
		m=4k	34.43	11.46	35.66	16.88	4.56	27.12	11.18	13.60	12.04	6.48	12.80	7.36	68.36	78.85	74.42	56.94	58.71	58.40
		m=5k	34.84	9.74	36.15	17.46	4.21	27.17	12.00	14.03	12.92	7.06	12.56	7.44	69.81	76.88	75.33	56.12	55.80	59.85
HPQ		m=ours	**38.80**	**15.21**	**39.97**	**23.30**	**9.14**	**32.87**	**12.52**	**14.91**	**13.44**	**8.26**	**14.94**	**9.13**	**76.29**	**82.27**	**80.81**	**61.24**	**59.52**	**62.69**
SPQ	50%	m=1k	24.23	2.30	25.18	14.37	1.57	24.79	9.20	10.70	10.08	5.98	10.78	6.16	62.12	66.35	66.63	53.41	56.05	54.69
		m=2k	34.87	7.83	34.87	18.88	5.68	29.75	10.77	12.81	11.71	6.12	11.31	6.32	66.96	74.15	71.43	50.97	55.58	54.41
		m=3k	34.59	8.96	34.59	19.27	6.21	29.56	11.17	13.31	11.93	6.61	11.68	6.76	67.03	73.99	72.04	55.62	58.39	57.25
		m=4k	34.98	9.44	35.74	20.13	7.01	30.80	11.24	13.51	12.17	6.54	11.74	6.68	70.21	77.23	74.64	58.47	58.69	59.24
		m=5k	34.25	9.22	36.55	19.04	6.01	31.08	11.33	13.25	12.16	6.68	11.89	7.43	70.53	76.28	75.53	61.01	58.45	62.83
HPQ		m=ours	**40.31**	**13.88**	**40.31**	**22.38**	**9.21**	**33.92**	**12.31**	**14.48**	**13.21**	**7.55**	**13.85**	**8.39**	**72.60**	**79.60**	**77.65**	**62.51**	**60.22**	**64.64**
SPQ	70%	m=1k	22.84	1.26	23.78	12.34	1.26	22.45	7.90	9.00	8.83	5.03	9.37	5.09	59.53	64.12	64.69	51.36	52.32	55.39
		m=2k	28.48	4.64	29.64	14.92	2.55	24.69	9.08	10.91	10.00	5.84	10.37	5.96	60.57	70.46	66.26	52.86	55.75	57.26
		m=3k	30.49	7.02	31.35	15.37	3.37	25.00	9.69	11.77	10.50	5.84	10.55	5.92	64.61	74.48	70.66	55.76	57.80	60.05
		m=4k	30.38	6.91	32.76	14.06	3.53	25.58	10.08	12.13	10.98	6.00	10.69	6.03	68.09	76.77	73.13	53.57	56.62	57.78
		m=5k	31.90	7.67	33.73	16.93	3.98	26.53	10.53	12.24	11.35	6.06	10.54	6.34	69.58	75.94	73.29	52.69	53.91	57.57
HPQ		m=ours	**35.04**	**11.54**	**37.30**	**21.46**	**8.36**	**31.49**	**11.18**	**13.35**	**11.96**	**6.82**	**12.69**	**7.55**	**71.05**	**79.19**	**76.75**	**60.59**	**58.77**	**63.18**

Convergence & Auxiliary Function Monotonicity



(a) BDGPFEA (b) NUSOBJECT (c) VGGFACEFIFTY

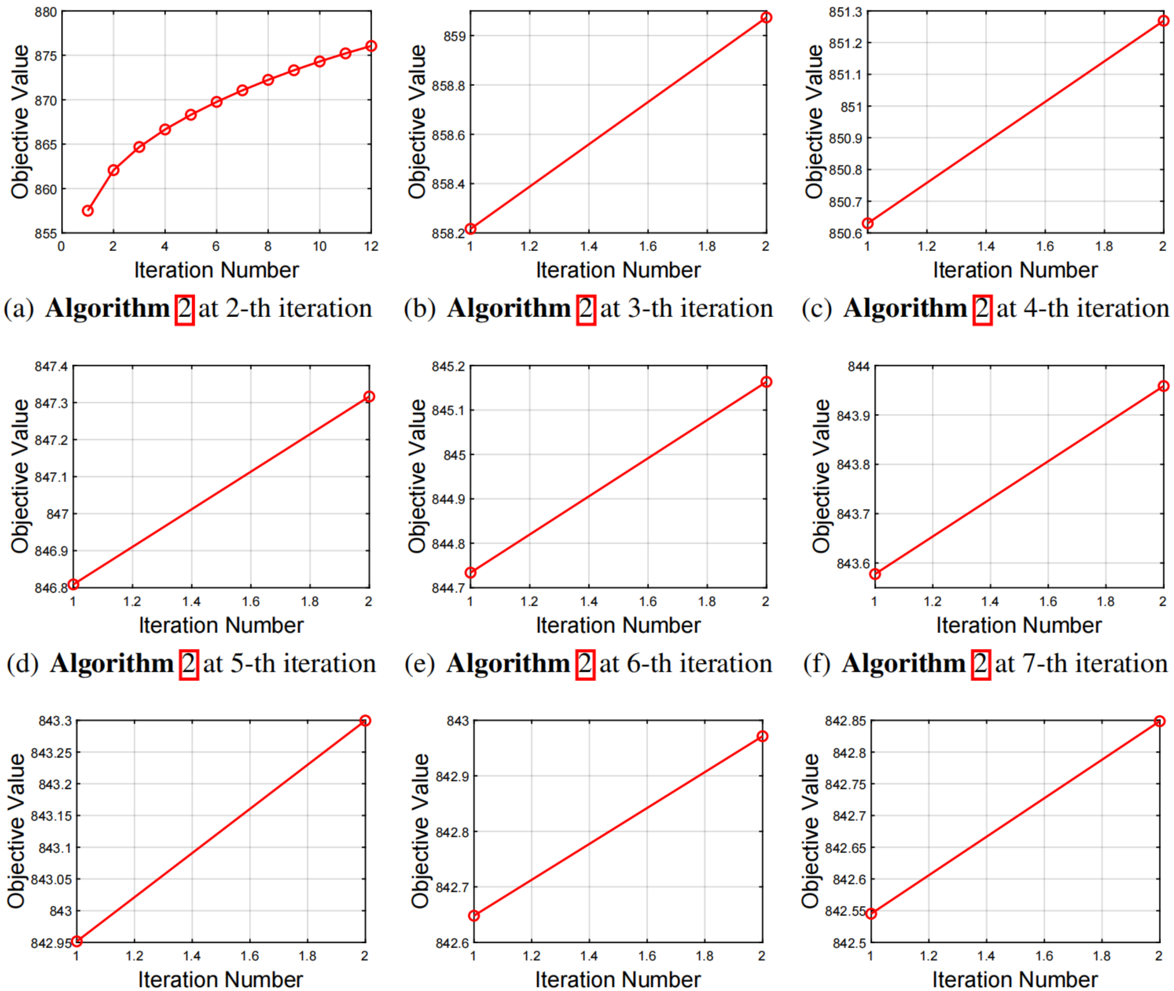
Figure 2: The objective value of **Algorithm** 2 on BDGPFEA, NUSOBJECT and VGGFACEFIFTY.



(a) BDGPFEA (b) NUSOBJECT (c) VGGFACEFIFTY

Figure 3: The value change of function g when **Algorithm** 2 is at the 1-th iteration.

Convergence & Auxiliary Function Monotonicity



(a) **Algorithm** 2 at 2-th iteration (b) **Algorithm** 2 at 3-th iteration (c) **Algorithm** 2 at 4-th iteration

(d) **Algorithm** 2 at 5-th iteration (e) **Algorithm** 2 at 6-th iteration (f) **Algorithm** 2 at 7-th iteration

(g) **Algorithm** 2 at 8-th iteration (h) **Algorithm** 2 at 9-th iteration (i) **Algorithm** 2 at 10-th iteration

Figure 4: The value change of function g on BDGPFEA.