Supplementary Material for
MPC: Multi-view Probabilistic Clustering

1. Additional Algorithm Details

Given a multi-view dataset of N samples with M views $S = \{V^{(1)}, V^{(2)}, ..., V^{(M)}\}$. $V^{(m)} \in R^{d^{(m)} * N}$ denotes the feature matrix in m-th view, where $d^{(m)}$ is the feature dimension of the m-th view. Let $W^{(m)} \in R^{N*N}$ calculated by $V^{(m)}$ using cosine similarity denotes the similarity matrix of the m-th view. The pairwise posterior probability of sample i and j can be expressed as:

$$P(i,j) = P(e_{ij} = 1 | w_{ij}^{(1)}, w_{ij}^{(2)}, ..., w_{ij}^{(M)}) \tag{1}$$

where e_{ij} indicates that the two samples belong to the same class and $w_{ij}^{(m)}$ denotes the similarity of the two samples in m-th view. As we all known, the Bayesian formula is:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{2}$$

Taking two views as an example and let $d_m = w_{ij}^{(m)}, e_1 = (e_{ij} = 1), e_0 = (e_{ij} = 0)$ for short. Based on Bayesian formula, Eq. (1) can be expressed as:

$$\begin{aligned} P(i,j) &= P(e_1|d_1,d_2) \\ &= \frac{P(d_1,d_2|e_1)P(e_1)}{P(d_1,d_2)} \\ &= \frac{P(d_1|e_1)P(e_1)}{P(d_1)} \frac{P(d_2|e_1)}{P(d_2|d_1)} \\ &= \frac{P(e_1|d_1)P(d_2|e_1)}{P(d_2|d_1)} \end{aligned} \tag{3}$$

Based on conditional independence assumption, $P(d_2|d_1)$ can be expressed as:

$$\begin{aligned} &P(d_2|e_1,d_1)P(e_1|d_1) + P(d_2|e_0,d_1)P(e_0|d_1) \\ &= P(d_2|e_1)P(e_1|d_1) + P(d_2|e_0)P(e_0|d_1) \end{aligned} \tag{4}$$

And then Eq. (1) can be expressed as:

$$P(i,j) = \frac{P(d_2|e_1)P(e_1|d_1)}{P(d_2|e_1)P(e_1|d_1) + P(d_2|e_0)P(e_0|d_1)} \tag{5}$$

Extending the formula to three views, the pairwise probability of sample i and j can be expressed as:

$$\begin{aligned} &P(i,j) = P(e_1|d_1,d_2,d_3) \\ &= \frac{P(d_2|e_1)P(e_1|d_1,d_3)}{P(d_2|e_1)P(e_1|d_1,d_3) + P(d_2|e_0)P(e_0|d_1,d_3)} \end{aligned} \tag{6}$$

Using Eq. (5), $P(e_1|d_1,d_3)$ and $P(e_0|d_1,d_3)$ can be expressed as:

$$\begin{aligned} P(e_1|d_1,d_3) &= \frac{P(d_3|e_1)P(e_1|d_1)}{P(d_3|e_1)P(e_1|d_1) + P(d_3|e_0)P(e_0|d_1)} \\ P(e_0|d_1,d_3) &= \frac{P(d_3|e_0)P(e_0|d_1)}{P(d_3|e_0)P(e_0|d_1) + P(d_3|e_1)P(e_1|d_1)} \end{aligned} \tag{7}$$

Naturally, Eq. (6) can be expressed as:

$$\begin{aligned} &P(i,j) = P(e_1|d_1,d_2,d_3) \\ &= \frac{P(d_2|e_1)P(e_1|d_1,d_3)}{P(d_2|e_1)P(e_1|d_1,d_3) + P(d_2|e_0)P(e_0|d_1,d_3)} \\ &= \frac{P(d_3|e_1)P(d_2|e_1)P(e_1|d_1)}{P(d_3|e_1)P(d_2|e_1)P(e_1|d_1) + P(d_3|e_0)P(d_2|e_0)P(e_0|d_1)} \end{aligned} \tag{8}$$

Thus, extending the formula to multiple views, the pairwise probability of sample i and j can be expressed as:

$$\begin{aligned} P(i,j) &= P(e_1|d_1,d_2,...,d_M) \\ &= \frac{(\prod_{m=2}^{M} P(d_m|e_1))P(e_1|d_1)}{\sum_{e\in\{e_0,e_1\}} (\prod_{m=2}^{M} P(d_m|e))P(e|d_1)} \end{aligned} \tag{9}$$

Finally, Eq. (1) can be expressed as:

$$\begin{aligned} &P(i,j) = P(e_{ij} = 1|w_{ij}^{(1)}, w_{ij}^{(2)}, ..., w_{ij}^{(M)}) \\ &= \frac{(\prod_{m=2}^{M} P(w_{ij}^{(m)}|e_{ij}=1))P(e_{ij}=1|w_{ij}^{(1)})}{\sum_{l\in\{0,1\}} (\prod_{m=2}^{M} P(w_{ij}^{(m)}|e_{ij}=l))P(e_{ij}=l|w_{ij}^{(1)})} \end{aligned} \tag{10}$$

which is the probability estimation algorithm we used in our proposed method.

Table 1. The incomplete multi-view clustering performance with different missing rates(0.1/0.3/0.7) on 100Leaves. The 1st/2nd best results are indicated in red/blue.

Methods	0.1				0.3				0.7			
	F_P	F_B	NMI	ARI	F_P	F_B	NMI	ARI	F_P	F_B	NMI	ARI
IMCCP [2]	26.52	39.50	71.86	25.48	21.47	34.15	67.84	20.30	27.81	38.67	70.71	26.86
OSLF [6]	57.65	61.16	83.75	57.23	37.64	42.48	73.32	36.99	27.12	33.13	69.27	26.38
UEAF [5]	51.26	56.98	81.79	50.76	43.30	50.76	78.43	42.70	36.51	44.76	75.16	35.81
EEIMC [3]	67.96	71.71	88.52	67.64	59.78	63.94	84.67	59.38	43.36	48.27	76.50	42.79
PIC [4]	75.79	79.14	91.27	75.55	63.64	67.44	86.22	63.28	38.46	44.77	76.12	37.84
MPC	79.18	81.18	92.45	79.00	66.25	68.98	87.18	65.95	44.54	48.96	78.13	44.08

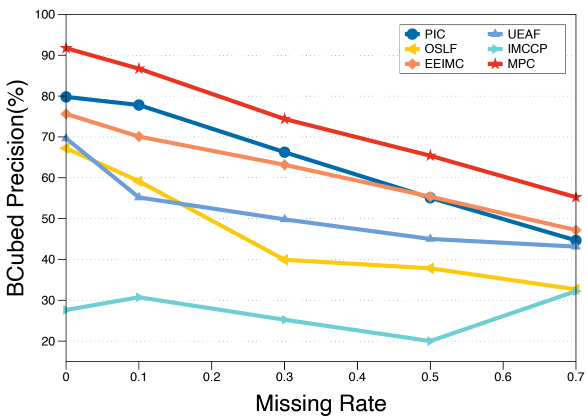

Figure 1. The clustering performance of BCubed Precision comparisons on 100Leaves with different missing rates.

Table 2. The clustering performance of BCubed Precision Pre_B and Fscore F_B comparisons on two additional datasets BUAA and BBCSport. MVC indicates complete multi-view clustering; IMVC indicates incomplete multi-view clustering with 0.5 missing rate. The 1st/2nd best results are indicated in red/blue.

Dataset	Type Methods	MVC		IMVC	
		Pre_B	F_B	Pre_B	F_B
BUAA	IMCCP [2]	39.29	39.74	32.50	32.94
	OSLF [6]	23.39	24.75	30.55	31.08
	EEIMC [3]	34.09	34.49	32.33	32.73
	UEAF [5]	28.46	29.59	29.02	30.05
	PIC [4]	44.25	43.65	35.02	35.46
	MPC	58.36	44.52	40.56	36.84
BBCSport	IMCCP [2]	28.67	35.42	25.13	34.20
	OSLF [6]	86.04	86.01	66.00	63.75
	EEIMC [3]	76.87	73.71	76.63	74.88
	UEAF [5]	82.69	83.88	87.51	87.20
	PIC [4]	90.41	90.39	86.80	86.96
	MPC	95.52	93.84	88.45	88.34

2. Additional Results and Analysis

Results on 100Leaves and Two Additional Datasets. The detailed incomplete clustering performance on 100Leaves is shown in Table 1. MPC surpasses the best baseline with different missing rates. As shown in Figure 1, our method surpasses PIC [4] by about 10% with different missing rates in terms of BCubed Precision, which further proves the accuracy of refined multi-view pairwise posterior matching probability in our proposed MPC. We conduct some experiments on two additional datasets BUAA and BBCSport. **BUAA-visnir face dataset (BUAA)** [1] contains 1350 visual images and 1350 near infrared images of the 150 volunteers. The feature dimension of both views used in experiments is 100. **BBCSport**[1] contains 544 samples of 5 categories. The feature dimensions of the two views used in experiments are 3181 and 3202 respectively. Table 2 lists the experimental results (BCubed Precision Pre_B and Fscore F_B) of different methods. Our method surpasses all tested baselines in terms of BCubed Precision and Fscore.

Analysis of Similarity Measures. To keep consistent with previous works PIC [4] and UEAF [5], we use cosine metric to estimate the similarity matrix. Overall, MPC is robust to the choice of metric. As listed in Table 3, we report the ARI obtained using similarity metric L_p, where $L_p(x_i, x_j) = (\sum_{l=1}^{n} |x_i^{(l)} - x_j^{(l)}|^p)^{\frac{1}{p}}, x_i = (x_i^{(1)}, ..., x_i^{(n)})$.

Table 3. Analysis of similarity measures on Humbi240 and Handwritten.

Similarity	**Cosine**	L_1	L_2	L_3	L_∞
Handwritten	**83.04**	81.06	81.28	81.37	55.77
Humbi240	**95.47**	95.33	95.45	95.34	91.26

Analysis of Parameter K. The K is the only input parameter for MPC. After we obtain the estimated probability using Eq. (10) on each view, we can find an approximate K to make the pairwise probability of farthest point in the KNN of most samples at 0.5. Table 4 lists the analysis of K on Humbi240 and the result shows that MPC is robust to parameter K.

Analysis of Convergence. As discussed in Section Fast Probabilistic Clustering (FPC), FPC is theoretically guaranteed to converge. The number of clustering set v.s. it-

[1] http://mlg.ucd.ie/datasets/segment.html

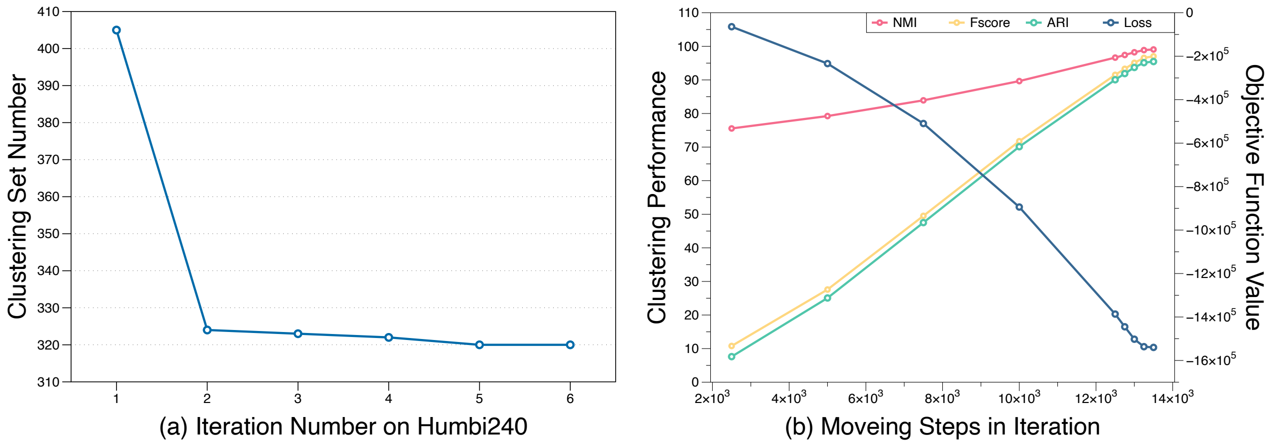

Figure 2. Analysis of Convergence. (a) The number of clustering set with increasing iterations on Humbi240. (b) The clustering performance of MPC with increasing moving steps on Humbi240. The x-axis denotes the moving steps in iteration, the left and right y-axis denote the clustering performance and corresponding objective function value, respectively.

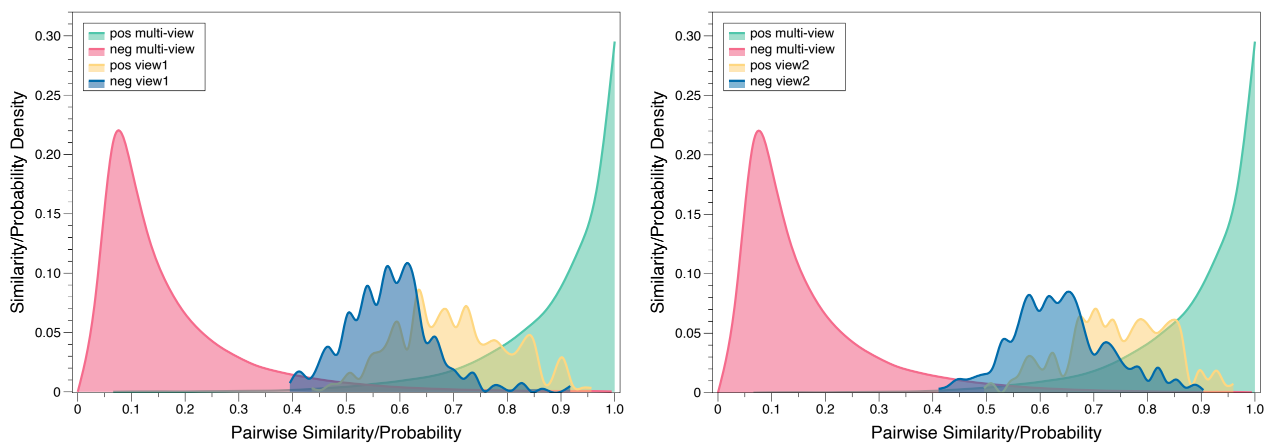

Figure 3. The pairwise similarity density in single view and the pairwise posterior probability density in MPC.

Table 4. Analysis of parameter K on Humbi240.

K	160	140	**120**	100	80	60
ARI	93.99	94.43	**95.47**	96.16	96.44	95.30

erations on Humbi240 is shown in Fig. 2(a) and it is clear that MPC converges in a few iterations, which sufficiently verifies the effectiveness of FPC. As shown in Fig. 2(b), the objective function value remarkably decreases during the moving steps in iteration, and meanwhile NMI, Fscore, and ARI continuously increase. And then the clustering performance keep stable in the last moving steps, which proves the good convergence property of our proposed FPC.

Visualization on MPC. As shown in Fig. 3, the experiments are conducted on Humbi240 dataset by visualizing the pairwise similarity density in single view and the pairwise posterior probability density in MPC. The pairsiwe similarity between positive pairs and negative pairs usually gather closely in view 1 and view 2, which makes it difficult to simply use a threshold to distinguish whether the pairwise samples are the same class. And using the probability estimation and refinement proposed in our method, the pairsiwe posterior probability between positive pairs and negative pairs naturaly become polarized and the overlap area between them is much smaller than that in single view, which demonstrates the success in multi-view information excavation of our MPC.

References

[1] Jia Sun Di Huang and Yunhong Wang. The buaa-visnir face database instructions. *Technical report*, 2012. 2

[2] Yijie Lin, Yuanbiao Gou, Zitao Liu, Boyun Li, Jiancheng Lv, and Xi Peng. Completer: Incomplete multi-view clustering

via contrastive prediction. In *CVPR*, pages 11174–11183, June 2021. 2

[3] Xinwang Liu, Miaomiao Li, Chang Tang, Jingyuan Xia, Jian Xiong, Li Liu, Marius Kloft, and En Zhu. Efficient and effective regularized incomplete multi-view clustering. *PAMI*, 43(8):2634–2646, 2021. 2

[4] Hao Wang, Linlin Zong, Bing Liu, Yan Yang, and Wei Zhou. Spectral perturbation meets incomplete multi-view data. In *IJCAI*, pages 3677–3683, 7 2019. 2

[5] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Hong Liu. Unified embedding alignment with missing views inferring for incomplete multi-view clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):5393–5400, Jul. 2019. 2

[6] Yi Zhang, Xinwang Liu, Siwei Wang, Jiyuan Liu, Sisi Dai, and En Zhu. One-stage incomplete multi-view clustering via late fusion. In *ACMMM*, page 2717–2725, 2021. 2