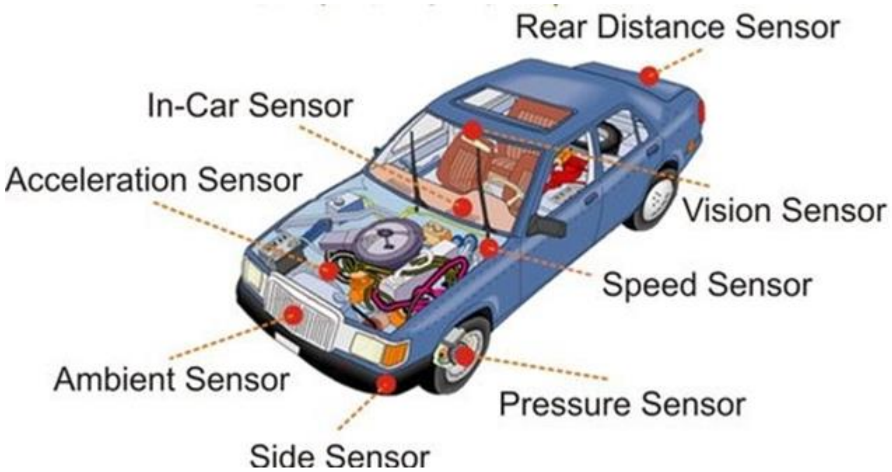
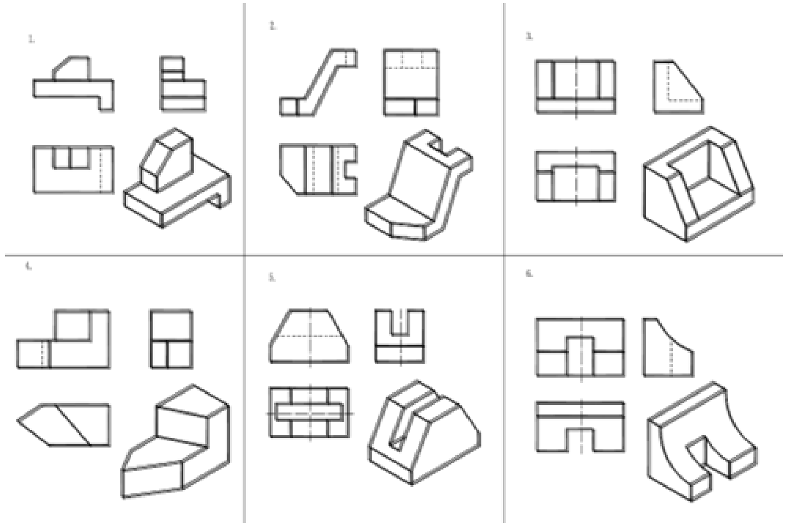
Deep Safe Incomplete Multi-view Clustering: Theorem and Algorithm

Huayi Tang, Yong Liu
Gaoling School of Artificial Intelligence, Renmin University of China

Introduction

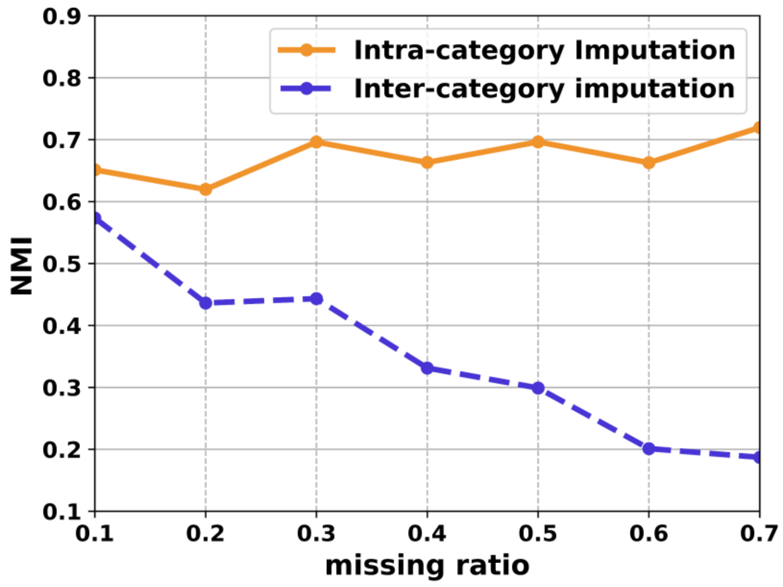




Multi-view data in real-world scenarios.

Introduction

- Imputed samples that are semantic consistent with the missing samples boost clustering performance
- semantic inconsistency between imputed views and missing views leads to degenerated clustering performance



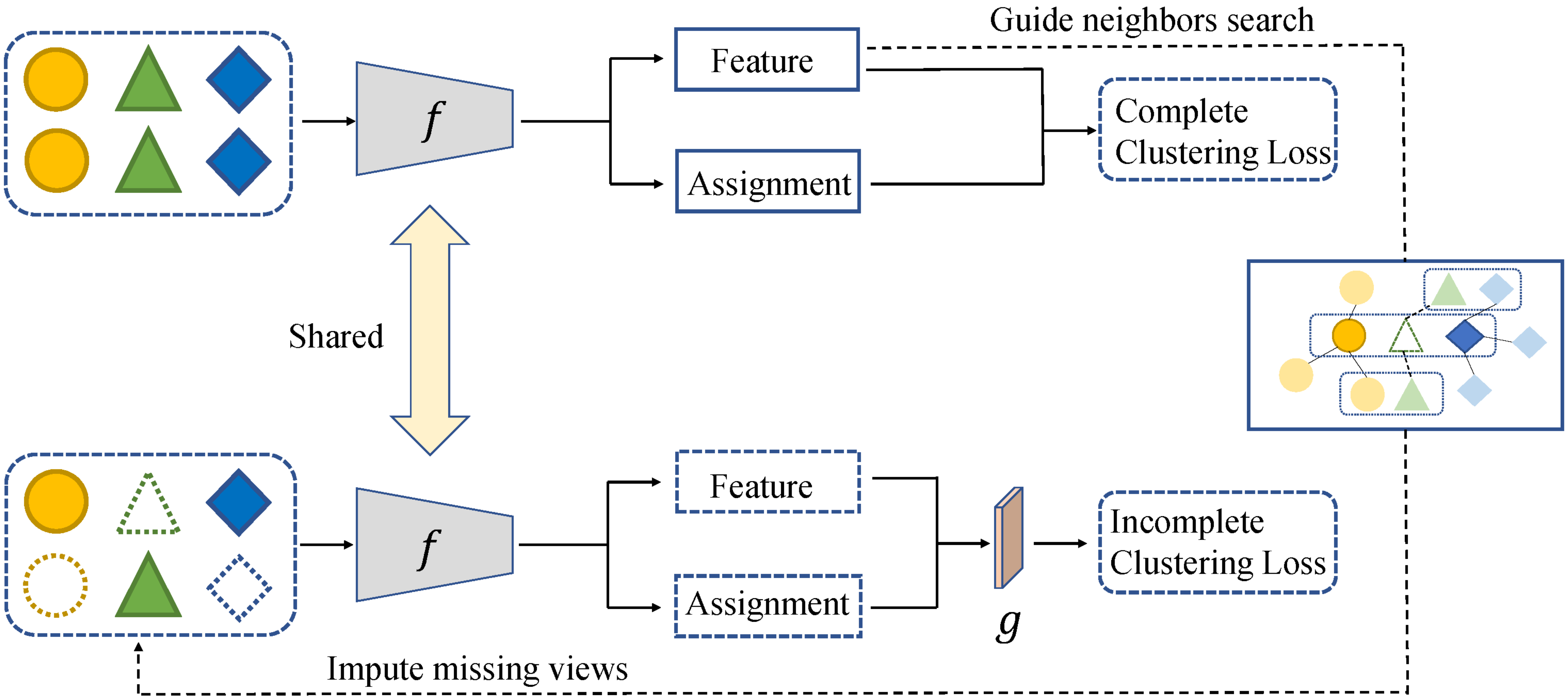
Clustering performance on incomplete data with intra-category and inter category imputations

Introduction

Challenges in IMVC:

- how to achieve semantic consistency between imputed views and missing views
- how to reduce the risk of clustering performance degradation caused by semantic inconsistency between imputed views and missing views

Methodology



Methodology

- features are dynamically updated in a moving-average manner

$$\mathbf{z}_i^{p,t} = \frac{(1-\gamma)\mathbf{z}_i^{p,t-1} + \gamma f_{\mathcal{Z}}\left(\mathrm{x}_i^{p}; w\right)}{\| (1-\gamma)\mathbf{z}_i^{p,t-1} + \gamma f_{\mathcal{Z}}\left(\mathrm{x}_i^{p,t}; w\right) \|_2}$$

- missing views are imputed from nearest neighbors inferred from features

$$\mathcal{N}_i^{p,t} := \bigcup_{q=1, q\neq p}^{m} \left\{\mathbf{x}_j^{p} \mid j \in \Psi_i^{q,t}\right\}, \hat{\mathbf{x}}_i^{p,t} = \frac{1}{\left|\mathcal{N}_i^{p,t}\right|} \sum_{\mathbf{x}^p \in \mathcal{N}_i^{p,t}} \mathbf{x}^p$$

Methodology

$$\min_{\phi,w} \mathcal{L}\left(f\left(\mathcal{D}^c; w\right)\right) \quad s.t. \quad w \in \mathcal{S}(\phi)$$

$$\mathcal{S}(\phi) = \underset{w}{\operatorname{argmin}} \mathcal{L}\left(f\left(\mathcal{D}^c; w\right)\right) + \mathcal{L}\left(f\left(\mathcal{D}^e; w\right), g\left(\mathcal{D}^e; \phi\right)\right)$$

- Lower-level problem: find the best multi-view model learning from both complete data and incomplete data with weights given by g
- Upper-level problem: g is optimized such that the model returned by the lower-level optimization task achieves the lowest empirical risk on complete data

Methodology

- empirical safe incomplete multi-view clustering

Let $\hat{\mathcal{L}}(f(\mathcal{D}^c; w)$ *be the empirical clustering risk on complete data* $\mathcal{D}^c$. *The parameters of the multi-view model learning only from complete data and the optimal solution of the bi-level optimization problem are denoted as* $w^* = \text{argmin}_{w \in \mathcal{W}} \hat{\mathcal{L}}\big(f(\mathcal{D}^c; w)\big)$ *and* $\hat{\phi}$ *respectively. We can prove that* $\hat{\mathcal{L}}\Big(f\big(\mathcal{D}^c; w^*(\hat{\phi})\big)\Big) \le \hat{\mathcal{L}}\big(f(\mathcal{D}^c; w^*)\big)$.

.

Methodology

- expected safe incomplete multi-view clustering

Suppose that $\left|\left|f_{\mathcal{Z}}(\mathbf{x}^p)\right|\right|_\infty \leq E$ *hold for all* $\{\mathbf{x}^p\}_{p=1}^m \in \mathcal{X}$*, where* $E > 0$ *is a constant. For any* $0 < \delta < 1$*, with at least probability* $1 - \delta$ *for any* $f \in \mathcal{F}$*, the following inequality holds*

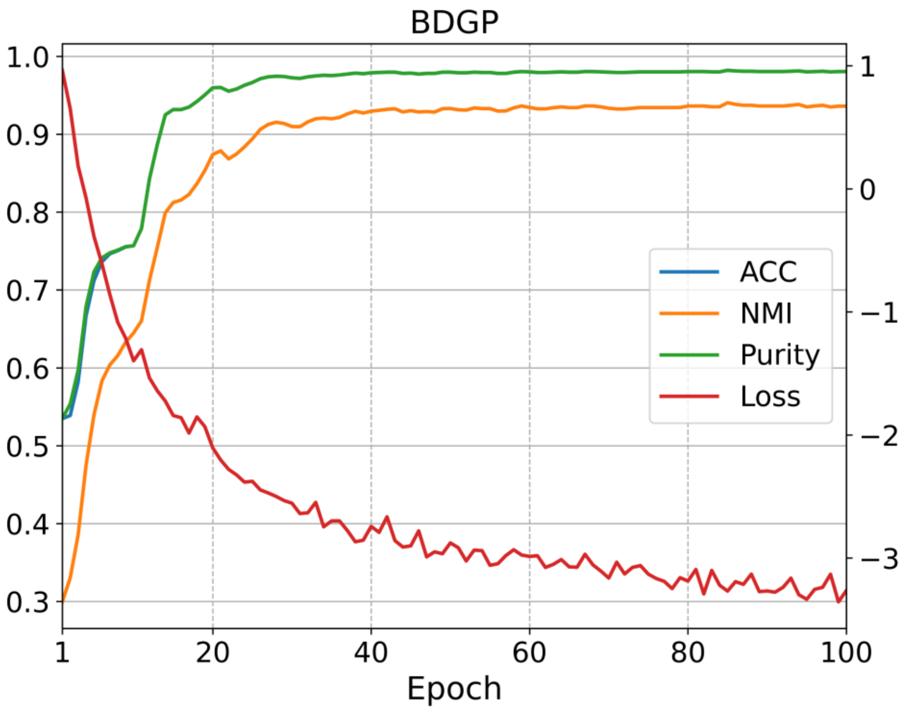
$$\mathcal{L}\left(f\left(w^*(\hat{\phi})\right)\right) + \varepsilon \leq \mathcal{L}\left(f(w^*)\right) + \frac{c_1}{\sqrt{n_c}} + c_2\sqrt{\frac{\log 12/\delta}{n_c}}$$

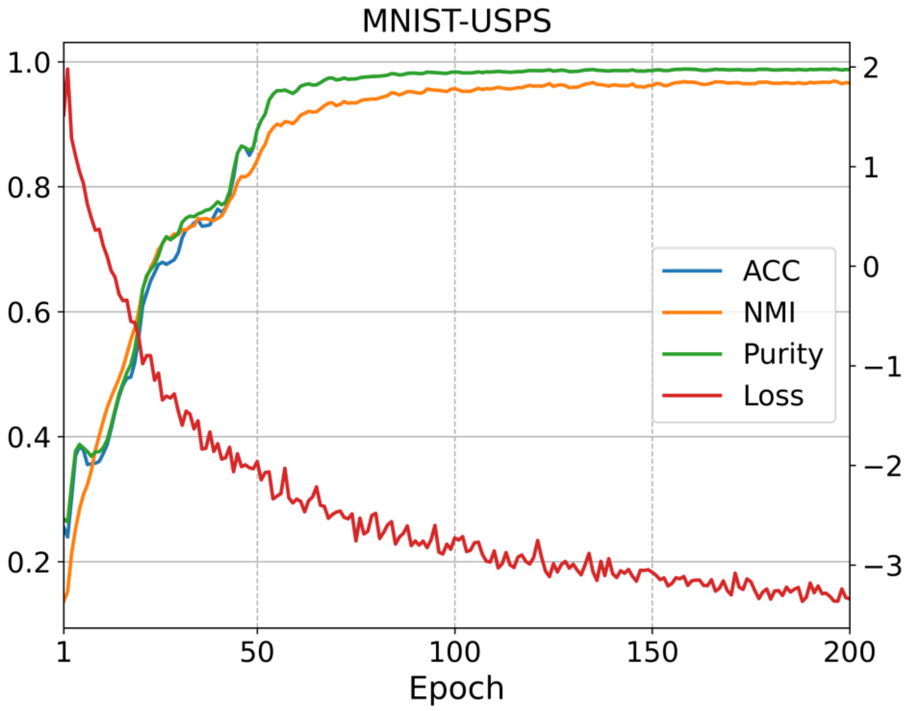
.

Experiments

		ACC				NMI			
Dataset	Method\ p	0.1	0.3	0.5	0.7	0.1	0.3	0.5	0.7
BDGP	BSV	59.64±1.43	54.67±1.34	44.49±0.59	35.96±0.32	47.21±1.54	42.53±1.40	32.24±0.63	22.97±0.37
	PVC	55.64±2.48	54.99±0.13	68.33±7.49	59.87±0.26	29.37±2.43	31.53±0.12	46.69±6.97	42.25±0.51
	UEAF	90.66±0.57	89.32±0.00	87.08±0.00	76.88±0.00	74.87±1.33	71.99±0.00	67.15±0.00	53.75±0.00
	CDIMC-net	80.47±0.82	74.67±0.53	67.71±1.05	56.11±4.80	70.08±0.36	67.64±0.78	54.51±1.12	39.70±4.85
	MKKM-IK	65.01±0.03	59.80±0.00	52.56±0.00	43.84±0.00	49.62±0.51	35.22±0.00	24.55±0.00	14.58±0.00
	EE-R-IMVC	65.28±0.00	57.36±0.00	42.48±0.00	34.85±2.48	43.82±0.00	31.79±0.00	21.39±0.00	11.87±1.99
	COMPLETER	40.91±7.04	41.80±4.12	41.54±7.64	39.63±2.78	33.19±4.33	31.15±5.43	32.62±5.58	27.47±3.37
	OS-LF-IMVC	82.78±2.18	74.34±1.16	59.71±3.22	45.34±1.39	60.25±4.69	48.27±2.33	30.56±3.97	18.54±1.31
	DSIMVC	**98.40±0.26**	**96.93±0.45**	**95.29±0.37**	**92.14±0.84**	**94.67±0.91**	**90.34±1.13**	**86.11±0.92**	**79.37±1.56**
MNIST-USPS	BSV	49.15±1.76	42.57±1.70	35.62±1.67	26.67±1.04	45.15±0.69	39.17±0.78	31.73±0.94	23.62±0.41
	PVC	64.57±2.73	63.04±3.69	52.56±1.14	50.24±2.84	58.74±1.66	55.63±1.03	46.35±0.47	44.34±1.33
	UEAF	71.27±0.97	66.08±1.26	61.94±0.00	54.18±0.00	66.75±1.81	58.04±2.14	57.84±0.00	49.77±0.00
	CDIMC-net	52.23±4.52	49.72±1.10	47.97±1.13	31.78±1.68	61.45±2.74	64.40±2.42	56.62±0.87	34.79±0.83
	MKKM-IK	72.25±0.61	64.44±0.00	49.74±1.04	35.70±0.00	61.64±0.18	52.01±0.00	37.67±0.59	24.68±0.00
	EE-R-IMVC	75.07±0.50	58.86±0.00	45.58±0.00	28.02±0.00	64.27±0.17	49.47±0.00	34.15±0.00	16.97±0.00
	COMPLETER	96.87±1.04	96.56±0.82	93.66±5.63	83.80±6.05	93.94±1.29	92.31±1.18	90.51±2.71	81.18±2.94
	OS-LF-IMVC	62.29±1.80	46.58±2.93	32.83±1.45	23.70±0.86	49.14±2.42	33.98±2.11	22.22±0.82	13.96±0.66
	DSIMVC	**98.88±0.09**	**97.89±0.14**	**96.78±0.25**	**93.34±0.64**	**96.91±0.21**	**94.50±0.36**	**91.98±0.55**	**85.64±0.93**
CCV	BSV	18.91±0.37	17.55±0.41	15.74±0.26	14.46±0.27	17.22±0.15	15.61±0.20	13.44±0.15	11.46±0.10
	PVC	16.48±0.40	15.54±0.27	14.75±0.33	14.01±0.24	13.86±0.36	10.12±0.28	9.67±0.27	8.66±0.18
	UEAF	26.38±0.00	24.82±0.00	22.63±0.00	14.92±3.20	23.64±0.00	23.10±0.00	21.34±0.00	10.42±3.66
	CDIMC-net	18.53±1.10	18.20±1.24	17.41±0.56	14.53±0.98	15.88±0.68	14.89±0.72	13.45±1.06	9.28±1.12
	MKKM-IK	19.71±0.38	18.29±0.00	15.46±0.00	14.13±0.00	14.78±0.06	12.61±0.00	10.30±0.00	8.00±0.00
	EE-R-IMVC	25.29±0.04	23.03±0.00	17.87±0.00	14.78±0.00	21.43±0.10	17.53±0.00	12.35±0.00	7.48±0.00
	COMPLETER	21.72±1.30	20.62±0.48	18.38±0.73	17.35±0.69	22.57±0.96	19.59±0.66	17.33±0.80	13.73±0.79
	OS-LF-IMVC	20.47±0.74	17.15±0.63	14.21±0.50	12.37±0.46	15.34±0.56	12.23±0.36	9.50±0.37	7.05±0.46
	DSIMVC	**30.90±1.22**	**29.33±1.24**	**27.07±0.81**	**24.87±0.49**	**29.76±0.71**	**28.18±0.65**	**25.72±0.61**	**22.96±0.56**
Multi-Fashion	BSV	49.81±2.60	42.97±2.01	34.83±1.32	26.59±0.83	48.32±0.99	40.85±0.60	32.46±0.64	23.73±0.41
	PVC	45.69±0.44	40.77±1.50	42.01±2.61	40.55±0.79	44.98±0.33	39.32±1.07	39.78±1.12	39.2±0.71
	UEAF	57.07±0.67	50.88±2.88	48.96±0.88	30.34±0.00	57.15±1.72	48.79±4.78	44.04±4.03	24.13±0.00
	CDIMC-net	51.00±4.89	44.73±2.23	42.10±3.00	37.61±3.68	62.52±1.94	54.67±1.94	44.85±4.19	46.05±1.29
	MKKM-IK	70.08±0.12	59.96±0.00	46.38±0.00	29.84±0.00	61.29±0.13	50.52±0.00	38.25±0.00	20.64±0.00
	EE-R-IMVC	72.83±0.97	63.32±0.00	51.16±0.00	20.24±0.00	65.78±0.36	57.28±0.00	43.50±0.00	14.61±0.00
	COMPLETER	78.63±0.33	71.68±3.70	70.76±5.62	69.33±4.51	82.23±1.18	77.12±0.57	74.76±1.35	70.23±2.73
	OS-LF-IMVC	62.54±1.01	50.10±2.68	37.47±1.38	27.67±1.25	52.36±1.11	38.74±2.05	30.04±1.48	19.98±1.70
	DSIMVC	**89.60±0.89**	**87.47±1.23**	**83.79±1.40**	**75.71±1.69**	**84.47±0.70**	**81.76±1.07**	**77.82±0.73**	**71.53±1.45**

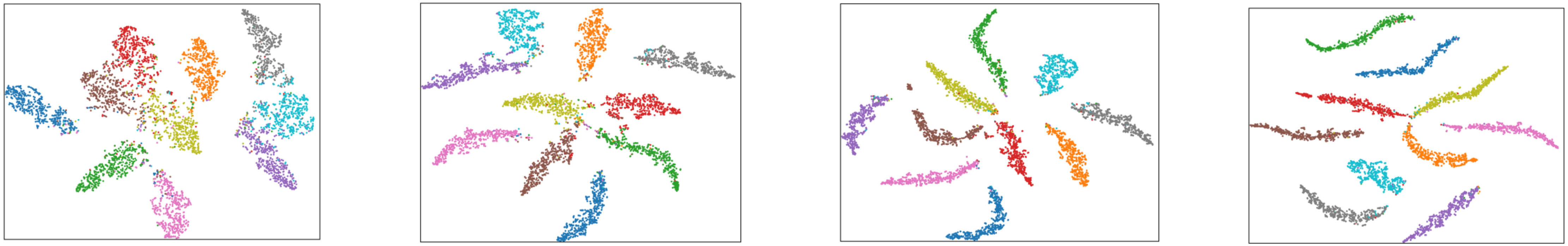
Experiments





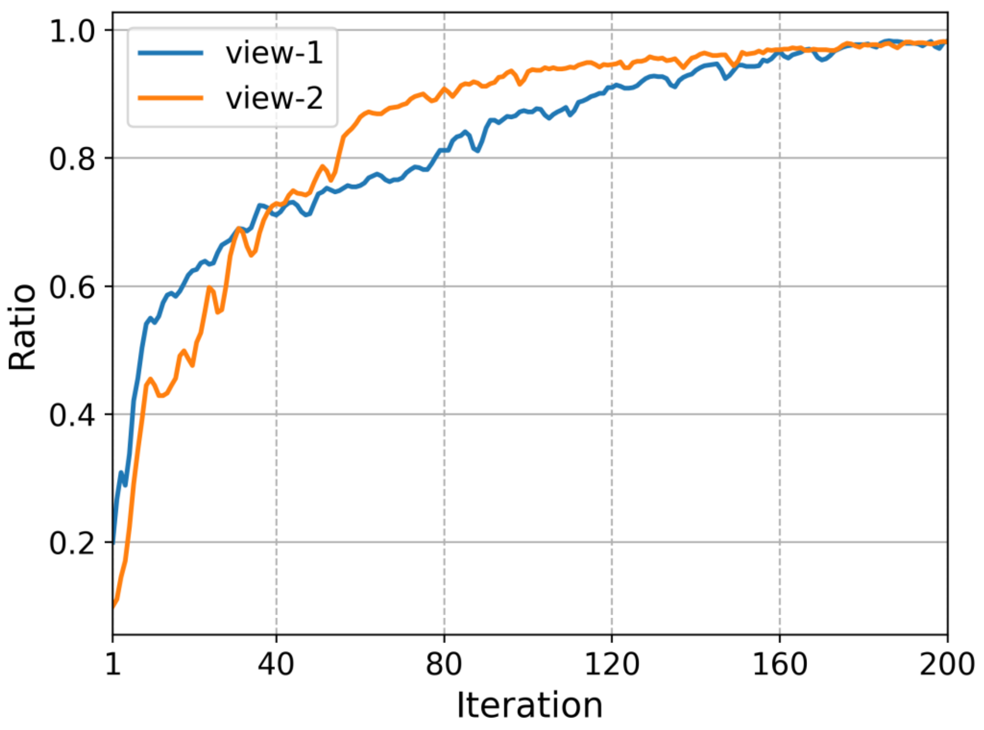
The objective value and clustering performance of DSIMVC with the increase of iterations on BDGP and MNIST-USPS

Experiments



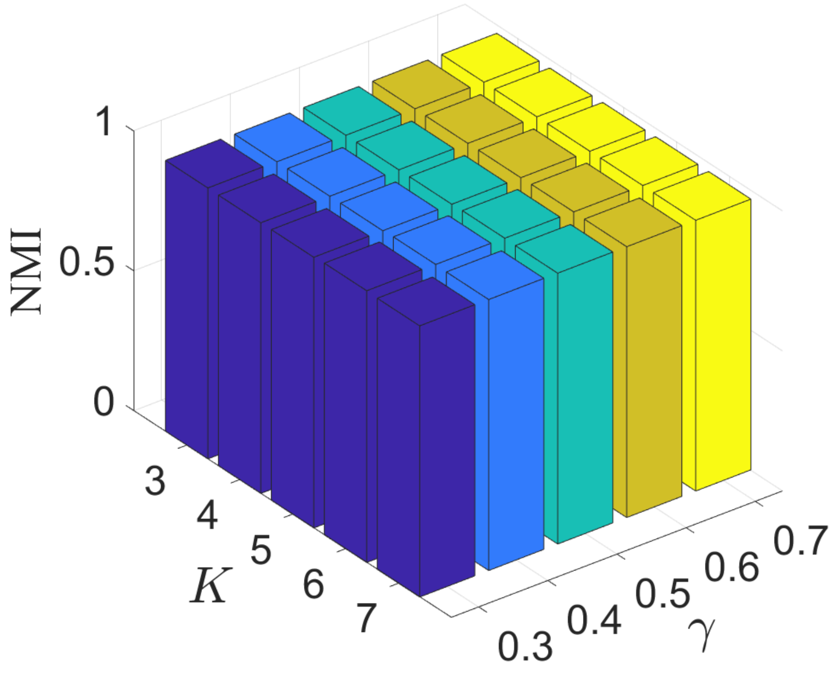
t-SNE visualization of the learned features on MNIST-USPS with increasing training iterations

Experiments

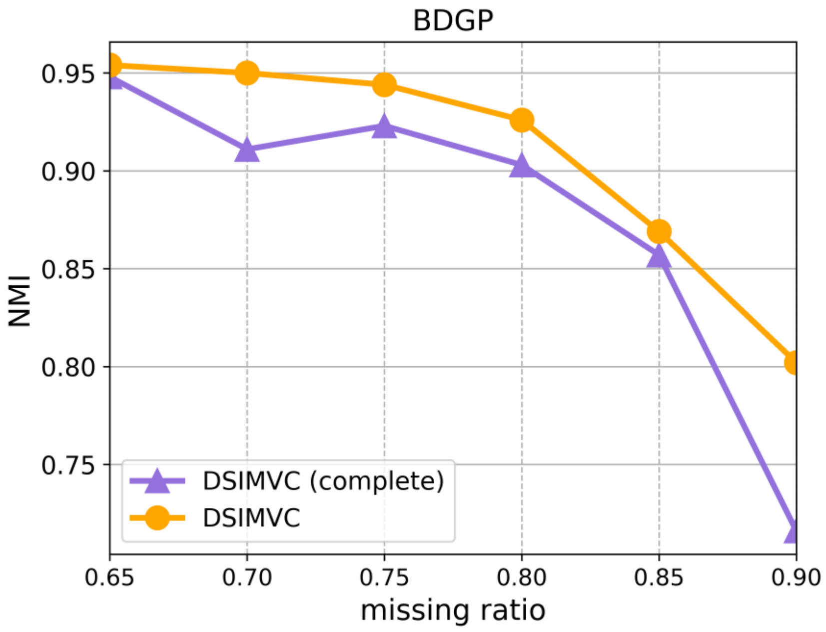


Sematic consistency ratio of the learned neighbors with the increase of iterations

Experiments



Parameters sensitivity analysis



Clustering performance in terms of NMI of DSIMVC and its variant with different missing ratios

Conclusion

- A novel IMVC framework to mine semantic consistent imputations and reduce the clustering performance degradation risk from semantic inconsistent imputations
- By the proposed bi-level optimization framework, missing views are dynamically imputed from semantic neighbors, and the incomplete samples are automatically selected for learning.
- The proposed framework is guaranteed to achieve no higher empirical and expected risk than the model learning only from complete data.

Thanks for your attention!