Generalized Information-theoretic Multi-view Clustering

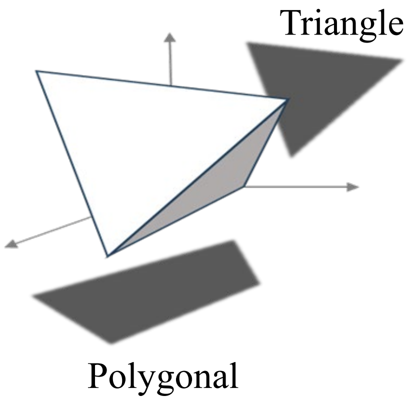
Weitian Huang, Sirui Yang, Hongmin Cai*

School of Computer Science and Engineering
South China University of Technology
Guangzhou, 510006, China

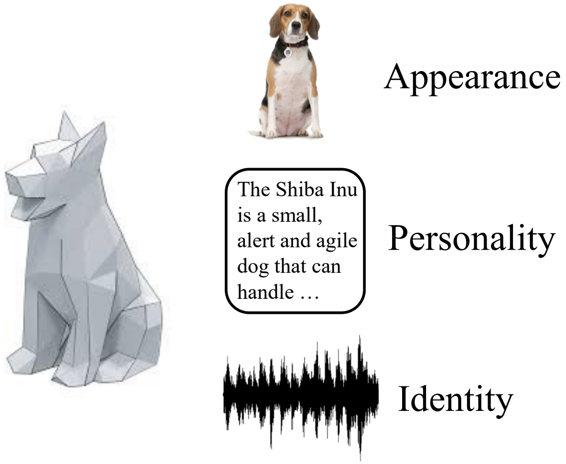
Background

Multi-view Clustering:

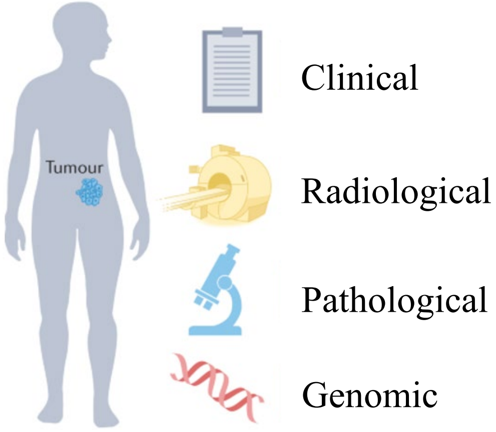
Exploring low-dimensional embeddings for describing flexible multi-view data and revealing the hidden patterns.



multiple views



multiple modalities



multiple levels

Minimal & Sufficient

- **Minimal:** eliminate superfluous information from each view.
- **Sufficient:** contain task-related information at different levels of data.

Variants of Information Bottleneck

- **Information Bottleneck:**

$$\mathcal{L}_{IB} = \max_{\mathbf{Z}} \underbrace{I(\mathbf{Z};\boldsymbol{Y})}_{\textit{Sufficient}} - \beta \underbrace{I(\mathbf{Z};\mathbf{X})}_{\textit{Minimal}}$$

- **Unsupervised Information Bottleneck:**

$$\begin{aligned}\mathcal{L}_{UIB} &= \max_{\mathbf{Z}} I(\mathbf{Z};\mathbf{X}) - \beta I(\mathbf{Z};i) \\ &\overset{(a)}{\geq} \mathbb{E}_{p(\mathbf{x})}\left[\underbrace{\mathbb{E}_{p(\mathbf{z}|\mathbf{x})}\left[\log q(\mathbf{x}|\mathbf{z})\right]}_{\textit{Reconstruction}} - \beta \underbrace{D_{KL}(p(\mathbf{z}|\mathbf{x})||q(\mathbf{z}))}_{\textit{Regularization}}\right]\end{aligned}$$

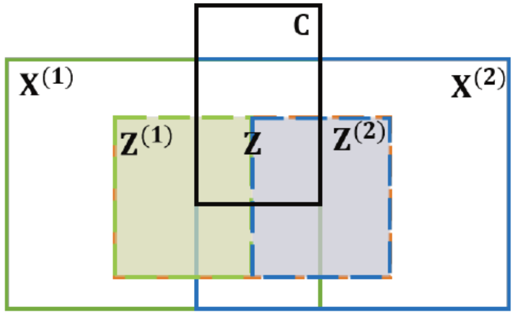
$\longrightarrow$ Connection to βVAE

- **Deep clustering:**

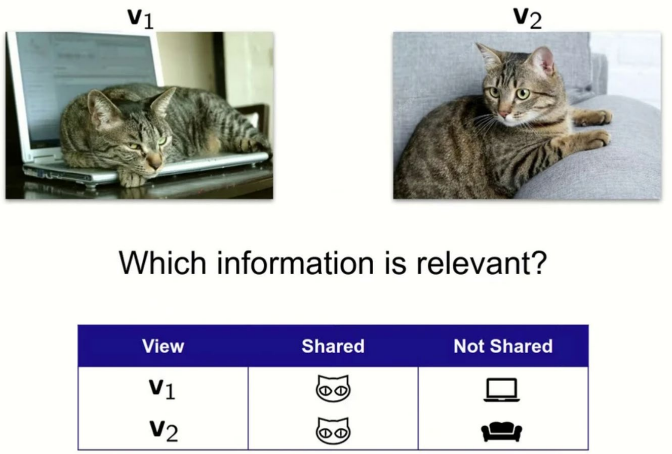
$$\begin{aligned}\mathcal{L}_{IBC} &= \max_{\mathbf{Z}} I(\mathbf{Z};\mathbf{X}) - \beta I(\mathbf{Z};i) \\ &\geq \mathbb{E}_{p(\mathbf{x})}\left[\mathbb{E}_{p(\mathbf{z}|\mathbf{x})}\left[\log q(\mathbf{x}|\mathbf{z})\right] - \beta D_{KL}(p(\mathbf{z}|\mathbf{x})||q(\mathbf{z}))\right. \\ &\quad \left. - \gamma \mathbb{E}_{p(\mathbf{z}|\mathbf{x})}\left[\underbrace{D_{KL}(p(\mathbf{c}|\mathbf{x})||q(\mathbf{c}|\mathbf{z}))}_{\textit{Cluster structure preserving}}\right]\right]\end{aligned}$$

$\longrightarrow$ Connection to *DEC* ($\beta = 0$ and $\gamma = 1$); *VaDE* ($\beta = 1$ and $\gamma = 1$)

Motivation



(a) Special case



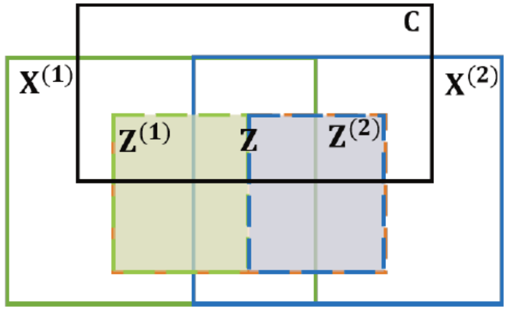
Related works:

[1] MIB (ICLR 2021)

Marco Federici, et al. “Learning robust representations via multi-view information bottleneck.” *In ICLR, 2021.*

[2] DCP (TPAMI 2022)

Yijie Lin, Yuanbiao Gou, Xiaotian Liu, Jinfeng Bai, Jiancheng Lv, and Xi Peng. “Dual contrastive prediction for incomplete multi-view representation learning.” *IEEE Trans. Pattern Anal. Mach. Intell., 2022.*



(b) General case

Generalized Information-theoretic Multi-view Clustering

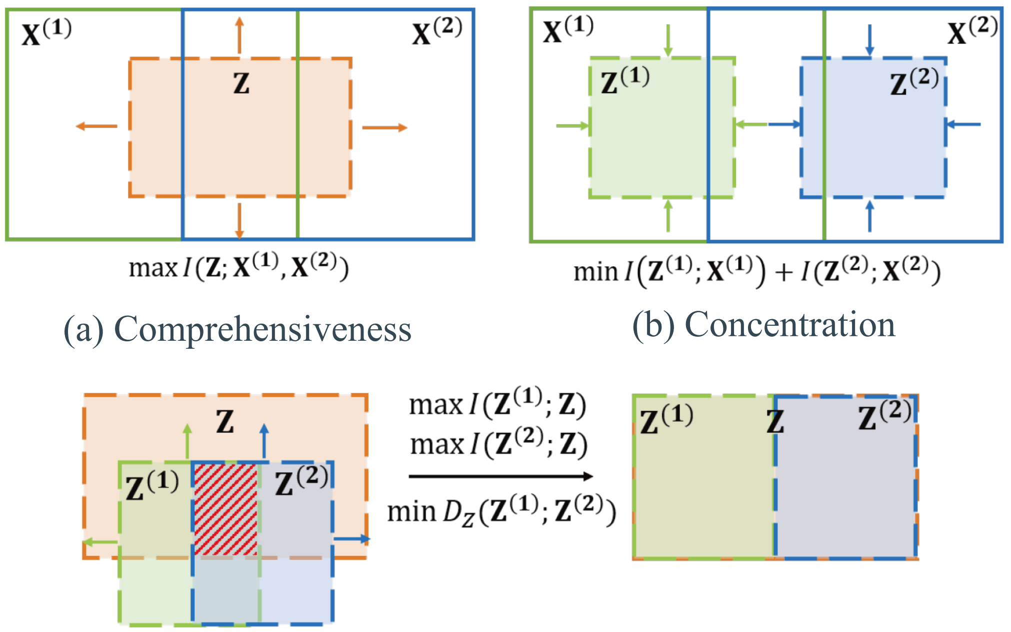
IMC: Information-based Multi-view Clustering

Comprehensiveness: $\mathbf{Z}^* = \arg\max_{\mathbf{Z}} I(\mathbf{Z}; \boldsymbol{\mathcal{X}})$

Concentration: $\mathbf{Z}^{(v)*} = \arg\min_{\mathbf{Z}^{(v)}} I(\mathbf{Z}^{(v)}; \mathbf{X}^{(v)})$

Cross-diversity: $\boldsymbol{\mathcal{Z}}^* = \arg\max_{\boldsymbol{\mathcal{Z}}} I(\mathbf{Z}; \boldsymbol{\mathcal{Z}})$

$$\mathcal{L}_{IMC} = \max_{\mathbf{Z},\boldsymbol{\mathcal{Z}}} I(\mathbf{Z}; \boldsymbol{\mathcal{X}}) - \sum_{v}^{m} I(\mathbf{Z}^{(v)}; \mathbf{X}^{(v)}) + \beta I(\mathbf{Z}; \boldsymbol{\mathcal{Z}})$$



(a) Comprehensiveness

(b) Concentration

(c) Cross-diversity

Cross-diversity is sufficient for comprehensiveness and concentration.

(*See Proposition 3.1 in the paper*)

Multi-VAE scheme to solve IMC

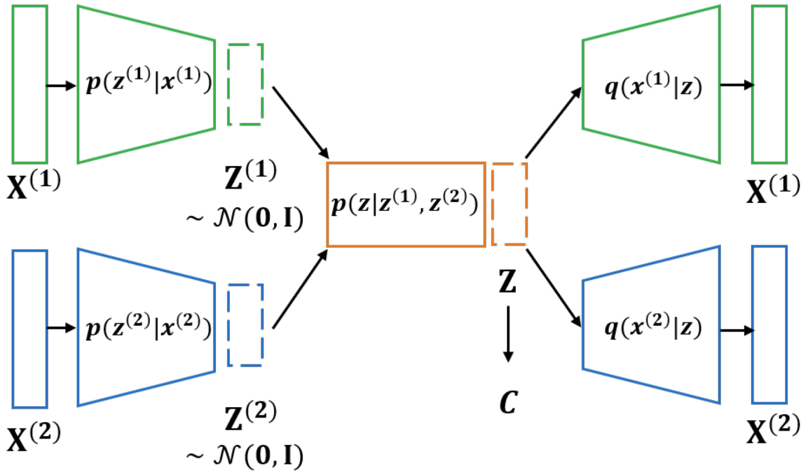
Following VAE, we instantiate the IMC using deep neural networks and optimize it by leveraging SGVB and Monte Carlo sampling.

$$
\begin{aligned}
\mathcal{L}_{IMC} &= \max_{\mathbf{Z},\mathcal{Z}} I(\mathbf{Z};\mathcal{X}) - \sum_{v}^{m} I(\mathbf{Z}^{(v)};\mathbf{X}^{(v)}) + \beta I(\mathbf{Z};\mathcal{Z}) \\
&\overset{(d)}{\geq} \mathbb{E}_{p(\mathcal{X})}\Big[\underbrace{\mathbb{E}_{p(\mathbf{z}|\mathcal{X})}\left[\log q(\mathcal{X}|\mathbf{z})\right]}_{\text{data reconstruction}} - \underbrace{\sum_{v}^{m} D_{KL}(p(\mathbf{z}^{(v)}|\mathbf{x}^{(v)})||q(\mathbf{z}^{(v)}))}_{\text{multi-regularization}} \\
&\quad - \gamma \underbrace{\mathbb{E}_{p(\mathbf{z}|\mathcal{X})}\left[D_{KL}(p(\mathbf{c}|\mathcal{X})||q(\mathbf{c}|\mathbf{z}))\right]}_{\text{clustering}}\Big] + \beta \underbrace{I(\mathbf{Z};\mathcal{Z})}_{\text{information shift}} .
\end{aligned}
$$

Encoders and Decoders:

$$
\begin{aligned}
p(\mathbf{z}|\mathcal{X}) &= \int\int \ldots \int p_{\psi}(\mathbf{z}|\{\mathbf{z}^{(v)}\}_{v=1}^{m}) \prod_{v=1}^{m} p_{\theta^{(v)}}(\mathbf{z}^{(v)}|\mathbf{x}^{(v)}) d_{\mathbf{z}^{(1)}} d_{\mathbf{z}^{2}} \ldots d_{\mathbf{z}^{m}}, \\
&= \mathbb{E}_{p_{\theta^{(1)}}(\mathbf{z}^{(1)}|\mathbf{x}^{(1)})} \mathbb{E}_{p_{\theta^{(2)}}(\mathbf{z}^{(2)}|\mathbf{x}^{(2)})} \ldots \mathbb{E}_{p_{\theta^{(m)}}(\mathbf{z}^{(m)}|\mathbf{x}^{(m)})} \left[p_{\psi}(\mathbf{z}|\{\mathbf{z}^{(v)}\}_{v=1}^{m})\right], \\
q(\mathcal{X}|\mathbf{z}) &= \prod_{v=1}^{m} q_{\phi^{(v)}}(\mathbf{x}^{(v)}|\mathbf{z}),
\end{aligned}
$$

illustration $\longrightarrow$



Results

Ablation models:

(1) **IMC-v1** reduces the item of information shift, i.e., $\beta = 0$.

(2) **IMC-v2** sets $\gamma = 0$ to discard the KL divergence of the clustering item and uses k-means to perform clustering.

Table 1: Clustering performance comparison on four datasets (mean±standard deviation). The optimal and suboptimal results are in bold and underlined, respectively.

Datasets	Metrics	DMVAE	MIB	CMIB-Nets	Completer	IMC-v1	IMC-v2	IMC
UCI-digits	ACC	90.95±0.62	83.30±1.27	85.70±1.15	91.28±1.41	90.01±0.60	84.01±1.15	**92.13±0.55**
	NMI	85.54±1.06	75.43±1.04	78.31±1.41	86.34±0.60	84.79±0.32	79.01±1.54	**88.01±0.73**
	ARI	85.40±1.54	76.16±1.55	76.97±1.64	86.67±0.86	84.78±0.43	78.18±0.88	**87.83±0.25**
Notting-Hill	ACC	76.22±1.21	81.65±1.37	85.40±2.36	80.17±2.79	77.79±2.00	84.83±1.90	**87.10±1.35**
	NMI	72.97±0.96	75.95±2.52	78.65±2.57	76.11±2.27	74.93±1.94	78.79±1.08	**80.67±1.42**
	ARI	69.50±0.77	71.91±1.61	**80.46±1.92**	71.48±3.29	70.30±2.15	79.10±2.32	80.19±1.74
BDGP	ACC	90.59±1.45	86.82±0.65	85.82±0.58	79.31±1.55	88.70±1.42	80.46±1.85	**91.46±0.82**
	NMI	**85.32±0.53**	80.82±0.86	81.60±0.66	74.25±0.69	81.47±1.12	73.31±2.45	84.40±1.20
	ARI	78.58±2.54	73.90±1.65	74.25±2.19	71.44±1.45	77.65±2.15	69.31±2.80	**80.25±1.62**
Caltech20	ACC	61.50±0.54	56.12±2.54	55.26±3.18	**62.31±2.65**	58.05±1.28	52.42±3.15	60.82±1.66
	NMI	68.32±1.23	63.28±2.66	62.44±2.56	**70.25±2.20**	65.75±2.20	61.76±3.40	69.20±1.48
	ARI	59.86±1.46	58.10±2.90	56.45±2.74	61.14±2.55	57.52±2.56	53.30±2.82	**61.42±2.12**

Limitations: The mathematical strategy for selecting the optimal trade-off parameters is a direction that can be studied in the future.

Thanks for your attention

Welcome to our poster session!