Supplementary Material for GCFAgg: Global and Cross-view Feature Aggregation for Multi-view Clustering

Weiqing Yan[1,5], Yuanyang Zhang[1], Chenlei Lv[2], Chang Tang[3*], Guanghui Yue[4], Liang Liao[5], Weisi Lin[5]
[1]School of Computer and Control Engineering, Yantai University, Yantai 264005, China
[2]College of Computer Science and Software Engineering, Shenzhen University, Shenzhen 518060, China
[3] School of Computer, China University of Geosciences, Wuhan 430074, China
[4] School of Biomedical Engineering, Health Science Center, Shenzhen University, Shenzhen 518060, China
[5] School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore.

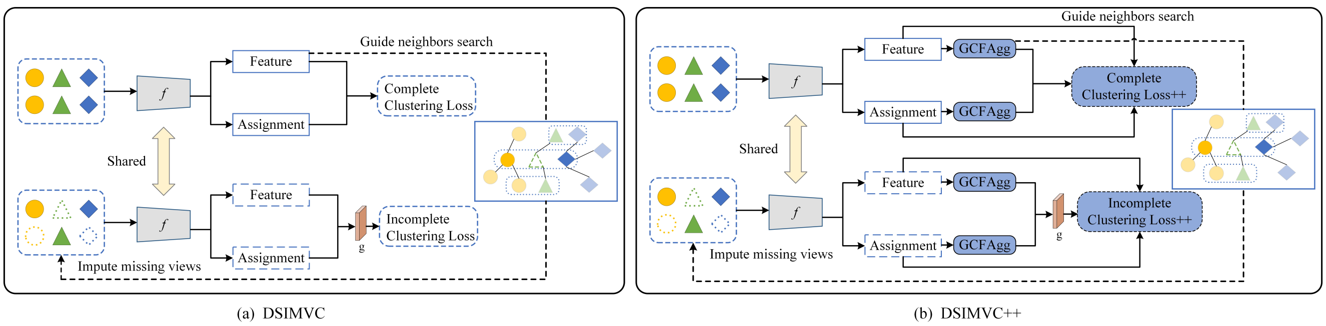

Figure 1. The caparison of DSIMVC [38] and DSIMVC++ framework for incomplete multi-view clustering. In DSIMVC++, the proposed GCFAgg module is integrated to obtain the consensus representation and consensus clustering assignment, and the clustering loss is enhanced by the proposed SgCL module.

1. The DSIMVC++ framework for incomplete multiview clustering

The proposed module is flexible multi-view data representation module for clustering, which can be also embedded to the incomplete multi-view data clustering task via plugging our module into other frameworks. In this section, the proposed module is embedded to the framework-DSIMVC for incomplete multi-view data clustering. The comparative figure is shown in Fig. 1. In DSIMVC++, we integrate the proposed GCFAgg module to obtain the consensus representation and the consensus clustering assignment. And then, they are input to the enhanced clustering loss module to achieve the alignment consistency.

In DSIMVC, the complete clustering loss includes the representation of each view alignment contrastive loss $\mathcal{L}_F\left(f(\mathcal{D}^c;w)\right)$, the cluster assignment loss $\mathcal{L}_C\left(f(\mathcal{D}^c;w)\right)$, and the clustering assignment regularization term $\mathcal{L}_R\left(f(\mathcal{D}^c;w)\right)$. In DSIMVC++, the first two losses are enhanced via the proposed Structure-guided Contrastive Loss (SgCL). Specifically, the enhanced $\mathcal{L}_{F_{++}}\left(f(\mathcal{D};w)\right)$ and $\mathcal{L}_{C_{++}}\left(f(\mathcal{D};w)\right)$ for complete clustering loss and incomplete clustering loss are shown in the following equation.

The view-specific representation alignment loss is enhanced via the proposed SgCL to achieve the structure-guided alignment among the consensus representations and the view-specific representations, which ensures that we only minimize the similarity between the view-specific representation and the consensus representation from different samples with low structure relationship. The loss is presented as follows:

$$\mathcal{L}_{F++}\left(f(\mathcal{D}^c;w)\right)=\sum_{i=1}^{n_c}\sum_{q=1}^{m}\Big[-\tfrac{2}{n_c}f_{\widehat{H}}(x_i;w_H)^T f_Z(x_i^q;w) + \tfrac{1}{n_c(n_c-1)}\sum_{j\neq i}\left(1-S_{i,j}\right)\left(f_{\widehat{H}}(x_i;w_H)^T, f_Z(x_j^q;w)\right)^2\Big] \tag{1}$$

where $\mathcal{D}^c$ denotes the complete sample set, $f_{\widehat{H}}(x_i;w_H)$ denotes the consensus representation feature of x_i from multiple views, $f_Z(x_i^q;w)$ denotes the q-th view representation of the i-th sample x_i, n_c denotes the number of complete samples, m denotes the number of views, $S_{i,j}$ denotes the global structure relationship obtained by the proposed GCFAgg module (See the Eq. (7)).

*Corresponding author(tangchang@cug.edu.cn)

The clustering assignment alignment loss is enhanced via the consensus clustering assignment probabilities and view-specific clustering assignment probabilities.

$$\mathcal{L}_{C++}\left(f(\mathcal{D}^c;w)\right) = -\frac{1}{K}\sum_{q=1}^{m}\sum_{j=1}^{K}\left[\log\frac{e^{\widehat{Q_j}^T Q_j^q}}{\sum\limits_{s\neq j} e^{\widehat{Q_j}^T Q_s^q}}\right] \quad (2)$$

where $\widehat{Q_j}$ denotes the j-th column of the learnt consensus clustering alignment probability $\widehat{Q}$ after our GCFAgg module, Q_j^q denotes the j-th column of the clustering alignment probability $Q^q \in \mathbb{R}^{n_c \times K}$ in the q-th view.

The clustering assignment regularization term in DSIMVC is used to prevent the trivial solution, it is presented as follows:

$$\mathcal{L}_R\left(f(\mathcal{D}^c;w)\right) = \sum_{p=1}^{m}\sum_{j=1}^{K}\left[\overline{Q}_j^p \log \overline{Q}_j^p\right] \quad (3)$$

where $\overline{Q}_j^p = \frac{1}{n_c}\sum\limits_{i=1}^{n_c} Q_{ij}^p$. In DSIMVC++, the total complete clustering loss is presented as follows:

$$\begin{aligned}\mathcal{L}_{++}\left(f(\mathcal{D}^c;w)\right) = \mathcal{L}_{F_{++}}\left(f(\mathcal{D}^c;w)\right) + \mathcal{L}_{C_{++}}\left(f(\mathcal{D}^c;w)\right) \\ +\mathcal{L}_R\left(f(\mathcal{D}^c;w)\right)\end{aligned} \quad (4)$$

In the incomplete clustering loss, a function $g(\mathcal{D}^e;\varphi)$ in the DSIMVC is added to dynamically select incomplete samples for training. The incomplete clustering loss is presented as follows:

$$\begin{aligned}&\mathcal{L}_{++}\left(f(\mathcal{D}^e;w), g(\mathcal{D}^e;\varphi)\right) = g\left(\widetilde{x^e};w\right) \\ &\left(L_{F++}\left(f(\mathcal{D}^e;w)\right) + L_{C++}\left(f(\mathcal{D}^e;w)\right) + L_R\left(f(\mathcal{D}^e;w)\right)\right)\end{aligned} \quad (5)$$

where $\mathcal{D}^e$ is the filled incomplete sample set, $\widetilde{x^e}$ denotes the imputed sample from $\mathcal{D}^e$ (Please refer the [38] for its computation in detail).

In DSIMVC++, the total loss is presented as follows:

$$\mathcal{L}_{++} = \mathcal{L}_{++}\left(f(\mathcal{D}^c;w)\right) + \mathcal{L}_{++}\left(f(\mathcal{D}^e;w), g(\mathcal{D}^e;\varphi)\right) \quad (6)$$

Compared with the loss of DSIMVC, there are the following advantages: we align the consensus presentation and view-specific representation, which makes the representations of these samples with high structure relationship be more similar. Further, the learnt global structure relationship obtained by the GCFAgg module is integrated to the contrastive learning, which ensures that we only minimize the similarity between the view-specific representation and the consensus representation from different samples with low structure relationship. The comparison results shown in Table 5 verify the effectiveness of our method.

2. Ablation study for the SgCL

In the experiment, we set the contrastive learning with the sample-level loss as Standard CL. That is, these interview presentations from the same sample are set as positive pairs, and view representations from different samples are set as negative pairs (such as the contrastive learning in [38,51]). The ablation study for the SgCL is show in Table 1. The experiment shows the effectiveness of our SgCL compared with the standard CL.

Table 1. The ablation study for the SgCL.

Datasets	Method	ACC	NMI	PUR
CCV	Standard CL	0.2711	0.2669	0.3046
	Standard CL with **S**	0.3046	0.3017	0.3363
	SgCL without **S**	0.2858	0.2833	0.3260
	SgCL	**0.3543**	**0.3292**	**0.3812**
MNIST-USPS	Standard CL	0.9562	0.9386	0.9562
	Standard CL with **S**	0.9768	0.9527	0.9768
	SgCL without **S**	0.9698	0.9327	0.9698
	SgCL	**0.9956**	**0.9871**	**0.9956**