Structure-aware Granular-Ball based Information Bottleneck for Multi-modal Clustering

Zhengzheng Lou, Yuhan Zhan, Mingyang Lv, Yingxuan Li,
Yuyang Du, Shizhe Hu†

School of Computer and Artificial Intelligence

Zhengzhou University

Zhengzhou, Henan, China

A scenic campus of Zhengzhou University in western suburbs

Zhengzhou： A grand ancient capital embracing modern prosperity

Outline

- Problem background
- Previous works
- Our proposal
- Experiments
- Conclusion

Outline

- **Problem background**
- Previous works
- Our proposal
- Experiments
- Conclusion

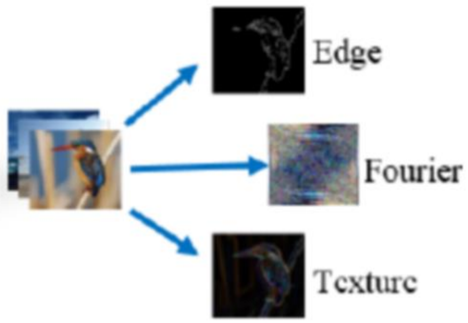
Characteristics of multi-modal datasets

Era: In the big data era, multi-modal data is diverse and heterogeneous.

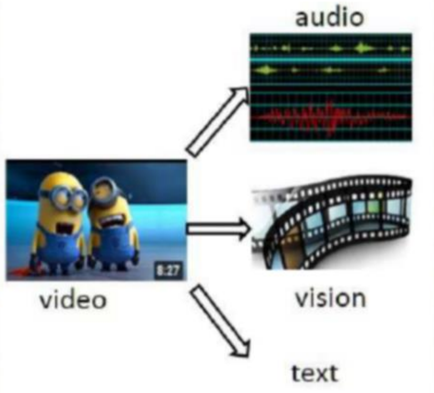
Multiple forms

Multiple data sources

Multiple perspectives



Multiple features



Multiple modes

Multiple formats

Heterogeneous

Outline

- Problem background
- **Previous works**
- Our proposal
- Experiments
- Conclusion

Previous multi-modal clustering methods

- Graph-based methods: Struggle with global structure modeling and are sensitive to modality-specific variations;
- Matrix decomposition/subspace methods: Lack explicit handling of inter-modal consistency and are often limited to linear assumptions;
- Information-theoretic paradigms: Lack coarse-grained modeling and depend on individual sample information.

Multi-modal Clustering

Prior Knowledge

The Information Bottleneck (IB) Principle:

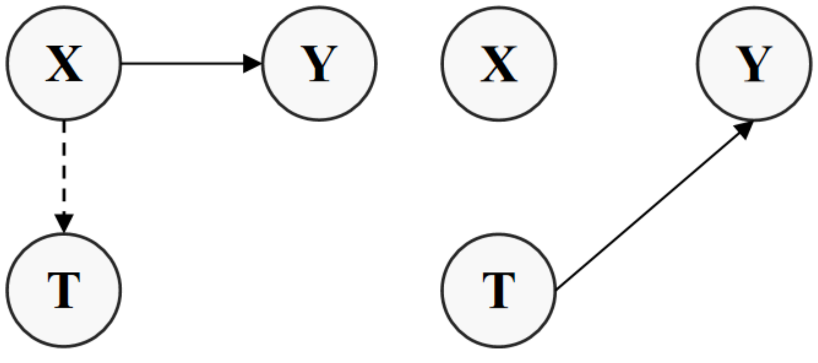
Core Idea:

Compresses X into compact T, maximizing I(T;Y) while minimizing I(T;X).

IB Objective:

max I(T;Y) - β^{-1}I(T;X)

- ○I(T;Y): Preserve discriminative information
- ○I(T;X): Compress redundant information
- ○β: Trade-off parameter (larger β → more preservation)



(a) Compression of X (b) Preservation of Y

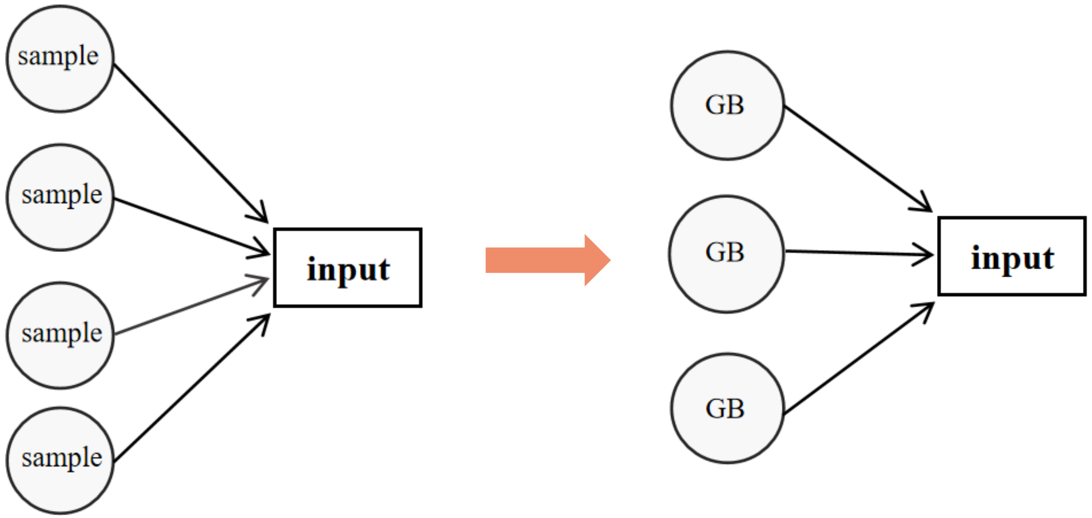
Granular-Ball(GB): The Right Abstraction

Core Idea:

Instead of operating at the individual sample level, we use GBs as the input.

Why?

GBs group local samples into coherent, structured units for multi-granularity modeling.



Outline

- Problem background
- Previous works
- **Our proposal**
- Experiments
- Conclusion

Our proposed method

- Structure-Aware Granular-Ball based Information Bottleneck (SGB-IB):
  - Granular-Ball generation;
  - Structure-aware Granular-Ball Constraints;
  - Information Preservation & Compression;
  - Structure-Aware IB Optimization;
  - Clustering Output.

Granular-Ball Generation

Goal: Bridge the semantic gap between samples and clusters

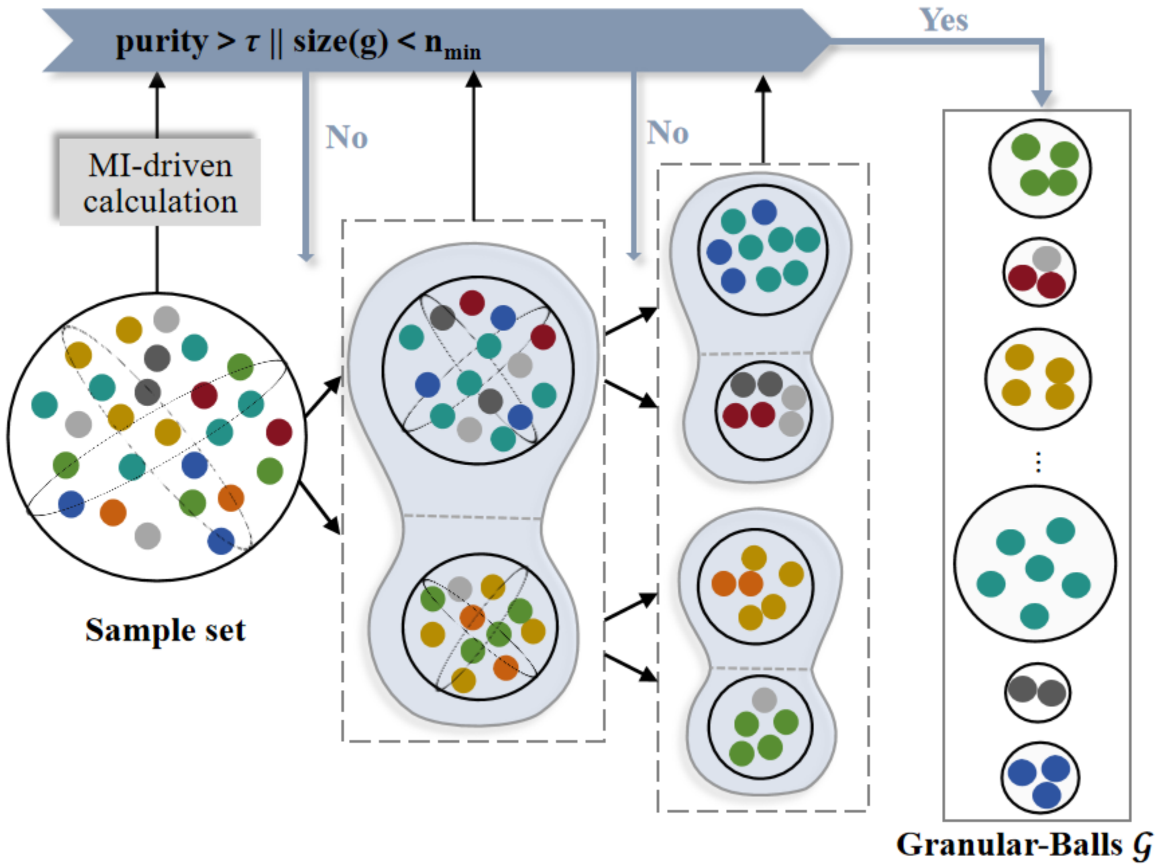
Process:

▫Initialize the entire dataset as a single GB;

▫Calculate purity of the GB using mutual information (MI);

▫Recursively split the GB into two if purity<τ and size>n_{min};

▫Stop when all GBs meet the purity or minimum size criteria.



Structure-aware Granular-Ball Constraints

For a GB g assigned to cluster t, its structural relevance SR is defined as:

$$\mathrm{SR}_g = \frac{\mathrm{sim}(g,t)}{\frac{1}{K-1}\sum_{t'\neq t}\mathrm{sim}(g,t')}$$

where sim(g,t) is the average cosine similarity between the multi-modal center of g and all GBs in cluster t.

- $SR_g > 1$ → High-quality GB (well-aligned with its cluster)
- $SR_g \leq 1$ → Low-quality GB (weak structural consistency)

Based on SR_g, we design the structure-aware term as:

$$S_{\mathcal{G}} = \sum_{g=1}^{|\mathcal{G}|}\left(1 - \alpha\left(1 - e^{-\mathrm{SR}_g}\right)\right)$$

Objective Function

We propose a novel Structure-aware Granular-Ball based Information Bottleneck (SGB-IB) method:

$$\mathcal{L} = S_{\mathcal{G}} \cdot \sum_{i=1}^{m} I(T; Y^{i}) - \beta^{-1} I(T; X).$$

Information Preservation & Compression

Preservation:

▫Keep discriminative cross-modal information

▫Multi-modal centers serve as stable anchors

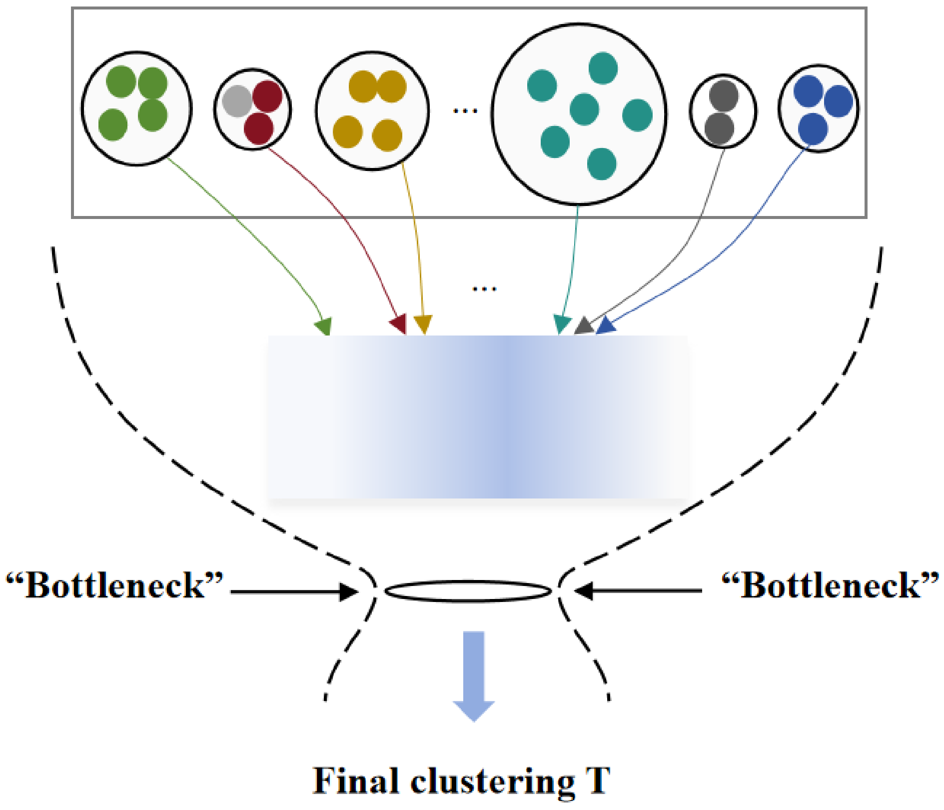
▫Enables T to capture commonalities and specificities

Compression:

▫Filter modality-specific redundancy

▫Local coherence of GBs naturally suppresses sample-level variations

▫Forces T to rely on compact, generalizable patterns



Advantages of the SGB-IB

- GB captures inherent data topology;
- It adopts multi-granularity strategies to fit complex distributions;
- Structure-aware weight balances the contribution of GBs;
- Structural quality explicitly embedded into IB objective;
- GB aggregation extracts coarse-grained structures.

Optimization Strategy

Algorithm 1 The Optimizing of SGB-IB

1: **Input:** Multi-modal data $\{X^{i}\}_{i=1}^{m}$, number of clusters K, hyperparameters α,β, thresholds n_{min} and τ.
2: **Output:** Final clustering results.
3: Initialize all sa mples as one single GB.
4: Perform recursive binary splitting until the final GB set $\mathcal{G}$= $\{g_1, g_2, ..., g_{|\mathcal{G}|}\}$ is obtained.
5: Randomly partition the GB set $\mathcal{G}$into K clusters.
6: **repeat**
7: Recompute $S_{\mathcal{G}}$ via Eq. (7).
8: **for** each GB g **do**
9: Extract g from t^{old} to form a singleton cluster $\{g\}$.
10: Compute the optimal target cluster t^{new} = arg $\min_{t\in T} \triangle\mathcal{L}(\{g\}, t)$.
11: Merge g into t^{new} (or back into t^{old} if unchanged).
12: **end for**
13: **until** Data partition unchanged or a fixed number of iterations finished.
14: **Return** the final clustering results.

Outline

- Problem background
- Previous works
- Our proposal
- **Experiments**
- Conclusion

Experimental Setup

Datasets:

Dataset	Type	#Modality	#Sample	#Cluster
SOCCER	Image	3	280	7
Reuters	Text	3	1200	6
BBC	Text	4	685	5
WVU	Video	4	650	10
COIL100	Image	2	7200	100

Metrics:

Accuracy (ACC)

Normalized Mutual Information (NMI)

$$Acc = \frac{\sum_{i=1}^{m} \delta(r_i, map(p_i))}{m}$$

$$NMI = \frac{I(V;C)}{\max(H(V), H(C))}$$

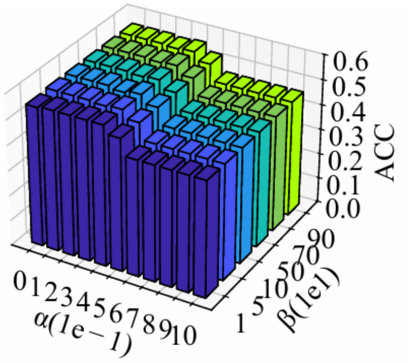
Compared methods

1) **Single-modal Clustering**:
   (1) K-Means (KM)
   (2) Ncuts.
2) **All-modal Clustering**:
   (1) KM-All
   (2) Ncuts-All.
3) **Multi-modal Clustering**:
   (1) LMSC (Zhang et al., 2017, CVPR)
   (2) MLAN (Nie et al., 2018, TIP)
   (3) GMC (Wang et al., 2020,TKDE)
   (4) SMVSC (Sun et al., 2021, ACM MM)
   (5) FPMVS-CAG (Wang et al., 2022, TIP)
   (6) OMSC (Chen et al., 2022,KDD)
   (7) FIMVC-VIA (Liu et al.,2024b, TNNLS)
   (8) AEVC (Liu et al.,2024a, CVPR)
   (9) INMKC (Feng et al.,2025,AAAI)
   (10)ASCR (Xu et al., 2025, AAAI)
   (11)ALPC (Chen et al., 2025, AAAI).

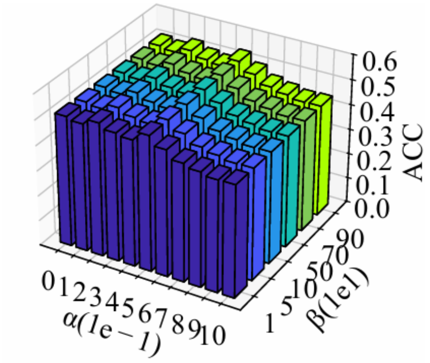
Clustering results

Methods	SOCCER		Reuters		BBC		WVU		COIL100	
	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
KM	26.24±5.0	19.37±7.2	25.22±6.8	10.89±7.1	37.80±4.8	11.59±8.2	27.42±4.3	33.87±5.9	27.58±1.7	57.57±2.0
IB	46.74±1.3	44.38±2.1	41.98±3.2	25.22±2.4	72.90±9.0	60.32±6.3	52.23±2.1	50.75±0.7	40.85±1.6	68.54±0.8
AmKM	21.73±2.1	7.06±2.3	21.50±6.0	4.73±5.3	40.73±6.7	12.79±11.0	27.22±5.4	24.28±7.0	30.23±3.4	51.56±3.0
AmIB	50.72±1.6	45.12±1.3	29.25±2.4	19.93±1.9	79.27±8.2	72.75±7.3	53.48±1.6	51.37±0.4	38.89±2.1	67.94±2.5
LMSC	47.31±6.3	44.56±5.1	47.39±3.1	27.48±2.3	71.96±6.4	53.21±5.3	57.31±4.2	58.68±3.0	50.37±0.8	67.97±0.3
MLAN	28.76±2.2	21.23±3.6	21.32±0.0	8.13±0.0	69.23±7.4	49.02±5.6	37.82±0.0	34.96±0.0	44.35±1.1	61.36±1.0
GMC	29.29±0.0	25.82±0.0	19.92±0.0	13.66±0.0	69.34±0.0	56.28±0.0	46.78±0.0	54.92±0.0	37.60±0.0	47.03±0.0
SMVSC	28.03±0.0	9.29±0.0	49.21±2.2	30.32±2.0	53.52±4.3	29.98±3.8	45.43±2.7	44.89±2.2	50.33±1.2	**70.13±0.8**
FPMVS-CAG	22.14±0.0	5.29±0.0	46.50±0.0	24.69±0.0	79.50±0.0	72.15±0.0	49.21±0.0	49.52±0.0	48.72±0.0	69.61±0.0
OMSC	24.29±0.0	6.70±0.0	50.58±0.0	25.62±0.0	65.11±0.0	54.86±0.0	52.69±0.0	56.22±0.0	48.10±0.0	67.77±0.0
FIMVC-VIA	50.81±0.2	44.14±0.2	48.44±1.0	27.24±0.6	67.32±0.5	41.63±0.3	53.83±0.8	48.66±0.7	39.14±0.6	56.56±0.3
AEVC	40.57±2.7	24.67±2.6	45.25±0.3	22.10±0.2	69.51±3.0	41.27±0.7	61.54±1.4	64.13±1.1	51.39±1.0	69.67±0.3
INMKC	34.13±0.0	15.36±0.0	51.42±0.0	29.29±0.0	65.99±0.0	43.45±0.0	62.52±0.0	59.68±0.0	51.69±0.0	68.72±0.0
ASCR	24.11±1.2	10.59±1.3	40.23±0.8	26.88±0.2	63.12±0.4	51.40±0.8	43.34±3.2	51.86±2.4	45.76±1.3	69.53±0.6
ALPC	49.29±0.0	34.20±0.0	43.58±0.0	20.69±0.0	74.60±0.0	56.25±0.0	59.94±0.0	65.06±0.0	38.07±0.0	55.02±0.0
Ours	**55.71±0.0**	**45.31±0.0**	**53.56±1.8**	**34.49±1.0**	**90.95±3.7**	**75.26±2.5**	**65.23±1.8**	**71.83±0.8**	**53.94±1.1**	69.86±0.6

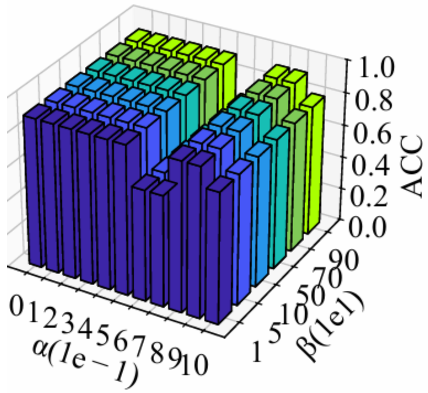
Parameter Sensitivity on five datasets



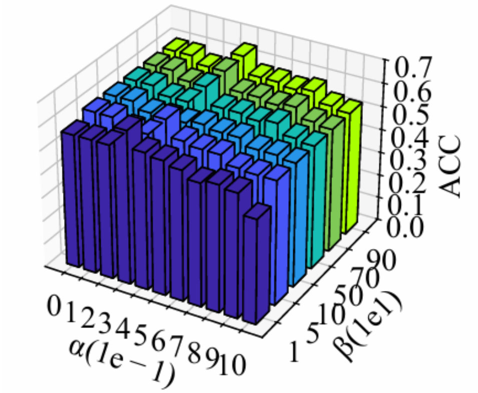
(a)



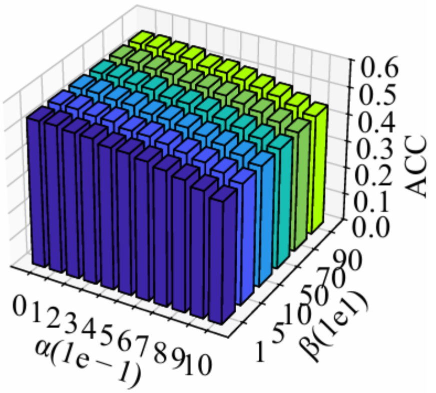
(b)



(c)



(d)

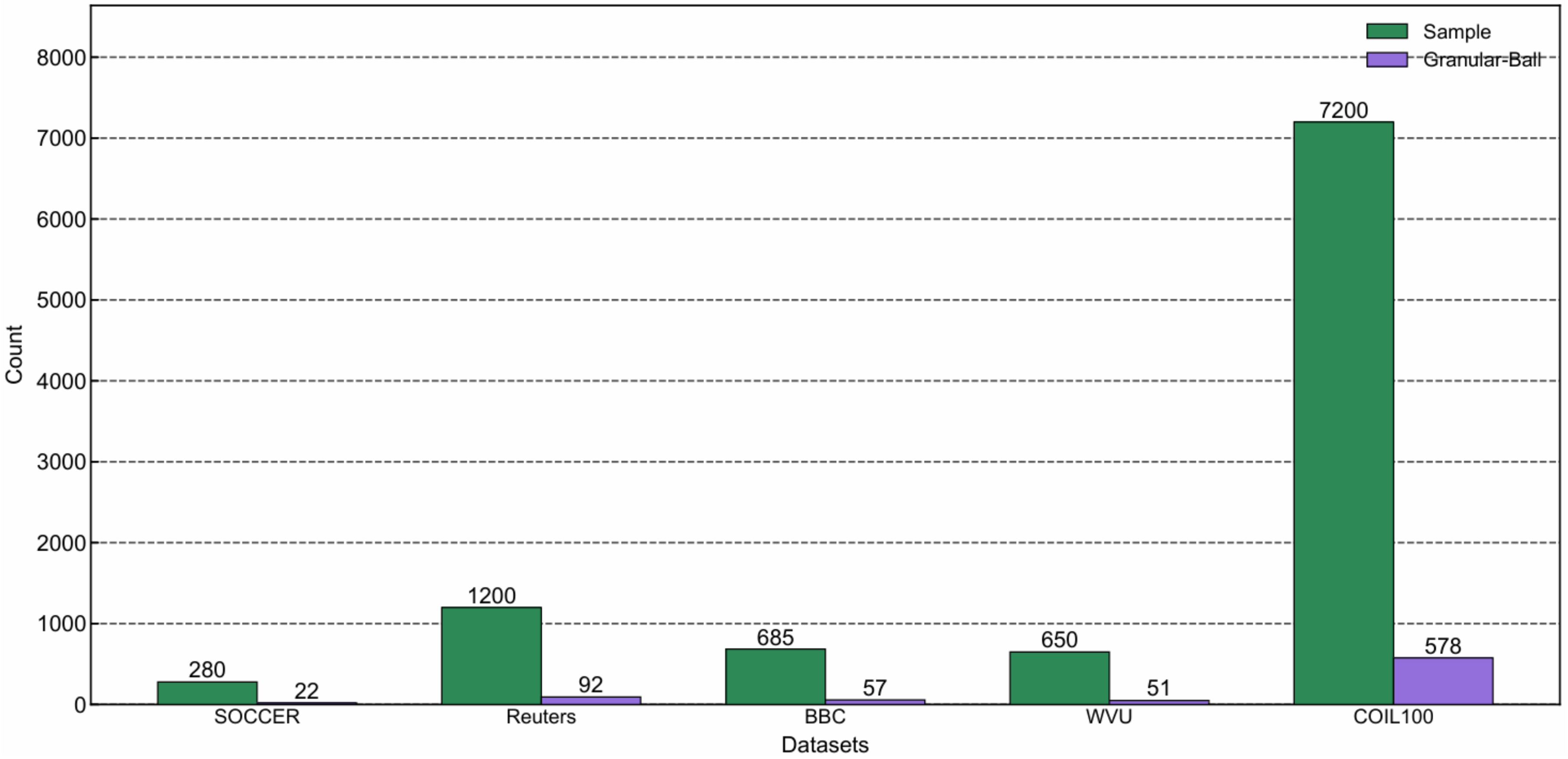


(e)

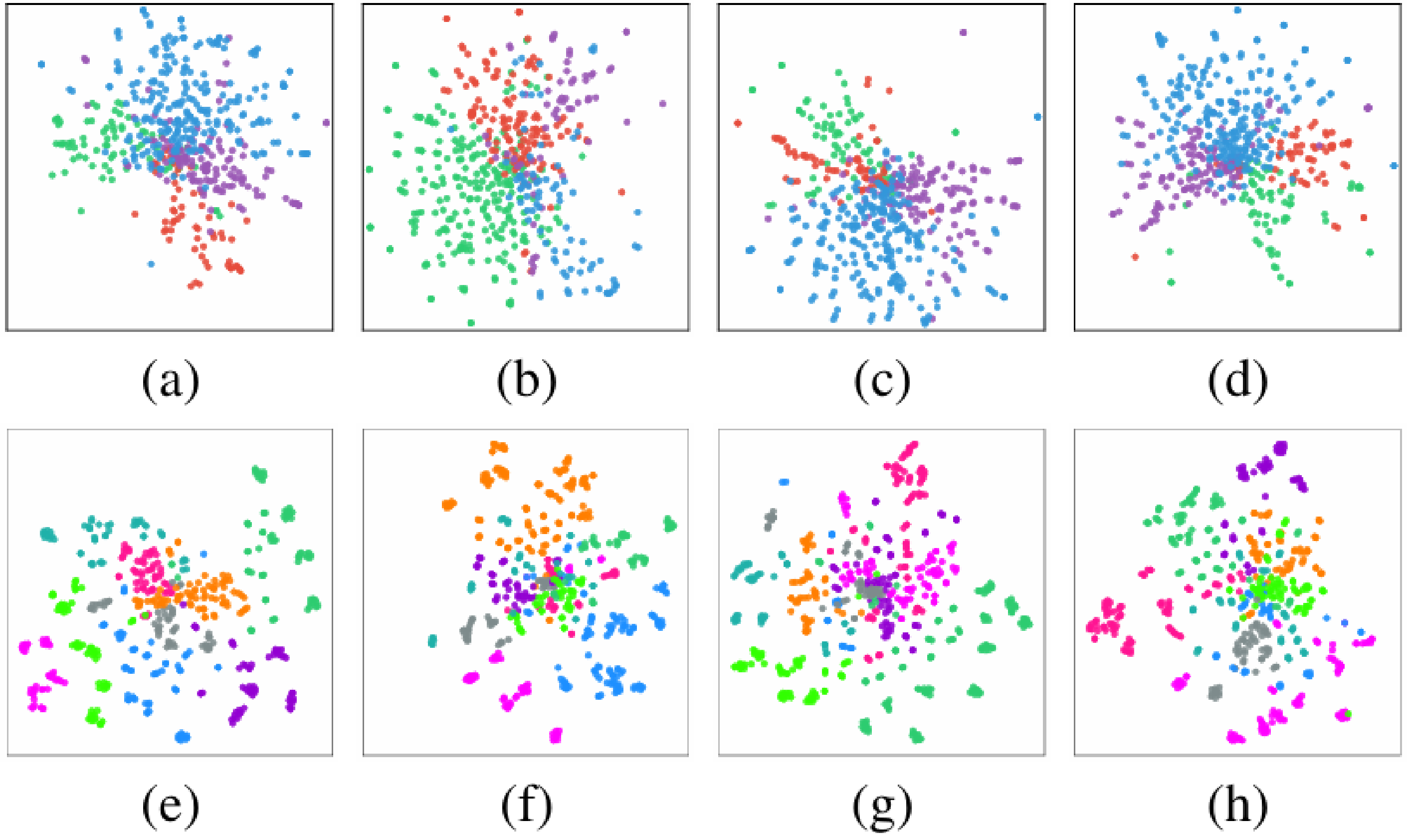
(a)–(e):SOCCER,Reuters,BBC,WVU,COIL100.

Count Statistics of samples and granular-balls

Statistical comparison between raw samples and generated granular-balls across different datasets.



T-SNE visualization of Clustering results on BBC and WVU datasets



(a)–(d) BBC Modal 1–4; (e)–(h) WVU Modal 1–4.

Outline

- Problem background
- Previous works
- Our proposal
- Experiments
- **Conclusion**

Summary

- Propose a novel Structure-aware Granular-Ball Information Bottleneck (SGB-IB) method for multi-modal clustering, integrating GB learning into the IB framework.
- Introduce a recursive splitting mechanism driven by purity to generate adaptive GB, serving as units for information learning.
- Design a structure-aware weighting mechanism to differentiate the contributions of GBs.
- Our approach achieves state-of-the-art performance.

Thank You!

Contact for communication:

ieshizhehu@gmail.com