A Peer-review Look on Multi-modal Clustering: An Information Bottleneck Realization Method

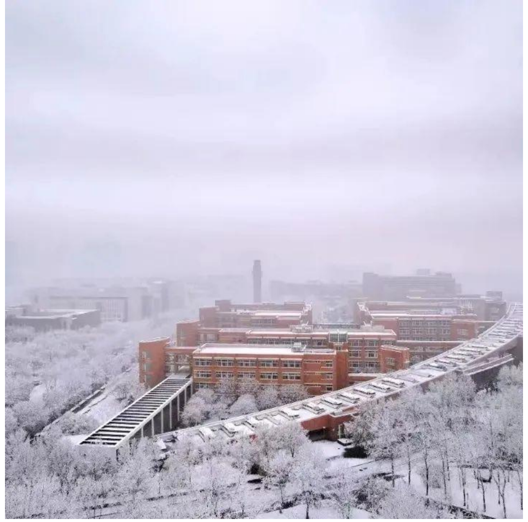
Zhengzheng Lou, Hang Xue, Chaoyang Zhang, Shizhe Hu†

School of Computer and Artificial Intelligence

Zhengzhou University

Zhengzhou, Henan, China

Zhengzhou University (Also called “Western Park of Zhengzhou”)



Tourist Spot

Outline

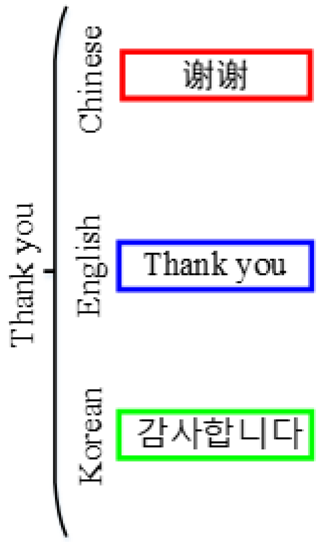
- Problem background
- Previous works
- Our proposal
- Experiments
- Conclusion

Outline

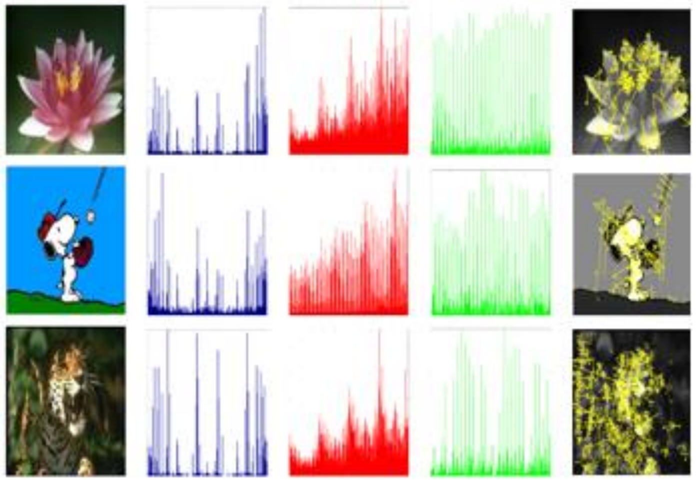
- **Problem background**
- Previous works
- Our proposal
- Experiments
- Conclusion

Characteristics of multi-modal datasets

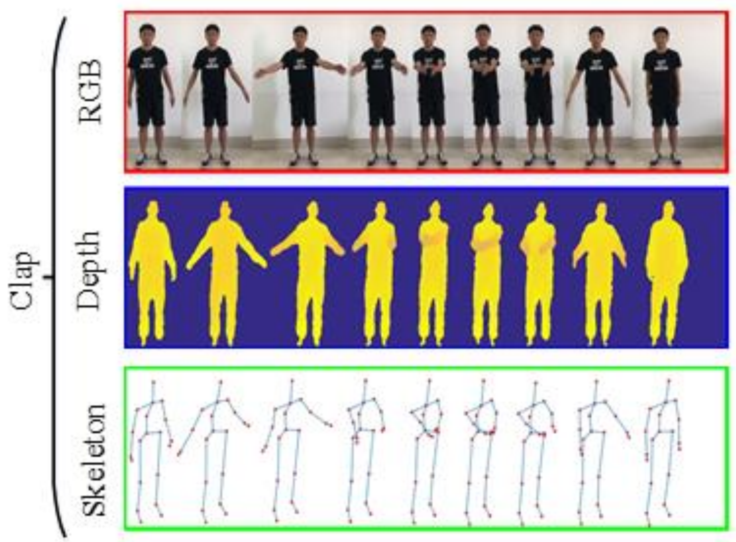
In Big Data era, many kinds of multi-modal data are emerging.



Multi-lingual Text



Multi-feature Image



Multi-modal human action video

Property: **Heterogeneous, Large-scale, Diversification, Complexity**

Limitations of supervised multi-modal classification methods

1. **Time-consuming and cost-expensive for labelling;**
2. **Over-reliance on the label information of trained data;**
3. **Ignoring the characteristics of the input data itself.**

Multi-modal Clustering

Outline

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

Previous multi-modal clustering methods

- **Weighted-based methods;**
- **Shared feature learning based methods;**
- **Tensor representation learning based methods;**
- **Multi-modal consensus clustering;**
- **Multi-modal co-clustering;**
- **Multi-modal subspace clustering.**

Previous multi-modal clustering methods

- **Weighted-based methods:**
  - *discovering the complementary relationship and learning the consistent clustering structure using the learned modal weights.*

Limitations:

- *Lack of trustworthiness in learned weights.*
- *Learning weights in an isolated view.*
- *Extra weight parameters for controlling the weight distribution.*

Outline

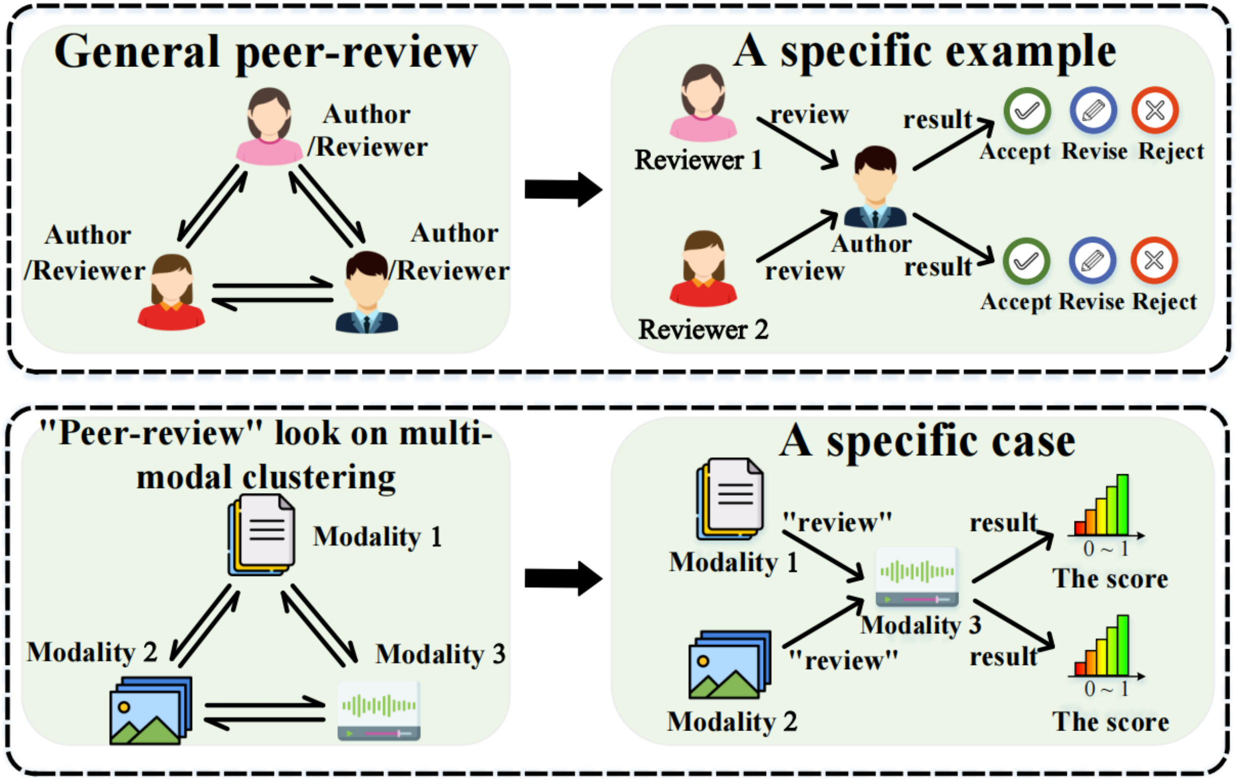
- Problem background
- Previous works
- **Our proposal**
- Experiments
- Conclusion

Our proposed method

- Peer-review Trustworthy Information Bottleneck (PTIB) :
  - Peer-review Score;
  - Trustworthy Score;
  - Modality Weight Learning;
  - Objective function.

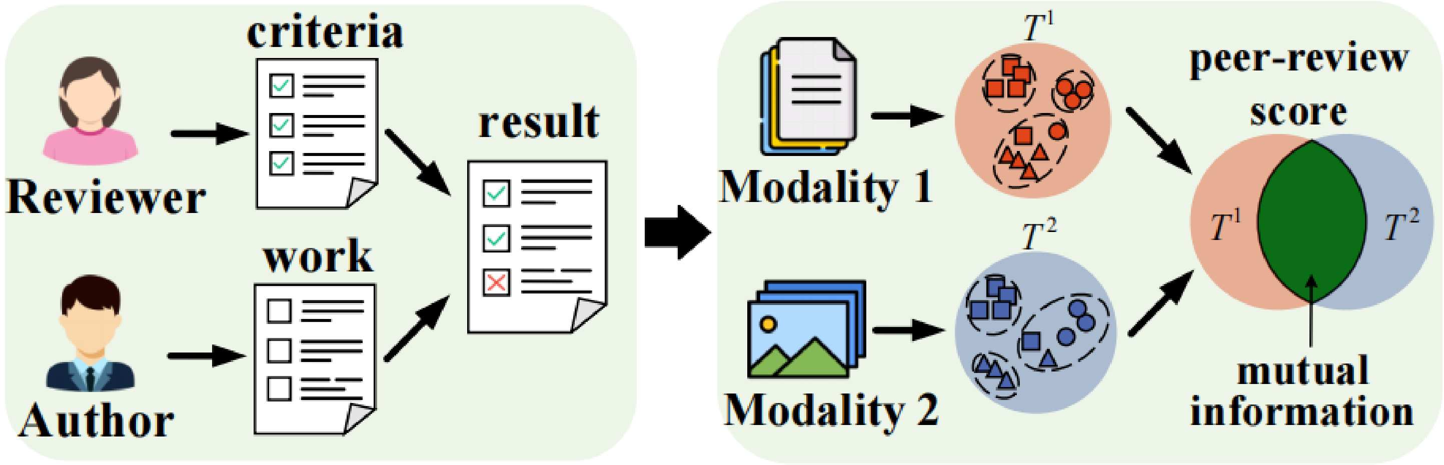
Peer-review Score

From the "peer-review" look on multi-modal clustering, one modality can either be an "author" or a "reviewer". The "reviewer" modalities review the work of the "author" modality and produces feedback review scores to evaluate the contribution.



Peer-review Score

It adopts the local clustering result of the modality as the "author" work or the "reviewer" criteria. The peer-review score depends on how similar the work is to the criteria, and the normalized mutual information is adopted to quantify this.



Peer-review Score

$$\mu_i^k = \frac{2 \times I(T^i, T^k)}{H(T^i) + H(T^k)}$$

$$\mu^k = \{\mu_1^k, \ldots, \mu_i^k, \ldots, \mu_m^k\}, i \neq k.$$

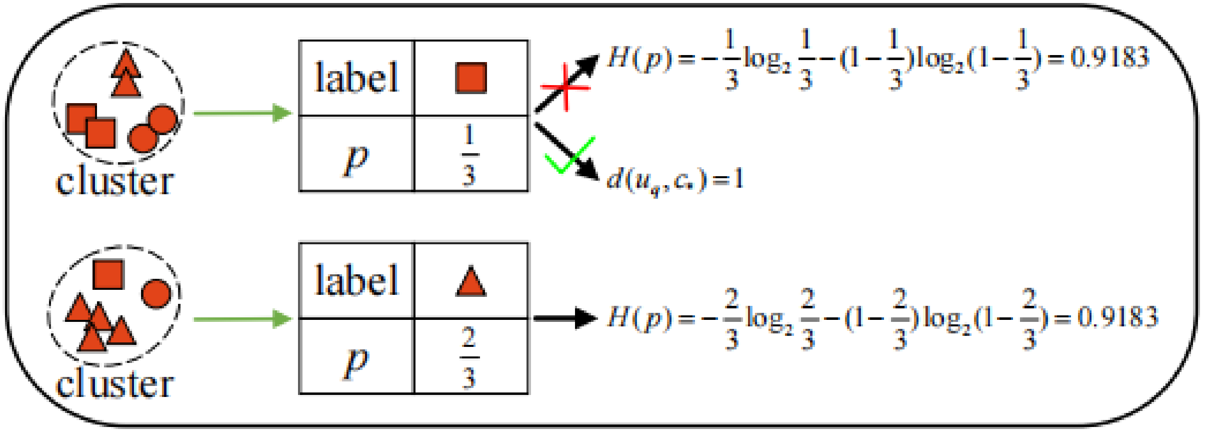
Trustworthy Score

We propose to regard the final clustering result as the "EIC / AE", which is then used to evaluate the trustworthiness of "reviewer" modalities in a self-supervision fashion.

Definition (Major/Minor Category):

Given a multi-modal dataset, if the local clustering result of modality is supervised by the final clustering result, the category of correctly assigned samples in a cluster of a specific local clustering result is called the major category, and the set of categories of incorrectly assigned samples in it is called the minor categories.

$$H(p) = -p\log_2 p - (1-p)\log_2(1-p).$$

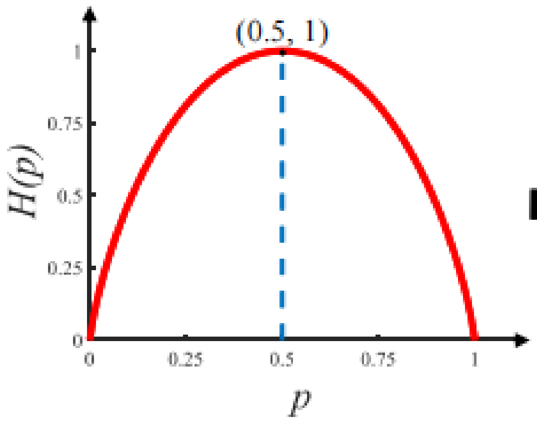


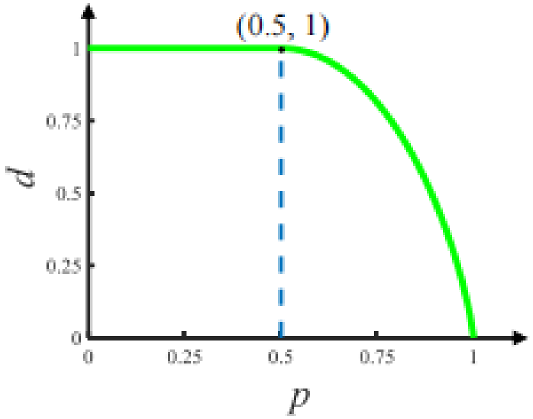
Trustworthy Score

Cluster Distortion

$$d(u_q, c_*) = \begin{cases} 1, & \text{if } 0 \le p < \frac{1}{2}, \\ H(p), & \text{if } \frac{1}{2} \le p \le 1. \end{cases}$$

$$D(U, C) = \frac{1}{|U|} \sum_{q=1}^{|U|} d(u_q, c_*).$$



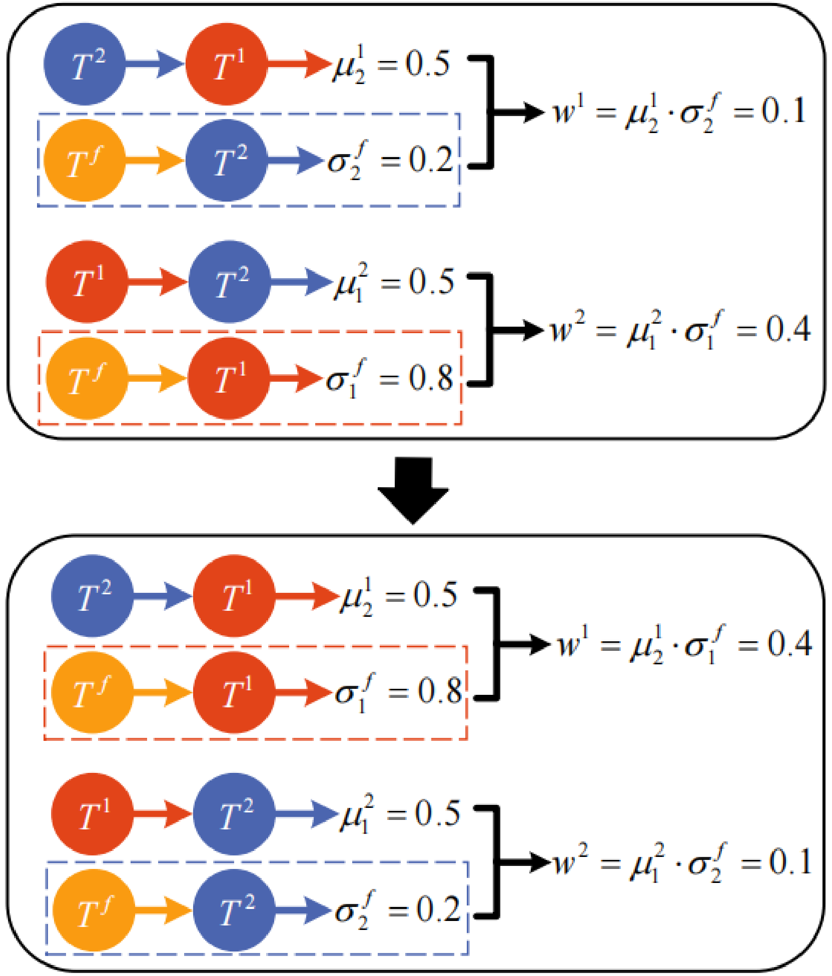


Trustworthy Score

$$\sigma_i^f = \frac{1}{D(T^i, T^f)}$$

$$\sigma^k = \{\sigma_1^f, \ldots, \sigma_i^f, \ldots, \sigma_m^f\}, i \neq k$$

Modality Weight Learning



$$w^k = \mu^k \bullet \sigma^k = \sum_{i=1, i \neq k}^{m} \mu_i^k \cdot \sigma_i^f, m > 2$$

$$w^k = \begin{cases} \sum_{i=1, i \neq k}^{m} \mu_i^k \cdot \sigma_k^f, & \text{if } m = 2, \\ \sum_{i=1, i \neq k}^{m} \mu_i^k \cdot \sigma_i^f, & \text{if } m > 2. \end{cases}$$

Objective function of PTIB

We propose a novel Peer-review Trustworthy Information Bottleneck method:

$$\mathcal{F}_{max}[p(t|x)] = \sum_{i=1}^{m} w^i \cdot [I(T;Y^i) - \beta^{-1} I(T;X)]$$

Advantages of the PTIB

- Trustworthy weight learning;
- Correlation quantization-based learning;
- Parameter-free weight learning;
- Self-supervision mechanism.

Optimization method

Algorithm 1 The Proposed PTIB

Input: m joint distributions $\{p(X, Y^i)\}_{i=1}^{m}$, the number of clusters $|T|$, the balance parameter β.
Output: Final clustering result $p(t|x)$.
Modality Weight Initialization: Compute the initial modality weights with initial peer-review and trustworthy score;
Random Clustering: $T \leftarrow$ Random partition of $\mathcal{X}$ into $|T|$ clusters;
repeat
  for all $x \in \mathcal{X}$ **do**
    Draw: Draw x from the "old" cluster t^{old} to become a separate cluster $\{x\}$;
    Merger: Select a "new" cluster t^{new} for the separate cluster $\{x\}$ to merge corresponding to the minimal merger cost in Theorem A.2;
  end for
  Update the trustworthy score using the clustering result in each iteration;
  Update the weight for each modality;
until Samples in different clusters remain unchanged or a fixed number of iterations.

Outline

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

Datasets

Dataset	Type	# Modality	# Samples	# Clusters
20NG	Text	3	500	5
COIL20	Image	3	1440	20
Event	Image	3	1579	8
Soccer	Image	3	280	7
17Flowers	Image	3	1360	17
75Flowers	Image	2	5514	75
COIL100	Image	2	7200	100
MMI	Video	2	1760	22


COIL20


Soccer


17Flowers


MMI

Compared methods

1) **Single-modal Clustering**: K-Means (KM) and Ncuts.

2) **All-modal Clustering**: KM-All, Ncuts-All.

3) **Multi-modal Clustering**:
   (1) MVIB: It is the first multi-view IB method proposed to address the document clustering problems.
   (2) Co(reg): It co-regularizes the data clustering hypotheses among views to learn consistent assignments.
   (3) MfIB: It is a weighted multi-feature IB method designed for solving the unsupervised image classification.
   (4) RMSC: It solves the noisy multi-view clustering problem by designing a robust spectral method.
   (5) LMSC: It learns latent shared representations among views to make the feature subspace more robust and accurate.
   (6) MLAN: It automatically tunes the view weights without using parameters.
   (7) GMC: It is a graph-based weighted multi-view clustering method by automatically tuning the parameters.
   (8) DMIB: It jointly considers the dual correlations about the cross-feature and cross-cluster view correlations.
   (9) FPMVS-CAG: It deals the multi-view subspace clustering with the guidance of selected consensus anchors.
   (10)MCMLE: It improves the traditional Ncuts method for multi-view clustering by Laplacian embedding to learn a shared binary assignment matrix among different modalities.
   (11)TBGL: It focuses on learning tensorized bipartite graphs and considering the intra/inter-view similarities.
   (12)TIM: It works by following three principles, i.e., contained, complementary and compatible principle.
   (13)SMVAGC-SF: It adaptively fuses multiple anchor graphs with different magnitudes.

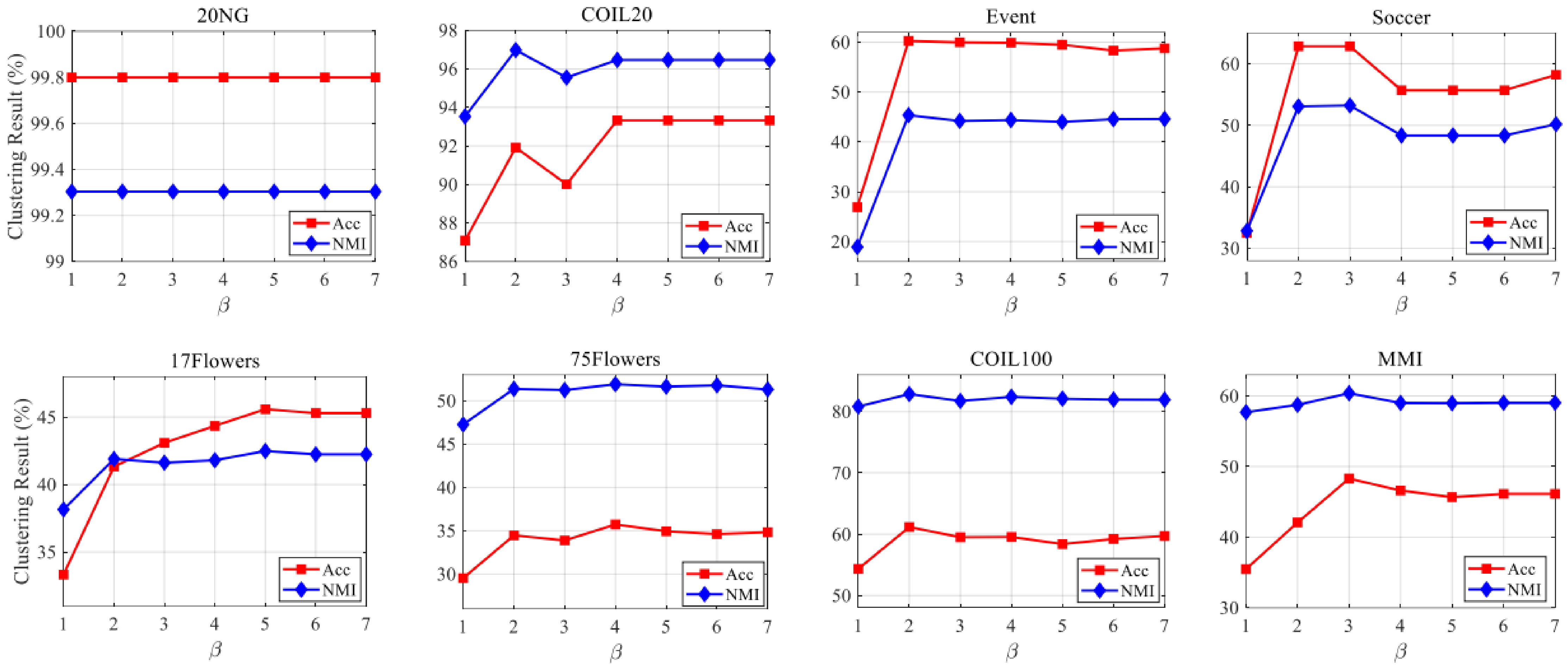
Clustering results

Method	20NG		COIL20		Event		Soccer	
	Acc	NMI	Acc	NMI	Acc	NMI	Acc	NMI
KM	22.28±1.48	4.45±2.32	53.06±3.20	65.06±2.12	33.93±4.10	19.84±2.69	25.82±5.06	18.70±7.97
Ncuts (TPAMI'00)	42.80±2.40	27.65±2.01	74.69±1.30	84.01±0.54	34.10±1.28	14.97±0.40	48.21±1.14	45.02±2.21
KM-All	21.46±0.68	1.76±0.65	46.14±6.58	60.70±4.51	28.85±2.29	11.37±2.10	22.46±3.94	8.14±3.59
Ncuts-All (TPAMI'00)	71.20±0.17	57.23±0.10	46.14±0.52	57.93±0.23	35.06±0.69	20.11±0.85	39.75±0.94	34.04±0.57
MVIB (DASFAA'07)	94.22±1.37	83.21±3.18	61.74±10.51	73.65±6.63	40.02±2.04	23.71±1.56	35.79±3.96	21.42±4.25
Co(reg) (NeurIPS'11)	20.02±0.62	3.15±0.54	64.33±1.68	83.79±0.45	38.58±0.92	24.30±0.55	24.13±0.53	11.43±0.39
MfIB (IJCAI'13)	93.76±2.89	85.11±4.54	83.81±4.29	92.39±1.97	48.58±1.50	33.41±1.35	53.64±2.76	49.74±3.44
RMSC (AAAI'14)	37.26±0.91	15.70±0.84	65.43±3.31	79.16±2.35	36.58±1.26	21.02±0.88	28.96±1.90	12.16±2.18
LMSC (CVPR'17)	96.16±0.57	88.37±1.54	71.94±2.72	82.18±2.37	43.92±2.84	27.53±2.58	31.25±6.53	15.85±8.71
MLAN (TIP'18)	96.40±0.11	89.18±0.17	87.22±2.30 ○	94.35±1.10 ○	19.90±0.72	6.66±0.80	28.21±0.01	21.27±0.17
GMC (TKDE'20)	98.20±0.00	93.92±0.00	60.90±0.00	84.67±0.00	18.11±0.00	10.74±0.00	29.29±0.00	25.82±0.00
DMIB (TCYB'22)	98.30±0.14	97.56±0.49	65.90±4.03	77.70±2.46	49.80±3.02	32.97±2.38	54.07±3.67	50.68±2.23 ○
FPMVS-CAG (TIP'22)	73.80±0.00	59.23±0.00	69.17±0.00	85.11±0.00	48.89±0.00	31.99±0.00	50.14±0.00	49.56±0.00
MCMLE (TPAMI'22)	77.40±0.00	69.96±0.00	85.83±0.00	93.48±0.00	44.46±0.00	30.24±0.00	56.07±0.00 ○	50.06±0.00
TBGL (TPAMI'23)	89.11±0.00	83.45±0.00	86.10±0.00	92.41±0.00	42.84±0.00	28.40±0.00	54.39±0.00	49.78±0.00
TIM (TIP'23)	99.40±0.00 ○	98.08±0.00 ○	56.70±4.08	71.39±0.29	54.60±2.50	36.86±1.75	48.93±0.51	41.42±4.09
SMVAGC-SF (TIP'24)	86.07±6.40	72.61±3.59	75.66±5.10	89.43±2.11	54.76±1.27○	36.97±0.65○	45.14±1.56	29.61±1.85
PTIB	**99.80±0.00** ●	**99.30±0.00** ●	**93.33±0.00** ●	**96.46±0.00** ●	**60.24±0.16** ●	**45.36±0.28** ●	**62.86±0.17** ●	**53.23±0.16** ●
Improve (● VS ○)	0.40 (↑)	1.22 (↑)	6.11 (↑)	2.11 (↑)	5.48 (↑)	8.39 (↑)	6.79 (↑)	2.55 (↑)

Clustering results

Method	17Flowers		75Flowers		COIL100		MMI	
	Acc	NMI	Acc	NMI	Acc	NMI	Acc	NMI
KM	22.41±1.67	24.31±1.14	19.48±0.85	35.21±0.75	27.96±1.78	58.13±1.52	26.89±2.95	44.15±1.60
Ncuts (TPAMI'00)	27.71±0.72	26.43±0.40	24.80±0.58	41.50±0.19	40.97±1.28	58.52±0.59	38.43±0.47	53.17±0.43
KM-All	17.63±1.27	13.55±1.86	21.13±0.88	32.57±0.71	29.25±1.57	50.55±2.15	27.11±1.81	38.76±1.59
Ncuts-All (TPAMI'00)	28.77±0.63	26.31±0.27	27.41±0.31	42.41±0.21	48.63±0.97	64.74±0.56	40.53±1.52	52.77±0.62
MVIB (DASFAA'07)	21.32±1.05	18.28±1.48	18.49±0.61	33.05±0.45	46.71±2.30	70.29±1.10	44.95±2.60 ○	54.65±1.49
Co(reg) (NeurIPS'11)	26.28±0.49	27.12±0.20	28.16±0.36	44.95±0.09	48.35±0.44	70.86±0.15	34.72±0.53	51.31±0.22
MfIB (IJCAI'13)	38.52±2.03	37.24±1.40 ○	24.57±0.32	40.79±0.37	50.52±0.08	72.81±0.46	40.14±2.09	52.50±1.69
RMSC (AAAI'14)	19.70±0.66	17.86±0.38	26.42±0.97	42.95±0.30	46.32±0.28	69.33±0.45	30.28±1.05	43.94±0.89
LMSC (CVPR'17)	33.29±2.29	31.49±1.60	24.58±0.90	42.50±0.59	48.76±1.45	66.74±0.85	40.17±1.88	51.62±1.29
MLAN (TIP'18)	24.32±1.91	22.21±1.24	25.58±0.53	34.16±1.15	45.05±0.41	59.55±0.53	38.15±0.05	52.68±0.04
GMC (TKDE'20)	6.76±0.00	4.78±0.00	18.52±0.00	30.96±0.00	38.86±0.00	67.55±0.00	35.60±0.00	55.65±0.00 ○
DMIB (TCYB'22)	35.48±6.04	32.56±5.47	26.72±1.13	43.13±0.79	50.33±1.88	72.57±0.87	41.10±2.65	52.96±2.10
FPMVS-CAG (TIP'22)	30.51±0.00	27.27±0.00	23.83±0.00	38.24±0.00	45.03±0.00	70.58±0.00	36.77±0.00	51.03±0.00
MCMLE (TPAMI'22)	32.13±0.00	32.11±0.00	28.76±0.00	47.03±0.00	50.47±0.00	74.59±0.00	42.04±0.00	52.97±0.00
TBGL (TPAMI'23)	31.07±0.00	32.46±0.00	26.52±0.00	47.09±0.00 ○	51.66±0.00	67.82±0.00	43.15±0.00	53.27±0.00
TIM (TIP'23)	32.98±3.28	29.36±3.60	21.83±0.60	26.23±1.24	51.43±1.72	74.98±0.70	28.98±1.57	39.56±2.96
SMVAGC-SF (TIP'24)	42.41±2.07○	36.40±1.43	31.92±0.63○	46.89±0.22	56.78±1.93○	76.78±0.55○	40.93±2.14	53.03±1.25
PTIB	**45.29±0.05** •	**42.49±0.08** •	**35.73±0.36** •	**51.91±0.20** •	**61.17±0.23** •	**82.86±0.19** •	**48.30±0.80** •	**60.39±0.49** •
Improve (• VS ○)	2.88 (↑)	5.25 (↑)	3.81 (↑)	4.82 (↑)	4.39 (↑)	6.08 (↑)	3.35 (↑)	4.74 (↑)

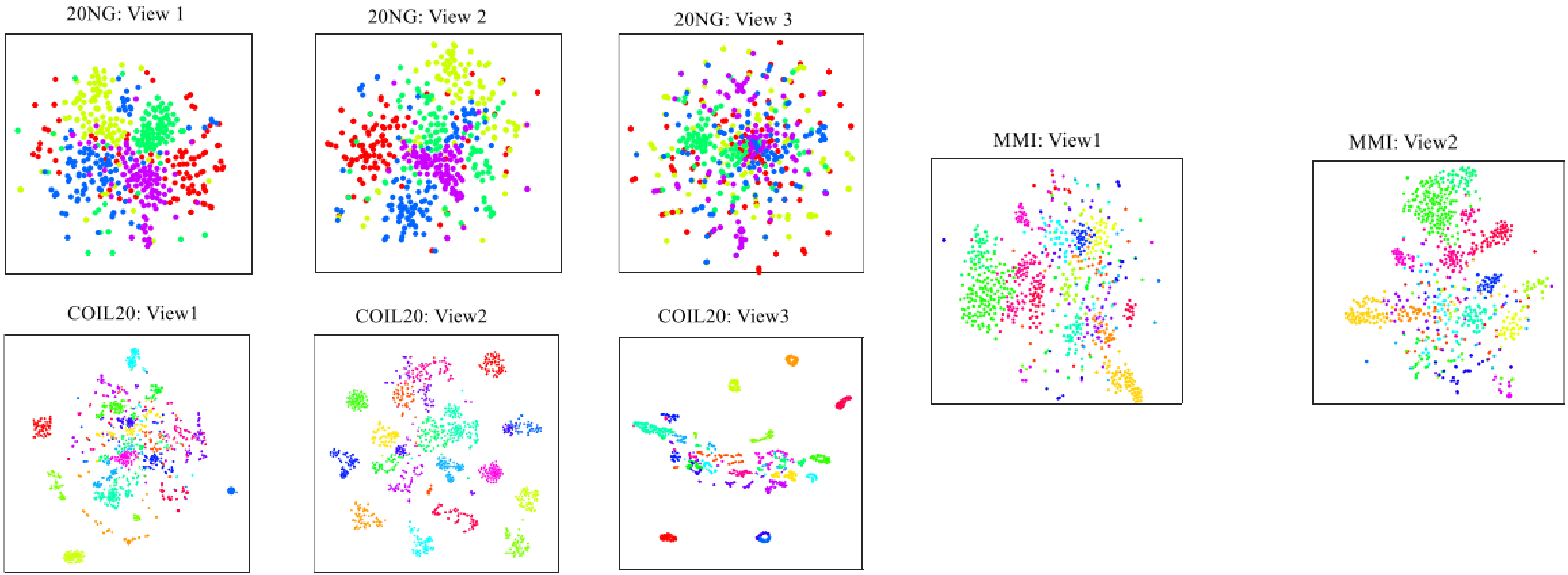
Parameter analysis of PTIB on eight datasets



Potential for Parameter-free Version

Datasets	PTIB		Parameter-free PTIB		Versus Margin	
	Acc	NMI	Acc	NMI	Acc	NMI
20NG	99.80±0.00	99.30±0.00	99.80±0.00	99.30±0.01	0.00	0.00
COIL20	93.33±0.00	96.46±0.00	86.46±0.00	93.80±0.00	-6.87	-2.66
Event	60.24±0.16	45.36±0.28	59.01±0.62	44.39±0.50	-1.23	-0.97
Soccer	62.86±0.17	53.23±0.16	59.64±0.00	51.65±0.01	-3.22	-1.58
17Flowers	45.29±0.05	42.49±0.08	42.74±1.38	40.92±0.82	-2.55	-1.57
75Flowers	35.73±0.36	51.91±0.20	34.57±0.36	51.23±0.24	-1.16	-0.68
COIL100	61.17±0.23	82.86±0.19	59.93±0.61	82.24±0.30	-1.24	-0.62
MMI	48.30±0.80	60.39±0.49	44.26±0.01	58.38±0.00	-4.04	-2.01

T-SNE visualization of Clustering results on 20NG, COIL20 and MMI datasets



Outline

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

Summary

- Propose a novel peer-review trustworthy information bottleneck (PTIB) method for addressing the weighted multi-modal clustering problem.;
- Give a new peer-review look on the multi-modal clustering problem, thus designing a peer-review score for evaluating the quality of each modality. A corresponding trustworthy score is newly designed to evaluate the trustworthiness of peer-review score, ensuring the reliability of multi-modal peer-review.
- Our approach achieves state-of-the-art performance.

Thank You!

Contact for communication:

ieshizhehu@gmail.com