PRIME: A Decoupled Multi-agent Actor-Critic for Multi-view Clustering

Jing Gao[1] , **Xinxin Liu**[1] , **Peng Li**[2,3,*] , **Jianing Zhang**[1] , **Meng Liu**[1] and **Qingchen Zhang**[4,*]
[1]School of Software Technology, Dalian University of Technology
[2]School of Computer Science and Technology, Dalian University of Technology
[3]Key Laboratory of Social Computing and Cognitive Intelligence (Dalian University of Technology)
[4]School of Computer Science and Technology, Hainan University
{gaojing, pli}@dlut.edu.cn, {xinxin_liu, zhang1234567893, liumeng}@mail.dlut.edu.cn,
zhangqingchen@hainanu.edu.cn

Abstract

Deep multi-view clustering draws plentiful attention in various domains, owing to remarkable performance in learning patterns from complementary information of multi-view data. However, previous methods encounter two challenges. They utilize a single pre-defined clustering strategy to perceive diverse structures from data in multiple views subject to heterogeneous distributions, failing to fully capture intricate complementary structures. They leverage either the feature fusion or the result fusion in clustering, which cannot fully integrate complementary information. Therefore, a decoupled multi-agent actor-critic (PRIME) is proposed via defining multi-view clustering as a partially observable Markov game, which establishes a dynamic trial-and-error assignment process of samples in a decentralized perception with centralized learning framework to adaptively learn the optimal clustering strategy. In PRIME, an actor leverages the policy gradient paradigm to independently implement a Markov decision process of data partition in a view, which fully explores data structures to tailor a clustering sub-policy for a view. Meanwhile, a critic utilizes the value function paradigm to centrally guide the Markov game among actors in different views, which constructs both feature and result fusion to progressively enhance complementary knowledge integration for robust clustering results. Extensive experiments on 6 benchmark datasets verify the superiority of PRIME against 10 methods.

1 Introduction

With advances in multimedia technologies, large amounts of multi-view data are collected from various applications [Ren *et al.*, 2025b] [Wang *et al.*, 2024]. In multi-view data, each view is comprised of information subject to heterogeneous distributions, and multiple views depict different aspects of objects of interest in a complementary manner. The in-depth analysis of multi-view data can reveal valuable knowledge and insights, which can fuel intelligent and transparent services on objective perception of the physical world. Multi-view clustering excavates complementary structure information from data in each view to capture partition patterns without need of label supervision, such that samples in the same group exhibit a higher degree of similarity in comparison with those from different groups [Gu *et al.*, 2024] [Xi *et al.*, 2025].

With the ongoing prevalence of deep representation learning, deep multi-view clustering that combines complementary information fusion and intrinsic structure discovery within deep neural networks has achieved considerable gains in data pattern mining [Wang *et al.*, 2025b] [Guan *et al.*, 2025]. Those cutting-edge methods are drawing a substantial amount of attention in various domains, and those methods can be roughly grouped into three branches, i.e., prototype-based clustering, graph-based clustering, and generative clustering. Albeit the previous methods have achieve substantial improvements in clustering performance, there still exist two critical challenges. In multi-view data, instances in different views of the same sample usually contain distinctive structures with different distributions. The previous methods utilize a single pre-defined clustering strategy to equivalently perceive diverse structures in different views, failing to exploit complementary information for pattern mining. Moreover, most of them often leverage either feature fusion or result fusion to integrate complementary information between views in mining patterns. The former captures correlations of representations between different views for information fusion, which is vulnerable to imbalance of view semantics. The latter utilizes view-consistent semantics to constrain the integration of clustering results in different views, which neglects interactive evolution of complementary information.

To address the challenges above, a decou**p**led multi-agent actor-criti**c** for **m**ulti-view clust**e**ring (PRIME) is proposed to adaptively learn the optimal partition policy of data. Specifically, a partially observable Markov game of data partition (POMG-DP) is designed within the reinforcement learning framework, which recasts multi-view clustering as a dynamic trial-and-error assignment process of samples. In POMG-DP, an agent leverages a Markov decision process of data partition to iteratively trial exploratory assignments between instances and clusters in a view, which tailors a clustering sub-policy to the view to fully explore complementary intrinsic structures. Meanwhile, multiple agents engage in a

*Corresponding authors: Peng Li and Qingchen Zhang

collaborative Markov game to dynamically negotiate view-consensus data partition from clustering sub-policies, which facilitates inter-view close cooperation on fusion of complementary information to support robust clustering results. Furthermore, a decoupled multi-agent actor-critic is devised to implement POMG-DP in a decentralized perception with centralized learning framework. In the decentralized perception, an actor utilizes the policy gradient paradigm to implement a Markov decision process of data partition in a view, which performs data perception, action selection, and reward feedback to directly learn the optimal clustering sub-policy. In the centralized learning, the critic leverages the action-value paradigm to guide the inter-view collaboration between actors in the game, which iteratively executes contribution quantification and action negotiation to enable both feature and result fusion of complementary intrinsic structures for robust clustering results. Thus, the contributions are summarized as:

- Multi-view clustering is defined as a partially observable Markov game of data partition to establish a dynamic trial-and-error assignment process of samples, which fully explores structure information of multi-view data.
- A decoupled multi-agent actor-critic is devised in a decentralized perception with centralized learning framework, which utilizes dynamic cooperation to fully fuse complementary information for partition policy learning.
- Extensive experiments are conducted on 6 benchmark datasets, and the results verify the superiority of PRIME in comparison with 10 state-of-the-art methods.

2 Related Work

2.1 Deep Multi-view Clustering

The previous methods can be roughly divided into three categories, i.e., prototype-based clustering, graph-based clustering, and generative clustering.

The prototype-based clustering captures accurate prototypes on fusion features derived from multi-view clustering networks, to minimize intra-cluster distance between samples to recognize patterns [Cui *et al.*, 2024][Ren *et al.*, 2025a]. For instance, [Wu *et al.*, 2024] propose a self-weighted contrastive strategy by measuring semantic discrepancies to adaptively prioritize view structure information in promoting view-consensus semantics of fusion features, which minimizes the squared distance between samples and the nearest prototype to unveil consistent cluster partition. [Zhu and Yin, 2025] propose a neighbor-guided contrastive clustering framework by pulling samples within intra-view and inter-view neighbors close in seeking cluster-friendly invariant information, which minimizes the squared intra-cluster distance between samples to extract clustering patterns from invariant information. The graph-based clustering constructs a data-driven similarity graph about the local adjacency between samples, which further leverages the graph cut to capture topological information for clustering [Liang *et al.*, 2022][Zhao and Li, 2023]. For example, [Wang *et al.*, 2025a] introduce a perturbation-driven learnable anchor mechanism via leveraging the variational inference to promote the topology distribution of view-consensus anchors, which maximizes similarities between samples with the nearest anchor to capture a clear relationship of samples in results. [Du *et al.*, 2025] devise an anchor-based deep unfolding architecture via decoupling the alternating optimization process of anchor clustering into multiple components, which utilizes the self-expressive constraint to extract view-consensus local topological information for clustering. The generative clustering constructs a data generation process from fusion features to samples within multi-view networks, to extract structure information of samples from the joint distribution between class variables and data variables for clustering pattern recognition [Yin *et al.*, 2020][Zhang *et al.*, 2025]. For example, [Xu *et al.*, 2024] propose a multi-expert generative clustering network by quantifying the precision of expert in each view to weight complementary information in feature fusion, which further utilizes Gaussian Mixture Model to shape the joint distribution between class variables and data variables for pattern recognition. [Gao and Pu, 2025] propose a multi-view permutation clustering framework by exploring invariant structures to facilitate semantics of fusion features in the sample generation process, which minimizes the squared distances between samples and their corresponding cluster centers to learn consensus partition.

The above methods utilize a single pre-defined clustering strategy to mine patterns of multi-view data, which is not flexible to handle diverse structures in heterogeneous views. Meanwhile, most of them utilize complementary information derived from either feature fusion or result fusion to mine patterns, which may be ineffective in extracting complementary information for clustering.

2.2 Reinforcement Learning Clustering

Recently, some pioneering research leverages reinforcement learning to capture an adaptive policy of data partition. For example, [Li *et al.*, 2023] propose a clustering agent that implements a Markov decision process of data partition in the encoding-decoding architecture, which maximizes structure rewards to learn a reinforcement clustering policy for single-view data. Then, [Du *et al.*, 2024] treat each sample as an agent that leverages decision-making processes to select a suitable centroid, which maximizes intra-cluster consistency to capture a clustering policy for single-view data. Simultaneously, [Gao *et al.*, 2024] introduce a multi-view clustering agent via implementing a Markov decision process of data partition in a semantics-mirror multi-view architecture, which learns a clustering policy on the fused features of views. Albeit the pioneering methods achieve promising performance, they still learn a single adaptive clustering policy to explore data structures in essence. Such a single policy is incompetent to perceive diverse structures in multiple views subject to heterogeneous distributions, leading to sub-optimal clustering results.

3 Problem Definition

Given a multi-view dataset $\mathrm{X} = \left\{\mathrm{X}^1, \mathrm{X}^2, ..., \mathrm{X}^V\right\}$ of N samples with V views, where $\mathrm{X}^v = \{x_i^v\}_{i=1}^N$ denotes the v-th view dataset and x_i^v denotes the v-th view instance of the i-th

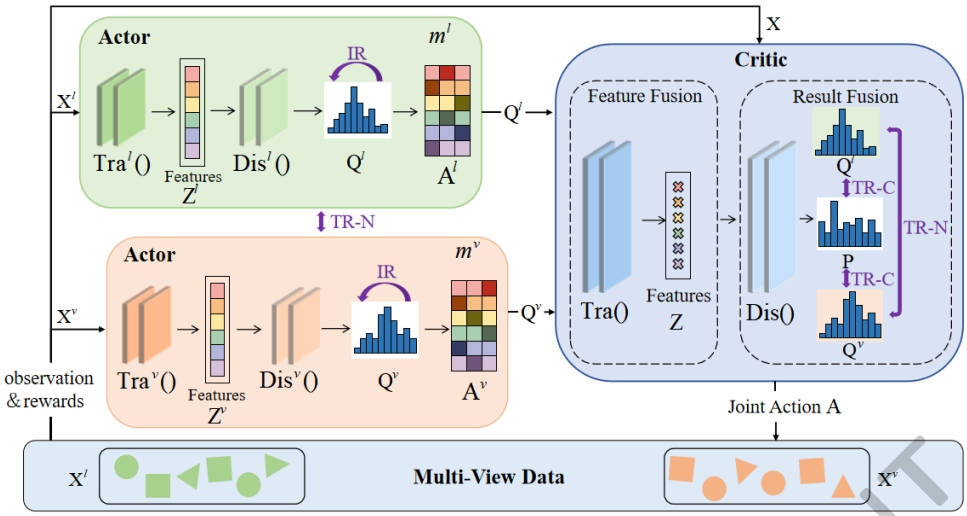


Figure 1: The framework of PRIME. Given a multi-view dataset $\{\mathbf{x}_i^v\}_{i=1,v=1}^{N,V}$, PRIME assigns an actor agent in each view to independently explore intrinsic structures of the view. In v-th view, actor v learns the compact representations Z^v from the observation instances X^v by the transform function $\mathrm{Tra}^v()$, i.e., $\mathrm{Z}^v = \mathrm{Tra}^v(\mathrm{X}^v)$. Subsequently, actor v utilizes the decision function $\mathrm{Dis}^v()$ to calculate the action probability distribution Q^v from Z^v, i.e., $\mathrm{Q}^v = \mathrm{Dis}^v(\mathrm{Z}^v)$, and outputs the action with the highest probability as the clustering assignment of Z^v from the policy network. Then, actor v receives an individual reward (IR) to reflect the fitness of partition. Meanwhile, a critic learns the representations Z from the observation samples X by the transform function $\mathrm{Tra}()$, i.e., $\mathrm{Z} = \mathrm{Tra}(\mathrm{X})$, and calculates the action probability distribution P from Z. Then, it outputs team rewards (TR) based on feature and result fusion, including contribution team rewards (TR-C) and negotiation team rewards (TR-N), to guide the collaborative game among actors. Finally, multiple actors collaborate with a critic in maximizing the cumulative game reward to learn clustering policy from the joint actions A of agents.

sample x_i, multi-view clustering is defined as a partially observable Markov game of data partition (POMG-DP) within the reinforcement learning paradigm to establish a dynamic trial-and-error assignment process of samples, which leverages spontaneous collaboration of agents to seamlessly fuse intrinsic structures in different views in learning the optimal clustering policy. In POMG-DP, an agent leverages a Markov decision process of data partition to iteratively trial exploratory assignments between instances and clusters in a view, which tailors the optimal clustering sub-policy to the view to fully explore complementary intrinsic structures. Meanwhile, multiple agents engage in a collaborative Markov game to dynamically negotiate view-consensus data partition, which enables close inter-view cooperation on fusion of complementary intrinsic structures for robust clustering results.

In detail, a five tuple $\mathbb{F} =< \mathrm{M}, \mathrm{S}, \mathrm{A}, \mathrm{T}, \mathrm{R} >$ is defined to describe POMG-DP, as follows:

- $\mathrm{M} = \{m^1, m^2, ..., m^V\}$ denotes the set of V agents where the v-th agent m^v learns the optimal clustering sub-policy $\pi^v(\cdot)$ to perceive intrinsic structures of data in the v-th view. Then, the V agents establish a dynamic game to jointly distill the optimal policy $\pi(\cdot)$ of data partition from their respective sub-policies.
- S is a finite set of states where $s \in \mathrm{S}$ denotes a state over multi-view data and cluster prototypes. At step i, the agent m^v utilizes an observation function $O : \mathrm{S} \times \mathrm{M} \to \mathrm{X}$ to receive the partial observation x_i^v of the i-th sample in the v-th view from the state s_i, i.e., $x_i^v = O(s_i, m^v)$.
- $\mathrm{A} = \mathrm{A}^1 \times \mathrm{A}^2 \times \cdots \times \mathrm{A}^V$ is the joint action space, where the action set A^v of the agent m^v contains K actions corresponding to K clusters. For the partial observation x_i^v, the agent m^v selects an action $a_i^v \in \mathrm{A}^v$ according to the clustering sub-policy, i.e., $a_i^v \sim \pi^v(\cdot|x_i^v)$, to assign x_i^v into a cluster. Each agent shares its action derived from the partial observation in the corresponding view to form the joint action $a_i = (a_i^1, a_i^2 ..., a_i^V)$, which further negotiates a view-consensus assignment p_i for the i-th sample to seamlessly fuse complementary information.
- $\mathrm{T} : \mathrm{S} \times \mathrm{A} \to \mathrm{S}$ denotes a deterministic state transition function that depicts the evolution of clustering structures distilled from multi-view data. The state transition function decides how to update the current state to the subsequent state given the joint action of agents.
- $\mathrm{R} = f(r_s, r_c)$ is a game reward function that translates evolution of clustering structures induced by a joint action of agents into an immediate reward. The individual reward r_s measures fitness of assignment actions between instances and clusters in a view. The team reward r_c quantifies the contribution of agents to the collaboration on learning the optimal clustering policy. After performing the joint action a_i, each agent receives an individual reward r_s and a team reward r_c, which guide agents to collaboratively explore view-consensus data structures with clear clusters from complementary information.

Specifically, at the step i of multi-view data partition, the agent m^v perceives a partial observation x_i^v of sample x_i, selects an assignment action a_i^v for x_i^v, and receives an individual reward r_s that measures intra-view partition fitness. In such a manner, the agent m^v performs a Markov decision pro-

cess of data partition in the v-th view to capture the optimal clustering sub-policy $\pi^v(\cdot|x_i^v)$. Meanwhile, the V agents negotiate the view-consensus assignment p_i of the i-th sample from the clustering sub-policies, and receive a team reward r_c that measures the contribution to collaboration. They achieve a close inter-view cooperation within the game paradigm to aggregate complementary information. In the game, V agents maximize cumulative rewards over a finite number of steps to adaptively learn the optimal clustering policy:

$$\begin{aligned} J &= \mathbb{E}_{a\sim\pi(\cdot|s)} \sum_{i=1}^{N} \gamma^{i-1} \mathrm{R}_i(r_c, r_s) \\ &\approx \sum_{v=1}^{V} \mathbb{E}_{a^v\sim\pi^v(\cdot|x^v)} (\sum_{i=1}^{N} \gamma^{i-1} \mathrm{R}_i^v(r_c, r_s)) \end{aligned} \tag{1}$$

where a and a^v denote the joint action and the action in the v-th view at a certain moment, respectively, and γ is a discount factor. $\mathrm{R}_i^v(r_c, r_s)$ is the game reward of the agent m^v at step i. $\mathbb{E}$ denotes the expectation operator.

In POMG-DP, an agent iteratively performs data perception, action selection, and reward feedback in the corresponding view to tailor the optimal clustering sub-policy to a view, which ensures full exploration of data structures in each view subject to heterogeneous distributions. Meanwhile, multiple agents leverage action negotiation and contribution quantification to conduct a dynamic Markov game over multi-view data partition, which establishes co-evolution of clustering sub-policies to capture view-consistent data structures with intra-cluster compactness and inter-cluster separation.

4 The Decoupled Actor-Critic

In PRIME, a decoupled actor-critic scheme is devised to implement POMG-DP in a decentralized perception with centralized learning framework, as shown in Figure 1, which establishes dynamic cooperation on seamless fusion of intrinsic structure information in each view to adaptively learn the optimal clustering policy.

In the decentralized perception, an actor independently utilizes the policy gradient paradigm to implement a Markov decision process of data partition in a view. It maximizes the cumulative individual rewards that measure intra-cluster compactness and inter-cluster separation to tailor a clustering sub-policy in terms of the view distribution, facilitating full exploration of intrinsic structures in each view. To this end, the actor stacks the heavy-tailed student's t distribution on a single-view network to construct a policy network. Then, it iteratively performs data perception, action selection, and reward feedback to directly learn the optimal sub-policy that assigns each instance into a suitable cluster in a view.

Specifically, given a sample $x_i = \{x_i^v\}_{v=1}^V$ at step i, the actor in the v-th view receives the high-dimensional observation instance x_i^v of the sample x_i, and then it utilizes the nonlinear transform function $\mathrm{Tra}^v(\cdot)$ of the single network to learn the compact representation z_i^v, i.e., $z_i^v = \mathrm{Tra}^v(x_i^v)$, alleviating interference of the curse of dimensionality in data perception. Afterwards, it utilizes the heavy-tailed student's t distribution that measures similarities between compact representations and implicit clusters to define a decision function $\mathrm{Dis}^v(\cdot)$, which generates a probability distribution q_i^v over K assignment actions $\{k\}_{k=1}^K$ for z_i^v to adaptively select a cluster assignment for the instance:

$$q_{ik}^v = \mathrm{Dis}^v\left(k|z_i^v\right) = \frac{\left(1 + \|z_i^v - \varepsilon_k^v\|^2\right)^{-1}}{\sum_j^K \left(1 + \left\|z_i^v - \varepsilon_j^v\right\|^2\right)^{-1}} \tag{2}$$

where ε_k^v denotes the learnable prototype of the k-th implicit cluster in the v-th view and q_{ik}^v denotes the probability to select the k-th action for z_i^v. Then, it outputs the action with the highest probability as the clustering assignment of z_i^v, i.e., $a_i^v = \arg\max_k \mathrm{Dis}^v(k|z_i^v)$ from the policy network.

After performing the selected action a_i^v, actor v receives a reward that utilizes the silhouette of data structures at step i to measure evolution of clustering. In implementation, the silhouette difference of structures between step i and step $i-1$ is defined as the immediate individual reward r_s^v for the action a_i^v in the maximization of cumulative rewards. That is, the immediate individual reward r_s^v is computed via:

$$r_{s,i}^v = \Phi(z_i^v, \mathbb{C}_i^v) - \Phi(z_{i-1}^v, \mathbb{C}_{i-1}^v). \tag{3}$$

In Eq. (3), $\Phi(z_i^v, \mathbb{C}_i^v)$ denotes the silhouette coefficient at step i , where $\mathbb{C}_i^v = \{C_k^v\}_{k=1}^K$ and C_k^v is the set of instances in the k-th cluster. Such a reward roots in the essence of clustering, i.e., intra-cluster compactness and inter-cluster separation, which encourages actors to capture clear clustering structures. The reason is three-fold. ① A positive reward reveals the improvement of clustering structures, which encourages actors to select similar actions. ② A negative reward manifests the degradation of clustering structures, making actors to avoid such harmful actions. ③ The zero reward indicates no change of clustering structures, which prompts agents to explore more promising actions.

Finally, the v-th actor maximizes the cumulative sum r_s^v of individual rewards (IR) to learn the optimal sub-policy, which ensures full exploration of structures in the v-th view via:

$$r_s^v = \sum_{i=1}^{N} r_{s,i}^v + \mathcal{C}_{\mathrm{rec}}^v \tag{4}$$

where $\mathcal{C}_{\mathrm{rec}}^v$ is a reconstruction constraint in v-th view of samples and is conducted via a symmetrical structure of the policy network to prevent the degradation of instance structures.

In the centralized learning, a critic utilizes the action-value paradigm to centrally guide the inter-view Markov game between actors. It maximizes cumulative team rewards that quantify the contribution of actors to view-consensus partition from both the feature and result perspectives, which enables close fusion of complementary intrinsic structures for robust clustering results. To this end, the critic combines a heavy-tailed student's t distribution with a multi-view network to construct an actor-value network, which iteratively performs contribution quantification and action negotiation to evaluate assignment actions of actors.

Specifically, given a sample $x_i = \{x_i^v\}_{v=1}^V$ at step i, the critic utilizes the nonlinear transform function $\mathrm{Tra}(\cdot)$ of the multi-view network to extract the compact representation z_i,

Algorithm 1 PRIME

Input: The multi-view dataset $\mathbf{X} = \{\mathbf{x}_i^v\}_{i=1,v=1}^{N,V}$, number of clusters K, hyperparameters α, β, η.
Output: Clustering partition $\overline{\mathbf{P}}$.

for $i = 1$ **to** N **do**
  Actor:
  Actor v extracts representation z_i^v of observation x_i^v.
  Compute probability distribution q_{ik}^v by Eq. (2).
  Select action $a_i^v = \arg\max_k q_{ik}^v$.
  Receive an individual reward $r_{s,i}^v$ by Eq. (3).
  Critic:
  Critic perceives the global structures of the sample z_i.
  Compute global target assignment p_{ik} by Eq. (5).
  Output contribution team rewards $\tilde{r}_{\mathrm{c},i}^v$ by Eq. (6).
  Output negotiation team rewards $\hat{r}_{\mathrm{c}}^v$ by Eq. (7)
  Output team rewards r_c^v by Eq. (8).
  Update parameters in actors and critic by maximizing cumulative reward J by Eq. (9).
end for
return Clustering partition $\overline{\mathbf{P}} = \frac{1}{V}\sum_{v=1}^{V} \mathbf{A}^v$.

i.e., $z_i = \mathrm{Tra}(x_i)$, which builds the feature fusion of complementary information to capture global structures in the sample x_i. Then, it leverages the heavy-tailed student's t distribution to produce a global probability distribution p_i over K assignment actions via:

$$p_{ik} = \frac{\left(1 + \|z_i - \varepsilon_k\|^2\right)^{-1}}{\sum_{\bar{k}=1}^{K}\left(1 + \left\|z_i - \varepsilon_{\bar{k}}\right\|^2\right)^{-1}} \tag{5}$$

where ε_k denotes the prototype of the k-th cluster and p_{ik} denotes the probability to select the k-th action for z_i.

Afterwards, the critic measures semantic divergence of assignment actions for the same sample between itself and actors to define a contribution team reward, which leverages global structures to quantify the collaboration contribution of actors in feature fusion. In detail, given the action probability distribution q_i^v of the v-th actor at step i and its own action probability distribution p_i, it leverages the Kullback-Leibler divergence to map semantic discrepancies into a contribution team reward (TR-C) $\tilde{r}_{\mathrm{c},i}^v$ for the action a_i^v of the v-th actor via:

$$\tilde{r}_{\mathrm{c},i}^v = -\mathrm{D}_{\mathrm{KL}}(\mathrm{p}_i||\mathrm{q}_i^v) = -\sum_{k=1}^{K} p_{ik}\log\frac{p_{ik}}{q_{ik}^v} \tag{6}$$

where $\mathrm{D}_{\mathrm{KL}}(\cdot)$ denotes the Kullback-Leibler divergence.

Then, the critic measures inter-view result discrepancies from the sample and cluster perspectives to define a negotiation team reward, which builds the result fusion of complementary information to encourage actors to collaborate on exploration of view-consistent clustering structures. Given observation representation matrices $\{\mathrm{Z}^1, \mathrm{Z}^2, ..., \mathrm{Z}^V\}$ and assignment matrices $\{\mathrm{Q}^1, \mathrm{Q}^2, ..., \mathrm{Q}^V\}$ in V views, the critic leverages the matrix trace to map the result discrepancies into a negotiation team reward (TR-N), as follows:

$$\begin{aligned}\hat{r}_{\mathrm{c}}^v = &\sum_{j=1}^{V}\mathrm{Tr}((\mathbf{Q}^v)^\top\mathbf{Q}^j) - ||\mathrm{OF}_{\mathrm{diag}}((\mathbf{Q}^v)^\top\mathbf{Q}^j)||_{\mathrm{F}}^2 \\ &+ \sum_{j=1}^{V}\mathrm{Tr}(\mathbf{Z}^v(\mathbf{Z}^j)^\top) - ||\mathrm{OF}_{\mathrm{diag}}(\mathbf{Z}^v(\mathbf{Z}^j)^\top)||_{\mathrm{F}}^2\end{aligned} \tag{7}$$

where $\mathrm{Tr}(\cdot)$ is the matrix trace, $\mathrm{OF}_{\mathrm{diag}}(\cdot)$ represents the non-diagonal part of a matrix, and $||\cdot||_{\mathrm{F}}$ is the Frobenius norm of a matrix. The critic gives the overall team reward (TR) to the action a_i^v of the v-th actor at step i via:

$$r_c^v = \alpha\hat{r}_{\mathrm{c}}^v + \sum_{i=1}^{N}\eta\tilde{r}_{c,i}^v \tag{8}$$

where α and η are trade-off parameters.

In PRIME, multiple actors collaborate with a critic in maximizing the following cumulative game reward to learn the optimal clustering policy:

$$J = \sum_{v=1}^{V}(\beta r_s^v + \hat{r}_c^v) \tag{9}$$

where β is a trade-off parameter. Algorithm 1 shows the detailed steps.

Dataset	Class	Sample	View	Type
BDGP	5	2500	2	Vector
MNIST-USPS	10	5000	2	Image
Digit-Product	10	30000	2	Vector
ALOI	100	10800	4	Vector
Caltech-4V	7	1400	4	Vector
Caltech-5V	7	1400	5	Vector

Table 1: The detail statistics of the 6 multi-view datasets.

5 Experiment

In the section, experiments are conducted on 6 benchmark datasets with 4 standard metrics against 10 baseline methods, and the results validate the superiority of PRIME.

5.1 Experiment Settings

Datasets: 6 benchmark multi-view datasets are used to validate the performance of PRIME, and statistical information is presented in Table 1.

Comparison Methods: The performance of PRIME is evaluated in comparison with 10 baseline methods from 4 categories. The prototype-based methods are SDMVC [Xu *et al.*, 2023], COST [Zhang *et al.*, 2024], DDMvC [Chen *et al.*, 2025], DIVIDE [Lu *et al.*, 2024], and SCMVC [Wu *et al.*, 2024]. The graph-based methods are DMAC [Wang *et al.*, 2025a] and hubREP [Xu *et al.*, 2025]. The generative methods are Multi-VAE [Xu *et al.*, 2021] and DVIMC [Xu *et al.*, 2024]. The reinforcement method is DMVC-SI [Gao *et al.*, 2024]. A grid search is conducted on the trade-off parameters for all baselines as recommended by the respective authors to ensure optimal performance.

Method	BDGP				MNIST-USPS				Digit-Product			
	ACC	NMI	ARI	PUR	ACC	NMI	ARI	PUR	ACC	NMI	ARI	PUR
Multi-VAE(2021)	0.8662	0.8439	0.6420	0.8662	0.9420	0.9693	0.9478	0.9420	0.8530	0.8320	0.8100	0.8530
SDMVC(2023)	0.9816	0.9447	0.9539	0.9548	0.9880	0.9815	0.9754	0.9880	0.9769	0.9506	0.9509	0.9767
SCMVC(2024)	0.9888	0.9609	0.9723	0.9888	0.9928	0.9793	0.9841	0.9928	0.9854	0.9620	0.9681	0.9854
DVIMC(2024)	0.9689	0.9341	0.9235	0.9689	0.9812	0.9761	0.9698	0.9812	0.7843	0.7279	0.6952	0.7931
DMVC-SI(2024)	0.9880	0.9634	0.9723	0.9880	0.9888	0.9835	0.9877	0.9888	0.9891	0.9744	0.9825	0.9891
DIVIDE(2024)	0.9604	0.9216	0.9307	0.9604	0.9740	0.9391	0.9431	0.9740	0.9890	0.9706	0.9759	0.9890
COST(2024)	0.9892	0.9648	0.9746	0.9892	0.9936	0.9815	0.9858	0.9936	0.9920	0.9782	0.9824	0.9920
DMAC(2025)	0.9012	0.8027	0.7750	0.9012	0.7416	0.6840	0.6041	0.7656	0.7253	0.7292	0.6450	0.7728
DDMvC(2025)	0.9776	0.9390	0.9455	0.9776	0.9950	0.9854	0.9889	0.9950	0.9909	0.9730	0.9798	0.9909
hubREP(2025)	0.9868	0.9538	0.9674	0.9868	0.9572	0.9090	0.9084	0.9592	0.7718	0.7364	0.6457	0.7745
PRIME	**0.9904**	**0.9716**	**0.9763**	**0.9904**	**0.9961**	**0.9890**	**0.9914**	**0.9961**	**0.9951**	**0.9853**	**0.9892**	**0.9951**
Method	ALOI				Caltech-4V				Caltech-5V			
	ACC	NMI	ARI	PUR	ACC	NMI	ARI	PUR	ACC	NMI	ARI	PUR
Multi-VAE(2021)	0.6515	0.8199	0.5449	0.6885	0.5693	0.4039	0.3902	0.5850	0.6190	0.5650	0.5420	0.6230
SDMVC(2023)	0.6389	0.7882	0.5123	0.6465	0.7692	0.6653	0.6218	0.7692	0.8928	0.7965	0.7862	0.8928
SCMVC(2024)	0.8115	0.9080	0.7586	0.8289	0.8450	0.7317	0.6975	0.8450	0.9100	0.8362	0.8137	0.9100
DVIMC(2024)	0.6812	0.6124	0.6072	0.6957	0.7823	0.6946	0.6889	0.7983	0.8957	0.8095	0.7920	0.8957
DMVC-SI(2024)	0.7579	0.6903	0.7288	0.7579	0.8205	0.6333	0.7964	0.8205	0.8813	0.7935	0.8365	0.8813
DIVIDE(2024)	0.7930	0.6829	0.7267	0.8095	0.6850	0.5543	0.4747	0.6850	0.7114	0.5847	0.5104	0.7114
COST(2024)	0.8368	0.8721	0.8854	0.8368	0.8211	0.7010	0.7752	0.8211	0.8750	0.7523	0.7786	0.8750
DMAC(2025)	0.6022	0.7326	0.5214	0.6486	0.6657	0.5497	0.4481	0.6757	0.6621	0.5465	0.4425	0.6750
DDMvC(2025)	0.6946	0.8259	0.6548	0.6946	0.8421	0.7779	0.7428	0.8421	0.8543	0.7893	0.7426	0.8543
hubREP(2025)	0.8020	0.8807	0.7834	0.8148	0.8814	0.7908	0.7632	0.8907	0.8907	0.8064	0.7814	0.9100
PRIME	**0.8624**	**0.9097**	**0.8983**	**0.8694**	**0.9016**	**0.8233**	**0.8004**	**0.9016**	**0.9132**	**0.8413**	**0.8387**	**0.9132**

Table 2: Experimental results of each method by ACC, NMI, ARI, and PUR. The best results are marked in bold.

Evaluation Metrics: Four standard clustering metrics, which include accuracy (ACC), normalized mutual information (NMI), adjusted rand index (ARI), and purity (PUR), are adopted to quantify clustering results. For the four metrics, higher values signify a stronger agreement between clustering results and true labels.

Implementation Details: PRIME adopts the fully connected network and the convolutional neural network to implement V actors and a critic for the vector datasets and the image dataset, respectively. The ReLU activation function and the Adam optimizer are utilized in the optimization.

5.2 Clustering Performance

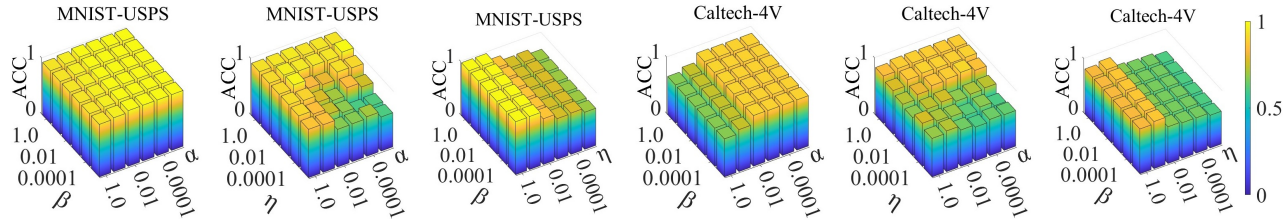
Table 2 illustrates numerical results on 6 standard datasets. In Table 2, PRIME significantly outperforms all the comparison methods on all the four metrics. This conclusion is supported by the following two observations. ① BDGP, MNIST-USPS and Digit-Product: The performance of state-of-the-art methods is close to the upper limit, resulting in a saturated competitive landscape. In such an intense competition, PRIME still achieves the first rank on all the metrics, demonstrating PRIME produces more precise assignment to recognize hard samples located at the partition boundary. ② ALOI, Caltech-4V and Caltech-5V: Each of the three datasets comprises multiple views of high heterogeneity, which poses great challenges in effective fusion of complementary information from various views. In the challenging scenario, PRIME outperforms the second best method by a significant margin. For example, PRIME surpasses the second best method hubREP on the Caltech-4V dataset by 2.02%, 3.25%, 3.72% and 1.09% in terms of ACC, NMI, ARI and PUR. The two observations verify the significant superiority of PRIME in the intrinsic structure exploration and complementary information fusion for clustering policy learning.

	BDGP		Caltech-5V	
	ACC	NMI	ACC	NMI
PRIME w/o IR	0.9837	0.9533	0.9086	0.8320
PRIME w/o TR	0.8944	0.7670	0.8292	0.7029
PRIME	**0.9904**	**0.9716**	**0.9126**	**0.8372**

Table 3: The ablation results on BDGP and Caltech-5V.

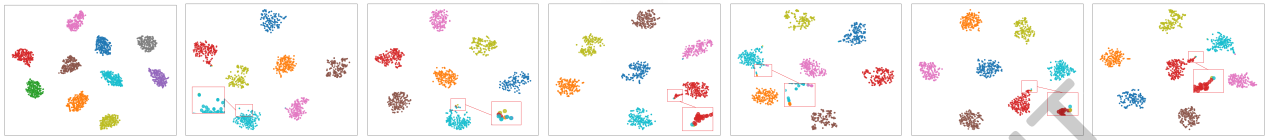
Ablation Study. Ablation studies are further conducted via comparing PRIME with two variants, i.e., PRIME w/o IR which removes the individual reward and PRIME w/o TR which eliminates the team reward. Table 3 shows the performance of ablation studies on two benchmarks in terms of ACC and NMI, demonstrating the importance of each component in PRIME. Three key observations can be drawn from Table 3. ① The performance of the PRIME w/o IR decreases on two datasets, suggesting that actors can tailor the optimal sub-policy for intrinsic structure exploration. ② PRIME w/o TR suffers a substantial performance drop, which underscores the indispensable role of the critic in both feature and result fusion. ③ PRIME achieves superior results on both datasets, thereby validating architecture rationality.

Hyperparameter Analysis. To evaluate the impact of hyperparameters, i.e., α, β and η, on the performance of PRIME,



(a) α-β (b) α-η (c) η-β (d) α-β (e) α-η (f) η-β

Figure 2: Sensitivity analysis of the hyperparameters, i.e., α, β and η, on Caltech-4V and MNIST-USPS.



(a) Digit-Product (b) Caltech-5V (c) view 1 (d) view 2 (e) view 3 (f) view 4 (g) view 5

Figure 3: The t-SNE visualization results. The sub-figures (a) and (b) illustrate the results of global cluster structure on Digit-Product and Caltech-5V, respectively. The sub-figures (c)-(g) demonstrate the results of local cluster structure of each view on Caltech-5V.

parameter sensitivity experiments are conducted on Caltech-4V and MNIST-USPS. In the experiment, the control variates method is employed to tune the three hyperparameters within the range of $\{1.0, 0.1, 0.01, 0.001, 0.0001, 0.00001\}$. As shown in Figure 2, the results are visualized in the form of heatmaps, to intuitively illustrate clustering performance over hyperparameter variations with the help of an interpretable color field. Three conclusions can be drawn. ① When the hyperparameter α is set to $\{0.01, 0.001, 0.0001, 0.00001\}$, PRIME obtains the best performance, as evidenced by the warmest color in the corresponding region of sub-figures (a) and (d). ② When the hyperparameter η takes values in $\{0.1, 0.01\}$, PRIME achieves the optimal performance, as indicated by the warmest colors in sub-figures (b) and (e). ③ The model performance can remain remarkably stable across a wide range of the hyperparameter β, as shown by the minimal color variation along the η-axis in sub-figures (c) and (f). Those conclusions demonstrate that PRIME is of strong robustness to hyperparameters, which substantially reduces the burden of parameter tuning in real-world deployments.

Visualization Analysis. The t-SNE visualization is utilized to qualitatively validate the model effectiveness in revealing cluster structures from data with heterogeneous views on Digit-Product and Caltech-5V. In detail, 1,400 samples randomly selected from each dataset are transformed into latent representations to construct the candidate sets, which are input into the t-SNE algorithm to produce visualizations of cluster structures. The results are shown in Figure 3.

In detail, the visualization experiments are designed as a two-tiered, mutually complementary paradigm to ensure a systematic and rigorous assessment of model performance. Firstly, the output representations of samples in the two candidate datasets are input into the t-SNE algorithm to visualize the cluster structure from the global perspective. As shown in sub-figures (a) and (b), the samples demonstrate remarkable intra-cluster cohesion and inter-cluster separation with no overlapping regions between clusters. The observation above confirms that PRIME is competent to learn an adaptive clustering strategy from multi-view data. Secondly, the view representations of samples in Caltech-5V of which each view with high variety provides a challenging test environment, are input into the t-SNE algorithm to visualize the cluster structure from the local perspective. As shown in sub-figures (c)-(g), instances in each view form well-defined compact aggregation regions without overlapping in the embedding space, which verifies PRIME can capture an adaptive clustering substrategy suitable for the heterogeneous distribution of each view. Furthermore, according to the comparisons between the sub-figure (b) and sub-figures (c)-(g), a few hard samples that are misclustered in local cluster structures of views can be well rectified by PRIME in global cluster structures, which validates the model effectiveness in the intrinsic structure exploration and complementary information fusion.

6 Conclusion

In the article, PRIME is proposed via defining multi-view clustering as a partially observable Markov game of data partition, which establishes a dynamic trial-and-error assignment process of samples to learn the optimal clustering policy. Furthermore, a decoupled multi-agent actor-critic is devised to implement the game of data partition in a decentralized perception with centralized learning framework. PRIME achieves two advantages. ① It tailors the optimal sub-policy to each view to guarantee full exploration of intrinsic structures in heterogeneous views. ② It builds a collaborative Markov game to dynamically distill view-consensus data partition from feature and result fusion, achieving close fusion of complementary intrinsic structures for robust clustering results. Finally, experiments on 6 benchmark datasets verify the superiority of PRIME against 10 state-of-the-art methods.

Acknowledgments

This work was supported in part by the Fundamental Research Funds for the Central Universities under Grants No. DUT25LAB124 and DUTZD25239, in part by the National Natural Science Foundation of China under Grant No. 62462022, in part by Outstanding Young Science and Technology Talent in Dalian (2024RY001), and in part by Central Guidance for Local Science and Technology Development Fund (ZY23CG29).

References

[Chen *et al.*, 2025] Zhe Chen, Xiaojun Wu, Tianyang Xu, Hui Li, and Josef Kittler. Deep discriminative multi-view clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 2025.

[Cui *et al.*, 2024] Jinrong Cui, Yuting Li, Han Huang, and Jie Wen. Dual contrast-driven deep multi-view clustering. *IEEE Transactions on Image Processing*, 2024.

[Du *et al.*, 2024] Peng Du, Fenglian Li, and Jianli Shao. Multi-agent reinforcement learning clustering algorithm based on silhouette coefficient. *Neurocomputing*, 596:127901, 2024.

[Du *et al.*, 2025] Shide Du, Chunming Wu, Zihan Fang, Wendi Zhao, Yilin Wu, Changwei Wang, and Shiping Wang. Largemvc-net: Anchor-based deep unfolding network for large-scale multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 1714–1723, 2025.

[Gao and Pu, 2025] Xin Gao and Jian Pu. Deep incomplete multi-view learning via cyclic permutation of vaes. In *International Conference on Learning Representations*, 2025.

[Gao *et al.*, 2024] Jing Gao, Meng Liu, Peng Li, Jianing Zhang, and Zhikui Chen. Deep multiview adaptive clustering with semantic invariance. *IEEE Transactions on Neural Networks and Learning Systems*, 35(9):12965–12978, 2024.

[Gu *et al.*, 2024] Zhibin Gu, Zhendong Li, and Songhe Feng. Topology-driven multi-view clustering via tensorial refined sigmoid rank minimization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 920–931, 2024.

[Guan *et al.*, 2025] Renxiang Guan, Wenxuan Tu, Siwei Wang, Jiyuan Liu, Dayu Hu, Chang Tang, Yu Feng, Junhong Li, Baili Xiao, and Xinwang Liu. Structure-adaptive multi-view graph clustering for remote sensing data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 16933–16941, 2025.

[Li *et al.*, 2023] Peng Li, Jing Gao, Jianing Zhang, Shan Jin, and Zhikui Chen. Deep reinforcement clustering. *IEEE Transactions on Multimedia*, 25:8183–8193, 2023.

[Liang *et al.*, 2022] Weixuan Liang, Xinwang Liu, Sihang Zhou, Jiyuan Liu, Siwei Wang, and En Zhu. Robust graph-based multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 36, pages 7462–7469, 2022.

[Lu *et al.*, 2024] Yiding Lu, Yijie Lin, Mouxing Yang, Dezhong Peng, Peng Hu, and Xi Peng. Decoupled contrastive multi-view clustering with high-order random walks. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 14193–14201, 2024.

[Ren *et al.*, 2025a] Fujian Ren, Wenlan Chen, Lu Gao, Fei Guo, and Cheng Liang. Dual-level distribution alignment for deep incomplete multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 2476–2485, 2025.

[Ren *et al.*, 2025b] Pengfei Ren, Jingyu Wang, Haifeng Sun, Qi Qi, Jing Wang, and Jianxin Liao. Rule meets learning: Confidence-aware multi-view fusion for self-supervised 3d hand pose estimation. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 1646–1655, 2025.

[Wang *et al.*, 2024] Zhecheng Wang, Rajanie Prabha, Tianyuan Huang, Jiajun Wu, and Ram Rajagopal. Skyscript: A large and semantically diverse vision-language dataset for remote sensing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 5805–5813, 2024.

[Wang *et al.*, 2025a] Bocheng Wang, Chusheng Zeng, Mulin Chen, and Xuelong Li. Towards learnable anchor for deep multi-view clustering. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 21044–21052, 2025.

[Wang *et al.*, 2025b] Zhenxi Wang, Zongyao Yin, Yujie Hou, and Xianchuan Yu. Robust multi-view clustering via pseudo label guided universum learning. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 1480–1489, 2025.

[Wu *et al.*, 2024] Song Wu, Yan Zheng, Yazhou Ren, Jing He, Xiaorong Pu, Shudong Huang, Zhifeng Hao, and Lifang He. Self-weighted contrastive fusion for deep multi-view clustering. *IEEE Transactions on Multimedia*, 26:9150–9162, 2024.

[Xi *et al.*, 2025] Yanwanyu Xi, Xiao Zheng, Chang Tang, Xingchen Hu, Yuanyuan Liu, Jun-Jie Huang, and Xinwang Liu. Lrgr: self-supervised incomplete multi-view clustering via local refinement and global realignment. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 6624–6632, 2025.

[Xu *et al.*, 2021] Jie Xu, Yazhou Ren, Huayi Tang, Xiaorong Pu, Xiaofeng Zhu, Ming Zeng, and Lifang He. Multi-vae: Learning disentangled view-common and view-peculiar visual representations for multi-view clustering. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9234–9243, 2021.

[Xu *et al.*, 2023] Jie Xu, Yazhou Ren, Huayi Tang, Zhimeng Yang, Lili Pan, Yang Yang, Xiaorong Pu, Philip S. Yu, and Lifang He. Self-supervised discriminative feature learning for deep multi-view clustering. *IEEE Transactions on Knowledge and Data Engineering*, 35(7):7470–7482, 2023.

[Xu *et al.*, 2024] Gehui Xu, Jie Wen, Chengliang Liu, Bing Hu, Yicheng Liu, Lunke Fei, and Wei Wang. Deep variational incomplete multi-view clustering: Exploring shared clustering structures. In *Proceedings of the AAAI conference on artificial intelligence*, volume 38, pages 16147–16155, 2024.

[Xu *et al.*, 2025] Zheming Xu, He Liu, Congyan Lang, Tao Wang, Yidong Li, and Michael C. Kampffmeyer. A hubness perspective on representation learning for graph-based multi-view clustering. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 15528–15537. IEEE, 2025.

[Yin *et al.*, 2020] Ming Yin, Weitian Huang, and Junbin Gao. Shared generative latent representation learning for multi-view clustering. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 6688–6695, 2020.

[Zhang *et al.*, 2024] Qian Zhang, Lin Zhang, Ran Song, Runmin Cong, Yonghuai Liu, and Wei Zhang. Learning common semantics via optimal transport for contrastive multi-view clustering. *IEEE Transactions on Image Processing*, 33:4501–4515, 2024.

[Zhang *et al.*, 2025] Jingwei Zhang, Mohammad Jalali, Cheuk Ting Li, and Farzan Farnia. Unveiling differences in generative models: A scalable differential clustering approach. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 8269–8278, 2025.

[Zhao and Li, 2023] Yang Zhao and Xuelong Li. Deep spectral clustering with regularized linear embedding for hyperspectral image clustering. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–11, 2023.

[Zhu and Yin, 2025] Kai Zhu and Jun Yin. Neighbor contrastive learning with weakened consensus graph for deep multi-view clustering. In *Proceedings of the 33rd ACM International Conference on Multimedia*, pages 1695–1703, 2025.