Fairly Incomplete Multi-View Clustering Inspired by the Information Bottleneck

Anonymous Author(s)
Affiliation
Address
email

Abstract

Nowadays, multi-source devices, which can obtain data from different views, generate massive and complex information. Missing data situations, which are common with multimodal data, pose a challenge for how to fuse incomplete data for model training. However, many data sets contain sensitive information (e.g., gender, marital status, age), which can lead machine learning models to learn biased features and make unfair decisions. The majority of existing models address these two issues independently, with no unified solution. To address the concerns, this paper employs a fair constraint strategy to encourage the model to train bias features independent of the data. Simultaneously, inspired by the information bottleneck theory, we extract common features of complete view data and align the distribution of incomplete and complete view data, significantly improving data utilization. The experiment demonstrates that our model not only overcomes the impact of sensitive features but also improves the performance of missing multi-view clustering tasks. The visual features indicate that our model ensures the fairness of clustering results.

1 Introduction

In the age of big data, the internet is teeming with numerous forms of information, including sensitive information. Machine learning algorithm models require a huge amount of training data to be trained, and sensitive information may cause the model to produce skewed findings toward a particular group during the training process, leading to unfair decision-making. One of the causes of the unfairness of machine learning models is the bias in training data. Especially when constructing labels for sensitive information such as gender, race, skin color, etc., artificially labeled data might occasionally include personal biases or stereotypes. For instance, a model that decides whether to offer someone a particular position may pick up on the gender bias in the training data and this may influence the judgment findings at the testing stage.

Removing sensitive features from the training data is a direct way to desensitize machine learning algorithms. In this way, sensitive attributes have no impact on the model training process. However, this strategy has certain downsides, as the erased sensitive attributes may be highly associated with other qualities, and immediate deletion is not favorable to model learning. Therefore, some effort has begun to examine how to train the model without bias while retaining sensitive variables, such that the final outcome is independent of sensitive attributes, i.e., fair machine learning. This idea was initially applied to supervised classification and regression [1]. However, unsupervised clustering tasks also present concerns with fairness. For instance, in recommendation systems, some establishments with shorter opening hours may be discriminated against due to the trait of having fewer followers while conducting cluster analysis on users and merchants. The likelihood of a store being exposed is quite minimal when the recommendation algorithm suggests a store to users. This lowers the rate of user and store matching, which is detrimental to both the user experience and the economy. To address this issue, Chierichetti et al.[2] originally suggested criteria for assessing the fairness of clustering tasks. They split the dataset into roughly balanced clusters of fairlets and used the k-means approach to establish the basic balance of sensitive qualities in each cluster. This approach works well on small-scale datasets, but it is computationally intensive and unsuitable for large-scale datasets. certain studies have suggested fairness-aware deep clustering models. These models typically include fairness constraints in clustering tasks in order to balance clustering accuracy and fairness. Inspired by these methods, we seek to add fairness constraints to the incomplete multi-view clustering work[3, 4, 5], thereby making the model independent of sensitive attributes.

On the other hand, many real-world circumstances involve data from various sources, types, or domains. These data can frequently be seen and described from a variety of perspectives, and each view may offer various features or dimensions. For example, when generating personalized suggestions for a user, one can gather varied data from the user's search history, purchase history, and browsing history. This can increase the accuracy of the recommendation system. However, the problem of multi-view clustering is how to properly combine information from several perspectives. Data between distinct views may differ and be redundant, so it is vital to determine how to deal with these concerns. Researchers have developed many techniques and methods to handle multi-view clustering problems, including matrix decomposition methods[6], kernel function methods[7], subspace learning methods[8], and generative adversarial network methods[9]. However, these methods have some limitations. The non-negative matrix decomposition method can be only applied to the non-negative data matrix. The kernel matrix filling approach cannot be extended to other clustering methods and can only be used for kernel matrix-based clustering methods. Although the subspace-based learning method does not require filling in missing data, there is a difficulty with the alignment of missing data. The method based on creating missing views may neglect their uniqueness and affect the performance of clustering.

To address the issue of incomplete multi-view clustering and ensure the fairness of the clustering results, we propose a novel fair deep multi-view clustering model based on subspace learning techniques. This model utilizes two information bottlenecks that not only remove redundant information from the complete view and extract the minimally sufficient common representation but also align the feature distribution between the incomplete view and the complete view, enhancing the consistency of multi-view data fusion. Moreover, the model imposes fairness constraints on the soft membership assignment of each cluster, guaranteeing that the proportion of different groups in each cluster is approximately equal to that in the entire dataset. Our contributions are summarized as follows:

- We consider the fairness issue of incomplete multi-view clustering tasks, which has received relatively little attention so far.
- We constrain the cluster's distribution of sensitive attributes to ensure that the distribution of sensitive groups in each cluster is close to the actual situation, preserving the fairness of the clustering.
- We apply information bottleneck theory to subspace learning, which not only learns the minimal common features required for a complete view but also aligns the distribution of missing view features and common features. By varying the degree of alignment, we can learn discriminative aspects of partial views.
- We evaluate the model on four datasets and observe significant improvements in clustering results and balance scores. The visualization experiment demonstrates the effectiveness of the fairness constraints in preserving the cluster's fairness distribution.

2 The Proposed FIMVC Framework

2.1 Symbols and Problem Definitions

The input multi-view data for this paper is set to $\mathbf{V_i}$ and $\mathbf{V_j}$, where $\mathbf{i}, \mathbf{j} \in \{0, 1, 2...\mathbf{v}\}$. Paired (complete) data in $\mathbf{V_i}$ and $\mathbf{V_j}$ are expressed as $\mathbf{X_C^i}$ and $\mathbf{X_C^j}$ and unpaired (incomplete) data as $\mathbf{X_I^i}$ and $\mathbf{X_I^j}$. $\mathbf{Z_{C+I}^i}$ represents the features that acquired from $\mathbf{V_i}$ ($\mathbf{X_C^i}$, $\mathbf{X_I^i}$) after passing through the encoder, and $\mathbf{Z_{C+I}^j}$ represents the features of $\mathbf{V_j}$ ($\mathbf{X_C^j}$, $\mathbf{X_I^j}$). $\mathbf{Z_C^i}$ and $\mathbf{Z_C^j}$ are the features of paired data $\mathbf{X_C^i}$ and $\mathbf{X_C^j}$, while $\mathbf{Z_I^i}$ and $\mathbf{Z_I^j}$ are the features of unpaired data $\mathbf{X_I^i}$ and $\mathbf{X_I^j}$. $\mathbf{Z_C}$ is the learned common feature

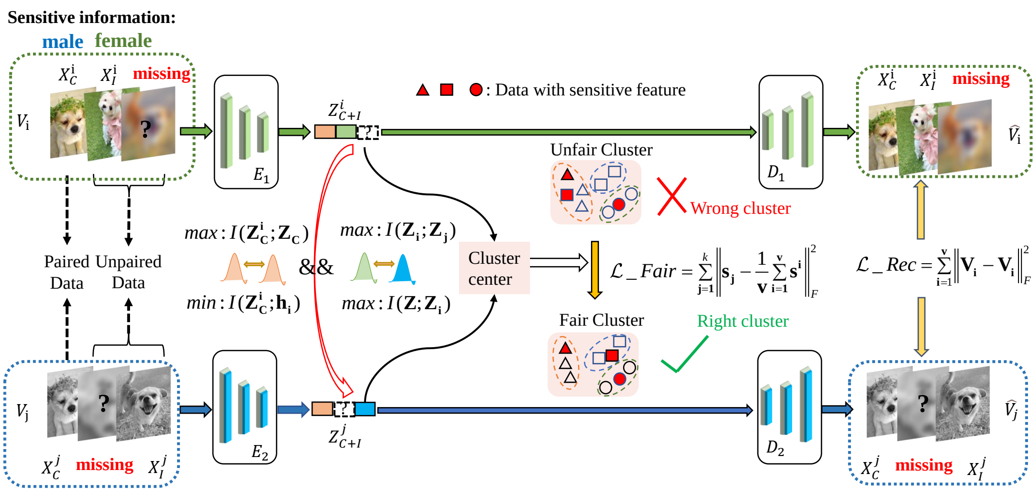


Figure 1: The framework of proposed FIMVC model. $max : I(\mathbf{Z^i_C}; \mathbf{Z_C})$ and $min : I(\mathbf{Z^i_C}; \mathbf{h_i})$ aim to learn common features in complete views. $max : I(\mathbf{Z_i}; \mathbf{Z_j})$ and $max : I(\mathbf{Z}; \mathbf{Z_i})$ are used to align the feature distribution between incomplete and complete views. $\mathcal{L}\_Fair$ is used to constrain the fairness of each cluster. $\mathcal{L}\_Rec$ is used to train encoders to ensure the degree of feature restoration to the original data.

of paired data. We created new features $\mathbf{Z_i}$ and $\mathbf{Z_j}$ by filling in each other's unpaired data features in the two views in order to align the distribution of paired and unpaired data, i.e., $\mathbf{Z_i} \supseteq \left\{\mathbf{Z^i_C}, \mathbf{Z^i_I}, \mathbf{Z^j_I}\right\}$, whereas $\mathbf{Z_j} \supseteq \left\{\mathbf{Z^j_C}, \mathbf{Z^i_I}, \mathbf{Z^j_I}\right\}$. The fused features of all data are represented as $\mathbf{Z}$, which integrates the features of $\mathbf{Z_i}$ and $\mathbf{Z_j}$, i.e., $\mathbf{Z} \supseteq \{\mathbf{Z_i}, \mathbf{Z_j}\}$. Finally, we utilize $\mathbf{s}$ to represent sensitive features and $\widehat{\mathbf{V_i}}, \widehat{\mathbf{V_j}}$ to represent the results of each view's reconstruction. Our work investigates fair clustering tasks in scenarios where missing multi-view data has sensitive attributes. Although our method can be used with multiple views, for the purposes of explanation, we only present theoretical statements in the case of two views. Then we can easily represent the model using the aforementioned symbol definitions.

2.2 Reconstruction Loss

Deep autoencoders, which exhibit outstanding feature extraction and robust learning capabilities, are frequently employed in clustering activities due to their ability to facilitate clustering operations[10, 11]. The loss function of the autoencoders is generally organized as follows:

$$\mathcal{L}\_Rec = \sum_{\mathbf{i}=1}^{\mathbf{v}} \|\mathbf{V_i} - \mathbf{D}_{\theta_\mathbf{i}}(\mathbf{E}_{\vartheta_\mathbf{i}}(\mathbf{V_i}))\|_F^2 = \sum_{\mathbf{i}=1}^{\mathbf{v}} \left\|\mathbf{V_i} - \widehat{\mathbf{V_i}}\right\|_F^2 \tag{1}$$

Where $\mathbf{V_i}$ is the i-th view; $\mathbf{E}_{\vartheta_\mathbf{i}}$ represents the encoder of the i-th view; $\mathbf{D}_{\theta_\mathbf{i}}$ represents the decoder of the i-th view. The original view is transformed by the encoder into latent features $\mathbf{Z^i_{C+I}}$, which are better suited for clustering tasks. The decoder tries to restore the latent features $\mathbf{Z^i_{C+I}}$ to the original data. This constraint ensures that the latent features do not deviate from the original data.

2.3 Fairness Constraint

A reasonable strategy is to ensure that the proportion of sensitive features acquired by clustering is consistent with the original data[3]. The distance from each data to its cluster centroid is given by:

$$\mathbf{d^i_j} = \frac{\exp(sim(\mathbf{z^i_{C+I}}, \mathbf{c_j}))}{\sum_{\mathbf{j'}} \exp(sim(\mathbf{z^i_{C+I}}, \mathbf{c_{j'}}))} \tag{2}$$

Where $\mathbf{z^i_{C+I}} \in \mathbf{Z^i_{C+I}}$, it is a sample from i-th view. $\mathbf{c_j}$ is the center of the j-th cluster. sim is the similarity function. Then, the proportion of sensitive features in each cluster is calculated as follows:

$$\mathbf{s_j} = \frac{\sum\limits_{\mathbf{i=1}}^{\mathbf{v}} \mathbf{d^i_j s^i}}{\sum\limits_{\mathbf{i=1}}^{\mathbf{v}} \mathbf{d^i_j}} \tag{3}$$

Where $\mathbf{s_j}$ denotes the number of sensitive features in the j-th cluster. $\mathbf{s^i}$ represents the data with sensitive attributes in the i-th view. $\mathbf{d^i_j}$ measures the distance between the sample in the i-th view and the center of the j-th cluster. The fairness constraint loss can then be defined as follows:

$$\mathcal{L}\_Fair = \sum_{\mathbf{j=1}}^{k} \left\| \mathbf{s_j} - \frac{1}{\mathbf{v}} \sum_{\mathbf{i=1}}^{\mathbf{v}} \mathbf{s^i} \right\|_F^2 \tag{4}$$

The second term in the equation is the average of sensitive features in the entire dataset. By optimizing this loss, we ensure the proportion of sensitive features in each cluster is close to the overall dataset distribution. This objective function can mitigate the model's bias problem on sensitive attributes.

2.4 Exploring Common Features Across Views

We first briefly review the information bottleneck. Naftali et al.[12] proposed information bottleneck to reveal the learning process of neural networks. Suppose the input is $\mathbf{X}$, and the neural network extracts its features to obtain $\widetilde{\mathbf{X}}$. Information bottleneck aims to maximize the information in $\widetilde{\mathbf{X}}$ about label $\mathbf{Y}$ and remove redundant information from $\mathbf{X}$. To balance between maximizing compressed information and preserving useful information, we formulate the objective function as follows:

$$\mathcal{L}[p(\widetilde{x}|x)] = I(\widetilde{\mathbf{X}}; \mathbf{X}) - \beta I(\widetilde{\mathbf{X}}; \mathbf{Y}) \tag{5}$$

Since the information in $\mathbf{X}$ is highly redundant, that is, $I(\widetilde{\mathbf{X}}; \mathbf{X}) \geq I(\widetilde{\mathbf{X}}; \mathbf{h})$, we can substitute encoder middle layer features $\mathbf{h}$ for $\mathbf{X}$:

$$\mathcal{L}[p(\widetilde{x}|x)] = I(\widetilde{\mathbf{X}}; \mathbf{h}) - \beta I(\widetilde{\mathbf{X}}; \mathbf{Y}) \tag{6}$$

Inspired by this idea, we apply this theory to extract common features from complete views. Using common features $\mathbf{Z_C}$ as anchors, we seek to increase the common information in paired data from multiple views while maximizing the compression of the original data. Our first objective function, as informed by information bottleneck theory is as follows:

$$\mathcal{L}\_Bottleneck1 = \sum_{i=1}^{\mathbf{v}} I(\mathbf{Z^i_C}; \mathbf{h^i}) - \beta \sum_{i=1}^{\mathbf{v}} I(\mathbf{Z^i_C}; \mathbf{Z_C}) \tag{7}$$

Where $\mathbf{Z^i_C}$ is the feature of the paired data encoded in the i-th view, $\mathbf{Z_C}$ is the learned common feature of the complete view data, and $\mathbf{h^i}$ is the data of the i-th view after passing through the fully connected layer of the encoder. However, for the missing multi-view data, the mutual information between the paired data features $\mathbf{Z^i_C}$ and the original data $\mathbf{h^i}$ of the above objective function is dimensionally inconsistent, which makes it impossible to compute. Therefore, we transform $\mathbf{Z^i_C}$ and $\mathbf{h^i}$, replacing them with $\sqrt{(\mathbf{Z^i_C})^T(\mathbf{Z^i_C})}$ and $\sqrt{(\mathbf{h^i})^T(\mathbf{h^i})}$. But for easy presentation, we still use $\mathbf{Z^i_C}$ and $\mathbf{h^i}$ in the following equations. Then, we use the Lagrange multiplier approach to solve Eq.(7). By introducing the Langrange multiplier, Eq.(7) becomes:

$$
\begin{aligned}
\mathcal{L} &= \sum_{\mathbf{i=1}}^{\mathbf{v}} I(\mathbf{Z_C^i};\mathbf{h^i}) - \beta \sum_{\mathbf{i=1}}^{\mathbf{v}} I(\mathbf{Z_C^i};\mathbf{Z_C}) - \sum_{\mathbf{i=1}}^{\mathbf{v}} \sum_{\mathbf{z_c^i},\mathbf{h_j^i}} \lambda(\mathbf{h_j^i}) p(\mathbf{z_c^i}|\mathbf{h_j^i}) \\
&= \sum_{\mathbf{i=1}}^{\mathbf{v}} \sum_{\mathbf{z_c^i},\mathbf{h_j^i}} p(\mathbf{z_c^i}|\mathbf{h_j^i}) \mathbf{p}(\mathbf{h_j^i}) \log[\frac{p(\mathbf{z_c^i}|\mathbf{h_j^i})}{p(\mathbf{z_c^i})}] - \beta \sum_{\mathbf{i=1}}^{\mathbf{v}} \sum_{\mathbf{z_c^i},\mathbf{z_c}} p(\mathbf{z_c^i}|\mathbf{z_c}) \log[\frac{p(\mathbf{z_c^i}|\mathbf{z_c})}{p(\mathbf{z_c^i})}] \\
&\quad - \sum_{\mathbf{i=1}}^{\mathbf{v}} \sum_{\mathbf{z_c^i},\mathbf{h_j^i}} \lambda(\mathbf{h_j^i}) p(\mathbf{z_c^i}|\mathbf{h_j^i})
\end{aligned} \tag{8}
$$

Where $\mathbf{z_c^i} \in \mathbf{Z_C^i}$, $\mathbf{h_j^i} \in \mathbf{h}^i$. β is used to constrain information and $\lambda(\mathbf{h_j^i})$ is the Langrange multiplier used for normalizing the conditional distribution $p(\mathbf{z_c^i}|\mathbf{h_j^i})$ at each $\mathbf{h_j^i}$. Finding the partial derivation of $p(\mathbf{z_c^i}|\mathbf{h_j^i})$ for Eq.(8), we can get:

$$
\begin{aligned}
\frac{\partial \mathcal{L}}{\partial p(\mathbf{z_c^i}|\mathbf{h_j^i})} = \sum_{\mathbf{i=1}}^{\mathbf{v}} p\left(\mathbf{h_j^i}\right) [\log \frac{p(\mathbf{z_c^i}|\mathbf{h_j^i})}{p(\mathbf{z_c^i})} + \beta \sum_{\mathbf{z_c}} p(\mathbf{z_c}|\mathbf{h_j^i}) \log \frac{p(\mathbf{z_c}|\mathbf{h_j^i})}{p(\mathbf{z_c}|\mathbf{z_c^i})} - \frac{\lambda(\mathbf{h_j^i})}{p(\mathbf{h_j^i})} \\
+ \beta \sum_{\mathbf{z_c}} p(\mathbf{z_c}|\mathbf{h_j^i}) \log \frac{p(\mathbf{z_c}|\mathbf{h_j^i})}{p(\mathbf{z_c})}]
\end{aligned} \tag{9}
$$

Let Eq.(9) equal to 0, we can obtain the optimal solution for Eq.(7):

$$
p(\mathbf{z_c^i}|\mathbf{h_j^i}) = \frac{p(\mathbf{z_c^i})}{\mathbf{\Gamma}(\mathbf{h_j^i}, \beta)} \exp(-\beta \mathbf{D_{KL}}[p(\mathbf{z_c}|\mathbf{h_j^i})|p(\mathbf{z_c}|\mathbf{z_c^i})]) \tag{10}
$$

where $\mathbf{D_{KL}}$ is the Kullback–Leibler divergence, $\mathbf{\Gamma}$ satisfies:

$$
\mathbf{\Gamma}(\mathbf{h_j^i}, \beta) = \sum_{\mathbf{i=1}}^{\mathbf{v}} \sum_{\mathbf{z_c^i}} \mathbf{p}(\mathbf{z_c^i}) \exp(-\beta \mathbf{D_{KL}}[\mathbf{p}(\mathbf{z_c}|\mathbf{h_j^i})|\mathbf{p}(\mathbf{z_c}|\mathbf{z_c^i})]) \tag{11}
$$

2.5 Unpaired Data Distribution Alignment

One major challenge in missing multi-view clustering is the fusion of incomplete data. Previous methods of imputing or recovering missing views[6, 7] often overlooked the uniqueness of missing data, which was not conducive to model generalization. We propose a fusion method that fully utilizes incomplete view data while aligning its distribution with complete data. The objective function for aligning partial and complete view distributions is as follows:

$$
\mathcal{L}\_\mathit{Bottleneck2} = - \sum_{\mathbf{i=1,i\neq j}}^{\mathbf{v}} I(\mathbf{Z_i};\mathbf{Z_j}) - \sum_{\mathbf{i=1}}^{\mathbf{v}} I(\mathbf{Z};\mathbf{Z_i}) \tag{12}
$$

The first mutual information term $I(\mathbf{Z_i};\mathbf{Z_j})$ aligns the distribution of unpaired data. Since $\mathbf{Z_i} \supseteq \left\{\mathbf{Z_C^i}, \mathbf{Z_I^i}, \mathbf{Z_I^j}\right\}$, $\mathbf{Z_j} \supseteq \left\{\mathbf{Z_C^j}, \mathbf{Z_I^i}, \mathbf{Z_I^j}\right\}$, maximizing the mutual information between two variables can be reformulated as maximizing the KL divergence of their joint distribution and marginal distribution:

$$
I(\mathbf{Z_i};\mathbf{Z_j}) = \mathbf{D_{KL}}[p(\mathbf{Z_i},\mathbf{Z_j})||p(\mathbf{Z_i})p(\mathbf{Z_j})] \tag{13}
$$

Hence, the upper bound on mutual information can be obtained by finding the maximum value of KL divergence. However, since the maximum value of KL divergence is divergent, we adopt the convergent Jensen-Shannon MI estimator[13] instead. The upper bound of mutual information $I(\mathbf{Z_i};\mathbf{Z_j})$ is expressed as:

$$
\begin{aligned}
\max I(\mathbf{Z_i};\mathbf{Z_j}) = \mathbf{D_{JS}}[p(\mathbf{Z_i},\mathbf{Z_j})||p(\mathbf{Z_i})p(\mathbf{Z_j})] = \max E_{(\mathbf{Z_i},\mathbf{Z_j})}[\log D(\mathbf{Z_i},\mathbf{Z_j})] \\
+ E_{(\mathbf{Z_i},\mathbf{Z_j})}[\log(1 - D(\mathbf{Z_i},\mathbf{Z_j}))]
\end{aligned} \tag{14}
$$

On the other hand, to prevent the distribution of incomplete data from deviating from common features, we introduced the latter mutual information terms $I(\mathbf{Z};\mathbf{Z_i})$ to align the distribution between paired data and unpaired data. The solution process is analogous to the first part:

$$\begin{aligned}\max I(\mathbf{Z};\mathbf{Z_i}) = \mathbf{D_{JS}}[p(\mathbf{Z},\mathbf{Z_i})||p(\mathbf{Z})p(\mathbf{Z_i})] = \max E_{(\mathbf{Z},\mathbf{Z_i})}[\log D(\mathbf{Z},\mathbf{Z_i})] \\ +E_{(\mathbf{Z},\mathbf{Z_i})}[\log(1-D(\mathbf{Z},\mathbf{Z_i}))]\end{aligned} \tag{15}$$

Where $\mathbf{Z} \supseteq \{\mathbf{Z_i}, \mathbf{Z_j}\}$, D is the discriminator used to calculate the input pair's probability. By solving the mutual information maximization term, we can prevent unpaired data from deviating from the overall distribution, achieving a consistent distribution of complete and incomplete view data.

2.6 Objective Function

In summary, the objective loss function of our model is as follows:

$$\mathcal{L}\_all = \lambda_1 \mathcal{L}\_Rec + \lambda_2 \mathcal{L}\_Bottleneck1 + \lambda_3 \mathcal{L}\_Bottleneck2 + \gamma \mathcal{L}\_Fair \tag{16}$$

Where $\mathcal{L}$*_Rec* denotes the reconstruction loss of the encoder and decoder. $\mathcal{L}$*_Bottleneck1* is the first loss of information bottleneck used to discover common representations across views. $\mathcal{L}$*_Bottleneck2* is the second bottleneck loss that is utilized to align the feature distribution between incomplete and complete view data. This constraint makes unpaired data closer to its cluster in the feature subspace. λ_3 is employed to control alignment loss, and excessive weights might impact feature resolution, resulting in the cluster being reduced to a single class. A weight that is too little can cause incomplete view data to depart from the cluster to which it belongs. $\mathcal{L}$*_Fair* is the fair constraint loss that ensures that the distribution of sensitive features in each cluster is near the true distribution.

3 Experiments

3.1 Experimental Setup

Datasets We validated the effectiveness of the model on four common fairness datasets, including Credit Card dataset[14], Zafar dataset[14], Bank dataset[14], Adult dataset[14]. The sensitive feature is gender in Credit Card and Adult, and marital status in Bank dataset, while in Zafar dataset, one binary value is generated as the sensitive feature. We preprocessed all four datasets for ease of observation. When the same number of samples are drawn from each category, the proportion of sensitive features in each category is approximately the same. For example, in the credit dataset, the number of male and female features in each category is roughly equal. We conducted our experiments on Windows10 systems with Python version 3.7 and Cuda 11.5. The hidden layer dimension for all autoencoder-related algorithms is set to 200 and every random seed has been set to 8.

Comparison Algorithm We compare the model with five clustering methods, which we will briefly discuss below:

- K-Means: a clustering method for assigning samples into nearest clusters.
- DEC[15]: a process for learning embedded features and clustering them with neural networks.
- CC[16]: a clustering technique that integrates contrastive learning and optimizes instance- and cluster-level contrastive losses.
- MvDSCN[17]: a subspace clustering model for learning multi-view self-expression matrices
- Fair-MVC[18]: a method that combines fair clustering with multi-view clustering and adds contrast regularization to improve feature representation.

Evaluation Criteria We evaluated two indicators of the model in clustering tasks: (1) normalized mutual information(NMI); (2) balance score. The NMI measures the coherence of clustering. The balance score assesses the fairness of clustering, which is defined as follows:

$$Balance = \min_i \frac{\min |C_i \cap s_j|}{|C_i|} \tag{17}$$

Table 1: Clustering result and balance score on the Credit Card dataset

Missing Rate	0		0.25		0.5	
Evaluation Metrics	NMI	Balance	NMI	Balance	NMI	Balance
K-Means	20.940 $\pm$ 1.14	35.530 $\pm$ 0.37	15.670 $\pm$ 1.48	36.020 $\pm$ 0.60	13.560 $\pm$ 0.63	36.320 $\pm$ 0.38
DEC	21.030 $\pm$ 2.09	35.960 $\pm$ 0.60	20.050 $\pm$ 0.79	36.400 $\pm$ 0.78	15.670 $\pm$ 1.21	36.260 $\pm$ 0.40
MvDSCN	21.920 $\pm$ 1.53	35.820 $\pm$ 0.41	20.340 $\pm$ 1.59	35.940 $\pm$ 0.48	16.340 $\pm$ 1.83	36.630 $\pm$ 0.69
CC	23.870 $\pm$ 1.28	35.740 $\pm$ 0.47	20.950 $\pm$ 0.68	36.160 $\pm$ 0.71	17.620 $\pm$ 1.01	36.800 $\pm$ 0.97
Fair-MVC	24.950 $\pm$ 0.41	41.980 $\pm$ 0.32	22.090 $\pm$ 0.75	38.200 $\pm$ 0.46	18.710 $\pm$ 0.79	39.070 $\pm$ 0.76
FIMVC(ours)	**26.890 $\pm$ 0.53**	**47.360 $\pm$ 1.54**	**24.360 $\pm$ 0.62**	**46.490 $\pm$ 1.53**	**20.390 $\pm$ 0.54**	**46.110 $\pm$ 1.56**

Where C_i denotes the the i-th cluster, and s_j denotes j-th protected subgroup. The distribution of sensitive features usually determines the upper limit of balance, and a higher balance value indicates a fairer outcome.

3.2 Experimental Results

Cluster Results and Fairness Scores Table 1 demonstrates that our method outperforms existing clustering methods on the Credit Card dataset, including the most recent fair clustering algorithm. Despite a certain trade-off between the balance score and clustering accuracy, the results show that our algorithm has high clustering accuracy. This is consistent with our second information bottleneck objective function, which aligns the distribution of partial and complete view data, allowing the learned features to be more useful for clustering tasks.

Table 2 presents the results of each algorithm on the Zafar dataset. Fair-MVC's fairness score is substantially higher than other comparison algorithms, as this method considers clustering fairness when completing the multi-view clustering problem. It maintains clustering fairness by using fairness restrictions and a balanced membership calculation approach. However, this method sacrifices higher fairness scores for lower clustering accuracy.

The results of each method on the Bank dataset are illustrated in Table 3. It can be seen that the MvDSCN model produces considerably better clustering results than K-Means and DEC. This is due to the fact that the MvDSCN model learns the multiview data in the subspace, fuses the multiview data by building a self-expression layer, and then uses the self-expression matrix for spectral clustering. This strategy improves the learning of multi-view information.

Table 4 displays the Adult dataset results for each algorithm. Our model exceeds Fair-MVC in clustering NMI by 2% on average, and the fairness score of clustering clusters is 5% higher on average. This result shows that our model has high feature extraction and fusion capabilities, which can improve the degree of balance while maintaining clustering accuracy.

Visualization Figure 2 depicts the visualization results of raw data and features with and without fairness constraints on Credit Card dataset. Males are represented by blue, while females are represented by red. The figure shows that male and female traits are evenly distributed in the original data due to the almost equal number of men and women in each category. However, in clustering results without fairness constraints, some clusters have significant bias, i.e., they consist of only male features, which are inconsistent with the original data distribution. The gender ratio in the cluster with fairness constraints remains balanced, similar to the distribution of the original data. This result indicates that our model can be independent of sensitive attributes; that is, the clustering results are unbiased, achieving the goal of fair clustering.

Table 2: Clustering result and balance score on the Zafar dataset

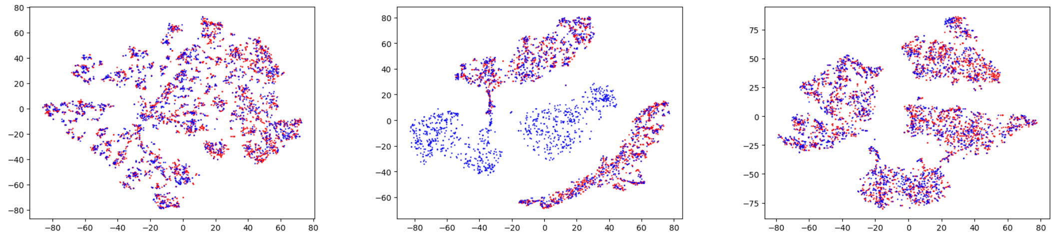
Missing Rate	0		0.25		0.5	
Evaluation Metrics	NMI	Balance	NMI	Balance	NMI	Balance
K-Means	70.320 ± 0.78	17.060 ± 0.76	64.830 ± 1.21	16.350 ± 0.44	60.240 ± 0.90	16.230 ± 1.11
DEC	72.550 ± 1.92	16.850 ± 0.73	70.890 ± 1.33	17.110 ± 0.84	64.930 ± 1.28	16.960 ± 0.93
MvDSCN	76.910 ± 0.42	17.130 ± 0.65	74.980 ± 1.01	18.320 ± 0.88	69.580 ± 1.66	16.870 ± 0.99
CC	78.950 ± 0.68	17.010 ± 0.71	75.220 ± 0.87	18.410 ± 0.67	72.130 ± 0.74	17.960 ± 0.83
Fair-MVC	81.610 ± 0.57	28.960 ± 0.59	79.890 ± 0.55	30.640 ± 0.89	76.970 ± 0.79	29.330 ± 0.80
FIMVC(ours)	**84.360 ± 0.81**	**35.210 ± 0.67**	**81.440 ± 0.66**	**35.330 ± 0.91**	**79.120 ± 0.73**	**34.670 ± 0.85**

Table 3: Clustering result and balance score on the Bank dataset

Missing Rate	0		0.25		0.5	
Evaluation Metrics	NMI	Balance	NMI	Balance	NMI	Balance
K-Means	28.670 ± 1.44	37.650 ± 0.66	24.770 ± 1.71	36.660 ± 0.45	21.080 ± 1.17	36.320 ± 0.89
DEC	30.930 ± 1.15	37.600 ± 0.96	28.970 ± 1.80	36.420 ± 0.69	24.370 ± 1.02	37.090 ± 0.78
MvDSCN	36.240 ± 0.55	37.590 ± 0.67	35.020 ± 1.21	36.110 ± 0.57	31.330 ± 1.76	36.960 ± 0.77
CC	36.230 ± 1.01	37.460 ± 0.97	34.880 ± 1.33	37.690 ± 0.95	31.130 ± 1.63	37.800 ± 0.85
Fair-MVC	38.990 ± 0.91	42.400 ± 0.75	36.660 ± 1.01	41.760 ± 0.83	32.900 ± 1.03	41.610 ± 0.67
FIMVC(ours)	**41.280 ± 0.80**	**46.640 ± 0.67**	**39.310 ± 1.05**	**45.640 ± 0.93**	**35.130 ± 0.91**	**45.330 ± 0.76**

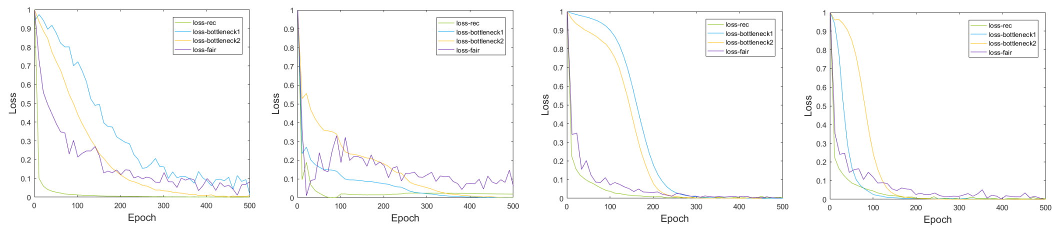
Table 4: Clustering result and balance score on the Adult datasets

Missing Rate	0		0.25		0.5	
Evaluation Metrics	NMI	Balance	NMI	Balance	NMI	Balance
K-Means	61.480 ± 0.61	28.530 ± 0.47	56.780 ± 1.23	27.440 ± 0.75	53.710 ± 1.04	26.690 ± 0.69
DEC	64.060 ± 0.67	27.810 ± 0.62	61.010 ± 1.10	27.870 ± 0.91	56.960 ± 0.78	28.110 ± 0.70
MvDSCN	68.120 ± 0.84	28.230 ± 0.79	65.660 ± 1.22	28.340 ± 0.69	61.660 ± 0.97	28.300 ± 0.35
CC	67.960 ± 0.66	26.380 ± 0.91	64.680 ± 0.81	27.580 ± 0.68	59.880 ± 1.11	26.480 ± 0.79
Fair-MVC	70.120 ± 0.84	35.810 ± 0.51	66.750 ± 0.78	36.020 ± 0.89	63.770 ± 0.73	35.780 ± 0.79
FIMVC(ours)	**72.710 ± 0.79**	**41.170 ± 0.83**	**68.680 ± 0.83**	**41.370 ± 0.82**	**65.110 ± 1.10**	**40.480 ± 0.92**



(a) raw data (a) without fair constraints (a) with fair constraints

Figure 2: Visualize sensitive features in Credit Card dataset using t-SNE. Blue represents the male feature, while red represents the female. The original data is visualized on the left, the result without fair constraint is shown in the middle, and the result with fair constraint is shown on the right.



(a) Credit Card (b) Zafar (c) Bank (d) Adult

Figure 3: Convergence curves on the four datasets

Convergence Analysis Figure 3 depicts the convergence curves of the model's objective functions on four datasets. We normalize the loss function values of the four parts because their values are not of the same order of magnitude. After training 500 batches, each part of the loss on the four data sets is reduced to a constant value. This suggests that our model has a high level of convergence.

4 Conclusion

In this paper, we proposed a fair clustering method for missing multi-view data that applies traditional information bottleneck methods to complete views and maximize mutual information on incomplete data. Our method ensures that the distribution of sensitive features within each cluster is similar to the distribution in the overall dataset, preventing biased information about sensitive features from being learned by the model. We also address the issue of fusion and alignment of incomplete view data. The results on four datasets show the effectiveness of our method, and the feature visualization results confirm the fairness of the model.

References

[1] Moritz Hardt, Eric Price, and Nathan Srebro. Equality of opportunity in supervised learning. *ArXiv*, abs/1610.02413, 2016.

[2] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, page 5036–5044, Red Hook, NY, USA, 2017. Curran Associates Inc.

[3] Matthäus Kleindessner, Samira Samadi, Pranjal Awasthi, and Jamie Morgenstern. Guarantees for spectral clustering with fairness constraints. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 3458–3467. PMLR, 09–15 Jun 2019.

[4] Imtiaz Masud Ziko, Eric Granger, Jing Yuan, and Ismail Ben Ayed. Clustering with fairness constraints: A flexible and scalable approach. *ArXiv*, abs/1906.08207, 2019.

[5] Peizhao Li, Han Zhao, and Hongfu Liu. Deep fair clustering for visual learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9067–9076, 2020.

[6] Shao-Yuan Li, Yuan Jiang, and Zhi-Hua Zhou. Partial multi-view clustering. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'14, page 1968–1974. AAAI Press, 2014.

[7] Piyush Rai and Scott L Duvall. Multiview clustering with incomplete views. In *NIPS*, 2010.

[8] Qiyue Yin, Shu Wu, and Liang Wang. Incomplete multi-view clustering via subspace learning. In *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, CIKM '15, page 383–392, New York, NY, USA, 2015. Association for Computing Machinery.

[9] Qianqian Wang, Zhengming Ding, Zhiqiang Tao, Quanxue Gao, and Yun Fu. Partial multi-view clustering via consistent gan. In *2018 IEEE International Conference on Data Mining (ICDM)*, pages 1290–1295, 2018.

[10] Jie Wen, Zheng Zhang, Yong Xu, Bob Zhang, Lunke Fei, and Guo-Sen Xie. Cdimc-net: Cognitive deep incomplete multi-view clustering network. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, IJCAI'20, 2021.

[11] Xi Peng, Jiashi Feng, Shijie Xiao, Wei-Yun Yau, Joey Tianyi Zhou, and Songfan Yang. Structured autoencoders for subspace clustering. *IEEE Transactions on Image Processing*, 27(10):5076–5086, 2018.

[12] Naftali Tishby, Cicero Pereira, and William Bialek. The information bottleneck method. *Proceedings of the 37th Allerton Conference on Communication, Control and Computation*, 49, 07 2001.

[13] Sebastian Nowozin, Botond Cseke, and Ryota Tomioka. f-gan: Training generative neural samplers using variational divergence minimization, 2016.

[14] Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P. Gummadi. Fairness constraints: Mechanisms for fair classification, 2017.

[15] Junyuan Xie, Ross Girshick, and Ali Farhadi. Unsupervised deep embedding for clustering analysis, 2016.

[16] Yunfan Li, Peng Hu, Zitao Liu, Dezhong Peng, Joey Tianyi Zhou, and Xi Peng. Contrastive clustering, 2020.

[17] Qianqian Wang, Zhiqiang Tao, Quanxue Gao, and Licheng Jiao. Multi-view subspace clustering via structured multi-pathway network. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–7, 2022.

[18] Lecheng Zheng, Yada Zhu, and Jingrui He. Fairness-aware multi-view clustering, 2023.