Common-Individual Semantic Fusion for Multi-View Multi-Label Learning

Gengyu Lyu¹, Weiqi Kang¹, Haobo Wang², Zheng Li¹, Zhen Yang¹*, Songhe Feng²*
¹Faculty of Information Technology, Beijing University of Technology
²School of Software Technology, Zhejiang University
{lyugengyu, lizhengcn, yangzhen} @bjut.edu.cn, weiqikang@email.bjut.edu.cn, wanghaobo@zju.edu.cn, shfeng@bjtu.edu.cn

Abstract
In Multi-View Multi-Label Learning, each instance is described by several heterogeneous features and associated with multiple valid labels simultaneously. Existing methods mainly focus on leveraging feature-level view fusion to capture a common representation for multi-label classifier induction. In this paper, we take a new perspective and propose a new semantic-level fusion model named Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). Different from previous feature-level fusion model, our proposed method directly focuses on semantic-level view fusion and simultaneously take both the common semantic across different views and the individual semantic of each specific view into consideration. Specifically, we first assume each view involves some common semantic labels while owns a few exclusive semantic labels. Then, the common and exclusive semantic labels are separately forced to be consensus and diverse to excavate the consistences and complementarities among different views. Afterwards, we introduce the low-rank and sparse constraint to highlight the label co-occurrence relationship of common semantics and the view-specific expression of individual semantics. We provide theoretical guarantee for the strict convexity of our method by properly setting parameters. Extensive experiments on various data sets have verified the superiority of our method.

1 Introduction
Multi-View Multi-Label Learning (MVML) learns from the training data, where each object is represented by several heterogeneous feature representations and associated with multiple class labels simultaneously [Bickel and Scheffer, 2004; Wu et al., 2019; Wu et al., 2020; Lyu et al., 2022a]. Recently, such learning paradigm has been widely used in many real-world applications. For example, in the task of news webpage classification, one news webpage can be represented by diverse channel information including video, image and text, while at the same time it is annotated with multiple class labels such as FIFA World Cup, Sports and Lionel Messi simultaneously.

The World Cup records Messi owns

Figure 1: An Example of Multi-View Multi-Label Learning. The news webpage is represented by several different channels including video, image and text, and associated with multiple class labels such as FIFA World Cup, Sports and Lionel Messi simultaneously.

The key to deal with multi-view multi-label data lies in how to effectively integrate these heterogeneous features while ensuring all relevant labels can be characterized comprehensively. A general strategy is to learn a latent multi-view subspace representation, which can characterize multiple semantic labels as much intact as possible. [Zhang et al., 2018a] proposes a matrix factorization based shared subspace representation method, which employs Hilbert-Schmidt independence criterion to strengthen its ability of consensus semantic characterization. [Zhao et al., 2023] also seeks for a latent low-dimensional representation, while it focuses on each specific view and employs the structural view-label consistency information to enhance the expressions of view-specific semantics. [Lu et al., 2023] proposes a bipartite graph based multi-view embedding representation method, and it imposes a joint low-rank constraint on both the embedding representation matrix and multi-label classifier matrix to enhance its robustness toward the labels with dependencies. Obviously, the above MVML methods just leverage the multi-view consensus and complementary relationship in the feature space and they formulate multi-view subspace as implicit semantic-aware representation.

*Zhen Yang and Songhe Feng are the corresponding authors.
In this paper, different from traditional feature-level multi-view fusion MVML methods, we take the first attempt to conduct multi-view fusion under the guidance of semantic fusion, and propose a Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). Specifically, we first assume each view corresponds to one view-specific label set and each of them contains two different kinds of semantics - Common Semantics & Individual Semantics. The common semantics refers to the core semantic labels shared by multiple views, which reflects the consensus information across different views. The individual semantics refers to the exclusive semantic labels owned by each specific view, which characterizes the complementary information of diverse views. Afterwards, we separately introduce the low-rank and sparse constraint to highlight the label co-occurrence relationship of the common semantics and the exclusive semantic representation of the individual semantics. Finally, we embed an adaptive global label correlations to enhance the semantic integrity for improving the performance of the final multi-label model.

In summary, the main contribution of our paper lie in the following aspects:

- We propose a new Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method (CISF). To the best of our knowledge, it is the first attempt to directly leverage multi-view fusion on the semantic space.

- Considering that single common semantics cannot characterize all relevant labels, we simultaneously consider the commonality and individuality of multi-view data, and introduce an adaptive global label correlations to enhance the semantic integrity of the final model.

- We provide theoretical guarantee for the strict convexity of CISF by properly setting parameters and develop an alternative optimization algorithm to solve it. Extensive results have verified the superiority of our method.

2 Related Work

2.1 Multi-View Learning (MVL)

Multi-View Learning aims to learn a desired multi-view representation from different views by leveraging the consensus and complementary information across heterogeneous features [Bickel and Scheffer, 2004; Wang et al., 2016; Zhang et al., 2017; Gu et al., 2023]. A core purpose of multi-view learning is to encapsulate multi-view information of different views to learn a share or common representation for clustering. Based on the way to exploit multi-view information, existing multi-view learning methods can be roughly divided into the following categories: canonical correlation analysis [Andrew et al., 2013; Wang et al., 2015], multi-view subspace clustering [Gao et al., 2015; Cao et al., 2015; Kang et al., 2020; Wang et al., 2021], multi-view matrix factorization [Liu et al., 2013; Zhao et al., 2017], and deep multi-view clustering [Li et al., 2019; Xu et al., 2023]. Besides, there are also many other multi-view learning methods for different tasks, such as retrieval [Yan et al., 2020], recommendation [Flanagan et al., 2021] and classification [Han et al., 2022; Lyu et al., 2022a], etc.

2.2 Multi-Label Learning (MLL)

In Multi-Label Learning, each instance is represented by a single feature vector and annotated with multiple valid labels [Wen et al., 2022]. Label correlation is a fundamental challenge to be utilized for improving the performance of multi-label learning. Based on the order of label correlations being exploited for model training, existing MLL methods can be roughly grouped into three categories: first-order correlations [Zhang et al., 2018b], second-order correlations [Majdandzic et al., 2013; Li et al., 2017] and high-order correlations [Burkhardt and Kramer, 2018]. The above methods mainly are formulated under full supervised settings while such phenomenon may not hold in real-world scenarios due to expensive annotation efforts. Recently, some weakly supervised MLL frameworks are proposed and have been widely used in many applications, such as semi-supervised MLL [Wang et al., 2020], MLL with missing labels [Zhu et al., 2018], partial multi-label learning [Lyu et al., 2020; Li et al., 2021; Lyu et al., 2022b; Wang et al., 2023], etc.

2.3 Multi-View Multi-Label Learning (MVML)

In Multi-View Multi-Label Learning, each instance is represented by several heterogeneous features and associated with multiple valid labels [Lyu et al., 2024]. Obviously, such paradigm can be regard as an integration of multi-view learning and multi-label learning, and the key to deal with MVML data lies in how to integrate heterogeneous features effectively while realize multi-label classification accurately. [Zhang et al., 2020] propose a sparse feature selection MVML method, which exploits both view relations and label correlations to select discriminative features for further multi-label model training. [Wu et al., 2019] propose a view-specific MVML method named SIMM, which simultaneously leverages shared subspace exploitation and view-specific information extraction to enhance the performance of multi-label classifier. Except for the above MVML methods, there are also some weakly supervised MVML methods, including MVML with missing labels [Huang et al., 2019], MVML with missing views [Tan et al., 2018], non-aligned MVML [Zhao et al., 2023; Zhong et al., 2024], etc.

3 The Proposed Method

3.1 Notations

Formally speaking, we denote $X = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \cdots \times \mathbb{R}^{d_v}$ as the feature space with $V$ views and $Y = \{c_1, c_2, \ldots, c_q\}$ as the label space with $q$ class labels, where $d_t (1 \leq t \leq T)$ is the feature dimension of $t$-th view. Given the multi-view multi-label training data $D = \{(X_i, y_i)\}_{1 \leq i \leq n}$ with $n$ instances, where each $X_i \in X$ is represented by $V$ feature vectors $[x_i^{(1)}; x_i^{(2)}; \ldots; x_i^{(V)}]$, and $y_i \in \{0, 1\}^{q \times 1}$ is the label vector associated with $X_i$, our proposed CISF aims to integrate these heterogeneous representations from different views to construct a robust multi-label classifier $f : X \mapsto 2^Y$ and further predicts some proper labels for unseen instances.

3.2 Formulation

Consistences and complementarities are two key ingredients for boosting multi-view multi-label learning. Existing
MVML methods generally advocates different view to predict the same label results to use consistent information across different views and simultaneously considers the different contribution weights of each specific view to learn the complementary information among different views. Eq. (1) illustrates a general MVML framework:

$$\min_{W} \sum_{i=1}^{V} \mu^{(i)} \| Y - W^{(i)} X^{(i)} \|_{F}^{2} + \Gamma(W^{(i)}),$$  \hspace{1cm} (1)$$

where $Y \in \mathbb{R}^{q \times n}$ is the label matrix, $X^{(i)} \in \mathbb{R}^{d_{i} \times n}$ is the feature matrix of $i$-th view, $W^{(i)} \in \mathbb{R}^{q \times d_{i}}$ is the mapping matrix from features to semantics, $\Gamma(\cdot)$ is the regularization term and $\mu^{(i)}$ is the weight of $i$-th view.

According to Eq.(1), we can easily see that existing MVML methods mainly formulate multi-view consistent and complementary information in the feature space. Basically, they target to seek one or several good view that can characterize all relevant labels as completely as possible. Obviously, such multi-view fusion strategy cannot respect every view sufficiently and even lead the semantic representation of some rare labels to be overwhelmed by the core labels.

Hence, in this paper, we directly take the multi-view fusion in the semantic space and explicitly measure multi-view consistences and complementarities in objective fusion model. Specifically, we assume that each view corresponds to a view-specific label set $Y^{(i)} \in \mathbb{R}^{q \times n}$ and each label set contains two different kinds of semantics - Common Semantics $C^{(i)} \in \mathbb{R}^{q \times n}$ and Individual Semantics $D^{(i)} \in \mathbb{R}^{q \times n}$, i.e.,

$$\min_{C,D} \sum_{i=1}^{V} \mu^{(i)} \| Y^{(i)} - \left( C^{(i)} + D^{(i)} \right) \|_{F}^{2} + \Phi(C) + \Psi(D).$$  \hspace{1cm} (2)$$

The common semantics $C^{(i)} \in \mathbb{R}^{q \times n \times V}$ refers to the shared semantic labels represented by all $V$ views and it reflects the consensus semantic information across different views. In our model, we introduce Hilbert-Schmidt Independence Criterion [Gretton et al., 2005] to constrain the semantic consistence of the $V$ different views, i.e.,

$$\Phi(C) = - \sum_{i,j=1}^{V} \mu^{(i)} \mu^{(j)} \mathbb{H}SIC(C^{(i)}, C^{(j)}).$$  \hspace{1cm} (3)$$

The individual semantics $D^{(i)} \in \mathbb{R}^{q \times n V}$ refers to the exclusive semantic labels owned by each specific view, which characterizes the diversities and complementarities among different views. Based on the assumption that the individual semantics is exclusive for each specific view and the diversity is also sparse across different views, we measure such semantic diversities and complementarities by minimizing the sum of the product of each pair of individual semantics, i.e.,

$$\Psi(D) = \sum_{i,j=1}^{V} \mu^{(i)} \mu^{(j)} Tr \left( D^{(i)} \cdot D^{(j)\top} \right).$$  \hspace{1cm} (4)$$

Except for the above semantics consistences and complementarities, another inherent property of learning from multi-view multi-label data is how to utilize label correlations. Different from previous fixed label co-occurrence relationships, we try to leverage a dynamic label correlations $S \in \mathbb{R}^{q \times q}$ and recover all relevant labels by minimizing

$$\min_{S} \sum_{i=1}^{V} \mu^{(i)} \| Y - SY^{(i)} \|_{F}^{2}.$$  \hspace{1cm} (5)$$

In addition, in order to construct the direct correspondences from features to semantics and obtain a desired multi-label classifier for unseen examples prediction simultaneously, we introduce two feature mapping matrices $W_{C}^{(i)}, W_{D}^{(i)} \in \mathbb{R}^{q \times d_{i}}$ that correspond to the common semantics $C^{(i)}$ and the individual semantics $D^{(i)}$ respectively, i.e.,

$$C^{(i)} = W_{C}^{(i)} X^{(i)}, \quad D^{(i)} = W_{D}^{(i)} X^{(i)}.$$
$W^{(i)}X^{(i)}$ and $D^{(i)} = W^{(i)}X^{(i)}$. By integrating the above functions (2)-(5), we can obtain the final framework of our proposed CISF method as follows:

$$\begin{align*}
\min_{W^{(i)}, W^{(i)}_D, S} & \sum_{i=1}^V \mu^{(i)} \left\| Y - S(W^{(i)} + W^{(i)}_D)X^{(i)} \right\|_F^2 \\
+ & \alpha \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} \left( -\mathcal{H}SI(C(W^{(i)}_C X^{(i)}, SW^{(j)} C X^{(j)})) \right) \\
+ & \beta \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} Tr \left( W^{(i)}_D X^{(i)} \cdot (W^{(j)}_D X^{(j)})^T \right) \\
+ & \gamma \sum_{i=1}^V \left\| W^{(i)}_C \right\|_*, + \eta \sum_{i=1}^V \left\| W^{(i)}_D \right\|_F^2.
\end{align*}$$

(6)

Here, the common feature mapping matrix $W^{(i)}$ is constrained with nuclear norm to preserve its low-rank property, since its represented shared semantic labels tend to have statistical co-occurrence. The exclusive feature mapping matrix is constrained with $F$-norm, since the diversities of exclusive semantics are always expressed as sparse. The weights $\mu^{(i)}$ are defined by inverse distance weighting strategy to avoid undesired hyperparameters [Nie et al., 2016].

### 3.3 Optimization

To optimize (6) conveniently, we introduce an additional variable constraint $A^{(i)} = SW^{(i)} C X^{(i)}$ and convert (6) to its Augmented Lagrange Multiplier (ALM) form as follows:

$$\begin{align*}
\min_{W^{(i)}, W^{(i)}_D, A, S} & \sum_{i=1}^V \mu^{(i)} \left\| Y - S(W^{(i)} + W^{(i)}_D)X^{(i)} \right\|_F^2 \\
+ & \alpha \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} \mathcal{H}SI(C(A^{(i)}, A^{(j)})) + \gamma \sum_{i=1}^V \left\| W^{(i)}_C \right\|_* \\
+ & \beta \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} Tr \left( W^{(i)}_D X^{(i)} \cdot (W^{(j)}_D X^{(j)})^T \right) \\
+ & \sum_{i=1}^V \lambda^{(i)} \frac{1}{2} \left\| A^{(i)} - SW^{(i)} C X^{(i)} \right\|_F^2 + \eta \sum_{i=1}^V \left\| W^{(i)}_D \right\|_F^2 \\
+ & \sum_{i=1}^V Tr \left( M^{(i)} - (A^{(i)} - SW^{(i)} C X^{(i)}) \right).
\end{align*}$$

(7)

Obviously, the above function involves four variables $W^{(i)}$, $W^{(i)}_D$, $A$ and $S$, which cannot be optimized simultaneously. Therefore, we adopt the alternating minimization strategy and update these variables iteratively.

**Update $W^{(i)}_C$ with other variables fixed.** We can calculate $W^{(i)}_C$ by minimizing the following objective function:

$$\begin{align*}
\min_{W^{(i)}_C} & \sum_{i=1}^V \zeta^{(i)} \left\| \Theta (\zeta^{(i)}) - SW^{(i)} C X^{(i)} \right\|_F^2 + \gamma \sum_{i=1}^V \left\| W^{(i)}_C \right\|_*.
\end{align*}$$

(8)

where $\Theta = 2\mu^{(i)} Y - SW^{(i)} C X^{(i)} + (\lambda^{(i)} C^{(i)} + M^{(i)})$ and $\zeta^{(i)} = (2\mu^{(i)} + \lambda^{(i)})/2$. According to [Zhu et al., 2010], (8) has the closed form solution and the variable $W^{(i)}_C$ can be optimized following $W^{(i)}_C = S \frac{\gamma}{2\zeta^{(i)}} (\Theta (\zeta^{(i)}))$, where $S$ is the singular value thresholding.

**Update $W^{(i)}_D$ with other variables fixed.** The variable $W^{(i)}_D$ can be updated following:

$$\begin{align*}
\min_{W^{(i)}_D} & \sum_{i=1}^V \mu^{(i)} \left\| Y - S(W^{(i)} + W^{(i)}_D)X^{(i)} \right\|_F^2 + \eta \sum_{i=1}^V \left\| W^{(i)}_D \right\|_F^2 \\
+ & \beta \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} Tr \left( W^{(i)}_D X^{(i)} \cdot (W^{(j)}_D X^{(j)})^T \right) \\
+ & \sum_{i=1}^V Tr \left( M^{(i)} - (A^{(i)} - SW^{(i)} C X^{(i)}) \right).
\end{align*}$$

(9)

We take the derivative of (9) with respect to $W^{(i)}_D$ to 0. Afterwards, based on KKT conditions, we can easily update $W^{(i)}_D$ in an iterative manner [Tan et al., 2021].

**Update $A$ with other variables fixed.** The optimization subproblem with regard to $A^{(i)}$ can be reformulated as:

$$\begin{align*}
\min_{A^{(i)}} & \sum_{i=1}^V \frac{1}{2} \left\| A^{(i)} - SW^{(i)} C X^{(i)} + \frac{1}{\lambda^{(i)}} M^{(i)} \right\|_F^2 \\
- & \alpha \sum_{i,j=1}^V \mu^{(i)} \mu^{(j)} \mathcal{H}SI(C(A^{(i)}, A^{(j)})).
\end{align*}$$

(10)

Here, $\mathcal{H}SI(C(A^{(i)}, A^{(j)})) = (n-1)^{-2} Tr (HK^{(i)} HK^{(j)})$, $K^{(i)} = A^{(i)} A^{(i)}$ is the Gram matrix and $H$ centers it to
have zero mean. Theorem 1 (in section 5.3) guarantees the subproblem (10) to be convex and the optimal solution could be obtained by setting its derivative with respect to $A^{(i)}$ to 0.

**Update $S$ with other variables fixed.** The variable $S$ can be updated by solving the following sub-problem:

$$
\begin{align*}
\min_S & \sum_{i=1}^{V} \mu^{(i)} ||Y - S(W_C^{(i)} + W_D^{(i)})X^{(i)}||^2_F \\
+ & \sum_{i=1}^{V} \frac{\lambda^{(i)}}{2} ||A^{(i)} - SW_C^{(i)}X^{(i)}||^2_F \\
+ & Tr\left(M^{(i)\top}\left(A^{(i)} - SW_C^{(i)}X^{(i)}\right)\right) \\
\end{align*}
$$

(11)

Similar to (9), we also take the derivative of (11) with respect to $S$ to 0, and then we can obtain its closed-form solution.

**Update $M^{(i)}$ and $\lambda^{(i)}$ with other variables fixed.** Finally, we update the Lagrange multiplier matrices $M^{(i)}$ and penalty scalars $\lambda^{(i)}$ following:

$$
M^{(i)\infty+1} = M^{(i)\infty} + \lambda^{(i)\infty} (A^{(i)\top} - SW_C^{(i)\top}X^{(i)\top})
$$

$$
\lambda^{(i)\infty+1} = \min(\lambda_{\max}, \tau \lambda^{(i)\infty})
$$

(12)

During the process of model training, we first initialize the required variables, and then repeat the above steps until the algorithm converges or reaches the maximum iterations. Finally, we make prediction for unseen instance following $y^* = \sum_{i=1}^{V} \mu^{(i)} S(W_C^{(i)} + W_D^{(i)})x_i^*$.

Algorithm 1 summarizes the whole procedure of our proposed CISF method.

## 4 Experiments

### 4.1 Experimental Setting

To evaluate the performance of our proposed CISF method, we implement experiments on seven widely-used MVML data sets, including Emotions, Scene, Corel5k, Espgame, Pascal, Iaprtc12 and Mirflickr data sets. Table 1 summarizes the detailed characteristics of the above data sets.

<table>
<thead>
<tr>
<th>Data sets</th>
<th>Instances</th>
<th>Views</th>
<th>$D_{\text{min}-\text{max}}$</th>
<th>Labels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Emotions</td>
<td>593</td>
<td>2</td>
<td>8.64</td>
<td>6</td>
</tr>
<tr>
<td>Scene</td>
<td>2407</td>
<td>2</td>
<td>98.196</td>
<td>6</td>
</tr>
<tr>
<td>Corel5k</td>
<td>4999</td>
<td>4</td>
<td>100.3895</td>
<td>260</td>
</tr>
<tr>
<td>Pascal</td>
<td>9963</td>
<td>5</td>
<td>512.4086</td>
<td>20</td>
</tr>
<tr>
<td>Iaprtc12</td>
<td>19627</td>
<td>6</td>
<td>100.3985</td>
<td>291</td>
</tr>
<tr>
<td>Espgame</td>
<td>20770</td>
<td>4</td>
<td>100.4096</td>
<td>268</td>
</tr>
<tr>
<td>Mirflickr</td>
<td>25000</td>
<td>5</td>
<td>100.4096</td>
<td>457</td>
</tr>
</tbody>
</table>

Table 1: Characteristics of our employed data sets. $D_{\text{min}-\text{max}}$ is the smallest-largest dimensions of features.

Meanwhile, we compare our proposed CISF with the following five state-of-the-art MVML methods, including LSPC [Szymanski et al., 2016], FIMAN [Wu et al., 2020], ICM2L [Tan et al., 2021], BEMVL [Lu et al., 2023] and NAIM3L [Li and Chen, 2022]. The configured parameters of the above methods are set according to the suggestions in their corresponding literature.

In addition, five popular multi-label evaluation metrics are employed to measure the performance of each comparing method, including Hamming Loss (H-L), Ranking Loss (R-L), One Error (O-E), Coverage (COV) and Average Precision (A-P) [Zhang and Zhou, 2013]. For each dataset, we randomly select 70% examples for training, 10% examples for parameter tuning and 20% examples for evaluation, where each algorithm is run 5 times independently. The codes and data sets are provided in https://gengyulyu.github.io/homepage/.

### 4.2 Experimental Results

Table 2 illustrates the experimental comparisons between our proposed CISF and other five comparing methods on all evaluation metrics, where the average metrics results and standard deviations are recorded respectively. According to Table 2, out of 210 (7 data sets × 6 methods × 5 metrics) statistical comparisons can make the following observations:

- Among all five comparing methods, our proposed CISF method is superior to LSPC, FIMAN and ICM2L in almost all cases, and it also outperforms BEMVL and NAIM3L in 91.42% and 88.57% cases, respectively.

- Among all employed evaluation metrics, our proposed CISF achieves the best performance in 97.14% cases on Hamming Loss, Ranking Loss and Average Precision metrics. And on One Error and Coverage metrics, it is also superior to other methods over 94% cases.

- Among all employed datasets, CISF outperforms almost all comparing methods on Emotions, Scene, Pascal, Iaprtc12 and Mirflickr datasets. And it also achieve superior performance against other comparing methods over 82% cases on Corel5k and Espgame data sets.

- Overall, our proposed CISF method can achieve competitive performance against other feature-fusion based MVML methods, which demonstrates the effective of our proposed multi-view semantic-fusion strategy.

In order to comprehensively evaluate the superiority of CISF, Friedman test [Demšar, 2006] is utilized as the statistical test to analyze the relative performance among the comparing algorithms. According to Table 3, the null hypothesis of distinguishable performance among the comparing algorithms is rejected at 0.05 significance level. Thus, we further employ the post-hoc Bonferroni-Dunn test [Demšar, 2006] to show the relative performance among the comparing algorithms. Figure 3 illustrates the CD diagrams on each evaluation metric, where the average rank of each algorithm is marked along the axis. According to Figure 3, it is observed that CISF always ranks 1st on all evaluation metrics.

## 5 Further Analysis

### 5.1 Ablation Study

In order to evaluate the effect of the each components of our proposed CISF, we conduct the Ablation Study between CISF and its three degenerated algorithms CISFnl, CISFn1 and CISFnL, where each degenerated algorithm ignores the common semantics, individual semantics and label correlations,
Table 2: Experimental comparisons of our proposed CISF with other comparing methods on six evaluation metrics, where the best performances on each metric are shown in bold face. “-” indicates that FIMAN needs over 128G of RAM on Mirflickr data set.

Table: |
<table>
<thead>
<tr>
<th></th>
<th>Emotions</th>
<th>Scene</th>
<th>Corel5k</th>
<th>Pascal</th>
<th>Iaprtc12</th>
<th>Espgame</th>
<th>Mirflickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISF</td>
<td>0.203 ± 0.020</td>
<td>0.169 ± 0.003</td>
<td>0.029 ± 0.000</td>
<td>0.073 ± 0.000</td>
<td>0.019 ± 0.000</td>
<td>0.017 ± 0.000</td>
<td>0.005 ± 0.000</td>
</tr>
<tr>
<td>LSPC</td>
<td>0.263 ± 0.017</td>
<td>0.253 ± 0.011</td>
<td>0.028 ± 0.001</td>
<td>0.239 ± 0.005</td>
<td>0.036 ± 0.000</td>
<td>0.029 ± 0.000</td>
<td>0.018 ± 0.002</td>
</tr>
<tr>
<td>FIMAN</td>
<td>0.240 ± 0.015</td>
<td>0.313 ± 0.009</td>
<td>0.019 ± 0.000</td>
<td>0.123 ± 0.000</td>
<td>0.035 ± 0.000</td>
<td>0.036 ± 0.000</td>
<td>-</td>
</tr>
<tr>
<td>ICML2</td>
<td>0.306 ± 0.026</td>
<td>0.279 ± 0.016</td>
<td>0.055 ± 0.005</td>
<td>0.153 ± 0.004</td>
<td>0.054 ± 0.000</td>
<td>0.020 ± 0.000</td>
<td>0.010 ± 0.000</td>
</tr>
<tr>
<td>BEMVL</td>
<td>0.386 ± 0.019</td>
<td>0.219 ± 0.004</td>
<td>0.025 ± 0.000</td>
<td>0.130 ± 0.008</td>
<td>0.031 ± 0.000</td>
<td>0.028 ± 0.000</td>
<td>0.013 ± 0.000</td>
</tr>
<tr>
<td>NAIM3L</td>
<td>0.242 ± 0.044</td>
<td>0.178 ± 0.021</td>
<td>0.013 ± 0.000</td>
<td>0.087 ± 0.007</td>
<td>0.019 ± 0.000</td>
<td>0.017 ± 0.000</td>
<td>0.006 ± 0.000</td>
</tr>
</tbody>
</table>

Table: |
<table>
<thead>
<tr>
<th></th>
<th>Emotions</th>
<th>Scene</th>
<th>Corel5k</th>
<th>Pascal</th>
<th>Iaprtc12</th>
<th>Espgame</th>
<th>Mirflickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISF</td>
<td>0.164 ± 0.010</td>
<td>0.115 ± 0.008</td>
<td>0.081 ± 0.000</td>
<td>0.123 ± 0.008</td>
<td>0.104 ± 0.001</td>
<td>0.184 ± 0.006</td>
<td>0.218 ± 0.014</td>
</tr>
<tr>
<td>LSPC</td>
<td>0.199 ± 0.026</td>
<td>0.251 ± 0.022</td>
<td>0.889 ± 0.015</td>
<td>0.881 ± 0.004</td>
<td>0.995 ± 0.001</td>
<td>0.992 ± 0.000</td>
<td>0.721 ± 0.008</td>
</tr>
<tr>
<td>FIMAN</td>
<td>0.185 ± 0.007</td>
<td>0.241 ± 0.013</td>
<td>0.141 ± 0.004</td>
<td>0.149 ± 0.002</td>
<td>0.138 ± 0.003</td>
<td>0.186 ± 0.004</td>
<td>-</td>
</tr>
<tr>
<td>ICML2</td>
<td>0.330 ± 0.048</td>
<td>0.349 ± 0.067</td>
<td>0.115 ± 0.064</td>
<td>0.216 ± 0.033</td>
<td>0.179 ± 0.004</td>
<td>0.204 ± 0.005</td>
<td>0.284 ± 0.002</td>
</tr>
<tr>
<td>BEMVL</td>
<td>0.408 ± 0.036</td>
<td>0.126 ± 0.008</td>
<td>0.394 ± 0.012</td>
<td>0.355 ± 0.062</td>
<td>0.272 ± 0.006</td>
<td>0.281 ± 0.004</td>
<td>0.469 ± 0.036</td>
</tr>
<tr>
<td>NAIM3L</td>
<td>0.244 ± 0.090</td>
<td>0.173 ± 0.052</td>
<td>0.113 ± 0.008</td>
<td>0.180 ± 0.037</td>
<td>0.144 ± 0.011</td>
<td>0.174 ± 0.010</td>
<td>0.266 ± 0.006</td>
</tr>
</tbody>
</table>

Table: |
<table>
<thead>
<tr>
<th></th>
<th>Emotions</th>
<th>Scene</th>
<th>Corel5k</th>
<th>Pascal</th>
<th>Iaprtc12</th>
<th>Espgame</th>
<th>Mirflickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISF</td>
<td>0.253 ± 0.018</td>
<td>0.326 ± 0.018</td>
<td>0.591 ± 0.252</td>
<td>0.450 ± 0.020</td>
<td>0.495 ± 0.005</td>
<td>0.549 ± 0.023</td>
<td>0.837 ± 0.044</td>
</tr>
<tr>
<td>LSPC</td>
<td>0.308 ± 0.037</td>
<td>0.429 ± 0.033</td>
<td>0.912 ± 0.018</td>
<td>0.936 ± 0.009</td>
<td>0.992 ± 0.001</td>
<td>0.989 ± 0.003</td>
<td>0.944 ± 0.004</td>
</tr>
<tr>
<td>FIMAN</td>
<td>0.279 ± 0.013</td>
<td>0.547 ± 0.028</td>
<td>0.602 ± 0.020</td>
<td>0.468 ± 0.005</td>
<td>0.559 ± 0.005</td>
<td>0.659 ± 0.008</td>
<td>-</td>
</tr>
<tr>
<td>ICML2</td>
<td>0.439 ± 0.064</td>
<td>0.665 ± 0.061</td>
<td>0.746 ± 0.143</td>
<td>0.585 ± 0.011</td>
<td>0.652 ± 0.022</td>
<td>0.709 ± 0.029</td>
<td>0.873 ± 0.007</td>
</tr>
<tr>
<td>BEMVL</td>
<td>0.334 ± 0.072</td>
<td>0.151 ± 0.011</td>
<td>0.602 ± 0.018</td>
<td>0.599 ± 0.061</td>
<td>0.272 ± 0.007</td>
<td>0.867 ± 0.032</td>
<td>0.873 ± 0.007</td>
</tr>
<tr>
<td>NAIM3L</td>
<td>0.500 ± 0.471</td>
<td>0.333 ± 0.235</td>
<td>0.842 ± 0.036</td>
<td>0.525 ± 0.244</td>
<td>0.829 ± 0.039</td>
<td>0.873 ± 0.054</td>
<td>0.978 ± 0.010</td>
</tr>
</tbody>
</table>

Table: |
<table>
<thead>
<tr>
<th></th>
<th>Emotions</th>
<th>Scene</th>
<th>Corel5k</th>
<th>Pascal</th>
<th>Iaprtc12</th>
<th>Espgame</th>
<th>Mirflickr</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISF</td>
<td>1.791 ± 0.118</td>
<td>0.724 ± 0.053</td>
<td>101.8 ± 4.736</td>
<td>3.100 ± 0.155</td>
<td>93.44 ± 0.995</td>
<td>120.2 ± 1.426</td>
<td>131.5 ± 9.822</td>
</tr>
<tr>
<td>LSPC</td>
<td>2.182 ± 0.129</td>
<td>1.398 ± 0.117</td>
<td>279.2 ± 9.160</td>
<td>19.57 ± 0.089</td>
<td>286.1 ± 0.036</td>
<td>281.3 ± 0.109</td>
<td>331.3 ± 2.831</td>
</tr>
<tr>
<td>FIMAN</td>
<td>1.948 ± 0.123</td>
<td>1.298 ± 0.076</td>
<td>83.26 ± 1.928</td>
<td>4.069 ± 0.059</td>
<td>116.2 ± 1.673</td>
<td>113.5 ± 1.892</td>
<td>-</td>
</tr>
<tr>
<td>ICML2</td>
<td>2.700 ± 0.177</td>
<td>1.852 ± 0.328</td>
<td>113.96 ± 6.985</td>
<td>5.625 ± 0.790</td>
<td>139.7 ± 2.300</td>
<td>128.6 ± 2.223</td>
<td>226.1 ± 1.326</td>
</tr>
<tr>
<td>BEMVL</td>
<td>2.944 ± 0.197</td>
<td>0.777 ± 0.046</td>
<td>185.9 ± 3.660</td>
<td>8.385 ± 1.268</td>
<td>193.6 ± 2.313</td>
<td>167.6 ± 1.329</td>
<td>320.5 ± 15.61</td>
</tr>
<tr>
<td>NAIM3L</td>
<td>2.173 ± 0.389</td>
<td>0.950 ± 0.270</td>
<td>65.59 ± 4.159</td>
<td>4.650 ± 0.828</td>
<td>112.7 ± 6.610</td>
<td>110.9 ± 6.533</td>
<td>154.9 ± 3.200</td>
</tr>
</tbody>
</table>

Figure 3: Experimental Comparisons of all comparing algorithms with the Bonferroni-Dunn test (CD = 2.576 at 0.05 significance level).
Figure 4: The parameter analysis of CISF on Emotions data set, where the Coverage results are normalized by the number of class labels (q). (Red: Hamming Loss, Brown: Ranking Loss, Pink: One Error, Green: Coverage, Blue: Average Precision.)

<table>
<thead>
<tr>
<th>Evaluation Metric</th>
<th>( \tau_F )</th>
<th>critical value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hamming Loss</td>
<td>5.536</td>
<td>2.533</td>
</tr>
<tr>
<td>Ranking Loss</td>
<td>10.705</td>
<td></td>
</tr>
<tr>
<td>One Error</td>
<td>6.927</td>
<td></td>
</tr>
<tr>
<td>Coverage</td>
<td>10.924</td>
<td>Methods: 6, Data sets: 7</td>
</tr>
<tr>
<td>Average Precision</td>
<td>7.125</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Friedman statics \( \tau_F \) in terms of each evaluation metric.

<table>
<thead>
<tr>
<th>Methods</th>
<th>H-L</th>
<th>R-L</th>
<th>O-E</th>
<th>COV</th>
<th>A-P</th>
</tr>
</thead>
<tbody>
<tr>
<td>CISFnC</td>
<td>0.296</td>
<td>0.174</td>
<td>0.293</td>
<td>1.905</td>
<td>0.778</td>
</tr>
<tr>
<td>CISFnI</td>
<td>0.264</td>
<td>0.172</td>
<td>0.286</td>
<td>1.895</td>
<td>0.789</td>
</tr>
<tr>
<td>CISFnL</td>
<td>0.208</td>
<td>0.175</td>
<td>0.290</td>
<td>1.852</td>
<td>0.785</td>
</tr>
<tr>
<td>CISF</td>
<td><strong>0.203</strong></td>
<td><strong>0.164</strong></td>
<td><strong>0.253</strong></td>
<td><strong>1.791</strong></td>
<td><strong>0.805</strong></td>
</tr>
</tbody>
</table>

Table 4: The ablation study of CISF on Emotions data set.

respectively. Table 4 reports the experimental comparison between these methods on Emotions data set. According to Table 4, we can find that CISFnI achieves better performance than CISFnC method, which shows that common semantics has greater contribution than individual semantics to the effectiveness of learning model. Besides, our proposed CISF is superior to both CISFnC and CISFnI methods, which demonstrates that the common semantics and individual semantics can jointly improve the performance of MVML model.

5.2 Parameter Sensitivity

We study the sensitivity analysis of our proposed CISF with respect to its four employed parameters \( \alpha, \beta, \gamma \) and \( \eta \). Figure 4 shows the performance of CISF under different parameter configurations on Emotions data set. According to Figure 4, we can find that \( \alpha \) and \( \gamma \) usually have great influence on the performance of the proposed model, and we select the optimal values of them from \( \{10^{-3}, 10^{-2}, \ldots, 10^{2}\} \) and \( \{0.01, 0.05, \ldots, 10\} \), respectively. Meanwhile, other parameters often follow the optimal configurations \( \beta = 0.1 \) and \( \eta = 100 \) but vary with minor adjustments on different data sets. In addition, in our experiments, the value of \( \lambda_{max} \) is set to \( 1e^{6} \) and the maximum iterations \( I_{max} \) is set to 50.

5.3 Complexity Analysis

At each iteration, the computational cost mainly comes from the derivative calculation and singular value decomposition (SVD) operations, and the complexity of our optimization procedure mainly comes from the optimization of three sub-problems with regard to \( W_{C}^{(i)}, W_{D}^{(i)} \) and \( A^{(i)} \). For simplicity, we suppose the dimensionality of each view is \( d \). And the complexity of these sub-problems are \( O(dnr) \), \( O(Tnqd) \) and \( O(q^{2}d + qdn + q^{3}) \) respectively, where \( r \) is the rank of \( W_{C}^{(i)} \) and \( T \) is the iteration number for updating \( W_{D}^{(i)} \).

5.4 Convergence Analysis

The convergence of the whole optimization problem (6) depends on how to guarantee the subproblem (10) is convex, especially its HSIC term is negative. Theorem 1 provides the theoretical guarantee for the convexity of (10) under proper parameter setting. Besides, Figure 5 illustrates the convergence curves on Emotions and Scene data sets, which also empirically demonstrates the convergence of our model.

**Theorem 1:** The problem (10) is convex given the parameter \( \lambda^{(i)} \geq 8q(V - 1)\alpha_{i}\mu_{i}\mu_{j} \), where \( V \) is the number of views.

**Proof:** Due to the page limitation, we provided the proof of Theorem 1 in https://gengyulyu.github.io/homepage/.

6 Conclusion

In this paper, we proposed a Common-Individual Semantic Fusion Multi-View Multi-Label Learning Method. Different from previous feature-fusion based MVML methods, it is the first attempt to conduct multi-view fusion under the guidance of the semantic fusion, where both common semantics and individual semantics are simultaneously incorporated into the multi-view fusion process to learn a desired multi-label classification model. Extensive experimental results on various MVML datasets has demonstrated the effectiveness of our proposed multi-view semantic-fusion strategy.
Acknowledgments
This work was supported by the National Key Research and Development Program of China (No. 2023YFB3107100), the National Natural Science Foundation of China (No. 62300620), the China Postdoctoral Science Foundation (No. 2022M720320), the Beijing Postdoctoral Science Foundation (No. 2023-zz-78), the Chaoyang Postdoctoral Science Foundation, the Fundamental Research Funds for the Central universities (No. 2022JBY019), the Beijing Natural Science Foundation (No. 4242046), the Major Research Plan of National Natural Science Foundation of China (No. 92167102).

References
[Nie et al., 2016] F. Nie, J. Li, and X. Li. Parameter-free auto-weighted multiple graph learning: a framework for


