Partial Multi-Label Learning via Multi-Subspace Representation[1]

Student: Guangfei Liang

Supervisor: Can Gao

Shenzhen University

December 11, 2022

Catalog

1. Background
2. MUSER
3. Modification of MUSER (mMUSER)
4. Experiment

Catalog

1. Background
2. MUSER
3. Modification of MUSER (mMUSER)
4. Experiment

Background

- Partial Multi-Label Learning (PML) is to learn the precise labels from the samples with redundant labels
- There are 6 ground-truth labels and 3 redundant labels in Figure.1

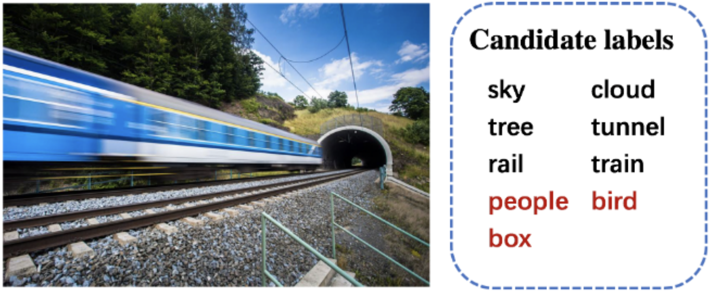


Figure: An example of PML

Background

- Existing PML methods can be divided into two groups:
  - Unified strategy: PML-fp, PML-lc, fPML, and PML-LRS.
  - Two-stage strategy: PARTICLE and DRAMA.
- Existing PML methods mainly focus on the noise in label space while the noise in feature space is ignored.

Catalog

1 Background

2 MUSER

3 Modification of MUSER (mMUSER)

4 Experiment

MUSER

The ground truth matrix $\widetilde{Y} \in \{0,1\}^{n \times q}$ is decomposed by a low-dimensional label subspace $U \in \mathbb{R}^{n \times c}$ and the label correlation matrix $P \in \mathbb{R}^{c \times q}$.

$$\widetilde{Y} \backsimeq UP \tag{1}$$

we minimize the reconstruction error between the candidate label matrix Y and the product of U and P as follows:

$$\min_{U,P} \frac{1}{2} \|Y - UP\|_F^2 + \mathcal{R}(U, P) \tag{2}$$

where $\mathcal{R}(U, P)$ denotes the regularization to control the model complexity.

MUSER

A graph Laplacian regularization is introduced to ensure such consistency between features and latent labels.
Definite pairwise similarity matrix $S \in \mathbb{R}^{n \times n}$:

$$S_{ij} = \begin{cases} exp(-||x_i - x_j||_2^2/\sigma^2) & ,i \ and \ j \ are \ k - nearest \ neighbours \\ 0 & ,otherwise \end{cases} \tag{3}$$

Then the graph regularization term is

$$\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} S_{ij}||\frac{u_i}{\sqrt{E_{ii}}} - \frac{u_j}{\sqrt{E_{jj}}}||_2^2 = Tr(U^T LU) \tag{4}$$

where $L = E^{-\frac{1}{2}}(E - S)E^{-\frac{1}{2}}$ is a graph Laplacian matrix and E is a diagonal matrix with $E_{ii} = \sum_{j=1}^{n} S_{ij}$.

MUSER

- In the real-world application, feature information is often corrupted by outliers and noise.
- We use a feature correlation matrix $Q \in \mathbb{R}^{m \times c}$ is introduced to map the original feature space to a low-dimensional feature subspace.
- The formulation of MUSER is

$$\begin{aligned} \min_{W,Q,U,P} & \frac{1}{2}||U - X^T QW||_F^2 + \frac{\alpha}{2}||Y - UP||_F^2 + \frac{\beta}{2}Tr(U^T LU) + \mathcal{R}(W,U,P) \\ s.t. \quad & Q^T Q = I \end{aligned} \tag{5}$$

  where $\mathcal{R}(W,U,P) = \frac{\gamma}{2}(||W||_F^2 + ||U||_F^2 + ||P||_F^2)$
- The prediction function is $\hat{Y} = X^{*T} QWP$

Optimization of MUSER

Step 1: Calculate P. With U, Q, W fixed, Eq.(5) can be reduced to:

$$\min_{p} \frac{\alpha}{2}||Y - UP||_F^2 + \frac{\gamma}{2}||P||_F^2 \tag{6}$$

and we can get the closed form solution:

$$P = (\alpha U^T U + \gamma I)^{-1} \alpha U^T Y \tag{7}$$

Step 2: Calculate U. With P, Q, W fixed, Eq.(5) can be reduced to:

$$\min_{U} \frac{1}{2}||U - X^T QW||_F^2 + \frac{\alpha}{2}||Y - UP||_F^2 + \frac{\beta}{2} Tr(U^T LU) + \frac{\gamma}{2}||U||_F^2 \tag{8}$$

Use the standard gradient descent algorithm to optimize U:

$$U := U - \lambda_U \nabla_U \tag{9}$$

where λ_U is the stepsize of gradient descent and

$$\nabla_U = (1 + \gamma)U + \beta LU + \alpha UPP^T - \alpha YP^T - X^T QW \tag{10}$$

Optimization of MUSER

Step 3: Calculate Q. With P, U, W fixed, Eq.(5) can be reduced to:

$$\min_{Q} \frac{1}{2}||U - X^T QW||_F^2 \\ s.t. Q^T Q = I \tag{11}$$

Similarity to **Step 2**, we can get Q as follows:

$$Q := Q - \lambda_Q(-XUW^T + XX^T QWW^T) \tag{12}$$

To satisfy the constraint $Q^T Q = I$, we map each row of Q onto the unit norm ball after each iteration:

$$Q_{i,:} \leftarrow \frac{Q_{i,:}}{||Q_{i,:}||} \tag{13}$$

where $Q_{i,:}$ is the i-th row of Q.

This could NOT ensure the orthogonal constraint actually!!!

Optimization of MUSERv

Step 4: Calculate W. With P, U, Q fixed, Eq.(5) can be reduced to:

$$\min_{W} \frac{1}{2}||U - X^T QW||_F^2 + \frac{\gamma}{2}||W||_F^2 \tag{14}$$

and we can get the closed form solution:

$$W = (Q^T X X^T Q + \gamma I)^{-1} Q^T X U \tag{15}$$

MUSER

Algorithm MUSER

Require: $X \in \mathbb{R}^d, Y \in \{0,1\}^{n \times q}, \alpha, \beta, \gamma, T_{max}$.
Ensure: W, Q, U, P.
  Initialize W, Q, U, P randomly, $t = 1$, $convergence = false$.
  while $t < T_{max}$ or $!convergence$ **do**
    Use Eq.(7) to update P
    Use Eq.(9) to update U
    Use Eq.(12) to update Q
    Use Eq.(15) to update W
    $t = t + 1$
    if objective function (5) is convergenced **then**
      $convergence = true$
    end if
  end while

Catalog

1. Background
2. MUSER
3. Modification of MUSER (mMUSER)
4. Experiment

mMUSER

The regularization of W may be redundant and degrade the performance of building the correlation between latent label subspace and latent feature subspace. Therefore, we modify the model (5) as

$$\min_{W,Q,U,P} \frac{1}{2}||U - X^T QW||_F^2 + \frac{\alpha}{2}||Y - UP||_F^2 + \frac{\beta}{2}Tr(U^T LU) + \mathcal{R}(U, P)$$
$$s.t. \quad Q^T Q = I \tag{16}$$

Optimization of mMUSER

Step 1: Calculate P**.** and **Step 2: Calculate** U**.** will be the same as optimization of MUSER in previous analysis.
Step 3: Calculate W**.** With P, U, Q fixed, Eq.(16) can be reduced to:

$$\min_{W} \frac{1}{2}||U - X^T QW||_F^2 \tag{17}$$

and we can get the closed form solution:

$$W = (Q^T X X^T Q)^{-1} Q^T X U \tag{18}$$

Optimization of mMUSER

Step 4: Calculate Q. With P, U, W fixed, Eq.(16) can be reduced to:

$$\begin{aligned} &\min_{Q} \frac{1}{2} ||U - X^T QW||_F^2 \\ &s.t. Q^T Q = I \end{aligned} \tag{19}$$

By substitute Eq.(18) into Eq.(19), the optimization problem of Q is transferred into

$$\begin{aligned} &\max_{Q} Tr[(Q^T XX^T Q)^{-1} Q^T XUU^T X^T Q] \\ &s.t. Q^T Q = I \end{aligned} \tag{20}$$

It is easy to see that the problem (20) is an **orthogonal LDA-like problem**, where $S_t = XX^T$ is the total scatter matrix and $S_w = XUU^T X^T$ is the within-class.

mMUSER

Algorithm mMUSER

Require: $X \in \mathbb{R}^d, Y \in \{0,1\}^{n \times q}, \alpha, \beta, \gamma, T_{max}$.
Ensure: W, Q, U, P.

Initialize W, Q, U, P randomly, $t = 1$, $convergence = false$.
while $t < T_{max}$ or $!convergence$ **do**
  Use Eq.(7) to update P
  Use Eq.(9) to update U
  Use Eq.(18) to update W
  Use Eq.(20) to update Q
  $t = t + 1$
  if objective function (5) is convergenced **then**
    $convergence = true$
  end if
end while

Catalog

1. Background
2. MUSER
3. Modification of MUSER (mMUSER)
4. Experiment

Experiment

Table: Comparison of MUSER(from original paper's experiment data), MUSER(from) and mMUSER on bibtex dataset with five evaluation metrics, where the direction of arrow points to represents the better and the best performances are shown in bold face.

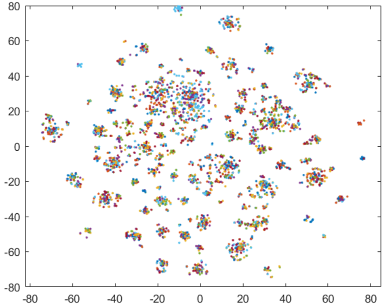
$r = 1$ (one redundant label for each instance)					
	Hamming↓	**Ranking↓**	**One-Error↓**	**Coverage↓**	**Average Precision↑**
MUSER(O)	**0.0090**	0.1230	0.3770	**0.2310**	0.5500
MUSER	0.9849	**0.0492**	**0.3262**	14.8798	0.3452
mMUSER	0.9849	0.3025	0.7812	66.4154	**0.6683**

Experiment

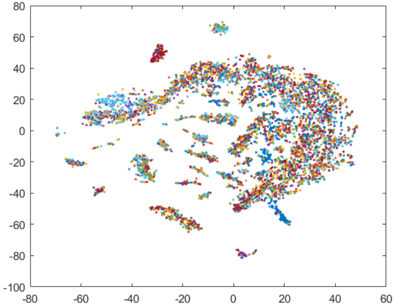
Table: Comparison of MUSER and mMUSER on BlogCatalog dataset with five evaluation metrics, where the direction of arrow points to represents the better and the best performances are shown in bold face.

$r = 1$ (one redundant label for each instance)					
	Hamming↓	**Ranking**↓	**One-Error**↓	**Coverage**↓	**Average Precision**↑
MUSER	0.2591	**0.0672**	**0.1673**	**4.1971**	**0.1568**
mMUSER	0.2591	0.0682	0.2132	4.2494	0.1473
$r = 2$ (one redundant label for each instance)					
	Hamming↓	**Ranking**↓	**One-Error**↓	**Coverage**↓	**Average Precision**↑
MUSER	0.2591	**0.0672**	0.1804	4.2266	0.1518
mMUSER	0.2591	0.0681	**0.1591**	**4.2249**	**0.1732**
$r = 3$ (one redundant label for each instance)					
	Hamming↓	**Ranking**↓	**One-Error**↓	**Coverage**↓	**Average Precision**↑
MUSER	0.2591	**0.0685**	0.1912	4.2889	**0.1390**
mMUSER	0.2591	0.0687	**0.1774**	**4.2850**	0.1251

Experiment



(a) MUSER



(b) mMUSER

Figure: The t-sne dimensionality reduction visualization on bibtex dataset of (a) MUSER and (b) mMUSER

References

[1] Z. Li, G. Lyu, and S. Feng, "Partial multi-label learning via multi-subspace representation," in *International Joint Conference on Artificial Intelligence*, 2020.