

基于实例的多视角多标签学习算法

苏可政¹, 肖燕珊¹, 刘波²

¹广东工业大学, 计算机学院, 广东 广州

²广东工业大学, 自动化学院, 广东 广州

收稿日期: 2022年3月1日; 录用日期: 2022年3月30日; 发布日期: 2022年4月6日

摘要

在多视角多标签学习中, 一个实例可以由多个数据视角表示, 并与多个类标签相关联。目前现有的多视角多标签学习算法研究的是变量与类标签之间的关系。在这篇文章中, 我们提出了一种基于实例的多视角多标签学习算法。与现有的多视角多标签算法不同, 我们利用训练实例和测试实例的相关性进行建模。在多视角多标签学习中, 数据来自多个视角。在每个视角中, 通过构建偏最小二乘回归模型来探索训练实例和测试实例之间的相关性, 而不是像传统的多视角多标签学习算法一样探索变量和类标签之间的映射函数。因此, 我们可以从多视角数据中学习到多个偏最小二乘回归模型。此外, 为了保证不同视角的一致性, 在多个回归模型中加入视角一致性约束。对比实验表明, 与现有的多视角多标签算法相比, 所提算法获得了更好的分类结果。

关键词

偏最小二乘(PLS)回归, 多视角多标签学习, 视角一致性

Multi-View Multi-Label Learning by Exploiting Instance Correlations

Kezheng Su¹, Yanshan Xiao¹, Bo Liu²

¹Department of Computer, Guangdong University of Technology, Guangzhou Guangdong

²Department of Automation, Guangdong University of Technology, Guangzhou Guangdong

Received: Mar. 1st, 2022; accepted: Mar. 30th, 2022; published: Apr. 6th, 2022

Abstract

In multi-view multi-label learning, an instance can be represented by multiple data views, and associated with multiple class labels. The existing multi-view multi-label learning works explore the

relationship between the variables and the class labels. In this paper, we present an instance-based multi-view multi-label learning algorithm. Different from the existing multi-view multi-label algorithms, we model the correlations of training instances and testing instances. In multi-view multi-label learning, the data is from multiple views. In each view, we build a partial least square regression model to explore the correlation between the training instances and the test instances, rather than the mapping function between the variables and the class labels. As a result, we can learn multiple partial least square regression models from multi-view data. Moreover, in order to ensure the mutual agreement on distinct views, a view consensus constraint is proposed based on the multiple regression models. The comparison experiments show that our algorithm obtains better classification results in comparing with the state-of-the-art multi-view multi-label algorithms.

Keywords

Partial Least Square (PLS) Regression, Multi-View Multi-Label Learning, View Consensus

Copyright © 2022 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

在多标签学习中, 每个对象都与多个类标签相关联。这些类标签是相关的, 每个类标签通常与其他类标签相互依赖。多标签学习的目标是基于类标签的相关性构建一个分类模型[1]。近几年, 多标签学习在很多场景中得到了广泛的应用, 例如图像分类[2], 文本识别[3], 情绪分类[4]和社交网络挖掘[5]。然而, 在多标签学习中仍然存在两个开放性的挑战: 1) 如何利用相关性有效地建立模型; 2) 如何有效地处理多标签数据的高维问题。目前, 已经有许多算法被提出用来解决第一个挑战[6] [7] [8] [9] [10], 然而这些算法的复杂度相对较高[11]。对于第二个挑战, 有两种解决方案: 特征选择和降维。但是特征选择忽略了标签之间的相关性, 传统降维的结果可解释性也很差。为了解决这些问题, Liu 等人提出了一种基于实例的稀疏的加权多标签学习算法 SWIM [12]。SWIM 采用偏最小二乘(PLS)回归来研究训练集实例和测试集实例的相关性。SWIM 将实例映射到一个潜在空间中, 可以更有效地降维, 从而解决高维问题[13]。实验结果已经证明 SWIM 的分类效果明显高于对比的其他多标签学习算法。

尽管 SWIM 有更好的性能, 但是它是用于解决单视角多标签学习问题的, 不能直接用在多视角多标签学习数据上。而在现实中, 可以从多个不同的方面来描述一个对象。例如, 在 web 图像检索中, 一个实例可以用两个特征集来描述: 图像的视觉特征以及用于描述图像的文字特征。另外, 在图像分类中, 我们可以用多个不同的特征集合来描述一幅图像, 例如 RGB, SIFT 和 GIST [14]。在这些应用中, 一个对象与多个特征集合(即多个视角)相关联, 并且每个视角表示了同一对象的不同方面。因此, 一个视角可能包含了其他视角所没有的一些独特的分类信息[15] [16]。这表示, 相较于在每个视图中分别训练单独的分类器, 利用不同视角的一致性和互补的信息来建立模型会拥有更好的性能[17] [18]。SWIM 在每个数据视角中单独学习对应的多标签分类器, 所以只能用于解决单视角多标签学习。它不能利用不同视角的一致性和互补信息来提高分类器的性能。

综上所述, 我们提出了一种包含了多个视角的一致性和互补信息的模型, 用来提高多标签分类器的性能。具体而言, 考虑到每个对象都用多个视角表示, 我们采用 PLS 回归算法将多视角实例映射到多个

潜在空间中，从而可以在降维的同时构建多标签分类器，从而更好地解决高维问题。此外，为了保证不同视角的一致性，基于 PLS 回归模型提出了视角一致性约束，提高了分类器的性能。我们工作的主要贡献如下：

- 1) 我们首次提出了一种新颖的算法，基于实例的多视角多标签学习算法，该算法能够将多视角的一致性和互补性信息整合到多标签分类器中，以提高多标签学习的分类性能。
- 2) 将模型转化为 LASSO (Least absolute shrinkage and selection operator)问题，并采用 NIPALS (Non-linear iterative Partial Least Squares)求解相应的优化问题。
- 3) 现有的多视角多标签模型研究变量与标签之间的相关性，而我们的算法却是对训练实例和测试实例的相关性进行研究和建模。大量的实验表明，提出的算法相较于现有的算法具有更好的性能。

2. 基于实例的多视角多标签学习算法

2.1. 算法框架

给定一个有 n 个已标记样本的多视角多标签学习训练集 $\{(X^A, X^B, Y)\} = \{(x_i^A, x_i^B, y_i)\}_{i=1}^n$ ，其中 $x_i^A \in \mathbb{R}^{d_1}$ 和 $x_i^B \in \mathbb{R}^{d_2}$ 分别是视角 A 和视角 B 的特征向量。 x_i 的上标 A 和 B 指的是第 i 个实例 ($1 \leq i \leq n$) 的两个视角。在对视角多标签学习中，一个样本 x_i 有多个标签，因此有 $y_i = \{y_{i1}, \dots, y_{ij}, \dots, y_{iq}\}$ ，其中 $y_{ij} \in \{0, 1\}$ 为样本 x_i 的第 j 个标签。此外， $X^A = \{x_i^A\}_{i=1}^n$ 和 $X^B = \{x_i^B\}_{i=1}^n$ 分别是 $n \times d_1$ 和 $n \times d_2$ 的矩阵。 $Y = \{y_i\}_{i=1}^n$ 是一个 $n \times q$ 的矩阵，其中 y_i 是样本 x_i 的类标签。 $D^A = \{(X_i^A, Y_i) | 1 \leq i \leq n\}$ 和 $D^B = \{(X_i^B, Y_i) | 1 \leq i \leq n\}$ 分别是两视角多标签数据集的视角 A 和视角 B 数据集。让 $D^A = D_t^A \cup D_s^A$ 和 $D^B = D_t^B \cup D_s^B$ ，其中数据集 D 的下标 s 和 t 分别表示训练集和测试集。

我们提出的基于实例的多视角多标签学习算法的框架图如图 1 所示。定义训练集为 $D_t^{A/B}$ 和测试集为 $D_s^{A/B}$ ，且 $D_t^{A/B} \cap D_s^{A/B} = \emptyset$ 。目前的多视角多标签学习方法根据变量和类标签之间的函数关系构造模型 [15] [16] [17] [18] [19]。在多视角多标签数据集中，训练集 $D_t^{A/B}$ 和测试集 $D_s^{A/B}$ 的数据分布是相同的，这意味着，如果可以得到训练集 $D_t^{A/B}$ 的变量和类标签之间的映射关系 (即图 1 的 f_A 和 f_B)，那么映射函数 f_A 和 f_B 可以分别被直接应用在数据集 D_s^A 和 D_s^B 里。综上所述，如果 $f_{A/B}$ 是 $X_t^{A/B}$ 和 Y_t 之间的映射函数 (即， $Y_t = f_{A/B}(X_t^{A/B})$)，我们就可以利用它来得到 $X_s^{A/B}$ 的预测标签，即 $\hat{Y}_s = f_{A/B}(X_s^{A/B})$ 。

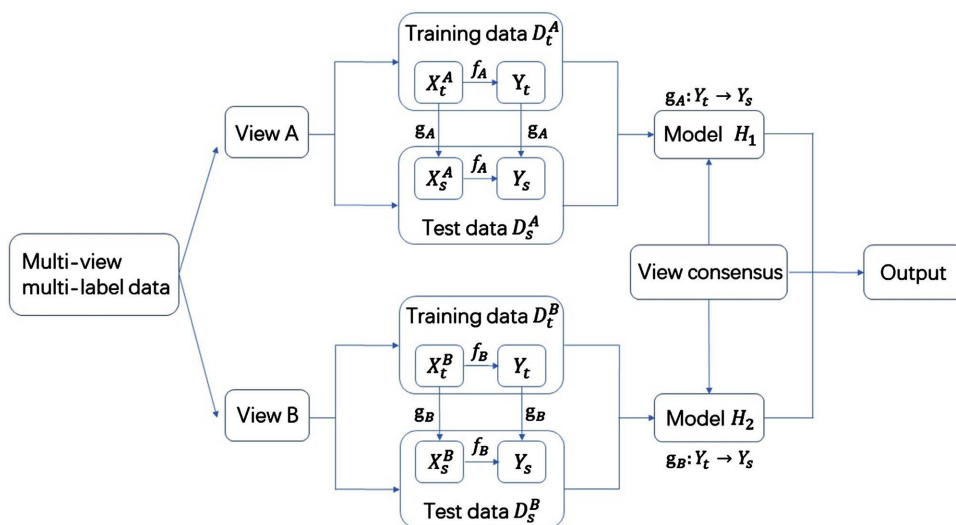


Figure 1. General diagram of algorithm
图 1. 算法框架图

与上述的多视角多标签不同，我们所提的算法利用了不同的映射方式来构造模型：确定训练示例 $X_t^{A/B}$ 和测试示例 $X_s^{A/B}$ 之间的函数关系 $g_{A/B}$ ，并通过 $g_{A/B}$ 来构造模型(即图 1 的 g_A 和 g_B)。该映射方法的工作原理如下：变量空间 $X^{A/B}$ 和类标签空间 Y 是同一个示例的输入空间和输出空间，它们具有相同的关联性，因此同一个数学定理既可以应用在变量空间 $X^{A/B}$ 上，也可以应用在类标签空间 Y 上。综上所述，当得到训练示例 $X_t^{A/B}$ 和测试示例 $X_s^{A/B}$ 之间的映射函数 $g_{A/B}$ 后(即 $\widehat{X_s^{A/B}} = g_{A/B}(X_t^{A/B})$)，映射函数也可以应用在 Y_t 和 Y_s 上，即： $\widehat{Y_s} = g_{A/B}(Y_t)$ 。因此，我们分别利用 g_A 和 g_B 为数据集 D_t^A 和 D_t^B 建立分类模型。

2.2. 构建模型

设 X^{A^T} 和 X^{B^T} 分别为 X^A 和 X^B 的转置矩阵， Y^T 为 Y 的转置矩阵。此外，设 X_t^A 和 X_s^A 分别为 $n_t \times d_1$ 和 $n_s \times d_1$ 的矩阵，其中 n_t 和 n_s 分别是训练集和测试集里示例的个数，而 d_1 是视角 A 里示例的维数。类似的， X_t^B 和 X_s^B 分别为 $n_t \times d_2$ 和 $n_s \times d_2$ 的矩阵，其中 d_2 是视角 B 里示例的维数。设 Y_t 和 Y_s 分别为 $n_t \times q$ 和 $n_s \times q$ 的矩阵，其中 q 是标签空间的维数。

$X_t^{A/B}$ 和 $X_s^{A/B}$ 拥有相同的数据分布，所以存在一个共同的潜在的实例空间，它们拥有相似的特征。因此，我们利用 PLS 来探索两个潜在空间，其中一个潜在空间来自于 X_t^A 和 X_s^A ，而另一个则来自于 X_t^B 和 X_s^B 。且 X_t^A ， X_s^A ， X_t^B 和 X_s^B 可以表示为如下形式：

$$X_t^{A^T} = T_A P_A^T + E_{tA} \tag{1}$$

$$X_t^{B^T} = T_B P_B^T + E_{tB} \tag{2}$$

$$X_s^{A^T} = T_A Q_A^T + E_{sA} \tag{3}$$

$$X_s^{B^T} = T_B Q_B^T + E_{sB} \tag{4}$$

其中 $T_A \in R^{d_1 \times m_1}$ 和 $T_B \in R^{d_2 \times m_2}$ 分别是 X_t^A 和 X_t^B 提取的 m_1 和 m_2 的潜在向量。加载矩阵 $P_A \in R^{n_t \times m_1}$ 和 $P_B \in R^{n_t \times m_2}$ 分别为 X_t^A 和 X_t^B 的潜在向量的对应系数。加载矩阵 $Q_A \in R^{n_s \times m_1}$ 和 $Q_B \in R^{n_s \times m_2}$ 则分别为 X_s^A 和 X_s^B 的潜在向量的系数。 E_{tA} ， E_{tB} ， E_{sA} 和 E_{sB} 则是随机误差。 T_A 和 T_B 可以表示为等价的线性变换形式，如下所示：

$$T_A = X_t^{A^T} W_A \tag{5}$$

$$T_B = X_t^{B^T} W_B \tag{6}$$

其中 $W_A \in R^{n_t \times m_1}$ 和 $W_B \in R^{n_t \times m_2}$ 是权重矩阵。这表示 X_t^A 和 X_t^B 可以分别由嵌套在 T_A 和 T_B 中 m_1 和 m_2 个独立的方向向量来表示。一旦得到 T_A 和 T_B ，我们就可以利用最小二乘回归方法根据公式(1)~(6)来计算 Q_A^T 和 Q_B^T ：

$$Q_A^T = (T_A^T T_A)^{-1} T_A^T X_s^{A^T} \tag{7}$$

$$Q_B^T = (T_B^T T_B)^{-1} T_B^T X_s^{B^T} \tag{8}$$

$X^{A/B}$ 和 Y 是相同示例的变量空间和类标签空间，所以它们拥有相同的属性。我们采用 PLS 判别分析来获得潜在成分 $T_{A/B}$ ，此时变量空间 $X_s^{A/B}$ 的投影比原始投影的误差更小。这里使用 PLS 是为了研究 $X_t^{A/B}$ 隐藏在 $X_s^{A/B}$ 的主要编码信息，即

$$\begin{aligned} & \max_{T_A, T_B} \text{cov}(T_A, X_s^{A^T}) + \text{cov}(T_B, X_s^{B^T}) \\ & \text{s.t. } T_A^T T_A = I, T_B^T T_B = I \end{aligned} \tag{9}$$

其中 $\text{cov}(\dots)$ 是协方差函数，而 I 是单位矩阵。

如果我们用暴力解法来解决优化问题(9), 那么将会相当耗时。根据 PLS 判别分析的特点[11], 可知潜在向量 T_A 和 T_B 是正交的, 并且其中的每一个方向向量 $t_{A/B}$ (即 $T_{A/B}$ 的列向量) 都与 $X_s^{A/B}$ 有最大的协方差。因此, 我们可以每次只考虑一个单一的方向向量 $t_{A/B}$ 来近似求解优化问题(9)。基于公式(5)和(6), 有 $t_A = X_t^{A^T} w_A$ 和 $t_B = X_t^{B^T} w_B$ 。在问题(9)中充分考虑协方差函数, 则可以将它转换为:

$$\begin{aligned} \max_{w_A, w_B} & w_A^T X_t^A X_s^{A^T} X_s^A X_t^{A^T} w_A + w_B^T X_t^B X_s^{B^T} X_s^B X_t^{B^T} w_B \\ \text{s.t.} & w_A^T X_t^A X_t^{A^T} w_A = I, w_B^T X_t^B X_t^{B^T} w_B = I \end{aligned} \quad (10)$$

我们通过在公式(10)上应用拉格朗日乘子来得到 w_A 和 w_B , 然后可以计算得到潜在成分 t_A 和 t_B 。因此, 目标函数可以改写为:

$$\begin{aligned} \max_{q_A, q_B, \lambda_1, \lambda_2} & \left\| X_s^{A^T} - t_A q_A^T \right\|^2 + \left\| X_s^{B^T} - t_B q_B^T \right\|^2 + \lambda_1 |q_A| + \lambda_2 |q_B| \\ \text{s.t.} & t_A^T t_A = 1, t_B^T t_B = 1 \end{aligned} \quad (11)$$

其中 q_A 和 q_B 分别是荷载矩阵 Q_A 和 Q_B 的列向量, λ_1 和 λ_2 是非负数参数。

在现实中, 观察值与有限的类标签相关联, 因此标签空间应该是稀疏的。然而, 公式(11)的 $t_{A/B}$ 和 $q_{A/B}$ 通常是密集且连续的, 这使得结果的可解释性变差。为了解决这个问题, 我们在 $q_{A/B}$ 上加入 L_1 范数惩罚约束(因为潜变量 $t_{A/B}$ 是正交的), 这为变量 $q_{A/B}$ 提供了变量选择并使它变得稀疏。同时, 我们在目标函数里加入了视角约束项 $\|Y_t^T w_A q_A^T - Y_t^T w_B q_B^T\|$ 来最大化不同视图的一致性。 $Y_t^T w_A q_A^T$ 和 $Y_t^T w_B q_B^T$ 分别是 Y_s^T 视角 A 和视角 B 的预测值。因此, 公式(11)可以写为:

$$\begin{aligned} \min_{q_A, q_B, \lambda_1, \lambda_2} \min_{q_A, q_B, \lambda_1, \lambda_2} & \left\| X_s^{A^T} - t_A q_A^T \right\|^2 + \left\| X_s^{B^T} - t_B q_B^T \right\|^2 + \lambda_1 |q_A| + \lambda_2 |q_B| + \left\| Y_t^T w_A q_A^T - Y_t^T w_B q_B^T \right\|^2 \\ \text{s.t.} & t_A^T t_A = 1, t_B^T t_B = 1 \end{aligned} \quad (12)$$

为了简化公式, 我们让 $t'_A = [t_A, \mathbf{0}]$, $t'_B = [\mathbf{0}, t_B]$, $E'_A = [\mathbf{1}, \mathbf{0}]$, $E'_B = [\mathbf{0}, \mathbf{1}]$, $q' = [q_A^T, q_B^T]$ 和 $Y' = [Y_t^T w_A, -Y_t^T w_B]$, 其中 $\mathbf{0}$ 和 $\mathbf{1}$ 是 $n_t \times 1$ 的向量, 这两个向量中的元素分别都是 0 和 1。因此, 公式(12)可以转换成:

$$\begin{aligned} \min_{q', \lambda_1, \lambda_2} & \left\| X_s^{A^T} - t'_A q'^T \right\|^2 + \left\| X_s^{B^T} - t'_B q'^T \right\|^2 + \lambda_1 |E'_A q'| + \lambda_2 |E'_B q'| + \|Y' q'^T\|^2 \\ \text{s.t.} & t'^T_A t'_A = 1, t'^T_B t'_B = 1 \end{aligned} \quad (13)$$

设 $f(q', \lambda_1, \lambda_2)$ 为公式(13)的优化函数:

$$f(q', \lambda_1, \lambda_2) = \min_{q', \lambda_1, \lambda_2} \left\| X_s^{A^T} - t'_A q'^T \right\|^2 + \left\| X_s^{B^T} - t'_B q'^T \right\|^2 + \lambda_1 |E'_A q'| + \lambda_2 |E'_B q'| + \|Y' q'^T\|^2 \quad (14)$$

问题(14)是一个传统的 LASSO 问题[20], 为了降低模型的复杂性, 我们使用软阈值操作来获得每个 q 的系数。因此, 对公式(14)求 q'^T 的偏导, 可得:

$$\frac{\partial f}{\partial q'^T} = 4q'^T - 2X_s^{A^T} t'_A - 2X_s^{B^T} t'_B + \lambda_1 \text{sgn}(E'_A q'^T) + \lambda_2 \text{sgn}(E'_B q'^T) + 2Y'^T Y' q'^T \quad (15)$$

其中 $\text{sgn}(a)$ 是 a 的符号函数, 如果 $a \geq 0$ 则 $\text{sgn}(a)$ 为 1; 否则 $\text{sgn}(a)$ 为 -1。设公式(15)为 0, 则有:

$$\begin{aligned} q'^T &= \frac{1}{2 + Y'^T Y'} (X_s^A t'_A + X_s^B t'_B) - \frac{1}{4 + 2Y'^T Y'} \{ \lambda_1 \text{sgn}(E'_A q'^T) + \lambda_2 \text{sgn}(E'_B q'^T) \} \\ &= S \left(\left[\frac{1}{2 + Y'^T Y'} (X_s^A t'_A + X_s^B t'_B) \right], \frac{1}{4 + 2Y'^T Y'} (\lambda_1 E'_A + \lambda_2 E'_B) \right) \\ &= S \left(\left[\frac{1}{2 + Y'^T Y'} (X_s^A t'_A + X_s^B t'_B) \right], \frac{1}{4 + 2Y'^T Y'} (\lambda_1 E'_A + \lambda_2 E'_B) \right) \end{aligned} \quad (16)$$

其中 $S(u, v)$ 是软阈值运算符，它的定义如下：

$$S(u, v) = \text{sgn}(u)(|u| - v)_+ = \begin{cases} u - v, & u > v \\ 0, & |u| \leq v \\ u + v, & u < -v \end{cases} \quad (17)$$

一旦求得 q^T ，就可以得到 q_A^T 和 q_B^T ，从而可以得到 Q_A^T 和 Q_B^T 。则 X_s^A 和 X_s^B 的回归模型为：

$$X_s^{A^T} = X_t^{A^T} W_A Q_A^T + E_{sA} \quad (18)$$

$$X_s^{B^T} = X_t^{B^T} W_B Q_B^T + E_{sB} \quad (19)$$

根据公式(18) (19)，视角 A 和视角 B 对应的类标签预测值如下：

$$\widehat{Y_s^{A^T}} = Y_t^T W_A Q_A^T \quad (20)$$

$$\widehat{Y_s^{B^T}} = Y_t^T W_B Q_B^T \quad (21)$$

则测试集最终的预测值为：

$$\widehat{Y_s^T} = \text{sgn}\left(\frac{1}{2}(Y_t^T W_A Q_A^T + Y_t^T W_B Q_B^T)\right) \quad (22)$$

预测值 Y_s^T 是连续的，但多视图多标签数据的标签空间应该是稀疏的。因此，我们引入 $\text{sgn}()$ 函数将预测值转换为二进制值：如果预测值大于 0，则将其重新赋值为 1；否则，它将被重新赋值为 0。

上述基于实例的多视角多标签学习算法模型，我们提出了该算法的完整过程，具体的算法实现步骤如表 1 所示。该算法的时间复杂度为 $\mathcal{O}(nks_i)$ ，而 $n = \max(n_i + n_s, d_1 + d_2)$ 。其中， m_1 和 m_2 是视角 A 和视角 B 中潜在变量的数量， ε 是非负参数。

Table 1. Algorithm implementation steps

表 1. 算法实现步骤

| |
|---|
| 输入：多视角多标签训练集数据 (X_i^A, X_i^B, Y_i) ，测试集数据 (X_s^A, X_s^B) ， |
| 参数 $(\lambda_1, \lambda_2, m_1, m_2, \varepsilon)$ |
| 输出： X_s^A 和 X_s^B 的标签集 Y_s |
| 1: 初始化: $Q_A = Q_B = \emptyset, W_A = W_B = \emptyset, k = m_1 = m_2$; |
| 2: for $i = 1, \dots, k$: |
| 3: $q_A = q_B = 1, s_i = 0$; |
| 4: 循环: |
| 5: $s_i = s_i + 1$; |
| 6: $q_{A(0)} = q_A, q_{B(0)} = q_B$; |
| 7: $u_A = X_s^{A^T} q_A, u_B = X_s^{B^T} q_B$; |
| 8: $w_A = X_i^A u_A, w_B = X_i^B u_B$; |
| 9: 计算 t : $t_A = X_i^{A^T} w_A, t_B = X_i^{B^T} w_B$; |
| 10: t 的标准化: $t_A = t_A / t_A^T t_A, t_B = t_B / t_B^T t_B$; |

Continued

- 11: 计算 p : $p_A = X_t^A t_A$, $p_B = X_t^B t_B$;
- 12: 根据公式(16)计算 q^{T^*} , 并从 q^{T^*} 得到 q_A^T 和 q_B^T ;
- 13: 直到满足 $(|q_A - q_{A(0)}| \leq \varepsilon \ \& \ |q_B - q_{B(0)}| \leq \varepsilon)$ 时停止循环
- 14: 更新 Q_A^T , Q_B^T , W_A 和 W_B :
- $$Q_A^T = [Q_A^T, q_A^T], \quad Q_B^T = [Q_B^T, q_B^T],$$
- $$W_A = [W_A, w_A], \quad W_B = [W_B, w_B];$$
- 15: 计算 X_t 和 Y_t 的误差信息:
- $$X_t^A = X_t^A - t_A p_A^T, \quad X_t^B = X_t^B - t_B p_B^T,$$
- $$X_s^A = X_s^A - t_A q_A^T, \quad X_s^B = X_s^B - t_B q_B^T;$$
- 16: 结束 for 循环
- 17: 输出测试集数据的标签: $Y_s^T = \text{sgn}\left(\frac{1}{2}(Y_t^T W_A Q_A^T + Y_t^T W_B Q_B^T)\right)$;

3. 实验与分析

3.1. 实验数据

为了验证模型的有效性和鲁棒性, 我们对多个多视角多标签数据集进行了对比实验, 这些数据集包括 mirickr, pascal07, corel5k, espgame, iapart12 [21], Health 和 Social [22]。在 mirickr, pascal07, corel5k, espgame 和 iapart12 数据集里, 我们使用 HSV 作为 A 视角的特征, 使用 RGB 作为视角 B 的特征。在 Health 和 Social 数据集里, 我们将特征分为两个相等的部分, 分别将它们视为视角 A 和视角 B 的特征。数据集的详情描述见表 2。

Table 2. Experimental data combination

表 2. 实验数据组合

| 视角 | mirickr | pascal07 | corel5k | espgame | iapart12 | Health | Social |
|-----|------------|------------|------------|------------|------------|-----------------|-----------------|
| 1 | HSV (4096) | HSV (4096) | HSV (4096) | HSV (4096) | HSV (4096) | Hea. Var. (306) | Soc. Var. (523) |
| 2 | RGB (4096) | RGB (4096) | RGB (4096) | RGB (4096) | RGB (4096) | Hea. Var. (306) | Soc. Var. (523) |
| q | 457 | 20 | 260 | 268 | 291 | 32 | 39 |
| n | 25000 | 9963 | 4999 | 20770 | 19627 | 5000 | 5000 |

3.2. 实验设置

为了验证算法的性能, 我们选择了四个评价指标, 分别是 Hamming loss、Ranking loss、One error 和 Average precision。其中 Hamming loss、Ranking loss 和 One error 的值越低, 则表示算法的性能越好; 而 Average precision 的值越高, 则算法的性能越好。为了证明算法的有效性, 我们选取了四种算法作对比, 如下所示:

- SWIM [12]: 该方法是一种单视角多标签学习算法, 它利用 PLS 来研究示例间的映射关系。

- **lrMMC [19]:** 该方法是一种多视角多标签学习算法, 它首先在所有数据视角中分别找到一个低秩的公共表示特征, 然后将这些低秩的公共表示特征嵌入到基于矩阵补全的图像分类器中来预测变量对应的标签。
- **SSDR-MML [23]:** 该方法是一种多视角多标签学习算法, 它利用了基于重构误差的框架去构建分类器。
- **GLMVML [24]:** 该方法是一种多视角多标签学习算法, 它从不同视角中结合了全局标签相关性和局部标签相关性, 然后提出了一种共识多视角表示方法来预测数据的类标签。

由于 SWIM 是一种单视角多标签学习算法, 它不能处理多视角多标签数据。因此, 我们首先利用 SWIM 算法为每个数据视角构建单视角多标签学习分类器(SWIM-A 和 SWIM-B), 然后再将视角 A 和视角 B 合并成一个单独的视角, 并基于该视角构建分类器(SWIM-all)。

在 SWIM 里, 参数 λ_w 和 λ_c 设为 0.1, m 的取值范围为 $\{1, \dots, 50\}$ 。对于 lrMMC, 公共子空间的维数 k 的取值范围为 $\{100, 200, \dots, 500\}$, 参数 γ 的取值范围为 $\{10^i \mid i = -4, \dots, 3\}$ 。对于 SSDR-MML, 我们将惩罚参数 C_0 设为 100, 让 $\lambda = 1$, 参数 $\alpha_1, \alpha_2, \beta_1$ 和 β_2 的取值范围为 $\{0, 0.05, \dots, 1\}$ 。对于 GLMVML, 我们让 $\lambda_0, \lambda_1, \lambda_2$ 和 λ_3 相等并设为 $\{2^{-5}, 2^{-4}, \dots, 2^0\}$, 让 $\lambda_4, \lambda_5, \lambda_6$ 和 λ_7 相等并设为 $\{10^{-6}, 10^{-5}, \dots, 10^0\}$, 让 λ_8 和 λ_9 相等并设为 $\{10^{-2}, 10^{-1}, \dots, 10^2\}$ 。对于我们的算法, 参数 λ_1 和 λ_2 的取值范围为 $\{0.05, 0.1, \dots, 1.0\}$, 让潜在变量的个数相等 ($m_1 = m_2$) 并将其取值范围设为 $\{1, 2, \dots, 50\}$ 。本文中所有的实验都是在 Conroe IV 上进行的, CPU 为 2.8 GHz, 主内存为 12 GB。我们用 MATLAB 2018b 实现了算法的性能对比, 在实验中, 我们采用 5 倍交叉验证和实验的平均值, 实验选取 80% 的数据作为训练集, 20% 的数据作为测试集。

3.3. 实验结果及分析

3.3.1. 算法的性能对比

在本节中, 我们将比较提出的方法和四个对比方法的性能。表 3~6 分别汇总了四个评价指标下的比较算法的性能。

Table 3. Hamming loss statistic
表 3. 汉明损失统计

| Hamming Loss | | | | | | | |
|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 算法 | mirflickr | pascal07 | corel5k | espgame | Iapart12 | Health | Social |
| SWIM-A | 0.041 | 0.097 | 0.066 | 0.072 | 0.073 | 0.044 | 0.024 |
| SWIM-B | 0.039 | 0.094 | 0.067 | 0.070 | 0.075 | 0.043 | 0.024 |
| SWIM-all | 0.036 | 0.089 | 0.063 | 0.067 | 0.068 | 0.040 | 0.023 |
| lrMMC | 0.034 | 0.093 | 0.051 | 0.056 | 0.060 | 0.043 | 0.024 |
| SSDR-MML | 0.036 | 0.082 | 0.059 | 0.069 | 0.072 | 0.043 | 0.024 |
| GLMVML | 0.043 | 0.107 | 0.061 | 0.098 | 0.111 | 0.043 | 0.024 |
| 所提方法 | 0.029 | 0.076 | 0.057 | 0.052 | 0.054 | 0.035 | 0.019 |

Table 4. Ranking loss statistic
表 4. 排序损失统计

| Ranking Loss | | | | | | | |
|--------------|-----------|----------|---------|---------|----------|--------|--------|
| 算法 | mirflickr | pascal07 | corel5k | espgame | Iapart12 | Health | Social |
| SWIM-A | 0.296 | 0.197 | 0.131 | 0.204 | 0.188 | 0.081 | 0.079 |

Continued

| | | | | | | | |
|----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| SWIM-B | 0.291 | 0.193 | 0.136 | 0.199 | 0.195 | 0.077 | 0.085 |
| SWIM-all | 0.269 | 0.186 | 0.123 | 0.190 | 0.182 | 0.064 | 0.071 |
| lrMMC | 0.287 | 0.371 | 0.205 | 0.257 | 0.236 | 0.082 | 0.095 |
| SSDR-MML | 0.308 | 0.168 | 0.152 | 0.219 | 0.194 | 0.060 | 0.062 |
| GLMVML | 0.393 | 0.162 | 0.118 | 0.246 | 0.211 | 0.056 | 0.078 |
| 所提方法 | 0.235 | 0.147 | 0.109 | 0.168 | 0.156 | 0.047 | 0.053 |

Table 5. One error statistic

表 5. 首标记错误统计

| One Error | | | | | | | |
|-----------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 算法 | mirflickr | pascal07 | corel5k | espgame | Iapart12 | Health | Social |
| SWIM-A | 0.968 | 0.491 | 0.565 | 0.588 | 0.661 | 0.370 | 0.342 |
| SWIM-B | 0.964 | 0.478 | 0.552 | 0.576 | 0.654 | 0.363 | 0.365 |
| SWIM-all | 0.949 | 0.455 | 0.534 | 0.559 | 0.644 | 0.339 | 0.328 |
| lrMMC | 0.972 | 0.637 | 0.816 | 0.789 | 0.756 | 0.377 | 0.386 |
| SSDR-MML | 0.895 | 0.422 | 0.677 | 0.663 | 0.657 | 0.342 | 0.319 |
| GLMVML | 0.906 | 0.524 | 0.481 | 0.574 | 0.553 | 0.335 | 0.331 |
| 所提方法 | 0.921 | 0.403 | 0.450 | 0.536 | 0.612 | 0.306 | 0.298 |

Table 6. Average precision statistic

表 6. 平均精度统计

| Average Precision | | | | | | | |
|-------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| 算法 | mirflickr | pascal07 | corel5k | espgame | Iapart12 | Health | Social |
| SWIM-A | 0.060 | 0.601 | 0.405 | 0.297 | 0.295 | 0.711 | 0.73 |
| SWIM-B | 0.063 | 0.605 | 0.402 | 0.303 | 0.286 | 0.714 | 0.721 |
| SWIM-all | 0.068 | 0.617 | 0.416 | 0.313 | 0.307 | 0.729 | 0.746 |
| lrMMC | 0.037 | 0.381 | 0.179 | 0.158 | 0.169 | 0.705 | 0.712 |
| SSDR-MML | 0.085 | 0.629 | 0.278 | 0.220 | 0.213 | 0.723 | 0.758 |
| GLMVML | 0.083 | 0.584 | 0.437 | 0.309 | 0.342 | 0.734 | 0.737 |
| 所提方法 | 0.089 | 0.657 | 0.459 | 0.332 | 0.338 | 0.756 | 0.772 |

根据实验结果, 我们提出的算法与其他对比算法相比, 拥有最好的性能。例如, 在表 6 得 pascal07 一列中, lrMMC、SSDR-MML 和 GLMVML 的 Average precision 分别为 0.381, 0.629, 0.584, 而提出的算法的 Average precision 为 0.657, 分别对比算法高出 0.276, 0.028 和 0.073。除 pascal07 数据集外, 在其他实验数据集上也可以发现类似的观察结果。

从实验结果可以看出, SWIM-all 的平均性能要优于 SWIM-A 和 SWIM-B。在所有数据集中, SWIM-all 的四项评价指标结果均高于单视角方法 SWIM-A 和 SWIM-B。以 Social 数据集为例, SWIM-all 的 Hamming loss, Ranking loss 和 One error 分别为 0.023, 0.071 和 0.328, 要明显低于 SWIM-A 的 0.024, 0.079 和 0.342, 以及 SWIM-B 的 0.024, 0.085 和 0.365。同时, SWIM-all 的 Average precision 为 0.746, 分别比 SWIM-A

的 0.733 和 SWIM-B 的 0.721 的值高 0.013 和 0.025。这些实验结果表明，不同视角可以直接提供编码在类标签中的互补信息，因此使用所有视角作为特征变量的分类器性能要优于使用单个视角为特征变量的分类器。

从表 3~6 可以看出，所提算法的性能要优于但是叫多标签算法的性能，包括 SWIM-all, SWIM-A 和 SWIM-B。在所有的数据集中，我们所提算法的性能都要明显优于 SWIM-all。以 pascal07 数据集为例，所提算法的 Hamming loss, Ranking loss 和 One error 分别为 0.076, 0.147 和 0.403, 明显低于 SWIM-A (0.097, 0.197, 0.491), SWIM-B (0.094, 0.193, 0.478) 和 SWIM-all (0.089, 0.186 和 0.455)。而所提算法的 Average precision 为 0.657, 分别比 SWIM-A (0.601), SWIM-B (0.605) 和 SWIM-all (0.617) 高 0.056, 0.052 和 0.040。这说明来自不同数据视角的数据具有不同的数据分布特征。与直接将两个视角连接成为一个单一视角相比，分别训练不同的数据视角再结合这些视角构建分类器分类器，会更有利于学习不同视角的特征，并且分类器的性能也会更好。

3.3.2. 参数影响

下面，我们将研究所提算法在不同参数值下的性能变化。我们提出的算法中，有两种参数影响算法的性能，即正则化常量 λ_1 和 λ_2 ，以及潜在变量的个数 m_1 和 m_2 。为了说明参数的作用，我们用不同参数值的来训练模型，并对比不同参数对模型性能的影响。

我们对数据集运行算法，得到 λ_1 和 λ_2 不同值时 Hamming loss 的变化，如图 2 所示。实验中，我们将 m_1 和 m_2 固定为 30，并令 $\lambda_1 = \lambda_2 = \{0.05, \dots, 1.0\}$ 。从图 2 可以看出，随着 λ 值的增加，Hamming loss 也在逐渐增大，但是增加值的变化很小。 λ 值越大模型的稀疏性越好，但性能相对较差。总的来说，算法的分类性能随着 λ 值的变化是相对稳定的。

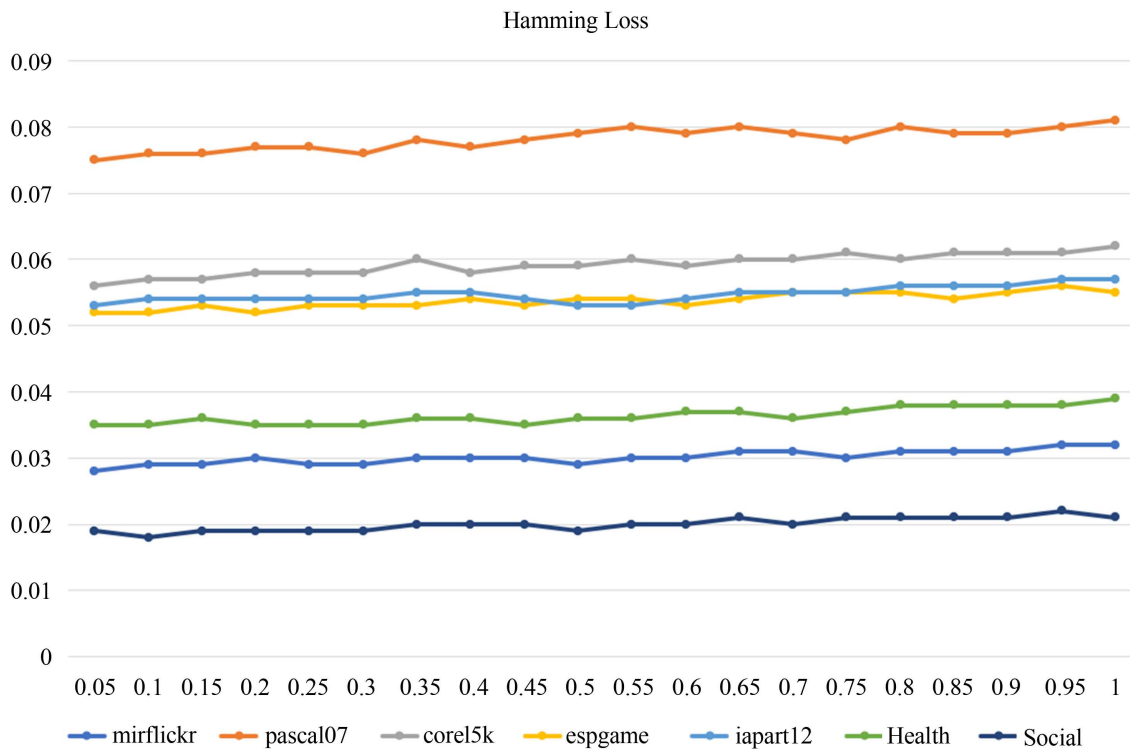


Figure 2. Hamming Loss with different values of λ

图 2. 不同 λ 值的 Hamming Loss 变化曲线

为了说明潜在变量数 m 的影响,我们用不同潜在变量数在数据集上进行了实验。我们令 $\lambda_1 = \lambda_2 = 0.1$, 令 $m_1 = m_2 = \{1, 2, \dots, 50\}$ 。 m_1 和 m_2 不同值时 Average precision 的变化如图 3 所示。显然,随着潜在变量数量的增加, Average precision 在开始时大幅上升,而当潜在变量数量 $m = 30$ 后 Average precision 的增加趋于缓慢,模型相对稳定。虽然 m_1 和 m_2 的值越大模型的平均精度越高,但模型的时间复杂度也在迅速增加。因此,潜在变量 m_1 和 m_2 的数量设置为 30 是相对合理的。

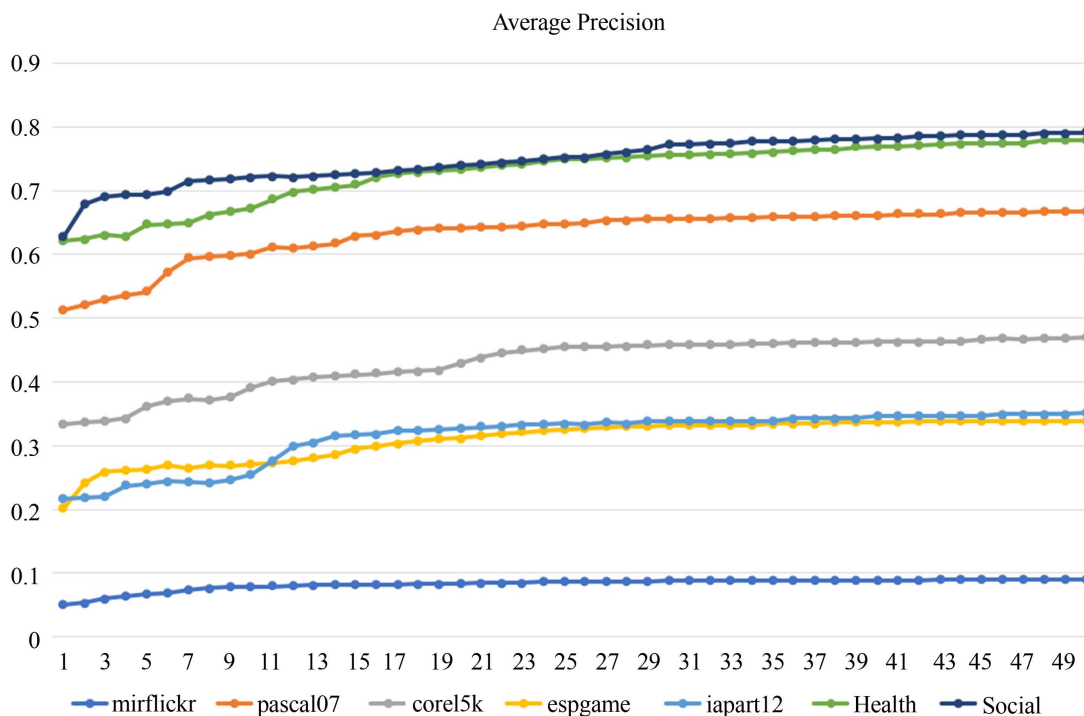


Figure 3. Average precision with different values of m

图 3. 不同 m 值的 Average precision 变化曲线

4. 总结

在本文中,我们提出了一种基于实例的多视角多标签方法,该方法通过映射训练示例和测试示例之间的关系来构建模型。简单来说,我们利用带 l_1 范数惩罚的 PLS 判别分析为每个数据视角构建多标签分类器,以使模型稀疏化并提高其鲁棒性。然后,根据视角一致性原则将所有数据视角的分类器结合起来,将不同视角的一致性和互补信息加入到多标签分类器中,从而提高模型的性能。我们在公共数据集上的实验结果表明,我们提出的算法表现出优异的性能。

基金项目

本文得到国家自然科学基金资助项目(No. 62076074)的资助。

参考文献

- [1] 唐静静, 田英杰. 多视角学习综述[J]. 数学建模及其应用, 2017, 6(3): 1-15+25.
- [2] 吕卫, 李德盛, 谭浪, 等. 基于多视角低秩表征的短视频多标签学习模型[J]. 激光与光电子学进展, 2020, 57(22): 139-146.
- [3] 黄家君. 情感分析中的方面类别识别研究[D]: [硕士学位论文]. 南京: 南京大学, 2016.

- [4] 温雯, 陈颖, 蔡瑞初, 等. 基于多视角多标签学习的读者情绪分类[J]. 计算机科学, 2018, 45(8): 191-197.
- [5] Feng, L., An, B. and He, S. (2019) Collaboration Based Multi-Label Learning. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, Honolulu, 27 January-1 February 2019, 3550-3557.
- [6] Fürnkranz, J., Hüllermeier, E., Loza Mencía, E., *et al.* (2008) Multilabel Classification via Calibrated Label Ranking. *Machine Learning*, **73**, 133-153. <https://doi.org/10.1007/s10994-008-5064-8>
- [7] Guo, Y. and Xue, W. (2013) Probabilistic Multi-Label Classification with Sparse Feature Learning. *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, Beijing, 3-9 August 2013, 1373-1379.
- [8] Su, H. and Rousu, J. (2013) Multilabel Classification through Random Graph Ensembles. *Machine Learning*, **99**, 231-256. <https://doi.org/10.1007/s10994-014-5465-9>
- [9] Enrique Sucar, L., Bielza, C., Morales, E.F., *et al.* (2014) Multi-Label Classification with Bayesian Network-Based Chain Classifiers. *Pattern Recognition Letters*, **41**, 14-22. <https://doi.org/10.1016/j.patrec.2013.11.007>
- [10] Zhu, S., Ji, X., Xu, W., *et al.* (2005) Multi-Labelled Classification Using Maximum Entropy Method. *Proceedings of the 28th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Salvador, 15-19 August 2005, 274-281. <https://doi.org/10.1145/1076034.1076082>
- [11] Liu, H., Ma, Z., Zhang, S., *et al.* (2015) Penalized Partial Least Square Discriminant Analysis with ℓ_1 -Norm for Multi-Label Data. *Pattern Recognition*, **48**, 1724-1733. <https://doi.org/10.1016/j.patcog.2014.11.007>
- [12] Liu, H., Li, X. and Zhang, S. (2016) Learning Instance Correlation Functions for Multilabel Classification. *IEEE Transactions on Cybernetics*, **47**, 499-510. <https://doi.org/10.1109/TCYB.2016.2519683>
- [13] Liu, H., Ma, Z., Han, J., *et al.* (2018) Regularized Partial Least Squares for Multi-Label Learning. *International Journal of Machine Learning and Cybernetics*, **9**, 335-346. <https://doi.org/10.1007/s13042-016-0500-8>
- [14] Yin, Q., Wu, S. and Wang, L. (2017) Unified Subspace Learning for Incomplete and Unlabeled Multi-View Data. *Pattern Recognition*, **67**, 313-327. <https://doi.org/10.1016/j.patcog.2017.01.035>
- [15] Tan, Q., Yu, G., Domeniconi, C., *et al.* (2018) Multi-View Weak-Label Learning Based on Matrix Completion. *Proceedings of the 2018 SIAM International Conference on Data Mining*, San Diego, 3-5 May 2018, 450-458. <https://doi.org/10.1137/1.9781611975321.51>
- [16] Tan, Q., Yu, G., Domeniconi, C., *et al.* (2018) Incomplete Multi-View Weak-Label Learning. *International Joint Conference on Artificial Intelligence*, Stockholm, 13-19 July 2018, 2703-2709.
- [17] Zhang, J., Li, C., Sun, Z., *et al.* (2019) Towards a Unified Multi-Source-Based Optimization Framework for Multi-Label Learning. *Applied Soft Computing*, **76**, 425-435. <https://doi.org/10.1016/j.asoc.2018.12.016>
- [18] Huang, J., Qu, X., Li, G., *et al.* (2019) Multi-View Multi-Label Learning with View-Label-Specific Features. *IEEE Access*, **7**, 100979-100992. <https://doi.org/10.1109/ACCESS.2019.2930468>
- [19] Liu, M., Luo, Y., Tao, D., *et al.* (2015) Low-Rank Multi-View Learning in Matrix Completion for Multi-Label Image Classification. *Twenty-Ninth AAAI Conference on Artificial Intelligence*, Austin, 25-30 January 2015, 2778-2784.
- [20] Friedman, J., Hastie, T. and Tibshirani, R. (2010) Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*, **33**, 1-22.
- [21] <http://lear.inrialpes.fr/people/guillaumin/data.php>
- [22] Learning from Multi-Label Data. MLKD. (2012) <http://mlkd.csd.auth.gr/multilabel.html>
- [23] Qian, B., Wang, X., Ye, J., *et al.* (2014) A Reconstruction Error Based Framework for Multi-Label and Multi-View Learning. *IEEE Transactions on Knowledge and Data Engineering*, **27**, 594-607. <https://doi.org/10.1109/TKDE.2014.2339860>
- [24] Zhu, C., Miao, D., Wang, Z., *et al.* (2020) Global and Local Multi-View Multi-Label Learning. *Neurocomputing*, **371**, 67-77. <https://doi.org/10.1016/j.neucom.2019.09.009>