

Research on Two-View Multi-Instance Image Classification Based on Similarity

Zijian Yin¹, Yanshan Xiao¹, Bo Liu²

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

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

Email: 229715148@qq.com

Received: Feb. 3rd, 2020; accepted: Feb. 18th, 2020; published: Feb. 25th, 2020

Abstract

In practice, some data contains a lot of privileged information, which can be used to train the classifier to improve classification performance. For example, in image classification, labels are used to describe images. These labels can be regarded as privileged information. The privileged information is complementary to the image and can be used for learning to improve the performance of image classification. The characteristics of multi-instance learning and two-view learning are suitable for image classification with privileged information. Therefore, a two-view multi-instance method based on similarity is proposed for image classification with privileged information. The proposed method considers one image as an instance, a collection of several images as a package, and privileged information as an instance. In order to solve the problem that the labels in the instances are unknown in practice, a similarity model is introduced. The proposed method first divides the image and privilege information into two different perspectives, then uses a clustering algorithm to construct the package, and finally trains a support vector machine classifier. The experimental results on four data sets show that the proposed method is more accurate than other similar models, and the two packet clustering algorithms are compared, and the sensitivity of each parameter is analyzed.

Keywords

Multi-Instance Learning, Two-View Learning, Image Classification, Support Vector Machine, Privileged Information

基于相似度的两视角多示例图像分类方法研究

尹子健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

Email: 229715148@qq.com

收稿日期: 2020年2月3日; 录用日期: 2020年2月18日; 发布日期: 2020年2月25日

摘要

在实际中,某些数据中包含许多特权信息,可用于训练分类器,从而提高分类性能。例如,在图像分类中,标签用于描述图像,这些标签可视为特权信息,特权信息与图像互补,可以用于学习以此提高图像分类性能。多示例学习和两视角学习的特性适用于带有特权信息的图像分类,因此提出了一种基于相似度的两视角多示例方法用于带有特权信息的图像分类。所提方法将一张图像视为一个示例,若干张图像的集合视为包,将特权信息视为示例。为解决实际中示例的标签是未知的问题,因而引入相似度模型。所提方法首先将图像和特权信息划分为两个不同的视角,然后使用聚类算法构造包,最后训练支持向量机分类器。在四个数据集上的实验结果表明,所提方法与其他相类似模型相比精确率更高,并比较了两种包的聚类算法,分析了各参数敏感度。

关键词

多示例学习, 两视角学习, 图像分类, 支持向量机, 特权信息

Copyright © 2020 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

多示例学习是从监督学习算法的基础上进化而来的,是为解决包(多个示例的集合)的分类[1]。在传统的多示例学习中,示例分为正示例和负示例,当包中至少有一个正示例时,该包的标签标记为正,称为正包;当包中的实例都是负示例时,该包的标签标记为负,称为负包。多示例学习的任务就是把已知标签的包进行训练,然后对未知标签的包的标签进行预测。

近年来,图像分类是计算机视觉领域的研究热点之一,也是其他图像应用领域的基础。因多示例学习二分类的特性,多示例学习也越来越多地应用在图像分类领域中,对于包与示例的定义有两种,一种定义是将一张图像视为一个包,对图像分割后产生的每个区域视为一个示例[2]。例如,Rao 等人[3]提出一种基于多尺度块的情感分类方法,该方法首先使用不同的图像分割方法提取多个比例的图像块,然后用多示例学习对图像的主要情感类型进行分类。另一种定义是将一张图像视为一个示例,把多张图像的集合视为包[4]。例如,Duan 等人[4]提出 GMI (generalized multi-instance)学习算法,该算法使用 k-means 聚类算法根据低级的视觉特征将相关的多张图像聚合成一个包,包中的图像视为示例,从而将图像分类问题转换成多示例学习问题。

尽管针对图像分类的多示例学习的研究有很多,但大多数都假设训练数据和测试数据具有相同数量的特征。但是,现实中可能会遇到这样的情况,即训练数据具有比测试数据更多的特征,而额外的特征通常称为特权信息[5]。在图像分类中,特权信息可以用作训练期间可用的附加特征,这些附加特征可以帮助训练处更准确地分类器并提高识别性能和泛化能力。目前,特权信息学习已在许多领域得到了广泛的研究。例如,Guo 等人[6]提出 ESVDD-neg (extended support vector data description with negative examples)方法,该方法使用特权信息学习来解决雷达自动目标识别问题。Wu 等人[7]提出 MT-PSVR (multi-target support vector regression)模型,该模型在训练每个目标模型时通过将其他目标视为特权信息来显示探索目标之间的相关性。

两视角学习的提出主要用于解决二维数据的学习问题[8]。这里的“视角”表示来自多个源或不同特征子集的数据，例如多媒体片段由视频信号和音频信号组成。Li 等人[9]提出两视角 TSVM (Two-view Transductive SVM)，用于已标记数据和未标记数据的分类。特权信息具有很多维度和互补性，例如网页包含内容文本和图像，它们相互补充并能够完整地描述对象。

特权信息的概念最初由 Vapnik 等人提出[10]，现实世界中的数据通常与丰富的描述相关联，并且这些丰富的描述在学习任务中被视为特权信息。例如，网络图像通常包含辅助信息(例如，文本描述，注释等)，这些信息被视为特权信息。实际中，特权信息可以是标签、属性、关键字等。随着在线图像共享平台(如 Flickr、Instagram 等)的发展，这些平台允许用户上传图像，并为图像贴上标签，如图 1 所示。这些标签(称为“特权信息”)能够描述图像内容，能提高图像检索的性能。这些图像和特权信息构成数据的二维性，因此可以应用于两视角学习。例如，Tang 等人[11]提出了多视角特权 SVM 模型。该模型在两视角学习和特权信息的基础上进行扩展。该方法将两种互补的特权信息分成两个不同的视角，并在两个视角之间添加约束来补偿它们之间的差距。



Figure 1. Web images with privileged information
图 1. 带有特权信息的 Web 图像

包和示例的特性适用于带有特权信息的图像分类，因此采用多示例学习来表示数据，采用两视角学习来划分数据并同时训练分类器。所提方法将一张图像视为一个示例，将若干图像的集合视为包，将一个关键词(特权信息)视为一个示例，将若干个关键词的集合视为包。使用特权信息提高图像分类的精确度的前提是包中示例的标签是已知的，但在现实中示例的标签是模糊的，而现有的研究较少提及这方面的处理，因此所提模型引入相似度模型来描述数据。所提模型可用于在线图像共享平台提高 Web 图像检索性能。

2. 基于相似度的两视角多示例模型

2.1. 相似度数据模型

在多示例学习中，每个正包至少包含一个正示例，负包里都是负实例。但是，正包中的示例标签是模糊的，即有可能是正的，有可能是负的。为此，引入相似度数据模型来描述多示例学习问题。

用 $D = \left\{ (B_1^+, Y_1^+), \dots, (B_{N^+}^+, Y_{N^+}^+), (B_1^-, Y_1^-), \dots, (B_{N^-}^-, Y_{N^-}^-) \right\}$ 来表示训练包的集合。其中， B_i^+ 表示一个正包， Y_i^+ 表示正包 B_i^+ 的标签， $Y_i^+ = +1$ ； B_i^- 表示一个负包， Y_i^- 表示正包 B_i^- 的标签， $Y_i^- = -1$ 。 N^+ 和 N^- 分别表示正包和负包的数量。该模型基于示例的相似度选择正候选，有以下定义：

定义 1: (基于单集的相似度): 给出一个示例 \bar{x} 和一个子集 S ，示例 \bar{x} 和子集 S 的相似度可以定义如下 [12]:

$$R(\bar{x}, S) = \frac{1}{2} \sum_{\bar{x}_i \in S} e^{-\|\varphi(\bar{x}) - \varphi(\bar{x}_i)\|} \quad (1)$$

其中， $\varphi(\cdot)$ 是一个非线性映射函数，将示例 \bar{x} 或 \bar{x}_i 映射到特征空间中。于是，两类成员计算如下：

$$m^+(\bar{x}_i) = \frac{1}{2} \left[R(\bar{x}, S_p^+) + 1 - R(\bar{x}, S_n^-) \right] \quad (2)$$

$$m^-(\bar{x}_i) = \frac{1}{2} \left[R(\bar{x}, S_n^-) + 1 - R(\bar{x}, S_p^+) \right] \quad (3)$$

让多示例包里的示例 \bar{x}_i 表示为 $\{\bar{x}_i, m^+(\bar{x}_i), m^-(\bar{x}_i)\}$ 。分别 $m^+(\bar{x}_i)$ 和 $m^-(\bar{x}_i)$ 表示示例 \bar{x}_i 的趋近于正类和负类的相似度, 有 $0 \leq m^+(\bar{x}_i) \leq 1$ 以及 $0 \leq m^-(\bar{x}_i) \leq 1$ 。如果示例 \bar{x}_i 是正示例, 则 $m^+(\bar{x}_i) = 1$ 且 $m^-(\bar{x}_i) = 0$ 。如果示例 \bar{x}_i 是负示例, 则 $m^+(\bar{x}_i) = 0$ 且 $m^-(\bar{x}_i) = 1$ 。如果示例 \bar{x}_i 是模糊示例, 其标签未知, 用 $\{\bar{x}_i, m^+(\bar{x}_i), m^-(\bar{x}_i)\}$ 表示, 有 $0 < m^+(\bar{x}_i) < 1$ 以及 $0 < m^-(\bar{x}_i) < 1$ 。其中, S_p^+ 存储正包中的正候选, S_a^+ 存储正包中除 S_p^+ 以外的示例, S_n^- 存储负包中的示例。对于 S_a^+ 中的示例, 有二成员 $\{\bar{x}_i, m^+(\bar{x}_i), m^-(\bar{x}_i)\}$ 分别趋近于正类和负类。根据以上定义, 设 $\{\bar{x}_i, m_A^+(\bar{x}_i), m_A^-(\bar{x}_i)\}$ 表示趋近于含有特权信息的正类和负类数据, 简称为 A 视角。进一步让 $S^A = S_p^{A+} \cup S_a^{A+}$ 和 $S^A = S_n^{A-} \cup S_a^{A-}$ 。类似的, 设 $\{\bar{x}_i, m_B^+(\bar{x}_i), m_B^-(\bar{x}_i)\}$ 表示趋近于剔除特权信息的正类和负类数据, 简称为 B 视角, 有 $S^{B+} = S_p^{B+} \cup S_a^{B+}$ 和 $S^{B-} = S_n^{B-} \cup S_a^{B-}$ 。

2.2. 对偶问题

所提方法分别对于 A 视角 (B_i^A, Y_i^A) 和 B 视角 (B_j^B, Y_j^B) 训练支持向量机。假设 $f_{A/B} = \omega_{A/B} \cdot \phi(\bar{x}) + b_{A/B}$ 为上述两个视角的超平面。所提方法是基于多示例学习, 引导出最小化问题, 其目标方程:

$$\begin{aligned} \min_{\omega, b, \xi} & \frac{1}{2} \|\omega_A\|^2 + \frac{1}{2} \|\omega_B\|^2 + C_1 \sum_{i=1}^n [\omega_A \cdot \phi(\bar{x}_i) + b_A] + C_2 \sum_{i=n+1}^{n+m} [\omega_B \cdot \phi(\bar{x}_j) + b_B] \\ & + D_A \left\{ \sum_{S^A} m_A^+ [\phi(\bar{x}_i)] \xi_i + \sum_{S^A} m_A^- [\phi(\bar{x}_j)] \xi_j \right\} \\ & + D_B \left\{ \sum_{S^{B+}} m_B^+ [\phi(\bar{x}_k)] \xi_k + \sum_{S^{B-}} m_B^- [\phi(\bar{x}_h)] \xi_h \right\} \end{aligned} \quad (4)$$

约束条件:

$$\begin{aligned} & \left| \omega_A \cdot \phi(\bar{x}_i) + b_A - \omega_B \cdot \phi(\bar{x}_j) - b_B \right| \leq \sum_{i=1}^{m+n} \eta_i + \varepsilon \\ & y_i \left[\omega_A \cdot \phi(\bar{x}_i) + b_A \right] \geq 1 - \left[\omega_B \cdot \phi(\bar{x}_j) + b_B \right] \\ & \omega_A \cdot \phi(\bar{x}_i) + b_A \geq 1 - \xi_i \\ & \omega_A \cdot \phi(\bar{x}_j) + b_A \leq -1 + \xi_j \\ & \omega_B \cdot \phi(\bar{x}_k) + b_B \geq 1 - \xi_k \\ & \omega_B \cdot \phi(\bar{x}_h) + b_B \leq -1 + \xi_h \\ & \xi_i \geq 0, \quad \xi_j \geq 0, \quad \xi_k \geq 0, \quad \xi_h \geq 0 \end{aligned}$$

其中, D_A 和 D_B 是控制两个视角的正则项。如果 $D_A > D_B$, 则 A 视角优于 B 视角, 反之, B 视角优于 A 视角。参数 C_1 和 C_2 是平衡边距和误差的参数。 $\xi_i, \xi_j, \xi_k, \xi_h$ 是松弛变量。在 A 视角中, 对于 S_p^{A+} 中的每个示例 \bar{x}_i , $m_A^+(\bar{x}_i) = 1$ 趋近于正类, 又因为有 $S^A = S_p^{A+} \cup S_a^{A+}$, 所以 S^A 有 $m_A^-(\bar{x}_j)$ 趋近于负类。B 视角中的 $m_B^+(\bar{x}_k), m_B^-(\bar{x}_h)$ 同理可得。约束 $y_i \left[\omega_A \cdot \phi(\bar{x}_i) + b_A \right] \geq 1 - \left[\omega_B \cdot \phi(\bar{x}_j) + b_B \right]$ 表示两视角间的约束。 f_A 和 f_B 分别表示 A 视角和 B 视角的决策函数。 η_i 是两视角间的共识变量。 ε 是允许某些实例违反约束的松弛变量。此外, $\phi(\cdot)$ 是映射函数, 它将数据从输入空间映射到特征空间, 可引入核函数来计算特征空间中两个向量的内积, 即 $K(\bar{v}, \bar{x}) = \phi(\bar{v}) \cdot \phi(\bar{x})$ 。

为解决公式(4)中的优化问题,用拉格朗日方法,引入拉格朗日乘子 $\alpha_i, \alpha_j, \alpha_k, \alpha_h$ 并对 ω_A 和 ω_B 求偏导得:

$$\omega_A = (\alpha_i + \beta_i - \beta_j + \theta_k y_i - C_1) \cdot \phi(\bar{x}_i) - \alpha_j \cdot \phi(\bar{x}_j) \quad (5)$$

$$\omega_B = \alpha_k \cdot \phi(\bar{x}_k) - \alpha_h \cdot \phi(\bar{x}_h) - (C_2 + \beta_i + \beta_j - \theta_k y_i) \cdot \phi(\bar{x}_j) \quad (6)$$

根据 Kuhn-Tucker Theorem 定理得:

$$L = \sum_{S^A} a_i + \sum_{S^A} a_j + \sum_{S^B} a_k + \sum_{S^A} a_h - \left(\sum_{i=1}^n \beta_i + \sum_{j=1}^n \beta_j \right) \varepsilon - \frac{1}{2} \|\omega_A\|^2 - \frac{1}{2} \|\omega_B\|^2 \quad (7)$$

将公式(5)和(6)代入(7)得到对偶问题:

$$\begin{aligned} \max F = & -\frac{1}{2} (\alpha_i + \beta_i - \beta_j + \theta_k y_i - C_1)^2 \cdot K(\bar{x}_i, \bar{x}_i) - \frac{1}{2} \alpha_j^2 \cdot K(\bar{x}_j, \bar{x}_j) \\ & + \alpha_j (\alpha_i + \beta_i - \beta_j + \theta_k y_i - C_1) \cdot K(\bar{x}_i, \bar{x}_j) - \frac{1}{2} \alpha_k^2 \cdot K(\bar{x}_k, \bar{x}_k) \\ & + \alpha_k \alpha_h \cdot K(\bar{x}_k, \bar{x}_h) - \frac{1}{2} (C_2 + \beta_i + \beta_j - \theta_k y_i)^2 \cdot K(\bar{x}_j, \bar{x}_j) \\ & - (C_2 + \beta_i + \beta_j - \theta_k y_i) \cdot \left[\alpha_h K(\bar{x}_h, \bar{x}_j) - \alpha_k K(\bar{x}_k, \bar{x}_j) \right] + \sum_{S^A} \alpha_i \\ & + \sum_{S^A} \alpha_j + \sum_{S^B} \alpha_k + \sum_{S^B} \alpha_h - \left(\sum_{i=1}^n \beta_i + \sum_{j=1}^n \beta_j \right) \cdot \varepsilon - \frac{1}{2} \alpha_h^2 \cdot K(\bar{x}_h, \bar{x}_h) \end{aligned} \quad (8)$$

约束条件:

$$\begin{aligned} 0 & \leq \beta_i \leq C_1 + C_2 \\ 0 & \leq \beta_j \leq C_1 + C_2 \\ 0 & \leq \alpha_i \leq D_A m_A^+ \cdot \phi(x_i) \\ 0 & \leq \alpha_j \leq D_A m_A^- \cdot \phi(x_i) \\ 0 & \leq \alpha_k \leq D_B m_B^+ \cdot \phi(x_i) \\ 0 & \leq \alpha_h \leq D_B m_B^- \cdot \phi(x_i) \end{aligned}$$

算法实现步骤如表 1 所示。

Table 1. Algorithm implementation steps

表 1. 算法实现步骤

输入: 带有特权信息的图像集

输出: $\omega_A, \omega_B, b_A, b_B$

- 1: 根据公式(2)和(3)初始化包 B_i 的标签 Y_i
- 2: 将文本和图像划分成两个视角
- 3: 初始化 $S_p^{A+}, S_p^{B+}, S_a^{A+}, S_a^{B+}, S_n^{A-}, S_n^{B-}$
- 4: 让 $t = 0$
- 5: 重复步骤 6 至 19
- 6: $t = t + 1$
- 7: 对于 A 视角和 B 视角中的每个正包 B_A^+, B_B^+ , 重复步骤 8 至 18

Continued

- 8: 对于正包 B_A^+, B_B^+ 中的每一个示例, 重复步骤 9 至 12
- 9: 令 x_i^A, x_i^B 为包 B_A^+, B_B^+ 的正候选
- 10: 用 S_p^{A+} 更新 $S_p^{A'}$, 用 S_p^{B+} 更新 $S_p^{B'}$ 。
- 11: 通过用 $x_{k'}^A, x_{k'}^B$ 替代 x_{k-1}^A, x_{k-1}^B 来更新 $S_p^{A'}, S_p^{B'}$
- 12: 计算 f_A, f_B 表示 $F(x_i)$
- 13: 通过用 $x_{k'}^A, x_{k'}^B$ 替代 x_{k-1}^A, x_{k-1}^B 来更新 $S_p^{A'}, S_p^{B'}$
- 14: 获得新的正候选并返回 $\arg \max F(x_i)$
- 15: 用 $D - S_p^{A+} - S_n^{A-}$ 来更新 S_a^{A+} ; 用 $D - S_p^{B+} - S_n^{B-}$ 来更新 S_a^{B+}
- 16: 根据公式(2)和(3)计算 $m^+(x_i), m^-(x_i)$
- 17: 基于 $S_p^{A+}, S_p^{B+}, S_a^{A+}, S_a^{B+}, S_n^{A-}, S_n^{B-}$ 解决 QP 问题得到 F
- 18: 用 F 更新 $F^{(l)}$
- 19: 直到 $|F^{(l-1)} - F^{(l)}| \leq \tau F^{(l-1)}$

3. 实验与结果分析

3.1. 数据集

实验数据来自 NUS-WIDE、Flickr30k、WebQuery 和 AwA2 (Animals with Attributes 2)这四个数据集:

- NUS-WIDE: 该数据集包括来自 Flickr 的 269648 张图像的 5018 个唯一标签和相关标签。
- Flickr30k: 该数据集已成为基于句子的图像描述的基准, 包含 31,783 张日常活动、事件和场景的照片以及 158,915 句文字描述。
- WebQuery: 该数据集包含从 353 个文本查询中检索到的 71,478 个 Web 图像。并且数据集中的每个图像都与多种语言的文字描述相关联。
- AwA2: 该数据集由 50 个动物类别的 37,322 张图像组成, 每个图像具有预先提取的特征表示, 并带有 85 个语义属性进行注释。

该实验将一张图像视为一个示例, 将一个标签、注释、一句文字描述视为一个示例, 都使用聚类算法构造多示例包。

3.2. 实验设计

3.2.1. 两种多示例包的聚类算法

在机器学习领域, 有很多聚类算法, 如 k-means [15], DBSCAN [16]算法等, 这些方法可以用来构造多示例包。实验中使用 k-means 和 DBSCAN 算法对于包进行构造。

k-means 聚类算法: 由于简单和效率高的特性, k-means 聚类算法成为聚类算法中使用最广泛的算法。根据前人的经验, 该实验设参数 $k = \lfloor T/15 \rfloor$, 其中 T 表示训练包的数量。该实验使用的 k-means 聚类算法是基于欧氏距离: $d(x_i, x_j) = \sqrt{\lambda^2 \|v_i - v_j\|^2 + \|t_i - t_j\|^2}$, 其中, v_i, t_i 和 v_j, t_j 分别是第 i 张和第 j 张图像的视觉和文本特征。 $d(x_i, x_j)$ 是两张图像之间的距离。

DBSCAN 聚类算法: DBSCAN 是基于密度的聚类算法, 可以应用于凸样本集和非凸样本集。DBSCAN 将聚类定义为按密度连接的最大点集, 将具有足够高密度的区域划分为聚类, 并可以在噪声的空间数据库中找到任意形状的聚类。

3.2.2. 比较算法和参数设定

由于所提方法是多示例学习方法，因此将其性能与以下四种多示例学习方法进行比较：

- sMIL-PI: 该方法将松散标记的 Web 数据合并到学习过程中，利用附加的文本特征来有效地处理相关训练图像的噪声标签。该方法基于最大平均差异准则添加了一个正则化器，以减少数据分布不匹配[7]。
- PSVM-2V: 该方法基于两视角学习用于特权信息学习，着重于将特权信息纳入多示例学习和训练 SVM 作为分类器[11]。
- SVM+: 该方法是一种基于特权信息的学习范式，用特权特征空间中定义的松弛函数替换标准 SVM 中的松弛变量[13]。
- SVM-2K: 该方法是最早提出的两视角学习方法，它构造两个带有标记和未标记数据的视角以训练分类器[14]。

实验将所提方法与 SVM+, SVM-2K, sMIL-PI 和 PSVM-2V 比较，采用平均精确度(average precision)来衡量各方法的性能。在 SVM-2K 和 PSVM-2V 中，设其参数 $C^A = C^B = C$ ；在 sMIL-PI 中，设其参数 $C_1 = C_2$ 。实验中将以上这些参数以及 sMIL-PI、PSVM-2V、SVM+ 中的权衡参数 γ 的值均选自范围 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ 。经验表明， ε 越小，分类的性能通常越好，因此使参数 $\varepsilon = 0.01$ 。对于所提方法，设参数 $C_1 = C_2, D_A = D_B$ ，并且这些参数选自范围 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ 。

对于所有基于 SVM 的方法，实验中使用高斯径向核函数： $K(\bar{x}_i, \bar{x}_j) = \exp\left(-\frac{\|\bar{x}_i - \bar{x}_j\|^2}{2\sigma^2}\right)$ ，核参数 σ 的值选自范围 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ 。

3.3. 实验结果与分析

3.3.1. 性能比较

实验使用 NUS-WIDE、Flickr30k、WebQuery 和 AwA2 数据集对 SVM+, SVM-2K, sMIL-PI, PSVM-2V 和所提方法进行测试与比较。首先将图像和特权信息划分成两个不同的视角，然后在每个视角中分别使用 k-means 和 DBSCAN 聚类算法构造多示例包，构造包的数量为 300、600、900、1200 和 1500 分别对各方法进行测试，结果如表 2~6 所示。

从表 2~6 可以看出，在 NUS-WIDE、Flickr30k、WebQuery 和 AwA2 数据集上，从整体上看所提方法略优于比其他多示例方法。随着包的数量从 300 到 900 的增加，SVM+、SVM-2K、sMIL-2V、PSVM-2V 和所提方法的平均精确度都有所上升。其中 PSVM-2V、SVM-2K 和所提方法在包的数量为 900 时达到峰值，随后开始下降，SVM+ 和 sMIL-2V 在包的数量为 1200 时达到峰值，随后也开始下降。随着包的增加，噪声也有所增加，导致分类精确度也随之下降。由此可以得出结论：基于相似度的多示例学习方法分类精度更高。同时由结果可以看出，整体上看使用 DBSCAN 比 k-means 算法效果要更优。

Table 2. Average precision when the bag numbers is 300 (unit: %)

表 2. 包的数量为 300 时的平均精确度(单位: %)

数据集	NUS-WIDE		Flickr30k		WebQuery		AwA2	
	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN
SVM+	72.83	74.46	71.02	72.97	69.48	72.51	75.34	77.38
SVM-2K	66.15	68.37	65.57	66.30	67.33	68.43	70.75	72.04
sMIL-PI	76.81	79.19	73.88	76.55	75.46	76.31	74.82	76.53
PSVM-2V	80.18	80.87	81.32	82.46	80.37	79.47	81.24	82.72
所提方法	81.24	82.60	80.28	81.04	82.61	83.53	81.73	83.59

Table 3. Average precision when the bag numbers is 600 (unit: %)**表 3.** 包的数量为 600 时的平均精确度(单位: %)

数据集	NUS-WIDE		Flickr30k		WebQuery		AwA2	
	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN
SVM+	76.64	75.09	75.52	75.03	74.64	72.57	78.57	78.79
SVM-2K	70.48	71.53	73.19	73.06	72.29	74.15	71.08	72.42
sMIL-PI	76.86	75.92	78.01	80.55	79.36	82.23	76.28	79.36
PSVM-2V	82.82	82.96	84.47	86.63	83.90	82.14	83.11	83.69
所提方法	83.49	83.02	84.93	85.95	85.81	85.60	84.03	85.76

Table 4. Average precision when the bag numbers is 900 (unit: %)**表 4.** 包的数量为 900 时的平均精确度(单位: %)

数据集	NUS-WIDE		Flickr30k		WebQuery		AwA2	
	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN
SVM+	77.22	79.04	80.43	82.68	78.38	79.63	79.18	80.81
SVM-2K	75.05	76.23	75.03	78.30	76.26	79.31	74.34	77.42
sMIL-PI	80.05	81.54	83.22	83.61	80.39	82.09	80.98	81.47
PSVM-2V	83.64	85.39	85.07	86.94	83.37	82.08	85.46	87.63
所提方法	86.21	88.72	87.40	88.46	85.09	86.29	88.22	88.35

Table 5. Average precision when the bag numbers is 1200 (unit: %)**表 5.** 包的数量为 1200 时的平均精确度(单位: %)

数据集	NUS-WIDE		Flickr30k		WebQuery		AwA2	
	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN
SVM+	79.80	81.31	80.84	84.52	79.41	81.06	82.07	81.19
SVM-2K	74.11	75.02	73.13	74.28	76.17	77.39	75.62	75.83
sMIL-PI	82.63	82.08	84.04	83.74	80.67	81.54	81.02	82.24
PSVM-2V	82.59	84.95	83.75	85.23	81.39	81.75	85.29	86.31
所提方法	84.62	84.54	86.02	86.36	83.70	84.01	87.25	86.49

Table 6. Average precision when the bag numbers is 1500 (unit: %)**表 6.** 包的数量为 1500 时的平均精确度(单位: %)

数据集	NUS-WIDE		Flickr30k		WebQuery		AwA2	
	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN	k-means	DBSCAN
SVM+	75.53	76.47	76.66	78.12	78.28	75.37	80.53	80.91
SVM-2K	72.02	73.64	70.15	69.58	69.82	73.26	71.47	72.24
sMIL-PI	78.35	79.74	76.21	77.49	78.37	80.36	74.51	75.83
PSVM-2V	80.33	81.07	82.96	81.47	79.04	80.73	82.29	83.65
所提方法	82.44	83.80	83.56	83.92	81.32	80.69	84.22	84.61

3.3.2. 参数敏感度分析

此外,在所提方法中有一些参数,实验分析了不同参数值下的性能变化。实验使用 NUS-WIDE 数据集做测试,实验结果如图 2 和图 3 所示。首先固定 D_A, D_B 测试参数 C_1, C_2 。由实验结果可得,如图 2 所示,在范围 $\{10^{-3}, 10^{-2}, 10^{-1}, 1, 10^1, 10^2, 10^3\}$ 中,参数 C_1, C_2 在 10^{-3} 到 1 时,性能上升,在 1 到 10^3 时,性能下降,当参数 $C_1 = C_2 = 1$ 时,效果最优。接着固定 C_1, C_2 测试 D_A, D_B ,如图 3 所示,参数 D_A, D_B 在 10^{-3} 到 10 时,性能上升,在 10 到 10^3 时,性能下降,当参数 $D_A = D_B = 10$ 时,效果最优。

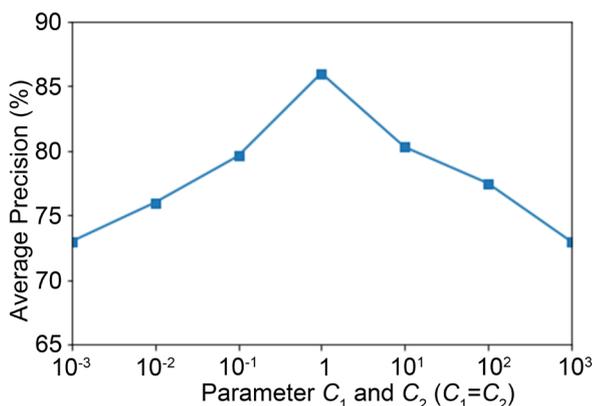


Figure 2. The Parameter Sensitiveness of C_1 and C_2

图 2. 参数 C_1 和 C_2 的敏感度

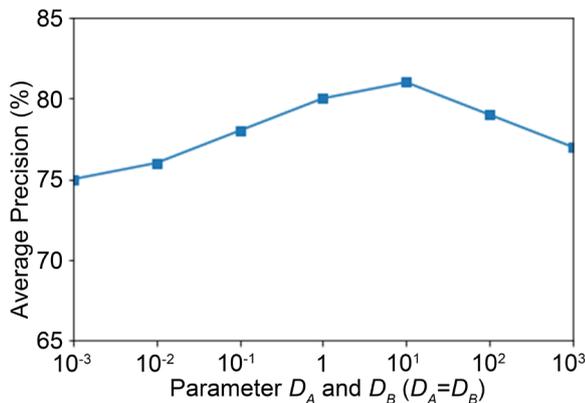


Figure 3. The Parameter Sensitiveness of D_A and D_B

图 3. 参数 D_A 和 D_B 的敏感度

3.3.3. 噪声敏感度测试

实验还测试算法性能对输入数据噪声的敏感程度,使用 NUS-WIDE 数据集做测试。将噪声添加到数据示例中方法的基本思想如图 4 所示。首先计算整个数据沿第 i 个维度的标准差 σ_i^0 ,然后获得高斯噪声的标准偏差 σ_i ,其范围随机选自 $[0, 2 \cdot \sigma_i^0]$ 。然后,对于第 i 维,我们将随机分布的噪声与标准差 σ_i 相加。这样,在数据示例 x_j 中添加了噪声,可以将其表示为向量: $\sigma^{x_j} = [\sigma_1^{x_j}, \sigma_2^{x_j}, \dots, \sigma_{n-1}^{x_j}, \sigma_n^{x_j}]$,其中 n 表示数据示例 x_j 的维数, $\sigma_i^{x_j}, i=1, \dots, n$ 表示添加到数据示例第 i 维的噪声。

实验中使数据噪声的百分比在 0% 到 30% 之间变化。这里以向量加偏移常数方法生成的数据为例,并将噪声添加到每个包的示例中。不同比例的噪声对性能的影响如图 5 所示,随着噪声水平的提高,所有方法的性能都会下降。但是与其他方法相比,所提方法可以获得更高的精度,对噪声的敏感度较低。

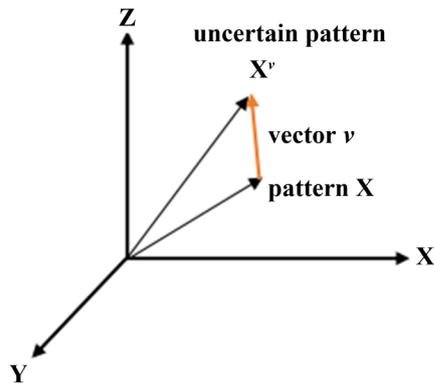


Figure 4. Add the noise to a data example
图 4. 将噪声添加到数据示例

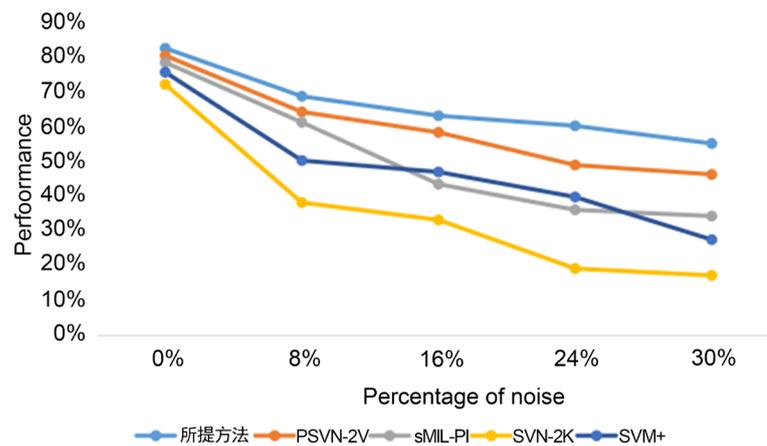


Figure 5. Effect of noise on algorithm performance
图 5. 噪声对算法性能的影响

3.3.4. 算法运行时间

各算法的平均运行时间比重如图 6 所示，所提方法比其他所有方法的时间开销要大得多。这是因为所提方法考虑了示例的相似性，并且同时利用约束条件来拟合原始特征和特权信息。因此，所提方法比其他方法花费更多时间。

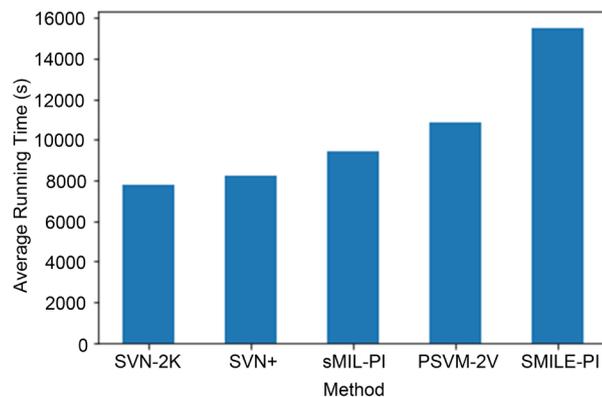


Figure 6. Average running time of different methods
图 6. 不同算法的平均运行时间

4. 总结与展望

针对带有特权信息的图像提出一种基于相似度的两视角多示例图像分类方法,该方法考虑了示例的相似性以及具有特权信息特征和原始特征的约束,提出了一个迭代框架来解决使用特权信息进行多示例学习的问题,该方法可用于带有文本描述的图像分类。实验使用两种聚类算法(k-means 和 DBSCAN)来处理图像并比较其性能。实验结果表明,所提方法的性能明显优于其他方法。但是,所提方法的时间开销比其他方法都大,优化模型以降低时间开销和降低对噪声的敏感程度是未来的工作。

基金项目

本文得到国家自然科学基金资助项目(No.61876044)的资助。

参考文献

- [1] Dietterich, T.G., Lathrop, R.H. and Lozano-Pérez, T. (1997) Solving the Multiple Instance Problem with Axis-Parallel Rectangles. *Artificial Intelligence*, **89**, 31-71. [https://doi.org/10.1016/S0004-3702\(96\)00034-3](https://doi.org/10.1016/S0004-3702(96)00034-3)
- [2] Lu, J. and Ma, S. (2009) Web Image Clustering Based-on Multiple Instance Learning. *Computer Research and Development*, **46**, 1462-1470.
- [3] Rao, T., Xu, M., Liu, H., Wang, J. and Burnett, I. (2016) Multi-Scale Blocks Based Image Emotion Classification Using Multiple Instance Learning. 2016 *IEEE International Conference on Image Processing (ICIP)*, Phoenix, AZ, 25-28 September 2016, 634-638. <https://doi.org/10.1109/ICIP.2016.7532434>
- [4] Duan, L., Li, W., Tsang, I.W. and Xu, D. (2011) Improving Web Image Search by Bag-Based Reranking. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, **20**, 3280-3290. <https://doi.org/10.1109/TIP.2011.2159227>
- [5] Yao, Y., Shen, F., Zhang, J., Liu, L., Tang, Z. and Shao, L. (2018) Extracting Privileged Information for Enhancing Classifier Learning. *IEEE Transactions on Image Processing*, **28**, 436-450.
- [6] Guo, Y., Xiao, H., Kan, Y. and Fu, Q. (2018) Learning Using Privileged Information for HRRP-Based Radar Target Recognition. *IET Signal Processing*, **12**, 188-197. <https://doi.org/10.1109/MSP.2018.2841413>
- [7] Wu, G., Tian, Y. and Liu, D. (2018) Privileged Multi-Target Support Vector Regression. 2018 *24th International Conference on Pattern Recognition*, Beijing, China, August 2018. <https://doi.org/10.1109/ICPR.2018.8545479>
- [8] Wang, S., Lu, J., Gu, X., et al. (2015) Canonical Principal Angles Correlation Analysis for Two-View Data. *Journal of Visual Communication and Image Representation*, **35**, 209-219. <https://doi.org/10.1016/j.jvcir.2015.12.001>
- [9] Guangxia, L.I., Hoi, S.C.H. and Chang, K. (2010) Two-View Transductive Support Vector Machines.
- [10] Vapnik, V. and Vashist, A. (2009) A New Learning Paradigm: Learning Using Privileged Information. *Neural Networks: The Official Journal of the International Neural Network Society*, **22**, 544-557. <https://doi.org/10.1016/j.neunet.2009.06.042>
- [11] Tang, J., Tian, Y., Zhang, P. and Liu, X. (2017) Multiview Privileged Support Vector Machines. *IEEE Transactions on Neural Networks and Learning Systems*, **29**, 3463-3477.
- [12] Xiao, Y.S., Liu, B., Hao, Z.F. and Cao, L.B. (2014) A Similarity-Based Classification Framework for Multiple-Instance Learning. *IEEE Transactions on Cybernetics*, **44**, 500-515. <https://doi.org/10.1109/TCYB.2013.2257749>
- [13] Li, W., Dai, D.X., Tan, M.K., Dong, X. and Van Gool, L. (2016) Fast Algorithms for Linear and Kernel Svm+. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, 27-30 June 2016, 2258-2266. <https://doi.org/10.1109/CVPR.2016.248>
- [14] Farquhar, J.D.R., Hardoon, D.R., Meng, H., Shawe-Taylor, J. and Szedmak, S. (2005) Two View Learning: SVM-2K, Theory and Practice. *Advances in Neural Information Processing Systems 18, Neural Information Processing Systems, NIPS 2005*, Vancouver, British Columbia, Canada, 5-8 December 2005.
- [15] Hartigan, J.A. and Wong, M.A. (2013) A K-Means Clustering Algorithm. *Applied Statistics*, **28**, 100-108. <https://doi.org/10.2307/2346830>
- [16] Louhichi, S., Gzara, M. and Ben Abdallah, H. (2014) A Density Based Algorithm for Discovering Clusters with Varied Density. 2014 *World Congress on Computer Applications and Information Systems*, Hammamet, Tunisia, 17-19 January 2014, 1-6. <https://doi.org/10.1109/WCCAIS.2014.6916622>