

基于全局和局部标签相关性的 MIMLSVM 改进算法^①



李村合, 张振凯

(中国石油大学(华东)计算机与通信工程学院, 青岛 266580)
通讯作者: 张振凯, E-mail: zhangzk1992@163.com

摘要: 多示例多标记学习是用多个示例来表示一个对象, 同时该对象与多个类别标记相关联的新型机器学习框架. 设计多示例多标记算法的一种方法是使用退化策略将其转化为多示例学习或者是多标记学习, 最后退化为传统监督学习, 然后使用某种算法进行训练和建模, 但是在退化过程中会有信息丢失, 从而影响到分类准确率. MIMLSVM 算法是以多标记学习为桥梁, 将多示例多标记学习问题退化为传统监督学习问题求解, 但是该算法在退化过程中没有考虑标记之间的相关信息, 本文利用一种既考虑到全局相关性又考虑到局部相关性的多标记算法 GLOCAL 来对 MIMLSVM 进行改进, 实验结果显示, 改进的算法取得了良好的分类效果.

关键词: 多示例多标记; 局部性; 全局性; 退化; MIMLSVM; GLOCAL

引用格式: 李村合, 张振凯. 基于全局和局部标签相关性的 MIMLSVM 改进算法. 计算机系统应用, 2019, 28(4): 131-138. <http://www.c-s-a.org.cn/1003-3254/6832.html>

Improved MIMLSVM Algorithm Based on Global and Local Label Correlations

LI Cun-He, ZHANG Zhen-Kai

(School of Computer & Communication Engineering, China University of Petroleum, Qingdao 266580, China)

Abstract: Multi-Instance Multi-Label (MIML) learning is new machine learning framework where an example is described by multiple instances and associated with multiple classes of labels. A method of designing MIML algorithm is to identify its equivalence in the traditional supervised learning framework, using multi-instance learning or multi-label learning as the bridge, then using an algorithm for training and modeling. But in the degradation process, there will be the loss of information, then affecting the classification accuracy. This study improves the MIMLSVM algorithm by using a multi-label algorithm GLOCAL that considers both global and local correlations. The experimental results show that the improved algorithm has achieved sound classification results.

Key words: Multi-Instance Multi-Label (MIML); locally; globally; degradation; MIMLSVM; GLOCAL

在利用机器学习中的传统监督学习^[1]解决实际问题时, 常见的做法是先对真实的对象进行特征提取, 用一个特征向量来描述这个对象, 这样就得到了一个示例 (instance), 然后把示例与该对象所对应的类别标记 (label) 关联起来, 就得到了一个例子 (example). 在拥有了一个较大的例子集合之后, 就可以利用某种学习算

法来学得示例空间与标记空间之间的一个映射, 该映射可以预测未见示例的标记. 在待学习对象具有明确的、单一的语义时, 上面的学习框架已经取得了巨大的成功. 但是真实世界中的对象往往具有复杂的含义, 只用一个示例和一个标记来表达过于简单, 在刚开始的表示阶段就失去了很多重要的信息.

① 基金项目: 山东省自然科学基金项目 (ZR2014FQ018)

Foundation item: Natural Science Foundation of Shandong Province (ZR2014FQ018)

收稿时间: 2018-10-10; 修改时间: 2018-10-30; 采用时间: 2018-10-31; csa 在线出版时间: 2019-03-28

为了解决上述问题,多示例多标签^[2,3]框架应运而生.在该框架下一个对象用一组示例来表示,并且和一组类别标记相关联.比如一篇文章中的每个部分可以用一个示例来表示,而这篇文章可能同时属于“娱乐”、“旅游”、“经济”等不同的类别.多示例多标签已成功应用到了场景分类、基因功能注释、图像标注等领域,彰显了其强大的学习能力和潜力.

支持向量机(Support Vector Machines, SVM)^[4]是一种高效的分类识别方法,在解决高维模式识别问题呈现出其独特的优势.它通过寻找一个最优分类超平面来实现分类.支持向量机以统计学习理论中的VC维(Vapnik-Chervonenkis Dimension)理论^[5]以及结构风险最小原理为基础,最终的决策函数由少量支持向量决定,其数学形式简单,同时有效克服样本空间的高维性带来的计算复杂问题,已经在系统辨识、人脸检测及生物信息等领域中被用到.多示例多标记框架下的许多算法或直接使用了SVM或对其改造使其适应多示例多标记的学习,这些算法在学习过程中已经取得了不错的分类效果.在真实的学习问题中,标记之间通常具有局部标记相关性及全局相关性,如果仅仅考虑其一有可能会影响到实验分类的准确率,因此如何兼顾考虑二者是能否取得良好实验效果的关键.

本文的整体组织结构如下:第1部分介绍相关工作,第2部分提出改进的算法,第3部分给出了实验及结果,最后进行了总结.

1 相关工作

传统的监督学习框架是一种单示例单标记(Single-Instance Single-Label, SISL)学习框架,形象化的来说,另 x 为示例空间, y 为标记空间,则学习的任务是从数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$ 中学得函数 $f: x \rightarrow y$,其中 $x_i \in X$ 为一个示例, $y_i \in y$ 为示例 x_i 所属的类别标记^[1].

多示例学习(Multi-Instance Learning, MIL)^[2]中的每个对象是用一个示例包来表示,同时该对象只与一个类别标记相关联.给定数据集 $\{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$,目标是学得 $f: 2^x \rightarrow y$.其中 $x_i \subset x$ 为一组示例 $\{x_{i1}, x_{i2}, \dots, x_{in_i}\}$, $x_{ij} \in x (j = 1, 2, \dots, n_i)$, $y_i \in y$ 为与 x_i 的类别标记. n_i 为 x_i 中所含示例的个数.近年来,也出现了很多多示例学习算法,比如Diverse Density^[6]、EM-DD^[7]、k-近邻,决策树算法RELIC^[8]、神经网络算法RBF-MIP^[9]等.多示例学习也被广泛的应用于自然场景分类、基于内容的图像检索、WEB目录页面推荐以及机器人领域中^[10].

多标记学习(Multi-Label Learning, MLL)^[2]中一个

对象是用一个示例来表示,同时该对象与多个类别标记相关联.多标签学习近年来得到了广泛的研究.根据所使用的标签相关程度,它可以分为三类^[11]:一阶,二阶和高阶.对于第一组,不考虑标签相关性,并将多标签问题转换为多个独立的二元分类问题.一个众所周知的例子是二元相关(BR)算法^[12],它为每个标签独立地训练一个分类器.对于第二组,考虑成对标签关系.例如,校准标签排名(CLR)^[13]将多标签学习问题转化为成对标签排序问题.最后,对于第三组,所有其他标签对每个标签施加的影响都被考虑在内.例如,分类器链(CC)^[14]将多标签学习问题转换为二元分类问题链,将标注标签编码为特征.另一种考虑所有标签相关性的方法是通过学习一个潜在的标签空间来捕捉更高级别的标签语义.通常,这是通过标签矩阵的低秩分解^[15]获得的.最近,Yeh CK等人^[16]也提出了深度学习方法来学习联合特征和标签嵌入.这些方法与典型相关分析(CCA)高度相关^[17],它学习了一个潜在的子空间,以便将实例表示与相应的标签对齐.

要发挥多示例多标记框架的能力,要设计有效的学习算法.传统的监督学习、多示例学习以及多标记学习都可以看作是多示例多标记的退化表示形式,因此设计多示例多标记算法的一种思路是使用退化策略,将多示例多标记问题转化为多示例或者是多标记学习问题,进而再进一步转化为传统监督学习问题,最后使用传统的机器学习来解决相应问题.但是在退化的过程中会有重要信息丢失,比如标记之间的联系信息、示例与标记之间的联系信息等,进而影响到了最后的分类效果.为了避免退化策略中信息丢失问题的发生,衍生了另一种算法思路,即直接改造机器学习中相应的算法使其适应多示例多标签框架.

基于退化策略,周志华等人提出了MIMLBOOST算法^[3]和MIMLSVM算法^[3],基于正则化机制以及最大化间隔策略提出了D-MIMLSVM算法^[1]和M³MIML算法^[18].MIMLBOOST算法首先将多示例多标记样本 (X_i, Y_i) 转化为 $|y|$ 个多示例单标记样本 $\{([X_i, y], \phi[X_i, y]) | y \in Y\}$ ($|y|$ 是标记空间中标记的数目),其中,每一个示例 $[X_i, y]$ 都是由示例集合 X_i 和类别标记 y 拼接而成,当类别标记 $y \in Y_i$ 时, $\phi[X_i, y] = +1$,否则 $\phi[X_i, y] = -1$.转化完成后,多示例多标记问题已经变为多示例学习问题了,然后使用多示例学习中的MIBOOSTING算法^[19]对转化后的问题进行求解.MIMLSVM算法则是以多标记为桥梁,首先利用聚类将多示例多标记转化为多标记学习问题,再借助于多标记学习的算法进行求解.

MIMLBOOST 算法和 MIMLSVM 算法都是使用的退化方法,因此均有信息丢失。D-MIMLSVM 算法假设类别标记集合 Y 中共含有 T 个类别标记,同时假设分类系统由 T 个函数 $f = (f_1, f_2, \dots, f_T)$ 构成,每一个函数都是由示例空间到实值空间的映射,即 $f_t: 2^X \rightarrow R, t = 1, 2, \dots, T$,也就是为每一个标记建立一个分类器。该算法在训练集上定义了一个损失函数 V ,使用了某种定义在包上的多示例核函数^[20],最后通过优化 V ,利用 CCCP^[21,22]的迭代优化策略对其求解。M³MIML 算法假设分类系统包含 T 个线性模型,每个线性模型对应于一个可能的概念类,一个测试样本是否属于第 i 类是由其包含的所有示例在上述对应的某个线性模型上的最大输出决定的。M³MIML 基于最大化间隔策略,将二次规划问题分解为一系列相对简单的线性规划问题求解,最后利用 KKT 条件^[23]求解对偶问题获得模型参数。

以前的大多数研究集中在全局标签相关性上,但是,有时标签关联只能由局部数据子集所共享。为了缓解这个问题,使用局部相关性的多标签学习 (MLLOC)^[24]通过嵌入代码来扩展每个实例的特征表示,所述代码编码实例标签对局部标签相关性的影响。MLLOC 在利用局部相关性方面取得了成功。但是像摘要所提到的那样,有时局部和全局同样重要,而 MLLOC 仅仅可以处理局部标签并且无法处理缺失标签的情况。为了解决这个问题,提出了本文的算法。

2 MIMLSVM 算法及改进算法

2.1 MIMLSVM 算法

MIMLSVM 算法是以多标记学习为桥梁,将多示例多标记学习问题退化为传统监督学习问题求解。首先, MIMLSVM 算法将每个多示例多标记样本 (X_i, Y_i) 转化为一个单示例多标记样本 $(\phi(X_i), Y_i)$ 。其中,函数 ϕ 是基于构造性聚类^[25]将每个示例包 X_i 转化为一个示例 Z_i 。该聚类过程将集合 $\Lambda = \{X_1, \dots, X_m\}$ 中的每个包看作一个原子对象,基于 Hausdorff 距离度量包之间的距离并利用 k-medoids 算法将集合 Λ 划分为若干个聚类。此时 Z_i 的每一维即为包 X_i 与各聚类中心的距离,这样,多示例多标记样本集就转换成了单示例多标记样本集,此过程完成后, MIMLSVM 算法利用 MLSVM 算法^[26]对转换得到的多标记学习问题进行求解。MLSVM 算法将多标签学习问题分解为多个独立的二元分类问题,从而为每一个类别标记建立一个 SVM 分类器,事实上,示例 z_i 参与了每一个分类器的构建,示例 Z_i 对应

的标签集合 Y_i 包含 y 时,即 $y \in Y_i$ 时 $\phi(Z_i, y) = +1$, 否则 $\phi(Z_i, y) = -1$ 。至此,已经将多标记学习转化为了传统的监督学习,可以使用 SVM 进行训练分类了。对于未见标记的示例包,利用构造性聚类转化为单个的示例,然后交给训练出来的 $|y|$ 个分类器,每个分类器在该示例上会有不同输出,用输出为正的分类器对应的类别来标记该示例即可。但是注意到,如果未见标记的示例在每个分类器上的输出都为负值,那该示例将会变得不可分,即没有类别标记与之对应。为了避免这种情况的出现,这里采用 T-Criterion 准则^[27]来进行预测,即当所有分类器的输出均是负值时,用输出值“最大”的分类器对应的类别来标记该示例。

但是在退化过程中存在一个明显的问题,那就是为每一个类别标记建立分类器的时候,忽略了标记之间的相关性。前文也提到过,针对此也提出了一些改进算法,比较有效的是 MLLOC 算法,虽然此算法有一定效果,但是仅仅只是从局部标签相关性角度出发,忽略了全局相关性。这里使用多标记学习中的一种既考虑局部相关性又考虑全局相关性并且可以处理缺失标签的算法 GLOCAL^[28]对 MIMLSVM 算法进行改进。

2.2 MIMLSVM 算法的改进算法

GLOCAL 算法^[28]是 Yue Zh、Kwok JT 和 Zhou ZH 在研究多标记学习时提出的一种全新的多标记算法,通过学习潜在标签表示和优化标签流形,同时处理全标签和缺失标签情况,并且同时利用全局和局部标签相关性。GLOCAL 算法之所以能够取得成功,主要归咎于以下四点:(1)它使用标签矩阵的低秩结构来获得更紧凑和更抽象的潜在标签表示,这也为丢失标签恢复提供了自然的解决方案(2.2.1 中介绍);(2)它利用全局和局部标签相关性,因此标签分类器可以利用来自所有标签的信息(2.2.2 中介绍);(3)它直接从数据中学习标签相关性,而不需要通过在关联矩阵中进行普通且困难的人工分辨(2.2.3 中介绍);(4)将上述问题综合为一个联合学习问题,采用高效交替最优化策略进行优化(2.2.4 中介绍)。

2.2.1 基本模型

基本 GLOCAL 模型对标签矩阵应用低秩分解以获得潜在标签,并学习从特征空间到潜在标签的映射。因此,我们可以得到一个更紧凑和更抽象的潜在标签表示,它是密集的,实值的,并且是低维的。学习从特征空间到潜在标签空间的映射也比学习原始标签空间(稀疏,二进制值和更高维)更容易。此外,它直接提供

缺少标签恢复的解决方案.

具体来说, 我们使用:

$$\tilde{Y} \approx UV \quad (1)$$

将标签矩阵 \tilde{Y} 分解为两个低秩矩阵 U 和 V , 其中 V 代表潜在标签, U 代表原始标签如何与潜在标签相关联. 矩阵 U 和 V 可以通过最小化重构误差 $\|\tilde{Y} - UV\|_F^2$ 来获得.

为了将示例映射到潜在标签, 我们学习了一个矩阵 $W \in R^{d \times k}$, 这个 W 可以通过最小化平方损耗 $\|V - W^T X\|_F^2$ 获得, 其中 $X = [x_1, \dots, x_n] \in R^{d \times n}$ 是包含所有示例的矩阵. 随后, 针对示例 x 所预测的标签是 $\text{sign}(f(x))$, 其中 $f(x) = UW^T x$. 让 $f = [f_1, \dots, f_l]^T$, 其中 $f_j(x)$ 是 x 的第 j 个预测标签, 可以将所有 $x \in X$ 的 $f(x)$ 连接在一起成为 $F_0 = [f(x_1), \dots, f(x_n)] = UW^T X$. 结合低秩矩阵分解的重构误差最小化和从示例到潜在标签的线性映射的平方损失最小化, 我们得到了基本 GLOCAL 模型的以下优化问题:

$$\min_{u,v,w} \left\| \prod_{\Omega} (Y - UV) \right\|_F^2 + \lambda \|V - W^T X\|_F^2 + \lambda_2 R(U, V, W) \quad (2)$$

其中, $R(U, V, W)$ 是正则化参数, λ_1, λ_2 是平衡参数, 尽管问题 (2) 使用了平方损失, 但它可以被任何可微分损失函数代替.

2.2.2 利用全局和局部相关性

利用标签相关性对多标签学习至关重要, 在这里, 我们使用标签相关性来规范模型. 需要注意的是全局和局部相关性可以共存, 本节, 引入标签流形正则化项来兼顾考虑二者. 全局流形正则化项的基本思想是从实例级流形正则化项^[29]中改进而来. 具体而言, 两个标签越正相关, 相应分类器输出越接近, 反之亦然.

回想一下, 对 n 个示例的预测都存储在 $l \times n$ 的矩阵 F_0 中, 其第 i 行 $f_{i,:}$ 包含第 i 个标签的预测. 如果第 i 个和第 j 个标签更具正相关性, 则 $f_{i,:}$ 应该更加相似于 $f_{j,:}$, 反之亦然. 类似于实例级流形正则化项^[29,30], 标签流形正则项被定义为:

$$\sum_{i,j} [S_0]_{i,j} \|f_{i,:} - f_{j,:}\|_2^2 \quad (3)$$

其中, S_0 是 $l \times l$ 的全局标签相关矩阵. 如果标签 i 和 j 是正相关的, 那么 $[S_0]_{i,j}$ 也是正数. 通过最小化式 (3), $\|f_{i,:} - f_{j,:}\|_2^2$ 将变小. 令 D_0 为对角线 S_0 1 的对角矩阵, 其中 1 是 1 的向量表示. 在式 (3) 中, 流行正则化项可以等

价写成 $\text{tr}(F_0^T L_0 F_0)$ ^[31], 其中 $L_0 = D_0 - S_0$ 是 S_0 的 $l \times l$ 标签拉普拉斯矩阵.

由于局部之间的标签相关性可能不同, 因此我们引入局部流形正则化式. 假设数据集 X 被分为 g 个组 $\{X_1, \dots, X_g\}$, 其中 $X_m \in R^{d \times n_m}$ 具有 n_m 个示例. 令 Y_m 为 Y 对应于 X_m 的标签子矩阵, $S_m \in R^{l \times l}$ 为组 m 的局部标签相关矩阵, 与全局标签相关性相似, 我们支持分类器输出正 (负) 相关标签上相似 (或不相似), 并最小 $\text{tr}(F_m^T L_m F_m)$, 其中 L_m 是 S_m 的拉普拉斯矩阵, $F_m = UW^T X_m$ 是组 m 的分类器输出矩阵. 对于问题 (2) 加入全局和局部标签流形正则化式后, 得到以下优化后的:

$$\min_{U,V,W} \left\| \prod_{\Omega} (Y - UV) \right\|_F^2 + \lambda \|V - W^T X\|_F^2 + \lambda_2 R(U, V, W) + \lambda_3 \text{tr}(F_0^T L_0 F_0) + \sum_{m=1}^g \lambda_4 \text{tr}(F_m^T L_m F_m) \quad (4)$$

其中, $\lambda_1, \lambda_2, \lambda_3, \lambda_4$ 是平衡参数.

标签流形正则化之所以能兼顾二者或者说促进作用是因为全局标签相关性用拉普拉斯矩阵 L_0 编码, 局部标签相关性用 L_m 编码, 直观来说, 一个大的局部组对全局标签相关性的贡献更大, 特别的, 当使用余弦相似度计算 S_{ij} 时, $S_0 = \sum_{m=1}^g \frac{n_m}{n} S_m$. 通常, 当全局标签相关矩阵是局部标签相关矩阵的线性组合时, 相应的全局标签拉普拉斯矩阵也是具有相同组合系数的局部标签拉普拉斯矩阵的线性组合. 综上所述, 问题 (4) 可以被重写为:

$$\min_{U,V,W} \left\| \prod_{\Omega} (Y - UV) \right\|_F^2 + \lambda \|V - W^T X\|_F^2 + \sum_{m=1}^g \left(\frac{\lambda_3 n_m}{n} \text{tr}(F_0^T L_m F_0) + \lambda_4 \text{tr}(F_m^T L_m F_m) \right) + \lambda_2 R(U, V, W) \quad (5)$$

2.2.3 学习标签相关性

标签流形正则化成功是取决于好的标签相关矩阵 (或等价于一个好的拉普拉斯矩阵), 在多标签学习中, 一种基本的方法是用余弦距离计算两个标签之间的相关系数^[32]. 但是, 由于某些标签在训练数据中可能只有极少数正示例, 因此估计值可能会很嘈杂. 当标签可能丢失时, 这个估计值甚至会引起很大的误差, 因为观察到的标签分布可能与真实的标签分布有很大不同. 在本算法中, 没有直接指定任何相关度量或标签相关矩阵, 而是直接学习拉普拉斯矩阵, 这样省去了在关联矩

阵中进行普通且困难的人工分辨. 需要注意的是, 拉普拉斯矩阵是对称正定的, 因此, 对于 $m \in \{1, \dots, g\}$, 我们将 L_m 分解为 $Z_m Z_m^T$, 其中 $Z_m \in R^{l \times k}$. 为了简单起见, 将 k 设置为潜在表示 V 的维度. 因此, 学习拉普拉斯矩阵转换为学习 $Z = \{Z_1, \dots, Z_g\}$. 注意, 优化关于 Z_m 可能会导致平凡解 $Z_m = 0$. 为了避免这种情况, 我们添加约束条件, 即每个对角线元素 $Z_m Z_m^T$ 为 1, 这也使我们能够获得 L_m 的归一化拉普拉斯矩阵^[33].

令 $J = [J_{ij}]$ 为指标矩阵, 当 $(i, j) \in \Omega$ 时 $J_{ij} = 1$, 否则为 0. $\prod_{\Omega}(Y - UV)$ 可以重写为 Hadamard 乘积 $J \circ (Y - UV)$. 结合拉普拉斯矩阵的分解和 Z_m 的对角约束, 我们得到了最优化问题:

$$\begin{cases} \min_{U, V, W, Z} \|J \circ (Y - UV)\|_F^2 + \lambda \|V - W^T X\|_F^2 \\ + \sum_{m=1}^g \left(\frac{\lambda_3 n_m}{n} \text{tr}(F_0^T Z_m Z_m^T F_0) + \lambda_4 \text{tr}(F_m^T Z_m Z_m^T F_m) \right) \\ + \lambda_2 R(U, V, W) \\ \text{s.t. } \text{diag}(Z_m Z_m^T) = 1, m = 1, 2, \dots, g. \end{cases} \quad (6)$$

此外, 将在后续中使用标准正则化式 $R(U, V, W) = \|U\|_F^2 + \|V\|_F^2 + \|W\|_F^2$.

2.2.4 通过交替最小化进行优化学习

问题 (6) 可以通过交替最小化来解决. 这使得我们能够迭代的调整变量以找到满意的解决方案. 在每次迭代中, 我们用梯度下降来更新 $\{Z, U, V, W\}$ 中的一个变量, 并固定其他变量. 整个优化问题进而简化为几个简单的子问题. 具体来说, MANOPT 工具箱^[34]被用来在欧几里得空间上用线搜索实现梯度下降以更新 U, V, W 以及更新 Z 的流形, 详细更新过程将在下面讨论.

2.2.4.1 更新 Z_m

随着 U, V, W 的被固定, 式 (6) 变为

$$\begin{cases} \min_{Z_m} \frac{\lambda_3 n_m}{n} \text{tr}(F_0^T Z_m Z_m^T F_0) + \lambda_4 \text{tr}(F_m^T Z_m Z_m^T F_m) \\ \text{s.t. } \text{diag}(Z_m Z_m^T) = 1, m \in \{1, \dots, g\} \end{cases} \quad (7)$$

由于约束 $\text{diag}(Z_m Z_m^T) = 1$, 它没有闭式解, 我们用投影的梯度下降来解决它, 关于 Z_m 的梯度为 $\nabla Z_m = \frac{\lambda_3 n_m}{n} U W^T X X^T W U^T Z_m + \lambda_4 U W^T X_m X_m^T W U^T Z_m$, 为了满足约束条件 $\text{diag}(Z_m Z_m^T) = 1$, 每次更新后我们将 Z_m 的每一行投影到单位标准球上:

$$Z_{m,j} \leftarrow Z_{m,j} / \|Z_{m,j}\|, \text{ 其中 } Z_{m,j} \text{ 是 } Z_m \text{ 的第 } j \text{ 行.}$$

2.2.4.2 更新 V

随着 Z_m, U, W 的被固定, 式 (6) 变为:

$$\min_V \|J \circ (Y - UV)\|_F^2 + \lambda \|V - W^T X\|_F^2 + \lambda_2 \|V\|_F^2 \quad (8)$$

注意, V 的列彼此独立, 因此可以逐列得到 V . 设 j_i 和 v_i 分别为 J 和 V 的第 i 列, v_i 的优化问题可以写为:

$$\min_{v_i} \|\text{diag}(j_i) y_i - \text{diag}(j_i) U v_i\|^2 + \lambda \|v_i - W^T x_i\|^2 + \lambda_2 \|v_i\|^2$$

设置梯度关于 v_i 到 0, 我们得到以下 v_i 闭式解: $v_i = (U^T \text{diag}(j_i) U + (\lambda + \lambda_2) I)^{-1} (\lambda W^T x_i + U^T \text{diag}(j_i) y_i)$. 这里涉及为每个 i 计算逆矩阵, 如果计算代价很大, 我们可以在式 (8) 上执行梯度下降, 在式 (8) 上, 关于 V 的梯度下降为: $\nabla_V = U^T (J \circ (UV - Y)) + \lambda (V - W^T X) + \lambda_2 V$.

2.2.4.3 更新 U

随着 Z_m, V, W 的被固定, 式 (6) 变为:

$$\begin{cases} \min_U \|J \circ (Y - UV)\|_F^2 + \lambda_2 \|U\|_F^2 \\ + \sum_{m=1}^g \left(\frac{\lambda_3 n_m}{n} \text{tr}(F_0^T Z_m Z_m^T F_0) + \lambda_4 \text{tr}(F_m^T Z_m Z_m^T F_m) \right) \end{cases} \quad (9)$$

我们再一次利用梯度下降, 关于 U 的梯度为:

$$\begin{aligned} \nabla U &= (J \circ (UV - Y)) V^T + \lambda_2 U \\ &+ \sum_{m=1}^g Z_m Z_m^T U \left(\frac{\lambda_3 n_m}{n} W^T X_m W + \lambda_4 W^T X X^T W \right) \end{aligned}$$

2.2.4.4 更新 W

随着 Z_m, U, V 的被固定, 式 (6) 变为:

$$\begin{cases} \min_W \lambda \|V - W^T X\|_F^2 + \lambda_2 \|W\|_F^2 \\ + \sum_{m=1}^g \left(\frac{\lambda_3 n_m}{n} \text{tr}(F_0^T Z_m Z_m^T F_0) + \lambda_4 \text{tr}(F_m^T Z_m Z_m^T F_m) \right) \end{cases} \quad (10)$$

关于 W 的梯度为:

$$\begin{aligned} \nabla W &= \lambda X (X^T W - V^T) + \lambda_2 W \\ &+ \sum_{m=1}^g \left(\frac{\lambda_3 n_m}{n} X X^T + \lambda_4 X_m X_m^T \right) W U^T Z_m Z_m^T U \end{aligned}$$

MIMLSVM 算法是针对多示例多标记框架设计的算法, 在基于构造性聚类将每一个包转化为一个示例之后就变成了单示例多标记问题, GLOCAL 算法针对的是单示例多标签的分类问题, 因此可以使用 GLOCAL 算法对原来的 MIMLSVM 算法进行改进, 替代原来的 MLSVM 算法得到 MIMLSVM-GLOCAL 算法, 像前文提到的那样, GLOCAL 算法并没有仅仅从局部或者全局一方面出发, 而是两者都考虑在内去解决相关问题,

并且也可以在标签缺失的情况下去解决问题. MIMLSVM-GLOCAL 的算法描述如表 1 所示.

表 1 MIMLSVM-GLOCAL 算法

算法 1. MIMLSVM-GLOCAL 算法

1. 将每一个包看作是原子对象, 则示例空间可划分为 $\Gamma = \{X_i | i=1, 2, \dots, n\}$.
2. 在 Γ 上使用 k-medoids 聚类将 Γ 分为 k 份, 每一份的聚类中心用 $M_t (t=1, 2, \dots, k)$ 来表示.
3. 将多示例多标签样本 (X_u, Y_u) 转换为单示例多标记样本 (z_u, Y_u) , 其中 $z_u = (z_{u1}, z_{u2}, \dots, z_{uk}) = (d_h(X_u, M_1), d_h(X_u, M_2), \dots, d_h(X_u, M_k))$.
4. 得到训练集 $D_y = \{(z_u, \varphi(z_u, y)) | u=1, 2, \dots, n\} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$.

建立 SVM 模型, 使用 GLOCAL 算法求解 $f_y = GLOCALTrain(D_y, J)$.

- 4.1 初始化 U, V, W, Z
- 4.2 repeat
- For $m=1, \dots, g$
- //学习标签相关性
- 固定 V, U, W , 通过解 (7) 来更新 Z_m .
- End for
- //学习潜在标签
- 固定 U, W, Z , 通过解 (8) 来更新 V ,
- //学习从原始标签到潜在标签的映射
- 固定 V, W, Z , 通过解 (9) 来更新 U ,
- //学习从示例到潜在标签的映射
- 固定 U, V, Z , 通过解 (10) 来更新 W ,
- Until 收敛

5. 对于未标记的样本集 X , $Y^* = \left\{ \arg \max_{y \in L} f_y(z^*) | f_y(z^*) < 0, \forall y \in L \right\} \cup \{y | f_y(z^*) \geq 0, y \in L\}$, 其中, $z^* = (d_h(X^*, M_1), d_h(X^*, M_2), \dots, d_h(X^*, M_k), U, V, W, Z)$.

3 实验

为了检验 MIMLSVM-GLOCAL 算法的分类效果, 将 MIMLSVM-GLOCAL 算法和其他的多示例多标记算法 MIMLBOOST、MIMLSVM、MIMLSVM+ 进行了比较. MIMLBOOST 算法以多示例为桥梁, 首先将多示例多标签样本转化为多示例单标签样本, 然后使用多示例学习中的 MIBOOSTING 算法对转化后的问题进行求解. MIMLSVM 算法则是以多标签为桥梁, 首先利用聚类将多示例多标签问题转化为多标签问题, 再借助于多标签学习中的 MLSVM 算法进行求解. MIMLSVM+ 算法利用退化策略, 将多示例转化成单示例, 再把多标签分解成一系列的两类分类问题. 实验中使用的 SVM 均使用高斯核函数, MIMLBOOST、MIMLSVM 算法中的参数根据文献[3]中的实验设置为最优, MIMLSVM+ 算法中的参数根据文献[1]中的实验设置为最优, GLOCAL

算法根据文献[28]中的实验设置为最优.

这里我们使用周志华等人提供的图像样本集^[3]和文本样本集^[19]进行实验. 实验中的算法均使用相同的样本集和测试集, 均采用 10 折交叉验证. 其中图像样本集由 2000 个自然场景图片构成, 分为 5 个类别, 分别是沙漠、山、海、日落和树. 属于一个以上的类(如山+日落)的样本的数目约占数据集的 22% 左右, 许多组合类(如山+海+树)约占 0.75% 左右, 单个标签的样本数目约占 77% 左右. 平均来说, 每个样本都与大约 1.24 个类标签相关联, 如表 2 所示.

表 2 场景样本集数据特征

标记集合	图片数量	标记集合	图片数量
沙漠	340	山+日落	19
山	268	山+树	106
海	341	海日落	172
日落	213	海+树	14
树	378	日落+树	28
沙漠+山	19	沙漠+山+日落	1
沙漠+海	5	沙漠+日落+树	3
沙漠+日落	21	山+海+树	6
沙漠+树	20	山+日落+树	1
山海	38	海+日落+树	4

每张图片通过 SBN 方法用包含 9 个示例、每个示例为 15 维的特征向量的示例包来表示. 文本数据集来自 Reuters 数据集, 此样本集被广泛利用, 包含 7 个类别的 2000 个文本样本, 这 2000 个样本是由去除没有标记和没有正文的文本, 再随机去除一些只有一个类别标记的文本后得到的. 具有多个标记的文本样本约占 15%, 平均每个文本样本具有 1.15 ± 0.37 个类别标记. 通过使用滑动窗口^[27]技术将每一篇文档用一个包来表示, 每个包中包含一组数量不等的特征向量, 每个特征向量都是 243 维的并且代表了这篇文档的某个部分. 本实验采用 10 折交叉验证的方式, 每次从 2000 个数据集中随机的抽取 1500 个数据作为训练集, 其余的 500 个数据作为测试集, 重复 10 次实验之后求得实验的平均值以及方差. 本实验中使用的场景样本集和文本样本集, 其结构特征如表 3 所示.

对于评级指标, 我们采用采用基本样本的也是广泛使用的 hamming loss、one-error、coverage、ranking loss 和 average precision^[1]来衡量学习效果. 简单来说, 对于 hamming loss、one-error、coverage 和 ranking loss 的值越小说明算法效果越好; 对于 average precision

的值越大说明算法效果越好。表4和表5分别显示了MIMLBOOST、MIMLSVM、MIMLSVM+和MIMLSVM-GLOCAL算法在场景样本集和文本样本集上的分类效果。

从表4和表5的实验结果可以看出,无论在场景样本集还是文本样本集,MIMLSVM-GLOCAL算法的性能都要高于其他算法,取得了良好的性能,这主要得益于文章2.2节所论述的优势,此外,全局标签流形提供有关标签如何作为一个整体相关的信息,并帮助学习少数标签,如果少数标签与其他标签正相关(相反,负相关),我们可以鼓励其标签分类器输出与其他标签

的输出更相似(相反)。局部标签流形还允许标签分类器的局部适应。拉普拉斯矩阵的学习发现了最适合全局和局部数据子集的标签相关性,并避免了手动指定标签相关性的常见任务和困难任务。从总体上看,MIMLSVM、MIMLBOOST、MIMLSVM+和MIMLSVM-GLOCAL在文本样本集上分类的各项指标明显优于场景样本集,这可能与样本集内部的结构有一定的关系,文本样本集示例的维数远远多于场景样本集,文本样本集可以更加准确的来表示对象,因此造成了两个样本集上的结果差异较大。

表3 文本样本集和场景样本集特征

样本集	样本数	类别数	包中示例数		示例维数	平均类别	训练集数目	测试集数目
			最小	最大				
场景	2000	5	9	9	9	1.24±0.44	1000	1000
文本	2000	7	2	26	243	1.15±0.37	1000	1000

表4 场景样本集实验结果

评价指标	MIMLBOOST	MIMLSVM	MIMLSVM ⁺	MIMLSVM-GLOCAL
hamming loss↓	0.1958±0.0232	0.1723±0.0131	0.2235±0.0107	0.1526±0.0137
one-error↓	0.6948±0.0310	0.5320±0.0412	0.4648±0.0430	0.4042±0.0228
coverage↓	1.8412±0.0560	1.5761±0.0910	1.4418±0.1037	1.1954±0.1021
ranking loss↓	0.2410±0.0187	0.2378±0.0169	0.2346±0.0140	0.2048±0.0236
average precision↑	0.6291±0.0352	0.6578±0.0232	0.7304±0.0138	0.8081±0.0251

表5 文本样本集实验结果

评价指标	MIMLBOOST	MIMLSVM	MIMLSVM ⁺	MIMLSVM-GLOCAL
Hamming loss↓	0.1970±0.0228	0.1021±0.0252	0.0958±0.0265	0.0446±0.0128
one-error↓	0.5210±0.0197	0.3215±0.0508	0.1851±0.0644	0.0831±0.0332
coverage↓	1.7891±0.1532	1.2162±0.2566	0.6121±0.0514	0.4891±0.0921
ranking loss↓	0.2644±0.0225	0.2271±0.0312	0.1542±0.0032	0.0682±0.0347
average precision↑	0.6512±0.0246	0.8483±0.0248	0.9412±0.0108	0.9725±0.0157

4 结束语

多示例多标记框架下的MIMLSVM算法已经取得了不错的分类效果,针对其第二步中由于忽略标签相关性而导致信息丢失的问题在之前也提出了一些解决方案,但都考虑的不周全,与以前的工作相比,本文将一种既考虑到全局又考虑到局部标签相关性的GLOCAL算法融入到了MIMLSVM算法中,整体提升了算法的性能,提高了分类准确率,取得了良好的实验效果。

参考文献

1 Zhou ZH, Zhang ML, Huang SJ, *et al.* Multi-instance multi-label learning. *Artificial Intelligence*, 2012, 176(1): 2291–2320. [doi: 10.1016/j.artint.2011.10.002]

2 Zhou ZH, Zhang ML, Huang SJ, *et al.* MIML: A framework for learning with ambiguous objects. arXiv: 0808.3231v2, 2008.

3 Zhou ZH, Zhang ML. Multi-instance multi-label learning with application to scene classification. *Proceedings of the 19th International Conference on Neural Information Processing Systems*. Canada. 2006. 1609–1616

4 丁世飞, 齐丙娟, 谭红艳. 支持向量机理论与算法研究综述. *电子科技大学学报*, 2011, 40(1): 2–10.

5 Kwok JT, Zhou ZH, Xu L. *Machine learning*. Kacprzyk J, Pedrycz W. Springer Handbook of Computational Intelligence. Berlin, Heidelberg: Springer, 2015. 495–522.

6 Maron O, Lozano-Pérez T. A framework for multiple-instance learning. *Proceedings of 1997 Conference on Advances in Neural Information Processing Systems*. Denver, CO, USA. 1998. 570–576.

- 7 Zhang Q, Goldman SA. EM-DD: An improved multiple-instance learning technique. Proceedings of the 14th International Conference on Neural Information Processing Systems: Natural and Synthetic. Vancouver, Canada. 2001. 1073–1080.
- 8 Ruffo G. Learning single and multiple instance decision trees for computer security applications[Ph. D. thesis]. Torino, Italy: University of Turin, 2000.
- 9 Zhang ML, Zhou ZH. Adapting RBF neural networks to multi-instance learning. Neural Processing Letters, 2006, 23(1): 1–26. [doi: [10.1007/s11063-005-2192-z](https://doi.org/10.1007/s11063-005-2192-z)]
- 10 刘大有. 知识科学中的基本问题研究. 北京: 清华大学出版社, 2006.
- 11 Zhang ML, Zhou ZH. A review on multi-label learning algorithms. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819–1837. [doi: [10.1109/TKDE.2013.39](https://doi.org/10.1109/TKDE.2013.39)]
- 12 Boutell MR, Luo JB, Shen XP, *et al.* Learning multi-label scene classification. Pattern Recognition, 2004, 37(9): 1757–1771. [doi: [10.1016/j.patcog.2004.03.009](https://doi.org/10.1016/j.patcog.2004.03.009)]
- 13 Fürnkranz J, Hüllermeier E, Mencia E, *et al.* Multilabel classification via calibrated label ranking. Machine Learning, 2008, 73(2): 133–153. [doi: [10.1007/s10994-008-5064-8](https://doi.org/10.1007/s10994-008-5064-8)]
- 14 Read J, Pfahringer B, Holmes G, *et al.* Classifier chains for multi-label classification. Machine Learning, 2011, 85(3): 333–359. [doi: [10.1007/s10994-011-5256-5](https://doi.org/10.1007/s10994-011-5256-5)]
- 15 Huang KH, Lin HT. Cost-sensitive label embedding for multi-label classification. Machine Learning, 2017, 106(9–10): 1725–1746. [doi: [10.1007/s10994-017-5659-z](https://doi.org/10.1007/s10994-017-5659-z)]
- 16 Yeh CK, Wu WC, Ko WJ, *et al.* Learning deep latent space for multi-label classification. Proceedings of the 31st AAAI Conference on Artificial Intelligence. San Francisco, CA, USA. 2017. 2838–2844.
- 17 Hastie T, Tibshirani R, Friedman J. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. New York: Springer, 2001.
- 18 Zhang ML, Zhou ZH. M3MIML: A maximum margin method for multi-instance multi-label learning. Proceedings of the 2008 8th IEEE International Conference on Data Mining. Pisa, Italy. 2008. 688–697.
- 19 Xu X, Frank E. Logistic regression and boosting for labeled bags of instances. Proceedings of the 8th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Sydney, Australia. 2004. 272–281.
- 20 Gärtner T, Flach PA, Kowalczyk A, *et al.* Multi-instance kernels. Proceedings of the 19th International Conference on Machine Learning. San Francisco, CA, USA. 2002. 179–186.
- 21 Cheung PM, Kwok JT. A regularization framework for multiple-instance learning. Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh, PA, USA. 2006. 193–200.
- 22 Smola AJ, Vishwanathan SVN, Hofmann T. Kernel methods for missing variables. Proceedings of the 10th International Workshop on Artificial Intelligence and Statistics. Barbodos. 2005. 325–332.
- 23 Cristianini N, Shawe-Taylor J. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge, UK: Cambridge University Press, 2000.
- 24 Huang SJ, Zhou ZH. Multi-label learning by exploiting label correlations locally. Proceedings of the 26th AAAI Conference on Artificial Intelligence. Toronto, Ontario, Canada. 2012. 949–955.
- 25 Zhou ZH, Zhang ML. Solving multi-instance problems with classifier ensemble based on constructive clustering. Knowledge and Information Systems, 2007, 11(2): 155–170. [doi: [10.1007/s10115-006-0029-3](https://doi.org/10.1007/s10115-006-0029-3)]
- 26 Boutell MR, Luo JB, Shen XP, *et al.* Learning multi-label scene classification. Pattern Recognition, 2004, 37(9): 1757–1771. [doi: [10.1016/j.patcog.2004.03.009](https://doi.org/10.1016/j.patcog.2004.03.009)]
- 27 Andrews S, Tsochantaridis I, Hofmann T. Support vector machines for multiple-instance learning. Advances in Neural Information Processing System. 2002. 577–584.
- 28 Zhu Y, Kwok JT, Zhou ZH. Multi-label learning with global and local label correlation. IEEE Transactions on Knowledge and Data Engineering, 2018, 30(6): 1081–1094. [doi: [10.1109/TKDE.2017.2785795](https://doi.org/10.1109/TKDE.2017.2785795)]
- 29 Belkin M, Niyogi P, Sindhvani V. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. The Journal of Machine Learning Research, 2006, 7: 2399–2434.
- 30 Melacci S, Belkin M. Laplacian support vector machines trained in the primal. The Journal of Machine Learning Research, 2011, 12: 1149–1184.
- 31 Luo DJ, Ding C, Huang H, *et al.* Non-negative Laplacian embedding. Proceedings of the 2009 9th IEEE International Conference on Data Mining. Miami, FL, USA. 2009. 337–346.
- 32 Wang H, Huang H, Ding C. Image annotation using multi-label correlated green's function. Proceedings of the 2009 IEEE 12th International Conference on Computer Vision. Kyoto, Japan. 2009. 2029–2034.
- 33 Chung FRK. Spectral Graph Theory. Providence, RI: American Mathematical Society, 1997.
- 34 Boumal N, Mishra B, Absil PA, *et al.* Manopt, a Matlab toolbox for optimization on manifolds. The Journal of Machine Learning Research, 2014, 15(1): 1455–1459.