软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2014,25(9):1967−1981 [doi: 10.13328/j.cnki.jos.004634]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel/Fax: +86-10-62562563

多标记分类和标记相关性的联合学习*

何志芬[1,2], 杨 明[1,2], 刘会东[2]

[1](南京师范大学 数学科学学院,江苏 南京 210023)

[2](南京师范大学 计算机科学与技术学院,江苏 南京 210023)

通讯作者: 杨明, E-mail: m.yang@njnu.edu.cn, http://www.njnu.edu.cn

摘 要: 提出了多标记分类和标记相关性的联合学习(JMLLC),在 JMLLC 中,构建了基于类别标记变量的有向条件依赖网络,这样不仅使得标记分类器之间可以联合学习,从而增强各个标记分类器的学习效果,而且标记分类器和标记相关性可以联合学习,从而使得学习得到的标记相关性更为准确.通过采用两种不同的损失函数:logistic 回归和最小二乘,分别提出了 JMLLC-LR(JMLLC with logistic regression)和 JMLLC-LS(JMLLC with least squares),并都拓展到再生核希尔伯特空间中.最后采用交替求解的方法求解 JMLLC-LR 和 JMLLC-LS.在 20 个基准数据集上基于 5 种不同的评价准则的实验结果表明,JMLLC 优于已提出的多标记学习算法.

关键词: 多标记学习;多标记分类;标记相关性;条件依赖网络;再生核希尔伯特空间;交替求解

中图法分类号: TP181

中文引用格式: 何志芬,杨明,刘会东.多标记分类和标记相关性的联合学习.软件学报,2014,25(9):1967−1981. http://www.jos.org.cn/1000-9825/4634.htm

英文引用格式: He ZF, Yang M, Liu HD. Joint learning of multi-label classification and label correlations. Ruan Jian Xue Bao/ Journal of Software, 2014,25(9):1967−1981 (in Chinese). http://www.jos.org.cn/1000-9825/4634.htm

Joint Learning of Multi-Label Classification and Label Correlations

HE Zhi-Fen[1,2], YANG Ming[1,2], LIU Hui-Dong[2]

[1](School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China)

[2](School of Computer Science and Technology, Nanjing Normal University, Nanjing 210023, China)

Corresponding author: YANG Ming, E-mail: m.yang@njnu.edu.cn, http://www.njnu.edu.cn

Abstract: In this paper, joint learning of multi-label classification and label correlations (JMLLC) is proposed. In JMLLC, a directed conditional dependency network is constructed based on class label variables. This not only enables joint learning of independent label classifiers to enhance the performance of label classifiers, but also allows joint learning of label classifiers and label correlations, thereby making the learned label correlations more accurate. JMLLC-LR (JMLLC with logistic regression) and JMLLC-LS (JMLLC with least squares), are proposed respectively by adopting two different loss functions: logistic regression and least squares, and are both extended to the reproducing kernel Hilbert space (RKHS). Finally, both JMLLC-LR and JMLLC-LS can be solved by alternating solution approaches. Experimental results on twenty benchmark data sets based on five different evaluation criteria demonstrate that JMLLC outperforms the state-of-the-art MLL algorithms.

Key words: multi-label learning; multi-label classification; label correlations; conditional dependency network; reproducing kernel Hilbert space; alternating solution

多标记学习是机器学习、模式识别等领域的研究热点之一.在多标记学习框架中,每个样本由一个特征向量表示,但可能同时隶属于多个类别标记,其目标是通过学习给定的多标记训练集有效地预测未知样本所属的

* 基金项目: 国家自然科学基金(61272222, 61003116); 江苏省自然科学基金(BK2011782, BK2011005)
收稿时间: 2014-01-29; 修改时间: 2014-04-22; 定稿时间: 2014-06-09

类别标记集合.然而,在传统的监督学习(也称为单标记学习)框架中,每个样本由一个示例表示,但只隶属于一个类别标记(不管标记是两个还是多个).当每个样本只与一个类别标记相关时,多标记学习问题将退化为单标记学习问题.因此,单标记学习问题本质上是多标记学习问题的一种特殊情况[1].

多标记学习概念的提出,源于研究文本分类时所遇到的歧义性问题[2].在文本分类问题中,一篇文档可能同时与多个预先定义的主题相关[2,3],例如"篮球赛"和"体育".近年来,多标记学习问题得到了广泛的关注[1−26].在现实世界中,多标记学习问题也涉及其他研究领域,例如自动图像标注[4]、视频自动标记[5]以及生物信息学[3,6]等.在自动图像标注中,每张图像可能同时标注为多个语义概念类[4],例如"草地"、"树木"和"建筑"等.在视频自动标记中,每段视频片段可标记为多个类别[5],例如"表演"、"唱歌"和"跳舞".在生物信息学中,每段基因序列可能具有多种功能[3,6],例如"转录"、"蛋白质合成"以及"新陈代谢".在情感分类中,每个音乐片段可能包含多种情感,例如"欢快的"和"轻松的".

近十几年来,许多学者致力于多标记学习问题的研究,提出了大量的多标记学习算法,并成功地应用于各个研究领域(如文本分类、图像和视频自动标记以及生物信息学等),例如基于 Boosting 的多标记文本分类算法(BoosTexter)[2]、多标记分类问题的核方法(RankSvm)[6]、多标记懒惰学习方法(MLkNN)[7]、校准标记排序算法(CLR)[8]以及集成的多标记分类器链算法(ECC)[9]等.一种最直接、最简单的方法是将多标记学习问题分成多个独立的二分类问题[4],其中每个对应一个标记,将属于该标记的样本看成正类,否则看成负类.对于未知样本,根据其在所有二分类器上的输出结果来预测所属的标记集合.该方法简单,但没有考虑标记之间的相关性.实际上,对于某个标记来说,如果充分利用其他相关的标记信息,可能更有利于其学习,尤其是当没有足够的训练样本时,标记之间的相关性可提供额外的有用信息[10].例如,一幅图像标记了"山"和"水",则很可能也标记为"树";一篇文档隶属于"奥斯卡"和"明星",则很可能也隶属于"娱乐";一段音乐标记为"欢乐的",则不太可能标记为"悲伤的".因此,如何有效地利用标记之间的相关性,是当前多标记学习问题的一个重要研究问题[1,10,11].

许多多标记学习算法试图利用标记之间的相关信息来提高多标记学习系统的泛化性能.例如,文献[3,6,8]考虑了标记对之间的相关性有效地提高系统的泛化性能.但是在真实世界中,一个标记可能与多个标记同时相关.文献[9]考虑了所有可能的标记之间的相关性,即对每个标记的学习都考虑了其他相关标记的影响.文献[12]考虑了随机标记子集之间的相关性.大多数多标记学习算法将标记之间的相关性作为先验知识或是计算标记的共现性,而标记之间的相关性可能事先很难准确获得.文献[11]提出多标记假设重用(MAHR)算法,可以自动地挖掘和利用标记之间的相关性,通过重用权重计算出标记之间的相关值,同时也发现标记之间的相关性是不对称的.据我们所知,很少有学者将多标记分类和标记相关性进行联合学习.

因此,本文提出了多标记分类和标记相关性的联合学习(joint learning of multi-label classification and label correlations,简称 JMLLC).首先,为每个标记构建二分类器,每个标记不仅依赖于输入特征向量而且也依赖于其他标记变量;其次,构建了多标记分类和标记相关性的联合学习模型,同时,根据选择不同的损失函数(即 logistic 回归损失函数和最小二乘损失函数),分别推导出 JMLLC-LR 和 JMLLC-LS 算法,并首先在原始特征空间中学习,然后拓展到再生核希尔伯特空间中;最后,JMLLC-LR 和 JMLLC-LS 模型都可转化为凸优化问题,并且可以采用交替迭代求解的方法进行模型求解.本文的主要贡献如下:

1) 构建了多标记分类和标记相关性的联合学习模型,自动挖掘和利用了高阶非对称的标记相关性,丰富了多标记学习和标记相关性等问题的研究;
2) 模型拓展性强,可以选择不同的凸损失函数,而且最终的优化问题可以转化为凸优化问题,并可以通过交替求解的方法进行模型求解;
3) 模型不仅可以在原始空间中学习,同时也可以拓展到再生核希尔伯特空间中学习.

1 相关工作

1.1 多标记学习

已提出的多标记学习算法大致可以分成两大类[1,12,13]:问题转化方法(problem transformation methods,简称

PTM)和算法改编方法(algorithm adaptation methods,简称 AAM).

1.1.1 问题转化方法(PTM)

PTM 将多标记学习问题转化为其他已知的学习问题,例如两类问题、多类问题和标记排序问题等.

BR(binary relevance)方法[4]是将多标记学习问题转化为若干个独立的二分类学习问题,该方法简单,而且每个二分类器可以单独学习,因此可以并行实现,但是忽略了标记之间的相关信息,系统的性能可能只达到次优而未达到最优.CC(classifier chains)[9]的基本思想是:将多标记学习问题转化为基于 BR 方法的分类器链,其中在分类器链中,后面分类器的构建是建立在前面的分类器基础上.该方法考虑了标记之间的相关性,实现了较高的预测性能,时间复杂度低,同时也保留了 BR 方法的优点.然而,链是随机排列的,随机地考虑了标记之间的相关性;而且当第 1 个分类器预测性能不好时,误差的影响可能随着链进行传播.为了克服这些不足,提出了多标记分类器链集成算法(ensembles of classifier chains,简称 ECC)[9].

LP(label powerset)方法直接将多标记学习问题转化为多类学习问题:首先,将训练集中存在的所有不同的类别标记子集进行二进制编码,每个编码值看成不同的类别值,即多标记数据集转化为多类数据集;然后,训练多类分类器.当给定一个未知样本时,首先根据训练得到的多类分类器对其进行预测;然后将该预测值转化为二进制编码,从而得到其所属的类别标记集合.LP 方法简单,但主要不足有:1) 当类别标记个数很多时,转化为多类数据集后,相应的新类别值个数会很多,从而导致有些新的类别值只有少量的训练样本以及训练时间开销大;2) 难以预测训练集以外的类别标记集合.为了保留 LP 方法的优点同时又克服其不足,提出了随机 k 标记集(random k-labelsets,简称 RAkEL)[12]算法.RAkEL 的基本思想是,将多标记学习问题转化为集成的多类学习问题:首先,从初始的类别标记集中随机地选择 k 个标记子集;然后采用 LP 方法,学习得到一个多类分类器;最后,建立一个集成的 LP 模型,通过阈值法或投票法预测未知样本的类别标记集合.

另一种被广泛应用的 PTM 方法是 LR(label ranking)[14],其基本思想是:通过标记成对比较,将多标记学习问题转化为标记排序问题.在为每个标记对(y_l,y_k)构建两类分类器时,将属于类别标记 y_l 但不属于类别标记 y_k 的样本看成是正类样本,将属于类别标记 y_k 但不属于类别标记 y_l 的样本看成是负类样本,忽略其他的样本.给定一个未知样本,对每个二分类器的预测值进行投票,通过阈值法将排序后的投票结果划分为该样本的相关标记和不相关标记.该方法的主要难点在于:如何确定阈值来尽可能正确地估计样本所属的类别标记集合.为了解决这个问题,文献[8]提出了校准的标记排序算法(calibrated label ranking,简称 CLR).与 LR 相比,给每个样本的类别标记集添加一个额外的虚拟标记 y_v,将其作为每个样本的相关标记和不相关标记的一个自然划分点;同时,也要将每个标记与虚拟标记进行成对比较,在对某个标记对(y_l,y_v)构建两类分类器时,将属于类别标记 y_l 的样本看成正类样本,否则看成负类样本.对于给定的未知样本,将所有二分类器的预测结果进行投票,然后排序,将那些投票次数大于虚拟标记 y_v 的类别标记看成该样本的相关标记,否则看成不相关标记.

1.1.2 算法改编方法(AAM)

AAM 是直接设计多标记学习算法处理多标记数据,即改编一些著名的算法来直接处理多标记数据.

文献[2]提出了多标记文本分类算法(BoosTexter),是著名的集成算法 Adaboost 的拓展.文献[2]采用了两种不同的方法来拓展 Adaboost,包括 Adaboost.MH 和 Adaboost.MR,其中:Adaboost.MH 目的是为了最小化汉明损失;而 Adaboost.MR 的目标是最小化排序损失,尽量使相关的标记排在前面.

RankSvm 算法[6]将经典的支持向量机(SVM)推广到多标记学习问题中.在 RankSvm 中,为每个类别标记构建一个 SVM 分类器,其中,经验损失项为排序损失.该方法利用排序损失考虑每个样本的相关标记和不相关标记,且目标优化问题可以转化为一个二次规划问题.由于需要计算大量的变量,故训练时间开销比较大.文献[15]基于 SVM 设计并实现了一个比 RankSvm 更高效的多标记分类算法(Rank-CVM).文献[16]通过引入近似的排序损失作为经验损失项,将传统的两类 SVM 拓展到多标记分类中,提出了拓展的一对多多标记 SVM 算法(OVR-ESVM).文献[17]在 RankSvm 模型的基础上增加了未标记样本的损失项以及未标记样本的预测值的均值与已标记样本的真实标记均值相等的约束项,提出了一种归纳的多标记分类算法(iMLCU).

文献[7]提出了一种懒惰的多标记学习方法(MLkNN),是由传统的 k 近邻算法衍生出来的.对于每个测试样

本,首先确定它的 k 个近邻样本;然后根据 k 个近邻样本的标记信息,用最大后验概率(MAP)准则预测它的类别标记集合.该方法简单,时间复杂度低;但独立地计算每个标记的先验概率和后验概率,没有考虑标记之间的相关信息.为此,文献[18]针对 MLkNN 中存在的不足,提出了一种新型多标记懒惰学习算法(IMLLA).该方法充分利用了训练数据的分布信息以及多个标记之间的相关信息.

文献[19]设计了基于朴素贝叶斯的多标记分类算法(MLNB),同时将特征选择机制加入到 MLNB 算法中.首先,用基于主成分分析(principal component analysis,简称 PCA)的特征提取技术,将原始特征空间投影到低维的特征空间中;然后,用基于遗传算法(genetic algorithm,简称 GA)的特征选择技术,选择最合适的特征子集来提高 MLNB 的预测精度.文献[20]将贝叶斯学习技术应用到多标记学习中,提出了用贝叶斯网络结构模型标记依赖性的多标记学习方法(LEAD).LEAD 是一个两阶段学习算法,首先通过贝叶斯网络学习标记依赖性,然后再进行分类.此外,文献[3]提出了基于后向传播神经网络的多标记学习算法(BPMLL);文献[21]提出了用决策树技术处理多标记数据的算法 ML-DT;文献[22]采用最大熵原理来处理多标记数据,提出了 CML 算法.

1.2 标记相关性

在多标记学习中,每个样本可能同时隶属于多个类别标记,标记之间的相关信息可能会为多标记问题的学习提供额外的有用信息,从而有利于提升多标记学习系统的性能.因此,如何有效地利用标记之间的相关性,是当前研究多标记学习问题的核心内容.

根据多标记学习算法中考虑的标记之间的相关性的阶,将存在的方法大致可以划分为 3 类[1]:

1) 一阶策略:每个标记独立地处理,完全不考虑标记之间的相关性.例如,将多标记学习问题转化为多个独立的二分类问题[4].该方法简单,但由于忽略了标记之间的相关信息,学习算法的性能可能没有达到最优;
2) 二阶策略:考虑了标记之间的成对关系.例如,将多标记学习问题转化为标记排序问题[8,14].然而在一些实际应用中,标记之间的相关性可能会超过二阶;
3) 高阶策略:考虑了每个标记与其他标记之间的相关性或者随机的标记子集之间的相关性.该方法挖掘了很强的标记相关性,但是可能会导致计算更复杂.例如,ECC[9]和 RAkEL[12]算法.

文献[23]假设多个标记共享一个子空间,标记之间的相关性通过多个标记所共享的低维子空间来获得,但并未显式地描述标记之间的相关性.很多利用标记之间相关性的方法都是假定标记之间的相关性是对称的或者事先确定的.然而,标记之间的相关性并非一定是对称的,且很难事先就确定.因此,文献[11]提出了标记之间的相关性是非对称的且可以正相关也可以负相关,并通过学习求解出了标记相关性矩阵,但求出的标记相关性矩阵并未用于最后的模型预测中.文献[24]采用标记协方差矩阵来显示地描述标记之间的相关性(可以正相关、不相关或负相关),而且标记相关性可以从数据中自动学习而不用事先确定,并可以同时学习标记相关性和模型参数,但是只能求出成对标记相关性.在文献[25,26]中,构建了基于类别标记变量的条件依赖网络,其中每个节点对应一个标记,且每个标记将其他的标记变量和输入特征变量看成为其父节点,用结构化的标记依赖特征来挖掘标记之间的相关性.然而对于每个测试样本,其在第 l 个概率预测函数上的预测值依赖于其他标记变量的值.而实际上,其他变量的值并非事先知道,不过可以采用 Gibbs 采样方法进行推理,但过程繁琐且非常耗时.

因此,本文构建基于类别标记变量的条件依赖网络,每个标记的学习依赖于输入变量和其他的类别标记;然后构建了多标记分类和标记相关性的联合学习模型,该模型最终可转化为凸优化问题,可采用交替求解的方法进行求解,且预测过程简单.与文献[23]相比,本文不需要假设多个标记共享一个子空间,可直接在原始空间中自动求解标记相关性,并显式地描述了标记之间的相关性.与文献[11]相比,本文的算法将求出的高阶非对称的标记相关性用于模型的预测中.与文献[24]一样,都是同时学习标记相关性和参数模型,但本文可求出高阶的标记相关性.与文献[25,26]相比,本文提出了将标记相关性和分类器进行联合学习,且预测过程非常简单.

2 多标记分类和标记相关性的联合学习(JMLLC)

假设给定训练数据集 $D=\{(x_i,Y_i)\}_{i=1}^n \subset R^d\times\{+1,-1\}^L$.其中,$x_i\in R^d$ 表示第 i 个训练样本,$Y_i\in\{+1,-1\}^L$ 表示其对应的类别标记向量,d 为特征空间维数,L 为类别标记个数.如果第 i 个样本属于第 l 个类别标记,则 $Y_{il}=+1$;否则,$Y_{il}=-1$.为了方便表示,记数据矩阵 $X=[x_1,\ldots,x_n]^T$,类别标记指示矩阵 $Y=[Y_1,\ldots,Y_n]^T$.

2.1 基本模型

首先,将多标记学习问题转化为 L 个独立的二分类问题,其中,第 l 个标记的分类判别函数为

$$g_l(x)=w_l^Tx+b_l \tag{1}$$

该方法简单,但主要的不足在于没有考虑标记之间的相关信息,而这些信息有利于提高系统的泛化性能.因此,如何将多标记学习问题转化为求解 L 个二分类问题的同时又考虑标记之间的相关性,是本文需要研究和解决的问题.受依赖网络(DN)模型[27]以及文献[25,26,28]的启发,构建基于类别标记变量的条件依赖网络,其中,每个节点代表一个类别标记,每个类别标记将输入特征变量和其他的类别标记作为其父节点.因此,每个分类器的构建依赖于输入特征变量和其他类别标记,则第 l 个类别标记最终所对应的预测函数表示如下:

$$f_l(x)=w_l^Tx+s_{\bar{l}}^TY_x^{-l}+b_l \tag{2}$$

其中,$w_l\in R^d$ 和 $b_l\in R$ 分别表示第 l 个预测函数所对应的权重向量和偏差,$s_{\bar{l}}=[s_{l,1},\ldots,s_{l,l-1},s_{l,l+1},\ldots,s_{l,L}]^T$ 表示标记相关权重向量.

文献[25,26]中,第 l 个概率预测函数虽然依赖于输入特征向量和其他标记变量,但 $Y_x^{-l}\in\{+1,-1\}^{L-1}$ 是已知的训练数据,在每次迭代学习的过程中是固定的,则学习 w_l 时,只依赖于标记相关权重向量 $s_{\bar{l}}$,而没有考虑其他的权重向量$\{w_1,\ldots,w_{l-1},w_{l+1},\ldots,w_L\}$.与文献[25,26]不同的是,本文令 $Y_x^{-l}=[g_1(x),\ldots,g_{l-1}(x),g_{l+1}(x),\ldots,g_L(x)]^T$,而不是一个固定的向量.这样,在学习某个 w_l 时,不仅考虑了其他的权重向量,而且也考虑了标记相关权重向量 $s_{\bar{l}}$,这样不仅可以使得原本各自独立的标记分类器可以联合起来同时学习,从而增强了各个标记分类器的学习效果,而且多标记分类和标记相关性也联合起来同时学习,使得学习得到的标记相关性更为准确.另外,本文为每个标记构建相应的预测函数,而非概率预测函数,因此可以选择不同的损失函数,模型可扩展性强.在预测阶段,文献[25,26]需要通过 Gibbs 采样方法来进行预测,而本文的预测过程非常简单.最后,公式(2)可以写成:

$$f_l(x)=w_l^Tx+\sum_{k\neq l}s_{l,k}(w_k^Tx+b_k)+b_l=\left(w_l+\sum_{k\neq l}s_{l,k}w_k\right)^Tx+b_l+\sum_{k\neq l}s_{l,k}b_k \tag{3}$$

令 $S=[s_1,\ldots,s_L]\in R^{L\times L}$,$s_l=[s_{l,1},\ldots,s_{l,l-1},s_{l,l+1},\ldots,s_{l,L}]^T\in R^L$,则公式(3)可表示为

$$f_l(x)=s_l^TW^Tx+bs_l \tag{4}$$

其中,$W=[w_1,\ldots,w_L]\in R^{d\times L}$,$b=[b_1,\ldots,b_L]\in R^{1\times L}$.

因此,多标记分类和标记相关性的联合学习模型如下:

$$\begin{cases}\min\limits_{\{w_l,b_l,s_l\}_{l=1}^L}\sum\limits_{l=1}^L\left(\sum\limits_{i=1}^nV(s_l^TW^Tx_i+bs_l,Y_{il})+\lambda_1\|Ws_l\|^2+\lambda_2\left(\|w_l\|^2+\|\sum\limits_{k\neq l}s_{l,k}w_k\|^2\right)+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1,l=1,\ldots,L\end{cases} \tag{5}$$

其中,V 代表损失函数,正则化项$\|Ws_l\|^2$ 用于控制模型的复杂度,$\|w_l\|^2$ 用于控制单个类别标记所含的信息量,$\sum\limits_{k\neq l}\|s_{l,k}w_k\|^2$ 用来控制其他相关标记所含的信息量,$\|s_l\|^2$用来控制相关系数的大小.λ_1,λ_2 和 λ_3 为正则化参数,用来权衡这 4 项.

在以下两节中,我们分别选择两种不同的损失函数,即logistic回归损失函数和最小二乘损失函数.对于每个模型,首先定义在原始特征空间中,然后拓展到再生核希尔伯特空间(RKHS)中.

2.2 JMLLC-LR (JMLLC with logistic regression)

首先,选择 logistic 回归损失函数,即: $V(s_l^T W^T x_i + bs_l, Y_{il}) = \log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l)))$.

2.2.1 在原始特征空间中学习

公式(5)可表示如下:

$$\begin{cases} \min\limits_{\{w_l, b_l, s_l\}_{l=1}^L} \sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l))) + \lambda_1 \| Ws_l \|^2 + \lambda_2\left(\| w_l \|^2 + \| \sum\limits_{k\neq l} s_{l,k} w_k \|^2\right) + \lambda_3 \| s_l \|^2\right) \\ \text{s.t. } s_{l,l} = 1, l = 1,...,L \end{cases} \tag{6}$$

公式(6)的求解,采用交替迭代求解的方法.

1) 固定 S,求 W 和 b

当 S 固定时,公式(6)中的第 4 项为常数项,因此可以忽略.同时,约束条件也可以忽略.则公式(6)可重新写成:

$$\min_{\{w_l, b_l\}_{l=1}^L} \sum_{l=1}^{L}\left(\sum_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l))) + \lambda_1 \| Ws_l \|^2 + \lambda_2\left(\| w_l \|^2 + \| \sum_{k\neq l} s_{l,k} w_k \|^2\right)\right) \tag{7}$$

公式(7)的目标函数是一个凸函数,并可重写成如下:

$$\min_{W,b} \sum_{i=1}^{n}\sum_{l=1}^{L}\log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l))) + \lambda_1 \| WS \|_F^2 + \lambda_2(\| W \|_F^2 + \| W(S - I_L) \|_F^2) \tag{8}$$

其中,I_L 为 $L\times L$ 的单位矩阵.

2) 固定 W 和 b,求 S

当 W 和 b 固定时,公式(6)中第 3 项的第 1 部分为常数项,因此可以忽略,则公式(6)可表示为

$$\begin{cases} \min\limits_{\{s_l\}_{l=1}^L} \sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l))) + \lambda_1 \| Ws_l \|^2 + \lambda_2 \| \sum\limits_{k\neq l} s_{l,k} w_k \|^2 + \lambda_3 \| s_l \|^2\right) \\ \text{s.t. } s_{l,l} = 1, l = 1,...,L \end{cases} \tag{9}$$

公式(9)的目标函数是一个凸函数,且可以进一步地分解成 L 个独立的优化问题.其中,第 l 个优化问题为

$$\begin{cases} \min\limits_{s_l} \sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^T W^T x_i + bs_l))) + \lambda_1 \| Ws_l \|^2 + \lambda_2 \| W(s_l - e_l) \|^2 + \lambda_3 \| s_l \|^2 \\ \text{s.t. } s_{l,l} = 1 \end{cases} \tag{10}$$

其中,e_l 为 L 维的列向量,其第 l 个元素为 1,其余为 0.

2.2.2 在 RKHS 中学习

现将 JMLLC-LR 拓展到再生核希尔伯特空间中.根据表示理论[29],可将 w_l 表示如下:

$$w_l = \sum_{i=1}^{n}\alpha_{li}\Phi(x_i) \tag{11}$$

其中,$\Phi: R^d \mapsto F$ 表示由核引导的特征映射.

将公式(11)代入到公式(6)中可得:

$$\begin{cases} \min\limits_{A,b,\{s_l\}_{l=1}^L} \sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^T A^T k_i + bs_l))) + \lambda_1 s_l^T A^T KAs_l + \lambda_2(\alpha_l^T K\alpha_l + (s_l - e_l)^T A^T KA(s_l - e_l)) + \lambda_3 \| s_l \|^2\right) \\ \text{s.t. } s_{l,l} = 1, l = 1,...,L \end{cases} \tag{12}$$

其中,$A=[\alpha_1,\ldots,\alpha_L]\in R^{n\times L}$,$\alpha_l=[\alpha_{l1},\ldots,\alpha_{ln}]^T\in R^n$ 表示第 l 个预测函数所对应的系数向量,K 为 $n\times n$ 的 Gram 矩阵,k_i 表示核矩阵 K 的第 i 列.

公式(12)的求解,采用交替迭代求解的方法.

1) 固定 S,求 A 和 b

当 S 固定时,公式(12)中的第 4 项为常量,因此可以忽略;同时,约束条件也可以忽略.则公式(12)可以重新表示为

$$\min_{A,b}\sum_{l=1}^{L}\left(\sum_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^TA^Tk_i+bs_l)))+\lambda_1 s_l^TA^TKAs_l+\lambda_2(\alpha_l^TK\alpha_l+(s_l-e_l)^TA^TKA(s_l-e_l))\right) \tag{13}$$

公式(13)的目标函数是一个凸函数,并且可进一步写成:

$$\min_{A,b}\sum_{l=1}^{L}\sum_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^TA^Tk_i+bs_l)))+\lambda_1 tr(S^TA^TKAS)+\lambda_2 tr(A^TKA+(S-I_L)^TA^TKA(S-I_L)) \tag{14}$$

其中,$tr(\cdot)$表示矩阵的迹.

2) 固定 A 和 b,求 S

当 A 和 b 固定时,公式(12)中第 3 项的第 1 部分为常数项,因此可以忽略,则公式(12)可重新写成如下:

$$\begin{cases}\min\limits_{\{s_l\}_{l=1}^L}\sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^TA^Tk_i+bs_l)))+\lambda_1 s_l^TA^TKAs_l+\lambda_2(s_l-e_l)^TA^TKA(s_l-e_l)+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1, l=1,\ldots,L\end{cases} \tag{15}$$

公式(15)的目标函数是一个凸函数,且可以进一步分解成 L 个独立的优化问题.其中,第 l 个优化问题如下:

$$\begin{cases}\min\limits_{s_l}\sum\limits_{i=1}^{n}\log(1+\exp(-Y_{il}(s_l^TA^Tk_i+bs_l)))+\lambda_1 s_l^TA^TKAs_l+\lambda_2(s_l-e_l)^TA^TKA(s_l-e_l)+\lambda_3\|s_l\|^2\\ \text{s.t. } s_{l,l}=1\end{cases} \tag{16}$$

2.3 JMLLC-LS (JMLLC with least squares)

为了简单起见,假设 X 和 Y 已经中心化,此时,偏差 $\{b_l\}_{l=1}^L$ 都为 0.我们选择最小二乘损失函数,即:

$$V(s_l^TW^Tx_i,Y_{il})=(s_l^TW^Tx_i-Y_{il})^2.$$

2.3.1 在原始特征空间中学习

当公式(5)中的损失项为最小二乘损失函数时,可以得到最终的模型如下:

$$\begin{cases}\min\limits_{\{w_l,s_l\}_{l=1}^L}\sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}(s_l^TW^Tx_i-Y_{il})^2+\lambda_1\|Ws_l\|^2+\lambda_2\left(\|w_l\|^2+\|\sum\limits_{k\neq l}s_{l,k}w_k\|^2\right)+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1, l=1,\ldots,L\end{cases} \tag{17}$$

定理 1. 采用交替迭代求解的方法求解公式(17).其中,

- 当 S 固定时,W 可通过求解一个 Sylvester 方程求得:

$$BW+WSS^TC=BX^TYS^TC \tag{18}$$

其中,$B=(X^TX+\lambda_1I_d)^{-1}$,$C=(\lambda_2I_L+\lambda_2(S-I_L)(S-I_L)^T)^{-1}$,$I_d$ 为 $d\times d$ 的单位矩阵;

- 当 W 固定时,S 可通过求解下式来求得:

$$\begin{aligned}&\min_{s_l}\|XWs_l-Y_{\cdot l}\|^2+\lambda_1\|Ws_l\|^2+\lambda_2\|W(s_l-e_l)\|^2+\lambda_3\|s_l\|^2\\ &\text{s.t. } s_{l,l}=1\end{aligned} \tag{19}$$

其中,$Y_{\cdot l}=[Y_{1l},\ldots,Y_{nl}]^T$,即标记指示矩阵 Y 的第 l 列;e_l 为 L 维的列向量,其第 l 个元素为 1,其余为 0.

证明: 用交替求解的方法来求解公式(17)中的优化问题.

1) 固定 S,求 W

当 S 固定时,公式(17)中的第 4 项为常数项,因此可以忽略;同时,约束条件也可以忽略,则公式(17)可重写成:

$$\min_{\{w_l\}_{l=1}^L}\sum_{l=1}^{L}\left(\sum_{i=1}^{n}(s_l^TW^Tx_i-Y_{il})^2+\lambda_1\|Ws_l\|^2+\lambda_2\left(\|w_l\|^2+\|\sum_{k\neq l}s_{l,k}w_k\|^2\right)\right) \tag{20}$$

公式(20)可重写成如下:

$$\min_{W}\|XWS-Y\|_F^2+\lambda_1\|WS\|_F^2+\lambda_2(\|W\|_F^2+\|W(S-I_L)\|_F^2) \tag{21}$$

公式(21)的目标函数是一个凸函数,对公式(21)中的目标函数关于 W 求导并令其等于 0,可得定理 1 中的公式(18).

2) 固定 W,求 S

当 W 固定时,公式(17)中第 3 项的第 1 部分为常数项,因此可以忽略.则公式(17)可表示为

$$\begin{cases}\min\limits_{\{s_l\}_{l=1}^L}\sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}(s_l^TW^Tx_i-Y_{il})^2+\lambda_1\|w_l+\sum\limits_{k\neq l}s_{lk}w_k\|^2+\lambda_2\|\sum\limits_{k\neq l}s_{l,k}w_k\|^2+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1,l=1,...,L\end{cases} \tag{22}$$

公式(22)的目标函数是凸函数,且可以进一步分解成 L 个独立的优化问题,可得定理 1 中的公式(19). □

2.3.2 在 RKHS 中学习

将公式(11)代入到公式(17)中可得:

$$\begin{cases}\min\limits_{A,\{s_l\}_{l=1}^L}\sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}(s_l^TA^Tk_i-Y_{il})^2+\lambda_1s_l^TA^TKAs_l+\lambda_2(\alpha_l^TK\alpha_l+(s_l-e_l)^TA^TKA(s_l-e_l))+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1,l=1,...,L\end{cases} \tag{23}$$

通过迭代求解的方法来求解 A 和 S,可总结如下定理:

定理 2. 采用交替迭代求解的方法求解公式(23).其中,

- 当 S 固定时,A 可通过求解一个 Sylvester 方程求得:

$$BA+ASS^TC=BYS^TC \tag{24}$$

其中,$B=(K+\lambda_1I_n)^{-1}$,$C=(\lambda_2I_L+\lambda_2(S-I_L)(S-I_L)^T)^{-1}$,$I_n$ 为 $n\times n$ 的单位矩阵;

- 当 A 固定时,S 可通过求解下式来求得:

$$\begin{cases}\min\limits_{s_l}\|KAs_l-Y_{\bullet l}\|^2+\lambda_1s_l^TA^TKAs_l+\lambda_2(s_l-e_l)^TA^TKA(s_l-e_l)+\lambda_3\|s_l\|^2\\ \text{s.t. } s_{l,l}=1\end{cases} \tag{25}$$

其中,$Y_{\bullet l}=[Y_{1l},\ldots,Y_{nl}]^T$,即标记指示矩阵 Y 的第 l 个列;e_l 为 L 维的列向量,其第 l 个元素为 1,其余为 0.

证明: 采用交替求解的方法来解公式(23)中的凸优化问题.

1) 固定 S,求 A

当 S 固定时,公式(23)中第 4 项为常量,因此可以忽略;同时,约束条件也可以忽略,则公式(23)可重新表示为

$$\min_A\sum_{l=1}^{L}\left(\sum_{i=1}^{n}(s_l^TA^Tk_i-Y_{il})^2+\lambda_1s_l^TA^TKAs_l+\lambda_2(\alpha_l^TK\alpha_l+(s_l-e_l)^TA^TKA(s_l-e_l))\right) \tag{26}$$

公式(26)可进一步写成:

$$\min_A\|KAS-Y\|_F^2+\lambda_1tr(S^TA^TKAS)+\lambda_2tr(A^TKA+(S-I_L)^TA^TKA(S-I_L)) \tag{27}$$

其中,$tr(\cdot)$表示矩阵的迹.

公式(27)的目标函数是一个凸函数,对其关于 A 求导并令其等于 0,可得定理 2 中的公式(24).

2) 固定 A,求 S

当 A 固定时,公式(23)中第 3 项的第 1 部分为常数项,因此可以忽略,则公式(23)可重新写成如下:

$$\begin{cases}\min\limits_{\{s_l\}_{l=1}^L}\sum\limits_{l=1}^{L}\left(\sum\limits_{i=1}^{n}(s_l^TA^Tk_i-Y_{il})^2+\lambda_1s_l^TA^TKAs_l+\lambda_2(s_l-e_l)^TA^TKA(s_l-e_l)+\lambda_3\|s_l\|^2\right)\\ \text{s.t. } s_{l,l}=1,l=1,...,L\end{cases} \tag{28}$$

公式(28)的目标函数是凸函数,并且可以进一步分解成 L 个独立的优化问题,可得定理 2 中的公式(25). □

2.4 预测阶段

在预测阶段,给定一个未知样本 x,其预测的类别标记集合为

$$h(x)=(h_1(x),\ldots,h_L(x))=sign(f_1(x),\ldots,f_L(x)) \tag{29}$$

其中,$sign(\cdot)$为符号函数.

3 实 验

3.1 数据集描述

为了验证本文提出的方法的性能,我们选取了 20 个多标记基准数据集(其中,Yahoo 包括 11 个数据集)进行实验测试.这些数据集涉及了文本分类、图像标记、音乐分类和生物信息学等应用领域,具体描述见表 1.

Table 1 Detailed descriptions of multi-label data sets

表 1 多标记数据集的详细描述

数据集	训练样本数	测试样本数	属性个数	标记个数	平均标记个数	标记密度	所属领域
medical	645	333	1 449	45	1.245	0.028	Text
langlog	751	502	1 004	75	1.375	0.018	Text
slashdot	2 269	1 513	1 079	22	1.181	0.054	Text
image	1 200	800	294	5	1.236	0.247	Images
scene	1 211	1 196	294	6	1.074	0.179	Images
emotions	391	202	72	6	1.868	0.311	Music
yeast	1 500	917	103	14	4.237	0.303	Biology
human	1 864	1 244	440	14	1.185	0.085	Biology
plant	558	390	440	12	1.079	0.090	Biology
Yahoo	2 000	3 000	438~1047	21~40	1.169~1.692	0.033~0.068	Text

其中,平均标记个数$=\frac{1}{n}\sum_{i=1}^{n}\sum_{l=1}^{L}[Y_{il}=+1]$,标记密度$=\frac{1}{nL}\sum_{i=1}^{n}\sum_{l=1}^{L}[Y_{il}=+1]$.[·]为 1,如果条件为真;否则为 0.

3.2 评价准则

多标记学习框架中,每个样本可能同时隶属于多个类别标记.因此,与单标记学习系统相比,多标记学习系统的评价准则要更加复杂些.到目前为止,已提出了许多多标记学习系统的性能评价准则[1,3,7,13,15,16,19].假设测试集$T=\{(x_i,Y_i)\}_{i=1}^{m}\subset R^d\times\{+1,-1\}^L$,并根据预测函数 $f_l(x)$,定义一个排序函数 $rank_f(x,l)\in\{1,\ldots,L\}$,如果 $f_l(x)>f_k(x)$,则 $rank_f(x,l)<rank_f(x,k)$.本文选取了 5 种常用的评价准则,即 Hamming Loss,One-Error,Coverage,Ranking Loss 和 Average Precision 来评价多标记学习系统的性能,具体定义如下:

1) Hamming Loss:用于度量样本在单个标记上的真实标记和预测标记的错误匹配情况:

$$hLoss(h)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{L}\sum_{l=1}^{L}[h_l(x_i)\neq Y_{il}];$$

2) One-Error:用来考察预测值排在第 1 位的标记却不隶属于该样本的情况:

$$oneError(f)=\frac{1}{m}\sum_{i=1}^{m}[Y_{i,l_i}=-1];$$

其中,$l_i=\arg\max_{k\in\{1,\ldots,L\}}f_k(x_i)$.

3) Coverage:用于度量平均上需要多少步才能遍历样本所有的相关标记:

$$coverage\ (f)=\frac{1}{m}\sum_{i=1}^{m}\max_{l\in R_i}rank_f(x_i,l)-1;$$

其中,$R_i=\{l|Y_{il}=+1\}$表示样本 x_i 的相关标记集合,而 $\bar{R}_i=\{l\,|\,Y_{il}=-1\}$ 表示样本 x_i 的不相关标记集合.

4) Ranking Loss:用来考察样本的不相关标记的排序低于相关标记的排序的情况:

$$rLoss(f)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|R_i||\bar{R}_i|}|\{(l,k)\,|\,rank_f(x_i,l)\geqslant rank_f(x_i,k),(l,k)\in R_i\times\bar{R}_i\}|;$$

5) Average Precision:用来考察排在隶属于该样本标记之前的标记仍属于样本的相关标记集合的情况:

$$avgPre(f)=\frac{1}{m}\sum_{i=1}^{m}\frac{1}{|R_i|}\sum_{l\in R_i}\frac{|\{k\,|\,rank_f(x_i,k)\leqslant rank_f(x_i,l),k\in R_i\}|}{rank_f(x_i,l)}.$$

上述 5 个评价准则中,前 4 个值越小系统性能越优,最优值为 0;最后 1 个值越大越好,最优值为 1.

3.3 参数选择

为了验证本文提出的方法 JMLLC(包括 JMLLC-LS 和 JMLLC-LR)的有效性,将 JMLLC 算法与 8 个多标记学习算法,即 BR[4],ML*k*NN[7],BPMLL[3],RankSvm[6],CLR[8],RAkEL[12],ECC[9]和 CDN-LR[26]进行实验比较.其中,JMLLC,ML*k*NN,BPMLL,RankSvm 和 CDN-LR 是基于 Matlab 实现的,其他的是基于 MULAN 软件包[30]实现的.对于 BR,CLR,RAkEL 和 ECC,都选择 C4.5 算法作为基分类器.对于 MLkNN,近邻个数 k 和平滑参数 s 分别设置为 10 和 1.在 BPMLL 中,神经网络的隐含神经元个数为特征个数的 20%,学习率 α=0.05 以及训练时最大的迭代次数为 100.对于 RankSvm,代价参数 c=1,同时选择 RBF 核函数,即 $K(x,y)=\exp(-\|x-y\|^2/2\sigma^2)$,其中,参数 σ 等于所有训练样本对之间的欧式距离的均值.对于 CDN-LR,也选择 RBF 核函数,λ,n,t_b 和 t_c 分别取 0.1,100,100 和 500.对于本文的 JMLLC-LS 和 JMLLC-LR,我们也选择 RBF 核函数,有 3 个正则化参数 λ_1,λ_2 和 λ_3 需要调整.对于 JMLLC-LS 和 JMLLC-LR,都在一些数据集上用 10 折交叉验证来选择 λ_1,λ_2 和 λ_3,它们的取值范围都为{0.0001,0.0005,0.001,0.005,…,1,5,10,50,100,500}.实验结果表明:当 λ_1=0.005,λ_2=0.005 以及 λ_3=50 时,JMLLC-LR 产生稳定的性能.因此,本文中对于 JMLLC-LR,在所有的数据集上,λ_1,λ_2 和 λ_3 都分别取 0.005,0.005 和 50.同样地,对于 JMLLC-LS,在所有的数据集上都取 λ_1=0.01,λ_2=0.5 以及 λ_3=5.

3.4 实验结果与分析

我们将 JMLLC-LR 和 JMLLC-LS 与其他 8 个常用的多标记学习算法——BR,MLkNN,BPMLL,RankSvm,CLR,RAkEL,ECC 和 CDN-LR 进行实验比较.根据 5 种不同的评价准则,表 2~表 6 分别列出了我们的方法与其他 8 种对比算法在表 1 中所列出的所有数据集上的详细实验结果,其中,将 Yahoo 的 11 个数据集上的平均结果作为 Yahoo 数据的结果.这些方法在每个数据集上的实验结果排序(其中,Yahoo 数据的排序结果为 Yahoo 数据的 11 个数据集上的平均排序)都在实验数据后面用下标形式写出,且最好的结果加粗表示;同时,每种方法在所有数据集上的平均排序结果列在最后一行,其中,平均排序越小,系统性能越优.

Table 2 Performance comparison based on Hamming Loss (the smaller, the better)

表 2 基于 Hamming Loss(值越小越好)的性能比较

数据集	JMLLC-LR	JMLLC-LS	BR	MLkNN	BPMLL	RankSvm	CLR	RAkEL	ECC	CDN-LR
medical	0.013 $6_{(6)}$	**0.011 7**$_{(1)}$	0.012 $7_{(4)}$	0.017 $1_{(7)}$	0.020 $2_{(9)}$	0.019 $6_{(8)}$	0.012 $8_{(5)}$	0.012 $5_{(3)}$	0.012 $3_{(2)}$	0.022 $5_{(10)}$
langlog	**0.017 3**$_{(1)}$	0.017 $7_{(2.5)}$	0.021 $0_{(7)}$	0.018 $1_{(4.5)}$	0.024 $7_{(9)}$	0.023 $5_{(8)}$	0.017 $7_{(2.5)}$	0.020 $3_{(6)}$	0.018 $1_{(4.5)}$	0.026 $8_{(10)}$
slashdot	0.041 $8_{(2)}$	**0.039 2**$_{(1)}$	0.046 $4_{(6)}$	0.052 $8_{(9)}$	0.045 $4_{(5)}$	0.047 $5_{(7)}$	0.048 $0_{(8)}$	0.043 $7_{(3)}$	0.044 $5_{(4)}$	0.066 $7_{(10)}$
image	0.166 $3_{(2)}$	**0.151 3**$_{(1)}$	0.225 $7_{(7)}$	0.172 $3_{(3.5)}$	0.247 $5_{(10)}$	0.178 $0_{(5)}$	0.228 $2_{(8)}$	0.183 $3_{(6)}$	0.172 $3_{(3.5)}$	0.244 $3_{(9)}$
scene	**0.083 9**$_{(1)}$	0.089 $6_{(2)}$	0.138 $9_{(8)}$	0.098 $9_{(3)}$	0.290 $3_{(10)}$	0.104 $8_{(5)}$	0.140 $9_{(9)}$	0.115 $0_{(6)}$	0.101 $6_{(4)}$	0.118 $0_{(7)}$
emotions	0.203 $8_{(2)}$	**0.189 8**$_{(1)}$	0.269 $0_{(10)}$	0.208 $7_{(3)}$	0.225 $2_{(6)}$	0.213 $7_{(4)}$	0.262 $4_{(9)}$	0.227 $7_{(7)}$	0.215 $3_{(5)}$	0.255 $8_{(8)}$
yeast	0.196 $8_{(2)}$	**0.193 8**$_{(1)}$	0.258 $8_{(9)}$	0.198 $0_{(3)}$	0.208 $9_{(5)}$	0.202 $4_{(4)}$	0.225 $7_{(7)}$	0.232 $8_{(8)}$	0.212 $4_{(6)}$	0.286 $9_{(10)}$
human	0.081 $4_{(2)}$	**0.079 6**$_{(1)}$	0.120 $3_{(10)}$	0.083 $1_{(3)}$	0.084 $4_{(4)}$	0.093 $0_{(6)}$	0.096 $8_{(8)}$	0.100 $0_{(9)}$	0.085 $3_{(5)}$	0.096 $6_{(7)}$
plant	0.088 $0_{(4)}$	**0.084 4**$_{(1)}$	0.130 $1_{(10)}$	0.087 $0_{(2)}$	0.089 $1_{(5)}$	0.106 $8_{(7)}$	0.105 $1_{(6)}$	0.109 $6_{(8)}$	0.087 $8_{(3)}$	0.117 $3_{(9)}$
Yahoo	0.039 $6_{(2.55)}$	**0.037 9**$_{(1.00)}$	0.049 $8_{(7.86)}$	0.043 $2_{(4.73)}$	0.051 $4_{(8.82)}$	0.049 $0_{(7.59)}$	0.043 $6_{(5.05)}$	0.043 $9_{(5.23)}$	0.040 $2_{(3.00)}$	0.054 $6_{(9.18)}$
平均排序	2.50	**1.13**	7.88	4.50	8.00	6.88	5.90	5.68	3.50	9.05

Table 3 Performance comparison based on One-Error (the smaller, the better)

表 3 基于 One-Error(值越小越好)的性能比较

数据集	JMLLC-LR	JMLLC-LS	BR	MLkNN	BPMLL	RankSvm	CLR	RAkEL	ECC	CDN-LR
medical	0.132 $1_{(2)}$	**0.117 1**$_{(1)}$	0.228 $2_{(5)}$	0.264 $3_{(7)}$	0.360 $4_{(9)}$	0.354 $4_{(8)}$	0.180 $2_{(4)}$	0.237 $2_{(6)}$	0.177 $2_{(3)}$	0.387 $4_{(10)}$
langlog	0.691 $2_{(2)}$	**0.681 3**$_{(1)}$	0.800 $8_{(7)}$	0.826 $7_{(8)}$	0.770 $9_{(6)}$	0.866 $5_{(10)}$	0.768 $9_{(5)}$	0.766 $9_{(4)}$	0.709 $2_{(3)}$	0.842 $6_{(9)}$
slashdot	0.439 $5_{(3)}$	**0.379 4**$_{(1)}$	0.544 $6_{(7)}$	0.664 $2_{(10)}$	0.405 $2_{(2)}$	0.530 $1_{(5)}$	0.568 $4_{(8)}$	0.510 $9_{(4)}$	0.530 $7_{(6)}$	0.655 $0_{(9)}$
image	0.285 $0_{(2)}$	**0.253 8**$_{(1)}$	0.460 $0_{(8)}$	0.323 $8_{(5.5)}$	0.552 $5_{(10)}$	0.297 $5_{(3)}$	0.357 $5_{(7)}$	0.323 $8_{(5.5)}$	0.308 $8_{(4)}$	0.475 $0_{(9)}$
scene	**0.208 2**$_{(1)}$	0.219 $1_{(2)}$	0.426 $4_{(9)}$	0.242 $5_{(4)}$	0.826 $9_{(10)}$	0.231 $6_{(3)}$	0.311 $0_{(8)}$	0.297 $7_{(6)}$	0.270 $1_{(5)}$	0.306 $9_{(7)}$
emotions	**0.257 4**$_{(1)}$	0.262 $4_{(2)}$	0.351 $5_{(9)}$	0.302 $0_{(5.5)}$	0.306 $9_{(7)}$	0.302 $0_{(5.5)}$	0.311 $9_{(8)}$	0.297 $0_{(4)}$	0.282 $2_{(3)}$	0.401 $0_{(10)}$
yeast	0.239 $9_{(4)}$	**0.231 2**$_{(1)}$	0.402 $4_{(9)}$	0.234 $5_{(2)}$	0.236 $6_{(3)}$	0.244 $3_{(5)}$	0.249 $7_{(7)}$	0.290 $1_{(8)}$	0.248 $6_{(6)}$	0.464 $6_{(10)}$
human	0.546 $6_{(2)}$	**0.517 7**$_{(1)}$	0.725 $1_{(10)}$	0.603 $7_{(5)}$	0.684 $1_{(9)}$	0.550 $6_{(3)}$	0.605 $3_{(6)}$	0.635 $0_{(8)}$	0.610 $9_{(7)}$	0.585 $2_{(4)}$
plant	**0.594 9**$_{(1)}$	0.607 $7_{(2)}$	0.794 $9_{(9)}$	0.664 $1_{(3.5)}$	0.943 $6_{(10)}$	0.664 $1_{(3.5)}$	0.725 $6_{(7)}$	0.738 $5_{(8)}$	0.705 $1_{(6)}$	0.669 $2_{(5)}$
Yahoo	0.398 $8_{(2.45)}$	**0.370 3**$_{(1.09)}$	0.527 $5_{(8.18)}$	0.471 $4_{(5.64)}$	0.523 $9_{(8.45)}$	0.507 $0_{(7.50)}$	0.430 $0_{(4.36)}$	0.452 $6_{(5.09)}$	0.412 $2_{(3.09)}$	0.541 $5_{(9.14)}$
平均排序	2.25	**1.20**	8.15	5.63	7.95	6.43	5.40	5.48	3.85	8.68

Table 4 Performance comparison based on Coverage (the smaller, the better)

表 4 基于 Coverage(值越小越好)的性能比较

数据集	JMLLC-LR	JMLLC-LS	BR	MLkNN	BPMLL	RankSvm	CLR	RAkEL	ECC	CDN-LR
medical	1.501 $5_{(2)}$	**1.198 2**$_{(1)}$	4.096 $1_{(9)}$	2.723 $7_{(6)}$	2.030 $0_{(4)}$	3.648 $6_{(7)}$	1.973 $0_{(3)}$	5.888 $9_{(10)}$	2.720 $7_{(5)}$	3.732 $7_{(8)}$
langlog	**12.37 9**$_{(1)}$	15.47 $8_{(4)}$	23.50 $8_{(9)}$	15.65 $9_{(5)}$	15.942$_{(7)}$	15.707$_{(6)}$	13.008$_{(2)}$	34.496$_{(10)}$	14.010$_{(3)}$	17.215$_{(8)}$
slashdot	2.407 $1_{(2)}$	2.519 $5_{(3)}$	3.828 $8_{(7)}$	4.262 $4_{(8)}$	**2.328 5**$_{(1)}$	3.076 $0_{(4)}$	3.103 $1_{(5)}$	5.833 $4_{(10)}$	3.343 $7_{(6)}$	4.269 $7_{(9)}$
image	0.886 $3_{(2.5)}$	**0.817 5**$_{(1)}$	1.465 $0_{(9)}$	0.966 $3_{(4)}$	1.636 $3_{(10)}$	0.886 $3_{(2.5)}$	1.050 $0_{(7)}$	0.992 $5_{(6)}$	0.971 $3_{(5)}$	1.271 $3_{(8)}$
scene	**0.441 5**$_{(1)}$	0.484 $1_{(3)}$	1.280 $9_{(9)}$	0.568 $6_{(4)}$	2.219 $9_{(10)}$	0.469 $9_{(2)}$	0.648 $8_{(6)}$	0.687 $3_{(7)}$	0.592 $8_{(5)}$	0.732 $4_{(8)}$
emotions	**1.747 5**$_{(1)}$	1.802 $0_{(2)}$	2.841 $6_{(10)}$	1.876 $2_{(3)}$	1.965 $3_{(6)}$	1.910 $9_{(5)}$	2.054 $5_{(7)}$	2.118 $8_{(8)}$	1.905 $9_{(4)}$	2.212 $9_{(9)}$
yeast	6.405 $7_{(3)}$	6.365 $3_{(2)}$	9.285 $7_{(10)}$	6.414 $4_{(4.5)}$	6.414 $4_{(4.5)}$	**6.321 7**$_{(1)}$	6.797 $2_{(7)}$	7.669 $6_{(8)}$	6.681 $6_{(6)}$	7.954 $2_{(9)}$
human	**2.037 8**$_{(1)}$	2.237 $1_{(3)}$	5.959 $8_{(10)}$	2.405 $1_{(5)}$	5.842 $4_{(9)}$	2.082 $8_{(2)}$	2.442 $1_{(6)}$	3.651 $9_{(8)}$	2.811 $9_{(7)}$	2.250 $8_{(4)}$
plant	**1.874 4**$_{(1)}$	1.892 $3_{(2)}$	4.956 $4_{(9)}$	2.423 $1_{(5)}$	5.653 $8_{(10)}$	2.066 $7_{(3)}$	2.515 $4_{(6)}$	3.351 $3_{(8)}$	2.882 $1_{(7)}$	2.123 $1_{(4)}$
Yahoo	**3.634 1**$_{(1.73)}$	4.654 $8_{(7.27)}$	8.069 $1_{(9.09)}$	4.097 $5_{(4.55)}$	3.980 $8_{(4.36)}$	3.992 $8_{(4.82)}$	3.688 $4_{(2.00)}$	9.878 $9_{(9.91)}$	3.847 $8_{(3.82)}$	4.756 $4_{(7.45)}$
平均排序	**1.68**	5.05	9.10	4.73	5.48	4.28	3.55	9.20	4.50	7.45

Table 5 Performance comparison based on Ranking Loss (the smaller, the better)

表 5 基于 Ranking Loss(值越小越好)的性能比较

数据集	JMLLC-LR	JMLLC-LS	BR	MLkNN	BPMLL	RankSvm	CLR	RAkEL	ECC	CDN-LR
medical	0.020 $8_{(2)}$	**0.015 6**$_{(1)}$	0.068 $1_{(9)}$	0.042 $5_{(5)}$	0.030 $6_{(4)}$	0.059 $8_{(7)}$	0.028 $6_{(3)}$	0.106 $3_{(10)}$	0.043 $4_{(6)}$	0.065 $2_{(8)}$
langlog	**0.130 2**$_{(1)}$	0.168 $4_{(4)}$	0.263 $2_{(9)}$	0.171 $1_{(6)}$	0.174 $7_{(7)}$	0.170 $2_{(5)}$	0.138 $2_{(2)}$	0.406 $0_{(10)}$	0.147 $2_{(3)}$	0.189 $9_{(8)}$
slashdot	0.096 $2_{(3)}$	0.095 $3_{(2)}$	0.158 $4_{(7)}$	0.178 $5_{(8)}$	**0.090 0**$_{(1)}$	0.124 $6_{(4)}$	0.127 $0_{(5)}$	0.240 $6_{(10)}$	0.136 $6_{(6)}$	0.180 $9_{(9)}$
image	0.155 $0_{(3)}$	**0.135 9**$_{(1)}$	0.295 $9_{(9)}$	0.175 $5_{(5)}$	0.346 $1_{(10)}$	0.153 $6_{(2)}$	0.193 $2_{(7)}$	0.181 $5_{(6)}$	0.171 $6_{(4)}$	0.259 $0_{(8)}$
scene	**0.068 0**$_{(1)}$	0.075 $1_{(3)}$	0.230 $7_{(9)}$	0.093 $1_{(4)}$	0.425 $3_{(10)}$	0.073 $9_{(2)}$	0.109 $6_{(6)}$	0.116 $6_{(7)}$	0.098 $0_{(5)}$	0.127 $5_{(8)}$
emotions	**0.140 4**$_{(1)}$	0.149 $0_{(2)}$	0.312 $9_{(10)}$	0.161 $5_{(3)}$	0.180 $8_{(6)}$	0.173 $0_{(5)}$	0.192 $6_{(7)}$	0.194 $0_{(8)}$	0.167 $9_{(4)}$	0.252 $6_{(9)}$
yeast	0.169 $8_{(2)}$	**0.167 9**$_{(1)}$	0.320 $6_{(10)}$	0.171 $5_{(3)}$	0.174 $8_{(5)}$	0.174 $4_{(4)}$	0.184 $7_{(6)}$	0.224 $0_{(8)}$	0.189 $4_{(7)}$	0.295 $5_{(9)}$
human	**0.133 5**$_{(1)}$	0.145 $9_{(3)}$	0.419 $1_{(10)}$	0.161 $1_{(5)}$	0.412 $9_{(9)}$	0.136 $9_{(2)}$	0.162 $1_{(6)}$	0.247 $5_{(8)}$	0.187 $8_{(7)}$	0.150 $6_{(4)}$
plant	**0.161 2**$_{(1)}$	0.162 $2_{(2)}$	0.441 $5_{(9)}$	0.211 $0_{(5)}$	0.498 $3_{(10)}$	0.180 $0_{(3)}$	0.221 $3_{(6)}$	0.295 $2_{(8)}$	0.252 $9_{(7)}$	0.185 $0_{(4)}$
Yahoo	**0.085 0**$_{(1.55)}$	0.106 $5_{(6.82)}$	0.209 $0_{(9.14)}$	0.102 $4_{(5.36)}$	0.094 $2_{(4.09)}$	0.095 $5_{(4.73)}$	0.087 $4_{(2.09)}$	0.259 $1_{(9.86)}$	0.091 $9_{(3.73)}$	0.117 $1_{(7.64)}$
平均排序	**1.60**	4.70	9.13	5.15	5.35	4.30	3.55	9.18	4.50	7.55

Table 6 Performance comparison based on Average Precision (the larger, the better)

表 6 基于 Average Precision(值越大越好)的性能比较

数据集	JMLLC-LR	JMLLC-LS	BR	MLkNN	BPMLL	RankSvm	CLR	RAkEL	ECC	CDN-LR
medical	0.894 $3_{(2)}$	**0.906 2**$_{(1)}$	0.793 $2_{(6)}$	0.795 $7_{(5)}$	0.758 $8_{(8)}$	0.717 $7_{(9)}$	0.865 $5_{(3)}$	0.770 $1_{(7)}$	0.853 $5_{(4)}$	0.703 $1_{(10)}$
langlog	**0.402 0**$_{(1)}$	0.398 $5_{(2)}$	0.281 $3_{(7)}$	0.285 $1_{(6)}$	0.334 $5_{(5)}$	0.265 $3_{(9)}$	0.345 $3_{(4)}$	0.272 $1_{(8)}$	0.375 $9_{(3)}$	0.254 $9_{(10)}$
slashdot	0.674 $7_{(3)}$	**0.712 1**$_{(1)}$	0.566 $6_{(7)}$	0.477 $5_{(10)}$	0.696 $8_{(2)}$	0.595 $7_{(4)}$	0.572 $6_{(6)}$	0.545 $5_{(8)}$	0.589 $6_{(5)}$	0.491 $1_{(9)}$
image	0.813 $8_{(2)}$	**0.834 3**$_{(1)}$	0.684 $1_{(9)}$	0.791 $6_{(5)}$	0.629 $3_{(10)}$	0.808 $8_{(3)}$	0.768 $6_{(7)}$	0.788 $8_{(6)}$	0.796 $0_{(4)}$	0.704 $3_{(8)}$
scene	**0.876 7**$_{(1)}$	0.867 $9_{(2)}$	0.711 $5_{(9)}$	0.851 $2_{(4)}$	0.451 $9_{(10)}$	0.864 $2_{(3)}$	0.812 $8_{(7)}$	0.815 $0_{(6)}$	0.836 $0_{(5)}$	0.808 $2_{(8)}$
emotions	**0.826 1**$_{(1)}$	0.815 $3_{(2)}$	0.693 $9_{(10)}$	0.790 $7_{(4)}$	0.778 $4_{(6)}$	0.789 $4_{(5)}$	0.767 $3_{(8)}$	0.774 $8_{(7)}$	0.797 $7_{(3)}$	0.722 $7_{(9)}$
yeast	0.759 $3_{(2)}$	**0.765 0**$_{(1)}$	0.616 $4_{(9)}$	0.758 $5_{(3)}$	0.750 $6_{(4.5)}$	0.750 $6_{(4.5)}$	0.739 $7_{(7)}$	0.710 $2_{(8)}$	0.741 $8_{(6)}$	0.612 $7_{(10)}$
human	0.628 $9_{(2)}$	**0.638 3**$_{(1)}$	0.401 $2_{(10)}$	0.581 $1_{(5)}$	0.410 $9_{(9)}$	0.623 $4_{(3)}$	0.574 $7_{(6)}$	0.527 $2_{(8)}$	0.562 $8_{(7)}$	0.598 $4_{(4)}$
plant	**0.593 2**$_{(1)}$	0.590 $9_{(2)}$	0.363 $4_{(9)}$	0.536 $5_{(5)}$	0.240 $5_{(10)}$	0.555 $1_{(3)}$	0.497 $8_{(6)}$	0.460 $6_{(8)}$	0.493 $4_{(7)}$	0.544 $6_{(4)}$
Yahoo	0.679 $9_{(2.36)}$	**0.692 7**$_{(1.18)}$	0.552 $7_{(9.27)}$	0.624 $7_{(5.09)}$	0.609 $3_{(6.82)}$	0.616 $1_{(6.18)}$	0.659 $0_{(4.09)}$	0.573 $0_{(8.45)}$	0.667 $3_{(3.09)}$	0.575 $4_{(7.73)}$
平均排序	2.05	**1.30**	8.90	5.15	6.98	5.58	4.95	7.95	3.90	7.85

在表 2 中,对于前 9 个数据集,JMLLC-LS 在 7 个数据集上的 Hamming Loss 值是最小的,即最好的,而 JMLLC-LR 在其余的 2 个数据集上的值是最小的.JMLLC-LR 和 JMLLC-LS 在绝大多数数据集上的 Hamming Loss 都要小于其他对比算法,而且,JMLLC-LS 算法的性能非常稳定.对于 Yahoo 数据集,JMLLC-LS 的平均结果最优,其次是 JMLLC-LR.根据 20 个数据集上的平均排序结果可以看出:JMLLC-LS 排在第 1,且在大多数数据集上都取得较优的结果;JMLLC-LR 排在第 2;而将其他标记变量值作为固定值的 CDN-LR 排在最后,性能最差.

从表 3 中列出的结果可以看出:对于前 9 个数据集,JMLLC-LR 和 JMLLC-LS 分别在 3 个和 6 个数据集上取得最好的结果,分别在 4 个和 3 个数据集上取得次优的结果,而且 JMLLC-LS 都是排在前两位;对于 Yahoo 数据集,JMLLC-LS 的平均结果最优,其次是 JMLLC-LR.根据各个算法的平均排序结果可以发现:JMLLC-LS 位居第 1,JMLLC-LR 排在第 2,考虑了高阶标记相关性的 ECC 算法位于第 3 位,而 CDN-LR 算法仍然位于最后.

表 4 中的实验结果显示:对于前 9 个数据集,在大多数数据集上,JMLLC-LR 都优于其他算法;在 slashdot 数据集上,BPMLL 取得了最好的性能,但与分别排第 2 和第 3 的 JMLLC-LR 和 JMLLC-LS 算法相比较,Coverage

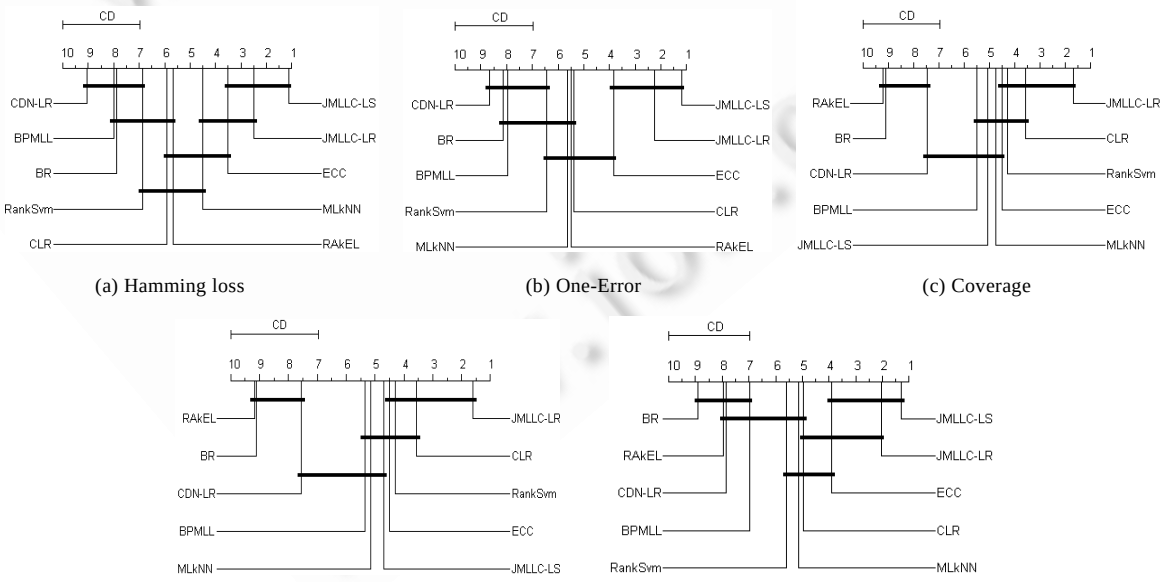
值减少得非常少,分别不到 0.1 和 0.2;在 yeast 数据集上,RankSvm 取得了最好的性能,但与次优的 JMLLC-LS 和 JMLLC-LR 在 Coverage 指标上的性能相当;对于 Yahoo 数据集,JMLLC-LR 的平均结果最优,其次是 CLR.从平均排序结果来看,在 Coverage 指标上,JMLLC-LR 的性能最稳定,排在第 1.

从表 5 可以看出:对于前 9 个数据集,JMLLC-LR,JMLLC-LS 和 BPMLL 分别在 5 个、3 个和 1 个数据集上性能最优;在 slashdot 数据集上,BPMLL 取得了最好的性能,但比 JMLLC-LR 和 JMLLC-LS 的 Ranking Loss 指标值只低不到 0.01,并且 BPMLL 在其他数据集上的性能都要比 JMLLC-LR 和 JMLLC-LS 差很多;对于 Yahoo 数据集,JMLLC-LR 的平均结果最优,其次是 CLR.从平均排序结果来看,JMLLC-LR 都要优于其他的对比算法.

根据表 6 中所列出的各种算法在以 Average Precision 作为性能评价指标的结果可以看出:对于前 9 个数据集,除了在 slashdot 数据集上 JMLLC-LR 排在第 3 位之外,在其他的数据集上,本文的算法 JMLLC(包括 JMLLC-LR 和 JMLLC-LS)性能都要优于其他的算法,JMLLC-LR 和 JMLLC-LS 性能很稳定并且分别在 4 个和 5 个数据集上性能最好;对于 Yahoo 数据集,JMLLC-LS 的平均结果最优,其次是 JMLLC-LR,而没有考虑标记相关性的 BR 算法平均结果最差.从平均排序结果可以看出:JMLLC-LS 位居第 1 位,JMLLC-LR 排在第 2 位.

根据以上的实验结果分析以及表 2~表 6 中列出的在 5 个评价准则上的实验结果,充分表明了本文的算法优于其他的 8 种对比算法.对于 Yahoo 数据集且评价准则为 Coverage 和 Ranking Loss 时,JMLLC-LR 要优于 JMLLC-LS;其他情况下,JMLLC-LR 和 JMLLC-LS 性能相当.这充分证明了本文算法 JMLLC 选择不同的凸损失函数作为经验损失项都能取得较好的性能,从而验证了本文算法的有效性和可扩展性.

为了在统计上比较本文算法与其他对比算法在 20 个数据集上的实验结果,我们采用显著性水平为 5%的 Friedman test[32].对于每个评价准则,由于都拒绝了零假设(所有算法性能相等),故需要结合特定的 post-hoc test 来进一步分析各个算法性能的差异[31,32].本文采用显著性水平为 5%的 Nemenyi test,若两个算法在所有数据集上的平均排序的差不低于临界差(critical difference,简称 CD),则认为它们有显著性差异.图 1 给出了在不同评价准则下所有算法之间的比较,其中,每个子图中最上行为临界值 *CD*=3.03(10 种算法、20 个数据集),坐标轴画出了各种算法的平均排序且最左(右)边的平均排序最高(低).用一根加粗的线连接性能没有显著差异的算法组.



(a) Hamming loss (b) One-Error (c) Coverage

(d) Ranking loss (e) Average precision

Fig.1 Performance comparison of all classifiers in terms of five evaluation criteria

图 1 所有分类器在 5 种评价准则上的性能比较

对于每种算法,都有 45 种实验比较结果(9 种对比算法、5 种评价准则).从图 1 可以发现:

- 对于 JMLLC-LR 算法,31.1%的情况下(有 14 种比较结果与其他算法没有显著性差异),在统计上与其他算法性能相当,即:基于 Hamming Loss 时,与 JMLLC-LS,ECC 和 MLkNN 性能相当(如图 1(a)所示);基于 One-Error 时,与 JMLLC-LS 和 ECC 和性能相当(如图 1(b)所示);基于 Coverage 时,与 CLR, RankSvm 和 ECC 性能相当(如图 1(c)所示);与 CLR,RankSvm 和 ECC 在评价准则为 Ranking Loss 时性能相当(如图 1(d)所示);与 JMLLC-LS,ECC 和 CLR 的 Average Precision 值没有显著性差异(如图 1(e)所示).而 68.9%的情况下,在统计上要优于其他算法;
- 对于 JMLLC-LS,在 40%的情况下,在统计上与其他算法性能相当;在 4.4%的情况下,在统计上比其他算法的性能要差一些;在 55.6%的情况,在统计上要优于其他算法;
- 对于 ECC,在 64.4%的情况下,在统计上与其他算法性能相当;在 35.6%的情况下,在统计上要优于其他算法.

总的来说:JMLLC-LR 性能最优,在 68.9%的情况下,在统计上优于其他比较算法,而且不比其他算法性能要差;其次是 JMLLC-LS 算法,在 55.6%的情况下,在统计上优于其他比较算法;排在第 3 的是 ECC 算法.

这些结果与分析进一步验证了本文算法的有效性.

4 结束语

本文提出了多标记分类和标记相关性的联合学习 JMLLC(包括 JMLLC-LR 和 JMLLC-LS).通过构建基于类别标记变量的条件依赖网络,其中每个标记变量将其他标记变量和输入特征变量当作其父节点,来充分挖掘标记之间的相关性.最终的目标优化问题可转化为一个双凸优化问题,并可采用交替求解的方法进行求解.

在 20 个多标记数据集基于 5 种不同的评价准则的实验结果表明,JMLLC-LR 和 JMLLC-LS 算法优于其他 8 种对比算法.

在现实世界中,获取有标记样本非常费时且代价很高,而大量的未标记样本却很容易获得.因此在未来的研究工作中,我们将致力于研究基于半监督的多标记学习算法.此外,获取大量标记不完整的弱标记样本也是相对比较容易的,故针对弱标记的多标记学习也是未来的研究方向之一.

致谢 我们衷心感谢各位老师(尤其是各位审稿专家)和同学给本文提出宝贵的建议以及对本文工作的支持.

References:

[1] Zhang ML, Zhou ZH. A review on multi-label learning algorithms. IEEE Trans. on Knowledge and Data Engineering, 2014,26(8): 1819−1837. [doi: 10.1109/TKDE.2013.39]

[2] Schapire RE, Singer Y. Boostexter: A boosting-based system for text categorization. Machine Learning, 2000,39(2):135−168. [doi: 10.1023/A:1007649029923]

[3] Zhang ML, Zhou ZH. Multilabel neural networks with applications to functional genomics and text categorization. IEEE Trans. on Knowledge and Data Engineering, 2006,18(10):1338−1351. [doi: 10.1109/TKDE.2006.162]

[4] Boutell MR, Luo J, Shen X, Brown CM. Learning multi-label scene classification. Pattern Recognition, 2004,37(9):1757−1771. [doi: 10.1016/j.patcog.2004.03.009]

[5] Lo HY, Wang JC, Wang HM, Lin SD. Cost-Sensitive multi-label learning for audio tag annotation and retrieval. IEEE Trans. on Multimedia, 2011,13(3):518−529. [doi: 10.1109/TMM.2011.2129498]

[6] Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: Dietterich TG, Becker S, Ghahramani Z, eds. Proc. of the Advances in Neural Information Processing Systems 14. Cambridge: MIT Press, 2002. 681−687.

[7] Zhang ML, Zhou ZH. ML-KNN: A lazy learning approach to multi-label learning. Pattern Recognition, 2007,40(7):2038−2048. [doi: 10.1016/j.patcog.2006.12.019]

[8] Fürnkranz J, Hüllermeier E, Mencía EL, Brinker K. Multilabel classification via calibrated label ranking. Machine Learning, 2008, 73(2):133−153. [doi: 10.1007/s10994-008-5064-8]

[9] Read J, Pfahringer B, Holmes G, Frank E. Classifier chains for multi-label classification. Machine Learning, 2011,85(3):333−359. [doi: 10.1007/s10994-011-5256-5]

[10] Huang SJ, Zhou ZH. Multi-Label learning by exploiting label correlations locally. In: Proc. of the 26th AAAI Conf. on Artificial Intelligence. Toronto, 2012. 949−955.

[11] Huang SJ, Yu Y, Zhou ZH. Multi-Label hypothesis reuse. In: Proc. of the 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2012. 525−533. [doi: 10.1145/2339530.2339615]

[12] Tsoumakas G, Katakis I, Vlahavas I. Random *k*-labelsets for multilabel classification. IEEE Trans. on Knowledge and Data Engineering, 2011,23(7):1079−1089. [doi: 10.1109/TKDE.2010.164]

[13] Tsoumakas G, Katakis I. Multi-Label classification: An overview. Int'l Journal of Data Warehousing and Mining, 2007,3(3):1−13. [doi: 10.4018/jdwm.2007070101]

[14] Hüllermeier E, Fürnkranz J, Cheng WW, Brinker K. Label ranking by learning pairwise preferences. Artificial Intelligence, 2008, 172(16):1897−1916. [doi:10.1016/j.artint.2008.08.002]

[15] Xu JH. Fast multi-Label core vector machine. Pattern Recognition, 2013,46(3):885−898. [doi: 10.1016/j.patcog.2012.09.003]

[16] Xu JH. An extended one-versus-rest support vector machine for multi-label classification. Neurocomputing, 2011,74(17): 3114−3124. [doi:10.1016/j.neucom.2011.04.024]

[17] Wu L, Zhang ML. Multi-Label classification with unlabeled data: An inductive approach. In: Proc. of the 5th Asian Conf. on Machine Learning. Canberra, 2013. 197−212.

[18] Zhang ML. An improved multi-label lazy learning approach. Journal of Computer Research and Development, 2012,49(11): 2271−2282 (in Chinese with English abstract).

[19] Zhang ML, Peña JM, Robles V. Feature selection for multi-label naive Bayes classification. Information Sciences, 2009,179(19): 3218−3229. [doi: 10.1016/j.ins.2009.06.010]

[20] Zhang ML, Zhang K. Multi-Label learning by exploiting label dependency. In: Proc. of the 16th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2010. 999−1008. [doi: 10.1145/1835804.1835930]

[21] Clare A, King RD. Knowledge discovery in multi-label phenotype data. In: Raedt LD, Siebes A, eds. LNCS 2168. Berlin: Springer-Verlag, 2001. 42−53. [doi: 10.1007/3-540-44794-6_4]

[22] Ghamrawi N, McCallum A. Collective multi-label classification. In: Proc. of the 14th ACM Int'l Conf. on Information and Knowledge Management. New York: ACM Press, 2005. 195−200. [doi: 10.1145/1099554.1099591]

[23] Ji SW, Tang L, Yu SP, Ye JP. Extracting shared subspace for multi-label classification. In: Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining. New York: ACM Press, 2008. 381−389. [doi: 10.1145/1401890.1401939]

[24] Zhang T, Yeung DY. Multilabel relationship learning. ACM Trans. on Knowledge Discovery from Data, 2013,7(2):Article 7. [doi: 10.1145/2499907.2499910]

[25] Guo YH, Xue W. Probabilistic multi-label classification with sparse feature learning. In: Rossi F, ed. Proc. of the 23rd Int'l Joint Conf. on Artificial Intelligence. AAAI Press, 2013. 1373−1379.

[26] Guo Y, Gu SC. Multi-Label classification using conditional dependency networks. In: Walsh T, ed. Proc. of the 22nd Int'l joint Conf. on Artificial Intelligence. AAAI Press, 2011. 1300−1305. [doi:10.5591/978-1-57735-516-8/IJCAI11-220]

[27] Heckerman D, Chickering DM, Meek C, Rounthwaite R, Kadie C. Dependency networks for inference, collaborative filtering, and data visualization. The Journal of Machine Learning Research, 2001,1:49−75. [doi:10.1162/153244301753344614]

[28] Godbole S, Sarawagi S. Discriminative methods for multi-labeled classification. In: Dai H, Srikant R, Zhang C, eds. Lecture Notes in Artificial Intelligence 3056, Berlin: Springer-Verlag, 2004. 22−30. [doi: 10.1007/978-3-540-24775-3_5]

[29] Schölkopf B, Herbrich R, Smola AJ. A generalized representer theorem. In: Helmbold D, Williamson B, eds. Lecture Notes in Artificial Intelligence 2111. Berlin: Springer-Verlag, 2001. 416−426. [doi: 10.1007/3-540-44581-1_27]

[30] Tsoumakas G, Xioufis ES, Vilcek J, Vlahavas I. MULAN: A Java library for multi-label learning. Journal of Machine Learning Research, 2011,12(7):2411−2414.

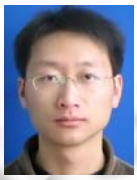
[31] Demšar J. Statistical comparisons of classifiers over multiple data sets. The Journal of Machine Learning Research, 2006,7(1): 1−30.

[32] Zhang ML. LIFT: Multi-Label learning with label-specific features. In: Walsh T, ed. Proc. of the 22nd Int'l joint Conf. on Artificial Intelligence. AAAI Press, 2011. 1609−1614. [doi: 10.5591/978-1-57735-516-8/IJCAI11-270]

附中文参考文献:

[18] 张敏灵.一种新型多标记懒惰学习算法.计算机研究与发展,2012,49(11):2271−2282.

何志芬(1988－),女,江西萍乡人,博士生,主要研究领域为模式识别,机器学习.
E-mail: hzfnjnu@gmail.com



刘会东(1987－),男,硕士,主要研究领域为机器学习,计算机视觉.
E-mail: h.d.liew@gmail.com

杨明(1964－),男,博士,教授,博士生导师,主要研究领域为数据挖掘,模式识别,机器学习与应用.
E-mail: m.yang@njnu.edu.cn