

基于标记因果顺序挖掘的多标记分类方法^{*}

陈加略, 姜远



(南京大学 计算机科学与技术系, 江苏 南京 210023)

通信作者: 姜远, E-mail: jiangyuan@nju.edu.cn

摘要: 在多标记学习(MLL)问题中, 每个示例都与一组标记相关联. 为了实现对未见示例的高效预测, 挖掘和利用标记之间的关系是至关重要的. 大多数已有的研究都将关系简化为标记之间的相关性, 而相关性又通常基于标记的共现性. 揭示了因果关系对于描述一个标记在学习过程中如何帮助另一个标记更为重要. 基于这一观察, 提出了两种策略来从标记因果有向无环图(DAG)中生成标记的因果顺序, 同时使得生成的因果顺序都遵循因标记应该在果标记之前的准则. 第1种策略的主要思想是对随机顺序进行排序, 使其满足DAG中的因果关系. 而第2种策略的主要思想是根据DAG的结构, 将标记放入许多不相交的拓扑层次中, 再通过它们的拓扑结构进行排序. 进一步, 通过将因果顺序纳入到分类器链(CC)模型中, 提出了一种有效的MLL方法, 从而从更加本质的角度来利用标记关系. 在多个数据集上的实验结果验证了该方法确实能够挖掘出有效的标记因果顺序, 并帮助提升学习性能.

关键词: 多标记学习; 因果; 标记顺序; 有向无环图; 分类器链

中图法分类号: TP181

中文引用格式: 陈加略, 姜远. 基于标记因果顺序挖掘的多标记分类方法. 软件学报, 2022, 33(4): 1267-1273. <http://www.jos.org.cn/1000-9825/6475.htm>

英文引用格式: Chen JL, Jiang Y. Multi-label Learning by Exploiting Causal Order of Labels. Ruan Jian Xue Bao/Journal of Software, 2022, 33(4): 1267-1273 (in Chinese). <http://www.jos.org.cn/1000-9825/6475.htm>

Multi-label Learning by Exploiting Causal Order of Labels

CHEN Jia-Lüe, JIANG Yuan

(Department of Computer Science and Technology, Nanjing University, Nanjing 210023, China)

Abstract: In multi-label learning (MLL) problems, each example is associated with a set of labels. In order to train a well-performed predictor for unseen examples, exploiting relations between labels is crucially important. Most exiting studies simplify the relation as correlations among labels, typically based on their co-occurrence. This study discloses that causal relations are more essential for describing how a label can help another one during the learning process. Based on this observation, two strategies are proposed to generate causal orders of labels from the label causal directed acyclic graph (DAG), following the constraint that the cause label should be prior to the effect label. The main idea of the first strategy is to sort a random order to make it satisfied the cause-effect relations in DAG. And the main idea of the second strategy is to put labels into many non-intersect topological levels based on the structure of the DAG, then sort these labels through their topological structure. Further, by incorporating the causal orders into the classifier chain (CC) model, an effective MLL approach is proposed to exploit the label relation from a more essential view. Experiments results on multiple datasets validate that the extracted causal order of labels indeed provides helpful information to boost the performance.

Key words: multi-label learning; causality; label-order; directed acyclic graph; classifier chain

在监督学习场景中, 每个示例都只有一个标记来描述其语义. 然而在很多实际应用中, 为了充分描述对象的语义, 需要为每个示例标注一个以上的标记. 为了处理这种应用场景, 多标记学习^[1,2]假设训练集中的每

* 基金项目: 国家自然科学基金(61673201, 61921006)

本文由“面向开放场景的鲁棒机器学习”专刊特约编辑陈恩红教授、李宇峰副教授、邹权教授推荐.

收稿时间: 2021-05-29; 修改时间: 2021-07-16; 采用时间: 2021-08-27; jos 在线出版时间: 2021-10-26

个示例都由一个特征向量表示,并与一组标记相关联,而其任务则是预测未见示例的标记集.从形式上看,多标记学习的关键是训练一个从特征空间到标记空间的映射.然而,从标记的幂集中考虑所有的组合是不切实际的,因此多标记学习的任务将变得相当具有挑战性.为了解决这个问题,如何有效利用标记之间的关系来促进多标记的学习是其中的关键.分类器链(CC)^[3]算法作为一种具有代表性的多标记学习方法,试图将每个示例的标记代入一个确定性的马尔科夫链模型中,以获取标记之间的相关性.在CC的训练过程中,会按照特定的顺序训练出一组分类器,其中,它们的训练输入的是示例的特征向量和之前标记的集合.由于分类器的性能对标记顺序的选择很敏感,这个问题的关键就成了如何确定CC的最佳顺序.

因果关系^[4,5]是一种描述随机变量之间潜在的数据生成过程的更加本质的关系.通过观察数据发现因果关系,在很多领域都发挥了举足轻重的作用,如医疗卫生^[4]和经济学研究^[6].在多标记学习中,通过将每个标记视为一个变量,标记之间可能存在一些自然的因果关系,而这种因果关系通常比标记之间的共现关系更加能够反映数据的结构信息.例如,仙人掌通常生长在沙漠地区,而沙漠上相对干燥的环境可能是仙人掌更加适合生活的原因.因此,在图像数据集^[7]中,标记沙漠也可以成为标记仙人掌的因.直观上,特征集作为标记的因是被普遍接受的,但这种常识即使在图像分类问题中也未必成立.例如,一个人想要写数字7,这个需求会导致一个运动模式产生数字7的图像.在这种情景下,作为需求表示的数字的类标记可以看成是其图像的因.这种因果顺序相反的情况,也被称为反向因果关系推断^[8].

在本文中,将使用有向无环图(DAG)来描述标记之间的因果关系^[4].在这个DAG中,每个顶点对应一个标记,每条边对应两个标记之间的因果关系.基于标记因果DAG的描述方式,我们提出了两种新颖的抽取标记顺序的方法,并将其与CC算法相结合.不幸的是,在大多数多标记任务中,因果关系(即DAG中的边)是未知的,所以必须探索标记之间的因果关系.为了简单起见,本文将使用广泛应用的PC算法^[9].通过计算标记之间的条件独立性,来探索DAG中的因果关系.当我们知道DAG中的所有边的方向后,通过DAG的结构和本文策略,将生成一个特定的标记顺序,这个标记顺序将一定满足因标记应该出现在果标记之前的合法性定义.最后,标记的顺序将决定CC算法的分类器的训练序列.更具体地说,在每一轮训练中,一个示例的特征向量和该轮之前顺序的标记都将被视为当前轮分类器的输入.与其他优化方法相比,本文的策略具有以下优点:1) 本文的策略是非启发式的优化方法,因此在期望上具有更强的自适应性;2) 本文的策略易于实现;3) 本文的策略可以同时适用于正向和反向因果推理.大量的实验验证了我们的方法可以为CC算法挖掘出有效的标记因果序列.

1 相关工作

现有的多标记学习方法可以根据其利用的标记之间相关性的阶数大致分为3类:一阶方法将MLL问题分解为若干个独立的二元分类问题^[10-12];二阶方法考虑的是标记之间的对偶关系^[13,14];高阶方法考虑的则是标记之间更高阶的关系^[15-18].分类器链是一种典型的高阶方法,即根据特定的标记顺序依次训练多个分类器.基于现在的研究,已经有一些优化方法被提出来,解决CC中标记顺序的问题.在文献[19]中,作者提出了一个概率框架,从风险最小化和贝叶斯最优预测的角度来看待MLL问题,然后将其应用到CC中,从而得到了得到probabilistic CC.在文献[20]中,作者提出了一个贝叶斯网络,来在不完全赋值的情况下,系统地获取不同标记之间的依赖关系.

因果发现是因果研究的工作中最关键的问题之一.基于约束的方法是因果发现方法的一种,它通过随机变量之间的条件独立性来探寻因果关系.PC算法是一种著名的基于约束的因果发现方法,它假设变量之间的所有条件独立性都是基于因果马尔科夫条件的.已经有很多研究工作在考虑在多标记学习问题中利用因果关系.在文献[21]中,作者提出了一种MLL算法,该算法使用贝叶斯网络结构来有效地编码标记和特征集之间的条件依赖性.在文献[22]中,作者从特征的角度出发,通过在因果特征的超平面上编码不同标记对应的边距来发现标记之间的因果关系.这些工作大多假设特征集是标记的共同原因.但遗憾的是,即使在一些通用的应用中,这个假设也未必成立.我们唯一能够利用的资源就只剩下标记之间的信息,因为此时特征集无法对

标记之间的因果推断有任何贡献. 以 PC 算法为例, 我们只能利用标记之间的条件独立性约束来生成 DAG.

2 预备知识

2.1 符号说明

设 $\mathcal{X}=\mathbb{R}^d$ 是 d 维的输入空间, 设 $\mathcal{Y}=\{1,2,\dots,q\}$ 是标记空间. 给定一个多标记训练集 $\mathcal{D}=\{(x^i,y^i)|1<i<m\}$, 其中, $x^i\in\mathcal{X}$ 是一个特征向量, 而 $y^i=\{y_1^i,\dots,y_q^i\}\in\mathcal{Y}$ 是示例 x^i 的标记集合. 多标记学习的目标是, 从数据集 \mathcal{D} 训练一个由未知示例到一个特定标记集合的映射函数 $h: X \rightarrow 2^{\mathcal{Y}}$.

2.2 分类器链算法

分类器链方法将函数 h 作为一个确定性的高阶马尔科夫链模型. 对于一个给定示例 x , 标记 y_p 只与示例特征以及标记 y_p 的所有前置标记 $PA(y_p)$ 有关, 其中, $PA(y_p)=\{y_i|y_i \text{ 先于 } y_p\}$, 所以 y_p 的后验概率可以写成 $P(y_p|x, PA(y_p))$. 在本文中, 后验概率由线性模型 $y_p = w_p^T * [x; PA(y_p)]$ 来实现. 给定一个固定顺序 $O=\{O_1, O_2, \dots, O_q| 1 \leq O_i \leq q\}$, 模型将通过 O 逐一训练. 以 x 为例, 标记集 y 的后验概率可以写成:

$$P(y|x) = P(y_{O_1}|x)P(y_{O_2}|x, y_{O_1}), \dots, P(y_{O_q}|x, y_{O_1}, y_{O_2}, \dots, y_{O_{q-1}}) \quad (1)$$

在测试阶段, 预测过程也应遵循 O 的顺序. 将标记 y_p 的预测值记为 \hat{y}_p , 其值也只与示例特征及其前置标记 $PA(\hat{y}_p)=\{y_i|y_i \text{ 先于 } \hat{y}_p\}$ 的有关. 于是, \hat{y}_p 的表述可以写成 $\hat{y}_p = w_p^T * [x; PA(\hat{y}_p)]$.

2.3 PC算法

如果变量 A, B, C 之间满足因果关系 $A \rightarrow B \rightarrow C$, 那么在给定变量 B 的情况下, 可以发现变量 A 和变量 C 条件独立. 相对的, 基于变量之间的条件独立性, 我们也可以反向推测他们之间的因果关系. 凡是基于这个思路来复原变量间因果结构的算法, 也被称为基于条件独立性(CI)的算法, 而 PC 算法正是其中被广泛使用的一种. PC 算法假设变量之间的条件独立性都必须遵循 Markov 条件, 并且假定数据中不存在混淆因子(潜在的双变量的共同因变量), 而其输出将会是一个有向无环图(DAG). 如果我们将每个标记看作是一个随机变量, 那么学习标记/变量之间的因果关系就变成了一个因果发现问题.

2.4 DAG因果结构

有向无环图(DAG)可以表示为 $G(V,E)$, 其中, V 是顶点集, E 是边集. 在我们的工作中, DAG 用于描述标记之间的因果结构. V 与标记相关, E 用于描述两个标记之间的因果关系. 在本文中, 我们只关心两种标记之间的关系: 第 1 种是独立关系, 例如, 如果 y_i 和 y_j 是独立的, 则边 $e_{i,j}$ 和 $e_{j,i}$ 都将被设为 0; 另一种关系是因果方向, 例如, 如果 y_i 是 y_j 的因, 那么边缘 $e_{i,j}$ 将被设为 1, 而边缘 $e_{j,i}$ 将被设为 0. 如果我们将每个标记看作是一个随机变量, 那么学习标记/变量之间的因果关系就变成了一个因果发现问题. 本文将采用 PC 算法, 通过计算标记/变量之间的条件独立性来探索 DAG 中的因果关系.

3 提出的算法

本文的动机是: 因标记可能对果标记的学习起到更大的作用, 所以训练时也应该排在果标记之前. 换句话说, 只有因标记排在果标记之前的顺序才会被定义为合法顺序. 所以对于 CC 算法, 很自然地想到: 如果我们需要训练 y_i 和 y_j 的分类器, 其中 y_i 是 y_j 的因, 那么 y_i 应该被加入到 y_j 的分类器的输入中. PC 算法可以用来发现 G 中标记的因果关系, 其中每个标记将被看作是一维的随机变量. 捕捉标记顺序的一个直接策略是随机生成一个顺序, 并检查 G 中的所有因果关系是否都遵循合法性假设. 但遗憾的是, 如果图 G 不是由许多相互不相交的连通分量组成的, 那么只会有少数的随机顺序是合法并可接受的, 因而这种策略可能会非常耗时. 为了避免这种情况, 本文提出了两种新颖而高效的启发式方法来获取标记的因果顺序, 然后结合 CC 算法来实现高性能的多标记分类模型.

3.1 DAG的预处理

理想情况下, 标记 DAG 中应该无环并且标记之间的所有因果关系都应该有明确的方向. 但遗憾的是, 并不是所有的因果关系都能通过标记之间的条件独立性检测出来. 更糟糕的是, 在许多现实世界的多标记任务中, 可能存在一些固有的未知标记, 从而导致因果发现算法得到的 DAG 中包含环. 在这些情况下, 我们只能使用局部有向图(PDG)来描述标记的因果结构. PDG 可以包含环或有向边. 一个有向边 (i,j) 意味着所采用的因果发现算法没有检测到因果关系的方向, 记为 $e_{i,j}=1$ 和 $e_{j,i}=-1$.

由于我们不能确定一个环内和无相边两端标记的顺序, 所以在本文中, 我们决定忽略这些不明确的关系, 而将其相关的标记以任意的顺序排列. 基于这个想法, 我们决定通过设置 $e_{i,j}=0$ 来删除所有环内边以及无向边. 由于此时图中只含有向边就成了严格的 DAG, 这个时候图上所有标记就不会存在无法确定的顺序关系了.

3.2 选择排序策略

选择排序策略(switchable sorting strategy, SSS)的主要思想: 先随机生成一个序列, 再通过迭代交换位置的方式来修正所有没有符合合法性要求的因果关系对. 图 1 展示了 SSS 的过程. 假设 y_3 是 y_1 的因. 在第 1 轮, 标记的顺序 O 为 $\{y_1, y_2, y_3, y_4, y_5\}$. 此时, y_3 的顺序在 y_1 之前, 这是非法的. 所以在下一轮, y_3 的顺序应该放在 y_1 的前面, O 就转化为 $\{y_3, y_1, y_2, y_4, y_5\}$.

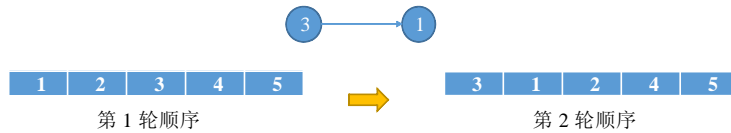


图 1 选择排序策略的过程说明

图 2 介绍了 SSS 的流程. SSS 的输入为随机顺序, 输出为 CC 的合法顺序. 在迭代的过程中, 将检查每一个因果关系对. 如果它们的顺序是不合法的, 则果标记应该放在因标记之后. 由此不难看出, SSS 生成的顺序对于 CC 来说一定是合法的.

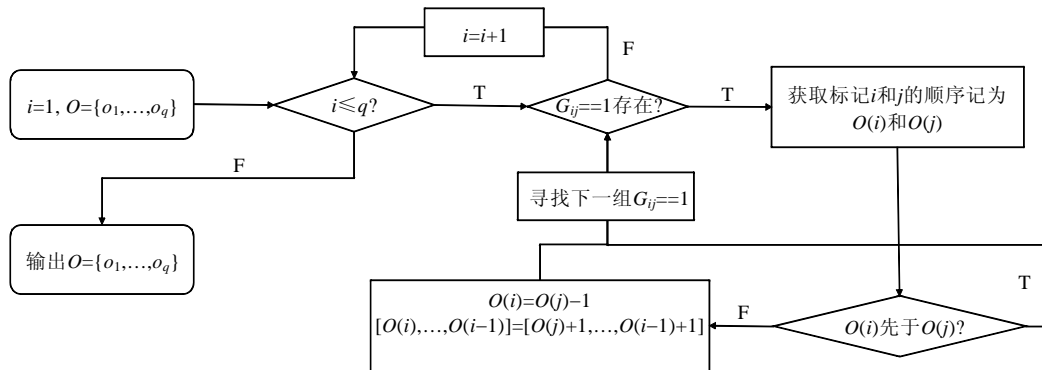


图 2 选择排序策略的流程图

3.3 拓扑生成策略

拓扑生成策略(topological generating strategy, TGS)的主要思想是: 根据图 G 的结构, 将标记放入许多不相交的拓扑层次中. 因标记应该在较先的层次, 而果标记应该在较后的层次, 所以较先低层次标记的顺序一定先于较后层次的标记.

图 3 介绍了 TGS 的流程. TGS 包括两个主要的步骤.

- 第 1 步, 根据 G 的拓扑结构生成一个非相交层级表 l_s , 然后扫描 G , 找到所有只有输出或独立(无输入输出)的标记并且没有被重复访问的标记. 由于这些标记的优先级最高, 所以将被添加到 l_s 的末尾.

此后, 这些标记就不需要再被访问了, 它们的输出也可以在 G 中被删除. 在反复迭代第 1 步之后, 最后就只剩下独立变量了, 它们就可以直接放置在 ls 的最后一层. 在流程图中, lc 和 vr_d 是两个辅助变量: lc 用于统计层数, 而 vr_d 则记录标记是否已经被访问;

- 第 2 步, 我们会逐层扫描 ls 来生成 O . 由于较先层级的标记会优先安排, 所以他们的顺序一定会排在较后层级的标记前面, 从而保证了顺序的合法性. 对于同一层的标记而言, 由于他们的优先顺序无法比较, 所以可以随机分配.

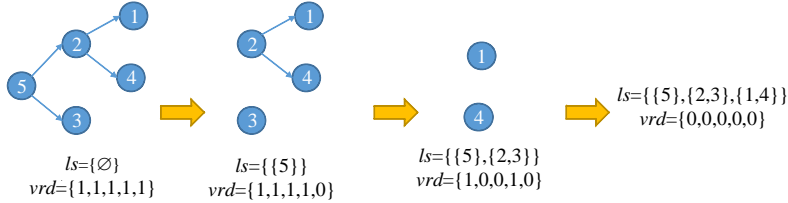


图 3 拓扑生成策略的流程图

图 4 展示了 TGS 的过程, 图中显示了标记 y_1 到 y_5 之间的因果关系. 在第一轮中, ls 是一个空集, vr_d 是全 1 的向量. 在这一轮中, y_5 处于最高优先级, 应该得到第一顺序, 所以 5 被添加到 ls 的末尾, $vr_d(5)$ 也应该被设为 0, 并且在图中 y_5 的输出边需要被删除. 在第二轮, y_2 是 y_1 和 y_4 的因, 所以 y_2 的级别应该高于 y_1 和 y_4 . 同时, y_3 已经是一个独立的标记, 它与其他标记没有顺序. 此时满足因果顺序合法性假设的顺序包括 $\{5,3,2,1,4\}$, $\{5,2,3,1,4\}$, $\{5,2,1,3,4\}$, $\{5,2,1,4,3\}$, $\{5,3,2,4,1\}$, $\{5,2,3,4,1\}$, $\{5,2,4,3,1\}$, $\{5,2,4,1,3\}$ 共 8 种顺序组合. 可想而知: 随着独立标记数目的增加, 如果想要输出全部满足条件的顺序, 那么算法的复杂度将是指数级别的, 这么做的话, 算法的效率就太低了. 所以为了减少随机性, 该策略不得不做一个假设, 假设方法每一轮出现的独立标记都拥有最高的顺位, 即这些所有当前独立的标记也放在 ls 的末尾. 因而, $\{2,3\}$ 被添加到 ls 的较后一级, 并在图中删除 y_2 与 y_3 的输出边; 在最后一轮, 只剩下独立的标记 y_1 和 y_4 , 所以可以直接置于 ls 的最后一层. 在得到 G 的拓扑层次结构 ls 后, 将执行 TGS 的第二个步骤, 通过逐层按序生成的方式从 ls 中获取一个符合 CC 方法的合法顺序.

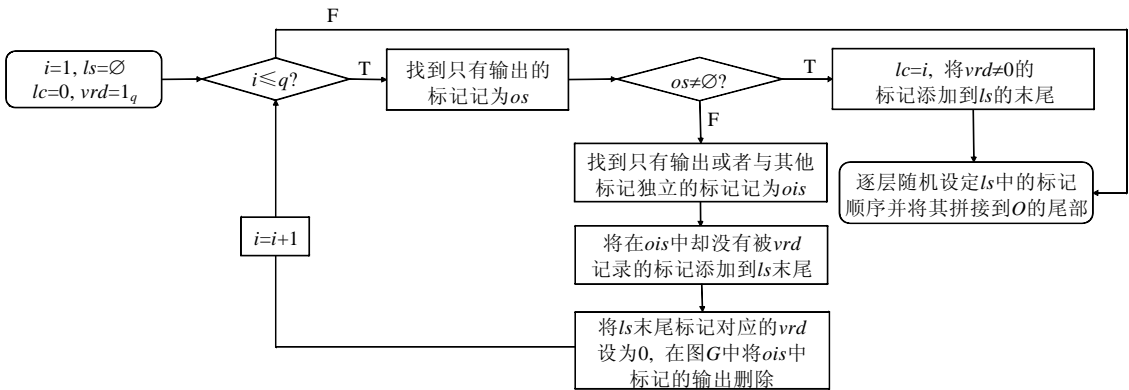


图 4 拓扑生成策略的过程说明

4 实验

4.1 实验设定

本文的实验旨在验证因果结构可以帮助基于顺序的多标记模型生成更好的模型. 由于文献[19,20]使用不同的算法技巧来生成的 CC 模型的, 而这些方法与本文的算法相互兼容, 可以相互组合. 不过, 由于本文的侧

重点集中于因果顺序的有效性, 所以决定不选取他们的方法作为对比方法. 此外, 文献[21,22]中没有采用反事实假设, 所以它们不能适应我们的任务, 也不会被选为对比算法. 综上所述, 我们只能将本文的策略与基线方法 Binary Relevance(BR)^[1]以及随机顺序的 CC 进行比较, 其标记为 CC-R. 而基于我们策略实现的 CC 方法, 将标记为 CC-SSS 和 CC-TGS. 为了公平起见, Libsvm^[23]将作为 BR 和 CC 的基分类器. 我们使用了 4 个真实世界的多标记数据集: (1) 电子邮件数据集 Enron, 有 53 个标记; (2) 临床文本数据集 Medical, 有 45 个标记; (3) 文档数据集 Reuters, 有 7 个标记; (4) 生物数据集 Yeast, 有 14 个标记. 70%/30% 的示例会被划分为训练/测试数据集.

对比算法采用各自文献中建议的参数. 对于基分类器, 模型是通过交叉训练策略学习的. 对于 CC 算法, 将生成 10 组随机顺序并通过它们的平均值来进行结果评价. 每个实验数据集都会进行十次交叉验证, 为了比较我们的方法与其他方法之间的结果, 本文选取了 3 个评价指标:

$$MicroF1(H) = \frac{2 \sum_{j=1}^q \sum_{i=1}^m y_{i,j} h_{i,j}}{\sum_{j=1}^q \sum_{i=1}^m y_{i,j} + \sum_{j=1}^q \sum_{i=1}^m h_{i,j}} \quad (2)$$

对于 $MicroF1(H)$, 越高越好;

$$MacroF1(H) = \frac{1}{q} \sum_{j=1}^q \frac{2 \sum_{i=1}^m y_{i,j} h_{i,j}}{\sum_{i=1}^m y_{i,j} + \sum_{i=1}^m h_{i,j}} \quad (3)$$

对于 $MacroF1(H)$, 越高越好;

$$HammingLoss(H) = \frac{1}{mq} \sum_{j=1}^q \sum_{i=1}^m [h_{i,j} \neq y_{i,j}] \quad (4)$$

对于 $HammingLoss(H)$, 越低越好.

4.2 实验结果

表 1 显示了 5 种方法在 4 个数据集上用 3 种不同的评估指标进行实验的性能, 其中, E- n 表示 10 的 n 次幂. 如果本文的其中一个策略比其他比较方法好, 那么比较方法的结果将被标记为“*”.

表 1 在多标记数据集上的结果

数据集	评价指标	策略			
		CC-SSS	CC-TGS	BR	CC-R
Enron	<i>MicroF1</i>	0.5623±6.7E-3	0.5657±3.4E-3	0.5552±7.3E-3*	0.5590±7.6E-3*
	<i>MacroF1</i>	0.1444±1.1E-2	0.1463±7.1E-3	0.1338±7.7E-3*	0.1422±5.8E-3*
	<i>HammingLoss</i>	0.0473±9.7E-4	0.04701±7.8E-4	0.0476±1.1E-3*	0.0473±1.2E-3*
Medical	<i>MicroF1</i>	0.7858±1.4E-2	0.7844±1.6E-2	0.7848±1.3E-2*	0.7856±1.2E-2*
	<i>MacroF1</i>	0.3132±1.2E-2	0.3140±1.2E-2	0.3053±1.3E-2*	0.3189±9.4E-3
	<i>HammingLoss</i>	0.0118±7.5E-4	0.0118±8.1E-4	0.0119±7.3E-4*	0.0118±0.6E-3
Reuters	<i>MicroF1</i>	0.8237±1.2E-2	0.8242±9.1E-3	0.7977±1.3E-2*	0.8128±8.7E-3*
	<i>MacroF1</i>	0.7376±1.5E-2	0.7380±1.4E-2	0.7007±1.8E-2*	0.7142±1.4E-2*
	<i>HammingLoss</i>	0.0527±3.2E-3	0.0527±2.5E-3	0.0583±3.5E-3*	0.0552±2.4E-3*
Yeast	<i>MicroF1</i>	0.6283±0.9E-2	0.6289±8.3E-3	0.6258±9.5E-3*	0.6271±9.6E-3*
	<i>MacroF1</i>	0.3363±1.1E-2	0.3372±6.6E-3	0.3294±7.6E-3*	0.3343±7.2E-3*
	<i>HammingLoss</i>	0.2027±4.4E-3	0.2026±4.4E-3	0.2024±4.8E-3	0.2033±5.1E-3*

可以观察到: 在大多数数据集上, 基于本文的策略生成顺序的 CC 算法获得了最好的性能, 他们只在 Medical 上输给了随机顺序. 所以可以得出结论, 因果关系确实有助于充分利用标记之间的信息. 除此以外可以发现: 在大多数情况下, TGS 的性能都比 SSS 更好. 我们考虑了以下可能的原因:

- 1) 在每一轮迭代中, TGS 策略倾向于给出独立标记的最优先的顺序. 因此, 这些独立标记的分类器不会包含任何其他标记的输入, 因而减少了无用的输入;

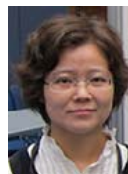
- 2) 在 TGS 中, 同一级别的标记会给出随机的顺序; 但在 SSS 中, 这些标记的顺序与它们初始序列中的顺序直接相关. 所以 TGS 与 SSS 相比, 对初始顺序更不敏感.

References:

- [1] Zhang ML, Zhou ZH. A review on multi-label learning algorithms. *IEEE Trans. on Knowledge and Data Engineering*, 2013, 26(8): 1819–1837.
- [2] Xie MK, Huang SJ. Partial multi-label learning with noisy label identification. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2021.
- [3] Jesse R, Bernhard P, Geo H, *et al.* Classifier chains for multi-label classification. *Machine Learning*, 2011, 85(3): 333.
- [4] Judea P. *Causality*. Cambridge University Press, 2009.
- [5] Jonas P, Dominik J, Bernhard S. *Elements of Causal Inference: Foundations and Learning Algorithms*. MIT Press, 2017.
- [6] James JH. Causal parameters and policy analysis in economics: A 20th century retrospective. *The Quarterly Journal of Economics*, 2000, 115(1): 45–97.
- [7] Feng JP, Wang XG, Liu WY. Deep graph cut network for weakly-supervised semantic segmentation. *Science China Information Sciences*, 2021, 64(3): 130105.
- [8] Bernhard S, Dominik J, Jonas P, *et al.* On causal and anticausal learning. arXiv: 1206.6471, 2012.
- [9] Peter S, Clark NG, Richard S, *et al.* *Causation, Prediction, and Search*. MIT Press, 2000.
- [10] Matthew RB, Luo JB, Shen XP, Christopher MB. Learning multi-label scene classification. *Pattern Recognition*, 2004, 37(9): 1757–1771.
- [11] Amanda C, Ross DK. Knowledge discovery in multi-label phenotype data. In: *Proc. of the European Conf. on Principles of Data Mining and Knowledge Discovery*. Springer, 2001. 42–53.
- [12] Jesse R, Bernhard P, Geoffrey H. Multi-label classification using ensembles of pruned sets. In: *Proc. of the 8th IEEE Int'l Conf. on Data Mining*. IEEE, 2008. 995–1000.
- [13] Andre E, Jason W. A kernel method for multi-labelled classification. In: *Proc. of the Advances in Neural Information Processing Systems*. 2002. 681–687.
- [14] Qi GJ, Hua XS, Rui Y, *et al.* Correlative multi-label video annotation. In: *Proc. of the 15th ACM Int'l Conf. on Multimedia*. 2007. 17–26.
- [15] Ji SW, Tang L, Yu SP, *et al.* Extracting shared subspace for multi-label classification. In: *Proc. of the 14th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*. 2008. 381–389.
- [16] Xu M, Guo LZ. Learning from group supervision: The impact of supervision deficiency on multi-label learning. *Science China Information Sciences*, 2021, 64(3): 1–13 (in Chinese with English abstract).
- [17] 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.
- [18] Zhang ML, Zhou ZH. MI-knn: A lazy learning approach to multi-label learning. *Pattern Recognition*, 2007, 40(7): 2038–2048.
- [19] Cheng WW, Eyke H, Krzysztof JD. Bayes optimal multilabel classification via probabilistic classifier chains. In: *Proc. of the 27th Int'l Conf. on Machine Learning (ICML 2010)*. 2010. 279–286.
- [20] Wang SF, Wang J, Wang ZY, *et al.* Enhancing multi-label classification by modeling dependencies among labels. *Pattern Recognition*, 2014, 47(10): 3405–3413.
- [21] 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*. 2010. 999–1008.
- [22] Xu C, Tao DC, Xu C. Large margin multi-label causal feature learning. In: *Proc. of the AAAI Conf. on Artificial Intelligence*. 2015. 1924–1930.
- [23] Chang CC, Lin CJ. Libsvm: A library for support vector machines. *ACM Trans. on Intelligent Systems and Technology (TIST)*, 2011, 2(3): Article 27.



陈加略(1991—), 男, 博士, CCF 学生会员, 主要研究领域为机器学习, 数据挖掘.



姜远(1976—), 女, 博士, 教授, 博士生导师, CCF 专业会员, 主要研究领域为机器学习, 数据挖掘.