Computer Science and Application 计算机科学与应用, 2025, 15(2), 200-208
Published Online February 2025 in Hans. https://www.hanspub.org/journal/csa
https://doi.org/10.12677/csa.2025.152047

Hans汉斯

基于条件随机场的多标签学习

张雨舒，陈琳琳，何　强

北京建筑大学理学院，北京

收稿日期：2025年1月22日；录用日期：2025年2月21日；发布日期：2025年2月28日

摘　要

多标签学习的目标是为每个样本分配一个或者多个标签的集合。在实际应用中，多标签之间通常存在复杂的依赖关系，这为模型的构建带来了挑战。通过将多标签学习问题转化为序列标注问题，能够充分利用标签之间的顺序依赖性，为多标签学习提供新思路。在此框架下，条件随机场(Conditional Random Fields, CRF)因其优异的序列建模能力和概率推断框架，被证明是一种有效的方法。CRF能够通过条件概率建模捕捉输入特征与标签之间的关系，并通过标签间的转移特征建模多标签间的依赖性。相比独立处理各标签的方法，CRF可以建模标签之间的相互影响，从而提高预测的准确性和一致性。通过进一步的理论探索和实践验证，CRF在多标签学习中的应用将变得更加广泛，为学习任务提供强有力的支持。

关键词

多标签学习，条件随机场，标签依赖性

Multi-Label Learning Based on Conditional Random Fields

Yushu Zhang, Linlin Chen, Qiang He

School of Science, Beijing University of Civil Engineering and Architecture, Beijing

Received: Jan. 22nd, 2025; accepted: Feb. 21st, 2025; published: Feb. 28th, 2025

Abstract

The goal of multi-label learning is to assign a set of one or more labels to each sample. In practical applications, there are often complex dependencies between multiple tags, which brings challenges to the construction of models. By transforming the multi-label learning problem into a sequential labeling problem, we can make full use of the order dependence between labels and provide a new idea for multi-label learning. Under this framework, Conditional Random Fields (CRF) proved to be an effective method due to its excellent sequence modeling ability and probabilistic inference

文章引用: 张雨舒, 陈琳琳, 何强. 基于条件随机场的多标签学习[J]. 计算机科学与应用, 2025, 15(2): 200-208.
DOI: 10.12677/csa.2025.152047

framework. CRF can capture the relationship between input features and labels through conditional probability modeling, and model the dependency between multiple labels through transition features between labels. CRF can model the interactions between labels to improve the accuracy and consistency of predictions compared to methods that treat each label independently. Through further theoretical exploration and practical verification, the application of CRF in multi-label learning will become more extensive and provide strong support for learning tasks.

Keywords

Multiple Label Learning, Conditional Random Fields, Label Dependence

1. 引言

多标签问题较于单标签学习更具有挑战性的原因有两个，一个是标签分配的可能性在标签数量上指数增长，另一个是不同的标签及其组合的出现并不是随意的，一些标签组合可能或多或少地反映了标签之间的依赖关系。这促使我们应该寻找更加简单并易于处理的形式来表示标签之间的依赖关系，并设计能够从数据中学习到这种依赖关系的算法。

目前主要的多标签学习方法中也有一些专注于挖掘和利用标签之间关系的方法。例如，考虑到标签之间潜在的相关性，Read 等人提出 Classifier Chains (CC) [1]算法，该方法利用问题转换方法将多标签问题转换为二分类问题，认为当前预测的标签的存在，对下一顺位标签的预测有一定的影响。该算法为每个类别训练一个分类器，再根据标签的顺序构成一个分类器链，每一个分类器都是在前一个分类器的预测结果基础上构建的。Ensembles of classifier chains (ECC) [2]算法是集成分类器链方法，可以对标签相关性进行建模，利用了机器学习中“集成”的想法来提高分类精度。Probabilistic classifier chains (PCC)算法[3]是通过穷举搜索最大后验概率来选择最佳预测变量，但穷举产生的指数级开销限制了它的应用。

多标签学习中，实例和标签间的对应关系是一对多的，这种对应关系隐含了不同标签之间必然存在着某种关系，关系可以分为正相关或者是负相关的关系，也就是一个标签对另一个标签的影响。因此深入了解多标签学习中标签间的依赖关系来提高模型性能是有必要的[4]-[6]。标签间的关系能为学习过程提供许多有效信息，给多标签学习提供更好的优化思想。所以本文将传统的多标签学习转化为序列标注问题，再利用条件随机场解决序列标注问题，提高模型效率。

2. 相关工作

根据以往的研究可知，多标签学习的解决方法可以大致分为两类，一类是问题转换方法，其中 BR [7]算法是根据标签将数据重新组成正负样本，针对每个类别的标签分别训练分类器，CC [1]算法是针对 BR 中标签关联性的问题，将这些分类器串联起来形成一条链，上一个分类器的输出作为下一个分类器的输入。另一类是算法适应法，其中 ML-kNN 算法[8]是基于 k 近邻形成的，对传统的 kNN 算法进行改造，通过使用最大后验规则来决定一个实例与每个标签是否相关，Rank-SVM 则是将经典的支持向量机扩展到多标签学习中。

本文通过问题转化法，提出了一种基于条件随机场的多标签学习算法(ML-CRF)，该方法通过将多标签问题转化为序列标注问题来提高算法的准确性，同时还能考虑到标签依赖性问题。条件随机场[9]

(CRF)是一种判别式概率模型(见图 1)，它是依据条件概率进行建模，这种建模方式使得条件随机场能够充分利用标签间的依赖关系来提高标注的准确性。与此同时，CRF 在训练过程中会考虑整个输入序列和输出标签序列的联合概率分布，从而找到全局最优的标签序列。条件随机场(CRF)的核心思想是，给定一组输入变量，模型会学习这些输入变量之间的条件概率分布。这与一些只考虑局部最优解的模型(如 HMM)相比，具有更高的准确性。如果不选用条件随机场来解决序列标注问题，也可以选用隐马尔可夫模型(HMM) [10]或者是最大熵马尔可夫模型(MEMM) [11]，但条件随机场与后两者相比还是具有很大优势的。

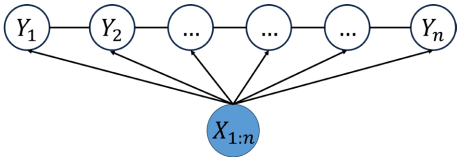

Figure 1. Schematic diagram of conditional random field model
图 1. 条件随机场模型示意图

对于隐马尔可夫模型(HMM) [10] (见图 2)而言，它需要满足一个很强的前提，也就是 HMM 的三大假设，其中比较重要的就是观测序列之间是独立的以及当前状态仅依赖于前一个的状态。这导致的问题是 HMM [10]只依赖于每一个状态和它对应的观察对象，但实际中序列标注问题不仅与单个词相关，而且和观察序列的长度，单词的上下文等等相关。HMM 学到的是状态和观察序列的联合分布，而预测问题中我们需要的是条件概率。这个算法本身就存在着一些不合理性。

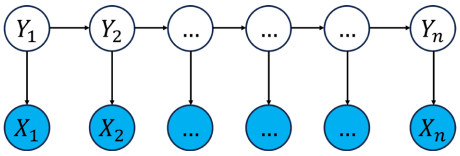

Figure 2. Diagram of hidden Markov model
图 2. 隐马尔可夫模型示意图

再观察最大熵马尔可夫模型(MEMM) [11] (见图 3)，它的思想是找到一个满足马尔可夫齐次性假设、观测不独立且熵最大的模型来解决序列标注问题。但是它存在标注偏差的问题，MEMM 对每个时刻都做局部归一化，每个节点的转移状态不同，每个节点的转移状态形成概率分布导致概率分布不均衡，转移状态越少的状态、转移概率就越大，因子最终概率最大路径中更可能出现转移状态较少的状态。条件随机场的去掉“有向性”就可以很好地避免这种标注偏差的状况发生。

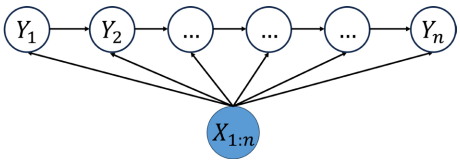

Figure 3. Schematic diagram of maximum entropy Markov model
图 3. 最大熵马尔可夫模型示意图

3. 模型说明

在多标签问题中，样本的标签实际上为一个固定长度的二进制序列(每个位置对应一个标签，1 表示

标签存在，0 表示标签不存在)，因此本文首先将多标签问题转化为一个序列标注问题。之后选用条件随机场(CRF)来解决这个问题，利用 CRF 模型中的状态概率和转移概率。状态概率表示每个标签的出现与输入特征之间的关系，转移概率则表示相邻标签之间的依赖性。在训练过程中，CRF 通过最大化训练数据的条件对数似然，学习标签与特征的关系以及标签之间的依赖关系，最终实现全局最优的多标签预测。

3.1. 多标签学习任务定义

对于输入实例 x，该实例可以被分配到一个标签集合 L 中。其中，$x\in\mathbb{R}^{\alpha\times1}$ 是维度为 α 的向量，标签集合 $L=\{l_1,l_2,\cdots,l_k\}$，k 为标签的总数。根据这些实例的标签可以获得指示向量 y，该向量是一个长度为 k 的二进制向量。若 x 属于第 i 个标签，那么 $y_i=1$ 否则 $y_i=0$。多标签学习问题即找到一个函数 $f:x\to\{0,1\}^k$，使 $f(x)=y$。

3.2. 多标签学习转换为序列标注问题

将 x 向量转换成 $\mathrm{X}=\{x_1,x_2,\cdots,x_k\}$ 序列。

3.2.1. 特征编码(神经网络)

实例 x 可以为一个输入向量 $x\in\mathbb{R}^{\alpha\times1}$，其中 α 为输入向量的维度。神经网络 $\mathbb{H}$ 为 β 层的网络。其中 $h^{(\beta)}=\sigma\left(W^{(\beta)}h^{(\beta-1)}+b^{(\beta)}\right)$。其中 $W^{(\beta)}$ 和 $b^{(\beta)}$ 分别为第 β 层的权重矩阵和偏置矩阵。σ 为激活函数，可以为 ReLU 或 Sigmoid。值得注意的是，对于第一层隐藏层，$h^{(0)}=x$。

对于最后一层 $\hat{y}=\sigma^*\left(W^oh^m+b^o\right)$，其中 $W^{(o)}$ 和 $b^{(o)}$ 为输出层的权重和偏置，h^m 是最后一个隐藏层的输出。σ^* 为 Sigmoid 函数，即 $\sigma^*(z)=\dfrac{1}{1+\mathrm{e}^{-z}}$。$\hat{y}\in[0,1]^k$ 为模型预测的概率向量。

为了训练上述网络，需要定义一个损失函数来衡量预测值与真实值之间的差异，这里使用二元交叉熵损失，如下式所示：

$$L(y,\hat{y})=-\sum_{i=1}^{k}\left[y_i\log(\hat{y}_i)+(1-y_i)\log(1-\hat{y}_i)\right] \tag{1}$$

L 为所有的标签的平均损失。

3.2.2. 输入向量变换

训练好的神经网络的最后一层参数构造 k 维序列 $x^*=W^oh^m+b^o$，其中 $x^*\in\mathbb{R}^{k\times1}$

$$Ax=x^* \tag{2}$$

其中 $x^*\in\mathbb{R}^{k\times1}$ 且 $x\in\mathbb{R}^{\alpha\times1}$，再利用 x 的伪逆矩阵求解矩阵 A

$$A=x^*x^+ \tag{3}$$

x^+ 是 x 的伪逆矩阵，容易得到 A 矩阵的维度为 $A\in\mathbb{R}^{k\times\alpha}$，把 A 看成一个特殊的特征提取器，求这个矩阵的特征向量 $\lambda_1,\lambda_2,\cdots$，所以构造下面的

$$X=x\cdot A_\lambda \tag{4}$$

其中 A_λ 是一个由 A 特征向量构建的一个长度为 k 的行向量。

4. 利用条件随机场求解

4.1. 条件随机场模型定义

条件随机场是一种用于序列标注的概率模型，我们的目标是学习给定输入序列 X 下标签序列 y 的条

件概率 $P(y|X)$。其形式为：

$$P(y|X)=\frac{1}{Z(X)}\exp\left(\sum_{i=1}^{k}\left(\sum_{s}\lambda_s f_s(y_i,X,i)+\sum_{t}\mu_t g_t(y_{i-1},y_i,X,i)\right)\right) \tag{5}$$

其中 $Z(X)$ 是归一化因子，用于确保所有可能标签序列的概率之和为 1：

$$Z(X)=\sum_{y}\exp\left(\sum_{i=1}^{k}\left(\sum_{s}\lambda_s f_s(y_i,X,i)+\sum_{t}\mu_t g_t(y_{i-1},y_i,X,i)\right)\right) \tag{6}$$

$f_s(y_i,X,i)$ 是状态特征函数，表示标签 y_i 与输入之间的关系。$g_t(y_{i-1},y_i,X,i)$ 是转移特征函数，表示相邻标签之间的依赖关系。λ_s 和 μ_t 是模型参数，需要通过训练数据学习。

状态特征函数 $f_s(y_i,X,i)$：该函数描述在位置 i 处标签 y_i 与输入 X 的关系，状态特征函数可以根据输入特征来定义，从而反映输入特征对标签的影响。

$$f_s(y_i,X,i)=y_i\cdot Indicator(\text{特征}_s) \tag{7}$$

其中 $Indicator(\text{特征}_s)$ 表示某个特征是否在位置 i 上出现。

转移特征函数 $g_t(y_{i-1},y_i,X,i)$：该函数描述标签之间的依赖关系，反映标签 y_{i-1} 和 y_i 的转移概率。在多标签问题中，转移特征函数允许模型学习标签之间的关联关系(如不同标签之间是否存在共现关系)。

$$g_t(y_{i-1},y_i,X,i)=Indicator(y_{i-1}=a,y_i=b) \tag{8}$$

4.2. 模型训练

概率建模中最常用的目标函数就是训练数据的对数似然函数。我们选用对数似然函数作为目标函数，目标是最大化似然值。通过最大化条件对数似然，学习最优的模型参数，使得模型参数可以最大化观测序列(标签序列)在给定输入序列下的概率，以便更好地描述输入特征和标签之间的关系。目标函数具体如下：

$$L(\theta)=\sum_{n=1}^{N}\log P\left(y^{(n)}|x^{(n)}\right) \tag{9}$$

其中 $\theta=\{\lambda_s,\mu_t\}$ 是模型参数。

通过参数估计，计算对数似然函数对参数的梯度

$$\frac{\partial L(\theta)}{\partial\theta}=\sum_{n=1}^{N}\left(\text{期望值}_{\text{数据}}-\text{期望值}_{\text{模型}}\right) \tag{10}$$

之后使用梯度上升法来求解参数。梯度上升法的核心思想是：通过沿着目标函数增长最快的方向(即梯度方向)来更新参数，使得对数似然函数逐步增大，直到达到最大值。梯度上升法本质上是一个增大对数似然函数值的逐步优化过程，用于找到 CRF 模型的最佳参数，使得模型在最大程度上符合训练数据。

算法 1 将多标签学习转换为序列标注问题并用 CRF 模型解决

```
# 输入：
#    X：输入特征矩阵    Y：多标签的标签矩阵
# 输出：
#    Pred_Y：预测的多标签的标签矩阵
# Main process
function main(X, Y):
    # 1. 数据转换
    将输入数据 X 和标签数据 Y 转换成适合于序列标注任务的数据格式
```

续表

```
即把每个样本的输入特征和对应的标签序列化成可以用于模型训练的序列数据
# 2. 训练 CRF 模型
调用一个 train_crf 函数来训练条件随机场模型，并返回训练好的模型

# 3. 使用 CRF 模型进行预测
利用训练好的 CRF 模型对输入特征序列进行预测，生成对应的标签序列

# 4. 转换预测序列为预测标签
将预测的标签序列转换回多标签的标签格式，以便与多标签学习问题的输出格式兼容

return Pred_Y
```

5. 实验设置

在本章节中，我们将介绍用于对比的算法以及评估 ML-CRF 模型性能的评价指标。

我们将由我们提出的利用条件随机场的多标签学习模型(ML-CRF)与简单的 Binary Relevance (BR)算法以及几种先进的多标签算法进行了比较。BR [12]算法的核心思想是将多标签学习问题分解为多个独立的二分类问题，每个标签被视为一个单独的二分类任务，该算法适用于标签之间关系较弱或可以忽略不计的多标签问题；Classifier Chains (CC) [13]算法通过将多标签问题转化为一个链式的分类问题来考虑标签之间的依赖关系，与 BR 不同的是 CC 算法会将每个标签的预测结果作为后续标签的输入特征，从而在预测时逐步构建出标签的依赖结构；Classifier Hierarchy Forest (CHF) [14]算法是一种集成学习方法，结合了层次结构与随机森林算法的思想，专门设计用于处理具有层次标签结构的多标签问题；Probabilistic Classifier Chains (PCC) [3]算法是在 CC 算法的基础上引入了概率模型，以更好地处理标签之间的依赖关系，它为每一个标签构建概率模型，并通过条件概率链来预测标签序列；Instance-Based Logistic Regression (IBLR) [15]算法结合了实例学习与逻辑回归，旨在利用实例学习的局部信息和逻辑回归的全局学习能力，从而增强多标签分类的性能，特别是在标签依赖性和非线性关系较强的数据集上效果良好；Multi-Label k-Nearest Neighbors (ML-kNN) [8]算法是一种扩展 kNN 算法，基于传统的 k 近邻算法，通过结合贝叶斯理论和 kNN 来处理多标签数据，特别适用于标签间关联较弱的多标签问题；Maximum Margin Output Coding (MMOC)算法的主要思想是通过最大化分类边界来处理标签的复杂依赖关系，基于输出编码的概念，将多标签问题转化为一系列单标签问题，通过最大化边界的方式使得分类器在处理多标签时更具鲁棒性。

对于上述方法，我们使用论文中建议的参数设置。对于 CC 算法，我们将类别顺序设置为 $Y_1 < Y_2 < \cdots < Y_d$；对于 MMOC 算法，λ 的参数设置为 1；对于 ML-kNN 和 IBLR 算法，使用欧氏距离来衡量实例的相似性，最近邻的数量设置为 10。

Micro-F1 和 Macro-F1 是评估多标签问题的两种直观测量方法。先介绍以下四个统计量：

$$TP_j(\text{“真”正例的个数}) = \left|\left\{x_i \mid y_i \in Y_i \wedge y_j \in h(x_i), (x_i, Y_i) \in S\right\}\right|$$

$$FP_j(\text{“伪”正例的个数}) = \left|\left\{x_i \mid y_i \notin Y_i \wedge y_j \in h(x_i), (x_i, Y_i) \in S\right\}\right|$$

$$TN_j(\text{“真”负例的个数}) = \left|\left\{x_i \mid y_i \notin Y_i \wedge y_j \notin h(x_i), (x_i, Y_i) \in S\right\}\right|$$

$$FN_j(\text{“伪”负例的个数}) = \left|\left\{x_i \mid y_i \in Y_i \wedge y_j \notin h(x_i), (x_i, Y_i) \in S\right\}\right|$$

以下性能指标可由以上四个统计量导出，例如：

$$\text{Accuracy} = \frac{TP_j + TN_j}{TP_j + FP_j + TN_j + FN_j} \tag{11}$$

$$\text{Precision} = \frac{TP_j}{TP_j + FP_j} \tag{12}$$

$$\text{Recall} = \frac{TP_j}{TP_j + FN_j} \tag{13}$$

Micro F1 评价指标将所有类别的预测和实际的标签合并在一起，之后计算整体的精确率、召回率和F1 分数。这个指标关注的是整体性能，而不是每个类别的性能。Micro F1 的计算公式如下：

$$\text{Micro F1} = \frac{1}{N}\sum_{i=1}^{C}\left(2 \times TP_j \times \text{Precision} \times \text{Recall}\right) \tag{14}$$

Macro F1 则是对每个类别分别计算精确率、召回率和 F1 分数，之后取所有类别的平均值。这个指标关注的是每个类别的性能，而不是整体性能。Macro F1 的计算公式如下：

$$\text{Macro F1} = \frac{1}{C}\sum_{i=1}^{C}\left(2 \times \frac{TP_j}{TP_j + FP_j + FN_j} \times \frac{TP_j}{TP_j + FN_j}\right) \tag{15}$$

Micro F1 更关注整体性能，Macro F1 更关注每个类别的性能。除以上两个指标外，我们也选用了其他指标来评价 ML-CRF 的性能。

此外，本文选用的对比算法中除了 MMOC 算法之外，其他方法都是元学习器，它们都可以使用几个基本分类器，通过对多个基学习器的性能和特征进行学习，来决定如何将这些基学习器集成为一个更强大的学习器。所以我们对这些方法使用 L2 惩罚逻辑回归，并通过交叉验证选择它们的正则化参数。通过在代价函数中添加权重的平方和来惩罚模型复杂度，这有助于减小权重的幅度，从而防止模型过于复杂。

6. 实验结果

6.1. 实验数据

为了评估 ML-CRF 的预测性能，我们在几组真实数据上进行了对比实验。这些数据集都来源于不同的领域，例如图像、生物或者文本等。需要特别注意的是，rcv1-top10 数据中的部分标签都是罕见的，所以我们选择了该数据集中最常见的几个标签进行研究。实验中所使用的数据集详细信息如表 1 所示，对于每个多标签数据集，我们分别展示出了样本数、特征数、标签数和数据类型。此外，我们还另外展示出了数据的两个统计数据，一个是显示出每个实例的平均标签数量的标签密度，另一个是衡量数据中出现的所有可能的标签组合的数量的标签分布。

Table 1. Multi-label data set information
表 1. 多标签数据集信息

名称	样本数	特征数	标签数	标签密度	标签分布	数据类型
Emotions	593	72	6	1.868	27	Music
Image	2000	135	5	1.24	20	Image
rcv1-top10	6000	8267	10	1.310	76	Text
Scene	2407	294	6	1.074	15	Image
Yeast	2417	103	14	4.237	198	biology

6.2. 实验结论

表 2 和表 3 中总结了 8 个算法在 Micro F1 和 Macro F1 两种评价指标下的实验结果。这两个指标在评估时数值越大，意味着算法的性能越出色。对所有方法应用标准的十倍交叉验证，并公布平均度量值。此外，为了在统计学上衡量差异的显著性，ML-CRF 与其他方法在 0.05 显著水平下进行配对 t 检验。

其中，在 Micro F1 评价指标下，本文算法 ML-CRF 在 5 个数据集中的 2 个数据集表现都是最好的，在另外 3 个数据集上的表现也不是最差的。在 Macro F1 评价指标下，本文算法 ML-CRF 在 5 个数据集中的 3 个数据集表现都是最好的，在另外 2 个数据集上的表现也不是最差的。足以说明条件随机场在解决多标签问题上有明显效果，给模型带来了显著提升。

Table 2. Micro F1 results for each method on the dataset (using a paired t-test at the 0.05 significance level)
表 2. 每种方法在数据集上的 Micro F1 实验结果(使用 0.05 显著性水平的配对 t 检验)

数据集	BR	CC	CHF	PCC	IBLR	ML-kNN	MMOC	ML-CRF
Emotions	0.645	0.621	0.672	0.664	0.692	0.656	0.687	0.697
Image	0.479	0.550	0.541	0.565	0.573	0.504	0.572	0.576
rcv1-top10	0.582	0.595	0.597	0.600	0.566	0.314	0.295	0.598
Scene	0.696	0.697	0.722	0.722	0.758	0.736	0.711	0.733
Yeast	0.635	0.628	0.637	0.645	0.661	0.646	0.651	0.629

Table 3. Multi-label data set information
表 3. 多标签数据集信息

数据集	BR	CC	CHF	PCC	IBLR	ML-kNN	MMOC	ML-CRF
Emotions	0.632	0.621	0.667	0.659	0.690	0.656	0.679	0.695
Image	0.486	0.562	0.546	0.575	0.581	0.516	0.578	0.585
rcv1-top10	0.500	0.537	0.526	0.534	0.487	0.257	0.385	0.525
Scene	0.703	0.709	0.730	0.729	0.765	0.743	0.721	0.776
Yeast	0.457	0.467	0.461	0.486	0.498	0.478	0.473	0.476

7. 结论及展望

在解决多标签学习问题时，将其转化为序列标注问题再采用条件随机场(CRF)是一种有效的策略，因为 CRF 擅长处理标签之间的复杂依赖关系，尤其是在多标签间具有关联性或序列依赖性时表现良好。CRF 不仅可以通过特征函数捕捉输入特征和标签之间的关系，还能通过转移特征建模标签间的依赖，从而提高多标签问题的准确性。

CRF 在标签数量较多或标签结构复杂的场景中，训练和推断的计算成本较高，尤其是当数据量大时。复杂的标签间依赖关系也可能导致模型训练过程中的优化难度增加。此外，CRF 模型对输入特征的表达能力高度依赖，如果特征选择不当，可能会限制模型的性能。

通过条件随机场解决多标签学习转化为序列标注问题在理论和实践上都具有重要意义。未来在模型优化、高效推断、迁移学习以及标签依赖结构建模上的进展将使得 CRF 在多标签问题中的应用更为广泛，处理能力和性能也将显著提升。

基金项目

国家自然科学基金(12301581)；北京市教育委员会科学研究计划项目(KM202210016002)。

参考文献

[1] Read, J., Pfahringer, B., Holmes, G. and Frank, E. (2011) Classifier Chains for Multi-Label Classification. *Machine Learning*, **85**, 333-359. https://doi.org/10.1007/s10994-011-5256-5

[2] Read, J., Pfahringer, B., Holmes, G. and Frank, E. (2009) Classifier Chains for Multi-Label Classification. In: Buntine, W., Grobelnik, M., Mladenić, D. and Shawe-Taylor, J., Eds., *Lecture Notes in Computer Science*, Springer Berlin Heidelberg, 254-269. https://doi.org/10.1007/978-3-642-04174-7_17

[3] Narassiguin, A., Elghazel, H. and Aussem, A. (2017) Dynamic Ensemble Selection with Probabilistic Classifier Chains. In: Ceci, M., Hollmén, J., Todorovski, L., Vens, C. and Džeroski, S., Eds., *Lecture Notes in Computer Science*, Springer International Publishing, 169-186. https://doi.org/10.1007/978-3-319-71249-9_11

[4] Madjarov, G., Gjorgjevikj, D., Dimitrovski, I. and Džeroski, S. (2016) The Use of Data-Derived Label Hierarchies in Multi-Label Classification. *Journal of Intelligent Information Systems*, **47**, 57-90. https://doi.org/10.1007/s10844-016-0405-8

[5] Spolaôr, N., Monard, M.C., Tsoumakas, G. and Lee, H.D. (2016) A Systematic Review of Multi-Label Feature Selection and a New Method Based on Label Construction. *Neurocomputing*, **180**, 3-15. https://doi.org/10.1016/j.neucom.2015.07.118

[6] Huang, S., Gao, W. and Zhou, Z. (2019) Fast Multi-Instance Multi-Label Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **41**, 2614-2627. https://doi.org/10.1109/tpami.2018.2861732

[7] Boutell, M.R., Luo, J., Shen, X. and Brown, C.M. (2004) Learning Multi-Label Scene Classification. *Pattern Recognition*, **37**, 1757-1771. https://doi.org/10.1016/j.patcog.2004.03.009

[8] Zhang, M. and Zhou, Z. (2007) ML-KNN: A Lazy Learning Approach to Multi-Label Learning. *Pattern Recognition*, **40**, 2038-2048. https://doi.org/10.1016/j.patcog.2006.12.019

[9] 洪铭材, 张阔, 唐杰, 等. 基于条件随机场(CRFs)的中文词性标注方法[J]. 计算机科学, 2006, 33(10): 148-151+155.

[10] 崔丽平, 古丽拉·阿东别克, 王智悦. 基于有向图模型的旅游领域命名实体识别[J]. 计算机工程, 2022, 48(2): 306-313.

[11] 喻鑫, 张矩, 邱武松, 等. 基于序列标注算法比较的医学文献风险事件抽取研究[J]. 计算机应用与软件, 2017, 34(12): 58-63.

[12] 方胜群. 基于机器学习的院感智能诊断技术研究[D]: [硕士学位论文]. 长沙: 国防科技大学, 2018.

[13] 张博晟. 基于深度学习和主题模型的多标签文本分类方法研究[D]: [硕士学位论文]. 杭州: 杭州电子科技大学, 2023.

[14] 刘凯, 龚辉, 曹晶晶, 等. 基于多类型无人机数据的红树林遥感分类对比[J]. 热带地理, 2019, 39(4): 492-501.

[15] 付彬. 基于标记依赖关系的多标记学习算法研究[D]: [博士学位论文]. 北京: 北京交通大学, 2016.