

## 基于标签关联的多标签演化超网络

王 进, 刘 彬, 孙开伟, 陈乔松, 邓 欣

(重庆邮电大学计算智能重庆市重点实验室, 重庆 400065)

**摘 要:** 针对多标签学习中如何有效挖掘利用高阶标签关联的问题, 提出了一种基于标签关联的多标签演化超网络模型. 该模型通过输入任意多标签学习方法的预测结果, 利用超边表征挖掘高阶标签关联, 并综合标签关联和特征信息作为最终的预测结果. 与3种传统多标签学习方法在6个多标签数据集上的对比实验表明, 本文提出模型不仅能够有效提升多个传统多标签学习方法的性能, 而且能够提供具有良好可读性的学习结果.

**关键词:** 机器学习; 多标签学习; 演化超网络; 标签关联

**中图分类号:** TP39 **文献标识码:** A **文章编号:** 0372-2112 (2018)04-1012-07

**电子学报 URL:** <http://www.ejournal.org.cn> **DOI:** 10.3969/j.issn.0372-2112.2018.04.032

## Multi-Label Evolutionary Hypernetwork Based on Label Correlations

WANG Jin, LIU Bin, SUN Kai-wei, CHEN Qiao-song, DENG Xin

(Chongqing Key Laboratory of Computational Intelligence, Chongqing University of Posts and Telecommunications, Chongqing 400065, China)

**Abstract:** In order to solve the problem that how to explore and exploit the high-order label correlations effectively in multi-label learning, a Multi-Label evolutionary HyperNetwork based on label Correlations (MLHNC) is proposed in this paper. In MLHNC, the predicting results obtained from any multi-label learning method are utilized as input of the model, the high-order correlations among labels are represented and explored by hyperedges, and the final prediction is made by integrating the label correlation and feature information. The experimental results on six multi-label datasets compared with three state-of-the-art multi-label learning methods show that the MLHNC not only improves the performance of various state-of-the-art multi-label learning methods, but also provides readable learning results.

**Key words:** machine learning; multi-label learning; evolutionary hypernetwork; label correlation

### 1 引言

传统有监督学习认为每个样本只有一个标签<sup>[1]</sup>, 但在现实应用中, 事物往往具有复杂的语义信息. 例如一副图片可能同时包括海洋、沙滩、天空等信息. 在多标签学习 (Multi-Label Learning) 中, 每个样本可能包含一个或多个标签, 含有更多标签信息的样本能够更好的表现事物语义信息的多样性<sup>[2,3]</sup>. 多标签学习的任务包括多标签分类和标签排序两方面: 前者旨在确定与测试样本关联的标签集合, 而后者旨在根据所有标签与测试样本的关联度确定标签序列<sup>[4]</sup>. 近年来, 多标签学习已逐渐成为机器学习及其相关领域的热点研究方向, 广泛应用在图像与视频标注<sup>[5-7]</sup>、文本分类<sup>[8-10]</sup>、生物信息学<sup>[11,12]</sup>、音乐情感分类<sup>[13]</sup>等诸多领域.

多标签数据包含大量的标签信息, 而标签之间常存在一定的关联性. 例如, “海洋”和“沙滩”往往同时出现在一张图片中. 因此, 如何有效利用标签关联, 尤其是高阶标签关联信息, 成为提升多标签学习方法性能的关键所在<sup>[3]</sup>. 此外, 部分多标签学习模型虽然关注和利用了标签关联, 但其学习结果缺乏可读性, 因此无法输出重要的标签关联信息.

超网络 (HyperNetwork, HN) 是受生物分子网络启发而建立的一种基于超图 (Hypergraph) 的认知学习模型<sup>[14]</sup>, 模型中的超边能够表征并挖掘顶点间的高阶关联关系, 并提供具有良好可读性的学习结果. 为了有效利用标签高阶关联信息, 解决现有方法学习结果可读性差的问题, 本文提出了一种基于标签关联的多标签演化超网络模型 (Multi-Label evolutionary HyperNetwork

收稿日期: 2016-05-12; 修回日期: 2017-07-14; 责任编辑: 孙瑶

基金项目: 重庆市基础与前沿研究计划 (No. cstc2014jcyjA40001, No. cstc2014jcyjA40022); 重庆教委科学技术研究项目 (自然科学类) (No. KJ1400436)

based on label Correlations, MLHNC). 在 MLHNC 中,任意传统多标签学习方法(本文称为基学习方法)的预测标签信息将转化为模型输入,顶点表示标签,超边表征高阶标签关联. MLHNC 模型通过演化学习挖掘和利用高阶标签关联,并综合标签关联和特征信息作为预测结果,提升基学习方法性能. 相对于传统多标签学习方法,MLHNC 的优势主要体现在:①在基学习方法的预测结果上,通过演化学习充分挖掘和利用了高阶标签关联信息;②通过超边表征重要的高阶标签关联,提供具有良好可读性的学习结果;③有效提升基学习方法分类性能,适用范围广泛.

## 2 多标签学习

### 2.1 多标签学习问题描述

设  $\mathbf{X} = \mathbf{R}^d$  是  $d$  维特征空间,  $\Lambda = \{\lambda_1, \lambda_2, \dots, \lambda_q\}$  为整个标签集合. 训练样本  $D^{(n)} = \{\mathbf{x}^{(n)}, Y^{(n)}\}$  由  $d$  维特征向量  $\mathbf{x}^{(n)}$  和标签子集  $Y^{(n)} \in 2^\Lambda$  组成, 其中  $Y^{(n)}$  可表示为一个  $q$  维向量  $\mathbf{y}^{(n)} = [y_1^{(n)}, y_2^{(n)}, \dots, y_q^{(n)}]$ , 其中若  $\lambda_j \in Y^{(n)}$  则  $y_j^{(n)} = 1$ , 否则  $y_j^{(n)} = 0$ . 对于测试样本  $D^{(t)}$ , 多标签学习方法需要输出与其相关的标签集合  $Y^{(t)*}$  (表示为  $\mathbf{y}^{(t)*} = [y_1^{(t)*}, y_2^{(t)*}, \dots, y_q^{(t)*}]$ ) 和每个标签与样本关联度  $\mathbf{f}\mathbf{y}^{(t)*} = [f_{y_1^{(t)*}}, f_{y_2^{(t)*}}, \dots, f_{y_q^{(t)*}}]$ .

### 2.2 多标签学习方法概述

根据利用标签关联信息的策略不用, 可将多标签分类方法划分为三类: 一阶策略 (First-Order Strategy), 二阶策略 (Second-Order Strategy) 和高阶策略 (High-Order Strategy)<sup>[3]</sup>. 一阶策略方法认为标签间是相互独立的, 没有利用任何标签间的关联关系. 典型的一阶策略方法有: 把多标签分类问题转化为  $q$  个独立二分类问题的二元关系法 (Binary Relevance, BR)<sup>[5]</sup> 和利用最大后验概率准则进行预测的多标签  $k$  近邻方法 (Multi-Label  $k$ -Nearest Neighbor, ML $k$ NN)<sup>[15]</sup>. 二阶策略方法利用了标签集合中成对标签间的关联信息. 把多标签学习问题转化为  $C_q^2$  个标签排序问题的基准标签排序 (Calibrated Label Ranking, CLR)<sup>[16]</sup> 和改进神经网络模型的反向传播多标签学习法 (Backpropagation for Multi-Label Learning, BPMLL)<sup>[11]</sup> 是典型的二阶策略方法. 高阶策略方法则利用了多个标签间的关联信息. 标签幂集法 (Label Powerset, LP)<sup>[5]</sup> 和随机  $k$  标签集 (Random  $k$ -labelsets, RakEL)<sup>[17]</sup> 把多标签问题转化为一个或多个多分类问题. 分类器链 (Classifier Chain, CC)<sup>[18]</sup> 和相关二元关系模型 (Dependent Binary Relevance Model, DBR)<sup>[19]</sup> 则是通过将标签信息转化为特征来挖掘高阶标签关联. 高阶策略考虑了更多的标签关联信息, 因此与前两种策略相比, 其时间开销往往更大.

多标签超网络 (Multi-Label HyperNetwork,

MLHN)<sup>[20]</sup> 将超网络模型拓展到多标签学习领域, 但与本文方法有显著不同: MLHN 中超边的顶点表示特征, 而本文方法则将顶点视为标签; 与 MLHN 相比, 本文方法以基学习方法的预测结果为模型输入, 深度挖掘和利用标签关联信息, 可提升任意基学习方法性能, 具有更广泛的适用性.

## 3 演化超网络

超网络是一种由大量超边组成的概率图模型, 通过超边库表达模式空间中数据的分布概率<sup>[14]</sup>. 超网络模型结构如图 1(a) 所示. 超边是超网络的基本组成元素, 其由顶点子集, 标签和权重三部分组成. 每条超边可以连接多个顶点, 而包含在同一条超边内的顶点往往具有一定的关联, 因此超边可以表征顶点间的高阶关联性. 超边所连接的顶点数称为超边的阶数 (Order), 所有超边阶数都为  $k$  的超网络称为  $k$  阶超网络<sup>[21]</sup>. 在分类模式下, 顶点表示特征, 超边可以看作决策规则, 超边的权重则表示其决策能力的强弱. 超网络通过演化学习调整超边库, 提高模型与输入数据在模式空间概率分布的拟合度. 超网络通过超边库得到输入样本  $\mathbf{x}$  与输出类别  $y$  的联合概率  $P(\mathbf{x}, y)$  以及  $\mathbf{x}$  的分布概率  $P(\mathbf{x})$ , 并通过后验概率最大化原则得到最终结果:

$$\mathbf{y}^* = \arg \max_y (P(y|\mathbf{x})) = \arg \max_y \left( \frac{P(\mathbf{x}, y; H)}{P(\mathbf{x}; H)} \right) \quad (1)$$

## 4 MLHNC

MLHNC 通过改进传统超网络结构, 将任意多标签学习方法的标签预测结果作为模型输入, 使超边能够表征标签间的高阶关联性, 进而有效挖掘和利用高阶标签关联信息, 提升多标签学习方法性能.

### 4.1 MLHNC 基本结构

在 MLHNC 中, 每个顶点表示一个标签, 超边中顶点的取值为对应标签的预测结果. 此外, 超边的标签和权重扩展为  $q$  维向量  $\mathbf{y} = [y_1, y_2, \dots, y_q]$  和  $\mathbf{w} = [w_1, w_2, \dots, w_q]$ , 其中  $w_j$  表示超边对标签  $\lambda_j$  的决策能力, 即超边中顶点表示的标签集合与标签  $\lambda_j$  的关联性的强弱. MLHNC 的结构如图 1(b) 所示.

对于训练样本  $D^{(n)}$ , 首先通过基学习方法获得其预测标签结果  $\mathbf{y}^{(n)*}$  和预测关联度  $\mathbf{f}\mathbf{y}^{(n)*}$ , 并将  $\mathbf{y}^{(n)*}$  转化为特征, 得到 MLHNC 的输入  $\tilde{D}^{(n)} = \{\mathbf{y}^{(n)*}, \mathbf{y}^{(n)}\}$ . 通过对模型结构和模型输入形式的改变, MLHNC 能够描述标签的预测值  $\mathbf{y}^*$  与标签真实值  $y_j$  的联合概率分布以及  $\mathbf{y}^*$  的概率分布:

$$P(\mathbf{y}^*, y_j; H) = \frac{1}{Z_j(H)} \exp(-\varepsilon(\mathbf{y}^*, y_j; H)) \quad (2)$$

$$P_j(\mathbf{y}^*; H) = \frac{1}{Z_j(H)} \exp(-\varepsilon(\mathbf{y}^*; H)) \quad (3)$$

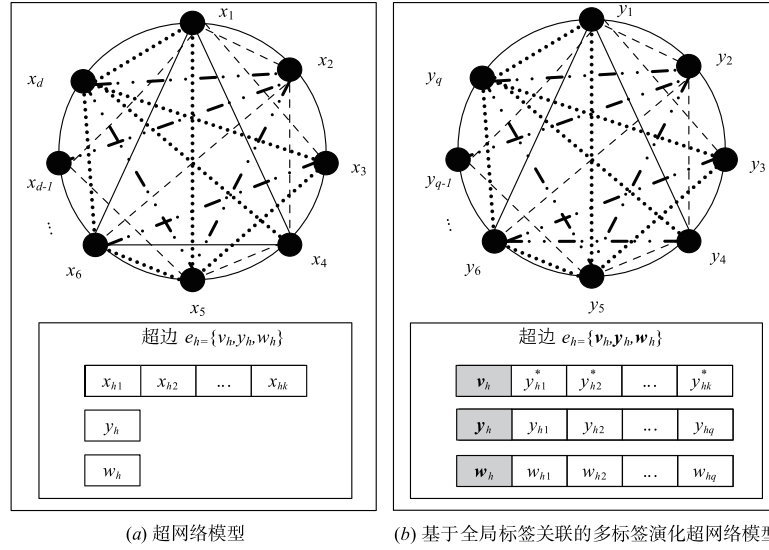


图1 传统超网络模型和基于标签关联的多标签演化超网络模型

$$\varepsilon(\mathbf{y}^*, y_j; H) = -\ln \sum_{h=1}^{|E|} w_{hj} I(\mathbf{y}^*, y_j; e_h) \quad (4)$$

$$\varepsilon_j(\mathbf{y}^*; H) = -\ln \sum_{h=1}^{|E|} w_{hj} I(\mathbf{y}^*; e_h) \quad (5)$$

$$Z_j(H) = \sum_{\mathbf{y}^*, y_j \in \bar{D}} \exp(-\varepsilon_j(\mathbf{y}^*; H)) \quad (6)$$

其中,  $|E|$  为超边总数,  $w_{hj}$  为超边  $e_h$  的第  $j$  个标签对应的权重;  $I$  为匹配函数, 当超边与样本匹配时取值为 1, 否则为 0;  $H$  为 MLHNC 模型,  $\varepsilon$  为能量函数,  $Z(H)$  是规范化因子。

对于测试样本  $D^{(t)} = \{\mathbf{x}^{(t)}\}$ , 通过基学习方法获得其预测标签结果  $\mathbf{y}^{(t)*}$  和预测关联度  $\mathbf{f}\mathbf{y}^{(t)*}$ , 并构建新的测试样本  $\bar{D}^{(t)} = \{\mathbf{y}^{(t)*}\}$ . MLHNC 通过  $\mathbf{y}^*$  与  $\mathbf{y}$  的联合概率分布以及  $\mathbf{y}^*$  的概率分布得到每个标签与  $D^{(t)}$  关联度:

$$f y_j^{(t)**} = P(y_j^{(t)} = 1 | \mathbf{y}^{(t)*}; H) = \frac{P(y_j^{(t)} = 1, \mathbf{y}^{(t)*}; H)}{P(\mathbf{y}^{(t)*}; H)}$$

$$= \frac{\sum_{h=1}^{|E|} w_{hj} I(\mathbf{y}^{(t)*}, y_j^{(t)}; e_h)}{\sum_{h=1}^{|E|} w_{hj} I(\mathbf{y}^{(t)*}; e_h)}, j=1, 2, \dots, q \quad (7)$$

其中,  $w_{hj}$  为超边  $e_h$  中对第  $j$  个标签的权重,  $f y_j^{(t)**}$  即为在已知  $\mathbf{y}^{(t)*}$  条件下标签  $\lambda_j$  与  $D^{(t)}$  相关联的概率. 通过式, MLHNC 能够挖掘和利用标签关联. 由于转化后的样本仅包含标签信息, 而  $\mathbf{f}\mathbf{y}^{(t)*}$  为已知特征信息  $\mathbf{x}^{(t)}$  下标签与测试样本  $D^{(t)}$  的关联度, 因此 MLHNC 最终预测关联度结果  $\mathbf{f}\mathbf{y}^{(t)\Psi}$  为特征信息  $\mathbf{f}\mathbf{y}^{(t)*}$  和标签关联  $\mathbf{f}\mathbf{y}^{(t)**}$  加权和:

$$\mathbf{f}\mathbf{y}^{(t)\Psi} = \alpha \mathbf{f}\mathbf{y}^{(t)*} + (1 - \alpha) \mathbf{f}\mathbf{y}^{(t)**} \quad (8)$$

其中,  $\alpha \in (0, 1)$  为特征信息的权重.  $\mathbf{f}\mathbf{y}_j^{(t)\Psi}$  可以看作  $y_j^{(t)} = 1$  的置信度, 而  $1 - f y_j^{(t)\Psi}$  可以看作  $y_j^{(t)} = 0$  的置信度.

模型最终的标签预测结果  $\mathbf{y}^{(t)\Psi}$  为:

$$y_j^{(t)\Psi} = \arg \max_{b \in \{0, 1\}} |b + f y_j^{(t)\Psi} - 1|, j=1, 2, \dots, q \quad (9)$$

MLHNC 的预测算法如算法 1 所示.

#### 算法 1 MLHNC 预测算法

输入: MLHNC;  $H$ , 基学习方法:  $TM$ , 测试样本:  $D^{(t)} = \{\mathbf{x}^{(t)}\}$ , 标签数:

$q$ ,  $TM$  权重:

输出:  $\mathbf{f}\mathbf{y}^{(t)\Psi}, \mathbf{y}^{(t)\Psi}$

1.  $(\mathbf{y}^{(t)*}, \mathbf{f}\mathbf{y}^{(t)*}) \leftarrow TM(D^{(t)});$

2. 构建新测试样本  $\bar{D}^{(t)} = \{\mathbf{y}^{(t)*}, \mathbf{f}\mathbf{y}^{(t)*}\};$

3. 超边匹配:

初始化  $q$  维向量  $\mathbf{W}_1, \mathbf{W}_0: \mathbf{W}_1 \leftarrow \mathbf{0}, \mathbf{W}_0 \leftarrow \mathbf{0};$

for  $h \leftarrow 1$  to  $|E|$  do

if  $I(\mathbf{y}^{(t)*}; e_h) = 0$  then

for  $j \leftarrow 1$  to  $q$  do

if  $y_{hj} = 1$  then

$\mathbf{W}_{1j} \leftarrow \mathbf{W}_{1j} + w_{hj};$

else

$\mathbf{W}_{0j} \leftarrow \mathbf{W}_{0j} + w_{hj};$

end if

end for

end if

end for

4. 计算输出结果:

$\mathbf{f}\mathbf{y}^{(t)**} = \mathbf{W}_1 / (\mathbf{W}_1 + \mathbf{W}_0);$

$\mathbf{f}\mathbf{y}^{(t)\Psi} = \alpha \mathbf{f}\mathbf{y}^{(t)*} + (1 - \alpha) \mathbf{f}\mathbf{y}^{(t)**};$

for  $j \leftarrow 1$  to  $q$  do

根据式计算  $y_j^{(t)\Psi};$

end for

return  $\mathbf{f}\mathbf{y}^{(t)\Psi}, \mathbf{y}^{(t)\Psi};$

## 4.2 MLHNC 演化学习

演化学习旨在通过调整超边组成和超边权重, 增

加超边库决策能力,降低模型与数据的拟合误差. MLHNC 演化学习方法主要包括三部分:超边初始化、超边替代、随机梯度下降. MLHNC 的演化学习算法如算法 2 步骤 3~5 所示.

#### 算法 2 MLHNC 训练算法

Input: 训练集:  $T$ , 样本数:  $N$ , 标签数:  $q$ , 基学习方法:  $TM$ , 每个样本生成的超边数:  $c$ , 超边替代迭代次数:  $tr$ , 初始超边替代率:  $r$ ,  $TM$  权重:  $\alpha$

Output: MLHNC:  $H$

1.  $(\{\mathbf{y}^{(n)*}\}_{n=1}^N, \{\mathbf{f}\mathbf{y}^{(n)*}\}_{n=1}^N) \leftarrow TM(T)$ ;
2. 构建新的训练集  $\tilde{T} = (\tilde{D}^{(n)})_{n=1}^N, \tilde{D}^{(n)} = (\mathbf{y}^{(n)*}, \mathbf{y}^{(n)})$ ;
3. 超边初始化:
  - 初始化  $H$  的超边库  $E \leftarrow \emptyset$ ;
  - for  $n \leftarrow 1$  to  $N$  do
    - $\tilde{D}^{(n)}$  产生  $c$  条超边并加入  $E$  中;
  - end for
4. 超边替代演化学习:
  - for  $j \leftarrow 1$  to  $q$  do
    - 根据式 (11) 计算标签不平衡度  $UB_j$ ;
    - end for
    - for  $h \leftarrow 1$  to  $|E|$  do
      - 根据式 (10) 计算  $e_h$  的适应值;
      - end for
      - for  $t \leftarrow 1$  to  $tr$  do
        - 根据适应值对  $E$  中的超边进行降序排序;
        - $r \leftarrow r * (tr - t + 1) / tr$ ;
        - for  $h \leftarrow |E|$  to  $(1 - r) * |E|$  do
          - 利用生成  $e_h$  的样本生成超边  $e_{new}$  来替代  $e_h$ ;
          - 根据式 (10) 计算  $e_{new}$  的适应值;
        - end for
      - end for
    - 5. 随机梯度下降演化学习:
      - do
        - for  $n \leftarrow 1$  to  $N$  do
          - 根据算法 1 得到  $\tilde{D}^{(n)}$  的预测结果  $\mathbf{f}\mathbf{y}^{(n)\Psi}, \mathbf{y}^{(n)\Psi}$ ;
          - 将于  $\tilde{D}^{(n)}$  匹配的超边加入超边集合  $M$ ;
          - for  $j \leftarrow 1$  to  $q$  do
            - if  $\mathbf{y}_j^{(n)\Psi} \neq \mathbf{y}_j^{(n)}$  then
              - for  $h \leftarrow 1$  to  $|M|$  do
                - 更新超超边权重  $w_{hj} \leftarrow w_{hj} + \Delta w_{hj}$ ;
              - end for
            - end if
          - end for
        - end for
        - until 满足迭代停止条件;
        - return  $H$ ;

#### 4.2.1 超边初始化

超边初始化即生成初始超边库的过程. 在超边初始化阶段, 每个新构建的训练集  $\tilde{D}^{(n)}$  生成  $c$  条超边, 每条超边的顶点集合通过随机选取预测标签结果子集得

到, 标签向量与生成该超边的样本的标签向量相同, 权重向量中每个权重的初始值设为 1.

#### 4.2.2 超边替代

超边初始化过程存在一定的随机性, 因此需要对初始超边库进行调整, 以确保模型性能. 超边替代演化学习方法以调整超边库组成为出发点, 旨在通过新生成的超边替代预测分类性能较差的超边, 得到具有较好决策能力的超边集合.

在超边替代中, 适应值是衡量超边决策能力及表征标签关联能力的指标. MLHNC 中将  $q$  标签的正确匹配系数的均值作为超边的适应值:

$$fitness(e_h) = \frac{1}{q} \sum_{j=1}^q cm_{hj} \quad (10)$$

其中,  $cm_{hj}$  为超边  $e_h$  对标签  $\lambda_j$  的正确匹配系数. 在多标签数据中, 每个标签相关联的样本的数量通常远远小于与其不相关的样本的数量, 因此标签  $\lambda_j$  的不平衡度  $UB_j$  定义为:

$$UB_j = \frac{|\{\tilde{D}^{(n)} \mid \mathbf{y}_j^{(n)} = 1, n = 1, \dots, N\}|}{|\{\tilde{D}^{(n)} \mid \mathbf{y}_j^{(n)} = 0, n = 1, \dots, N\}|}, j = 1, \dots, q \quad (11)$$

则考虑了不平衡度的  $cm_{hj}$  计算公式为:

$$cm_{hj} = \begin{cases} \frac{cn_j(e_h) * UB_j}{wn_j(e_h) + cn_j(e_h) * UB_j}, & y_{hj} = 1 \\ \frac{cn_j(e_h)}{wn_j(e_h) * UB_j + cn_j(e_h)}, & \text{其他} \end{cases} \quad (12)$$

$$cn_j(e_h) = \sum_{n=1}^N (I(\mathbf{y}^{(n)*}; e_h) \llbracket \mathbf{y}_j^{(n)} = y_{hj} \rrbracket) \quad (13)$$

$$wn_j(e_h) = \sum_{n=1}^N (I(\mathbf{y}^{(n)*}; e_h) \llbracket \mathbf{y}_j^{(n)} \neq y_{hj} \rrbracket) \quad (14)$$

其中,  $y_{hj}$  为超边  $e_h$  的标签向量中第  $j$  个元素的值. 在每轮迭代中删除适应值最低的  $r * |E|$  条超边 ( $r$  为超边替代率,  $r \in (0, 1)$ ), 并生成等量新超边. 随着迭代次数的增加, 超边库中性能较好超边的数量增加,  $r$  逐渐降低. 由于超边搜索的空间范围取决于标签数量  $q$ , 因此本文中超边替代迭代次数设为  $\lceil \log_2 q \rceil$ .

#### 4.2.3 随机梯度下降

随机梯度下降演化学习方法旨在通过随机梯度下降法调整超边权重以减少模型在训练集上的误差. MLHNC 在训练样本  $\tilde{D}^{(n)}$  的标签  $\lambda_j$  上的误差定义为:

$$J_{nj}(\mathbf{W}) = \frac{1}{2} [P^*(\mathbf{y}_j^{(n)\Psi} \mid \tilde{D}^{(n)}) - P(\mathbf{y}_j^{(n)\Psi} \mid \tilde{D}^{(n)})]^2 \cdot \llbracket \mathbf{y}_j^{(n)} \neq \mathbf{y}_j^{(n)\Psi} \rrbracket \quad (15)$$

其中,  $P^*(\mathbf{y}_j^{(n)\Psi} \mid \tilde{D}^{(n)})$  为  $\mathbf{y}_j^{(n)} = \mathbf{y}_j^{(n)\Psi}$  的目标概率,  $P(\mathbf{y}_j^{(n)\Psi} \mid \tilde{D}^{(n)}) = f_{\mathbf{y}_j^{(n)\Psi}}$ , 仅当  $\mathbf{y}_j^{(n)} \neq \mathbf{y}_j^{(n)\Psi}$  时, 模型会产生误差. MLHNC 在训练集  $\tilde{T}$  上的整体误差为:

$$J(\mathbf{W}) = \sum_{\tilde{D}^{(n)} \in \tilde{T}} \sum_{j=1}^q J_{nj}(\mathbf{W}) \quad (16)$$

在随机梯度下降方法中,针对样本  $\tilde{D}^{(n)}$  的误分标签  $\lambda_j$ , 超边  $e_h$  的权重  $w_{hj}$  的变化量  $\Delta w_{hj}$ :

$$\Delta w_{hj} = -\eta \frac{\partial J_{hj}(\mathbf{W})}{\partial w_{hj}} \quad (17)$$

$$\frac{\partial J_{hj}(\mathbf{W})}{\partial w_{hj}} = [P^*(y_j^{(n)\Psi} | \tilde{D}^{(n)}) - fy_j^{(n)\Psi}] \frac{\partial fy_j^{(n)\Psi}}{\partial w_{hj}} \quad (18)$$

$$\frac{\partial fy_j^{(n)\Psi}}{\partial w_{hj}} = (1 - \alpha) \frac{\partial fy_j^{(n)**}}{\partial w_{hj}} \quad (19)$$

$$\frac{\partial fy_j^{(n)**}}{\partial w_{hj}} = \frac{I(\mathbf{y}^{(n)*}; e_h)}{\sum_{m=1}^{|E|} w_{mj} I(\mathbf{y}^{(n)*}; e_m)} \cdot (I(\mathbf{y}^{(n)*}, y_j^{(n)*}; e_h) - fy_j^{(n)**}) \quad (20)$$

其中,  $\eta$  为学习速率. 当训练集的预测结果  $\mathbf{y}^\Psi$  不再改变时, 梯度下降算法终止.

### 4.3 MLHNC 整体框架

如图 2 所示, MLHNC 的整体框架包括实线表示的训练过程和虚线表示的测试过程.

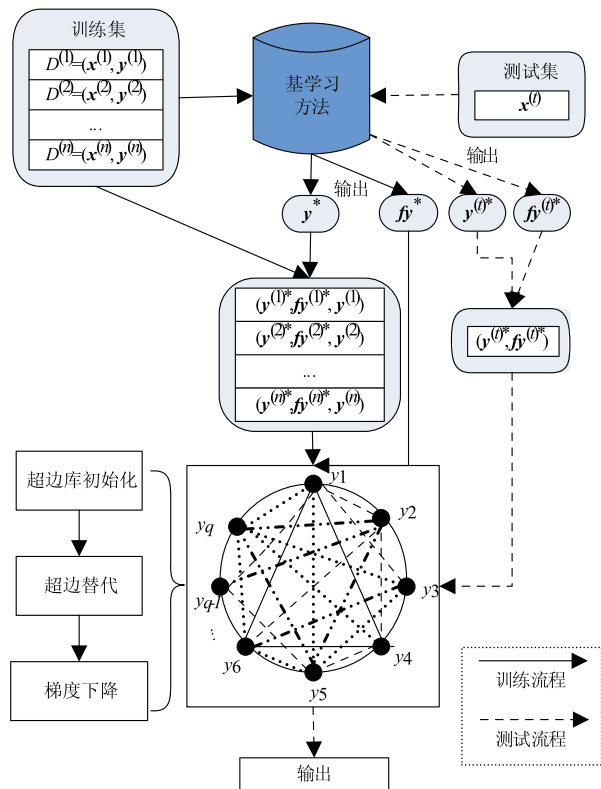


图2 MLHNC训练和测试过程示意图

训练阶段对应算法 2. 首先将训练集  $T$  输入到基学习方法, 获得预测结果  $\mathbf{y}^*$  和  $\mathbf{f}\mathbf{y}^*$  (步骤 1); 其次将  $\mathbf{y}^*$  转化为特征, 得到新训练集  $\tilde{T}$ , 并连同  $\mathbf{f}\mathbf{y}^*$  输入到 MLHNC 模型中 (步骤 2); 最后通过演化学习得到训练好的 MLHNC 模型 (步骤 3 ~ 5).

测试阶段对应算法 1. 首先将测试样本  $D^{(l)}$  输入到基学习方法, 获得预测结果  $\mathbf{y}^{(l)*}$  和  $\mathbf{f}\mathbf{y}^{(l)*}$ , 并构建新测试样

本  $\tilde{D}^{(l)}$  (步骤 1 ~ 2); 然后将  $\tilde{D}^{(l)}$  连同  $\mathbf{f}\mathbf{y}^{(l)*}$  输入到训练好的 MLHNC 中, 通过步骤 3 ~ 4 得到最终预测结果.

## 5 实验

### 5.1 实验设置与数据

为验证 MLHNC 能够有效挖掘利用高阶标签关联信息, 提升传统多标签学习方法性能, 本文选择一阶策略方法 BR<sup>[5]</sup>、二阶策略方法 BPMLL<sup>[11]</sup> 和高阶策略方法 RAKEL<sup>[17]</sup> 作为 MLHNC 的基学习方法. 为简便起见, 在本文中 BR-C、BPMLL-C 和 RAKEL-C 表示 MLHNC 利用相应方法作为基学习方法.

在实验中, BR、BPMLL 和 RAKEL 三种传统多标签学习方法采用 Mulan 多标签学习开源项目<sup>[22]</sup> 提供的算法实现, 其参数为 Mulan 中的推荐参数. MLHNC 中, 每个样本生成的超边数为 5, 初始的超边替代率为 0.4, 阶数由对训练集的 5-折交叉验证确定. 本文中所有的实验结果均为对数据集进行 10-折交叉验证得到的平均结果.

本文使用 6 个来源于 Mulan 的多标签数据集, 其详细信息如表 1 所示.

表 1 数据集信息表

| 数据集      | 样本数  | 属性数  | 标签数 |
|----------|------|------|-----|
| emotions | 593  | 72   | 6   |
| scene    | 2407 | 294  | 6   |
| yeast    | 2417 | 103  | 14  |
| medical  | 978  | 1449 | 45  |
| enron    | 1702 | 1001 | 53  |
| CAL500   | 502  | 68   | 174 |

### 5.2 评价指标

多标签学习方法的评价指标分为两类: 一类用于多标签分类, 一类用于标签排序. 本文选用汉明损失 (Hamming Loss), 微观 F1 值 (Micro F1-measure) 作为多标签分类性能的评价指标, 覆盖率 (Coverage) 和排序损失 (Ranking Loss) 作为标签排序性能的评价指标<sup>[2]</sup>.

### 5.3 实验结果与分析

BR、BPMLL 和 RAKEL3 种传统多标签学习方法以及 MLHNC 采用 3 种方法作为基学习方法在 4 个评价指标上的性能如表 2 ~ 表 5 所示, 其中“·”表示通过 t-检验验证 ( $p < 0.05$ ), MLHNC 显著优于其对应的基学习方法.

表 2 不同方法在汉明损失指标下的实验结果

|          | BR     | BR-C · | BPMLL  | BPMLL-C | RAkEL  | RAkEL-C · |
|----------|--------|--------|--------|---------|--------|-----------|
| emotions | 0.2524 | 0.2521 | 0.2069 | 0.2064  | 0.2236 | 0.2162    |
| scene    | 0.1291 | 0.1287 | 0.2406 | 0.2911  | 0.1026 | 0.1021    |
| yeast    | 0.2438 | 0.2422 | 0.2262 | 0.2261  | 0.2288 | 0.2090    |
| medical  | 0.0101 | 0.0099 | 0.5224 | 0.7170  | 0.0100 | 0.0097    |
| enron    | 0.0519 | 0.0512 | 0.2330 | 0.1930  | 0.0487 | 0.0478    |
| CAL500   | 0.1607 | 0.1573 | 0.2493 | 0.1613  | 0.1663 | 0.1430    |

表 3 不同方法在微观 F1 值指标下的实验结果

|          | BR     | BR-C   | BPMLL  | BPMLL-C · | RAkEL  | RAkEL-C |
|----------|--------|--------|--------|-----------|--------|---------|
| emotions | 0.5890 | 0.5892 | 0.6838 | 0.6936    | 0.6255 | 0.6308  |
| scene    | 0.6385 | 0.6391 | 0.5256 | 0.5366    | 0.6936 | 0.6981  |
| yeast    | 0.5910 | 0.5926 | 0.6472 | 0.6521    | 0.6186 | 0.6322  |
| medical  | 0.8137 | 0.8151 | 0.0447 | 0.0704    | 0.8160 | 0.8169  |
| enron    | 0.5377 | 0.5358 | 0.3195 | 0.3577    | 0.5634 | 0.5460  |
| CAL500   | 0.3474 | 0.3487 | 0.4594 | 0.4738    | 0.3749 | 0.3285  |

表 4 不同方法在覆盖率指标下的实验结果

|          | BR     | BR-C · | BPMLL   | BPMLL-C · | RAkEL  | RAkEL-C · |
|----------|--------|--------|---------|-----------|--------|-----------|
| emotions | 2.6084 | 2.0561 | 1.7869  | 1.7330    | 1.9785 | 1.8037    |
| scene    | 1.2623 | 0.9378 | 0.8668  | 0.8212    | 0.6749 | 0.6396    |
| yeast    | 9.2105 | 6.9544 | 6.4867  | 6.3733    | 7.0016 | 6.5117    |
| medical  | 3.8078 | 2.3852 | 21.7725 | 7.3590    | 4.5665 | 2.2016    |
| enron    | 22.829 | 13.529 | 16.076  | 13.109    | 20.621 | 12.683    |
| CAL500   | 167.87 | 130.85 | 127.88  | 127.78    | 152.39 | 130.75    |

表 5 不同方法在排序损失指标下的实验结果

|          | BR     | BR-C · | BPMLL  | BPMLL-C · | RAkEL  | RAkEL-C · |
|----------|--------|--------|--------|-----------|--------|-----------|
| emotions | 0.3597 | 0.2330 | 0.1656 | 0.1582    | 0.2191 | 0.1819    |
| scene    | 0.2886 | 0.1682 | 0.1596 | 0.1480    | 0.1357 | 0.1105    |
| yeast    | 0.3577 | 0.2099 | 0.1817 | 0.1764    | 0.2333 | 0.1826    |
| medical  | 0.0691 | 0.0396 | 0.4633 | 0.1409    | 0.1440 | 0.0405    |
| enron    | 0.1728 | 0.0916 | 0.1556 | 0.0974    | 0.3256 | 0.0847    |
| CAL500   | 0.3088 | 0.1967 | 0.1774 | 0.1759    | 0.3580 | 0.2749    |

多数情况下, MLHNC 通过挖掘和利用标签关联, 可以有效提升基学习方法在标签分类和标签排序两方面的性能. 此外, MLHNC 对于三种采用不同标签关联策略的多标签学习方法均是有效的, 具有较好的适用性. 对于 scene 和 medical 数据集, BPMLL-C 在汉明损失指标上的性能低于 BPMLL. 这主要是因为相对于 BR 和 RAkEL, BPMLL 对以上两个数据集在汉明损失指标上的表现比较差, 即误分类的标签数量过多(在 medical 数据集中, 误分标签的数量超过了 50%), 进而影响了 MLHNC 利用 BPMLL 的预测结果挖掘和利用标签关联的有效性和准确性. 对于 enron 数据集, BR-C 和 RAkEL-C 在微观 F1 值上的性能要稍差于 BR 和 RAkEL. 这主要是因为 BR-C 和 RAkEL-C 将较多的相关标签错分为不相关标签, 即把“1”错分为“0”, 这导致模型在微观召回率指标上性能的显著下降, 进而影响了模型在微观 F1 值指标上的性能.

MLHNC 不仅可以有效提升传统多标签学习方法的性能, 还能够提供描述重要高阶标签关联关系的可读性学习结果. 在 MLHNC 中, 超边能够有效表征高阶标签关联信息, 超边权重越高, 其对应标签间的关联性越强. 表 6 列出了在阶数设定为 2 和 3 时, BR-C 在 emotions 数据集上通过演化学习得到的权重最高的 5 条超边, 其中第一条 3 阶超边表示当一个样本与  $y_3$  关联并

且与  $y_1$  和  $y_5$  无关时, 那么该样本很可能与  $y_6$  无关.

表 6 BR-C 在 emotions 数据集上获得的权重最高的 5 条 2 阶和 3 阶超边

| 2 阶超边              |           | 3 阶超边                       |           |
|--------------------|-----------|-----------------------------|-----------|
| 顶点集合               | 标签        | 顶点集合                        | 标签        |
| $y_3 = 1, y_5 = 0$ | $y_1 = 0$ | $y_2 = 0, y_3 = 1, y_5 = 0$ | $y_6 = 0$ |
| $y_1 = 0, y_3 = 1$ | $y_6 = 0$ | $y_2 = 0, y_3 = 1, y_5 = 0$ | $y_3 = 1$ |
| $y_2 = 1, y_4 = 0$ | $y_6 = 0$ | $y_2 = 0, y_3 = 1, y_4 = 0$ | $y_1 = 0$ |
| $y_3 = 0, y_5 = 0$ | $y_6 = 1$ | $y_2 = 0, y_3 = 0, y_5 = 0$ | $y_6 = 1$ |
| $y_2 = 1, y_4 = 0$ | $y_5 = 0$ | $y_2 = 0, y_3 = 1, y_4 = 0$ | $y_5 = 0$ |

## 6 结论

本文提出了一种基于标签关联的多标签演化超网络模型. 该模型通过改变传统超网络模型结构, 使超边具备了表征高阶标签关联关系的能力, 并通过演化学习充分挖掘和利用高阶标签关联, 最终输出综合标签关联和特征信息的预测结果. 通过在 6 个数据集上的实验结果表明, MLHNC 不仅能够有效利用高阶标签关联信息, 提升采用不同标签关联策略的多标签学习方法性能, 而且可以通过超边提供表示重要高阶标签关联关系的具有良好可读性的学习结果.

## 参考文献

- [1] 罗会兰, 郭敏杰, 孔繁胜. 一种基于多级空间视觉词典集体的图像分类方法[J]. 电子学报, 2015, 43(4): 684-693.
- [2] LUO Hui-lan, GUO Min-jie, KONG Fan-sheng. An image classification method based on multiple level spatial visual dictionary ensemble[J]. Acta Electronica Sinica, 2015, 43(4): 684-693. (in Chinese)
- [3] TSOUMAKAS G, KATAKIS I, VLAHAVAS I. Mining Multi-Label Data[A]. Data Mining and Knowledge Discovery Handbook[M]. Springer, 2010. 667-85.
- [4] ZHANG M L, ZHOU Z H. A review on multi-label learning algorithms[J]. IEEE Transactions on Knowledge and Data Engineering, 2014, 26(8): 1819-1837.
- [5] 李思男, 李宁, 李战怀. 多标签数据挖掘技术: 研究综述[J]. 计算机科学, 2013, 4(4): 14-21.
- [6] LI Si-nan, LI Ning, LI Zhan-huai. Multi-label data mining: a survey[J]. Computer Science, 2013, 4(4): 14-21. (in Chinese)
- [7] BOUTELL M R, LUO J B, SHEN X P, et al. Learning multi-label scene classification[J]. Pattern Recognition, 2004, 37(9): 1757-1771.
- [8] YU Y, PEDRYCZ W, MIAO D. Neighborhood rough sets based multi-label classification for automatic image annotation[J]. International Journal of Approximate Reasoning, 2013, 54(9): 1373-1387.
- [9] 刘杰, 杜军平. 基于潜在主题融合的跨媒体图像语义标

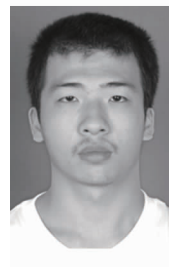
- 注[J]. 电子学报, 2014, 42(5): 987 - 991.
- LIU Jie, DU Jun-ping. Latent topic fusion-based cross-media image semantic annotation[J]. Acta Electronica Sinica, 2014, 42(5): 987 - 991. (in Chinese)
- [8] SCHAPIRE R E, SINGER Y. BoosTexter; a boosting-based system for text categorization[J]. Machine Learning, 2000, 39(2-3): 135 - 68.
- [9] ÎANCULEF R, FLAOUNAS I, CRISTIANINI N. Efficient classification of multi-labeled text streams by clashing[J]. Expert Systems with Applications, 2014, 41(11): 5431 - 5450.
- [10] 江雨燕, 李平, 王清. 基于共享背景主题的 Labeled LDA 模型[J]. 电子学报, 2013, 41(9): 1794 - 1799.
- JIANG Yu-yan, LI Ping, WANG Qing. Labeled LDA model based on shared background topics[J]. Acta Electronica Sinica, 2013, 41(9): 1794 - 1799. (in Chinese)
- [11] ZHANG M L, ZHOU Z H. Multilabel neural networks with applications to functional genomics and text categorization[J]. IEEE Transactions on Knowledge and Data Engineering, 2006, 18(10): 1338 - 1351.
- [12] YU G X, RANGWALA H, DOMENICONI C, et al. Protein function prediction with incomplete annotations[J]. IEEE Transactions on Computational Biology and Bioinformatics, 2014, 11(3): 579 - 591.
- [13] LO H-Y, WANG J-C, WANG H-M, et al. Cost-sensitive multi-label learning for audio tag annotation and retrieval [J]. IEEE Transactions on Multimedia, 2011, 13(3): 518 - 529.
- [14] ZHANG B T. Hypernetworks: a molecular evolutionary architecture for cognitive learning and memory[J]. IEEE Computational Intelligence Magazine, 2008, 3(3): 49 - 63.
- [15] ZHANG M L, ZHOU Z H. ML-KNN: A lazy learning approach to multi-label learning [J]. Pattern Recognition, 2007, 40(7): 2038 - 2048.
- [16] RNKRANZ J F, LLERMEIER E H, MENC A E L, et al. Multilabel classification via calibrated label ranking[J]. Machine Learning, 2008, 73(2): 133 - 153.
- [17] TSOUMAKAS G, KATAKIS I, VLAHAVAS L. Random k-labelsets for multilabel classification[J]. IEEE Transactions on Knowledge and Data Engineering, 2011, 23(7): 1079 - 1089.
- [18] READ J, PFAHRINGER B, HOLMES G, et al. Classifier chains for multi-label classification [J]. Machine Learning, 2011, 85(3): 333 - 359.
- [19] MONTA ES E, SENGE R, BARRANQUERO J, et al. Dependent binary relevance models for multi-label classification[J]. Pattern Recognition, 2014, 47(3): 1494 - 1508.
- [20] SUN K W, LEE C H, XIE X F. MLHN: a hypernetwork model for multi-label classification[J]. International Journal of Pattern Recognition & Artificial Intelligence, 2015, 26(6): 1550020.
- [21] KIM S J, HA J W, ZHANG B T. Bayesian evolutionary hypergraph learning for predicting cancer clinical outcomes [J]. Journal of Biomedical Informatics, 2014, 49(6): 101 - 111.
- [22] TSOUMAKAS G, SPYROMITROS-XIOUFIS E, VILCEK J, et al. Mulan: A java library for multi-label learning [J]. The Journal of Machine Learning Research, 2011, 12(7): 2411 - 2414.

#### 作者简介



王进男, 1979年1月出生于重庆, 工学博士, 2008年在韩国仁荷大学获得工学博士学位, 现为重庆邮电大学教授. 主要研究方向为数据挖掘、机器学习、智能交通.

E-mail: wangjin@cqupt.edu.cn



刘彬男, 1989年11月出生于河北保定, 2016年在重庆邮电大学获得工学硕士学位, 现为重庆邮电大学助教. 主要研究方向为数据挖掘.