IJCAI24 Learning Label-Specific Multiple Local Metrics for Multi-Label Classification
202200130267 张宇欣

1. 问题陈述

现有的多标签度量学习方法可能并非最优,因为现有的多标签度量学习方法通常学习一个统一的度量，应用于所有标签空间的实例。然而，多标签实例在每个标签空间内的分布通常是非线性的。例如：

　　在多媒体标注中，标签“太阳”可能同时出现在描述森林和城市的图像中。

　　在文本分类中，标签“总统”可能同时出现在政治和经济语境中。

在这种情况下,使用单一的一致性度量标准在同时将相似实例拉近和将不相似实例推远方面会面临巨大挑战。

2. 方法概述

提出了 LSMM，这是一种用于多标签分类的标签特定多度量学习框架，通过在全局度量的基础上学习不同实例的标签特定多个局部度量，来考虑多标签实例的非线性分布特征。将每个标签空间内的多标签实例划分为若干不相交的子集（基于语义或聚类策略），并为每个子集训练一个局部度量。引入一个全局度量，捕捉所有标签之间的高阶相关性，用以补充局部度量的局限性。

LSMM 的核心思想是通过构建标签特定的局部度量和全局度量相结合的方式来区分多标签样本的相似性和差异性。

1. 标签特定辅助信息的构建

首先将数据集中的样本分为正样本集合和负样本集合

$$\mathcal{P}_p = \{\boldsymbol{x}_i | (\boldsymbol{x}_i, Y_i) \in \mathcal{D}, l_p \in Y_i\},$$
$$\mathcal{N}_p = \{\boldsymbol{x}_i | (\boldsymbol{x}_i, Y_i) \in \mathcal{D}, l_p \notin Y_i\}.$$

为了降低计算成本，并提升训练效率，选择了部分最近邻的样本对作为 targets 和 imposters。

模型的调节方式：

较大的 λ1/λ2：倾向于学习单一的全局度量 L0。

较小的 λ1/λ2：倾向于学习不相关的局部度量（即全局度量 L0 接近零）。

2.分区策略：

在语义分区策略中，对每个标签 l_p，正样本和负样本分别分配不同的局部度量。样本对的平方距离定义为：

$$\mathrm{Dis}^2_{\mathbf{L}_0+\mathbf{L}_p^{s(\boldsymbol{x}_i)}}(\boldsymbol{x}_i, \boldsymbol{x}_j) = ||(\mathbf{L}_0 + \mathbf{L}_p^{s(\boldsymbol{x}_i)})^\top (\boldsymbol{x}_i - \boldsymbol{x}_j)||_2^2. \quad (3)$$

优化目标：

$$\min_{\mathbf{L}_0,\left\{\mathbf{L}_p^1,\mathbf{L}_p^0\right\}_{p=1}^q} \frac{1}{q}\sum_{p=1}^{q}\left(\ell_{\text{pos}}^p+\ell_{\text{neg}}^p\right)+\frac{\lambda_1}{2q}\sum_{p=1}^{q}\sum_{t=0}^{1}||\mathbf{L}_p^t||_F^2 + \lambda_2||\mathbf{L}_0||_F^2. \quad (4)$$

3.聚类分区

聚类分区策略使用 k-means 算法，将样本划分为 C 个不相交的簇 D，样本对的平方距离定义为

$$\text{Dis}^2_{\mathbf{L}_0+\mathbf{L}_p^{c(\boldsymbol{x}_i)}}(\boldsymbol{x}_i,\boldsymbol{x}_j)=||(\mathbf{L}_0+\mathbf{L}_p^{c(\boldsymbol{x}_i)})^\top(\boldsymbol{x}_i-\boldsymbol{x}_j)||_2^2. \quad (6)$$

优化目标，簇内 C 被设为 3

$$\ell_m^p=\frac{1}{|\mathcal{C}^m|}\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\sim\mathcal{T}_p\\ \boldsymbol{x}_i\in\mathcal{C}^m}}\ell\left(\theta_{ij}^p\left(\gamma-\alpha\text{Dis}^2_{\mathbf{L}_0+\mathbf{L}_p^m}(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)\right)$$

$$\min_{\mathbf{L}_0,\left\{\mathbf{L}_p^1,\mathbf{L}_p^2,\ldots,\mathbf{L}_p^C\right\}_{p=1}^q}\frac{1}{q}\sum_{p=1}^{q}\sum_{m=1}^{C}\ell_m^p+\frac{\lambda_1}{Cq}\sum_{p=1}^{q}\sum_{t=1}^{C}||\mathbf{L}_p^t||_F^2 + \lambda_2||\mathbf{L}_0||_F^2. \quad (7)$$

4.优化过程：

通过 L-BFGS 算法解决语义划分和聚类划分策略中的非线性优化问题。

对于基于语义的分区策略

For the semantic-based partition strategy, the first-order derivatives of $\mathcal{L}_{se}$ w.r.t $\mathbf{L}_0$ and $\left\{\mathbf{L}_p^1,\mathbf{L}_p^0\right\}_{p=1}^q$ respectively are

$$\frac{\partial\mathcal{L}_{se}}{\partial\mathbf{L}_0}=\frac{2}{q}\sum_{p=1}^{q}\left(\frac{1}{|\mathcal{P}_p|}\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\sim\mathcal{T}_p\\ \boldsymbol{x}_i\in\mathcal{P}_p}}\sigma_{ij}^p\theta_{ij}^p\mathbf{A}_{ij}\left(\mathbf{L}_0+\mathbf{L}_p^1\right)\right.$$
$$\left.+\frac{1}{|\mathcal{N}_p|}\sum_{\substack{(\boldsymbol{x}_i,\boldsymbol{x}_j)\sim\mathcal{T}_p\\ \boldsymbol{x}_i\in\mathcal{N}_p}}\sigma_{ij}^p\theta_{ij}^p\mathbf{A}_{ij}\left(\mathbf{L}_0+\mathbf{L}_p^0\right)\right)+2\lambda_2\mathbf{L}_0$$

$$\frac{\partial \mathcal{L}_{se}}{\partial \mathbf{L}_p^1} = \frac{2}{q|\mathcal{P}_p|} \sum_{\substack{(\boldsymbol{x}_i, \boldsymbol{x}_j) \sim \mathcal{T}_p \\ \boldsymbol{x}_i \in \mathcal{P}_p}} \sigma_{ij}^p \theta_{ij}^p \mathbf{A}_{ij} \left(\mathbf{L}_0 + \mathbf{L}_p^1\right) + \frac{\lambda_1}{q} \mathbf{L}_p^1 \tag{9}$$

$$\frac{\partial \mathcal{L}_{se}}{\partial \mathbf{L}_p^0} = \frac{2}{q|\mathcal{N}_p|} \sum_{\substack{(\boldsymbol{x}_i, \boldsymbol{x}_j) \sim \mathcal{T}_p \\ \boldsymbol{x}_i \in \mathcal{N}_p}} \sigma_{ij}^p \theta_{ij}^p \mathbf{A}_{ij} \left(\mathbf{L}_0 + \mathbf{L}_p^0\right) + \frac{\lambda_1}{q} \mathbf{L}_p^0 \tag{10}$$

对于基于聚类的分区策略：

For the cluster-based partition strategy, the first-order derivatives of $\mathcal{L}_{cl}$ w.r.t $\mathbf{L}_0$ and $\left\{\mathbf{L}_p^1, \mathbf{L}_p^2, \ldots, \mathbf{L}_p^C\right\}_{p=1}^q$ respectively are

$$\frac{\partial \mathcal{L}_{cl}}{\partial \mathbf{L}_0} = \frac{2}{q} \sum_{p=1}^{q} \sum_{m=1}^{C} \frac{1}{|\mathcal{C}^m|} \sum_{\substack{(\boldsymbol{x}_i, \boldsymbol{x}_j) \sim \mathcal{T}_p \\ \boldsymbol{x}_i \in \mathcal{C}^m}} \zeta_{ij}^p \theta_{ij}^p \mathbf{A}_{ij} \left(\mathbf{L}_0 + \mathbf{L}_p^m\right) + 2\lambda_2 \mathbf{L}_0 \tag{13}$$

$$\frac{\partial \mathcal{L}_{cl}}{\partial \mathbf{L}_p^m} = \frac{2}{q|\mathcal{C}^m|} \sum_{\substack{(\boldsymbol{x}_i, \boldsymbol{x}_j) \sim \mathcal{T}_p \\ \boldsymbol{x}_i \in \mathcal{C}^m}} \zeta_{ij}^p \theta_{ij}^p \mathbf{A}_{ij} \left(\mathbf{L}_0 + \mathbf{L}_p^m\right) + \frac{2\lambda_1}{Cq} \mathbf{L}_p^m \tag{14}$$

在利用上述 LSMM 框架学习到全局度量和特定标签的多个局部度量之后，可依据公式（3）和公式（6）结合全局度量和相应的特定标签局部度量来测量多标签实例之间的距离。利用 KNN 方法，可对多标签数据进行分类预测。

该框架具有以下特点：
作者设计了两种划分策略：
语义分割: 根据样本是否为某标签的正例或负例进行划分，并为每个分区训练独立的度量。
聚类分割: 使用 k-means 将样本分为多个簇，并为每个簇分配不同的度量。
全局与局部度量学习相结合:
每个标签空间分别训练局部度量，以捕获与特定标签相关的局部特性。

学习一个全局度量，用于捕获跨标签的高阶相关性。全局度量结合局部度量，共同衡量样本间的距离。
优化方法：
优化目标是通过最小化样本间的距离误差来推近相似样本，同时拉远不相似样本。损失函数还包含正则化项，用于控制局部与全局度量的平衡。提升模型的泛化能力

作者的方法基于前人的研究成果进行了改进。以下是部分基于的核心方法：
度量学习:使用了 Mahalanobis 距离及其分解形式。特别是利用矩阵分解的方式，简化了正定约束并加速了优化。
目标点和干扰点选择:通过选择部分邻居（目标点和干扰点）减少计算量。
多标签分类方法:使用了基于 KNN 的经典方法，这些方法广泛应用于多标签分类任务。

3. 实验

采用了九个具有多样化属性的基准多标签数据集。

Dataset	$\|\mathcal{D}\|$	$dim(\mathcal{D})$	$L(\mathcal{D})$	$LCard(\mathcal{D})$	Domain
emotions	593	72	6	1.869	Music[1]
birds	645	258	19	1.014	Audio[1]
medical	978	1449	45	1.245	Text[1]
enron	1702	1001	53	3.378	Text[1]
image	2000	294	5	1.236	Image[2]
scene	2407	294	6	1.074	Image[1]
slashdot	3782	1079	22	1.177	Text[3]
arts	5000	462	26	1.636	Text[1]
education	5000	550	33	1.461	Text[1]

[1] http://mulan.sourceforge.net/datasets.html
[2] http://palm.seu.edu.cn/zhangml/Resources.htm#data
[3] https://waikato.github.io/meka/datasets/

在性能评估方面，采用了六种广泛用于多标签 分类的评估指标，包括汉明损失、排序损失、覆盖率、平均精度、宏 F1 值和宏平均 AUC。
实验验证 LSMM 框架的有效性，通过对比其在多标签分类中的表现。使用两种 KNN 基础多标签分类方法：BR-KNN 和 ML-KNN。不同的多标签度量学习算法与 KNN 方法结合，形成对比方法。

Compared Algorithms	Datasets								
	emotions	birds	medical	enron	image	scene	slashdot	arts	education
					Hamming Loss ↓				
BR-KNN	0.263±0.023	0.056±0.007	0.016±0.002	0.055±0.002	0.167±0.016	0.088±0.008	0.065±0.001	0.073±0.002	0.038±0.001
BR-KNN-LM	0.270±0.019	0.065±0.009	**0.011±0.002**	0.048±0.003	0.180±0.016	0.088±0.012	0.044±0.003	0.055±0.002	0.038±0.001
BR-KNN-LJE	0.219±0.022	0.055±0.007	0.022±0.003	0.059±0.003	0.184±0.018	0.110±0.011	0.060±0.002	0.061±0.001	0.043±0.002
BR-KNN-COMMU	0.263±0.023	0.056±0.007	0.016±0.002	0.055±0.002	0.167±0.016	0.088±0.008	0.056±0.001	0.072±0.002	0.038±0.001
BR-KNN-LIMIC	0.212±0.008	0.053±0.006	0.014±0.002	0.049±0.003	0.165±0.016	0.083±0.008	0.045±0.002	0.058±0.002	0.039±0.001
BR-KNN-LSMM-SE	0.204±0.018	0.051±0.006	0.014±0.003	0.045±0.003	0.162±0.013	0.081±0.007	**0.035±0.002**	0.054±0.001	0.037±0.002
BR-KNN-LSMM-CL	**0.202±0.017**	**0.050±0.007**	0.012±0.002	**0.044±0.003**	**0.157±0.017**	**0.079±0.009**	0.037±0.002	**0.052±0.002**	0.038±0.001
ML-KNN	0.262±0.022	0.054±0.006	0.016±0.002	0.052±0.003	0.171±0.013	0.084±0.009	0.058±0.001	0.060±0.001	0.038±0.001
ML-KNN-LM	0.254±0.017	0.054±0.007	0.013±0.002	0.048±0.002	0.174±0.015	0.086±0.010	0.045±0.003	0.054±0.001	0.038±0.001
ML-KNN-LJE	0.227±0.022	0.054±0.007	0.023±0.003	0.058±0.002	0.184±0.017	0.109±0.009	0.056±0.001	0.060±0.001	0.042±0.002
ML-KNN-COMMU	0.262±0.022	0.054±0.006	0.015±0.002	0.052±0.003	0.171±0.013	0.084±0.009	0.050±0.001	0.059±0.001	0.038±0.001
ML-KNN-LIMIC	0.215±0.009	0.053±0.006	0.013±0.002	0.050±0.003	0.164±0.017	0.082±0.007	0.050±0.003	0.057±0.002	0.038±0.002
ML-KNN-LSMM-SE	**0.204±0.016**	**0.050±0.005**	**0.012±0.003**	**0.043±0.002**	0.159±0.014	0.080±0.007	0.041±0.002	0.053±0.001	**0.037±0.001**
ML-KNN-LSMM-CL	0.209±0.018	0.051±0.006	**0.012±0.002**	0.045±0.003	**0.155±0.013**	**0.078±0.007**	**0.039±0.003**	**0.052±0.001**	0.038±0.001
					Ranking Loss ↓				
BR-KNN	0.272±0.048	0.516±0.064	0.081±0.028	0.228±0.023	0.180±0.020	0.095±0.014	0.417±0.026	0.354±0.029	0.244±0.013
BR-KNN-LM	0.256±0.030	0.527±0.066	0.105±0.033	0.201±0.011	0.191±0.020	0.097±0.016	0.221±0.021	**0.257±0.015**	0.222±0.014
BR-KNN-LJE	0.202±0.035	0.497±0.067	0.114±0.037	0.244±0.027	0.205±0.020	0.124±0.019	0.345±0.024	0.265±0.015	0.234±0.012
BR-KNN-COMMU	0.272±0.048	0.516±0.064	0.088±0.031	0.230±0.023	0.180±0.020	0.095±0.014	0.407±0.027	0.357±0.033	0.244±0.013
BR-KNN-LIMIC	0.182±0.028	0.513±0.066	0.077±0.023	0.245±0.020	0.171±0.024	0.102±0.011	0.393±0.023	0.348±0.021	0.220±0.015
BR-KNN-LSMM-SE	**0.170±0.033**	**0.472±0.061**	**0.065±0.018**	0.194±0.026	0.157±0.021	0.092±0.011	0.201±0.018	0.263±0.010	0.210±0.010
BR-KNN-LSMM-CL	0.178±0.034	0.487±0.058	0.071±0.023	**0.192±0.022**	**0.150±0.025**	**0.090±0.013**	**0.190±0.025**	0.258±0.027	**0.205±0.012**
ML-KNN	0.258±0.038	0.295±0.035	0.032±0.009	0.091±0.010	0.173±0.018	0.075±0.011	0.182±0.013	0.148±0.008	0.097±0.004
ML-KNN-LM	0.240±0.026	0.298±0.047	0.034±0.015	0.090±0.008	0.177±0.018	0.078±0.012	0.112±0.008	0.129±0.007	0.078±0.005
ML-KNN-LJE	0.201±0.030	0.299±0.047	0.053±0.017	0.111±0.011	0.201±0.022	0.110±0.017	0.198±0.013	0.145±0.009	0.089±0.003
ML-KNN-COMMU	0.258±0.038	0.295±0.035	0.038±0.013	0.091±0.010	0.173±0.018	0.075±0.011	0.159±0.012	0.147±0.008	0.097±0.004
ML-KNN-LIMIC	0.184±0.020	0.263±0.039	**0.031±0.012**	0.088±0.010	0.163±0.022	0.072±0.007	0.133±0.009	0.137±0.006	0.079±0.004
ML-KNN-LSMM-SE	**0.166±0.033**	**0.256±0.034**	0.034±0.015	0.075±0.007	0.157±0.021	0.072±0.004	0.095±0.008	0.123±0.006	0.078±0.004
ML-KNN-LSMM-CL	0.170±0.036	0.260±0.038	0.038±0.013	**0.073±0.009**	**0.146±0.020**	**0.069±0.005**	**0.083±0.007**	**0.120±0.005**	**0.074±0.004**

实验结果中可以看到，LSMM-SE 和 LSMM-CL 显著提升了 BR-KNN 和 ML-KNN 的分类性能。在绝大多数情况下优于 LM、LJE、COMMU 和 LIMIC。LSMM 框架增强的 KNN 方法（如 BR-KNN-LSMM 和 ML-KNN-LSMM）与其他多标签分类方法（如 LIFT、WRAP、HOMI 等）对比，表现出更优异或至少可比的性能。
结果表明，尽管简单的 BR-KNN 和 ML-KNN 的性能不如最先进的多标签分类方法，但 LSMM 增强版在统 计上优于或至少与最先进的多标签分类方法相当。这些结果再次验证了 LSMM 框架在学习多标签分类的有效相似性度量方面的优越性。

4. 思考

论文提出了新型的标签特定多指标学习框架（LSMM）。针对多标签分类问题，LSMM 框架通过学习标签特定的多个局部指标，捕获了多标签实例的非线性分布特性。在此基础上还保留了一个全局指标，从而实现了全局与局部特性的综合平衡。实验验证表明，LSMM 在学习有效的相似性指标方面优于现有的多标签一致性指标学习算法。作者明确指出这种标签特定方法可能面临在极端多标签学习场景下的泛化难题，并提出进一步探索这一方向的重要性。

LSMM 需要为每个标签学习特定的局部指标，当标签数量非常多时（如在极端多标签学习中），计算成本和存储成本都会显著增加，这可能影响其实际应用场景的可行性。LSMM 的多指标学习框架可能引入额外的超参数，这可能会增加模型调试的复杂性。

未来可以针对标签数量极高场景下的扩展性问题，可探索如何通过标签嵌入、标签聚类或稀疏学习策略减少计算成本，同时保持性能。针对极端多标签学习中性能与复杂度的权衡，提出多目标优化框架，平衡精度、可扩展性与计算开销。

5. 其它（选填）

需要特别记录的其它笔记