

基于自编码器及超图学习的多标签特征提取

唐朝辉^{1,2} 朱清新¹ 洪朝群² 祝峰³

摘要 在实际应用场景中越来越多的数据具有多标签的特性,且特征维度较高,包含大量冗余信息.为提高多标签数据挖掘的效率,多标签特征提取已经成为当前研究的热点.本文采用去噪自编码器获取多标签数据特征空间的鲁棒表达,在此基础上结合超图学习理论,融合多个标签对样本间几何关系的影响以提升特征提取的性能,构建多标签数据样本间几何关系所对应超图的 Laplacian 矩阵,并通过 Laplacian 矩阵的特征值分解得到低维投影空间.实验结果证明了本文所提出的算法在分类性能上是有效可行的.

关键词 深度学习, 自编码器, 多标签, 超图, 特征提取

引用格式 唐朝辉, 朱清新, 洪朝群, 祝峰. 基于自编码器及超图学习的多标签特征提取. 自动化学报, 2016, 42(7): 1014–1021

DOI 10.16383/j.aas.2016.c150736

Multi-label Feature Selection with Autoencoders and Hypergraph Learning

TANG Chao-Hui^{1,2} ZHU Qing-Xin¹ HONG Chao-Qun² ZHU William³

Abstract In practical application scenarios, more and more data tend to be assigned with multiple labels and contain much redundant information in the high dimensional feature space. To improve the efficiency and effectiveness of multi-label data mining, multi-label data feature selection has become a hotspot. This paper utilizes denoising autoencoders to obtain a more robust version of multi-label data feature representation. Furthermore, based on hypergraph learning theory, a hypergraph Laplacian matrix corresponding to multi-label data is constructed by fusing the effects of all labels on geometrical relationship among all the samples, and then a projection space with lower dimension is obtained by conducting eigenvalue decomposition of the Laplacian matrix. Experimental results demonstrate the effectiveness and feasibility of the proposed algorithm according to its multi-label data classification performance.

Key words Deep learning, autoencoders, multi-label, hypergraph, feature selection

Citation Tang Chao-Hui, Zhu Qing-Xin, Hong Chao-Qun, Zhu William. Multi-label feature selection with autoencoders and hypergraph learning. *Acta Automatica Sinica*, 2016, 42(7): 1014–1021

多标签学习是数据挖掘和信息检索中很重要的主题^[1–9]. 多标签数据中的每个样本都会对应一个标签集合,这在实际应用中非常普遍,比如蛋白质功能分类^[6]、图像标注^[10]以及图像情景分类^[11]等.

多标签学习面临两个重要的挑战. 首先,传统的单标签学习中样本的分类是互斥的,而多标签学习中的分类类型相互依赖、相互关联. 比如图像标注应用中,一个图像可能同时具有“树”、“雨水”、“彩虹”以及“湖水”等标签,而另一个图则具有“树”、“太阳”、“彩虹”以及“沙漠”等标签,即不同的样本可能具有部分相同的标签. 其次,多标签数据通常具有较高维度的特征向量. 比如一张图像数据的维度可能是几兆,一个文本的维度通常可以 10 k 以上,而高维度的数据很容易导致“维度灾难”. 为了解决这个问题,研究者们已经提出了一些多标签降维算法^[1, 12–14],虽然这些算法可以在一定程度上有效地融合多个标签之间关系以实现高维多标签数据的降维,但这些算法忽略了多标签数据内含的噪声以及样本间几何关系对于多标签数据特征空间降维的影响,而这对于提高多标签特征提取算法的性能至关重要.

为了有效提取高维多标签数据的低维表达性能,本文首先利用去噪自编码器对原始特征空间进行多层

收稿日期 2015-11-09 录用日期 2016-05-03
Manuscript received November 9, 2015; accepted May 3, 2016
国家自然科学基金 (61300192, 61472110, 61573297, 61379049),
中央高校基本科研项目 (ZYGX2014J052), 福建省自然科学基金
(2014J01256, 2015J01277) 资助
Supported by National Natural Science Foundation of China
(61300192, 61472110, 61573297, 61379049), the Fundamental
Research Funds for the Central Universities (ZYGX2014J052),
and the Natural Science Foundation of Fujian Province
(2014J01256, 2015J01277)

本文责任编辑 柯登峰
Recommended by Associate Editor KE Deng-Feng
1. 电子科技大学信息与软件工程学院 成都 611731 2. 厦门理工学院
计算机与信息工程学院 厦门 361024 3. 闽南师范大学粒计算实验
室 漳州 363000
1. School of Information and Software Engineering, University
of Electronic Science and Technology of China, Chengdu 611731
2. School of Computer and Information Engineering, Xiamen
University of Technology, Xiamen 361024 3. Laboratory of
Granular Computing, Minnan Normal University, Zhangzhou
363000

次抗干扰处理, 以便提取出比原始特征空间更鲁棒的表达; 其次, 利用超图理论来挖掘多标签特征空间样本之间的几何关系, 并有效融合多个标签对样本间几何关系的影响, 构建出完备的 Laplacian 矩阵并通过矩阵的标准特征值分解获得低维特征空间。

1 相关工作

1.1 多标签学习

本文多标签学习算法中, X 表示特征空间, C 与 Y 表示标签空间, 并且它们都是非空有限集。

传统学习机 $L = (X, C, T)$ 的目标是通过学习获得一个特征空间 X 与标签空间 C 的映射, 其中 $|C| = 1$, 即单标签分类器。大量的学者对单标签分类器进行了深入研究, 也取得了良好的分类性能, 但单标签分类器基于一个分类样本只有一个特定的标签的假设, 而这个假设在很多实际应用场景中并不成立^[6, 11]。因此越来越多的学者通过构建多标签学习机来处理实际应用中越来越多的多标签数据^[4, 9]。

$ML = (X, Y, T)$ 定义了一个多标签学习机, 其中 X 是特征空间, Y 是标签空间, $T = \{(x_i, l_i) | 1 \leq i \leq n\}$ 是一个训练集。 $x_i \in X$ 是其中的一个训练样本, 包含了一个对象所有的特征值, $l_i \subseteq Y$ 是 x_i 对应的标签集合, n 是训练集中样本的个数, 通常 $|Y| \geq 2$ 。多标签学习机 $ML = (X, Y, T)$ 的目标是通过学习获得一个映射函数 $map: X \rightarrow 2^Y$, 通常又称为一个多标签分类器。

多标签分类器 $map: X \rightarrow 2^Y$ 的目标函数可以被形式化地定义为: $Conf: X \times Y \rightarrow \mathbf{R}$ 。这意味着给定一个无标签的训练样本 $x_i \in X$, 实值函数 $Conf: X \times Y \rightarrow \mathbf{R}$ 的输出是样本 x_i 拥有标签 $y_i \in Y$ 的置信度。另外, $Conf(x_i, y_i)$ 可以转换成标签排序函数 $rank(x_i, y_i)$ 的形式, 它们之间的对应关系为: 如果 $rank(x_i, y_1) \leq rank(x_i, y_2)$, 那么 $Conf(x_i, y_1) \geq Conf(x_i, y_2)$, 其中 $y_1 \subseteq Y, y_2 \subseteq Y$ 。

多标签特征提取算法假设所有的标签共享一个特征子集。ARMLNRS^[13] 采用邻域粗糙集技术以及贪婪策略对多标签数据进行特征约简, 但不适用于图像和文本处理; MLFSIE (Multi-label feature selection algorithm based on information entropy)^[12] 通过挖掘标签与特征之间的信息增益来提取有效特征子集; Yu 等^[15] 使用基于无监督学习的潜在语义索引技术来获取多标签数据的低维特征空间, 从而不能充分利用标签与特征之间的关系来提升多标签学习性能; Zhang 等^[1] 通过最大化特征与标签之间的依赖来识别最优特征子集; Sun 等^[7] 利用超图理论分别构建特征空间与标签空间的超图

并计算出保留超图信息的特征子空间, 该方法充分利用了特征与标签之间的关系来提取特征, 但忽略了样本间几何结构对提取特征子空间的影响。

1.2 超图学习

传统采用图与子空间的机器学习理论通常基于流形假设。首先, 假设存在一个低维流形空间, 在该空间上的一个较小的局部邻域内样本具有相似的性质, 建立在此流形空间上的决策函数也具有局部平滑性; 其次, 在传统图模型中, 样本之间的关系是成对的, 没有考虑多个样本之间存在一致的关系^[16-17]。但在多标签数据中多个样本具有相同的性质, 即包含相同的标签, 则需要构建多条边来表达^[18]。

在超图中, 具有相同性质的多个顶点共享一条边, 因而可以使用超图来提高样本间几何关系表达的效率 and 可靠性^[19]。基于超图的样本几何关系表达已经用于多种应用, 比如分类^[20-21]、图像分割^[22] 以及信息检索^[23-24]。

2 基于自编码器与超图学习的多标签特征提取

2.1 自编码器

深度学习在挖掘图像潜在表达上非常有用, 已经成为计算机视觉领域的研究热点。自编码器基于深度学习理论, 是一种无监督的特征学习方法, 编码器的内层可以有效抽取图像的内蕴表达, 其学习策略可以抽象成一个最小化重构误差的凸优化问题:

$$\sum_{i=1}^n \|x_i - \hat{x}_i\|^2 \quad (1)$$

对于去噪自编码器, 数据输入 x_1, \dots, x_n 的局部数据被随机地替换, 从而对原始数据加入了人为的噪声。 \hat{x}_i 表示 x_i 加入噪声后的结果, $W: \mathbf{R}^d \rightarrow \mathbf{R}^d$ 表示用来重构带有人为噪声输入数据的一种映射。因此重构误差的目标函数可以定义为:

$$\frac{1}{2n} \sum_{i=1}^n \|x_i - W\hat{x}_i\|^2 \quad (2)$$

最小化目标函数 (2) 的解在很大程度上取决于输入数据的哪些特征被随机损坏, 为了降低重构过程的方差, 本文采用 MDA (Marginalized denoising auto-encoder)^[25] 对训练数据进行多次处理, 并且在每次处理过程中对输入数据加入不同的随机噪声。

最小二乘的损失函数可以重新定义为:

$$\frac{1}{2mn} \sum_{j=1}^m \sum_{i=1}^n \|x_i - W\hat{x}_{i,j}\|^2 \quad (3)$$

其中, $\hat{x}_{i,j}$ 表示输入数据 x_i 第 j 次加入随机噪声后的值, m 是对同一输入处理的次数, 即 MDA 的层数.

为了提高算法的处理效率, 本文对输入数据进行矩阵化处理. 其中 $X = [x_1, \dots, x_n] \in \mathbf{R}^{d \times n}$ 为输入数据矩阵, $\bar{X} = [X, \dots, X]$ 表示 X 的 m 次拷贝, 而 \hat{X} 是 \bar{X} 加入随机噪声后的数据. 因为 $1/(2mn)$ 是常量, 对最小化损失函数无影响, 因此损失函数 (3) 可简化为式 (4):

$$\begin{aligned} \text{loss}(W) &= \sum_{j=1}^m \sum_{i=1}^n \|x_i - W\hat{x}_{i,j}\|^2 = \\ &\text{tr} \left((\bar{X} - W\hat{X})^T (\bar{X} - W\hat{X}) \right) \end{aligned} \quad (4)$$

其中 $\text{tr}(\cdot)$ 代表矩阵的求迹操作. 最小化式 (4) 是一个具有全局最优解的最小二乘问题, 其解为:

$$W = PQ^{-1} \quad (5)$$

其中, $Q = \hat{X}\hat{X}^T$, $P = \bar{X}\hat{X}^T$.

证明. 因为式 (4) 是关于 W 的凸函数, 只需要满足目标函数的极值必要条件, 即将其方向导数所有分量置零, 便可以求出问题的全局最优解.

$$\begin{aligned} \frac{\partial \text{loss}(W)}{\partial W} &= \left(\frac{\partial \text{tr}[\bar{X}^T \bar{X}]}{\partial W} - \frac{\partial \text{tr}[\bar{X}^T W \hat{X}]}{\partial W} - \right. \\ &\left. \frac{\partial \text{tr}[\hat{X}^T W^T \bar{X}]}{\partial W} + \frac{\partial \text{tr}[\hat{X}^T W^T W \hat{X}]}{\partial W} \right) = \\ &- 2\bar{X}\hat{X}^T + 2W\hat{X}\hat{X}^T \end{aligned} \quad (6)$$

由于 $\hat{X}\hat{X}^T$ 是可逆矩阵, 则由 $\frac{\partial \text{loss}(W)}{\partial W} = 0$, 可以求出最优解 W^* 为:

$$W^* = \bar{X}\hat{X}^T(\hat{X}\hat{X}^T)^{-1} \quad (7)$$

□

自编码阶段对原始多标签数据特征空间进行了特征提取, 提取的特征空间抗干扰能力更强, 但由于没有考虑标签与特征空间之间的关联, 且特征空间维度没有减小, 故在此基础上构建的多标签分类算法的学习精度和时间性能都会受到一定的制约. 基于以上考虑, 本文在自编码的基础上采用基于监督的多标签超图学习以降低多标签数据的特征维度.

2.2 基于超图的多标签特征提取

为了更加清晰地描述本文提出的方法, 首先定义几个重要的标记, 如表 1 所示. 超图中每个顶点对应一个样本, 每条超边描述了多个样本的共同属性. 为了求解超图在平滑约束下的 Laplacian 矩阵, 可以将问题近似为一个实值函数的优化问题^[18, 22]:

$$\arg \min_f \sum_{e \in E} \sum_{u, \nu \in e} \frac{\omega(e)}{\delta(e)} \left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(\nu)}{\sqrt{d(\nu)}} \right)^2 \quad (8)$$

为提高计算效率, 将式 (8) 转换成矩阵的形式:

$$\arg \min_f (fL f^T) \quad (9)$$

其中, $fL f^T$ 可视为 f 相对于超图 Laplacian 矩阵 L 的平滑性度量, 由式 (8) 可得 L :

$$L = I - \frac{1}{\sqrt{D_v}} H \Omega D_e H^T \frac{1}{\sqrt{D_v}} \quad (10)$$

Zhang 等^[26] 首次提出了批对齐框架 (Patch alignment framework, PAF), 该框架统一了基于谱分析的流形学习方法, 是一个强大的流形学习分析、开发工具, 它包含两个阶段: 局部批构建阶段与全局对齐阶段. 超图的几何结构可以由超图对应的 Laplacian 矩阵表征, 对于多标签数据的每个标签 l , 可以构建出一个 Laplacian 矩阵.

表 1 重要的标记定义

Table 1 Definitions of important notations

标记	标记语义
n	训练集中训练样本的个数
u, ν	顶点
$f \in \mathbf{R}^n$	所有样本的得分向量
$f(u)$ 或者 f_u	样本 u 的得分函数
e, E	超边, 超边集合
$\delta(e)$	超边 e 的度
$d(u)$	顶点 u 的度
D_v	顶点集的度矩阵
D_e	超边集的度矩阵
$\omega(e)$	超边 e 的权重
H	超图对应的邻接矩阵
Ω	$\Omega_{(i,i)}$ 是第 i 条超边的权重, 其他取 0
r	约简后的特征维度
I	超图学习前的原特征空间
S	超图学习后的语义特征空间
p_i	基于样本 x_i 的局部批
P_i	基于样本 x_i 的局部批特征投影矩阵

在局部批构建阶段, PAF 将高维的特征空间 I 映射到低维特征空间 S , 同时尽可能保持超图的局

部几何结构不变, 本文中 I 是通过自编码器抽取的特征空间. 对于每个标签 l 以及每个单个的训练样本 $x_i \in I$, 通过计算出 k_s 个与 $x_i \in I$ 具有相同 l 标签值 (0 或 1) 的最近邻的样本集合 knn_s 以及 k_d 个与 $x_i \in I$ 具有不同 l 标签值的最近邻的样本集合 knn_d , 可以构建出局部批 $p_i = x_i \cup knn_s \cup knn_d$. p_i 形成了一条超边, 对应一个代表局部几何结构的子超图, 可以通过式 (10) 来构建局部 Laplacian 矩阵 L_i . 相对应地, 在低维特征空间 S 中, 可以用相同的方式计算出 $x'_i \in S$ 在低维特征空间的局部批 p'_i , 其中 x'_i 是 $x_i \in I$ 对应于低维特征空间的值. 为了维持 p_i 与 p'_i 集合中包含的样本以及样本排序的最大一致性, 局部批优化的目标函数可以表示成:

$$\arg \min_{p'_i} \text{tr}(p'_i L_i (p'_i)^T) \quad (11)$$

全局对齐阶段, PAF 利用样本选择矩阵 P_i , 满足 $p_i = IP_i$, 来建立全局低维特征空间 S 与局部批 p'_i 的映射关系, 即 $p'_i = SP_i$. 将 $p'_i = SP_i$ 代入式 (11) 得到单个标签 l 下的目标优化函数:

$$\arg \min_S \text{tr}(SL^l S^T) \quad (12)$$

其中 L^l 定义为:

$$L^l = \sum_i^n P_i L_i P_i^T \quad (13)$$

为了有效融合多个标签对特征选择的影响, 本文将所有标签对应的 Laplacian 矩阵 L^l 进行加权求和作为全局超图几何结构的表征矩阵, 记为 L_g . 本文假设每个标签的贡献是均等的, 因此 L_g 融合了多个标签对样本间局部几何结构的影响, 如式 (14) 所示.

$$L_g = \sum_{l=1}^{|Y|} L^l \gamma_l \quad (14)$$

s.t. $\gamma_l = \frac{1}{|Y|}$

其中 $|Y|$ 是多标签数据集标签的个数.

最后, 通过 L_g 矩阵的标准特征值分解, 得到 r 个最小非零特征值对应的特征向量 (最小特征值可能为零), 便得到了训练集特征空间约简后的特征空间.

2.3 算法结构

本文新提出的多标签特征提取算法称之为 MLFS-AH. MLFS-AH (Multi-label feature selection with autoencoders and hypergraph learning) 首先采用去噪自编码器对原始多标签数据进行

预处理, 然后再使用基于超图的多标签学习融合多个标签对多标签数据特征提取的影响, 并采用 MLKNN^[27] 对特征提取的结果进行分类检验. MLFS-AH 的具体步骤如下所示:

步骤 1. 构建 m 层的去噪自编码器, 该自编码器接受多标签数 X 为输入, 通过 m 层的自编码处理, 求解一个具有全局最优解的最小二乘优化问题提取出样本空间的鲁棒表达, 从而有效提高多标签数据的抗干扰性. 重构数据的映射矩阵由式 (5) 得出, 重构后的数据记为 X' .

步骤 2. 对于单个标签 $l \in Y$ 以及单个训练样本 $x_i \in X'$, 计算出 k_s 个与 $x_i \in X'$ 具有相同标签的最近邻的样本集合 knn_s 以及 k_d 个与 $x_i \in I$ 具有不同标签的最近邻的样本集合 knn_d . $x_i \cup knn_s \cup knn_d$ 形成了一条超边, 对应的局部批可以由式 (10) 计算得出, 全局对齐则由式 (13) 给出, 记为 L^l .

步骤 3. 重复步骤 2 直到遍历所有的标签.

步骤 4. 对于所有的标签 $\{l \in Y\}$, 融合多个标签对样本间几何结构的影响, 由式 (14) 得到全局超图样本间几何结构的表征矩阵 L_g .

步骤 5. 求解 Laplacian 矩阵 L_g 的标准特征值分解, 得到 r 个最小的非零特征值对应的特征向量集合, 该集合构成的特征空间即是约简后的特征空间, 该特征空间的样本维度是 r .

步骤 6. 使用 MLKNN 多标签分类算法对提取出来的特征空间进行多标签分类, 验证特征提取算法的分类性能.

2.4 算法分析

算法的复杂度包含空间复杂度和时间复杂度. 算法主要分为两个阶段: 去噪自编码和特征约简. 去噪自编码是具有全局最优解的凸优化问题, 所以算法复杂度相对较小, 本节主要讨论特征约简算法的复杂度.

特征约简阶段的空间复杂度主要在于超图的矩阵表达, 包含 D_v , D_e , Ω , H . 其中超图中顶点数目 $|V|$ 等于多标签数据的样本个数, 超边数目 $|E| = |V| \times |Y|$, 其中 $|Y|$ 表示标签的个数. 矩阵表达的空间复杂度如表 2 所示.

表 2 算法空间复杂度
Table 2 Space consumption

矩阵	空间复杂度
D_v	$ V \times V $
D_e	$ E \times E $
H	$ E \times V $
Ω	$ E \times E $
L_g	$ E \times E $

特征约简阶段的时间复杂度主要在 Laplacian 矩阵的特征值分解, 为 $O(|V|^3)$.

3 实验结果

3.1 数据集与度量指标

本文提出的多标签特征提取算法在 5 个公开数据集上测试, 数据集的具体信息如表 3 所示.

表 3 数据集信息
Table 3 Information of data sets

编号	名称	样本数	特征数	标签数
1	Emotions	593	72	6
2	Yeast	2417	103	14
3	Scene	2407	294	6
4	Birds	645	260	19
5	Computer	5000	681	33

实验在如表 3 所示的数据集上做十折交叉验证. 在此过程中, 每个数据集都被分成了 10 个近似大小的子集, 在每次的交叉验证中, 其中的 1 个子集被保留作为测试集, 剩下的作为算法的训练集, 所以这个过程会重复 10 次以提高测试的准确性和结果的可推广性.

检验一个多标签算法的有效性通常比检验单标签算法更加复杂. 给定一个测试集 $T = \{(x_i, l_i) | 1 \leq i \leq n\}$, 可以采用目前流行和有效的多标签评价指标来验证算法的有效性, 本文采用其中 4 个评价指标^[27]: *OneError*, *Coverage*, *RankingLoss* 和 *AveragePrecision*.

OneError 用来衡量所有单个训练样本 $x_i \in X$ 中具有最高 $rank(x_i, y_i)$ 值的标签 $y_i \in Y$ 被分类器预测为不属于该样本的次数. $OneError(Conf, T)$ 值越小, 多标签分类器的分类性能越好.

Coverage 用来计算所有训练样本以及预测出来的标签序列 T , 需要平均移动多少步才可以遍历完所有正确预测的标签. $Coverage(Conf, T)$ 的值越小, 表明多标签分类器的分类性能越好.

RankingLoss 用来度量对于所有的单个训练样本, 预测结果中出现不相关标签比相关标签具有更低 $rank$ 值的次数. $RankingLoss(Conf, T)$ 的值越小, 表明多标签分类器的分类性能越好.

假设 *Precision* 是某个样本 $x_i \in X$ 预测标签中 $rank$ 值低于某个特定被正确预测标签的所有标签的百分比, *AveragePrecision* 则是所有训练样本的 *Precision* 的平均值. $AveragePrecision(Conf, T)$ 值越大, 表明多标签分类器的分类性能越好.

本文所有的实验均在 4GB RAM 以及

2.40 GHz Intel 至强 4 核处理器的主机上完成.

3.2 参数优化

MLFS-AH 中有三个重要的参数 k_s , k_d 与 r . 参数的取值对算法的性能有直接的影响, 所以在对比 MLFS-AH 与其他的算法性能前, 需要对 k_s , k_d 与 r 进行一定的调优. 由于不同数据集特征维度的差异较大, 本文实验中 r 分别取 $r \in [5 : 5 : 350]$, 通过充分的实验表明, r 对于数据集 Emotions, Yeast, Scene, Birds, Computer 分别取 15, 30, 20, 45, 60 时算法 MLFS-AH 的训练结果最好, 其值继续增加不能显著提高算法的分类性能, 故以下实验中对于表 3 中给出的 5 个数据集 r 分别取 15, 30, 20, 45, 60. 对于 k_s , k_d , 在算法训练阶段对参数进行调优过程中, 先固定其中一个参数的值, 然后调整另一个参数. 结果分别如图 1 和图 2 所示.

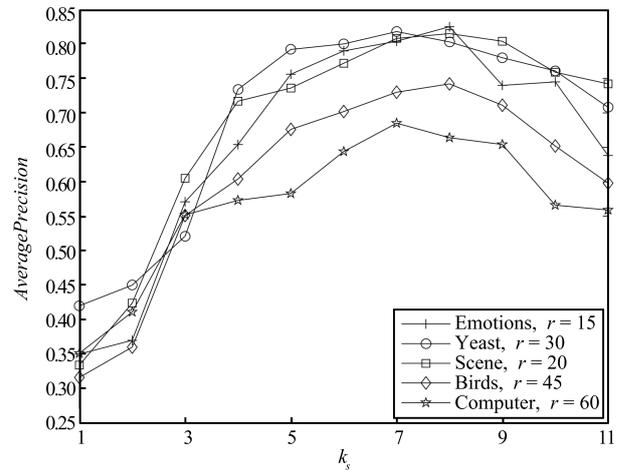


图 1 参数 k_s 对 *AveragePrecision* 的影响 ($k_d = 3$)

Fig. 1 The influences of k_s to *AveragePrecision* ($k_d = 3$)

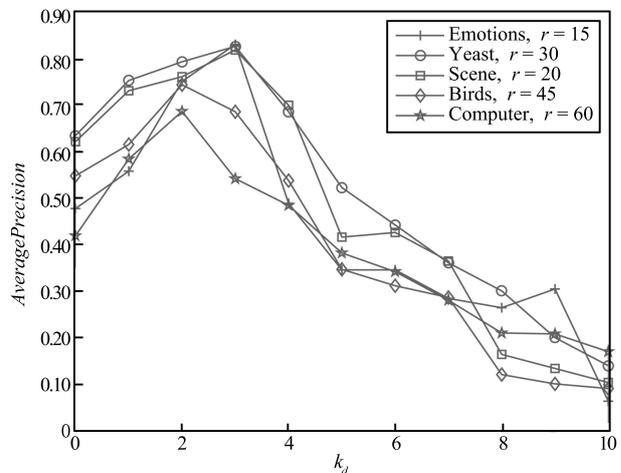


图 2 参数 k_d 对 *AveragePrecision* 的影响 ($k_s = 8$)

Fig. 2 The influences of k_d to *AveragePrecision* ($k_s = 8$)

由图 1 可知, 当 k_s 分别取 8, 7, 8, 8, 7 且 $k_d = 3$ 时, 算法 MLFS-AH 在 *AveragePrecision* 上的性能最优; 由图 2 可知, 当 k_d 分别取 3, 3, 3, 2, 2 且 $k_s = 8$ 时, 算法 MLFS-AH 在 *AveragePrecision* 上的性能最优.

3.3 算法比较

本文将 MLFS-AH 与 ARMLNRS^[13], MLF-SIE^[12], FMLFS^[28] 三个多标签特征提取算法进行性能对比. 为保证可对比性, 本文均采用 MLKNN 作为分类器, 对提取出来的特征进行分类检验. *OE*, *Cov*, *RL* 以及 *AP* 分别代表 *OneError*, *Coverage*, *RankingLoss* 和 *AveragePrecision*. 实验结果如表 4~8 所示, 其中 $params = (k_s, k_d, r)$ 是一个三元组, 且每个评价指标的最优结果用粗体标出. 需要指明的是表中只列出了十折交叉验证的均值, 没有列出十次结果的标准方差; 对于每一个评价指标, \downarrow 表示越小越好, \uparrow 则相反. 下文中对比较算法采用代号, 分别是 a0: MLKNN; a1: ARMLNRS; a2: MLFSIE; a3: FMLFS.

表 4 数据集 Emotions 测试结果 ($params = (8, 3, 15)$)

Table 4 Results on Emotions ($params = (8, 3, 15)$)

指标	a0	a1	a2	a3	MLFS-AH
<i>OE</i> \downarrow	0.290	0.277	0.489	0.265	0.256
<i>Cov</i> \downarrow	1.893	1.842	2.791	1.733	1.756
<i>RL</i> \downarrow	0.173	0.181	0.349	0.168	0.152
<i>AP</i> \uparrow	0.770	0.784	0.658	0.811	0.825

为了更清晰地描述算法性能差异的显著性, 本文定义一个偏序 \succ . 对于每个数据集以及选中某个评价指标, 如果一个算法 $alg1$ 在统计上显著优于算法 $alg2$, 那么可以表示为: $alg1 \succ alg2$. 首先, 本文采用十折交叉验证测试某个算法的性能可获得的一组数值结果, 包含 10 个数值; 然后采用双尾成对 t 检验, 验证对比的两个算法性能是否在给定评价指标上存在显著的差异, 其中显著性水平置为 5%, 即如果统计出两个算法平均性能相等的概率小于 5%, 则可以等价地认为两个算法的平均性能在特定显著性水平下存在显著差异.

对数据集 Emotions 上的实验结果做双尾成对 t 检验, 得到以下性能差异显著性结果: 对于 *OE*, *RL*, *AP*, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, MLFS-AH \succ a3; 对于 *Cov*, 除 a3 \succ MLFS-AH 外, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2.

对数据集 Yeast 上的测试结果做双尾成对 t 检验, 得到以下性能差异显著性结果: 对于 *OE*, *Cov*, *AP*, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, MLFS-AH \succ a3; 对于 *RL*, a3 稍优于 MLFS-AH, 且 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, 如表 5 所示.

表 5 数据集 Yeast 测试结果 ($params = (7, 3, 30)$)

Table 5 Results on Yeast ($params = (7, 3, 30)$)

指标	a0	a1	a2	a3	MLFS-AH
<i>OE</i> \downarrow	0.283	0.274	0.289	0.268	0.243
<i>Cov</i> \downarrow	6.452	6.331	6.538	6.245	6.121
<i>RL</i> \downarrow	0.174	0.168	0.203	0.156	0.160
<i>AP</i> \uparrow	0.760	0.758	0.717	0.782	0.811

对数据集 Scene 上的测试结果做双尾成对 t 检验, 得到以下性能差异显著性结果: 对于 *OE*, *RL*, *AP*, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, MLFS-AH \succ a3; 对于 *Cov*, a3 稍优于 MLFS-AH, 且 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, 如表 6 所示.

表 6 数据集 Scene 测试结果 ($params = (8, 3, 20)$)

Table 6 Results on Scene ($params = (8, 3, 20)$)

指标	a0	a1	a2	a3	MLFS-AH
<i>OE</i> \downarrow	0.275	0.261	0.318	0.260	0.248
<i>Cov</i> \downarrow	0.573	0.536	0.619	0.425	0.429
<i>RL</i> \downarrow	0.163	0.165	0.230	0.166	0.157
<i>AP</i> \uparrow	0.776	0.795	0.723	0.792	0.815

对数据集 Birds 上的测试结果做双尾成对 t 检验, 得到以下性能差异显著性结果: 对于 *OE*, *RL*, *AP*, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, MLFS-AH \succ a3; 对于 *Cov*, 除 a3 稍优于 MLFS-AH 外, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, 如表 7 所示.

对数据集 Computer 上的测试结果做双尾成对 t 检验, 得到以下性能差异显著性结果: 对于 *OE*, *Cov*, *AP*, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, MLFS-AH \succ a3; 对于 *RL*, 除 a3 稍优于 MLFS-AH 外, 均有 MLFS-AH \succ a0, MLFS-AH \succ a1, MLFS-AH \succ a2, 如表 8 所示.

综上, MLFS-AH 的总体性能显著优于 MLF-SIE, ARMLNRS 以及 FMLFS. FMLFS 和 MLFS-AH 都考虑了样本之间几何关系对特征提取性能的影响, 但由于 MLFS-AH 对原始的特征空间做了抗干扰处理, 且对于同一个标签在构建超边时同时考

虑了同类近邻与异类近邻对样本几何关系的影响,并对多个标签的影响进行融合,所以几何关系的表达更加精确,特征提取的性能总体上更优.

表 7 数据集 Birds 测试结果 ($params = (8, 2, 45)$)

Table 7 Results on Birds ($params = (8, 2, 45)$)

指标	a0	a1	a2	a3	MLFS-AH
$OE \downarrow$	0.379	0.371	0.369	0.352	0.344
$Cov \downarrow$	3.411	3.426	3.724	3.385	3.389
$RL \downarrow$	0.129	0.125	0.138	0.124	0.121
$AP \uparrow$	0.712	0.718	0.705	0.727	0.742

表 8 数据集 Computer 测试结果 ($params = (7, 2, 60)$)

Table 8 Results on Computer ($params = (7, 2, 60)$)

指标	a0	a1	a2	a3	MLFS-AH
$OE \downarrow$	0.434	0.438	0.439	0.432	0.425
$Cov \downarrow$	4.435	4.439	4.532	4.378	4.339
$RL \downarrow$	0.108	0.104	0.106	0.089	0.091
$AP \uparrow$	0.641	0.643	0.647	0.649	0.675

4 总结

本文提出了一个基于自编码器与超图学习的多标签数据特征提取算法. 首先该算法采用去噪自编码器提取原特征空间的鲁棒表达,使得特征提取算法抗干扰性更强;然后基于超图理论和 PAF 框架构建每个标签产生的样本之间的几何结构,并融合多个标签对几何结构的影响得到全局 Laplacian 矩阵;最后通过 Laplacian 矩阵的特征值分解得到约简的特征空间. 针对公开数据集的实验结果表明本文的算法优于对比算法,是有效可行的.

References

- Zhang Y, Zhou Z H. Multi-label dimensionality reduction via dependence maximization. In: Proceedings of the 23rd AAAI Conference on Artificial Intelligence. Chicago, USA: AAAI Press, 2008. 1503–1505
- Fu Zhong-Liang. Cost-sensitive ensemble learning algorithm for multi-label classification problems. *Acta Automatica Sinica*, 2014, **40**(6): 1075–1085
(付忠良. 多标签代价敏感分类集成学习算法. 自动化学报, 2014, **40**(6): 1075–1085)
- Zhang Chen-Guang, Zhang Yan, Zhang Xia-Huan. Normalized dependence maximization multi-label semi-supervised learning method. *Acta Automatica Sinica*, 2015, **41**(9): 1577–1588
(张晨光, 张燕, 张夏欢. 最大规范化依赖性多标记半监督学习方法. 自动化学报, 2015, **41**(9): 1577–1588)
- Zhang M L, Zhang K. Multi-label learning by exploiting label dependency. In: Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. Washington, USA: ACM, 2010. 999–1008
- Hariharan B, Zelnik-Manor L, Vishwanathan S V N, Varma M. Large scale max-margin multi-label classification with priors. In: Proceedings of the 27th International Conference on Machine Learning. Haifa, Israel: Omnipress, 2010. 423–430
- Elisseeff A, Weston J. A kernel method for multi-labelled classification. In: Proceedings of the 2001 Advances in Neural Information Processing Systems 14. British Columbia, Canada: MIT Press, 2001. 681–687
- Sun L, Ji S W, Ye J P. Hypergraph spectral learning for multi-label classification. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Las Vegas, USA: ACM, 2008. 668–676
- Zhang M L, Zhou Z H. A review on multi-label learning algorithms. *IEEE Transactions on Knowledge and Data Engineering*, 2014, **26**(8): 1819–1837
- Gibaja E, Ventura S. A tutorial on multi-label learning. *ACM Computing Surveys*, 2015, **47**(3): Article No. 52
- Tian Feng, Shen Xu-Kun. Large scale web image online annotation by learning label set relevance. *Acta Automatica Sinica*, 2014, **40**(8): 1635–1643
(田枫, 沈旭昆. 基于标签集相关性学习的大规模网络图像在线标注. 自动化学报, 2014, **40**(8): 1635–1643)
- Boutell M R, Luo J B, Shen X P, Brown C M. Learning multi-label scene classification. *Pattern Recognition*, 2004, **37**(9): 1757–1771
- Zhang Zhen-Hai, Li Shi-Ning, Li Zhi-Gang, Chen Hao. Multi-label feature selection algorithm based on information entropy. *Journal of Computer Research and Development*, 2013, **50**(6): 1177–1184
(张振海, 李士宁, 李志刚, 陈昊. 一类基于信息熵的多标签特征选择算法. 计算机研究与发展, 2013, **50**(6): 1177–1184)
- Duan Jie, Hu Qing-Hua, Zhang Ling-Jun, Qian Yu-Hua, Li De-Yu. Feature selection for multi-label classification based on neighborhood rough sets. *Journal of Computer Research and Development*, 2015, **52**(1): 56–65
(段洁, 胡清华, 张灵均, 钱宇华, 李德玉. 基于邻域粗糙集的多标记分类特征选择算法. 计算机研究与发展, 2015, **52**(1): 56–65)
- Sun L, Ji S W, Ye J P. *Multi-label Dimensionality Reduction*. Britain: Chapman and Hall/CRC Press, 2013. 34–49
- Yu K, Yu S P, Tresp V. Multi-label informed latent semantic indexing. In: Proceedings of the 28th Annual International ACM SIGIR Conference on Research & Development in Information Retrieval. Salvador, Brazil: ACM, 2005. 258–265
- Tao D C, Li X L, Wu X D, Maybank S J. General tensor discriminant analysis and Gabor features for gait recognition. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2007, **29**(10): 1700–1715

- 17 Tao D C, Li X L, Wu X D, Maybank S J. Geometric mean for subspace selection. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2009, **31**(2): 260–274
- 18 Zhou D Y, Huang J Y, Schölkopf B. Learning with hypergraphs: clustering, classification, and embedding. In: *Proceedings of the 2007 Advances in Neural Information Processing Systems*. Vancouver, Canada: MIT Press, 2007, 1601–1608
- 19 Berge C. *Hypergraphs: Combinatorics of Finite Sets*. Amsterdam: North-Holland, 1989. 83–96
- 20 Gao Y, Chua T S. Hyperspectral image classification by using pixel spatial correlation. In: *Proceedings of the 19th International Conference on Advances in Multimedia Modeling*. Huangshan, China: Springer, 2013. 141–151
- 21 Yu J, Tao D C, Wang M. Adaptive hypergraph learning and its application in image classification. *IEEE Transactions on Image Processing*, 2012, **21**(7): 3262–3272
- 22 Shi J B, Malik J. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 2000, **22**(8): 888–905
- 23 Gao Y, Wang M, Tao D C, Ji R R, Dai Q H. 3-D object retrieval and recognition with hypergraph analysis. *IEEE Transactions on Image Processing*, 2012, **21**(9): 4290–4303
- 24 Hong C Q, Zhu J K. Hypergraph-based multi-example ranking with sparse representation for transductive learning image retrieval. *Neurocomputing*, 2013, **101**: 94–103
- 25 Chen M M, Weinberger K, Sha F, Bengio Y. Marginalized denoising auto-encoders for nonlinear representations. In: *Proceedings of the 31st International Conference on Machine Learning*. Beijing, China, 2014. 1476–1484
- 26 Zhang T H, Tao D C, Li X L, Yang J. Patch alignment for dimensionality reduction. *IEEE Transactions on Knowledge and Data Engineering*, 2009, **21**(9): 1299–1313
- 27 Zhang M L, Zhou Z H. ML-KNN: a lazy learning approach to multi-label learning. *Pattern Recognition*, 2007, **40**(7): 2038–2048
- 28 Lee J, Kim D W. Fast multi-label feature selection based on information-theoretic feature ranking. *Pattern Recognition*, 2015, **48**(9): 2761–2771



唐朝辉 电子科技大学信息与软件工程学院博士研究生. 2009 年获得湖南大学硕士学位. 主要研究方向为机器学习与计算机视觉.

E-mail: chhtang@xmut.edu.cn

(**TANG Chao-Hui** Ph.D. candidate at the School of Information and Software Engineering, University of

Electronic Science and Technology of China. He received his master degree from Hunan University in 2009. His research interest covers machine learning and computer vision.)



朱清新 电子科技大学信息与软件工程学院教授. 1993 年获得渥太华大学博士学位. 主要研究方向为生物信息学, 信息检索. E-mail: qxzhu@uestc.edu.cn

(**ZHU Qing-Xin** Professor at the School of Information and Software Engineering, University of Electronic Science and Technology of China. He received his Ph. D. degree from University of Ottawa in 1993. His research interest covers bioinformatics and information

retrieval.)



洪朝群 厦门理工学院计算机与信息工程学院副教授. 2011 年获得浙江大学博士学位. 主要研究方向为计算机视觉与机器学习.

E-mail: cqhong@xmut.edu.cn

(**HONG Chao-Qun** Associate professor at the School of Computer and Information Engineering, Xiamen University of Technology. He received his Ph. D. degree from Zhejiang University in 2011. His research interest covers

computer vision and machine learning.)



祝峰 闽南师范大学教授. 2006 年获得奥克兰大学博士学位. 主要研究方向为数据挖掘与人工智能. 本文通信作者.

E-mail: williamfengzhu@gmail.com

(**ZHU William** Professor at Minnan Normal University. He received his Ph. D. degree from Oakland University in 2006. His research interest covers

data mining and artificial intelligence. Corresponding author of this paper.)