

# 基于 I2C 距离和标记相关性的多标记场景分类

郝 虹 计 华 张化祥 刘 丽

(山东师范大学信息科学与工程学院 济南 250014)

(山东省分布式计算机软件新技术重点实验室 济南 250014)

**摘 要** 将改进的 ML-I2C 与基于标记相关性的方法结合,提出一种改进的多标记场景分类方法。首先提取所有图像的 SURF 特征,将每个类用一个特征集来表示;然后采用改进的 I2C 方法来计算待测图像与已知类之间的距离,根据距离进行标记排序;最后根据排序,利用标记相关性来预测待测图像的所有可能标记。实验结果表明,该方法对多标记场景分类的准确率较高。

**关键词** 多标记学习,场景分类,I2C 距离,卡方检验

中图法分类号 TP181 文献标识码 A

## Multi-label Scene Classification Based on I2C Distance and Label Dependency

HAO Hong JI Hua ZHANG Hua-xiang LIU Li

(School of Information Science and Engineering, Shandong Normal University, Jinan 250014, China)

(Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Jinan 250014, China)

**Abstract** Combining improved ML-I2C and the correlation between labels, we proposed a modified multi-label scene classification method. First, the SURF feature of all images is extracted, and each class is represented with a feature set. Second, the improved I2C method is adopted to calculate the distance between a query image and each class, getting a label rank based on the distance. Last, label correlation is used for label prediction according to the label rank. Experiment shows that this method achieves a higher accuracy rate on multi-label scene classification.

**Keywords** Multi-label learning, Scene classification, I2C distance, Chi-square test

## 1 引言

多标记学习是机器学习的一个重要分支,在传统的单标记学习中,一个样例只对应一个标记,但是,在真实世界中,往往不止一个标记。比如,一篇关于游泳运动员孙杨的报道文本,它可能属于几个类别:运动类、传记类等;电影《1942》既可以标记为历史片,也可以标记为灾难片;一幅场景图像可以属于海洋同时也属于日落。上述问题仅使用单标记学习方法是难以解决的,此时多标记学习方法就显示出它的优势了。

多标记学习最初是用于解决文档分类时的歧义性问题,在多标记学习框架下,每个对象由一个示例表示而且对应于多个类标记。目前,多标记学习已经广泛应用于多个领域,其中除了最初的文档分类,还包括生物信息领域以及场景分类。

Boutell<sup>[1]</sup>等人首次将多标记学习应用于场景分类,构建了一种用于场景分类的多标记学习框架。此后相继出现了几种不同的方法,具有代表性的方法有 ML-kNN<sup>[2]</sup>、InsDif<sup>[3]</sup>等,也有很多研究人员将这些基本方法进行改进,使其效果更优。近些年,周志华等<sup>[4,5]</sup>提出了一种多示例多标记学习框

架(Multi-instance Multi-label,后文简称为 MIML),并将其应用于场景分类,他们将多标记多示例学习分成多标记学习和多示例学习,相应地提出了分别以这两种学习为桥梁的 MIML-SVM 算法和 MIML-Boost 算法。为了防止分解缺失有用信息,他们又提出了一种 M3MIML<sup>[6]</sup>方法,优化了多标记场景分类。

在多标记学习中,研究人员经常把多标记问题转化为多个独立的二类分类问题,这样就忽略了标记之间的相关性,在真实样例中,考虑到标记相关性对于多标记学习是非常重要的<sup>[7-9]</sup>。

场景分类是图像分类的一部分,图像分类对于预测未知图像类标来说是很重要的,一幅图像一般包含的对象很多,所以可能其标记也会很多,但是二类分类或多类分类只会赋予图像一个标记,而多标记学习正适合这种图像分类。同时,近些年基于 I2C(image-to-class)距离的相关算法在图像分类问题上效果显著<sup>[10,11]</sup>。

Oren Boiman 和 Eli Shechtman<sup>[10]</sup>第一次提出 I2C 距离概念,证明了它在图像分类上优于 I2I(image-to-image)距离。

到稿日期:2013-05-08 返修日期:2013-06-17 本文受国家自然科学基金(61170145),教育部博士点基金(20113704110001),山东省自然科学基金(ZR2010FM021),山东省科技攻关计划(2013GGX10125)及泰山学者项目资助。

郝 虹(1989—),女,硕士生,主要研究方向为机器学习、模式识别,E-mail:haohong925@163.com;计 华(1969—),女,副教授,硕士生导师,主要研究方向为机器学习、模式识别等;张化祥(1966—),男,教授,博士生导师,主要研究方向为机器学习、模式识别及 Web 挖掘等;刘 丽(1979—),女,副教授,硕士生导师,主要研究方向为计算机图形学、图形图像处理和模式识别等。

Zhengxiang Wang 和 Yiqun Hu<sup>[11]</sup> 在原有 I2C 距离基础上加入权重优化了算法,并且将 I2C 距离运用到多标记场景分类和图像标注中,接着,他们改进了自己的方法,加快了寻找最近邻的速度<sup>[12]</sup>。I2C 距离应用比较广泛,除了应用在图像分类、图像标注以外,还可以应用在人类行为识别上<sup>[13]</sup>。

运用 I2C 距离进行图像分类,和其他方法的不同之处主要在于此方法是用一个特征集来表示一个对象(此处和多示例学习相似),而其他大部分方法是用一个向量来表示一个对象。

本文主要运用改进的 I2C 距离,并结合多标记学习中重要的标记相关性,重点对多标记场景进行分类,提出了一种新的改进方法,本文称之为 ML-kI2C(Multi-label knn Image-to-class)方法。整个过程的流程如图 1 所示。

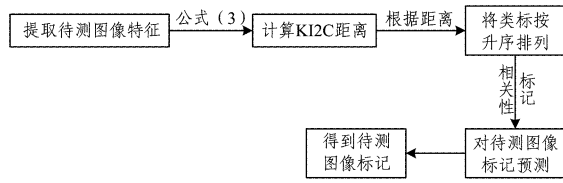


图 1 本文方法对待测图像分类的流程

## 2 改进 I2C 距离

传统的 I2C 距离是计算新图像中的特征与其在某类中的最近邻的距离之和,其值越小证明新图像属于该类的可能性越大。其计算公式如下:

$$Dist(X_i, c) = \sum_{m=1}^{M_i} \|f_m^i - f_n^c\|^2 = \sum_{m=1}^{M_i} d_i^c(m, n) \quad (1)$$

式中,  $X_i$  为待测图像,  $c$  为某一个类,  $M_i$  是图像的特征数目,  $f_m^i$  表示图像的第  $m$  个特征,  $f_n^c$  是  $f_m^i$  在类  $c$  中的最近邻。

文献<sup>[2, 11]</sup>中改进的方法是在传统的 I2C 距离基础上加了权重,该方法最重要的步骤是:建立一个学习框架学习每个类中所有特征(训练集中所有具有该类标的图像的特征)的权重。对于待测图像某一特征与类中最近邻特征的距离是传统 I2C 距离乘以该最近邻特征的权重。具体公式表示为:

$$Dist(X_i, c) = \sum_{m=1}^{M_i} \omega_n^c \|f_m^i - f_n^c\|^2 = \sum_{m=1}^{M_i} \omega_n^c d_i^c(m, n) \quad (2)$$

式中,  $\omega_n^c$  表示类  $c$  中特征  $f_n^c$  的权重。

传统方法是采用待测图像所有特征与其在类中的最近邻特征之间的欧式距离之和来量化 I2C 距离。该方法的弊端在于:当待测图像特征在某类中的最近邻并不是该类的显著特征时,那么计算的距离的准确度就不高,从而影响分类精度。一般地,解决上述问题的办法有两种:加权重和取  $k$  近邻。第二种方法正是对类中所有特征加了权重,从而在计算 I2C 距离时一定程度上减弱了非显著特征的影响。本文采用待测图像所有特征与其在类中的  $k$  近邻的加权欧式距离之和来量化 I2C 距离,其算式如下:

$$Dist(X_i, c) = \sum_{m=1}^{M_i} \sum_{e=1}^k \omega_e^c \|f_m^i - f_e^c\|^2 \quad (3)$$

式中,  $f_e^c$  是  $f_m^i$  对应类  $c$  中的  $k$  近邻,  $\omega_e^c$  是量化  $f_m^i$  与  $f_e^c$  相似性的权值,越相似  $\omega_e^c$  的值越大,当  $f_e^c$  是  $f_m^i$  在类  $c$  中的  $k$  近邻时,  $\omega_e^c = \frac{1}{s_e} \cdot \exp(-\frac{\|f_m^i - f_e^c\|^2}{2\sigma^2})$ , 否则为 0, 其中  $s_e =$

$\sum_{f_e^c \in F_k^c} \exp(-\frac{\|f_m^i - f_e^c\|^2}{2\sigma^2})$ , s. t.  $\sum_e \omega_e^c = 1$ ,  $\sigma$  取特征与其  $k$  个

近邻欧式距离的平均值。

图 2(a)、(b)中左边矩形框表示一幅图像  $X_i$ , 右边圆表示某一类  $c$ , 左边黑色的符号表示提取的图像的特征, 右侧对应的表示图像每个特征在该类中的最近邻。图 2(c)为本文改进的 KNN-I2C, 本图与图 2(a)、(b)的不同之处在于: 图像中每个特征在类中对应的都是一个集合, 该集合是由该特征在类中的  $k$  个近邻组成的(图中对应黑色的符号), 图示只表示了图像中一个特征的情况, 其他特征类似。

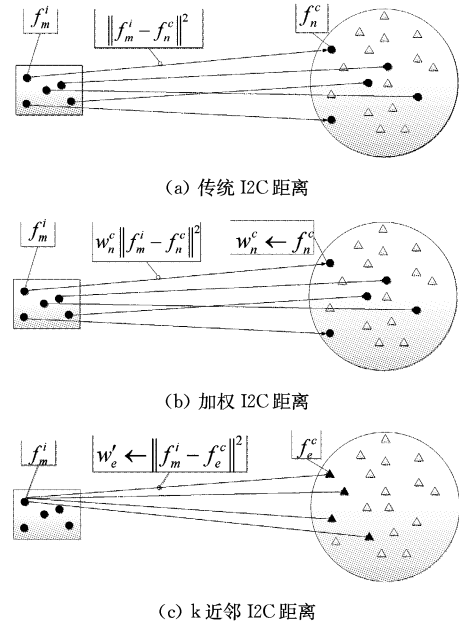


图 2

上文所述的类的特征集由属于该类的图像的所有特征组成, 由于在多标记学习模型中, 每幅图像都可能属于多个类, 因此不同类的特征集之间会存在交集。另一方面, 一幅图像中会存在几个不同的示例, 这些示例的属性都会被提取出来作为该图像的特征, 这样的图像往往就属于多个类, 但是该图像的特征在不同的类中不一定是显著特征。如图 3 所示, (a) 图既有沙漠类的标记也有树类的标记, 即沙漠和树两类的特征集中都包含该图像的特征; 当需要对图 3(b) 进行分类时, 如果运用最近邻 I2C 方法, 就有可能把该图像标记为沙漠和树, 很显然该图像不应该被标记为沙漠类, 本文利用  $k$  近邻 I2C 方法正是为了减小这种错分类的概率。

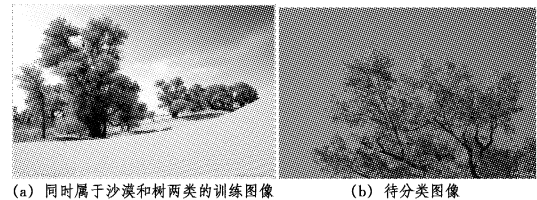


图 3

## 3 标记预测

利用标记之间的相关性进行标记预测是多标记学习中应用比较广泛的方法, 而且效率较高。计算 I2C 距离是为了根据待标记图像到已知类的距离进行排序, 初步确定属于某类的可能性大小。为了得到精确的图像标记, 本文结合 I2C 排序并利用标记之间的相关性对图像标记进行进一步预测。

对于标记相关性的大小, 采用卡方独立性检验来计算, 文

献[14]采用此方法检测标记之间的相关性。具体计算两个标记相关性的方法为:给定两个标记  $c_i$  和  $c_j$ ,统计两个标记同时出现在一幅图像的次数为  $a$ ,不同时出现的次数为  $b$  和  $c$ ,同时不出现的次数为  $d$ ,如表 1 所列。

表 1

	$c_j$	$\neg c_j$	总和
$c_i$	$a$	$b$	$a+b$
$\neg c_i$	$c$	$d$	$c+d$
总和	$a+c$	$b+d$	$a+b+c+d$

两者之间的相关性大小用卡方表示为:

$$M_{ij} = \chi^2 = \frac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)} \quad (4)$$

式中,  $M_{ij}$  的值反映标记相关性的 大小,定义所有的  $M_{ij}$  组成的标记相关性矩阵为  $M$ 。根据相关性矩阵,将排好序的标记分为正标记集和负标记集,正标记集就是待测图像的预测标记。正标记集的确定过程是:首先根据排好的序列,确定序列最顶标记为正标记,其他为负标记(保持原序列顺序不变);然后计算负标记集中第一个标记与正标记集所有标记的相关性大小,当其中存在相关性值大于阈值  $\theta$  时,就将该标记加入到正标记集中,同时将其从负标记集中删掉;重复第二步,直至负标记集中第一个负标记与正标记集中每个标记的相关性值都小于阈值  $\theta$  停止。阈值  $\theta$  根据训练集的相关性值的平均来确定。

#### 4 实验结果及分析

本文采用含有 2000 幅自然场景图像的多标记场景数据集,该数据集是经典的多标记学习的数据集,其共分 5 类: desert, mountains, sea, sunset 和 trees,每个类中对应多个标记的图像占数据集总数的 22%,平均每幅图像具有 1.24 个标记,实验中提取的图像特征是 SURF 特征。实验时把数据集平均分为 10 组,采用十折交叉验证方法进行实验,每次 9 组作为训练集,1 组作为测试集,最后结果取 10 次的平均。

用作对比的实验方法包括 InsDif, ML-kNN, MIML-SVM 和 MIML-Boost,以及基于最近邻的 ML-I2C。算法的评价指标有: Hamming loss, ranking loss, one-error, coverage, average precision 5 项,前 4 个指标值越小最后一个值越大,表示算法越好。实验结果如表 2 所列。

表 2 分类算法及效果比较

算法	评测方法				
	hamming loss	ranking loss	one-error	coverage	average precision
InsDif	0.1521	0.1404	0.2586	0.8342	0.8298
ML-kNN	0.1692	0.1680	0.3001	0.9385	0.8030
MIML-Boost	0.1887	0.1809	0.3514	0.9892	0.7767
MIML-SVM	0.1796	0.1874	0.3272	1.0220	0.7833
ML-I2C	0.1591	0.1558	0.3113	0.8827	0.8044
ML-kl2C	0.1452	0.1280	0.2031	0.8114	0.8324

由表 2 中实验数据可以看出,本文的方法在各个指标上较其他方法都有所提升,尤其在 one-error 指标上,提升效果比较显著。

**结束语** 本文主要运用改进的全新的 I2C 距离计算方法

计算图像到类的距离,同时结合标记相关性,将其应用到多标记场景分类上,使分类结果更精确。另外,该方法也存在一定的问题,如搜索  $k$  近邻的时间复杂度较高,有待改进。

#### 参 考 文 献

- [1] Boutell M R, Luo J, Shen X, et al. Learning multi-label scene classification[J]. Pattern recognition, 2004, 37(9): 1757-1771
- [2] Zhang Min-ling, Zhou Zhi-hua. ML-KNN: A lazy learning approach to multi-label learning[J]. Pattern Recognition, 2007, 40(7): 2038-2048
- [3] Zhang Min-ling, Zhou Zhi-hua. Multi-label learning by instance differentiation[J]. Proceedings of the 22<sup>nd</sup> National Conference on Artificial Intelligence, AAAI, 2007, 22(1): 669-674
- [4] Zhou Zhi-hua, Zhang Min-ling. Multi-instance multi-label learning with application to scene classification [J]. Advances in Neural Information Processing Systems, 2007, 19: 1609
- [5] Zhou Z H, Zhang M L, Huang S J, et al. Multi-instance multi-label learning[J]. Artificial Intelligence, 2012, 176(1): 2291-2320
- [6] Zhang M L, Zhou Z H. M3MIML: A maximum margin method for multi-instance multi-label learning[C]// Eighth IEEE International Conference on Data Mining, 2008, ICDM' 08. IEEE, 2008: 688-697
- [7] Kang F, Jin R, Sukthankar R. Correlated label propagation with application to multi-label learning [C]// 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE, 2006, 2: 1719-1726
- [8] Tsoumakas G, Dimou A, Spyromitros E, et al. Correlation-based pruning of stacked binary relevance models for multi-label learning[C]// Proceeding of ECML/PKDD 2009 Workshop on Learning from Multi-Label Data, Bled, Slovenia. 2009: 101-116
- [9] Wang Hua, Huang Heng, Ding C. Image annotation using multi-label correlated Green's function[C]// 2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 2029-2034
- [10] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2008, CVPR 2008. IEEE, 2008: 1-8
- [11] Wang Zheng-xiang, Hu Yi-qun, Chia Liang-tien. Multi-label learning by Image-to-Class distance for scene classification and image annotation [C]// Proceedings of the ACM International Conference on Image and Video Retrieval. ACM, 2010: 105-112
- [12] Wang Zheng-xiang, Hu Yi-qun, Chia Liang-tien. Improved learning of I2C distance and accelerating the neighborhood search for image classification[J]. Pattern Recognition, 2011, 44(10): 2384-2394
- [13] Wang Zheng-xiang, Hu Yi-qun, Chia Liang-tien. Learning instance-to-class distance for human action recognition[C]// 2009 16th IEEE International Conference on Image Processing (ICIP). IEEE, 2009: 3545-3548
- [14] Tenenboim-Chekina L, Rokach L, Shapira B. Identification of label dependencies for multi-label classification[C]// Proceedings of the second International Workshop on Learning from Multi-Label data. 2010: 53-60