

LDA-CRF: 一种基于概率图模型的目标检测方法

郭乔进 李 宁 杨育彬 武港山

(计算机软件新技术国家重点实验室(南京大学) 南京 210093)

(南京大学计算机科学与技术系 南京 210093)

(ln@nju.edu.cn)

LDA-CRF: Object Detection Based on Graphical Model

Guo Qiaojin, Li Ning, Yang Yubin, and Wu Gangshan

(National Key Laboratory for Novel Software Technology (Nanjing University), Nanjing 210093)

(Department of Computer Science and Technology, Nanjing University, Nanjing 210093)

Abstract Object detection and recognition is actively studied in computer vision and machine learning. Particularly, in recently years, topic models such as latent Dirichlet allocation (LDA) has achieved great success in unsupervised recognition and localization of objects. However, LDA ignores the spatial relationships among image regions. To address this issue, conditional random field (CRF) introduces local dependence to improve the classification accuracy of image patches. In this paper, we propose a latent Dirichlet allocation-conditional random field (LDA-CRF) model by combining LDA with CRF. CRF is trained with topic features generated by LDA, while LDA generates topic information by utilizing structured class labels provided by CRF. Experimental results show that LDA-CRF performs better than CRF in object detection and recognition.

Key words LDA; CRF; object detection; variational inference; graphical model

摘 要 目标检测和识别是计算机视觉和机器学习的研究热点. 近年来, 主题模型(如 LDA 等)在无监督的图像识别和定位等应用中获得了巨大的成功. 然而, LDA 忽略了图像区域之间的空间关系, 并且不能处理连续值的视觉特征或特征向量. 而条件随机场(CRF)能够利用图像区域之间的局部相关性来提高分类准确性. 基于 LDA 和 CRF 提出了一种 LDA-CRF 模型. 通过利用 LDA 生成的主题信息来辅助 CRF 的分类, 同时结合图像区域之间结构化的类别信息来改进 LDA 的主题生成机制. 实验结果表明, LDA-CRF 模型的检测效果要优于 CRF.

关键词 LDA; CRF; 目标检测; 变分推论; 概率图模型

中图法分类号 TP391.41

目标检测(object detection)是计算机视觉、模式识别和人工智能中的一个研究热点, 其目标是对图像中的各个物体进行识别和语义标注, 例如人脸检测、行人检测、医学图像中的肿块检测等. 近年来,

主题模型(topic models), 如 PLSA^[1]和 LDA^[2]在图像的分割和识别等领域中得到了广泛的应用, 并取得了很大的成功. 文献[3]中通过利用 PLSA 来进行图像中目标的检测和定位, 并利用主题信息来进

收稿日期: 2011-05-05; 修回日期: 2011-08-19

基金项目: 国家“九七三”重点基础研究发展计划基金项目(2010CB327903); 国家自然科学基金项目(60875011, 60723003, 61035003, 60975043); 江苏省自然科学基金重点项目(BK2010054)

通信作者: 李 宁(ln@nju.edu.cn)

行无监督的图像分类和聚类. 文献[4]通过结合 LDA 和图像过分割技术来对图像中物体区域进行有监督和无监督的识别与分割. 由于主题信息可以通过无监督的过程来生成, 因此利用主题信息来辅助有监督学习的也可以提供额外的特征信息, 从而提高分类的效果^[5-6]. 文献[5]通过利用 CRF 生成的主题信息作为特征来生成文摘. 针对图像视觉特征的量化造成的影响, 文献[6]中, 我们将 PLSA 生成的主题信息与图像的视觉特征相结合进行分类. 然而使用主题模型通常会有一定的局限性: 一是主题模型忽略了图像中各个区域之间的空间关系, 二是在利用主题模型处理图像数据时, 首先需要对每个区域的视觉特征进行量化, 而量化后的特征对图像区域的描述能力会明显下降^[7]. 针对主题模型不能处理空间关系的问题, 文献[8-9]通过利用 MRF 对 PLSA 和 LDA 的主题信息建立空间关系, 并用于图像的分割和目标的识别, 然而两种方法都是直接对主题信息建立空间关系. 在分类的过程中, 直接使用对应的主题作为类别标签, 并且都只使用了量化后的视觉关键词来进行识别和分类. 为了能够处理图像区域的连续值特征(或特征向量), Blei 等人^[10]提出了基于 GM-Mixture, GM-LDA 的模型, 利用混合高斯模型来处理连续值特征(或特征向量). 然而上述模型并没有同时解决空间关系和特征量化的问题, 因此本文提出了一种 LDA-CRF 模型, 通过结合 LDA 和 CRF 来解决上述问题. 利用 LDA 来生成图像中各个区域的主题信息, 同时利用 CRF 来结合图像区域的视觉特征和主题信息来进行目标检测和识别. 与 LDA 相比, LDA-CRF 模型通过利用 CRF 来处理连续值的视觉特征(或特征向量)和区域之间的空间关系; 与 PLSA-MRF^[8]和 LDMRF^[9]相比, 我们将主题节点和类别节点分开描述, 同时能够处理连续值特征(或特征向量); 文献[5-6]通过直接利用主题信息作为特征向量来提高分类的效果, 而本文中 LDA-CRF 针对如何利用主题信息来辅助分类建立了一个完整的概率模型.

1 LDA 和 CRF 模型简介

1.1 LDA 模型介绍

LDA^[2]是自然语言处理中的一种常用的主题模型, 通过统计文档中的关键词出现频率等信息来得到不同关键词在不同文档中的主题信息 $P(w_n | z_k, d)$. 在利用 LDA 来处理图像数据时, 首先需要

将图像进行分块, 然后对每个分块区域提取特征, 并量化为若干关键词. 这里假设词汇表中共有 N 个关键词 $w_n, n=1, \dots, N, z_k, k=1, \dots, K$ 表示第 k 个主题, K 为主题的数量, $d=1, \dots, D$ 代表 D 个文档. LDA 的概率图结构如图 1 所示, 其中 α 为 K 维向量, $P(\theta | \alpha)$ 满足 Dirichlet 分布, $P(z | \theta)$ 满足多项式分布, $\beta_{kn} = P(w=n | z=k)$. $P(w_n | z_k, d)$ 表示在文档 d 中对应主题 z_k 的关键词为 w_n 的概率, 可以看出 LDA 忽略了文档中各个关键词之间的空间关系, 在同一文档中, 相同的关键词所对应的概率是完全相同的, 与其空间位置没有关系.

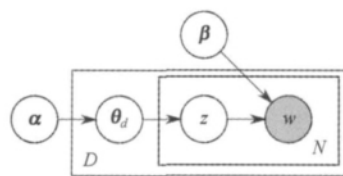


Fig. 1 Graphical model representation of LDA.

图 1 LDA 模型结构

1.2 CRF 模型介绍

CRF^[11]模型同样是最早用于自然语言的标注, 处理图像数据需要用到二维的 CRF(或者叫作 DRF^[12]), 其结构如图 2 所示, 其中 x_i, c_i 分别代表图像中每个区域的视觉特征和类别, 图像的一组标注序列对应的概率为

$$P(c | x) = \exp\left(\sum_{i \in N} uF(x_i, c_i) + \sum_{ij \in E} vF(x_i, x_j, c_i, c_j)\right) / Z(x), \quad (1)$$

$$Z(x) = \sum_c \exp\left(\sum_{i \in N} uF(x_i, c_i) + \sum_{ij \in E} vF(x_i, x_j, c_i, c_j)\right), \quad (2)$$

其中 N 和 E 分别代表 CRF 中的节点(node)和边(edge)的集合, u 为每个节点的视觉特征对应的权重, v 为边界特征对应的权重, $F(x_i, c_i)$ 对应当前节点的特征, $F(x_i, x_j, c_i, c_j)$ 对应当前边界的特征.

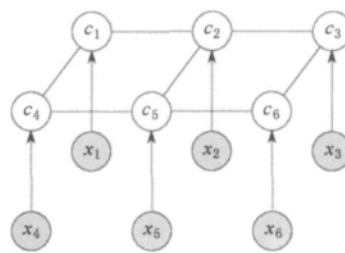


Fig. 2 Graphical model representation of CRF.

图 2 CRF 模型结构

2 LDA-CRF 模型

2.1 LDA-CRF 模型结构

在利用 LDA 模型对图像中的目标进行检测、识别和定位时,忽略了图像中各个区域之间的空间位置信息,同时需要对每个区域的视觉特征进行量化,只能处理离散化的视觉关键词,而 CRF 能够处理连续值的视觉特征(或特征向量),同时可以处理区域之间的空间关系.因此,本文提出了一种结合 LDA 和 CRF 的目标检测模型 LDA-CRF,利用 LDA 生成的主题信息来改进 CRF 的检测效果,同时利用 LDA 生成每个节点的主题信息时,不仅考虑每个节点对应的关键词,还加入了对应的类别信息,其结构如图 3 所示:

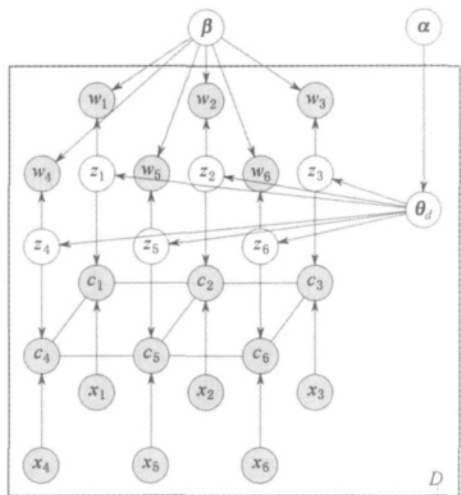


Fig. 3 Graphical model representation of LDA-CRF.

图 3 LDA-CRF 模型结构

这里我们使用 CRF 来处理空间关系和连续值特征(或特征向量),同时每个节点的类别还会受到主题信息的影响,LDA-CRF 的条件概率公式为

$$p(c | z, x, u, v) = \exp \left(\sum_{k=1}^K \sum_i \delta(z_i - k) u_k F(x_i, c_i) + \sum_{ij} v F(x_i, x_j, c_i, c_j) \right) / Z(u, v, x, z), \quad (3)$$

$$\delta(x - a) = \begin{cases} 1, & \text{if } x = a, \\ 0, & \text{o. w.} \end{cases} \quad (4)$$

从图 3 中可以看出,与 LDA 模型相比,在生成每个区域的主题信息时,不仅仅考虑到每个区域对应的视觉关键词,还会利用每个区域的类别标签,而区域之间的类别标签存在着空间关系,从而弥补了 LDA 无法利用空间位置信息的不足;与 CRF 模型

相比,在对每个区域进行分类时,不仅仅考虑到当前节点和邻域节点的视觉特征,还会利用当前区域的主题信息来辅助进行分类.从而通过结合两种模型的优点来改进目标检测的效果.

2.2 参数估计

在训练 LDA-CRF 时,如图 3 所示,每个区域对应的类别 c_i 是已知的,从而我们需要求解的参数包括 $\alpha, \beta, u_k, k=1, \dots, K, v$. 观察变量包括每个节点对应的关键词 w_i 、类别 c_i 以及对应区域的视觉特征 x_i . 为了方便,这里我们定义 $s = \{u_k, k=1, \dots, K, v\}$. 从而对于每幅图像,其似然(Log-Likelihood)为

$$L_{\alpha, \beta, s} = \ln P(w, c, x | \alpha, \beta, s) = \int P(\theta | \alpha) \left\{ \prod_{n=1}^N P(z_n | \theta) P(w_n | z_n, \beta) \right\} \cdot P(c | z, x, s) d\theta. \quad (5)$$

对于整个数据集需要最大化似然:

$$\ln P(D | \alpha, \beta, s) = \sum_{d=1}^D \ln P(w_d, c_d, x_d | \alpha, \beta, s). \quad (6)$$

训练 LDA-CRF 的目标是为了寻找参数 $\{\alpha, \beta, s\} = \arg \max_{(\alpha, \beta, s)} \ln P(D | \alpha, \beta, s)$,为此,我们首先介绍 E-CRF(expanded-CRF)模型,然后利用变分方法(variational methods)^[13]和 E-CRF 来进行求解参数.

2.3 E-CRF 模型

首先我们定义 E-CRF(expanded-CRF)的条件概率分布为

$$P(c | x, \phi, s) = \exp \left(\sum_i \begin{bmatrix} \phi_{i1} \\ \vdots \\ \phi_{ik} \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} F(x_i, c_i) + \sum_{ij} v F(x_i, x_j, c_i, c_j) \right) / Z(x, \phi, s), \quad (7)$$

其中 $\phi_{ik} = P(z_i = k)$ 为节点 i 对应第 k 个主题的概率.可以看出,E-CRF 利用主题的概率对 CRF 的节点特征进行了扩展.

本文中我们通过最大似然来进行训练 E-CRF,给定一组训练样本 $(x^d, c^d), d=1, \dots, D$,我们需要最大化:

$$L_s = \ln \prod_{d=1}^D P(c^d | x^d, \phi^d, s) = \sum_{d=1}^D \left\{ \sum_i \begin{bmatrix} \phi_{i1}^d \\ \vdots \\ \phi_{ik}^d \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} F(x_i^d, c_i^d) + \sum_{ij} v F(x_i^d, x_j^d, c_i^d, c_j^d) - \ln Z(x^d, \phi^d, s) \right\}. \quad (8)$$

我们通过感知器(perceptron)方法,每次考虑一个错误分类样本来计算更新权重向量 s 的梯度方向,文献[14]中指出,利用下面的公式来更新权重向量会遇到过配(overfitting)的问题,因此文献[14]通过引入投票机制对多轮迭代生成的权重的平均来得到最后的权重向量,相比感知器方法,投票感知器(voted perceptron)能够获得更好的训练和分类效果,因此本文中使用了投票感知器来训练 E-CRF.

$$\nabla_{u_k} \approx \sum_i \phi_{ik}^d (F(x_i^d, c_i^d) - F(x_i^d, c_i^{d'})), \quad (9)$$

$$\nabla_v \approx \sum_{ij} (F(x_i^d, x_j^d, c_i^d, c_j^d) - F(x_i^d, x_j^d, c_i^{d'}, c_j^{d'})), \quad (10)$$

$$u_k^{t+1} = u_k^t + \nabla_{u_k}, v^{t+1} = v^t + \nabla_v, \quad (11)$$

$$c^{d'} = \arg\max_c \left\{ \sum_i \begin{bmatrix} \phi_{i1}^d \\ \vdots \\ \phi_{ik}^d \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} F(x_i^d, c_i^d) + \sum_{ij} v F(x_i^d, x_j^d, c_i^d, c_j^d) \right\}. \quad (12)$$

本文中,我们使用 LBP(loopy belief propagation)^[15]来求解 $c^{d'}$. 与 CRF 相比, E-CRF 在计算梯度方向时的耗时要增加 K 倍,然后由于大部分时间都是用来计算 $c^{d'}$,所以 E-CRF 和 CRF 每轮迭代的时间几乎相同,因此,训练 CRF 和训练 E-CRF 具有几乎相同的时间复杂度.

2.4 变分推论(variational inference)

针对 LDA-CRF 模型,由于参数的耦合,直接最大化其似然是非常困难的,因此这里使用变分方法^[2,13]来最大化 $L_{\alpha,\beta,s}$ 的下界(lower bound)(式(14)):

$$q(\theta, z | \gamma, \phi) = q(\theta | \gamma) \prod_{n=1}^N q(z_n | \phi_n). \quad (13)$$

$$L_{\alpha,\beta,s} \geq L(\gamma, \phi; \alpha, \beta, s) =$$

$$E_q[\ln P(\theta, z, w, c, x | \alpha, \beta, s)]. \quad (14)$$

通过引入变分分布(variational distribution) $q(\theta, z | \gamma, \phi)$,从而可以利用变分 EM 来迭代求解 (γ, ϕ) 和 (α, β, s) ,其中 $E_q(f)$ 表示随机变量 f 的期望值, f 满足分布 q . 利用变分分布 $q(\theta, z | \gamma, \phi)$,我们可以通过最大化 $L_{\alpha,\beta,s}$ 的下界 $L(\gamma, \phi; \alpha, \beta, s)$ 来求解参数:

$$L_{\alpha,\beta,s} \geq L(\gamma, \phi; \alpha, \beta, s) =$$

$$E_q[\ln P(\theta, z, w, c, x | \alpha, \beta, s)] =$$

$$E_q[\ln p(\theta | \alpha)] + E_q[\ln p(z | \theta)] +$$

$$E_q[\ln p(w | z, \beta)] + E_q[\ln p(c | z, x, s)] -$$

$$E_q[\ln q(\theta)] - E_q[\ln q(z)]. \quad (15)$$

这里我们将公式右端的每项简记为 L_i , 其中 $L_i, i=1,2,3,5,6$ 的计算过程可以参考文献[2],从而 α, β, γ 的更新公式与文献[2]相同,这里不再赘述. 因此我们主要关注 $L_4 = E_q[\ln p(c | z, x, s)]$,将式(12)代入式(3)可以得到:

$$\begin{aligned} p(c | z, x, u, v) \geq & \exp\left(\sum_k \sum_i \delta(z_i - k) u_k F(x_i, c_i) + \sum_{ij} v F(x_i, x_j, c_i, c_j)\right) / \\ & A \exp\left(\sum_k \sum_i \delta(z_i - k) u_k F(x_i, c_i') + \sum_{ij} v F(x_i, x_j, c_i', c_j')\right), \end{aligned} \quad (16)$$

其中 A 为所有可能的标注序列的个数,针对给定的训练集, A 为常数,将式(16)代入 L_4 ,可以得到:

$$\begin{aligned} L_4 \geq E_q \left[\sum_i \begin{bmatrix} \delta(z_i - 1) \\ \vdots \\ \delta(z_i - K) \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix} F(x_i, c_i) + \sum_{ij} v F(x_i, x_j, c_i, c_j) \right] - \\ E_q \left[\sum_i \begin{bmatrix} \delta(z_i - 1) \\ \vdots \\ \delta(z_i - K) \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_K \end{bmatrix} F(x_i, c_i') + \sum_{ij} v F(x_i, x_j, c_i', c_j') \right] - \ln A. \end{aligned} \quad (17)$$

根据变分分布的定义式(13)以及指数族分布(exponential family)^[16]的性质,有

$$E_{q(z|\phi)} \left[\begin{bmatrix} \delta(z_i - 1) \\ \vdots \\ \delta(z_i - k) \end{bmatrix} \right] = \begin{bmatrix} \phi_{i1} \\ \vdots \\ \phi_{ik} \end{bmatrix}. \quad (18)$$

将式(18)代入式(17),可以得到:

$$L_4 \geq L_{4'} - \ln A. \quad (19)$$

$$\begin{aligned} L_{4'} = & \sum_i \begin{bmatrix} \phi_{i1} \\ \vdots \\ \phi_{ik} \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} F(x_i, y_i) + \\ & \sum_{ij} v F(x_i, x_j, c_i, c_j) - \\ & \sum_i \begin{bmatrix} \phi_{i1} \\ \vdots \\ \phi_{ik} \end{bmatrix}^T \begin{bmatrix} u_1 \\ \vdots \\ u_k \end{bmatrix} F(x_i, c_i') - \\ & \sum_{ij} v F(x_i, x_j, c_i', c_j'). \end{aligned} \quad (20)$$

在求解参数时, α, β, γ 的更新公式可以参考文献[2],这里我们只需要考虑 ϕ, s 的优化过程. 将式(19)(20)代入式(15),对参数 ϕ_{ni} 求偏导,可以得到:

$$L_{\alpha, \beta, s} \geq L(\gamma, \phi; \alpha, \beta, s) \geq L_1 + L_2 + L_3 + L_4' - L_5 - L_6 - \ln A = \hat{L}. \quad (21)$$

将式(21)进行展开, 因为这里需要优化参数, 所以我们只保留其中包含 ϕ_{ni} 的项 $\hat{L}_{[\phi_{ni}]}$, 同时考虑到 $\sum_{i=1}^K \phi_{ni} = 1$, 从而有:

$$\begin{aligned} \phi_{ni} &= \operatorname{argmax}_{\phi_{ni}} \hat{L} = \operatorname{argmax}_{\phi_{ni}} \hat{L}_{[\phi_{ni}]}. \quad (22) \\ \hat{L}_{[\phi_{ni}]} &= \phi_{ni} (\Psi(\gamma_i) - \Psi(\sum_{j=1}^K \gamma_j)) + \phi_{ni} \ln \beta_{i0} + \\ &\quad \phi_{ni} (u_i F(x_n, c_n) - u_i F(x_n, c_{n'})) - \\ &\quad \phi_{ni} \ln \phi_{ni} + \lambda_n (\sum_{j=1}^K \phi_{nj} - 1). \quad (23) \end{aligned}$$

式(23)对参数 ϕ_{ni} 求偏导, 从而 ϕ_{ni} 的更新公式为:

$$\begin{aligned} \phi_{ni} &\propto \beta_{i0} \exp(u_i F(x_n, c_n) - \\ &\quad u_i F(x_n, c_{n'})) \exp(\psi(\gamma_i) - \psi(\sum_{j=1}^K \gamma_j)). \quad (24) \end{aligned}$$

参数 s 的优化目标是:

$$s = \operatorname{argmax}_s \sum_{d=1}^D L_4^d. \quad (25)$$

这里通过利用投票感知器^[14]来求解 s , 根据 L_4' 对参数 u_k, v 求偏导数, 得到权重向量的梯度方向与式(9)(10)相同, 从而参数 s 的更新可以通过训练 E-CRF 来完成.

2.5 LDA-CRF 模型的训练和分类步骤

在训练过程中, 为了加速 LDA-CRF 模型的收敛, 需要为参数 α, β, s 寻找较好的初始值, 本文中我们使用 LDA 模型训练生成的 α, β 作为初始值, 同时利用 LDA 生成的各个区域的主题信息 ϕ 来训练 E-CRF, 得到参数 s 作为 LDA-CRF 的初始参数. LDA-CRF 在训练过程中, 每个区域的类别标签是已知的, 而在分类过程中, 其类别是未知的, 从而 LDA-CRF 模型的训练和分类过程略有不同, 具体步骤如算法 1 和算法 2 所述.

算法 1. LDA-CRF 模型训练步骤.

输入: 训练图像 $x_d^i, c_d^i, w_d^i, d=1, \dots, D, i$ 为图像中的每个区域;

输出: LDA-CRF 参数 α, β, s .

1) 训练 LDA, 生成 $\alpha^0, \beta^0, \phi^0$;

2) 根据 ϕ^0 训练 E-CRF, 生成 s^0 ;

3) LDA-CRF 模型迭代:

① 针对每个图像 d , 根据 $\alpha^t, \beta^t, s^t, c_d$ 生成 $\gamma_d^{t+1}, \phi_d^{t+1}$;

② 根据 $\gamma_d^{t+1}, \phi_d^{t+1}$ 生成 $\alpha^{t+1}, \beta^{t+1}, \phi^{t+1}$;

③ 根据 ϕ_d^{t+1} 训练 E-CRF, 生成 s^{t+1} ;

④ 循环①②③直至收敛.

输出 LDA-CRF 参数 α, β, s .

算法 2. LDA-CRF 模型分类步骤.

输入: 测试图像 d , 参数 α, β, s ;

输出: 检测结果.

1) 初始化 $\phi^0 = \frac{1}{K}$;

2) 利用 ϕ 和 E-CRF 对图像进行分类得到 c^t ;

3) 根据 α, β, s, c^t 生成 $\gamma_d^{t+1}, \phi_d^{t+1}$;

4) 根据 $\gamma_d^{t+1}, \phi_d^{t+1}$ 生成 ϕ^{t+1} ;

5) 重复 2)3)4) 直至收敛.

输出检测结果 $c=c^t$.

3 实验结果和分析

3.1 实验数据

本文中, 我们使用人造 (man-made)^[12,17] 数据集来检验 LDA-CRF 的效果. 该数据集包括 108 幅训练图像和 129 幅测试图像, 每幅图像为 256×384 像素的彩色图像. 根据文献[17], 针对图像中每个不重叠的 16×16 的方块提取特征, 并聚类为 300 个视觉关键词. 由于数据集中包含两个类别: 人造 (建筑物等) 和非人造, 同时, 两个类别存在一定的数据不平衡性, 所以这里我们使用 AUC 来作为评价标准.

3.2 E-CRF 模型实验结果

这里我们通过实验来验证不同主题数量时 E-CRF 的效果, 图 4 和图 5 描述的是 E-CRF 在人造数据集上的分类效果图, 其中横坐标为 E-CRF 迭代的轮数, 纵坐标为 AUC 值. 其中图 4 没有使用任何主题信息来辅助分类, 图 4 中主题数量为 1 时, 对应于 CRF 的 AUC, 从图 4 可以看出, 随着主题数量的增加, E-CRF 的效果会缓慢上升, 当主题数量为 5 时, 其 AUC 值要略高于 CRF. 随着主题数量的上升, E-CRF 的初始 AUC 会不断上升, 但是其随着训练轮数的增加而提升的幅度会越来越小, 当主题数量为 30 时, 随着训练轮数的增加, E-CRF 的 AUC 值并没有上升, 这主要是因为随着主题数量的增加, 特征的维度也会随之增加, 对应的权重数量也会增加, 从而加大了 E-CRF 的训练难度, 导致 AUC 值上升缓慢.

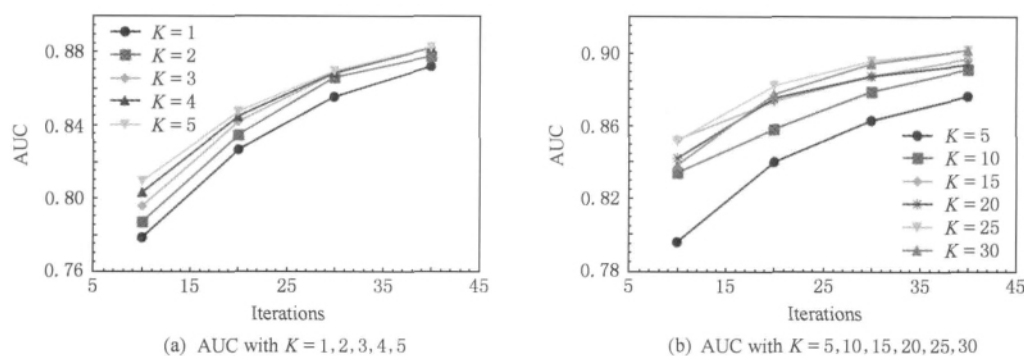


Fig. 4 AUC of E-CRF without topic information.

图 4 不使用主题信息的 E-CRF 模型的 AUC 图

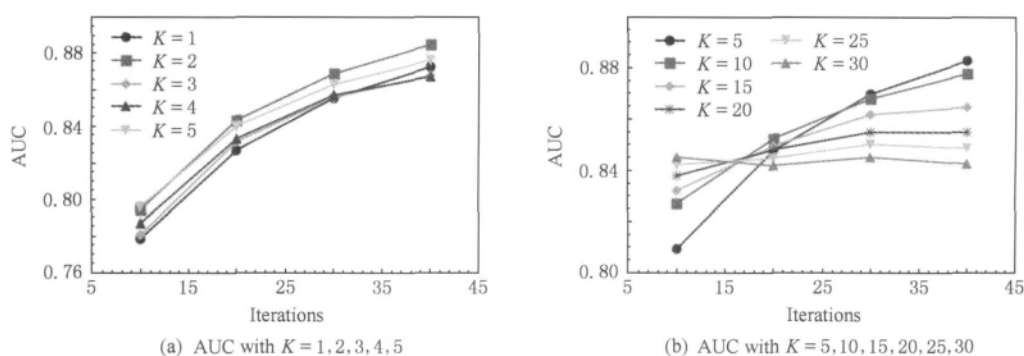


Fig. 5 AUC of E-CRF with topic information.

图 5 使用主题信息的 E-CRF 模型的 AUC 图

图 5 中将 LDA 生成的主题信息代入 E-CRF 进行训练,对比图 4 可以看出,在加入了主题信息时, E-CRF 的 AUC 会得到提升,其中最明显的就是主题数量为 25 和 30 时,在没有加入主题信息时, E-CRF 的 AUC 值增长缓慢,在迭代 40 轮时, AUC 不超过 0.85,而加入了主题信息后, E-CRF 的 AUC 增长更为明显,当迭代 40 轮时,其 AUC 可以达到 0.90. 实验表明,利用 LDA 生成的主题信息可以有效地帮助 E-CRF 的训练.

3.3 LDA-CRF 模型实验结果

参考文献 [17], 我们对 Logistic, CRF 以及

LDA-CRF 3 种算法在人造数据集上的 AUC 进行了实验对比. 在利用 Logistic 进行目标检测时,其 AUC 为 0.822,对比图 6 和图 7,可以看出 CRF 和 LDA-CRF 可以明显地提高检测效果. CRF 相比 Logistic 增加了相邻区域之间的空间关系,使用 CRF 来进行目标检测,其 AUC 最高能达到 0.9 左右,而 LDA-CRF 在 CRF 的基础上利用 LDA 引入图像区域之间的共生关系,从而进一步提高目标检测的效果,其 AUC 最高可以达到 0.96 左右. 这里使用 E-CRF 迭代 10 轮来生成参数 s 的初始值, LDA-CRF 模型中每个 E-CRF 迭代 10 轮. 相比 CRF,

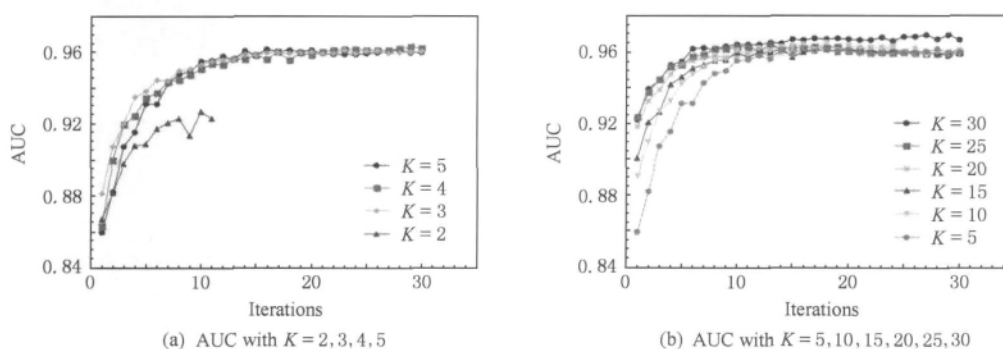


Fig. 6 AUC of LDA-CRF.

图 6 LDA-CRF 模型的 AUC 图

使用 LDA-CRF 可以加速收敛过程. 从图 7 可以看出, CRF 大约需要 100 轮才能达到最佳的 AUC, 而 LDA-CRF 迭代 10 轮达到最佳 AUC 要明显高于 CRF, 考虑到每一轮中的 E-CRF 需要训练 10 轮, 也就是说需要训练 100 轮 E-CRF, 再加上初始化参数 s 的 10 轮 E-CRF, 只需要大约 110 轮 AUC 就可以达到 0.96, 而达到 0.9 所需要的轮数则更少. 当主题数量为 25 时, LDA-CRF 迭代 1 轮 AUC 就能达到 0.92. 随着主题数量的增加, 初始的 AUC 会不断增加, 在 $K=2$ 时, LDA-CRF 迭代 1 轮的 AUC 为 0.866, 当 K 增长到 30 时, LDA-CRF 迭代 1 轮的 AUC 为 0.922. 随着迭代轮数的增加, LDA-CRF 的 AUC 会收敛到 0.96~0.97 之间($K=2$ 除外).

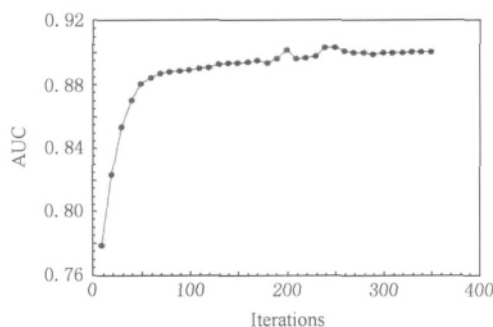


Fig. 7 AUC of CRF.

图 7 CRF 模型的 AUC 图

图 8 对比了 CRF 和 LDA-CRF 迭代 10 轮的平均时间, 其中横坐标表示主题的数量, 主题数量为 1 时对应的是 CRF 迭代 10 轮的时间. 可以看出, LDA-CRF 和 CRF 的时间复杂度相差不多, 并且不会随着主题数量的增加而明显增加.

图 9 为 CRF 和 LDA-CRF 在数据集上的检测

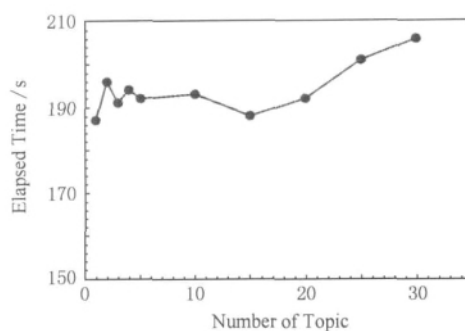


Fig. 8 Elapsed time of LDA-CRF with different number of topics.

图 8 LDA-CRF 不同主题数量的迭代时间

效果图, 这里使用热度图来对比两种模型的检测效果, 其中图 9(a)(d)为 Logistic 的检测效果, 图 9(b)(e)为 CRF 的检测效果, 图 9(c)(f)为 LDA-CRF 的检测效果, 高亮度对应高概率, 低亮度对应低概率, 其中, 主题的数量 $K=30$. Logistic 针对每个区域单独进行分类, 没有考虑区域之间的空间位置信息, 因此其检测结果会出现很多的假阳性区域. CRF 在进行分类时, 会考虑每个区域以及相邻区域的视觉特征, 利用局部的相关性来进行分类; 而 LDA 由于忽略了空间关系, 同一幅图像中具有相同视觉关键词的图像区域是等价的, 会分类为相同的类别; LDA-CRF 则结合了上述特点, 在对每个区域进行分类时, 不仅会考虑到相邻区域之间的相关性, 同时还会考虑到整幅图像中的所有区域. 如果 CRF 会倾向于图像类别标记的局部平滑性, 那么 LDA-CRF 则会在局部平滑性的基础上, 同时要求整幅图像内相似的区域具有相同的类别标签, 通过训练来寻求两者之间的平衡. 从图 9 可以看出, LDA-CRF 模型对于

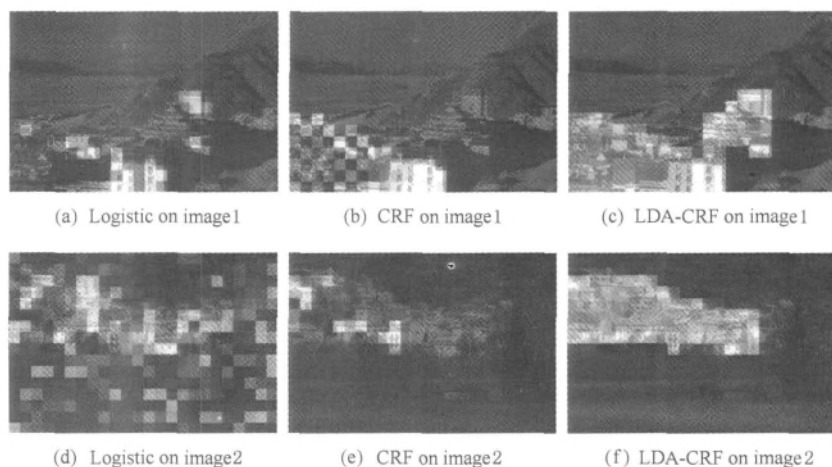


Fig. 9 Detection results of Logistic, CRF and LDA-CRF.

图 9 Logistic, CRF 和 LDA-CRF 检测效果图对比

图像中人造区域的检测效果要优于 Logistic 和 CRF.

4 结束语

本文通过结合 LDA 和 CRF 提出了一种用于目标检测的 LDA-CRF 模型, 通过利用 LDA 生成的主题信息来辅助 CRF 的分类, 同时利用图像区域之间结构化的类别信息来改进 LDA 的主题生成机制, 以及利用 CRF 来处理连续值的视觉特征(或特征向量). 实验结果表明, LDA-CRF 模型的检测效果要优于 CRF. 本文中, 我们只考虑了两类的分类问题, 虽然 LDA-CRF 模型在理论上能够处理多类分类问题, 但是由于在处理图像的多类分类时, CRF 模型的参数数量迅速增加, 从而导致计算量的明显增加. 在下一步的工作中, 我们将进一步地改进 LDA-CRF 模型, 使其能够有效地处理多类分类问题.

参 考 文 献

- [1] Hofmann T. Probabilistic latent semantic indexing [C] // Proc of Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 1999: 50-57
- [2] Blei D M, Ng A Y, Jordan M I. Latent Dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3(4/5): 993-1022
- [3] Sivic J, Russell B C, Efros A A, et al. Discovering objects and their location in images [C] // Proc of IEEE Int Conf on Computer Vision, vol 1. Piscataway, NJ: IEEE, 2005: 370-377
- [4] Li Cao, Li Feifei. Spatially coherent latent topic model for concurrent segmentation and classification of objects and scenes [C] // Proc of Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2007: 1-8
- [5] Wu Xiaofeng, Zong Chengqing. An approach to automatic summarization by integrating latent Dirichlet allocation in conditional random field [J]. Journal of Chinese Information Processing, 2009, 23(6): 39-45 (in Chinese)
(吴晓峰, 宗成庆. 一种基于 LDA 的 CRF 自动文摘方法[J]. 中文信息学报, 2009, 23(6): 39-45)
- [6] Guo Qiaojin, Ding Yi, Li Ning. A context based method for ROI detection in digitized mammograms [J]. Journal of Shandong University: Natural Science, 2010, 45(7): 70-75 (in Chinese)
(郭乔进, 丁轶, 李宁. 一种基于上下文信息的乳腺肿块 ROI 检测方法[J]. 山东大学学报: 理学版, 2010, 45(7): 70-75)
- [7] Boiman O, Shechtman E, Irani M. In defense of nearest-neighbor based image classification [C] // Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2008: 1-8
- [8] Verbeek J, Triggs W. Region classification with Markov field aspect models [C] // Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2007: 1-8
- [9] Mackey L. Latent Dirichlet Markov random fields for semi-supervised image segmentation and object recognition [R]. Berkeley: Computer Science at University of California, Berkeley, 2007
- [10] Blei D M, Jordan M I. Modeling annotated data [C] // Proc of the ACM SIGIR Conf on Research and Development in Informaion Retrieval. New York: ACM, 2003: 127-134
- [11] Lafferty J, McCallum A, Pereira F. Conditional random fields: Probabilistic models for segmenting and labeling sequence data [C] // Proc of Int Conf on Machine Learning. Piscataway, NJ: IEEE, 2001: 282-289
- [12] Kumar S, Hebert M. Discriminative fields for modeling spatial dependencies in natural images [C] // Advances in Neural Information Processing Systems. Cambridge, MA: MIT press, 2004: 1531-1538
- [13] Jordan M I. Learning in Graphical Models [M]. Cambridge, MA: MIT Press, 1998
- [14] Collins M. Discriminative training methods for hidden Markov models: Theory and experiments with perceptron algorithms [C] // Proc of Empirical Methods on Natural Language Processing. San Francisco: Morgan Kaufmann, 2002: 1-8
- [15] Murphy K, Weiss Y, Jordan M I. Loopy belief propagation for approximate inference: An empirical study [C] // Proc of Uncertainty in AI. San Francisco: Morgan Kaufmann, 1999: 467-475
- [16] Wainwright M J, Jordan M I. Graphical Models, Exponential Families, and Variational Inference [M]. Boston: Now Publishers Inc, 2008
- [17] Kumar S, Hebert M. Man-made structure detection in natural images using a causal multiscale random field [C] // Proc of IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2003: 119-126



Guo Qiaojin, born in 1986. PhD candidate of the Department of Computer Science and Technology in Nanjing University. His main research interests include machine learning and image annotation.



Li Ning, born in 1968. Associate professor in Nanjing University. Member of China Computer Federation. Her main research interests focus on machine learning and semantic-based image retrieval.



Yang Yubin, born in 1977. Associate professor in Nanjing University. Member of China Computer Federation. His main research interests include semantic-based image retrieval and large-scale data

mining.



Wu Gangshan, born in 1967. Professor and PhD supervisor of the Department of Computer Science and Technology at Nanjing University. Member of China Computer Federation. His main research

interests include image processing, multimedia information retrieval and digital museum(gswu@nju.edu.cn).

研发动态

科学家组合上千纳米机器可模拟人体肌肉运动

法国斯特拉斯堡大学教授尼古拉斯·朱塞波尼领导的研究小组首次成功合成了一种长长的聚合链,通过超分子键把成千上万的纳米机器结合在一起,每个纳米机器都能产生约 1 nm 的线性伸缩运动.在 pH 值影响下,它们的模拟运动能使整个聚合链产生 10 μm 的收缩或舒张,因此各种运动就被相应地放大了 1 万倍,就像肌肉组织中那样.巴黎狄德罗大学材料与复杂系统实验室用生物计量方法对这种纳米聚合链进行了精确测量.研究人员指出:这一成果将为许多应用领域打开广阔的前景,如微型机器人、纳米信息存储、人工合成肌肉等.这些可组装的纳米机器还能用来设计拥有新奇机械性能的其他材料(摘自: <http://www.sciencenet.cn/>, 2012-10-26, 科学网).

英国宣布成立网络安全研究机构

英国宣布成立一个新的学术研究机构,以提高人们对与日俱增的网络安全威胁的科学认识.该机构由英国外交部情报司(英国政府通讯总部)牵头设立,英国研究理事会全球不确定项目以及英国商业、创新与技能部参与合作,资助款额为 380 万英镑,是英国政府跨部门承诺提高在网络安全领域内学术能力的举措之一.该研究将最终使商业界、个人和政府作出正确的决定,以求更好地实施网络防护措施,从而安全地从网络空间所提供的海量的机会里受益.英国网络安全领域内的学术带头人,包括社会科学家、数学家和电脑科学家将在该机构内协同工作(摘自: <http://www.most.gov.cn/>, 2012-10-16, 科技部网站).

IBM 与 CloudShare 合作推出敏捷开发平台

IBM 和 CloudShare 最近宣布他们的伙伴关系,提供一个基于云的在几分钟内可以设置完成的敏捷开发环境.用户可以在包括 Windows XP, CentOS, Ubuntu, Xubuntu, Windows 7, Windows 8, Windows Server 2003, Windows Server 2008, Windows Server 2012 等 60 多个虚拟机模板中进行自主选择. IBM 和 CloudShare 提供 Rational Team Concert(RTC): 一个团队的完整的开发环境,结合了敏捷方法,在单个的平台上正式和混合规划及报告. CloudShare 预配置的 RTC 4.0 环境提供预先安装 RTC 的虚拟机、客户机与 Eclipse 的集成开发环境、一个 RTC 插件和代码示例. CloudShare 表示这可以让开发者在几分钟内开始编码,也更容易启动新的虚拟机进行测试和评估(摘自: <http://www.zdnet.com/>, 2012-10-25, ZDNet).

Gartner 发布未来 5 年的 10 大关键性技术趋势

在 Gartner 日前举办的年度 IT 研讨会上,分析师 David Cappuccio 总结了一些不容易被 IT 控制而又在推动 IT 支出的技术趋势.云计算、社交媒体网络、移动性和信息管理都正以迅猛的速度发展,这些发展完全不受 IT 部门的控制, Cappuccio 表示:“这些年来,IT 部门被迫支持平板电脑,而最终用户又迫使他们支持 IM 以及无线网络,IT 部门被迫支持的技术正越来越多地开始涌现.”未来 5 年影响 IT 的关键性趋势包括:与消费类应用程序和服务相同水平的 IT 性能和服务;软件定义网络;更大的数据和存储;混合云;客户端服务器;物联网;虚拟数据中心等(摘自: <http://www.cnw.com.cn/>, 2012-10-24, 网界网).

2013 大预测:对顶尖技术进行 beta 测试

根据 Computerworld 年度预测调查显示,在过去几年中,正在进行 beta 测试或者试点计划的受访者比率正在稳步上升,在接受调查的 334 名 IT 高管中,超过 3/4(77%)的受访者表示,他们的团队正在 beta 测试或者试行新技术,这些新技术将于明年推出.而 4 年前,这个数字仅为 43%(摘自: <http://www.cnw.com.cn/>, 2012-09-27, 网界网).