

基于条件随机场的图像语义分割

张春芳

(北京工业大学, 北京 100124)

摘要: 本文研究了一种基于 CRF 的判别模型来学习不同类别标签的后验条件概率分布, 并通过最大化该后验概率来得到最佳标签。CRF 能量函数纹理模块利用形状滤波器来描述图像的底层纹理特征以及基于周围像素纹理特征的上下文信息; 采用 Joint-boosting 算法迭代构造一个强分类器对图像进行初步的标注。在 MRSC 21-object class 数据集上进行实验来评估我们建立的模型的性能。实验结果表明, 该模型基于像素点的分割精度能达到 71.6%, 在识别准确率以及分割效果方面, 具有较强的竞争力。

关键词: 语义分割; 条件随机场; 分段训练

中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1007-9599 (2012) 09-0021-03

一、引言

近些年, 关于对象识别与分割的领域相当活跃。CRF 模型是一种基于无向图的概率模型, 用来对序列数据进行标记, 具有很强的概率推理能力。与经典的马尔可夫随机场模型 (MRF) 相比, CRF 模型主要有两个优点: 第一, 对观测序列没有独立性要求; 第二, 能对长间隔序列的复杂上下文关系进行描述。因此在本文中我们使用 CRF 模型来解决多类对象的识别与分割问题。我们的实验是在 MSRC 21-object class 数据集上完成的。该数据集由分属于 21 类对象 (包括: 建筑物、草地、树、牛、羊、天空、花、标志、鸟、书、椅子、路、猫、狗等) 的 591 张图像组成。我们在数据库随机地选取 45% 的图像作为训练集, 10% 的图像作为验证集, 10% 的图像作为测试集。

二、基于条件随机场的多特征模型

给定一幅图像, 我们使用 CRF 模型来学习类别标签的条件分布。使用 CRF 我们能将形状、纹理、颜色、位置和边缘信息结合在整个的模型里。为了计算方便, 我们对公式两边同时取对数:

$$\log P(c|X, q) = \sum_{(i,j) \in E} y_i(c_i, X; q_y) + \sum_i p(c_i, X; q_p) + l(c_i, i; q_l) + \sum_{(i,j) \in E} f(c_i, c_j, g_y(X); q_f) - \log Z(q, X)$$

其中 y 表示纹理模块, p 表示颜色模块, l 表示位置模块, f 表示边缘模块, q 是对应于各部分的参数集。

(一) 模块描述与参数训练

该模型综合考虑纹理、边缘、颜色、位置四个特征, 采用分段训练^[1]的方法对模型中各模块的参数进行单独的学习, 我们通过为各个模块设置固定的权值来减轻各模块组合构成整个模型时出现的过拟合的问题。在本文中, 我们为位置和颜色模块增加了权值参数, 并且对它们分别进行优化。

1. 纹理模块

在本文中, 纹理特征使用一种新颖的描述方法——形状滤波器来描述像素点的底层纹理特征以及周围像素点的基于纹理上下文信息, 并对基于该特征的多个弱分类器通过 Joint-boosting 算法得到一个强分类器。纹理模块作为单独的一元模块进行处理。纹理模块的特征函数描述如下:

$$y_i(c_i, X; q_y) = \log P_i(c_i | X)$$

其中参数集的定义以及学习过程在第三节将会有详细描述。

2. 边缘模块

在本文中, 我们使用 Potts 模型来对边缘特征建模。边缘模块的函数表达为:

$$\sum_{(i,j) \in E} f(c_i, c_j, g_y(X); q_f) = -q_f \sum_y g_y(X) d(c_i - c_j)$$

其中, $g_y = [\exp(-b \|x_i - x_j\|^2), 1]$ 是边缘特征的特征函数, 用来描述相邻像素间颜色值的差异。

q_f 通过在整数集中人工将两个相关参数进行比对并选择的

方法, 以使在验证集上的错误率尽可能低来获得, 在本实验中, 取 $q_f = [45, 10]^T$ 。

3. 颜色模块

在本文中, 使用混合高斯模型 (HMM) 对颜色特征进行建模, 采用 EM 算法将图像中颜色特征近似的像素点聚类为 k 个高斯部分。像素点 x 的颜色特征的条件概率由下式给出:

$$P(x|c) = \sum_k \hat{a}_k p(k|c) N(x | \bar{x}_k, \Sigma_k)$$

其中, k 是用来表示像素点 x 所属高斯部分的随机变量, \bar{x}_k

与 Σ_k 分别是第 k 个高斯部分的均值与方差。颜色模块的函数表达式为:

$$p(c_i, x; q_p) = \log \sum_k \hat{a}_k q_p(c_i, k) P(k|x_i)$$

其中, 参数 q_p 表示第 k 个高斯部分属于标记为类别 c_i 的概率分布; $P(k|x_i)$ 表示像素点 x_i 属于第 k 个高斯部分的概率分布。

由于我们需要对每张图像的颜色模块的参数分别进行学习, 每张图像对应得到的参数都是互不影响的, 因此颜色模块参数的学习工作不需要在训练集上完成。在测试时, 使用条件迭代模式 (ICM) 对每幅图像的颜色模块参数 q_p 迭代地进行学习。首先给

定出一个初始的类别标签 c_i^* , 然后颜色参数 q_p 使用下式进行更新。参数被重新设定后, 会推断出新的类别标签, 并反复迭代这个过程。

$$q_p(c_i, k) = \left(\frac{\hat{a}_i d(c_i = c_i^*) P(k|x_i) + a_p}{\sum_i \hat{a}_i P(k|x_i) + a_p} \right)^{w_p}$$

需要说明的是, 为了避免过拟合的发生, 我们对颜色模块参数 q_p 和位置模块参数 q_l 进行狄利克雷平滑。在实验中, q_p 经过 2 次迭代得到。其中, 对颜色特征进行高斯混合建模时, 通过比对, 我们设置高斯部分的数目 $K=15$; 设置公式中狄利克雷平滑项 $a_p=0.1$, 权值项 $w_p=3$ 。

4. 位置模块

在本文中, 我们使用位置特征来描述类别标签对像素绝对位置的弱依赖关系, 不同类别的对象在图像中的绝对位置信息用位置特征来表示。位置模块的函数表达式为:

$$l_i(c_i, i; q_l) = \log q_l(c_i, i)$$

其中, i 是像素点在归一化图像中的位置, 由于我们采用相同尺寸大小的图像集来进行实验, 因此, i 即图像中像素点的绝对位置 i 。参数 q_l 表示在归一化位置 i 处的各类别标签的概率分布。

对位置模块参数的学习,我们采用统计的方法来得到:

$$q_{i(c,i)} = \frac{N_{c,i} + a_i}{N_i + a_i}$$

其中, $N_{c,i}$ 为训练集里所有图像中在位置 i 处且类别标签为 c_i 的像素点的数目; N_i 为训练集里所有图像中在位置 i 处的像素点的数目; 同样的, 通过手动比对, 我们设置狄利克雷平滑项 $a_i=1$, 权值项 $w_i=0.1$ 。

(二) 使用 CRF 模型进行推理

使用 CRF 模型对参数集进行学习后, 我们通过最大化条件概率来得到各像素的最佳标签。在我们研究的 CRF 模型中, 初始的类别标签由纹理模块的标注结果来给出, 然后使用图割理论^[10]对模型进行优化, 从而实现准确的分割。图割是一种基于图论的组合优化技术, 能用求在对象识别中求解能量函数最小化的问题。利用图割理论能将图像映射为网络图, 并建立关于类别标签的能量函数, 运用最大流最小割算法对图像进行准确分割。图割能使能量函数的解收敛到全局最小。

CRF 模型的条件后验概率 $P(c|X,q)$ (公式 3-5) 的最大化等价于 CRF 模型能量值的最小化。因此我们将能量函数设置为:

$$E(Y) = -\sum_i \mu_i y_i(c_i, X, q_i) + \sum_{(i,j) \in E} f_{ij}(c_i, c_j, g_{ij}(X), q_i, q_j)$$

其中 μ 为权重参数。对上式中能量函数 $E(Y)$ 的最小化的工作, 可用图割理论中的 alpha-expansion 算法来高效解决。

三、纹理模块的特征描述与分类器构造

在 CRF 能量函数中, 最重要的部分是我们描述纹理特征的模块, 在这个模块中, 不仅包含图像的纹理特征, 还包括基于纹理的上下文信息。

(一) 纹理特征

使用纹理特征来实现图像的语义分割。首先, 用一个 17 维的滤波器组对训练集中的所有图像进行卷积运算来进行滤波, 也就是对图像进行纹理编码。然后对训练集中的所有图像使用马氏距离根据滤波器响应进行 K 均值聚类, 图像中的每个像素都被分配到最近的聚类中心, 对每幅图像中所有像素点标记其相应的聚类中心编号, 得到相应的纹理图像。

像素点所属的类别不只取决于本身的纹理信息, 还取决于周围像素的纹理信息, 即纹理上下文信息。比如, 某个像素周围有类别 “bird” 的纹理特征出现, 那么该像素分类为 “sky” 的概率就比较大。

因此, 我们建立一个既包括第 i 个位置的像素点自身纹理特征, 又包括与周围像素纹理上下文信息的一个特征函数, 定义如下:

$$y_i(c_i, x) = \log P_i(c_i | x)$$

其中,

$$P_i(c_i | x) = \frac{\exp(H(c_i))}{\sum_{c_i} \exp(H(c_i))}$$

其中, 分类器 $H(c_i)$ 由多个弱分类器相加得到的。

(二) 特征提取与分类器构造方法

我们使用 Joint-boosting 算法来实现形状滤波器参数的优化选取以及强分类器 $H(c_i)$ 的构造。该算法迭代地产生多个弱分类器, 并求和得到一个强分类器。

利用形状滤波器来描述像素点自身的底层纹理特征, 以及周

围像素点的纹理特征的上下文信息, 从而得到强分类器 $H(c_i)$ 。每个形状滤波器由一个三元组 $v(i, r, t)$ 来描述, 其中 t 纹理编号, r 为矩形区域, 能用一个四元组 $r(width, height, x, y)$ 来表示这个矩形区域, 其中 $width$ 和 $height$ 为矩形的宽度和长度, x 和 y 为矩形左上角相对于像素点 i 的偏移量, 矩形 r 的四个顶点都是从固定的边界集中随机选取的。对应于像素点 i 的形状滤波器的响应值为:

$$v_{(i,r,t)}(i) = \frac{\sum_{i \in r} [T(i) = t]}{s(r)}$$

其中, $T(i)$ 表示对应于像素点 i 的纹理编号, $s(r)$ 表示矩形区域的面积。

强分类器 $H(c_i)$ 的构造, 是通过 Joint-boosting 算法来得到的。

我们将 M 个带权值的弱分类器求和得到强分类器 $H(c_i)$:

$$H(c_i) = \sum_{m=1}^M h_m(c_i)$$

其中, 每个弱分类器 $h_m(c_i)$ 都是一个决策函数:

$$h_m(c_i) = \begin{cases} ad(v(i, r, t) > q) + b & \text{if } c_i \in N \\ k_{c_i} & \text{if } c_i \notin N \end{cases}$$

其中, $v(i, r, t)$ 为优化选取得到的纹理特征, d 为控制特征值的阈值函数, N 为共享该纹理特征的类别标签的集合。 k_{c_i} 在这里用来减弱各类别在样本中分布不均匀的影响。每次迭代都会产生一个弱分类器 $(a, b, \{k_{c_i}\}, q, N, r, t)$ 。我们能对下面的目标函数进行优化:

$$h_m = \arg \min_{\{a, b, \{k_{c_i}\}, q, N, r, t\}} \sum_{c=1}^C \sum_{i=1}^n w_{i,m-1}^c (l_i^c - h(c, v))$$

其中, n 为训练集中所有像素点的个数, l_i^c 为训练样本中像素点 i 的类别标签, 当属于共享标签集 N 时, 其值为 1; 当不属于共享标签集 N 时, 其值为 -1。 w 权重控制项, 对分错的样本相应设置一个较大的权值, 对分对的样本相应设置一个较小的权值, 从而提高整个分类器的准确率。

通过 Joint-boosting 算法就能迭代得到强分类器 $H(c_i)$, 并优化得到参数集 $q_i = (a, b, \{k_{c_i}\}, q, N, r, t)$ 。

在训练阶段, 逐个像素进行训练需要大量的存储空间和训练时间, 这是不切实际的。因此, 有效的取样和随机特征选择是必不可少的。通过设置 $\Delta \times \Delta$ 网格 (本文中选取 5×5 网格), 计算网格的滤波器响应将会大大减少计算开销。在测试阶段, 形状滤波器能逐个像素进行计算以使每个像素都能实现准确的分类。上述采样过程会产生小程度的平移, 将会导致对象分割时边界的不准确性。但是, 但应用在 CRF 中时, 边缘模块和颜色模块都会准确确定对象的边缘。

即使采用子采样方式, 在 boosting 的每一次迭代中对所有的纹理特征进行穷举搜索都是不可能实现的。但是, 我们的算法只对特征中的一小部分 ($\tau \ll 1$) 进行搜索, 并在每一次迭代中随机选取。设定 $\tau=0.003$, 几千次的迭代后, 基本上能保证对所有的特征都至少测试过一次。

四、实验结果及分析

在采用 Joint-boosting 算法进行实验时, 我们通过最小化在验证集上的误差来选取弱分类器的数目即迭代次数 M 以及纹理特征的数目 N_i , 在实验中, 我们取 $M=5000$, $N_i=400$ 。在纹理模块的 Boosting 分类器得到初步标注的基础上, 我们将纹理、边缘、颜色、位置四个模块合并到 CRF 模型中, 使用训练阶段得到的各模块参数, 在测试集上进行再次标注。

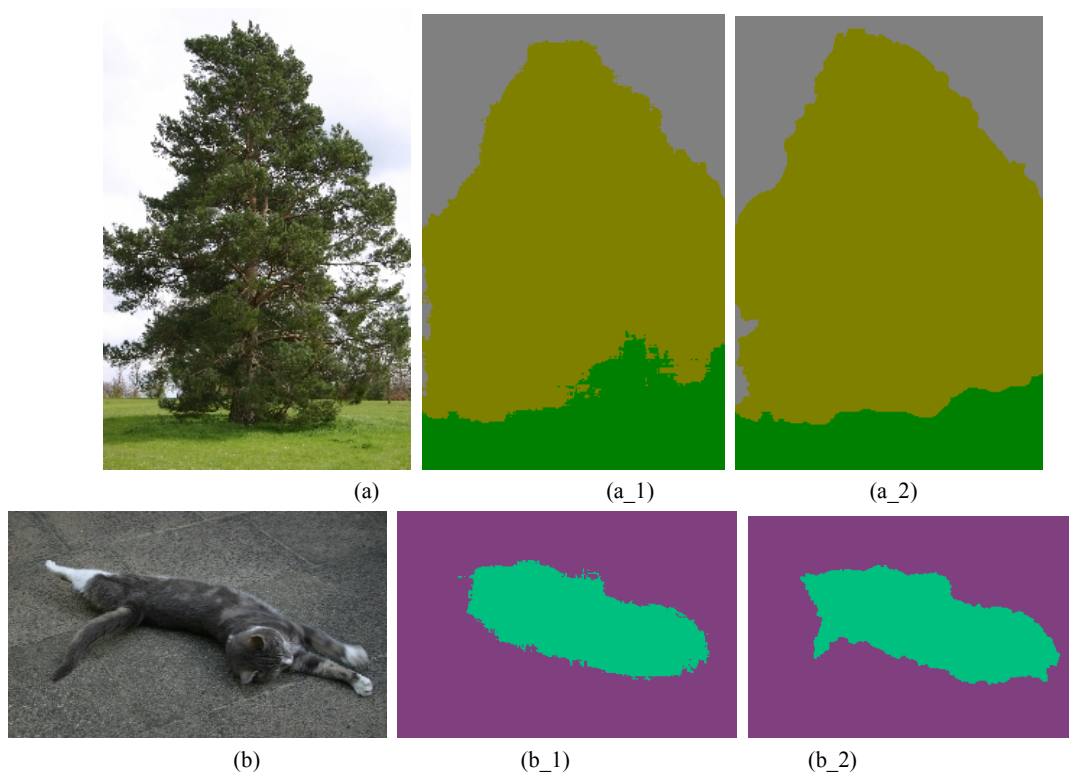


图1 CRF 模型实验结果样图

在图1中,(a)和(b)为原始图像,(a_1)和(b_1)为使用 Boosting 分类器得到的图像,(a_2)和(b_2)为使用整体模型得到的图像。能看出,图像中大部分像素在 Boosting 分类器下都被正确地识别与分割,取得较好的效果。如,图(a_1)天空与树的纹理特征差异较大,树与天空的边缘被准确地分割;树与草地的纹理特征较为接近,但是也得到了很好的分割结果;图(b_1)中,猫的轮廓也较为清晰。与 Boosting 分类器的结果相比,整体 CRF 模型得到的标注图像中,对各对象的分割得到的边缘变得比较平滑,轮廓更加清晰,这是由于加入了边缘特征使得分割更加准确,另外,颜色特征与位置特征的引入也为识别与分割的准确率做出了贡献。

对整个测试集分类得到的标注图像进行分割精度的计算,实验结果显示,基于像素点的识别与分割准确率能达到 71.6%。实验结果表明,21 类中,大部分类别的分割精度都在 60%以上,其中草、树、书、路等类别的分割精度较高,能达到 80%以上,其中以草地的分割精度最高,达到了 92.9%。但是,鸟、狗、船等类别的分割精度比较低,都在 20%以下,其中以船的分割精度最

低,只有 14.2%。

五、结论

本文研究了一种基于 CRF 的判别模型来解决语义分割的问题。模型综合考虑了纹理、边缘、颜色、位置这四个特征,将各个特征变换成 CRF 模型中能量函数的各个势(potential),采用分段训练的方法对模型参数进行学习,使用 α -expansion 算法对模型进行优化,通过最大化每个类别标签的条件概率来得到最佳标签。从而实现对图像的自动分割与标注。实验结果表明,在 MRSC 的 21-object class 数据集上进行实验,我们的模型基于像素点的分割精度能达到 71.6%。在识别准确率以及分割效果方面,我们的模型得到了不错的分割效果。

参考文献:

- [1]R.Fergus,Object class recognition by unsupervised scale-invariant learning.In CVPR,2009,3
- [2]A.C.Berg,Shape matching and object recognition using low distortion correspondences.In CVPR,2005

(上接第 51 页)

由于各业务属性数据具有多时相、多种类和空间分布性等特点,根据数据管理的可行性和空间分析的需要,把各业务属性数据存储在关系数据库中,其中具有空间分布性特点的数据通过 ID 和其空间位置信息相关联。空间位置数据采用 ESRI 的空间数据引擎 (ARCSDE) 进行管理。

三、系统实现

本系统采用 ArcSDE9.3+ORACLE 10G 存储管理道路客运业务数据、基础信息空间和属性数据,并在 MyEclipse 开发中实现基于 Hibernate 的数据库访问。采用面向服务的体系架构 (SOA) 和基于 ARCGIS Server 以及 JAVA 的各项 WEBGIS 平台开发模式,应用 ARCGIS Server 强大的 Web 开发功能实现道路客运管理服务系统的地图操作以及空间分析功能。使用 AJAX 技术解决页面局部刷新问题,大大减少了服务器的负担并提高了客户的响应速度。

四、结语

本系统采用 ARCGIS Server 平台以及 AJAX 技术实现了对道路客运信息的统计查询、空间查询和定位、以及动态监控和统计测算等功能,为道路客运管理部门们对道路客运经济运行信息的动态监控和运力统计测算工作提供技术支撑。本系统还将和道路运输信息服务系统其他业务系统进行集成,进一步实现交通运输信息整合和服务共享。

参考文献:

- [1]胡丽琴,刘明柱.基于 Web 的分布式交通运输空间信息服务系统应用研究[J].北京城市学院学报,2006(4):79-82.
- [2]李季涛,杨俊锋,荣文等.WEBGIS 及其在交通运输中的应用研究[J].现代情报,2004(8):195-197
- [3]关昆,冯涛,王建国.ARCGIS Server 在一体化地籍管理信息系统中的应用[J].城市勘测,2008(4):78-81