

基于迭代神经网络的图像结构表示和分类

屈伸¹, 王庆¹, 池哲儒²

(1. 西北工业大学 计算机学院, 陕西 西安 710072; 2. 香港理工大学 电子与资讯工程学系, 香港)
(ququss@163.com)

摘要: 由于缺少结构化的表示, 基于内容的图像分类存在一定的问题, 据此提出了一种基于迭代神经网络的自然图像表示和分类的方法。利用 Berkeley 分割算法将图像分割成不同的区域, 采用基于人工的多叉树或基于邻接区域的二叉树的方法进行区域合并, 同时提取区域统计特征, 得到图像的树型结构表示。根据 BPTS 算法对网络进行训练, 训练好的网络就具备了图像分类的功能。实验结果表明, 基于迭代神经网络的结构表示和分类方法具有很强的结构学习能力, 同时人工生成的多叉树涵盖更多的语义信息且能得到较好的分类结果。

关键词: BPTS; 图像分割; 区域合并; 迭代神经网络

中图分类号: TP391.41; TP183 **文献标识码:** A

Image structure representation and classification based on recursive neural network

QU Shen¹, WANG Qing¹, CHI Zhe-ru²

(1. School of Computer Science, Northwestern Polytechnical University, Xi'an Shaanxi 710072, China;
2. Department of Electronics and Information Engineering, Hong Kong Polytechnic University, Hong Kong, China)

Abstract: Due to the lack of structural information representation, there still exist key issues in the research domain of content-based image retrieval. In the paper, a recursive neural network-based approach for structure representation and pattern classification was presented. The natural scene images were segmented using Berkeley algorithm and integrated by multi-branch directed trees manually or binary trees automatically. The statistical features were also extracted and got the structure representation. After the training by BPTS algorithm, the structural networks can be employed into image classification. Experimental results show that the recursive network-based structure model has significant ability to represent complicated patterns and the multi-branch tree has more semantic meanings to achieve higher classification rates.

Key words: BPTS; image segmentation; region mergence; recursive neural network

0 引言

近来,在商标注册、数字图书馆等领域,基于内容分类图像的需求迅速增加。一些系统使用关键字来描述图像内容,然而,基于分类学的关键字方法在减少系统复杂度的同时也增加了用户知识的需求。另一方面,研究图像内容可视特征的表示也取得了很大的进展,大多数应用通过较低级的可视特征描述图像,例如局部或全局的高维特征矢量,包括颜色、形状、纹理和空间位置等。但是,如果图像内容很相似甚至它们的特征非常接近,那么在高维特征空间中使用的距离或相似性测度很难区分这些特征点。

通过分析人类的视觉感知机理,我们发现人类是以结构的方式认知现实世界的,因此现实世界中的对象和其相互关系对内容表示都起到了作用。图像中的区域或对象的空间位置及其相互关系包含了很多语义信息,所以它们在描述图像内容的时候都担当了重要的角色。通过把这些有意义的区域组织成树型或图型结构,图像内容就可以清楚细致的表现出来。这个过程对下一步的图像分类和图像检索都是有帮助

的。例如区域间的空间信息在平移和缩放的条件下是不变的,这样的表示方法可以避免系统用高维的视觉特征来描述图像内容。

文献[3~5]等研究了用神经网络表示和处理数据结构的方法,通过结构的反传算法(BPTS)对树型和图型结构进行学习和分类。本文的主要工作是根据图像的分割结果,通过人工和自动合并得到结构表示的多叉树和二叉树,引入基于结构化的图像内容分析方法,用 BPTS 算法训练迭代网络,最终完成图像的分类。

1 图像库和图像分割

随着图像处理和计算机视觉的发展,提出了许多图像分割方法。由于本文的重点是结构信息的表示和处理,所以直接采用了 UC Berkeley 的图像库和人工分割结果。图像库包括训练和测试样本图像库,每幅图像都有对应的人工分割结果,以分割文件的形式表示。图像库中的每幅图像一般对应多个分割文件。每个分割文件有不同的分割块数和分割方法。为简单起见,本文选择了分割块数最少的那个分割文件

收稿日期:2004-09-28;修订日期:2004-12-07

基金项目:国家自然科学基金资助项目(60403008);陕西省自然科学基金资助项目

作者简介:屈伸(1980-),男,陕西西安人,硕士研究生,主要研究方向:模式识别、神经网络;王庆(1969-),陕西人,副教授,主要研究方向:图像处理、模式识别;池哲儒(1961-),福建人,副教授,主要研究方向:模式识别、人工智能。

对应源图像。主要原因是如果分割块数过多,会增加结构表示的复杂度,即对应的生成树结点数和深度过大,影响网络的学习。在图1中,源图像被分割成六个区域,每个区域每点的像素值都被赋予该区域所有像素值的平均。



图1 源图像和区域分割后的图像

2 自动二叉树和人工多叉树结构的建立

在图像分割后,采用自底向上的原则,通过合并相似的邻接领域,形成了结构的二叉树表示。构造的过程如下:

先将相似的邻接区域合并成一个较大的区域。在分割区域集中,每次合并具有最小差异的区域对,生成一个新的区域集。然后继续在这个集合上执行上面的操作,直到这个集合中只有一个区域为止,参见图2(a)。邻接领域 S_i 和 S_j 之间的差异 D_{ij} 定义为:

$$D_{ij} = 0.299 \times |\bar{R}_i - \bar{R}_j| + 0.587 \times |\bar{G}_i - \bar{G}_j| + 0.114 \times |\bar{B}_i - \bar{B}_j| \quad (1)$$

其中, $(\bar{R}, \bar{G}, \bar{B})$ 是一个区域的平均颜色值。从图2(a)可以看出,结点5是右边的小马,生成结点M3代表整个背景。然而,并不是每个生成结点都是有意义的。如结点M1,因为此时的最小差异区域就只有M1和4了。除此之外,为了避免整个树的深度过大,在自动合并区域的时候,如果遇到小区域(区域中包含的像素数小于整个图像的百分之一),直接和邻接区域合并但不生成新结点。

人工生成的过程可以采用多区域合并的方法,即生成多叉树型结构。其中每个生成结点基本上都是有意义的。如M1和M0分别代表左边的马和两匹马。显而易见,人工生成树的显著缺点就是每幅图像的结构特征提取都需要人的参与。

生成树型结构的同时,也产生了BPTS所能处理的特征文件。即图像中的每个结点对应特征文件中每行的参数。众所周知,视觉特征,颜色,纹理和形状对于描述图像内容来说都是非常重要的。在我们的实验中,每个区域结点有六个属性,分别是区域所包含的像素点的个数,区域中 R, G, B 颜色所占的比例,区域的亮度均值和方差。这样就表述了区域的颜色和统计特征。

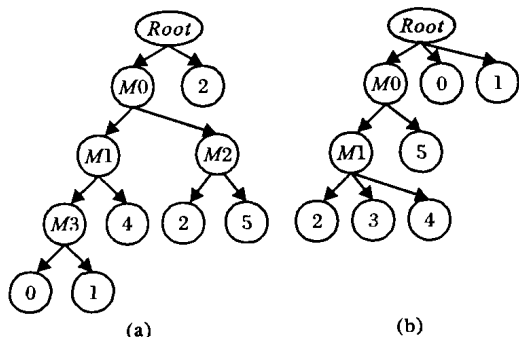


图2 自动生成的二叉树(a)与人工生成的多叉树(b)

3 自适应处理数据结构

一般来说,连接模型已经成功解决了由简单数据类型如静态模型或序列描述的学习任务。然而,现实世界中的大多数结构信息无法用简单的序列来表示。尽管早期用基于语义模式识别的方法来学习结构信息。但由于领域知识不够充分完整,推导一个合适的语法通常是很困难的。相反,在图(树)的表示中,结点数或是输入单元的大小是可以改变的。图3所示的编码过程是与分类任务无关的。近来,文献[4]已提出了用神经网络来处理数据结构,其算法就叫做通过结构的反传(BPTS)。这个算法在序列的情况下拓展到时间的延伸上就是通过时间的反传(BPTT)。文献[5]和文献[8]介绍了自适应处理数据结构的一般框架。根据图3,统一的图形编码公式是:

$$x = F_n(Aq^{-1}y + Bu) \quad (2)$$

$$y = F_p(Cx + Du) \quad (3)$$

这里 x, u 和 y 分别代表 n 个隐层神经元的 n 维输出矢量,输入层神经元的 m 维输入和输出层的 p 维输出。 q^{-1} 表示结点的部分输入来自它的孩子。矩阵 A 定义如下:

$$A = [A^1 \ A^2 \ \dots \ A^c] \quad (4)$$

这里 c 是图的最大出度。 $A^i (i = 1, 2, \dots, c)$ 是 $n \times p$ 的矩阵,由矢量 a_j^i 构成; $j = 1, 2, \dots, n_0$ 。 A 是一个 $c \times (n \times p)$ 矩阵。而 B, C 和 D 的维数分别是 $n \times m, p \times n$ 和 $p \times m$ 。 $F_n(g)$ 这个 n 维矢量定义如下:

$$F_n(\alpha) = [f(\alpha) \ f(\alpha) \ \dots \ f(\alpha)]^T \quad (5)$$

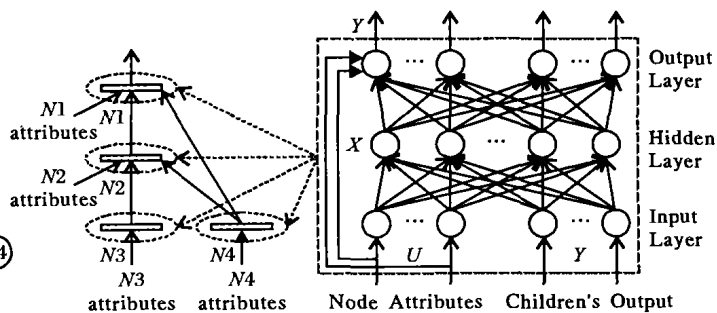


图3 单隐层树型编码网络的框图

这里 $f(g)$ 是一个非线性函数,如Sigmoid函数。这里仅仅假设这个模型只有一个隐层。

训练过程就是通过最小化代价准则从输入样本集中估计参数矩阵 A, B, C 和 D :

$$J = \frac{1}{2} \sum_{i=1}^{N_T} N_T \|d_i - y_i\|^2 \quad (6)$$

这里 y_i 表示第 i 个样本的根结点输出, d_i 表示第 i 个样本的理想输出, N_T 是样本的数量。训练算法的最小化代价准则(6)的微分将遵循梯度学习的模式,通过根据 A, B, C 和 D 计算代价 J 的偏微分。

4 实验结果及分析

以UC Berkeley图像库为实验对象,建立训练和测试集。为了简单起见,将图像库分为两类((a):动物类;(b):风景类,如图4)。从图像库中各选32幅有代表性的图像分别作为两类的训练样本,再各选12幅分别作为测试样本。采用Berkeley分割图像库中的分割结果文件对图像进行分割,分割结果选择了分割块数较少的类别,目的是减少树的结点数

和深度。于是,就可以根据第3节的方法自动生成二叉树或人工生成多叉树的结构表示,用于训练和分类。

实验中,迭代神经网络的参数是(16, C5, S4, O2),即单隐层、6个输入特征、最多5个孩子和2个输出结点。迭代网络处理树型结构的某一结点时,输入层的数据是结点自身的属性和该结点的孩子结点的输出(见公式(2))。当处理叶子结点时,结点的孩子就是迭代的初始状态。然后根据公式3计算该结点的输出。自适应处理整个树型结构(图2)的时候,以自底向上的方式首先处理终端(叶子)结点,在处理完所有叶子结点后,再处理其父结点。重复这个处理过程直至树的根结点。这样根结点的输出就是该样本的输出。迭代网络的目标代价误差设置为0.005,其他为缺省设置。实验结果列于表1。

表1 自动二叉树和人工多叉树训练和识别结果比较

		a	s	正确	错误	拒识	总数	正确率(%)	
自动生成 二叉树	训练集	a	32	0	32	0	0	32	100.00
		s	0	32	32	0	0	32	100.00
	集			64	0	0	64	100.00	
	测试集	a	9	3	9	3	0	12	75.00
		s	4	8	8	4	0	12	66.67
	集			17	7	0	24	70.83	
人工生成 多叉树	训练集	a	32	0	32	0	0	32	100.00
		s	0	32	32	0	0	32	100.00
	集			64	0	0	64	100.00	
	测试集	a	10	2	10	2	0	12	83.33
		s	2	10	10	2	0	12	83.33
	集			20	4	0	24	83.33	



图4 动物类(a)和风景类(b)的样本图像

从上面实验可以看出,这两种方法各有优缺点。基于区域合并自动生成二叉树的方法可以快速的批处理大量的训练样本,但是如前所述,由于二叉树表示相对缺少完整的语义信息,所以导致有些样本的分类结果不是很理想。而人工生成的多叉树包含了较多的语义信息,减少了结构树的深度,降低了收敛时间,从实验结果可以看出有更好的分类效果。但需要人工干预生成的样本树型结果有一些不确定性和人的主观性,在一定程度上会影响分类结果。另外就是人工合并区域的工作量较大。

虽然两种算法各有优缺点,但最终影响分类结果的因素还是多方面的,其中最重要的一点就是图像分割算法的选择。因为不同的分割结果可能导致不同区域块的划分,从而直接影响生成树的结构。本文使用了 Berkeley 分割算法中分割块数最少的那个结果,当然,在人工生成的过程中,也给用户提供了分割块数的选择,每个区域的属性也是提取了一些简单的颜色和统计特征。如果将区域的形状和纹理特征以及空间位置等可视特征考虑进去,可能会提供更多的语义信息,提高分类的准确性。最后,自动生成二叉树的区域差异 D_v 的定义也是可以选择的。将来,找寻更好的分割算法以及提取更有意义的区域特征是要进一步研究的工作。

5 结语

本文提出了一种基于树型结构的自然图像表示和分类的方法,在图像分割的基础之上,用基于人工生成多叉树和基于邻接区域自动生成二叉树的方法得到了图像的结构信息表示,结合区域的统计特征,实现了基于结构和统计信息的图像内容表示和分类算法。实验结果表明,迭代神经网络在树型结构学习上表现出较好的能力,人工生成的多叉树的方法涵

盖了更多的语义信息,分类性能较好。

参考文献:

- [1] WANG Z, CHI Z, FENG D, *et al.* Adaptive processing of tree-structure image representation[A]. IEEE Pacific-Rim Conference on Multimedia(PCM2001)[C]. Beijing, China, 2001. 989-995.
- [2] SALEMBIER P, GARRIDO L. Binary partition tree as an efficient representation for image processing, segmentation, and information retrieval[J]. IEEE Trans. on Image Processing, 2000, 9(4): 561-576.
- [3] GOLLE C, KUHLER A. Learning task-dependent distributed representations by backpropagation through structure[A]. IEEE International Conference on Neural Networks[C]. 1996. 347-352.
- [4] SPERDUTI A, STARITA A. Supervised neural networks for the classification of structures[J]. IEEE Transactions on Neural Networks, 1997, 8(3): 714-735.
- [5] FRASCONI P, GORI M, SPERDUTI A. A general framework for adaptive processing of data structures[J]. IEEE Transactions on Neural Networks, 1998, 9(5): 768-786.
- [6] WANG Z, CHI Z, FENG D, *et al.* Content based image retrieval with relevance feedback using adaptive processing of tree-structure image representation[J]. Int. Journal of Image and Graphics, 2003, 3(1): 119-143.
- [7] MA W, MANJUNATH B. Edgeflow: a technique for boundary detection and image segmentation[J]. IEEE Transactions on Image Processing, 2000, 9(8): 1375-1388.
- [8] TSOI AC. Adaptive processing of data structures: an expository overview and comments[R]. Technique report, Faculty of Informatics, University of Wollongong, Australia, 1998.
- [9] WANG Z, FENG D, CHI Z. Region-based binary tree representation for image classification[A]. IEEE International Conference Neural Networks & Signal Processing[C]. Nanjing, China, 2003.