

基于模糊理论的场景图像情感语义标注方法研究*

曹建芳^{1,2}, 陈俊杰², 李海芳², 赵涓涓²

(1. 忻州师范学院 计算机科学与技术系, 山西 忻州 034000;

2. 太原理工大学 计算机科学与技术学院, 太原 030024)

摘要:电子技术和成像技术的发展导致数字图像迅速增长,依靠先进的技术提取图像蕴含的情感语义并对其自动标注正是当前各行业急需解决的问题。为此提出了一种基于模糊理论的场景图像情感语义标注方法,通过计算模糊隶属度描述图像的情感程度,使用 Adaboost 算法和 RBF 神经网络实现,解决了图像自动标注中的语义模糊问题。使用百度图片频道上下载的 869 张场景图像进行训练和测试,实验通过与人工标注结果相比较,取得了良好的标注效果,可为更多类型的图像情感自动标注打好基础,具有一定的实用价值。

关键词:图像情感语义标注;模糊理论;情感模型;RBF 神经网络;模糊隶属度

中图分类号:TP391

文献标志码 A

文章编号:1672-6693(2014)02-0067-05

随着电子技术和互联网技术的发展,以及近年来数码相机、可拍照移动设备等图像获取设备的迅速普及,使得数字图像的规模急剧膨胀,每天都会有数以 GB 甚至 TB 的数字图像产生、发布及共享。面对庞大的图像数据,用户通过现有的搜索引擎查找自己想要的图像,但结果并不尽如人意。因为图像数据携带丰富的情感语义信息,而目前的图像标注大多基于图像低层视觉特征,无法体现图像本身的语义。图像情感语义标注作为一项新兴的技术逐渐成为人们研究的热点,它能够自动为图像标注情感类形容词,提取图像内涵的语义信息,反映人们对图像的感情,是数字图像理解研究的重要内容,将会对多媒体信息检索、非法信息过滤和拦截等方面的研究提供便利。

1 相关研究

自上个世纪 90 年代末以来,图像情感语义分析的研究在国内外迅速发展起来,研究者们对各种类型图像数据的情感标识展开研究,并取得了一定的成果。例如,毛峡等^[1]使用 2 维波动的数学模型,提出了一种图像波动分析方法,证明了符合“1/f 波动”规律的图像能够给人以和谐和美的感觉。李娉婷等^[2]提出了一种基于颜色特征的家居设计图情感分类方法,通过人们对色彩的感知和理解,建立图像颜色特征与情感语义的关系模型,使用径向基网络分类器实现了对家居设计图像的风格分类。王上飞等^[3]从心理学的“维量”思想出发,采用语义量化技术和因子分析的方法建立情感空间,并分析情感空间的相似性度量方式,抽取图像的颜色和形状特征作为图像的感性特征,采用径向基函数神经网络将图像由特征空间映射到情感空间,在情感空间内实现图像的感性检索。李海芳等^[4]通过研究性格与情感、心情衰减的关系,构建心情和情感衰减函数,得到情感状态波动时各种情感相互影响的数据,建立了一种多层情感模型。Cho 等^[5]使用离散小波变换方法提取图像特征,然后通过基于交互式遗传算法实现了情感图像的检索。黄少荣^[6]提出了一种基于随机参数的粒子群优化算法,冀亚丽等^[7]提出一种基于重量化多特征提取的彩色图像检索方法,Liu 等^[8]提出了一种基于局部特征的语义聚类方法,大大缩小了图像检索范围,提高了图像检索效率。

大部分的情感语义研究都集中在建立情感表述词汇及其与图像低层视觉特征的映射关系上,但对于情感词汇与图像蕴含语义的概率关联程度研究较少。本文从模糊理论的角度出发,结合概率论的方法,提出了一种场景图像的模糊语义特征提取方法,旨在解决人们对图像评价存在的模糊性和主观性问题。

* 收稿日期:2013-01-22 修回日期:2013-03-30 网络出版时间:2014-03-10 19:23

资助项目:国家自然科学基金(No. 61202163);山西省自然科学基金(No. 2013011017-2);山西省高校科技创新项目(No. 2013150);忻州师范学院重点学科专项课题(No. XK201308)

作者简介:曹建芳,女,副教授,博士,研究方向为情感计算、数字图像理解等,E-mail:kcxdj122@126.com;通讯作者:陈俊杰,E-mail:chenjj@xztu.edu.cn

网络出版地址:http://www.cnki.net/kcms/detail/50.1165.N.20140310.1923.015.html

2 基于模糊理论的场景图像情感语义标注

2.1 基本定义

定义 1 模糊集: 设 A 是集合 X 在论域 U 上到 $[0, 1]$ 的一个映射, $A: X \rightarrow [0, 1], x \rightarrow A(x)$, 则称 A 是 X 上的模糊集, $A(x)$ 称为模糊集 A 的隶属函数, 或称 $A(x)$ 为 x 对模糊集 A 的隶属度。

定义 2 情感变量: 用一个 5 维向量表示: $\langle x, E(x), U, G, T \rangle$ 。其中, x 是变量名称, $E(x)$ 是 x 的情感值集合, U 是论域(本文指图像特征提取空间), G 是产生 $E(x)$ 情感值的语法规则, T 是产生情感隶属程度的语义规则。

定义 3 基本情感值集合: 是指在语义上不可再分割的情感值集合。

定义 4 扩展情感值集合: 是对基本情感值程度描述的情感值集合。

2.2 基于模糊理论的情感模型建立

建立情感模型, 确定基本情感值集合 $E(x)$ 是关键。一般分 3 步: 第一, 收集表达情感的形容词, 确定情感值; 第二, 做语义量化实验, 由调查对象对图像进行评价, 建立情感数据库; 第三, 对数据库中的数据分析, 建立情感空间。

2.2.1 建立情感值集合 首先针对场景图像, 精心选择了 7 个形容词: $E(x) = \{\text{清新自然的, 浪漫温馨的, 柔和幽雅的, 轻松愉快的, 充满生机的, 清爽开阔的, 富于变幻的}\}$, 进而又构建了扩展情感值集合 $\{\text{非常, 中性, 几乎不}\}$ 。整个情感变量值集合如表 1 所示。

表 1 场景图像的情感变量集合

基本情感值		扩展情感值	
清新自然	非常清新自然	清新自然	几乎不清新自然
浪漫温馨	非常浪漫温馨	浪漫温馨	几乎不浪漫温馨
柔和幽雅	非常柔和幽雅	柔和幽雅	几乎不柔和幽雅
轻松愉快	非常轻松愉快	轻松愉快	几乎不轻松愉快
充满生机	非常充满生机	充满生机	几乎不充满生机
清爽开阔	非常清爽开阔	清爽开阔	几乎不清爽开阔
富于变幻	非常富于变幻	富于变幻	几乎不富于变幻

2.2.2 语义量化和情感空间建立 从百度图片频道下载了 869 张场景图像, 从中选取了 100 张有代表性的场景图像作为样本提供给用户测评, 这些图像具有不同的颜色构成、空间布局和内容。用户是 60 名大学一年级的学生, 年龄在 18~20 周岁之间, 25 名女生, 35 名男生, 这样形成了情感数据库。

1) 图像特征提取。常用的图像特征有: 颜色、纹理和形状等, 但对于场景图像来说, 很不规则, 纹理和形状特征的提取比较困难, 而颜色是反映场景图像语义最为关键的特征, 直接作用于视觉, 给人以不同的感觉, 因此, 本文主要提取图像全局颜色特征作为反映语义信息的视觉特征。由于 HSV (Hue, saturation, and value) 颜色空间能更好地

反映人类对色彩的认知, 所以选择 HSV 为工作空间。相对于颜色的饱和度 (Saturation) 和色调 (Value) 来说, 人类的视觉系统对色度 (Hue) 更敏感, 因此将 HSV 空间量化为 10 种色度、2 种饱和度和 3 种色调, 这样形成了一个 60 维的颜色特征向量, 构成了情感变量中的论域 U 。

2) 扩展情感值的确定。构建了情感变量的语法规则 G :

$\langle \text{情感表达式} \rangle ::= \langle \text{扩展情感值} \rangle | \langle \text{基本情感值} \rangle$

$\langle \text{扩展情感值} \rangle ::= \langle \text{隶属变量} \rangle | \langle \text{基本情感值} \rangle$

$\langle \text{隶属变量} \rangle ::= \text{非常} | \text{中性} | \text{几乎不}$

$\langle \text{基本情感值} \rangle ::= \text{清新自然} | \text{浪漫温馨} | \text{柔和幽雅} | \text{轻松愉快} | \text{充满生机} | \text{清爽开阔} | \text{富于变幻}$

根据实验, 3 个扩展情感值“非常 (very)、中性 (neutrally)、几乎不 (hardly)”被量化为

$$\begin{aligned} V_e(x) &= \{T_e^2(x) | x \in U\} \\ N_e(x) &= \{\sin(T_e(x) * \pi) | x \in U\} \\ H_e(x) &= \{1 - T_e(x) | x \in U\} \end{aligned} \quad (1)$$

其中, e 是基本情感值, $T_e(x)$ 是 RBFNN 训练得到的模糊隶属度。

3) 情感空间的建立。对于一个样本集 $\{(V_1, y_1), (V_2, y_2), \dots, (V_n, y_n)\}$, 其中, V_i 是上述提取的 60 维的颜色特征向量, $V_i \in U (i=1, 2, \dots, 60)$, $y_i (i=1, 2, \dots, n)$ 是基本情感值隶属于扩展情感值的隶属度, 需要构造一个映射 $T_e: V \rightarrow y, e \in E(x)$, 这个映射就是 5 维情感变量中的语义规则 T 。

鉴于径向基神经网络 (Radial basis function neural network, RBFNN)^[9] 结构简单、训练速度快, 具有较强的学习能力, 而且便于处理模糊集; 而 Adaboost 算法的思想是合并多个弱预测器的输出以产生有效预测, 本文将 Adaboost 算法与 RBF 神经网络相结合, 将 RBF 神经网络作为弱预测器, 反复训练 10 个 RBF 神经网络预测样

本输出,通过 Adaboost 算法得到 10 个 RBF 神经网络弱预测器组成的强预测器,使用该网络构造情感空间,它的算法流程和单个 RBFNN 学习过程分别如图 1 和封二彩图 2 所示。

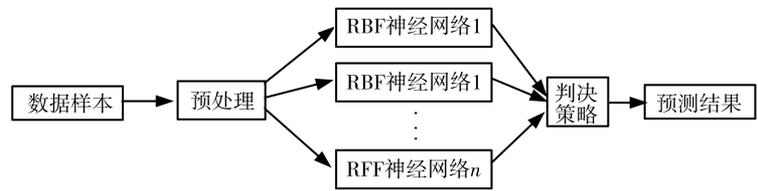


图 1 算法流程

RBFNN 是一个标准的包含输入层、隐含层和输出层的 3 层前馈神经网络,提取的 60 维的低层视觉颜色特征向量作为网络的输入,实验选取隐含层单元数目是 20,使用 Gaussian 函数作为隐含层的激活函数:

$$\Phi(x) = e^{-\frac{(x-\mu)^2}{\sigma^2}} \quad (2)$$

使用 Adaboost 算法构成强预测器,算法步骤:

1) 数据选择和网络初始化。从样本空间中随机选择 m 组训练数据,初始化测试数据的分布权值 $D_t(i) = \frac{1}{m}$,根据样本输入输出维数确定神经网络结构,初始化 RBF 神经网络的权值和阈值;

2) 弱预测器预测。训练第 t 个弱预测器时,用训练数据训练 RBF 神经网络并预测训练数据输出,得到预测序列 $g(t)$ 的预测误差和 e_t ,它的计算公式为

$$e_t = \sum_i D_t(i), i = 1, 2, \dots, m (g(t) \neq y)$$

其中, $g(t)$ 为预测结果, y 为期望结果;

3) 计算预测序列权重。根据预测序列 $g(t)$ 的预测误差 e_t ,计算序列的权重 a_t ,计算公式为

$$a_t = \frac{1}{2} \ln\left(\frac{1-e_t}{e_t}\right)$$

4) 测试数据权重调整。根据预测序列权重 a_t ,调整下一轮训练样本的权重,调整公式为

$$D_{t+1}(i) = \frac{D_t(i)}{B_t} \times \exp[-a_t y_i g_t(x_i)], i = 1, 2, \dots, m$$

其中, B_t 是归一化因子,目的是在权重比例不变的情况下使分布权值和为 1;

5) 强预测函数。将训练 T 轮后得到 T 组弱预测器的权重 a_t 归一化: $a_t = \frac{a_t}{\sum_{t=1}^T a_t}$,预测函数预测结果 $y(x)$ 为

$$y(x) = a_t h(x)$$

其中, $h(x)$ 为 T 个弱预测器得到的预测样本的预测值。

经过训练学习,输出是 7 个情感变量值的模糊隶属度。例如,对于基本情感值“清新自然的”,训练输出得到的隶属度为 0.85,这就表明此输入图像给人清新自然感觉的程度是 0.85。

这样,图像的特征就被表示成为一个 7 维向量: $F = \{y_1, y_2, \dots, y_7\}$ 。其中, $y_i, i = 1, 2, \dots, 7$ 是基本情感变量值集合的模糊隶属度,文中相应的在 F 空间上的场景图像的基本情感值集合是{清新自然(qz),浪漫温馨(lw),柔和幽雅(ry),轻松愉快(qy),充满生机(sj),清爽开阔(qk),富于变幻(bh)}。

例如,如果通过训练得到了一幅场景图像的语义特征向量为 $F = \{0.93, 0.31, 0.45, 0.89, 0.81, 0.78, 0.26\}$,则根据公式(1),它的扩展情感值如表 2 所示。

表 2 图像语义特征扩展情感值

基本情感值	语义特征值	扩展情感值隶属度		
		非常 (Ve)	中性 (Ne)	几乎不 (He)
清新自然	0.93	0.865	0.220	0.07
浪漫温馨	0.31	0.096	0.827	0.69
柔和幽雅	0.45	0.203	0.988	0.55
轻松愉快	0.89	0.792	0.340	0.11
充满生机	0.81	0.656	0.563	0.19
清爽开阔	0.78	0.608	0.638	0.22
富于变幻	0.28	0.078	0.770	0.72

2.3 场景图像情感语义标注

图像情感语义标注就是使用大量的实例图像数据,通过某种机器学习方法学习建立语义概念模型,实现图像自动语义标注。本文使用带有模糊隶属度的基本情感形容词标注场景图像,即计算每个基本情感值对应于 3 种扩展情感值的隶属度,从中选择隶属度最大的两个扩展情感值对应的基本情感形容词加模糊隶属度对图像进行标注。为使得标注结果更易于理解,将扩展情感值“几乎不”所对应的 7 个基本情感值标注为与其意义相反的形容词,即{模糊不清,呆板清凉、不柔和幽雅、死气沉沉、荒凉的、视野狭窄、单调乏味}。对于表 2 中的例子,这幅图像的标注结果为:柔和幽雅(0.988),非常清新自然(0.865)。

3 实验步骤及结果分析

3.1 实验步骤

首先对收集的 869 张场景图像进行预处理,去除了风格和语义模糊不清的 89 张图像,将剩余的 780 张图像随机分成 4 组,每组 195 张;将 60 名参与评价的用户也随机分成 4 组,每组 15 人,分别对上述 4 组图像做人工标注。

以 Matlab 为平台,开发了一个图像情感语义自动标注系统,分训练和测试两个阶段。在训练阶段,从 780 张图像中选择 550 张风格不同的场景图像作为训练集;在测试阶段,将剩余的 230 张图像作为测试集,计算它们的模糊隶属度和语义标注情感形容词。

3.2 实验结果及分析

本文对 230 张场景图像进行了测试,封三彩图 3 是部分图像自动情感语义标注的结果。

另外,为验证系统性能,将 230 张测试图像的标注结果与人工标注结果进行了比较,以人工标注结果为基准,本文让测评用户做了 2 种人工标注结果:1)仅使用基本情感值标注图像;2)使用扩展情感值+基本情感值标注图像。对一张图像,选择用户对其评价最多的两个形容词作为人工标注结果,人工标注结果不含隶属度。因此,只要自动标注结果的基本情感值与人工标注结果相同,就认为是进行了正确的自动标注。首先,如果仅使用 RBF 神经网络建立的模型进行测试,以基本情感值人工标注为基准,通过比较,197 张图像的自动标注结果含有人工标注结果中的基本情感值,标注的准确率达到 $197/230=85.7\%$;然后,以扩展情感值+基本情感值人工标注为基准,169 张图像的自动标注结果与人工标注结果一致,标注的准确率达到 $169/230=73.5\%$ 。但如果结合 Adaboost 算法构造强预测器,标注准确率可分别达到 91.2% 和 82.4% 。为验证系统标注效率,表 3 和表 4 分别列出了自动标注结果在基本情感值和扩展情感值下相对于所有用户人工标注的准确率的最高值、最低值和平均值。

使用两种不同的模型,表 3 是不含模糊隶属度的基本情感值的标注结果,系统的标注准确率较高,表 4 列出的是在加上扩展情感值后系统的标注结果,相对来说标注准确率较低。可以得出结论:不同的用户对于同一张图像的情感理解在总体方向上是一致的,但理解程度深浅差别比较大;将 Adaboost 算法与 RBF 神经网络方法相结合,系统标注准确率有明显提高,体现了 Adaboost 算法的优势;另外,对于情感形容词“浪漫温馨”、“富于变幻”,标注准确率较低,说明与它们相关的视觉元素在情感理解上是比较模糊的。

表 3 7 种基本情感值标识准确率

用户标注	模型	清新自然	浪漫温馨	柔和幽雅	轻松愉快	充满生机	清爽开阔	富于变幻
最高值	RBF	93.2	85.7	84.1	85.7	90.6	91.4	83.5
	Adaboost-RBF	95.4	88.5	87.8	89.7	93.6	93.9	85.1
最低值	RBF	87.1	77.6	80.2	78.2	86.1	85.9	76.9
	Adaboost-RBF	89.8	80.1	83.7	82.3	89.4	89.3	78.2
平均值	RBF	91.4	82.0	81.7	83.8	88.4	87.8	81.9
	Adaboost-RBF	93.2	84.6	84.0	86.9	90.5	91.7	83.6

表 4 7 种基本情感值在扩展情感值“非常”下的标识准确率

用户标注	模型	清新自然	浪漫温馨	柔和幽雅	轻松愉快	充满生机	清爽开阔	富于变幻
最高值	RBF	83.7	73.6	78.0	82.7	80.4	84.4	71.8
	Adaboost-RBF	90.3	80.1	83.4	85.9	87.4	89.2	76.9
最低值	RBF	76.1	68.8	70.2	74.2	71.9	74.9	64.3
	Adaboost-RBF	81.8	72.1	76.1	80.1	78.5	80.3	72.0
平均值	RBF	79.2	70.6	73.4	78.4	75.8	78.1	68.3
	Adaboost-RBF	85.2	75.5	79.2	82.5	83.1	84.8	74.1

4 结束语

如何让计算机能像人类一样具有识别和表达情感的能力,实现人机和谐交互是当前人工智能所面临的重要挑战。本文应用模糊理论的原理,探索场景图像与人类主观理解图像语义之间的关系,并使用基于 Adaboost 算法和 RBF 神经网络相结合的方法建模,实现场景图像的情感自动标注。实验证明将模糊理论与 RBF 神经网络、Adaboost 算法相结合,在解决主观性和不确定性较强的情感理解方面的问题具有较大的优势,实验取得了较好的效果。本文的不足之处及下一步研究方向有:

实验证明将模糊理论与 RBF 神经网络、Adaboost 算法相结合,在解决主观性和不确定性较强的情感理解方面的问题具有较大的优势,实验取得了较好的效果。本文的不足之处及下一步研究方向有:

1)参加测评的用户对象类型单一,均为一年级大学生,这势必给研究结果带来一定的误差。进一步扩大测评用户的选择范围,会使得研究结果更加真实和客观。

2)神经网络模型需要大量的训练集,训练数据越多,测试结果就越准确,这会耗费大量的时间和人力资源。如何在小样本集上获得较高的标注准确率,这是下一步研究的重点内容。

3)场景图像蕴含的语义内容特别丰富,如何合理地规范图像情感值的类别,这也是有待进一步研究的内容。

参考文献:

- [1] 毛峡,丁玉宽,牟田一弥. 图像的情感特征分析及其和谐感评价[J]. 电子学报,2001,29(12A):1923-1927.
Mao X, Ding Y K, Mu T Y M. Analysis of affective characteristics and evaluation on harmonious feeling of image [J]. Chinese of Journal Electronics, 2001, 29(12A): 1923-1927.
- [2] 李娉婷,石跃祥,戴皇冠. 基于颜色特征的家居设计图分类[J]. 计算机工程,2011,37(16):224-226,229.
Li P T, Shi Y X, Dai H G. Classification of house-designing image based on color feature[J]. Computer Engineering, 2011, 37(16): 224-226, 229.
- [3] 王上飞,陈恩红,王胜惠,等. 基于情感模型的感性图像检索[J]. 电路与系统学报,2003,8(6):48-52.
Wang S F, Chen E H, Wang S H, et al. Kansei image retrieval based on emotion model[J]. Journal of Circuits and Systems, 2003, 8(6): 48-52.
- [4] 李海芳,何海鹏,陈俊杰. 性格、心情和情感的多层情感建模方法[J]. 计算机辅助设计与图形学学报,2011,23(4):725-729.
Li H F, He H P, Chen J J. A multilayer affective model based on personality, mood and emotion[J]. Journal of Computer Aided Design & Computer Graphics, 2011, 23(4): 725-729.
- [5] Cho S B, Lee J Y. A human-oriented image retrieval system using interactive genetic algorithm[J]. IEEE Trans on Systems, Man and Cybernetics, 2002, 32(3): 452-458.
- [6] 黄少荣. 基于随机参数的粒子群优化算法[J]. 重庆师范大学学报:自然科学版,2013,30(6):123-127.
Huang S R. A new particle swarm optimization with random parameters[J]. Journal of Chongqing Normal University: Natural Science Edition, 2013, 30(6): 123-127.
- [7] 冀亚丽,程小平. 基于重量化多特征提取的彩色图像检索[J]. 西南师范大学学报:自然科学版,2005,30(2):254-257.
Ji Y L, Cheng X P. Color image retrieval based-on multi-feature extraction of re-quantization[J]. Journal of Southwest China Normal University: Natural Science Edition, 2005, 30(2): 254-257.
- [8] Ying L, Xin C, Zhang C C, et al. Semantic clustering for region-based image retrieval[J]. J Vis Commun Image R, 2008,20:157-166.
- [9] 孙鑫,谢元旦,任地成. 基于小波变换和子图的图像配准方法[J]. 计算机工程与设计,2010,31(21):4653-4654.
Sun X, Xie Y D, Ren D C. Study on image registration technique based on wavelet transform and subgraph [J]. Computer Engineering and Design, 2010, 31(21): 4653-4654.
- [10] Semeulders A W M. Content-based image retrieval at the end of early years[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2000, 22(12): 1349-1380.
- [11] Goleman D. Emotional intelligence[M]. New York, USA: BantamBooks, 1995.

Sentiment Annotation of Scene Image Based on Fuzzy Theory

CAO Jian-fang^{1,2}, CHEN Jun-jie², LI Hai-fang², ZHAO Juan-juan²

(1. Department of Computer Science & Technology, Xinzhou Teachers' University, Xinzhou Shanxi 034000;

2. School of Computer & Software, Taiyuan University of Technology, Taiyuan 030024, China)

Abstract: The development of electronic technology and imaging technology has resulted in the rapid growth of digital images. It has become an urgent problem to rely on advanced technology to extract sentiment of images and annotate them automatically. A method of sentiment annotation for scene images based on fuzzy theory is proposed. The method describes emotional level of image by computing fuzzy membership degree, uses Adaboost algorithm and RBF neural network to implement it and solve semantic ambiguity on automatic image annotation. Using 869 scene images downloaded by Baidu photo channel to train and test, experiments achieved good effect compared with manual annotation results. The proposed method can lay a good foundation for more types of image sentiment automatic annotation and has some practice value.

Key words: image sentiment annotation; fuzzy theory; sentiment model; RBF neural network; fuzzy membership degree