

基于卷积神经网络的多标签图像自动标注

黎健成^{1,2} 袁春¹ 宋友²

(清华大学计算机科学与技术系 北京 100084)¹ (北京航空航天大学软件学院 北京 100191)²

摘要 如今生活中,图像资源无处不在,海量的图像让人应接不暇。如何快速有效地对这些图像信息进行查询、检索和组织,成为了当前亟需解决的热门问题。而图像自动标注是解决基于文本的图像检索的关键。文中提出的这套基于深度学习模型中的卷积神经网络模型的多标签图像自动标注系统,实现了多标签损失排名函数,完成了多标签数据的训练与测试。在实验验证上,先选取 CIFAR-10 数据集进行算法的有效性测试,然后选取多标签图像数据集 Corel 5k 进行定量测试比较,结果表明,该算法的综合性能指标与现有算法相比有较大的提升。

关键词 图像自动标注,多标签,深度学习,卷积神经网络

中图分类号 TP391.41 文献标识码 A DOI 10.11896/j.issn.1002-137X.2016.7.006

Multi-label Image Annotation Based on Convolutional Neural Network

LI Jian-cheng^{1,2} YUAN Chun¹ SONG You²

(Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)¹

(School of Software, Beihang University, Beijing 100191, China)²

Abstract In today's life, the image resource is almost ubiquitous. An ocean of images make people overwhelmed. How to query, retrieve and organize these image information quickly and effectively is an urgent hot issue. The automatic image annotation is the key of text-based image retrieval solution. A multi-label image annotation system based on a well-known deep learning model, convolutional neural network, was proposed in this paper, together with a multi-label loss ranking function to complete, the training and testing of multi-label image dataset. In the experiments, firstly, CIFAR-10 dataset were selected to test the effectiveness of the algorithm, and then quantitative test comparison was conducted on multi-label image dataset Corel 5k. The proposed solution shows superior performance over the conventional algorithm.

Keywords Image annotation, Multi-label, Deep learning, CNN

1 引言

随着多媒体设备和网络技术的高速发展和日益普及,以图像形式呈现出来的信息越来越多。大量丰富的图像资源吸引了越来越多的用户,用户可以在图像上找到需要的信息。然而,海量大规模的图像信息容易让人感到不知所措,如何较快并且有效地对这些图像信息进行查询、检索和组织,成为当下亟需解决的热门问题。于是,图像检索技术应运而生,并且因此受到了较为广泛的关注。

图像由于往往包含复杂多样的语义信息,通常并不是只由一个标签来标注,因此非常有必要考虑多标签标注的情况。此外,多标签标注相对单标签而言也更能检索带来更多的信息,能更快速、更准确地检索到相应的图像。本文主要基于深度学习模型,对多标签图像自动标注问题进行研究。

多标签图像自动标注问题主要关注包含多个标签的图像的自动标注,即对于给定包含多个语义的图像,计算机需要在

经过一定的计算后,对其进行多个标签的标注。而这些标签则包含了这张图像所带来的大部分信息,人们也可以通过这些标签快速找到这张图像。这与互联网上海量包含大规模语义信息的图像的搜索需求是相吻合的。

2 相关工作

多标签图像自动标注的方案主要可以分为3类:生成模型、判别模型和最近邻居模型。

生成模型(Generative Model)可以随机生成观测数据,特别是在某些隐含参数已经给定的情况下。它可以进一步对主题模型和混合模型进行分类。主题模型把图像作为一个特定混合主题的样本进行标注,每个主题是图像的视觉特征及其标注的标签的分布。如文献[1]中的翻译模型、文献[2]中的潜在狄利克雷分配模型等都是主题模型。混合模型则考虑了图像的视觉特征与标签的联合分布。对于给定的一张新图像,模型会根据图像的视觉特征和归一化的联合分布概率计

到稿日期:2015-12-22 返修日期:2016-02-27 本文受国家自然科学基金(U1433112,61170253),国家核高基项目(2013ZX01039001-002),国家高技术研究发展计划("863计划")(2011AA01A205),清华-腾讯合作项目(人体虚拟形象建模)资助。

黎健成(1993-),男,硕士生,主要研究方向为计算机视觉、模式识别,E-mail:lijiancheng0614@gmail.com;袁春(1969-),男,博士,副研究员,主要研究方向为计算机视觉、视频编码和处理、密码学和隐私保护;宋友(1973-),男,博士,教授,主要研究方向为程序设计、算法分析与设计、信号与信息处理、生物计算等。

算每个标签的条件概率。混合模型可以使用固定数量的标签及其对应视觉特征的混合组件,如文献[3],也可以使用训练数据集的视觉特征与标签的关系作为混合组件,如文献[4]中的 CRM(Cross-media Relevance Model)模型、文献[5]中的 MBRM(Multiple Bernoulli Relevance Model)模型。

判别模型(Discriminative Model)对每个待标注的标签学习一个独立的分类器,并使用分类器对测试的图像进行预测,从而判断这张图像是否能被这个标签标注,如文献[6]。与生成模型不同,判别模型不考虑图像视觉特征与标签之间的联合分布。然而,与生成模型一样,判别模型也是提前选择好图像的视觉特征,并且没有分析特征之间的差异性。

最近邻居模型(Nearest Neighbor Model)随着训练数据的增加而越来越受欢迎。文献[7]介绍了最近邻居标签的传递机制,在这个方法中,图像标注被当作检索问题来解决。最近邻居取决于几个根据视觉特征来计算的距离的平均,也称之为联合平等贡献(Joint Equal Contribution, JEC)。对于待测试的图像,标签是通过邻居传递的。基本颜色和纹理的视觉特征都用于比较和测试,特征选择的正则化也根据标签的相似度来考虑。然而,它并没有增加稀疏性,也没有大幅度提高准确率。文献[8]中的 TagProp 模型是一种新型的最近邻居模型,通过邻居标签存在与不存在的状况来对权重进行结合,取得了较好的效果。

近年来,随着深度学习不断备受关注,一些学者开始研究如何把深度学习运用在计算机视觉上。2012年, Hinton 等人利用多层卷积神经网络对目前图像识别使用广泛的大规模数据库 ImageNet^[9]进行图像分类,取得了非常惊人的识别效果^[10]。之后,大量研究对卷积神经网络进行了各种结构、性能等各方面的改进,如 Google 提出的 GoogLeNet 网络^[11]在 2014 年的图像大规模识别比赛中获得了冠军,微软亚洲研究院视觉计算组开发的基于深度卷积神经网络的计算机视觉系统在 ImageNet 1000 挑战中首次超越了人类进行对象识别与分类的能力^[12]。不过他们解决的是图像分类问题,并未考虑图像多标签的情况。

通过比较多标签图像自动标注的方法,分析目前深度学习模型的性能特点,本文设计了一套基于卷积神经网络模型的多标签图像自动标注解决方案。

3 基于卷积神经网络的多标签图像自动标注算法

卷积神经网络是一种在二维图像数据集中有很好分类效果的模型,而单标签图像分类的成功,不禁让人想到可以把它推广到多标签图像自动标注中。此前,有不少学者会把多标签图像自动标注当成图像分类来进行处理,如判别模型把多标签作为相互独立的标签来考虑,然后对每个互不相关的标签进行分类,最后把分类的结果进行综合后作为每张图像的多标签标注。但是由于每个标签并非对整张图像的所有区域都相关,而且对于一个特定的图像特征来说,它在图像中的位置、姿态、大小都可能会有各种不同的情况,这对简单地使用图像特征进行分类的方法有着非常大的影响,故对整张图像进行标签分类并判定某一个标签是否标注存在非常大的不合理性;并且使用这种方法进行相关实验后,得到的结果也不是十分理想。

既然不改动卷积神经网络而只在最后阶段进行处理的方法不可行,那么,可以考虑对卷积神经网络进行修改,使其能

对多标签进行标注。若是在卷积神经网络中把标签作为一个向量进行训练,然后使用多标签的损失函数来进行后向传播不断更新,那么测试时就能得到较为准确的多个标签。于是,本文提出的解决方案为使用包含多个卷积层和全连接层作为基础架构的卷积神经网络,结合多标签排名损失函数进行训练,最后对测试数据集进行测试。

3.1 模型结构

文献[10]中的卷积神经网络是近年来效果较好的一个模型,许多在其之后提出的方法(如文献[11]和文献[12])也是基于其上进行的研究与改进。本文提出的解决方案基础结构也使用文献[10]中的卷积神经网络模型结构,并对其进行了改进。模型将多标签图像及其标签作为输入,经过 5 个卷积层得到图像的视觉特征,然后经过 3 个全连接层,最后根据多标签排名损失函数得到多个标签。网络进行训练时根据损失函数的结果进行反向传播,测试时根据损失函数选取最佳的一定数量的标签,并与实际的标签进行比较。

模型的参数基本与文献[10]中提出的 AlexNet 类似,在卷积层之后依然使用 ReLU 作为非线性激活函数对其输出进行修正,全连接层的输出大小同为 4096,使用 Dropout 层且其 dropout 比例为 0.6。而其它部分参数经过笔者进行对比实验并分析其结果,决定做出一些修改:1)所有的卷积层中卷积核的间隔(stride)与池化间隔都为 2,之所以不用 4,是因为大的粒度虽会加快速度但会降低精度;2)为了优化整个网络,加入了一个动量项为 0.9 的异步随机梯度下降;3)网络初始的学习速率设为 0.0001,设置较小的学习速率是为了防止欠拟合。

3.2 多标签损失排名函数

虽然网络的结构以及参数等与图像分类常用的卷积神经网络相比有了一定的改进与优化,但基本结构还是比较相似,至于如何设计初始的参数使得网络能得到最好的结果,这并不是本文的讨论重点。而本文一个比较重要的关注点是在损失层上。损失层需要设计一个能对多个标签与实际标签进行比较的损失函数。结合单标签卷积神经网络的损失函数,并考虑多标签损失函数的各种情况,最后设计了一种新的基于 Softmax 回归的多标签排名损失函数。

Softmax 回归可以看作是 logistic 回归在分类问题上的一个推广。因为在分类问题中,logistic 回归是二分类的,即类标签 y 只能取两个值 $\{0, 1\}$,而 Softmax 回归是可以取 k 个不同的值,当 $k=2$ 时,Softmax 回归就变成了 logistic 回归,所以 Softmax 回归其实是 logistic 回归的一般形式。

假设需要对图像标注 c 个标签,若 $f_j(x_i)$ 表示图像 x_i 与标签 j 的激活值,那么对于图像 x_i 和标签 j 的后验概率 p_{ij} ,有

$$p_{ij} = \frac{e^{f_j(x_i)}}{\sum_{k=1}^c e^{f_k(x_i)}} \quad (1)$$

然后要最小化模型预测的标签与实际的标签的 KL 散度。KL 散度(Kullback-Leibler divergence,又称为相对熵)是两个概率分布差别的非对称性度量,它可以用来度量两个概率分布进行编码时相差的平均位元数。一般这两个概率分布一个为理论分布,一个为近似分布。而这与提出的模型得到的概率分布和实际标签的概率分布的情况是相吻合的。

对于每张图像的多个标签,可以假设每个标签之间都是相互独立的,那么这些标签可以组成一个标签向量 $y \in R^{1 \times c}$,

$y_j=1$ 就表示对这张图像标注第 j 个标签;对应地, $y_j=0$ 则表示没有标注第 j 个标签。于是可以通过正则化 y 得到图像 x_i 与标签 j 的概率 $\overline{p_{ij}}$:

$$\overline{p_{ij}} = \frac{y_j}{\|y\|_1} \quad (2)$$

那么要最小化的代价函数为:

$$J = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^c \overline{p_{ij}} \log(p_{ij}) \quad (3)$$

而实际上,如果 $c=1$, j 从 0 开始,那么就与 logistic 回归的代价函数十分类似了,只不过 Softmax 损失函数对标签的 k 个可能值进行了累加。设 c_+ 是对每张图像的标注正确的标签数量,那么可以进一步将代价函数化简为:

$$J = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{c_+} \frac{1}{c_+} \log(p_{ij}) \quad (4)$$

而在这个代价函数的最小化问题上,至今仍未找到比较好的解决办法,不过可以利用梯度下降法等迭代的优化算法。而在具体实现 Softmax 损失函数时,其实权重衰减(weight decay)可以一起使用,即在代价函数中添加一个权重衰减项,这样代价函数就是严格的凸函数了,从而保证收敛并得到全局最优解。

4 实验和分析

代码的实现主要基于 CAFFE(Convolutional Architecture for Fast Feature Embedding)框架^[13],这套基于 BSD 许可包含 C++ 库的框架是目前为止效果较好的卷积神经网络结构的代码实现。它被 Python 和 Matlab 调用进行运算,并且能使用 CUDA 调用 GPU 来计算。根据作者介绍,使用一个 K40 或 Titan 的 GPU 一天可以完成 4 千万张图像的处理,平均 2.5ms 处理一张图像,完全能满足工业和互联网规模的媒体需求。由于并不需要大规模的实验,并且也没有特别的时间需求,故本文使用普通的型号为 NVIDIA GeForce GT 420M 的 GPU 来进行实验。

对于代码中单个标签表示的变量,可以用一个向量来代替,以此来表示多个标签。在 C++ 代码中,原本用 int 类型来表示标注的是第几个标签,而现在需要用一个 vector<int> 类型来表示多个标签,其中 vector 中每一个值表示该标签是否被标注,如 1 为标注, -1 为不标注, 0 为不考虑。而 vector 的大小即是标签总数量,如对于 Corel 5k 数据集即是 260。更改一个变量的类型,需要把用到这类变量的所有代码均进行相应的修改,由于代码修改过多,这里不一一列出。

对于多标签排名损失函数层,则参考 Softmax 损失函数的实现,以及上一节所列出的公式。此外,在这个框架中还需要实现 Accuracy 层,用于计算多个标签是否为实际标签,以及相关的衡量指标。

4.1 基于数据集 CIFAR-10 的算法可行性验证

为了证明卷积神经网络运用在多标签图像自动标注上是可以兼容图像分类的,也为了从侧面证明具体代码实现没有太大的问题,首先把图像分类数据集 CIFAR-10 的单标签扩展成多标签(即把该图像类别标签标注为相关的标签,其它标签标注为无关标签,所有的图像均只有一个标签)进行训练与测试,然后再使用多标签图像数据集 Corel 5k 进行训练与测试,并与其它解决方案进行比较。

CIFAR-10 是一个比较经典、常用的图像分类数据集,它是由 Alex Krizhevsky 等人收集的大小为 32×32 的彩色图像

的集合,共有 60000 张。一般随机选择其中的 50000 张图像作为训练集,而测试集则是剩下的 10000 张图像。为了保证合理地与其它使用该数据集的研究进行对比,本研究使用了其它大部分研究使用的训练集与测试集。每张图像的前景基本都由一个物体占据,故也只能标注一个标签。标签共有 10 个,每个标签都有 6000 张图像。因此数据集可以分为 10 类,且每一类都是相互独立的。

这个实验是用 CIFAR-10 数据集来对卷积神经网络运用在多标签图像标注上的可行性进行定性的验证。首先是使用 CAFFE 框架提供的图像分类模型构建一个 3 个卷积层 + 1 个全连接层的卷积神经网络。之所以只选择 3 个卷积层,是因为 CIFAR-10 中的图像并不是很大,经过多次实验发现卷积层为 3 层时网络基本达到饱和状态,而且 3 层既不会使网络太浅得不到较好的效果,也不至于使网络太深需要运行太长的时间。

首先对本方法的可行性进行数据集测试验证,所以没有对网络进行过多的优化,也没有进行数据增强,仅使用了一个快速得到结果的模型进行对比。对 CIFAR-10 数据集的分类与多标签标注结果如表 1 所列。

表 1 CIFAR-10 实验结果

	准确率 P	召回率 R	F1-Score
分类	0.80	0.63	0.70
多标签标注	0.65	0.69	0.67

从表 1 可以看到,虽然在准确率上多标签标注没有直接分类得到的效果好,但综合考虑得到的 F1-Score 是相近的。因此,基本可以认为卷积神经网络运用在多标签图像自动标注上没有太大的问题。至于多标签标注没有得到与分类更接近的效果,原因可能是多标签标注需要更多的训练参数,那么就需要更多的训练数据才能取得更好的效果。而在实际实验中也发现,多标签标注需要更多的迭代才能达到收敛。

由于使用该数据集只是为了验证本文提出的解决方案的可行性,因此没有必要对其进行过多的优化,也没有与现有的图像分类方法进行比较。

4.2 基于数据集 Corel 5k 的算法性能比较

Corel 5k 是一个常用的多标签图像标注数据集,最早在文献[14]中使用,之后便成为多标签图像标注的一个基准。这个数据集包含大小为 192×168 的 5000 张彩色图像,每张图像都被人为地标注 1 至 5 个标签,平均每张图像能被 3.5 个标签标注,总共有 260 个标签出现。一般随机选择其中的 4500 张用于训练,剩下的 500 张用于测试。为了保证合理地与其它使用该数据集的研究进行对比,本研究使用了其它大部分研究使用的训练集与测试集。虽然该数据集使用得比较多,但经过研究发现,部分图像的标签有误,而且语义存在一定的含糊性,如 lake(湖)和 sea(海)。然而目前暂时未找到更好的多标签标注数据集,且对这个数据集进行实验的研究比较多,于是依然使用该数据集进行实验。作为对比,本研究选了一些具有代表性的模型进行比较,包括 2003 年文献[4]中的 CRM 模型、2004 年文献[15]中的 InfNet 模型、文献[5]中的 MBRM 模型、2005 年文献[16]中的 NPDE 模型、2007 年文献[17]中的 SML 模型、2008 年文献[7]中的 LASSO 模型和 JEC 模型、2009 年文献[18]中的 TGLM 模型和文献[8]中的 TagProp 模型、2012 年文献[19]中的 Group Sparsity 模型。

从表2中可以看到,本文提出的解决方案(Multi-label CNN)与其它几种方案相比,准确率有了较大的提高,进一步证明了该解决方案的可行性。召回率虽然没有 TagProp 模型的高,但也是达到较高的水平,而综合指标 F1-Score 则是最高。准确率比 TagProp 高而召回率没有 TagProp 高,说明了本文提出的多标签卷积神经网络对某一些特征明显的标签几乎都能标注正确,但对于一些包含其语义的图像数量较少的标签,则没有取得很好的效果,因此本文的解决方案更倾向于对某一些标签正确标注而并不是所有标签都存在正确标注。出现这种情况可能是因为对于部分标签,包含其语义的图像数量较少,其出现的频率较低,而卷积神经网络的训练需要大量的数据才能得到较好的特征,因此召回率相对较低。虽然本文的解决方案在 F1-Score 综合衡量上与 TagProp 相比并没有大幅度的提高,但准确率的大幅度提高说明了解决方案在常用的标签上具有较强的实用性;此外,若是使用更多大规模的数据集,本文提出的解决方案理论上能取得较好的效果。

表2 Corel 5k 实验结果

	准确率 P	召回率 R	F1-Score
CRM ^[4]	0.16	0.19	0.17
InfNet ^[15]	0.17	0.24	0.20
MBRM ^[5]	0.24	0.25	0.24
NPDE ^[16]	0.18	0.21	0.19
SML ^[17]	0.23	0.29	0.26
LASSO ^[7]	0.24	0.29	0.26
JEC ^[7]	0.27	0.32	0.29
TGLM ^[18]	0.25	0.29	0.27
TagProp ^[8]	0.33	0.42	0.37
GS ^[19]	0.30	0.33	0.31
Multi-label CNN	0.41	0.35	0.38

从图1中部分图像的标注结果可以看出,本文提出的方法在第一张图像中比 TagProp 的准确率更高。本文提出的方法在第一张图像中把所有实际标签都标注出来了,而 TagProp 并未完全把所有实际标签标注出来。对于第二张图像中的标签 arctic, den, fox, 由于训练集中标注这几个标签的图像较少(如只有几张或几十张),本文提出的方法不容易对其提取出独特的特征,因此不容易标注正确。而对于这样的标签, TagProp 也没有取得较好的效果。

图像	实际标签	本文 预测标签	TagProp 预测标签
	iguana	<u>iguana</u>	<u>iguana</u>
	lizard	<u>lizard</u>	<u>marine</u>
	marine	<u>marine</u>	<u>lizard</u>
	rocks	<u>rocks</u>	<u>water</u>
		water	sky
	arctic	<u>grass</u>	<u>tree</u>
	den	<u>tree</u>	<u>grass</u>
	fox	leaf	rocks
	grass	bush	flowers
		reflection	tiger

图1 本文方法与 TagProp 方法对 Corel 5k 数据集部分图像标注结果对比

图2示出本文提出的方法与 JEC 方法的对比,虽然本文提出的方法只有最后一张图像标注正确标签更多,但从其它几张图像也可以看到本文提出的方法更具合理性。第一张图像两种方法都把所有实际标签都标注出来了,而本文预测的

第5个标签 hills 显然比 JEC 预测的第5个标签更贴近图像的语义。第二张图像中的标签 white-tailed 虽然在训练集中标注该标签的图像并不多,但本文能标注正确,这是因为该标签的特征比较独特,不易与其它标签的特征重合。第三张图像中的标签 fox 与标签 elk, 由于二者的特征较为相近,而训练集中标注该标签的图像较少,无法更进一步获得更多的特征,因此出现了本文提出的方法对某些标签标注不准确的情况。

图像	实际标签	本文预测 标签	JEC 预测 标签
	rocks	<u>rocks</u>	grass
	sand	<u>sand</u>	<u>rocks</u>
	valley	<u>valley</u>	<u>sand</u>
	canyon	<u>canyon</u>	<u>valley</u>
		hills	<u>canyon</u>
	tree	<u>tree</u>	water
	forest	<u>deer</u>	<u>tree</u>
	deer	<u>white-tailed</u>	grass
	white-tailed	grass	<u>deer</u>
		ground	<u>white-tailed</u>
	tree	<u>tree</u>	bear
	snow	<u>snow</u>	<u>snow</u>
	wood	<u>wood</u>	<u>wood</u>
	fox	elk	deer
		rocks	white-tailed

图2 本文方法与 JEC 方法对 Corel 5k 数据集部分图像标注结果对比

结束语 本文提出了一套基于深度学习模型中的卷积神经网络模型的多标签图像自动标注系统,并在经典、常用的多标签图像数据集 Corel 5k 上进行了定量的验证,也与现有的相关文献中叙述该数据集的结果进行了比较。总体而言,与以往的解决方案相比,本文提出的解决方案取得的效果有了一定幅度的提高,有一定的可行性。但限于时间、数据、计算机性能等因素,本文并没有进行过多的实验来证明解决方案的稳定性,且仍然有许多方面值得改进与进一步的研究。以下提出几点较为值得思考的地方:

1) 使用更大规模的数据集进行测试。在现实生活中,随着科技不断发展,信息呈现出爆炸性增长的趋势,图像数据的规模也不断增大。如何能高效快速地解决大规模的图像处理问题,是当今研究的热门方向,也是对解决方案的一个较大挑战。接下来可以使用更大规模的数据集对解决方案进行测试,并根据所得到的结果进行相应的改进,进一步提高解决方案的稳定性。而使用更大规模的图像数据集,对网络的训练有一定的好处,也不容易导致过拟合。

2) 选择更好的损失函数与图像视觉特征,尝试不同的结构。在时间等条件允许的情况下,可以对不同的损失函数、图像视觉特征进行比较,选择效果更好的配置。此外,卷积神经网络也可以选择不同的结构,增加不同的优化。而根据实际需要,还可以研究在限定时间、限定计算机性能等情况下,选择哪一种网络能取得最好的效果。

3) 使用具备噪声的图像数据集。在互联网中,图像是可能存在遮挡、变形等噪声的数据,标签也可能是存在不严谨的标注。虽然卷积神经网络提取的图像视觉特征对于平移与缩放具有一定的不变性,但对旋转等一些变形会受到影响。而标签会存在语义重合、信息冗余等情况,是否需要先对标签进行一定的处理也值得考虑。

参考文献

- [1] Duygulu P, Barnard K, de Freitas J F G, et al. Object recognition as machine translation; Learning a lexicon for a fixed image vocabulary[C]// Computer Vision (ECCV 2002). Springer Berlin Heidelberg, 2002; 97-112
- [2] Barnard K, Duygulu P, Forsyth D, et al. Matching words and pictures[J]. The Journal of Machine Learning Research, 2003, 3 (2): 1107-1135
- [3] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410
- [4] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[C]// Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval. ACM, 2003; 119-126
- [5] Feng S L, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation[C]// Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004 (CVPR 2004). IEEE, 2004, 2
- [6] Grangier D, Bengio S. A discriminative kernel-based approach to rank images from text queries[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(8): 1371-1384
- [7] Makadia A, Pavlovic V, Kumar S. A new baseline for image annotation[M]// Computer Vision (ECCV 2008). Springer Berlin Heidelberg, 2008; 316-329
- [8] Guillaumin M, Mensink T, Verbeek J, et al. Tagprop: Discriminative metric learning in nearest neighbor models for image auto-annotation [C] // IEEE 12th International Conference on Computer Vision, 2009. IEEE, 2009; 309-316
- [9] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]// IEEE Conference on Computer Vision and Pattern Recognition, 2009 (CVPR 2009). IEEE, 2009; 248-255
- [10] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. 2012; 1097-1105
- [11] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]// CVPR. 2015; 1-9
- [12] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]// ICCV. 2015; 1026-1034
- [13] Jia Y, Shelhamer E, Donahue J, et al. Caffe: Convolutional architecture for fast feature embedding[C]// Proceedings of the ACM International Conference on Multimedia. ACM, 2014; 675-678
- [14] Duygulu P, Barnard K, de Freitas J F G, et al. Object recognition as machine translation; Learning a lexicon for a fixed image vocabulary[M]// Computer Vision (ECCV 2002). Springer Berlin Heidelberg, 2002; 97-112
- [15] Metzler D, Manmatha R. An inference network approach to image retrieval[M] // Image and video retrieval. Springer Berlin Heidelberg, 2004; 42-50
- [16] Yavlinsky A, Schofield E, Ruger S. Automated image annotation using global features and robust nonparametric density estimation[M] // Image and video retrieval. Springer Berlin Heidelberg, 2005; 507-517
- [17] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(3): 394-410
- [18] Liu J, Li M, Liu Q, et al. Image annotation via graph learning [J]. Pattern recognition, 2009, 42(2): 218-228
- [19] Zhang S, Huang J, Huang Y, et al. Automatic image annotation using group sparsity[C]// 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2010; 3312-3319
- [54] Baniassad E, Clarke S. Finding Aspects in Requirements with Theme/Doc[C]// Proceedings of Early Aspects 2004: Aspect-oriented Requirements Engineering and Architecture Design. 2004
- [55] Elrad T, Filman E R, Bader A. Aspect-oriented programming: Introduction[J]. Communications of the ACM, 2001, 44(10): 29-32
- [56] Sampaio A, Rashid A, Chitchyan R. EA-Miner: Towards Automation in Aspect-Oriented Requirements Engineering [J]. Transactions on Aspect-Oriented Software Development III, 2007, 4620; 4-39
- [57] Mylopoulos J. Goal-Oriented Requirements Engineering [C]// 12th Asia-Pacific Software Engineering Conference (APSEC). 2005; 3
- [58] Anton A I, McCracken W M, Potts C. Goal Decomposition and Scenario Analysis in Business Process Reengineering[M]// Advanced Information Systems Engineering. Springer-Verlag, 1994; 94-104
- [59] Liu Wei, He Ke-qing, Wang Jian, et al. Heavyweight Semantic Inducement for Requirement Elicitation and Analysis [C] // Third International Conference on Semantics, Knowledge and Grid. 2007; 206-211
- [60] Rebelo H, Lima R, Leavens G T. Modular contracts with procedures, annotations, pointcuts and advice[C]// Proceedings of the 2011 Brazilian Symposium on Programming Languages. 2011; 1-16
- [61] Khatchadourian R, Greenwood P, Rashid A, et al. Pointcut Rejuvenation; Recovering Pointcut Expressions in Evolving Aspect-Oriented Software[J]. Sactions on Software Engineering, 2012, 38(3): 642-656
- [62] Wang L, Aotani T, Suzuki M. Interaction awareness for aspect refactoring[C]// Proceedings of the 8th International Workshop on Advanced Modularization Techniques-AOAsia/Pacific. New York: ACM, 2013; 15-18
- [63] Lee J, Hsu K. GEA: A Goal-Driven Approach to Discovering Early Aspect[J]. IEEE Transactions on Software Engineering, 2014, 40(6): 584-602

(上接第 12 页)

何成万, 张立军, 张慧. 基于元数据和反射的面向方面软件演化方法[J]. 电子学报, 2011, 39(8): 1771-1777