

doi:10.3772/j.issn.1000-0135.2016.012.007

# 医学图像模态特征表达及其比较研究<sup>1)</sup>

丁恒 陆伟 林霞 曹光超

(武汉大学信息管理学院, 信息检索与知识挖掘研究所, 武汉 430072)

**摘要** 医学图像模态的自动标注是建立检索词和医学图像间准确关联, 实现医学图像检索的基础。当前医学图像的模态特征表达多采用基于单一表达模型的方法, 没有充分运用不同表达模型的互补优势, 医学图像的语义描述框架与医学图像检索需求不匹配。针对这些不足, 本文在介绍医学图像模态分类框架和标准数据集的基础上, 分析了文本特征和视觉特征对医学图像模态自动标注的影响, 对比了向量空间模型(VSM)、潜在语义索引(LSI)、视觉词袋模型(BOW)、卷积神经网络模型(CNN)等多种特征表达模型在医学图像模态自动标注上的效果, 研究发现文本特征与视觉特征各适用于部分医学模态的识别与区分, 组合多种表达模型能够获得最佳的标注效果。

**关键词** 医学图像检索 图像标注 ImageCLEFmed 模型组合

## A Research On Medical Image Modality Feature Representation and Its Comparison

Ding Heng, Lu Wei, Lin Xia and Cao Guangchao

(School of Information Management, Wuhan University, Wuhan 430072)

**Abstract** Automatic image annotation is the foundation of medical image search. The current methods are mostly based on single feature representation model, thus information on different feature spaces of medical image is not fully considered. And also the semantic description framework of medical images does not meet the requirement of medical image retrieval. In this paper, we introduced the category of medical image modality and a benchmark dataset for automatic medical image modality annotation. We designed an experiment to compare and analyze different feature representation model such as vector space model, latent semantic indexing, bog of words and convolutional neural networks, our experiment shows that combination of feature representation models obtains definite improvements.

**Keywords** medical image retrieval, image annotation, ImageCLEFmed, model combination

### 1 引言

医学图像在生物医学文献中具有重要地位, 往往被视作“解释现象的图例”、“证明观点的证据”, 是临床医生和医学研究人员重点关注的信息资源。然而当前成熟的文献信息检索系统大多关注于文本

内容, 甚少涉及医学文献中的图像信息, 难以满足医学工作者的实际检索需求<sup>[1]</sup>。图像自动标注技术是建立用户查询与图像语义关联的重要方法, 是实现图像语义检索的核心基础性问题<sup>[2,3]</sup>, 在医学信息检索领域, 利用各种相关信息实现医学图像的语义自动标注已成为医学信息检索的前沿。

目前, 医学图像自动标注多采用基于单一表达

收稿日期: 2015年12月5日

作者简介: 丁恒, 男, 1988年生, 博士研究生, 主要研究方向: 多媒体信息检索、数据挖掘, E-mail: chrisdy2014@gmail.com。陆伟, 男, 1974年生, 博士, 副院长, 教授, 主要研究方向: 信息检索、知识管理、数据挖掘等。E-mail: weilu@whu.edu.cn。林霞, 女, 1990年生, 硕士研究生, 主要研究方向: 信息服务。曹光超, 男, 1991年生, 硕士研究生, 主要研究方向: 专家检索。

模型的图像视觉特征方法<sup>[4-7]</sup>,这一系列研究为对该领域的发展产生了巨大的推进作用。然而时至今日也未能催生出成熟的相应应用产品,究其原因在于图像视觉特征与图像高层语义之间普遍存在语义鸿沟,多数视觉特征描述符仅仅刻画了医学图像某一方面的信息(如颜色、纹理、形状),而医学图像的高层语义往往是多方面信息的综合体现,医学工作者的实际检索需求正是面向医学图像的高层语义展开。此外,由于领域内研究人员的学科背景局限性(大多数为计算机科学、情报学),使得现有的研究者缺乏对医学信息检索用户需求的深刻认识(医学工作者是如何运用专业知识进行医学信息检索),因此多数研究中构建的关键词词表(分类目录)不能满足医学图像检索的实际需求。

本文认为,医学图像的语义分类框架理应与医学领域的知识组织体系紧密结合,由此产生的医学图像标注结果才能够适应于医学图像的检索需求,因此本研究中引入了医学图像的模态分类框架。其次,由于文本信息与“语义”之间存在的天然联系,挖掘医学图像的相关文本描述内容与医学图像高层语义之间的映射关系,将文本语义信息与图像物理信息相融合,是有效解决“语义鸿沟”问题的一种途径,因此本文从语义关联的角度切入,对比分析了文本特征表达模型(向量空间模型、潜在语义索引)和视觉特征表达模型(视觉词袋模型、卷积神经网络模型)在医学图像模态特征表达上的能力,并探讨了文本和视觉特征模型组合对医学图像模态自动标注的影响。本文的章节结构如下,第二节对相关研究进行评述;然后第三节简要介绍 ImageCLEFmed 的医学图像模态分类框架演化和标准数据集变化;第四节说明和设计了医学图像模态自动标注实验,探讨了不同特征表达模型在该任务上的效果,并分析说明了相应的原因;最后对本文研究进行了分析和总结,并提出了进一步的研究工作。

## 2 相关研究述评

医学图像模态自动标注(分类)研究从本质上看是一个关于医学图片的模式识别问题,即“通过一个学习算法对某标准数据集进行学习,从而构建一个估计模型,当向该模型输入一个新的医学图片时,模型能够判断图片的具体含义,为其匹配合理的语义标签”,其研究深受图像识别、机器学习等学科领域的影响。如文献[5]提出了使用小波变换纹理

特征对正常脑部 CT 图像和含有病灶的异常脑部 CT 图像进行识别分类,克服了旋转角度对于分类识别的影响;文献[8]探讨了 LBP 特征和 ModSIFT 特征对医学图像分类的影响,实验结果证明采用高层特征融合方式融合 LBP 特征和 ModSIFT 特征可以达到较高的分类正确率;文献[9]运用 SURF 局部纹理特征和 HSV 颜色直方图特征对包括人物、汽车、恐龙、大象、花卉、马群、山川、美食的八类图像进行了分类标注实验;文献[10]结合 LBP 特征描述符和 BP 神经网络对腹部、胸部、头部、其他四类医学图像进行自动分类标注;文献[4]提出了组合运用贝叶斯和神经网络的医学图像标注模型,对正常 CT 肝脏图像和异常 CT 肝脏图像进行分类;文献[11]研究了支持向量机(SVM)在医学图像分类中的运用;文献[3]探讨了使用医学图像模态类别作为查询拓展对医学信息检索的影响,研究发现精确的医学图像模态分类有利于提高检索的准确性,提升医学信息检索系统的效果。文献[12]提出了稀疏自编码仿射不变特征描述符(sparse coded affine-invariant descriptors),并将其运用于八类别的医学图像模态分类,报告结果显示该方面具有较高的正确率。

上述研究一方面致力于寻找适合医学图像分类的视觉特征描述符,另一方面关注于机器学习分类算法在医学图像分类上的效果差异,究其本质都是挖掘医学图像视觉内容信息与高级语义(模态)之间的映射关系。然而不论是局部图像视觉特征描述符或者全局图像视觉特征描述符往往揭示的是图像低层语义信息(颜色、纹理、形状、空间关系),都面临低层语义与高层语义之间普遍存在的语义鸿沟问题,因而难以有效的运用于面向高层语义的医学图像检索领域(基于医学图像模态的医学图像检索需求<sup>[13]</sup>)。尤其需要注意的是,随着医学信息检索研究的深入,医学模态分类框架为了满足用户的实际检索需求,其类目体系的划分并非严格依照视觉形态的差异,例如肝脏 CT 图像和脑部 CT 图像在医学图像模态分类上都属于 CT 图像类别,然而现有的视觉特征描述符都是从图像低层语义信息(颜色、纹理、形状)上对图像内容加以描述,因此往往只适用于形态相异的图像分类,即图像低层语义如颜色、形状、纹理上有较大差异的图像分类问题。有鉴于此,笔者认为解决这一问题的主要途径有二,一是继续寻找能够表征医学图像模态语义信息的视觉特征,从而构建医学图像模态视觉特征与医学图像模态语义之间的映射关系,二是利用医学图像相关的

文本信息与其模态语义之间的天然联系,挖掘医学图像的文本描述与医学图像模态分类之间的关联关系,并将其与视觉特征描述符所代表的图像低层语义信息相结合,利用不同特征表达模型的优势。本研究正是沿着第二条思路加以展开。

### 3 医学图像模态及其分类 框架和标准数据集

ImageCLEF ( Cross-language image retrieval evaluations forum)从2010年开始设立医学图像模态分类任务,关注利用医学图像模态分类促进医学信息检索的研究,是国内外专家学者、相关研究人员了解、跟踪医学图像模态分类研究进展的重要窗口。正如前文所述,长期以来有相当一部分研究者根据自我对医学图像分类研究的认识和独特的需求,对该领域进行了研究探索。然而,生物医学学科领域的专业性为医学信息检索研究增添了很多有别于传统信息检索的独特性,这导致医学信息检索的发展需要多学科、跨领域的深刻交流,才能够推进该领域的研究真正面向实际运用。有鉴于此,ImageCLEF广泛邀请各界学者参与该领域的研究,征集相关研究论文,经过多年探讨,于2010年首次确立了一个被认可的、面向医学信息检索的医学图像模态分类框架,并发布了相关的标准数据集“ImageCLEFmed 2010”。笔者认为,从ImageCLEF 2010开始,医学图像模态分类研究踏入了一个崭新的时期,一方面医学图像模态分类框架的发布破除了过去研究缺乏统一的分类目录标准的困境,推进该研究进入以检索需求驱动的研究阶段,另一方面标准数据集的发布使得相关

研究有了共同的研究素材、评判基准,有利于该领域研究成果的相互比较和学习。ImageCLEFmed 发展至今,其医学图像模态分类框架和标准数据集共经历了三次变更,其具体变化如下表1所示。

ImageCLEFmed 2010 至 ImageCLEFmed 2011,从模态分类框架来看类目个数由8增加至18,类目中增加了诸如 Photograph、Graphic、Flow Chart 这样的非医学专有图像类别,主要是考虑到此类图像在医学文献中往往用于“解释实验设计框架”、“说明实验结果”,具有检索价值,其类目层次体系从一层变为二层,更贴近医学领域的医学图像常用的组织形式。ImageCLEFmed2012,将类目层次体系从二层拓展到了三层,首先考虑医学文献中图像的功能性,将复合图像、诊断图像、一般生物医学图例区分开来;其次,对于专业医学图像(诊断图像)根据成像技术的不同划分为 Radiology、Microscopy、3D reconstructions、Visible light phototgraphy、Printed signals & waves 等5种类型,充分考虑了临床医学工作者对医学图像的认知习惯。一般生物医学图例的次级分类目录主要依据图像的内容和功能差别,例如程序列表(Program listing)、基因序列(Gene sequence)、数学公式(Mathematics, formulae)、化学结构(Chemical structure)等。ImageCLEF不断完善医学图像模态分类框架,建立如此详细的类目体系(医学图像信息组织体系),其主要目的在于推进医学图像检索向医学图像语义理解层面发展,构建面向临床医学工作者检索需求的医学信息检索系统。本文采用最新的类目框架“ImageCLEFmed 2013”,如图1所示。

表1 ImageCLEFmed 模态分类框架和数据集变化表

版本	医学图像模态分类框架和数据集
ImageCLEFmed2010	CT、MR、XR 等八个模态类目,训练集 2000 张图片,测试集 2000 张图片
ImageCLEFmed2011	Radiology、Microscopy、Photograph、Graphic、Other 等五大类,AN、CT、MR、US 等 18 小类,二层分类目录体系
ImageCLEFmed2012 (ImageCLEFmed2013,由于13版和12版分类框架无变化,且数据集变换很小,因此本文中将其视作同一版本)	Compound images、Diagnostic images、Generic biomedical illustrations 三大类目,Diagnostic images 又分为 Radiology、Microscopy、3D reconstructions 等 5 个二级目录和 PET、Endoscopy、X-Ray 等 17 个三级目录,共 31 个类目。训练集 2901 张图片,测试集 2500 多张图片

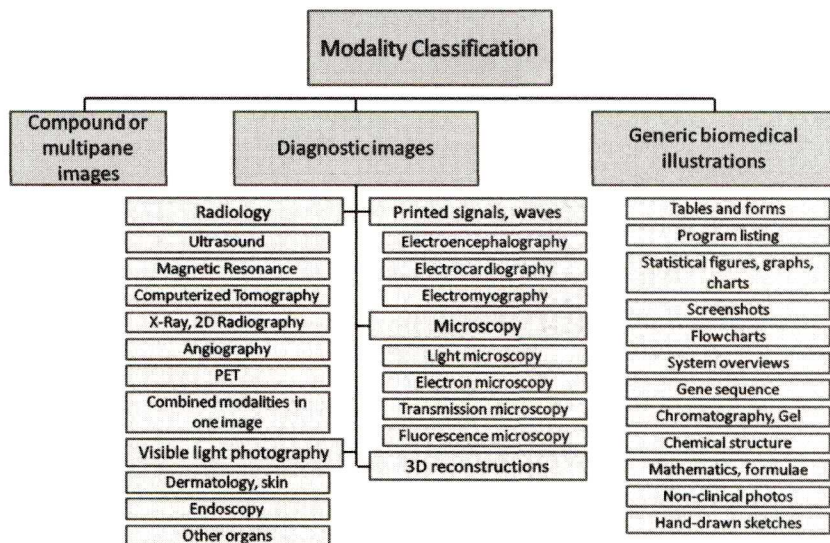


图1 ImageCLEFmed2013 医学图像模态分类体系<sup>[14]</sup>

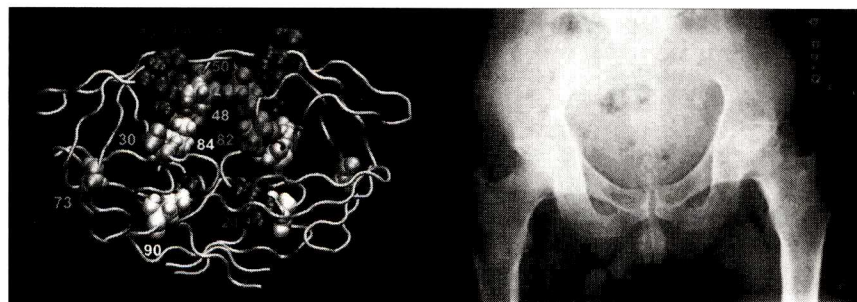


图2 左 D3DR(3D 重构),右 DRXR(X 射线)

纵观医学图像模态分类目录可知,医学图像的模态分类主要参考两个依据,一是模态类别尽量与现实临床医学诊断的图像资源相匹配,即按照生物医学研究和工作领域的习惯方式组织医学图像信息,这一部分图像资源都是由专业医疗设备所产生;二是从图像对医学研究、工作的作用层次上看,将与医学研究和工作有关的统计图表、流程图、化学结构图等图像也视为医学图像的模态类别,这部分图像资源往往是有医学研究工作者生产。因此本文将医学图像模态归纳为“医学图像模态是指由用于描述专业医疗设备和医学工作者生产创造的不同内容形态、不同作用的医学图像信息资源的分类目录”。

#### 4 医学图像模态自动标注

医学图像模态自动标注研究的主体思想是:

(1)“医学图像上下文中包含众多词汇,这些词汇潜藏的语义信息往往揭示了医学图像描述的具体内容,一定程度上体现了医学图像在模态层面的主

题”。例如一幅图像的相关文本描述为“SNP validation from CGP panels 1 and 2. Flowchart showing the output resulting from testing of the 3072 selected SNPs. The 1620 predicted alleles correspond to the set of validated SNPs”,此描述中出现了“Flowchart showing”这样的模态特征信号,很容易猜测该医学图像对应的模态是流程图的可能性很高。再比如“section of the tumor revealed ‘fish-meat’ like greyish-white color inside with spotted bleeding and cystic necrosis cavities ranging 1.2-3.0 cm in diameter filled with brown color fluid (figure 2)”一句中出现了“spotted bleeding and cystic necrosis(出血和囊性坏死)”,而其图像对应真实的类别是“Angiography”(血管造影),这体现了某些专业术语、专有名词和医学图像模态类别之间的存在对应关系,在大规模数据集上挖掘这些关系,是实现医学图像模态自动标注的关键。

(2)“医学图像的视觉特征有利于区分视觉差异较大的临床诊断医学图像”。例如图2左右分别

为3D重构图(D3DR)和X射线图(DRXR),其颜色、纹理、形状上皆存在较大差异,可以利用其这些视觉特征的差异性,建立低层视觉特征与高层医学图像模态语义直接的联系。

(3)对于文本(或图像)信息,不同特征表达模型能够表达不同层次的有效信息,组合多种特征表达模型,有利于区分单一表达模型无法识别的差异,联合文本特征和视觉特征有利于结合两者的优势,从而提高医学图像模态自动标注效果。

#### 4.1 问题描述

参照文献[14]和文献[15]有关监督学习任务 and 分类任务的描述,本文将医学图像模态自动标注方法作如下形式化定义:

对于给定的医学图像  $t_i = (x_i, y_i)$ , 其中  $x_i = (x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)})$  是该医学图像在某特征表达模型下的向量表达,  $y_i \in C = \{c_1, c_2, \dots, c_j\}$  表示该医学图像所属的模态类别。利用某种学习算法,对已知分类结果的数据集  $T = \{t_1, t_2, \dots, t_N\}$  进行学习,得到某个函数映射关系(模型)  $f: x \rightarrow c$ , 使得对于任意未知模态类别的医学图像  $t' = (x', ?)$ , 可通过模型  $f$  估计其模态类别  $y'$ 。求解该类问题两个核心是特征向量  $x_i$  和模型  $f$ , 分别对应于模式识别领域的特征表达和分类算法两大问题。

#### 4.2 特征表达模型及模型组合

医学图像的特征信息包括文本信息和图像信息。文本信息来源于医学图像相关的语句,包括医学图像对应的文献标题(title)、图头(caption)和上下文(context,指除图头外用于解释说明医学图像的语句),笔者认为以上三种信息与图像语义内容的相关性排序依次为图头 > 上下文 > 文献标题。图像信息来源于医学图像本身,包括其颜色、纹理、形状等。不同的特征表达模型将产生不同的特征空间,体现不同层次的特征信息。本文的研究核心是比较和分析不同特征表达模型在医学图像模态特征表达上的能力,并探讨多种特征表达模型组合(联合特征空间)对于医学图像模态自动标注效果的影响。因此,研究中采用了向量空间模型(Vector Space Model)<sup>[16]</sup>、潜在语义索引(Latent Semantic Indexing)<sup>[17]</sup>、词袋模型(Bag of features)<sup>[18]</sup>、卷积神经网络(Convolutional Neural Networks)<sup>[19]</sup>,从不同的层次上分别对医学图像的文本信息和视觉信息进

行特征表达建模,详见表2。

表2 特征表达统计表

特征类别	特征表达模型
文本	VSM (Vector space model) + TF-IDF
	LSI (Latent semantic indexing)
图像	BOF (Bag of features) + SIFT (Scale-invariant feature transforms)
	CNN (Convolutional Neural Networks)

模型组合是指将不同特征表达模型得到的特征向量连接起来,形成新的特征向量,其形式化描述如下:对于给定的医学图像  $t_i = (x_i, y_i)$ , 其在联合特征空间  $S = \{s_1, s_2, \dots, s_k\}$  上的特征向量为  $x_i^S = (x_i^{s_1}, x_i^{s_2}, \dots, x_i^{s_k})$ , 其中  $x_i^{s_j} = (x_i^{s_j(1)}, x_i^{s_j(2)}, \dots, x_i^{s_j(n)})$ 。

#### 4.3 实验过程及结果分析

##### (1) 数据处理

本文选取“ImageCLEFmed2013”标准数据集的部分数据作为实验数据。该数据集包括已标注模态类别的医学图像5482张,以及74654篇生物医学文献的文本数据,医学图像均源于这74654篇文献。实验首先通过图片编号(iri)寻找图像对应的医学文献文本,建立对应关系。然后抽取了医学图像对应的文献标题(title)、图头(caption)和上下文(context)语句,按图片编号(iri)、图头(caption)、文献标题(title)、上下文(context)、模态分类数字编号的格式重新组织。

由于“ImageCLEFmed2013”标准数据集中COMP(复合图表,2119条)的样本量远大于其他类别,且可采用复合图表切割技术将其转化为多个其他类别的子图像<sup>[20]</sup>,因此本研究中将该类数据去除,保留其他30个类别,共3363条数据的数据集。为了后续实验方便,本文对医学图像模态类别进行了数字编码,类别编码及样本统计详见表3,其中编号19缺省(留给COMP),编码含义详见文献[20]。

##### (2) 特征表达

向量空间模型:本研究使用nlk<sup>[21]</sup>自然语言处理工具包进行分词和停用词处理,采用波特词干器(Porter Stemmer)<sup>[22]</sup>作词干提取,最终将医学图像对应的文本语句转化为词的集合,用tf-idf表示词的权重,进而构建医学图像的向量空间表达。



表 3 类别编码和样本统计表

数字编码	英文编码	样本数	数字编码	英文编码	样本数
1	GTAB	94	16	DRCO	23
2	DMFL	66	17	DSEM	19
3	GGEL	85	18	GPLI	50
4	GFIG	204	20	GMAT	25
5	DVOR	162	21	DSEC	125
6	GCHE	81	22	DRXR	414
7	DMLI	212	23	DRMR	187
8	DRAN	72	24	GFLO	114
9	GSCR	111	25	DVDM	107
10	DVEN	84	26	DMEL	71
11	GHDR	100	27	DRCT	299
12	GGEN	89	28	DRPE	19
13	GSYS	105	29	GNCP	133
14	DSEE	30	30	DMTR	65
15	DRUS	145	31	D3DR	72

潜在语义索引:研究使用 *gensim*<sup>[23]</sup> 主题模型工具包将图像对应的文本语句投影到 200 维的主题分布向量上,将该向量作为医学图像在 LSI 空间的特征表达。

词袋模型:本文首先使用 *vlfeat*<sup>[24]</sup> 开源机器视觉库抽取每张医学图像的 SIFT 特征<sup>[25]</sup>,然后运用 *kmeans++* 聚类算法将所有的 sift 特征向量聚类到 1000 个类(考虑到聚类的时间消耗,实际只使用了 90 万个 sift 特征向量),对于每张医学图像将其所有的 sift 特征投影到 1000 个类别维度上,从而形成该医学图像在 BOF 空间的特征表达。

卷积神经网络特征:鉴于深度学习(deep learning)在机器视觉领域的广泛成功<sup>[20]</sup>,本文采用 *overfeat*<sup>[26]</sup> 预训练的卷积神经网络提取医学图像的特征,该特征为 4096 维特征向量。

### (3) 实验设置及结果分析

为了探析不同特征表达模型及模型组合对最终标注结果的影响,研究设置了多组实验方案,具体实验方案设计和实验结果如表 4 所示。笔者使用 *liblinear* 工具包<sup>[27]</sup> 对本文提出的思路进行分类实验,以准确率、召回率、*F* 值作为评测标准,实验结果以五折交叉检验作最终估计。

表 4 实验结果表

实验编号	平均正确率	平均召回率	平均 <i>F</i> 值
VSM	0.7369	0.6679	0.6922
LSI	0.6684	0.6380	0.6475
BOF	0.5879	0.5363	0.5375
CNN	0.7161	0.6884	0.6971
VSM + LSI	0.7342	0.6824	0.7002
BOF + CNN	0.7667	0.7348	0.7459
CNN + LSI	0.8118	0.7801	0.7926
BOF + CNN + VSM + LSI	0.8695	0.8195	0.8374

第一组实验(VSM/LSI/BOF/CNN)选择单一特征表达模型表达医学图像的模态特征。结果对比表明,使用文本特征时,向量空间模型优于潜在语义索引主题模型(0.6922 > 0.6475)。向量空间模型是从词的层面对医学图像的文本信息进行建模,而潜在语义索引主题模型则从主题层面上表达文本信息特征。当仅以文本信息为依据区分医学图像模态类别时,大部分关键的有效特征信息往往是相关句子中的某些特殊词汇,例如“SNP validation from CGP panels 1 and 2. Flowchart showing the output resulting

from testing of the 3072 selected SNPs. The 1620 predicted alleles correspond to the set of validated SNPs”此句中,大多数词没有贡献出有效的模态信息,但“Flowchart”一词提供了强烈的模态特征信息。向量空间模型虽然使得特征空间维度极大的增加,但是同时考虑到句子中每一个词的出现对医学图像模态识别的影响。而潜在语义索引主题模型则将不同的词映射到某一主题下,虽然降低了特征空间的维度,但是同时也可能导致关键词信息的丢失,从而影响识别效果。使用图像特征时,卷积神经网络模型优于视觉词袋模型(0.6971 > 0.5375),这一结果并不令人意外。正如卷积神经网络在其他计算机视觉任务中表现出的优势一样,它能够更好的模拟人类视觉神经系统对事物的识别和认知方式,特别是对某些细节纹理差异的区分。而视觉词袋模型首先将图像分割成很多小的区块,然后通过聚类将成千上万个区块根据视觉相似性聚类成有限个数的视觉词汇,这一过程会丢失一部分有效特征信息。此外,最优单一特征表达模型是卷积神经网络模型(0.6971 > 0.6475)。这主要是因为大多数医学图像模态始终能够通过视觉内容区分,而很多时候医学图像的文本信息中可能不包含任何用于识别其模态类型的特征信息。

为了考虑特征表达模型对于每个类别的效用,研究计算了每个类别的正确率、召回率和  $F$  值,结果表明有些类别在文本特征中较容易区分,另一些类别在图像特征空间中较易区分,详见表 5,其中类别 13、16、20、31 等 4 各类别在上述 4 个空间中的区分度都较低,最佳  $F$  值小于 0.6。值得注意的是,在那些文本信息优于图像信息的类别上,有时潜在语义索引主题模型优于向量空间模型,在图像信息优于文本信息时,在类别 22(DRXR)下视觉词袋模型优于卷积神经网络模型。这说明在运用文本信息区分医学模态类型时,并非所有的模态区分都需要考虑到词的层面,有一些医学模态仅从主题层面就能够很好的区分,深入词层面之后增加了特征空间的维度,从而引发了“维度灾难”问题,抵消了细节特征对医学图像模态自动标注的积极影响。同理,本文中视觉词袋模型特征向量为 1000 维,而卷积神经网络模型特征向量为 4096 维,高维视觉空间能够更好的表示细节差异,同时也面临更严重的维度灾难。对于类别 22(DRXR)的医学图像模态来说,其与其他模态类型的视觉差异非常明显,仅仅使用一个较低维的特征空间就能够很好的表达模态差异性,因此视觉词袋模型反而优于卷积神经网络模型。

表 5 最佳单一特征空间统计表

类别	最佳空间	最佳 $F$ 值	数字编码	最佳空间	最佳 $F$ 值
1	CNN	0.6595	16	CNN	0.5946
2	CNN	0.8732	17	VSM	0.9730
3	VSM	0.8639	18	CNN	0.7931
4	CNN	0.7563	20	VSM	0.3889
5	CNN	0.7105	21	LSI	0.9291
6	CNN	0.8313	22	BOF	0.9147
7	CNN	0.8986	23	LSI	0.8122
8	CNN	0.7733	24	VSM	0.6325
9	CNN	0.7545	25	CNN	0.7488
10	CNN	0.8075	26	VSM	0.7101
11	CNN	0.7526	27	LSI	0.8235
12	VSM	0.7513	28	CNN	0.6842
13	LSI	0.5619	29	VSM	0.6638
14	LSI	0.9492	30	VSM	0.6226
15	CNN	0.8919	31	LSI	0.5517

第二组实验选择将不同特征表达模型组合在一起用于表达医学图像的模式特征。结果表明同类特征表达模型(文本特征之间或者图像特征之间)的组合能够提高医学模态类别的区分能力,而异类特征表达模型(文本特征与图像特征之间)的组合可较大的提升分类标注结果(5%的提升)。组合4个特征表达模型将获得最佳结果,说明加入有效特征表达模型有利于提升医学模态分类标注的效果。此外,类别13和20的 $F$ 值低于0.7,分别为0.6598和0.5263,其在单一特征表达模型下最佳 $F$ 值分别为0.5619和0.3889,这表明模型组合在13和20两个类别上效果不佳的原因在于:“没有寻找到一个合适的单一特征表达模型能够有效的区分这两个类别”,即VSM、LSI、BOF、CNN四个特征表达模型在13和20两个类别上区分能力较弱。寻找能够有效表征这两个类别的差异信息的特征表达模型,将有利于提高这两个类别的分类效果。

## 5 结 语

本文通过在国际标准数据集“ImageCLEFMed2013”上设计医学图像模态自动标注实验,分析比较了向量空间模型、潜在语义索引、视觉词袋模型和卷积神经网络模型等四个特征表达模型在医学图像模态特征表达上的能力,并探讨了模型组合对医学图像模态自动标注的影响。实验结果表明“不同空间的特征表达了医学图像不同层次的信息,同时考虑多特征空间,有利于区分单一特征无法识别的差异,从而提高医学图像模态自动标注效果”。研究主要有以下三个发现:①有些医学模态类别在文本特征中较容易区分,另一些类别在图像特征空间中较易区分;②异类特征模型组合优于同类特征模型组合;③若所有特征表达模型都不利于区分某类别,则其组合模型在该类别的区分能力较弱。此外,本文发现了某些医学图像模态适合于使用高维特征空间,突出细节差异性,而某些医学图像模态则适用于低维特征空间,降低“维度灾难”的影响。研究的不足在于,虽然指出了VSM、LSI、BOF、CNN特征空间在13和20两个类别上区分能力较弱,但是没有进一步寻找有效特征表达模型。

## 参 考 文 献

[1] Han X H, Chen Y W. Biomedical imaging modality classification using combined visual features and textual terms [J].

- Journal of Biomedical Imaging, 2011, 2011:1-7.
- [2] 陆泉,陈静,丁恒. 基于社会标签的图像情感自动分类标注研究[J]. 图书情报工作, 2014, 58(12): 118-123.
- [3] Castellanos A, Benavent X, García-Serrano A, et al. Multimedia retrieval in a medical image collection; results using modality classes [C]// Hayit Greenspan. MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support. Nice: Springer Berlin Heidelberg, 2012:133-144.
- [4] 陈健美,宋顺林,朱玉全,等. 一种基于贝叶斯和神经网络的医学图像组合分类方法[J]. 计算机科学, 2008, 35(3):244-246.
- [5] 宋余庆,刘博,谢军. 基于Gabor小波变换的医学图像纹理特征分类[J]. 计算机工程, 2010, 36(11): 200-202.
- [6] 张翼. 基于局部特征医学图像分类中关键技术研究[D]. 成都:电子科技大学, 2013.
- [7] 李林,吴跃,叶茂. 一种基于HOG-PCA的高效图像分类方法[J]. 计算机应用研究, 2013, 30(11): 3476-3479.
- [8] 武京相. 融合全局和局部特征的医学图像分类[D]. 成都:电子科技大学, 2010.
- [9] 梁进,刘泉,艾青松. 基于SURF和全局特征融合的图像分类研究[J]. 计算机工程与应用, 2013(17): 174-177.
- [10] 孙君顶,李琳. 基于BP神经网络的医学图像分类[J]. 计算机系统应用, 2012(3):160-162.
- [11] 华翔,孙蕾. 基于SVM的医学图像分类器的设计[J]. 微电子学与计算机, 2011, 28(6):171-175.
- [12] Gál V, Solt I, Kerre E, et al. Modality classification for medical images using sparse coded affine-invariant descriptors [C]// Takashi Washio; Emerging Trends in Knowledge Discovery and Data Mining. Kuala Lumpur: Springer Berlin Heidelberg, 2013:1-10.
- [13] Pauly O, Mateus D, Navab N. ImageCLEF 2010 Working Notes on the Modality Classification Subtask [C]// Working notes of CLEF. CLEF, 2010:1-9.
- [14] 李航. 统计学习方法[M]. 北京:清华大学出版社, 2012:6-10.
- [15] Manning C D, Raghavan P, Schütze H. 信息检索导论[M]. 北京:人民邮电出版社, 2010:177-178.
- [16] Salton G, Wong A, Yang C S. A vector space model for automatic indexing [J]. Communications of the ACM, 1975, 18(11):613-620.
- [17] Dumais S T. Latent semantic analysis [J]. Annual review of information science and technology, 2004, 38(1): 188-230.



- [18] Tirilly P, Claveau V, Gros P. Language modeling for bag-of-visual words image categorization [ C ]// Proceedings of the 2008 international conference on Content-based image and video retrieval. New York: ACM, 2008;249-258.
- [19] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[ C ]// Fernando Pereira; Advances in neural information processing systems. Lake Tahoe;2012;1097-1105.
- [20] de Herrera A G S, Kalpathy-Cramer J, Fushman D D, et al. Overview of the ImageCLEF 2013 medical tasks [ C ]// Working notes of CLEF. CLEF, 2013;1-15.
- [21] Bird S, Klein E, Loper E. Natural language processing with Python[ M ]. Beijing: O'Reilly Media, Inc., 2009; 111-112.
- [22] Van Rijsbergen C J, Robertson S E, Porter M F. New models in probabilistic information retrieval [ M ]. London: Computer Laboratory, University of Cambridge, 1980;98-102.
- [23] Rehurek R, Sojka P. Software framework for topic modelling with large corpora[ C ]// Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks. Valletta; LREC, 2010;46-50.
- [24] Vedaldi A, Fulkerson B. VLFeat: An open and portable library of computer vision algorithms[ C ]// Proceedings of the international conference on Multimedia. New York: ACM, 2010;1469-1472.
- [25] Lowe D G. Distinctive image features from scale-invariant keypoints[ J ]. International journal of computer vision, 2004, 60(2):91-110.
- [26] Sermanet P, Eigen D, Zhang X, et al. Overfeat: Integrated recognition, localization and detection using convolutional networks [ J/OL ]. <http://arxiv.org/abs/1312.6229>, 2013-9-21/2016-5-15.
- [27] Chang C C, Lin C J. LIBSVM: a library for support vector machines [ J ]. ACM Transactions on Intelligent Systems and Technology (TIST), 2011, 2(3):27.

(责任编辑 车尧)