

# 时空域深度卷积神经网络及其在行为识别上的应用

刘 琮 许维胜 吴启迪

(同济大学电子与信息工程学院 上海 201804)

**摘 要** 近年来深度卷积神经网络在静态图像识别上取得了较大进展,但在行为视频上建模运动信息的能力较弱。但是,运动信息是行为识别区别于静态图像识别的关键。基于滤波器响应积提出了时空域深度卷积神经网络。该网络先将相邻帧对应的卷积核分为两组,近似地形成傅里叶基函数对,后续的乘法层将不同帧产生的响应两两相乘后再输入加法层求和,从而将相邻帧映射到变换矩阵的特征值对应的不变子空间上,依靠相邻帧在不变子空间上的旋转角度检测它们之间的运动特征。理论分析证明,网络既对运动敏感,又对内容敏感。实验表明,该网络能对行为视频做出更准确的分类,并与近年出现的其他 6 种算法进行比较,结果体现了本算法的优越性。

**关键词** 时空域,卷积神经网络,深度学习,动作特征,行为识别

**中图法分类号** TP301 **文献标识码** A **DOI** 10.11896/j.issn.1002-137X.2015.7.052

## Spatiotemporal Convolutional Neural Networks and its Application in Action Recognition

LIU Cong XU Wei-sheng WU Qi-di

(School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

**Abstract** The key thing that distinguishes action recognition from other recognition tasks is to encode motion explicitly. But, so far, most works based on convolutional neural networks (CNN) cannot properly handle the spatiotemporal interaction in video. We developed a spatiotemporal-CNN that explicitly exploits this important cue provided by video. Instead of summing filter responses, responses are multiplied and our approach is based on that. Specifically, the spatiotemporal-CNN divides convolutional kernels into two groups forming sinusoidals of Fourier Transform. Then, the responses of convolutional kernels are multiplied by multiplicative layer as calculating covariance and the outputs are put into sum layer. In this way, the inputs and adjacent frames are mapped into the subspaces spanned by the eigenvectors, and the special geometric transformations or motion features can be checked by the rotating angles in that space. Additional theoretical analysis proves that spatiotemporal-CNN is sensitive to both motion and content. The experiment shows that our approach produces more accurate classification than current algorithms.

**Keywords** Spatiotemporal, Convolutional neural networks, Deep learning, Motion feature, Action recognition

## 1 引言

从视频中学习人类行为是一项有挑战性的机器视觉任务,近年来受到了研究人员的关注<sup>[1]</sup>。静态图像识别只需要对一张图像中的静态特征进行学习,行为识别则需要对视频中帧与帧之间的运动特征进行学习。视频识别不仅需要考虑空间域(视频的单帧、静态图像)上相邻像素之间的相互关系,还需要考虑时间域上相邻帧之间的相互关系,加之视频的数据量比单个图像大得多,所以算法复杂性和时间复杂性都较高,行为识别的困难性来源于此。当前,尚无一种较为完善的算法能够有效、鲁棒地将人类行为视频准确分类。目前,基于手工设计特征的传统算法<sup>[2]</sup>采用的处理流程是:首先使用稀疏兴趣点(sparse interest points)检测算法在视频帧中确定有可能出现运动的位置;然后使用人工设计的各种特征描述子(如 HOG、HOF 等)在视频帧间抽取运动特征,接着使用 k-means 算法将上一步取得的特征进行矢量化,而后分别在时

域和空域上进行池化(pooling),从而得到用于分类的特征;最后将特征输入各种分类器(如 SVM、神经网络等)中对行为视频进行分类。然而,这种基于人工设计的特征与计算神经学理论相悖,因为它依赖于特定数据集,而人类大脑能自主地从大量无标签数据中学习得到特征。于是,近年来在机器学习领域出现了以深度学习为突破点的纯数据驱动的特征学习算法<sup>[3]</sup>。深度学习算法与传统的浅层学习算法(如独立分量分析(ICA)、等度规映射(ISOMAP)、拉普拉斯特征映射(Laplacian Eigenmaps)<sup>[4]</sup>等)不同,它舍弃了依靠人工精心设计特征的方法,通过构建深度神经网络(拥有数十隐层、数千万甚至过亿的网络参数)来自主地学习特征。卷积神经网络(Convolutional Neural Networks, CNN)又引入了模拟人脑视觉处理机制的网络结构(卷积层、池化层间隔叠加),自 2012 年以来连续刷新了多个静态图像识别基准测试集识别率的新纪录(例如 ImageNet<sup>[5]</sup>)。所以,研究人员最近也尝试将 CNN 应用到行为识别上<sup>[6]</sup>,但是它的表现并不令人满意,尤其是与人

到稿日期:2014-09-12 修稿日期:2014-12-22

刘 琮(1982-),男,博士生,主要研究方向为机器视觉和机器学习,E-mail:lizhliu\_zhong@163.com;许维胜(1966-),男,博士生导师,主要研究方向为智能计算;吴启迪(1947-),女,教授,博士生导师,主要研究方向为智能计算。

工设计特征的传统算法相比较。

朱旭东等<sup>[7]</sup>提出了基于主题隐马尔科夫模型的人体异常行为识别方法,但该方法建立的低层人体运动特征表示仍是人工设计的。文献[8]用一种三维 CNN 来处理运动视频,三维是指空域上的两维和时域上的一维,通过同时学习时空域相关性来获得帧之间的关联特征,但三维卷积能否真正学习到运动特征仍不明确,三维 CNN 依然要靠输入间接的光流特征而不是直接输入帧来引入运动信息。文献[9]将缓慢特征分析(slow feature analysis, SFA)引入 CNN,此网络的目标函数偏好帧之间缓慢变化信息,惩罚快速变化信息。所以, SFA 可能更适合于识别目标的静态信息。例如,两个在空中做抛物线运动的足球和篮球,它们的运动轨迹是快速变化的, SFA 可能会基于静态信息把足球和篮球分为两类,而基于行为识别的分类标准,更应将它们都视为做抛物线运动的一类。文献[10]建立了一个拥有一百万 youtube 视频的行为识别数据集,并在其上应用了 3 种不同结构的 CNN,通过将多帧图像同时输入以上网络来学习帧之间的运动特征。但是,他们的实验结果却表明,不管哪种类型的 CNN,其在单帧图像上的识别率和在多帧图像上的识别率差别不大。这说明了传统的深度卷积神经网络并不能适应学习运动特征的需求。

针对以上算法的不足,本文设计了一种基于 CNN 的网络结构来显式地建模时空域上的行为信息。此网络可以同时学习运动信息和静态图像内容信息,本文将其称为时空域深度卷积神经网络(Spatiotemporal-CNN)。

## 2 时空域深度卷积神经网络

当三维空间里的人体运动投影到摄像机的二维底片上时,三维运动信息会被转换为相邻帧之间的二维图像变换(image warp)关系,我们期望通过捕捉二维上的变换关系来间接地学习三维空间里的运动信息。对于相邻两帧  $X$  和  $Y$ ,它们之间存在变换关系  $Y=LX$ 。所有变换矩阵  $L$  都可用正交矩阵来近似<sup>[11]</sup>,且有特征分解  $L=UDU^T$ ,  $U$  和  $D$  中的元素都是复数,代入上式得  $U^T Y=DU^T x$ 。因为不同变换关系的  $L$  都有相同的  $U^{[11]}$ ,所以不同的变换关系可由  $D$  来唯一确定。又因为离散傅里叶变换的基函数也是复数,所以可以考虑把  $X$  和  $Y$  分别投影到傅里叶基函数对(使用欧拉公式将复数写成  $\cos$  和  $\sin$  函数对的形式)上来确定  $D$ 。据此,可以设计时空域深度卷积神经网络的结构:首先将每个帧对应的卷积核(或滤波器)分为二组,经过训练,组间会近似地构成傅里叶基函数对,此层产生的滤波器响应又称作特征映射(feature maps),后续层将对应不同帧的特征映射两两求积,再两两求和,从而得到能同时建模运动信息和静态图像内容的特征,最后将得到的特征进行空域池化,以便减少图像形变带来的噪音。下面先给出时空域深度卷积神经网络的详细结构,再给出这种网络结构能同时学习到运动和内容特征的证明。

### 2.1 网络结构

虽然 CNN 不能很好地对运动信息进行建模,但它仍有很多优点,比如:能保持空域上相邻像素之间的二维关联结构、能处理大尺寸的图片(例如  $200 \times 200$ )、能通过权值共享减少参数的总量从而减少训练时的计算量等。时空域深度卷积神经网络的结构继承了 CNN 的优点,但又有许多不同,如图 1 所示。

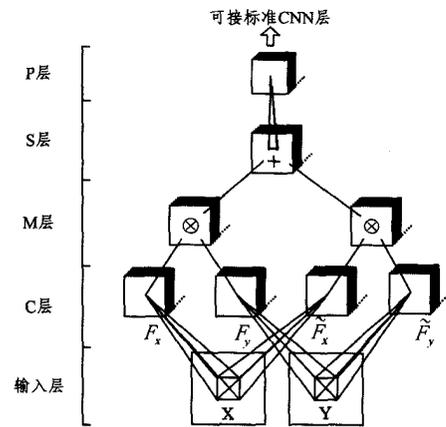


图 1 时空域深度卷积神经网络

时空域深度卷积神经网络由 5 层组成。输入层使用两个相邻帧  $X$  和  $Y$  作为整个网络的输入,  $X$  和  $Y$  是矩阵,网络要能捕捉到它们之间的变换关系。C 层是卷积层,由 4 组卷积核的响应(即特征映射)组成。如果将每组中的一个卷积核写作矩阵形式:  $F_x, \tilde{F}_x, F_y$  和  $\tilde{F}_y$ , 则经过训练后  $F_x$  和  $\tilde{F}_x, F_y$  和  $\tilde{F}_y$  之间会自动形成 2 对傅里叶基函数对,此层输出的 4 个特征映射可表示为:  $(X * F_x), (X * \tilde{F}_x), (Y * F_y)$  和  $(Y * \tilde{F}_y)$ , 其中  $*$  表示卷积操作。M 层是乘积层,用来计算输入的相邻帧对应的两个特征映射之间的元素积(element-wise product),即  $(X * F_x) \otimes (Y * F_y)$  和  $(X * \tilde{F}_x) \otimes (Y * \tilde{F}_y)$ ,  $\otimes$  表示元素乘,将 M 层的输出称为积映射。S 层是加法层,计算两个积映射的元素和,即  $(X * F_x) \otimes (Y * F_y) + (X * \tilde{F}_x) \otimes (Y * \tilde{F}_y)$ , 这里的  $+$  表示元素求和,称 S 层输出为和映射。P 层是与传统深度卷积网络类似的池化层,取和映射上相邻空域的最大值(可能还要再经过一个线性变换)来代表此相邻域,得到比和映射空域大小要小得多的输出。池化层的作用是减轻微小形变对特征的影响并降低计算难度。在 C 层和 P 层的线性映射参数中还可以添加偏置参数,从而用仿射来代替线性映射,但为了保持推导的简洁性,省去了它在公式中的表达。S 层还可以使用传统神经网络的非线性激活函数(activation function): sigmoid 或者 tanh。

### 2.2 和映射上的节点值

将 2D 离散卷积操作转换成向量与矩阵相乘的形式,以便利用 2.3 节的结论。在时空域深度卷积神经网络中,相邻两帧输入网络后会在 S 层上得到多个和映射响应。考虑某个和映射上的某个节点  $s_k$ , 它的值可计算如下:

$$s_k = (X * F_x)_{ij} \otimes (Y * F_y)_{ij} + (X * \tilde{F}_x)_{ij} \otimes (Y * \tilde{F}_y)_{ij} \quad (1)$$

其中,  $i, j$  用来索引对节点  $s_k$  的值有贡献的图像  $X$  或  $Y$  的空域范围。为方便起见,定义一个将矩阵映射为向量的函数  $vec(\cdot)$ , 此函数按照列顺序将输入矩阵  $\cdot$  的列连接成一个向量输出。2D 离散卷积操作可用一个特殊的块循环矩阵(block circulant matrix)来实现,例如:卷积操作  $X * \tilde{F}_x$  可以写作  $vec(X * \tilde{F}_x) = A_x vec(X)$ , 其中  $A_x$  是一个块循环矩阵,用来实现  $X$  和  $\tilde{F}_x$  之间的二维卷积,  $\tilde{F}_x$  是由原卷积核  $F_x$  在周边增补 0 值直到和图像  $X$  同样大小。注意,卷积操作一般有两个类型:循环卷积(circular convolution)和可用卷积(valid convolution),这里考虑的是循环卷积,但结论对可用卷积也近似成立。式(1)可重新写作:

$$s_k = (f_x^T x) (f_y^T y) + (\tilde{f}_x^T x) (\tilde{f}_y^T y) \quad (2)$$

其中,  $x = \text{vec}(X)$ ,  $y = \text{vec}(Y)$ ,  $f_x^T, f_y^T, \tilde{f}_x^T$  和  $\tilde{f}_y^T$  分别是对应的块循环矩阵  $F_x, \tilde{F}_x, F_y, \tilde{F}_y$  中的一行。所以, 卷积可以看作在输入图像的一定空域范围内做线性变换。

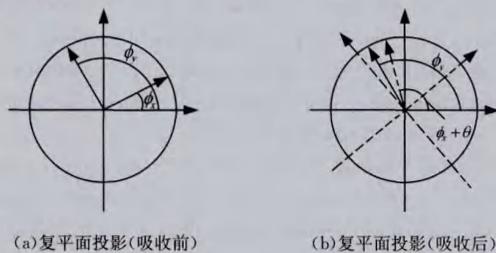
### 2.3 不变子空间上的旋转角度检测子

先在图像变换的一种——平移变换上证明结论, 再推广到所有变换。对于一对相邻帧之间的几何变换, 可用变换矩阵  $L$  表示为  $y = Lx$ ,  $x$  和  $y$  是向量。对于平移变换而言, 它的变换矩阵是正交的, 且在次对角线上的值为 1, 其他各处均为 0。这种矩阵的特征分解  $L = UDU^T$  有如下特性<sup>[12]</sup>: 每一个特征值都是模为 1 的复数, 其对应的特征向量也是复数, 且对应于傅里叶变换中的相位差为  $90^\circ$  的傅里叶基函数对 (cos 和 sin 函数), 而一对傅里叶基函数可支撑一个针对特定特征值的不变子空间。又因为乘以一个模为 1 的复数相当于在复平面进行一个模不变的旋转, 所以将图像  $x$  映射为图像  $y$  可以分解为 3 步: 首先将  $x$  投影到每个特征值确定的不变子空间里, 再根据特征值在每个不变子空间里做旋转, 最后再映射回图像空间得到  $y$ 。

文献[11]证明了自然视频中的大多数的图像变换都可以用正交变换矩阵  $L$  表示, 而且它们对应的不变子空间是相同的。所以, 正交变换矩阵是由特征值唯一确定的, 两个正交变换矩阵之间唯一的差异是它们在不变子空间上旋转的角度不同。为了获得相邻帧的几何变换信息, 只需要将两帧分别投影到不变子空间里, 再读出旋转角度差即可。也即, 考虑某个特征值  $v$  的实部  $v_R$  和虚部  $v_I$ ,  $v = v_R + iv_I$ , 则  $x$  和  $y$  在不变子空间上的投影为  $v_R^T x, v_I^T x$  和  $v_R^T y, v_I^T y$ 。如果再将投影的结果归一化, 则两个投影之间的角度差的余弦值为

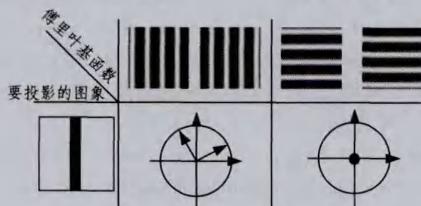
$$\cos(\phi_y - \phi_x) = (v_R^T x)(v_R^T y) + (v_I^T x)(v_I^T y) \quad (3)$$

其中,  $\phi_x$  和  $\phi_y$  是  $x$  和  $y$  在复平面上的投影与实轴之间的角度, 参见图 2(a)。



(a) 复平面投影(吸收前)

(b) 复平面投影(吸收后)



(c) 复平面投影对图像内容敏感

图 2 不变子空间里的旋转角度检测子

进一步, 还可以把特征值吸收进特征向量中, 定义  $v^\theta = v(\cos\theta + i\sin\theta)$ , 其中  $v$  是任意特征向量,  $v^\theta$  是吸收了旋转角度的复数滤波器, 则可以定义旋转角度检测子  $r^\theta$ :

$$r^\theta = (v_R^T x)(v_R^T y) + (v_I^T x)(v_I^T y) \quad (4)$$

当  $\phi_y - \phi_x = \theta$  时,  $r^\theta$  最大, 参见图 2(b)。所以, 判断图像变换是否相同, 就可以用  $r^\theta$  的最大值来检测。注意到式(4)和式(2)是等价的, 所以和映射上的一个节点  $s_k$  其实就是在不变子空间的旋转角度检测子  $r^\theta$ 。

### 2.4 旋转角度检测子的图像内容敏感性

因为每个  $s_k$  都是定义在特定的傅里叶基函数对上的, 当某帧的图像内容缺少此傅里叶基函数对时, 该帧在不变子空间中的投影的模是 0。例如: 图 2(c)中只有竖向条纹的图像(第 2 行第 1 列), 投影到竖直方向的 2D 傅里叶基函数对(第 1 行第 2 列)组成的不变子空间时模不为 0(第 2 行第 2 列), 而投影到水平方向的傅里叶基函数对(第 1 行第 3 列)时模为 0(第 2 行第 3 列)。也就是说, 同一种变换并不能每次都激活  $s_k$ 。这在视觉系统中被称为光圈问题(aperture problem)。每个和映射都有着不同的旋转角度检测子和对应的傅里叶基函数对, 所以对图像变换和图像内容都具有敏感性。而这恰是行为识别所需要的, 因为图像变换是组成运动信息的关键, 而某些特定运动与图像的静态内容密切相关。所以和映射通过卷积操作对相邻帧的图像进行空域上的扫描, 响应大的位置说明有特定的运动特征和内容特征出现。

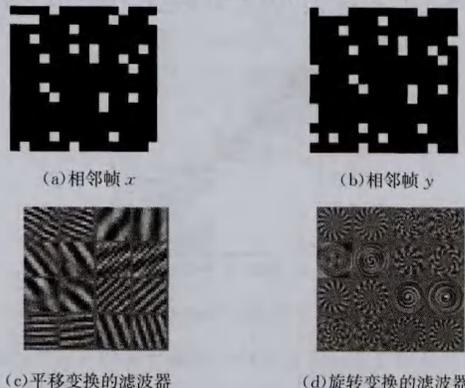
## 3 实验和讨论

为了测试时空域深度卷积神经网络的性能, 基于 Python 的深度学习库 Theano 实现了基于 GPU 并行加速的算法代码, 并在 1 个合成的数据集和 1 个行为识别数据集上验证了算法的有效性。

有别于手工设计特征的传统算法<sup>[12]</sup>, 实验并没有采用先将时空域特征矢量化, 再用支持向量机进行分类的流程, 而是先在 P 层后面接上传统卷积神经网络的卷积层和池化层以使用深度建模来代替矢量化建模; 再接上全连接层, 此层的每个节点都与上层的全部输出映射相连接, 以便将网络学到的时空域特征映射为一个一维特征向量; 最后, 再接上 softmax 分类层, 此层的每个节点与上层的所有节点相连接, 节点的数量等于要分类的类数目, 每个节点的输出值等于该类的概率。在网络的训练阶段, 行为视频中的每 5 对相邻帧被同时输入。为了在特征抽取阶段学习更能代表每个分类的特征, 使用了基于有监督训练的反向传播算法来学习网络的参数(权重和偏置), 使用交叉熵作为目标函数, 采用基于小批量(mini-batch)策略的随机梯度下降算法来更新参数, 并对相邻帧组成的数据集进行随机排序, 以保证每次迭代时采用的小批量数据能达到类均衡。

### 3.1 合成数据集

为了定性地验证时空域深度卷积神经网络对运动信息的建模能力, 人工合成了一个相邻帧(两个图像)。因为像素点是随机生成的, 所以图像都不含静态内容信息, 但是两个图像之间存在变换, 例如, 图 3(a)和(b)之间存在平移运动。



(a) 相邻帧  $x$

(b) 相邻帧  $y$

(c) 平移变换的滤波器

(d) 旋转变换的滤波器

图 3 合成数据集

使用时空域深度卷积神经网络训练后得到的卷积层滤波器可以被可视化,图3(c)、(d)显示了16个随机挑选的滤波器可以清楚地看到图中每两个滤波器构成一个相位差为 $90^\circ$ 的傅里叶基函数对,如图3(c)中的第一行第一个和第二个滤波器。还可以发现旋转变换的滤波器为了提取旋转运动信息而呈现出一种花环状。平移和旋转两种运动有着完全不同的滤波器,说明了本文的时空域深度卷积神经网络的确学到了运动信息。

### 3.2 KTH 数据集

在人类行为识别领域,KTH数据集是使用较为广泛的测试集。由4种场景下的25个人分别作出6种动作:走、慢跑、跑、拳击、挥手、拍手,共录制了2391个短视频。使用其中8个人的视频作为训练集,8个人的视频作为验证集,9个人的视频作为测试集。

图4(a)、(b)显示了其中的两个动作:拳击、慢跑。时空域深度卷积神经网络在此实验中的各项参数为:输入视频保持25fps的帧率,视频尺寸保持 $160 \times 120$ ,C层有64个尺寸为 $18 \times 18$ 的卷积核,第二个卷积层有256个大小为 $9 \times 9$ 的卷积核,全连接层有500个节点,小批量数据集的容量为128个相邻帧,其他超参数用交叉验证法确定。

图4(c)、(d)分别显示了训练后图像 $x$ 和图像 $y$ 对应的16个随机选择的滤波器,可以发现自然数据集上,时空域深度卷积神经网络学习到的特征具有Gabor滤波器的特性,即不同的滤波器对于位置、频率、方向和相位有着不同的选择性。

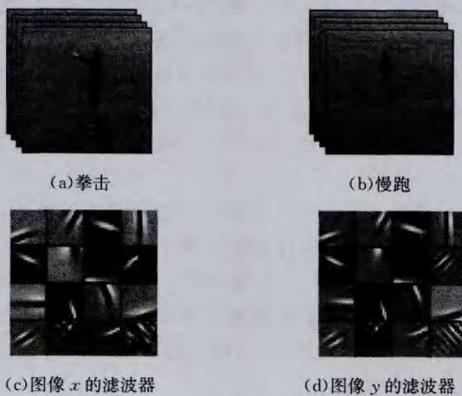


图4 KTH数据集和滤波器

图5显示了网络在KTH数据集上的混淆矩阵(Confusion matrix),行代表正确的类别,列代表算法的分类结果。从图中可以发现总体识别率较好,在挥手、拍手、拳击3类上的错误率比走、慢跑、跑3类稍高,而这也与我们的直觉相似,因为人类要正确判断这3个类也稍显困难。

走	91	4	2	0	0	0
慢跑	2	98	0	0	0	0
跑	1	2	97	0	0	0
拳击	0	0	0	90	7	3
拍手	0	0	0	0	98	2
挥手	0	0	0	4	3	93
	走	慢跑	跑	拳击	拍手	挥手

图5 KTH数据集的混淆矩阵

图6是时空域深度卷积神经网络与其他6种算法的平均正确率的比较,可以看出时空域深度卷积神经网络在6类上的平均正确率为95.1%,优于基于深度卷积神经网络的

GRBM<sup>[11]</sup>、深度神经网络的ISA MODEL<sup>[14]</sup>,也优于另外3个手工设计特征的传统算法:HOG 3D<sup>[12]</sup>、HOG/HOF<sup>[12]</sup>、BoW-MKL<sup>[15]</sup>。

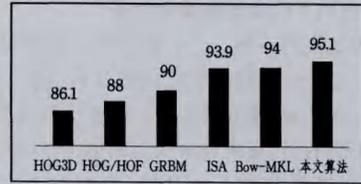


图6 6种算法的正确率

**结束语** 本文提出时空域深度卷积神经网络,将类似于傅里叶基函数滤波器的卷积核引入传统深度卷积神经网络中,通过卷积层分组、乘积层、加法层将相邻帧映射到不变子空间中,从而得到相邻帧之间的运动特征和内容特征。此网络可显性地学习运动特征,克服了原深度卷积神经网络在行为识别上效果差的问题,基于卷积层的特征抽取方法能更好地处理大尺寸图像,还减少了参数,提高了训练效率,有监督训练使学到的特征具有更强的分类判别性。理论分析与实验结果都表明,此网络在基准测试数据集上的正确率较高,可以广泛应用于需要捕捉图像序列间相互关系的视觉任务中,如人类行为识别、双目视觉深度图、视频跟踪等。

### 参考文献

- [1] 胡琼,秦磊,黄庆明. 基于视觉的人体动作识别综述[J]. 计算机学报,2013,36(12):2512-2524  
Hu Qiong, Qin Lei, Huang Qing-ming. A survey on visual human action recognition[J]. Chinese Journal of Computers, 2013, 36(12):2512-2524
- [2] 孔邵颖,郭宏亮. 基于可伸缩语义网络距离的Web多维信息识别算法[J]. 科技通报,2013,29(4):33-35  
Kong Shao-ying, Guo Hong-liang. Web multi-dimensional information identification algorithm based on the stretch of the distance from the semantic network[J]. Bulletin of Science and Technology, 2013, 29(4):33-35
- [3] Bengio Y, Courville A, Vincent P. Representation Learning: A Review and New Perspectives[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8):1798-1828
- [4] Zhao Zeng-shun, Zhang Li, Zhao Meng, et al. Gabor face recognition by multi-channel classifier fusion of supervised kernel manifold learning[J]. Neuro-computing, 2012, 97:398-404
- [5] 范瑞娟,王倩,罗强. 改进DV-HOP输电线路上的WSN节点定位[J]. 计算机仿真,2013,30(9):131-134  
Fan Rui-juan, Wang Qian, Luo Qiang. Improved DV-HOP algorithm to WSN node localization on transmission line[J]. Computer Simulation, 2013, 30(9):131-134
- [6] 郑胤,陈权崎,章毓晋. 深度学习及其在目标和行为识别中的新进展[J]. 中国图像图形学报,2014,19(2):175-184  
Zheng Yin, Chen Quan-qi, Zhang Yu-jin. Deep learning and its new progress in object and behavior recognition[J]. Journal of Image and Graphics, 2014, 19(2):175-184
- [7] 朱旭东,刘志镜. 基于主题隐马尔科夫模型的人体异常行为识别[J]. 计算机科学,2012,39(3):251-255  
Zhu Xu-dong, Liu Zhi-jing. Human abnormal behavior recognition based on topic hidden markov model[J]. Computer Science, 2012, 39(3):251-255
- [8] Ji S, Xu W, Yang M, et al. 3D convolutional neural networks for

human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231

[9] Mobahi, Hossein, Collobert R, et al. Deep learning from temporal coherence in video[C]//Proceedings of the 26th Annual International Conference on Machine Learning. ACM, 2009; 737-744

[10] Karpathy, Andrej, et al. Large-scale video classification with convolutional neural networks[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2014

[11] Memisevic R. Learning to relate images[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1829-1846

[12] Pollock D S G. Circulant matrices and time-series analysis[J].

International Journal of Mathematical Education in Science and Technology, 2002, 33(2): 213-230

[13] Taylor G W, Fergus R, Lecun Y, et al. Convolutional learning of spatio-temporal features[M]//Computer Vision--ECCV 2010. Springer, 2010; 140-153

[14] Le Q V, Zou W Y, Yeung S Y, et al. Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis[C]//2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2011; 3361-3368

[15] Bregonzio M, Xiang T, Gong S. Fusing appearance and distribution information of interest points for action recognition[J]. Pattern Recognition, 2012, 45(3): 1220-1234

(上接第 239 页)

如图 5 所示, 30 组查询语句在输出结果数量分别为 top-5、top-10、top-20、top-50 时, 3 种不同模型的平均 NDCG 值各有差异, 各个模型的参数均被设为最优值, 如本模型中权重参数  $\alpha$  取 0.9,  $\lambda$  取 0.25; LMs 模型中  $\beta$  取 0.9; BANKS 系统中  $\gamma$  取 1。从图 5 中得出, 应用相同的现实世界数据集, 本文中 SimLM 方法的 NDCG 值明显高于其他模型及系统, 排序效果较为优异, 例如在 IMDb 数据集条件下, 进行单侧配对  $T$  测试 ( $p$  值 = 0.047), SimLM 方法的 NDCG 统计值比 BANKS 系统高 17% 以上。

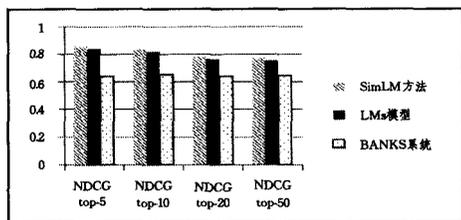


图 5 相同数据集的平均 NDCG 值

**结束语** 关键词查询是 RDF 图数据管理的重要组成部分。针对传统查询模型效率低下、结果冗余与偏离等问题, 本文提出一种基于迭代的图查询算法 ISGR 对所查询关键词进行子图匹配, 在降低时间复杂度的同时, 能保证得到的每个结果子图唯一且最大; 在结果排序阶段, 提出一种结合 RDF 图数据结构和内容性的综合排序方法 SimLM, 考虑关键词图与结果子图之间的结构信息, 并借助统计语言模型 SLM 对结果子图进行综合打分, 按分数降序输出 Top- $k$  个的结果。实验结果表明, 本模型能够较快速、有效地完成用户对数据中的所有关键词 (包括实体和关系名) 的查询需求。通过对比, SimLM 方法的结果可信度符合实际需求, 排序效果明显优于 LM 模型及 BANKS 系统。本文的后续工作考虑将 RDF 图的关键词查询和元组模式查询相结合, 并设计统一框架来解决针对 RDF 图数据的查询问题。

### 参 考 文 献

[1] 李江华, 时鹏, 胡长军. 本体搜索与排序方法研究综述[J]. 小型微型计算机系统, 2013, 34(10): 2396-2406  
Li Jiang-hua, Shi Peng, Hu Chang-jun. Overview of ontology search and ranking[J]. Journal of Chinese Computer Systems, 2013, 34(10): 2396-2406

[2] 杜方, 陈跃国, 杜小勇. RDF 数据查询处理技术综述[J]. 软件学报, 2013, 24(6): 1222-1242

Du Fang, Chen Yue-guo, Du Xiao-yong. Survey of RDF query processing techniques[J]. Journal of Software, 2013, 24(6): 1222-1242

[3] Tran T, Wang H, Rudolph S, et al. Top-k exploration of query candidates for efficient keyword search on graph-shaped (rdf) data[C]//IEEE 25th International Conference on Data Engineering (ICDE'09). IEEE, 2009; 405-416

[4] Ladwig G, Tran T. Combining query translation with query answering for efficient keyword search[M]//The Semantic Web: Research and Applications. Springer Berlin Heidelberg, 2010; 288-303

[5] Bhalotia G, Hulgeri A, Nakhe C, et al. Keyword searching and browsing in databases using BANKS[C]//Proceedings of 18th International Conference on Data Engineering, 2002. IEEE, 2002; 431-440

[6] Zou L, Özsu M T, Chen L, et al. gStore: a graph-based SPARQL query engine[J]. The VLDB Journal, 2013, 23(4): 565-590

[7] Nie Z, Ma Y, Shi S, et al. Web object retrieval[C]//Proceedings of the 16th international conference on World Wide Web. ACM, 2007; 81-90

[8] Elbassuoni S, Ramanath M, Schenkel R, et al. Language-model-based ranking for queries on RDF-graphs[C]//Proceedings of 18th ACM Conference on Information and Knowledge Management. ACM, 2009; 977-986

[9] Elbassuoni S, Blanco R. Keyword search over RDF graphs[C]//Proceedings of 20th ACM International Conference on Information and Knowledge Management. ACM, 2011; 237-242

[10] Wernicke S. A faster algorithm for detecting network motifs[M]//Algorithms in Bioinformatics. Springer Berlin Heidelberg, 2005; 165-177

[11] Zager L A, Verghese G C. Graph similarity scoring and matching[J]. Applied Mathematics Letters, 2008, 21(1): 86-94

[12] Kleinberg J M. Authoritative sources in a hyperlinked environment[J]. J ACM, 1999, 46(5): 614-632

[13] Blondel V D, Gajardo A, Heymans M, et al. A measure of similarity between graph vertices: Applications to synonym extraction and web searching[J]. SIAM review, 2004, 46(4): 647-666

[14] Wallenstein S, Zucker C L, Fleiss J L. Some statistical methods useful in circulation research[J]. Circulation Research, 1980, 47(1): 1-9

[15] Järvelin K, Kekäläinen J. IR evaluation methods for retrieving highly relevant documents[C]//Proceedings of 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2000; 41-48