DOI：10.11830/ISSN.1000-5013.201508047

深度学习与一致性表示空间学习的跨媒体检索

邹辉，杜吉祥，翟传敏，王靖

（华侨大学 计算机科学与技术学院，福建 厦门 361021）

摘要： 提出一种基于深度学习与一致性表示空间学习的方法，针对图像与文本 2 种模态，分别采用卷积神经网络模型和潜在狄利克雷分布算法学习图像的深度特征和文档的主题概率分布；通过一个概率模型将两个高度异构的向量空间非线性映射到一个一致性表示空间；采用中心相关性算法计算不同模态信息在此空间的距离. 在 Wikipedia Dataset 上的实验结果表明：在单模态输入检索中，文中方法的平均准确率为 38.43%，相比于其他方法有明显提高.

关键词： 跨模态；跨媒体；深度学习；卷积神经网络；一致性表示空间；中心相关性

中图分类号： TP 391 **文献标志码：** A **文章编号：** 1000-5013(2018)01-0127-06

Cross-Modal Multimedia Retrieval Based Deep Learning and Shared Representation Space Learning

ZOU Hui, DU Jixiang, ZHAI Chuanmin, WANG Jing

(College of Computer Science and Technology, Huaqiao University, Xiamen 361021, China)

Abstract: A new learning method based deep learning and shared representation space learning is proposed in this paper. Using image and text as an example, we learn the deep learning features of images by convolution neural networks, and learn the text topic probability distribution by a latent Dirichlet allocation model respectively. Then nonlinear mapping the two features spaces into a shared presentation space by a probability model. At last, we adopt centered correlation to measure the distance between them. The experimental results in the Wikipedia Dataset show that our approach is better than that of the similar experiments for single mode input retrieval in recent years and its mean average precision reaches 38.43%.

Keywords: cross-modal; cross-media; deep learning; convolution neural networks; shared presentation space; centered correlation

互联网高速发展的今天，多媒体数据量正在日益增多，而事物的知识需要由所有与其相关的各种媒体信息（如文本、声音、视频和图像等）综合起来表示. 在过去的 20 年间，许多学者致力于跨媒体信息检索的研究，取得了不少成果[1-4]. 多种不同模态信息的特征空间之间往往是高度异构的关系. 近年来，更多学者专注于多媒体信息间关联关系的研究[5-6]. Rasiwasia 等[7]提出将典型关联分析（canonical correlation analysis，CCA）用于分析文本特征空间与图像特征空间的相关关系，结合语义分析，提出了语义

收稿日期： 2015-08-26

通信作者： 杜吉祥（1977-），男，教授，博士，主要从事模式识别、数字图像处理的研究. E-mail：jxdu@hqu.edu.cn.

基金项目： 国家自然科学基金资助项目（61673186，61175121）；福建省自然科学基金资助项目（2013J06014）；华侨大学中青年教师科研提升计划项目（ZQN-YX108）

关联匹配算法,但是其采用的SIFT特征无法很好地表达图像丰富的全局内容.深度学习的概念是由Hinton等[8]提出,而卷积神经网络(convolutional neural networks,CNNs)[9]在2012年之后被广泛地用于图像识别、声音识别等领域[10-11],并且取得了很多突破性的成果.潜在狄利克雷分布(latent Dirichlet allocation,LDA)是由Blei等[12]设计的主题模型,被广泛用于文档分类中[13-15],效果显著.因此,本文结合CNN网络、LDA及一个概率模型构建一个图像与文本的媒介空间,在此媒介空间中进行语义匹配(semantic matching,SM),采用Wikipedia Dataset验证提出的方法,并将其与其他的跨媒体检索方法对比.

1 跨媒体检索模型

针对图像与文本两种形式的信息,提出一种新的一致性表达空间学习模型,用于文本对未标注图像的检索,以及图像对相关文章的搜索,模型框架如图1所示.

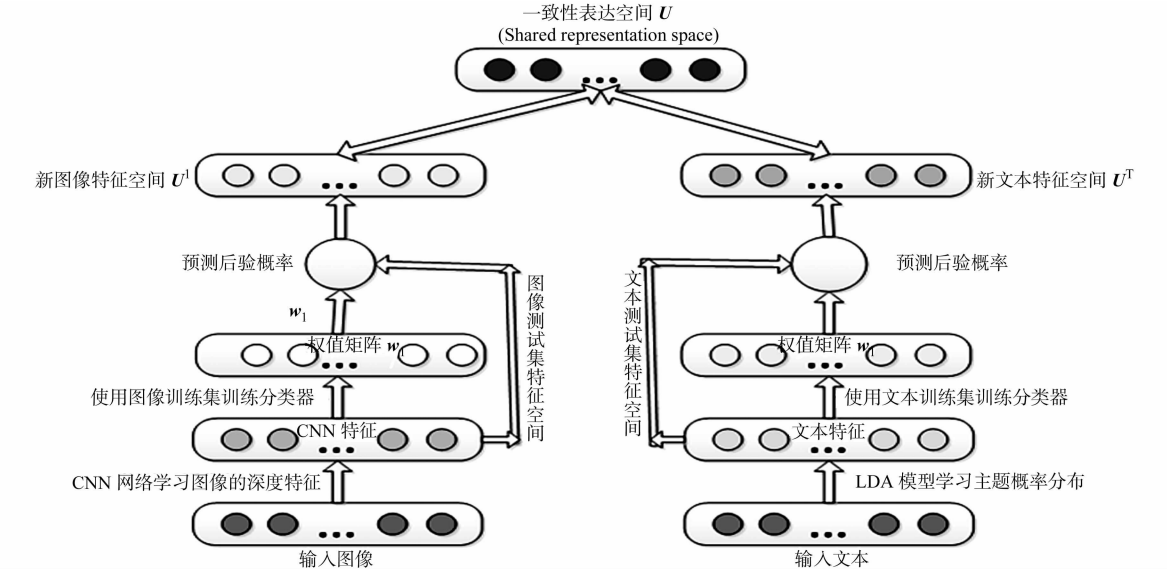

图1 跨媒体检索模型结构

Fig. 1 Cross-media retrieval model

1.1 图像的卷积特征表示

Caffe框架[16]重现了Krizhevsky等[17]提出的CNN模型(bvlc_reference_caffenet.caffe model),针对Wikipedia Dataset,调整该网络模型用于获取图像的CNN特征.CNN的结构,如图2所示.该CNN网络模型有8层,图像数据经过多次卷积、池化和非线性变换后,再经过最后一层softmax层进行分类.卷积层通过权值共享的方法减少训练参数,池化层对卷积的结果进行处理,使之具有平移、旋转及伸缩不变性,并且还起到降维的作用.网络中使用ReLUs(rectified linear units)[18]作为激活函数,即$g(x)=\max(0,x)$.

标准的sigmoid输出不具备稀疏性,而ReLUs是线性修正.Li等[19]证明,训练后的网络完全具备适度的稀疏性,从函数的二维坐标图形看,ReLUs比sigmoid更接近生物学的激活模型.Hinton等[20]提出了一种dropout机制,在训练神经网络时,如果训练样本较少,为了防止过拟合,需要以一定概率将隐含节点清零,提取网络的第6层或第7层的数据$\mathcal{R}^I$表达图像,并用于实验中.

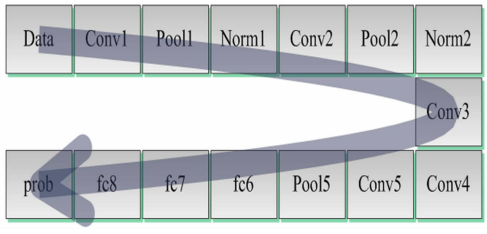

图2 CNN结构

Fig. 2 CNN structure

1.2 文本的主题概率分布

LDA是一种文档主题生成模型,它包含单词、主题和文档3层结构,语料库中的每一篇文档与T个主题的一个多项分布相对应,每个主题又与词汇表中的V个单词的一个多项分布相对应,即

$$D = \{d_1, d_2, \cdots, d_m\},$$
$$d_i = \{w_1, w_2, \cdots, w_n\},$$
$$w_j \in V = \{v_1, v_2, \cdots, v_l\},$$
$$T = \{t_1, t_2, \cdots, t_T\}.$$

式中：D 为文档集，每个文档可看作由多个单词组成的一个序列；词典 V 是由所有不同单词组成的一个集合；T 为主题集. 由每篇文档 d 对应的主题 t 的概率 $p(t|d)$ 和该主题 t 生成单词 w 的概率 $p(w|t)$，可以得到文档中出现单词 w 的概率，即

$$p(w \mid d) = p(w \mid t) \times p(t \mid d).$$

计算每个文档中一个单词在某一个文档中的概率 $p(w|d)$. 然后，根据结果修改该单词应该属于哪个主题. 如果该单词所属的主题改变了，就会反过来影响 $p(t|d)$ 的值. 将文档中所有主题分布的概率值 $p(t|d)$ 作为文本文档的特征描述，文本特征空间表示为 $\mathscr{R}^{\mathrm{T}}$.

1.3　一致性表达空间与相似性度量

获取图像与文本的特征表示后，将 2 种模态特征进行匹配，以完成两者间的相互检索. 给出一个图像（文本）查询 $\boldsymbol{I}_{\mathrm{q}} \in \mathscr{R}^{\mathrm{I}}$（$\boldsymbol{T}_{\mathrm{q}} \in \mathscr{R}^{\mathrm{T}}$），跨媒体系统将返回 $\mathscr{R}^{\mathrm{T}}$（$\mathscr{R}^{\mathrm{I}}$）中与 $\boldsymbol{I}_{\mathrm{q}}$（$\boldsymbol{T}_{\mathrm{q}}$）语义最接近的文本（图像）.

1.3.1　*一致性表达空间*　传统的检索问题一般寻找一个线性映射，即

$$P : \mathscr{R}^{\mathrm{I}} \rightarrow \mathscr{R}^{\mathrm{T}},$$

使得 P 是可逆的. 在跨媒体检索中，由于文本与图像的表示形式往往不一样，在 $\mathscr{R}^{\mathrm{I}}$ 与 $\mathscr{R}^{\mathrm{T}}$ 之间不存在某种自然的对应. 因此，简单的映射或者求最近邻值无法挖掘两个异构空间之间内在的语义关联. 采用一个机制将两个高度异构的特征空间映射到新的空间，即

$$P_{\mathrm{I}} : \mathscr{R}^{\mathrm{I}} \rightarrow \boldsymbol{U}^{\mathrm{I}}, \qquad P_{\mathrm{T}} : \mathscr{R}^{\mathrm{T}} \rightarrow \boldsymbol{U}^{\mathrm{T}},$$

使得 P_{I} 和 P_{T} 都是可逆的非线性映射，$\boldsymbol{U}^{\mathrm{I}}$ 和 $\boldsymbol{U}^{\mathrm{T}}$ 两个新的特征空间是同构的，并且有 $\boldsymbol{U}^{\mathrm{I}} = \boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}$. 通过以上两个映射，将 $\mathscr{R}^{\mathrm{I}}$ 和 $\mathscr{R}^{\mathrm{T}}$ 映射到了一个共享的语义空间 $\boldsymbol{U}$，称此共享空间为一致性表示空间.

1.3.2　*概率模型*　用一个概率模型将两个特征空间映射到同一个特征空间. 语义概念词汇表 $C = \{c_1, c_2, \cdots, c_k\}$ 表示文档的 k 类语义. 用线性分类器分别训练图像与文本的训练集，学习得到相应的权值矩阵 $\boldsymbol{W}_{\mathrm{I}}$ 与 $\boldsymbol{W}_{\mathrm{T}}$，用多类逻辑回归预测图像与文本的测试数据中每一个样本属于类别 r 的概率，即

$$\boldsymbol{P}_{C|X}(r \mid \boldsymbol{x}; \boldsymbol{w}) = \frac{1}{Z(\boldsymbol{x}, \boldsymbol{w})} \exp(\boldsymbol{w}_j^{\mathrm{T}} \boldsymbol{x}).$$

式中：C 表示语义概念（也就是类别标签）；r 表示 k 类中的第 r 类；$\boldsymbol{x}$ 表示 $\boldsymbol{I} \in \mathscr{R}^{\mathrm{I}}$ 和 $\boldsymbol{T} \in \mathscr{R}^{\mathrm{T}}$；$Z(\boldsymbol{x}, \boldsymbol{w}) = \sum_r \exp(\boldsymbol{w}_j^{\mathrm{T}}, \boldsymbol{x})$ 是一个归一化常数. 完成映射的概率公式为

$$P_{\mathrm{I}} : \mathscr{R}^{\mathrm{I}} \rightarrow \boldsymbol{U}^{\mathrm{I}},$$
$$P_{\mathrm{T}} : \mathscr{R}^{\mathrm{T}} \rightarrow \boldsymbol{U}_{\mathrm{T}}.$$

两个映射分别将每个图像特征 $\boldsymbol{I} \in \mathscr{R}^{\mathrm{I}}$ 映射到后验概率向量 $\boldsymbol{P}_{C|\mathrm{I}}(r|\boldsymbol{I})$，将每个文本特征 $\boldsymbol{T} \in \mathscr{R}^{\mathrm{T}}$ 映射到后验概率向量 $\boldsymbol{P}_{C|\mathrm{T}}(r|\boldsymbol{T})$，$r = \{1, 2, 3, \cdots, k\}$，得到图像与文本的语义空间 $\boldsymbol{U}^{\mathrm{I}}$ 和 $\boldsymbol{U}^{\mathrm{T}}$. 两个语义空间是原特征空间更高层次的抽象，并且是同构的，表示的都是语义概念的概率空间. 因此，可以把两个语义空间看成是同一个向量空间（$\boldsymbol{U}^{\mathrm{I}} = \boldsymbol{U}^{\mathrm{T}} = \boldsymbol{U}$），$\boldsymbol{U}$ 即为一致性表达空间. 跨媒体检索实验中，两个模态的相似性比较将在此一致性表示空间度量.

1.3.3　*距离度量*　实验衡量的是两个不同模态特征向量的相似度，而向量的相似度与向量的方向也有关系. 文中采用的距离度量方法是经修正调整后的中心相关性，即

$$d_{i,j} = -\frac{\sum_{i=1}^{m} \sum_{j=1}^{n} (x_i - \bar{x})(y_j - \bar{y})}{n}, \qquad m = n.$$

中心相关性度量方法主要考虑向量 $\boldsymbol{X}$ 与 $\boldsymbol{Y}$ 的线性相关性. 在计算相似度时，先减去向量平均值，再计算两个向量的内积，m 和 n 分别是两个向量的长度. 用相关性的负数表示两个向量的距离，相关性越大，距离 $d_{i,j}$ 就越小.

2 实验结果及分析

采用提出的检索框架进行跨媒体检索实验,验证模型及所采用的距离度量方法的有效性,并与近年同类研究结果作对比,分析实验结果.

2.1 数据集

使用维基百科数据集(Wikipedia Dataset),该数据集是从维基百科的专题文章《Featured Articles》中收集得到的.文章被分成10类,每篇文章分章节,每个章节有一个对应的图片,最终得到包含2 866个文本图像对的文档集.每个文本图像对都标有相应的语义类别,标签包括Art & Architecture,Biology,Geography & Places,History,Literature & Theatre,Media,Music,Royalty & Nobility,Sport & Recreation和Warfare.与文献[7]相同,将数据集划分成2 173个训练样本和693个测试样本.

2.2 caffe微调与CNN特征提取

使用8层结构的caffe模型(bvlc_reference_caffenet.caffe model),将像素大小为256 px×256 px的图像作为模型的输入,使用该CNN网络模型进行训练.由于样本数量比较少,采用Li等[19]提出的dropout机制,将其设置为0.5,在训练样本的时候,以50%的概率将隐含节点清零,防止过拟合.分别提取第6层(fc6)和第7层(fc7)的数据表示图像特征,特征维数为4 096.

2.3 实验结果分析

平均精度均值(mean average precision,MAP)是反映搜索性能的评价指标,其大小与检索效果的排名情况有关,系统检索出来的相关文档越靠前(rank越高),MAP就越高.因此,文中采用MAP作为跨媒体检索算法的评价指标.

2.3.1 *跨媒体检索模型有效性验证* 文中的CNN与概率模型相结合的跨媒体检索结果与Rasiwasia等[7]提出的3种模型对比,距离度量方法使用标准的皮尔逊相关性度量(normalized correlation,NC).

多模态检索结果的MAP评估,如表1所示.由表1可知:文中的CNN与SM相结合模型的检索平均准确率(MAP)比文献[7]使用的3种模型高,验证了提出模型的有效性;相比于人工选择的SIFT特征,采用CNN网络学习得到的深度特征可以更有效地表达图像的抽象概念、描述图像的深层语义;对于采用的多样性较高的Wikipedia Dataset,CNN特征的优势表现得更加明显.

表1 多模态检索结果的MAP评估

Tab.1 MAP evaluation of multimodal retrieval

实验模型	图搜文本	文本搜图	平均值
CCA+SIFT[9]	0.249 0	0.196 0	0.223 0
SM+SIFT[9]	0.225 0	0.223 0	0.224 0
SCM+SIFT[9]	0.277 0	0.226 0	0.252 0
SM+CNN(fc6)	0.401 9	0.315 1	0.358 5
SM+CNN(fc7)	0.398 5	0.323 0	0.360 8

2.3.2 *相似性度量方法的有效性验证* 采用第一范式(L1)、第二范式(L2,欧氏距离)、KL散度(KL-divergence)、标准相关性(NC)、余弦相似度(cos)和中心相关性(centred correlation,CC)等多种算法度量两种模态特征向量间的距离,并对比验证实验采用的CC算法的有效性.

与其他相似度度量方法不同,CC算法在计算相似度时,不仅考虑了两种不同模态特征向量的方向,而且在中心化后消除了指标量纲的影响.不同度量方法对检索结果影响对比,如表2所示.由表2可知:不论是CNN网络第6层的特征还是第7层的特征,CC算法都能计算得到更准确的相似度.

此外,在使用相同距离度量算法的情况下,图像检索相关文本的模式中,CNN网络中第6层的特征能取得较好的结果;而用文本检索相关图像的模式中,第7层的特征表现出更好的检索结果.如果同时考虑图搜文本与文本搜图两种模式网路第6层的特征要比第7层的特征表现的更好些.

2.3.3 *跨媒体检索模型与现有模型实验结果的对比分析* 为了进一步证明文中提出模型的优势,将提出方法与其他跨媒体检索模型进行对比,结果如表3所示.表3中:Random为随机排序的MAP值;SCM是Rasiwasia等[9]提出的模型;MSAE是Wang等[21]使用的模型,采用栈自动编码器学习图像与文本的深度特征;CML2R是Wu等[22]提出的模型,他们将图像与文本特征联合编码为一个共享的特征向量作为两种模态的连接点;TSRtext和SRimg是Ling等[23]提出的方法.

由表3可知:无论是图像搜索相关文本还是文本搜索相关图像,文中提出的跨媒体检索方法比其他

跨媒体检索方法表现出更好的检索结果，充分验证了所设计系统的有效性.

表 2　不同度量方法对检索结果影响对比

Tab. 2　Comparison of retrieval results between different measurement methods

实验模型	距离度量	图搜文本	文本搜图	平均值
SM+CNN (fc6)	L1	0.400 6	0.293 3	0.347 0
	L2	0.388 5	0.279 3	0.333 9
	KL	0.387 5	0.272 9	0.330 2
	NC	0.401 9	0.315 1	0.358 5
	CS	0.399 7	0.325 6	0.362 7
	CC	0.416 9	0.351 7	0.384 3
SM+CNN (fc7)	L1	0.396 9	0.300 2	0.348 6
	L2	0.386 4	0.287 3	0.336 9
	KL	0.383 9	0.272 3	0.328 1
	NC	0.398 5	0.323 0	0.360 8
	CS	0.396 9	0.329 5	0.363 2
	CC	0.408 8	0.351 8	0.380 3

表 3　提出的跨媒体检索方法与现有方法的实验结果对比

Tab. 3　Comparison with similar experiments in recent years

实验模型	图搜文本	文本搜图	平均值
Random	0.118 0	0.118 0	0.118 0
SCM[9]	0.277 0	0.226 0	0.252 0
MSAE[21]	0.187 0	0.179 0	0.183 0
CML2R[22]	0.233 0	0.215 2	0.224 1
TSRtext[23]	0.295 0	0.207 0	0.251 0
TSRimg[23]	0.322 0	0.251 0	0.287 0
SM+CNN(fc6)	0.416 9	0.351 7	0.384 3
SM+CNN(fc7)	0.408 8	0.351 8	0.380 3

3　结束语

针对跨媒体检索中不同模态间的相似度匹配问题，提出了卷积神经网络与语义匹配相结合的方法，得到文本与图像的共享语义特征空间. 在此特征空间中，进行两种模态特征的相似性度量，较大地提高了跨媒体检索结果的平均准确率. 中心相似性距离度量方法也进一步改善了检索结果. 由于 Wikipedia 数据集所收集的数据多样性较大，比较贴近现实中的检索环境. 当前的实验结果还没有达到令人满意的准确率，今后需要努力的方向是进一步研究深度学习，获取更有效的图像与文本的特征表示，建立数据量更大的多媒体数据库，并把跨媒体检索的研究扩展到语音、视频等其他模态数据上.

参考文献：

[1] YANG Yi, XU Dong, NIE Feiping, *et al*. Ranking with local regression and global alignment for cross media retrieval [C]// Proceedings of the 17th International Conference on Multimedia. Vancouver: ACM, 2009: 175-184. DOI: 10.1145/1631272.1631298.

[2] SRIVASTAVA N, SALAKHUTDINOV R R. Multimodal learning with deep boltzmann machines[J]. Journal of Machine Learning Research, 2014, 24(8): 1967-2006.

[3] LU Xinyan, WU Fei, TANG Siliang, *et al*. A low rank structural large margin method for cross-modal ranking[C]// Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval. Dublin: ACM, 2013: 433-442. DOI: 10.1145/2484028.2484039.

[4] WU Fei, LU Xinyan, ZHANG Zhongfei, *et al*. Cross-media semantic representation via bi-directional learning to rank [C]// Proceedings of the 21st ACM International Conference on Multimedia. New York: ACM, 2013: 877-886. DOI: 10.1145/2502081.2502097.

[5] ZHANG Yanyan, LI Guorong, CHU Lingyang, *et al*. Cross-media topic detection: A multi-modality fusion framework[C]// IEEE International Conference on Multimedia and Expo. San Jose: IEEE Press, 2013: 1-6. DOI: 10.1109/ICME.2013.6607487.

[6] LI Liang, JIANG Shuqiang, HUANG Qingming. Learning image vicept description via mixed-norm regularization for large scale semantic image search[C]// IEEE Conference on Computer Vision and Pattern Recognition. Providence RI: IEEE Press, 2011: 825-832. DOI: 10.1109/CVPR.2011.5995570.

[7] RASIWASIA N, COSTA P J, COVIELLO E, *et al*. A new approach to cross-modal multimedia retrieval[C]// Proceedings of the International Conference on Multimedia. Firenze: ACM, 2010: 251-260. DOI: 10.1145/1873951.1873987.

[8] HINTON G E, OSINDERO S, TEH Y W. A fast learning algorithm for deep belief nets[J]. Neural Computation, 2006, 18(7): 1527-1554. DOI: 10.1162/neco.2006.18.7.1527.

[9] RASIWASIA N, COSTA P J, COVIELLO E, *et al*. A new approach to cross-modal multimedia retrieval[C]// Proceedings of the 18th International Conference on Multimedia. Firenze: ACM, 2010: 251-260.

[10] JI Shuiwang, XU Wei, YANG Ming, *et al*. 3D convolutional neural networks for human action recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2013, 35(1): 221-231. DOI: 10.1109/TPAMI.2012.59.

[11] RAZAVIAN A S, AZIZPOUR H, SULLIVAN J, *et al*. CNN features off-the-shelf: An astounding baseline for recognition[C]// IEEE Conference on Computer Vision and Pattern Recognition Workshops. Columbus: IEEE Press, 2014: 512-519. DOI: 10.1109/CVPRW.2014.131.

[12] BLEI D M, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3: 993-1022. DOI: 10.1162/jmlr.2003.3.4-5.993.

[13] ROSEN-ZVI M, GRIFFITHS T, STEYVERS M, *et al*. The author-topic model for authors and documents[C]// Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence. Pittsburgh: AAAI Press, 2004: 487-494.

[14] RAMAGE D, HALL D, NALLAPATI R, *et al*. Labeled LDA: A supervised topic model for credit attribution in multi-labeled corpora[C]// Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing. Singapore: Association for Computational Linguistics, 2009: 248-256. DOI: 10.3115/1699510.1699543.

[15] LIU Yan, NICULESCU-MIZIL A, GRYC W. Topic-link LDA: Joint models of topic and author community[C]// Proceedings of the 26th Annual International Conference on Machine Learning. Quebec: ACM, 2009: 665-672. DOI: 10.1145/1553374.1553460.

[16] JIA Yangqing, SHELHAMER E, DONAHUE J, *et al*. Caffe: Convolutional architecture for fast feature embedding [C]// Proceedings of the ACM International Conference on Multimedia. Orlando: ACM, 2014: 675-678. DOI: 10.1145/2647868.2654889.

[17] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[C]// Advances in Neural Information Processing Systems. South Lake Tahoe: NIPS, 2012: 1097-1105. DOI: 10.1145/3065386.

[18] NAIR V, HINTON G E. Rectified linear units improve restricted boltzmann machines[C]// Proceedings of the 27th International Conference on Machine Learning. Haifa: [s. n.], 2010: 807-814. DOI: 10.1.1.165.6419.

[19] LI Jun, LUO Wei, YANG Jian, *et al*. Why does the unsupervised pretraining encourages moderate-sparseness[C]// The 31st International Conference on Machine Learning. Beijing: [s. n.], 2014: 1-6.

[20] HINTON G E, SRIVASTAVA N, KRIZHEVSKY A, *et al*. Improving neural networks by preventing co-adaptation of feature detectors[J]. ArXiv Preprint ArXiv, 2012, 3(4): 212-223.

[21] WANG Wei, OOI B C, YANG Xiaoyan, *et al*. Effective multi-modal retrieval based on stacked auto-encoders[J]. Proceedings of the VLDB Endowment, 2014, 7(8): 649-660. DOI: 10.14778/2732296.2732301.

[22] WU Fei, JIANG Xinyang, LI Xi, *et al*. Cross-modal learning to rank via latent joint representation[J]. IEEE Transactions on Image Processing, 2015, 24(5): 1497-1509. DOI: 10.1109/TIP.2015.2403240.

[23] LING Li, ZHAI Xiaohua, PENG Yuxin. Tri-space and ranking based heterogeneous similarity measure for cross-media retrieval[C]// 21st International Conference on Pattern Recognition. Ibaraki: IEEE Press, 2012: 230-233.

(**责任编辑**: 黄晓楠 **英文审校**: 吴逢铁)