基于弱监督深度学习的文本聚类算法及应用

谭 敏 张宏源 张海超
(杭州电子科技大学计算机学院 浙江 杭州 310018)

摘 要 围绕基于用户点击数据的文本聚类展开研究。利用点击数据将查询文本表征为图像点击特征图,并在此上训练深度点击模型。为了应对文本噪声,引入可刻画文本可靠性的权重,提出基于弱监督深度学习的文本聚类算法来迭代更新文本权重和深度模型。将该算法应用于基于点击特征的图像识别中,通过合并相似文本,为图像构建紧凑的文本集点击特征向量,实现高效的图像识别。在 Clickture-Dog 和 Clickture-Bird 两个公开点击数据集上进行验证,结果表明:用图像点击特征图来表征查询文本可有效解决原始点击特征向量的稀疏和不连续性,帮助获得优秀识别率;弱监督深度聚类模型不仅帮助学习强大的文本表征,还能有效选择高质量文本数据训练模型,进一步提高性能。

关键词 图像识别 深度聚类 用户点击数据 查询合并 弱监督学习

中图分类号 TP3 **文献标识码** A DOI:10.3969/j.issn.1000-386x.2019.04.027

TEXT CLUSTERING ALGORITHM AND ITS APPLICATION BASED ON WEAKLY-SUPERVISED DEEP LEARNING

Tan Min Zhang Hongyuan Zhang Haichao
(*School of Computer Science and Technology, Hangzhou Dianzi University, Hangzhou 310018, Zhejiang, China*)

Abstract The research is based on the text clustering from user-click data. With click data, a query-text was represented as a smooth image-click-graph, and a deep click model was trained. In order to deal with heavy noise in the clicked query-text set, a weight vector that could measure the reliability of the query-text was introduced, and a text clustering algorithm based on weakly-supervised training method was proposed to iteratively update the weight vector and deep model. The text clustering algorithm was applied to click-feature-based image recognition. After combining similar query-text, a compact click-frequency-vector for images was constructed to achieve accurate image recognition. The proposed method was verified on public Clickture-Dog and Clickture-Bird datasets. The experimental results show that representing each query as an image-click-graph can deal with the non-smoothness and sparseness in the original click vectors, which helps to improve image recognition accuracy. Weakly-supervised deep learning not only helps to learn powerful representations, but also can effectively select queries of high quality, which further improved the recognition performance.

Keywords Image recognition Deep clustering User-click data Query clustering Weakly-supervised learning

0 引 言

图像识别一直是计算机视觉领域中最受关注的问题之一。尽管近年来在相关技术方面有了较大的突破和进展,但是如何克服“语义鸿沟”依然是一个巨大的挑战。为了解决这个问题,近年来一些学者开始使用用户点击数据来代替视觉特征表示图像[1-5]。利用点击数据,一张图片可以被表示为一个文本点击频率向量,即文本点击特征[2]。由于点击数据是从商业搜索

收稿日期:2018-11-28。国家自然科学基金青年基金项目(61602136)。**谭敏**,讲师,主研领域:深度学习。**张宏源**,硕士生。**张海超**,硕士生。

引擎中爬取的用户反馈数据,与传统的视觉特征相比,文本点击特征有更丰富的语义信息,在许多计算机视觉任务上表现更为出色[1-5]。

尽管点击特征有诸多优势,但直接将这种点击特征用于图像识别仍然面临很多的挑战。由于查询文本集的规模庞大,噪声较多,原始的点击特征非常稀疏和冗余。针对此问题,许多学者提出了利用点击特征进行文本合并的方法[3]来应对传统自然语言处理方法中的"语义鸿沟"问题。然而这些工作都是利用图像点击次数向量来表征文本。这种特征尽管简单,但无法刻画文本的层次化的深度语义特征。为此,我们提出利用深度网络学习文本的深度点击特征表达,并基于深度点击特征表达合并相近语义的查询文本。

随着深度模型在视觉分类领域的广泛应用,近年来,学者们也开始研究基于深度学习的图像聚类模型[6-7]。基于此类模型,本文提出了面向点击特征的深度文本聚类框架来合并语义相似的查询文本,其中深度特征和查询类别通过网络自主迭代学习。为了克服点击特征向量的稀疏性,本文提出构建平滑的结构化的点击特征图来表征查询文本,并以此作为深度网络的输入来学习查询文本的深度点击特征。本文将杨等提出的无监督深度聚类框架 JULE[6](Joint Unsupervised LEarning of deep representations and image clusters)扩展到点击数据上,并融合弱监督学习策略对文本进行加权,利用迭代优化交替地学习文本权重和深度点击特征,从而实现在噪声文本数据中的自动样本选择。

1 JULE 模型简介

JULE 是一个端到端的深度图像聚类模型,它通过迭代更新深度图像特征和类别标号实现无监督的图像聚类。与传统的深度图像识别模型相比,该模型不需要精确的图像类别信息,只需要为模型初始化粗糙的类标号。鉴于这些优势,我们将此模型扩展到基于点击数据的文本聚类上,以应对原始查询文本缺乏类别标号的特点。该模型的特点是在训练过程中联合更新图像的聚类结果和深度特征实现完全自主学习。

该模型通过一个三元加权的损失函数组进行训练。实验证明,该模型在许多图像识别数据集中都具有优秀的特征学习能力和图像聚类效果,如 MNIST、USPS、COIL、UMist、FRGC、CMU-PIE、YTF 等。

除了 JULE 外,关于如何将深度学习应用到聚类任务中也有许多其他的研究。如 Dizaji 等提出了 DEPICT 模型,它通过将数据映射到一个具有差异性的子空间来获得更好的聚类效果[7];Tian 等提出了一种简单的深度学习方法来进行图片聚类,该方法首先通过堆叠自动编码器得到图片的视觉特征,然后用 K-means 算法对这些特征进行聚类[8]。

尽管近些年深度聚类的研究工作越来越多,但已有模型都是针对图像数据设计的,而本文研究的查询文本与图像本质上具有较大差距。为此,本文基于光滑性假设,为查询文本构建了点击特征图,从而将 JULE 扩展到文本聚类任务上。此外,本文结合弱监督学习策略提出了可对抗文本噪声的深度聚类网络。

2 算法设计及应用

本文提出了一种基于弱监督深度学习的文本聚类方法来进行查询文本合并,并利用合并后的文本集为图像构建紧凑的点击特征,从而实现高效的图像识别。本文所提出的图像识别算法流程如图 1 所示。在本节中,首先将简介点击数据及对应的图像(文本)点击特征,接着详细介绍基于弱监督深度学习的文本聚类框架,最后介绍算法在图像识别中的具体应用。

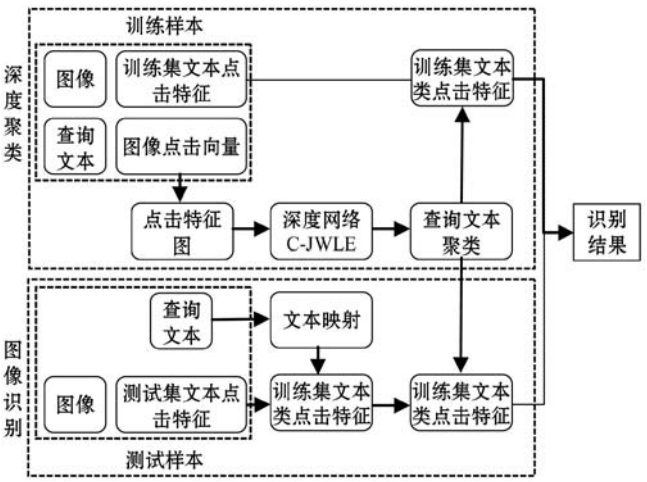

图 1 基于弱监督深度学习的文本聚类与图像识别框架

2.1 点击数据及点击特征向量

假设包含 n 张图片的训练图片集为 $\{x_i \mid 1,2,\cdots,n\}$,图片所对应的类别标签为 $\{y_i \mid i=1,2,\cdots,n\}$。该图像集在一个包含 m 条查询的文本集 $\{q_j \mid j=1,2,\cdots,m\}$ 上有非零的用户点击次数,且相应点击矩阵为 $\boldsymbol{C} \in \boldsymbol{R}^{n\times m}$(其中 $c_{i,j}$ 表示第 i 张图片在查询 j 下的点击次数),每张图片可以用查询文本下的用户点击频率向量来表示。

具体而言,利用点击数据,任意图片可表示为 $\boldsymbol{u}_i=(c_{i,1},c_{i,2},\cdots,c_{i,m})$。类似地,查询文本可表示为 $\boldsymbol{v}_j=$

$(c_{1,j},c_{2,j},\cdots,c_{n,j})$。注意到原始的点击向量 $\boldsymbol{u}_i$ 和 $\boldsymbol{v}_j$ 的特征维度分别由点击数据涉及的图像和查询文本集大小决定,而高维的点击数据容易导致维度灾难。

2.2　基于弱监督深度学习的文本聚类算法

本文将查询文本表征为图像点击特征,并在此上学习它的深度点击特征。

2.2.1　点击特征图的构建

如前文所述,本文将利用深度学习网络学习查询文本的深度点击特征。与文献[1,3,9]中类似,利用用户点击数据,输入的查询文本可表示为图像点击向量。然而,由于互联网图像集庞大,原始的图像点击特征往往过于稀疏。为了解决该特征的不平滑性和稀疏性,本文利用原始图像点击向量,每个查询文本构建了点击特征图 G。

点击特征图的构建流程如图 2 所示。首先将查询文本的原始点击特征转化为图像类点击特征矩阵,再利用视觉相似性将此矩阵转化为平滑的点击特征图。如下将展开介绍这两个过程。

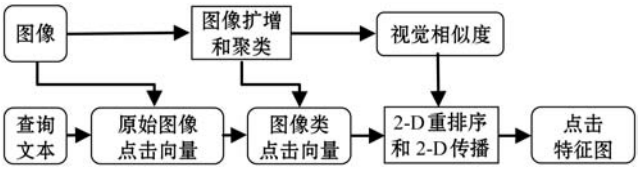

图 2　点击特征图构建流程

1)图像类点击特征矩阵。构建图像类点击特征矩阵要利用到上文所述的点击向量 $\boldsymbol{v}_j$ 及真实标签 y_i。利用类别的真实标签对 $\boldsymbol{v}$ 进行重排列得到矩阵 $(\boldsymbol{M}_j)_i$,使得 $(\boldsymbol{M}_j)_i$ 的每一行对应同一类图像下的点击特征向量。由于 Clickture-Dog 和 Clickture-Bird 数据集类内不平衡,有些种类的图片过少。为了平衡数据,本文首先利用图像扩增算法对图片数量少的类别进行扩充操作。对于每一张图片 x_i,它的扩充图像 L_i 定义如下:

$$L_i=\{\tau(x_i)\mid\tau(\cdot)\in\Gamma(\cdot)\}\tag{1}$$

式中:$\tau(\cdot)$ 是一种图像变换,包括遮挡、加噪、改变颜色及其混合。L 是增强后的数据集,变换后的图片与原始图片共享点击特征。

得到增强过的数据后,本文将每个种类的图片集聚类到 N_I 个子类,这样文本在同一类图像下的点击向量就可以转化为一个维度 N_I 的类点击向量。具体来说,对于第 j 类图片集,实现基于深度视觉特征的聚类,得到对应的子类图像集索引 $\{A_{j,1},A_{j,2},\cdots,A_{j,N_I}\}$。

聚类完成后,更新后的点击特征矩阵定义如下:

$$(M_j)_i=\Big(\sum_{(i,t)\in A_i^1}c_{i,j},\cdots,\sum_{(i,t)\in A_i^{N_I}}c_{i,j}\Big)\tag{2}$$

式中:(i,t) 和下文的 I_i^t 都表示第 i 张图片用了第 t 种图像变换得到的图片。本文将更新后得到的点击矩阵称为图像类点击特征矩阵。

相比于利用原始点击特征构建的点击特征矩阵,经过图像扩增后聚类操作后得到的结构化的类点击特征矩阵有效克服了数据集中的类别不平衡。

2)点击特征图。为了改善图像类点击特征矩阵稀疏不连续的缺点,本文利用排序和传播算法将图像类点击特征矩阵转化为平滑的点击特征图。受到文献[3]启发,本文提出了 2-D 的重排序和 2-D 点击传播算法。该方法将点击量在各图像类和同类不同图像中传播,有效改善了点击矩阵不连续性和稀疏性的问题。

以上两种算法都是基于图像相似度图进行的。相似度图分为类间相似度图和类内相似度图。类间相似度图 S 用来衡量不同图像类间的距离,它定义两个图像集的深度视觉特征的 Hausdorff 距离[12]。类内相似度矩阵 $\widetilde{\boldsymbol{S}}_i$ 用同一类别下不同子类的聚类中心视觉特征来衡量。基于相似度图,进行如下的排序和传播算法:

(1)*点击重排序*　重排序算法是为了增加点击特征图的平滑性,使得相似的大类或者小类在点击矩阵坐标空间中更接近。重排序包括基于 S 类的类间排序和基于 S_i 的类内排序。具体而言,选择任意图像类/图像子类为参考类,根据 S/S_i 进行降序排列,从而得到了排列后的图像类点击特征矩阵 $(\overline{\boldsymbol{M}}_j)_i$,该矩阵在维度空间上依照视觉相似性排列,因而具有类似于图像的局部相似性特性。

(2)*点击传播*　传播算法主要是为了解决点击特征稀疏的问题。通过在相似样本间分享点击量,使得点击特征更加平滑均匀。与重排序过程类似,传播分为类间传播和类内传播两过程。类间传播是指一个图像类的点击量和按照比例分享给其他相似类。

由上文可知,查询 j 的重排后点击特征为 $\overline{\boldsymbol{M}}_j$,将图像类别的点击总量特征记为 μ,则 μ 定义如下:

$$\mu=(\|(\overline{\boldsymbol{M}}_j)_1\|,\|(\overline{\boldsymbol{M}}_j)_2\|,\cdots,\|(\overline{\boldsymbol{M}}_j)_N\|)\tag{3}$$

类间传播的公式如下:

$$\begin{aligned}&\overline{\mu}=\mu(\rho S+(1-\rho)\boldsymbol{E}-\rho\Lambda(S))\\&\Lambda(S)=diag(s_{1,1},s_{2,2},\cdots,s_{N',N'})\\&(\widetilde{\boldsymbol{M}}_j)_i=\frac{\overline{\mu}_i}{\mu_i}(\overline{\boldsymbol{M}}_j)\end{aligned}\tag{4}$$

式中:ρ 为传播率,$\boldsymbol{E}$ 是单位矩阵。

类内传播与类间传播相似,是指将同一个类图像里各子类的点击量根据类内相似度相互分享。本文采

用 K-近邻传播方法分享类内点击量,此过程基于式(4)得到的$(\tilde{\boldsymbol{M}}_j)_i$矩阵进行,其公式如下:

$$\boldsymbol{G}=(\hat{\boldsymbol{M}}_j)_i=(\tilde{\boldsymbol{M}}_j)_i(\rho\tilde{\boldsymbol{S}}^i+(1-\rho)\boldsymbol{E}-\rho\boldsymbol{\Lambda}(\tilde{\boldsymbol{S}}^i)) \tag{5}$$

式中:$\boldsymbol{E}$ 和 $\boldsymbol{\Lambda}(\cdot)$同式(4)一样分别代表单位矩阵和对角化矩阵。

通过 2-D 重排和 2-D 传播算法,得到最终的点击矩阵$(\hat{\boldsymbol{M}}_j)_i$,本文称其为点击特征图,因为经过处理后的矩阵有类似图的平滑性,可以作为理想的深度网络的输入。通过近邻的传播点击量,最大程度地保留了各子类点击数据的独特性。

2.2.2 弱监督深度文本聚类框架

弱监督深度学习的文本聚类框架旨在学习文本的深度点击特征。受到文献[6]中图像深度聚类网络"JULE"的启发,我们构建了面向点击特征图的深度聚类模型。

除了构建点击特征图作为输入外,本文还将弱监督学习引入到训练过程中,使得深度网络在训练的过程中能自动选择可靠性较高的文本进行训练。具体地,我们引入了权重向量 $\boldsymbol{\omega}$ 来衡量查询文本的可靠性,并使用弱监督学习方法使得网络在训练过程中自动更新权重 $\boldsymbol{\omega}$。设网络的参数为 θ,则整个模型可形式化为求解如下问题:

$$(\theta^*,\boldsymbol{w}^*)=\operatorname{argmin}\frac{1}{2}\|\theta\|_2^2+\frac{c}{n}\sum_{j=1}^{m}w_j l(y_j,o_j)+\beta P(\boldsymbol{w})+\gamma S(Z,\boldsymbol{w}) \tag{6}$$

$$\text{s.t.}\begin{cases}\sum_{j=1}^{m}w_j=m\\ l(y_j,o_j)=-\log\left(\dfrac{\mathrm{e}^{o_{y_j}}}{\sum_{k=1}^{m}\mathrm{e}^{o_k}}\right)\\ w_j>0\quad\forall j\end{cases}$$

式中:y_j是查询文本的类别,它被初始为 k-means 算法得到的类别标号,并随着网络迭代逐步更新类标号,o_j为网络输出结果。$l(o,y)$是样本分类损失项,$P(\boldsymbol{w})$是权重先验项,依据文献[6],本文用文本被点击的次数总和来估计相应的权重,即:

$$P(\boldsymbol{w})=\|\boldsymbol{w}-\boldsymbol{w}^C\|_2^2 \tag{7}$$

式中:$\boldsymbol{w}^C$是每个查询文本点击次数和构成的向量。式(6)中 $S(Z,\boldsymbol{w})$是平滑项,与文献[6]中类似,它是根据特征一致性假设构建的。由于式(6)是个过于复杂的非凸优化问题,因此本文仿照文献[6],分两步来训练整个网络。首先固定权重向量 $\boldsymbol{\omega}$ 更新网络参数 θ,之后利用新的网络所提取出的特征和产生的新聚类结果来更新权重 $\boldsymbol{\omega}$。

整个网络的构造如图 3 所示。

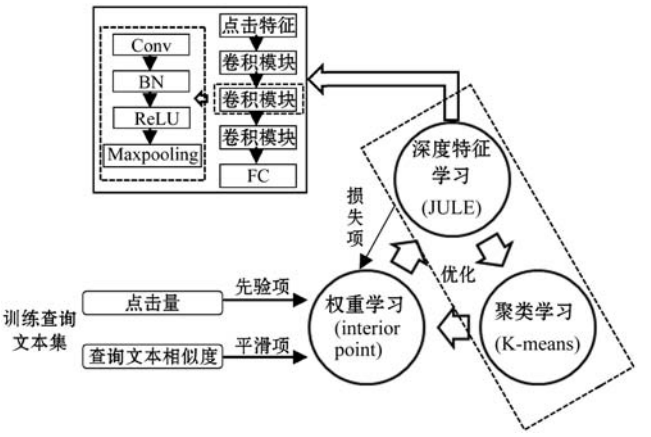

图 3 基于弱监督深度学习的文本聚类框架

与文献[6]中"JULE"网络的结构不同,本文特别为点击输入构建了文本深度网络结构。由于点击的稀疏性,该框架采用相对较少的卷积层。表 1 列出文本深度聚类网络的结构。

表 1 网络结构细节

模块	类型	卷积核,步长,填充	输出
输入	–	–	$n\times m$
卷积模块	Conv	$p\times q,1,\lceil\frac{p}{2}\rceil\times\lceil\frac{q}{2}\rceil$	$n\times m\times 32$
	Maxpooling	$2\times 2,2,0$	$\lceil\frac{n}{2}\rceil\times\lceil\frac{m}{2}\rceil\times 32$
卷积模块	Conv	$p\times q,1,\lceil\frac{p}{2}\rceil\times\lceil\frac{q}{2}\rceil$	$\lceil\frac{n}{2}\rceil\times\lceil\frac{m}{2}\rceil\times 64$
	Maxpooling	$2\times 2,2,0$	$\lceil\frac{n}{4}\rceil\times\lceil\frac{m}{4}\rceil\times 64$
卷积模块	Conv	$p\times q,1,\lceil\frac{p}{2}\rceil\times\lceil\frac{q}{2}\rceil$	$\lceil\frac{n}{4}\rceil\times\lceil\frac{m}{4}\rceil\times 128$
	Maxpooling	$2\times 2,2,0$	$\lceil\frac{n}{8}\rceil\times\lceil\frac{m}{8}\rceil\times 128$
全连接模块	FC	1×160	1×160
损失函数	Softmax	–	1×160

2.3 基于点击数据的图像识别

如上文所述,本文对原始查询文本进行聚类来合并相似查询,并用合并后的查询文本来表示每张图片。利用合并后的查询文本,原始的图片表征 u_i则可更新为 $\hat{u}_i$,其定义如下:

$$\hat{u}_i=\left(\sum_{j\in\vartheta_1}c_{i,j},\sum_{j\in\vartheta_2}c_{i,j},\cdots,\sum_{j\in\vartheta_K}c_{i,j}\right) \tag{8}$$

式中:K 是查询文本的聚类个数,$\vartheta=\{\{i\mid\hat{l}_i=1\},\{\{i\mid\hat{l}_i=2\},\{\{i\mid\hat{l}_i=K\}\}$是每个查询聚类的查询索引集。

当图像均被表征为文本点击特征向量后,利用如下最近邻算法得到预测测试图像集标签 $\hat{y}_i$:

$$\hat{y}_i = y_{i^*} \qquad i^* = \arg\min \| \hat{u} - \hat{u}_i \|_2^2 \tag{9}$$

值得注意的是,训练和测试集中的查询文本往往区别很大,即在训练图像上点击过的查询有可能在测试集上点击次数为零。为了解决这个问题,本文通过寻求查询文本在训练-测试集中映射关系,并利用此关系将测试图像也表征为训练文本集上的点击特征。

在构建文本映射时,需要衡量两个查询之间的距离,本文利用文本点击的图像视觉特征相似度来度量文本间距离。训练集与测试集中的查询文本对(q_i, q_j)之间的距离公式如下:

$$d(q_i, q_j) = \| f(\boldsymbol{\phi}_t, \boldsymbol{v}_i) - f(\boldsymbol{\phi}_s, \boldsymbol{v}_j) \|$$
$$f(\boldsymbol{\phi}, \boldsymbol{v}) = \boldsymbol{\phi} \cdot \boldsymbol{v} \tag{10}$$

式中:$\boldsymbol{v}_i$、$\boldsymbol{v}_j$是 q_i、q_j的图像类点击特征向量,$\boldsymbol{\phi}_t$、$\boldsymbol{\phi}_s$是训练(测试)图像集的深度视觉特征矩阵。

3　实　验

和文献[9]一样,本文在 Clickture-Dog 和 Clickture-Bird 两个公开的点击数据集上进行了实验。Clickture 数据集是从商业图像搜索引擎必应的一年点击日志中抽取的,该数据集包含了一系列(图像、查询文本、点击次数)三元组,是目前最为主流和完善的点击数据集。在本节中,将首先介绍实验的相关设置;之后通过图像识别精度展现点击特征图及深度聚类网络的优势;最后将本文方法与一些经典算法进行对比验证。本文利用基于文本类点击特征的图像识别精度来度量文本聚类算法的效果,所列出的实验结果为多次实验后的平均结果。

3.1　实验设置

和文献[10]一样,本文首先对 Clickture-Dog 和 Clickture-Bird 数据集进行了预处理。并用与文献[11]同样的方式划分数据集。

在表 2 中,我们详细列出了实验数据的相关信息,包括在上文中提到的图像扩增操作。下文中,如无特别说明,所列数据是在 Clickture-Dog 上的结果。

表 2　数据集详细信息

图像	类别	图像			文本	
Clickture	–	训练集	扩增后	测试集	训练集	测试集
Dog	129	13 812	27 493	4 922	12 728	56 614
Bird	75	7 498	20 821	2 461	16 335	6 524

3.2　点击特征图实验

首先实验研究各参数对于点击特征图构建的影响,然后对比原始点击特征向量和点击特征图的识别率,以此验证点击特征图的有效性。

3.2.1　参数实验

1) 聚类个数　本文对聚类个数做了大量实验,结果如表 3 所示。对比不同取值的聚类个数 N_I后,可发现:(1) 图像识别精度与聚类个数 N_I间呈负相关关系。这种现象表明把某一图像类细分为太多的子类会打破这类样本集间的相关性,从而消除相邻元素间的本征联系。(2) 子类个数太少则使点击特征矩阵维度过低,而子类个数过多又会丧失点击数据的特点。因此,本文选择了一个适中的聚类个数,即 N_I为 30。

表 3　聚类个数对精度的影响

N_I	10	20	30	40	60
精度/%	73.36 ±0.54	71.53 ±0.32	69.90 ±0.14	67.48 ±0.66	63.25 ±0.80

2) 近邻传播参数　本文测试了不同的近邻传播参数为 K'与传播率为 ρ 对构建点击特征图的影响,如图 4 所示。

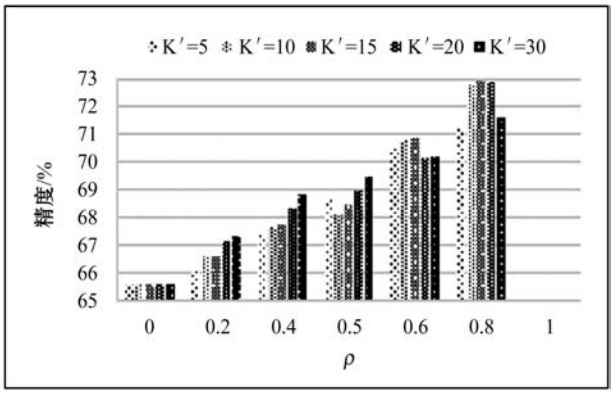

图 4　不同参数构建的点击特征图效果对比

由图 4 可知:(1) 除传播率 $\rho = 1$ 以外,识别精度与传播率 ρ 间呈正相关关系。可能有两方面原因:一是将自身点击量全部传播出去将打破原始点击信息的有效性,降低图像识别精度;二是适当的传播操作可以改善点击数据的稀疏性,令点击数据更加平滑、图像识别精度更高。(2) 当传播率 $\rho < 0.5$ 时,识别精度随 K'的增加而增加;当传播率 $\rho > 0.5$ 时,$K' = 10$ 或 $K' = 15$ 条件下的识别精度较优。

经过以上实验,我们选择 $N_I = 30$、$\rho = 0.8$、$K' = 15$。

3.2.2　点击特征图有效性

本文通过不同点击特征形式的精度验证构建点击特征图的有效性。

表 4 中的“V”、“VP”、“M”、“G”分别表示点击特

征向量、传播的点击特征向量($\rho=0.8$)、点击特征矩阵(传播前)、点击特征图(传播后),对比结果可知:(1)"VP"远优于"V"的结果,证实了 K 近邻传播操作能有效地解决点击数据过于稀疏的问题;(2)"VP"与"M"的识别精度相当,说明图像聚类操作对文本聚类结果的影响并不明显;(3)M 的效果好于"V"也说明了增强图片和聚类表达点击特征具有一定的效果;(4)综合对比"V"、"VP"、"M"、"G"下的识别精度,可以发现使用点击特征图"G"的图像识别效果明显优于聚类其他类型的点击特征的识别结果。

表 4 点击特征图构造过程结果对比

特征类型	V	VP	M	G
精度/%	42.80 ±1.87	65.49 ±0.28	65.61 ±1.64	72.95 ±0.95

3.3 弱监督深度聚类模型

如上文所述,本文的输入为点击特征图,而传统的深度网络的输入为图像。为了寻找最适合于点击数据的深度模型,本文充分研究了几个主要网络结构参数的影响,即卷积核大小和网络层数,结果如表 5 和表 6 所示。根据实验数据,最终确定卷积核大小为 7×7,网络结构为 3 个卷积层加 1 个全连接层。

表 5 卷积核大小对精度影响对比

卷积核大小	3×3	3×5	5×3	5×5	7×7
精度/%	73.06 ±0.82	73.11 ±0.34	73.15 ±0.72	73.80 ±0.72	73.87 ±0.33

表 6 网络层数对精度影响对比

网络层数	2	3	4	5	6
精度/%	72.93 ±0.95	73.87 ±0.33	73.42 ±0.76	73.06 ±1.00	72.49 ±0.56

对于弱监督参数 β、γ 和权重更新次数 T,本文进行了如图 5 所示的对比实验。由图可知,T 也对结果有很大影响,权重更新次数越多,学习到的特征表征能力越强。在最优性能下我们设定 $\beta=0.1$、$\gamma=0.001$。

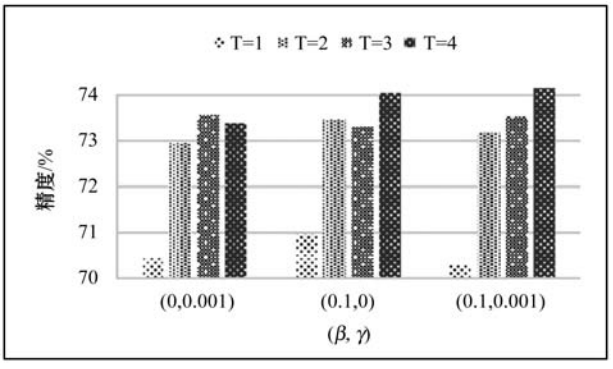

图 5 弱监督中参数不同值的效果对比

3.4 与相关方法的对比

本小节将本文提出的方法和其他常用的深度特征模型进行对比,利用不同模型获得文本的深度点击特征,再利用 K-means 进行文本聚类。本文最终设定查询聚类的聚类个数 $K=500$。

本文采用 VGG、JULE 和 DEPICT[7] 作为对比网络。VGG 是经典的卷积神经网络,而 JULE 和 DEPICT 是深度聚类网络。由于本文的输入是点击特征图,因此我们对 JULE 和 DEPICT 进行了调整,将点击特征图作为输入。调整后的模型我们称为 C-JULE 和 C-DEPICT。本文提出的方法使用点击特征图作为输入,并融合了弱监督的训练方法,因此将本文的方法称为 C-JWLE(Click-data guided Joint Weakly-supervised LEarning of deep representations)。

3.4.1 识别精度

我们在 Clickture-Dog 和 Clickture-Bird 上进行对比实验,结果如表 7 和表 8 所示。

表 7 在 Clickture-Dog 上的不同深度模型对比

方法	Graph	VGG	JULE	C-DEPICT	C-JULE	C-JWLE
精度/%	72.95 ±0.95	71.22 ±1.39	72.83 ±0.90	73.87 ±0.33	72.49 ±0.56	74.16 ±0.42

表 8 在 Clickture-Bird 上的不同深度模型对比

方法	Graph	VGG	JULE	C-JULE	C-JWLE
精度/%	36.68 ±0.37	34.19 ±1.08	36.85 ±0.50	37.95 ±0.85	38.42 ±0.42

从上述结果可知:

(1) C-DEPICT/C-JULE 优于 VGG/JULE 的性能,说明传统的图像深度模型(VGG 和 JULE)是依据图像的视觉特点搭建的,并不适用于点击数据。与之相比,C-DEPICT/C-JULE 是专门针对点击数据设计的浅层深度模型。同时,C-JULE 明显优于 JULE 方法的识别精度,也证明了基于点击数据设计专属模型的必要性。

(2) 与 C-JULE 相比,C-JWLE 由于融合了弱监督学习策略,取得了更好的效果。说明弱监督的学习策略可以更好地消除点击数据中的噪声,进而提升模型的整体性能。

3.4.2 聚类可视化分析

进一步地,为了更加直观地分析弱监督算法的效果,本文对基于 C-JULE 和 C-JWLE 的聚类结果进行了可视化对比,图 6 和图 7 分别展示了 C-JULE 和 C-JWLE 产生的若干个文本聚类结果,图中每个查询类

cluster 中一行表示一条查询文本。

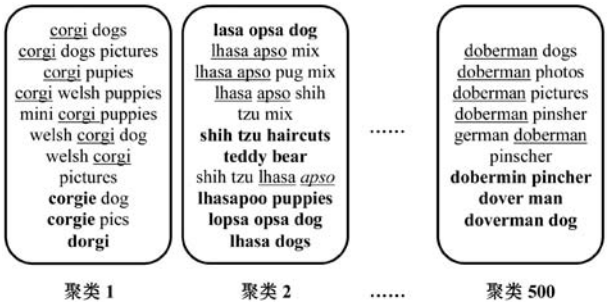

图 6　C-JULE 聚类效果

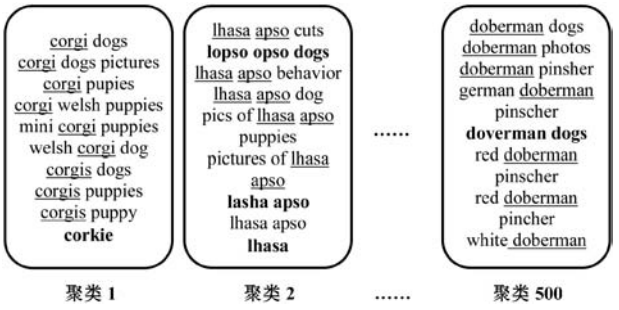

图 7　C-JWLE 聚类效果

由图可知,基于 C-JWLE 得到的每个聚类中,更多的查询文本拥有相同的主题词根(黑色划线),而 C-JULE 更容易将含有不同意义词根(黑色加粗)的文本聚成一类。这种现象说明 C-JWLE 由于能更好地应对文本噪声,从而产生优于 C-JULE 的文本聚类效果。

4　结　语

本文利用点击数据将图像表征为文本点击特征向量进而实现鲁棒的图像识别。针对查询文本集的规模庞大、冗余的问题,本文提出面向点击特征的深度文本聚类框架来合并语义相似的查询文本。特别地,本文提出了一种新颖的 2-D 重排和 2-D 点击传播方法来构建一个平滑的结构化的点击特征图来表示查询文本。此外,本文将深度学框架扩展到点击数据上,学习查询文本的深度表征。本文还结合弱监督学习策略自动学习查询文本权重,利用迭代优化的方法交替更新文本权重和深度点击特征。本文在公共数据集 Clickture-Dog 和 Clickture-Bird 上进行了实验。结果表明:(1)点击特征图的构建有效地解决了查询文本的稀疏性和不平滑性问题;(2)通过引入弱监督学习策略,有效地克服了查询文本中的噪声问题。今后,将继续对该算法进行改进,以获得更好的聚类效果。同时,也在考虑利用迁移学习的思想,将点击数据应用到其他公共数据集中,辅助完成其他计算机视觉任务。

参 考 文 献

[1] Zheng G, Tan M, Yu J, et al. Fine-grained image recognition via weakly supervised click data guided bilinear CNN model[C]//IEEE International Conference on Multimedia and Expo. IEEE, 2017:661 - 666.

[2] Tan M, Yu J, Zheng G, et al. Deep Neural Network Boosted Large Scale Image Recognition Using User Click Data [C]//International Conference on Internet Multimedia Computing and Service. ACM, 2016:118 - 121.

[3] Wu W, Tan M, Zheng G, et al. Query Modeling for Click Data Based Image Recognition Using Graph Based Propagation and Sparse Coding[C]//International Conference on Internet Multimedia Computing and Service. Springer, Singapore, 2017:191 - 199.

[4] Yu J, Rui Y, Chen B. Exploiting Click Constraints and Multi-view Features for Image Re-ranking[J]. IEEE Transactions on Multimedia, 2013, 16(1):159 - 168.

[5] Yu J, Rui Y, Tao D. Click prediction for web image reranking using multimodal sparse coding[J]. Image Processing IEEE Transactions on, 2014, 23(5):2019 - 2032.

[6] Yang J, Parikh D, Batra D. Joint Unsupervised Learning of Deep Representations and Image Clusters[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016:5147 - 5156.

[7] Dizaji K G, Herandi A, Deng C, et al. Deep Clustering via Joint Convolutional Autoencoder Embedding and Relative Entropy Minimization[EB]. arXiv:1704.06327, 2017.

[8] Tian F, Gao B, Cui Q, et al. Learning Deep Representations for Graph Clustering[C]//Twenty-eighth Aaai Conference on Artificial Intelligence. AAAI Press, 2014.

[9] Tan M, Yu J, Yu Z, et al. User-Click-Data-Based Fine-Grained Image Recognition via Weakly Supervised Metric Learning[J]. ACM Transactions on Multimedia Computing, Communications,and Applications(TOMM),2018,14(3):70.

[10] Feng W, Liu D. Fine-Grained Image Recognition from Click-Through Logs Using Deep Siamese Network[C]//International Conference on Multimedia Modeling. Springer, Cham, 2017:127 - 138.

[11] Hua X S, Yang L, Wang J, et al. Clickage: Towards Bridging Semantic and Intent Gaps via Mining Click Logs of Search Engines[C]//Acm International Conference on Multimedia. ACM, 2013:243 - 252.

[12] Huttenlocher D P, Klanderman G, Rucklidge W J. Comparing images using the Hausdorff distance[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 1993, 15(9):850 - 863.