

基于深度学习耦合稀疏语义度量的商标检索算法

梁平¹, 柴建伟¹, 裴圣华²

(1.廊坊燕京职业技术学院, 廊坊 065200; 2.吉安职业技术学院, 吉安 343000)

摘要: **目的** 针对当前商标图像检索中的语义鸿沟问题, 提出一种深度学习耦合稀疏语义度量的商标图像检索方案, 有效抑制噪声干扰, 降低冗余特征维数。**方法** 首先, 根据由卷积与池化组成的无监督学习机制, 对输入商标图像进行多层特征提取, 输出一维特征向量。随后, 通过 L2-支持向量机 (L2-SVM) 进行分类, 利用特征向量进行训练, 获得多级联特征。然后, 根据商标图像的多级联特征和用户标签信息的异构数据结构, 设计一种稀疏语义度量方法进行相似检索, 减少语义鸿沟。此外, 引入一种混合范数作为相似度量的稀疏约束, 以抑制原始输入空间中的冗余特征维数和噪声, 优化检索结果。**结果** 实验表明, 与当前流行的商标检索方案相比, 所提算法具有更高的检索精度, 其输出的结果中仅有 1 幅无关图像。**结论** 该方案具有较高的检索精度和较强的鲁棒性, 在商标检测、商标保护等方面中具有良好的应用价值。

关键词: 商标检索; 语义鸿沟; 深度学习; 稀疏语义度量; L2 支持向量机; 混合范数

中图分类号: TP391.4 **文献标识码:** A **文章编号:** 1001-3563(2019)03-0237-09

DOI: 10.19554/j.cnki.1001-3563.2019.03.036

Trademark Retrieval Algorithm Based on Deep Learning Coupled Sparse Semantic Measure

LIANG Ping¹, CHAI Jian-wei¹, PEI Sheng-hua²

(1.Langfang Yanjing Vocational Technical College, Langfang 065200, China; 2.Ji'an College, Ji'an 343000, China)

ABSTRACT: The work aims to propose a trademark image retrieval scheme with deep learning coupled sparse semantic measure to effectively restrain noise interference and reduce redundant feature dimension, with respect to the problem of semantic gap in trademark image retrieval. Firstly, according to the unsupervised learning mechanism composed of convolution and cistern, the multi-layer feature extraction of input trademark image were carried out to output the one-dimensional feature vector. Then, the L2-support vector machine (L2-SVM) was used to classify the feature vectors, and the multilevel features were obtained based on the training with feature vectors. Then, according to the multilevel feature of the trademark image and the heterogeneous data structure of the user's label information, a sparse semantic measure method was designed for similar retrieval, which reduced the semantic gap. In addition, a mixed norm was introduced as a sparse constraint of similarity measure to suppress the redundant feature dimension and noise in the original input space and optimize the retrieval results. The experiment showed that, compared with the current popular trademark retrieval scheme, the proposed algorithm had higher retrieval accuracy, whose output results only had one irrelevant image. The proposed scheme has higher retrieval precision and stronger robustness, and it has good application value in trademark detection, trademark protection and so on.

KEY WORDS: trademark retrieval; semantic gap; deep learning; sparse semantic measure; L2-SVM; mixed norm

收稿日期: 2018-05-01

基金项目: 河北省高等学校科学技术研究项目 (Z2015085); 江西省教育厅自然科学研究项目 (GJJ171375)

作者简介: 梁平 (1984—), 女, 硕士, 廊坊燕京职业技术学院讲师, 主要研究方向为计算图像、信息检测、软件技术。

随着市场经济的飞速发展,商标申请与注册也急剧增长,到目前为止,国内注册的商标已超 1200 万个^[1]。面对如此巨大的商标数量,在商标注册时,对商标唯一性审查的难度不断升级,同时给商标的保护与维权也带来一定麻烦,因此,对商标的自动化识别与检索受到广泛的关注。注册商标受法律维护的一种标志,一般是通过数字、文字、形状与颜色组合形成了一种图案,具有专有性与唯一性^[2-3]。

近年来,学者经过研究建立了一些商标检索方案,刘剑英等^[4]设计了一种角点描述子与局部区域特征相结合的商标检索方案。该方案首先利用角点的邻域特征和分布特征,构建一种角点描述子。其次,通过区域分割的交叉角点匹配完成粗略选择,得到结构相似的图形。并利用匹配结果分割出候选相似区域。通过候选相似区域之间的区域特征进行匹配,确保细节的相似性。该算法对商标图像局部相似性的检索有一定的检索性能,对平移、伸缩、扭曲等有较好鲁棒性,但角点描述符对颜色与噪声的辨别较弱,因此其对商标检索适应性不强。Akriti 等^[5]定义了一种 SIFT 耦合 HSV 相关图加权组合的商标图像检索方案。通过结合颜色与形状特征,利用 SIFT 对商标进行描述,可增强算法对旋转、缩放、平移等各种变换影响的稳定性。SIFT 主要考虑局部像素的梯度方向,容易使主方向计算出现偏差,从而影响特征向量提取,导致对形状表示能力不足,且 SIFT 特征维度巨大,需占用较大内存和时间成本。Yiju 等^[6]针对商标检索,定义了颜色和空间描述符的自适应融合方法。首先,将颜色量化和 K-均值结合用于有效的主颜色提取。对于每个提取的主导颜色,组件的空间描述符是针对局部特征导出的。通过对图像直方图的分析,实现了这颜色与空间特征的融合,以实现更有效的标志提取和更精确的图像检索。该方法提高了商标图像检索性能,但是,当图像质量因噪声等各种原因而劣化时,颜色和空间特征的可靠性可能以不同的方式变化,而其权重无法自适应变化,导致检索效果不是特别满意。

商标检索中的核心技术是对商标特征的提取与度量。商标特征提取得准确与否,直接影响了后续的检索结果。对此,文中利用深度学习与稀疏语义度量作用于商标检索。通过卷积与池化组成的无监督学习对输入商标图像进行多层特征提取,并通过 L2-SVM 进行分类,利用特征向量进行训练,获得多级联特征,获得了商标特征的深层次表示,得到了局部与全局空间特征。为了避免语义鸿沟,构建了一种稀疏语义度量,通过对商标图像的多级联特征和用户标签信息的异构数据结构进行探索,在文本空间的语义结构中进行相似度量。引入一种混合范数作为相似度量的稀疏约束,抑制原始输入空间中的冗余特征维数与噪声,提高了相似度量能力,改善了商标检索性能。最后,

对所提算法的检索性能进行验证。

1 深度学习

在图像检索中,机器视觉的关键技术是如何在无监督学习中对获得鲁棒特征。当前,更多学者将目标转移到深度学习中,利用多层、自动特征提取方法,获得像素级特征。深度学习的核心是通过大量数据的训练来建立含多个隐含层的深层模型,让每层学习获得简单且有效的特征,从而获得低维、稀疏、但更具区分性的高层特征,提高了深度学习模型的表达能力和分类准确^[7]。文中设计了一种深度学习方法,包括多层的特征提取与多级联分类器。多层特征提取通过无监督学习实现,无需图像的标签,其输出为 1D 的特征向量,为图像的抽象表示。通过 L2-SVM 进行分类,利用特征向量与标签信息进行训练。特征提取是通过若干个相似的多层组成完成的,每个层由卷积与池化操作组成^[8]。

为了提高提取特征的鲁棒性,文中引入了降噪自动编码器,通过稀疏约束控制隐含单元含有的信息,抑制噪声,并定义损失函数,使编码层具有稀疏性。如图 1 中,在执行中,寻找隐含信息 $h = f_{\theta}(\bar{x}) = \sigma_1(W\bar{x} + b)$, 并反计算输入信息 $y = f_{\phi}(h) = \sigma_2(Wh + b)$ 。 $\bar{x}$ 表示输入信息 x 的失真信息, $E_H(x, y) = \|y - x\|^2$ 为训练集的重建误差,文中的目的是使 y 与原始信息 x 最大限度保持一致。

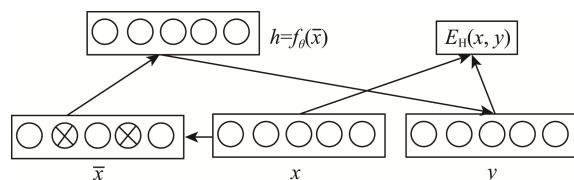


图 1 降噪自动编码器

Fig.1 Automatic encoder for noise reduction

为优化性能,在执行过程中加入了惩罚因子 v , 表示如下:

$$v = \sum_{j=1}^{l_{k+1}} \rho \lg \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \lg \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (1)$$

式中: ρ 为稀疏参数,一般取值很小。 $\hat{\rho}_j$ 为隐含信息 h_j 的均值。如果 $\hat{\rho}_j$ 与 ρ 相差越大,那么 v 的值也越大。

通过降噪自动编码器得到的特征后,将其与原始图像卷积处理,获得不同图像坐标的特征响应值。假定大小为 $m \times n$ 的图像,区域为 $p \times p$ 的特征响应值表示为 $f_s = \sigma(Wx_s + b)$, 卷积特征图的为 $(m - p + 1) \times (n - p + 1)$ 。 σ 为线性激活方程。采取 ON-OFF 方法^[9] 激活函数 $\begin{pmatrix} \max(0, x) \\ \max(0, -x) \end{pmatrix}$, 通过 2 个通道将数值输入到

下个步骤。通过池化操作,将特征图的值聚集在一起,求取平均值。利用池化,可得到更低的特征图,降低过拟合,提高平移不变性,因此,即便图像产生了一定的平移,通池化操作后仍可保持相同的特征。

经过上述过程后,可获得鲁棒的特征,但面对如此复杂多样的图像,需要足够的深度来学习,通过不同层级特征,表示局部与全局的表示。如第 1 级可能为边缘、颜色等特征。第 2 级可能为组合特征等更深层次的表示。对此,针对不同层级的特征,设计了一个多级分类器,将池化特征作为输入量。在分类器中,每个特征得到一个得分 S , 传统的分类器中输出结果可能不是很准确,真实的特征可能隐藏在相似特征中。通过多级分类器后,一幅图像可得到 $2K$ 个得分,对于得到的多个得分结果,由于高层特征的预测效果好于底层特征,为了显示其优势,引入权重因子,多级分类器表示如下^[10]:

$$P = \sum_{j=1}^K \delta_j (P_{\max j} + P_{\text{mean}j}) \quad (2)$$

式中: P 为最终得分; K 为层数; δ 为权重因子; $P_{\max}$, P_{mean} 分别为最大与平均池化特征得分。

2 稀疏语义度量

2.1 问题的定义

假设给出了一组 n 个数据点 $X=[x_1, x_2, \dots, x_n]$, 对任何 2 个数据点 x_i, x_j , 令 x_i, x_j 之间的距离为 $d_M(x_i, x_j)$, 定义如下:

$$d_M(x_i, x_j) = \|x_i - x_j\|_M^2 = (x_i - x_j)^T M (x_i - x_j) \quad (3)$$

式中: M 为 Mahalabobis 测量, 是一个对称矩阵, $M \in R^{d \times d}$ 。为满足度量的性质, 如非负性和三角不等式, M 必须为正半定。注意, 当 M 等于单位矩阵 I 时, 式 (3) 中的距离减少到 Euclidean 距离。在实践中, 每一个有效的度量 M 可以分解为 $M=UU^T$, 其中 $U=[u_1, u_2, \dots, u_r] \in R^{d \times r}$, $r \leq d$, 因此, 2 个目标之间的距离可以表示为:

$$d_U(x_i, x_j) = \|U^T(x_i - x_j)\|^2 \quad (4)$$

在文中任务中, 尝试通过学习度量来搜索相关图像到给定的查询图像。对于查询图像, 只有视觉特征是已知的, 虽然从网上收集的索引图像具有用户标签信息以及它们的视觉特征。目标是通过联合探索视觉特征 (记录为 X) 与图像的标签信息 ($Y=[y_1, y_2, \dots, y_n]$) 来学习最佳距离度量, y_i 为第 i 个图像标签矢量。然后, 将学习问题转化为以下优化框架:

$$M = \arg \min_M F(M, X, Y) + R(M), M \geq 0 \quad (5)$$

式中: $F()$ 为关于给定数据收集 X 和 Y 上定义的距离度量 M 的目标函数; $R()$ 为 M 的正则项。距离学习问题的关键是制定这 2 个函数, 在下面的小

节中, 将讨论定义优化框架的一些原则, 在此基础上, 定义一种求解优化问题的有效算法。

2.2 模型建立

对于文中提出的度量方法, 其主要思想是: 学习度量通过计算视觉相似度与文本域中的语义相似度一致, 且不应偏离 Euclidean 空间中的原始相似度, 以防止在真实数据中过度拟合噪声标记信息。对此, 文中引入概念 $F(M, X, Y)$ 来定义目标函数。对于 $R(M)$, 其动机是获得紧凑的特征表示稀疏约束 M 。具体解释如下:

首先, 为了消除语义鸿沟, 鼓励映射空间和文本空间中的两两相似度保持一致性。为此, 提出以下成本函数作为异质空间之间不一致的度量^[11]:

$$\begin{aligned} f(M) &= \frac{1}{2} \sum_{i,j=1}^n \|U^T x_i - U^T x_j\|_2^2 W_{ij} = \\ &= \sum_{k=1}^r u_k^T X(D-W)X^T u_k = \sum_{k=1}^r u_k^T XLX^T u_k = \\ &= Tr(U^T XLX^T U) = Tr(XLX^T UU^T) = Tr(XLX^T M) \end{aligned} \quad (6)$$

式中: W 为标签相似度矩阵, 其元素定义为

$$W_{ij} = \frac{y_i^T y_j}{\|y_i\|_2 \|y_j\|_2}。D \text{ 为对角矩阵, 其对角线元素定义为 } D_{ij} = \sum_j W_{ij}。L=D-W \text{ 称为 Laplacian 矩阵。通过}$$

最小化上述函数, 可以得到语义一致的度量。在实际生活中, 图像标签是不完整的, 具有嘈杂性和主观性, 因为它们通常是由用户随时随地创建的。对于式 (6) 中目标函数的学习度量 M 的最小解, 可能导致噪声语义相似性的过度拟合。为了避免过拟合, 期望学习度量能够保留原始视觉相似性。具体来说, 目的是使 M 尽可能接近身份矩阵 I , 对此, 文中采用 Bregman 散度量 M 和 I 之间的紧密度^[12], 表示为:

$$D(M\|I) = g(M) - g(I) - \langle \nabla_g(I), M - I \rangle \quad (7)$$

式中: $g()$ 为凸连续可微函数。在文中, 使用 logdet 函数^[13]定义 $g()$, 例如 $g() = -\log \det(M)$, 因此, 上述公式可演化为:

$$D(M\|I) = Tr(M) - \log \det(M) - n \quad (8)$$

结合式 (4) 与式 (6) 可得到:

$$\begin{aligned} F(M) &= f(M) + \alpha D(M\|I) = \\ &= Tr(XLX^T M) + \alpha [Tr(M) - \log \det(M)] \end{aligned} \quad (9)$$

式中: α 为权衡正参数; n 为常数。显然, 由于第 1 项和 BReGeman 散度分别是凸的, 上述函数为凸性。

对于正则化项 $R(M)$, 约束其为行稀疏, 定义 M 的 $\ell_{2,1}$ 范数:

$$R(M) = \|M\|_{2,1} = \sum_{l=1}^d \|m^l\|_2 \quad (10)$$

其中, $\mathbf{m}^l$ 为 $\mathbf{M}$ 的行向量。首先通过计算 $\mathbf{M}$ 行上的 2-范数得到混合范数, 向量的 1-范数表示为 $v(\mathbf{M}) = (\|\mathbf{m}^1\|_2, \|\mathbf{m}^2\|_2, \dots, \|\mathbf{m}^d\|_2)$ 。很容易地观察到, 由 1-范数所施加的稀疏性要求 $v(\mathbf{M})$ 的一些元素为 0。 $\mathbf{m}^l=0$, 相当于 $\mathbf{U}^T$ 第 l 列向量为 0, 例 $(\mathbf{U}^T)_l=0$ 。因此, 混合 $\ell_{2,1}$ -范数的行稀疏可以消除原始空间中的一些噪声或冗余特征维数。

用上述定义将 $F(\mathbf{M})$ 与 $R(\mathbf{M})$ 代入式 (5), 得到优化问题为:

$$\min_{\mathbf{M} \geq 0} \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}] + \alpha(\text{Tr}[\mathbf{M}] - \log \det(\mathbf{M})) + \frac{\beta}{2} \|\mathbf{M}\|_{2,1} \quad (11)$$

其中, 权衡参数 $\alpha > 0, \beta > 0$ 。

2.3 模型优化

对于 2.2 节建立的模型中涉及的非光滑的 $\ell_{2,1}$ -范数, 其不能以封闭形式解决。对此, 文中开发了一种迭代算法来求解。通过定义对角线元素

$\mathbf{D}_{ll} = \frac{1}{2\|\mathbf{m}^l\|_2}$ 的对角矩阵 $\mathbf{D}$, 式 (11) 可表示为:

$$\min_{\mathbf{M} \geq 0} \lambda(\mathbf{M}) = \min_{\mathbf{M} \geq 0} \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}] + \alpha(\text{Tr}[\mathbf{M}] - \log \det(\mathbf{M})) + \frac{\beta}{2} \text{Tr}[\mathbf{M}^T \mathbf{D} \mathbf{M}] \quad (12)$$

使用式 (13) 来得到 $\lambda(\mathbf{M})$ 的导数, 见式 (14)。

$$\begin{cases} \frac{\partial \log \det(\mathbf{M})}{\partial \mathbf{M}} = \mathbf{M}^{-1} \\ \frac{\partial \text{Tr}[\mathbf{M}]}{\partial \mathbf{M}} = \mathbf{I} \end{cases} \quad (13)$$

$$\frac{\lambda(\mathbf{M})}{\partial \mathbf{M}} = \mathbf{X}\mathbf{L}\mathbf{X}^T + \alpha(\mathbf{I} - \mathbf{M}) + \beta \mathbf{D} \mathbf{M} \quad (14)$$

通过将上面的导数设为 0, 得到如下:

$$\begin{aligned} \mathbf{X}\mathbf{L}\mathbf{X}^T + \alpha(\mathbf{I} - \mathbf{M}) + \beta \mathbf{D} \mathbf{M} &= 0 \\ \Leftrightarrow \alpha \mathbf{M}^{-1} &= \mathbf{X}\mathbf{L}\mathbf{X}^T + \alpha \mathbf{I} + \beta \mathbf{D} \mathbf{M} \\ \Leftrightarrow \mathbf{M} &= \alpha(\mathbf{X}\mathbf{L}\mathbf{X}^T + \alpha \mathbf{I} + \beta \mathbf{D} \mathbf{M})^{-1} \end{aligned} \quad (15)$$

为了计算其逆矩阵, 利用 Taylor 近似求解矩阵逆问题^[14], 为了简单起见, 采用一阶 Taylor 展开, 表示为:

$$\begin{aligned} \mathbf{M} &= \mathbf{I} - \frac{1}{\alpha} \mathbf{X}\mathbf{L}\mathbf{X}^T - \frac{\beta}{\alpha} \mathbf{D} \mathbf{M} \\ \Leftrightarrow \mathbf{M} &= (\alpha \mathbf{I} + \beta \mathbf{D})^{-1} (\alpha \mathbf{I} - \mathbf{X}\mathbf{L}\mathbf{X}^T) \end{aligned} \quad (16)$$

从上述方程中可以看出, $\mathbf{M}$ 的解与输入的 $\mathbf{D}$ 有关, 因此, 获得 $\mathbf{M}$ 是不简单的。为了处理这个问题, 提出了一种迭代算法, 见表 1。为了效率, 在每个迭代步骤中, 不强制半正定 (positive semi-definite, PSD) 约束, 而是在整个算法结束时对最终矩阵 $\mathbf{M}$ 执行 PSD 投影。

算法 1, 迭代优化。

输入: 数据 $\mathbf{X}$ 和 $\mathbf{Y}$; 参数 α, β 。

1) 令 $\tau=1$, 初始化 $\mathbf{M}_\tau$ 。

2) 根据 $\mathbf{Y}$ 计算 Laplacian 矩阵。

3) 重复。

4) 更新对角矩阵 $\mathbf{D}$, 为 $\mathbf{D}_{\tau+1} =$

$$\begin{bmatrix} \frac{1}{2\|\mathbf{m}_{\tau+1}^1\|_2} & & \\ & \dots & \\ & & \frac{1}{2\|\mathbf{m}_{\tau+1}^d\|_2} \end{bmatrix}。$$

5) 更新 $\mathbf{M}$: $\mathbf{M}_{\tau+1} = (\alpha \mathbf{I} + \beta \mathbf{D}_\tau)^{-1} (\alpha \mathbf{I} - \mathbf{X}\mathbf{L}\mathbf{X}^T)$ 。

6) $\tau=\tau+1$ 。

7) 直到满足收敛条件。

8) $\mathbf{M}=\mathbf{M}_\tau$ 。

9) PSD 映射: $\mathbf{M}=\text{PSD}(\mathbf{M})$ 。

输出: 度量 $\mathbf{M}$ 。

算法 1 中所提出的迭代过程可以通过以下定理 1 验证收敛到最优 $\mathbf{M}$ 。

定理 1: 算法 1 中的交替更新规则在每次迭代中单调地降低目标函数值。正如上面的分析, 可以得出:

$$\begin{aligned} \mathbf{M}_{\tau+1} &= \arg \min \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}] + \\ &\alpha(\mathbf{M} - \log \det(\mathbf{M})) + \frac{\beta}{2} \text{Tr}[\mathbf{M}^T \mathbf{D}_\tau \mathbf{M}] \end{aligned} \quad (17)$$

也就是说, 对于矩阵 $\mathbf{M}_{\tau+1}$, $\mathbf{M}_\tau$, 可表示为:

$$\begin{aligned} \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}_{\tau+1}] + \alpha(\mathbf{M}_{\tau+1} - \log \det(\mathbf{M}_{\tau+1})) + \\ \frac{\beta}{2} \text{Tr}[\mathbf{M}_{\tau+1}^T \mathbf{D}_\tau \mathbf{M}_{\tau+1}] \leq \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}_\tau] + \\ \alpha(\mathbf{M}_\tau - \log \det(\mathbf{M}_\tau)) + \frac{\beta}{2} \text{Tr}[\mathbf{M}_\tau^T \mathbf{D}_\tau \mathbf{M}_\tau] \end{aligned} \quad (18)$$

为了便于表达, 定义 $h(\mathbf{M}) = \text{Tr}[\mathbf{X}\mathbf{L}\mathbf{X}^T \mathbf{M}] + \alpha(\mathbf{M}) - \log \det(\mathbf{M})$, 因此, 可得到:

$$h(\mathbf{M}_{\tau+1}) + \frac{\beta}{2} \text{Tr}[\mathbf{M}_{\tau+1}^T \mathbf{D}_\tau \mathbf{M}_{\tau+1}] \leq h(\mathbf{M}_\tau) + \frac{\beta}{2} \text{Tr}[\mathbf{M}_\tau^T \mathbf{D}_\tau \mathbf{M}_\tau] \quad (19)$$

根据 $\mathbf{D}$ 的定义, 上述方程等价于:

$$\begin{aligned} h(\mathbf{M}_{\tau+1}) + \frac{\beta}{2} \sum_{l=1}^d \frac{\|\mathbf{m}_{\tau+1}^l\|_2^2}{2\|\mathbf{m}_{\tau+1}^l\|_2} \leq h(\mathbf{M}_\tau) + \frac{\beta}{2} \sum_{l=1}^d \frac{\|\mathbf{m}_\tau^l\|_2^2}{2\|\mathbf{m}_\tau^l\|_2} \\ \Leftrightarrow h(\mathbf{M}_{\tau+1}) + \frac{\beta}{2} \|\mathbf{m}_{\tau+1}^l\|_{2,1} - \frac{\beta}{2} \left(\|\mathbf{m}_{\tau+1}^l\|_{2,1} - \sum_{l=1}^d \frac{\|\mathbf{m}_{\tau+1}^l\|_2^2}{2\|\mathbf{m}_{\tau+1}^l\|_2} \right) \leq \\ h(\mathbf{M}_\tau) + \frac{\beta}{2} \|\mathbf{m}_\tau^l\|_{2,1} - \frac{\beta}{2} \left(\|\mathbf{m}_\tau^l\|_{2,1} - \sum_{l=1}^d \frac{\|\mathbf{m}_\tau^l\|_2^2}{2\|\mathbf{m}_\tau^l\|_2} \right) \end{aligned} \quad (20)$$

对于任意 2 个非零常量 a 和 b , 其关系有

$\sqrt{a} - \frac{a}{2\sqrt{b}} \leq \sqrt{b} - \frac{b}{a\sqrt{b}}$ ，因此，可得到：

$$\sum_{l=1}^d \|m'_{\tau+1}\|_2 - \frac{\|m'_{\tau+1}\|_2^2}{2\|m'_{\tau+1}\|_2} \leq \sum_{l=1}^d \|m'_\tau\|_2 - \frac{\|m'_\tau\|_2^2}{2\|m'_\tau\|_2} \quad (21)$$

此外，还可得到：

$$h(\mathbf{M}_{\tau+1}) + \frac{\beta}{2} \|m'_{\tau+1}\|_{2,1} \leq f(\mathbf{M}_\tau) + \frac{\beta}{2} \|m'_\tau\|_{2,1} \quad (22)$$

通过上述函数可表明目标函数在算法 1 中使用的更新规则为单调递减。

3 算法实现

文中提出的深度学习耦合稀疏语义度量的商标图像检索方案见图 2。对于查询商标图像，为了提取更深层次的商标特征，采取了一种深度学习的无监督学习机制。为了提高提取特征的鲁棒性，文中引入了降噪自动编码器，通过稀疏约束控制隐含单元含有的信息，抑制噪声，并定义损失函数，使编码层具有稀疏性。通过降噪自动编码器得到的特征后，将其与原始图像卷积处理，获得不同图像坐标的特征响应值。在利用池化操作，将特征图的值聚集在一起，求取平均值。从而可得到更低的特征图，降低过拟合，提高平移不变性。针对不同层级的特征，设计了一个 L2-SVM 多级分类器，将池化特征作为输入量。在分类器中，每个特征得到一个得分 S ，传统的分类器中输出结果可能不是很准确，真实的特征可能隐藏在相似特征中。通过多级分类器后，一幅图像可得到 2K 个得分，对于得到的多个不同得分的多级联特征。再根据商标图像的多级联特征和用户标签信息的异构数据结构，设计了一种稀疏语义度量方法进行相似检索，减少语义鸿沟。此外，引入一种混合 $\ell_{2,1}$ -范数作为相似度量的稀疏约束，以抑制原始输入空间中的冗余特征维数与噪声，并开发了一个迭代算法进行迭代优化。

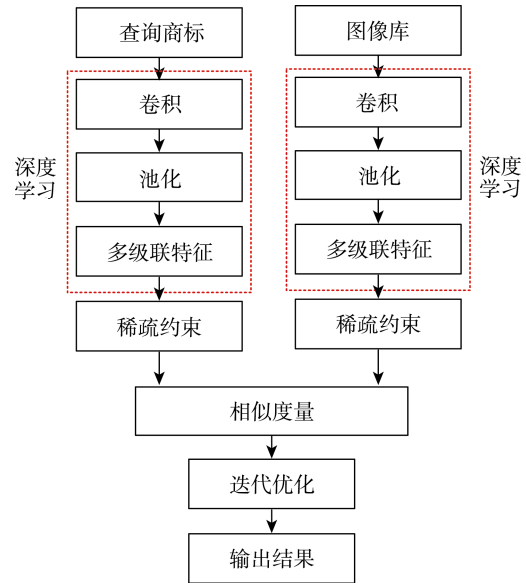


图 2 文中算法框架

Fig.2 Framework of the proposed algorithm

4 实验与分析

为验证算法的有效性，选取 2000 幅不同的商标组成测试集。测试环境为：Intel (R) Core (TM) I5-4590CPU@3.30 GHz, 4 GB ROM, Windows8 系统。同时借助利用 Matlab 2012a 进行仿真分析。为体现算法的先进性，利用常用商标检索算法作对照组：文献 [4]、文献 [5]、文献 [6]。对于参数 α 、 β ，通过试验得出了 α 、 β 与平均精度 F_1 的曲线，见图 3。根据图 3 中显示，当 $\alpha=1$ ， $\beta=100$ ，迭代次数 10，权重因子 $\delta=0.32$ ，常数 $n=1.5$ ，

4.1 评价标准

为了准确衡量算法的性能，引入 Precision-Recall 以及 F_1 测量作为评价标准，其中，Precision 和 Recall 分别定义为^[15]：

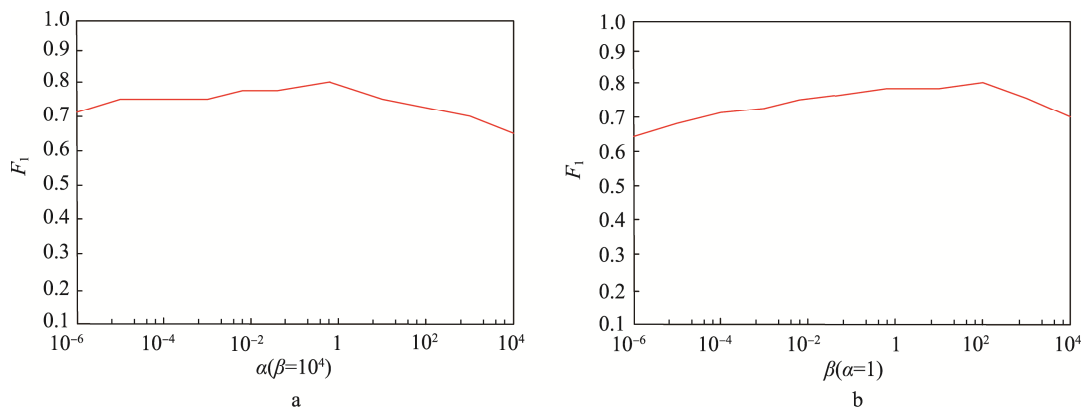


图 3 参数 α 、 β 与 F_1 的关系

Fig.3 Relationship between parameters α , β and F_1

$$P = \frac{T_p}{T_p + F_p} \quad (23)$$

$$R = \frac{T_p}{T_p + F_N} \quad (24)$$

式中: T_p 为检索相关商标; F_p 为检索不相关商标; F_N 为漏检的商标。

P-R 曲线简单直观, 既具有检索结果的覆盖度, 又呈现了检索结果的排序状况。单个的 P-R 曲线虽很直观, 但难以准确反映 2 个查询结果的好坏。为了综合 Precision-Recall 的评估指标, 反映算法的整体的性

能, 称之为 F_1 -测量, 定义为^[16]:

$$F_1 = \frac{2PR}{P+R} \quad (25)$$

4.2 实验结果

实验中选择 12 幅商标作为检索对象, 见图 4, 每幅商标分别执行有差异性的操作(旋转、平移、噪声、模糊、遮挡、缩放)等变化, 每幅商标生成了 20 种基准图像。实验过程中, 返回 12 幅图像作为检索结果。



图4 商标示例

Fig.4 Trademark example

实验中以图 5“李宁”为检索目标, 利用对照算法与文中算法返回了 12 幅图像, 见图 6—8。依据实验结果看出, 文献[4]返回的 12 个对象中, 与检索目标相关的个数为 9 个, 出现了 3 个不相关的商标, 见图 6d。文献[5]中得到的结果中, 相关的商标图像个数为 10 个, 不相关图像为 2 个(“安踏”、耐克)。文献[6]得到的结果中, 相关商标个数为 10 个, 不相关商标为 2 个(“特步”、“安踏”)。文中法返回的相关商标个数为 11 个, 不相关商标为 1 个(“耐克”)。文中法取得良好性能主要归功于深度学习耦合稀疏语义度量技术, 通过由卷积与池化组成的无监督学习对输入商标图像进行多层特征提取, 得到特征向量。通过 L2-SVM 进行分类, 利用特征向量进行训练, 获得多级联特征。在相似度量阶段, 定义了一种稀疏语义度量算子, 通过对商标图像的多级联特征和用户标签信息的异构数据结构进行探索。此外, 引入一种混合范

数作为相似度量的稀疏约束, 以抑制原始输入空间中的冗余特征维数与噪声, 并进行优化, 从而获得了较优的商标检索性能。文献[4]采用的角点描述符对颜色与噪声的辨别较弱, 因此其对商标检索适应性不强。文献[5]中采用的 SIFT 主要依靠于局部像素的梯度方向, 易使主方向计算不精确, 从而影响特征向量提取, 导致对形状表示能力不足。文献[6]中易受到噪声等各种影响时, 颜色和空间特征的可靠性可能以不同的方式变化, 而其权重无法自适应变化, 导致检索效果有所下降。



图5 检索目标

Fig.5 Retrieval target

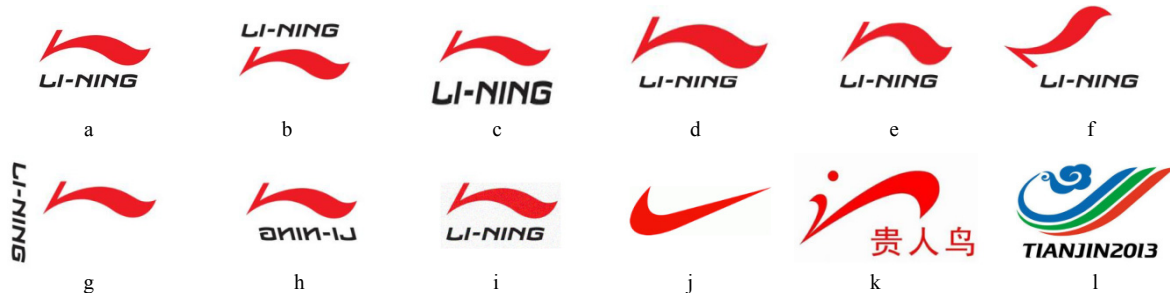


图6 文献[4]检索结果

Fig.6 Document [4] retrieval results

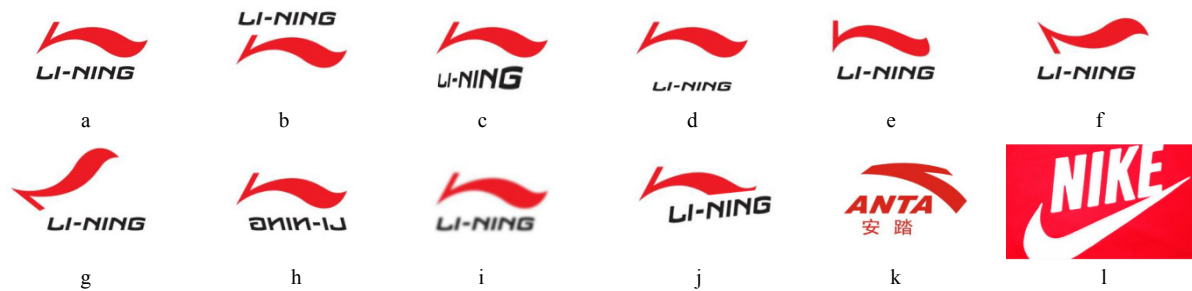


图 7 文献[5]检索结果
Fig.7 Document [5] retrieval results

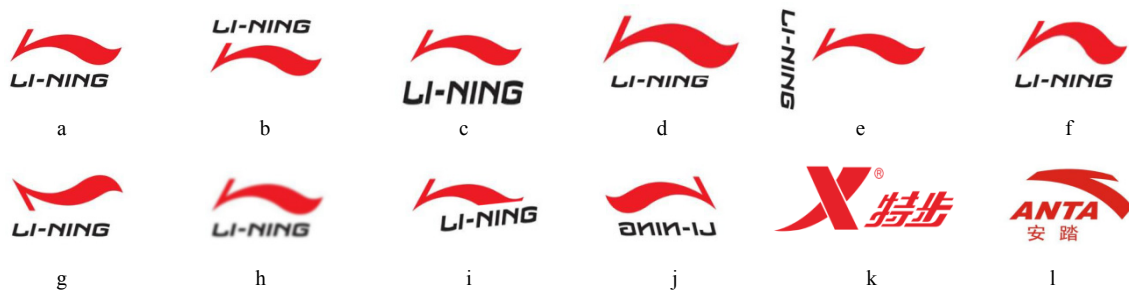


图 8 文献[6]检索结果
Fig.8 Document [6] retrieval results

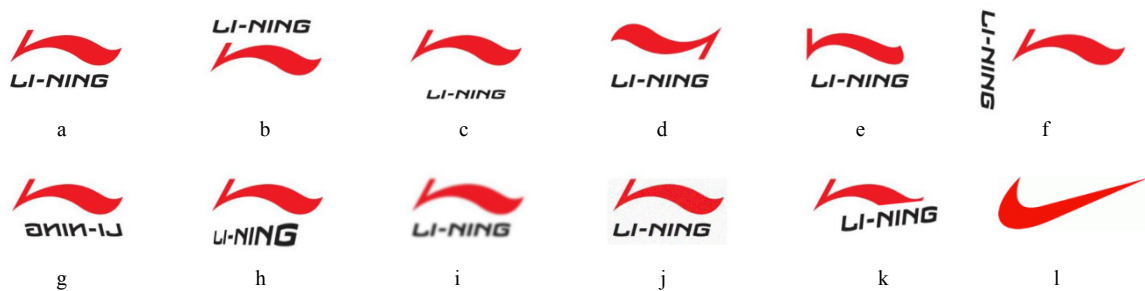


图 9 文中算法检索结果
Fig.9 Retrieval results of the proposed algorithm

图 10 显示了不同算法的 P-R 曲线图与 F_1 值，从图 10 中看出，文中算法的 P-R 曲线相对较稳定，表现更优异。图 10b 中的 F_1 值反应了算法的综合检索性能，可得出，文中算法的检索能力更强，具有更高的检索精度与鲁棒性，能够适应各种变化的商标。

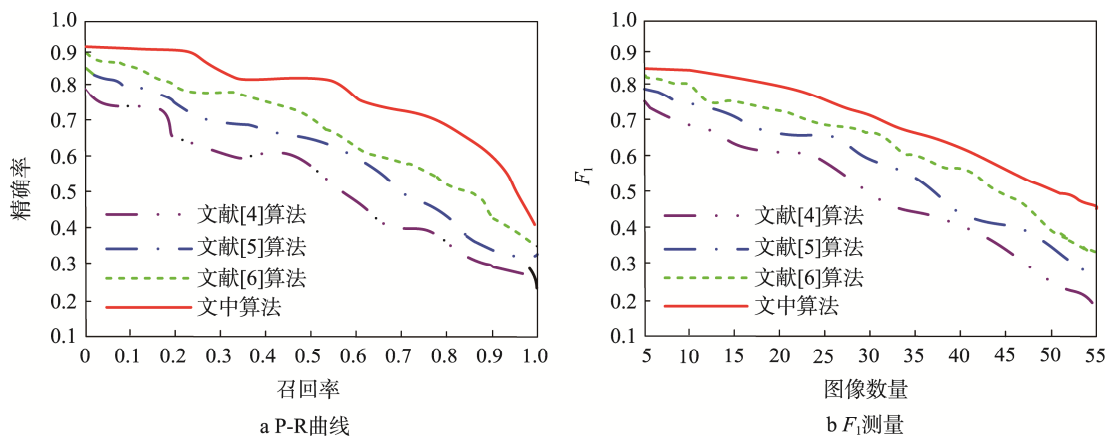


图 10 不同算法测量结果
Fig.10 Measurement results of different algorithms

为进一步测量商标检索的性能,对其时间成本进行统计,见表1。表1中分别给出了不同算法在返回商标数 $N=12$ 与 $N=24$ 的耗时情况,当 $N=12$ 时,耗时为 1.56 s,当 $N=24$ 时,耗时为 3.48 s。文中算法取得较好效率的主要原因是采用了混合 $\ell_{2,1}$ -范数的行稀疏,消除了原始空间中的一些噪声或冗余特征维数。文献[4]算法中对角点的邻域特征与分布特征计算时需消耗较多时间。文献[5]中 SIFT 特征维度巨大,需占用较大内存与时间成本。文献[6]中对颜色与空间特征的融合时效率较低。

表 1 不同算法的检索时间
Tab.1 Retrieval time of different algorithms

算法	时间/s	
	$N=12$	$N=24$
文献[4]	2.14	5.86
文献[5]	11.36	36.58
文献[6]	5.69	16.37
文中算法	1.56	3.48

5 结语

文中提出一种深度学习耦合稀疏语义度量的商标图像检索方案,保持视觉空间和文本空间之间的一致性,有效降低了语义鸿沟。通过深度学习,摒弃了不完整的或个性化的用户标签,提取了商标的多级联特征,并利用稀疏语义度量,抑制某些噪声与降低冗余特征维度。然后通过迭代算法证明了优化问题的收敛性。最后,通过实验表明了提出的算法抑制了噪声、降低冗余特征维数,具有优异的检索精度与鲁棒性。

参考文献:

- [1] GIUSEPPE A, FABRIZO F, LUCIA V. Aggregating Binary Local Descriptors for Image Retrieval[J]. Multimedia Tools and Applications, 2018, 77(5): 5385—5415.
- [2] 黄友文, 黄雅兰. 一种基于 Harris 与 SIFT 算子结合的商标搜索方法[J]. 电视技术, 2014, 38(3): 35—38.
HUANG You-wen, HUANG Ya-lan. Method of Trademark Search Based on Harris and SIFT[J]. Video Engineering, 2014, 38(3): 35—38.
- [3] LU Wei-ding, HENG Jiang. A Document Expansion Framework for Tag-based Image Retrieval[J]. Aslib Journal of Information Management, 2018, 70(1): 47—65.
- [4] 刘剑英. 基于角点描述和区域特征的商标图像检索[D]. 西安: 西安电子科技大学, 2017.
- [5] AKRITI N. Trademark Image Retrieval Using Weighted Combination of Sift and HSV Correlogram[J]. International Journal of Computer Applications in Technology, 2016, 54(1): 61—67.
- [6] YAN Yi-jun, REN Jin-chang, LI Yin-sheng. Adaptive Fusion of Color and Spatial Features for Noise-robust Retrieval of Colored Logo and Trademark Images[J]. Multidimensional Systems and Signal Processing, 2016, 27(4): 945—968.
- [7] 王伟凝, 王励, 赵明权. 基于并行深度卷积神经网络的图像美感分类[J]. 自动化学报, 2016, 42(6): 904—914.
WANG Wei-ning, WANG Li, ZHAO Ming-quan. Image Aesthetic Classification Using Parallel Deep Convolution Neural Networks[J]. Acta Automatic Sinica, 2016, 42(6): 904—914.
- [8] 罗富贵, 李明珍. 基于卷积核分解的深度 CNN 模型结构优化及其在小图像识别中的应用[J]. 井冈山大学学报(自然科学版), 2018, 39(2): 31—38.
LUO Fu-gui, LI Ming-zhen. Structural Optimization of Deep CNN Model Based on Convolution Kernel Decomposition and Its application in Small Image Recognition[J]. Journal of Jinggangshan University(Natural Science Edition), 2018, 39(2): 31—38.
- [9] 房卫东, 石志东, 单联海. 一种基于 BETA 分布抗 On-off 攻击的信任机制[J]. 系统仿真学报, 2015, 27(11): 2722—2728.
FANG Wei-dong, SHI Zhi-dong, SHAN Lian-hai. Trusted Scheme for Defending On-Off Attack Based on BETA Distribution[J]. Journal of System Simulation, 2015, 27(11): 2722—2728.
- [10] 王维智. 基于特征提取和特征选择的级联深度学习模型研究[D]. 哈尔滨: 哈尔滨工业大学, 2015.
WANG Wei-zhi. Research on Cascaded Deep Learning Model Based on Feature Extraction and Feature Selection[D]. Harbin: Harbin Institute of Technology, 2015.
- [11] LIU Jing, LI Ze-chao, LU Han-qing. Sparse Semantic Metric Learning for Image Retrieval[J]. Multimedia Systems, 2014, 20(6): 635—643.
- [12] ESKANDAR N. Existence and Convergence Theorems for Bregman Best Proximity Points in Reflexive Banach Spaces[J]. Journal of Fixed Point Theory and Applications, 2018, 20(1): 453—468.
- [13] 郭强. 基于 JBLD 度量的快速记忆梯度距离度量方法[J]. 光电技术应用, 2016, 31(5): 51—54.
GUO Qiang. Fast Memory Gradient Algorithm Based on Jensen-Bregman LogDet Metric[J]. Electro-optic

- Technology Application, 2016, 31(5): 51—54.
- [14] 李瑞雪, 夏斌, 袁文浩. 基于多元变量 Taylor 级数展开模型的定位算法[J]. 计算机应用研究, 2016, 33(6): 1853—1856.
- LI Rui-xue, XIA Bin, YUAN Wen-hao. Localization Algorithm Based on Multivariable Taylor Series Expansion Model[J]. Application Research of Computers, 2016, 33(6): 1853—1856.
- [15] 柯圣财, 赵永威, 李弼程. 基于卷积神经网络和监督核哈希的图像检索方法[J]. 电子学报, 2017, 45(1): 157—163.
- KE Sheng-cai, ZHAO Yong-wei, LI Bi-cheng. Image Retrieval Based on Convolutional Neural Network and Kernel-Based Supervised Hashing[J]. Acta Electronica Sinica, 2017, 45(1): 157—163.
- [16] 相子喜, 吕学强, 张凯. 基于有向图模型的多模态新闻图像检索研究[J]. 科学技术与工程, 2016, 16(3): 78—84.
- XIANG Zi-xi, LYU Xue-qiang, ZHANG Kai. Research on Multimodal News Image Retrieval Based on Directed Graph Model[J]. Science Technology and Engineering, 2016, 16(3): 78—84.