**引用格式**:徐炎,曹春萍. 语义核 SVM 结合改进 EMD 跨越语义鸿沟[J]. 轻工学报,2019,34(3):77 – 83.
**中图分类号**:TP391.1 **文献标识码**:A
**DOI**:10.3969/j.issn.2096 – 1553.2019.03.009
**文章编号**:2096 – 1553(2019)03 – 0077 – 07

# 语义核 SVM 结合改进 EMD 跨越语义鸿沟

## Crossing semantic gap by semantic kernel SVM combined with improved EMD

徐炎,曹春萍
XU Yan, CAO Chunping

上海理工大学 光电信息与计算机工程学院,上海 200093
*School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai* 200093, *China*

**关键词**:
语义鸿沟;语义核;支持向量机;改进 EMD

**Key words**:
semantic gap; semantic kernel; SVM; improved EMD

**摘要**:针对跨越语义鸿沟方法中未考虑文本间语义相关性和样本数量增加时计算量过大的问题,提出了一种语义核 SVM 结合改进 EMD 跨越语义鸿沟方法. 该方法首先考虑到文本特征间的语义关系,提取与图像共生的文本关键词,结合 HowNet 通用本体库和内部统计特征构造语义核函数,然后将语义核函数嵌入 SVM 进行关键词分类,得到最佳候选关键词,从而解决文本间语义相关性问题;再通过最佳减小矩阵对 EMD 算法进行改进,从而减小计算量. 对比实验结果表明,该方法充分利用了与图像共生的文本特征间的语义关系,标注准确率明显高于其他 3 种方法,且标注时间缩短为其他方法的 1/5 左右.

**收稿日期**:2018 – 12 – 16
**基金项目**:国家自然科学基金资助项目(61472256)
**作者简介**:徐炎(1993—),男,江苏省南通市人,上海理工大学硕士研究生,主要研究方向为跨媒体关联理解与深度挖掘.
**通信作者**:曹春萍(1968—),女,甘肃省兰州市人,上海理工大学副教授,硕士,主要研究方向为信息管理与决策支持系统.

**Abstract**: Aiming at the problem that the semantics relation among texts is not considered and the amount of computation is too large while samples increases in crossing semantic gap methods, a method of crossing semantic gap was proposed based on semantic kernel SVM combined with improved EMD. Firstly, to solve the semantic relation problem among texts, the method constructed the semantic kernel function based on taking semantic relations of text features into consideration, extracting text features coexisting with images and combining HowNet common ontology repository. Then the semantic kernel function was embedded into the SVM to classify keywords for obtaining best candidate words. Secondly, the algorithm improved EMD with best decreasing matrix to cut down the amount of computation. The experiment result showed that the method proposed takes full consideration of semantic relation in the texts related, the annotation accuracy rate was obviously higher than the other 3 methods and the annotation time was cut down to 1/5 of before.

## 0　引言

跨媒体语义研究针对的是混合在一起的语义内容相同但模态不同的信息对象[1]. 共生的文本和图像作为常见的跨媒体语义研究对象一直是业界研究的热点,相关研究主要集中在图像语义标注、语义图像检索 SIR(semantic image retrieval)等方面[2-4]. 无论是图像语义标注还是图像检索,其关键步骤都是描述图像内容及其语义,其中语义鸿沟作为主要的难点一直制约着标注和检索的准确率. A. W. M. Smeulders 等[5]将语义鸿沟定义为:在给定的情形下,用户从视觉数据中获取的信息与其对视觉数据的理解间存在的不一致. 为了消除视觉数据的多义性,以便获得准确的语义表达,研究人员利用跨媒体特征,从不同模态的数据中寻找语义相关信息,从而对视觉特征的语义进行限定. Y. Cao 等[6]提出,深度视觉语义哈希模型(DVSH)通过端到端的深度学习架构生成图像和语句的哈希码可用以跨越语义鸿沟. 该方法虽然考虑到图像与文本间的语义特征,但忽略了文本内在的语义关联,且深度学习架构对计算力要求较高. B. Wang 等[7]提出的对抗跨模检索方法 ACMR,采用基于对抗性学习的方法获得可以将图像与文本直接进行比较的共享子空间,用以跨越语义鸿沟. 该方法的缺点在于:采用三重限制获取共享子空间中的统一表征,而当样本数据量增大时计算量增长过快. V. S. Tseng 等[8]提出的 FMD 模型,通过结合 Web 页面中与带标注信息的图像内容相关的文字描述部分,实现用户特定需求的语义理解. FMD 模型分为三步:首先构建基于图像分割的视觉特征模型,识别图像中的对象并进行特征提取,通过比较不同图像间对象的相似度,得到已标注的词与对象的关系;然后构建基于决策树的文字模型,利用 C4.5 算法对由相关文本中提取的关键词进行分类,得到最合适的候选词;最后融合两种模型,根据最短视觉距离将最合适的候选词标注图像,从而跨越语义鸿沟. 在提取关键字并利用决策树分类的过程中,FMD 模型仅采用去停用词和词干化处理相关文本得到待分类关键词,忽略了文本特征间的语义关系. 此外,由于 C4.5 决策树算法存在因训练集规模小而结果不可靠的缺点,导致最终候选词不可靠[9]. 同时,通过计算不同图像中对象的相似度得到最短视觉距离的方法会因训练样本数量不足而导致准确率不高.

支持向量机 SVM(support vector machine)在解决小样本、非线性和高维度模式识别问题方面具有良好的性能,基于语义核函数的 SVM 可以结合文本特征间的语义关系对文本进行分类[10]. EMD(earth mover's distance)算法是计算跨模态数据相似度的常用方法之一,其定义为:将一个分布变换为另一个分布所需的最小

工作量,作为距离函数有一个非常好的特点是存在下界. 基于此,本文拟对 EMD 进行改进以减小计算量,并将语义核 SVM 结合改进 EMD,以解决未考虑文本特征间语义关系的问题,进而提高标注的准确率,在样本数据量增大时减小所需计算量.

## 1 EMD 算法及其改进

### 1.1 EMD 基本原理

EMD 是使用最广泛的衡量两种模态对象相似度的方法之一,当样本量较小时,其计算次数少的优势尤为明显. 设 $\boldsymbol{P}=\{p_1,p_2,\cdots,p_n\}$ 和 $\boldsymbol{Q}=\{q_1,q_2,\cdots,q_n\}$ 为 $n$ 维直方图,其中 $p_i$ 和 $q_i$ 是直方柱(也称权重或概率). 矩阵 $\boldsymbol{D}=[d]$ 称为地面距离矩阵,其中 $d_{ij}$ 是 $p_i$ 与 $q_i$ 之间的地面距离. 地面距离可以由任意度量距离定义,例如欧氏距离和曼哈顿距离. 当直方柱由 $p_{ij}$ 转换为 $q_{ij}$ 时,直方柱中发生变化的部分为 $f_{ij}$,矩阵 $\boldsymbol{F}=[f_{ij}]$ 称为流矩阵. 这一转换过程定义为 $f_{ij}$ 与 $d_{ij}$ 的乘积. $\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 之间的 EMD 定义为分布 $\boldsymbol{P}$ 转换为分布 $\boldsymbol{Q}$ 所需的最小工作量,即

$$EMD(\boldsymbol{P},\boldsymbol{Q})=\min\Big\{\sum_{i=1}^{n}\sum_{j=1}^{n}f_{ij}\cdot d_{ii}\Big\}$$

其中,$\forall i,j:f_{ij}\geqslant 0$,$\forall i:\sum_{j=1}^{n}f_{ij}=p_i$,且 $\forall j:\sum_{i=1}^{n}f_{ij}=q_i$.

### 1.2 EMD 算法改进

EMD 的投影下界是影响计算复杂度的主要因素. 设 $\boldsymbol{S}_j$ 为 $\boldsymbol{R}^d$ 中的单位向量,那么

$$EMD(\boldsymbol{P},\boldsymbol{Q})\geqslant EMD(proj\boldsymbol{S}_j(\boldsymbol{P}),proj\boldsymbol{S}_j(\boldsymbol{Q}))$$

其中,$proj\boldsymbol{S}_j(\boldsymbol{P})$ 为 $\boldsymbol{P}$ 在向量 $\boldsymbol{S}_j$ 上的投影(也是 $\boldsymbol{Q}$ 的相似),即

$$proj\boldsymbol{S}_j(\boldsymbol{P})=(p_1,\cdots p_n,t_1,\cdots,t_n),t_i=\boldsymbol{S}_j^t\cdot b_i$$

其中,$p_i$ 为权重,$t_i$ 为投影直方条. 设 $\boldsymbol{S}=(S_1,\cdots,S_{d'})$ 为 $\boldsymbol{R}^d$ 中的正交轴,那么

$$EMD(\boldsymbol{P},\boldsymbol{Q})\geqslant\frac{1}{\sqrt{d'}}\sum_{i=1}^{d'}EMD(proj\boldsymbol{S}_j(\boldsymbol{P}),proj\boldsymbol{S}_j(\boldsymbol{Q}))$$

$\boldsymbol{P}$ 和 $\boldsymbol{Q}$ 之间的 EMD 投影下界为一组正交向量的 EMD 之和除以向量数的平方根,单一投影的计算时间复杂度为 $O(n)$. 当所需计算对象数量增加时,EMD 算法计算更耗时[13]. 为了解决这一问题,结合 EMD 算法计算低维直方图比高维直方图快速这一特点,利用 $n\times n'(n'<n)$ 维矩阵变换降低矩阵维度的思路,对 EMD 算法进行如下改进.

改进的 EMD 算法的计算方法为:投射矩阵 $\boldsymbol{G}^{(0)}$ 的初始值为随机生成的正交矩阵,在第 $k$ 次迭代训练中,先通过固定矩阵 $\boldsymbol{G}$ 得到变换矩阵 $\boldsymbol{E}$,其中

$$\boldsymbol{E}^{(k)}=\min_{E^{(k)}\in J(p^S,p^T)}\sum_{s=1}^{|V^S|}\sum_{t=1}^{|V^T|}\boldsymbol{E}_{st}^{(k)}d(\boldsymbol{G}^{(k-1)}\boldsymbol{w}_s,\boldsymbol{w}_t)$$

$$\boldsymbol{G}^{(k)}=\min_{G^{(k)}\in R^{d\times d}}\sum_{s=1}^{|V^S|}\sum_{T=1}^{|V^T|}\boldsymbol{E}_{st}^{(k)}d(\boldsymbol{G}^{(k)}\boldsymbol{w}_s,\boldsymbol{w}_t)$$

$$\text{s.t.}\ (\boldsymbol{G}^{(k)})^{\mathrm{T}}\boldsymbol{G}^{(k)}=\boldsymbol{I}$$

然后,给定变换矩阵 $\boldsymbol{E}$ 再计算得到投射矩阵 $\boldsymbol{G}$. $\boldsymbol{V}^S$ 为变换前向量空间,$\boldsymbol{V}^T$ 为变换后的向量空间,$\boldsymbol{w}_s$ 和 $\boldsymbol{w}_t$ 分别为变换前后的特征向量. 由于正交的限制,计算有一定难度. 但是如果选择平方欧几里得距离作为地面距离函数 $d$,则目标问题转化为奇异值分解问题:

$$\boldsymbol{G}^{(k)}=\boldsymbol{U}\boldsymbol{V}^T$$

式中的矩阵 $\boldsymbol{U}$ 和 $\boldsymbol{V}$ 可由下式计算得到:

$$\sum_{s=1}^{|V^S|}\sum_{t=1}^{|V^T|}\boldsymbol{E}_{st}^{(k)}\boldsymbol{w}_s\boldsymbol{w}_t^{\mathrm{T}}=\boldsymbol{U}\sum\boldsymbol{V}^T$$

投射矩阵 $\boldsymbol{G}$ 经过 $n\times n'$ 维矩阵变换得到最佳减小矩阵 $\boldsymbol{D}'=[d'_{i'j'}]$,最后得到降维后直方图的地面距离. 最佳减小距离矩阵 $\boldsymbol{D}'$ 可以确保 EMD 中降维后的直方图永远小于原始直方图,因此计算量大大减少.

## 2 基于语义核函数的 SVM

语义核函数可通过将线性不可分问题中的

数据点映射到高维空间,把问题转化为线性可分问题,然后通过计算高维空间中数据点间的距离来实现分类. 此方法的优势在于通过在原始空间中的计算即可得到高维空间中数据点间距离,且文本数据的稀疏性使得计算距离非常有效. 语义核函数的概念由 G. Siolas[11] 首次提出,实际上是利用文档维度正交地调整原始空间中的数据向量. 对于任意向量 $\boldsymbol{x},\boldsymbol{z}\in\boldsymbol{X}$,当核函数 $K(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{\Phi}(\boldsymbol{x})\cdot\boldsymbol{\Phi}(\boldsymbol{z})$ 时 $K$ 有效,其中 $\boldsymbol{X}$ 是输入向量空间,$\boldsymbol{\Phi}$ 是向量空间 $\boldsymbol{X}$ 到特征空间 $F$ 的映射. 对于任意文档向量 $\boldsymbol{x},\boldsymbol{z}\in\boldsymbol{X}$,语义核函数 $K(\boldsymbol{x},\boldsymbol{z})=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{M}\boldsymbol{z}$,其中 $\boldsymbol{M}$ 为对称矩阵,称为度量矩阵,每个矩阵代表输入空间的 $X$ 维之间的语义相似度[12]. 语义核函数依赖于度量矩阵 $\boldsymbol{M}$ 的构造. 按知识来源,语义核函数可分为基于外部知识源和内部统计特征的语义核函数两部分. 此处的外部知识源为本体,内部统计特征为语义相似度. 因此,语义核函数的构造方法为

$$\overline{\overline{\varphi_{t_i,t_j}}}=\begin{cases}\overline{\varphi_{t_i,t_j}} & t_j \text{ 的范围是与 } t_j \text{ 最相似的}\\ & k \text{ 个特征值}\\ 0 & \text{其他}\end{cases}$$

$$\overline{\varphi_{t_i,t_j}}=k_1\times Rel_{\mathrm{HowNet}}(t_i,t_j)+k_2\times Sim_{\mathrm{syn}}(t_i,t_j)$$

其中,$\overline{\overline{\varphi_{t_i,t_j}}}$ 对应特征矩阵中的元素; $k$ 用来控制与 $t_i$ 相关联的词数; $Rel_{\mathrm{HowNet}}$ 为 $t_i$ 和 $t_j$ 在本体库 HowNet 中的语义相似度; $Sim_{\mathrm{syn}}$ 为 $t_i$ 与 $t_j$ 的同义词间的相似度; $k_1$, $k_2$ 为平滑系数,且 $k_1+k_2=1$. 当训练文本充足时,基于统计特征的语义核函数性会更好. 反之,当训练文本数量不足或者无法依据统计特征得到语义关系时,基于本体的语义核函数性会更好. 因此,结合这一特性可得

$$\boldsymbol{N}=\boldsymbol{I}+\gamma_1\times\boldsymbol{N}_{\mathrm{stat}}+\gamma_2\times\boldsymbol{N}_{\mathrm{ontology}}$$

其中,$\boldsymbol{I}$ 为单元矩阵, $\boldsymbol{N}_{\mathrm{stat}}$ 为基于统计特征的特征词相关矩阵, $\boldsymbol{N}_{\mathrm{ontology}}$ 为基于本体的特征词相关矩阵, $\boldsymbol{N}$ 为最终与特征词相关的矩阵, $\gamma_1$, $\gamma_2\in(0,1)$. 分类过程如图 1 所示.

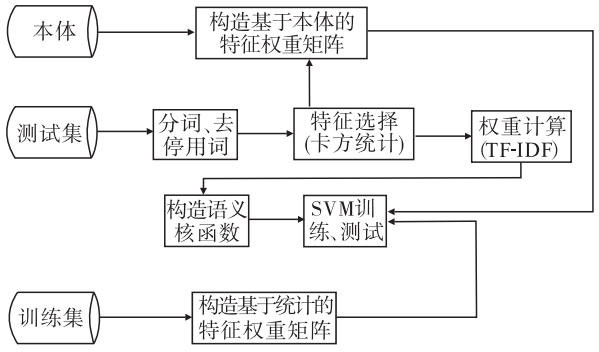


图 1　基于语义核 SVM 的中文文本分类过程

Fig. 1　The Chinese text classifying process based on semantic kernel SVM

## 3　语义核 SVM 结合改进的 EMD 跨越语义鸿沟方法

为解决最短视觉距离因训练样本数量不足而导致准确率不高的问题,采用改进 EMD 计算关键词和图像间的距离. 为解决待分类关键词忽略文本特征间语义关系,以及 C4.5 算法因训练集规模小而导致最终候选词不可靠的缺点,采用基于语义核函数的 SVM 对关键词进行分类.

语义核 SVM 结合改进的 EMD 跨越语义鸿沟的具体过程如下:模型输入为共生图文,首先构建基于图像分割的视觉特征模型,提取图像的颜色、纹理和形状特征,识别图像中的对象. 其次,对相关文本分词、去停用词,通过卡方统计选择语义相似度较高的候选词,由 TF-IDF 计算得到候选词的权重,构造语义核函数. 再次,结合 HowNet 通用本体库构建基于语义核函数的 SVM 对候选词分类,得到最佳候选词. 最后,由改进 EMD 计算出最佳候选词与图像中对象的距离,选择距离最短的候选词描述图像,从而跨越语义鸿沟. 该方法流程如图 2 所示.

## 4 图像标注实验结果与分析

为了测试本文方法的图像描述准确率,进行图像标注实验. 实验数据来源于百度搜索得到的 15 000 个网页中的图文共生数据,网页中的图片均带有一个或多个标签,共 10 个类别的主题,分别为熊猫、狗、猫、汽车、轮船、飞机、冰激凌、面条、电话和杯子. 在 15 000 张图片中取 10 000 张作为训练数据,其余 5000 张作为测试数据. 实验环境为 64 位 Windows10 操作系统,12 G 内存和 GNU Octave, version 4.4.1,图像分割程序采用 Normalized Cut[14]. 图 3 为部分实验图像分割结果. 从图 3 可以看出,图像主体部分能被很好地识别出来.

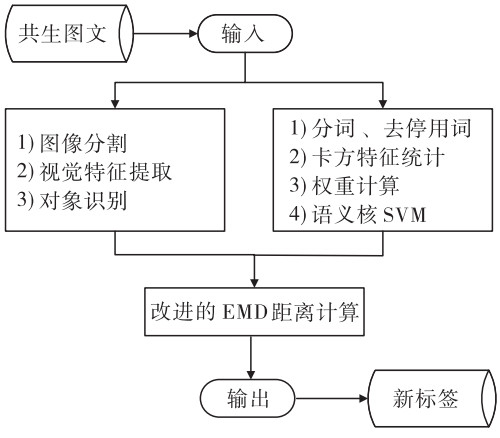


图 2 本文提出方法流程

Fig. 2 Process of the method proposed in this paper

为了验证本文方法的有效性,将本文模型 SVMemd 分别与基于文本的 SVM(SVMt)、基于视觉的 SVM(SVMv)和 FMD 模型进行对比. 在此引入精度、召回率、$F$ 值作为评价指标,其定义如下:

$$精度 = \frac{|正确标注标签数量|}{|标注结果标签数量|}$$

$$召回率 = \frac{|正确标准标签数量|}{|实际标签数量|}$$

$$F\ 值 = \frac{2 \times 精度 \times 召回率}{精度 + 召回率}$$

测试样本数量为 500 ~ 5000 张图片时,4 种方法的 $F$ 值对比结果见图 4. 由图 4 可以看出,当训练数目一定、测试数目增加时,准确率均呈一定下降的趋势;本文方法相比较基于传统 SVM 的标注方法,$F$ 值提升 18.7%,比 FMD 模型提高了 8.1%.

选择实际标签数量为 1 ~ 6 个的图片为测

a)熊猫

b)汽车

图 3 部分实验图像分割结果

Fig. 3 Some image segmentation results

试数据,每种标签数量的图片均为 200 张时,4 种方法的 $F$ 值对比结果见图 5. 由图 5 可以看出,当实际标签数量增加,即图像内容更复杂时,SVMt,SVMv 和 FMD 的 $F$ 值迅速下降,由本文方法产生的 $F$ 值虽然也呈下降趋势,但仍保持较高位,标注准确率明显高于其他 3 种方法.

对 10—100 张图片进行标注,4 种方法所花费的时间结果见图 6. 由图 6 可见,本文提出的方法有效减小了计算量,缩短了标注时间,标注时间缩短为其他标注方法的 1/5 左右.

## 5　结语

本文采用基于语义核的 SVM 结合改进 EMD 的跨越语义鸿沟方法对图像进行标注. 该方法考虑与图像共生文本间的语义特征,通过基于外部知识源和内部统计特征的语义核函数得到文本特征间的语义关系,采用基于语义核的 SVM 对相关文本的文本特征分类,得到最佳候选关键词,同时为了减少计算时间,提出减小变换矩阵改进 EMD 算法衡量文本特征与图像对象的距离,选取与图象距离最小的关键词作为标注词. 实验结果表明,本文方法有效提高了标注准确率,减小了计算量. 然而当图像内容变得复杂时,本文方法由于对象识别准确率下降,可能导致标注准确率下降,因此今后的研究方向将集中在提高图像内容识别的准确率上.

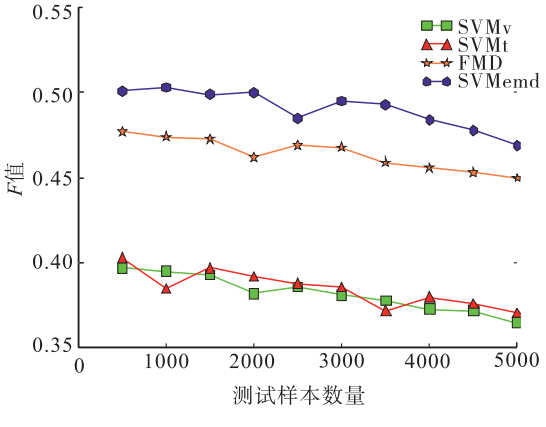


图 4　不同测试样本数量时 4 种方法的 $F$ 值对比

Fig. 4　Comparison of $F$ values of 4 methods for different test samples

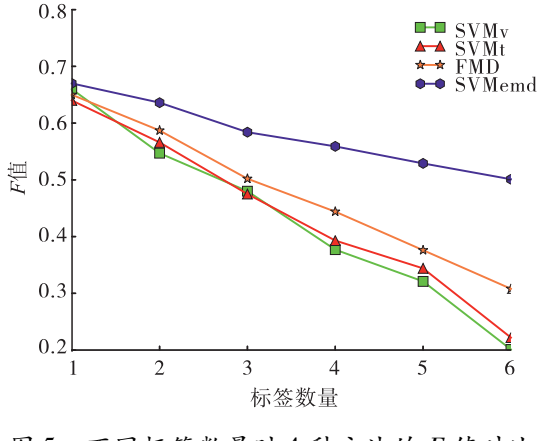


图 5　不同标签数量时 4 种方法的 $F$ 值对比

Fig. 5　Comparison of $F$ values of 4 methods for different amount of labels

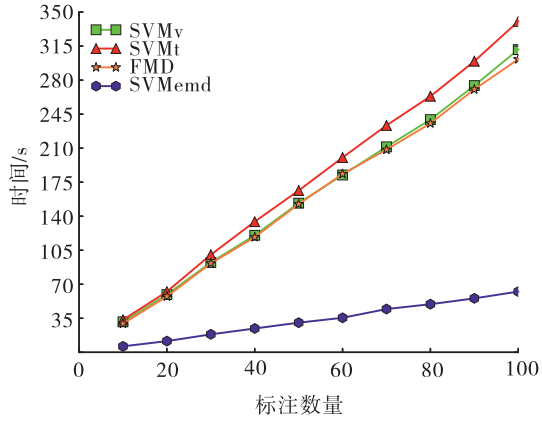


图 6　图像标注时间对比结果

Fig. 6　Comparison result of image label time

**参考文献:**

[1] WEI S, WEI Y, ZHANG L, et al. Heterogeneous data alignment for cross-media computing [C]// International Conference on Internet Multimedia Computing and Service. New York: ACM, 2015: 84.

[2] LU Z, PENG Y. Image annotation by semantic sparse recoding of visual content [C] // ACM International Conference on Multimedia. New York: ACM, 2012: 499.

[3] 许红涛,周向东,向宇,等. 一种自适应的 Web

图像语义自动标注方法[J]. 软件学报,2010,21(9):2183.

[4] VIKHAR P A, SHINKAR D V, MISHRA N. Improving the performance of CBIR system using relevance feedback [C] // International Conference and Workshop on Emerging Trends in Technology. New York: ACM, 2010: 554.

[5] SMEULDERS A W M, WORRING M, SANTINI S, et al. Content-based image retrieval at the end of the early years [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2000, 22(12): 1349.

[6] CAO Y, LONG M, WANG J, et al. Deep visual-semantic hashing for cross-modal retrieval [C] // Proceedings of the 22nd ACM SIGKDD International Comference. New York: ACM, 2016: 1445.

[7] WANG B, YANG Y, XU X, et al. Adversarial cross-modal retrieval[C] // ACM on Multimedia Conference. New York: ACM, 2017: 154.

[8] TSENG V S, SU J H, WANG B W, et al. Web image annotation by fusing visual features and textual information [C] // ACM Symposium on Applied Computing. New York: ACM, 2007: 1056.

[9] ELAIDI H, BENABBOU Z, ABBAR H. A comparative study of algorithms constructing decision trees: ID3 and C4. 5 [C] // LOPAL '18 Proceedings of the International Conference on Learning and Optimization Algorithms: Theory and Applications. New York: ACM, 2018: 26.

[10] MAI F, HUANG L, TAN J, et al. The research of semantic kernel in SVM for chinese text Classification[C] //2017 International Conference on inteligent Information Processing. New York: ACM, 2017: 1.

[11] SIOLAS G. Support vector machines based on a semantic kernel for text categorization [C] // Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IEEE Computer Society. Piscataway: IEEE, 2000: 5205.

[12] 张玉峰,王志芳. 文本分类中的语义核函数研究[J]. 情报科学,2010(7):970.

[13] WAN X J, PENG Y X. The earth mover's distance as a semantic measure for document similarity[C] // Proceedings of the 14th ACM International Conference on Information and Knowledge Management. New York: ACM, 2005: 301.

[14] SHI J, MALIK J. Normalized cuts and image segmentation[J]. IEEE Transactions on Pattern Analysis and Maching Intelligence, 2000, 22(8): 888.