基于双向学习排序的跨媒体语义相似性度量方法

刘 爽 白 亮 于天元 贾玉华
(国防科学技术大学信息系统工程重点实验室 长沙 410073)

摘 要 随着互联网技术的迅猛发展,网络信息的呈现形式不断从简单的文本扩展到图像、声音、视频等多媒体表达形式。在多媒体信息检索领域中,传统方法往往在同一个特征空间中表示所有的媒体模式,并采取一对一的配对数据,或者利用单向排序实例作为训练样本进行检索。在此背景下,考虑了学习双向排序实例,进而实现了跨媒体检索的方法。在 Wikipedia 数据集上进行测试,实验结果表明,基于双向排序的跨媒体语义相似性度量方法具有更好的性能。

关键词 跨媒体表示,双向学习排序,隐空间,相似性度量
中图法分类号 TP393 **文献标识码** A

Cross-media Semantic Similarity Measurement Using Bi-directional Learning Ranking

LIU Shuang BAI Liang YU Tian-yuan JIA Yu-hua
(Key Laboratory of Information System Engineering, National University of Defense Technology, Changsha 410073, China)

Abstract With the rapid development of Internet technology, the presented forms of network information have extended from simple text to images, voice, video and other multimedia expression. In the field of multimedia information retrieval, the traditional methods often represent all of the media mode in the same feature space model. Existing methods take either one-to-one paired data or uni-directional ranking examples. In this paper, we considered learning bi-directional ranking examples in the cross-media retrieval. By analyzing the experimental results basing on the Wikipedia dataset, it is demonstrated better performance of the proposed method.

Keywords Cross-media representation, Bi-directional learning ranking, Latent space, Similarity measurement

1 引言

随着计算机、互联网和多媒体技术的迅速发展,网络信息的呈现方式层出不穷。用户更渴望通过信息搜索得到不仅仅是单一类型的内容,而是与之有关的包括图像、声音以及视频等多媒体数据。跨媒体是在多媒体的基础上,利用各种媒体的形式和特征,对相同或相关的信息用不同的媒体表述形式进行处理,由此产生的存储、处理、检索和共享等活动。因而,跨媒体检索是指实现文本、图像、视频等多模态信息之间的相互检索。

跨媒体检索技术研究中最大的挑战是多模态数据之间的"异构鸿沟"问题。即不同类型的多媒体数据虽然可以在语义层面上统一起来,但是它们对信息的表现形式各异,底层特征也不同。现有方法主要是通过利用特定的方法将不同媒体的形式内容进行融合,使得不同模态的数据表达同一语义,从而解决该问题。

传统的跨媒体检索方法按时间顺序可分为 6 类:1)采用 CCA [1] 的检索方法,提取特征对应的线性空间,讨论线性空间之间的相关性;之后再对不同模态的数据进行逻辑回归[2]。2)DL 方法[3]通过为多模态数据生成不同的字典来捕获异构特征。3)多模态的 LDA 模型[4]实现了从相关度较小的文本描述检索出相应的图像拓展为学习多模态数据的联合分布,例如多模态文档的随机域。4)PAMIR[5]是解决直接由文本查询到图像排序问题的首次尝试。PAMIR 规划跨媒体检索问题的方式类似于 RankingSVM,并且应用 Passive-Aggressive 算法得到一个有效的训练过程。5)Bi-CMSRM[6]是优化双向排序的损失,实现跨媒体双向检索的过程。6)C^2MLR[7]是将基于配对的多模态数据通过局部比对和全局比对的训练方法实现跨媒体检索的方法。

常用的跨媒体搜索方法假定训练集是一一配对的,或者只进行了单向排序实例。由于受训的模型不能够用于反方向的跨媒体检索,因此泛化性能有限,不能捕捉查询方式的潜在结构,并且检索效率低下。本文假定双向排序实例可供训练,目标是学习一个可以应用于图像到文本、文本到图像的双向查询检索隐空间。该空间作为一个潜在的语义空间,能够使两个具有相似语义的数据对象彼此接近。本文以图片和文字两种媒体数据为例,把两类媒体数据映射到共同空间,并在一个统一算法框架下利用结构化的支持向量机实现最优的优化排名评价方法。

2 基于双向学习排序的跨媒体相似性度量

利用支持向量机算法训练一个可以将相似的文本图像对

刘 爽(1995-),女,硕士生,主要研究方向为多媒体信息处理;白 亮(1978-),男,博士,副教授,主要研究方向为多媒体信息处理与应用,E-mail:xabpz@163.com;于天元(1992-),男,硕士生,主要研究方向为多媒体信息处理;贾玉华(1992-),男,硕士生,主要研究方向为多媒体信息处理。

与非相似的文本图像对进行最大程度分离的超平面。其目标是把原来高维的文本和图像向量通过线性映射函数对其进行降维，得到一个嵌入的隐空间。在隐空间中，必定存在着一个超平面，使得原始空间中同类的文档在隐空间中被划为同一类。

2.1　特征向量提取

计算机处理信息的过程类似于人的大脑，但在进行模式识别前要通过一定的处理使其变成可以利用计算机进行处理的结构化信息。

2.1.1　文本特征向量的提取

利用 TF-IDF 权重算法的 BoW 模型得到文本特征向量。首先将文本看作由一些词汇构成的集合，这些词汇彼此之间是独立的，通过词汇将其分类。继而利用 TF-IDF 算法为词典中的词赋予一定的权重，即将文档中出现频率较高的词汇以及出现频率低但却可以代表某一类文章的词汇均赋予较高的权重。

2.1.2　图像特征向量的提取

将文本检索领域的技巧直接应用到图像检索领域，采用将一张图片拆分为视觉单词的方式。BoVW 通过用视觉单词的方式表示图像，大大提高了图像的检索效率。该过程共分为 4 个步骤：1）自动检测特征兴趣的区域，图像的主要内容集中于少部分关键区域；2）计算该特征区域/点的描述子，用 SIFT 进行特征提取；3）利用视觉词典将特征描述子映射为视觉单词；4）统计每个视觉在图像中的出现频率，生成视觉单词直方图。

2.2　线性映射函数

本文假设所有的向量都是列向量，并且上标 T 代表矩阵或者向量的转置。m 代表文本特征空间的维度，n 代表图像特征空间的维度。给定训练集中有 $N+M$ 个实例，其中有 N 个文本查询的实例和 M 个图像查询的实例。设一次查询为 q，可能代表一个图像查询 p 或者一个文本查询 t。对应的检索集合 d 可以是图片集 p 或者是文本集 t。

每个文本查询实例包含一个文本 $t_i \in R^m (i=1,\cdots,N)$、相对应的检索图片的集合 p_i 以及关于该文本查询的图像真实排名 $y_i^* \in Y$，其中 Y 代表所有可能的排列集合。类似地，一次图片查询实例包含一个图片 $p_j \in R^n (j=N+1,\cdots,N+M)$和相对应的检索文本的集合 t_j 以及在这个文本集上的真实排名 $y_j^* \in Y$。为简单起见，省略下标 i 和 j，用 (q,d,y) 表示一个通用的训练实例。

用矩阵表示配对的关系，$Y \subset \{-1,0,+1\}^{|d|\times|d|}$，其中运算符 $|\cdot|$ 表示一个集合中的元素数。对于任何 $y \in Y$，如果文档 d_i 排在文档 d_j 之前，则令 $y_{ij}=+1$；若文档 d_j 排在文档 d_i 之前，则令 $y_{ij}=-1$；若文档 d_i 和 d_j 有同样的级别，则令 $y_{ij}=0$。假设矩阵只对应于有效的排名，服从反对称性和传递性。本文假设真实排名是只有两个等级值的弱排名，即相关的或者不相关的排名。对于任意的查询 q，令 d^+ 和 d^- 分别表示相关和不相关的文档集 d。例如，相关的文本文档集用 t^+ 表示，不相关的文本文档集用 t^- 表示。

学习分别把文本和图像映射到一个共同的隐空间的映射矩阵 U 和 V，要求相近语义的文本和图像在这个隐空间中的距离较小。给定一个文本 $t \in R^m$ 和图像 $p \in R^n$，用线性相似性函数来度量 t 和 p 之间的关联性，如式(1)所示：

$$f(t,p)=(Ut)^{\mathrm{T}}Vp \tag{1}$$

其中，$U \in R^{k\times m}$，$V \in R^{k\times n}$。U 是指将文本 t 从 m 维文本空间通过线性变换映射到 k 维的隐空间，V 是指将图片 p 从 n 维图像空间映射到 k 维的隐空间。因此，文本和图像映射到一个共同的 k 维隐空间，两者之间的相似性由 k 维空间中两个向量的点积测定，点积通常是用来测量文本向量之间的匹配程度[14]。

直观地说，式(1)中的线性模型尤其有助于解决发生在文本空间和图像空间中同义和多义词的问题。注意到潜在语义索引(LSI)[8]是用无人监管的单一方式对文本词语(同义词和多义词)之间的相关性进行刻画，而式(1)中的线性模型是尝试用监管的方式捕捉两个不同模态之间的相关性。通过约束方程(1)的形式的优点类似于 LSI：U 和 V 不仅引入了 k 维隐空间以便于更快地计算，并且用 k 维向量表示图片和文本比它们原始的维数 m 和 n 所用存储空间更小(k 的选择远小于 m 或 n)。类似于文献[9]，U 和 V 是不同的，并不需要假设文本和图像应以同样的方式嵌入到隐空间。本文中最重要的一点就是对 U 和 V 的学习。

通过监督的方式学习 U 和 V，特别是从训练实例的两个方向学习，可以充分提高检索性能。例如给定一个文本查询，对应一组图像的排序；给定一个图像查询，对应一组文本的排序。相似性函数 f 也可以考虑排序功能：给定文本查询 t 和图像集 q，通过 $f(t,p)$ 降序值排序，获得 q 的排列预测 y。如果 $f(t,p_i)>f(t,p_j)$，则 $y_{ij}=1$，否则 $y_{ij}=-1$。因此，本文旨在通过最小化下列各项实验排序风险，获得 U 和 V 的值，如式(2)所示。

$$R^{\Delta}(f)=\frac{1}{N}\sum_{i=1}^{N}\Delta(y_i^*,y_i)+\frac{1}{M}\sum_{j=N+1}^{N+M}\Delta(y_j^*,y_j) \tag{2}$$

其中，Δ 为损失函数，这里用 AP(平均精度)来定义损失函数：

$$\Delta_{ap}(y^*,y)=1-AP(rank(y^*),rank(y)) \tag{3}$$

2.3　基于双向学习排序的跨媒体相似性度量方法

基于双向学习排序的跨媒体相似性度量方法是在基于结构化的支持向量机框架下进行的[10]。该方法的目的是学习一个跨媒体的排序函数 $h: X \to Y$，输入空间 X 包括一次查询 q 和所有检索的目标文档 d，输出空间 Y 是基于检索的所有文档集的排序。类似于结构化的支持向量机，通过最大化判别函数 h，得出预测排名 y：

$$h(q,d)=\arg\max_{y\in Y} F(q,d,y;U,V) \tag{4}$$

其中，F 是由 U 和 V 确定参数的兼容性函数，用来测定 q,d,y 三者之间的兼容性。以文本查询图像为例，结合文献[11]中的偏序，定义兼容函数如下：

$$F(t,p,y)=\sum_{i\in p^+}\sum_{j\in p^-} y_{ij}\frac{(Ut)^{\mathrm{T}}V(p_i-p_j)}{|p^+|\cdot|p^-|} \tag{5}$$

由于 F 可以理解为对若干个成对的文档线性求和，因此可以分别优化每个 y_{ij} 来实现 F 的最大化。对于固定的 U 和 V，最大化函数 F 所得到的排序 y 与对函数 $f(t,p)=(Ut)^{\mathrm{T}}Vp$ 降序的原理相同。

因为 U 和 V 在求和式(4)中都是独立的，将 F 写成 U 和 V 的线性函数：

$$F(t,p,y)=\langle U^{\mathrm{T}}V,\Psi(t,p,y)\rangle \tag{6}$$

其中

$$\Psi(t,p,y)=t\sum_{i\in p^+}\sum_{j\in p^-} y_{ij}\frac{p_i^{\mathrm{T}}-p_j^{\mathrm{T}}}{|p^+|\cdot|p^-|} \tag{7}$$

其中，组合特征函数 $\Psi(t,p,y)$ 是对所有有关或无关的图像对的矢量差的求和。学习跨媒体排名函数 F，自然地扩展到结构化的 SVM 的想法，该方法要求映射到 U 和 V 空间的排序损失值和原始空间的排序损失值相近，这样才能保证隐空间是有效的，即：

$$\forall y\in Y:\delta F(t_i,p_i,y)\geqslant\Delta(y_i^*,y)-\zeta_{1,i} \tag{8}$$

其中：

$$\delta F(t_i,p_i,y)=F(t_i,p_i,y_i^*)-F(t_i,p_i,y) \tag{9}$$

同样，另一方向的检索过程是类似的。定义兼容函数：

$$F(p,t,y)=\sum_{i\in t^+}\sum_{j\in t^-}y_{ij}\frac{(Vp)^{\mathrm{T}}U(t_i-t_j)}{|t^+|\cdot|t^-|} \tag{10}$$

重写 F 作为 $V^{\mathrm{T}}U$ 的线性函数：

$$F(p,t,y)=\langle V^{\mathrm{T}}U,\Psi(p,t,y)\rangle \tag{11}$$

其中：

$$\Psi(p,t,y)=p\sum_{i\in t^+}\sum_{j\in t^-}y_{ij}\frac{t_i^{\mathrm{T}}-t_j^{\mathrm{T}}}{|t^+|\cdot|t^-|} \tag{12}$$

同理：

$$\forall y\in Y:\delta F(p_j,t_j,y)\geqslant\Delta(y_i^*,y)-\zeta_{2,j} \tag{13}$$

其中：

$$\delta F(p_j,t_j,y)=F(p_j,t_j,y_j^*)-F(p_j,t_j,y) \tag{14}$$

基于双向学习排序的跨媒体相似性度量方法主要适应于结构化的 SVM 学习最优的 U^* 和 V^*，分别把文本和图像映射到一个共同的隐空间，用 $\frac{\lambda}{2}\|U\|_F^2+\frac{\lambda}{2}\|V\|_F^2$ 取代标准的二次正则化 $\frac{\lambda}{2}\|\omega\|_2^2$，其中 λ 表示正则化参数。

优化问题表示如下：

$$\min_{U,V,\zeta_1,\zeta_2}\ \frac{\lambda}{2}\|U\|_F^2+\frac{\lambda}{2}\|V\|_F^2+\frac{1}{N}\sum_{i=1}^{N}\zeta_{1,i}+\frac{1}{M}\sum_{j=N+1}^{N+M}\zeta_{2,j} \tag{15}$$

对于在训练集中的每次查询 p，将式(8)－式(14)的一系列约束都添加到优化问题上。在预测过程中，给定 U 和 V，模型选择排序 $\bar{y}$ 是最大化 $F(q,d,y)$ 得到的。如果预测的排名是不正确的排名 $\bar{y}$，即 $F(q,d,\bar{y})<F(q,d,y^*)$，其中 y^* 是真正的排名，相应的松弛变量 ξ 最多是 $\Delta(y^*,\bar{y})$ 就可以满足约束。考虑到所有三元组 (t_i,p_i,y_i) 和 (p_j,t_j,y_j)，松弛变量的所有加权和 $\frac{1}{N}\sum_{i=1}^{N}\zeta_{1,i}+\frac{1}{M}\sum_{j=N+1}^{j=N+M}\zeta_{2,j}$ 就是方程(2)中实验风险 $R^{\Delta}(f)$ 的上限。

2.4 算法设计

本文学习 U 和 V 的方法改编自切割平面算法[13]。算法在两个步骤之间交替，第一个步骤是优化模型参数，即模型中的 U 和 V；另一个步骤是用一批最违反当前模型的排名即最差排名 $(\hat{y}_1,\cdots,\hat{y}_{N+M})$ 更新约束设置。$\hat{y}_i(i=1,\cdots,N)$ 是一次文本查询图像实例的一个排名，$\hat{y}_j(j=N+1,\cdots,N+M)$ 是一次图像查询文本实例的一个排名。当达到基于经验风险的准确性的停止准则时，算法终止，新一批约束的经验风险不会超过约束在公差范围内的当前设置 $\varepsilon(\varepsilon>0)$。

算法 1　基于双向学习排序的跨媒体相似性度量方法

输入：文本查询图像实例 (t_i,p_i,y_i^*)，$i=1,\cdots,N$，图像查询文本实例 (p_j,t_j,y_j^*)，$j=N+1,\cdots,N+M$，正则化参数 $\lambda>0$，精度公差阈值 $\varepsilon>0$

输出：映射参数 U 和 V，松弛变量 $\zeta_1\geqslant0$ 和 $\zeta_2\geqslant0$

step1　$W_1\leftarrow\Phi$，$W_2\leftarrow\Phi$

step2　重复

step3　找到最优的 U，V 和松弛变量 ζ_1,ζ_2：

$$\min_{U,V,\zeta_1,\zeta_2}\ \frac{\lambda}{2}\|U\|_F^2+\frac{\lambda}{2}\|V\|_F^2+\zeta_1+\zeta_2$$

s.t. $\forall(y_1,\cdots,y_N)\in W_1$：

$$\frac{1}{N}\sum_{i=1}^{N}\delta F(t_i,p_i,y_i)\geqslant\frac{1}{N}\sum_{i=1}^{N}\Delta(y_i^*,y_i)-\zeta_1$$

$\forall(y_{N+1},\cdots,y_{N+M})\in W_2$：

$$\frac{1}{M}\sum_{j=N+1}^{N+M}\delta F(p_j,t_j,y_j)\geqslant\frac{1}{M}\sum_{j=N+1}^{N+M}\Delta(y_j^*,y_j)-\zeta_2$$

step4　对于 i=1,…,N 实现循环

step5　$\hat{y}_i\leftarrow\underset{y\in Y}{\operatorname{argmax}}\Delta(y_i^*,y)+F(t_i,p_i,y)$

step6　$W_1\leftarrow W_1\cup(\hat{y}_1,\cdots,\hat{y}_N)$

step7　对于 j=N+1,…,N+M 实现循环

step8　$\hat{y}_j\leftarrow\underset{y\in Y}{\operatorname{argmax}}\Delta(y_j^*,y)+F(p_j,t_j,y)$

step9　$W_2\leftarrow W_2\cup(\hat{y}_{N+1},\cdots,\hat{y}_{N+M})$

step10　直到

$$\frac{1}{N}\sum_{i=1}^{N}\Delta(y_i^*,\hat{y}_i)-\frac{1}{N}\sum_{i=1}^{N}\delta F(t_i,p_i,\hat{y}_i)\leqslant\zeta_1+\varepsilon$$

并且

$$\frac{1}{M}\sum_{j=N+1}^{N+M}\Delta(y_j^*,\hat{y}_j)-\frac{1}{M}\sum_{j=N+1}^{N+M}\delta F(p_j,t_j,\hat{y}_j)\leqslant\zeta_2+\varepsilon$$

step11　返回 U，V，ζ_1,ζ_2

在问题中，梯度下降迭代执行，每个迭代来自集合 w_1 中最违反的排名元组 $(\hat{y}_1,\cdots,\hat{y}_N)$ 和集合 w_2 中最违反的排名元组 $(\hat{y}_{N+1},\cdots,\hat{y}_{N+M})$，同时尽量减少松弛变量。利用 Pegasos 算法[12]对 t 和 U 进行更新：

$$U_{t+\frac{1}{2}}\leftarrow(1-\eta_t\lambda)U_t+\frac{\eta_t}{N}\sum_{i=1}^{N}V_t(\delta\psi(t_i,p_i,\hat{y}_i))^{\mathrm{T}}+\frac{\eta_t}{M}\sum_{j=N+1}^{N+M}V_t\delta\psi(p_j,t_j,\hat{y}_j) \tag{16}$$

其中，$\delta\Psi(t_i,p_i,\hat{y}_i)=\Psi(t_i,p_i,y_i^*)-\Psi(t_i,p_i,\hat{y}_i)$，$\delta\Psi(t_j,p_j,\hat{y}_j)=\Psi(t_j,p_j,y_j^*)-\Psi(t_j,p_j,\hat{y}_j)$，$\eta_t$ 是迭代 t 可调的学习率。U_{t+1} 可以通过把 $U_{t+1/2}$ 映射到加速集中得到：

$$B=\{U:\|U\|_F\leqslant\frac{1}{\sqrt{\lambda}}\} \tag{17}$$

V 的更新可以类似地获得，但是要除去最违反的排名元组。

$$V_{t+\frac{1}{2}}\leftarrow(1-\eta_t\lambda)V_t+\frac{\eta_t}{N}\sum_{i=1}^{N}U_{t+1}\delta\psi(t_i,p_i,\hat{y}_i)+\frac{\eta_t}{M}\sum_{j=N+1}^{N+M}U_{t+1}(\delta\psi(p_j,t_j,\hat{y}_j))^{\mathrm{T}} \tag{18}$$

此外，本文与文献[13]的不同之处在于：本文的目标函数由 $\frac{\lambda}{2}\|U\|_F^2+\frac{\lambda}{2}\|V\|_F^2$ 来处罚以控制模型的复杂度。需要指出的是，最优 U 和 V 必须满足条件 $\|U\|_F=\|V\|_F$，由于预测规则只使用了 U^TV 的内积，因此每次梯度下降后，更新的 U 和 V 分别被迫乘以一个常数以确保 $\|U^{\mathrm{T}}V\|_F$ 不变，令 $\alpha=\sqrt{\|U\|_F\|V\|_F}$。

$$U\leftarrow\alpha U/\|U\|_F \tag{19}$$

$$V\leftarrow\alpha V/\|V\|_F \tag{20}$$

实验结果表明，这种策略可达到更快的收敛速度。固定容忍 $\varepsilon=0.01$，算法 1 中的循环通常在 200 次迭代终止。

3　实验结果与分析

本节针对提出的方法进行实验验证与结果分析。实验的

主要目的是评价提出的基于双向学习排序的跨媒体相似性度量方法的有效性，并与其他主要的跨媒体检索方法 CCA，PAMIR，SSI 进行比较分析。

3.1　实验设置

3.1.1　数据集

在对比实验中使用了公共的真实世界数据集，Wikipedia 数据集[15]是图像和相关联的文本一一配对的双模态数据集。Wikipedia 数据集包括 2866 幅图像，每幅图像都有对应的一篇文章描述它。图像用 10 个不同语义类进行分类，用一个 10 维向量来表示数据集的分类。

查询列表生成如下：每次查询时，在训练集中随机选择 40 个检索集合作为候选集，根据候选集的相关或不相关形成一个排名实例。

3.1.2　性能评价

通过构造两个由 PAMIR，CCA 和 SSI 分别优化两个方向的检索任务的模型，以及一个基于双向学习排序的跨媒体相似性度量方法同时优化两个方向的检索任务的统一模型，进行模型的评价。

效用评估是评价信息检索的重要指标。本文利用 MAP 平均精度作为效用的衡量尺度，同时用查准率与查全率进行形象化说明。

3.1.3　参数调整

为了获得最佳性能，所有的方法都需要进行参数调整。对于基于双向学习排序的跨媒体相似性度量方法，调整嵌入的隐空间维数 k 的值以及模型的复杂性和经验风险之间的折中参数 λ。k 和 λ 的取值范围分别是：{5，10，25，50，100，200} 和 {0.01，0.1，1，10}，选择这些数据中性能最好的参数。

3.2　结果与分析

基于双向学习排序的跨媒体相似性度量方法和其他模型的性能比较结果如表 1 所列。

表 1　5 种检索方法的实验结果

方法	文本查询图像		图像查询文本	
	$R=50$	$R=$all	$R=50$	$R=$all
CCA	0.2341	0.1422	0.2218	0.1478
PAMIR	0.3034	0.1747	0.1738	0.1739
SSI	0.2819	0.1638	0.2337	0.1740
单向查询	0.3649	0.2036	0.2549	0.2226
双向查询	0.3981	0.2123	0.2599	0.2528

表 1 中，在 Wikipedia 数据集上通过 MAP@R 的值来比较性能，最好的结果以粗体显示，可以看出本文所提方法在两个方向上的检索性能优于所有对比方法。

进一步将欧氏距离、标准欧氏距离、曼哈顿距离等多种距离度量方法与本文的相似性函数作比较，分别对 5 种方法进行实验得到的 MAP 值绘制柱状图，结果如图 1 所示。

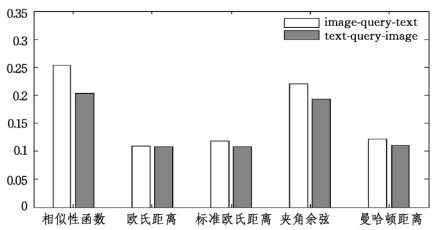

图 1　5 种方法的 MAP 值

此外，5 种方法的 Precision-Recall 曲线结果如图 2 所示。

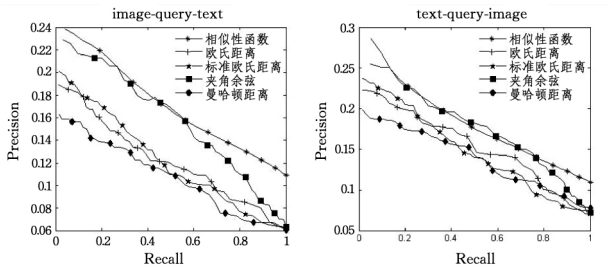

图 2　PR 曲线

通过对 MAP 值和 PR 曲线的比较可以发现，本文采用的相似性函数的效果最好，夹角余弦方法的效果次之，欧氏距离和标准欧氏距离的相似性度量方法的效果最差。

两个向量间的欧氏距离的计算方便，在数学表达上也更容易理解。但是，它将向量的各分量等同看待，往往不能满足实际需求。正是因为欧氏距离的缺点，所以引进了标准欧氏距离的测量方法，但是通过实验结果表明，标准欧氏距离的结果也不能令人满意。原因是空间上两个向量可以有多种形式达到标准欧氏距离相等的情况，但是两个向量之间的夹角也同样决定着两向量的相似程度。

采用曼哈顿距离的结果与欧氏距离的结果类似，其原因也是两个向量之间的差别不能只关注距离，还必须关注两向量之间的夹角。

由此，引出了夹角余弦的方法，事实证明夹角余弦的实验效果确实优于欧氏距离和标准欧氏距离的方法，与本文提出的相似性函数有异曲同工之妙，也是一种比较好的衡量两个向量相似度大小的方法；但是夹角余弦的结果次于本文的相似性函数，这是因为本文的相似性函数在考虑夹角的同时还考虑了向量的大小，这更能说明两个向量的相似程度。

上述实验的结果证明了本文提出的基于双向学习排序的跨媒体相似性度量方法的优越性，可以通过改变其他距离度量方法进一步优化 Wikipedia 数据集上的实验结果。

结束语　本文从跨媒体的信息检索问题上着手，分析了现存的一些经典检索方法的弊端，提出了基于双向学习排序的相似性度量方法以实现跨媒体检索。将 SVM 算法应用于双向学习排序的跨媒体检索的方法中作为约束条件，每次迭代利用 Pegasos 机器学习算法对参数 U 和 V 进行更新。最终学习得到的 U 和 V 降低了文本和图像空间的维度。利用训练集得到的最终参数，对测试集进行仿真实验，结果表明基于双向学习排序的跨媒体相似性度量方法的性能优越。本文利用现实世界中的真实数据集 Wikipedia 对本文的相似性度量方法进行检验，进而通过与经典的方法进行比较，证明了本文所提方法确实可行。与此同时，通过不同的相似性函数对实验数据集进行训练，利用平均正确率均值 MAP 和 PR 曲线等指标，证明了本文算法的正确性。最后，通过研究距离度量方法进一步优化实验结果。

在下一步的工作中，将重点解决以下几个方面的问题：

(1)跨媒体数据关联网络的构建

在跨媒体数据关联网中，综合考虑跨媒体数据的发布时间和相似度，提出一个指标来计算每个跨媒体数据的重要程度。在这个指标中，定义每个跨媒体数据点的入度和出度。一个跨媒体数据的入度越大，该媒体数据是后关注主要语义并且与主要语义偏离的可能性也就越大；反之，一个跨媒体数据的出度越大，该跨媒体为先关注主要语义并且与主要语义

(下转第 118 页)

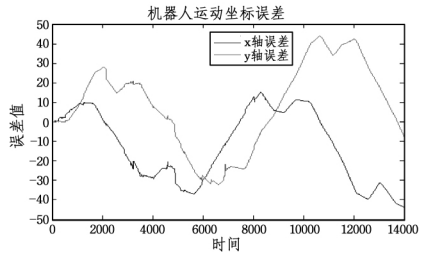

图 4 标准 EKF 模型模拟误差

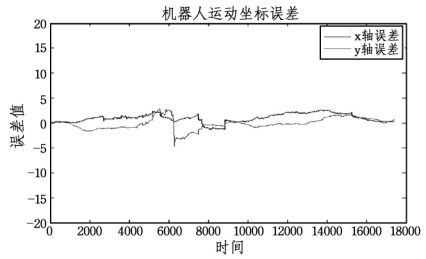

图 5 抗差 EKF 模型模拟误差

从图 2 中可以看到，加入粗差以后，标准的 EKF 算法得到的机器人运行轨迹与实际有较大偏差；从图 4 可以看出误差很大，并且沿同一方向行进时，误差会产生累积，即逐渐偏离真实航向，无法达到定位的要求。从图 3 和图 5 中可以看到，对于观测向量中含有较大粗差的模型，运用改进的抗差 EKF 算法仍然能有效剔除粗差的影响，并计算得到较精确的机器人的运行轨迹，各方向的误差控制在 5m 以内，其中 93% 的点误差在 3m 以内。

由模拟程序实验可以得到以下结论：在观测向量含有较大粗差的系统中，应用抗差 EKF 模型能有效剔除和减弱粗差对系统结果处理的影响。

参 考 文 献

[1] HUGH D W, TIME B. Simultaneous localization and mapping: Part I [J]. Robotics & Automation Magazine, 2006, 13(2): 99-110.

[2] SMITH R, SELF M, CHESSEMAN P. Estimating uncertain spatial relationships in robotics[M]// Autonomous Robot Vehicles. Springer-Verlag, New York, 1990: 167-193.

[3] 浙江大学数学系. 概率论与数理统计[M]. 北京：科学出版社，1965.

[4] 李漳南，吴荣. 随机过程教程[M]. 北京：高等教育出版社，1987.

[5] 罗荣华，洪炳熔. 移动机器人同时定位与地图创建研究进展[J]. 机器人，2004，26(2)：182-186.

[6] 陈卫东，张飞. 移动机器人的同步自定位与地图创建研究进展[J]. 控制理论与应用，2005，22(3)：455-460.

[7] DISSANAYAKE G, NEWMAN P M, et al. A solution to the simultaneous localization and map building problem[J]. IEEE Transactions on Robotics and Automation, 2001, 17(3): 229-241.

[8] 余学祥，陆伟才. 抗差卡尔曼滤波模型及其在 GPS 监测网中的应用[J]. 测绘学报，2001，30(1)：27-31.

[9] 李德仁，袁修孝. 误差处理与可靠性理论[M]. 武汉：武汉大学出版社，2002.

[10] 王坚，王金岭，高井祥. 基于抗差 EKF 的 GNSS 导航模型研究[J]. 中国矿业大学学报，2008，37(4)：473-477.

[11] BHATTI U I, OCHIENG W Y, FENG S J. Integrity of an integrated GPS/INS system in the presence of slowly growing errors. PartI: A critical review[J]. GPS solution, 2007, 11(3): 173-181.

(上接第 87 页)

相近的可能性也就越大。

(2)互联网用户社会网络的构建

如果两个跨媒体数据在跨媒体数据关联网络中有联系，那么这两个跨媒体数据的发布者都会在互联网用户社会网络中相关联。在用户网络中，每个点代表一个用户，每条边代表两个用户的联系，边的权值表示两个用户的关联度。

(3)用户网络潜在应用

通过用户网络和跨媒体数据网络选出的重要互联网用户和重要跨媒体数据可以在相近语义的挖掘中发挥巨大的作用。例如，对于一个包含同一具体事件的大数据集，可以通过两种网络的分析快速确定重要的跨媒体数据，这些数据可能能够完整地概述整个数据集或表达民众的态度。

参 考 文 献

[1] HOTELLING H. Relations between two sets of variates[J]. Biometrika, 1936, 28(3/4): 321-377.

[2] RASIWASIA N, PEREIRA J C, COVIELLO E, et al. A new approach to cross-modal multimedia retrieval[C]// Proceedings of the International Conference on Multimedia. 2010: 251-260.

[3] JIA Y, SALZMANN M, DARRELL T. Factorized latent spaces with structured sparsity[C]// Advances in Neural Information Processing Systems. 2010, 23: 982-990.

[4] JIA Y, SALZMANN M, DARRELL T. Learning cross-modality similarity for multinomial data[C]// IEEE International Conference on Computer Vision. 2011: 2407-2414.

[5] GRANGIER D, BENGIO S. A discriminative kernel-based approach to rank images from text queries[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2008, 30(8): 1371-1384.

[6] WU F, LU X, ZHANG Z, et al. Cross-media semantic representation via bi-directional learning to rank[C]// MM'13. 2013: 877-886.

[7] JIANG X, WU F, LI X, et al. ACM International Conference on Multimedia [J]. Computer & Graphics, 1994, 18(4): 611-612.

[8] DEERWESTER S, DUMAIS S, FURNAS G, et al. Indexing by latent semantic analysis[J]. Journal of the American Society for Information Science, 1990, 41(6): 391-407.

[9] BAI B, WESTON J, GRANGIER D, et al. Learning to rank with (a lot of) word features[J]. Information Retrieval, 2010, 13(3): 291-314.

[10] TSOCHANTARIDIS I, JOACHIMS T, HOFMANN T, et al. Large margin methods for structured and interdependent output variables[J]. Journal of Machine Learning Research, 2006, 6(2): 1453-1484.

[11] JOACHIMS T. A support vector method for multivariate performance measures[C]// Proceedings of the 22nd International Conference on Machine Learning. 2005: 377-384.

[12] SHALEV-SHWARTZ S, SINGER Y, SREBRO N. Pegasos: Primal estimated sub-gradient solver for svm[C]// Proceedings of the 24th International Conference on Machine Learning. 2007: 807-814.

[13] JOACHIMS T, FINLEY T, YU C. Cutting-plane training of structural svms[J]. Machine Learning, 2009, 77(1): 27-59.

[14] BAEZA-YATES R, RIBEIRO-NETO B. Modern Information Retrieval[M]. Addison-Wesley, 1999.

[15] Ada[OL]. http://www.svcl.ucsd.edu/projects.