

基于内容的图像检索中相关反馈算法综述^{*}

Review of Relevance Feedback Algorithms in Content-based Image Retrieval

杨娜 罗航哉 薛向阳 吴立德

(复旦大学计算机科学与工程系 上海200433)

Abstract Relevance feedback is a power tool to increase the retrieval performance in content-based image retrieval system. In this paper, we review three relevance feedback algorithms, i. e. spatial transform model, clustering model and Bayes model, and discuss their advantages and disadvantages.

Keywords Relevance feedback, Content-based image retrieval

1 引言

从70年代开始,人们就进行图像检索技术的研究,那时候主要工作是对图像进行人工标注,然后应用文本检索技术实现对标注图像的检索。然而图像内容是复杂多样的,不同人对图像的理解不完全一致,一般难用准确的词汇描述图像内容;另外,对大量图像做标注也是一项费时费力的工作,这些因素导致基于文本标注的图像检索方法在实际应用中很难达到预期效果。进入90年代,全球网络通讯技术和光盘等大容量信息存储技术迅速发展,图像、视频和语音等多媒体信息的数据量急剧增加,如何从这些海量的多媒体信息中寻找感兴趣的内容已成为目前急需解决的、有挑战性的难题。近几年来,研究人员企图用基于内容的信息检索技术解决这个难题。基本路线是自动或半自动提取多媒体对象的底层特征、对图像而言可以提取颜色直方图、主色调、形状、纹理等特征,然后基于这些特征定义图像之间的相似度,从而实现基于内容或特征的检索。

现实世界的图像是丰富多彩的,人们在观察图像时往往从语义出发,而颜色、形状和纹理等底层特征和图像的语义没有直接的联系。一些特征可以反映部分语义,而一些特征与语义无关。因此,图像特征以及基于特征计算出来的相似度并不能很好地体现人对图像的感知,如何选取和组合这些特征以更好地模拟人对图像的主观感受成为图像检索中的重要研究课题。相关反馈技术是其中一种解决方案,基本思想是让用户引导计算机完成图像检索。在查询过程中,系统返回给用户一组结果图像,用户对这些结果图像进行评价,系

统从用户反馈信息中自动分析最能表征用户查询目标的特征,自动调整相似度的度量方法,然后进行新的查询,如此多次反馈,最终得到满意的查询结果。

相关反馈技术最早应用于文本检索中,它有效提高检索性能。最近,相关反馈的基本思想被借用到图像检索中,由于图像检索在许多方面不同于文本检索,因此,文本检索中使用的相关反馈技术往往不能直接应用于图像检索,需要进行改进或创新。Yong Rui等利用优化查询矢量和特征权矩阵的方法改进查询矢量和相似度计算公式^[1,2],Catherin S. Lee等采用聚类的方法把图像特征空间进行划分和聚类,通过对训练图像的学习找到最接近用户目标的特征空间类^[3],Nuno Vasconcelos等则利用贝叶斯法则和最大似然度方法找到最佳目标图像类^[4,5]。本文首先介绍基于内容的图像检索的原理,然后具体介绍图像检索中三种相关反馈技术的模型和实现。

2 基于内容的图像检索原理

基于内容的图像检索系统将图像看作高维空间的矢量 $\vec{p} \in R^n$, $\vec{p}$ 是从图像中提取的颜色、纹理、形状等底层特征或它们的组合, R^n 通常被称为特征空间。在特征空间上可以定义矢量间的距离函数以衡量图像之间的差异。距离函数可以取不同的形式,只要满足以下条件:1) $D(x, y) = 0 \Leftrightarrow x = y$; 2) $D(x, y) = D(y, x)$; 3) $D(x, y) \leq D(x, z) + D(y, z)$; 4) $D(x, y) \geq 0$ 。图像检索中常用的有 L_1 (Manhattan)、 L_2 (Euclidean)、 L_∞ (MAX距离)、City-block距离等。

图像间的相似度则可以用其距离的单调递减函数

^{*} 本文得到自然科学基金课题(69935010, 60003017)和国家863计划等资助。杨娜 硕士研究生,现从事基于内容的图像检索研究。罗航哉 助教、在职研究生,现从事基于内容的图像和视频检索研究。薛向阳 副教授,现从事基于内容的图像和视频检索研究。吴立德 博导,现从事文本处理和基于内容的图像和视频检索研究。

定义,相似度 $S(p, q)$ 通常采用以下函数: $S(p, q) = \frac{a}{D(p, q) + 1}$ 或 $S(p, q) = e^{-aD(p, q)}$ (a 为常数)。

由于特定特征空间中的距离并不能反映人对不同图像的感受的差异,采用以上方法衡量图像间的相似程度在图像检索中往往不能得到满意的结果。为改善查询结果,可以通过改变特征空间、改变距离的计算方法以及相似度的度量公式等使相似度更接近于人的感受,相关反馈技术便是通过与用户交互达到上述目标的。

3 三种主要的相关反馈模型

3.1 基于空间变换模型

空间变换的核心思想是通过各种空间变换改变特征空间中图像点的分布,把在一个特征空间中不聚类的点集,映射到另一特征空间中成为聚类的点集。由于用底层特征表示的图像与人的视觉感受之间存在差异,在一个特征空间中分散的图像集在人看来可能是相似的,可以通过空间变换减小它们之间的距离,使图像特征的描述更接近人的视觉感受。

设图像 q 在特征空间 A 中的坐标为 $q^A = (q_1^A, q_2^A, \dots, q_n^A)$, 经过变换 $F: A \rightarrow B$, 在空间 B 中的坐标为 $q^B = (q_1^B, q_2^B, \dots, q_n^B)$, 即 $q^B = F(q^A)$ (m, n 为特征的维数)。

设训练图像集中与用户目标相关的图像集为 $\{r_1, r_2, \dots, r_s\}$, 空间变换的目标是解最优化问题:

$$\min_F \sum_{i=1}^s D(F(r_i), F(r_j)) \quad (1)$$

以得到最优变换 F 。

若 F 采用线性变换,即 $q^B = Mq^A$ (M 为变换矩阵), 图像间距离采用欧氏距离, $\tilde{x}$ 和 $\tilde{y}$ 为两幅图像的特征矢量, 则 $\tilde{x}$ 和 $\tilde{y}$ 在空间 A 中距离为: $D(\tilde{x}^A, \tilde{y}^A) = (\tilde{x}^A - \tilde{y}^A)^T (\tilde{x}^A - \tilde{y}^A)$ 。在变换后的空间 B 中距离为 $D(\tilde{x}^B, \tilde{y}^B) = (\tilde{M}\tilde{x}^A - \tilde{M}\tilde{y}^A)^T (\tilde{M}\tilde{x}^A - \tilde{M}\tilde{y}^A) = (\tilde{x}^A - \tilde{y}^A)^T \tilde{M}^T \tilde{M} (\tilde{x}^A - \tilde{y}^A)$ 。

令 $M' = M^T M$, 则(1)式的最优化问题成为求满足:

$$\min_{M'} \sum (\tilde{x}^A - \tilde{y}^A)^T M' (\tilde{x}^A - \tilde{y}^A) \quad (2)$$

的变换矩阵 M' 的最优解。

Yong Rui 等^[1,2]采用的查询矢量与特征权重最优化的方法是线性空间变换模型的例子。在该方法中,图像用特征向量表示,由纹理、颜色、形状等组合而成。距离计算分两步完成,如下公式所示:

$$\begin{cases} d_n = \Phi(g_n, U) & i = 1, 2, \dots, I; \\ g_n = \Psi(\tilde{x}_n, \tilde{q}, W_i) & n = 1, 2, \dots, N. \end{cases}$$

这里 N 为训练图像的数目,每幅图像由 I 个特征分量

表示, $\tilde{q}$ 和 $\tilde{x}_n$ 分别为查询图像 q 与训练图像 x_n 的一种特征(颜色、纹理、形状等),函数 Ψ 表示它们的距离,记为 g_n , d_n 表示 q 与 x_n 之间的总距离。 W_i 和 U 是距离公式中的参数。

实现过程如下:

1) 选取函数 Φ 和 Ψ , 均用线性函数实现。

$$\begin{cases} g_n = (\tilde{x}_n - \tilde{q})^T W_i (\tilde{x}_n - \tilde{q}) \\ d_n = \tilde{u}^T g_n \end{cases}$$

$W_i, \tilde{u}^T$ 为各分量的权重,这里的 W_i 相当于(2)式中的 M' 。

2) 用户给出训练图像的一组评价值 $(\pi_1, \pi_2, \dots, \pi_n)$, n 为训练图像的数目。

3) 确定优化的目标: $\min J = \tilde{\pi}^T \times \tilde{d}$, 约束条件:

$$\sum_{i=1}^I \frac{1}{u_i} = 1, \det(W_i) = 1.$$

利用 Lagrange 数乘方法得到最优查询矢量和相似度公式中各特征的权重矩阵,使相关图像在特征空间中聚类,并使查询矢量移至相关图像的集合中心,且远离不相关图像。

4) 利用新查询矢量和相似度公式进行新一轮查询。

Yong Rui 等的方法不需要用户提供非常精确的查询图像,而通过与用户的交互,对经用户评价的训练图像进行学习,使新的查询矢量逐步趋近用户的目标图像,并把图像特征矢量映射到与人视觉感受相近的空间中,在此空间中计算新查询矢量与图像库中图像的相似度。这是一种理论上较完美的方法,但空间变换非常复杂,在图像特征维数较高,训练图像的数目小于特征维数时,优化过程中空间变换矩阵的解不唯一。在这种情况下,很难找到最优解,给实现带来一定的困难。

3.2 基于聚类分析模型

该方法不需对图像特征空间进行变换,而是在原有的特征空间中对图像进行聚类。根据用户对训练样本图像的评价,把特征空间进行划分,形成不同的图像类,使相关图像落入同一类或距离较近的类中,而不相关图像落入距离较远的不同类中。在检索中根据类与类之间的距离确定图像间相似性的大小。

Catherin S. Lee 等^[3]采用此模型研制的图像检索系统,对图像特征空间进行聚类,并计算每两类之间的权重作为类间相似度的度量。通过调整类的大小和类间权重,把相似的类拉近,并使不相似的类远离。

在实现中,首先利用 K 均值算法对图像库的特征空间进行初始聚类,根据类间底层特征的距离得到初始权重,形成一个权矩阵 $M_{N \times r}$ (r 为图像类的数目)来表示不同类之间的相似度,然后根据用户对查询结果

图像的评价进行类的调整。步骤如下:

1. 利用下面公式把特征向量 x 变换为 $F(x)$ 。

$$[F(x)]_i = \exp\left\{\frac{-1}{2\sigma^2}\|x - c_i\|^2\right\},$$

其中 c_i 为第 i 个图像类的均值, σ^2 为第 i 个图像类的方差, 该变换的结果表示了特征向量 x 和各个图像类的相似度。

2. 根据用户从结果图像中选出的相关图像 $\{p_i | 1 \leq i \leq n\}$ 和不相关图像 $\{n_i | 1 \leq i \leq m\}$, 把同时包含 $p_i (1 \leq i \leq n)$ 与 $n_i (1 \leq i \leq m)$ 的类分裂成两类, 并利用以下迭代算法对权矩阵进行调整。

$$M_{i+1} = M_i + \sum_{j=1}^n F(q)F(p_i)^T - \sum_{j=1}^m F(q)F(n_i)^T$$

q 为初始查询向量, M_i 为当前权矩阵, M_{i+1} 为新的权矩阵。

3. 根据调整后的权矩阵, 把权重很大的类合并成一个新的类, 并重新计算权矩阵。

最后确定每幅图像与查询图像的相似度, 并按顺序排列。

1. 计算查询向量和各图像类的距离, 找到距离最小的类作为目标类, 并根据其它类与目标类间的权重对各个类排序。

2. 同一类中的向量与查询向量的相似度由底层特征间的距离来衡量。

聚类方法在不进行空间变换的情况下, 有效地调整不同图像类间的距离, 其效果类似于特征空间的线性变换, 能较好地改善查询结果。在检索时, 只需确定目标图像类及计算查询图像与该图像类中图像的相似度, 而不需要与图像库中的每幅图像进行比较, 减少了检索的计算量, 而且由于聚类过程不改变特征空间, 有利于建立图像库的索引结构。另外, 特征空间的重新聚类结果可以供下一次查询时使用, 这样, 系统具有简单的记忆功能, 有效地减少了同类查询的计算时间。

但是, 在聚类过程中没有进行特征的选取, 各种视觉特征在检索中的重要性被同等看待, 这不符合人对图像的观察。当特征维数较多时, 不仅计算复杂, 而且由于训练样本数目较少使聚类的准确性及有效性下降。

3.3 基于统计概率分类模型

空间变换和聚类的方法均是在图像类为几何可分(线性或非线性)的前提下展开的。实际应用中的图像库内容丰富, 用户对图像的分类要求千差万别, 几何分类有时并不能满足要求, 当图像类不满足几何可分时, 可用概率统计的方法进行分类。

基于 Bayes 法则的分类: 设有 K 类样本, 分别为 $\omega_1, \omega_2, \dots, \omega_k$, 若每类的先验概率为 $P(\omega_i), i=1, 2, \dots, k$, 对一随机矢量 X , 每类的条件概率为 $P(x|\omega_i)$, 根据

Bayes 公式, 后验概率:

$$P(\omega_i|x) = \frac{P(x|\omega_i)P(\omega_i)}{\sum_{i=1}^k P(x|\omega_i)P(\omega_i)} \quad (3)$$

从后验概率出发, 有 Bayes 法则: $P(\omega_i|x) > P(\omega_j|x) \Rightarrow x \in \omega_i$, 其中 $i, j=1, 2, \dots, k$, 且 $\forall j \neq i$ 。

因为(3)式的分母仅是使后验概率小于1的系数, 消去分母, 可得等价的判决法则:

$$P(x|\omega_i)P(\omega_i) > P(x|\omega_j)P(\omega_j) \Rightarrow x \in \omega_i, \quad i, j=1, 2, \dots, k; \forall j \neq i \quad (4)$$

在图像检索中, x 作为查询矢量, 检索的目标是找到使 $P(\omega_i|x)$ 最大的图像类 ω_i 。根据 Bayes 法则和判决法则(4), 问题可转化为满足 $\max_i g(x) = P(x|\omega_i)P(\omega_i)$ 的最优解。各图像类 ω_i 的先验概率密度 $P(\omega_i)$ 可由预先统计整个数据库得到。

相关反馈过程中, $x_1, x_2, \dots, x_i$ 为一组经用户评价过的训练图像, 反馈的目标是找到满足:

$$\max_i P(\omega_i|x_1, x_2, \dots, x_i)P(\omega_i) \quad (5)$$

的第 i 类图像。

Nuno Vasconcelos 等人^[4,5]利用 Bayes 分类器进行检索和相关反馈。检索的最终目标是在图像库中查找用户所需的某一幅图像, 在初次查询时, 用户提交一幅近似于目标的图像, 并对返回的检索结果进行评价, 找到与目标最接近的图像作为新查询图像, 重新计算每一类图像的后验概率找到满足(5)式的第 i 类图像, 经多次反馈直到满意为止。

在系统实现过程中, 假定满足以下三个条件: a) 每类图像作为检索目标的概率是一样的, 即先验概率 $P(\omega_i) = 1/N$, N 为图像类的个数。这样, 问题简化为求满足 $\max_i P(\omega_i|x_1, x_2, \dots, x_i)$ 的第 i 类图像。b) 用户当前的评价结果独立于以前的评价。c) 各图像类的条件概率分布满足混合高斯分布。

由假定条件(a)和(b), 并根据 Bayes 公式, $g(x)$ 可写为以下形式:

$$g(x) = P(\omega_i|x_1, x_2, \dots, x_i) = \gamma_i P(x_i|\omega_i)P(\omega_i|x_1, \dots, x_{i-1}), (\gamma_i \text{ 为一常数})$$

用对数形式表示并进一步展开得到:

$$g(x) = \log P(\omega_i|x_1, x_2, \dots, x_i) = \sum_{k=0}^{i-1} \log \gamma_{i-k} + \sum_{k=0}^{i-1} \log P(x_{i-k}|\omega_i) + \log P(\omega_i) \quad (6)$$

由于(6)式中首项 $\sum_{k=0}^{i-1} \log \gamma_{i-k}$ 和末项 $\log P(\omega_i)$ 均为常数, 因此只需求出满足 $\max_i \sum_{k=0}^{i-1} \log P(x_{i-k}|\omega_i)$ 的第 i 类图像。

(下转第105页)

结束语 本文基于合作式共同进化计算模型,提出一种求解0-1背包问题的共同进化遗传算法。通过仿真实验比较了对非法解三种处理方式的优劣,以及不同种群数下共同进化遗传算法对不同规模问题的求解性能。研究表明:合作式共同进化遗传算法的收敛速度和问题求解质量均优于传统单种群遗传算法。本文中种群的划分采用的是手工方式,进一步工作是研究问题的自动分解方法和针对具体问题的有效合作方式,以及对0-1背包问题的扩充问题的求解,以期应用于实际工程中。

致谢 感谢博士后流动站的谢涛副教授在课题研究过程中给予的中肯建议和无私的帮助,也感谢博士生葛建坤同学的有益讨论。

参考文献

- 1 Martello S, Toth P. Knapsack Problems: Algorithms and Computer Implementations. John Wiley & Sons Ltd., Chichester, England, 1990
- 2 Martello S, Toth P. An upper bound for the zero-one knapsack problem and a branch and bound algorithm. European Journal of Operational Research, 1977(1): 169~175
- 3 Andonov, Rajopadhye. A Sparse Knapsack Algorithm and its Synthesis. In: Int Conf. on Application-Specific Array Processors (ASAP-94). San Francisco, CA. IEEE Press, 1994. 302~313
- 4 Holland J H. Adaptation in Natural and Artificial Systems. Ann Arbor: The University of Michigan Press, 1975
- 5 Goldberg D E. Genetic Algorithms in Search, Optimization and Machine Learning. Addison-Wesley Publishing Company, Inc. 1989
- 6 Dolan A. A general GA toolkit implemented in Java. Available at: Http://www.aridolan.com/ga/gaa/gaa.html
- 7 Potter M A. The Design and Analysis of a Computational Model of Cooperative Coevolution. [PhD. Dissertation]. George Mason University, 1997
- 8 Paredis J. Coevolutionary Algorithms. The Handbook of Evolutionary Computation. In: B. T. Fogel, D. Michalewicz, eds. Oxford University Press, 1998
- 9 钟求喜. 网络计算中任务分配中调度的遗传算法研究. [博士学位论文]. 国防科技大学研究生院
- 10 谢涛, 陈火旺. 基于函数分解的可伸缩进化算法. 自然科学进展, 2001

(上接第111页)

类图像。

根据假定(c): $p(x|\omega_i)$ 采用混合高斯密度函数, 其中的参数由期望最大(EM)算法得到。

统计分类是一种理论上理想的分类方法, 由于利用了最大似然决策作为图像距离的度量方法, 用 Bayes 判决方法构造的分类器在性能上优于采用欧氏距离和聚类等几何划分的方法。然而利用最大似然估计对于一些小概率事件不可避免地造成判断错误, 影响分类结果的准确性。在实际应用中, 条件概率的精确估计是难以解决的问题, 图像在特征空间中分布非常复杂, 对于不同用户的要求, 图像的分类要求也不一致, 造成条件概率估计不准确。对于特定的图像库, 经过多次训练可能找到较好的条件概率分布, 但不同类型的图像库间不存在通用的条件分布, 因此很难实现通用的检索。

结论 相关反馈技术是在图像检索领域中经常采用的重要方法之一。本文具体介绍了近几年来一些主要的相关反馈技术模型和算法, 可以作为借鉴, 在它们的基础上作更进一步研究。

人的图像理解过程往往是基于图像语义的, 因而图像的各种特征和语义之间的联系以及人们如何评价图像之间的相似度是图像检索中的难点。每幅图像都有丰富的内容, 任何特征均不能准确地描述人对图像的认识。相关反馈技术的目标就是从系统与人的交互

过程中学习, 让系统获得人对图像理解的信息, 使系统模拟人类观察和理解图像的方式, 提高图像检索的性能。我们认为图像检索领域发展的趋势之一是利用相关反馈过程提取用户查询目标图像的语义, 找到图像特征与语义之间的对应关系, 并对特征进行自动选取, 滤除与检索目标语义无关的特征。并且根据用户对样本图像相似性的评价找到最佳特征空间和距离度量方法, 在此空间中, 特征点间的距离能很好地反映人对不同特征点对应的图像间差异的感受。

参考文献

- 1 Rui Y, Huang T S. A Novel Relevance Feedback Technique in Image Retrieval. In: ACM MM, Orlando, Florida, October 30—November 5, 1999
- 2 Rui Y, et al. Relevance feedback: a power tool for interactive content based image retrieval. IEEE Trans. Circuits and systems for video technology, 1988, 8(5): 644~655
- 3 Lee C, Ma W Y, Zhang H J. Information Embedding Based on User's Relevance Feedback for Image Retrieval. [Technical Report]. HP Labs, 1998
- 4 Vasconcelos N, Lippman A. Learning From User Feedback in Image Retrieval System. NIPS'99, Denver, Colorado, 1999
- 5 Vasconcelos N, Lippman A. Bayesian Representations and Learning Mechanisms for Content-Based Image Retrieval. SPIE Storage and Retrieval for Media Databases 2000, San Jose, California, 2000. A short version appears in Wkshp on CATVL, CVPR'00, South Carolina, 2000