

使用基于 SVM 的否定概率和法的图像标注

路 晶,金奕江,马少平,茹立云

(清华大学 计算机科学与技术系,智能技术与系统国家重点实验室,北京 100084)

摘 要:在基于内容的图像检索中,建立图像底层视觉特征与高层语义的联系是个难题.对此提出了一种为图像提供语义标签的标注方法.先建立小规模图像库为训练集,库中每个图像标有单一的语义标签,再利用其底层特征,以 SVM 为子分类器,“否定概率和”法为合成方法构建基于成对耦合方式(PWC)的多类分类器,并对未标注的图像进行分类,结果以 N 维标注向量表示,实验表明,与一对多方式(OPC)的多类分类器及使用概率和法的 PWC 相比,“否定概率和”法性能更好.

关键词:语义标签;否定概率和法;成对耦合;标注向量

中图分类号: TP3 **文献标识码:** A **文章编号:** 1673-4785(2006)01-0062-05

Image annotation using the summation of negative probability based on SVM

LU Jing, JIN Yi-jiang, MA Shao-ping, RU Li-yun

(The State Key Lab of Intelligence Technology System, Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China)

Abstract: In the approach of content-based image retrieval, a critical point is to provide maximum support in bridging the semantic gap between low-level visual features and high-level concepts. An annotation procedure for providing images with semantic labels was proposed. The annotation procedure started with labeling a small set of training images, each with one semantic label. An ensemble of support vector machines (SVMs) based on the summation of negative probability, which was constructed by pairwise coupling (PWC), was then got with the content-based image features. It was applied to give the unlabeled image an N dimension label-vector, thus providing users with a conceptualized annotation. The ensemble is better than the one per class (OPC) scheme and the PWC based on the summation of probability.

Keywords: semantic label; the summation of negative probability; pairwise coupling; label-vector

伴随着多媒体数据的大规模运用,对于基于内容的图像检索(CBIR)的研究成为当前的一个热点,已经进入实用阶段. Smeulders 等人发表了一篇总结性的文章^[1],回顾了近年来 200 多篇相关的论文.由于图像中物体的多样性和复杂性,多数的 CBIR 系统是以例图或草图为查询方式的图像检索,并通过对底层特征的比较实现检索.使用的底层特征,是从图像中利用算法自动抽取的颜色、纹理和形状等.但低级特征对图像的描述与人类的描述存在较大差

异.日常生活中,人类使用的是高层语义概念,如“太阳”、“高楼”、“海浪”等.在特定的应用领域如人脸及指纹识别中,可以建立从底层特征到高层语义(人脸或指纹)的联系,但在一般情况下,它们并没有直接的联系,而是一种复杂的多对多的映射关系,一个语义往往对应着多种特征,反之亦然.

为了建立起这种联系并能更准确地描述图像语义,学者们进行了多方面的研究.斯坦福大学的 SIMPL Icity 系统^[2]通过统计的方法将图像大致分为纹理—非纹理,图表—照片等大的语义类以帮助缩小图像检索的范围.纽约大学的 Show & Tell 系统^[3]利用区域分割对整幅图像进行自动的标注,该系统使用物体之间的空间位置关系和目标识别技术

收稿日期:2006-03-02.

基金项目:国家重点基础研究基金资助项目(2004CB318108);国家自然科学基金资助项目(60223004,60321002,60503064,60303005);教育部科学技术研究重点项目(104236).

来建立高层语义和底层特征的联系,主要用于分析地图。但是,目前图像分割仍是计算机视觉中的难点问题,图像中的对象难以从背景中准确分割出来,因此 H. L. Tang 和 R. Hanka 以固定的方式分割图像^[4]。他们将图像分成 $m \times n$ 的子图,再使用 3 层的神经网络对每个子图分类,并根据子图分类结果和知识库产生整个图像的语义标注。该方法适用于子图类别数目不多的情况,如医学图像等,对于一般图像却很难奏效。

针对这种情况,有些研究者使用图像整体的底层特征来进行标注或分类。Vailaya 等人对图像集采取了多级分类以降低分类的难度^[5],在每一级分类时,采取了贝叶斯分类的方法。它假设在每一级分类时,图像集包含的图像类别是固定的;又假设每类图像的先验概率是已知的,但实际上这 2 个理论假设在实际应用中是难以满足的。为了避免对显式先验知识的需求,并获得较好的推广能力,G. Sychay 和 E. Chang 等人使用了以 SVM 为子分类器,一对多(one per class)方式构建的多类分类方法^[6]对图像特征进行分类,并将输出结果转化为概率形式^[7],性能较好。

为了进一步提高图像标注的准确性,文中采用了如下的方法:首先建立初始的训练集,为其中每个图像类人工标上单一的语义标签,例如 forest, land, sky 等;然后根据训练集数据,以 SVM 为子分类器,采用文中提出的否定概率和法,以成对耦合的组合方式(PWC)构建多类分类器,并通过该多类分类器对未标注的图像进行分类处理。最后,系统为每个图像提供一个对应其本身内容的 N 维标注向量作为标注结果,如 (forest: 0. 1, land: 0. 9, sky: 0. 7, ...)。该标注向量由多个语义标签和相关的确信度组成,上例表示图像语义上标注为 forest, land 和 sky 的可能性分别为 10 %, 90 % 和 70 %。从而使得用户可以进行概念化的图像查询。

1 系统组成

1. 1 图像特征

本文选取 6 种性能较好的底层特征来描述图像^[8],分别为颜色直方图、颜色矩、颜色一致性向量、Gabor 纹理、MRSAR 纹理和 TWT 小波纹理,详细描述如表 1 所示。对于图像 P_k ,其特征组为 $x_k = [x_{k1}, \dots, x_{k6}]$ 。向量 x_k 的每一维分别代表表 1 中某种特征,例如颜色直方图,它们仍然是向量。可见图像特征向量(组) x_k 的维数是很高的,不妨设为 d 。特征归一化后,再对 x_k 进行 PCA 主成分分析,得到新

的 m 维特征 Y ,经实验,取 $m = d \times 0. 3$ 就能得到很好的检索性能^[8], Y 即为图像最终的特征向量。

表 1 6 种图像底层特征
Table 1 6 kinds of visual features

底层图像特征	维数	底层图像特征	维数
HSV 空间上的颜色直方图 (HSV- Histogram)	64	Gabor 变换纹理特征 (Gabor)	48
RGB 空间上的三阶颜色矩 (RGB- Moments)	9	树型小波变换纹理特征 (TWT)	104
RGB 空间上的颜色一致性向量 (RGB- CCV)	128	多尺度同步自回归纹理特征 (MRSAR)	16

1. 2 学习算法

文中使用的学习算法是以 SVM 为子分类器,以成对耦合的组合方式(PWC)构建的多类分类器,采用的合成方法为否定概率和法,最后的结果以标注向量表示。

1. 2. 1 支持向量机(SVM)

选择 SVM 作为子分类器,是因为 SVM 不仅有统计学习理论作为其推广性的保证,而且在很多实际问题中获得了优于其他方法的结果。对于样本集 $(X_i, Y_i), i = 1, 2, \dots, n$, 向量 $X_i \in R^m, Y_i \in \{-1, +1\}$, 其 SVM 分类函数为

$$f(X) = \operatorname{sgn}(g(X)) = \operatorname{sgn}\left(\sum_{i=1}^n \alpha_i^* Y_i K(X_i, X) + b^*\right),$$
$$0 \leq \alpha_i^* \leq C, i = 1, 2, \dots, n. \tag{1}$$

式中: α_i^* 为 SVM 训练得到的参数,为了取得最佳性能,选择核函数 $K(U, V) = \exp[-\sum_{i=1}^m |u_i - v_i|]$ 为拉普拉斯径向基函数(LRBF)^[7],并取 $\gamma = 0. 001, C = 100$ 。

注意到 SVM 的决策输出 $f(X)$ 并不是图像标注需要的概率,对式(1)中 $g(X)$ 进行概率化,得到样本 X 被判别为正类的概率 $P(Y = 1 | g)$,采用 sigmoid 参数模型^[9]:

$$f(X) = P(Y = 1 | g) = 1 / [1 + \exp(Ag + B)].$$
$$\tag{2}$$

固定 A, B 的值分别为 $A = - 3, B = - 0. 2^{[7]}$,保证 $g = 1$ 时, $P(Y = 1 | g) = 0. 95, g = - 1$ 时 $P(Y = 1 | g) = 0. 05$,以及 $g = 0$ 时 $P(Y = 1 | g) = 0. 5$ 。

1. 2. 2 多分类器合成

文中采用 2 类判别的手段来解决图像多类分类

的问题,需要经过问题分解和结果综合 2 个步骤。

问题分解是指采用一定的手段将单一的复杂的多类问题划分成多个相关的 2 类问题的过程。常用的分解方法有 OPC 方式与 PWC 方式^[9]。OPC 方法将一个 K 类问题分解成 K 个子问题,每个子问题将原始问题中的一个类作为正类,而此类之外所有其他类合并在一起作为负类,共需 $L = K$ 个二分类的子分类器。PWC 方式方法即成对耦合方法,针对多类分类问题中的每一对类别分别建立一个二分类器 (i, j) ,该二分类器以该对类别中的一类为正类,另一类为负类,在假设样本 $x \in \{x_i, x_j\}$ (x 为 m 维向量)的前提下,判别 $x \in x_i$ 或 $x \in x_j$,将原始问题分解为 $L = \binom{K}{2} = K(K-1)/2$ 个子问题。不同的问题分解方式对应着不同的结果综合方法。

对于 OPC 方式,每个待识别样本有 K 个决策输出, $0 \leq f_k \leq 1$,最终的判别结果为

$$k(x) = \arg \max_{1 \leq k \leq K} f_k(x). \quad (3)$$

对于 PWC 方式,以不同的综合方法对子分类器的判别结果进行处理,会得到不同的最终判别结果。一般会使用 PWC 方式综合方法中的“概率和”法(PWC-SP),其公式表示如下:

$$k(x) = \arg \max_{1 \leq k \leq K} f_k^{(kl)}(x). \quad (4)$$

式中: $f_k^{(kl)} = P(x \in k | x \in k, l)$, 显然有 $f_l^{(kl)} = 1 - f_k^{(kl)}$. 不妨设 $P(x \in k, l)$ 为常数,则式(4)中 $\sum_{l=1}^K f_k^{(kl)}(x)$ 与样本 x 属于类别 k 的概率 $P(x \in k)$ 成正比。

如前所述,OPC 方法需要的子分类器数目相对较少(需要 K 个),但每个子分类器均需要所有样本参与训练,复杂性很高;同时样本数目不平衡,正例样本数远远小于负例样本数(正例与负例的比例为 $\frac{1}{K-1}$),会影响分类器性能。对于 PWC-SP 方法,虽然需要的子分类器数目大大增加(从 K 到 $K(K-1)/2$),但每个子分类器的训练样本相应变少,参数估计相对简单,而且正例与负例数目相当。不足之处是在对样本 $x(x \sim)$ 进行判别的子二分类器集合中,绝大多数二分类器都与样本真正的类别无关,会带来无用甚至有害的干扰信息。也就是说,由于事先不知道样本 x 所属的类别,即使一个二分类器 (k, l) 将样本 x 判别属于 k ,也不能完全相信这个判断。因为该分类器根本不“认识”其他的类别,它完全可能将属于其他类别的样本随意地判断为 k 。但是,从该判断可以确切知道 $x \in l$,因为该二分类器

是“认识”类别 l 的,若 $x \in l$,则它会给出肯定的判断。基于上述观点,应使用判别方法:

$$k(x) = \arg \min_{1 \leq k \leq K} \sum_{l=1, l \neq k}^K f_l^{(kl)}(x). \quad (5)$$

式中: $f_l^{(kl)} = P(x \in l | x \in k, l)$. 然而,由于 2 类判别的对称性,式(5)会给出与式(4)完全相同的判别结果。因此在概率求和时,文中只累加被子分类器否定的那些类的概率值,最终得到了更好的 PWC 方式的综合方法——“否定概率和”法(PWC-SNP),公式如下

$$k(x) = \arg \min_{1 \leq k \leq K} \sum_{l=1, l \neq k}^K \vartheta(f_k^{(kl)}, f_l^{(kl)}). \quad (6)$$

式中: $\vartheta(f_k, f_l) = \begin{cases} f_l, & \text{if } f_k < f_l \\ 0, & \text{if } f_k > f_l \end{cases}$. 即按照子分类器的否定概率 $f_l^{(kl)}$ 选择性的投票,并赋予待识别样本以累加值最小的类别。忽略 $P(x \in k, l)$ 的影响,式(6)中 $\sum_{l=1, l \neq k}^K \vartheta(f_k^{(kl)}, f_l^{(kl)})$ 与概率 $P(x \in k)$ 成正比,即与 $P(x \in k)$ 成反比。实验表明,否定概率和法确实能提高结果合成的正确率,从而提高整体的分类性能。

1.2.3 图像标注

分类系统将待识别图像判别为 K 类中的某一类时,通常是给出每一类概率意义上的“确信度”,再从中选出极大值对应的类别作为分类结果,而忽略其余的信息。文中关心的则是所有的确信度。在人工指定 K 个语义标签之后,对于每幅待识别图像,分类系统产生相应的 K 个确信度,语义标签和确信度一起组成 K 维标注向量。式(3)中 $f_k(x)$ 即为 OPC 方法第 k 类的确信度,它表示样本 x 具有第 k 种语义的概率。对于 PWC-SP 方法,先取式(4)中的 $\sum_{l=1}^K f_k^{(kl)}$ 作为第 k 维数据,再对标注向量进行归一化;对于 PWC-SNP 方法,取标注向量的第 k 维数据为 $K - \sum_{l=1, l \neq k}^K \vartheta(f_k^{(kl)}, f_l^{(kl)})$,再对其归一化。这样就建立起了图像的底层特征与高层语义之间的联系。

2 实验结果

实验使用的图像来自于 Corel Image Gallery^[10],从中挑选 25 个语义类,每个类包含 100 张图像,共计 2 500 张图像作为实验数据,这些图像的内容涵盖人物、自然风景、动物、植物、建筑等。

为了评价多类分类器的性能,从实验数据中随机选取每个语义类图像的 80 % 作为训练数据,余下的 20 % 作为测试数据。取 10 次随机实验的平均值作为其最终实验结果,如图 1 所示,纵坐标为分类精度

或召回率,横坐标为图像类别. 定义精确度 (precision)和召回率 (recall) 如下:

$$\text{precision} = \frac{a}{a + b}, \text{recall} = \frac{a}{a + d}$$

式中: a 为被正确分到类别 c 的样本数, b 为被错误分到类别 c 的样本数, d 为属于 c 类而未被分到 c 类的样本数.

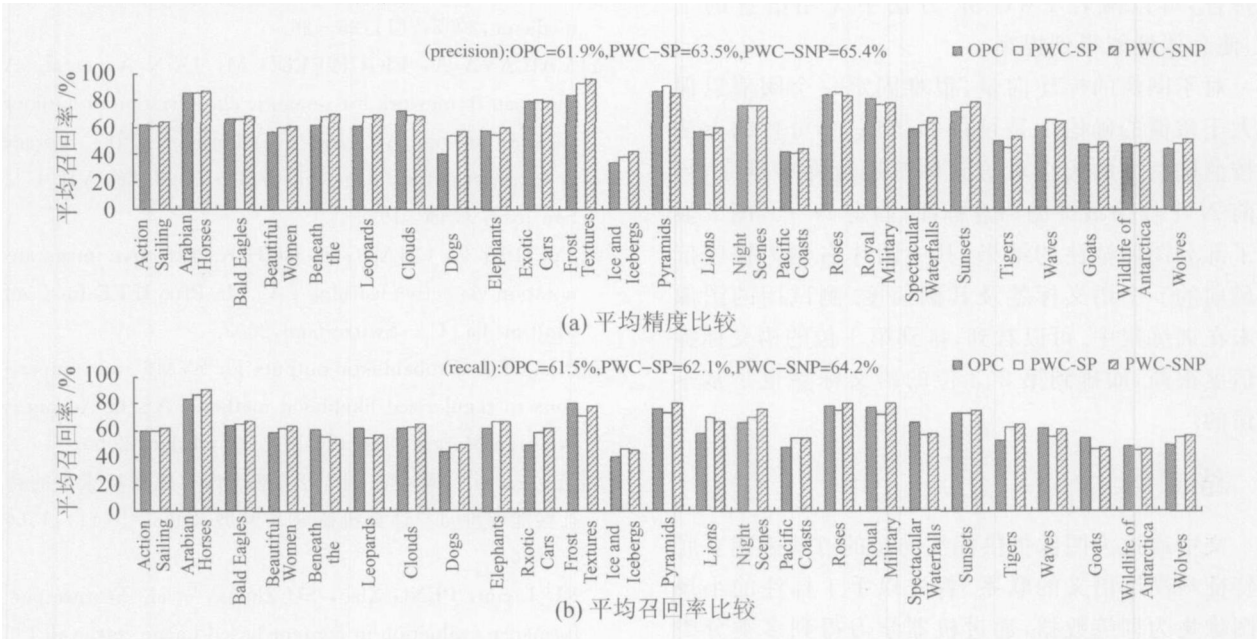


图 1 分类器性能评价

Fig. 1 Performance of multi-class classifiers

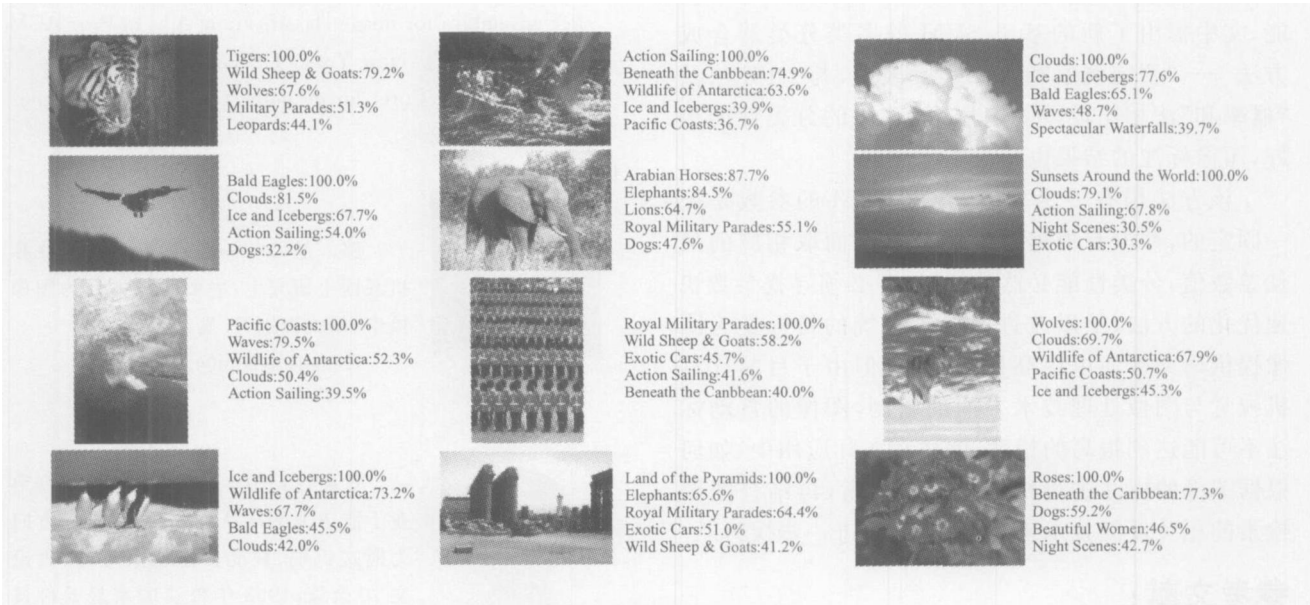


图 2 部分标注图像

Fig. 2 Some automatic annotated images

对于第 i 类,其分类精确度表示为 A_i ,则平均精度用下式计算:

$$P_{Ave} = \frac{1}{K} \sum_{i=1}^K A_i$$

平均召回率计算与之类似.就平均精度而言, PWC-SP 比 OPC 方法相对提高了 2.58%, PWC-

SNP 比 OPC 相对提高了 5.65%,其中语义类 “Dogs”性能相对提高了将近 38%;对于平均召回率, PWC-SP 比 OPC 相对提高了 1.00%, PWC-SNP 比 OPC 相对提高了 4.39%.无论精度还是召回率,对于 25 个语义类中大多数的类别, PWC 方式的性能优于 OPC, PWC-SNP 优于 PWC-SP,对余下的少数语义类, PWC-SNP 虽然略差于其他方法,但

性能相差得并不明显.这说明 PWC 方式在总体上可以减少 OPC 方式中训练样本数目不平衡带来的分类错误,而 PWC-SNP 方法也与前面的理论分析相符合,可以减轻 PWC-SP 方法中无用信息的干扰,使总体性能得到提高.

对于图像的标注向量,很难固定一个阈值以保证大于该值的确信度是可信的,但一个可行的方法是按照确信度的大小对语义标签进行降序排序,再取前 N 个作为最终的语义标注.固定 $N=5$,图 2 显示了部分图像标注的结果,其中图片右侧为排序位于最前的 5 个语义标签及其确信度.测试用的图像并未在训练集中.可以看到,排到第 1 位的语义标签可信度很高,而排到第 2、3 位的语义标签也是基本可信的.

3 结束语

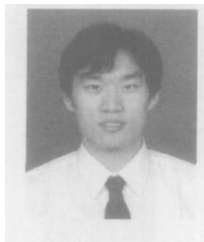
文中通过为图像提供标注向量的方式来建立底层特征与高层语义的联系.首先以手工标注的小规模图像集为训练数据,通过机器学习得到多类分类器,再对未标注的图像进行分类,获得以标注向量方式表示的图像高层语义.在标注向量中,确信度表示了语义与图像内容的相关程度.为了获得更好的性能,文中提出了新的基于 SVM 的多类分类器合成方法——“否定概率和”法.实验表明,与 OPC 法及“概率和”法相比较,“否定概率和”法的分类性能更好,图像标注的结果也是令人满意的.

该方法仍有不足之处,其中 SVM 的参数是统一固定的,并没有根据不同的图像类而取相应的最佳参数值,分类性能必然受到影响.必须寻找参数快速优化的方法,予以改进.另外,系统的目标是为图像提供语义标注以提高检索性能,但由于目前计算机视觉与图像处理技术方面的限制,图像的自动标注不可能达到很高的精度,因此在实际应用中,如何根据图像的标注进行快速的初步检索,再结合图像检索的相关反馈技术进行精化,值得进一步探讨.

参考文献:

- [1] 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 and Machine Intelligence, 2000, 22(12): 1349 - 1380.
- [2] WANG J, LI J, WIEDERHOLD G. Simplicity: semantics - sensitive integrated matching for picture libraries [J]. IEEE Transactions on Pattern Analysis Machine Intelligence, 2001, 23(9): 947 - 963.
- [3] SRIHARI R K, ZHANG Z F. A semi-automated image annotation system[J]. IEEE Multimedia, 2000, 7(3): 61 - 71.
- [4] TANG H L, HANKA R, HORACE H S. Histological image retrieval based on semantic content analysis[J]. IEEE Transactions on Information Technology in Biomedicine, 2003, 7(1): 26 - 36.
- [5] VAILAYA A, FIGUEIREDO M, JAIN A, et al. A Bayesian framework for semantic classification of outdoor vacation images [A]. In Proceedings of SPIE: Storage and Retrieval for Image and Video Databases VII [C]. San Jose: USA, 1999.
- [6] SYCHAY G, CHANGE E, GOH K. Effective image annotation via active learning [A]. In Proc IEEE Int Conf Multimedia [C]. Switzerland, 2002.
- [7] PLATT J. Probabilistic outputs for SVMs and comparisons to regularized likelihood methods [A]. In Advances in Large Margin Classifiers [C]. Cambridge, 2000.
- [8] 茹立云, 彭 潇, 苏 中, 等. 基于内容图像检索中的特征性能评价[J]. 计算机研究与发展, 2003, 40(11): 1566 - 1570.
- RU Liyun, PENG Xiao, SU Zhong, et al. Feature performance evaluation in content-based image retrieval[J]. Journal of Computer Research and Development, 2003, 40(11): 1566 - 1570.
- [9] GOH K, CHANGE E, CHENG K T. SVM binary classifier ensembles for image classification [A]. In Proc ACM CIKM 2001 [C]. New York, USA, 2001.
- [10] WANG J Z. Corel image gallery [OL]. <http://wang.ist.psu.edu>. 2000.

作者简介:



路 晶, 1980 年生, 清华大学计算机系博士研究生, 主要研究方向为图像检索、模式识别等. 发表论文 3 篇.

E-mail: postal @263.net



金奕江, 1970 年生, 助研, 博士, 毕业于清华大学计算机系, 主要研究方向为模式识别、智能信息检索等. 发表论文 10 余篇, 1998 年曾获国家教委科技进步二等奖.



马少平, 教授, 博士生导师, 毕业于清华大学计算机系, 主要研究方向为智能信息检索、模式识别等. 发表论文 50 余篇, 出版专著 3 部. 承担多项国家自然科学基金, “863” “973” 和国际合作项目. 1998 年获国家教委科技进步二等奖.