# 基于集成分类算法的自动图像标注

蒋黎星 [1] 侯 进 [1]

**摘 要** 基于语义的图像检索技术中, 按照图像的语义进行自动标注是一个具有挑战性的工作. 本文把图像的自动标注过程转化为图像分类的过程, 通过有监督学习对每个图像区域分类并得到相应关键字, 实现标注. 采用一种快速随机森林 (Fast random forest, FRF) 集成分类算法, 它可以对大量的训练数据进行有效的分类和标注. 在基于 Corel 数据集的实验中, 相比经典算法, FRF 改善了运算速度, 并且分类精度保持稳定. 在图像标注方面有很好的应用.

**关键词** 自动图像标注, 机器学习, 集成分类器, 快速随机森林算法

**引用格式** 蒋黎星, 侯进. 基于集成分类算法的自动图像标注. 自动化学报, 2012, **38**(8): 1257−1262

**DOI** 10.3724/SP.J.1004.2012.01257

## Image Annotation Using the Ensemble Learning

JIANG Li-Xing[1] HOU Jin[1]

**Abstract** Automatic image annotation is an important but highly challenging problem in semantic-based image retrieval. In this paper, we formulate image annotation as a supervised learning image classification problem under the region-based image annotation framework. In region-based image annotation, keywords are usually associated with individual regions in the training data set. This paper presents a novel ensemble fast random rorest algorithm (FRF), which can classify a large number of training data effectively by bootstrap aggregation (Bagging) algorithm building multiple tree component classifier. The final result is obtained by component classifier voting. The proposed FRF algorithm is experimented on image annotation Corel data sets. Compared to classical algorithms, the FRF accelerates the operation speed of the algorithm, and the classification accuracy remains robust. It has a good application in automatic image annotation system.

**Key words** Automatic image annotation, machine learning, ensemble learning, fast random forest (FRF)

**Citation** Jiang Li-Xing, Hou Jin. Image annotation using the ensemble learning. *Acta Automatica Sinica*, 2012, **38**(8): 1257−1262

随着互联网通讯技术与数字影像技术的快速发展, 用户可以通过各种方式轻松地获取大量的网络数字图像. “语义清晰” 是大规模数字图像管理的重要前提, 因此通过计算机信息技术自动获取图像语义内容的研究具有十分重要的理论与实践意义[1]. 自动图像标注作为图像检索的关键技术是获取图像语义信息的一种有效方法. 早期, 人们通过手工标注的方式来实现对图像内容的标注, 但这是一项既耗时又费力的工作[2]. 首先, 在面对海量的网络图像资源时, 完全用手工标注已经无法胜任; 其次, 人们在手工标注图像时会带有自己的主观感情色彩, 对于同一幅图像标注内容有时会不同, 很难保证标注内容的一致性和完整性; 最后, 对于实时的流媒体处理, 必须应用计算机软件进行实时的图像内容分析, 用手工标注是不可行的. 目前自动图像标注的方法主要分为两大类: 基于概率模型的自动图像标注方法和基于统计分类的图像标注方法[3]. 在基于概率模型的方法中, 主要是从已有标注信息的训练集中, 间接或直接地获取图像或区域的视觉特征 (如颜色、纹理、形状等), 然后学习图像视觉特征和高级语义特征 (文本关键词) 之间的联合概率分布, 最后利用所得出的概率分布对未标注的图像或区域进行标注. 如 Duygulu 等[4] 提出了机器翻译模型, 把图像区域和高层语义之间的过程看作是一个机器翻译的过程, 最终通过期望最大化 (Expectation-maximization, EM) 算法将区域翻译成关键字. 随后 Mori 等提出的 Co-occurrence 图像标注模型[5], Blei 等提出的线性判别式分析 (Linear discriminant analysis, LDA) 模型[6], Jeon 等提出的跨媒体相关模型 (Cross-media relevance modal, CMRM) 模型[7], 2003 年 Lavrenko 等提出的基于连续特征的相关模型 (Continuous-space relevance model, CRM)[8] 以及使用非参数高斯核的多伯努

收稿日期 2011-05-30 录用日期 2011-07-04
Manuscript received May 30, 2011; accepted July 4, 2011
教育部留学回国人员科研启动基金, 高等学校博士学科点专项科研基金 (20090184120022), 中央高校基本科研业务费专项资金科技创新项目 (SWJTU09CX036) 资助
Supported by the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry, the Specialized Research Fund for the Doctoral Program of Higher Education (20090184120022), and the Fundamental Research Funds for the Central Universities (SWJTU09CX036)
本文责任编委 刘一军
Recommended by Associate Editor LIU Yi-Jun
1. 西南交通大学信息科学与技术学院 成都 610031
1. School of Information Science and Technology, Southwest Jiaotong University, Chengdu 610031

利相关模型 (Multiple-Bernoulli relevance model, MBRM)[9] 都是基于概率的自动图像标注方法. 图像的标注字能够很好地表达图像的语义内容, 缩小图像的高级语义和低级视觉特征之间的差距, 使得大规模基于语义图像检索应用更具现实性. 通过对每个区域进行分类可以得到相应的类标签 (标注字), 最后实现整幅图像的自动标注. 这种方法把图像的标注转化成了图像的分类过程[10−15]. 在基于统计的图像分类方法中, Chang 等提出了基于支持向量机 (Support vector machines, SVM) 的分类方法[16−17], Lu 等在文献 [18] 中提出了使用基于多例学习的启发式 SVM 算法的图像自动标注方法, Hou 等提出了基于决策树 (Decision tree, DT) 分类算法的图像标注[11]. 本文中, 把图像的自动标注过程转化为图像分类的过程. 一幅图像可以包含若干个区域, 每个区域有不同的语义内容和不同的颜色纹理等底层特征. 首先可以把图像分割成具有一定视觉特征相似的各个区域 (Region). 提取每个区域的视觉图像特征信息, 构造训练集, 然后用集成算法对训练集进行分类训练, 每一个类对应一个类标签, 类标签可以是文本关键字, 最后实现图像的自动标注. 在系统中, 使用快速随机森林 (Fast random forest, FRF) 分类算法. FRF 是集成分类算法, 它使用多个简单分类器进行分类, 最后进行投票选择, 选出分类率最高的那个结果. FRF 算法比起传统的 SVM 算法和决策树算法 C4.5, 速度快、分类精度高, 在图像的自动标注中具有实用性.

## 1 FRF 算法

由分量分类器集合和组合规则构成的分类算法称为集成分类算法 (Ensemble learning). 随机森林 (Random forests, RF) 是一种集成分类算法[19], 它是由 Breiman 在 2001 年提出的. RF 采用 DT 作为元分类器, 用自主聚类 (Bootstrap aggregation, Bagging)[20] 方法产生各个不同的训练集, 用以生成各分量分类器, 最后通过简单多数投票决定最终分类结果, 如图 1 所示. Bagging 方法和单 DT 算法的结合, 再加上随机选择特征进行属性分裂, 使得 RF 能较好容忍噪声, 并具有较好的分类性能[21−22]. FRF 是 RF 的改进算法, 它通过增加一个空间变量来单独存储训练结果, 并在 Bagging 过程中对每个特征属性的实例进行排序, 在实验当中, 新算法的分类精度和运行时间都有所提高.

在 FRF 算法中, 每一棵树分类器都要对输入的特征向量进行投票选取. 当对输入的特征变量进行分类时, 首先要把该特征变量输入到森林中去, 该森林包含了多个 DT, 该特征变量要输入到每棵 DT 中, 每棵 DT 对输入的特征变量单独进行分类, 最后按分类的结果进行投票 (Voting)[23], FRF 算法选择得票率最多的作为分类结果输出. FRF 算法主要包含三个步骤: 1) 树的生长; 2) 投票的过程; 3) 分类的误差估计.

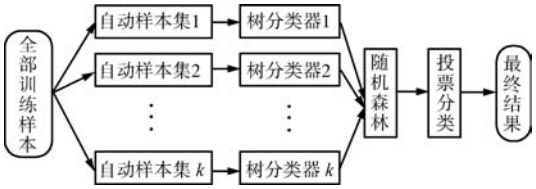


图 1 FRF 算法框图

Fig. 1 FRF algorithm diagram

### 1.1 树的生长

随机森林通过 Bagging 采样技术生成了多个树型分类器[21]. 假设原始训练集为 $N$, 新的自主样本集可以通过 Bagging 方法有放回地随机选择, 并可以构成 $k$ 棵分类树, 每次没有被抽取到的样本可以组成 $k$ 个袋外数据 (Out-of-bag, OOB). 主要操作步骤如下:

1) 从容量为 $N$ 的原始训练样本集合中, 采用有放回抽样的方法随机抽取 Bagging 样本集, 重复 $k$ 次.

2) 将每个自助样本集作为训练集生长成单棵 DT. 设训练集有 $n$ 个输入特征, 从 $n$ 个输入特征中随机选择 $t$ ($t \le n$) 个特征, 则在 DT 的每个结点处, 按照 $n$ 个特征的信息熵[11] 选出最佳拆分点进行树的分支生长.

a) 设训练数据集 $Train$ 一共有 $m$ 类样例, 每类样例数为 $|C_i|$ ($i = 1, 2, \cdots, m$). 以属性 $A$ 作为 DT 的根, 具有 $V$ 个值 $v_1, v_2, \cdots, v_v$, 假设子集中任意元组属于类 $C$ 的概率用 $p_i$ 表示, $p_i = \frac{|C_{i,Train}|}{|Train|}$. $|C_{i,Train}|$ 表示训练集中属于 $|C_i|$ 类的样本个数, $|Train|$ 为样本集 $Train$ 中的样本个数, 那么, 该子集的信息量定义如式 (1) 所示:

$$I(e_r) = -\sum_{i=1}^{m} p_i \log_2(p_i), \qquad r = 1, 2, \cdots, v \quad (1)$$

以 $A$ 为根分类后所需的信息期望如式 (2) 所示.

$$\mathrm{E}(A) = \sum_{j=1}^{v} \frac{|e_r|}{|Train|} \times I(e_r), \qquad r = 1, 2, \cdots, v \quad (2)$$

b) 以 $A$ 为根的信息增益如式 (3) 所示.

$$IG(A) = I(e_r) - \mathrm{E}(A) \quad (3)$$

通过选择信息增益最大的属性作为当前的根结点, 迭代此过程, 使每棵 DT 充分生长, 不用剪枝操作.

### 1.2　投票过程

由于 FRF 算法融合了一系列 DT 的分类结果, 在 FRF 中应该选择效果最好的分类结果, 需要在众多分类中采用投票的方法来预测结果, 让每一棵 DT 都对输入的特征向量进行投票, 计算所有投票数, 找出其中投票数最多的类别作为该特征向量的分类标签.

### 1.3　分类的误差估计

在上述方法中, 采用了放回抽样的方法, 在随机产生训练集的时候, 大概只有 2/3 的原始样本会被抽中作为 DT 的训练集, 剩余 1/3 没有被抽中的数据为袋外数据 OOB. 当一个分类树生成后, 用该树对其对应的全部 OOB 数据进行分类预测. 把这个分类预测结果和实际的类别相比较, 就得到了 FRF 算法的分类误差估计.

FRF 在处理过程中有很多优点, 如不需要进行预处理、不会产生过拟合现象 (Overfitting) 等. 比起传统的方法 RF, 对内存进行优化, FRF 算法增加一个 DataCatche 空间变量来单独存储训练结果; 在 Bagging 的方法中应用排序的方法首先对每个特征属性的实例进行排序, 有序地查找加快了算法的运算速度.

## 2　基于 FRF 算法的图像标注

基于 FRF 的自动图像过程如图 2 所示, 主要分为两个步骤: 标注模型的训练和模型的测试.

首先, 训练标注模型. 如图 2 虚线左侧所示. 在训练阶段首先对训练集中的每一幅图像进行分割, 由于一幅图像包含了若干个区域, 每个区域有不同的语义, 那么一个最直接的方法就是按照视觉特征对图像进行分割, 分割成若干个区域后提取每个区域的底层特征信息. 分割算法的好坏决定了区域的底层特征和图像标注的准确度. Duygulu 等采用 Normalized-cuts 分割算法[4] 提取每个区域的底层特征. 王向阳等提出的一种基于视觉权值的分块颜色直方图图像检索方法中对图像首先进行分块处理[24]. 在本文中, 使用 J 值分割法 (J-value segmentation, JSEG)[25] 对图像进行分割. 虽然目前图像的自动分割技术并不完善, 但已经可以提供图像中基于视觉特征的每个区域的信息. 有关文本关键字信息、分割后的图像和提取的区域实例见图 3.

对分割后的区域, 它们有大致相同的视觉特征, 可以用一组表示颜色、纹理等底层特征的向量值来描述. 图像的自动标注就是找到最佳的预定义文本关键词来描述这组特征值, 这组特征值又映射到特定的区域, 使图像的标注问题转换成图像的自动分类问题.

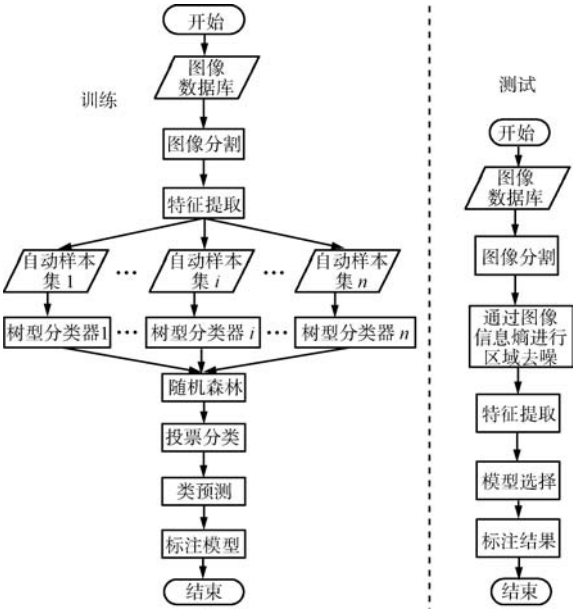


图 2　FRF 标注模型框图

Fig. 2　FRF algorithm annotation diagram

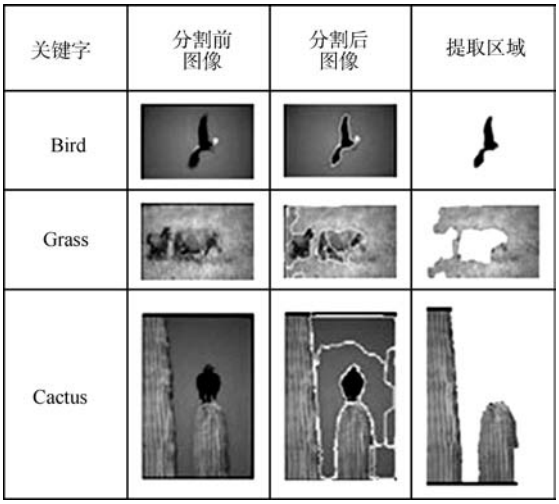


图 3　文本关键字、分割前后的图像和所提取区域实例

Fig. 3　Examples of keyword, image before and after segmentation and extracted regions

用 $Test$ 代表未标注的测试图像集, $Train$ 代表已经分好类的训练数据集. 通过对图像库中每一张图像进行分割, 可以得到多个视觉特征相似的区域 $R$ (Region).

每张训练区域 $R_i \in Train$ 可以通过一些区域视觉特征 $f_i = x_1, x_2, \cdots, x_n$ 和一组关键词 $k_i \subseteq K$ 来表示. 如 $x_1 = Color$, $x_2 = Texture$, $x_3 = Shape, \cdots$, $Color(C)$ 为颜色特征, $Texture(T)$ 为纹理特征, $Shape(S)$ 为形状特征. 每张测试区域 $R_j \in Test$, 也可以用一组特征值

$f_j = x_1, x_2, \cdots, x_n$ 来表示, 但测试域中关键字是未知的. $f_i$ 和 $f_j\,(i, j = 1, 2, \cdots, n)$ 是图像区域 $i$ 或 $j$ 的特征值. $K = k_1, k_2, \cdots, k_m$ 是关键词词典, 是词库 $k_i\,(i = 1, \cdots, m)$ 中的第 $i$ 个关键词.

图像自动标注的目的是从词库 $K$ 中选择多个关键词 $k$ 来最准确地描述图像区域 $R$, 对于训练集 $Train$, 包含了 $N$ 个图像区域和相对应的关键字 $Train = \{(R_1, k_1) \cdots (R_N, k_N)\}$. 如果把每一个关键字 $k_i$ 看作一个类标签, 那么图像的标注问题就可以转换成图像的分类问题, 这里首先要考虑的是对于给定的一组表示区域 $R$ 的特征向量, 哪一个类是属于包含在区域 $R$ 当中的.

尽管如此, 图像标注并不是一个传统的监督机器学习算法, 因为训练图像样本中并没有提供整个图片和类别的准确对应关系, 仅仅是关键词和图片中某个区域之间的对应关系. 如图 4 所示, 图像可以被标注为 “bird”, 但可以包含很多与 “bird” 不相关的其他区域, 如 “sky”, “mountain”, “sea”, 这些区域的底层特征值可以完全不同于 “bird” 区域. 大量的噪声数据会出现在训练集当中, 针对这些问题, 需要分类精度更高、适应性更强的分类算法.

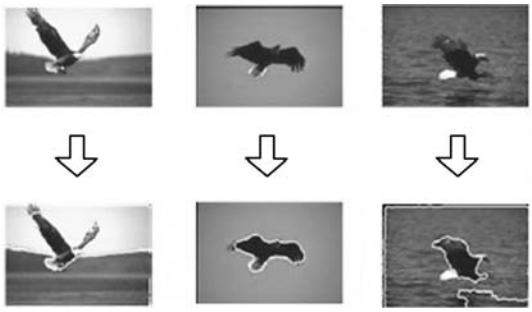

图 4 三幅图像中包含与 “bird” 不相关的区域: “sky”, “mountain”, “sea”

Fig. 4 Non-related regions of “sky”, “mountain”, and “sea”

接着提取每个区域视觉底层特征, 对特征进行前期处理, 用 Bagging 方法产生各个不同的训练集, 用以生成各分量分类器, 最后通过简单多数投票决定最终分类结果, 得到分类模型, 将标注词汇与各区域特征建立对应关系.

其次, 模型测试标注阶段. 如图 2 虚线右侧部分所示. 首先分割每一幅待标注的图像, 然后提取待测试区域的视觉特征, 通过图像信息熵进行区域去噪.

图像信息熵是把图像的颜色直方图看成概率密度函数, 假设 $h_i$ 表示被量化颜色 $i$ 的像素在整幅图像中所占的百分比, 从而图像的信息熵可以表示为

$$\mathrm{E}(H) = -\sum_{i=1}^{n} h_i \log_2(h_i) \tag{4}$$

在本文对 $\mathrm{E}(H)$ 取一个经验值, 过滤掉过分割的区域. 最后利用样本训练中的分类模型对待标注区域进行分类, 得到各个区域的类标签值实现自动图像标注.

## 3 实验结果与分析

本文在实验环境 Intel Core 2.0 GHz, 2.0 GB 内存, Window 7, Matlab 7 下进行. 实验采用 Corel[4] 图像库进行图像自动分类和标注, Corel 数据库包括了 5 000 张图片, 共有 50 个类, 每个类有 100 个图片实例. 对数据库中的每幅图像, 通过分割算法将其分成多个区域. 对每一个区域, 可以提取不同特征, 如颜色 $(Color, C)$、纹理 $(Texture, T)$、颜色纹理特征 $(Color\text{-}Texture, CT)$、形状 $(Shape, S)$ 或空间特征 $(Space, SP)$ 等, 用这些特征值来表示区域的内容. 特征属性的取值是一些能表示图像内容的数值, 用它们作为分类器的输入, 通过集成分类器可以对这些输入数据进行分类训练.

为了验证本文所提算法的有效性, 从 Corel 图像库中选取 5 组数据样本对算法进行训练比较. 实验中 FRF 与其他经典机器学习算法如 SVM, 朴素贝叶斯 (Native Bayes, NB), K-最近邻法 (K-nearest neighbor, KNN) 等算法进行比较. 使用十折交叉验证 (10-fold cross validation) 来评估分类算法的准确率.

如表 1 所示, 有不同的 5 组数据: $225C\_T\_CT\_SP$ 代表共 225 个训练样本, 提取图像的 $C$, $T$, $CT$ 和 $SP$ 特征值, 在训练样本中包含有 9 个类 (blue sky, bird, plane, firework, flower, grass, horse, mountain, and night); $270C\_T\_CT\_SP$ 代表共 270 个训练样本, 提取图像的 $C, T, CT$ 和 $SP$ 特征值, 在训练样本中有 9 个类; $380C\_T\_CT\_SP$ 代表共有 380 个训练样本, 提取图像的 $C$, $T$, $CT$ 和 $SP$ 特征值, 在训练样本中有 19 个类 (grass, forest, sky, sea, flower, sunset, sand, firework, tiger, fur, bird, building, snow, rock, bear, night, people, butterfly, and mountain); $1200C\_T\_S$ 代表共 1 200 个训练样本, 提取图像的 $C$, $T$, $CT$ 和 $SP$ 特征值, 在训练样本中有 40 个类 (mountain, grass, horse, plane, bird, blue sky, firework, night sky, flower, butterfly, cactus, car, bonsai, helicopter, building, balloon, light, sunset, cloud, tree, sculpture, house, orangutang, sea, boat, snow, road, sand, rock, stone, leaf, river, ground); Corel5K 是采用 ECCV2002 中 4 500 张样本图像进行训练.

使用以上 5 组数据样本训练随机森林. 最终 FRF 和其他分类算法的分类精度如表 1 所示. 分类精度和分类时间趋势比较可见图 5 和图 6.

表 1　基于 Corel 图像库中 5 组数据使用多种机器分类算法进行实验比较

Table 1　The classification comparison of a variety of classification algorithms based on 5 groups Corel training data

| 分类精度 (%)/时间 (s) | 225$C_T_CT_SP$ | 270$C_T_CT_SP$ | 380$C_T_CT_SP$ | 1200$C_T_S$ | Corel 5K |
|---|---|---|---|---|---|
| C4.5 | 81.33/0.03 | 79.25/0.01 | 76.32/0.04 | 72.68/0.01 | 64.00/1.08 |
| SVM | 83.56/0.82 | 79.26/1.12 | 75.53/4.18 | 80.93/12.2 | 49.80/26.1 |
| KNN | 72.44/0.01 | 69.63/0.01 | 71.58/0.01 | 80.93/0.01 | 73.20/0.07 |
| RF | 80.89/0.05 | 76.67/0.02 | 73.95/0.30 | 78.48/0.13 | 33.60/0.01 |
| FRF | 83.11/0.01 | 77.78/0.01 | 75.73/0.01 | 80.93/0.07 | 79.90/3.01 |

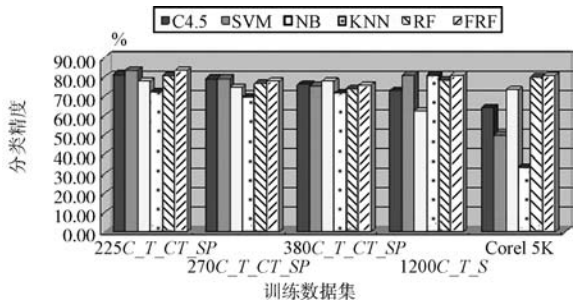


图 5　基于 5 组训练数据多种分类算法的分类精度比较

Fig. 5　Classification accuracy comparison of a variety of classification algorithms based on 5 groups training data

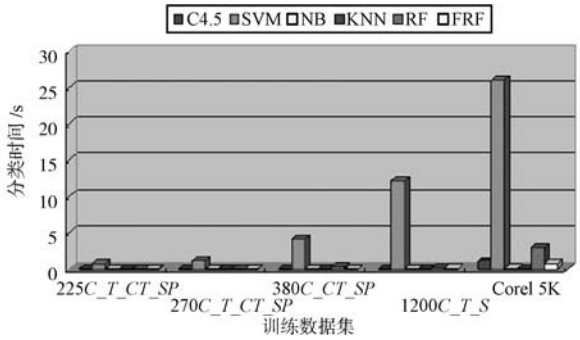


图 6　基于 5 组训练数据多种分类算法的分类时间比较

Fig. 6　The classification time comparison of a variety of classification algorithms based on 5 groups training data

从实验中可以看出, FRF 算法随着数据量的增加平均分类精度最高, 且平均训练时间为 0.182 s. 对同样的样本集用 SVM 或传统的 RF 算法进行训练, 分类精度相当, 但需要花费较多的时间, 平均训练时间分别为 8.884 s 和 0.702 s. 总结以上实验可以得出, 相比 SVM 和 RF 算法, 集成分类 FRF 算法有效地提高了训练速度, 并保持稳定的分类精度.

最后进行自动图像标注系统的实现和分析. 在 Corel 数据库中定义了 40 个关键词 (mountain, grass, horse, plane, bird, blue sky, firework, night sky, flower, butterfly, cactus, car, bonsai, helicopter, building, balloon, light, sunset, cloud, tree, sculpture, house, orangutang, sea, boat, snow, road, sand, rock, stone, leaf, river, ground), 每个词从 Corel 数据库中挑选了 30 个相关的模板区域, 这样形成一个共 1 200 个区域图像的训练数据组. 用集成算法 FRF 和其他机器学习算法进行图像的自动标注. 其结果如图 7.

| Example |  |  |  |  |  |
|---|---|---|---|---|---|
| C4.5 | grass, horse | sea, bird | firework, night | sea, nountain, sky cloud | sky, plane |
| SVM | grass, horse | cactus, bird | firework, cactus, tiger | mountain, butterfly, firework | sky, plane |
| NB | grass, bird | grass, bird | firework, night | sea, mountain, sky | sky, plane |
| KNN | grass, bird | grass, bird | firework, night | sea, mountain, sky | sky, plane |
| RF | grass, horse | sea, bird | firework, night | sea, mountain, sky | sky, plane |
| FRF | grass, horse | sea, bird | firework, night | sea, mountain, sky | sky, plane |

图 7　基于多种分类器算法的图像标注结果

Fig. 7　The annotation results based on a variety of classification algorithms

## 4　结论

本文采用了一种改进的集成分类算法 FRF, 该算法对基于 Corel 图像库的训练集, 能够很好地完成图像的分类和标注. 而且实验结果表明, 该算法具有稳健性, 随着数据量的增大, 分类精度基本保持不变. 此外, 改进后的集成算法 FRF 与 SVM、RF 相比, 分类精度相当, 但运算速度有很大的提高, 并且能满足用户及时性的需求, 在图像标注方面也有广泛的应用.

## References

1 Ulges A, Worring M, Breuel T. Learning visual contexts for image annotation from flickr groups. *IEEE Transactions on Multimedia*, 2010, **13**(2): 330−341

2 Rui Y, Huang T S, Chang S F. Image retrieval: past, present, and future. *Journal of Visual Communication and Image Representation*, 1997, **10**: 1−23

3 Xu Hong-Tao, Zhou Xiang-Dong, Xiang Yu, Shi Bai-Le. Adaptive model for web image semantic automatic annotation. *Journal of Software*, 2010, **21**(9): 2183−2195

(许红涛, 周向东, 向宇, 施伯乐. 一种自适应的 Web 图像语义自动标注方法. 软件学报, 2010, **21**(9): 2183−2195)

4 Duygulu P, Barnard K, Freitas J F G, Forsyth D A. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary. In: Proceedings of the 7th European Conference on Computer Vision. Copenhagen, Denmark: Springer, 2002. 97−112

5 Mori Y, Takahashi H, Oka R. Image-to-word transformation based on dividing and vector quantizing images with words. In: Proceedings of the 1st International Workshop on Multimedia Intelligent Storage and Retrieval Management. Orlando, Florida: MISRM, 1999. 1−9

6 Blei D, Jordan M. Modeling annotated data. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto, Canada: ACM, 2003. 127−134

7 Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models. In: Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Toronto, Canada: ACM, 2003. 119−126

8 Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures. In: Proceedings of the Advances in Neural Information Processing Systems. Vancouver, Canada: MIT Press, 2003. 553−560

9 Feng S L, Manmatha R, Lavrenko V. Multiple Bernoulli relevance models for image and video annotation. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). Washington D. C., USA: IEEE, 2004. 1002−1009

10 Chow T W S, Rahman M K M. A new image classification technique using tree-structured regional features. *Neurocomputing*, 2007, **70**(4−6): 1040−1050

11 Hou J, Chen Z, Qin X, Zhang D. Automatic image search based on improved feature descriptors and decision tree. *Integrated Computer-Aided Engineering*, 2011, **18**(2): 167−180

12 Fan J, Shen Y, Yang C, Zhou N. Structured max-margin learning for inter-related classifier training and multilabel image annotation. *IEEE Transactions on Image Processing*, 2011, **20**(3): 837−854

13 Nezamabadi-pour H, Kabir E. Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval. *Expert Systems with Applications*, 2008, **36**(3): 5948−5954

14 Pourghassem H, Ghassemian H. Content-based medical image classification using a new hierarchical merging scheme. *Computerized Medical Imaging and Graphics*, 2008, **32**(8): 651−661

15 Tsai C F. Image mining by spectral features: a case study of scenery image classification. *Expert Systems with Applications*, 2007, **32**(1): 135−142

16 Chang E, Kingshy G, Sychay G, Wu G. CBSA: content-based soft annotation for multimodal image retrieval using Bayes point machines. *IEEE Transactions on Circuits and Systems for Video Technology*, 2003, **13**(1): 26−38

17 Yang C B, Dong M, Hua J. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). New York, USA: IEEE, 2006. 2057−2063

18 Lu Jing, Ma Shao-Ping. Region-based image annotation using heuristic support vector machine in multiple-instance learning. *Journal of Computer Research and Development*, 2009, **46**(5): 864−871
(路晶, 马少平. 使用基于多例学习的启发式 SVM 算法的图像自动标注. 计算机研究与发展, 2009, **46**(5): 864−871)

19 Breiman L. Random forests. *Machine Learning*, 2001, **45**(1): 5−32

20 Efron B. Bootstrap methods: another look at the Jack-knife. *The Annals of Statistics*, 1979, **7**(1): 1−26

21 Rokach L. Ensemble-based classifiers. *Artificial Intelligence Review*, 2010, **33**(1−2): 1−39

22 Breiman L. Bagging predictors. *Machine Learning*, 1996, **24**(2): 123−140

23 Xu L, Krzyzak A, Suen C Y. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Transactions on Systems, Man, and Cybernetics*, 1992, **22**(3): 418−435

24 Wang Xiang-Yang, Yang Hong-Ying, Zheng Hong-Liang, Wu Jun-Feng. A color block-histogram image retrieval based on visual weight. *Acta Automatica Sinica*, 2010, **36**(10): 1489−1492
(王向阳, 杨红颖, 郑宏亮, 吴俊峰. 基于视觉权值的分块颜色直方图图像检索算法. 自动化学报, 2010, **36**(10): 1489−1492)

25 Deng Y, Manjunath B S. Unsupervised segmentation of color-texture regions in images and video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2001, **23**(8): 800−810

**蒋黎星** 西南交通大学信息科学与技术学院硕士研究生. 主要研究方向为机器学习与图像标注.
(**JIANG Li-Xing** Master student at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers machine learnig and image annotation.)

**侯 进** 西南交通大学信息科学与技术学院副教授. 主要研究方向为多媒体检索与人机交互. 本文通信作者.
E-mail: jhou@swjtu.edu.cn
(**HOU Jin** Associate professor at the School of Information Science and Technology, Southwest Jiaotong University. Her research interest covers multimedia information retrieval and human computer interaction. Corresponding author of this paper.)