

刘邱云,王璐璐,黄涛.一种基于 Logistic 回归的基本信度分配函数的构造新方法[J].江西师范大学学报(自然科学版),2022,46(3):277-281.

LIU Qiuyun, WANG Lulu, HUANG Tao. The new method based on Logistic regression of constructing basic belief assignment function [J]. Journal of Jiangxi Normal University( Natural Science) 2022 46( 3) : 277-281.

文章编号: 1000-5862( 2022) 03-0277-05

# 一种基于 Logistic 回归的基本信度分配函数的构造新方法

刘邱云<sup>1</sup>, 王璐璐<sup>2</sup>, 黄涛<sup>1</sup>

( 1.江西师范大学数学与统计学院,江西 南昌 330022; 2.江西经济管理干部学院财务与金融学院,江西 南昌 330088)

**摘要:** 结合 Logistic 回归分类,该文提出一种新的构造证据理论基本信度分配函数的方法,并将其应用于多特征图像分类.该方法首先以多类 Logistic 回归分类法输出的后验概率与样本分类正确率建立证据权重系数,然后构造出加权的基本信度分配函数,最后利用加权 D-S 证据融合判别所属类别.实验结果显示:该方法既能提高图像分类的正确率,又能改正使用单特征分类导致的分类正确率的不稳定的缺点.

**关键词:** Logistic 回归; 基本信度分配函数; 加权 D-S 证据融合; 图像分类

中图分类号: TP 18 文献标志码: A DOI: 10.16357/j.cnki.issn1000-5862.2022.03.10

## 0 引言

Logistic 回归<sup>[1-3]</sup>模型分类方法是在统计分析领域中基于具体模型的分类方法,它在处理分类问题时,既能对样本所属的类别进行预测,又能对分类的相关概率信息进行计算.此方法常被用于疾病诊断、经济预测及数据挖掘等领域<sup>[4-5]</sup>.

证据理论<sup>[6]</sup>作为不确定性推理的一种重要方法,能较好地表达及融合决策层的不确定信息,较广泛地应用在决策分析、信息融合以及模式识别等领域<sup>[7-9]</sup>中.使用证据理论做不确定性推理,首先要解决如何表示不确定性信息,即构造基本信度分配(BBA)函数问题.

在运用证据理论时,基本信度分配函数作为一种集值随机变量即随机集<sup>[10]</sup>,它的构造方式与具体运用紧密相关.李世诚等<sup>[11]</sup>利用马尔可夫随机场(MRF)构造 BBA 函数,建立了图像邻域证据场.童涛等<sup>[12]</sup>借助 SVM 分类结果作为独立证据生成 BBA 函数,解决了对 SAR 图像目标的有效分类.李新德

等<sup>[13]</sup>结合 PNN 网络生成目标识别矩阵以获取 BBA 函数,提出了一种面向多特征飞机图像目标的信息融合方法.S. Petit-Renaud 等<sup>[14]</sup>基于非参数回归分析针对已知的输入向量,预测输出变量的值,得到模糊信任分配(FBA)函数,并构造为集值模糊集上的一个 BBA 函数.XU Peida 等<sup>[15]</sup>用一种非参数方法生成 BBA 函数,以处理测试样本与概率间的关系模型中的分类问题.M.C. Garrido 等<sup>[16]</sup>针对不完全信息和异构的分类问题,基于回归方法构造了不完整数据集上的证据函数.就以上研究趋势可以发现:基本信度分配函数的构造侧重利用问题内在的不确定性<sup>[17]</sup>,尤其是对无法用概率描述的不确定性建立模型,以体现证据理论在处理不确定性问题中的优势.与以上研究相比较,Logistic 回归是一种更简单的模型,尤其对于大规模的线性分类,操作更方便.Logistic 回归利用非线性映射,将离分类平面比较远的点的权重大大降低,而与分类更相关的数据点的权重就得到了相对提高.基于此,本文结合 Logistic 回归分类,提出一种新的证据理论基本信度分配函数的构造方法.

收稿日期: 2021-11-04

基金项目: 江西省教育厅科学技术研究( GJJ181392, GJJ191687) 资助项目.

作者简介: 刘邱云( 1976-),女,江西铜鼓人,讲师,主要从事不确定性推理和描述逻辑的研究.Email: lqyxinxiang@ 126.com

本文的主要贡献有如下 2 方面: (i) 首先以多类 Logistic 回归分类法输出的样本后验概率和分类正确率建立证据权重系数, 其次再建立加权的基本信度分配函数, 最后通过加权 D-S 证据融合进行决策, 判别样本所属的类别; (ii) 将构造的新方法应用于多特征图像分类.

## 1 Logistic 回归分类

### 1.1 2 类 Logistic 回归分类法

假设  $n$  个训练样本是  $\{x, y\}$ , 其中  $x = (x_1, x_2, \dots, x_n)$ ,  $x_i$  是  $d$  维的样本特征向量  $y = \{y_1, y_2, \dots, y_n\}$ , 此处用 1 和 0 作为类别标签, 分别表示正类和负类, 则将样本  $x$  归入类别中正类的“概率”为

$$p(y = 1 | x; \theta) = g(\theta^T x) = 1 / (1 + e^{-\theta^T x}),$$

这里  $\theta$  是  $d$  维的模型参数, 也就是回归系数  $g(\theta^T x)$  是 Logistic 函数.

本文使用极大似然估计方法来求解模型参数, 由于直接对似然函数进行最大化通常较麻烦, 因此先将似然函数取对数, 再最大化. 这里,

$$l(\theta) = \ln \prod_{i=1}^n P(y_i | x_i, \theta) = \ln \prod_{i=1}^n f(x_i)^{y_i} (1 - f(x_i))^{1-y_i}.$$

进一步, 转换成代价函数:

$$J(\theta) = -\frac{1}{n} \sum_{i=1}^n y_i \ln f(x_i) + (1 - y_i) \ln(1 - f(x_i)).$$

事实上, 对代价函数最小化就相当于对似然估计最大化, 再结合梯度下降法对  $l(\theta)$  的极大值进行求解, 最终得到参数  $\theta$ .

### 1.2 多类 Logistic 回归分类法

为了将 2 类回归分类法推广为多类回归分类法, 这里使用 one-vs-all 策略, 即假设分类问题中有  $c$  个类, 将其中某一个类作为一类, 而其余的类作为另一类, 以此建立一个 2 类分类器. 按这种方式一共可以建立  $c$  个 2 类分类器. 假设类别标签  $y \in \{1, 2, \dots, c\}$ ,  $c$  个参数  $\theta_i (i = 1, 2, \dots, c)$ , 则

$$p(y = i | x) = e^{\theta_i^T x} / \sum_{k=1}^c e^{\theta_k^T x}. \quad (1)$$

## 2 权重 BBA 函数的构造及分类

### 2.1 权重 BBA 函数的构造

定义 1 已知识别框架为  $\Theta$ , 若函数  $m: 2^\Theta \rightarrow [0, 1]$  满足  $\sum_{A \subseteq \Theta} m(A) = 1$ ,  $m(\emptyset) = 0$ , 则称  $m$  为  $\Theta$  上

的基本信度分配函数.

定义 2 设同一识别框架  $\Theta$  上有 2 个独立证据, 其基本信度分配函数分别为  $m_1$  和  $m_2$ , 则用 D-S 组合规则融合后的证据  $m = m_1 \oplus m_2$  为

$$m(C) = \frac{1}{K} \sum_{A, B \in \Theta, A \cap B = C} m_1(A) m_2(B), \quad (2)$$

其中  $C \neq \emptyset$ ,  $K = \sum_{A, B \in \Theta, A \cap B \neq \emptyset} m_1(A) m_2(B)$ , 当  $C = \emptyset$  时  $m(C) = 0$ .

定义 3 假设有  $Q$  个样本, 其中所属类分类正确的样本个数为  $q$ , 则样本分类正确率为

$$E = q/Q. \quad (3)$$

定义 4 假设分类器包含  $c$  个类, 其识别框架为  $\Theta = \{L_1, L_2, \dots, L_c\}$ ,  $E(L_i)$  为分类器对第  $i$  类的分类正确率, 则第  $i$  类的权重系数分配函数为

$$W(L_i) = nE^k(L_i) / \sum_{j=1}^c E^k(L_j), \quad (4)$$

其中  $n = |\Theta|$ , 这里正整数  $k$  越大表明分类正确率越高, 即赋予的权重越大. 由此可见权重系数  $W(L_i)$  体现了各证据的可靠程度.

定理 1 假设分类训练样本集的样本共有  $n$  个特征, 使用 Logistic 回归分类法对第  $j$  个特征进行训练, 由式 (3) 得分类正确率为  $E_j(L_i)$ , 由式 (4) 得权重系数为  $W_j(L_i) (j = 1, 2, \dots, n; i = 1, 2, \dots, c)$ ,  $x^s$  为待识别样本, 利用式 (1) 代入第  $j$  个特征可求得  $p_j(y = i | x^s)$ , 引入折扣因子  $\alpha (\alpha \geq 1)$  有

$$p_j(L_i) = \begin{cases} \alpha p_j(y = L_i | x^s) & L_i = \operatorname{argmax} p_j(y = L_i | x^s) \\ p_j(y = L_i | x^s) & \text{否则} \end{cases}$$

于是建立第  $j$  个特征对应的权重 BBA 函数  $mW_j$  为

$$\begin{cases} mW_j(L_i) = W_j(L_i) p_j(E_j(L_i) / (1 - \\ p_j(E_j(L_i) + \sum_{i=1}^n W_j(L_i) p_j(E_j(A_i))) \\ mW_j(\Theta) = 1 - \sum_{i=1}^n mW_j(L_i) \end{cases} \quad (5)$$

### 2.2 分类步骤

步骤如下:

- (i) 提取样本的若干个特征 (假设有  $n$  个);
- (ii) 利用训练样本集中类别已知的样本, 求解回归参数  $\theta_i (i = 1, 2, \dots, c)$ ;
- (iii) 通过式 (5) 建立加权的基本信度分配函数  $mW(L_i)$ ;
- (iv) 利用式 (2) 对这些基本信度分配函数做

D-S 证据融合  $mW_1 \oplus mW_2 \oplus \cdots \oplus mW_n(L_i)$  最后根据  $\arg\max mW_1 \oplus mW_2 \oplus \cdots \oplus mW_n(L_i) = L_{i_0}$  做决策,即将待分类样本  $x^s$  归入第  $i_0$  类。

### 3 应用于图像分类

为了将上述新方法应用于图像分类,本文选取在 Scene 15 场景分类数据集集中的 kitchen( $K$ )、livingroom( $L$ ) 及 bedroom( $B$ ) 3 类室内场景图像作为分类对象。每个类型含 200 幅图像,随机选取 100 幅图像作为训练样本,另外 100 幅图像作为验证样本。对比模型为 Hu 不变矩、单一纹理特征和加权 D-S 证据融合特征。

由于在图像分类中,图像的 Hu 不变矩与纹理特征相互独立,所以使用加权 D-S 证据融合的做法是合理的。

模型应用于图像分类的流程图如图 1 所示,其步骤如下:

(i) 提取场景图像样本的特征量——Hu 不变矩和纹理特征;

(ii) 根据 2 个特征量,对类别已知的 100 幅图像进行训练,求解出 Logistic 回归参数  $\theta_i (i = 1, 2, 3)$ , 并依次计算出 2 个特征量的正确率和权重系数,然后借助式(1)将验证样本中提取的特征代入,得到  $p(y = L_i | x)$  其中  $L_i \in (K, L, B)$ ;

(iii) 通过式(5)构造加权的基本信度分配函数  $mW(L_i)$  其中  $L_i \in (K, L, B)$ ;

(iv) 最后通过式(2)做 D-S 证据融合,并根据  $\arg\max mW_1 \oplus mW_2(L_i)$  判定验证样本的类别。

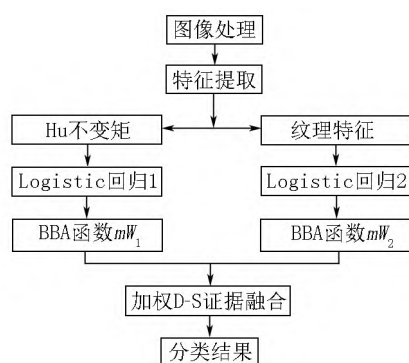


图1 图像分类流程图

#### 3.1 提取图像特征

对于图像分类,本文提取在图像矩特征量中的 Hu 不变矩及在纹理特征中的灰度共生矩阵。

3.1.1 Hu 不变矩 图像的矩特征是以图像分布的各阶矩来描述灰度的统计特征的方法,它具有较好的平移、旋转和比例不变性。借助 2 阶和 3 阶中心矩, M.K. Hu 构造了如下表达式的 7 个不变矩:

$$f_1 = \eta_{20} + \eta_{02},$$

$$f_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2,$$

$$f_3 = (\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2,$$

$$f_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2,$$

$$f_5 = (\eta_{30} - 3\eta_{12})(\eta_{30} - \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{30} + \eta_{12})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2),$$

$$f_6 = (\eta_{20} - \eta_{02})((\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2) + \eta_{11}(\eta_{30} + \eta_{12})(\eta_{03} + \eta_{21}),$$

$$f_7 = (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 - 3(\eta_{03} + \eta_{21})^2) + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{03})(3(\eta_{30} + \eta_{12})^2 - (\eta_{03} + \eta_{21})^2),$$

其中  $\eta_{pq} (p, q = 0, 1, 2, 3)$  表示归一化中心矩。用 100 个训练样本和 100 个验证样本的 7 个 Hu 不变矩特征构造出图像的矩特征向量为

$$F_A = (f_1, f_2, f_3, f_4, f_5, f_6, f_7).$$

3.1.2 纹理特征 图像的纹理特征通过灰度共生矩阵的 4 个指标(熵  $T_1$ 、能量  $T_2$ 、对比度  $T_3$ 、相关性  $T_4$ )来描述。其中熵  $T_1$  描述了纹理的复杂程度和非均匀程度,能量  $T_2$  描述了图像的纹理粗细度和灰度分布的均匀程度,对比度  $T_3$  描述了某像素值的亮度与其邻域像素值的亮度的对比情况,相关性  $T_4$  描述了纹理的一致性。它们的表达式分别为

$$T_1 = \sum_{i,j=0}^{L-1} P_{i,j}(-\ln p_{i,j}), T_2 = \sum_{i,j=0}^{L-1} P_{i,j}^2,$$

$$T_3 = \sum_{i,j=0}^{L-1} P_{i,j}(i-j)^2,$$

$$T_4 = \sum_{i,j=0}^{L-1} P_{i,j}(i-\mu_i)(j-\mu_j) / \sqrt{\sigma_i^2 \sigma_j^2},$$

其中  $P_{i,j}$  即为  $P((i,j)/d, \theta)$ 。这里  $\theta$  表示方向;  $d$  表示距离;  $P$  为灰度共生矩阵;  $\mu_i$  表示灰度共生矩阵各行的均值,  $\mu_j$  表示各列的均值;  $\sigma_i$  表示灰度共生矩阵各行的标准差,  $\sigma_j$  表示各列的标准差。构造出图像的纹理特征向量为  $F_B = (T_1, T_2, T_3, T_4)$ 。

#### 3.2 实验与结果分析

首先从提取到的 3 类室内场景图像样本的 Hu 不变矩和纹理特征中随机选出 100 个训练样本,利

用多类 Logistic 回归分类法进行参数学习,以获得优化参数  $\theta_i (i = 1, 2, 3)$ ; 再把训练样本代入训练好的模型中以获取上述 3 类样本的分类正确率  $E_j(L_i)$ , 结果如表 1 所示。

表 1 各特征的分类正确率

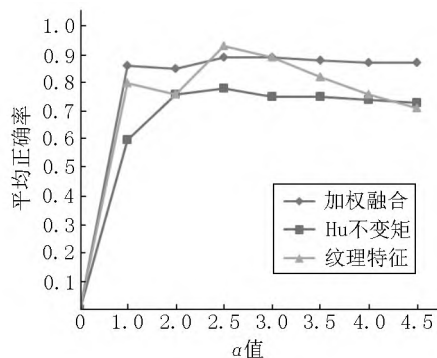
特征	$E(K)$	$E(L)$	$E(B)$
Hu 不变矩	0.74	0.88	0.67
纹理	0.76	0.87	0.83

然后以上述特征为证据体( $k = 2$ ), 按照定义 4 处理得到 2 个特征的加权系数, 结果如表 2 所示。

表 2 各特征的加权系数

特征	$W(K)$	$W(L)$	$W(B)$
Hu 矩	0.926 8	1.061 0	1.012 2
纹理	0.969 4	1.152 8	0.877 7

最后对 100 个验证样本进行测试, 针对不同的折扣因子  $\alpha$  值, 分别得到样本单一特征的平均分类正确率与加权融合的平均分类正确率, 结果如图 2 所示。

图 2  $\alpha$  取不同值的各特征的分类正确率

由图 2 可知: 通过加权 D-S 证据融合后的平均分类正确率高于单一特征的平均分类正确率; 根据单一的纹理特征或 Hu 不变矩特征分类的正确率不稳定, 尤其是单一的纹理特征起伏更加明显, 而加权 D-S 证据融合特征分类的正确率高且相当稳定, 因此可信度较高。

当折扣因子  $\alpha = 3.5$  时, 各特征的分类正确率和  $F$  均值结果如表 3 所示。其中图像 bedroom( $B$ ) 样本采用 2 种单一特征分类的正确率分别为 0.64、0.72, 采用加权 D-S 证据融合特征后正确率提高到 0.86; 图像 livingroom( $L$ ) 样本的分类结果类似, 采用 2 种单一特征分类的正确率分别为 0.79、0.78, 采用加权 D-S 证据融合特征后正确率提高到 0.84; 而图像 kitchen( $K$ ) 样本的 Hu 不变矩的分类正确率低于加权 D-S 证据融合特征后的分类正确率,  $K$  样本的纹

理特征的分类正确率高于加权 D-S 证据融合特征后的分类正确率, 恰好与 D-S 证据理论的特点相符, 即降低证据之间的冲突程度, 增大可信度。根据  $F$  均值分析数据可得类似结果。

表 3 各特征的正确率混合矩阵

特征	类别	$K$	$L$	$B$	$E(L_i)$	$F$ 均值
Hu 不变矩	$K$	81	0	36	0.81	0.75
	$L$	0	79	0	0.79	0.88
	$B$	19	21	64	0.64	0.63
纹理特征	$K$	97	16	15	0.97	0.85
	$L$	3	78	13	0.78	0.81
	$B$	0	6	72	0.72	0.81
加权融合	$K$	93	0	0	0.93	0.96
	$L$	0	84	14	0.84	0.85
	$B$	7	16	86	0.86	0.82

## 4 结论

本文提出了一种新的基于 Logistic 回归分类模型的基本信度分配函数的构造方法, 并且在多特征图像分类上验证了该方法的有效性。这种新方法利用多类 Logistic 回归分类法, 分别基于单一纹理特征和 Hu 不变矩对图像样本进行初步识别, 并以 Logistic 回归分类法输出的后验概率与分类正确率建立证据权重系数, 从而构造出加权的基本信度分配函数; 最后利用加权 D-S 证据理论对样本的各个特征信息进行有效融合, 并根据融合后的最大值对样本所属的类别做决策。实验结果显示: 本文提出的新方法实现了多特征的有效融合, 既提高了分类的正确率, 又改正了使用单特征导致的分类正确率的不稳定的缺点。

## 5 参考文献

- [1] 任康, 李刚. Logistic 回归模型在判别分析中的应用 [J]. 统计与信息论坛, 2007, 22(6): 71-73.
- [2] OWEN A B. Empirical likelihood for linear models [J]. The Annals of Statistics, 1991, 19(4): 1725-1747.
- [3] PATRICK G O W, WANG Chengguan, WU Congxin. Nearly best linear estimations of logistic parameters based on complete ordered statistics [J]. Journal of Harbin Institute of Technology (New Series), 2001, 8(2): 178-183.
- [4] 杨柏辰, 张怡, 郭伟. 基于 Logistic 回归模型的企业财务风险诊断研究 [J]. 系统科学学报, 2021, 29(3): 84-87.
- [5] 李振杰, 韩杰. 基于 Logistic 回归模型的农户土地流转

- 意愿实证分析 [J]. 统计与决策, 2019, 35(13): 110-114.
- [6] YAGER R R. On the Dempster-Shafer framework and new combination rules [J]. Information System, 1989, 41(2): 93-137.
- [7] SEVASTIANOV P, DYMOVA L, BARTOSIEWICZ P. A framework for rule-based evidential reasoning in the interval setting applied to diagnosing type 2 diabetes [J]. Expert Systems with Applications, 2012, 39(4): 4190-4200.
- [8] MOKHTARI K, REN Jun, ROBERTS C, et al. Decision support framework for risk management on sea ports and terminals using fuzzy set theory and evidential reasoning approach [J]. Expert Systems with Applications, 2012, 39(5): 5087-5103.
- [9] 张春英, 冯晓泽, 刘凤春, 等. 基于 D-S 证据理论的三支决策推理模型 [J]. 模糊系统与数学, 2020, 34(6): 85-98.
- [10] 韩崇昭, 朱洪艳, 段战胜. 多源信息融合 [M]. 2 版. 北京: 清华大学出版社, 2010.
- [11] 李世诚, 韩德强, 杨艺, 等. 基于邻域证据场的图像分割算法 [C]//中国自动化学会控制理论专业委员会. 第 32 届中国控制会议论文集(C 卷). 西安: 西北工业大学出版社, 2013: 3828-3833.
- [12] 童涛, 杨桃, 李昕, 等. 基于 D-S 证据理论的多特征融合 SAR 图像目标识别方法 [J]. 国土资源遥感, 2013, 25(2): 37-41.
- [13] 李新德, 杨伟东, DEZERT J. 一种飞机图像目标多特征信息融合识别方法 [J]. 自动化学报, 2012, 38(8): 1298-1307.
- [14] PETIT-RENAUD S, DENOEUX T. Nonparametric regression analysis of uncertain and imprecise data using belief functions [J]. International Journal of Approximate Reasoning, 2004, 35(1): 1-28.
- [15] XU Peida, SU Xiaoyan, MAHADEVAN S, et al. A non-parametric method to determine basic probability assignment for classification problems [J]. Applied Intelligence, 2014, 41(3): 681-693.
- [16] GARRIDO M C, CADENAS J M, BONISSONE P P. A classification and regression technique to handle heterogeneous and imperfect information [J]. Soft Computing, 2010, 14(11): 1165-1185.
- [17] 韩德强, 杨艺, 韩崇昭. DS 证据理论研究进展及相关问题探讨 [J]. 控制与决策, 2014, 29(1): 1-11.

## The New Method Based on Logistic Regression of Constructing Basic Belief Assignment Function

LIU Qiuyun<sup>1</sup>, WANG Lulu<sup>2</sup>, HUANG Tao<sup>1</sup>

(1. School of Mathematics and Statistics, Jiangxi Normal University, Nanchang Jiangxi 330022, China;

2. School of Accounting and Finance, Jiangxi Institute of Economic Administrators, Nanchang Jiangxi 330088, China)

**Abstract:** Combined with Logistic regression classification, the new method of constructing the basic belief assignment function of evidence theory is presented and it is applied in the multi-feature image classification. Firstly, the weight coefficient of the evidence is established based on the posterior probability output by multi-class Logistic regression classification method and the samples' classification accuracy. Secondly, the weighted basic belief assignment function is constructed. Finally, the weighted D-S evidence fusion is used to distinguish the category. Experimental results show that the new method can not only improve the accuracy of image classification, but also overcome the instability of classification accuracy caused by single feature classification.

**Key words:** Logistic regression; basic belief assignment function; weighted D-S evidence fusion; image classification

(责任编辑: 曾剑锋)