

# 基于随机森林的自适应特征选择算法

刘凯, 郑山红, 蒋权, 赵天傲

(长春工业大学 计算机科学与工程学院, 吉林 长春 130012)

**摘要:** 为了解决传统的随机森林算法在随机特征选择时, 导致少数比较重要的特征变量被过滤掉的问题, 以及没有考虑特征变量相关性对预测应变量准确性带来的影响, 提出了一种基于随机森林的自适应特征选择算法 SARFFS。该算法首先利用卡方检验样本间关联程度后自助采样, 并设计出一种特征对类代表强弱程度的计算方法; 然后引入自适应稀疏约束机制 Group LASSO 优化特征的选择; 最后在 Spark 分布式计算平台利用 UCI 数据集进行实验。结果表明, 相比传统的 RF 算法, SARFFS 算法在特征子集选择上具有更好的性能, 在  $F_1$  上提升将近 9%。从最终排名靠前的特征分析, 该算法能够考虑特征间相关性, 对预测结果确实有影响, 并有效地提高了随机属性权值的可靠性和稳定性。

**关键词:** 随机森林; 自适应; 特征选择; Group LASSO 方法

中图分类号: TP301.6

文献标识码: A

文章编号: 1673-629X(2018)09-0101-04

doi: 10.3969/j.issn.1673-629X.2018.09.021

## A Self-adaptive Feature Selection Algorithm Based on Random Forest

LIU Kai ZHENG Shan-hong JIANG Quan ZHAO Tian-ao

(School of Computer Science and Engineering, Changchun University of Technology,  
Changchun 130012, China)

**Abstract:** In order to solve the problem that a small number of important variables are filtered out in the selection process of random feature adopted by the method of traditional random forest algorithm, and without considering the influence of characteristic variable correlation on the accuracy of prediction variables, we propose a self-adaptive feature selection algorithm SARFFS based on random forests. It first uses the Chi square to test the degree of association between samples and then bootstrap sampling, and we design a method for calculating the intensity and degree of the class represented by the feature. Then an adaptive sparse constraint mechanism Group LASSO is introduced to optimize the selection of the features. Finally, the experiments are carried out on the Spark distributed computing platform using UCI data sets, which shows that compared with the traditional RF algorithm, the SARFFS has better performance in feature subset selection, and the efficiency has been increased nearly 9% in the  $F_1$ . From the analysis of important characteristics of the final ranking, the proposed algorithm can consider the correlation having an effect on the prediction results definitely, and improves the reliability and stability of random attribute weights effectively.

**Key words:** random forest; self-adaptive; feature selection; Group LASSO method

## 0 引言

随机森林(random forest, RF)是在数据挖掘、生物信息学、管理学、医学等领域中解决高维度和非线性样本的一种分类器<sup>[1]</sup>。如何从高维度的数据选择最优的特征选择算法<sup>[2]</sup>、构建分类和回归算法,从而达到更好的预测精度,是众多学者研究的重要方向。

Tang F 等<sup>[3]</sup>针对海量数据集中数据严重缺失的问题,基于 RF 算法提出了 mForest 算法,利用 RF 算法

的插补性能进一步增强了随机特征之间的相关性,实验表明 mForest 算法相比于 missForest 算法在效率上提升高达 10 倍;Vrushali Y Kulkarni 等<sup>[4]</sup>通过对随机森林算法中样本进行不同维度的划分并赋予相应的权值,在一定程度上提升了随机森林算法的分类精度和运行效率,但对于多类数据集的分类效果不是非常明显;Oshiro T M 等<sup>[5]</sup>通过大量的数学理论解释了随机森林在重复抽样中选择不同的训练集能得到更准确

收稿日期: 2017-08-09

修回日期: 2017-12-20

网络出版时间: 2018-04-28

基金项目: 吉林省自然科学基金资助项目(20130101060JC);吉林省教育计划“十二五”科学技术研究基金资助项目(2014131, 2014125)

作者简介: 刘凯(1989-),女,硕士研究生,研究方向为智能计算、数据挖掘;郑山红,教授,CCF 会员(40111M),通信作者,研究方向为智能计算、语义 Web。

网络出版地址: <http://kns.cnki.net/kcms/detail/61.1450.TP.20180427.1630.022.html>

确度; Bernard S 等<sup>[6]</sup>重点研究了随机森林算法中每个子分类器的强度与相关性的关系,但不能在构建子树之前对森林的强度和相关度进行有效预估; Gall J 等<sup>[7]</sup>采用序列和广义序列的后向选择方法进行特征选择,并提出了一种基于随机森林的封装式特征选择算法,但该算法在高维数据集中并不能确定广义后向搜索方法中的  $L$  值。

综上所述,上面都对随机森林算法进行了有意义的改进,但本质上没有很好地解决随机森林算法自身的缺陷。所以,针对 RF 在随机选择属性时属性权重可信度低以及在构造单棵决策树时算法整体分割能力差的问题,提出了一种基于随机森林的自适应特征选择算法 SARFFS。

## 1 随机森林算法

定义 1: 随机森林<sup>[1,8]</sup>是一个在多个决策树分类器  $\{h(X, \Theta_k) \mid k = 1, 2, \dots, K\}$  构建 Bagging 集成的基础上进一步组合而成的大型集成分类器。构建 RF 算法的基本思想: 首先,从原始数据集中利用 bootstrap 方法有放回地随机抽取  $K$  个新样本集,并用新样本构建  $K$  棵分类回归树;其次,假设有  $n$  个特征,在每棵树的每个节点随机选择  $m_{\text{try}}$  个特征,其中  $m_{\text{try}} < n$ ,计算每个特征的信息熵,并通过概率值大小选择分类能力最强的特征进行节点的分裂;最后,无需裁剪,直接将生成的多棵树组成随机森林对新数据进行分类,分类结果采用简单多数投票机制。

## 2 SARFFS 算法

算法 SARFFS 是基于 Spark 分布式大数据计算平台<sup>[9-10]</sup>。本节主要分成三部分: 一种新的多元特征提取方法的提出、Group LASSO 对 RF 特征选择的优化和 SARFFS 算法生成及核心推理过程。

### 2.1 新的多元特征提取方法

随机森林构建 CART 树的整个过程中会有两点不足: 一是特征随机选择会影响属性权重计算的不准确,二是随着迭代次数的增加,数据集逐步增大,导致算法的运行效率大大降低。文中提出一种新的多元特征提取方法,充分考虑到每个特征对最终分类能力影响的差异性<sup>[11]</sup>,使用卡方检验<sup>[12]</sup>计算样本之间的关联度权重,并提出一种计算特征代表类的权重的计算方法,最后利用线性加权在内的两种权重计算方法来代替传统算法中依靠随机性进行特征选择的方法。

新的多元特征提取方法特征值对类的代表能力程度 = 特征值在全部分类中的分布的集中程度 \* 特征值在某个类中出现的次的计算方法来给特征赋予有意义的权重

最终的特征代表类的能力大小实质上就是给每个特征赋予一个区分程度的权重,样本间的关联程度可以通过卡方检验公式计算得出。

这种融合的方法充分利用了两个分类算法的特性, SARFFS 算法不仅很好地克服了传统决策树在多元特征时整体优化的不足,而且大大提升了随机森林区分子决策树并进行分割的能力。

### 2.2 Group LASSO 优化特征的选择

传统的随机森林算法存在对随机选择最优的特征对决策树进行切分而导致没有考虑特征之间关联性的问题,比如每次从总的特征集中随机抽取一组特征时,一方面会导致相关性的特征会重复抽取,另一方面具有代表性的一组特征又出现被分开的问题<sup>[13]</sup>。对此, Friedman 等<sup>[14]</sup>指定一组变量同时选进或者选出,提出稀疏 Group LASSO 方法来保留样本内重要的变量。该方法不仅去除了 Group LASSO 不重要的组,同时让组中的变量选择更有灵活性,其公式如下:

$$\hat{\theta} = \operatorname{argmin} \left\{ \frac{1}{2n} \|y - X\theta\|^2 + \alpha \|\theta\|_{c_2} + \beta \|\theta\|_{c_1} \right\} \quad (1)$$

其中,当  $\alpha = 0$  时,式 1 就是 LASSO;当  $\beta = 0$  就是 Group LASSO。文中采取的方法是设置,用该方法能够在每次随机抽取的部分特征中保持内部特征对分类结果的一致性,即多分类作业非常强的特征会挑选在同一组,而相反比较弱的特征就会进入到下次的采样过程中,这样很好地保证了每次选择出的特征是最优的,从而达到特征选择优化的效果。

### 2.3 SARFFS 算法的生成过程

利用 2.1 节提出的特征提取方法,并引用 2.2 中建立的架构算法代替 RF 算法中的传统决策树算法,建立了以传统随机森林为基本架构的优化算法 SARFFS,保留了传统算法的业务解释和稳定性。

SARFFS 算法的构建主要分两步: 一是通过多元特征提取方法对训练集筛选出候选集;二是把候选集放入到逻辑回归和决策树融合的算法中,从而产生多棵随机生成树。其中采用 Bagging 抽样获得样本的个数  $n$ ,多元特征提取方法采用线性加权样本的卡方检验关联度  $W_1$  来度量,计算如式 2:

$$\omega_1 = \chi^2 = \sum_{i=1}^m \frac{(A_i - np_i)^2}{np_i} \quad i = 1, 2, \dots, m \quad (2)$$

其中,  $A_i$  为实际特征的频数;  $np_i$  为理论特征的频数;  $W_1$  反映特征间的相关联程度。

计算特征代表类的能力程度  $W_2$ , 如式 3 所示:

$$\omega_2(i, A_i) = \frac{|A_i^n|}{m_i} * \left( 1 - \frac{1}{\log k} * \sum_{j=1}^k p(A_j^n) \log(p(A_j^n)) \right) \quad (3)$$

其中,  $|A_i^n|$  表示每一个特征出现的频数, 后面的因子表示用信息熵度量属性值在所有类中分布的集中程度。通过  $W = \lambda_1 W_1 + \lambda_2 W_2$  (其中  $\lambda_1$ 、 $\lambda_2$  是权重系数) 对特征的重要性  $W$  进行排序, 作为算法特征选择重要性的依据。

通过 Group LASSO 优化特征的选择后放入算法中训练出的算法函数为:

$$S(x) = \operatorname{argmax} \sum_{i=1}^k \left( \frac{1}{2} \sum_{i=1}^N (y_i - \sum_{j=1}^J \omega_2(i, A_j) \theta_j)^2 + \lambda \sum \|\theta_j\|_2 \right) \quad (4)$$

最后通过反复调参得到最优分类器, 对算法的分类精确度、最优特征的集合进行输出, 并在验证集上进行算法的验证和评估。综上所述, 列出 SARFFS 算法的生成过程, 如图 1 所示。

SARFFS算法:  
输入: 原始数据集S  
构建随机森林中的选用样本树的个数 $N_f$   
样本特征 $A=(A_1, A_2, \dots, A_m)$   
构建分类树的个数 $T$   
输出: 随机森林R  
测试集上预测的最大分类正确率TGMRate及对应的经过权重排序后特征集合FGSort  
算法:  
1 初始化数据集S, 设置TGMRate=0  
2 计算每个特征的卡方检验关联度并给每个特征赋予权重 $W_1$   
$$w_1 = \chi^2 = \sum_{i=1}^m \frac{(A_i - np_i)^2}{np^i} \quad (i=1, 2, \dots, m)$$
  
3 计算每个特征对分类的弱弱的代表能力, 其权重为 $W_2$   
$$w_2(i, A_j) = \frac{|A_i^n|}{m'} * (1 - \frac{-1}{\log k} * \sum_{j=1}^k p(A_j^n) \log(p(A_j^n)))$$
  
4 令 $W=\lambda_1 W_1 + \lambda_2 W_2$  ( $0 < \lambda_1, \lambda_2 < 1$ ), 并根据权值的大小进行特征排序  
5 for( $i=1$  to  $T$ )  
5.1 采用Bagging抽样, 获得样本个数 $n_i$   
5.2 根据特征权重 $W$ 挑选出特征子集  
5.3 通过Group Lasso方法选择出的组特征训练出的分类模型 $S(x)$   
$$S(x) = \operatorname{argmax} \sum_{i=1}^k \left( \frac{1}{2} \sum_{i=1}^N (y_i - \sum_{j=1}^J \omega_2(i, A_j) \theta_j)^2 + \lambda \sum \|\theta_j\|_2 \right)$$
  
6 for( $i=1$  to  $N_f$ ) {  
6.1 在S上运行NRF创建分类器 $C_i$   
6.2 在测试集上执行predict进行分类, 并计算预测分类精度ResultLRate  
if(TGMRate<ResultLRate){  
TGMRate=ResultLRate[i]  
对排序后的特征集合保存在FGSort中  
}  
}  
7 输出随机森林R、最大分类正确率TGMRate以及对应的特征集合FGSort

图1 SARFFS 算法的生成过程

### 3 实验

#### 3.1 实验数据与实验环境

实验数据均来自 UCI 机器学习数据库, 挑选了 Sonar\_lisan、Ionosphere、Glass\_lisan、Vehicle\_lisan 等 10 个数据集进行测试, 如表 1 所示。

表1 实验数据

样本	样本数 $n$	条件属性 $a$	类别个数 $c$
Sonar_lisan	208	60	2
Ionosphere	315	34	2
Glass_lisan	214	9	7
Vehicle_lisan	846	18	7
Ecoli	336	7	8
Breast	699	9	2
Wine	178	13	3
Wpbe	198	33	2
Car	1 728	6	4
Haberman	306	3	2

实验采用 3 台虚拟机搭建集群环境, 每台虚拟机的硬件配置为 8 cores、8 G 内存、50 G 磁盘, 操作系统版本为 Centos6.5, Spark 版本为 1.6.5, Hadoop 版本为 2.6.2。

#### 3.2 实验结果分析

在 Spark 集群上运行 SARFFS 算法, 并配置好 SparkR 的环境,  $n_{tree}$  的个数从 10 棵树到 100 棵树并以每次 10 棵树的规模递增。通过实验得出,  $n_{tree} = 80$  时算法效果最佳, 特征选取  $m_{try}$  参数为  $\sqrt{n}$  (其中  $n$  为训练集中特征的个数) 最佳, 数据集划分比例为 90% 进行训练, 10% 进行测试, 实验重复 50 次。

选择最佳的参数设置<sup>[15-16]</sup>, 根据 SARFFS 算法中新的多元特征提取方法, 可以得到特征的选择过程和结果以及特征个数与分类精度之间的关系, 如图 2 所示。

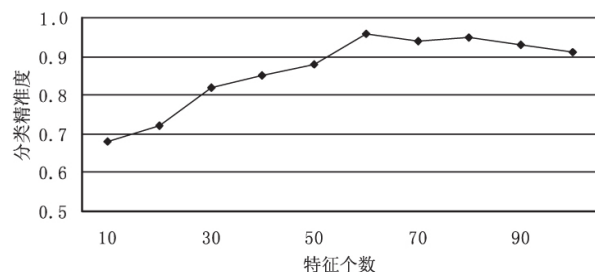


图2 特征个数与分类精度关系曲线

从图 2 中可以得出结论, 当特征数为 60 时, 分类精度达到最高, 这主要是因为 SARFFS 在迭代计算过程中会对一些不重要和不相关的特征进行动态消除或更替, 这样可以确保权重高的特征被优先选择; 当分类准确率达到最高峰值后会呈现下降趋势, 这主要是因为算法删除了一些有用特征之后, 可以根据实验效果人为地控制特征选择, SARFFS 算法可以在每次迭代过程中输出当前最优的特征和分类准确率。如表 2 所示, 经过多元特征提取方法提取的最优特征比随机选择的特征具有更好的分类能力, 这里的最优特征并不是指所有的特征。这 8 组实验结果是在最高分类点时不断删除或替换最优特征而得到的, 此外可以通过最优特征之间的变化情况及时调整特征的参数。这充分说明 SARFFS 算法具有高效的最优特征提取率, 并可以判断哪些特征对其相应的类别信息代表程度高一些, 进而提升分类器的分类精度。

表 3 列出了 SARFFS 算法与 RF 算法在相同数据集上分类精度的对比。从中可以明显看出, 在不同的数据集上 SARFFS 算法相比于传统 RF 算法的分类性能都要好一点, 这主要是因为 SARFFS 算法采用整体分类能力强的逻辑回归与决策树融合替代 RF 算法中的传统决策树算法, 克服了 RF 算法整体分割能力差的问题。

表 2 最优特征下对应的分类正确率

$L$	选择出的最优特征	分类正确率
1	V10 ,V15 ,V30 ,V20 ,V27 ,V9 ,V8 ,V11 ,V15 ,V7 ,V4 ,V2 ,V3 ,V25 ,V18	0.967 6
2	V15 ,V20 ,V9 ,V8 ,V4 ,V15 ,V7 ,V25 ,V13 ,V17 ,V18	0.933 2
3	V10 ,V15 ,V30 ,V20 ,V27 ,V9 ,V8 ,V11 ,V15 ,V7 ,V4 ,V3	0.929 4
4	V15 ,V8 ,V27 ,V11 ,V9 ,V8 ,V15 ,V7 ,V24 ,V25 ,V10	0.934 5
5	V10 ,V15 ,V9 ,V27 ,V18 ,V13 ,V25 ,V4 ,V1 ,V3	0.950 1
6	V10 ,V15 ,V11 ,V20 ,V30 ,V9 ,V8 ,V11 ,V15 ,V18 ,V4 ,V2 ,V28 ,V25	0.985 6
7	V10 ,V8 ,V27 ,V15 ,V30 ,V7 ,V4 ,V2 ,V3 ,V11 ,V15 ,V25 ,V18	0.963 2
8	V15 ,V20 ,V15 ,V9 ,V27 ,V5 ,V23 ,V3 ,V25 ,V18	0.956 3

表 3 SARFFS 与 RF 的分类误差对比

数据集	SARFFS	RF
Sonar_lisan	8.95	13.23
Ionosphere	11.32	34.28
Glass_lisan	7.63	4.31
Vehicle_lisan	10.32	9.56
Ecoli	18.64	15.65
Breast	6.87	10.65
Wine	15.76	14.34
Wpbe	3.54	3.45
Car	8.65	9.53
Haberman	22.65	14.27

表 3 采用  $F_1$  分数<sup>[17]</sup> 分析 SARFFS 和 RF 分类算法的对比效果,进一步通过平衡  $F_1$  分数比较 SARFFS 与 RF 的分类性能。

文中选择了 Sonar\_lisan、Ionosphere、Glass\_lisan 等 8 个数据集,分别计算了分类算法的准确率、召回率和  $F_1$  的值。从表 4 中可以得出, SARFFS 算法比传统 RF 算法在分类精度上提升了将近 9%,说明 SARFFS 算法使用逻辑回归和决策树融合代替决策树分类器在一定程度上解决了传统决策树整体分割能力不足的问题。

表 4 SARFFS 和 RF 在不同数据集下的效果对比

数据集	SARFFS			RF		
	precision	recall	$F_1$	precision	recall	$F_1$
Sonar_lisan	67.55	73.33	70.32	64.49	70.05	67.15
Ionosphere	70.50	73.12	71.79	65.03	74.12	69.28
Glass_lisan	67.90	78.72	72.91	75.83	61.59	67.97
Vehicle_lisan	68.93	78.03	73.20	69.26	70.82	70.04
Ecoli	69.51	76.88	73.01	67.39	70.00	68.67
Breast	71.91	73.51	72.70	64.48	70.52	67.36
Wine	72.64	72.36	72.50	72.34	67.83	70.01
Wpbe	72.98	71.78	72.38	72.34	67.83	70.01

## 4 结束语

提出了一种新的随机森林算法 SARFFS,该算法主要采用一种新的多元特征提取计算方法代替以前的随机特征选取策略,并利用 Group LASSO 方法对特征选择进行优化,通过 UCI 数据进行实验,结果表明 SARFFS 算法的线性加权计算方法能选择出更优的特征,并在分类性能上使平衡  $F_1$  分数提高近 9%。

未来的研究工作将会从两方面着手:第一,在特征选择上继续优化选择的策略;第二,与梯度提升决策树 (gradient boost decision tree,GBDT) 算法进行对比分析。

## 参考文献:

- [1] BREIMAN L. Random forests [J]. Machine Learning, 2001, 45(1): 5-32.
- [2] 荣盘祥, 曾凡永, 黄金杰. 数据挖掘中特征选择算法研究 [J]. 哈尔滨理工大学学报, 2016, 21(1): 106-109.
- [3] TANG F, ISHWARAN H. Random forest missing data algorithms [J]. Statistical Analysis & Data Mining, 2017, 10(6): 363-377.
- [4] KULKARNI V, SINHA P. Efficient learning of random forest classifier using disjoint partitioning approach [C] // Proceedings of the world congress on engineering. [s.l.]: [s.n.], 2013.
- [5] OSHIRO T M, PEREZ P S, BARANAUSKAS J A. How many trees in a random forest [C] // Proceedings of the 8th international conference on machine learning and data mining in pattern recognition. Berlin, Germany: Springer - Verlag, 2012: 154-168.
- [6] BERNARD S, HEUTTE L, ADAM S. Towards a better understanding of random forests through the study of strength and correlation [C] // International conference on emerging intelligent computing technology & applications. Ulsan, South Korea: [s.n.], 2009: 536-545.
- [7] GALL J, YAO A, RAZAVI N et al. Hough forests for object detection, tracking and action recognition [J]. IEEE Transac-

(下转第 111 页)

- 240.
- [36] GIONIS A ,INDYK P ,MOTWANI R. Similarity search in high dimensions via hashing [C]//Proceedings of the 25th international conference on very large data bases. [s.l.]: Morgan Kaufmann Publishers Inc. ,1999:518-529.
- [37] VADREVU S ,TEO C H ,RAJAN S ,et al. Scalable clustering of news search results [C]//ACM international conference on web search and data mining. Hong Kong ,China: ACM , 2011:675-684.
- [38] 卞艺杰,陈超,马玲玲,等.一种改进的 LSH/MinHash 协同过滤算法[J].计算机与现代化,2013,29(12):19-22.
- [39] DAS A S ,DATAR M ,GARG A ,et al. Google news personalization: scalable online collaborative filtering [C]//Proceedings of the 16th international conference on world wide web. Banff ,Alberta ,Canada: ACM ,2007:271-280.
- [40] REZAEIAN N ,NOVIKOVA G M. Detecting near-duplicates in russian documents through using fingerprint algorithm sim-hash [J]. Procedia Computer Science, 2017, 103:421-425.
- [41] 龙志禧,程蔚.基于词聚类的热点话题检测算法[J].计算机工程与设计,2011,32(6):2214-2216.
- [42] 李婷玉.基于语义的文本事件信息抽取方法的研究与实现[D].上海:上海交通大学,2012.
- [43] 许旭阳,李弼程,张先飞,等.基于事件实例驱动的新闻文本事件抽取[J].计算机科学,2011,38(8):232-235.
- [44] 赵江江.开放域事件抽取与微博事件检测跟踪[D].哈尔滨:哈尔滨工业大学,2013.
- [45] 张阔,李涓子,吴刚,等.基于关键词元的话题内事件检测[J].计算机研究与发展,2009,46(2):245-252.
- [46] 杨文漪.面向微博的事件检测算法研究[D].北京:北京邮电大学,2013.
- [47] 李艳,郝身刚,赵卫东,等.时间敏感的社交网络热点话题检测[J].计算机工程与设计,2014,35(12):4324-4328.
- [48] 郭跬秀,吕学强,李卓.基于突发词聚类的微博突发事件检测方法[J].计算机应用,2014,34(2):486-490.
- [49] 冯永,韩楠,贾东风.云计算环境下基于代表点增量层次密度聚类的微博事件检测及跟踪[J].计算机应用,2013,33(12):3559-3562.
- [50] 费绍栋,杨玉珍,刘培玉,等.融合情感过滤的突发事件检测方法[J].计算机应用,2015,35(5):1320-1323.
- [51] NING Yue ,MUTHIAH S ,TANDON R ,et al. Uncovering news-Twitter reciprocity via interaction patterns [C]//Proceedings of the 2015 IEEE/ACM international conference on advances in social networks analysis and mining 2015. Paris , France: ACM ,2015:1-8.
- [52] HUA Ting ,CHEN Feng ,ZHAO Liang ,et al. STED: semi-supervised targeted-interest event detection in twitter [C]//Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. Chicago ,Illinois ,USA: ACM ,2013:1466-1469.
- +++++
- (上接第104页)
- tions on Pattern Analysis and Machine Intelligence, 2011, 33(11):2188-2202.
- [8] 邓生雄,雒江涛,刘勇,等.集成随机森林的分类模型[J].计算机应用研究,2015,32(6):1621-1624.
- [9] ZAHARIA M ,CHOWDHURY M ,FRANKLIN M J ,et al. Spark: cluster computing with working sets [C]//Proceedings of the 2nd USENIX conference on hot topics in cloud computing. Boston ,MA: ACM ,2010:10.
- [10] ZHAO Bo ,ZHOU Hucheng ,LI Guoqiang ,et al. ZenLDA: an efficient and scalable topic model training system on distributed data parallel platform [C]//Proceedings of computer science. [s.l.]: [s.n.], 2015.
- [11] 马友忠,贾世杰,张永新.基于卡方分布的高维数据相似性连接查询算法[J].计算机应用,2016,36(7):1993-1997.
- [12] YANG Chun ,MOHAMMADI A ,CHEN Qingwei. Multi-sensor fusion with interaction multiple model and Chi-Square test tolerant filter [J]. Sensors, 2016, 16(11):1835.
- [13] 唐亮,段建国,许洪波,等.基于互信息最大化的特征选择算法及应用[J].计算机工程与应用,2008,44(13):130-133.
- [14] FRIEDMAN J ,HASTIE T ,TIBSHIRANI R. A note on the group lasso and a sparse group lasso [J]. Statistics, 2010, 4(1):58-63.
- [15] SMOLISKI U ,BARKALOV A ,TIARENKO L. Adaptation of the two sources of code and one-hot encoding method for designing a model of microprogram control unit with output identification [J]. Intelligent Control & Automation, 2015, 6(2):242-261.
- [16] 姚登举,杨静,詹晓娟.基于随机森林的特征选择算法[J].吉林大学学报:工学版,2014,44(1):137-141.
- [17] SEPULVEDA J ,VELASTIN S A. F1 score assesment of gaussian mixture background subtraction algorithms using the MuHAVi dataset [C]//International conference on imaging for crime prevention & detection. London ,UK: IEEE ,2015:1-6.