

文章编号: 2095-2163(2020)06-0137-05

中图分类号: TP391.41

文献标志码: A

基于棒状像素与随机森林的道路场景理解

柳明, 黄影平, 胡福志

(上海理工大学 光电信息与计算机工程学院, 上海 200093)

摘要: 本文针对棒状像素(Stixel)在表达障碍物时缺少语义信息的问题,提出了一种生成棒状像素级类别标签的方法。通过颜色、纹理和几何高度特征的联合,推断出像素级语义标签,即障碍物大类、植物与天空。将棒状像素位置信息融合至像素级语义图,用类别标签来优化每个棒状像素的顶点估计后,后以其内部像素的优势类来判定类别,得到语义棒状像素图。分类器选用能够快速推理且具抗过拟合能力的随机森林(RF)分类器。实验结果表明,本文提出的方法能够有效地对棒状像素进行语义表达,在一定程度上优化棒状像素在纹理缺失区域的高度估计。

关键词: 棒状像素; 像素级语义标签; 随机森林

Road Scene Understanding Using Stixel and Random Forest Classifier

LIU Ming, HUANG Yingping, HU Fuzhi

(School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China)

[Abstract] In order to solve the problem of lacking semantic information when Stixel is used to express obstacles, this paper proposes a method to generate category labels of rod-like pixel level. Through the combination of color, texture and geometric height features, the pixel-level semantic labels, namely obstacle categories, plants and the sky, are inferred. The position information of the bar pixel is fused to the semantic bar pixel graph. After the vertex estimation of each bar pixel is optimized by the category label, the category is determined by the advantage class of its internal pixels, and the semantic bar pixel graph is obtained. The random forest (RF) classifier with fast reasoning and resistance to overfitting was selected. Experimental results show that the method proposed in this paper can effectively express the semantic meaning of rod-like pixels and optimize the height estimation of rod-like pixels in the texture missing region to some extent.

[Key words] Stixel; Scene semantic representation; Random Forest

0 引言

棒状像素(Stixel)模型是表达交通场景的一种有效方法。每个棒状像素宽度占据几个像素,垂直于地平面且具有一定高度。Badino等人首次提出棒状像素模型,依据立体视觉图像计算占据栅格图,获取棒状像素的基点信息,使用灰度代价函数计算前景与背景边界,以确定棒状像素的高度^[1]。Chen等为了避免视差模糊带来的误差,提出了一种基于U视差的棒状像素建立算法,有效提高了部分场景下的估计精度^[2]。棒状像素模型以其紧凑的3D表达优势,能很好地实现对道路场景中各种障碍物的检测,因此被广泛采用。尽管如此,现有的棒状像素模型只是解决了对障碍物的检测任务,没有对各类障碍物进行辨识并进行类别标注。本文的主要工作是在棒状像素检测模型的基础上,进一步对棒状像素进行语义标注。

对于视觉方法,交通场景理解的目的是将语义

类标签分配给图像中的像素或者区域。如Carreira等、Ladický等的自由形态区域分类和Cheng等的区域像素级分类,采用自下而上的方式操作且通常不会恢复对象级的场景表示^[3-5]。在这种情况下,密集特征和无序模型(BOF)被证明有效^[6],其步骤包括:自由形状区域内的局部特征提取,给定区域内所有特征的编码和空间池化,以及随后的混合特征向量的判别分类。

在优化棒状像素估计方面,Sanberg等扩展了一个在线的自监督颜色模型,以便更好地分离地面和障碍物^[7]。与他们的工作不同,本文提出的棒状像素级语义标签生成算法使用类标签来增强棒状像素表示。两种方法都能在一定程度上解决视差图误差或稀疏时,棒状像素算法产生错误估计的问题。但本文的方法在棒状像素优化的同时,还为其添加了语义信息。

作者简介: 柳明(1994-),男,硕士研究生,主要研究方向:机器视觉;黄影平(1966-),男,博士,教授,博士生导师,主要研究方向:无人驾驶与机器视觉;胡福志(1994-),男,硕士研究生,主要研究方向:无人驾驶与机器视觉。

收稿日期: 2020-03-22

1 方法描述

1.1 总体框架

图1为算法的总体框架图。首先从原始彩色图中获取色彩与纹理特征,从原始图像的视差图中获得像素级的高度特征。以颜色、纹理和高度为特征,使用随机森林分类器来实现像素级特征通道至语义标签的映射,将可行驶区域以上的场景分为障碍物(包括建筑物、车辆、行人)、植物和天空三大类,实现像素级语义表达。像素级语义图中的可行驶区域(道路区域)通过棒状像素图^[2]获得。将棒状像素

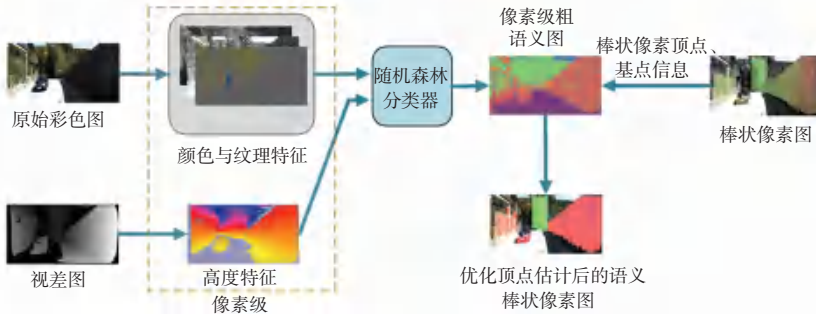


图1 算法框架图

Fig. 1 Framework of the algorithm

$$I = \log(G) - \alpha \log(B) - (1 - \alpha) \log(R), \quad (1)$$

其中, α 的值须根据每个相机传感器的特点确定。其证明与直接使用 RGB 图相比,该方法具有更强的鲁棒性。图 2(b)为其可视化结果。

RG 色度: 对每个像素的 RGB 值进行归一化处理。此转换方法减弱了阴影等强度变化,加强了植被和天空的颜色。图 2(c)展示了经过这种标准化处理后的示例。

$$r = \frac{R}{R + G + B}, \quad g = \frac{G}{R + G + B}. \quad (2)$$

光强: CIE Lab 图相比于 RGB 图更加符合人类的视觉特点,选用其中的 L 分量来编码一些在日光下具有典型光强等级的类别。

1.2.2 纹理特征

Gabor 滤波器^[9]由一组基本滤波器线性组合而成,简单哺乳动物大脑视觉皮层的细胞可以通过 Gabor 函数进行建模。Gabor 基本滤波器 $g_{mn}(x, y)$ 由它的尺度和方向确定:

$$\begin{aligned} g_{mn}(x, y) &= a^{-2m} g(x', y'), \quad a \geq 1, \\ \begin{pmatrix} x' \\ y' \end{pmatrix} &= a^{-m} \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}, \\ \theta &= \frac{n\pi}{L}; \quad m = 0, 1, \dots, K-1; \quad n = 0, 1, \dots, L-1. \end{aligned} \quad (3)$$

位置信息融合至像素级语义图,用类别标签来优化每个棒状像素的顶点估计,然后以其内部像素的优势类来判定类别,得到语义棒状像素图。

1.2 像素级特征提取

1.2.1 颜色特征

为了去除图像的绝对强度,只保留相对的颜色信息,对 RGB 图像应用以下两个转换:

光照不变性: Ratnasingham 等^[8]编码了特定日光光谱的色度。在光谱为黑体光谱的前提下,定义了从 RGB 到单通道光照不变图像的映射。

其中: $g(x', y')$ 为二维 Gabor 函数, a 为一个固定的尺度因子, m 是尺度参数, n 是方向参数, K 是总的尺度数目, L 是总的方向数目。

选用 4 个方向参数 $n(n = 0^\circ, 45^\circ, 90^\circ, 135^\circ)$ 和 3 个尺度参数 $m(m = 3, 4, 5)$, 采集窗口尺寸设为 5×5 个像素,应用 Gabor 滤波器至彩色图 YCbCr 颜色空间的明度分量 Y 上,取得到的 12 个子带系数中绝对值最大的系数值作为相应像素的纹理特征,结果见图 2(d)。

1.2.3 高度特征

几何高度: 将像素点的图像坐标 (u, v) 转换为世界坐标 (X_w, Y_w, Z_w) ^[10], 并选用其中的 Y_w 分量作为像素距离道路面的高度特征。

$$\begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} = \begin{pmatrix} \cos \varphi_p & 0 & 0 \\ \sin \varphi_p & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X_c \\ Y_c \\ Z_c \end{pmatrix} \quad (4)$$

φ_p 为双目相机的安装俯仰角, $(C_x, C_y, C_z)^T$ 为双目相机在世界坐标系下的三维坐标。 $(X_c, Y_c, Z_c, 1)^T$ 为相机坐标系下的齐次坐标。图像坐标到相机坐标的投影关系, Z_c 为目标至相机的距离,与视差 d 成反比,满足 $Z_c = f_x b / d$ 。其余参数与双目相机的内部结构有关,称为相机内部参数,必须通过相机标定才能得到。 $f_x = f / d_x$ 与 $f_y = f / d_y$ 为经过像素尺寸归一化后的相机焦距, (u_0, v_0) 为图像坐标原点:

$$\begin{pmatrix} \frac{Z_c}{f_x} & 0 & 0 \\ \frac{Z_c}{f_y} & -\frac{Z_c}{f_y} & 0 \\ 0 & 0 & Z_c \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} \frac{Z_c}{f_x} & 0 & 0 \\ -\frac{Z_c}{f_y} & 0 & 0 \\ 0 & 0 & Z_c \\ 0 & 0 & 1 \end{pmatrix} \quad (5)$$

特征提取结果如图 2(e)。

像素位置: 选用像素距道路平面的垂直高度作为高度特征的补充, 道路平面由 U 视差探测的可行驶区域^[2]确定。检测到的可行驶区域如图 2(f)。

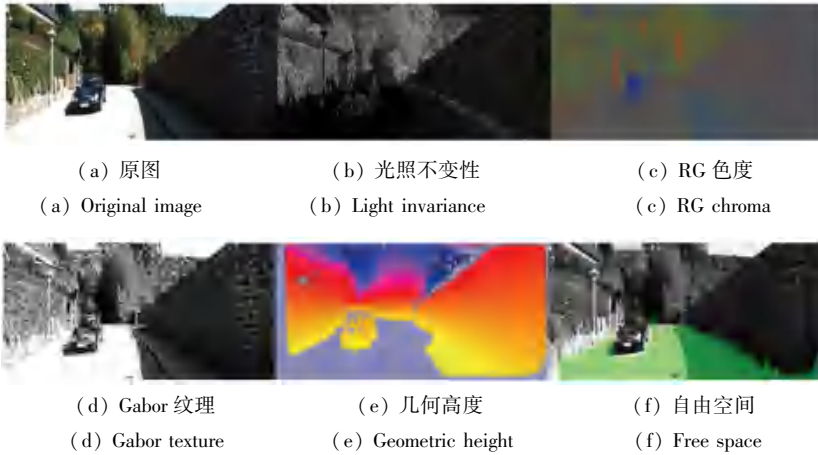


图 2 部分特征通道可视化

Fig. 2 Visualization of some feature channels

1.4 分类方法

所用算法为基于 CART 的随机森林分类器^[11]。节点处选取分裂特征的准则为 Gini 指数。其作为熵的一阶近似, 值越小表明划分越正确。 p_k 为某一节点的样本集 D 在按某一特征的某个特征值划分后, 类别 k 的样本所占的比率, K 为总的类别数:

$$Gini(D) = \sum_{k=1}^K p_k(1 - p_k) = 1 - \sum_{k=1}^K p_k^2. \quad (6)$$

具体训练过程如下:

- (1) 采用随机且有放回的方式(Boostrap^[9])从总训练集 S 中抽取总数 2/3 左右的样本, 生成子集 S_i 。
- (2) 使用子集 S_i 生成决策树:
 - ① 随机从总特征集 M 中选取 n 个特征 ($n \leq M$);
 - ② 每个节点处, 对每个特征的数值型取值采用“二分法”划分样本。依据式公式(6), 从 n 个特征中选最优分裂特征以及其最优的划分值;
 - ③ 以“完全分裂”方式生成决策树。
- (3) 重复(2)步骤使得所有样本均到达叶子节点或子样本集不能再分裂为止。
- (4) 重复(1)~(3)建立预设数量的决策树, 形成随机森林。

1.5 棒状像素拓展

以像素级语义标签作为输入, 对原始 Stixel^[2]表达进行拓展。单纯依据深度信息进行 Stixel 的估计

是具有局限性的, 在一些视野较宽的道路场景中, 稍远区域的障碍物与天空由于纹理信息缺失, 双目图像立体匹配会出现错误, 此时视差信息变得不可靠, 前后景无法有效区分, 导致 Stixel 的高度估计出现明显的失真。本文在文献[2]算法的基础上引入语义信息来更好的进行高度估计。

U 视差图 I_1 中的一点 $M^*(u, d, n)$ 在原始视差图 I_2 中有多个对应点 $M^*_1(u_1, v_1, d_1), M^*_2(u_2, v_2, d_2), \dots, M^*_s(u_s, v_s, d_s)$ 。在原方法的投影条件最后加入语义信息约束:

$$\begin{cases} \uparrow u_1 = u_2 = \dots = u_s = u, \\ \uparrow \varphi(u, \Delta(u, v_i)) = n \neq 0, \\ \uparrow d_1 = d_2 = \dots = d_s = d, \\ \uparrow s = n, \\ \uparrow L(u_i, v_i) = C_{obstacle}, i \in [1, s]. \end{cases} \quad (7)$$

$L(u_i, v_i)$ 为语义标签图 L 对应位置 (u_i, v_i) 的语义类别, $C_{obstacle}$ 为障碍物, 即语义标签中的障碍物大类与植物。通过在 I_2 的第 p 列搜索具有视差为 b 的点, 并借助语义信息排除明显属于非障碍物的对应点, 即可得到全部前景障碍物相关的图像坐标点集。

2 实验结果与分析

选用 KITTI 数据集^[12]进行算法评估, 实验硬件配置为 Intel Xeon(R) Silver 4110, 实验软件环境为

VS2013+Python3.0。像素级 RF 分类器的训练与测试选用 KITTI 数据集中的 pixel-level semantics 序列,此序列提供了带有像素级语义标签的测试集与训练集。对于光照不变性特征,设置 $\alpha = 0.48$ 。

像素级分类结果采用交并比评价指标:

$$IoU = \frac{TP}{TP + FP + FN} \quad (8)$$

其中, TP (True Positive) 为将正例判断为正例的样本数, FP (False Positive) 为将反例判断为正例的样本数, FN (False Negative) 为将正例判断为反例的样本数。

2.1 RF 分类器的参数确定与像素级分类结果

RF 分类器执行时间与分类性能取决于两个主要参数:树棵数与树深度。从像素级训练集中随机选取 10% 的平衡样本作为调节参数的验证集。实验选用所有特征通道,针对所有类别,在保持其中一个参数不变,逐渐变动另一个参数的策略下进行参数选取。当使用 35 棵数、树深 25 时,分类性能达到一个稳定水平。

对于像素级分类器,为了定量分析每个特征通道对总体分类性能的影响,首先使用所有特征通道来测试模型的分类性能,由于总体像素过多,此处将分类结果化为占对应实际类别总数的百分比,得到了表 1 的混淆矩阵,其中使用了交并作为评价指标。从表中可见,有 7.1% 属于障碍物类别的像素被误分为了天空,这是因为有些建筑或者白色车辆的光照具有高饱和度,导致错误分类。有些场景下部分植物区域与障碍物区域的阴影较深,导致分类器无法进行有效区分。

表 1 使用了所有特征通道的像素级分类混淆矩阵 (%)

Tab. 1 Confusion matrix using all pixel-level feature channels

		预测		
		天空	障碍物	植物
实际	天空	98.2	1.5	0.3
	障碍物	7.1	78.8	14.1
	植物	0.7	18.8	81.5
	IoU	88.2	71.1	79.0

采取每次只排除一个特征或只留下一个特征的策略来测试特征的显著性。从表 2 可以看出一旦只剔除像素的几何高度特征后,相比剔除其他的特征,平均 IoU 下降最明显,即说明几何高度特征明显提高了分类性能,是最显著的特征。关于颜色特征,由表 2 可见,只剔除 RG 色度通道相比只剔除光照不变性通道后下降的平均 IoU 更多,且只保留前者的平均 IoU 更高,即 RG 色度通道携带了更重要的分

类信息。以同样方式分析可知,纹理通道明显对分类贡献最低,将其归因于整合多个纹理通道时损失了一定的有效信息。

在 Stixel 估计方面,本文算法优化了 Stixel 算法^[2]的顶点估计。根据文献[2]对棒状像素顶点误差的定义,选用 KITTI 数据集^[10]中城市、校园、居民区、道路四个场景下的图像序列进行平均顶点误差的计算,两种方法的平均顶点误差对比如表 3 所示。由表 3 可知,道路场景下的棒状像素顶点估计精度有所提高。

表 2 像素级分类不同特征通道组合的混淆矩阵 (%)

Tab. 2 Confusion matrix of different feature channel combinations

Pixel-level classification results				
特征通道	天空	障碍物	植物	平均 IoU
保留所有通道				
	88.2	71.1	79.0	79.4
只剔除此通道				
纹理	88.0	70.7	78.2	79.0
光照不变性	87.3	70.5	78.8	78.9
光强	88.6	69.7	73.9	77.4
像素位置	84.1	69.2	77.0	76.8
RG 色度	83.2	68.7	76.5	76.1
几何高度	66.1	65.4	77.0	69.5
只保留此通道				
RG 色度	56.2	41.8	73.1	57.0
光照不变性	56.7	30.8	72.0	53.2
几何高度	51.2	41.5	31.3	41.3
光强	41.2	20.4	48.5	36.7
纹理	40.7	21.0	39.3	33.7
像素位置	20.9	14.9	28.3	21.4

表 3 拓展 Stixel 与原始 Stixel 的顶点估计误差比较

Tab. 3 Comparison of top-point estimation error between extended Stixel and original Stixel

场景类别	顶点误差 (像素)	
	本文算法	Stixel 算法[2]
城市	5.0	5.2
校园	6.7	6.8
居民区	6.3	6.6
道路	3.1	5.7

图 3 为本算法应用于 KITTI 中两个场景的实验结果。每个场景从左至右分别为原图、初始 Stixel 图(颜色表示深度,不代表类别)、像素级语义标签图(颜色代表类别,分为天空、植物、障碍物大类)、语义 Stixel 图(颜色代表类别,分为背景、植物、障碍物大类)。在上述图中,浅蓝色为天空,绿色为植物,土红色代表障碍物大类,紫色代表根据原 Stixel^[2]得到的自由空间,灰色为 Stixel 分类后确定的背景区域。

3 结束语

本文在前期棒状像素检测的基础上,进一步提出了一种能够实现棒状像素语义标注的方法。此算

法既充分利用了棒状像素的障碍物检测优势,又实现了在道路场景中对各类前景障碍物的语义理解,优化了原有的棒状像素估计。

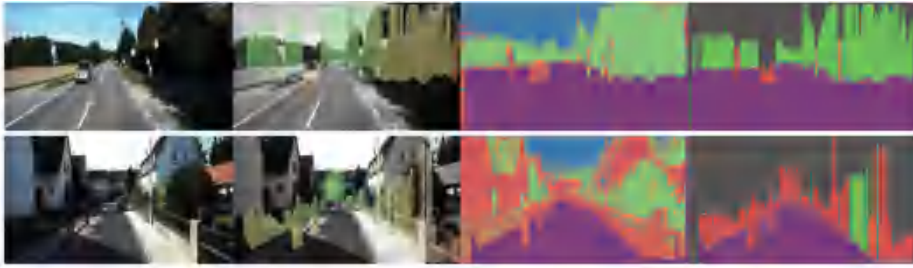


图3 两个典型交通场景的棒状像素语义分割结果

Fig. 3 Semantic segmentation of two typical traffic scenarios

参考文献

- [1] BADINO H, FRANKE U, PFEIFFER D. The stixel world – a compact medium level representation of the 3d – world [C]// Proceedings of Joint Pattern Recognition Symposium. Springer, Berlin, Heidelberg, 2009: 51–60.
- [2] CHEN L, HUANG Y, HU X, et al. Stixel-based Traffic Scene Representation Using U-disparity [C]// Proceedings of Journal of Physics: Conference Series. IOP Publishing, 2019, 1237 (3) : 032082.
- [3] CARREIRA J, CASEIRO R, BATISTA J, et al. Semantic segmentation with second – order pooling [C]// Proceedings of European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2012: 430–443.
- [4] Lubor Ladicky, Paul Sturgess, Chris Russell, Sunando Sengupta, Yalin Bastanlar, William Clocksin, Philip H. S. Torr. Joint Optimization for Object Class Segmentation and Dense Stereo Reconstruction [J]. International Journal of Computer Vision, 2012, 100(2):122 – 133.
- [5] CHENG H Y, WENG C C, CHEN Y Y. Vehicle detection in aerial surveillance using dynamic Bayesian networks [J]. IEEE transactions on image processing, 2011, 21(4): 2152–2159.
- [6] XIE L, TIAN Q, WANG M, et al. Spatial pooling of heterogeneous features for image classification [J]. IEEE Transactions on Image Processing, 2014, 23(5): 1994–2008.
- [7] SANBERG W P, DUBBELMAN G, de WITH P H N. Extending the stixel world with online self – supervised color modeling for road – versus – obstacle segmentation [C]// Proceedings of 17th International IEEE Conference on Intelligent Transportation Systems (ITSC). IEEE, 2014: 1400–1407.
- [8] RATNASINGAM S, COLLINS S. Study of the photodetector characteristics of a camera for color constancy in natural scenes [J]. JOSA A, 2010, 27(2): 286–294.
- [9] 盛建瓴,朱冬梅. 指纹识别预处理算法 [J]. 电脑学习, 2008(2): 12–13.
- [10] 陈磊,黄影平,胡兴,等. 用于智能车的棒状像素场景表达方法研究 [J]. 小型微型计算机系统, 2019, 40(12): 2698–2704.
- [11] 陈静,张艳新,姜媛媛. 融合多特征与随机森林的纹理图像分类方法 [J]. 传感器与微系统, 2019, 38(12): 58–61.
- [12] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? the kitti vision benchmark suite [C]. Proceedings of 2012 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2012: 3354–3361.
- [5] BRAY F, FERLAY J, SOERJOMATARAM I, et al. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: a cancer journal for clinicians, 2018, 68(6): 394–424.
- [6] 林燕,马蕾,布力布·吉力斯汉,等. 中医药辅助治疗结肠癌的应用研究进展 [J]. 中国药师, 2018, 21(9): 1637–1640.
- [7] CORTESC, VAPNIK V. Support – vector networks [J]. Machine learning, 1995, 20(3): 273–297.
- [8] 史双睿. 异质集成学习器在鸢尾花卉分类中的应用 [J]. 电子制作, 2019(2): 45–47, 79.
- [9] 阙红星,张璐瑶,董昌武. 一种 2 型糖尿病中医证型的舌图像识别方法 [J]. 中国生物医学工程学报, 2016, 35(6): 658–664.
- [10] 沈鑫. 中医辨证分型的糖尿病食疗系统的设计与实现 [J]. 电脑知识与技术, 2015, 11(14): 83–84.
- [11] 中华中医药学会. 中华中医药学会标准: 肿瘤中医诊疗指南 [M]. 北京: 中国中医药出版社, 2008: 26–29.
- [12] 李玲玲. 基于 B/S 中医营养食疗系统的设计 [J]. 软件工程师, 2015, 18(2): 12–13.

(上接第 136 页)

3 结束语

本文通过机器学习实现结直肠癌的中医智能辨证分型,提出结直肠癌食疗药膳系统的设计方案。描述了系统设计采用的关键技术和方法、系统的框架及其主要的功能设计,实现了中医辨证分型、食疗药膳推荐、健康膳食及生活方式指导等功能^[12]。

参考文献

- [1] 董国华,朱莉乐,李璟健,等. 粗糙集结合神经网络的哮喘辨证分型法 [J]. 计算机与数字工程, 2015, 43(9): 1622–1626.
- [2] MORPURGO R, MUSSI S. An intelligent diagnostic support system [J]. Expert Systems, 2001, 18(1): 43–58.
- [3] SELA R J, SIMONOFF J S. RE – EM trees: a data mining approach for longitudinal and clustered data [J]. Machine learning, 2012, 86(2): 169–207.
- [4] 周璐,李光庚,孙燕,等. 复合结构智能化辨证选方模型的构建 [J]. 世界中医药, 2018, 13(2): 479–483.