A New Better-Fit Decision Features Selection Method for C5.0 Decision Tree

Hong-wei YANG* and Sheng-jie SUN

College of Information Science & Technology, Hainan University, No. 58, Renmin Avenue, Haikou City, Hainan Province, 570228, P.R. China.

*Corresponding author

Keywords: C5.0 decision tree, Better-fit decision features selection (BFDFS), Rubber woods classification from high resolution remote sensing image.

Abstract. C5.0 decision tree method has a disadvantage that it’s difficult to select better-fit decision features, so a better-fit decision features selection (abbreviated as BFDFS) methods are proposed in this paper. The procedure of BFDFS is as follows: (1) all decision features are pre-processed and then integrated into one image file; (2) interesting regions of typical kinds of land objects are chosen; (3) values of all features in interesting regions are computed according the "3σ" theory; (4) all ratios of different objects features’ values are computed and sorted to obtain the better-fit decision features; (5) using the better-fit decision features to build C5.0 decision tree rules. The BFDFS is tested by the use of classifying the rubber woods from high resolution remote sensing images, experimental area is selected in Guangba Farm, Dongfang City, Hainan Island, China. The results indicate that the producer accuracy, user accuracy, total accuracy of rubber woods, and the Kappa coefficient are 88%, 91.67%, 92%, and 0.89, respectively. All four indices are better than the other classification methods, proving the feasibility and efficiency of the BFDFS method.

Introduction

Among the classification methods based on remote sensing images, decision tree (DT)[[1],[2],[3],[4]] (Kandrika et al., 2008; Yang et al., 2003; Punia et al., 2011; Sesnie et al., 2008) is a good and popular one, which conveniently adding new decision conditions, but it is difficult to select better-fit decision features. The most common way of choosing better-decision features of the DT method is based on users’ experiences and by the computation of the maximum information gain ratio. However, the information gain ratio cannot be achieved easily. Especially, in case of multiple features, it is more difficult because of extensive computation [[5],[6]](Galathiya, 2012; Quinlan, 1986). Therefore, the improved selection method of better decision features needs to be studied further.

In this paper, a method of better-fit decision features selection (BFDFS) has been presented. Following all features benefiting the classification of land objects have been collected and integrated into one image, the BFDFS method has been proposed to choose the better-fit decision features, and the fit value of decision features has been achieved also.

The structure of this paper is as follows: following the introduction, the study areas and data sources have been described in section 2. In section 3, an auto-selecting method for the decision features have been described. The experiment has been described in section 4, and the classification accuracy has also been evaluated. In section 5, conclusions have been drawn at length.

Experimental Area and data

In order to prove the efficiency of our method, an experiment of selecting better fit decision tree features for C5.0 was done, which has been used to classify the rubber woods from high resolution images, the experimental area and data will be described as follows.
Study Area

Guangba Farm, Dongfang City, was selected as the study area. It is located to the southwest of Hainan Island, China (figure 1). Rubber trees are the major vegetation type in the area; grasses, shrubs, bare area, and surface water also characterize the land-cover type of the study area. A typical area was selected as shown in Scene B of figure 1.

Data

1) High-resolution remote sensing images

The high-resolution remote sensing images of QuickBird, taken on July 6, 2012, were used in this experiment, a total of 5 spectral images were included; the spatial resolution of the panchromatic image is 0.6 m and that of the multi-spectral regions (red, green, blue, and near-infrared) is 2.4 m.

2) Digital Elevation Model (DEM)

The Advanced Space-borne Thermal Emission and Reflection Radiometer Global Digital Elevation Model (ASTER GDEM), which has a resolution of 30 m, was used in our experiment.

3) Ground truth data

Some ground truth data were derived by GPS between May 2 and May 6, 2015, when the derived time was different from the time taken. However, the derived time and the time taken were the same in a different year, and the vegetation had a similar growing status during the same season. Hence, the ground truth data can be used as the reference data for geometry correction and classification accuracy evaluation of the remote sensing images.

Method

Introduction of C5.0 Decision Tree

C5.0 is a traditional, efficient, and widely used method. Its procedure has been described briefly as follows[7] (Quiland, 1993).
For a sample set called $S$, the target variable $C$ has $k$ types, $fq(C_i, S)$ is the sample number of $S \in S_i$, $|S|$ is the sample number of sample set $S$. We can define the information entropy of set $S$ as

$$\text{info}(s) = -\sum_{i=1}^{k} \left( \frac{fq(C_i, S)}{|S|} \right) \times \log_2 \left( \frac{fq(C_i, S)}{|S|} \right)$$

(1)

For an attribute variable $T$, there are $n$ class types of $S$, the condition entropy of the attribute $T$ can be defined as:

$$\text{Cinfo}(T) = -\sum_{i=1}^{n} \left( \frac{|T_i|}{|T|} \times \text{info}(T_i) \right)$$

(2)

The information gains of $T$ can be expressed as

$$\text{Gain}(T) = \text{info}(s) - \text{Cinfo}(T)$$

(3)

Hence, the ratio of information gains can be expressed as

$$\text{Gainrto}(T) = \frac{\text{Gain}(T)}{\text{info}(s)}$$

(4)

When we use C5.0 decision tree algorithm, the maximum of $\text{Gainrto}(T)$ is the judgment index of the decision features being chosen.

Though the ideal decision tree method is easy, selecting better-fit decision features, especially multiple decision features, is extremely difficult. This is because of two main reasons: 1) we have to scan all features sets when we build a decision tree. This procedure will make the algorithm of C5.0 decision tree inefficient. 2) The C5.0 decision tree method is fit for the data sets that can reside in the memory of computer. when the training sets’ volume is bigger than the computer’s, it is difficult to use them[5][6] (Galathiya, 2012; Quinlan, 1986). In addition, the value of better-fit decision features is difficult to be evaluated. Obtaining a suitable value of better decision features is another key problem.

**Better-Fit Decision Features Selection (BFDFS) Method**

The Sketch framework of BFDFS method is shown as figure 2, can be divided into five steps to be described, which is shown as follows.
Step one: integrating multi-features into one image file.

All collected features should be integrated together as required for the decision tree method. The process of integrating the multi-features includes standardizing the spatial resolution of all the images, correcting the images according to the same standard, and finally, cutting all the images. Then, all the decision features are stacked into one file, which has the same resolution, coordinate system, and location at last. In our study, we unified all the spectral feature images, vegetation index images, and spatial feature images into the World Geodetic System-1984 coordinate system, interpolated the values in the DEM image according to the spatial resolution of the remote sensing image, corrected the DEM image according to the remote sensing image, cut all the images according to the same area, and stacked all the images into one file using ENVI software.

Step two: selecting interesting areas of every land objects

Users classify the objects into key representative classes based on the users’ experience after integrating multi-features of different images, and select locations of different representative objects for supervised classification (our classification method is a supervised classification method), all are prepared well for the next computing.

Step three: computing values of key representative objects

Users compute the maximum, minimum, mean, and standard deviation of the different representative objects. In addition, users compute the "3σ" value and get the fit intervals of representative objects of each location. The computation theory and process of "3σ" value is as follows.

If X is a normal distribution, which can be expressed as $X \sim N(\mu, \sigma^2)$, X satisfies the rule expressed in equation (15).

$$P\{\mu - 3\sigma \leq X \leq \mu + 3\sigma\} = 0.9974$$

(5)

where P is the probability of X, $\mu$ is the mean, and $\sigma$ is the standard deviation.
First, we need to compute the “$3\sigma$” (3 multiply $\sigma$) value of every representative object, and then compare the minimum value with the $\mu - 3\sigma$ value of the representative objects of each location to get their maximum value. The maximum value was used as the lower limit of the value of the representative objects. We also compared the maximum value with the $\mu + 3\sigma$ value of the representative objects of the area of interest to obtain the minimum value, which is used as the upper limit of the value of the representative objects. Hence, the value of every representative object of a location is between its lower and upper limit values. This is also the better-fit intervals of the representative objects of the area of interest.

Step four: Comparing the intervals between two different objects and getting the better-fit decision features according to the reality condition.

All the procedures of the decision tree method help the users distinguish between two objects. Distinguishing two objects is a key and difficult job in all procedures. Hence, distinguishing two land objects is prominently described as follows.

Figure 3. Detailed frameworks of BFDFS method used for distinguishing between object A and B.

We suppose two land objects named A and B, as described in figure 3, are representative objects in an area of interest, Max is the maximum of the object statistical data, Min is the minimum of the object statistical data, Mean is the mean value of the object statistical data (it is also expressed by $\mu$), Stdev is the standard deviation of the object statistical data (it is also expressed by $\sigma$), $U_a$ is the upper limit of object A, $U_b$ is the upper limit of object B. We supposed $U_a \geq U_b$ during the procedure. If $U_a \leq U_b$, we adjusted the order of object A and B in the same procedure.

The distinguishing procedures between the land objects A and B are as follows:

We selected the area of interest, A and B, which is based on the experiences of users. Hence, users should pay more attention to this procedure, which directly affects the accuracy of the final result. The area around A and B contains all typical features, and all these typical features are prepared well for the decision tree features.

We computed the Max, Min, Mean, and Stdev of every feature of A and B. Then, we systematically computed the “$3\sigma$” values of the corresponding objects. After comparing the $\mu - 3\sigma$ value with the Min of the corresponding object, we selected the maximum value as the lower limit of the corresponding object (the lower limit of A is called $L_a$, the lower limit of B is $L_b$). We compared the $\mu + 3\sigma$ value with the Max of the corresponding objects, selected the minimum value as the upper limit of the corresponding object (the upper limit value of A is called $U_a$, the upper limit of B is called $U_b$). For example, the procedure of obtaining $L_a$ and $U_a$ of the green spectrum is as follows. We computed the Min value of the green spectrum of A in its area of interest, and then computed the $\mu - 3\sigma$ value of green spectrum of A. After comparing the Min with the $\mu - 3\sigma$ value, we obtained the maximum value as $L_a$. We also computed the Max value of the green spectrum of A, and then computed the $\mu + 3\sigma$ value of the green spectrum of A. After comparing the Max with the $\mu + 3\sigma$ value, we obtained their minimum as $U_a$. 

103
We derived equation $R_{ab} = (U_a - L_b)/(U_b - L_a)$, computed the $R_{ab}$ value of all the features of A and B, where the features of A and B are the same. For example, the procedure of obtaining $R_{ab}$ of the green spectrum between objects A and B is as follows: We determined the $L_a$, $U_a$, $L_b$, and $U_b$ values of the green spectrum of A and B, $R_{ab}(\text{green}) = (U_a(\text{green}) - L_b(\text{green}))/ (U_b(\text{green}) - L_a(\text{green}))$, where $R_{ab}(\text{green})$ is the $R_{ab}$ value of the green spectrum, $U_a(\text{green})$ is the $U_a$ value of the green spectrum, $L_b(\text{green})$ is the $L_b$ value of the green spectrum, $U_b(\text{green})$ is the $U_b$ value of the green spectrum, and $L_a(\text{green})$ is the $L_a$ value of the green spectrum.

After determining the $R_{ab}$ and $R_{ba}$ values of all the features, we sorted the $R_{ab}$ and $R_{ba}$ values in ascending order, selected the front of $R_{ab}$ and $R_{ba}$ (which is also smaller than other feature values) as an option (the count of selected $R_{ab}$ and $R_{ba}$ should be defined according the users’ reality needs). The features corresponding to the selected $R_{ab}$ and $R_{ba}$ are the better-fit decision features of the C5.0 decision tree algorithm if the features’ intervals satisfy case 1 or 2. Here, cases 1, 2, and 3 have been described as follows.

![Figure 4. Three distribution cases of interval between object A and B.](image)

The relationship of the interval value of A and B can be grouped into three types, and their distributions are described in figure 4. The first is case 1, in which the intervals of A and B have a separation relationship. In case 2, the intervals of A and B have an intersection relationship. In case 3, the intervals of A and B have an inclusive relationship.

We can get the corresponding features of $R_{ab}$, and the relationship of the features’ interval value will satisfy one of the cases noted in figure 4. When the relationship agrees with case 1, the feature is the better-fit decision feature of the decision tree, and the decision feature’s value of A is between $L_a$ and $U_a$, and the decision feature’s value of B is between $L_b$ and $U_b$. When the relationship satisfies case 2, the feature is also the better-fit decision feature of the decision tree, and the decision feature’s value of object A is between $L_a$ and $G_o \cdot \text{Stddev}(A)$, and the decision feature’s value of object B is between $G_o \cdot \text{Stddev}(A)$ and $U_b$, where $G_o = (U_a - L_b)/(\text{Stddev}(A) + \text{Stddev}(B))$. When the relationship satisfies case 3, the feature cannot be the better fit decision feature of the decision tree. In this condition, object A should be divided into two types of new representative objects (new objects A and B). Then, we redid the entire procedure described above until the relationship of the features interval value satisfies case 1 or 2. After distinguishing the two objects, we unified the interval value of the new objects A and B.

Step five: users use the better-fit decision features to build the decision rules for the C5.0 decision tree.

After obtaining the better decision features and their intervals, we can build the decision tree of rubber woods classification based on C5.0 decision tree method.

**Experiment and Results**

In this experiment, we fuse the red, NIR and green of the remote sensing image, and the DEM data that spatial resolution is 30 m. After being transformed, all the images reserve the land information well while the resolution attains 0.6 m, and all data are in the same coordinate system (GS-1984) in the end. Then we cut the image to satisfy the experimental area, and select NAVI, SAVI (the L value is 0.5), RVI, red, NIR, green, blue to obtain the texture feature data. When all the data (spectral feature, texture feature, and DEM) were collected, we integrated all the data by stacking all the feature data into one file.

The BFDFS method is displayed using the sample of distinguishing rubber woods from shrubs and grass, because the procedure of distinguishing rubber woods from shrubs and grass is extremely
After selecting interesting areas of every key representative land objects, we compute the Min, Max, Mean and Stdev of every key representative land objects (are displayed in table 1), and use the "3σ" theory to obtain the results of fit values of rubber woods, which are displayed in table 2.

<table>
<thead>
<tr>
<th>Base data</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>Stdev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubber(BLUE)</td>
<td>0.000</td>
<td>255.000</td>
<td>127.648</td>
<td>55.359</td>
</tr>
<tr>
<td>Rubber(contrast(GREEN,19))</td>
<td>32.000</td>
<td>255.000</td>
<td>122.970</td>
<td>32.373</td>
</tr>
<tr>
<td>Rubber(DEM)</td>
<td>6.000</td>
<td>70.000</td>
<td>20.914</td>
<td>16.393</td>
</tr>
<tr>
<td>Rubber(GREEN)</td>
<td>0.000</td>
<td>255.000</td>
<td>118.700</td>
<td>51.483</td>
</tr>
<tr>
<td>Rubber(homogeneity(RVI,9))</td>
<td>0.000</td>
<td>255.000</td>
<td>181.793</td>
<td>62.211</td>
</tr>
<tr>
<td>Rubber(mean(BLUE,15))</td>
<td>30.000</td>
<td>195.000</td>
<td>103.342</td>
<td>20.757</td>
</tr>
<tr>
<td>Rubber(mean(GREEN,19))</td>
<td>44.000</td>
<td>145.000</td>
<td>88.356</td>
<td>17.153</td>
</tr>
<tr>
<td>Rubber(mean(NIR,19))</td>
<td>101.000</td>
<td>215.000</td>
<td>151.693</td>
<td>23.186</td>
</tr>
<tr>
<td>Rubber(mean(RED,19))</td>
<td>31.000</td>
<td>100.000</td>
<td>61.010</td>
<td>11.201</td>
</tr>
<tr>
<td>Rubber(NDVI)</td>
<td>-0.081</td>
<td>1.000</td>
<td>0.160</td>
<td>0.155</td>
</tr>
<tr>
<td>Rubber(NIR)</td>
<td>3.000</td>
<td>255.000</td>
<td>144.070</td>
<td>42.809</td>
</tr>
<tr>
<td>Rubber(RED)</td>
<td>0.850</td>
<td>82.000</td>
<td>1.538</td>
<td>2.130</td>
</tr>
<tr>
<td>Rubber(RVI)</td>
<td>-0.122</td>
<td>1.493</td>
<td>0.239</td>
<td>0.231</td>
</tr>
<tr>
<td>Rubber(SAVI)</td>
<td>11.000</td>
<td>228.000</td>
<td>85.256</td>
<td>30.979</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Base data</th>
<th>3σ Low Bound</th>
<th>3σ Upper Bound</th>
<th>Fit low bound</th>
<th>Fit upper bound</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubber(BLUE)</td>
<td>-38.429</td>
<td>293.725</td>
<td>0.000</td>
<td>255.000</td>
</tr>
<tr>
<td>Rubber(contrast(GREEN,19))</td>
<td>25.851</td>
<td>220.088</td>
<td>32.000</td>
<td>220.088</td>
</tr>
<tr>
<td>Rubber(DEM)</td>
<td>-28.265</td>
<td>70.093</td>
<td>6.000</td>
<td>70.000</td>
</tr>
<tr>
<td>Rubber(GREEN)</td>
<td>-35.749</td>
<td>273.149</td>
<td>0.000</td>
<td>255.000</td>
</tr>
<tr>
<td>Rubber(homogeneity(RVI,9))</td>
<td>-4.840</td>
<td>368.426</td>
<td>0.000</td>
<td>255.000</td>
</tr>
<tr>
<td>Rubber(mean(BLUE,15))</td>
<td>41.072</td>
<td>165.612</td>
<td>41.072</td>
<td>165.612</td>
</tr>
<tr>
<td>Rubber(mean(GREEN,19))</td>
<td>36.897</td>
<td>139.816</td>
<td>44.000</td>
<td>139.816</td>
</tr>
<tr>
<td>Rubber(mean(NIR,19))</td>
<td>82.135</td>
<td>221.251</td>
<td>101.000</td>
<td>215.000</td>
</tr>
<tr>
<td>Rubber(mean(RED,19))</td>
<td>27.407</td>
<td>94.613</td>
<td>31.000</td>
<td>94.613</td>
</tr>
<tr>
<td>Rubber(NDVI)</td>
<td>-0.305</td>
<td>0.624</td>
<td>-0.081</td>
<td>0.624</td>
</tr>
<tr>
<td>Rubber(NIR)</td>
<td>15.643</td>
<td>272.497</td>
<td>15.643</td>
<td>255.000</td>
</tr>
<tr>
<td>Rubber(RED)</td>
<td>-34.213</td>
<td>260.195</td>
<td>0.000</td>
<td>255.000</td>
</tr>
<tr>
<td>Rubber(RVI)</td>
<td>-4.852</td>
<td>7.929</td>
<td>0.850</td>
<td>7.929</td>
</tr>
<tr>
<td>Rubber(SAVI)</td>
<td>-0.454</td>
<td>0.930</td>
<td>-0.122</td>
<td>0.930</td>
</tr>
<tr>
<td>Rubber(variance(RED,19))</td>
<td>-7.681</td>
<td>178.192</td>
<td>11.000</td>
<td>178.192</td>
</tr>
</tbody>
</table>

(105)
Grass(mean(BLUE,15))     0.000  255.000  Rubber(mean(BLUE,15))   41.072  165.612  0.838933  0.649459
Grass(mean(GREEN,19))   -0.900  1.221  Rubber(mean(GREEN,19))  44.000  139.816  -0.30401  1
Grass(mean(NIR,19))     0.000  8.230  Rubber(mean(NIR,19))  101.000  215.000  -0.43149  1
Grass(mean(RED,19))     -0.613  0.830  Rubber(mean(RED,19))  31.000  94.613  -0.29683  1
Grass(NIR)              3.495  157.000  Rubber(NIR)         -0.081  0.624  1  -0.01828
Grass(RED)              6.000  255.000  Rubber(RED)         15.643  255.000  1  0.929412
Grass(RVI)              1.895  226.500  Rubber(RVI)        0.850  7.929  1  0.026741
Grass(SAVI)             0.000  255.000  Rubber(SAVI)      -0.122  0.930  1  0.003645
Grass(variance(RED,19)) 8.897  255.000  Rubber(variance(RED,19))  11.000  178.192  0.991455  0.687903

Table 4. Comparison of the partly fit spectral and spatial features of shrubs and rubber woods.

According the BFDFS results, the experimental status satisfies case 3. Therefore, we divided the shrub and grass land objects into shrub land object and grass land object, and reselected the area of interest. Then, we computed and obtained the results of the fit values of the shrub land object and grass land object as in tables 3 and 4.

After building the decision tree (figure 5), we can choose the better-fit decision features and their value using the BFDFS method, and the decision tree rules of the land object classification are described in table 5.

![Decision Tree Model](image)

Figure 5. Classifying decision tree model of rubber woods.

After executing the decision tree rules, we obtained the classified image and used the majority analyst method for de-noising. The kernel of the majority analyst method is the same as the rubber
woods’ texture computation window size (here, 19 pixels). The post-processed image is displayed in figure 6.

![Figure 6. Classified result image of rubber woods.](image)

Ground truth ROIs (regions of interest) are generated randomly from the QuickBird high-resolution remote sensing images by visual interpretation and GCPs. By comparing the classification results with the ground truth ROIs and GCPs, the error matrix of rubber woods classification is prepared as displayed in table 6. From table 6, we can see that the total accuracy is 92.00%, the Kappa efficient is 0.89, the user accuracy of rubber woods is 91.67%, and the producer accuracy of rubber woods is 88.00%.

<table>
<thead>
<tr>
<th>Reference image</th>
<th>Evaluated image</th>
<th>Producer accuracy/ %</th>
<th>User accuracy/ %</th>
<th>Total accuracy</th>
<th>Producer accuracy/ %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rubber woods</td>
<td>88</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Shrub and grass</td>
<td>8</td>
<td>87</td>
<td>5</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Bare area</td>
<td>0</td>
<td>6</td>
<td>94</td>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>Water</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>99</td>
<td>100</td>
</tr>
<tr>
<td>Total samples</td>
<td>96</td>
<td>105</td>
<td>100</td>
<td>99</td>
<td>400</td>
</tr>
<tr>
<td>User accuracy</td>
<td>91.67</td>
<td>82.86</td>
<td>94.00</td>
<td>100</td>
<td>--</td>
</tr>
</tbody>
</table>

Many other methods are used to extract rubber woods. All the classified images and accuracy evaluation results are displayed in table 7.
Table 7 Classified results and accuracy evaluation of the other method.

<table>
<thead>
<tr>
<th>Method</th>
<th>Result image of classification</th>
<th>Evaluation of classification</th>
</tr>
</thead>
</table>
| Object oriented    | ![Image](image1.png)           | Total accuracy: 86.25%  
Kappa coefficient: 0.82  
User accuracy of rubber: 78.49%  
Producer accuracy of rubber: 73.00% |
| Neural network     | ![Image](image2.png)           | Total accuracy: 74.50%  
Kappa coefficient: 0.66  
User accuracy of rubber: 68.29%  
Producer accuracy of rubber: 56.00% |
| Support vector     | ![Image](image3.png)           | Total accuracy: 82.50%  
Kappa coefficient: 0.77  
User accuracy of rubber: 83.00%  
Producer accuracy of rubber: 65.35% |

In table 7, 🟢 denotes rubber woods, 🟣 denotes shrubs and grass, 🟡 denotes water, 🟠 denotes bare land.

In this paper, we used the classification accuracy value as the assessment index to evaluate the feasibility of the BFDFS method. According to the theory of the C5.0 decision tree method (described in 3.1 section), the maximum of Gainrto(T) is used to assess the efficiency of the BFDFS method. When the data set is discrete, Gainrto(T) is easy to be computed. However, when the data set is continuous, it is difficult to compute. In our experiment, the values are continuous. Hence, using the maximum Gainrto(T) to evaluate the efficiency of the BFDFS method is not advisable. The classification accuracy of the BFDFS method is proportional to the value of Gainrto(T), which implies a higher accuracy and a bigger value of Gainrto(T). Thus, using the classification accuracy value to evaluate the efficiency of the BFDFS method will be helpful. In addition, the running time of classification in our experiment is less than one minute, which also proves that the BFDFS method is efficient.

Conclusions

The proposed BFDFS is a feasible and efficient method for selecting better-fit decision features. The traditional method of decision features selection is based on the experiences of users, while the BFDFS method is based on computation. Hence, the BFDFS method has a sounder theoretical basis. Furthermore, the classification accuracy of rubber woods proves its feasibility, and the time taken for the classification procedure is less than one minute, which also proves its efficiency.

On the other hand, the procedure of automatic building C5.0 decision tree model for rubber woods classification needs to be studied further. The C5.0 decision tree model of rubber woods classification is prepared based on the experiences of users. Hence, the procedure of automatically building the C5.0 decision tree model also needs to be studied further.
Acknowledgements

This paper was substantially supported by the Natural Science Foundation of Hainan province, China (Project no. 614225).

References


