



# Information classification algorithm based on decision tree optimization

Hongbin Wang<sup>1</sup> · Tong Wang<sup>2</sup> · Yucai Zhou<sup>3</sup> · Lianke Zhou<sup>1</sup> · Huafeng Li<sup>1</sup>

Received: 28 November 2017 / Revised: 9 January 2018 / Accepted: 2 February 2018 / Published online: 26 February 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

With the rapid development of information technology, the efficiency of information management has drawn increasing importance with its broadening application. Hence, a new information classification algorithm has proposed in this paper so as to improve information management of the limited resources by reducing its complexity. However, ID3 algorithm is a classical and imprecise algorithm in data mining, because traditional ID3 algorithm selects the attribute that has the maximum information gain according to the data set as that of the split node. Then the data subset is further divided according to the number of attribute values, and the information gain of each subset is calculated recursively. Decision Tree Optimization Ratio is the core approach in this algorithm, whose basic ideas have been introduced and analysed, proving to be more complex. Therefore, the authors propose a relatively precise RLBOR algorithm which takes the number of nodes in the decision tree model into consideration. The experiment show more precise of RLBOR algorithm.

**Keywords** ID3 algorithm · Decision tree model · Decision tree optimization ratio · Leaf nodes

## 1 Introduction

Information management includes extracting information and knowledge from data, managing data warehouses, mining and visualizing data [1]. Much work has been carried out on the information management in recent years. As a result, various information management technologies are applicable for managing data that needs to be examined now. For example, data has to be visualized so that one can better understand them. Besides, visualization tools for the data should also be developed. At the same time, aggregating the data as well as building repositories and warehouses may also be needed. However, much of them may

be transient data, therefore we also need to determine which data to be stored and which data to be discarded. Apart from that, those data may also have to be processed in real time so that some of the data may be stored, possibly warehoused and analyzed for conducting researches and predicting trends. That is, the data emanating from surveillance cameras have to be processed within a certain time. The data may also be warehoused so that one can analyze them later.

Data mining is becoming an important area recently [2], which requires a close examination of data mining techniques such as association rule mining, clustering, and link analysis for data. At the same time, we need to manage data in real time as stressed in the above. Therefore, the data may also need to be mined in real time, which not only means building models ahead of time so that we can analyze the data in real time, but also building them in real time for the future reference. In other words, the models have to be flexible and dynamic as well, which poses a major challenge. Furthermore, training examples may also be greedily needed to build such models. For example, the data emanating should be mined and detected so that they may possibly prevent terrorist attacks, which means that we need practising examples to train the neural networks,

---

✉ Tong Wang  
jinxinhou@21cn.com

<sup>1</sup> College of Computer Science and Technology, Harbin Engineering University, Harbin, Heilongjiang Province, China

<sup>2</sup> Information and Communication Engineering College, Harbin Engineering University, Harbin, Heilongjiang Province, China

<sup>3</sup> School of Energy and Power, Changsha University of Science and Technology, Changsha 410076, China

classifiers, and other tools, thus they can recognize it in real time when a potential anomaly occurs. Since data mining is a fairly new research area, a research program for data management and data mining may be wanted.

Information classification is the foundation of information management, data visualizing and data mining. The decision tree method has been widely researched and applied for its systematic clearness and wide availability. ID3 algorithm [3], one of the most important method in decision tree learning, is also a classical classification algorithm [4–8]. The improvement of ID3 algorithm is also a hot research topic. This algorithm can directly reflect characteristics of the data besides to be easily understood. Moreover, the decision tree model is efficient in classification and prediction. Therefore, the decision rule can be conveniently drawn. At present, many experts and scholars have put forward a number of algorithms using decision tree to categorize large-scale data, of which ID3 algorithm was advanced by Quinlan in 1986, was the most typical one. For improving the quality of vocational education evaluation validity, the Ref. [9] introduces the research progress of teaching evaluation, uses data mining technology to put forward a feasible model of higher vocational teaching quality evaluation, and constructs a decision tree which has practical significance. By the method used, the traditional algorithm has improved the decision tree to improve the teaching contents to the location of the second master node, reflecting the characteristics of vocational teaching practice. A new algorithm based on attribute similarity for multivalued bias of ID3 algorithm was proposed in the Ref. [10].

On one hand, some scholars consider fuzziness [11–15]. For example, the Ref. [9] introduces a fuzzy consciousness function and further provides generalized fuzzy partition entropy for the attribute-selecting heuristic of a fuzzy decision tree. The reference subsequently proposes a generalized fuzzy partition entropy-based fuzzy ID3 algorithm (abbreviated as GFID3) that can support decision making and analyze the performance of the GFID3 through several case-based examples.

On the other hand, some studies focus on the framework of information classification [16–18]. Ref. [16] presents a framework for classification of EMG signals using principal component analysis (MSPCA) for de-noising, discrete wavelet transform (DWT) for feature extraction and decision tree algorithms for classification. The presented framework automatically classifies the EMG signals into

myopathy, ALS or normal, using CART, C4.5 and random forest decision tree algorithms.

Nevertheless, the decision tree algorithm has one major shortcoming that it is biased in favour of those attributes who own more values while these are not always the best. As a result, the improved algorithm is described in the following sections to solve these problems.

The remainder of this paper is organized as follows. Section 2 provides the preliminary works that are improvement ideas. Section 3 gives the improved RLBOR (Reduce Leaf Based on Optimization Ratio) algorithm based on the decision tree. The algorithm is evaluated and analyzed against the reference algorithm in Sect. 4. Finally, Sect. 5 concludes the paper.

## 2 Preliminary

Hong Jiarong and other scholars have proved that in order to find the optimal decision tree [8, 11, 16–20] we need solve three problems [21]: (1) The least number of leaves. (2) The minimized depth of each leaf. (3) Meeting the above two conditions at the same time.

Based on the above analysis in Sect. 1, the generation of decision tree in ID3 algorithm whose core idea is adopting the information gain to select attributes-is based on the rule (2). The principle of this algorithm will be further introduced and improved according to the rule (1) in the following sections. However, although ID3 algorithm is the most typical one of decision tree learning algorithm, it also needs to be improved.

In the decision tree classification, we assume that  $S$  is the training set and  $|S|$  is the number of training samples. The samples have been divided into  $N$  different types of  $C_1, C_2, \dots, C_n$  and the size of these classes are labeled as  $|C_1|, |C_2|, \dots, |C_n|$ .

In the training sample set  $S$ , the calculation formula of sample probability for class  $C_i$  is shown as below.

$$p(S_i) = \frac{|C_i|}{|S|} \quad (1)$$

Set  $a_1, a_2, \dots, a_n$  for all occurrence possibilities of an event, then the calculation formula of entropy of information is as follows [22].

$$\text{Entropy}(S) = \sum_{i=1}^n p(a_i) \log_2 \frac{1}{p(a_i)} \quad (2)$$

In the formula,  $Entropy(S)$  indicates the entropy of information of training sample set  $S$ . In the case of a certain attribute, the calculation formula of the entropy of information is as follows.

$$Entropy(S,A) = \sum_{v \in V} \left( \frac{|S_v|}{|S|} \right) * Entropy(S_v) \quad (3)$$

$Entropy(S, A)$  represents the entropy of information of the attribute  $A$ ,  $A$  represents a value of the attribute  $A$ ,  $V$  represents all sets of values of  $v$ , and  $S_v$  represents a subset of  $S$  in which the value of  $A$  is  $v$ ,  $|S_v|$  is the number of elements in the  $S_v$ , and  $|S|$  is the number of elements in  $S$ .

The formula for calculating the information gain is as follows.

$$Gain(S,A) = Entropy(S) - Entropy(S,A) \quad (4)$$

$Gain(S, A)$  represents the information gain of the attribute  $A$  in the data set  $S$ ,  $Entropy(S)$  represents information entropy of the sample set  $S$ , while  $Entropy(S, A)$  represents the entropy of information of the attribute  $A$ .

In the ID3 algorithm, the entropy of information is understood as the degree of uncertainty. So if  $Entropy(S)$  is the uncertainty in the training set  $S$ , and  $Entropy(S, A)$  is the uncertainty of the attribute  $A$  in the training set  $S$ , the difference between them is that even the attribute  $A$  has been selected, the amount of uncertainty is still undetermined. From the information gain formula (4), it can be seen that the larger the information gain, the more information you can choose to provide for the classification.

ID3 algorithm selects the attribute that has the maximum information gain according to the data set, and uses it as the attribute of the split node. Then, according to the value of this attribute, the data subset is divided, and the information gain of each subset is calculated recursively. The decision tree trained according to the information gain method is huge and deep. Therefore, although this division reduces the depth, the approach does not take the number of leaf nodes into account. Hence, in this paper, the problem of how to reduce the number of nodes in the decision tree is studied.

### 3 The RLBOR algorithm based on decision tree

Jiarong Hong et al. have proved that it is an NP problem to find the optimal decision tree. Therefore, the purpose of this paper is to find a better decision tree than ID3.

### 3.1 The RLBOR algorithm

As can be seen from the above analysis, the larger the information gain, the more conducive it is to select the property; the fewer the number of leaf nodes, the more tendency it has to select the attribute. So it can be drawn that the value of a property is proportional to the value of the information gain, and it is inversely proportional to the number of leaf nodes. Therefore, in this paper we propose the definition of decision tree optimization rate as follows.

**Definition 1 (Decision Tree Optimization Ratio)** Decision Tree Optimization Ratio is the ratio. The ratio is the information gain of the number of leaf nodes in the decision tree. The decision tree is generated by the current node. The formula is shown as the following.

$$DTOR(S,A) = \frac{Gain(S,A)}{LeafNum(S,A)} \quad (5)$$

In the above formula, DTOR represents Decision Tree Optimization Ratio,  $Gain(S, A)$  represents the information gain of the attribute  $A$  in the data set  $S$ , and  $LeafNum(S, A)$  represents that  $A$  is taken as the attributes, that is, the number of leaf nodes of the decision tree formed according to the data set  $S$ .

It can be seen from the formula (5) that the formula has taken the depth of the decision tree as well as the number of leaf nodes consideration. When the denominators of the two DTORs are equal, the DTOR with larger  $Gain(S, A)$  is larger; when the numerators of the two DTORs are equal, the DTOR with smaller  $LeafNum(S, A)$  is larger. This is consistent with the rule (3) of the optimal decision tree proposed by Hong Jiarong et al. In this paper, the decision tree ratio is regarded as the basis for selecting attributes.

Class attributes are attributes in the data set used to identify the class of the sample.

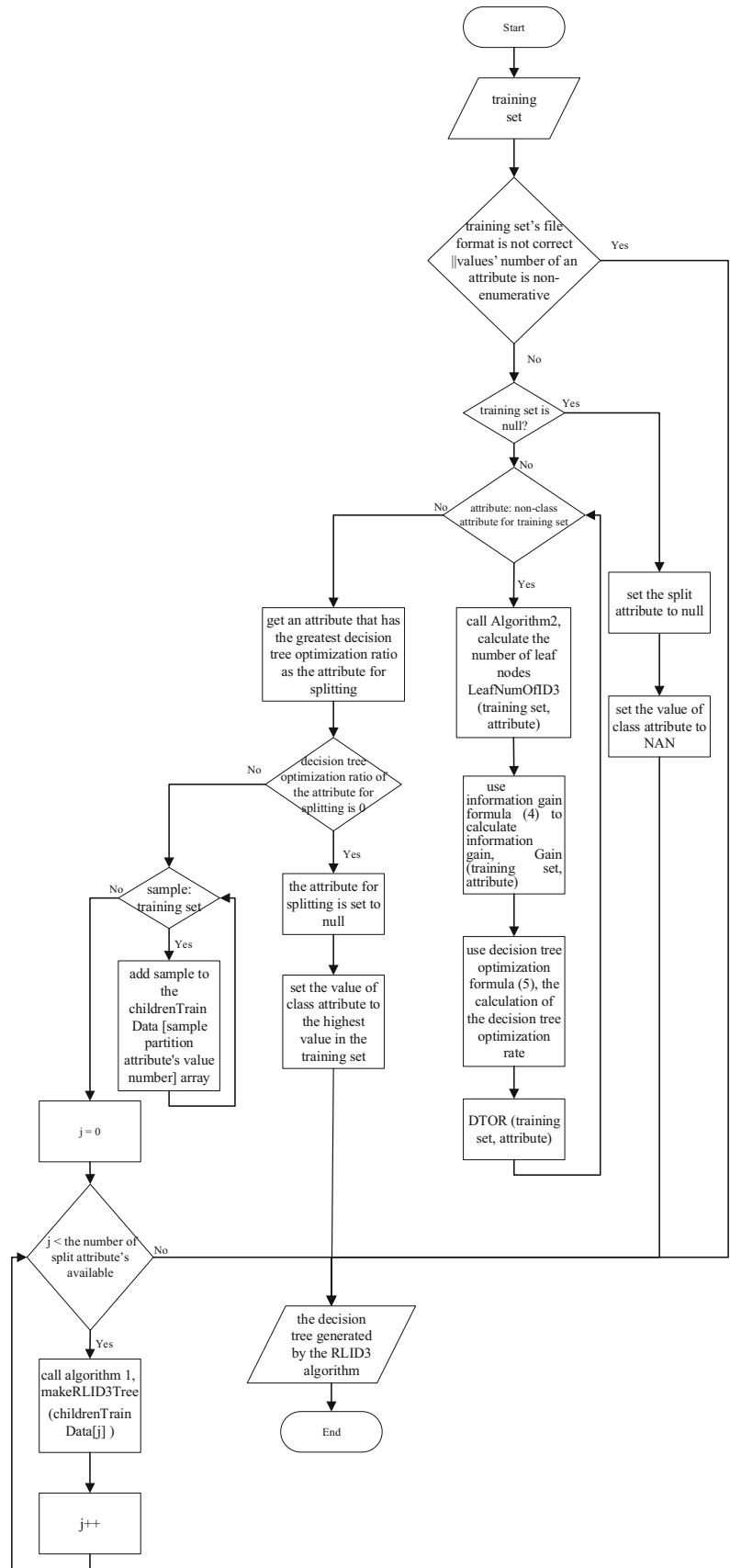
Non-class attributes refer to the attributes in the data set used to identify the condition of a certain aspect of the sample.

For example, in Table 1, outlook and humidity denote non-class attributes; play represents a class attribute. Examples of data collection are shown in Table 1.

**Table 1** Examples of data collection

No.	Outlook	Humidity	Play
1	Sunny	High	No
2	Sunny	High	No
3	Overcast	High	Yes
4	Rainy	High	Yes

**Fig. 1** The flow chart of algorithm 1



On the above theoretical basis, the proposed RLBOR algorithm is shown as below; the flow chart of algorithm 1 is shown in Fig. 1.

---

**Algorithm 1:** makeRLBORTree (training set)  
 Input: training set (TS)  
 Output: the decision tree generated by the RLBOR algorithm (DT)

---

```

01: if (the format of a TS is incorrect || the number of
values of an attribute is non-enumerative)
02: return;
03: if (TS == null) {
04: attribute = null | attribute ∈ divide attribute
05: class attribute = NAN
06: return;
07:}
08:while (attribute | ∃attribute == non-class attribute
∈ TS) {
09: call Algorithm2, return LeafNumOfID3 (TS,
attribute)
10: use formula (4), return Gain (TS, attribute)
11: use formula (5), return optimization rate, DTOR
(TS, attribute)
12:}
13: define SA (attribute for splitting) ⊆ {attribute set
| the greatest optimization ratio}
14:if (SA's optimization ratio == 0) {
15: SA = null
16: define CA (class attribute) ⊆ {class value | the
highest value in the TS}
17:}else{
18: while (sample: TS) {
19: add (sample, childrenTrainData) [numbers of
values that divide attributes in the sample]
20: }
21: for (j = 0; j < |SA|; j++)
22: call algorithm 1
23: makeRLBORTree ( childrenTrainData[j] )
24: }
25:}

```

---

The algorithm for calculating the number of leaf nodes in the ID3 decision tree is shown in the following algorithm 2.

---

**Algorithm 2** LeafNumOfID3 (TS, attribute)  
 Input: training set (TS), attribute  
 Output: Input attribute as the attribute of root nodes,  
 and the number of leaf nodes of ID3 decision tree is  
 generated by TS.

---

```

01:if current attribute = input attribute
02:if (TS == null) {
03: SA = null
04: CA = NAN
05: return;
06:}
07:if (CA == 0) { CA ∈ information gain for the CA
} } {
08: current attribute = null
09: CA = CA for the highest value in the TS
10:}else{
11:while (sample: TS) {
12: add (sample, childrenTrainData) [numbers of
values that divide attributes in the sample]
13:}
14:for (j = 0; j < |SA|; j++){
15: call ID3 algorithm,
16: makeID3Tree (childrenTrainData[j])
17:}
18:Init(count) = 0 // count = global variable count of
leaf nodes' number counter
19:call computeLeafNum(0){
20: return |leaf nodes|
21:}
22:leafCount = count
23:count = 0 // for calculating the next number of
leaf nodes
24: return leafCount;

```

---

The  $m\_Successors[j]$  shows that the current attribute value is the successor node of the node corresponding to the  $j$  value of the current attribute. Use makeID3Tree (training set) to represent ID3 algorithm, the flow chart of algorithm 2 is shown in Fig. 2.

### 3.2 The RLBOR algorithm example analysis

The example selects the nursery data set in UCI database, the total amount of the samples is 12960. In this paper, we only present the difference between the decision trees obtained by ID3 algorithm and RLBOR algorithm. The structures of the trees obtained by ID3 and RLBOR are shown in Figs. 3 and 4.

Figures 3 and 4 describe a branch of the decision tree, which forms one of the two branches of the decision tree.

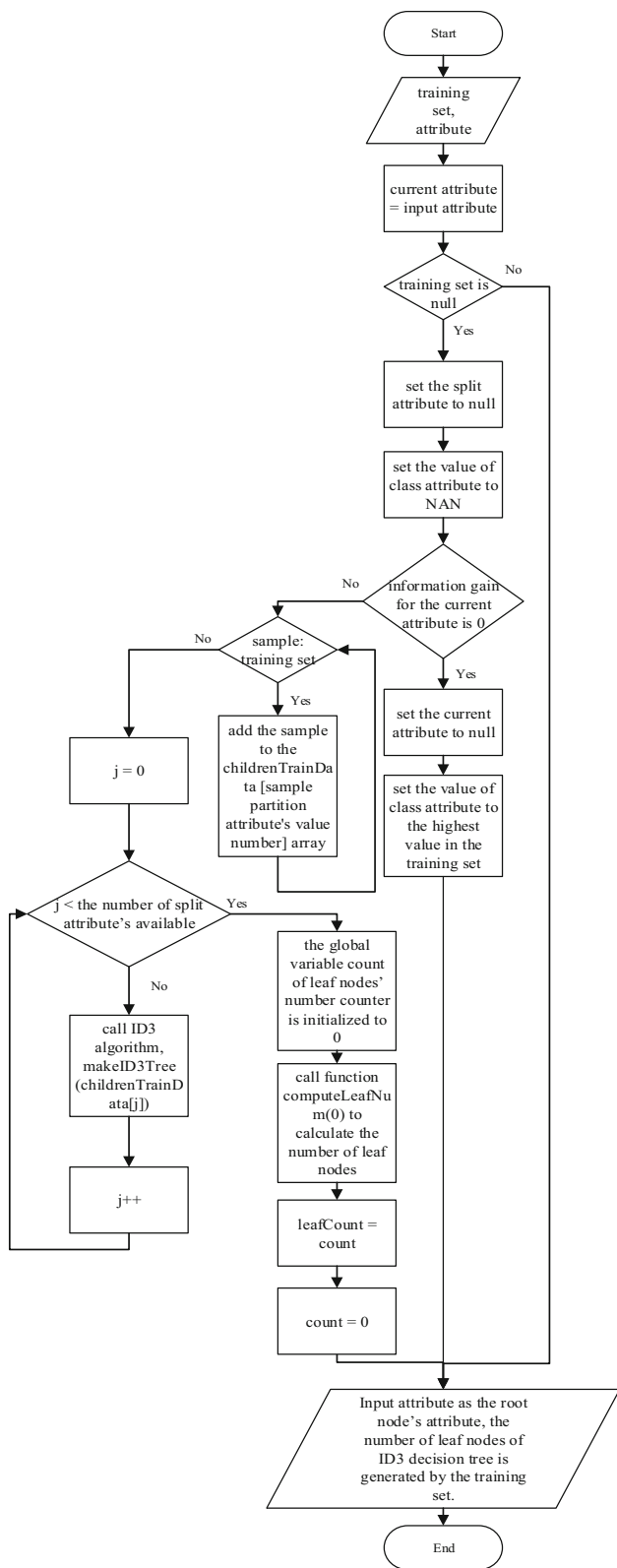


Fig. 2 The flow chart of algorithm 2

In Figs. 3 and 4, the branch is the one that takes the parents attribute as a split attribute and valued as a great\_pret. In the decision trees generated by ID3 algorithm and RLBOR algorithm, parents are on the third layer of the decision tree.

From the shape of the branches of above two decision trees.

1. The depth of RLBOR tree is less than that of ID3 tree.
2. The numbers of leaves contained in the RLBOR are less than that in the ID3 tree.
3. The structure of the RLBOR tree is more simplified and does not employ the finance attribute.

Number of leaves contained in the ID3 tree and RLBOR are 839 and 807 respectively.

### 4 Experiment results and analysis

#### 4.1 Selection of data set

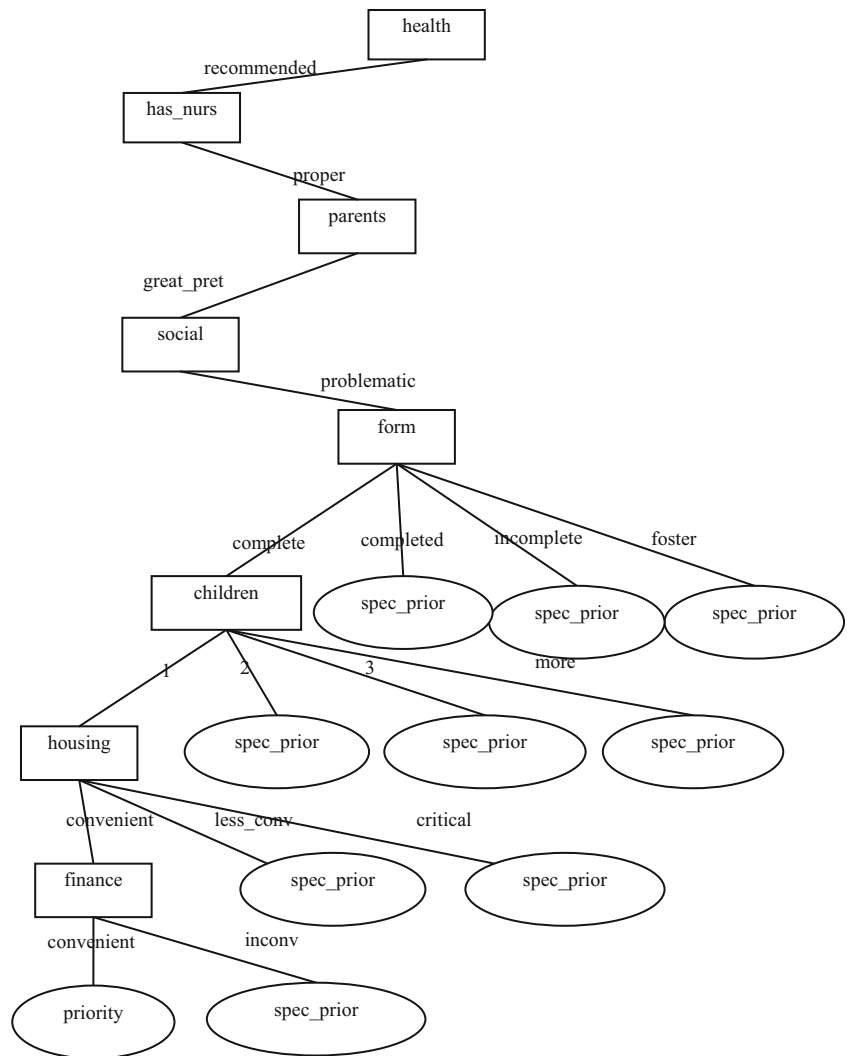
In this section, the classification problem is studied and discussed, with 10 wide-range data sets related to classification covering the fields of life, computer, social matters and games in the UCI database employed in the experiment. The basic information of the data sets is shown in Table 2.

Because the algorithm in this paper can only processes the samples with non-missing values, the Mushroom and Breast-w are processed. As for methods to preprocess the data set with missing values, two operations are involved in this paper: the first one is to remove the stalk-root from the data set; the second one is to remove the data with missing values from the data set. In the Mushroom data set, only the stalk-root attribute has missing values, and the correspondent proportion of missing data is large. Therefore, the paper removes the stalk-root attribute from the Mushroom directly. After the above process, there are no data set with samples of missing values, so that the situation where no categorical attributes, attributes are too little, or the attribute value is not enumerable can be avoided.

#### 4.2 Evaluation method

In this paper, precision is used as the metric of the algorithms, and the formula is as follows.

**Fig. 3** Branch of decision tree based on ID3 algorithm



Precision

$$= \frac{\text{The numbers of samples that can be correctly classified}}{\text{The numbers of samples}}, \tag{6}$$

**4.3 Results and analysis**

In the following experiments, the number of the leaf nodes and the depth of the decision tree are generated from the whole data set. Hence, it is more conducive to compare the differences between decision trees generated by different algorithms. The classification accuracy of ID3 algorithm and RLBOR algorithm on different data sets is listed below.

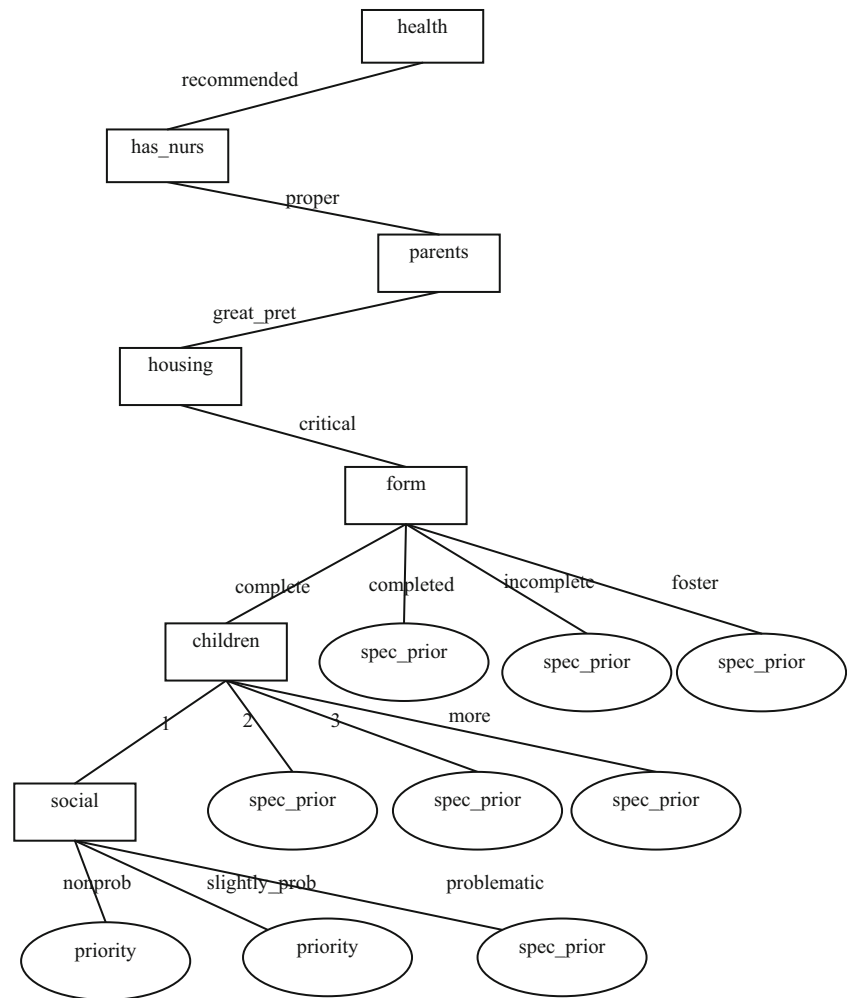
As can be seen from Fig. 5, RLBOR is more precise than ID3 for the above sets. Since the decision trees of ID3

algorithm and RLBOR algorithm are affected by the training set, it is possible that the training set is too small to be fully trained. However, in this paper a larger data set, UCI data set, is employed, thus the above possibility can be avoided.

In the above data sets, the Letter and Connect-4 data set are much larger than the others, which causes differences in the size of the number of leaves generated from the data sets. Hence, the data sets with fewer leaf nodes are compared in order to better analyze the comparative effect with the results shown in Fig. 6.

In the above data set, the number of leaves in the RLBOR tree is less than that in the ID3 tree, and the larger the size of the data set, the less the leaves. The comparison results of the depth of decision tree are also shown in Fig. 7.

**Fig. 4** Branch of decision tree based on RLBOR algorithm



**Table 2** Basic information of the data set

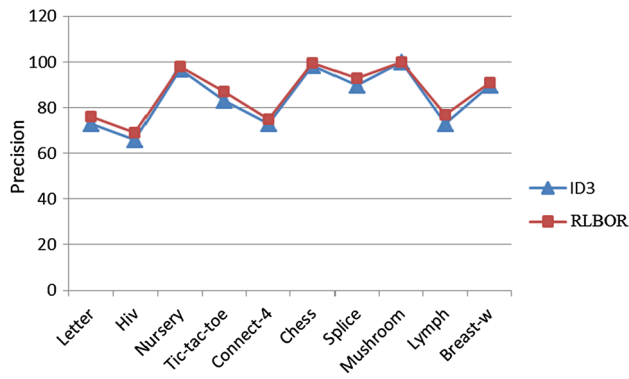
Name	Instances	Attributes	Classifications	Missing values
Letter	20,000	16	26	No
HIV	6590	8	2	No
Nursery	12,960	8	5	No
Tic-Tac-Toe	958	9	2	No
Connect-4	67,557	42	3	No
Chess	3196	36	2	No
Splice	3190	61	3	No
Mushroom	8124	22	2	Yes
Lymph	148	18	4	No
Breast-w	699	10	2	Yes

In the above data set, the depth of the RLBOR tree is almost the same as that in the ID3 tree, which is often the case. Even if there is difference in the depth, it won't be much. The conclusion is proposed that RLBOR is more precise.

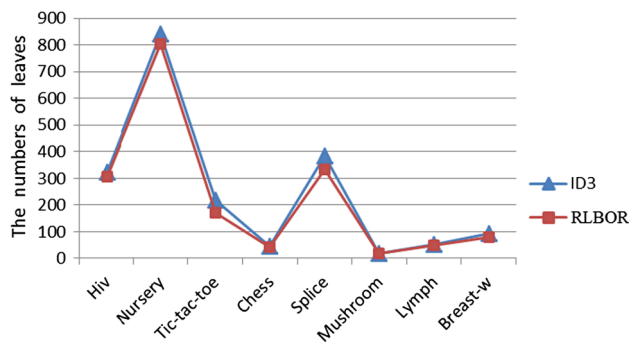
### 5 Conclusion

It is important to manage data. One important work is information classification. ID3 algorithm, a classical classification algorithm, the research of which has been a hot

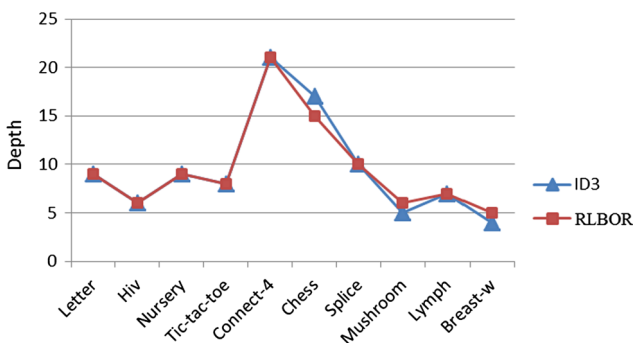




**Fig. 5** Precision comparison diagram between ID3 algorithm and RLBOR algorithm



**Fig. 6** Leaves comparison diagram between ID3 algorithm and RLBOR algorithm



**Fig. 7** Depth comparison diagram between ID3 algorithm and RLBOR algorithm

topic and mainly focused on the improvement and application of ID3 algorithm at present. Many experts and scholars have studied how to improve the ID3 algorithm from various angles, some of them focused on how to apply the ID3 algorithm into the actual production and the real life. As for the structure of the paper, it first summarizes the research status of ID3 algorithm, and then puts forward the method to reduce the complexity of ID3. In view of the problem that ID3 algorithm does not consider the number of leaf nodes of decision tree, RLBOR algorithm based on the decision tree optimization ratio is proposed in this

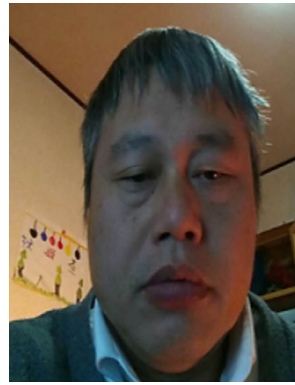
paper. The experiment shows more precise of RLBOR algorithm. Then the difference between ID3 tree and RLBOR tree is compared through examples. Future research will consider how to further improve the construction speed of ID3 algorithm, and whether it can be built in parallel. Besides, the possibility of further reducing the spatial storage of ID3 algorithm may also be studied.

**Acknowledgements** This work was funded by the National Natural Science Foundation of China under Grant (Nos. 61772152, 61502037), and the Basic Research Project (Nos. JCKY2016206B001, JCKY2014206C002 and JCKY2017604C010).

## References

- Dobra, A., Garofalakis, M., Gehrke, J. and Rastogi, R.: Processing complex aggregate queries over data streams. In: Proc. 2002 ACM Sigmod Int. Conf. Management of Data, Madison, WI, pp. 90–98 (2002)
- Rajgopal Kannan, K., Sarangi, S., Ray, S. and Sitharama Iyengar, S.: Minimal sensor integrity in sensor grids. In: Proc. Int. Conf. Parallel Processing, pp. 21–27 (2002)
- Quinlan, J.R.: Induction of decision trees, Machine Learning, pp. 257–264 (1986)
- Liu, Y., Pi, D., Cheng, Q.: Ensemble kernel method: SVM classification based on game theory. J. Syst. Eng. Electr. **27**, 251–259 (2016)
- Antal, M., Bokor, Z., Szabó, L.Z.: Information revealed from scrolling interactions on mobile devices. Pattern Recognit. Lett. **56**, 7–13 (2015)
- Khatami, R., Mountrakis, G., Stehman, S.V.: A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: general guidelines for practitioners and future research. Remote Sens. Environ. **177**, 89–100 (2016)
- Lee, J., kim, D.W.: Memetic feature selection algorithm for multi-label classification. Inf. Sci. **93**, 80–96 (2015)
- Kavzoglu, T., Colkesen, I., Yomralioglu, T.: Object-based classification with rotation forest ensemble learning algorithm using very-high-resolution WorldView-2 image. Remote Sens. Lett. **6**, 834–843 (2015)
- Jiang, W.X., Zhong, Y.Z., Liang, H.: An Evaluation model of polytechnic teaching quality based on ID3 decision making Tree. In: 3rd International Conference on Advanced Engineering Materials and Architecture Science (ICAEMAS), Huhhot, China, pp. 2437–2440 (2014)
- Li, J.F., Lei, J.H., Zhao, X.X.: An improved ID3 algorithm. In: 2nd International Conference on Advances in Computational Modeling and Simulation (ACMS 2013), Kunming, China, pp. 723–727 (2014)
- Ludwig, S.A., Jakobovic, D., Picek, S.: Analyzing gene expression data: fuzzy decision tree algorithm applied to the classification of cancer data, Fuzzy Systems (FUZZ-IEEE). In: 2015 IEEE International Conference on. IEEE, pp. 1–8 (2015)
- Chhipi-Shrestha, G., Mori, J., Hewage, K., et al.: Clostridium difficile infection incidence prediction in hospitals (CDIIPH): a predictive model based on decision tree and fuzzy techniques. Stoch. Environ. Res. Risk Assess. **31**, 417–430 (2017)
- Jin, C.X., Li, F.C., Li, Y.: A generalized fuzzy ID3 algorithm using generalized information entropy, knowledge-based systems, pp. 13–21(2014)

14. Ahmadi, E., Javadi, H., Khansefid, A., et al.: Fuzzy decision tree learning for preoperative classification of adnexal masses. In: 4th International Conference on Health Informatics (HEALTHINF 2011), Rome, ITALY, pp. 364–375 (2011)
15. Zeinalkhani, M., Eftekhari, M.: Fuzzy partitioning of continuous attributes through discretization methods to construct fuzzy decision tree classifiers. *Inf. Sci.* **278**(26), 715–735 (2014)
16. Gokgoz, E., Subasi, A.: Comparison of decision tree algorithms for EMG signal classification using DWT. *Biomed. Signal. Process. Control* **18**, 138–144 (2015)
17. Vlahovic, N.: An evaluation framework and a brief survey of decision tree tools. In: 39th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, CROATIA, pp. 1299–1304 (2016)
18. Hameed, A., Dai, R., Balas, B.: A decision-tree-based perceptual video quality prediction model and its application in FEC for wireless multimedia communications. *IEEE Trans. Multimed* **18**, 764–774 (2016)
19. Ronowicz, J., Thommes, M., Kleinebudde, P., et al.: A data mining approach to optimize pellets manufacturing process based on a decision tree algorithm. *Eur. J. Parm. Sci.* **73**, 44–48 (2015)
20. Hashemi, S.A.H., Ghodrati Amiri, A.C., Hamed, F.: Steel buildings damage classification by damage spectrum and decision tree algorithm. *J. Rehabil. Civil Eng.* **3**, 24–42 (2015)
21. Hong, J.R., Ding, M.F., Li, X.Y., Wang, L.W.: A new algorithm of decision tree induction. *Chin. J. Comput.* **18**, 470–474 (1995)
22. Wang, X.H., Wang, L.L., Li, N.F.: An application of decision tree based on ID3. *Phys. Procedia* **25**, 1017–1021 (2012)



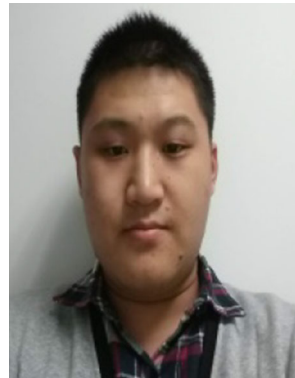
**Yucai Zhou** (1971-) received her Ph.D. degree in mechanical and electrical Engineering from central south university in 2012, he is a vice professor in the Energy and Power Institute of Changsha University of Science and Technology, his research in signal processing, intelligent control and intelligent system, embed system.



**Lianke Zhou** (1977-) received his Ph.D. degree in computer architecture from Harbin Institute of Technology in 2011. He is currently a lecturer in the college of Computer Science and Technology of Harbin Engineering University. His research interests include data visualization, dataspace, distributed computing and mobile computing.



**Hongbin Wang** (1979-) received his Ph.D. degree in computer application technology from Harbin Engineering University in 2010. He is currently an associate professor in the college of Computer Science and Technology of Harbin Engineering University. His research interests include information management, data integration, data space, semantic web, ontology and information system design.



**Huafeng Li** (1991-) received his master degree in software engineering from Harbin Engineering University in 2016. His research interests include information management, data integration, and data classification.



**Tong Wang** (1977-) received his Ph.D. degree in computer application technology from Harbin Engineering University in 2007. He is currently an associate professor in the college of Information and Communication Engineering of Harbin Engineering University. His research interests include Wireless Network, Vehicular adhoc Network and Internet of Things.