



Decision Tree Algorithm based on Granular Computing and Important Degree of Attribute Value

Liu Ping^a, Wu Zhengang^a, Ge Licheng^b, Wang Haochong^c, Yang Junping^{*d}

^a School of information engineering, Nanchang University, Nanchang 330031

^b Institute of Advanced study, Nanchang University, Nanchang330031

^c Institute of electrical and mechanical engineering, Nanchang University, Nanchang 330031

^d Affiliated Hospital of Jiangxi College of Traditional Chinese Medicine, Nanchang 330006, China
yfyjp0507@126.com

Conventional decision tree algorithm employs information gain and information gain ratio to select splitting attribute, and avoid attribute value significance. According to the analysis on a single attribute decision-making problem, it is found that, different value of the same condition attribute has different influence on decision-making results. Based on this preliminary conclusion, proportion matrix and Euclidean norm are introduced to quantitatively describe the important degree of attribute value and a decision tree algorithms proposed based on granular computing. Experimental results show that, compared with ID3 algorithm, the proposed algorithm has higher accuracy when applied to classification problems with multiple attribute values.

1. Introduction

Decision tree algorithm is an effective nonparametric method of learning which can be used for classification and regression. As early as in 1979, Quinlan proposed the famous ID3 algorithm, a decision tree classification algorithm based on information entropy [Quinlan (1986) reported]. After the concept was proposed, many scholars developed an interest on ID3 algorithm and made much improvement in respond to its weaknesses. One of the research directions is from the perspective of threshold setting and pruning strategy [Chen et al. (2011) reported]. Due to distinct features of data sets, and the fact that the decision tree algorithm resembles hill-climbing, such perspective has received wide discussion which never ends. Another direction is from the perspective of taxonomy, for example, weigh the information entropy of ID3 algorithm to get new information gain [Chen et al(2011),Yu(2012)reported, DING et al (2010) reported, ZHOU et al(2010) reported, Chenxia (2014) reported]. Other directions are engaged in using rough set [Qiu et al. (2009)reported, DING et al (2010) reported, Ran et al(2014) reported] or fuzzy mathematics [Eyke (2008) reported, Chenxia (2014) reported, V. Srinivasan(2013) reported], or employing modern artificial intelligence algorithms [FENG et al.(2009) reported].

The traditional ID3 algorithm and the improved algorithm are all considered to select the best splitting attribute, without considering the influence of different values of the same attribute on results. Through the analysis of a single attribute decision-making problem, this paper figures out the direct influence of the condition attribute value on decision-making results and uses the proportion matrix and the Euclidean norm to conduct a quantitative analysis so as to confirm new classifications.

2. Relevant concepts of granule in the information system

Granular computing [Han et al.(2010)reported, Qiu et al.(2009)reported, Zhang et al.(2011)reported] takes on a micro perspective. By roughly or detailed computing the granule, people can observe the granule from different aspects, analyze and solve the problem.

An information system can be formalized as $S = (U, A, V, f)$, $U = \{u_1, u_2, \dots, u_N\}$ is a set of objects; it is a non-empty finite set of objects, called the universe, $A = C \cup D$ is a set of attributes. C as the condition attributes set, D as the decision attribute set. $A = \{a_1, a_2, \dots, a_M\}$, $V = \{V_{a_1}, V_{a_2}, \dots, V_{a_M}\}$ is a set of attribute values, $V_{a_i} = \{V_{a_i,1}, V_{a_i,2}, \dots, V_{a_i,j}\}$ is the domain of a_i attribute, $f(u, a_i) : U \times A \rightarrow V_{a_i}$ f is a mapping, $\forall u \in U, a_i \in A, f(u, a_i) \in V_{a_i}$

Definition 1(Elementary granule)

For $a_i(a_i \in A)$, Function $h(a_i, v_j)$ reflects the set of objects whose value is $v_j(v_j \in V_{a_i})$. In the information system $S = \langle U, A, V, f \rangle$, $G_r = ((a_i, v_j), h(a_i, v_j))$, $a_i \in A, G_r$ is named elementary granule of the information system. Size of information granule can be defined as the cardinality of the extension of the information granule, it is expressed as $|h(a_i, v_j)|$. ($i = 1, \dots, |A|; j = 1, \dots, |V_{a_i}|$).

Definition 2(Elementary condition granule and Elementary decision granular)

For the information system $S = \langle U, A, V, f \rangle$, attribute sets are divided into two parts, in which $A = C \cup D$, $C = \{c_1, c_2, \dots, c_s\}$ is named the condition attribute set and $D = \{d_1, d_2, \dots, d_T\}$ is named the decision attribute set. The concept of elementary condition granular:

$$G_{rc} = ((c_i, v_j), h(c_i, v_j)), c_i \in C, v_j \in V_{c_i} (i = 1, \dots, s; j = 1, \dots, |V_{c_i}|)$$

The concept of elementary decision granular: $G_{rd} = ((d_i, v_j), h(d_i, v_j)), d_i \in D, v_j \in V_{d_i} (i = 1, \dots, T; j = 1, \dots, |V_{d_i}|)$

Definition 3 (combination granule)

Suppose ω is the logical combination of many elementary granules using logical words $\wedge, \vee, \rightarrow, \sim, \leftarrow$. $h(\omega)$ is used to show the set the objects which meet the requirement of the logical combination ω . $G_r = ((\omega), h(\omega))$ is named the combination granule. Size of combination granule [Qiu et al.(2009)reported] can be defined as the cardinality of the extension of the combination granule, it is expressed as $L(G_r) = |h(\omega)|$

Definition 4 (intersection \otimes)

For two combination granules: $G_{r1} = ((\omega_1), h(\omega_1)), G_{r2} = ((\omega_2), h(\omega_2))$ $G_{r1} \otimes G_{r2} = ((\omega_1 \vee \omega_2), h(\omega_1 \wedge \omega_2))$,

Definition 5(The condition granular set)

The concept of junior condition granular set is: $CS_i = \{G_{rc_{i,1}}, G_{rc_{i,2}} \dots G_{rc_{i,j}}\}$, $G_{rc_{i,j}} = ((a_i, v_j), h(a_i, v_j)), a_i \in C, v_j \in V_{c_i}, (i = 1, \dots, |C|; j = 1, \dots, |V_{c_i}|)$, it shows a set of all elementary condition granules of a condition attribute, and we name it junior set of condition granular. We name the combination granule set after several intersections of junior condition sets as a set of condition granular, and every combination granule is one combination of the condition granular set. If two combinations belong to the same condition granular set, we name it consanguinity association. If two combinations do not belong to the same condition set, we name it non-consanguinity association.

Definition 6(Intersection of condition granular set : \otimes)

For any two junior condition sets

$$CS_m = \{G_{rc_{m,1}}, G_{rc_{m,2}} \dots G_{rc_{m,s}}\} \quad CS_n = \{G_{rc_{n,1}}, G_{rc_{n,2}} \dots G_{rc_{n,t}}\}$$

$$CS_m \otimes CS_n = \left\{ \begin{array}{l} G_{rc_{m,1}} \otimes G_{rc_{n,1}}, G_{rc_{m,1}} \otimes G_{rc_{n,2}} \dots G_{rc_{m,1}} \otimes G_{rc_{n,t}} \\ G_{rc_{m,2}} \otimes G_{rc_{n,1}}, G_{rc_{m,2}} \otimes G_{rc_{n,2}} \dots G_{rc_{m,2}} \otimes G_{rc_{n,t}} \\ \vdots \\ G_{rc_{m,s}} \otimes G_{rc_{n,1}}, G_{rc_{m,s}} \otimes G_{rc_{n,2}} \dots G_{rc_{m,s}} \otimes G_{rc_{n,t}} \end{array} \right\}$$

Definition 7 (The set of decision granular)

The concept of decision granular set is: $D_K = \{G_{rd_{k,1}}, G_{rd_{k,2}} \dots G_{rd_{k,t}}\}$, it shows a set of all elementary decision granules about a decision attribute.

Definition 8(Intersection of condition granular set and decision granular set)

Without loss of generality, we assume that there is only one decision attribute.

$$CS_{\omega_m} = \{G_{r\omega_{m,1}}, G_{r\omega_{m,2}}, \dots G_{r\omega_{m,s}}\} \quad D_K = \{G_{rd_{k,1}}, G_{rd_{k,2}} \dots G_{rd_{k,t}}\}$$

$$CS_m \otimes D_K = \left\{ \begin{array}{l} G_{r\omega_{m,1}} \otimes G_{rd_{k,1}}, G_{r\omega_{m,1}} \otimes G_{rd_{k,2}}, \dots, G_{r\omega_{m,1}} \otimes G_{rd_{k,t}} \\ G_{r\omega_{m,2}} \otimes G_{rd_{k,1}}, G_{r\omega_{m,2}} \otimes G_{rd_{k,2}}, \dots, G_{r\omega_{m,2}} \otimes G_{rd_{k,t}} \\ \vdots \\ G_{r\omega_{m,s}} \otimes G_{rd_{k,1}}, G_{r\omega_{m,s}} \otimes G_{rd_{k,2}}, \dots, G_{r\omega_{m,s}} \otimes G_{rd_{k,t}} \end{array} \right\}$$

Definition 9 (proportion matrix)

For a decision attribute $d_k \in D, V_{d_k} = \{V_{d_{k,1}}, V_{d_{k,2}} \dots V_{d_{k,t}}\}$ is the range of the decision attribute. All combinations of a condition set CS is $cs_i (i = 1, 2, 3 \dots s)$. We name the ratio of decision granular set and condition granular set after intersection as proportion matrix p ,

$$P = \begin{pmatrix} \frac{L(G_{r\omega_{m,1}} \otimes G_{rd_{k,1}})}{L(G_{r\omega_{m,1}})} & \frac{L(G_{r\omega_{m,1}} \otimes G_{rd_{k,2}})}{L(G_{r\omega_{m,1}})} & \dots & \frac{L(G_{r\omega_{m,1}} \otimes G_{rd_{k,t}})}{L(G_{r\omega_{m,1}})} \\ \frac{L(G_{r\omega_{m,2}} \otimes G_{rd_{k,1}})}{L(G_{r\omega_{m,2}})} & \frac{L(G_{r\omega_{m,2}} \otimes G_{rd_{k,2}})}{L(G_{r\omega_{m,2}})} & \dots & \frac{L(G_{r\omega_{m,2}} \otimes G_{rd_{k,t}})}{L(G_{r\omega_{m,2}})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{L(G_{r\omega_{m,s}} \otimes G_{rd_{k,1}})}{L(G_{r\omega_{m,s}})} & \frac{L(G_{r\omega_{m,s}} \otimes G_{rd_{k,2}})}{L(G_{r\omega_{m,s}})} & \dots & \frac{L(G_{r\omega_{m,s}} \otimes G_{rd_{k,t}})}{L(G_{r\omega_{m,s}})} \end{pmatrix}$$

In particularly, if CS is a junior condition set, we name the proportion matrix as junior proportion matrix. Meanwhile, we name the j line of the proportion matrix P as ratio vector of condition combination granule cs_j .

3. Decision tree algorithm based on granular computing and important degree of attribute value

Definition10 Important degree of attribute value. (The correlation degree about the condition granule combination and the decision granule)

We take the correlation degree between a combination of condition granule cs_j and the decision attribute D as important degree of attribute value.

Hypothesis that the proportion of vector cs_j as follows: $P_j = \begin{bmatrix} p_{1j} \\ p_{2j} \\ \vdots \\ p_{mj} \end{bmatrix}$

$$\textcircled{1} \text{Normalization } P_j' = \frac{p_j}{\text{sum}(p_j)} = \begin{bmatrix} p'_{1j} \\ p'_{2j} \\ \vdots \\ p'_{mj} \end{bmatrix} \quad \textcircled{2} \text{Euclidean norm } R = \sqrt{p'^2_{1j} + p'^2_{2j} + \dots + p'^2_{mj}}$$

R expresses the important degree of attribute value.

Decision tree algorithm description:

Algorithm 1: Produce the Elementary granule space

Input: Decision information system $S = (U, A, V, f)$, $A = C \cup D$, object $x \in U$, attribute $a \in A$.

Output: Elementary granule space (Elementary condition granule set CS and Elementary decision granular D).

$CS = \emptyset; D = \emptyset;$

for $(a_i \in A(1 \leq i \ll |A|))$ do

for $(v_j \in V_i(1 \leq j \ll |V_{a_i}|))$

for $u_k \in U(1 \leq k \leq |U|)$ do

if $(f(u_k, a_i) = v_j) \quad j = 1, \dots, |V_{a_i}|$

{ $\varphi = (a_i, v_j); \quad h(\varphi) = h(\varphi) + \{u_k\};$ //end if

if $(a_i \in C)$

{ $G_{rc} = ((c_i, v_j), h(c_i, v_j)); \quad CS_i = CS_i + \{G_{rc}\}$

else { $G_{rd} = ((d_i, v_j), h(d_i, v_j)); \quad D_i = D_i + \{G_{rd}\}$

Algorithm 2: Construct the decision tree

Repeat

{Subject each combination and each non-consanguinity association at the right:

Step1: According to the Definition 6-10, calculate important degree of attribute value.

Compute $CS_m \otimes CS_n$ and $CS_m \otimes D_k$ according to definition 6, 8.

// CS_m, CS_n are non-consanguinity association.

Get the proportion matrix according to definition 9.

Calculate the important degree of attribute value according to definition 10.

Step2: ranking from the maximum of the important degree of attribute value to the minimum.

Step3: Store the decision result in the node.}

Until there is no non-consanguinity association at the right of this node.

4. Case study

Table 1: An example data set

No.	A	B	C	Class
1.	R	1	0	Y
2.	R	2	0	Y
3.	S	1	1	N
4.	R	2	1	Y
5.	R	2	1	Y
6.	S	2	1	N
7.	S	1	0	Y
8.	S	1	1	Y
9.	R	2	0	N
10.	R	2	0	N
11.	S	1	1	N
12.	R	1	1	Y
13.	R	1	1	Y

Step1: According to Algorithm 1, set up elementary granule space.
 Elementary condition granule set CS:

$$\{(A, R) [1, 2, 4, 5, 9, 10, 12, 13]; (A, S) [3, 6, 7, 8, 11]\}^{\cup}$$

$$(B, 1) [1, 3, 7, 8, 11, 12, 13]; (B, 2) [2, 4, 5, 6, 9, 10]\}^{\cup}$$

$$(C, 0) [1, 2, 7, 9, 10]; (C, 1) [3, 4, 5, 6, 8, 11, 12, 13]\}^{\cup}$$

$$\{(class, Y) [1, 2, 4, 5, 7, 8, 12, 13] (class, N) [3, 6, 9, 10, 11]\}^{\cup}$$

Figure1: Example of granule set

Step2: According to Algorithm 2, construct the decision tree

Step2.1 According to the Definition 6-10, calculates important degree of attribute value.
 Compute $CS_m \otimes CS_n$ and $CS_m \otimes D_K$ according to definition 6, 8.

\otimes	(class, Y)	(class, N)
(A, R)	[1, 2, 4, 5, 12, 13]	[9, 10]
(A, S)	[7, 8]	[3, 6, 11]

\otimes	(class, Y)	(class, N)
(B, 1)	[1, 7, 8, 12, 13]	[3, 11]
(B, 2)	[2, 4, 5]	[6, 9, 10]

\otimes	(class, Y)	(class, N)
(C, 0)	[1, 2, 7]	[9, 10]
(C, 1)	[4, 5, 8, 12, 13]	[3, 6, 11]

Figure 2: Calculation process of combination granule

$$A: \begin{bmatrix} 6/13 & 2/13 \\ 2/13 & 3/13 \end{bmatrix} P_A = \begin{bmatrix} 0.7906 & \\ & 0.7211 \end{bmatrix} \quad B: \begin{bmatrix} 5/13 & 2/13 \\ 3/13 & 3/13 \end{bmatrix} P_B = \begin{bmatrix} 0.7693 & \\ & 0.7071 \end{bmatrix} \quad C: \begin{bmatrix} 3/13 & 2/13 \\ 5/13 & 3/13 \end{bmatrix} P_C = \begin{bmatrix} 0.7606 & \\ & 0.7289 \end{bmatrix}$$

Figure3: Calculation process of proportion matrix and important degree of attribute value

Step2.2 ranking from the maximum of the important degree of attributes value to the minimum.



Figure 4: The 1-st layer of decision tree

Step3: The method is same as step2.
For example, we construct the next layer of A.R.

\otimes	(class,Y)	(class,N)	$\begin{bmatrix} 3/8 & 0 \\ 3/8 & 2/8 \end{bmatrix}$	$P_1 = \begin{bmatrix} 1 \\ 0.7211 \end{bmatrix}$
A.R \oplus B.1	[1,12,13]	[]		
A.R \oplus B.2	[2,4,5]	[9,10]		

\otimes	(class,Y)	(class,N)	$\begin{bmatrix} 2/11 & 2/11 \\ 4/11 & 3/11 \end{bmatrix}$	$P_2 = \begin{bmatrix} 0.7071 \\ 0.7143 \end{bmatrix}$
A.R \oplus C.0	[1,2]	[9,10]		
A.R \oplus C.1	[4,5,12,13]	[3,6,11]		

Figure 5: Calculation process about the 2-st layer of decision tree

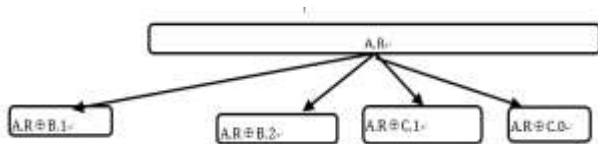


Figure 6: The 2-st layer of decision tree

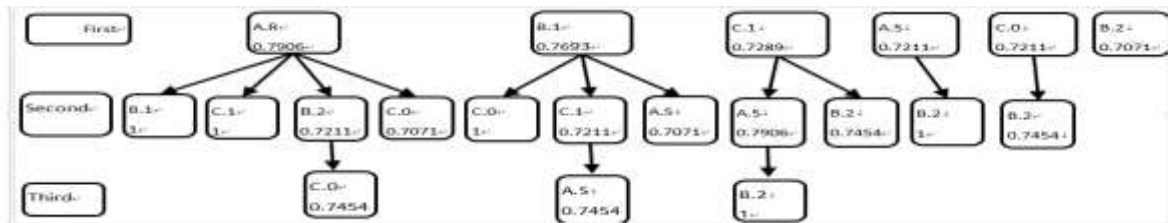


Figure 7: An example of decision tree

5. Data test

Referring to test plan in the article (Tao, 2012), this paper compares the accuracy of the algorithm with that of ID3.

Test plan: divide the data set “Nursery” randomly into 8 shares. Each one share is named a training set, while the rest seven are named test sets. Divide the data set Car Evaluation randomly into 4 shares. Each one share is named a training set, while the rest three are named test sets.

Table 2: Two test sets

Data set	Nursery	Car Evaluation
Number of data set	12960	1278
Number of training set	1620	432
Number of test set	11340	1296
Number of condition attribute	8	6
Number of decision types	5	4

Experimental results are shown as Table 3:

Table 3: Accuracy of the test sets

Algorithm	Nursery	Car Evaluation
ID3	78.90%	76.62%
The proposed method	88.83%	74.42%

Take the average of test set accuracy as the test results and compare it with ID3 algorithm:

For the “Nursery” data set, the classification accuracy of the proposed algorithm is higher than that of ID3 algorithm. But the “Car Evaluation” data set, the classification accuracy of the proposed algorithm is lower than that of ID3 algorithm. The author explains that the “Nursery” data set has more condition attributes and junior combinations. The proposed algorithm has high requirement on condition attributes and junior combinations. So the “Nursery” data set achieves better classification effect. For the Car Evaluation data set, as the samples are in small amount, it would be meaningless to divide into more shares. So the proposed algorithm is not suitable for small data.

6. Conclusion

The proposed algorithm that makes the classification from the perspective of granular is quite easy as it is integrated with mathematics. It is adapted to multi-attribute decision-making problem and problems with multiple conditions. Still, there is much room for improvement.

Acknowledgments

This research is supported by the National Natural Science Foundation of China under grant No. 81260578 and No. 81460769.

Reference

- Chen J.J., Su S.B., Xu H.L., 2011, Decision tree optimization algorithm based on multiscale rough set, *Journal of Computer Application*, VOL.31 No.12, 3243-3246, DOI:10.3724/SP.J.1087.2011.03243.
- Ding C.R., Li L.S., 2010, Application of the Variable Precision Rough Set Model in Building Decision Trees, *COMPUTER ENGINEERING & SCIENCE*, VOL.32 No.7, 88-125, DOI:10.3969/j.issn.1007-130X.2010.07.024.
- Feng S.R., Xiao W.J., 2009, Improved Decision Tree Algorithm Based on Samples Selection, *JOURNAL OF SOUTHWEST JIAOTONG UNIVERSITY*, VOL.44, No.5, 643-647, DOI 10.3969/j.issn.0258-2724.2009.05.003.
- Han J.C., Lin T.Y., 2010, Granular Computing: Models and Applications. *INTERNATIONAL JOURNAL OF INTELLIGENT SYSTEMS*, VOL.25, 111-117, DOI 10.1002/int.20390
- Hullemeier E., 2011, Fuzzy sets in machine learning and data mining, *Applied Soft Computing Journal*, 11 (2) 1493–1505, DOI:10.1016/j.asoc.2008.01.004.
- Jin C.X., Li F.C., li Y., 2014, A generalized fuzzy ID3 algorithm using generalized information Entropy, *knowledge-Based Systems* 64 (2014) 13-21 <http://dx.doi.org/10.1016/j.knosys.2014.03.014>0950-7051.
- Jin R., Kou C.H., Liu R.J., 2014, Biclustering algorithm of differential co-expression for gene data. *REVIEW OF COMPUTER ENGINEER STUDIES* 1(1):7-14.
- Qiu T.R., Liu Q., Huang H.K., 2009, A Granular Computing Approach to Knowledge Discovery in Relational Databases. *ACTA AUTOMATICA SINACA*, 35(8):1071-1079, DOI:10.3724/SP.J.1004.2009.01071.
- Quinlan J.R., 1986, Induction of Decision Tree. *Machine Learning*, 1(1):81-106, DOI: 10.1023/A:1022643204877.
- Srinivasan V., Rajenderan V., Kuzhali J.V., Aruna M., 2013, Fuzzy fast classification algorithm with hybrid of ID3 and SVM, *Journal of Intelligent & Fuzzy Systems* 24 555–561, DOI:10.3233/IFS-2012-0574
- Tao D.Q., Ma L.L., Peng C., 2012, Decision tree Algorithm based on Classification Matrix [J]. *Computer Engineering and Design*, 2012, 33(6):2309-2313, DOI: 1000-7024(2012) 06-2309-05.
- Yu J.P., Huang X.M., Li K.S., 2012, Improved ID3 Algorithm based on New Attributes Selection Criterion . *Computer Engineering and Design*, 29(8):2895-2908, DOI:10.3969/j.issn.1001-3695.2012.08.024.
- Zhou L., Yan L., 2010, Decision tree algorithm using archetype abstraction and attribute classification value. *Application Research of Computers*, Vo1. 27 No. 8, 2899-2901, DOI:10.3969/j.issn.1001-3695.2010.08.023.
- Zhang Q.H., Xing Y.K., Zhou Y.L., 2011, The Incremental Knowledge Acquisition Algorithm Based on Granular Computing. *Journal of Electronics & Information Technology*, 33(2):435-441, DOI:10.3724/SP.J.1146.2010.00217.