Mutiple Driving Behavior Analysis

Mao Tiezheng
Institute of Computer
South China University of technology
Guangzhou, China, 510001
1150344295@qq.com

Wen Guihua
Institute of Computer
South China University of technology
Guangzhou, China.510001
1102221491 @qq.com

***Abstract*—this article hopes to learn rules of driver behavior habit, which is useful for early warning. Those rules are divided into single behavior rules and multiple behavior rules. C4.5 is used to learn single behavior rules; resolution principle and correlation principle are combined and used to learn multiple behavior rules. At last rules that have conflict are filtered by a multi-attribute compositional method. Experiments show that the obtained rules are both efficient and comprehensive.**

Keywords-C4.5;resolution principle; correlation principle; filtering rules

I. INTRODUCTION

Driving behavior analysis is becoming more and more important. Research area about it is as followed: 1) Car following model based on mathematic model[1][2]; 2) Preview optimal curvature model[3] ;3) Analysis about the relation between driving safety and driving decision [4][5][6][7]. Paper [4] introduces initiative security control; Paper [5] uses single factor method to analyze driving behavior dangerous degree. Paper [6] introduces driving behavior decision based on decision tree; Paper [7] mix Bayes model, FCM model and neural network to analyze driving behavior dangerous degree.

This paper divides driving behavior into single behavior and multi-behavior. Firstly, we use resolution principle to extract multi-behavior rules. Then, we use C4.5 to extract the single behavior rule. At last, conflict resolution is solved in a multi-attribute compositional method. By this way, we can study driver's usual behavior, and warn him when his behavior is unusual.

II. REALIZE METHOD AND DATA STRUCTURE

Driving Simulator is used in our experiment which can set the experiment condition and safer. Data structure is as followed: 5 behavior attributes are used to describe driver's behavior. 10 status attributes describe driver's behavior.

y_1 is steering wheel signal：When steering wheel is turn rapidly, y_1=1, others y_1=0; y_2 is accelerator signal, the value is set to [0,1] according to accelerator status; y_3 is clutch signal，if clutch is on, y_3=1; y_4 is brake signal，if brake is on then y_4=1;

x_1: If side front car and our car is in safety distance, it is set to 0, else 1.x_2：If side rear car and our car is in safety distance, it is set to 0, else 1. x_3：If front car and our car is in safety distance, it is set to 0, else 1. x_4：If rear car and our cars are in safety distance, it is set to 0, else 1.x_5 is the speed difference between side front car and our car. x_6 is the speed difference between side rear car and our car. x_7 is about speed difference between front car and our car. x_8 is about speed difference between rear car and our car. If the speed difference decreases, x_5-x_8 is set to -1,If no changes x_5-x_8 is set to 0,if increases, set to 1.x_9 is traffic sign, if there is traffic sign, x_9=1 else x_9=0. x_{10} is gear signal, the value can be 0, 0.2, 0.4, 0.6, and 0.8,1.

III. MULTI-BEHAVIOR RULE

A. Definition

Definition 1：T is a two dimension table set, T[i][j]is the element of T, T[i][*] means the row i of T，T[*][j] is column j of T, δ（T）is the row number.

Definition 2：$\Re$ (T, K) is a function to get the subset of T that satisfies condition K.

Definition 3：ω(T) is a column name set of T, ξ (T, W) can get sub-column that satisfies condition W, W$\subset$ ω(T).

Definition 4：$\aleph$（A）is a combination of set A.

Theorem1：(resolution principle), if C_1=P $\vee$ R and C_2=~P $\vee$ Q are true, then C_{12}= R $\vee$ Q is true. It means when P and ~P appear in those different formulas, the result has nothing to do with P and ~P.

Theorem 2: Relation between set A and B is

$$\frac{P(A\cap B)*P(\sim A\cap \sim B)-P(\sim A\cap B)*P(A\cap \sim B)}{\sqrt{P(A)*P(B)*P(\sim A)*P(\sim B)}} \qquad (1)$$

B. Relation analyse and attribute resolution

1) Get behavior combination: for $\forall$ A$\in$ $\aleph$ (ω(T，Y)), and i=δ (A)we get sub-column A[1],A[2]…..A[i].

2) To enhance accuracy, we firstly calculate the attribute that is more relative to behavior.

a) For $\forall$ attribute $x_i \in$ ω(T，X)，(i=1..9), B=ξ (T, A+ x_i).

b) Get the value of attribute A[1], A[2]…A[i] {w_1,w_2,….w_i}, value of attribute x_i is {x_{i1},x_{i2}….x_{ij}}

c) From formula (1) ,we can calculate

$$f(x_i,A)=\sum_{k=1}^{j}\frac{\delta(x_i=x_{ik},w_1=1,..w_i=1)*\delta(x_i\neq xi_1,w_1\neq 1..w_i\neq 1)}{\sqrt{\delta(x_i=x_{ik})*\delta(w_1=1.w_i=1)*\delta(x_i\neq x_{ik})*\delta(x_i\neq x_{i1},w_1\neq 1,..w_i\neq 1)}}/j \qquad (2)$$

© the authors

d) Repeat b)-d) until all attributes related with set A are calculated. The result set is x_j

e) For $x_j \in \aleph$（X）, we get T`=ξ(T, x_j+A)

f) get sub-row of T” = $\Re$ *(T, K).(K is A[1]=w_1, A[2]=w_2, A[i]=w_i).*

3) Using resolution principle to calculate:

a) set H=T”[0][*]

b) Compare H[k] with T”[i][k], if they are the same then k=k+1, goto b), else H[k]= ϕ

c) Repeat b) until every row of T’’ are compared.

d) Repeat b)-c) until every row of T’’ are compared.

4) Get all attribute which is not empty, in this way to get the rules.

5) Repeat 1)- 5) to get all behavior rules.

C. Example

TABLE I RAW DATA SET

No	x_1	x_2	x_3	X_4	x_5	x_6	x_7	x_8	x_9	x_{10}	y_1	y_2	y_3	Y_4
1	0	1	1	0	1	-1	0	1	0	0.4	1	0	0	0
2	1	1	0	0	1	0	1	0	1	0.6	1	0	1	1
3	0	0	1	1	0	-1	-1	-1	0	0.6	1	1	0	0
4	0	1	1	0	0	1	1	1	1	0.2	1	0	1	1
..														

1) Table I is raw data set, for us to get combination of behavior.For example, if we should definite combination behavior of y_1 and y_3, we select sub-column y_1 and y_3.

2) Using formula (1) to calculate the most relative attributes. As result, we get a result that x_2,x_3,x_6,x_7,x_8,x_9 are relative to y_1 and y_3.

3) Get column x_2,x_3,x_6,x_7,x_8,x_9,y_1,y_3 to create new table T’. where y_1=1,y_3=1, we get sub-table T’’(TABLE II）

TABLE II T’’ IS SINGLE BEHAVIOR SUB-TABLE OF T’

No.	X_2	X_3	X_6	X_7	X_8	X_9	Y_1	Y_3
2	1	1	0	1	0	1	1	1
4	1	0	1	1	1	1	1	1

4) Attribute x_2 value in row 1 and row 3 is different, this mean no matter what the value of x2 is, y1=1,y3=1 always set up. By the same way, attribute x_4 x_7 and x_8 can be filtered. The result rule is: if(x_2=1) and (x_7=1) and (x_9=1) then (y_1=1) and (y_3=1)

IV. USING C4.5 TO CACULATE SINGLE BEHAVIOR

C4.5 gains ratio from information to calculate the decision tree and then turn the tree into rules. ID3 can lead to decision tree too, but in ID3, the attribute which has multiple values has much bigger entropy. In our case, because attribute x_5-x_{10} has multiple value, we use C4.5 instead of ID3.In this way we can get trees about y1-y4 more accurately.

The step is as followed:

1) T is training set, attribute C_i has m value, C={C_1,C_2,….,C_m}.

2) If the frequency of C_i is p_i(i=1,2,…m),then entropy of set T is

$$Entropy\ (S) = \sum_{i=1}^{m} p_i \log_2 p_i \qquad (3)$$

3) If we divide the training set T with attribute A, information gain-ratio based on A is calculated as follow:

a) If attribute A has K different values, then A will divide set S into K sub-set {S_1,S_2,…,S_k}, information entropy of A is

$$Entropy_A(S) = \sum_{i=1}^{k} \frac{|S_i|}{|S|} Entropy\ (S_i) \qquad (4)$$

Here $|S_i|$ and $|S|$ is the count of set S_i and S.

b) Information gain of A is:

$$Gain(A)=Entropy(S)-Entropy_A(S) \qquad (5)$$

c) Split information of A is:

$$SplitE(A)= \sum_{i=1}^{k} \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|} \qquad (6)$$

d) Information gain ratio is:

$$GainRatio(A)=Gain(A)/ SplitE(A) \qquad (7)$$

4) After calculating information gain ratio of each attribute, we select the attribute that has biggest information gain-ratio as the root of decision tree. Then divide the tree by this attribute value. By this way, set T is divided to sub-set T_1,T_2,…T_m, for each sub-set repeat 2)-4), till every attribute don’t have any child in this sub-set.

For example, we get Figure 1 decision tree of y_4: The tree in Figure 1 can be changed to rules.

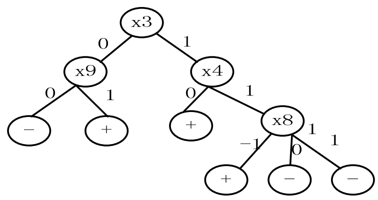

Figure 1. decision tree

- If (x3=0) and (x9=0) then y4=0
- If (x3=0) and (x9=1) then y4=1
- If (x3=1)and (x4=0) then y4=1
- If (x3=1) and (x4=1) and (x8=-1) then y4=1
- If (x3=1) and (x4=1) and (x8=0) the y4=0
- If (x3=1) and (x4=1) and (x8=1) the y4=0

V. RULE CONFLICT RESOLUTION RESOLVE

In the experiment we find some conflicts between rules.

1) Decision tree some time become partly best instead of overall best, we call this overfitting. The rules fit better in training set than in test set.

2) Rules got from different method perhaps conflict.

a) When the rule has same conclusion, as it’s

© the authors

prerequisite is different ,such as: “if p and q then r” ; “if p and ~q then r” ,we can get rule “if p then r”。

b) “if p then q” and “if p then ~q”,there is contradiction.

c) If two rules have same conclusion, but one’s prerequisite is included in the other’s, we call rule 1 is sub-rule of rule 2.

We show the way to solve the problem in 1),b) and c) through an example. We use the method based on multi-attribute compositional as table 4 shows: There are 10 records in test set in table 4, and now we get a rule: if x1=0 and x3=1 and x4=1 then y1=1; in table 4,“+” represent positive record (record fitting this rule), “-“ represent negative record, there are 3 positive records, and 2 negative records, so the rule’s precision is 3/5, if we put off x_1 from the rule, the rule becomes ”if x_3=1 and x_4=1 then y_1=1”, the new rule’s precision turns to 4/7. In table III, after putting off x_3 and x_4, we get the best precision（5/7）,then at last the rule become “if x_1 then y_1”.

VI. EXPERIMENTAL RESULTS

We have collected 100-1000 action and status records from drivers and divide them equally to a train set and a test set. Experiment result base on ID3, Bayes and our method shows as Figure 2.

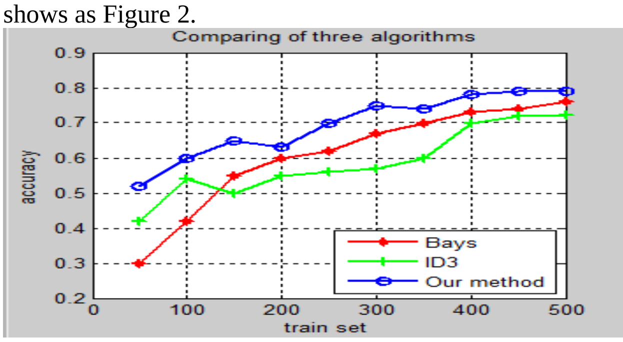

Figure 2. accuracy rate

The accuracy of Bayes is lower when there are less test records. But its curve is smoothing, and the accuracy increases rapidly.

ID3’s curve fluctuates. And with the train set increase, accuracy increases.

Although our method also curve fluctuates, it is better than ID3 on precision. It performs better than Bays when there is less data. Our method is the best overall.

VII. CONCLUSION

The innovations of this paper are:

1) Using C4.5 to solve attribute problem, this method is more accurate than ID3.

2) Using resolution principle to extract multi-behavior rules

3) Use method based on multi-attribute compositional to solve conflict between rules.

ACKNOWLEDGMENT

This paper is funded by national natural science foundation of China(60973083)

REFERENCES

[1] Cao Baogui, Yang Zhaosheng. An Improved Car- Following Dynamic Model [J]. Journal of South China University of Technology, 2011.39(10):96-99.

[2] Wang Xiaoyuan,Wang Lei.Coperative simulation Model for Driving Behavior Based on cognition Activity Chain [J].Journal of Southwest University, 2007.42(2): p238- 242.

[3] Guo Konghui . Driver Model Parameter Identification of the Driver-vehicle-road Closed-loop System [J] . Automotive engineering, 2002, 24(1) : 20-24.

[4] Chu Ying,Xiao Xianqiang,Zhu jiacheng. Study on Automotive Active Safety Technology Based on Driving Behavior and Intention.[J],Machinery Design &Manufacture. 2011.1:266-269.

[5] Guo Zizheng,Chen Chongshuang,Wang Xin. Risk Identification for Driving Behaviors based on Bayesian Discrimination [J]. Journal of Southwest University, 2009.44(5):772-775.

[6] Wang Xiaoyuan,Yang Xinyue. Study on Decision Mechanism of Driving Behavior Based on Decision Tree [J]. Journal of System Simulation, 2008.20(2): 415-419

[7] Guo Zizheng,Chen Chongshuang,Wang Xin. Identification of risk states in driving behavior based on evidence theory [J], Journal of Southeast university, 2010.40(4):866-870

TABLE III RULE CONFLICT RESOLVED

x_1	x_2	x_3	x_4	x_5	y_1		Cut x_1	Cut x_3	Cut x_4	Cut x_1x_3	Cut x_3x_4	Cut X_1x_4
0	0	1	1	0	1	+	+	+	+	+	+	+
0	1	0	1		1			+		+	+	
0	1	1	1		0	-	-	-	-	-	-	-
0	0	1	1	1	1	+	+	+	+	+	+	+
0	0	1	0		0				-		-	-
0	0	1	1	0	0	-	-	-	-	-	+	-
0	1	1	1		1	+	+	+	+	+	+	+
1	1	1	1		1		+			+		+
1	1	0	0		0							
1	0	1	1		0		-			-		-
						3/5	4/7	4/6	4/6	5/8	5/7	4/8

© the authors