

# 決定木による自動車運転時の危険予知に関する知識獲得

## Knowledge acquisition by decision trees from dangerous scenes at car-driving

竹植 正和

Masakazu Takeue

広島市立大学大学院

情報科学研究科知能工学専攻

Email: takeue@cm.info.hiroshima-cu.ac.jp

高橋 健一

Kenichi Takahashi

広島市立大学大学院

情報科学研究科知能工学専攻

Email: takahasi@hiroshima-cu.ac.jp

**Abstract**—The objective of this study is to obtain predictive rules and knowledge regarding dangerous scenes that drivers encounter when driving a car. First, dangerous objects and elements in scenes included in still and moving pictures are extracted as classes and attributions. Then, decision trees are made using those data, and their accuracy is examined through the K-cross validation method. In order to improve the accuracy, a data sort algorithm considering distributions of attributes and classes is proposed. Experiments are conducted to confirm that effective rules and knowledge can be obtained.

### I. はじめに

平成 16 年度から交通事故の発生件数は減少傾向にあるが、平成 23 年度は 69 万 2 千件以上と 1 日に 1900 件近く発生している[1]。また、負傷者に関しては約 85 万 4 千人、事故発生から 30 日以内の死者は 5450 人となっている。その対策として、ドライバーの安全意識や運転技術の向上を目指す啓蒙活動や自動車の安全装置の開発が行われている。啓蒙活動に関しては自動車学校での講習や、街頭でのポスターの設置などが行われたりしている。また、近年ではドライブレコーダーを装着している車も増えており、運転中の走行場面を録画でき、事故の瞬間を録画することができる。その映像をウェブサイト上にアップロードして一般のユーザ間で情報を共有できる。また、JAF(日本自動車連盟)では、JAFMATE という月刊の会員誌で交通安全に関する様々な情報や危険予知トレーニングに関する問題を掲載している。自動車の安全装置に関しては、これまで任意装着であった ESC(横滑り防止装置)と BAS(ブレーキアシストシステム)を 2012 年 10 月 1 日以降の新型生産の登録車に義務化することが決まった[2]。また、富士重工業が開発した EyeSight(アイサイト)が注目を集めている。これは障害物や前方車に衝突する危険があると自動的にブレーキをかけて車を停止させるシステムである。他にも、車間距離警報や、車線逸脱警報、ふらつき警報が実用化されている。しかし、これらの先進技術はステレオカメラや

ミリ波レーダーを利用しており、製造や取り付けにコストがかかる。

本研究は自動車運転時の様々な事例における危険予知に関する規則や知識を獲得することを目的とする。このため、まずカメラを用いて撮影された危険予知問題の静止画や、ドライブレコーダーなどで撮影された動画において、事故が起こる 3 秒前の画像を用いて、危険な箇所と、危険な箇所を構成する周囲の要因を属性として抜き出す。次に、データマイニングソフトである Weka[3]の J48 アルゴリズムを用いて決定木学習を行い、決定木の精度を K-交差検定法を用いて検証する。K-交差検定法では、トレーニングデータを K 個に分割し、K-1 組のデータを用いて決定木を生成し、残りの 1 組のデータで精度を検証するため、事例の属性分布が重要となる。そこで、事例の並び順を変えるソートプログラムを作成して、学習実験、精度検証を行う。実験により得られた決定木からそれぞれの場面に共通するルールを抽出する。

### II. 決定木

決定木とは、意思決定や物事の分類を多段階で繰り返し実行する場合、その多段の分岐過程を階層化して樹形図で表現したグラフ表現である[4]。本研究で使用する決定木生成アルゴリズム J48 は、C4.5 を Weka に実装したアルゴリズムである。C4.5 は ID3(Iterative Dichotomiser 3)を改良・発展させたもので、最も広く用いられている手法である。ID3 は利得基準(Gain)が用いられているが、C4.5 では利得比基準(Gain Ratio)を用いている。これにより多くの属性値を持つ属性が高い評価をされてしまうという ID3 の欠陥が修正されている。また、決定木学習では、学習を重ねていくと訓練事例に対して木が細かく分岐し、訓練事例内の例外的な値や誤りに対して学習しすぎて、予測精度が悪化するという問題がある。これを過学習、またはオーバーフィッティング(over fitting)という。この問題を解決する手法を枝刈り(pruning)という。C4.5 では pessimistic pruning[5]を用いて枝刈りを行う。これは、葉であるノードに含まれる  $n$  個のデータを母集団から取り出したサンプルとみなし、サンプルのエラー率から母集団のエラー率を統計

的に推定し、その推定値に基づいて枝刈りを行う手法である。また、枝刈りの基準を決める値として信頼度(confidence factor)を用い、この値が小さいほど多く枝刈りが行われる。

K-交差検定法は、生成された決定木の評価方法である。この手法は事例集合 S を K 個に分割し、K-1 個を訓練事例として学習を行い、残り 1 つをテスト事例として K 回評価を行う方法である。そして、得られた K 回の評価結果を平均して 1 つの評価を得る手法である。そのため、分割する際に事例中の属性の分布が均一である方が、精度が良くなると考えられる。

本研究で使用した Weka は世界中で使われているデータマイニングツールで、機械学習の研究者によって Java 言語で開発された。Weka は前章で紹介した C4.5 以外にもベイズ分類器、ニューラルネットワーク、最近傍法等の様々なデータマイニングアルゴリズムが使用できる。

### III. 入力データ

本研究では事例として、JAFMATE 危険予知問題集、JAFMATE ドライブレコーダー動画投稿ページ、独立行政法人自動車事故対策機構、HONDA 危険予知トレーニング、神奈川県警察、交通安全マップから合計 200 種類の危険予知に関する静止画像・動画を収集した。これらのデータから、属性・クラスを抽出し、Weka を用いて決定木を生成する。なお、動画を事例とする場合、事故を起こす瞬間の状況を解答画像、解答画像から 3 秒前の状況を問題画像とした。

Weka にデータを入力するため、属性と属性値、クラスを決定する必要がある。本研究では収集した画像内の危険な箇所を構成する要素を抜き出し、属性とする。そして、抜き出した属性を用いて属性値を決定する。また、危険予知問題の解答画像における危険な箇所をクラスとする。

本研究で使用した属性・属性値の一部とクラスをそれぞれ表 1、表 2 に示す。また、図 1 から図 3 に車の位置に関する属性を示す。

例えば、図 4 のような問題画像[6]が与えられた場合には、「天気」と「歩行者」、「対向車 2 の行動」、「道路」が属性となる。また、属性値は、「天候」に関しては「曇り」、「歩行者」は「左歩道の歩行」、「対向車 2 の行動」は「直進」、「道路」は「カーブ」となる。そして、図 5 の解答画像[6]から自動車を運転する際に注意すべき点をクラスとして抽出する。この解答画像の場合はカーブ先に注意しなければならないので、図 5 におけるクラスは「カーブの先」となる。

表 1 使用した属性・属性値(一部)

属性	属性値
前方車 1 の行動	直進, 左折, 右折, 停止
対向車 1 の行動	直進, 左折, 右折, 停止
道路	直線, カーブ, 上り坂
歩行者	右歩道の歩行, 停止
天候	晴れ, 雨, 雪, 曇り

表 2 使用したクラス

前方車	後方車	対向車	歩行者
自転車	カーブの先	上り坂	交差車
標識	自車の死角		



図 1. 前方車, 隣接車, 後方車の位置関係



図 2. 対向車の位置関係

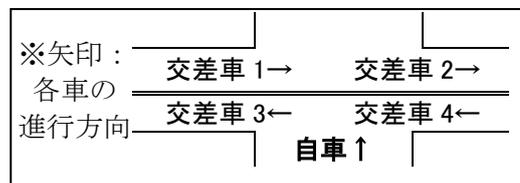


図 3. 交差車の位置関係



図 4. JAFMATE の危険予知問題



図 5. 危険予知問題の解答

#### IV. 事例ソートプログラム

先行研究[7]では決定木による分類の最高精度が73.4%であった。これは、属性数28、クラス数10、精度71.5%を示した入力データを、 $K$ -交差検定法の分割数  $K$  にあわせてクラス分布のみを考慮して人手により事例をソートし、 $K$  が10、枝刈り率  $C$  が0.19のときに得られた精度である。そこで、本研究ではクラスの分布だけでなく、属性の分布も考慮してデータをソートする方法を考えた。この方法では異なる属性値に違う評価値を与え、事例ごとに評価値を合計する。そして、合計した評価値をもとに事例を昇順にソートし交差検定数別に事例を振り分ける。以下に、 $K$  個のグループの中に同じ属性値またはクラスを可能な限り同じように分散させる事例ソートアルゴリズムを示す。

1.  $n$  個の属性(便宜上、 $n$  番目の属性をクラスの集合とする)を持つ、 $X$  個の事例の集合  $S$  を生成する。 $i$  番目の事例の属性値を  $a_{i0}, \dots, a_{in}$  (便宜上、 $n$  番目はクラス)とし、その属性の評価値(初期値=0)を同様に  $e_{i0}, \dots, e_{in}$  とする。
2.  $j$  番目の属性について考える、 $i$  番目の事例の  $j$  番目の属性の属性値  $a_{ij}$  が、0から  $i-1$  番目までの事例で、 $j$  番目の属性の属性値  $a_{kj}$  ( $k=0, \dots, i-1$ ) のいずれかと同じ値を持つならば、対応する評価値  $e_{kj}$  と同じ値を  $e_{ij}$  に代入する。もし、同じ値を持たないならば、 $e_{0j}$  から  $e_{i-1j}$  の中の最大値に1を加えた値を  $e_{ij}$  に代入する。この操作をすべての事例について行う。
3.  $i$  番目の事例について、事例評価値  $E_i$  を 
$$E_i = \sum_{j=0}^n e_{ij}$$
 のように計算する。この操作をすべての事例について行う。
4. 事例評価値  $E$  をもとに事例を昇順に並び替える。
5.  $i$  番目の事例を交差検定数  $K$  にあわせて  $i \bmod K$  番目のグループに割り振り、交差検定に用いる  $K$  個のグループを生成する。

このアルゴリズムにより事例をソートし実験を行う。

#### V. 実験と考察

先行研究[7]で使用した属性28個、クラス10個の事例200個のクラスの分布を考慮していない入力データにおいて、同じ属性値を持つ事例が160個以上ある属性を削除して新たな入力データを作成した。事例の属性を少なくすることにより無駄な木の分岐をなくすことができ、小さい木が出力される可能性が高くなると考えられる。このデータを前述の事例ソートアルゴリズムにより、事例を並び替え、枝刈り率  $C=0.25, 0.16, 0.07$ ,  $K$ -交差検定において  $K=4, \dots, 10$  に変化させて実験を行った。その結果、先行研究において73.5%であった精度が最大で  $K=8$ ,

$C=0.25, 0.16$  のときに74.0%に向上した。このことより、先行研究で使用した属性の中に必要のない属性が含まれており、結果として精度を悪化させていたことがわかった。そのうえ、先行研究で8であった決定木の高さが7になった。表3に、実験の結果を示す。

表3 生成された決定木の精度(%)

枝刈り率 $C$ 交差検定数 $K$	$C=0.25$	$C=0.16$	$C=0.07$
$K=4$	68.5	66.5	64.5
$K=5$	67.5	67.0	64.0
$K=6$	70.5	70.5	65.5
$K=7$	68.0	69.0	66.0
$K=8$	74.0	74.0	67.0
$K=9$	71.0	71.0	68.0
$K=10$	69.0	69.5	67.0

次に、誤分類された事例を先程生成された決定木から探し、削除した属性のみで決定木を作成する。このことにより、属性値の数に偏りのある属性でも決定木の作成に重要である属性を見つけることができる。そして、削除した属性のみで作った決定木に現れた属性を上記データに戻して、ソートプログラムに入力して事例をソートし、再度実験をする。表4に最も精度の高かった実験の結果を示す。

表4 生成された決定木の精度(%)

枝刈り率 $C$ 交差検定数 $K$	$C=0.25$	$C=0.16$	$C=0.07$
$K=4$	69.0	68.5	66.0
$K=5$	69.5	66.0	64.0
$K=6$	71.0	71.0	67.5
$K=7$	76.0	75.0	74.0
$K=8$	70.5	70.0	67.5
$K=9$	72.0	73.0	69.5
$K=10$	71.5	70.5	68.5

表4に示すように、 $K=7, C=0.25$  のとき最も高い精度76.0%を得ることができた。これは、作成したソートプログラムと、決定木の生成に必要な属性が減って学習が進んだためと考えられる。なお、このとき使用した属性は以下の表5のようになった。

表5. 決定木作成に使用した属性

前方車2 の行動	前方車5 の行動	隣接車1 の行動
後方車1 の行動	対向車1 の行動	対向車2 の行動
対向車3 の行動	交差車1 の行動	歩行者 の行動
自転車の行動	道路の種類	自車の行動
信号	自車の走行車線	

先行研究で最高精度のとき、作成された決定木の高さは8であったが、本研究では7に減少した。このことにより、ルールの抽出が容易になったと考えられる。また、決定木から道路の種類が交差点である場合に注目すべき点は「交差車」や「自転車」、直線の道路の場合では、「前方車」が多く葉の部分に現れていたため、これらに注意すべきということがわかった。そのほか、自動車を運転している時に、始めに注目すべき点は「道路」の種類で、道路の種類が「交差点」の場合は、「前方車5の行動」つまり、直前の車の動きに注意するということがわかった。以下に本研究で得られたルールの一部を示す。

1. 片側1車線の交差点で、自車が直進している際には交差車に注意。
2. 交差点で、直前の前方車が左折していれば、前方車に注意。
3. 直線道路で、自転車が右車道を走行していれば、対向車に注意。
4. 直線道路ならば、前方車に注意。
5. カーブならば、カーブの先に注意。
6. 片側1車線の交差点で、自車が減速している場合には自転車に注意。
7. 直線道路で、歩行者が右歩道を歩行していて、直前の前方車が停止していれば、標識に注意。
8. 直線道路で、歩行者が右歩道に停止していれば、標識に注意。

1~5のルールは、理解しやすいルールであり、自動車を運転する際に有用である。6~8のルールは、決定木に現れた特殊なルールである。このように、決定木の中には特殊なものも含まれているが、有用なルールの方が多く見られる。これらの抽出されたルールより、それぞれの事例に共通するルールや規則がある程度発見できたとと言える。

## VI. おわりに

本研究は自動車運転時の様々な事例における危険予知に関する規則や知識を獲得することを目的として、危険予知問題の静止画や動画において、危険な箇所と、危険な箇所を構成する周囲の要因を属性として抜き出し、属性とクラスの分布を考慮したソートプログラムにより、データを並び替え、決定木学習を行った。そして得られた決定木からそれぞれの場面に共通するルールを抽出した。

先行研究では全200種類の事例、28種類の属性、10種類のクラスを用いて、クラス分布を考慮した場合は73.5%の精度であり、クラスの分布を考慮しない場合は71.5%の精度であった。これに対して本研究では、属性とクラスの両方の分布を考慮するために、事例をソートするプログラムを作成して、実験では、属性値の数に偏りのある属性を削除したところ、先行研究を超える74.0%の精度が得られた。これは、属性

値の数に偏りのある属性の中に決定木を作成する際に必要のない属性があったためと考えられた。削除した属性の中にも決定木作成に重要となる属性がある可能性があるため、削除した属性を一部戻して再度実験を行った。ここでは、最高で76.0%の精度を得ることができた。76.0%を示した決定木より、交差点においては自転車や交差車に注目して、直線道路においては歩行者や前方車に注目するという知識が得られた。これらの実験結果から、危険予知に関するある程度の知識を発見できたといえる。

本研究における今後の課題として、精度の向上のために属性数を増やしたり、属性値・クラスをまとめたりすることが考えられる。また、ある程度の知識の発見ができたので、これらの知識を応用したシステムの構築などが挙げられる。

## 参考文献

- [1] “平成24年度交通安全白書 第1部道路交通事故の長期的推移”，共生社会政策統括官。  
[http://www8.cao.go.jp/koutu/taisaku/h24kou\\_haku/pdf/gaiyo/gen1\\_1\\_1.pdf](http://www8.cao.go.jp/koutu/taisaku/h24kou_haku/pdf/gaiyo/gen1_1_1.pdf) (2012/7/1 現在アクセス可能)
- [2] 「道路運送車両の保安基準の細目を定める告示」等の一部改正について～横滑り防止装置及びブレーキアシストシステムの義務化～，国土交通省 報道発表資料。  
[http://www.mlit.go.jp/report/press/jidosha07\\_hh\\_000070.html](http://www.mlit.go.jp/report/press/jidosha07_hh_000070.html)  
(2012/6/27 現在アクセス可能)
- [3] “Weka3 - Data Mining with Open Source Machine Learning Software in Java”，  
<http://www.cs.waikato.ac.nz/ml/index.html>  
(2012/6/27 現在アクセス可能)
- [4] 元田浩，津本周作，山口高平，沼尾正行：データマイニングの基礎，株式会社オーム社，平成20年第1版。
- [5] J.Ross Quinlan “C4.5 Programs for Machine Learning”，Morgan Kaufmann,1993
- [6] 岩越和紀 “JAF MATE2009年6月号”  
株式会社 JAF MATE 社，pp31-33，2009年6月1日発行
- [7] 村上祐大，高橋健一，上田祐彰：“決定木を用いた自動車の運転時における危険予知に関する基礎研究”，平成22年度(第61回)電気・情報関連学会中国支部大会，No.1236, 2010