

# 人工知能 中間レポート

学籍番号：045713C

氏名：大城和也

提出日：平成 18 年 1 月 31 日

# 1 C4.5 プログラムにより, 課題の例題を実行し, 結果を得る.

プログラムをコンパイルし, Sample のデータで決定木を作成してみた.

```
-----
[nw0413:report-mid/R8/Src] j04013% ./c4.5 -f ../../Sample.1/sample

C4.5 [release 8] decision tree generator      Sun Jan 29 16:50:19 2006
-----

Options:
  File stem <../../Sample.1/sample>

Read 9 cases (3 attributes) from ../../Sample.1/sample.data

Decision Tree:

eye-color = brown: - (4.0)
eye-color = blue:
|  hair-color = blond: + (2.0)
|  hair-color = black: - (2.0)
|  hair-color = red: + (1.0)

Tree saved

Evaluation on training data (9 items):

      Before Pruning      After Pruning
-----
Size      Errors      Size      Errors      Estimate
6      0( 0.0%)      6      0( 0.0%)      (43.6%)  <<
-----
```

テストデータを評価してみると以下のようになった.

```
-----
[nw0413:report-mid/R8/Src] j04013% ./c4.5 -f ../../Sample.1/sample -u

C4.5 [release 8] decision tree generator      Tue Jan 31 14:39:51 2006
-----

Options:
  File stem <../../Sample.1/sample>
  Trees evaluated on unseen cases

Read 9 cases (3 attributes) from ../../Sample.1/sample.data

Decision Tree:

eye-color = brown: - (4.0)
eye-color = blue:
|  hair-color = blond: + (2.0)
|  hair-color = black: - (2.0)
|  hair-color = red: + (1.0)

Tree saved

Evaluation on training data (9 items):

      Before Pruning      After Pruning
-----
Size      Errors      Size      Errors      Estimate
6      0( 0.0%)      6      0( 0.0%)      (43.6%)  <<
-----
```

Evaluation on test data (2 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
6	2(100.0%)	6	2(100.0%)	(43.6%) <<

(a)	(b)	<-classified as
1	1	(a): class + (b): class -

## 2 自身で分類問題を設計し，その問題に対する決定木を生成する．

### 2.1 問題意図

今回は F1 の GP のレースにおけるポイントを獲得できる車体 (レースで 6 位内) を分類する木を作成することにした．実際，F1 自体に興味があったわけではないけど，レース系をやってみたかったのでこのテーマにした．

### 2.2 属性及び属性値とクラス

属性は以下のような設定となっている．

- 天気 (Weather) : 晴れ (s), 雨 (r), 曇り (c)
- 路面状態 (Road) : ウェット (w), ドライ (d)
- 気温 (Temperature) :  
30 度以上 (hot), 20 度<(mid)>30 度, 20 度以下 (cold)
- タイヤ (Wheel) : プリヂストン (B), ミシュラン (M)
- ピットイン数 (Stop) : 1, 2, 3, 4
- グリッド位置 (Grid) : [1to5], [6to11], [12to17], [18to22]
- チーム名 (team) :  
FER, BMW, MCL, REN, SAU, JOR,  
JAG, BAR, MIN, TOY, ARR

これらを属性としてそれらの最終判断となるクラスは順位が 6 位内のときには '+' をそれ以外なら '-' とした．

## 2.3 C4.5 プログラムの実行結果

トレーニング用のデータ, テストデータ, 属性設定データなどは  
<http://www.ie.u-ryukyu.ac.jp/j04013/file/F1.tgz> に置く.

これらのデータを用いて C4.5 プログラムを動かす.

```
*****
[nw0413:report-mid/R8/Src] j04013% ./c4.5 -f ../Data/F1

C4.5 [release 8] decision tree generator      Tue Jan 31 00:17:43 2006
-----

Options:
  File stem <../Data/F1>

Read 66 cases (7 attributes) from ../Data/F1.data

Decision Tree:

Grid = [18to22]: - (10.0)
Grid = [1to5]:
| Road = w: - (3.0)
| Road = d:
| | team = FER: + (9.0)
| | team = MCL: - (1.0)
| | team = REN: + (0.0)
| | team = SAU: + (0.0)
| | team = JOR: + (0.0)
| | team = JAG: + (0.0)
| | team = BAR: - (1.0)
| | team = MIN: + (0.0)
| | team = TOY: + (0.0)
| | team = ARR: + (0.0)
| | team = BMW:
| | | Weather = s: - (3.0/1.0)
| | | Weather = r: + (0.0)
| | | Weather = c: + (3.0)
Grid = [6to11]:
| team = FER: + (0.0)
| team = BMW: + (1.0)
| team = MCL: + (3.0/1.0)
| team = REN: + (3.0/1.0)
| team = SAU: - (3.0)
| team = JOR: + (2.0)
| team = JAG: + (0.0)
| team = BAR: + (1.0)
| team = MIN: + (0.0)
| team = TOY: - (3.0)
| team = ARR: + (0.0)
Grid = [12to17]:
| team = FER: - (0.0)
| team = BMW: - (0.0)
| team = MCL: + (1.0)
| team = REN: - (2.0)
| team = SAU: - (1.0)
| team = JOR: - (3.0)
| team = JAG: - (2.0)
| team = BAR: - (5.0)
| team = MIN: - (1.0)
| team = TOY: - (3.0)
| team = ARR: + (2.0/1.0)

Simplified Decision Tree:

Grid = [12to17]: - (20.0/3.7)
Grid = [18to22]: - (10.0/1.3)
Grid = [1to5]:
| Road = w: - (3.0/1.1)
| Road = d: + (17.0/5.9)
Grid = [6to11]:
| team = FER: + (0.0)
```

```

| team = BMW: + (1.0/0.8)
| team = MCL: + (3.0/2.1)
| team = REN: + (3.0/2.1)
| team = SAU: - (3.0/1.1)
| team = JOR: + (2.0/1.0)
| team = JAG: + (0.0)
| team = BAR: + (1.0/0.8)
| team = MIN: + (0.0)
| team = TOY: - (3.0/1.1)
| team = ARR: + (0.0)

```

Tree saved

Evaluation on training data (66 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
43	4 ( 6.1%)	18	8 (12.1%)	(31.5%) <<

+++++

## 2.4 決定木の分析

まず、Decision Tree より

決定木の始めに分けられているのは Grid で、このことより今回のデータを分けるには Grid が一番大事ということが判る。Grid が大きい時には小さい時と比べれば 6 位に入ることは少ない。特に 18 から 22 においてはその条件だけで負けると決めている。

Grid が 1 から 5 の場合にはさらに路面状態 (Road) で分岐している。ウェット (w) の場合には負けを意味し、ドライの場合にはそれぞれのチームによって 6 位内に入るか否かが決まっている。特にフェラーリ (FER) の値は高い。なお、(0.0) となっているのはルールによって生成されはしたが、それに当てはまるデータがなかったことを表している。

上記の Grid が 1 から 5 の場合、分岐の後、BMW はさらに天候によって左右する。晴れでは負けることが多いようだ。

他の Grid の時にはそれぞれはチームによって分けられている。それぞれのチームによってクラスが '+' になる率は変化している。

次に Simplified Decision Tree を見てみる。

これは先の決定木を枝切りした物であるが、簡潔になっているのが見て判る。木としては Grid で別れ、Grid が 1 から 5 の時で路面によって変化する。ドライの時に '+' となるがこのとき 17 個のうち 5.9 個は間違えることがある。

Grid が 6 から 11 の時にはチーム名 (team) によって変化する。これらの状態から見て全体的に '+' の判断は間違えることが多く、 '-' の判断は間違いが少ないと思われる。

## 2.5 テストデータによる評価

```
+++++
[nw0413:report-mid/R8/Src] j04013% ./c4.5 -f ../Data/F1 -u

C4.5 [release 8] decision tree generator      Tue Jan 31 00:34:46 2006
-----

Options:
  File stem <../Data/F1>
  Trees evaluated on unseen cases

Read 66 cases (7 attributes) from ../Data/F1.data

Decision Tree:

Grid = [18to22]: - (10.0)
Grid = [1to5]:
| Road = w: - (3.0)
| Road = d:
| | team = FER: + (9.0)
| | team = MCL: - (1.0)
| | team = REN: + (0.0)
| | team = SAU: + (0.0)
| | team = JOR: + (0.0)
| | team = JAG: + (0.0)
| | team = BAR: - (1.0)
| | team = MIN: + (0.0)
| | team = TOY: + (0.0)
| | team = ARR: + (0.0)
| | team = BMW:
| | | Weather = s: - (3.0/1.0)
| | | Weather = r: + (0.0)
| | | Weather = c: + (3.0)
Grid = [6to11]:
| team = FER: + (0.0)
| team = BMW: + (1.0)
| team = MCL: + (3.0/1.0)
| team = REN: + (3.0/1.0)
| team = SAU: - (3.0)
| team = JOR: + (2.0)
| team = JAG: + (0.0)
| team = BAR: + (1.0)
| team = MIN: + (0.0)
| team = TOY: - (3.0)
| team = ARR: + (0.0)
Grid = [12to17]:
| team = FER: - (0.0)
| team = BMW: - (0.0)
| team = MCL: + (1.0)
| team = REN: - (2.0)
| team = SAU: - (1.0)
| team = JOR: - (3.0)
| team = JAG: - (2.0)
| team = BAR: - (5.0)
| team = MIN: - (1.0)
| team = TOY: - (3.0)
| team = ARR: + (2.0/1.0)

Simplified Decision Tree:

Grid = [12to17]: - (20.0/3.7)
Grid = [18to22]: - (10.0/1.3)
Grid = [1to5]:
| Road = w: - (3.0/1.1)
| Road = d: + (17.0/5.9)
Grid = [6to11]:
| team = FER: + (0.0)
| team = BMW: + (1.0/0.8)
| team = MCL: + (3.0/2.1)
| team = REN: + (3.0/2.1)
| team = SAU: - (3.0/1.1)
```

```

| team = JOR: + (2.0/1.0)
| team = JAG: + (0.0)
| team = BAR: + (1.0/0.8)
| team = MIN: + (0.0)
| team = TOY: - (3.0/1.1)
| team = ARR: + (0.0)

```

Tree saved

Evaluation on training data (66 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
43	4 (6.1%)	18	8 (12.1%)	(31.5%) <<

Evaluation on test data (16 items):

Before Pruning		After Pruning		
Size	Errors	Size	Errors	Estimate
43	5 (31.2%)	18	4 (25.0%)	(31.5%) <<

(a)	(b)	<-classified as
6	1	(a): class +
3	6	(b): class -

+++++

## 2.6 考察

テストデータは2004年のレースの値を用いた。16個のデータの内、正確に分類できたのは12個、間違いの4つ内3つは'+'と判断したが実際は'-'だったという物だ。自分の考えではエラーは'-'と判断したが実際は'+'というものが多量だと思っていたが、そうはならなかった。これはランダムに取ったデータがたまたまそうだったと考えられる。さらに年度によってレーサーが変わることもあるのでそれが影響したのかもしれない。

決定木をみて判るように、決定木を作成する際に使われていない属性が多々ある。つまりこれらはデータを評価するにあたってあまり必要ない属性とも考えられる。あるいはデータを増やすことによって変化したかもしれない。実際、データが少ないときに決定木を作成したときは一分岐しかしなかったが、データを増やすことによってこのようになった。それと属性値の数にも気をつけないと分岐しにくくなることがわかった。

ここからは感想だが、今回、F1というデータを決定木で評価しようとする際、それぞれの車の値はほとんど同じで属性を持たすことが難しく、さらに優勝するチーム(選手)がほとんど決まっているということで決定木には向いていないと感じた。勝利の鍵となる選手も属性として加えることもできなかった。まあ、それでもGridが重要なことはわかったりしたので良かったです。

## 参考文献

[1] F1-Page.com ”<http://www.f1-page.com/>”

[2] The Official Formula 1 Website ”<http://www.formula1.com/>”