

# 人工知能エンジンと決定木

10NM733X 林華

## 3.1.1 知能の表現

- 80年代—エキスパートシステム
  - 保険のエキスパートシステム
  - 採掘のエキスパートシステム
  - 医療診断エキスパートシステム  
(MYCIN、CASNET、INTERNIST)
- しかし、維持や保守や管理にはコスト高
  - 新条件に対応しない
  - 知識の変化にプログラムを書き直し
  - ネットワークの発展より、データ量の爆発

## 3.1.2 知能の獲得に

- エキスパートシステムは時代遅れ
- 知識獲得するアルゴリズム—機械学習に注目
  - 利点：保守、管理コスト抑える
  - 特点：統計学の多段階層別手法
  - 統計的な相関関係より、生成した規則の精度が高い

## 3.2 タイタニックの運命

等級/立場	大人/子供	性別	生死	人数
一等乗客	大人	男性	生還	57
一等乗客	大人	男性	死亡	118
一等乗客	大人	女性	生還	140
一等乗客	大人	女性	死亡	4
一等乗客	子供	男性	生還	5
一等乗客	子供	女性	生還	1
二等乗客	大人	男性	生還	14
二等乗客	大人	男性	死亡	154
二等乗客	大人	女性	生還	80
二等乗客	大人	女性	死亡	13
二等乗客	子供	男性	生還	11
二等乗客	子供	女性	生還	13
三等乗客	大人	男性	生還	75
三等乗客	大人	男性	死亡	387
三等乗客	大人	女性	生還	76
三等乗客	大人	女性	死亡	89
三等乗客	子供	男性	生還	13
三等乗客	子供	男性	死亡	35
三等乗客	子供	女性	生還	14
三等乗客	子供	女性	死亡	17
乗務員	大人	男性	生還	192
乗務員	大人	男性	死亡	670

## 3.3 決定木

女性 ( 470 人の中 73 %) → 生還

- 1等・2等・乗務員 ( 274 人の中 93 %) → 生還
- 3等乗客 ( 196 人の中 54 %) → 死亡

男性 ( 1731 人の中 79 %) → 死亡

- 大人 ( 1667 人の中 80 %) → 死亡
- 子供 ( 64 人の中 55 %) → 死亡
  - 1等・2等乗客 ( 16 人の中 100 %) → 生還
  - 3等乗客 ( 48 人の中 73 %) → 死亡

## 3.4 生死の特徴

### 決定木から抽出されたルール

#### 死亡規則

- もし 女性 かつ 3等乗客 → 死亡(196人中54%)
- もし 男性 かつ 大人 → 死亡(1667人中80%)
- もし 男性 かつ 子供かつ3等乗客→死亡(48人中73%)

#### • 生還規則

- もし女性&(1等乗客|2等乗客|乗務員)→生還(93%)
- もし男性&子供&(1等乗客|2等乗客)→生還(100%)

## 3.5 CART モデル

- 1970 ごろから研究し始め、1980 年代前半に公開
- 説明変数を 2 進分岐させ、決定木を生成
- 分岐基準に、ジニ係数を利用
- 情報利得も利用される

## 3.6 分岐基準ージニ係数

- $A$ : 親ノード、 $C=(c_1, c_2, \dots, c_j)$ : 予測変数
- $p_{Aj} = A$  に属するオブザベーションが  $c_j$  である確率
- 不純度ージニ係数 :  $A$  における  $C$  の総分散

$$I(A) = i(p_A) = \sum_{j=1}^J p_{Aj}(1 - p_{Aj}) = 1 - \sum_{j=1}^J p_{Aj}^2$$

- 分岐基準 :  $\Delta I$  の最大化

$$\Delta I = P(A)I(A) - P(A_L)I(A_L) + P(A_R)I(A_R)$$

## 3.7 分岐基準の計算例

- ルートノートのジニ係数  $I(A)=1-(711/2201)^2+(1490/2201)^2=0.43737$

- 「性別」のジニ係数

$$I(A_{\text{男}})=1-[0.212^2+0.788^2]=0.33411$$

$$I(A_{\text{女}})=1-[0.732^2+0.268^2]=0.39235$$

- 「性別」の分岐基準  $\Delta I=0.43737-(1731/2201)\times 0.33411-(470/2201)\times 0.39235=0.09082$

- 「大人 / 子供」ジニ係数

$$I(A_{\text{大人}})=1-[0.313^2+0.687^2]=0.43006$$

$$I(A_{\text{子供}})=1-[0.523^2+0.477^2]=0.49894$$

- 「大人 / 子供」の分岐基準

$$\Delta I=0.43737-(2092/2201)\times 0.43006-(109/2201)\times 0.49894=0.0039$$