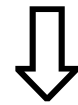


決定木との連動

清水 雄太

今回の概要

SOMと決定木の組み合わせ



分類の基準がない状態で、ツボを抑えた
変数で観測対象を効果的に分類するの
が目的

分析結果の不定性

- 伝統的な統計解析手法では、データが与えられると結果が一意に定まる。
- データマイニングの手法では、同一のデータから、同一の分析結果・解釈になるとは限らない。

決定木

- 分岐の基準を正確に定めれば、同一のデータから、同一の分析が可能である。
- 分岐の基準を変え、根源付近で分岐に利用する変数を入れ替えると、以後の枝振りが全く異なり、解釈が異なる可能性がある。

逆伝播学習型のニューラルネット

- 同一のデータで同一のトポロジーのモデルを学習させても、乱数の数が異なり、重みの初期値が異なると、学習された結果の重みは、学習ごとに異なる。
- 乱数の種という分析手段によって、ユニットの重みが変わるのが原因。
- しかし、学習結果の本質的な解釈や含意まで変化させることは少ない。⇒この点が決定木との違い

SOM

- 同一のデータを用いても乱数の種が異なると、競合学習された重みも異なる。
⇒この点では逆伝播学習と同じ
- SOMは重みが直に観測対象の布置を決めるので、マップそのものが変わってくる。

演習

- kouri.csvを使って10 × 10のSOMを作成し、決定木と組み合わせる。

解説(マップを作る)

```
library(som)
```

```
小売業データ <- read.csv("kouri.csv",header=T)
```

```
標準化小売業データ <- normalize(小売業データ[,2:9], byrow=F) #データの  
標準化
```

```
小売業SOM <- som(標準化小売業データ, xdim=10, ydim=10, topol="rect")
```

```
乱数 <- cbind(rnorm(nrow(小売業データ), 0, 0.15), rnorm(nrow(小売業デー  
タ), 0, 0.15))
```

```
小売業マップ <- 小売業SOM $visual[,1:2] + 乱数 + 0.5
```

```
plot(小売業マップ, xlim=c(0,10), ylim=c(0,10))
```

```
text(小売業マップ[,1], 小売業マップ[,2], 小売業データ$業種名)
```


解説(決定木と組み合わせる)

```
library(mvpart)
```

```
小売業SOM2 <- som(標準化小売業データ, xdim=5, ydim=1,  
topol="rect") //決定木用にSOMを5×1に書き換える
```

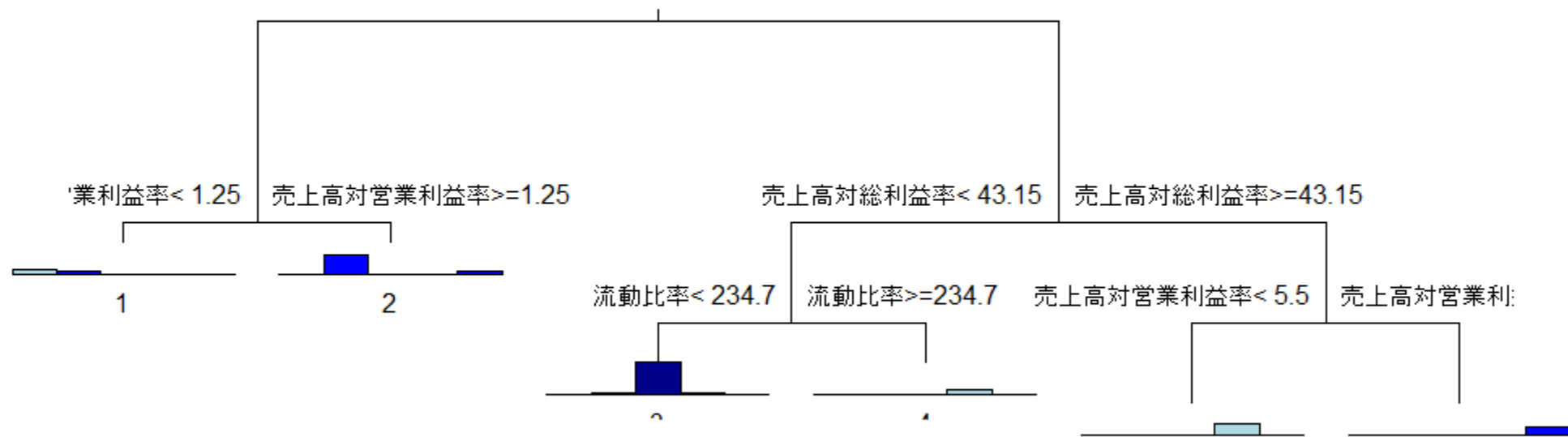
```
決定木用小売業データ <- cbind(小売業データ, 小売業  
SOM2$visual[,1]+1) //データの縦の結合
```

```
colnames(決定木用小売業データ)[10] <- "セグメント" //名前付け
```

```
小売業分類木 <- rpart(セグメント~., data=決定木用小売業データ  
[,2:10], method="class")
```

```
print(小売業分類木)
```

結果



宿題

前回の課題で扱ったpokemon2.csvで作ったマップを使って、決定木と組み合わせる。

まとめ

SOMと決定木を組み合わせてすることで便利になるなと思った。もう少し理解してから臨みたかったです。