

シングルモルトの特徴

山田 道洋

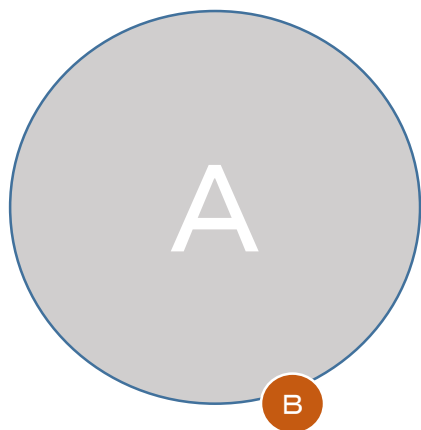
前回の最後

有用な連関規則を自動で見つけたい！！

では有用な連関規則とはどんなものなのか？

前回の発表より

- ・条件付き確率と比較して事前確率の低い連関規則は良い規則である。

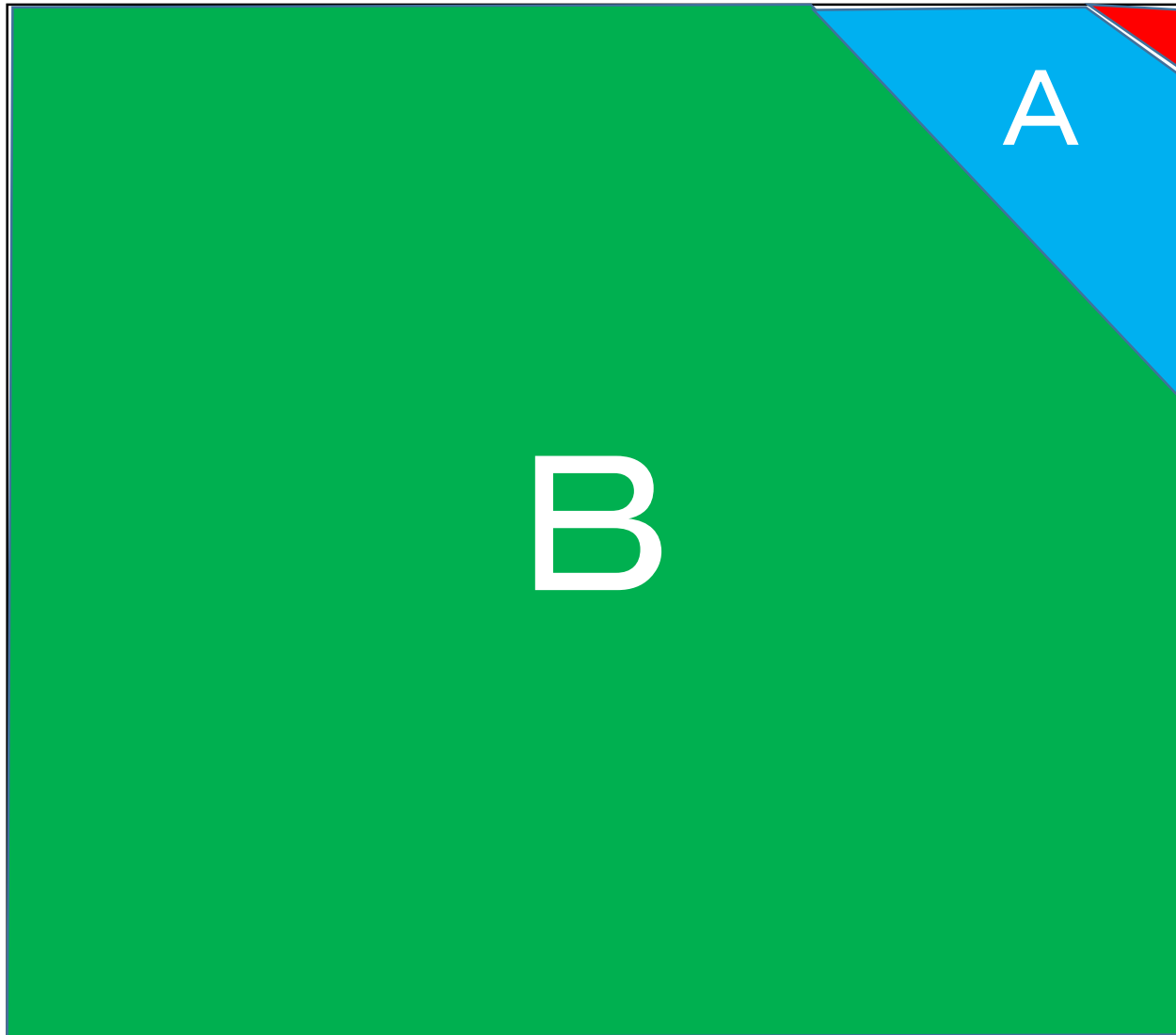


P:クジを買う人	1000万人
A:1000枚以上買う人	1万人
B:1等にあたる人	15人

条件付き確率 $P(B | A) = 0.01\%$ 1人

事前確率 $P(B) = 0.00015\%$

$P(B | A)$ のほうが $P(B)$ より高いので有用である



P: 風邪をひいた人	1000人
A: 薬を飲まなかった人(青+赤)	100人
B: 死ななかった人(緑+青)	999人

条件付き確率 $P(B | A) = 99\%$ 99人

事前確率 $P(B) = 99.9\%$

$P(B | A)$ のほうが $P(B)$ より低いので有用でない

このように条件付き確率がいくら高くても事前確率よりも低い場合は存在意義がない(前提確率の吟味が必要なくなる)

1. 口当たり(軽やか)	→味(甘い)(0.31,0.56)
2. 色(濃金色)	→味(甘い)(0.25,0.56)
3. 香り(芝生)	→味(甘い)(0.15,0.56)
8. 香り(甘い)	→味(甘い)(0.06,0.43)
9. 色(白ワイン)	→味(甘い)(0.07,0.38)
10. 後味(余韻ある)	→味(甘い)(0.22,0.38)

シングルモルト全体で味が甘いのは51銘柄($P(B)=0.46$)

1~3は $P(B | A) > P(B)$ なので有用

8~10は $P(B | A) < P(B)$ なので有用でない

このように規則が有用であるかないかは改善率 (lift) を計算することで判断する

改善率 = $P(B | A) / P(B) = P(B, A) / (P(A)P(B))$ で求められ ($P(A)P(B | A) = P(B, A)$)

改善率 > 1 ならばその連関規則が有用 ($P(B | A) > P(B)$)

改善率 < 1 ならばその連関規則が有用でないと言える ($P(B | A) > P(B)$)

演習

scotch.csvの甘い味のスコッチについて改善率が1より大きいデータを表示する

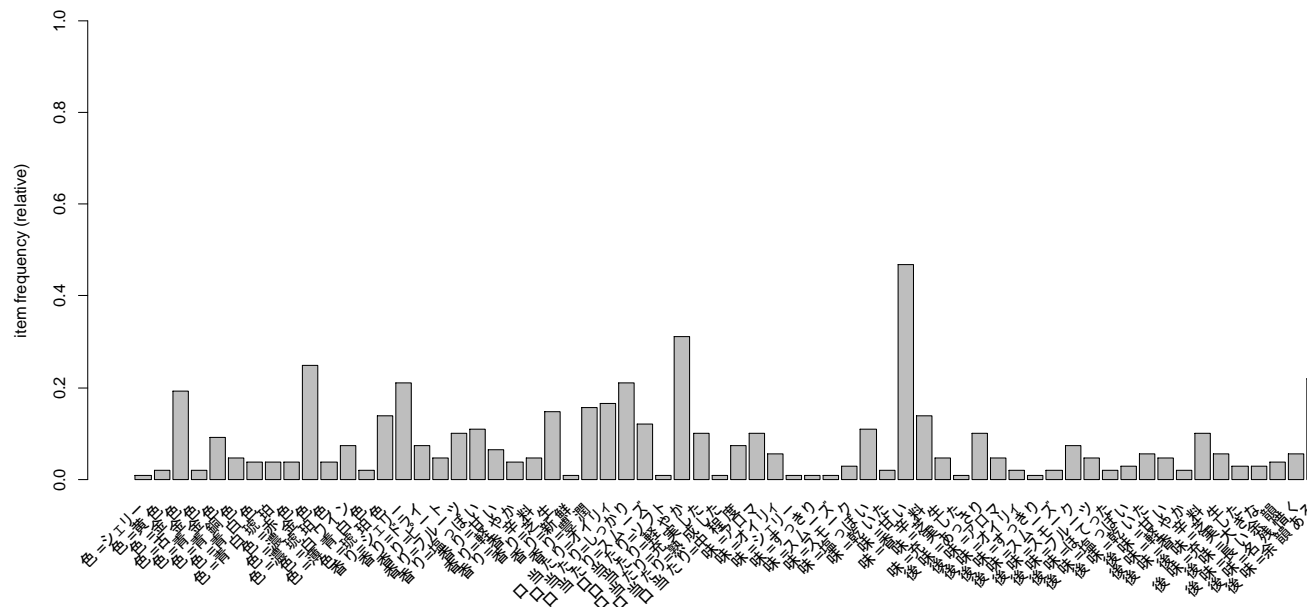

```
setwd("../Dropbox/Share/DL02045/Program/chap5")
library(arules)
scotchData <- read.csv("scotch.csv", header=T, row.names=1)
```

※scotch2.csv はデータが英語

今回は数値データを使用したため as.matrix() を使用したが、今回は文字列のデータを使うので as() を使用する

※教科書と同じ表示にするには options(digits=3)

```
scotchTrans <- as(scotchData, "transactions")
summary(scotchTrans)
inspect(scotchTrans)
itemFrequency(scotchTrans)
itemFrequencyPlot(scotchTrans, ylim=c(0,1))
```



```
scotchRule <- apriori(scotchTrans, parameter=list(maxlen=3, sup=0.02, conf=0.35, ext=T))
```

※教科書と同じ表示にするにはoptions(digits=2)

```
sweetBig <- subset(scotchRule, subset=(rhs %in% "味=甘い") & (lift>1.0))           ※taste=swe  
inspect(sweetBig)
```

subset() で一度構築された連関規則(ここではscotchRule)からさらに条件を設定してルールを抽出する
()ないのsubset=の後に条件を記述
rhs(ライトハンドサイド)はルールボディを指す “味=甘い” を指定
lift>1.0 で改善率が1より大きいものを抽出

```
inspect(sweetBig [c(16,15,11,5,35,33),])
```

改善率が1より小さいものを抽出

```
sweetSmall <- subset(scotchRule, subset=(rhs %in% "味=甘い") & (lift<1.0))  
inspect(sweetSmall)  
inspect(sweetSmall [c(8,1,2,9),])
```

Lift>1.0

lhs	rhs	support	confidence	lhs.support	lift
1 {色=青白色}	=> {味=甘い}	0.02752294	0.7500000	0.03669725	1.602941
2 {香り=ピート}	=> {味=甘い}	0.03669725	0.8000000	0.04587156	1.709804
3 {後味=甘い}	=> {味=甘い}	0.03669725	0.8000000	0.04587156	1.709804
4 {色=青銅色}	=> {味=甘い}	0.02752294	0.6000000	0.04587156	1.282353
5 {後味=名残惜く}	=> {味=甘い}	0.03669725	0.6666667	0.05504587	1.424837
6 {後味=芝生}	=> {味=甘い}	0.02752294	0.5000000	0.05504587	1.068627
7 {口当たり=中程度}	=> {味=甘い}	0.03669725	0.5000000	0.07339450	1.068627
8 {香り=ドライ}	=> {味=甘い}	0.03669725	0.5000000	0.07339450	1.068627
9 {後味=スモーク}	=> {味=甘い}	0.03669725	0.5000000	0.07339450	1.068627
10 {後味=香辛料}	=> {味=甘い}	0.05504587	0.5454545	0.10091743	1.165775
11 {香り=芝生}	=> {味=甘い}	0.08256881	0.5625000	0.14678899	1.202206
12 {香り=豊潤}	=> {味=甘い}	0.08256881	0.5294118	0.15596330	1.131488
13 {色=金色}	=> {味=甘い}	0.09174312	0.4761905	0.19266055	1.017740
14 {口当たり=しっかり}	=> {味=甘い}	0.10091743	0.4782609	0.21100917	1.022165
15 {色=濃金色}	=> {味=甘い}	0.13761468	0.5555556	0.24770642	1.187364
16 {口当たり=軽やか}	=> {味=甘い}	0.17431193	0.5588235	0.31192661	1.194348
17 {色=濃金色, 香り=ピート}	=> {味=甘い}	0.03669725	1.0000000	0.03669725	2.137255
18 {香り=ドライ, 口当たり=軽やか}	=> {味=甘い}	0.02752294	0.7500000	0.03669725	1.602941
19 {色=琥珀色, 後味=あっさり}	=> {味=甘い}	0.02752294	0.7500000	0.03669725	1.602941

lift<1.0

lhs	rhs	support	confidence	lhs.support	lift
1 {香り=甘い}	=> {味=甘い}	0.02752294	0.4285714	0.06422018	0.9159664
2 {色=白ワイン}	=> {味=甘い}	0.02752294	0.3750000	0.07339450	0.8014706
3 {後味=あっさり}	=> {味=甘い}	0.04587156	0.4545455	0.10091743	0.9714795
4 {香り=フルーツ}	=> {味=甘い}	0.04587156	0.4545455	0.10091743	0.9714795
5 {口当たり=充実した}	=> {味=甘い}	0.04587156	0.4545455	0.10091743	0.9714795
6 {口当たり=スムーズ}	=> {味=甘い}	0.05504587	0.4615385	0.11926606	0.9864253
7 {色=琥珀色}	=> {味=甘い}	0.06422018	0.4666667	0.13761468	0.9973856
8 {香り=シェリー}	=> {味=甘い}	0.09174312	0.4347826	0.21100917	0.9292413
9 {後味=余韻ある}	=> {味=甘い}	0.08256881	0.3750000	0.22018349	0.8014706
10 {香り=豊潤, 後味=余韻ある}	=> {味=甘い}	0.02752294	0.4285714	0.06422018	0.9159664
11 {色=金色, 口当たり=軽やか}	=> {味=甘い}	0.03669725	0.4444444	0.08256881	0.94989

感想

知りたいものに対する関連規則を簡単に見つけられるようになり
有用そうな関連規則の絞り込みもできるようになった

しかし

結局、最も有用かどうかは自分で考えるしかなかった。

宿題

chap5にあるpokem3.csvかpokemon4.csvを使って好きな項目についての関連規則をしてみる

Pokemon3は英語表記pokemon4は日本語表記