データサイエンスの諸相
（アソシエーション分析）

前田利之（まえだ　としゆき）
maechan@hannan-u.ac.jp

阪南大学　経営情報学部

(2024.11.20)

k-means【復習】

1. 以下のデータが I:￥rdata￥km01.csv に保存されているはず（されてない人は打ち込む）

```
name,x,y
C1,1,1
C2,2,1
C3,1,3
C4,4,5
C5,5,5
C6,5,3
```

2. R を立ち上げ、以下を実行する

```
km1 <- read.csv('I:/rdata/km01.csv'
 ,header=T,row.names=1) # 本当は１行
km2 <- kmeans(km1,2)   # ２クラスター
km2   # 内容を確認
plot(km1,pch=km2$cluster)          # プロット
points(km2$centers,pch=8) # ポイント
```

アソシエーション分析

- ２種類のデータの関係…共分散・相関
- しかし、これらは関係があることは示せても、[　　]影響を与えていくかは表現していない
- そこで[　　]＝「もしＡならＢである」という関係の分析を試みる

アソシエーション分析

- ２種類のデータの関係…共分散・相関
- しかし、これらは関係があることは示せても、どちらがどちらに影響を与えていくかは表現していない
- そこで [　　] =「もしＡならＢである」という関係の分析を試みる

アソシエーション分析

- ２種類のデータの関係…共分散・相関
- しかし、これらは関係があることは示せても、**どちらがどちらに**影響を与えていくかは表現していない
- そこで **因果関係**＝「もしＡならＢである」という関係の分析を試みる

有名な例：POSシステム

- POS = [ ] …在庫管理, 販売記録
- あるお店で客の購入履歴を調べたところ、週末の夕方に [ ] で同時に [ ]
  →全ての購買記録＝大量のデータを分析することで興味深いルールを抽出できた
- サンプル調査とは根本的に違う [ ]

有名な例：POSシステム

- POS = Point Of Sales …在庫管理, 販売記録
- あるお店で客の購入履歴を調べたところ、週末の夕方に で同時に

  →全ての購買記録＝大量のデータを分析することで興味深いルールを抽出できた
- サンプル調査とは根本的に違う

有名な例：POSシステム

- POS = Point Of Sales …在庫管理, 販売記録
- あるお店で客の購入履歴を調べたところ、週末の夕方に「紙おむつ」を買った客で同時に

  →全ての購買記録＝大量のデータを分析することで興味深いルールを抽出できた
- サンプル調査とは根本的に違う

有名な例：POSシステム

- POS = Point Of Sales …在庫管理, 販売記録
- あるお店で客の購入履歴を調べたところ、週末の夕方に「紙おむつ」を買った客で同時に「ビール」を買った人が多かった
  →全ての購買記録＝大量のデータを分析することで興味深いルールを抽出できた
- サンプル調査とは根本的に違う

有名な例：POSシステム

- POS = Point Of Sales …在庫管理, 販売記録
- あるお店で客の購入履歴を調べたところ、週末の夕方に「紙おむつ」を買った客で同時に「ビール」を買った人が多かった
  →全ての購買記録＝大量のデータを分析することで興味深いルールを抽出できた
- サンプル調査とは根本的に違う
  （全体データからの分析）

アソシエーションルール

- 事象:前例での「◯△を買う」というような出来事のこと
- 事象「紙おむつを買う」を A, 事象「ビールを買う」を B とすると、「紙おむつを買う人はビールを買う」を $A \Rightarrow B$ と書く
-  : このような条件文のこと

アソシエーションルール

- 事象:前例での「◯△を買う」というような出来事のこと
- 事象「紙おむつを買う」を A, 事象「ビールを買う」を B とすると、「紙おむつを買う人はビールを買う」を $A \Rightarrow B$ と書く
- アソシエーションルール: このような条件文のこと

アソシエーションルール（続き）

- $A \Rightarrow B$ の矢印の向きは重要
  - 「紙おむつを買う人はビールを買う」と、「ビールを買う人は紙おむつを買う」は意味が違う、ということ
  - 例：松原市民は大阪府民であるが、大阪府民であっても松原市民とは限らない（大阪市民、堺市民…の可能性あり）
- A：
- B：

アソシエーションルール（続き）

- $A \Rightarrow B$ の矢印の向きは重要
  - 「紙おむつを買う人はビールを買う」と、「ビールを買う人は紙おむつを買う」は意味が違う、ということ
  - 例：松原市民は大阪府民であるが、大阪府民であっても松原市民とは限らない（大阪市民、堺市民…の可能性あり）
- A：条件部 (rule head)
- B：

アソシエーションルール（続き）

- $A \Rightarrow B$ の矢印の向きは重要
  - 「紙おむつを買う人はビールを買う」と、「ビールを買う人は紙おむつを買う」は意味が違う、ということ
  - 例：松原市民は大阪府民であるが、大阪府民であっても松原市民とは限らない（大阪市民、堺市民…の可能性あり）
- A：条件部 (rule head)
- B：結論部 (rule body)

集合と指標

- $n(X)$ を事象 X が起こった回数とする (X は A であったり B であったりする)
- $n(X, Y)$ を事象 X, Y が同時に起こった回数とする
- 全部の事象を Ω とする

指標（続き）

- B が起こる確率＝ [ ]: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ [ ]:

  $$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ [ ]:

  $$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- [ ] を B が起こる確率で割ったもの＝ [ ]:

  $$\frac{p(B|A)}{p(B)} = \frac{p(A, B)}{p(A)p(B)}$$

指標（続き）

- B が起こる確率＝ **期待信頼度**: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ [　　]:

  $$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ [　　]:

  $$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- [　　]を B が起こる確率で割ったもの＝ [　　]:$\dfrac{p(B|A)}{p(B)} = \dfrac{p(A, B)}{p(A)p(B)}$

指標（続き）

- B が起こる確率＝ **期待信頼度**: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ **支持度**:

$$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ [　　]:

$$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- [　　] を B が起こる確率で割ったもの＝ [　　]:

$$\frac{p(B|A)}{p(B)} = \frac{p(A, B)}{p(A)p(B)}$$

指標（続き）

- B が起こる確率＝ **期待信頼度**: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ **支持度**:

$$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ **信頼度**:

$$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- [　　]を B が起こる確率で割ったもの＝ [　　]:$\dfrac{p(B|A)}{p(B)} = \dfrac{p(A, B)}{p(A)p(B)}$

指標（続き）

- B が起こる確率＝ **期待信頼度**: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ **支持度**:

$$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ **信頼度**:

$$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- **信頼度** を B が起こる確率で割ったもの＝ [　　]:

$$\frac{p(B|A)}{p(B)} = \frac{p(A, B)}{p(A)p(B)}$$

指標（続き）

- B が起こる確率＝ **期待信頼度**: $p(B) = \dfrac{n(B)}{n(\Omega)}$
- A, B が一緒に起こる確率＝ **支持度**:

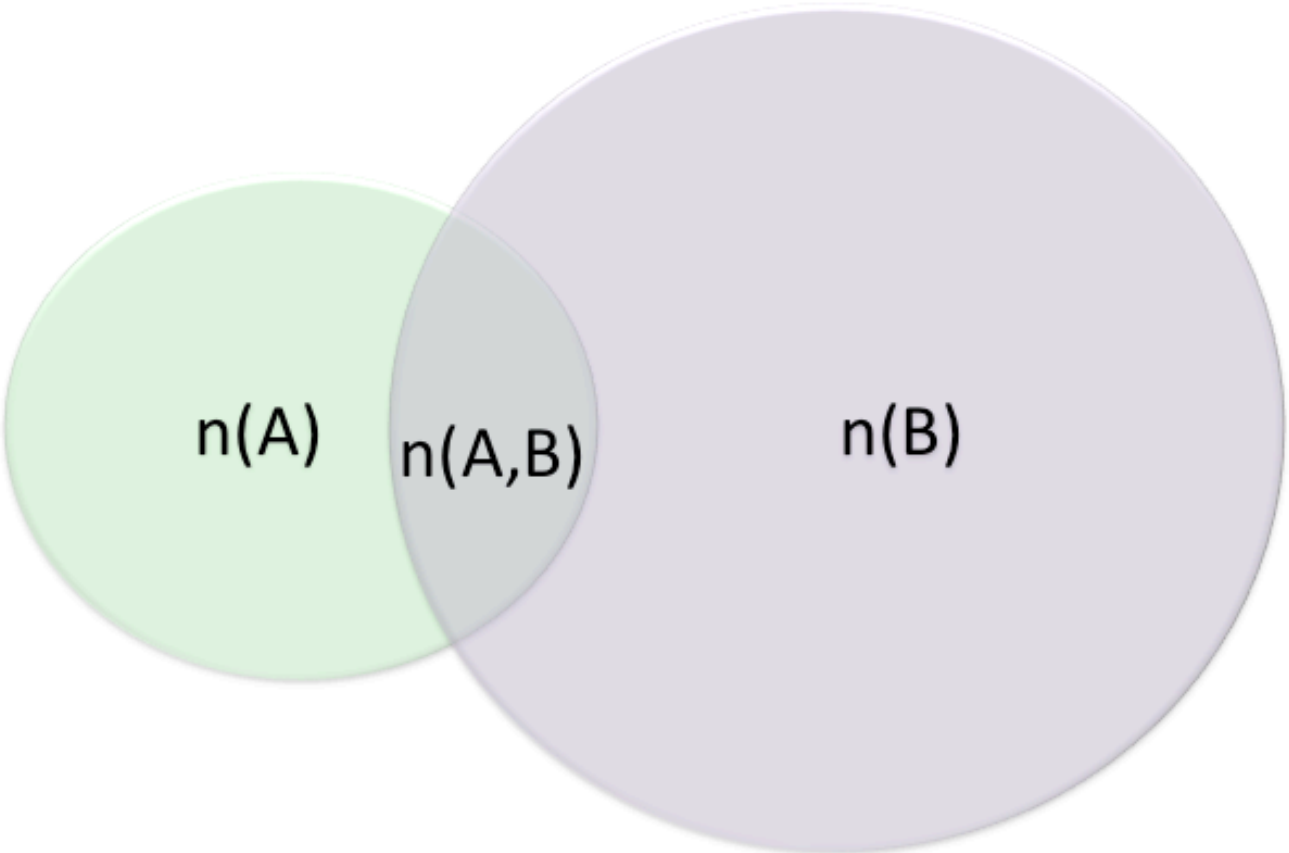
$$p(A, B) = \frac{n(A, B)}{n(\Omega)}$$

- A が起こった前提で B が起こる確率＝ **信頼度**:

$$p(B|A) = \frac{n(A, B)}{n(A)} = \frac{p(A, B)}{p(A)}$$

- **信頼度** を B が起こる確率で割ったもの＝ **リフト値**: $\dfrac{p(B|A)}{p(B)} = \dfrac{p(A, B)}{p(A)p(B)}$

集合（ベン図）



指標（例）

- ２つの店 Shop1, Shop2 があるとする
- それぞれの店での、$A\{$ ミントタブレット $\}$, $B\{$ のど飴 $\}$ を買った人の人数をデータとする
- ここで、$A \Rightarrow B$ (条件) を考える

指標（例）

Shop1		Shop2	
$n(\Omega)$	100	$n(\Omega)$	100
$n(A)$	25	$n(A)$	25
$n(B)$	40	$n(B)$	80
$n(A,B)$	20	$n(A,B)$	20

指標（例）

- Shop1 では
  - 支持度 =
  - 期待信頼度 =
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop1 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 =
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop1 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 40/100 = 0.4
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop1 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 40/100 = 0.4
  - 信頼度 = 20/25 = 0.8
  - リフト値 =

指標（例）

- Shop1 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 40/100 = 0.4
  - 信頼度 = 20/25 = 0.8
  - リフト値 = 0.8/0.4 = 2

指標（例）

- Shop2 では
  - 支持度 =
  - 期待信頼度 =
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop2 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 =
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop2 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 80/100 = 0.8
  - 信頼度 =
  - リフト値 =

指標（例）

- Shop2 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 80/100 = 0.8
  - 信頼度 = 20/25 = 0.8
  - リフト値 =

指標（例）

- Shop2 では
  - 支持度 = 20/100 = 0.2
  - 期待信頼度 = 80/100 = 0.8
  - 信頼度 = 20/25 = 0.8
  - リフト値 = 0.8/0.8 = 1

指標（例）

- どちらも信頼度は高い [　　] →ミントタブレットを買った人は [　　] のど飴も買う確率が高い、と言える
- しかし Shop2 ではもともとのど飴を買う確率が高い→ミントタブレットを買うついでに買ったのかも
- 一方 Shop1 ではミントタブレットを買った人がのど飴を買う確率 [　　] は単にのど飴を買う確率 [　　] より大きい→ミントタブレットを買った人が特にのど飴も買う傾向があると言える

指標（例）

- どちらも信頼度は高い (0.8) →ミントタブレットを買った人は [　　] のど飴も買う確率が高い、と言える
- しかし Shop2 ではもともとのど飴を買う確率が高い→ミントタブレットを買うついでに買ったのかも
- 一方 Shop1 ではミントタブレットを買った人がのど飴を買う確率 [　　] は単にのど飴を買う確率 [　　] より大きい→ミントタブレットを買った人が特にのど飴も買う傾向があると言える

指標（例）

- どちらも信頼度は高い (0.8) →ミントタブレットを買った人は 同時に のど飴も買う確率が高い、と言える
- しかし Shop2 ではもともとのど飴を買う確率が高い→ミントタブレットを買うついでに買ったのかも
- 一方 Shop1 ではミントタブレットを買った人がのど飴を買う確率 [　　] は単にのど飴を買う確率 [　　] より大きい→ミントタブレットを買った人が特にのど飴も買う傾向があると言える

指標（例）

- どちらも信頼度は高い (0.8) →ミントタブレットを買った人は 同時に のど飴も買う確率が高い、と言える
- しかし Shop2 ではもともとのど飴を買う確率が高い→ミントタブレットを買うついでに買ったのかも
- 一方 Shop1 ではミントタブレットを買った人がのど飴を買う確率 (0.8) は単にのど飴を買う確率 [ ] より大きい→ミントタブレットを買った人が特にのど飴も買う傾向があると言える

指標（例）

- どちらも信頼度は高い (0.8) →ミントタブレットを買った人は 同時に のど飴も買う確率が高い、と言える
- しかし Shop2 ではもともとのど飴を買う確率が高い→ミントタブレットを買うついでに買ったのかも
- 一方 Shop1 ではミントタブレットを買った人がのど飴を買う確率 (0.8) は単にのど飴を買う確率 (0.4) より大きい→ミントタブレットを買った人が特にのど飴も買う傾向があると言える

分析（実践練習）

- arules というパッケージにある [　　] という関数を使う
  - 複数の品目の関係性を調べる場合、個々の指標の計算は単純でも組み合わせの場合の数は大量となる
  - そこで、支持度に着目し、この値が基準より小さいものを無視することで調べる組み合わせを減らす
  - （支持度が小さい→あまり起きない→影響は小さい、ということ）
  - その上で、ある値以上の信頼度について意味のあるルールを探す

分析（実践練習）

- arules というパッケージにある apriori という関数を使う
  - 複数の品目の関係性を調べる場合、個々の指標の計算は単純でも組み合わせの場合の数は大量となる
  - そこで、支持度に着目し、この値が基準より小さいものを無視することで調べる組み合わせを減らす
  - （支持度が小さい→あまり起きない→影響は小さい、ということ）
  - その上で、ある値以上の信頼度について意味のあるルールを探す

分析（実践練習）

- 準備：パッケージ “arules” をインストールする
  - R を起動し、メニューの「パッケージとデータ」からインストールを選択する
  - 一番最初は、どのミラーサイトからパッケージを取得するかを選択しないといけないかもしれない（状況依存）
- “ ” を実行する。すると；

```
>library(arules)
要求されたパッケージ Matrix をロード中です。
...
```

のように表示される

分析（実践練習）

- 準備：パッケージ “arules” をインストールする
  - R を起動し、メニューの「パッケージとデータ」からインストールを選択する
  - 一番最初は、どのミラーサイトからパッケージを取得するかを選択しないといけないかもしれない（状況依存）
- “ library(arules) ” を実行する。すると；

```
>library(arules)
要求されたパッケージ Matrix をロード中です。

...
```

のように表示される

分析（実践練習）

- データはパッケージ付属の [　　] ＝仮想の食料品店でのPOSデータを使う
- 以下を実行する。

```
> data(Groceries)
> Groceries
transactions in spare format with 9835
transactions (rows) and 169 items
(columns)
> g0 <- Groceries
```

(以降、g0を使う)

分析（実践練習）

- データはパッケージ付属の Groceries ＝仮想の食料品店での POS データを使う
- 以下を実行する。

```
> data(Groceries)
> Groceries
transactions in spare format with 9835
transactions (rows) and 169 items
(columns)
> g0 <- Groceries
```

(以降、g0 を使う)

分析（実践練習）

- 以下を実行し、データをまずデータフレーム形式のテキストと csv に保存する

```
> gfrm0 <- as(g0,"data.frame")
> write.table(gfrm0,"I:/rdata/gdat1.txt")
> gmat0 <- as(g0,"matrix")
> write.csv(gmat0,"I:/rdata/gdat2.csv")
```

- それぞれのデータの形式を確認する

分析（実践練習）

gdat1.txt の中身は以下のようになっている（はず）

```
"items"
"1" "{citrus fruit,semi-finished bread,margarine,
"2" "{tropical fruit,yogurt,coffee}"
"3" "{whole milk}"
"4" "{pip fruit,yogurt,cream cheese ,meat spreads
...(以下略)
```

分析（実践練習）

gdat2.csv の中身は以下のようになっている（はず）

```
"","frankfurter","sausage","liver loaf",...(以下略)
"1",FALSE,FALSE,FALSE,FALSE,...(以下略)
"2",FALSE,FALSE,FALSE,FALSE,...(以下略)
"3",FALSE,FALSE,FALSE,FALSE,...(以下略)
"4",FALSE,FALSE,FALSE,FALSE,...(以下略)
...(以下略)
```

分析（実践練習）

`grule1 <- apriori(g0)` を実行する。すると、以下が得られる（はず）

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.8    0.1    1 none FALSE            TRUE       5     0.1      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 983

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [8 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 done [0.00s].
writing ... [0 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

分析（実践練習）

- apriori はパラメータを指定しないと信頼度 0.8 以上、支持度 0.1 以上のものを調べる
- この条件で、ルールは見つからなかった、ということ
- なので、これら最低支持度や最低信頼度を変えてみる

分析（実践練習）

```
grule2 <- apriori(g0,
p=list(support=0.01,confidence=0.5))
```

（本当は１行）を実行する。すると、以下のようになる（15個のルールが得られる）

```
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.5    0.1    1 none FALSE            TRUE       5    0.01      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 98

set item appearances ...[0 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [88 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [15 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

分析（実践練習）

- 抽出されたルールをみるには [　　] 関数を使う
- ルールが多い場合は見辛いので、以下の関数を使う
  - 先頭からの数行を表示する: [　　] 関数
  - 並び替えをおこなう: [　　] 関数

分析（実践練習）

- 抽出されたルールをみるには inspect() 関数を使う
- ルールが多い場合は見辛いので、以下の関数を使う
  - 先頭からの数行を表示する: [　　] 関数
  - 並び替えをおこなう: [　　] 関数

分析（実践練習）

- 抽出されたルールをみるには inspect() 関数を使う
- ルールが多い場合は見辛いので、以下の関数を使う
  - 先頭からの数行を表示する: head() 関数
  - 並び替えをおこなう: [ ] 関数

分析（実践練習）

- 抽出されたルールをみるには inspect() 関数を使う
- ルールが多い場合は見辛いので、以下の関数を使う
  - 先頭からの数行を表示する: head() 関数
  - 並び替えをおこなう: sort() 関数

分析（実践練習）

- 以下を実行する。sort() 関数の引数の “by” は arules パッケージに含まれているほうで、並べ替えの項目を指定できる。

```
> grule3 <- sort(grule2,by="confidence")
> grule4 <- head(grule3)
> inspect(grule4)
```

分析（実践練習）

下のように表示される

```
    lhs                                  rhs                support    confidence lift     count
[1] {citrus fruit,root vegetables}   => {other vegetables} 0.01037112 0.5862069  3.029608 102
[2] {tropical fruit,root vegetables} => {other vegetables} 0.01230300 0.5845411  3.020999 121
[3] {curd,yogurt}                    => {whole milk}       0.01006609 0.5823529  2.279125  99
[4] {other vegetables,butter}        => {whole milk}       0.01148958 0.5736041  2.244885 113
[5] {tropical fruit,root vegetables} => {whole milk}       0.01199797 0.5700483  2.230969 118
[6] {root vegetables,yogurt}         => {whole milk}       0.01453991 0.5629921  2.203354 143
```

分析（実践練習）

- 前の結果より、例えばcitrus fruit と root vegetables の両方を買った人は同時に [ ] を買っている割合が高いということが分かる
- リフト値が3.029608ということは、通常よりもかなり確率が上ということが言える

分析（実践練習）

- 前の結果より、例えばcitrus fruit と root vegetables の両方を買った人は同時に other vegetables を買っている割合が高いということが分かる
- リフト値が3.029608ということは、通常よりもかなり確率が上ということが言える

分析（実践練習）

- ルールを見つけるためには、適切なパラメータ指定が必要
- →元々のデータをもう少し詳しく見る
- → [　　]: 個々のアイテムの割合 $p(A)$ を表示
- [　　]: ヒストグラムを表示する（品目が多いと字が潰れる）

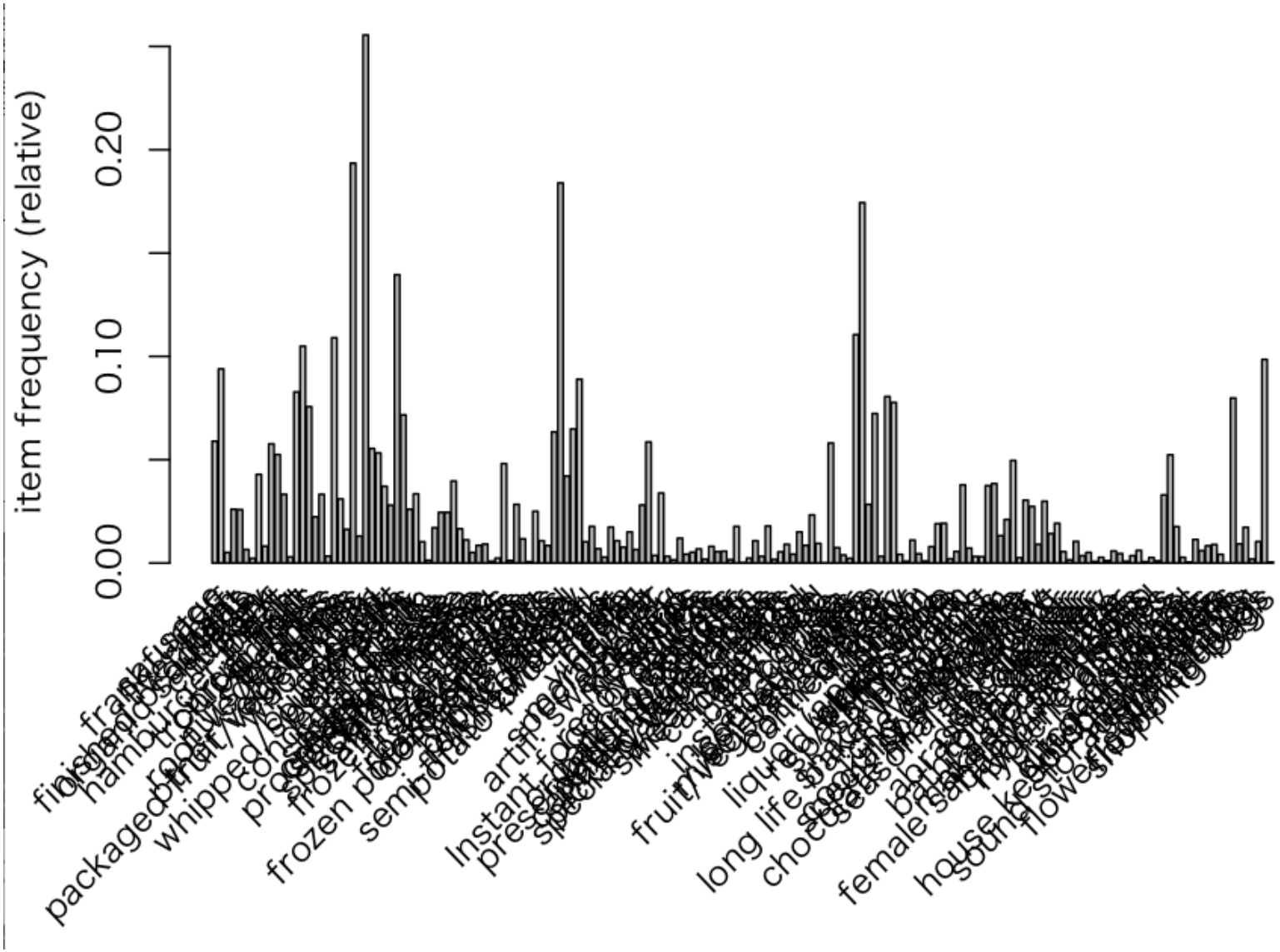
分析（実践練習）

- ルールを見つけるためには、適切なパラメータ指定が必要
- →元々のデータをもう少し詳しく見る
- → itemFrequency(): 個々のアイテムの割合 $p(A)$ を表示
- [ ]: ヒストグラムを表示する（品目が多いと字が潰れる）

分析（実践練習）

- ルールを見つけるためには、適切なパラメータ指定が必要
- →元々のデータをもう少し詳しく見る
- → itemFrequency(): 個々のアイテムの割合 $p(A)$ を表示
- itemFrequencyPlot(): ヒストグラムを表示する（品目が多いと字が潰れる）

分析（実践練習）



分析（実践練習）

- 以下を実行する。sort() 関数の引数の “d=T” は arules パッケージに含まれていないほうで、大きい順に並べる指定である。tail() は最後の数行をみる関数である。

```
> gdat1 <- itemFrequency(g0)
> head(sort(gdat1))
> head(sort(gdat1,d=T))
> tail(sort(gdat1))
> write.csv(gdat1,"I:/rdata/item1.csv") # 結果を保存
```

分析（実践練習）

下のように表示される

```
> gdat1 <- itemFrequency(g0)
> head(sort(gdat1))
           baby food  sound storage medium preservation products       kitchen utensil
        0.0001016777          0.0001016777          0.0002033554          0.0004067107
                bags        frozen chicken
        0.0004067107          0.0006100661
> head(sort(gdat1,d=T))
      whole milk other vegetables       rolls/buns             soda           yogurt
       0.2555160        0.1934926        0.1839349        0.1743772        0.1395018
   bottled water
       0.1105236
> tail(sort(gdat1))
   bottled water           yogurt             soda       rolls/buns other vegetables
       0.1105236        0.1395018        0.1743772        0.1839349        0.1934926
      whole milk
       0.2555160
```

分析（実践練習）

- ルールを抽出するときに、前提部 (lhs) や結論部 (rhs) にくる項目を指定したい場合もある
- → [　　] で指定
- 例：ルールの結論部に ”whole milik” を含んだルールだけを抽出したい場合には appearance=list(rhs=”whole milk”,default=”lhs”) と指定する

分析（実践練習）

- ルールを抽出するときに、前提部 (lhs) や結論部 (rhs) にくる項目を指定したい場合もある
- → appearance=list() で指定
- 例：ルールの結論部に ”whole milik” を含んだルールだけを抽出したい場合には appearance=list(rhs=”whole milk”,default=”lhs”) と指定する

分析（実践練習）

```
grule5 <- apriori(g0,
parameter=list(support=0.005,confidence=0.7),
appearance=list(rhs="whole milk",
default="lhs"))
```
を実行すると、以下のようになる（２行目の＋は継続行のプロンプト）

```
> grule5 <- apriori(g0,parameter=list(support=0.005,confidence=0.7),
+ appearance=list(rhs="whole milk", default="lhs"))
Apriori

Parameter specification:
 confidence minval smax arem  aval originalSupport maxtime support minlen maxlen target   ext
        0.7    0.1    1 none FALSE            TRUE       5   0.005      1     10  rules FALSE

Algorithmic control:
 filter tree heap memopt load sort verbose
    0.1 TRUE TRUE  FALSE TRUE    2    TRUE

Absolute minimum support count: 49

set item appearances ...[1 item(s)] done [0.00s].
set transactions ...[169 item(s), 9835 transaction(s)] done [0.00s].
sorting and recoding items ... [120 item(s)] done [0.00s].
creating transaction tree ... done [0.00s].
checking subsets of size 1 2 3 4 done [0.00s].
writing ... [1 rule(s)] done [0.00s].
creating S4 object  ... done [0.00s].
```

分析（実践練習）

- inspect(grule5) を実行すると、以下のようになる（ルールが得られる）

```
> inspect(grule5)
     lhs
     rhs                support    confidence
     lift      count # ほんとは１行
[1] {tropical fruit,root vegetables,yogurt}=>
     {whole milk} 0.00569395 0.7
     2.739554 56     # ほんとは１行
```

- つまり「熱帯フルーツと根菜とヨーグルトを買った人は牛乳もよく買う」ということ

おしまい

質問があれば、お気軽に！