Turning *tricot* rankings into Plackett-Luce rankings

Kauê de Sousa

2021

Content

- Beans data
- Getting the ordering
- Sparse matrix
- Comparison with the local
- rank_tricot() from gosset package

Lecture

This lecture is available on Youtube

Set up

For this lecture we need to packages PlackettLuce and gosset. We install it from GitHub. You just need to run this command once.

```
library("remotes")
install_github("hturner/PlackettLuce", upgrade = "never")
install_github("agrdatasci/gosset", upgrade = "never")
```

Then we load the packages with `library()`. We run this command every time we open a new R section

```
library("PlackettLuce")
library("gosset")
```

Beans data

Beans data

Load the `beans` data from PlackettLuce, a data frame with 842 entries and 14 variables

```
data("beans", package = "PlackettLuce")
str(beans)
```

```
## 'data.frame':    842 obs. of  14 variables:
##  $ variety_a    : chr  "BRT 103-182" "INTA Rojo" "INTA Ferroso" "INTA Centro Sur" ...
##  $ variety_b    : chr  "SJC 730-79" "INTA Centro Sur" "INTA Matagalpa" "INTA Rojo" ...
##  $ variety_c    : chr  "PM2 Don Rey" "INTA Sequia" "BRT 103-182" "ALS 0532-6" ...
##  $ best         : chr  "C" "B" "A" "B" ...
##  $ worst        : chr  "A" "A" "C" "C" ...
##  $ var_a        : chr  "Worse" "Worse" "Better" "Better" ...
##  $ var_b        : chr  "Worse" "Better" "Worse" "Better" ...
##  $ var_c        : chr  "Better" "Better" "Worse" "Better" ...
##  $ season       : Factor w/ 5 levels "Po - 15","Ap - 15",..: 2 2 2 2 2 2 2 2 2 2 ...
##  $ year         : num  2015 2015 2015 2015 2015 ...
##  $ maxTN        : num  19.4 18.9 18.4 18.9 18.9 ...
##  $ lon          : num  -85.7 -85.4 -85.4 -85.4 -85.4 ...
##  $ lat          : num  13.1 13.3 13.3 13.3 13.3 ...
##  $ planting_date: Date, format: "2015-12-18" "2015-12-18" ...
```

A subset of the data

Get the first three rows and only the columns important for this process

```
beans2 <- beans[1:3, c("variety_a", "variety_b", "variety_c", "best", "worst")]
beans2
```

```
##      variety_a       variety_b   variety_c best worst
## 1  BRT 103-182      SJC 730-79 PM2 Don Rey    C     A
## 2    INTA Rojo INTA Centro Sur INTA Sequia    B     A
## 3 INTA Ferroso  INTA Matagalpa BRT 103-182    A     C
```

Getting the ordering

This gives an ordering of the three varieties assigned to each farmer. The names of these varieties are stored in separate columns

```
##      variety_a       variety_b   variety_c best worst
## 1  BRT 103-182      SJC 730-79 PM2 Don Rey    C     A
## 2    INTA Rojo INTA Centro Sur INTA Sequia    B     A
## 3 INTA Ferroso  INTA Matagalpa BRT 103-182    A     C
```

With this ordering we can run the function `as.rankings(..., input = "orderings")` from PlackettLuce and get the rankings

```
##              best         middle       worst
## 1     PM2 Don Rey     SJC 730-79 BRT 103-182
## 2 INTA Centro Sur    INTA Sequia   INTA Rojo
## 3    INTA Ferroso INTA Matagalpa BRT 103-182
```

Sparse matrix

Internaly what it does is decode the orderings and create a Sparse matrix.

```
##              best         middle       worst
## 1     PM2 Don Rey     SJC 730-79 BRT 103-182
## 2 INTA Centro Sur    INTA Sequia   INTA Rojo
## 3    INTA Ferroso INTA Matagalpa BRT 103-182
```

Here each variety is a column, each row is the evaluation from a given farmer.

Varieties not assigned to a given farmer are set with 0.

```
##   BRT103-1 INTACntS INTAFrrs INTAMtgl INTARojo INTASequ PM2DonRy SJC730-7
## 1        3        0        0        0        0        0        1        2
## 2        0        1        0        0        3        2        0        0
## 3        3        0        1        2        0        0        0        0
```

PlackettLuce rankings

PlackettLuce has a printing method to see these rankings

```
##                                                  1
##      "PM2 Don Rey > SJC 730-79 > BRT 103-182"
##                                                  2
##   "INTA Centro Sur > INTA Sequia > INTA Rojo"
##                                                  3
## "INTA Ferroso > INTA Matagalpa > BRT 103-182"
```

Comparison with the local

Comparison with the local

Comparison with the local variety are organized as paired comparisons and added as additional rankings. We do this avoid the need to reorder the three varieties into the new ranking of four.

Lets subset the data again to see it better

```
beans3 <- beans[1:3, c("variety_a", "var_a")]
beans3
```

```
##      variety_a  var_a
## 1  BRT 103-182  Worse
## 2    INTA Rojo  Worse
## 3 INTA Ferroso Better
```

Which brings us to this ordering

```
##      [,1]           [,2]
## [1,] "Local"        "BRT 103-182"
## [2,] "Local"        "INTA Rojo"
## [3,] "INTA Ferroso" "Local"
```

Comparison with the local

From this ordering

```
##      [,1]           [,2]
## [1,] "Local"        "BRT 103-182"
## [2,] "Local"        "INTA Rojo"
## [3,] "INTA Ferroso" "Local"
```

To the Sparse matrix

```
##      BRT 103-182 INTA Ferroso INTA Rojo Local
## [1,]           2            0         0     1
## [2,]           0            0         2     1
## [3,]           0            1         0     2
```

rank_tricot()

rank_tricot() from gosset

The package gosset has the function `rank_tricot()` to handle this process. This returns an object of class "rankings" with 842 rows (one for each farmer) and 10 columns (one for each variety).

- data, the name of the data frame
- items, the column names with the varieties concatenated with `c()`
- input, the column names with the tricot rankings concatenated with `c()`

```
R <- rank_tricot(data = beans,
                 items = c("variety_a", "variety_b", "variety_c"),
                 input = c("best", "worst"))

print(R[1:6], width = 100)
```

```
## [1] "PM2 Don Rey > SJC 730-79 > BRT 103-182"
## [2] "INTA Centro Sur > INTA Sequia > INTA Rojo"
## [3] "INTA Ferroso > INTA Matagalpa > BRT 103-182"
## [4] "INTA Rojo > INTA Centro Sur > ALS 0532-6"
## [5] "PM2 Don Rey > INTA Sequia > SJC 730-79"
## [6] "ALS 0532-6 > INTA Matagalpa > INTA Rojo"
```

Add comparison with local

We pass the comparison with the local with the argument additional.rank which points to a subset of the data with the columns representing the evaluation with the local. This returns an object of class "rankings" with 3368 rows (nfarmers x 4) and 11 columns (10 tricot varieties + local).

```
R <- rank_tricot(data = beans,
                 items = c("variety_a", "variety_b", "variety_c"),
                 input = c("best", "worst"),
                 additional.rank = beans[c("var_a", "var_b", "var_c")])

print(R[c(1:3, 843:845)], width = 100)
```

```
## [1] "PM2 Don Rey > SJC 730-79 > BRT 103-182"
## [2] "INTA Centro Sur > INTA Sequia > INTA Rojo"
## [3] "INTA Ferroso > INTA Matagalpa > BRT 103-182"
## [4] "Local > BRT 103-182"
## [5] "Local > INTA Rojo"
## [6] "INTA Ferroso > Local"
```

Thank you!

@desousakaue

k.desousa@cgiar.org

Back to the main page