Ensemble packages

useR2017

05/07/2017

Martin Otava
Janssen Pharmaceutical Companies of Johnson & Johnson

Rudradev Sengupta, Ewoud de Troyer and Ziv Shkedy
Hasselt University
The purpose of ensemble
R packages landscape

- CRAN, Bioconductor, R-Forge, Github, private websites

- CRAN: 10890 packages & growing

- CRAN Task Views
  - Listings per topic
  - Depends on proper annotation of author
  - Limited information only to keep it concise
  - Still dozens of packages per topic
  - Often overlap of methods
  - Some packages highly specific
  - Not compatible input/output
Our motivation

PhD project:
1. Develop methodology
2. Program the method in software
3. Explore what is available
4. Compare with competing methods

• Quite some programming effort needed, if the competing packages are very different in terms of input and output
User perspective

• User/data analyst perspective: “all-included” approach

• Creation a complete R package for data analysis on the same topic

• Instead of browsing CRAN, user can find the most updated materials available in R for a specific topic in one package.
Ensemble package

My definition:
• Connects multiple R packages
• No new methodology needed
• Unified input and output

Extension:
• Graphical user interface
• Method comparison
• Integration of results
Basics

• Not just referencing, but actual combination of packages into unified framework
• Concise notation, clear explanation of methods and avoid duplication (pick the best)

• Methodology should be explored carefully
• Additional peer review for the user
• Collaboration with authors of original package

• No data management needed for various packages
• Unified output format
• Unified plotting options
Extensions

• Use friendly layer may be useful
• GUI, R commander, Shiny, ...

• Comparison of methods performance can be part of the vignette
• Advantages of each package and which are the situations that it works best

• Robust methods for sensitivity analysis
• Simply overlap/union of results obtained using various approaches
Case studies
IsoGeneGUI: overview

- Original CRAN IsoGene package
- Graphical user interface extension IsoGeneGUI
- Ensemble layer added: multiple packages added and standardized

- Dose-response R packages developed for analysis of high-dimensional data
  - Inference: **IsoGene, orQA**
  - Clustering: **ORCME, ORIClust**
  - Model selection: **goric**
- Packages developed independently by different groups

- orQA: fast algorithm, but very specific purpose
- For general use, lot of thinking about input was needed
- IsoGeneGUI: orQA & IsoGene input aligned
IsoGeneGUI: example
IsoGeneGUI: resources

- Most recent version on R-Forge

- Package website

- *IsoGeneGUI: Multiple Approaches for Dose-Response Analysis of Microarray Data Using R*

- Among “Accepted articles” of The R Journal

- R-forge maintainer: Rudradev Sengupta (rudradev.sengupta at uhasselt.be)
BiclustGUI: overview

- CRAN contains many biclustering packages
- BiclustGUI: R commander tool that integrates them
- No new methodology as such
- Framework to add new packages easily in future

- Biclustering: biclust, fabia, isa2, iBBiG, rqubic, BicARE, s4vd, BiBitR
- Diagnostics: BCDiag, superbiclust
BiclustGUI: example
BiclustGUI: adding future methods

- Coding GUI is not “standard R”
- Idea: authors of new method can by themselves add it in BiclustGUI
- Templates for the GUI of BiclustGUI standard windows
- Templates for the function specification
- No knowledge on the GUI programming is needed to add your method to the package
- Final curation of package maintainer
BiclustGUI: adding future methods

1. Making the Frames
2. Configuring the Grid
3. Combining Rows

- Combine row 1 ("Plaid Specifications" Title)
- Combine row 2 & 3 ("Layer Specifications" Title)
**BiclustGUI: resources**

- Available on CRAN
- Developmental version on GitHub
- [Package website](#)
- [Related materials and resources](#) (book website)

- Shiny version available online or locally

- REST R Package:
  - Spin-off package
  - Generalized Template Scripts to quickly create small R Commander Plug-Ins

- Maintainer: Ewoud de Troyer
  (ewoud.detroyer at uhasselt.be)
Conclusion
Create good ensemble package

• Evaluate & integrate methods
• Standardize input & output
• Add your own contribution
• Final package to help navigate R packages landscape
Thank you

motava@its.jnj.com

05/07/2016