Introduction to the TAF package

3 TAF features

Arni Magnusson and Colin Millar

Overview

The boot procedure

Similar to booting a computer, the TAF boot procedure readies the data and software components that are required for subsequent computations.

The boot procedure takes place inside the boot folder, where the `taf.boot()` function looks for files called DATA.bib (required) and SOFTWARE.bib (optional).

In the `linreg` example, the DATA.bib file contains a single metadata entry:

```
@Misc{ezekiel.txt,
  originator = {Mordecai Ezekiel},
  year       = {1930},
  title      = {Speed of automobile and distance to stop after signal},
  source     = {file},
}
```

The boot procedure

The source field specifies where data or software originate from. The following types of values can be used in the source field:

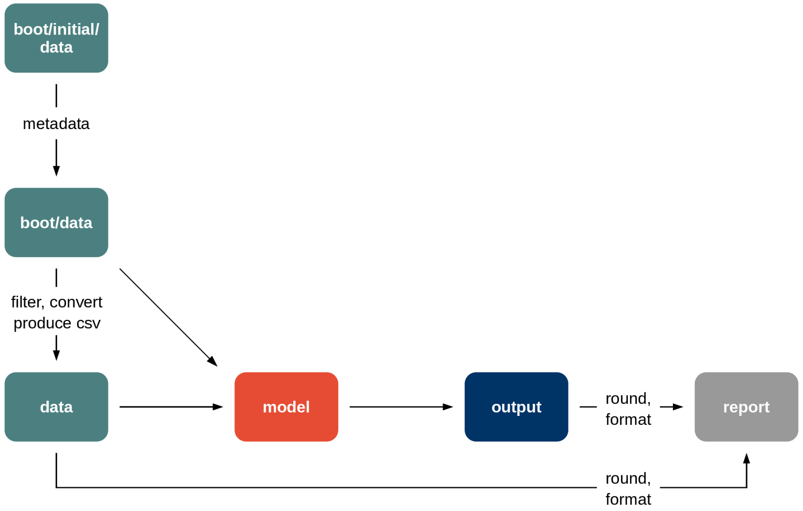
1. GitHub reference of the form `owner/repo[/subdir]@ref`, identifying a specific version of a GitHub resource.
2. URL, identifying a file to download.
3. Special value `script`, indicating that a boot script (a custom R script) should be run to fetch or produce files, e.g., by querying local or online databases.
4. Relative path starting with `initial`, identifying the location of a file or directory somewhere inside the `boot/initial` folder.
5. Special value `file` or `folder`, indicating that the file or folder is inside `boot/initial/data` or `boot/initial/software`.

The flow of data

There are several important differences between the `boot/initial/data` folder and the `boot/data` folder:

- The `boot/initial/data` folder is where the scientist can make initial data files available that are not coming from online data repositories.
- The `boot/data` folder is machine-generated and its contents should not be manually edited by the user. This folder will be regenerated and overwritten whenever the boot procedure is run.
- The contents of the `boot/data` folder are guaranteed to come with descriptive metadata that are declared in the `DATA.bib` file. The purpose of the metadata is to elevate the level of data quality and transparency.
- The `data.R` script reads from the `boot/data` folder and not from `boot/initial/data`.

The flow of data



Creating a new analysis

When authoring a TAF analysis, one can either start with a new workflow or from a similar workflow and adapt it to the current analysis.

The `taf.skeleton()` function creates a new workflow, consisting of an empty `boot/initial/data` folder and the four TAF scripts: `data.R`, `model.R`, `output.R`, and `report.R`.

Each script provides a starting point for that step of the analysis, for example, a new `data.R` script contains the following lines:

```r
# Prepare data, write CSV data tables

# Before:
# After:

library(TAF)

mkdir("data")
```

Creating a new analysis

After running `taf.skeleton()` to create a new TAF workflow, the scientist can populate the `boot/initial/data` folder with initial data files and run `draft.data(file=TRUE)` to produce a `DATA.bib` file.

The next step is then to run `taf.boot()` to populate the `boot/data` folder and start editing the `data.R` script.

Overview of functions

Initial TAF steps

- `draft.data`
- `draft.software`
- `taf.boot`
- `taf.example`
- `taf.skeleton`

Running scripts

- `source.all`

File management

- `mkdir`
- `read.taf`
- `write.taf`

Plots

- `taf.png`

Overview of functions

The TAF package provides many other functions that can be useful but are not required for authoring or running TAF workflows.

Several TAF functions are designed to support running the same analysis across different operating systems and locales, and every function comes with a help page that includes examples and cross-references.

Furthermore, typing `?TAF` opens a package help page that gives an overview of all the functions in the package, grouped by functionality.

Summary

1 **Background** *objectives, design*

2 **Running a TAF analysis** *linear regression, boot and run, structured scripts*

3 **TAF features** *boot procedure, data flow, new analysis, overview of functions*

4 **The TAF community** *browsing an existing analysis, related R packages*

5 **Discussion** *contents of a TAF analysis, benefits of TAF*

6 **Online examples** *ICES, FAO, SPC, various*