Using the package skeleton for patient-level prediction studies

Jenna M. Reps

2019-11-08

Contents

1 Introduction

This vignette describes how one can use the package skeleton for patient-level prediction studies to create one's own study package. This skeleton is aimed at patient-level prediction studies using the `PatientLevelPrediction` package. The resulting package can be used to execute the study at any site that has access to an observational database in the Common Data Model. It will perform the following steps:

1. Instantiate all cohorts needed for the study in a study-specific cohort table.
2. The main analysis will be executed using the `PatientLevelPrediction` package, which involves development and internal validation of prediction models.
3. The prediction models can be exported into a network study package ready to share for external validation.

The package skeleton currently implements an examplar study, predicting various outcomes in multiple target populations. If desired (as a test), one can run the package as is.

1.1 Open the project in Rstudio

Make sure to have RStudio installed. Then open the R project downloaded from ATLAS by decompressing the downloaded folder and clicking on the .Rproj file (where is replaced by the study name you specified in ATLAS). This should open an RStudio session.

1.2 Installing all package dependencies

Before you can build the package you downloaded from ATLAS you need to make sure you have all the dependencies:

```r
source('./extras/packageDeps.R')
```

1.3 Building the package

Once you have the dependencies installed you can now build the R package. This creates a library you can load to run the prediction study. To build the package click 'Build' on the top right hand side tab menu (there are tabs: ‘Environment’, ‘History’, ‘Connections’, ‘Build’, ‘Git’). Once in ‘Build’ click the ‘Install and Restart’ button. This will now build your package and create the R library. If it succeeds you will see ‘* DONE ()’, if it fails you will see red output and the library may not be created. Please report an issue to: https://github.com/OHDSI/PatientLevelPrediction/issues if your library does not get created.

1.4 Running the package

To run the study, open the extras/CodeToRun.R R script (the file called `CodeToRun.R` in the `extras` folder). This folder specifies the R variables you need to define (e.g., outputFolder and database connection settings). See the R help system for details:

```r
library(SkeletonpredictionStudy)
?execute
```

By default all the options are set to F for the execute fuction:

```r
execute(connectionDetails = connectionDetails,
        cdmDatabaseSchema = 'your cdm schema',
        cohortDatabaseSchema = 'your cohort schema',
        cdmDatabaseName = 'Name of database used in study',
        cohortTable = "cohort",
        oracleTempSchema = NULL,
        outputFolder = './results',
        createProtocol = F,
        createCohorts = F,
        runAnalyses = F,
        createResultsDoc = F,
        packageResults = F,
        createValidationPackage = F,
        #analysesToValidate = 1,
        minCellCount= 5,
        createShiny = F,
        createJournalDocument = F,
        analysisIdDocument = 1)
```

If you run the above nothing will happen as each option is false.

To create a study protocol set:

```
    createProtocol = T
```

This uses the settings you specified in ATLAS to generate a protocol for the study.

To create the target and outcome cohorts (cohorts are created into cohortDatabaseSchema.cohortTable)

```
    createCohorts = T
```

To develop and internally validate the models run the code:

```
    runAnalyses = T
```

If the study runs and you get results, you can then create a result document with the full results appended to the protocol document by running (this will fail if you have not run the study first):

```
    createResultsDoc = T
```

To package the results ready for sharing with others you can set:

```
    packageResults = T
```

To create a new R package that can be used to externally validate the models you developed set (this will fail if you have not run the study first to create models):

```
createValidationPackage = T
#analysesToValidate = 1
```

If you do not set analysesToValidate then all the developed models will be transproted into the validation R package. To restrict to Analysis 1 and 5 models set: `analysesToValidate = c(1,5)`. The validation package will be found in your outputFolder directory with the same name as your prediction package but with Validation appended (e.g., outputFolder/Validation). You can add this valdiation package directory to the studyProtocol on the OHDSI github to share the model with other collaborators.

Once you run the study you can view the results via a local shiny app by running:

```
viewMultiplePlp(outputFolder)
```

however, if you want to create a shiny app that you can share with the OHDSI community, you can run:

```
createShiny = T
```

This will create a directory in outputFolder named 'ShinyApp'. You can add this directory to the shinyDeploy OHDSI github to add it to the website data.ohdsi.org . Any sensitive data are removed from the shiny results and you can specify the minimum cell count for the shiny results with 'minCellCount', for example to only show results when there are 10 or more patients `minCellCount = 10`.

To create a template journal document for Analysis 3 run execute with:

```
createJournalDocument = T
analysisIdDocument = 3
```

You ill then find the document in the outputFolder directory.

1.5 extras/PackageMaintenance.R

This file contains other useful code to be used only by the package developer (you), such as code to generate the package manual, and code to insert cohort definitions into the package. All statements in this file assume the current working directory is set to the root of the package.