HypeRSpark vignette

Jarno E. Smit

Introduction

Installing the package

Can be installed directly from GitHub, using the *devtools* package. Building the vignettes relies on the *markdown* package. The code below illustrates how the packages can be downloaded and installed from GitHub.

```
install.packages("devtools")
install.packages("rmarkdown")
devtools::install_github("jarnos97/rPackageHypeRSpark", build_vignettes = TRUE)
```

Why was the package created. What is the use? Which functions are there?

Functions

1. dependencies

HyperSpark is Scala-Based, consequently Java Runtime Environment (JRE) (version1.6 or later) is required. Moreover, Scala requires the Java Developer Kit (JDK) to be installed, which includes JRE. Consequently, users exclusively need to install JDK. HyperSpark is packaged to a jar using Apache Maven. Maven is a Java tool and thus also requires a Java installation. The R package thus has two external dependencies:

- Java SE Development Toolkit (JDK): the package was build using version 8 - update 281. However, it keeps compatibility with newer versions. This requires a (free) Oracle account. Download links here.
- Apache Maven: the package was build using version 3.8.1. Download links here. However, it keeps compatibility with newer versions.

Users should install these dependencies and add both Java and Maven to the system path. This tutorial shows how to add Java to the system path for Windows 10, the same method is used for Maven.

2. configureHyperSpark

HyperSpark parameters

Parameter	Options	Description
setProblem*	PfsProblem, NrProblem	An implemented problem, such as the PFSP. Should be the full name of the class.
data*	Data file	-
setStoppingCondition*	TimeExpired	-
stoppingValue*	Int (milliseconds)	-

Parameter	Options	Description
setAlgorithms*	Problem specific	Are problem specific. All algorithms are defined as their name + Algorithm, i.e. 'ACOAlgorithm'.
numOfAlgorithms	Int	-
setRandomSeed	Int	Ensures reproducibility of results.
setInitialSeeds	?	-
setNInitialSeeds	Array	-
setNDefaultInitialSeeds	Int	-
setSeedingStrategy	Problem specific	-
setNumberOfIterations	Int	-
setProperty	?	?
setMapReduceHandler	?	?

* mandatory parameter

3. packageHyperSpark

The first time running this command can take a while, as all dependencies need to be downloaded from the Maven repository. Subsequent packaging is much faster.