RapidProM 2 User Guide

Data Science Center Eindhoven
Department of Mathematics and Computer Science
Eindhoven University of Technology
*Written by Alfredo Bolt.** *
March 2015
Version 2.0

RapidMiner is a Scientific Workflow Management System, it contains more than 500 operators altogether for all tasks of professional data analysis, i.e. operators for input and output as well as data processing, modelling and other aspects of data mining. Within the process mining domain, there was not any support for the construction and execution of a workflow which describes all analysis steps and their order. For this reason we integrated the ProM 6 Process Mining Framework (`http://www.promtools.org`) within RapidMiner 5. The goal of the integration is to use the knowledge from the scientific workflow domain for the design and execution of process mining workflows.

Contents

---

* rapidprom@promtools.org

1   Introduction

Within the process mining domain there is currently no support for the construction and execution of a workflow which describes all analysis steps and their order. However, within the scientific workflow domain, special kinds of workflow management systems (e.g. KNIME, RapdMiner, Taverna) exist which are designed to compose and execute a series of computational or data manipulation steps, or a workflow, in a scientific application. All together, knowledge from the scientific workflow domain can be used for the design and execution of workflows within the process mining domain.

In particular, in case concepts of the scientific workflow domain are applied within the process mining domain we can think of the following advantages:

- **Process mining analyses can be easily repeated:** Due to scientific workflow capabilities, process mining workflow executions can be easily repeated with just one click.

- **Reproducibility of results:** Since the analysis is constructed using scientific workflow operators, anyone can reproduce the same results as the original experiment if they use the same workflow and input data. This is very important for experiment and analysis validation.

- **An earlier performed experiment can be easily repeated:** Within a scientific workflow management system the settings of each step are saved. Experimentation: Within a scientific workflow management system, large scale scientific experiments can be performed. For example, a process mining algorithm can be tested against different parameter settings.

- **Data Mining and Process Mining combined:** Usage of data access, data transformation, analysis, mining, data visualization, and data exploitation techniques already available within a particular scientific workflow management system. Therefore, they can be used in combination with process mining techniques.

Based on these advantages we decided to connect the ProM Process Mining Framework and the RapidMiner Scientific Workflow and Data Analysis Solution. As such any discovery, conformance, or enhancement algorithm of ProM could be used within a RapidMiner analysis process or a dedicated process mining analysis can be constructed.

2   Installation and Setup

2.1   Installing RapidMiner

For using the RapidProM extension, you first need to install the **RapidMiner 5.3.015** Suite. You can find it in `http://www.rapidprom.org`, or use your favorite search engine to find it elsewhere. Download the software and follow the provided installation instructions. Check the official forum for advanced troubleshooting: `http://rapid-i.com/rapidforum`

2.2   Installing the RapidProM extension

There are two ways to install the RapidProM extension in RapidMiner, the first is using the Marketplace, and the second is doing it manually.

2.2.1 Install RapidProM through RapidMiner Marketplace

The easiest way to install RapidProM is by using the RapidMiner *Marketplace*. In order to do this you will need an internet connection and have port 80 open, as RapidMiner contacts its Marketplace Server through that port. If you don't know anything about port 80 status (or ports in general) just try to install it anyway, as in the worst case it will just not work, and you would have to make a manual installation (Sec 2.2.2).

To access the marketplace, go to the *Help* menu on the toolbar, then click on *Updates and Extensions (Marketplace)*. Then, the Marketplace window will pop up (shown in Fig 1).
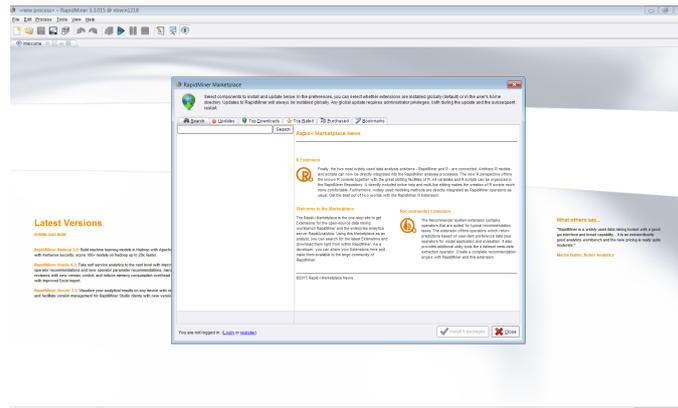

Figure 1: RapidMiner Marketplace

You can search for the keyword *rapidprom* in the search textfield, and the RapidProM extension will show up. When you click on the extension, the description panel will show the extension details as shown in Fig 2. All you have to do is click on the *Select for installation* checkbox in the bottom of the description panel, and then click on the *Install 1 package* button.
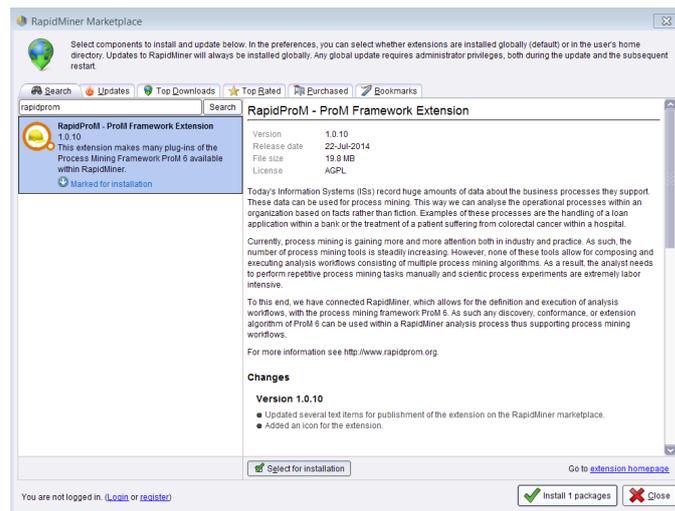

Figure 2: RapidProM extension

After this, RapidMiner will download and install the extension, and then ask you to restart RapidMiner so changes are applied, click on *yes* and

---

restart RapidProM. If everything worked as expected, now you should proceed to the *Initial setup of RapidMiner* described in Sec 2.3.

2.2.2 Install RapidProM manually

If your Marketplace installation went wrong, you can always perform a manual installation. For this you have to download the extension from `http://www.rapidprom.org`. There you will find a .zip file. Extract it and there you will find the file that you need: *rapidminer-RapidProM-2.0.1.jar* All you have to do now is put this .jar file in the right folder. Before doing this, please make sure that RapidMiner is **not** open. Now, go to the RapidMiner installation folder (normally this is C:/Program Files/RapidMiner). There you will find a single folder named *Rapid-Miner5* which contains the RapidMiner files. Go to the *lib → plugins* folder and put the *rapidminer-RapidProM-2.0.0.jar* file there. Now you can start RapidMiner and proceed to the *Initial setup.*

2.3 Initial setup

The RapidProM extension uses packages from the ProM framework. These packages contain the process mining algorithms and techniques provided by ProM. The RapidProM extension needs to download and install these packages in order to use the provided operators, and then change some settings in RapidMiner in order to use them. There are two ways of doing the Initial setup, the first is by using the Setup Wizard, and the second is doing a Manual Setup.

2.3.1 Initial Setup using the Setup Wizard

Using the setup wizard is the easiest way to make the initial setup of RapidProM. Once you have correctly installed RapidProM (either manually or using the marketplace) and start RapidMiner, the setup wizard will appear as a popup asking if the packages have been installed before (shown in Fig. 3).
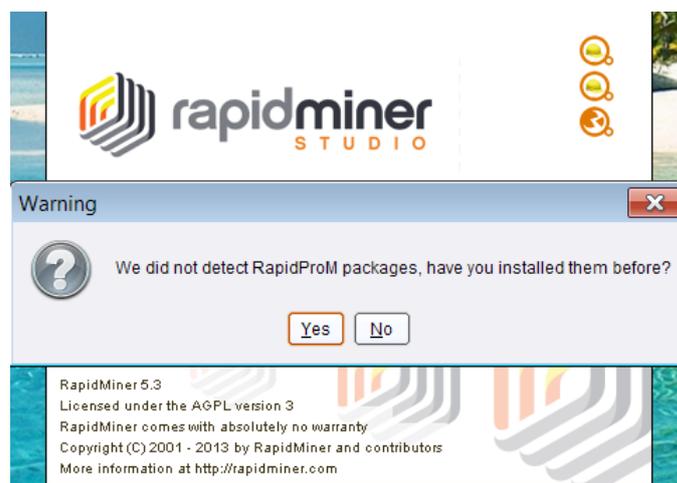

Figure 3: Initial setup wizard

If you have installed RapidProM before, and you want to use the downloaded packages (and not to download them again), then click *Yes* and go the manual setup section (2.3.2). Otherwise, if this is the first time

that you install RapidProM or you just want to use the wizard and download the packages again, then click *No*. Please note that normal ProM packages are **not** equivalent to RapidProM packages.

If you clicked *No* then RapidProM will need a folder for installing the packages. Please select any folder or your convenience, although we recommend to use the RapidMiner installation folder (just to keep things ordered). After this step, RapidProM will try to download the required packages. Since some of the libraries used by RapidProM are different for each Operative System, RapidProM will ask your OS version in order to download the proper package libraries. Once you have selected the right OS version, RapidProM will download and install the packages. To check if the packages were correctly installed, please refer to section X. The installation might be obstructed by your firewall. In that case, you can always perform the Initial Setup manually (Sec. 2.3.2).

2.3.2 Manual Initial Setup

If you had issues with your internet connection such as closed ports or no connection at all during the initial setup, and the wizard could not download the required packages, you can still set up RapidProM manually in your computer.

Before setting up RapidProM you need to download the correct package libraries used by RapidProM to execute its operations. These package libraries are available in `http://www.rapidprom.org`. There you will find specific distributions of these package libraries for your OS. These packages are compressed in a .zip file which you will need to extract in any folder you choose (the new extracted folder containing the packages is named **ProMRM**), although we recommend that you extract it to the RapidMiner installation folder. If you already downloaded the newest RapidProM packages before, then obviously you do not need to do this again.

After downloading and extracting these packages, you need to download a file named *ProM.ini*. This file gives RapidProM the location of the required packages that you just downloaded, and it is available in `http://www.rapidprom.org`. In this file you will need to manually set the `PROM_USER_FOLDER` variable by writing the url of the location of the *ProMRM* folder that you extracted from the .zip file that you just downloaded. For example, we could use:

`PROM_USER_FOLDER = C:/Program Files/RapidMiner/ProMRM`

After these steps, we will proceed with the actual set up by starting RapidMiner. Now, the dialog shown in Fig 3 will appear. This time, you have to select the **Yes** option since you already downloaded the packages. Now another dialog will show up asking for the location of the *ProM.ini* file that you recently downloaded and edited. Find the file and select it. After this, RapidMiner will start normally.

2.3.3 Checking if RapidProM was correctly installed

The most natural way to see if the RapidMiner extension was correctly installed and set up is to look in the operators for a *ProM* folder that contains the recently installed process mining operators. The second way is to go to the *Help* menu then *Installed Extensions*. There you should see the RapidProM extension in that list. Click on it for version details.

2.3.4 Troubleshooting

If something went wrong and you need to re-setup or re-install Rapid-ProM, you will need to *reset* the setup. For this, you need to close Rapid-Miner, then you will have to find the folder where RapidMiner saves all its configurations. Normally this folder is called ".RapidMiner5" (with a dot) and it is located in the *Users* folder. In windows this would be `C:/Users/xxx/.RapidMiner5` where `xxx` is the user that installed Rapid-Miner. Inside this folder you will find all the saved configurations made by RapidMinner, including parameter settings, workflows, etc. So, be careful not to delete random things around here! The only file you will have to delete is named `5_3_YYY_rapidminerrc.ZZZZZZZ` where `YYY` is the revision number of your installed version (009 or 015 are the most common) and `ZZZZZZZ` is your OS (Windows 7, Mac OSX, etc.). After you delete this file, start RapidMiner again and perform the initial setup (this time you already have the packages installed from your previous attempt).

3   Acknowledgements

I would like to thank Ronny Mans for his hard work on the first version of RapidProM, which provided a nice framework for this version.