Optimization with Rapidminer

Robert Meusel, Heiko Paulheim Christian Bizer
Outline

1. Introduction
2. Parameter Optimization
3. Attribute Selection Optimization
Introduction

- Why should we optimize?
  - Default configuration does not work always best
  - Operators come with a (large) set of different parameters
  - Data-Set attributes might be useful or useless

- Why not manual?
  - Testing all combinations of kernel type and svm type (of LibSVM) means rerunning the process 25 times.
  - Playing around with epsilon in addition (5 different values) leads to 125 possible set-ups
  - But, yes you can do it manually, if you are patient

- What can be optimized?
  - Parameters of Operators
  - Selection of Attributes from the Data-Sets
Parameter Optimization

Main Idea:
Let Rapidminer test possible operator parameter combinations for you.
Parameter Optimization

List of nested operators

List of parameters of selected operator

Parameters to Optimize

Selection setup:
Range to Test
Values to Test

Final number of combinations!
Attribute Selection Optimization

Main Idea:

Let Rapidminer selects optimal configuration of attributes for the process and the selected classifier/clustering algorithm.
Interim Results

Using the Log Operator

- Column Name
- Operator
- parameter or port
- Selection of output
Need more information?

• Parameter Optimization YouTube Video:
  • http://www.youtube.com/watch?v=R5vPrTLMzng

• Attribute Selection Optimization YouTube Video:
  • Part 1: http://www.youtube.com/watch?v=7IC3IQEdWxA
  • Part 2: http://www.youtube.com/watch?v=j5vhwbLIZWg
Questions?