Adaptation of Scientific Workflows by Means of Process-Oriented Case-Based Reasoning

Christian Zeyen, Lukas Malburg, Ralph Bergmann

Universität Trier
www.wi2.uni-trier.de

Background & Motivation

- Developing scientific workflows is a demanding task that is supported by various assistance approaches
- Previous case-based approaches mainly focus on the retrieval of reusable workflows, leaving the necessary adaptation up to the user
- Our previous work focused on adaptation of business workflows on the example of cooking recipes. However, significant differences between the workflow types prevent immediate application:

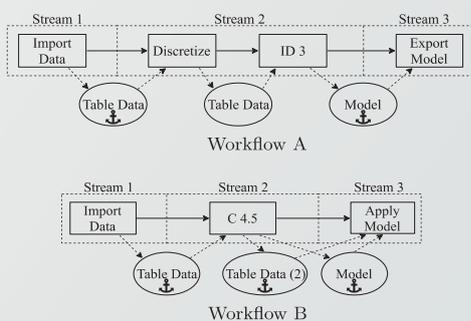|	Business Workflows	Scientific Workflows
Goal	automation - determined before execution	exploratory - validate hypotheses and thus not necessarily known before execution
Usage	mainly executed by humans	designed for computerized execution
Modeling Focus	control-flow of tasks; typically no port connections are used	data-flow between computational steps; port connections are used to ensure valid data exchange

Adaptation Approaches

Substitutional Adapt. by Generalization and Specialization

- Solely applied to tasks, i.e., computational steps, that consume and produce the same type of data
- Retaining common port connections prevents over-generalization
- Known specializations are preferred for improving executability
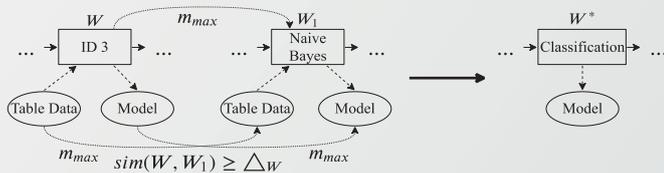
Structural Adaptation with Workflow Streams

- Streams are meaningful and substitutable sub-workflows that produce a partial output
- Automatic partitioning and heuristic search for best substitute stream w.r.t. a given query

Workflow A

Workflow B

Implementation

The approaches are fully implemented in the ProCAKE framework (procake.uni-trier.de) and applied to data mining workflows in the RapidMiner scientific workflow system:

Total View of the Workflow

Nested Sub-workflows of the Cross Validation Task
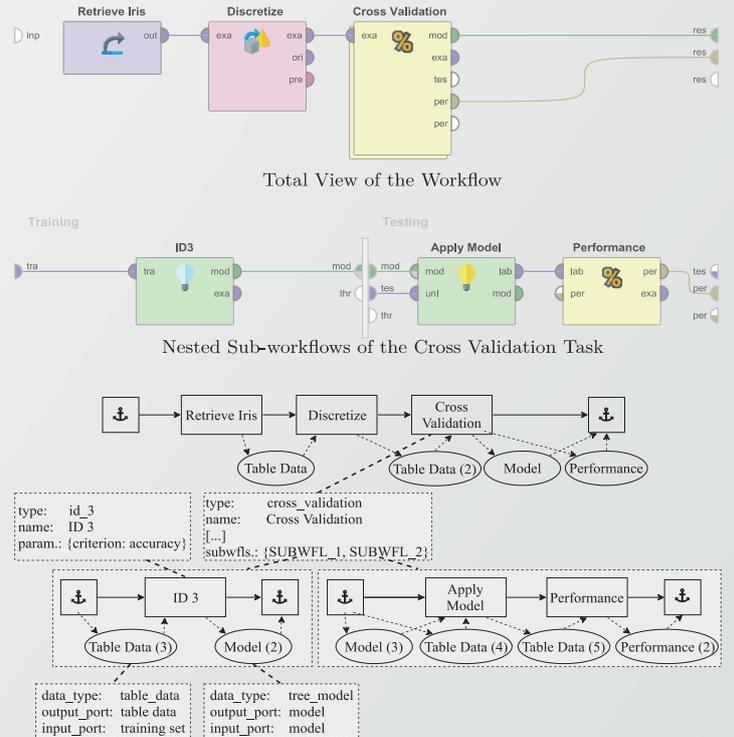
Experimental Evaluation

Hypotheses

H1 Each adaptation method provides at least as good results as the sole retrieval

H2 The structural adaptation method outperforms the substitutional adaptation method

H3 The combined adaptation method provides better results as both, the structural and the substitutional adaptation method

H4 The adaptation methods produce executable and semantically correct workflows

Setup

- Experiments with 20 cases and 10 queries
- k-NN retrieval with subsequent adaptation

Results

Adaptation approaches significantly improve workflows w.r.t. the queries while mostly maintaining executability and semantic correctness

		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Avg
w/o		0.82	0.60	0.85	0.85	0.91	0.81	0.78	0.85	0.71	0.72	0.79
struct.	$k = 1$			1.00			1.00 ✗	1.00			0.86 ✗	0.86
	$k = 2$	1.00		1.00	1.00		1.00 ✗	1.00		0.91	0.86 ✗	0.91
	$k \geq 3$	1.00	0.69	1.00	1.00		1.00 ✗	1.00		0.91	0.86 ✗	0.92
subst.	$k \geq 1$	1.00	0.68		1.00		1.00 ✗	1.00		0.91	0.91	0.91
comb.	$k \geq 1$	1.00	0.69	1.00	1.00		1.00 ✗	1.00		0.91	1.00 ✗	0.94