

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques

Andreas Rentschler, Dominik Werle, Qais Noorshams, Lucia Happe, Ralf Reussner

3rd Workshop on the Analysis of Model Transformations Monday, September 29, 2014

SOFTWARE DESIGN AND QUALITY GROUP INSTITUTE FOR PROGRAM STRUCTURES AND DATA ORGANIZATION, FACULTY OF INFORMATICS



KIT - University of the State of Baden-Wuerttemberg and National Research Center of the Helmholtz Association

sdq.ipd.kit.edu

www.kit.edu





Performance model of a component-based software architecture

Approach 000000

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques - Andreas Rentschler et al.



Performance data: Execution time Throughput **Resource utilisation**

Validation O



Approach 000000

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques - Andreas Rentschler et al.

Validation O



Approach 000000

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Validation O



Motivation \odot \bigcirc

Approach 000000

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques - Andreas Rentschler et al.

Validation O



Approach 0 0 0 0 0 0 0 0

Validation O s Rentschler et al.

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Approach 0 0 0 0 0 0 0 0

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation O



Motivation

Approach 0 0 0 0 0 0 0 0

Validation O s Rentschler et al.

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Motivation

Approach 0 0 0 0 0 0 0 0

Validation O

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Motivation

Approach 0 0 0 0 0 0 0 0

Validation O s Rentschler et al.

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



- How can we support typical transformation designs?
- What dependence information is required?

Motivation

Approach 000000

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Validation O

- How can we support typical transformation designs?
- What dependence information is required?

Validation O Motivation Approach 000000 Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Design Rules

What's makes model transformations different from GPL programs?

- Data-centric operations
- Data is hierarchically structured
- Data models extrinsically defined

Common decompositional styles [Lawley04]:

Source-driven



one-to-many mappings

Target-driven



many-to-one mappings, M2T templates

Motivation
•

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.







a mixture of both

Validation O

Motivation

Approach $\bullet \bullet \circ \circ \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation O

Source Model



Motivation • •

Approach $\bullet \bullet \circ \circ \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation O

Source Model



Motivation
•

Approach $\bullet \odot \circ \circ \circ \circ \circ$

Validation O s Rentschler et al.

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Target Model





Motivation • •

Approach $\bullet \odot \circ \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Validation O



Motivation •

Approach $\bullet \odot \circ \circ \circ \circ \circ$

Validation O

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Motivation • •

Approach $\bullet \odot \circ \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

- Validation O

ActivityModel Activity2Process mapping⁻ in Activity mapActivity2Process : call Action2Step actions mapping in Action mapAction2Step successors v i ▲ call mapping in mapAction2Step **StopAction** ♥:: mapping in **StartAction** mapAction2Step CompositeAction2Step-CompositeActions mappingactions in mapAction2Step Composite Action helper ; in createProcess

Motivation • •

Source Model

Approach $\bullet \odot \circ \circ \circ \circ \circ$

Validation O

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Transformation



Target Model



The Bunch Clustering Approach

(Bunch v3.5 - May 2013	
F	ile Utility Help		
	Options:		
	Basic Clustering C	Options Libraries Omnipresent Use	r Directed Clus
	Input Graph File:	0-10-OFF W5=W6)/Activity2ProcessModule.r	ndg
	Clustering Method:	Hill Climbing	-
	Output Cluster File:	90-10-OFF W5=W6)/Activity2ProcessModule	.dot
	Output File Format:	Dotty	-
			Gener
	Action:		
	Agglomerative Cluste	ring 🗨	

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, ... [Shtern12]

Motivation

Motivation

Approach $\bullet \bullet \odot \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

ז*

Legacy Transformation Deper Analy	Manual Decomposition Indency sis	Modularized Transformation
Control and Data Dependencies	Cluster Analysis	Clusters



Validation O

The Bunch Clustering Approach

O File Utility Help	Bunch	v3.5 - May 2013		
Options: Basic Clustering C Input Graph File: Clustering Method: Output Cluster File: Output File Format: Action: Agglomerative Cluster	Dptions Libraries 10-10-OFF W5=W6)/A Hill Climbing 90-10-OFF W5=W6)/ Dotty	Omnipresent Activity2ProcessMo Activity2ProcessM	User Direct dule.mdg odule.dot	cted Clus Genera
\				

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, ... [Shtern12]

Motivation

Motivation

Approach $\bullet \bullet \odot \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

ר*

Legacy Transformation Deper Analy	Manual Decomposition Indency sis	Modularized Transformation
Control and Data Dependencies	Cluster Analysis	Clusters



Validation O

The Bunch Clustering Approach*

O File Utility Help	Bunch	v3.5 - May 2013		
Options: Basic Clustering C Input Graph File: Clustering Method: Output Cluster File: Output File Format: Action: Agglomerative Cluster	Dptions Libraries 10-10-OFF W5=W6)/A Hill Climbing 90-10-OFF W5=W6)/ Dotty	Omnipresent Activity2ProcessMo Activity2ProcessM	User Direct dule.mdg odule.dot	cted Clus Genera
\				

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, ... [Shtern12]

Motivation

Motivation

Approach $\bullet \bullet \odot \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation O

The Bunch Clustering Approach*

Clustering methods: hill climbing, genetic, exhaustive

	 File Utility Help Options: 	Bunch	v3.5 - May 2013	3	
15	📔 🛛 🖉 Basic 🕺 Clustering	Options Libraries	Omnipresent	User Direct	ted Clus
	Input Graph File: Clustering Method: Output Cluster File: Output File Format:	0-10-OFF W5=W6)/A Hill Climbing 90-10-OFF W5=W6)/	Activity2ProcessMo	odule.mdg	
					Genera
	Action: Agglomerative Clust	ering 🔻			
	\				

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, ... [Shtern12]

Motivation
•

Approach $\bullet \bullet \odot \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.







Validation O

The Bunch Clustering Approach*

Clustering methods: hill climbing, genetic, exhaustive

		Bunch	v3.5 - May 2013	}
	File Utility Help			
	Options:			
15	Basic Clustering	Options Libraries	Omnipresent	User Directed Clus
	Input Graph File:	0-10-OFF W5=W6)/A	Activity2ProcessMo	dule.mdg
	Clustering Method:	Hill Climbing		-
	Output Cluster File:	90-10-OFF W5=W6)/	Activity2ProcessM	lodule.dot
	Output File Format:	Dotty		~
				Genera
	Action:			
	Agglomerative Cluste	ering 🔽		

Modularity quality index: favor low coupling & high cohesion

$$MQ = \sum_{i=1}^{k} CF_{i}, \quad CF_{i} = \begin{cases} 0, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{1}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,i})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}{2} \sum_{\substack{j=1 \ j \neq i}}^{k} (\epsilon_{i,j} + \epsilon_{j,j})}, & \frac{\mu_{i}}{\mu_{i} + \frac{\mu_{i}}$$

* by [Mitchell06]; alternative approaches are ARCH, ACDC, LIMBO, ... [Shtern12]

Motivation
•

Approach $\bullet \bullet \odot \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.







$$\mu_i = 0$$
otherwise

Validation O



Motivation • •

Approach $\bullet \bullet \bullet \bullet \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Dependence Analysis



Motivation • •

Approach $\bullet \bullet \bullet \bullet \circ \circ \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation O

Weight Configuration

Depending on the style...

Source-driven





...initialize a configuration vector for weighting the various types of edges:

 $WC := \langle W_{write}, W_{read}, W_{call}, W_{package} \rangle \in \mathbb{N}_0^4$

Method invokes another method

Method reads model element

Method creates/updates model element

Motivation

Approach $\bullet \bullet \bullet \bullet \bullet \odot \circ \circ$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Model element is contained in package

Validation O

Assigning Weights







Assigning Weights







Assigning Weights









Motivation • •

Approach $\bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet \bullet$

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques - Andreas Rentschler et al.

Validation O

- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

Motivation
•

Approach • • • • • •

Validation

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Legacy Transformation	Manual Decomposition	Modularized
Depe	ndency sis	
Control and Data Dependencies	Cluster Analysis	Clusters

- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics



Motivation • •

Approach • • • • • •

Validation

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

Configuration		atistics	Similarity to expert clustering			
	*	There indet	Precision	Recall	Fidee St	n Neci
Expert clustering Derived manually	3	1.067	100%	100%	100	100%
Method-call dependencies only Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$	2	1.214	20.00%	100%	54.54	60%

Motivation
•

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation

- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

Configuration		atistics	Similarity to expert clustering			
		inters indet	arecision	ecall	- deest	n NeCi
	*	Ż,	<i>Ş</i> *	\$~		<i>L</i> ,
Expert clustering Derived manually	3	1.067	100%	100%	100	100%
Method-call dependencies only Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$	2	1.214	20.00%	100%	54.54	60%
Class-level dependencies Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$	2	1.083	33.33%	100%	72.72	85%

Motivation
•

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation

- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

Configuration	Sta	atistics	Similarity to expert clustering			
	*	No indet	Precision	Recall	Ridse Si	n Neci
Expert clustering Derived manually	3	1.067	100%	100%	100	100%
Method-call dependencies only Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$	2	1.214	20.00%	100%	54.54	60%
Class-level dependencies Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$	2	1.083	33.33%	100%	72.72	85%

Motivation
•

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation

- Compare derived clustering with a manual expert clustering
- Using three similarity/distance metrics

Configuration	Sta	atistics	Similarity to expert clustering			
	X	onsters indet	Precision	Recall	Rideesi	n Neci
Expert clustering Derived manually	3	1.067	100%	100%	100	100%
Method-call dependencies only Hill Climbing, $WC = \langle 0, 0, 1, 0 \rangle$	2	1.214	20.00%	100%	54.54	60%
Class-level dependencies Hill Climbing, $WC = \langle 1, 15, 5, 15 \rangle$	2	1.083	33.33%	100%	72.72	85%

Significantly better results with model elements considered

Motivation
•

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.



Validation

Problem: Model transformation programs are often badly structured.

Motivation
•

Approach • • • • • •

Validation

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Problem: Model transformation programs are often badly structured. **Approach**: Apply automatic clustering algorithms to do the job.



Motivation
•

Approach • • • • • •

Validation

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Problem: Model transformation programs are often badly structured. **Approach:** Apply automatic clustering algorithms to do the job.



- By including the model structure, we can derive clusterings that Findings: follow source and target-driven decompositional styles.
 - Selecting 'good' weights requires expertise and/or experimenting.

Motivation

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Validation

Problem: Model transformation programs are often badly structured. **Approach:** Apply automatic clustering algorithms to do the job.



- By including the model structure, we can derive clusterings that Findings: follow source and target-driven decompositional styles.
 - Selecting 'good' weights requires expertise and/or experimenting.

Open challenges:

- Can we support aspectual decompositions?
 - How do the results compare in maintenance scenarios?

Motivation

Approach • • • • • •

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Rentschler et al.

Validation

Problem: Model transformation programs are often badly structured. **Approach:** Apply automatic clustering algorithms to do the job.



- By including the model structure, we can derive clusterings that Findings: follow source and target-driven decompositional styles.
 - Selecting 'good' weights requires expertise and/or experimenting.

Open challenges:

- Can we support aspectual decompositions?
 - How do the results compare in maintenance scenarios?

Motivation • •

Approach • • • • • •



schlei

Remodularizing Legacy Model Transformations with Automatic Clustering Techniques – Andreas Re