Optimizing the Trade-Off between Complexity and Conformance in Process Reduction

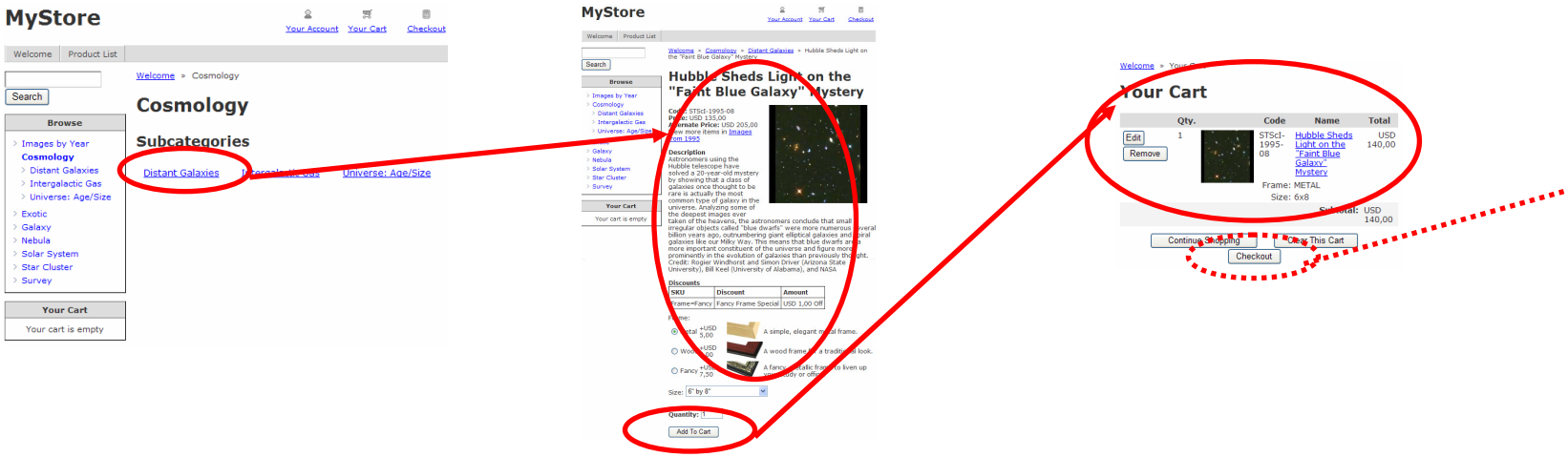
Alessandro Marchetto, Chiara Di Francescomarino, Paolo Tonella
Fondazione Bruno Kessler – SE unit
{marchetto, dfmchiara, tonella}@fbk.eu

Outline

- Process recovery for web applications
- Process reduction and (multi-objective) optimization
- Case Study
- Conclusions, Ongoing and Future works

Process Recovery: *the idea*

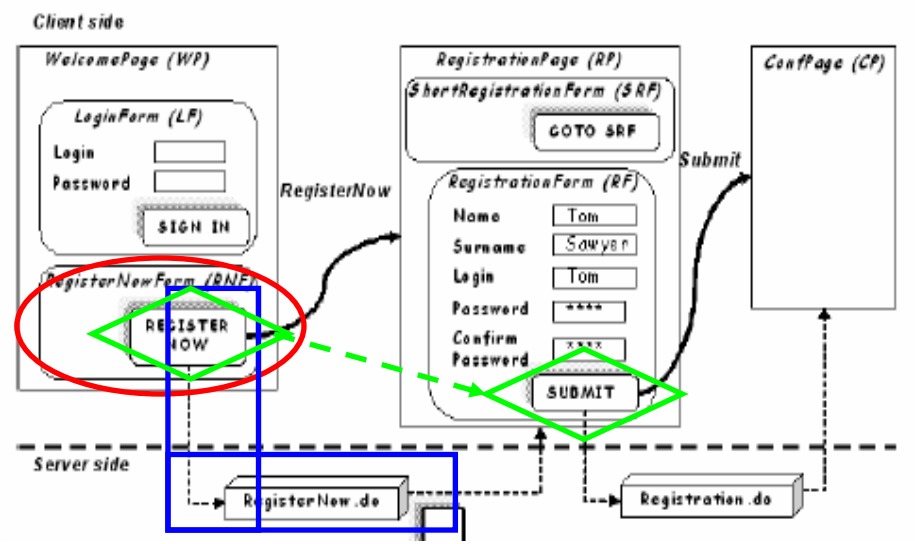
- Web Applications (WAs) are often the preferred way to expose business processes;
- The implemented process is however rarely documented;
- WAs are UI-oriented → The process underlying a WA is guided by the user in her interaction with the WA GUI



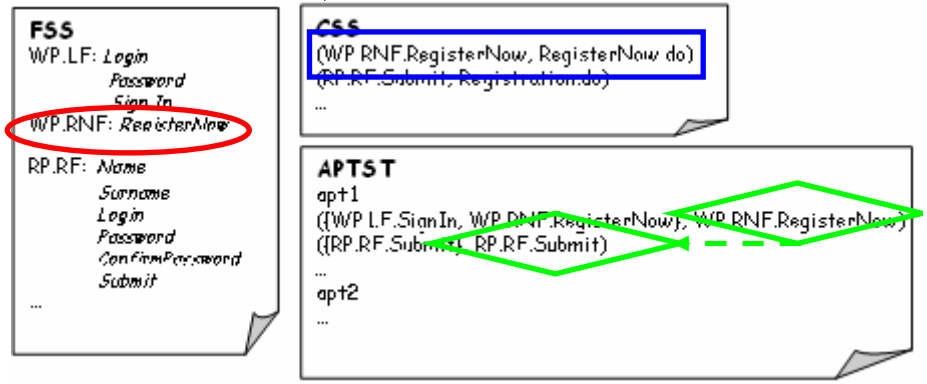
→ By capturing the information about the activated buttons, links and forms the application process can be inferred.

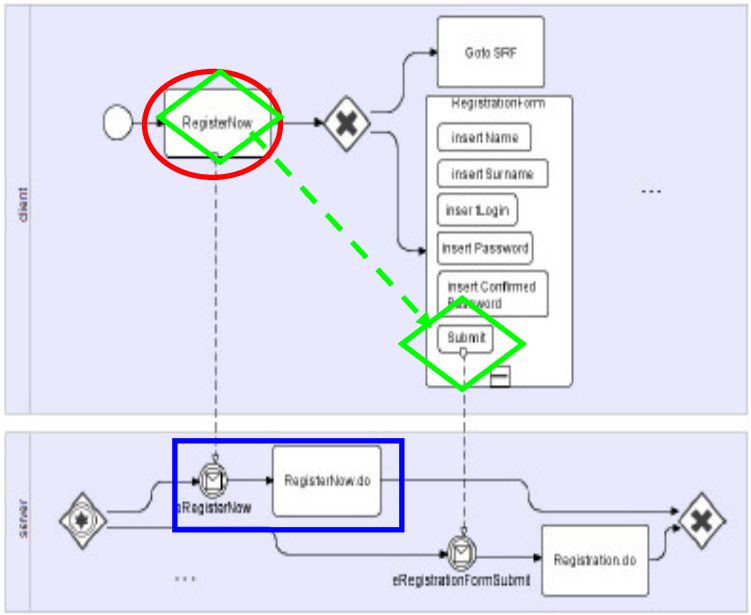
[CSMR09, JSME10]

Process Recovery: *by example*



Execution tracing to collect information



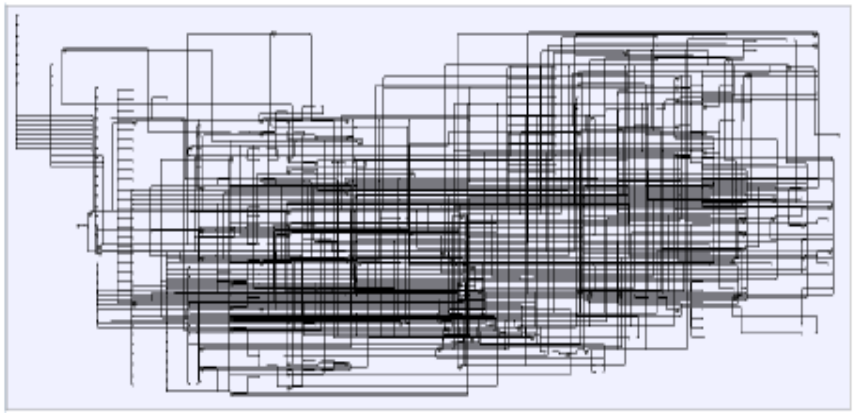


Process extraction

[CSMR09, JSME10]

Limit and existing solutions

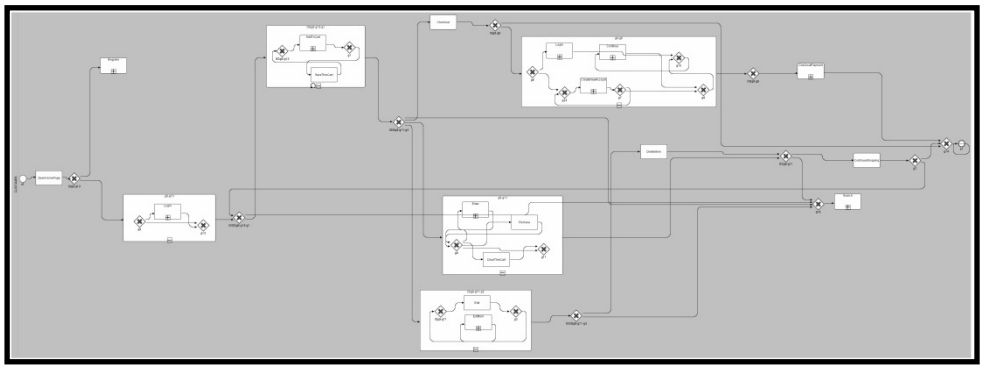
(1) Under-generalization → consider more traces
(2) Over-generalization → improve the model recovery algorithm
(3) Size and complexity of recovered processes



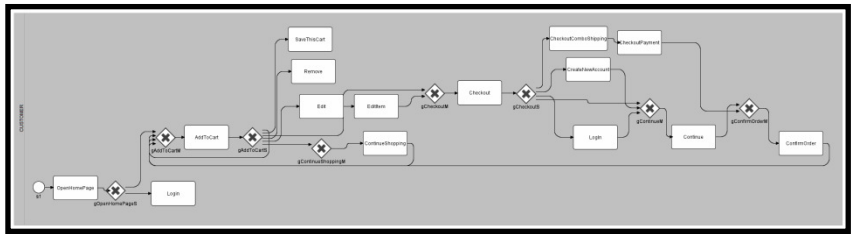
(1) they reproduce all traced behaviors
(2) they generalize

→ large, complex and intricate processes

Modularization/Clustering



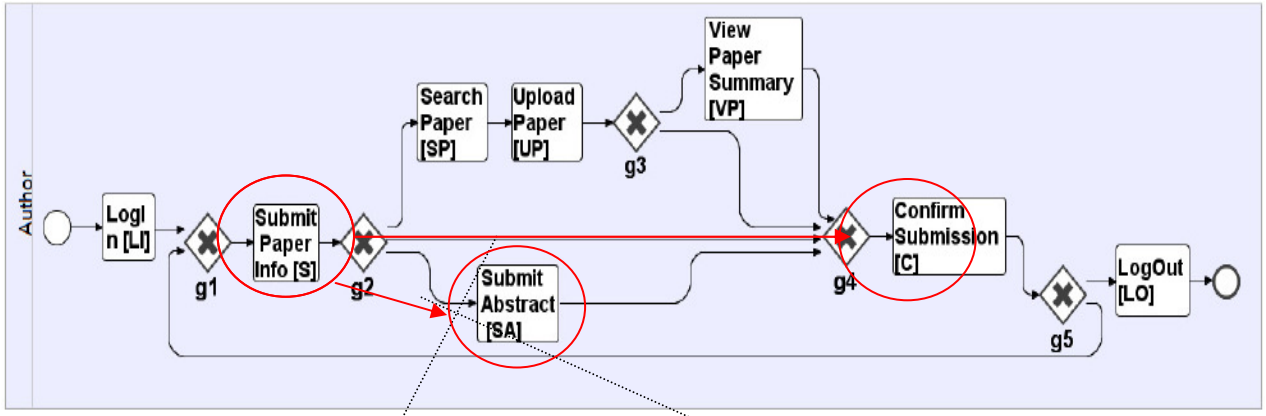
Frequency-based reduction (FBR)



[CSMR09, JSME10]

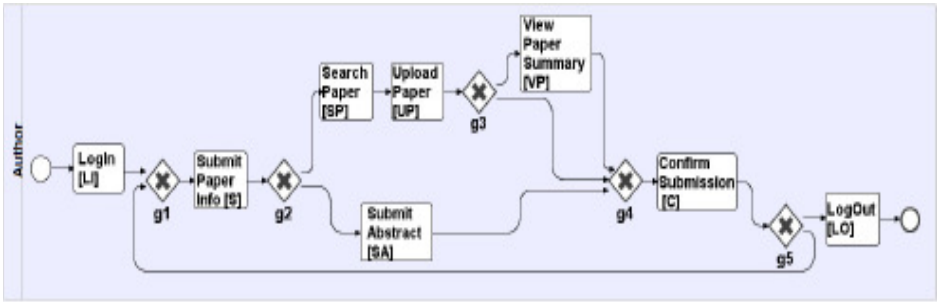
Process Reduction: *by example*

Recovered process models are **reduced** by means of **atomic reduction operations** applied to the process elements.

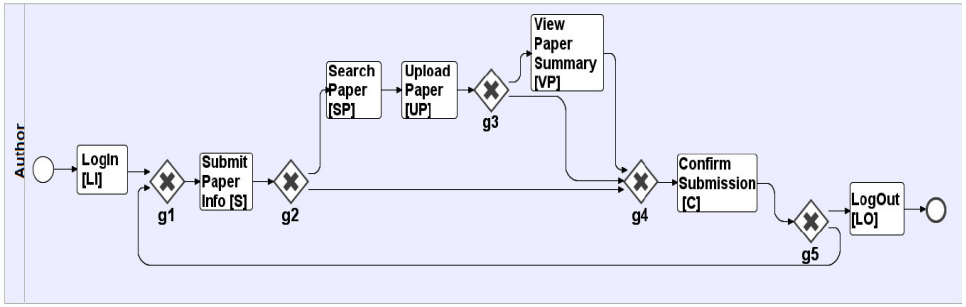


An **atomic reduction** operation is the removal of a transition from the process, followed by the removal of unreachable elements.

removal of the transition (S,C)

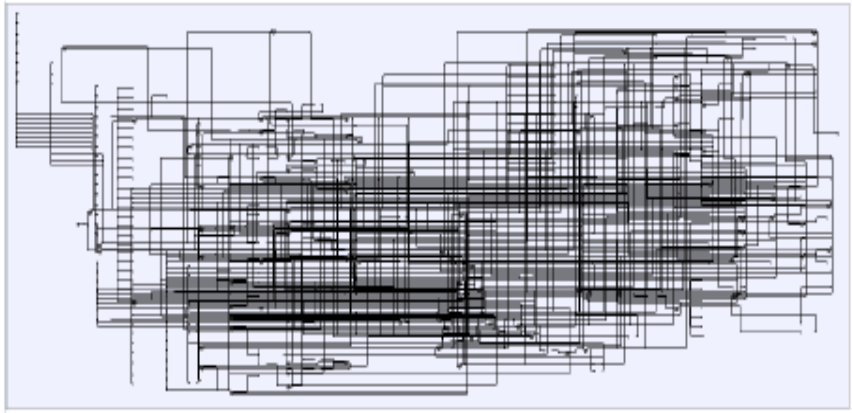


removal of the transition (S, SA)



Multi-objective optimization

Which elements can be removed ?



Reducing recovered processes is a BI-DIMENSIONAL problem

Improving process model comprehensibility (i.e., reducing process model size and complexity)

… without loosing business information

→ Multi-objective Optimization Problem

Multi-objective Optimization of the process reduction

Non-dominated Sorting Genetic Algorithm II (NSGA-II)

Solution Encoding: a solution is a process in which some edges are kept while other removed. A standard binary encoding (binary vector) is used.

Initialization: (a) random; and (b) frequency-based reduction (FBR)

Genetic Operators: bit-flip mutation, one-point crossover, and binary tournament for the selection.

Fitness Functions: each process is evaluated in terms of process complexity and non-conformance.

Process quality: a bi-dimentional problem

The **process control flow complexity**

$$CFC(P) = \sum_{g \in G(P) \wedge FOUT(g) > 1} FOUT(g)$$

A high value of CFC indicates a high number of alternative execution flows, thus denoting a process difficult to read and understand.

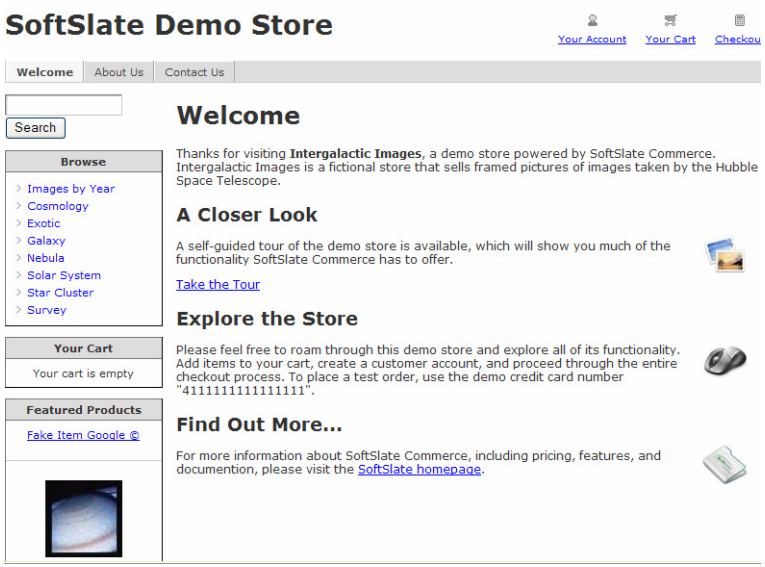
The **process non-conformance** is its inability to reproduce the execution traces

$$NConf(P) = \left| \bigcup_{t \in T} \{(a,b) | (a,b) \in t \wedge dc(a,b) \notin P\} \right|$$

A high value of NConf indicates that the process is not able to reproduce many transitions in the traces.

Case Study: *the application*

Softslate Commerce is a Java-based (>200k LOCs) shopping cart application for managing on-line stores. It implements, e.g., catalogue, cart, order form, payment and checkout management;



Research questions

quantitative

RQ1 : Does the shape of the Pareto fronts offer a set of solutions which includes a wide range of tunable trade-offs between complexity and conformance?

RQ2 : Does the genetic algorithm improve the initial solutions (both random and frequency-based reduction)?

qualitative

RQ3 : Are the reduced processes in the Pareto front understandable and meaningful for business analysts?

RQ4 : Do the processes obtained by applying multi-objective optimization offer qualitative improvements over those obtained by applying the frequency-based reduction (FBR)?

Procedure

(1) We **trace some executions** of Softslate Commerce by exercising each application functionality at least once.

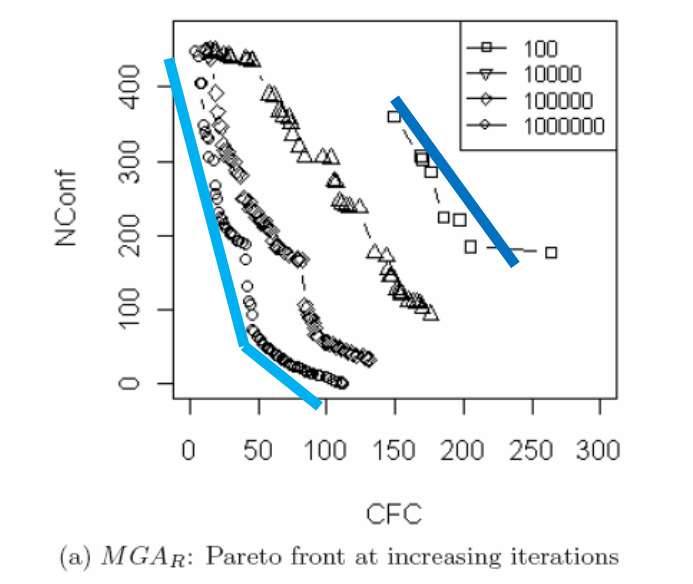
(2) We recover the **unreduced process model**.

(3) We reduce the unreduced process model by applying **FBR**. A set of solutions is obtained by varying the frequency threshold.

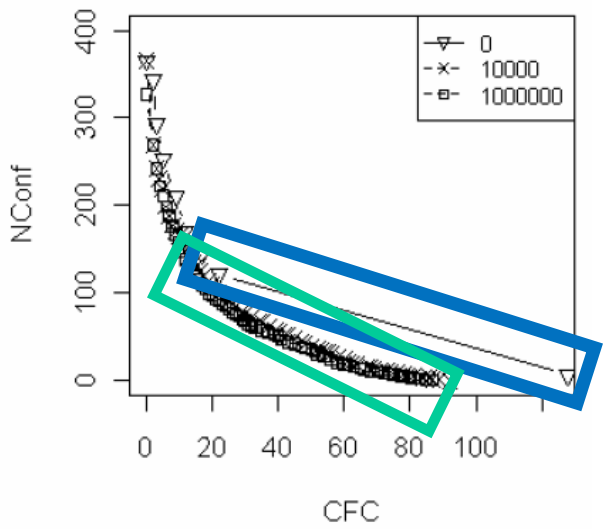
(4) We reduce the unreduced process model by applying the **multi-objective reduction** (MGA):

- MGA_R: random initial population;
- MGA_F: FBR-based initial population.

Quantitative Analysis



(a) MGA_R: Pareto front at increasing iterations



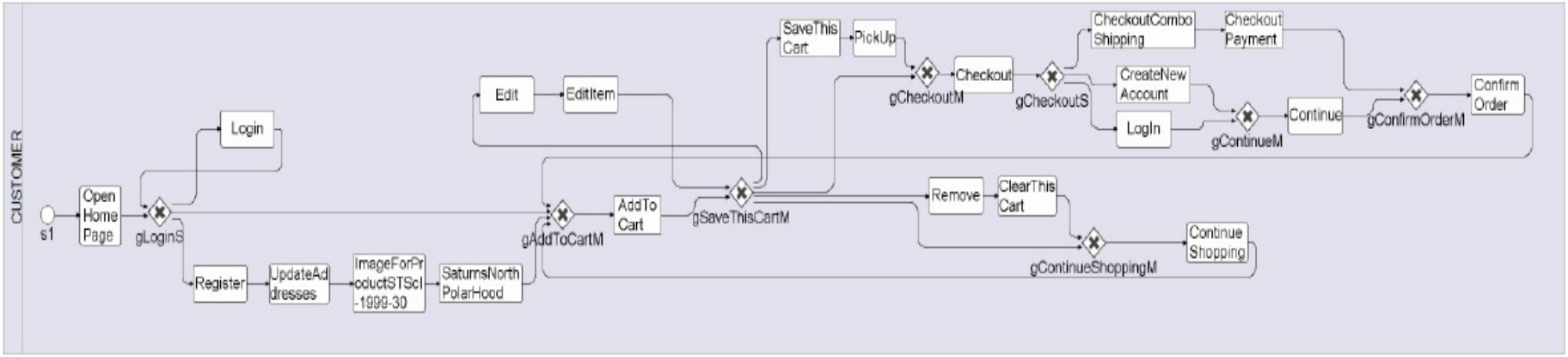
(b) MGA_F: Pareto front at increasing iterations

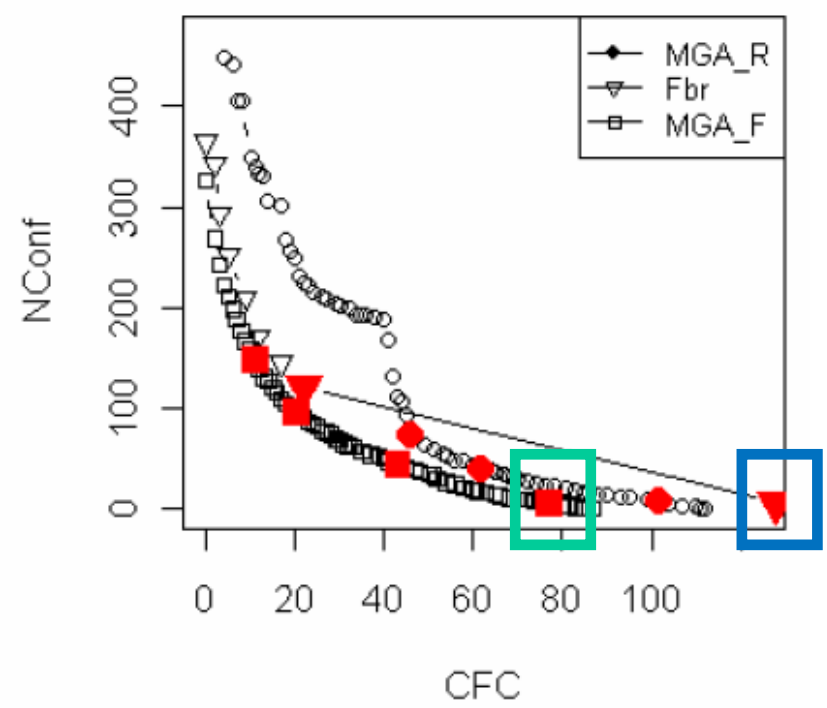
Algo	Iter	Time (min.)
MGA_R	100	1.18
MGA_R	10000	39.3
MGA_R	100000	294.2
MGA_R	1000000	2097.9
MGA_F	0	<0.2
MGA_F	100	1.24
MGA_F	10000	12.1
MGA_F	1000000	1077.2
Fbr	-	-

RQ1: MGA produces a Pareto front which includes a wide range of tunable solutions. However, this requires a high enough number of iterations and a carefully initialized starting population (via FBR).

RQ2: MGA improves both random and FBR solutions, the latter to a lower degree.

Qualitative analysis





(c) Selected processes in the final Pareto fronts

Pr. id	#Act	#G	#SF	CFC	NConf
unreduced					
-	213	105	581	502	0
MGA_R					
p24	31	25	104	46	74
p25	39	37	125	62	40
p66	48	50	168	102	7
MGA_F					
p23	21	8	52	11	148
p27	23	15	60	20	97
p67	36	25	97	43	44
p41	46	38	142	77	6
Fbr					
p4	19	16	51	22	122
p15	46	44	188	128	6

RQ3: meaningful processes

Business activities	
... a3.Add To Cart; ... a7.Clear This Cart; a8. Check-out; ... a10.Delete Item; a11.Log In; a12. Create New Account; ... a14.Confirm Order; ... a19.Reorder	
Business properties	
pr2	Remove has to follow Add To Cart, Pick Up or Re-order
pr3	After Edit it has to be possible to Edit Item or to Delete Item
pr5	Delete Item has to follow Add To Cart, Pick Up or Reorder
pr15	After Pick Up it has to be possible to choose among Edit, Remove, Continue Shopping, Save This Cart, Clear This Cart and Checkout
pr19	Confirm Order has to follow Checkout

unreduced	MGA_R			*Fbr*		MGA_F			
	p24	p25	p66	p4	p15	p23	p27	p41	p67
Business Activities									
19	17	18	19	15	19	15	16	19	18
Business Properties									
19	19	21	25	18	24	19	22	25	25

19 activities & 28 properties

RQ3 :

- Few missing business activities and properties in the reduced processes.
- Reduction is almost always beneficial thanks to its implicit capability of reducing over-generalization.

RQ4: (comparing "similar" processes generated by FBR and MGA_F)

- MGA is more effective in both the the two dimensions.
- MGA offers also a wider range of alternative solutions.

Conclusions, Ongoing and Future Works

Multi-objective optimization for process model reduction by balancing complexity and conformance.

Results of the case study:

(1) MGA produces a rich, fine grained, evenly distributed set of alternatives;

(2) though reduced, processes produced by MGA include relevant business activities and properties;

Future works will be devoted to perform further experiments, involving additional case studies.

Thanks …

[SSBSE11] Marchetto A., Di Francescomarino C., Tonella P.,
Optimizing the Trade-Off between Complexity and Conformance in Process Reduction.
In proc. of Search Based Software Engineering (2011)

[JSME10] Di Francescomarino, C., Marchetto, A., Tonella, P.,
Cluster-based Modularization of Processes Recovered from Web Applications.
Journal of Software Maintenance and Evolution: Research and Practice (2010)

[CSMR09] Di Francescomarino, C., Marchetto, A., Tonella, P.,
Reverse Engineering of Business Processes exposed as Web Applications.
In proc of European Conference on Software Maintenance and Reegnineering (2009)