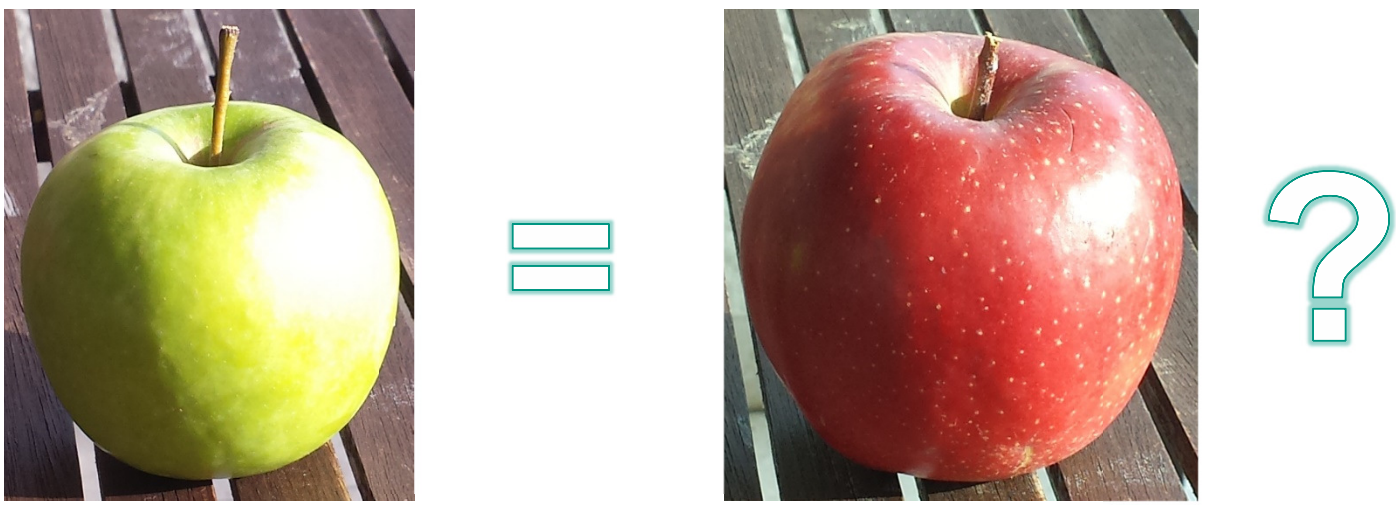
A Comparative Analysis of Business Process Model Similarity Measures

Tom Thaler, *Andreas Schoknecht*, Peter Fettke, Andreas Oberweis, and Ralf Laue

1st International Workshop on Process Querying

Karlsruhe Institute of Technology (KIT),
Institute of Applied Informatics and Formal Description Methods



Motivation

- Similarity measures are used for various reasons
  - Conformance checking
  - Reuse
  - Similarity-based search

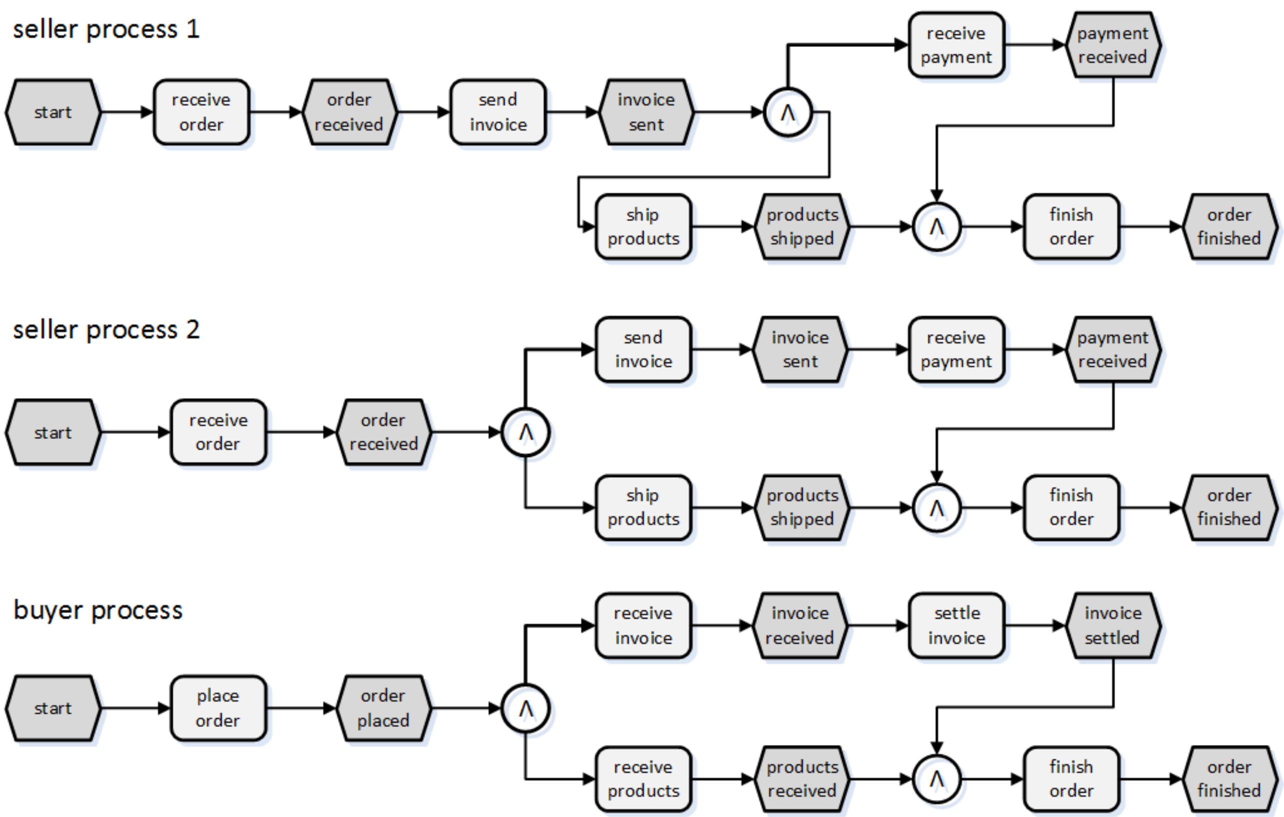
- **Interpretation** of similarity is **quite different**

- Research Questions

  (1) How do the values of existing similarity measures correlate?

  (2) How do existing implementations perform and what does that imply for their practical usage?

Dimensions of Similarity Measurement



- Natural Language
- Graph Structure
- Behavior
- Human Estimation
- Other

Analysis Methodology (1)

- Field models
  - No restrictions regarding labeling
  - University Admission and Birth Registration models from Process Model Matching Contest 2013 (18 models)

Cayoglu, U. et al.: The Process Model Matching Contest 2013. BPM Workshops, pp. 442-463 2014

- Controlled modelling environment
  - Models based on natural language text description
  - Student exercise (8 models)

http://rmm.dfki.de

- Mined models
  - Linguistically harmonized labels
  - Dutch governance models (80 models)

Vogelaar, J. et al.: Comparing Business Processes to Determine the Feasibility of Configurable Models: A Case Study. BPM Workshops. pp. 50-61 2011

Analysis Methodology (2)

- 8 similarity measure implementations could be used
- Dimensions used by similarity measures

Dimension / Ref.	[1]	[7]	[10]	[17]	[23]	[15]	[16]	[12]
Natural language	syn	syn	syn	syn	syn	syn	syn+sem	syn
Graph structure		x	x	x	x	x	x	
Behavior								x

- All measures base on matches between process models
- Varying complexity of similarity calculation

Analysis Results – Underlying Matching Quality

	field						controlled						mined					
approach	TP	FP	FN	P	R	F	TP	FP	FN	P	R	F	TP	FP	FN	P	R	F
[1][10][17]	152	17	962	0.9	0.14	0.24	6	0	284	1.00	0.02	0.04	9,767	0	0	1.00	1.00	1.00
[7]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[23]	315	906	799	0.27	0.28	0.27	125	228	165	0.35	0.43	0.39	9,554	50,354	213	0.16	0.98	0.27
[15]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[16]	289	205	825	0.59	0.26	0.36	96	33	194	0.74	0.33	0.46	9,767	9,187	0	0.52	1.00	0.68
[12]	175	20	939	0.90	0.16	.027	19	1	271	0.95	0.07	0.12	9,767	1,257	0	0.89	1.00	0.94

sim = similarity measure, TP = true positives, FP = false positives, FN = false negatives, P = μ-average of precision, R = μ-average of recall, F = μ-average of f-measure.

Analysis Results – Correlation Values

		[1]	[7]	[10]	[17]	[23]	[15]	[16]	[12]
[1]	F	1.00	0.93	0.77	0.97	0.70	0.96	0.94	0.94
	C	1.00	0.94	0.80	0.98	0.43*	0.97	0.96	0.92
	M	1.00	0.96	0.85	0.98	0.60	0.95	0.93	#
[7]	F	0.93	1.00	0.91	0.98	0.62	0.97	0.98	0.98
	C	0.94	1.00	0.91	0.95	0.49*	0.98	0.94	0.97
	M	0.96	1.00	0.87	0.99	0.55	0.94	0.98	#
[10]	F	0.77	0.91	1.00	0.86	0.55	0.85	0.87	0.85
	C	0.80	0.91	1.00	0.95	0.49	0.98	0.89	0.96
	M	0.85	0.87	1.00	0.85	0.49	0.82	0.84	#
[17]	F	0.97	0.98	0.86	1.00	0.65	0.98	0.98	0.97
	C	0.98	0.95	0.95	1.00	0.56*	0.99	0.97	0.91
	M	0.98	0.99	0.85	1.00	0.57	0.94	0.97	#
[23]	F	0.70	0.62	0.55	0.65	1.00	0.75	0.63	0.63
	C	0.43*	0.49*	0.49	0.56*	1.00	0.55	0.65	0.55*
	M	0.60	0.55	0.49	0.57	1.00	0.76	0.57	#
[15]	F	0.96	0.97	0.85	0.98	0.75	1.00	0.97	0.96
	C	0.97	0.98	0.98	0.99	0.55	1.00	0.97	0.95
	M	0.95	0.94	0.82	0.94	0.76	1.00	0.93	#
[16]	F	0.94	0.98	0.87	0.98	0.63	0.97	1.00	0.98
	C	0.96	0.94	0.89	0.97	0.65	0.97	1.00	0.96
	M	0.93	0.98	0.84	0.97	0.57	0.93	1.00	#
[12]	F	0.94	0.98	0.85	0.97	0.63	0.96	0.98	1.00
	C	0.92	0.97	0.96	0.91	0.55*	0.95	0.96	1.00
	M	#	#	#	#	#	#	#	#

p-value $\leq 1\%$, F = field models, C = controlled models, M = mined models, # = calculation aborted because of memory overflow, * = p-value $> 1\%$.

Pearson correlation coefficients

High correlation between all measures except [23]

High correlation although different dimensions used

Matching quality does not influence correlation

Similarity measures seem to be exchangeable

Analysis Results – Run time

Measure	Dutch Governance	Student exercises	Birth registration	University admission
[1]	3:28 min	0:00 min	0:02 min	0:02 min
[7]	8:40 min	0:01 min	0:04 min	0:05 min
[10]	n/a[1]	0:37 min	9:32 min	26:30 min
[17]	8:40 min	0:02 min	0:04 min	0:05 min
[23]	45:37 min	0:03 min	0:23 min	0:56 min
[15]	40:21 min	0:03 min	0:15 min	0:36 min
[16]	39:22 min	0:14 min	0:20 min	0:22 min
[12]	memory overflow	0:03 min	0:07 min	4:52 min

[1] For [10] the Dutch Governance processing had to be split because of a memory overflow. Since summing up the partial run times might have led to a corruption in comparison to the other calculations, it was decided to state it as not available.

Fast calculation for small model sets

Bigger model sets problematic

Depending on practical application calculation time might be to high

Discussion and Limitations

- Limited availability of implementations
- Underlying matching
  - Similarity values depend on matches determined by the different measures
  - Possibly repeat experiment with consistent matching
- No comparison with similarity measures not requiring matching

Conclusion

Karlsruhe Institute of Technology

- Similarity calculation is basically a two step process: (1) determine node matches, (2) calculate similarity value

- Analysis results
  - High correlation values between all analyzed measures except one
  - Run time for bigger model sets partially quite high

- Open questions
  - Do two measures measure the same pragmatic aspects?
  - How do automatic similarity measures resemble human similarity estimation?

Thank you for your attention!

QUESTIONS?