

Predicting the Quality of Making Process Model Matching Work Process Model Matching

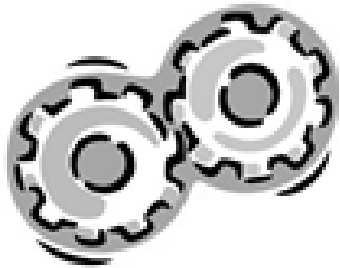
Matthias Weidlich, Tomer Sagi, Henrik Leopold,
Avigdor Gal, and Jan Mendling

Technion - Israel Institute of Technology, Haifa, Israel

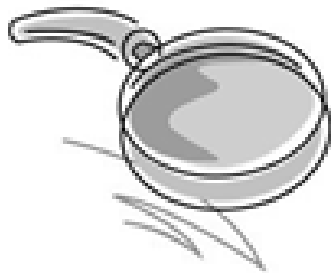
Humboldt-Universität zu Berlin, Germany

Wirtschaftsuniversität Wien, Austria

Matching Use Cases



Validation of process implementation

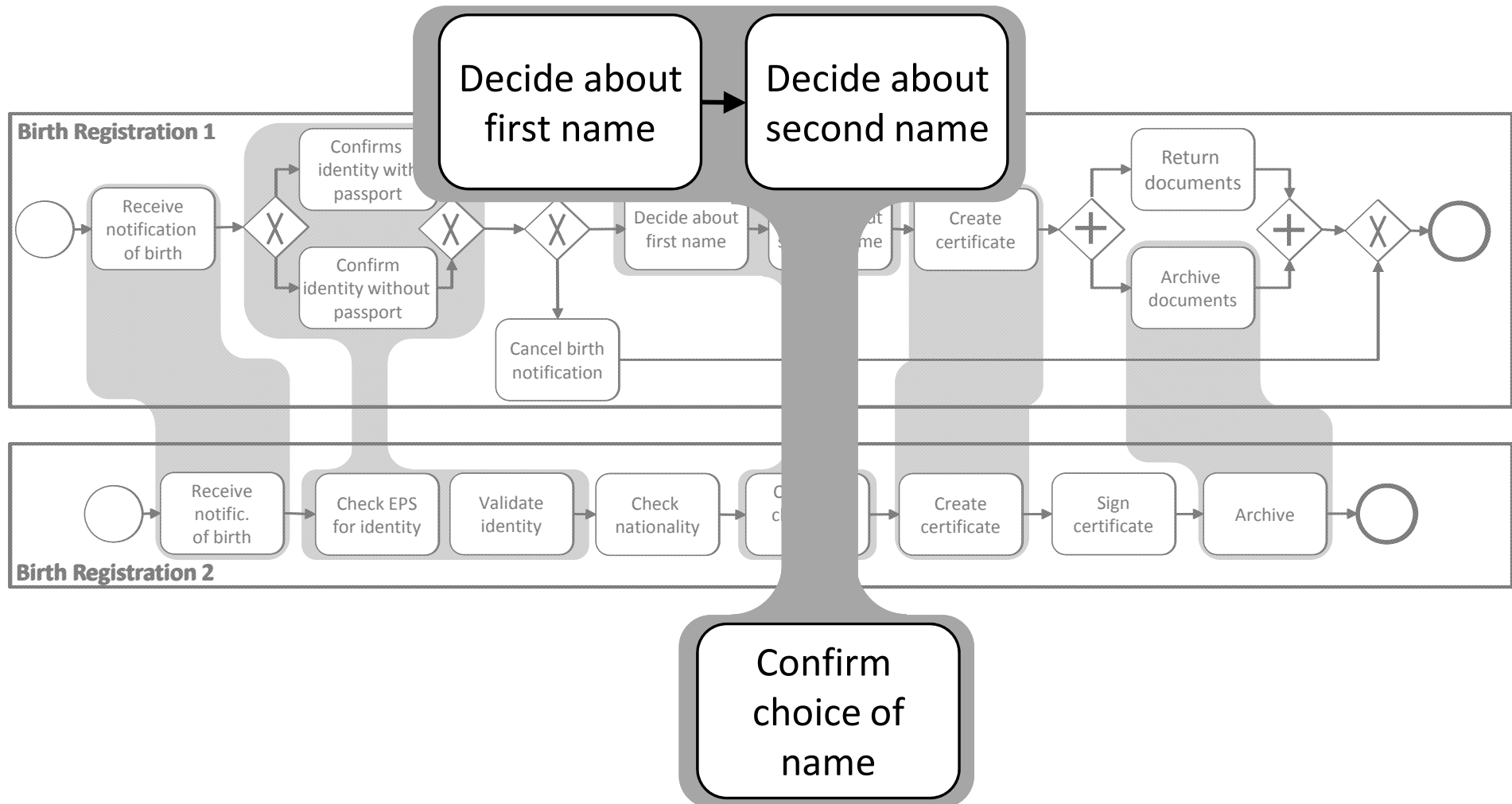


Similarity search



Managing process model collections

The Matching Problem



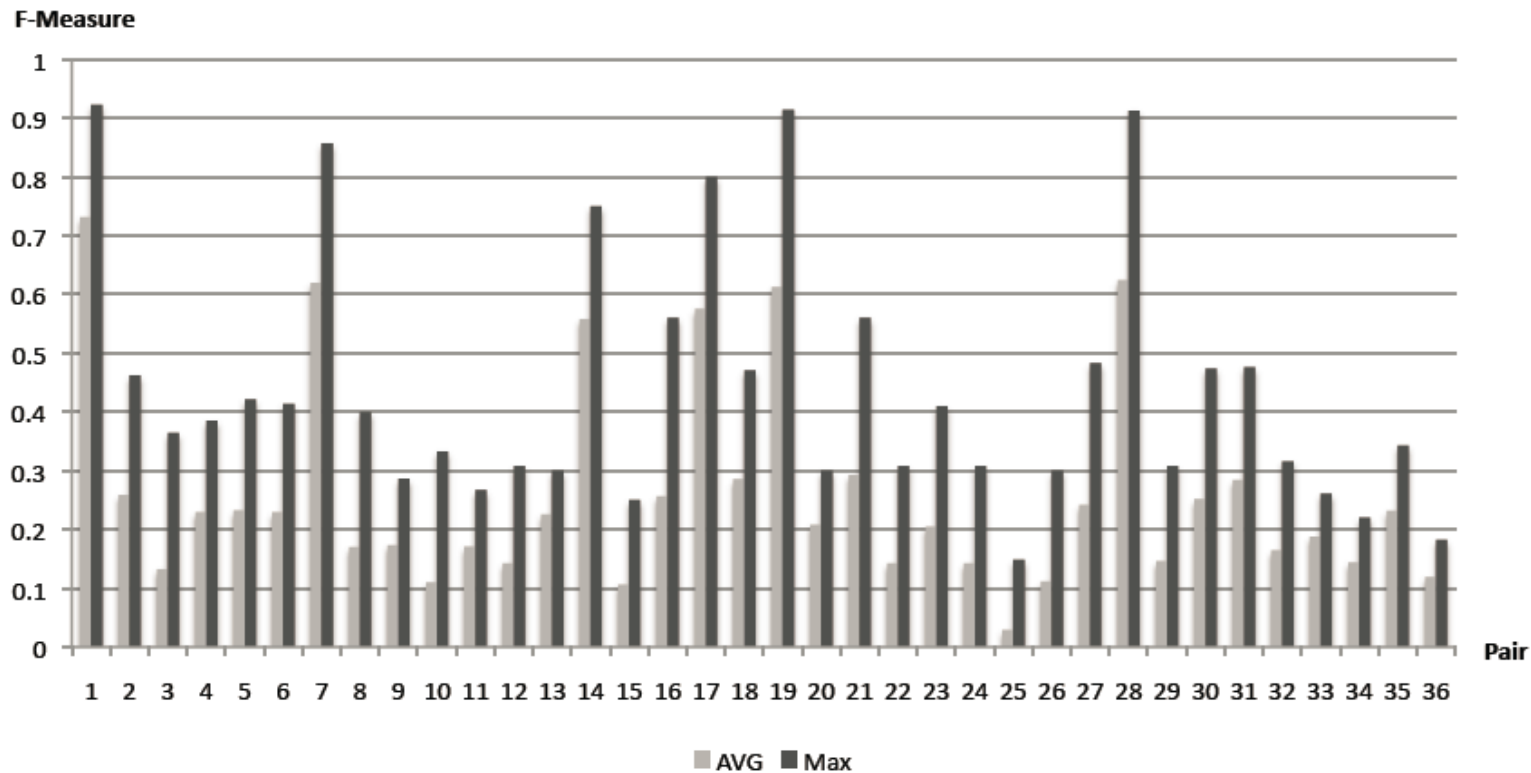
Some Observations

No.	Approach	Precision		Recall		F-Measure	
		AVG	STD	AVG	STD	AVG	STD
1	Triple-S	0.19	0.21	0.25	0.33	0.22	0.23
2	BP Graph Matching	0.55	0.48	0.19	0.28	0.28	0.30
3	RefMod-Mine/NSCM	0.68	0.19	0.33	0.22	0.45	0.18
4	RefMod-Mine/ESGM	0.25	0.28	0.18	0.26	0.21	0.23
5	Bag-of-Words Similarity	0.29	0.35	0.22	0.30	0.25	0.31
6	PMLM	0.19	0.09	0.60	0.20	0.29	0.12
7	ICoP	0.42	0.27	0.28	0.23	0.33	0.24

“Given these average values, should we give up on process matching?”



The Prediction Problem

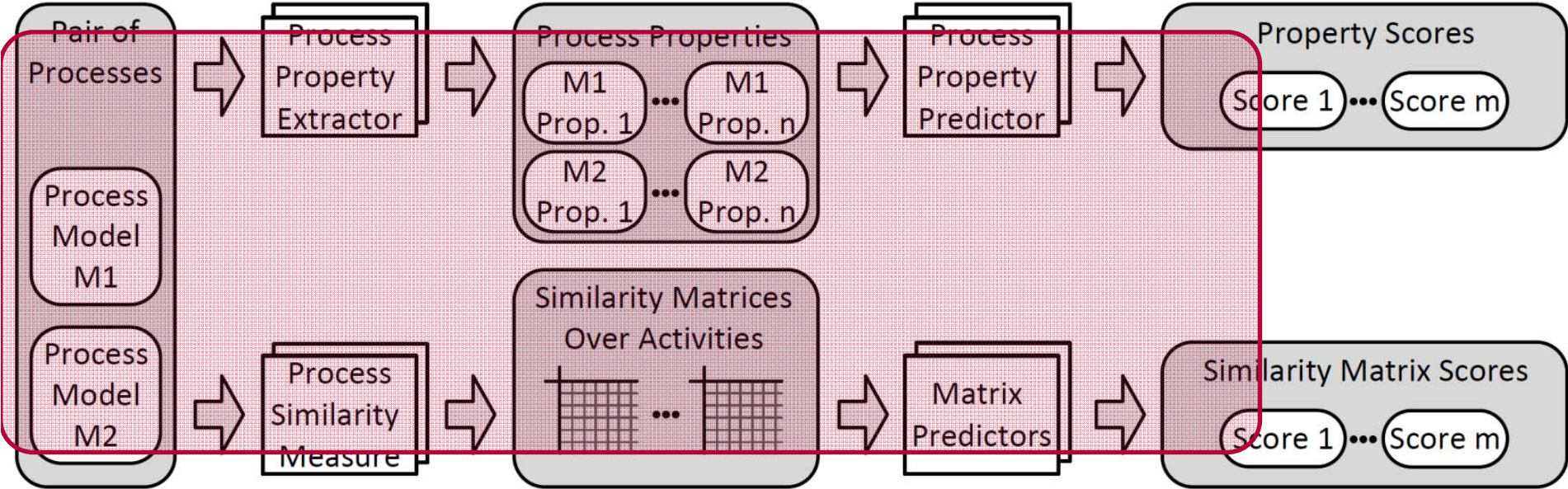


How to distinguish matching tasks for which matchers yield high quality matches from those for which results are poor?

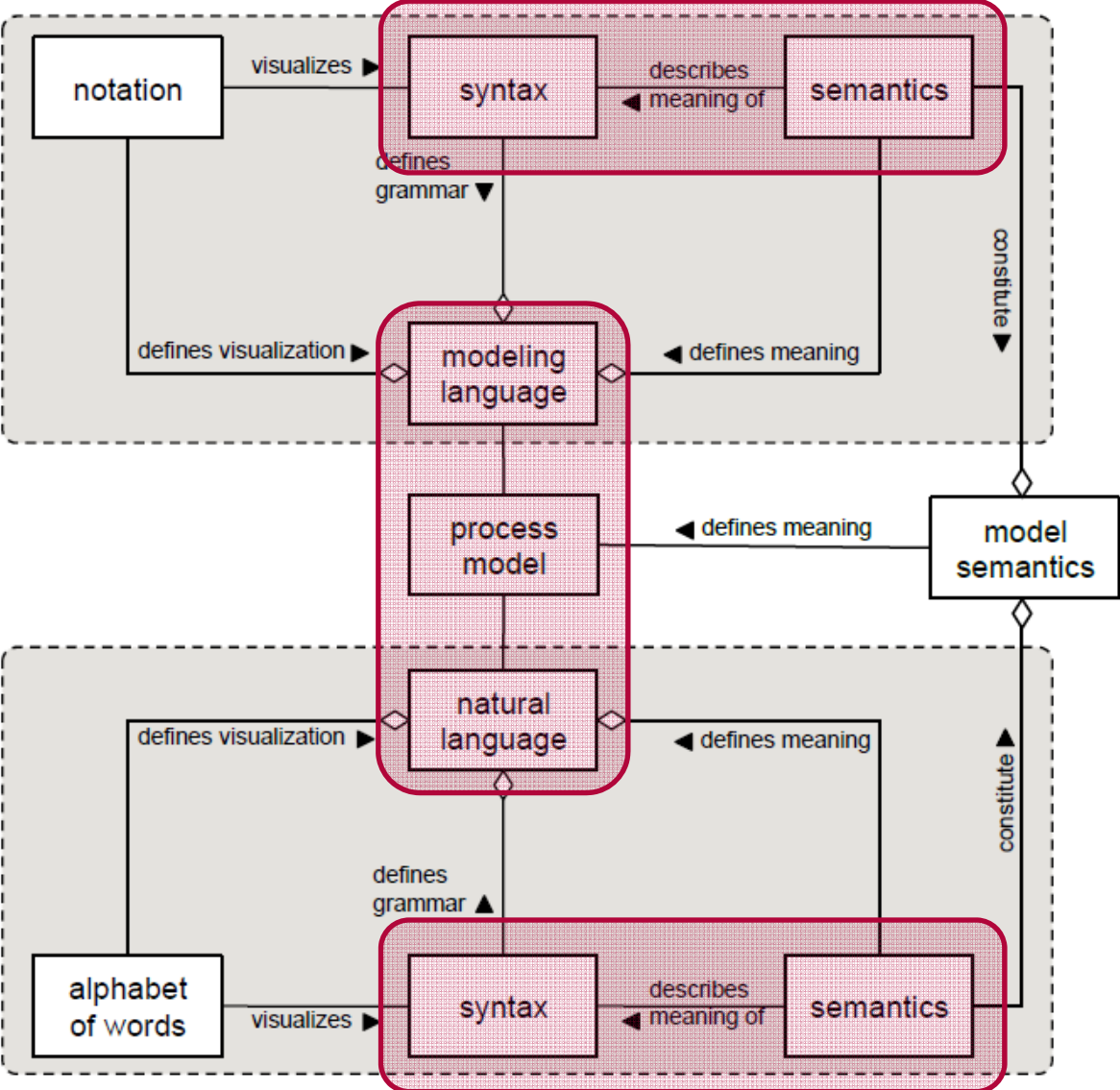
Agenda

- 1) Background
- 2) Process Model Match Prediction
 - Architecture
 - Property-based prediction
 - Similarity-based prediction
 - Prediction models
- 3) Experimental Results
- 4) Take Away

Prediction Architecture



Basis for Matching



Property Predictors

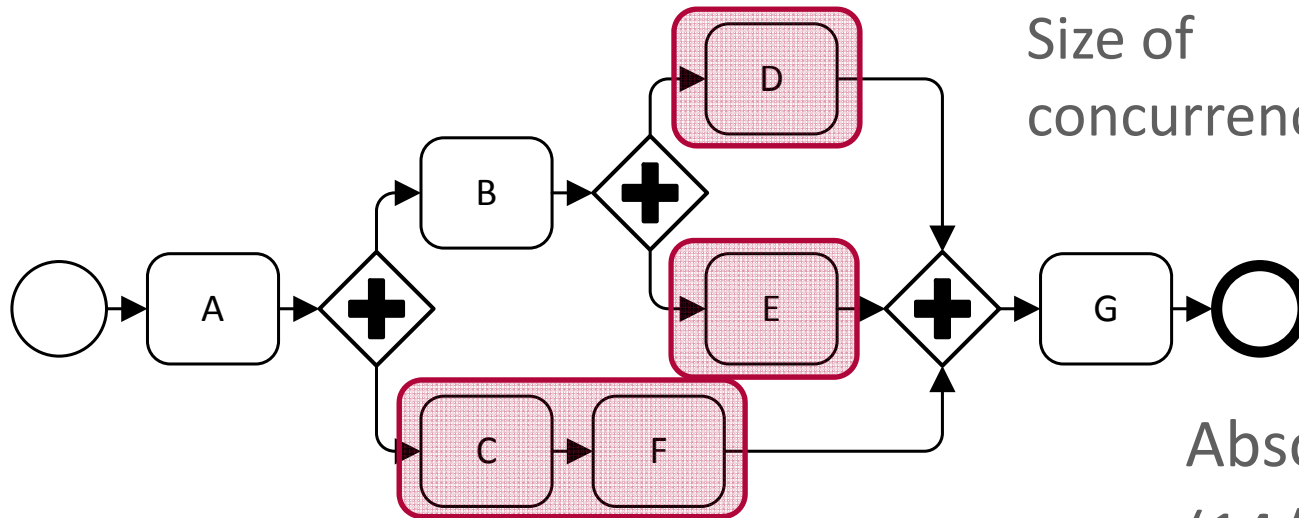
Characteristics of process models hint at challenges for a matcher

- Absolute values of models
- Relative comparison of models

Examples:

- Number of labels with action in a lexical database
- Number of nodes in cycle
- Size of the concurrency relation

Property Predictor cont.



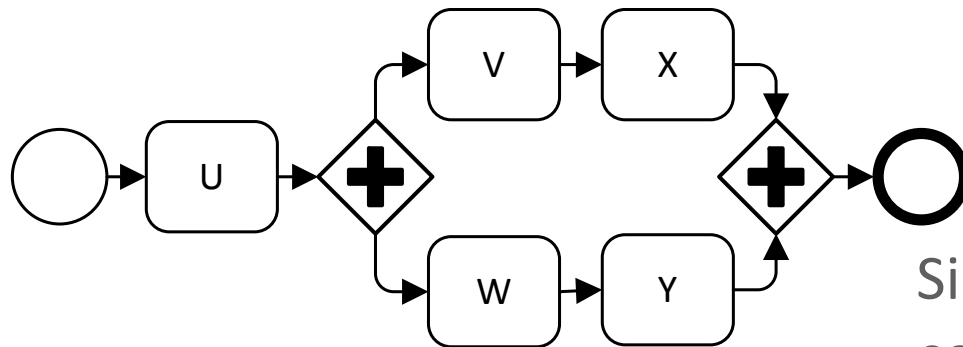
Size of
concurrency relation: 14

Absolute:

$$(14/49 + 8/25)/2 \approx 0.30$$

Relative:

$$1 - (14-8)/49 \approx 0.88$$



Size of
concurrency relation: 8

Similarity Measures

Similarity values obtained for pairs of activities
hint at challenges for a matcher

- Obtain a similarity matrix
- Apply predictors over the matrix

Example measures:

- Optimal string edit distance over activity labels
- Distance of virtual documents in a vector space
- Lin distance between activity labels

Similarity Matrix Predictors



Dominants predictor

- Ratio of dominant values in the matrix

Binary matrix predictor

- Distance to closest binary matrix in vector space

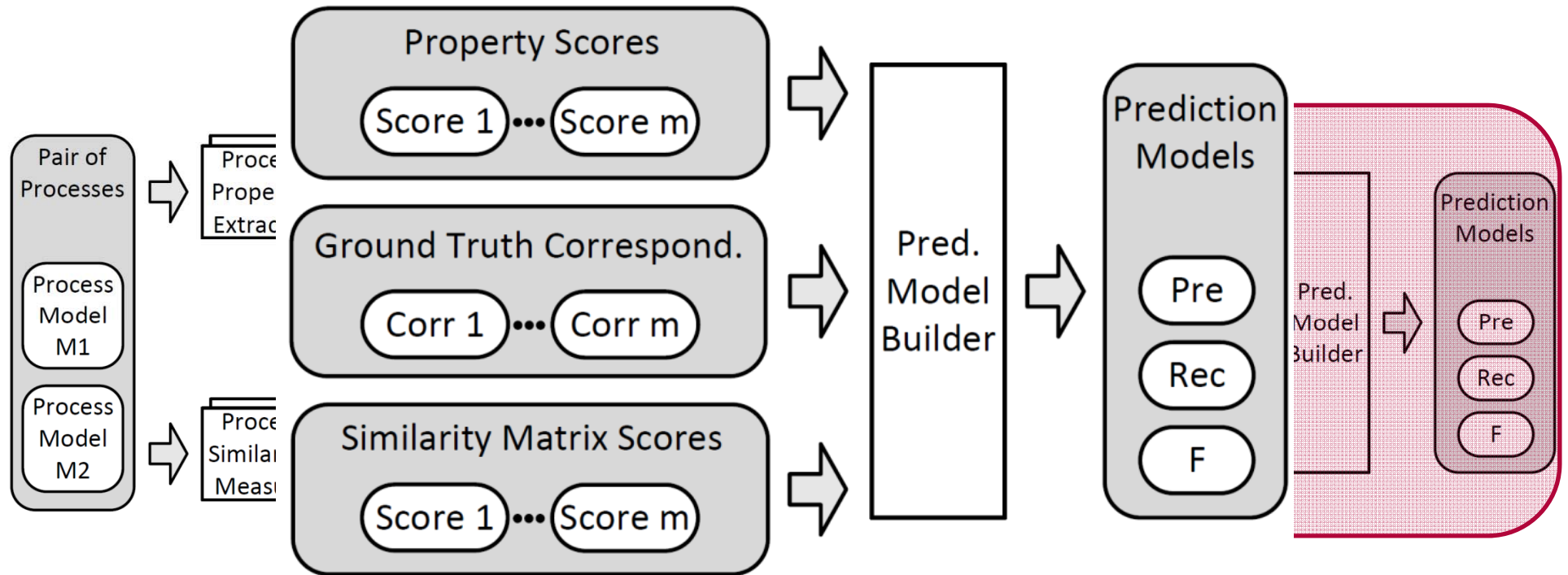
	Receive Loan Request in Call Center	Record Request Details	Get Stored Loan Request	Conduct Liability Check
Get Loan Application	0.67	0.34	0.91	0.11
Register Application	0.21	0.87	0.29	0.09
Check Credit History	0.18	0.54	0.05	0.48

1	0	1	0
0	1	0	0
0	1	0	0

⋮

1	0	1	0
0	1	0	0
0	0	0	1

Building a Prediction Model



Building a prediction model

- Select quality criterion
- Conduct step-wise regression over the prediction scores

Agenda

- 1) Background
- 2) Process Model Match Prediction
 - Architecture
 - Property-based prediction
 - Similarity-based prediction
 - Prediction models
- 3) Experimental Results
- 4) Take Away

Three model collections

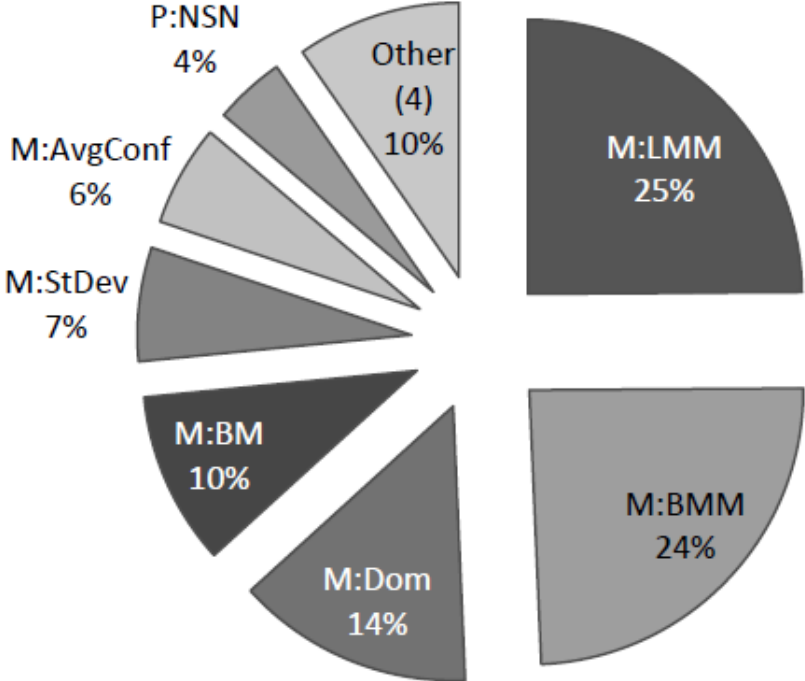
- University admission processes, 36 pairs (English)
- Birth registration processes, 36 pairs (English)
- Municipality processes, 17 pairs (Dutch)

Large set of predictors

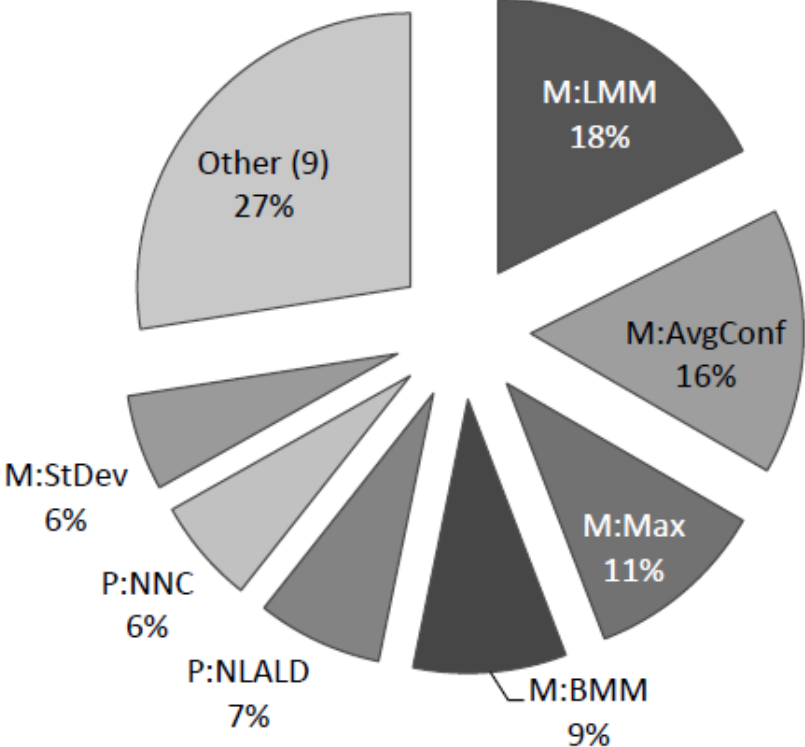
- 30 property predictors
- 10 similarity measures with six matrix predictors

Prediction Models

Precision



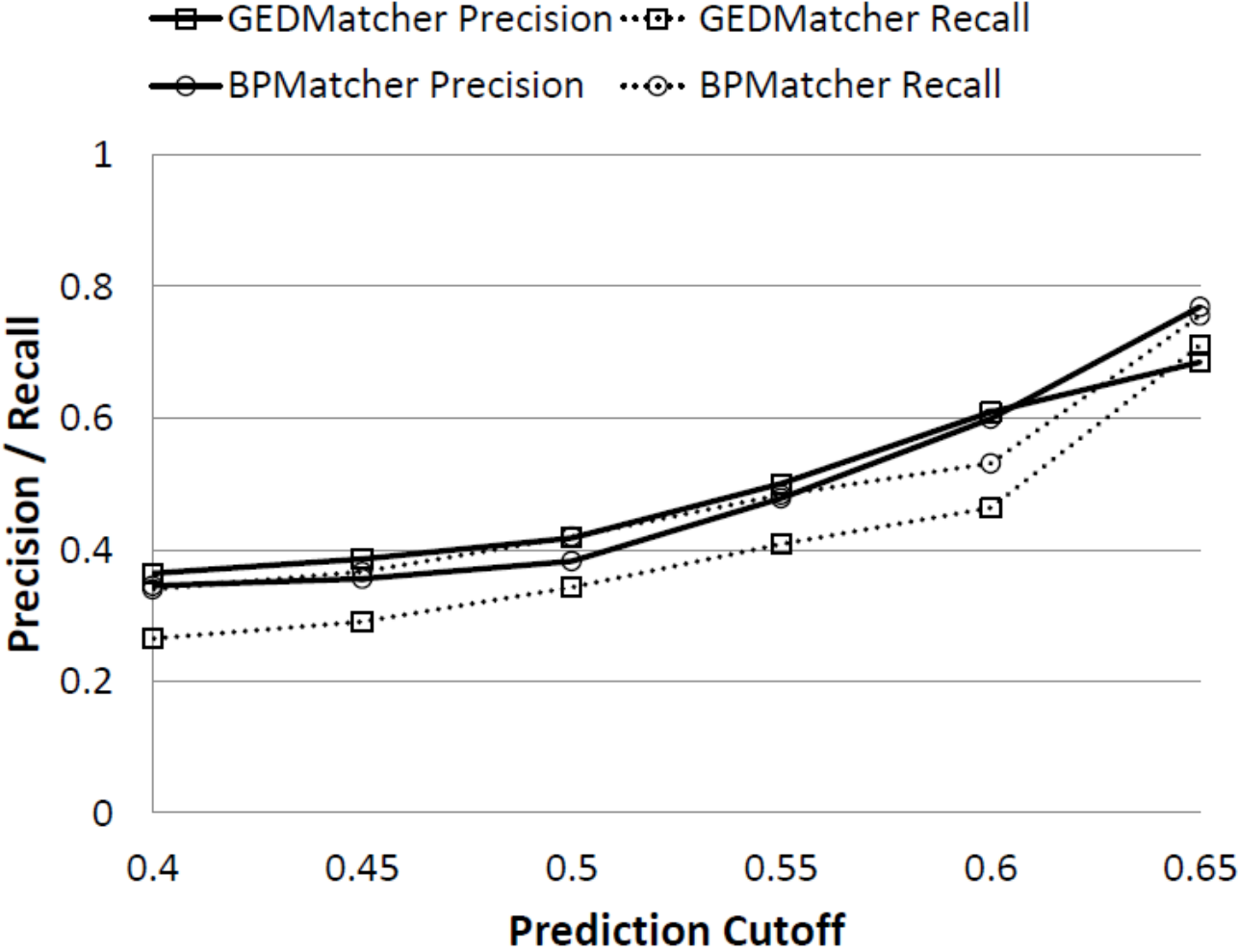
Recall



Prediction Quality

Quality Measure	Prediction Model	# Predictors	Adj. R^2	Std. Error of Estimate
NB-Precision	Property only	8	.113	.155
	Matrix only	6	.527	.113
	Property & Matrix	11	.563	.109
NB-Recall	Property only	8	.157	.379
	Matrix only	6	.807	.181
	Property & Matrix	16	.865	.151
NB-F-Score	Property only	6	.114	.221
	Matrix only	5	.418	.179
	Property & Matrix	20	.478	.170

Using Prediction



Take Away

Prediction of the match quality

- Distinguish tasks for which matchers yield high quality matches from those with poor results
- Considering process model properties and elementary similarities
- Yield high prediction quality
- Available in the Ontobuilder Research Environment

Future work: predictions for improving the actual matching

