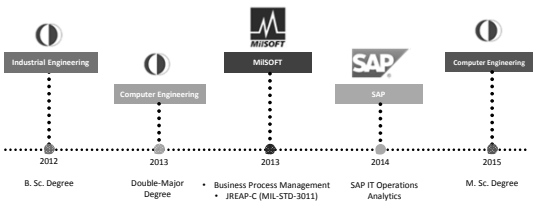
Recommendation Generation for Performance Improvement by using Cross-Organizational Process Mining

Onur Yılmaz
Supervisor: Assoc. Prof. Pınar Karagöz

September 1, 2015

Onur Yılmaz



Agenda

1. Introduction
2. Related Work
3. Background
4. Methodology
5. Results & Discussions
6. Conclusion & Future Work
7. Demonstration

Introduction
Process Mining

Relatively young and developing research area and main idea is to

- discover,
- monitor and;
- improve processes by extracting information from event logs. [1]

Events recorded and available

Competitive business life

Introduction
Cross- Organizational Process Mining

Cloud computing and shared infrastructures

Event logs of multiple organizations [2]

Analyze the big picture

Work together to execute the same process

Learn from each other

Execute the same task on shared infrastructure [3]

Introduction
Focus of this Study

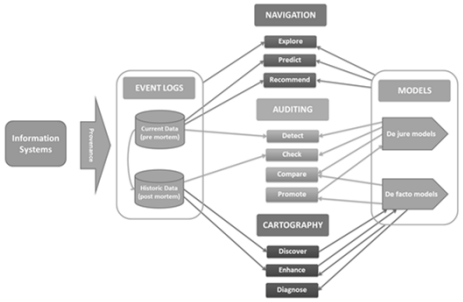
- A hybrid approach
  - Using different process mining subfields to create a new point of view
- Cross-organizational process mining
  - Processes are executed on several organizations,
  - Unsupervised learning using performances of organizations

Agenda

1. Introduction
2. Related Work
3. Background
4. Methodology
5. Results & Discussions
6. Conclusion & Future Work
7. Demonstration

Related Work

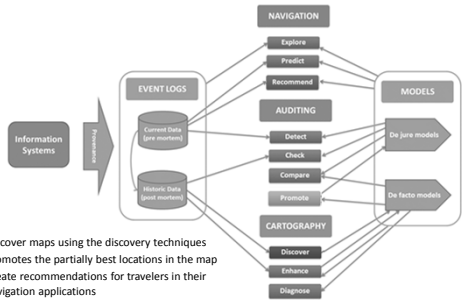
State of the Art in Process Mining



[4]

Related Work

State of the Art in Process Mining



- Discover maps using the discovery techniques
- Promotes the partially best locations in the map
- Create recommendations for travelers in their navigation applications

Related Work
Contributions of This Study

- Cross-organizational process mining approach for process performance improvement
- Generic, noise-capable process mining method for mining process models of different organizations
- Clustering of organizations based on their performance indicators
  - Unlike the clustering methods based on process structures in the literature [9]

Related Work
Contributions of This Study

- Mismatch analysis for spotting differences between processes of organizations
  - Formulation and implementation of patterns and analyzers
- Recommendation generation to show how organizations can learn from other organizations which perform better
- Open-source, extensible and configurable set of plugins in ProM framework

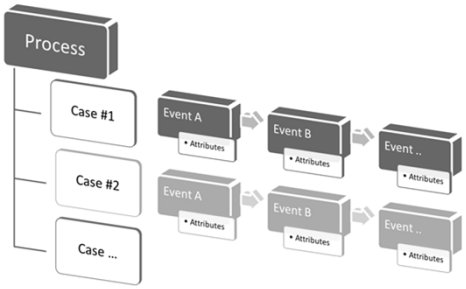
Agenda

Background
Event Log

- Outputs of the software systems like Enterprise Resource Planning (ERP) or Business Process Management (BPM)

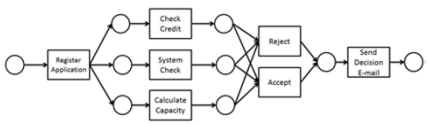
Event Log				
		Attributes		
	Event	Date	Time	Transition
Case #1	Register Application	16.04.2013	14:37:27	Complete
	Check Credit	16.04.2013	14:41:19	Complete
	Check System	16.04.2013	14:47:35	Complete
	Calculate Capacity	16.04.2013	14:50:21	Complete
	Accept	16.04.2013	14:53:22	Complete
	Send decision e-mail	16.04.2013	14:55:11	Complete
Case #2	Register Application	16.04.2013	16:28:19	Complete
	Check Credit	16.04.2013	16:36:22	Complete

Background
Event Log



Background
Process Modeling

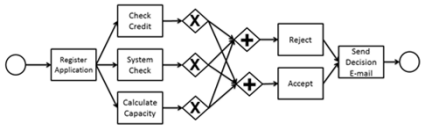
- Workflow Net
  - Petri net with a start node, end node and connectedness



  - Mathematical background

Background
Process Modeling

- Business Process Modelling Notation (BPMN)
  - Standardized and easy to understand by stakeholders



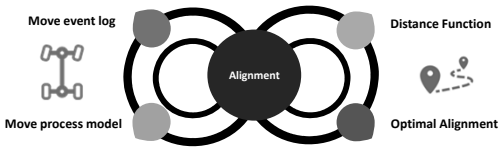
  - Business oriented

Background
Process Discovery

- One of the most challenging tasks is to construct a process model based on the behavior in the event logs
- Inductive Miner Infrequent (*IMi*) is used since it is simple, highly applicable and configurable to handle noise [20]
  - Block-structured Workflow Nets
  - Rediscoverability

Background
Process Performance Analysis

- Discover relationships between event logs and process models for conformance and performance analysis [21]



Background
Mismatch Patterns in Process Models

- Patterns for frequent mismatches between the similar process models by Dijkman [14]

Authorization

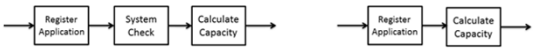
Activity
- Skipped Activity
- Refined Activity

Control Flow
- Activities at Different Moments in Processes
- Different Conditions for Occurrence
- Different Dependencies
- Additional Dependencies

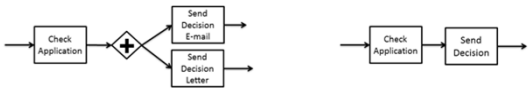
Background
Mismatch Patterns in Process Models

Activity Mismatch Patterns [14]

- Skipped Activity



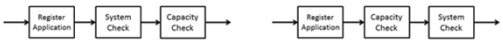
- Refined Activity



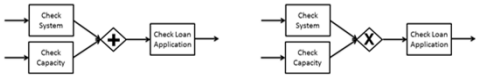
Background
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns [14]

- Activities at Different Moments in Processes



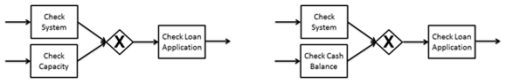
- Different Conditions for Occurrence



Background
Mismatch Patterns in Process Models

Control Flow Mismatch Patterns [14]

- Different Dependencies

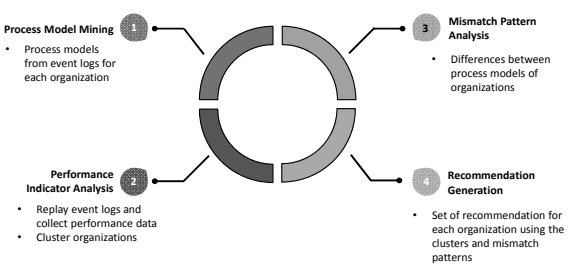


- Additional Dependencies

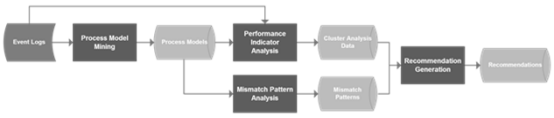
Agenda

1. Introduction
2. Related Work
3. Background
4. Methodology
5. Results & Discussions
6. Conclusion & Future Work
7. Demonstration

Methodology
Approach Overview

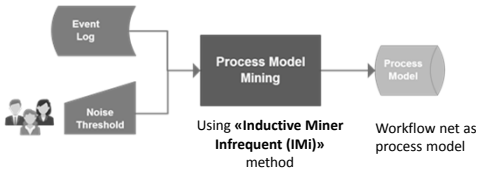


Methodology
Approach Overview



Methodology
Process Model Mining

Applied for each organization:

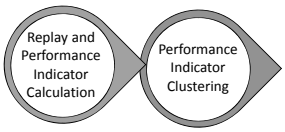


Using **«Inductive Miner Infrequent (IMi)»** method

Workflow net as process model

Methodology
Performance Indicator Analysis

- Two steps:

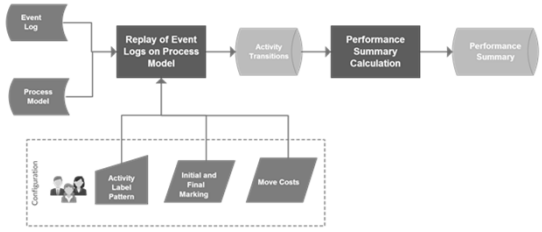


- Performance Indicators:
  - Average Time Between Activities ⇩
  - Standard Deviation of Time Between Activities ⇩

Methodology

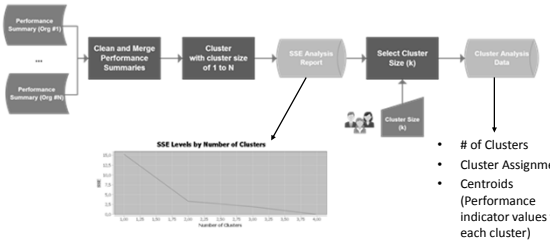
Performance Indicator Analysis - Replay and Performance Indicator Calculation

Applied for each organization:



Methodology

Performance Indicator Analysis – Performance Indicator Clustering

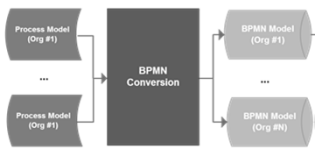


- # of Clusters
- Cluster Assignment
- Centroids (Performance indicator values for each cluster)

Methodology

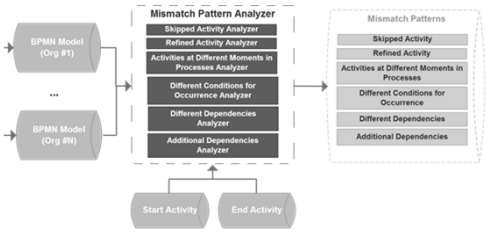
Mismatch Pattern Analysis

- Spot the differences between process models of different organizations as mismatch patterns
- BPMN used since notation is more appropriate to formulate mismatch patterns



Methodology
Mismatch Pattern Analysis

- Mismatch patterns and analyzers are developed

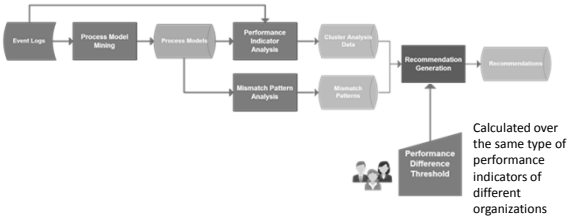


Methodology
Recommendation Generation

- Providing a set of mismatch patterns for each organization to enhance their processes.
  - Mismatch patterns between organizations, which are **performing better** in terms of their performance indicator values.

$$\text{Recommendation} = (\text{Organization}, A_{start}, A_{end}, \text{Mismatch Patterns})$$

Methodology
Recommendation Generation



Calculated over the same type of performance indicators of different organizations

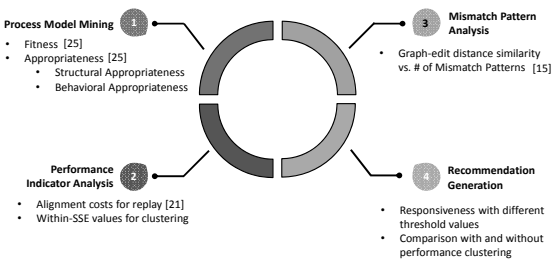
Methodology
Recommendation Generation

```
RecommendationGeneration
input: O organization, C Cluster Analysis Data, P performance
difference threshold
output: Recommendations a set of recommendations

Recommendations ← {}
i ← Cluster of organization O
for each centroid for cluster i
  Get other cluster j with the centroids of A_start and A_end; and
  value difference larger than P
  for each organization O' in the cluster j
    MismatchPatterns ←  Mismatch Pattern Analysis(O, O',A_start, A_end)
    Recommendations ←  Recommendation(O, A_start,A_end, MismatchPatterns)
```


Agenda

Results & Discussions
Evaluation Metrics



Results & Discussions
Dataset Selection

- **Loan Application Process** [26]
  - Synthetically generated
  - 4 variants of a simple loan application in a financial institute
- **Environmental Permit Application Process** [27]
  - Real-life event log from "Configurable Services for Local Governments (CoSeLoG)" project [13]
  - «Environmental Permit Application Process» of 5 municipalities in Netherlands

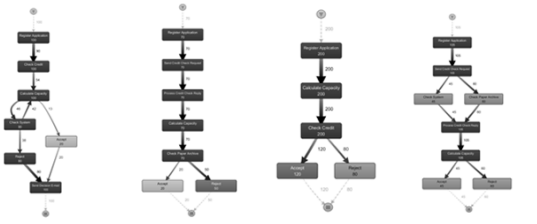
Results & Discussions
Loan Application Process

	Cases	Events	Percentage
Variant #1	100	590	24 %
Variant #2	70	420	17 %
Variant #3	200	800	33 %
Variant #4	105	630	26 %

Cases 475 Events 2440

- These variants are used as organizational logs

Results & Discussions
Loan Application Process – Process Model Mining

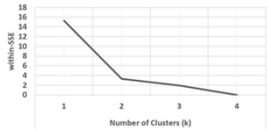


Variant #1	Variant #2	Variant #3	Variant #4
100 % (Fitness)	100 % (Fitness)	100 % (Fitness)	100 % (Fitness)
84.2 % (Avg. App.)	100 % (Avg. App.)	100 % (Avg. App.)	99.1 % (Avg. App.)

Results & Discussions

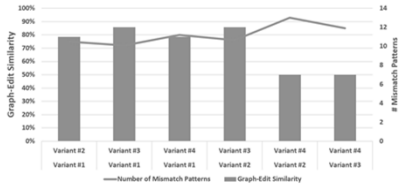
Loan Application Process – Performance Indicator Analysis

- **Clustering:**



Results & Discussions

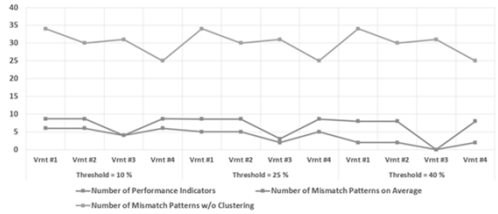
Loan Application Process – Mismatch Pattern Analysis



Correlation between graph-edit similarity and number of mismatch patterns

Results & Discussions

Loan Application Process – Recommendation Generation



- Responsiveness and degree of helping the user to focus on the performance improvement

Results & Discussions

Environmental Permit Application Process

	Cases	Events	Percentage
Municipality #1	54	131	6.1 %
Municipality #2	302	586	27.3 %
Municipality #3	37	73	3.4 %
Municipality #4	340	507	23.7 %
Municipality #5	481	845	39.4 %

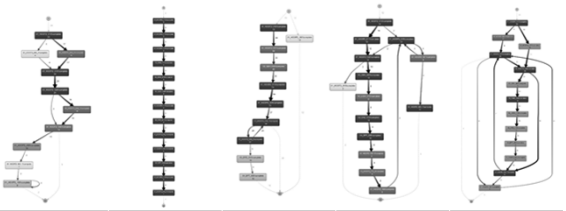
Cases 1214 Events 2142

- Preprocessing is undertaken on the raw dataset [28]
- These municipalities are used as organizational logs

Results & Discussions

Environmental Permit Application Process – Process Model Mining

- 10 to 20 times simplified process models

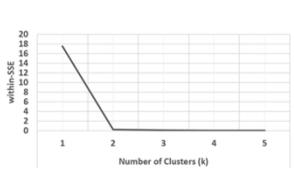


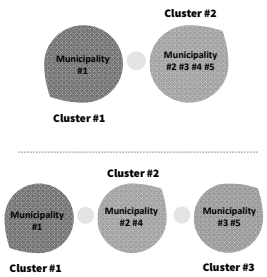
Municipality #1	Municipality #2	Municipality #3	Municipality #4	Municipality #5
86 % (Fitness)	100 % (Fitness)	92.3 % (Fitness)	96.8 % (Fitness)	94.5 % (Fitness)
76 % (Avg. App.)	100 % (Avg. App.)	69.1 % (Avg. App.)	64.9 % (Avg. App.)	49.3 % (Avg. App.)

Results & Discussions

Environmental Permit Application Process – Performance Indicator Analysis

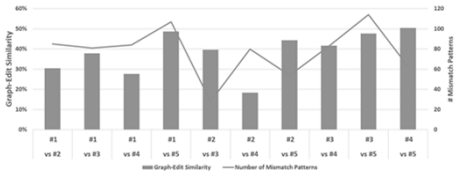
- **Clustering:**





Results & Discussions

Environmental Permit Application Process – Mismatch Pattern Analysis

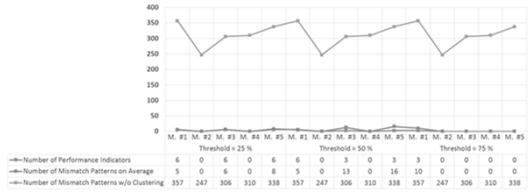


- Correlation between graph-edit similarity and number of mismatch patterns except Municipality #4 and #5

Results & Discussions

Environmental Permit Application Process – Recommendation Generation

(with 3 clusters)



	M. #1	M. #2	M. #3	M. #4	M. #5	M. #1	M. #2	M. #3	M. #4	M. #5	M. #1	M. #2	M. #3	M. #4	M. #5
Number of Performance Indicators	6	0	6	0	6	6	0	3	0	3	3	0	0	0	0
Number of Mismatch Patterns on Average	5	0	6	0	8	5	0	13	0	16	10	0	0	0	0
Number of Mismatch Patterns w/o Clustering	357	247	306	310	338	357	247	306	310	338	357	247	306	310	338

- Learning opportunities increases as number of cluster increases
- Number of mismatch patterns to check significantly decreases with performance clustering

Results & Discussions

Discussions

- Mismatch analysis stage:
  - Differences in accordance with similarity metrics
  - Information value of mismatch patterns are not equal
- Recommendation generation stage:
  - Performance clustering helps to focus on differences
    - 3 times less in Loan Application Process Dataset
    - 100 times less in Environmental Permit Application Process Dataset

Results & Discussions
Discussions

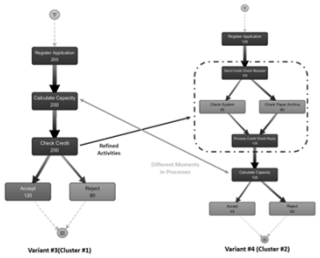
- Business value of generated recommendations:
  - Results may be
    - important or
    - infeasible and irrelevant for business environment
  - Some insights about results can be provided but business environment and case related assessment is important.

Results & Discussions
Discussions

In Loan Application Process, Variant #4 performs better
- 27 % on average time and
- 12 % on standard deviation time

between activities

Calculate Capacity → Accept



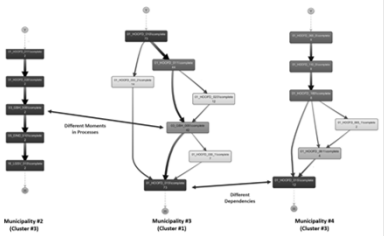
Process models between activities «Calculate Capacity» and «Accept»

Results & Discussions
Discussions

In Environmental Permit Application Process, Municipality #3 performs worse
- 40 % on average time and
- 53 % on standard deviation time

between activities

01_HOOFD_010 → 01_HOOFD_015



Simplified process models between activities «01_HOOFD_010» and «01_HOOFD_015»

Agenda

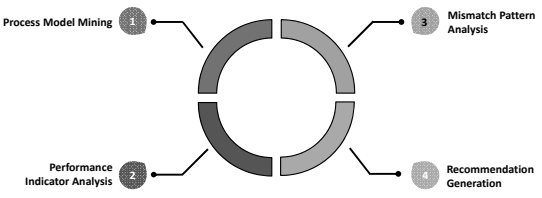
1. Introduction
2. Related Work
3. Background
4. Methodology
5. Results & Discussions
6. Conclusion & Future Work
7. Demonstration

Conclusion

- Cross-organizational process mining is applied
  - Unsupervised learning with predictor variables as performances of organizations
  - In an environment where processes are executed on several organizations
- Results show that it is possible to use cross-organizational process mining and mismatch patterns for performance improvement recommendations

Conclusion

- A four-stage solution is presented and their performances are explained



Conclusion
Future Work

- Process mining stage:
  - Different techniques can be used to mine complex process models
- Performance indicator stage:
  - New indicators based on business environment and needs

Conclusion
Future Work

- Mismatch pattern analysis:
  - New patterns can be introduced
- Recommendation generation:
  - Domain or BPM expertise to assess the quality of recommendations

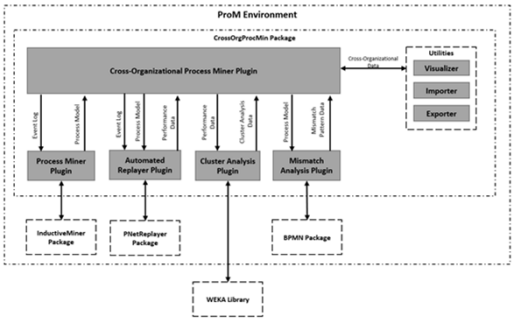
Agenda

ProM Implementation

- ProM, extensible framework that supports a wide variety of process mining techniques in form of plugins [24]
- Widely accepted in industry and academia with an active community
- Developed set of plugins are packaged with the name of **«CrossOrgProcMin»** and published on Github.

onuryilmaz/cross-orc-proc-min

ProM Implementation



Demonstration

ProM process mining workbench

YouTube

Also available on YouTube
http://youtu.be/T92UrRfl3r0

References

[1] W. M. P. van der Aalst, A. Adriansyah, A.K.A. de Medeiros, and et al. Process mining manifesto. In Business process management workshops, pages 169–194. Springer, 2012.
[2] J. C. Buijs, B. F. van Dongen, and W. M. P. van der Aalst. Towards crossorganizational process mining in collections of process models and their executions. In Business Process Management Workshops, pages 2–13. Springer, 2012.
[3] W. M. P. van der Aalst. Intra-and inter-organizational process mining: Discovering processes within and between organizations. In The Practice of Enterprise Modeling, pages 1–11. Springer, 2011.
[4] W. M. P. van der Aalst. Process mining: discovery, conformance and enhancement of business processes. Springer Science & Business Media, 2011.
[5] A. K. A. de Medeiros, B. F. van Dongen,W. M. P. van der Aalst, and A. J. M. M. Weijters. Process mining: Extending the -algorithm to mine short loops. 2004.
[6] W. M. P. van der Aalst, T.Weijters, and L. Maruster. Workflow mining: Discovering process models from event logs. Knowledge and Data Engineering, IEEE Transactions on, 16(9):1128–1142, 2004.
[7] J. Herbst. Dealing with concurrency in workflow induction. In European Concurrent Engineering Conference. SCS Europe. Citeseer, 2000.

References

[8] J. Herbst and D. Karagiannis. Integrating machine learning and workflow management to support acquisition and adaptation of workflow models. In Database and Expert Systems Applications, 1998. Proceedings. Ninth International Workshop on, pages 745–752. IEEE, 1998.
[9] G. Greco, A. Guzzo, and L. Pontieri. Mining hierarchies of models: From abstract views to concrete specifications. In Business Process Management, pages 32–47. Springer, 2005.
[10] W. M. P. van der Aalst, A.K.A. de Medeiros, and A. J. M. M. Weijters. Genetic process mining. In Applications and theory of Petri nets 2005, pages 48–69. Springer, 2005.
[11] E. Esgin, P. Senkul, and C. Cimenbicer. A hybrid approach for process mining: using from-to chart arranged by genetic algorithms. In Hybrid Artificial Intelligence Systems, pages 178–186. Springer, 2010.
[12] E. Esgin and R. Senkul. A hybrid approach to process mining: Finding immediate successors of a process by using from-to chart. In Machine Learning and Applications, 2009. ICMLA'09. International Conference on, pages 664–668. IEEE, 2009.
[13] W. M. P. van der Aalst. Business process configuration in the cloud: how to support and analyze multi-tenant processes? In Web Services (ECOWS), 2011 Ninth IEEE European Conference on, pages 3–10. IEEE, 2011.

References

[14] R. Dijkman. Mismatch Patterns in Similar Business Processes. Beta, Research School for Operations Management and Logistics, 2007.
[15] R. Dijkman, M. Dumas, B. van Dongen, R. Käärik, and J. Mendling. Similarity of business process models: Metrics and evaluation. Information Systems, 36(2):498–516, 2011.
[16] J. C. Buijs and H. A. Reijers. Comparing business process variants using models and event logs. In Enterprise, Business-Process and Information Systems Modeling, pages 154–168. Springer, 2014.
[17] E. Esgin and P. Karagoz. Sequence alignment adaptation for process diagnostics and delta analysis. In Hybrid Artificial Intelligent Systems, pages 191–201. Springer, 2013.
[18] E. Esgin and P. Senkul. Delta analysis: a hybrid quantitative approach for measuring discrepancies between business process models. In Hybrid Artificial Intelligent Systems, pages 296–304. Springer, 2011.
[19] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst. Discovering blockstructured process models from event logs-a constructive approach. In Application and Theory of Petri Nets and Concurrency, pages 311–329. Springer, 2013.
[20] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst. Discovering block-structured process models from event logs containing infrequent behaviour. In Business Process Management Workshops, pages 66–78. Springer, 2014.

References

[21] W. M. P. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying history on process models for conformance checking and performance analysis. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2(2):182–192, 2012.
[22] D. Arthur and S. Vassilvitskii. k-means++: the advantages of carefull seeding. In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms, pages 1027–1035, 2007.
[23] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten. The weka data mining software: An update. SIGKDD Explor. Newsl., 11(1):10–18, November 2009.
[24] H. M. W. Verbeek, J. C Buijs, B. F. van Dongen, and W. M. P. van der Aalst. Prom 6: The process mining toolkit. Proc. of BPM Demonstration Track, 615:34–39, 2010.
[25] A. Rozinat and W. M. P. van der Aalst. Conformance checking of processes based on monitoring real behavior. Information Systems, 33(1):64–95, 2008.
[26] J. C. Buijs. Loan application example, 2013.
[27] J.C. Buijs. Environmental permit application process ('wabo'), coselog project, 2014.
[28] J. C. Buijs. Flexible Evolutionary Algorithms for Mining Structured Process Models. PhD thesis, PhD thesis. Eindhoven, The Netherlands: Technische Universiteit Eindhoven, 2014 (cit. on p. 179), 2014.

Questions & Comments

Thank you for your attention!

Recommendation Generation for Performance Improvement by using Cross-Organizational Process Mining

Onur Yılmaz
Supervisor: Assoc. Prof. Pınar Karagöz

September 1, 2015